TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Huy Tran

Environment-sensing Mechanisms of Gene Expression and their Effects on the Dynamics of Genetic Circuits across Cell Generations

Julkaisu 1367 • Publication 1367

Tampere 2016

Tampereen teknillinen yliopisto. Julkaisu 1367
Tampere University of Technology. Publication 1367

Huy Tran

Environment-sensing Mechanisms of Gene Expression and their Effects on the Dynamics of Genetic Circuits across Cell Generations

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Sähkötalo Building, Auditorium SA203, at Tampere University of Technology, on the 25th of January 2016, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2016

Supervisor:	Assoc. Prof. Andre S. Ribeiro Tampere University of Technology, Finland
Pre-examiners:	Assoc. Prof. Ertuğul Özbudak Albert Einstein College of Medicine, U.S.A.
	Dr. Georg Fritz LOEWE Center for Synthetic Microbiology Germany
Opponent:	Prof. Kim Sneppen University of Copenhagen, Denmark

ISBN 978-952-15-3678-6 (printed)
ISBN 978-952-15-3686-1 (PDF)
ISSN 1459-2045

Abstract

In genetic circuits, the constituent genes do not interact only between themselves, they are also affected by regulatory molecules of the host cells that support the circuits' operation and by the environmental conditions. These factors, along with the intrinsic noise in gene expression, affect the functioning of the circuits. As such, to understand the structure of natural circuits and to engineer functional synthetic circuits, one needs to characterize thoroughly how external factors and perturbations from the environment may affect their behavior.

This thesis focused on two cellular mechanisms through which the dynamics of gene expression becomes environment dependent: the intake of gene expression regulatory molecules from the media and the σ factor competition. The first mechanism determines the dynamics by which inducer molecules in the media enter the cell cytoplasm and trigger or repress the expression of the target gene. The second mechanism allows cells to change its gene expression profile to adapt to specific stress conditions.

Following the characterization of the effects of these mechanisms on the expression dynamics of individual genes from live, single cell measurements, we then performed *in silico* assessments on how these effects at the single gene level propagate to the circuit level. Here, the dynamics of genetic circuits was observed in both non-dividing and dividing cell populations, where errors in the partitioning of molecules in cell division occur and introduce significant variance between sister cells.

From these studies, with the knowledge on the factors of the host cells and their environment sensing mechanisms, more predictive models of the circuits' dynamics are expected to emerge. The models would further help in identifying what circuit composition, properties of the host strains and environmental conditions are needed for the circuits to exhibit the desired behavior.

Preface

The studies in this thesis are conducted in the Laboratory of Biosystem Dynamics (LBD), department of Signal Processing, faculty of Computing and Electrical Engineering, Tampere University of Technology (TUT), under the supervision of Associate Professor Andre S. Ribeiro.

I would like to express my deepest gratitude toward my supervisor, Andre S. Ribeiro, without whom none of these studies would have been viable and I would still be swimming in my random thoughts, with full of complacency and ignorance. You have always been a dedicated teacher and a loyal friend, who did not hesitate to point out my flaws and to share your knowledge and experiences. I could not expect more than that from a supervisor.

I would like to thank the thesis pre-examiners, Dr. Ertuğul Özbudak and Dr. Georg Fritz, for their valuable comments. I do believe that the thesis has been improved tremendously following your suggestions and I find myself still having a lot to learn to become an independent researcher.

I would like to thank Antti Häkkinen for introducing me into the world of modeling and statistics. Your concentration and rigor in science have always been admirable. I am also grateful to all LBD's members, whom I spent most of my time with in the last 4 years. Thanks to you, my experiences in Finland have been very pleasurable and fruitful. I did enjoy the conversations with Jason Lloyd-Price, Abhishekh Gupta, Teppo Annila and Sofia Startceva on modeling and signal processing techniques. Samuel Oliveira shared with me his enthusiasm in the field of synthetic biology. Meenakshisundaram Kandhavelu, Jarno Mäkelä, Vinodh Kandavalli, Jerome Geoffrey, Nádia Gonçalves, Barathi Muthukrishnan and Ramakanth Neeli familiarized me with the laboratory experiments. I also learned a lot about music "gears" through the coffee-time discussions with Eero Lihavainen. The companionship of you all in science, in Hervanta's "cuisine" and in music has been memorable.

I am grateful to the secretaries and the personnel of the Department of Signal Processing for creating such favorable and friendly environment that allows researchers to fully focus on their studies. I had the opportunities to work directly with Vivre Larmila, Noora Rotolla-Pukkila and Ulla Siltaloppi, who are always there to help me navigate

through the problems of bureaucracy. I am also thankful to my study coordinator, Elina Orava, who guided me through the formalities of the doctoral program.

Last, but far from the least, are my deepest appreciations to my family. Thank you, Mom and Dad, for always supporting me and giving me the freedom of the mind and of the heart. Thank you, my sister, for showing me the idea of true happiness and a path toward it. May the smiles always be on your lips wherever you go.

An exit from the highway has just come into sight! Are those years studying in the LBD really over soon? I could not believe it, and certainly Andre would not believe it as our conversations on the future projects never seem to cease. I will now take this exit and soon enter another highway. However, I am sure that its direction will not deviate from the previous inspiring one!

Tampere 5th, January, 2016

Huy Tran

Contents

List of Abbreviations

E. coli	*Escherichia coli*
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
RNAP	RNA polymerase
RNAPσ	RNAP holoenzyme
TSS	transcription start site
IPTG	isopropyl β-D-1-thiogalactopyranoside
aTc	Anhydrotetracycline
cAMP	cyclic adenosine monophosphate
CRP	cAMP receptor protein
GFP	green fluorescent protein
FISH	fluorescence *in situ* hybridization
smFISH	single molecule FISH
CME	Chemical Master Equation
ODE	Ordinary Differential Equations
SSA	Stochastic Simulation Algorithm
DSSA	Delayed Stochastic Simulation Algorithm

List of Figures

List of Publications

This thesis is based on studies I, II, III, IV, and V. The first four have been published and are referred to in the text as "**Publication I**", "**Publication II**", "**Publication III**" and "**Publication IV**" respectively. Study V, currently under review, is referred to as "**Study V**". All publications are reproduced with kind permissions from the publishers.

I. A Häkkinen, **H Tran**, O Yli-Harja, and AS Ribeiro (2013) "Effects of rate-limiting steps in transcription initiation on genetic filter motifs", *PLoS ONE*, 8(8): e70439. doi: 10.1371/journal.pone.0070439.

II. J Lloyd-Price, **H Tran**, AS Ribeiro (2014), "Dynamics of small genetic circuits subject to stochastic partitioning in cell division", *Journal of Theoretical Biology*, 356, pg. 11-19. doi: 10.1016/j.jtbi.2014.04.018.

III. **H Tran** and AS Ribeiro (2013), "Effects of inducer intake kinetics on the dynamics of gene expression", *Advances in Artificial Life, ECAL 2013: Proceedings of the 12th European Conference on the Synthesis and Simulation of Living Systems*, pg. 1174-1181. Taormina – Sicily, Italy, September 2013. doi: 10.7551/978-0-262-31719-2-ch181.

IV. **H Tran**, SMD Oliveira, NM Goncalves, AS Ribeiro (2015), "Kinetics of the cellular intake of a gene expression inducer at high concentrations", *Molecular BioSystems*, 11, pg. 2579. doi: 10.1039/C5MB00244C.

V. VK Kandavalli[+], **H Tran**[+], JG Chandraseelan, AS Ribeiro (2015), "Degree of regulation by σ factor competition in *Escherichia coli* depends on the kinetics of transcription initiation", Submitted. ([+]equal contributions)

In **Publication I**, we studied the effects of the multi-step nature of transcription initiation on the dynamics of small genetic networks. In particular, we explored how the number and durations of the rate-limiting steps in transcription initiation affect the response of genetic filter motifs. In this work, the author designed and tested, along with Antti Häkkinen, the topology of filter motifs, based on the behavior of their logical analogues. The author also participated in the composition of the manuscript. This publication is also included in the dissertation of Antti Häkkinen (2016).

In **Publication II**, we studied the effects of partitioning errors in cell division on the dynamics of two small genetic circuits, namely the Toggle Switch and the Repressilator. The author's contribution to this work is as follows: First, the author derived stochastic models of partitioning schemes to, by tuning the models' parameters, generate partitioning errors of different degrees. Next, the author investigated how the degrees of partitioning errors affect the stability of the Toggle Switch in randomly chosen cell lineages and the consequent phenotypic variability in a cell population with a Toggle Switch in each cell. Finally, the author contributed actively to the composition and revision of the manuscript, along with Jason Lloyd-Price and Andre Ribeiro. This publication has also been included in the dissertation of Jason Lloyd-Price (2015).

In **Publication III**, the author assessed the role of the kinetics of inducer intake on the response of the inducible gene's expression upon the introduction of inducers in the media. In particular, the author studied how the nature of the intake process of inducers (whether it is pure diffusion, with positive feedback or with negative feedback) determines the transient and steady-state dynamics of gene expression. In this work, the author conceived the study with Andre Ribeiro and was fully responsible for the design and the simulation of stochastic models of intake mechanisms as well as the analysis of the simulation results, under the supervision of Andre Ribeiro. The author was also actively involved in the writing of the manuscript.

In **Publication IV**, from single-cell single-RNA measurements of the transcription activity of the lac-ara-1 promoter, we characterized the intake kinetics of a lactose analogue in the regime of high media concentrations. The author conceived this work. Also, the author designed the models of intake and of gene expression. The author is also responsible for analysis of the measurements (conducted by Samuel Oliveira and Nadia Goncalves) and the interpretation of the experimental results. Finally, the author, along with Andre Ribeiro, wrote the manuscript.

Finally, in **Study V**, we investigated how the kinetics of transcription initiation affects the sensitivity of gene expression to perturbations in the σ factor population of a cell. In this work, the author derived the model of gene expression combined with the model of σ factor competition for RNA polymerase core enzymes and tested how the dependence of gene expression on the σ factor population changes with kinetic parameters of transcription initiation. The author is partially responsible for the analysis of the measurement data, which are obtained by Vinodh Kandavalli and Jerome Chandraseelan. The author also participated actively in preparation of the manuscript.

1 Introduction

1.1 Background motivation

The discovery of the gene regulatory network (Jacob and Monod, 1961), responsible for many cellular complex behaviors (Kauffman, 1969), foresaw the prospect of single-cell machines: by inserting into cells synthetic genetic circuits, a module of genes interacting with one another in a predefined motif, one can create or modify cellular functions or responses to external signals. However, despite the numerous proposals for circuit motifs with attractive theoretical dynamics (Samad et al., 2005; Wolf and Arkin, 2003), only a limited number of these were successfully implemented (Anderson et al., 2007; Elowitz and Leibler, 2000; Gardner et al., 2000; Stricker et al., 2008). Furthermore, the stability and tunability of these circuits still has much left to be desired (Chandraseelan et al., 2013; Elowitz and Leibler, 2000), particularly when compared with genetic circuits existing in nature (Ishiura et al., 1998; Shea and Ackers, 1985), which were tailored merely by trial and error.

The behavior of a genetic circuit depends not only on its composition but also on the surroundings, namely the host and the environmental context of this host (Cardinale and Arkin, 2012). A circuit's composition includes its components and how they are wired to one another. These wires govern the flow of information between the genes and define the logical behavior of the circuit. Meanwhile, the host organism supports the operation of the circuit via e.g. its numbers of native transcription factors, RNA polymerases and ribosomes, the growth dynamics, etc. These factors cause the behavior of genetic constructs to vary even between cells of identical genotype (Elowitz and Leibler, 2000). Lastly, as the environmental conditions change or as cells navigate in space, the environmental cues constantly permeate through the membrane and affect some of the

host factors (Farewell et al., 1998; Jensen et al., 1993), leading to potential changes in the circuits' operation.

In bacteria, the regulatory mechanisms of gene expression to adapt to the environmental conditions target mostly at the process of transcription. For example, when *E. coli* cells are placed in poor media conditions, the production of stress-responding sigma factors, subunits of RNA polymerases (Gruber and Gross, 2003), are induced. These factors bind to the RNA polymerase core enzymes to form holoenzymes and assist the binding of these holoenzymes to the promoter region of the stress-responding genes, thus enhancing their transcription (Farewell et al., 1998). Also, when lactose appears in the media and is absorbed by cells, lactose molecules bind to the *lac* repressors in the cells and reduce their affinity to the *lac* promoter drastically (Lewis, 2005). As the promoter is cleared from the repressors, the transcription of the genes of the *lac* operon is triggered (Jacob and Monod, 1961). It is expected that these regulatory mechanisms will interfere directly with the interactions between genes of the circuits, where the regulations of one gene over another also occur primarily at the transcription level. Therefore, to thoroughly assess how genetic circuits respond to changes in the environmental conditions, one needs to characterize the transcription kinetics of the constituent genes, the impacts of the host factors on this kinetics and the mechanisms through which the host factors sense fluctuations in the environment (Cardinale and Arkin, 2012).

With the recent advances in fluorescent imaging and image analysis techniques, the process of gene expression can now be monitored in great detail (Golding et al., 2005; So et al., 2011; Yu et al., 2006). From *in vivo* measurements of the transcription dynamics of individual genes at the single event level, the key intermediate processes in transcription (Kandhavelu et al., 2012a; Muthukrishnan et al., 2012), i.e. the closed complex formation, isomerizations and the open complex formation, can be discerned. The findings derived from the observations are contributing to the design of more realistic models of gene expression dynamics. Furthermore, by observing this dynamics in various environmental conditions and host strains, one can expect to learn about the mechanisms through which cells regulate the expression of individual genes in response to media variations (Dong and Schellhorn, 2009a; Farewell et al., 1998; Mäkelä et al., 2013; Marbach and Bettenbrock, 2012). For example, in (Farewell et al., 1998), from comparisons of the gene expression profiles in wild type *E. coli* and in a *rpoS*-deletion mutant strain, it was shown that sigma factors (Gruber and Gross, 2003) compete for the binding to a limited pool of RNA polymerase core enzymes. As a result, increasing the abundance of one sigma factor

negatively regulates the expression of genes recognized by the other sigma factors (Dong and Schellhorn, 2009a). This knowledge will aid in understanding the operation of natural genetic circuits and in designing synthetic circuits with desired behaviors by selecting the right components from a well-characterized promoter – regulator library. It also helps in identifying what properties of the host organisms and the environment ensure robust operations of the circuits.

1.2 Thesis objective

This thesis aims to characterize cellular mechanisms through which the transcription dynamics of individual genes becomes environment-dependent and how the impact of this external regulation on individual genes propagates to the dynamics of the circuits they constitute.

For this, the author and colleagues first investigated the dynamics of genetic circuits in a fixed context. Here, the context's impacts are defined explicitly in the kinetic rates of the constituent genes' expression. Recent works have shown that the initiation of transcription, where most regulation of gene expression occurs, consists of multiple sequential rate-limiting steps (Kandhavelu et al., 2012b; Muthukrishnan et al., 2012), namely the closed complex formation, isomerizations and the open complex formation. Changes in the context (i.e. promoter – regulator pairs, host strains and environmental conditions) affect the number and durations of the rate limiting steps in the transcription initiation process of individual genes (Kandhavelu et al., 2012b; Muthukrishnan et al., 2012, 2014). Using stochastic modeling, we made use of this knowledge to study how the behaviors of genetic circuits are affected by these changes in context by varying the parameters of transcription initiation of the component genes of the circuit. This objective was addressed in **Publication I.**

Next, we observed the operation of genetic circuits in the context of proliferating cell populations, where cell growth and cell division are significant factors. Recently, it has been recognized that the partitioning of plasmids, RNAs, proteins and other macromolecules during cell division is not without errors (Huh and Paulsson, 2011a; Lindner et al., 2008; Lloyd-Price et al., 2012b), resulting in differences in the number of molecules inherited by each daughter cell after division. Further, in stress conditions, cells may employ higher-variance partitioning schemes (Gupta et al., 2014; Männik et al., 2012) to

diverse sister cells even further. Here, using a stochastic model of gene expression (McClure, 1985) and various disordered and ordered partitioning schemes of regulatory molecules during cell division (Huh and Paulsson, 2011b), we explored the effects of errors in partitioning on the behavior of genetic circuits across cell generations. The results should inform on how this phenomenon causes qualitatively different effects from those caused by other sources of noise (e.g. in transcription, RNA degradation, etc.). This objective was addressed in **Publication II.**

Finally, two cellular mechanisms through which gene expression becomes environment dependent were studied: the intake of gene expression inducer (Boezi and Cowie, 1961) and the σ factor competition for RNA polymerase (Farewell et al., 1998).

The first mechanism, featured in **Publication III** and **Publication IV**, determines how inducer molecules enter cell cytoplasm from the media and trigger the expression of the target gene. Here, we theoretically investigated how the nature of the inducer intake system (whether it is purely diffusive, with positive feedback or with negative feedback) affects the expression dynamics of the target gene in response to the introduction of inducers in the media. We next applied this knowledge to characterize the intake kinetics of a lactose analogue from measurements of the *in vivo* transcription dynamics of the lac-ara1 promoter in the *Escherichia coli* strain DH5-αPRO (Lutz and Bujard, 1997) at the single event level (Golding et al., 2005).

The second mechanism allows cells to switch between gene expression profiles to adapt to specific stress conditions. This is achieved through the environment-dependent population of σ factors (Farewell et al., 1998; Jishage and Ishihama, 1995; Jishage et al., 1996), which are subunits of the RNA polymerase that are required for specific binding of the RNA polymerase to gene promoters (Gruber and Gross, 2003). Due to the limited pool of RNA polymerases (Grigorova et al., 2006; Maeda et al., 2000), the appearance of a stress-induced σ factor is expected to cause correlated reductions in the expression of all genes recognized by other σ factors. In practice, some of such genes were found to be impervious to this appearance (Dong and Schellhorn, 2009b; Rahman et al., 2006; Tani et al., 2002), for unknown reasons. In **Study V**, using simulations and *in vivo* measurements, we addressed this phenomenon by studying the transcription kinetics of two RpoD (housekeeping σ factor)-dependent promoters, P_{BAD} and P_{tetA}, differing in their sensitivity to the appearance of the stress responding σ factor RpoS in cells.

1.3 Thesis outline

The thesis is organized as follows:

Chapter 2 provides the biological background of the thesis's topics. First, it presents information on the process of gene expression and the sub-process transcription. Next, the chapter discusses the actors that contribute to the transcription dynamics of individual genes. The concept of genetic circuits is also presented.

Chapter 3 presents the relevant methods and approaches used to perform the data acquisition, extraction and analysis. Discussed are experiment systems, statistical tools, as well as the modeling and simulations of cellular processes.

Chapter 4 is the summary of the results presented in the publications composing this thesis.

Chapter 5 provides a discussion of the outcomes and future directions following the presented works.

2 Biological background

This chapter provides an overview of the biological concepts relevant to the thesis. Namely, it presents what is presently known about the dynamics and regulatory mechanisms of the sub-processes of gene expression in *Escherichia coli*. It also discusses qualitatively how the effects of these regulatory mechanisms on the expression dynamics of a single gene can propagate to that of small gene networks.

2.1 The biology of *Escherichia coli*

Being a member of the prokaryotic genus *Escherichia,* the bacterium *Escherichia coli* (*E. coli*) is Gram-negative and rod-shaped, typically 2 μm in length and 0.25-1 μm in diameter (Figure 1), commonly found in the lower-intestine of warm-blooded animals. Despite being an enteric bacterium, *E. coli* can survive for an extended period of time outside the host organism, e.g. in soil, sand, and sediments (Ishii and Sadowsky, 2008). These extra-intestinal conditions can be extremely cold and aerobic, far different from the favorable warm and damp environment inside the gastrointestinal tract. Even its primary habitat, *E. coli* cells constantly have to face fluctuations in the pH level and abrupt fluxes of nutrients and water due to the host's dietary activities. The fact that *E. coli* is able to accommodate itself to such a life between feast and famine makes it an interesting model organism to study the elasticity of cellular processes, such as intake of substances from the environment (Hansen et al., 1998; Jensen et al., 1993), gene expression (Dong and Schellhorn, 2009a; Muthukrishnan et al., 2014), and catabolism and anabolism (Shimizu, 2013; Wang et al., 2010), to environmental cues.

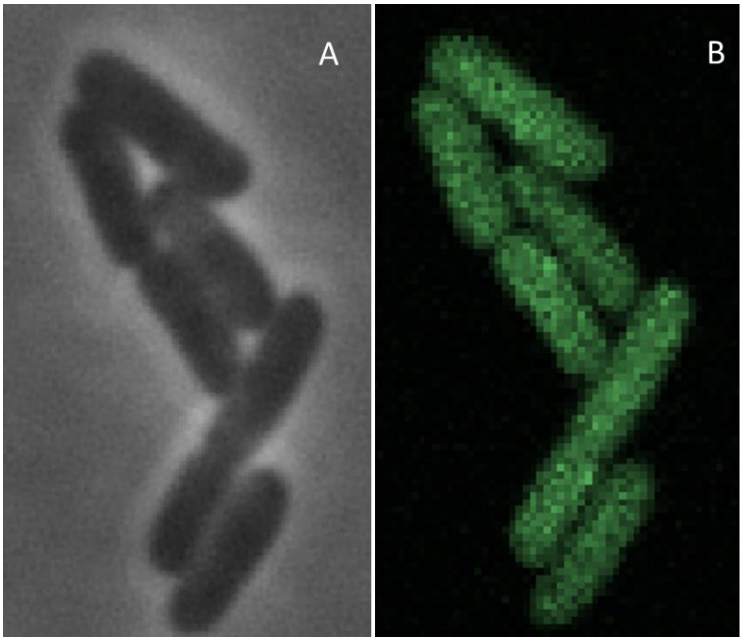


Figure 1. Images of *E. coli* DH5-αPRO cells (Lutz and Bujard, 1997) under (A) phase contrast microscope and (B) confocal microscope. Each rod-shaped object corresponds to a single cell. The cells are filled with fluorescent molecules (MS2-GFP), which makes them visible under the confocal microscope.

Cultivated *E. coli* strains *(E. coli* K-12) can be grown and cultured efficiently and inexpensively in a laboratory setting, and have served as a prokaryotic model organism for over 60 years. Under favorable conditions, *E. coli* can reproduce as fast as 20 minutes by means of binary fission (Cooper and Helmstetter, 1968). Due to the ease of its genetic manipulation, *E. coli* is commonly chosen as the host organism for works involving recombinant DNA (Cohen et al., 1972) to produce a wide range of heterologous proteins, from human insulin (Goeddel et al., 1979), bovine chymosin (Emtage et al., 1983), to vaccines (Zhou et al., 2004), or to study prokaryotic processes, such as RNA and protein synthesis (Golding et al., 2005; Laursen et al., 2005; Muthukrishnan et al., 2012), degradations (Bernstein et al., 2002; Chen et al., 2015).

2.2 Gene expression

Upon division, the daughter cells of an *E. coli* cell inherit not only the materials from the parent cell but also its instructions to survive. These instructions are encoded in the macromolecule DNA (Deoxyribonucleic Acid). They are written in a four-letter-alphabet and

stored in units of heredity called genes. In *E. coli* MG1655's genome, over six hundred essential genes and three thousand non-essential genes (dispensable for cell growth) have been identified (Gerdes et al., 2003).

The information encoded in the genes are used to synthesize functional gene products through a process called gene expression, which consists of two main sub-processes: the first being transcription and the second being translation (Alberts et al., 1994) (Figure 2A). In transcription, the RNA polymerase holoenzyme (RNAPσ) transcribes a gene on DNA and assembles a messenger RNA (mRNA) as a single-stranded complementary copy of that gene. In translation, the mature mRNA is bound and read by ribosomes to synthesize proteins, chains of amino acid residues that, upon folding, are able to perform a vast array of functions in cells. In *Escherichia coli*, translation and transcription are coupled (Figure 2B) (Miller et al., 1970) and determine the dynamics of gene expression and consequently the abundance of gene products (i.e. mRNAs, proteins).

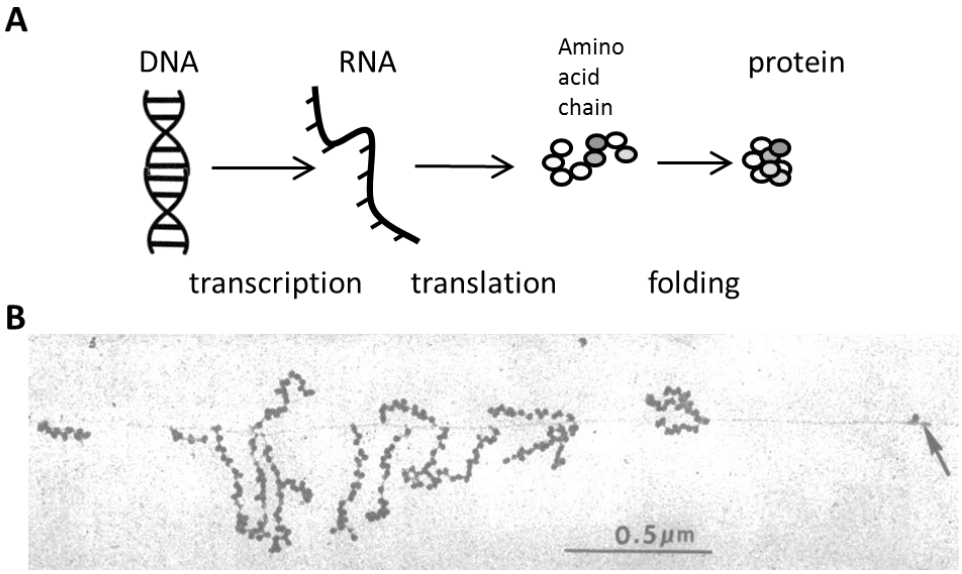


Figure 2: (A) Diagram of sub-processes of gene expression. (B) Visualization of the coupling between transcription and translation: The ribosomes attach to and translate the elongating mRNA segments of increasing lengths (from right to left). The short ribosome-mRNA complexes (also called polyribosomes) farthest to the left may be caused by mRNA degradation. The arrow indicates putative RNA polymerase molecules presumably on or very near the transcription initiation sites. Adapted from (Miller et al., 1970).

Gene expression can be highly responsive to environment cues. For example, it was observed that the expression rate of the genes in the *lac* operon, responsible for the transport and metabolism of lactose in *E. coli* can vary over the order of a hundred fold with the presence of lactose in the media (Jacob and Monod, 1961). The variation of smaller orders was also observed, e.g., in the *L-arabinose* operon (Johnson and Schleif, 1995). Another study conducted on the synthetic lac-ara1 promoter (Lutz and Bujard, 1997) showed that not only the mean rate but also the noise in transcription can be regulated by external factors (Kandhavelu et al., 2012b).

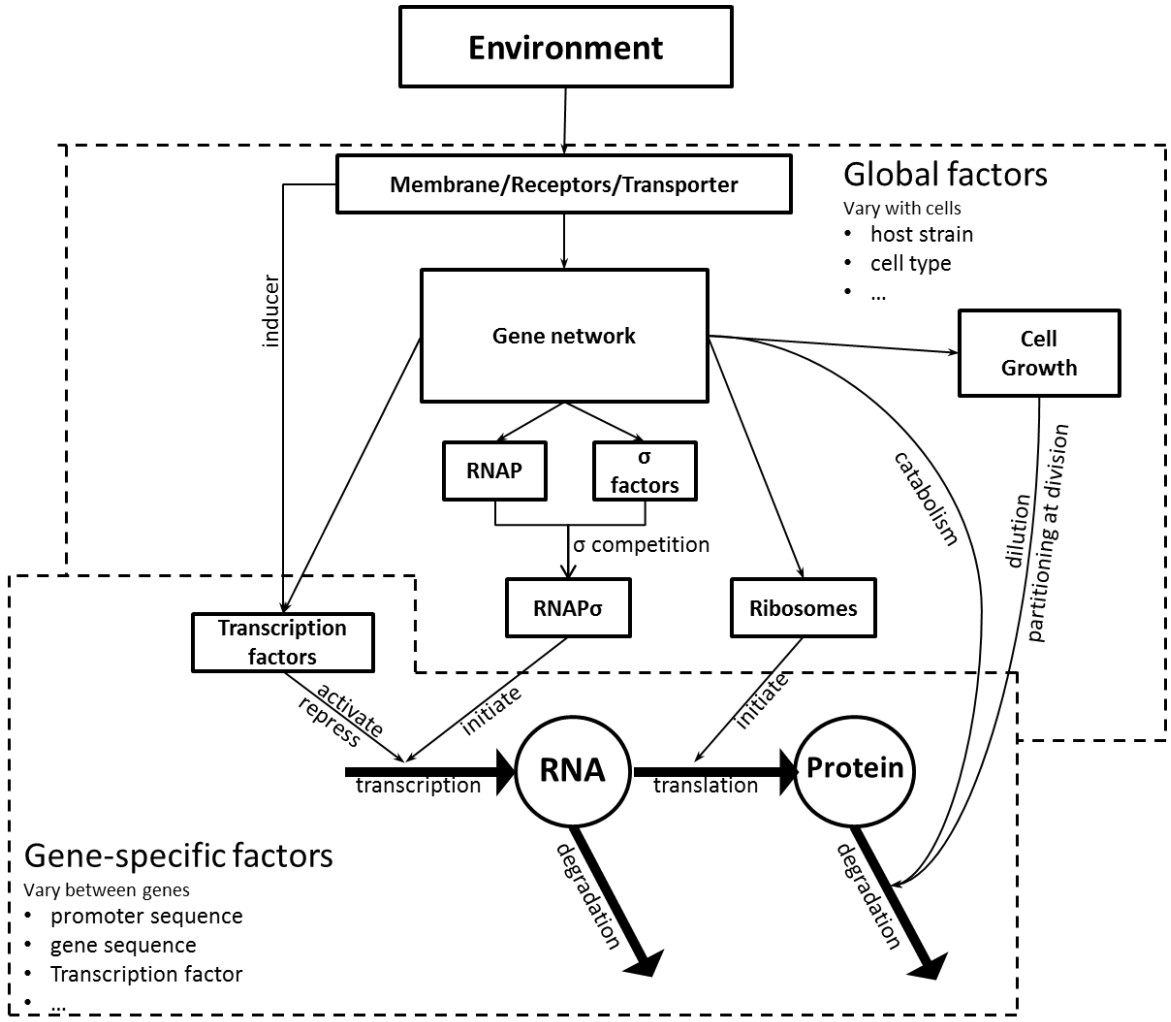


Figure 3. Signal pathways from the environment to the expression of a generic gene, indicated by the abundance of its RNA and proteins. The intermediate factors can have global effects on gene expression or can be gene-specific.

The responsiveness is achievable since, upon fluctuations in the environment (e.g. changes in nutrient availability, osmolarity, temperature, etc.), cells are able to undergo changes that directly or indirectly affect several components of the transcription, translation and degradation machineries (Figure 3). Considering that regulations of bacterial gene expression, at least of the RNA numbers, occur mostly in transcription (Chen et al., 2015; Taniguchi et al., 2010), we only focus on the components of the transcription machinery, which are presented in the following sections.

2.3 Mechanism of transcription

The interaction between RNAP polymerase (RNAP) and the gene segment during the process of transcription is not an elementary reaction; rather it contains several subprocesses (Buc and McClure, 1985; DeHaseth and Helmann, 1995; Lutz and Bujard, 1997; Saecker et al., 2011). As *E. coli*'s DNA consists of two coiled biopolymer strands and exists in double helix form in most of the cell cycle (Cooper and Keasling, 1998; Gordon et al., 1997), the information on the gene segments are not readily assessable for RNAP to read and assemble a new RNA template from it. Instead, RNAP can only bind to a specific region of the gene segment (transcription starting site (TSS)), called the promoter region. Upon binding, the RNAP must perform several modifications on promoter region to render the gene segment accessible and gather energy for the following promoter escape (Hsu, 2002), RNA elongation (Burns and Minchin, 1994; Saecker et al., 2011) and transcription termination (Henkin, 2000).

In vitro measurements suggest that, in *E. coli*, the initiation of transcription is a sequential process (Buc and McClure, 1985; Lutz et al., 2001; McClure, 1985) (described in Figure 4). First, RNAP finds and binds to the promoter region to form the promoter - RNAP closed complex (RP_c). This complex undergoes isomerizations (I_1, I_2 and I_3), followed by DNA melting to form the promoter open complex (RP_o) (i.e. open complex formation (DeHaseth and Helmann, 1995)). Upon forming of the open complex, the promoter may undergo several rounds of abortive initiation (Gralla et al., 1980) before forming the initiating complex (RP). It should be noted that during the initiation of transcription, the promoter region is occupied and modified by the bound RNAP, which prevents the binding of another RNAP that would trigger of another round of transcription.

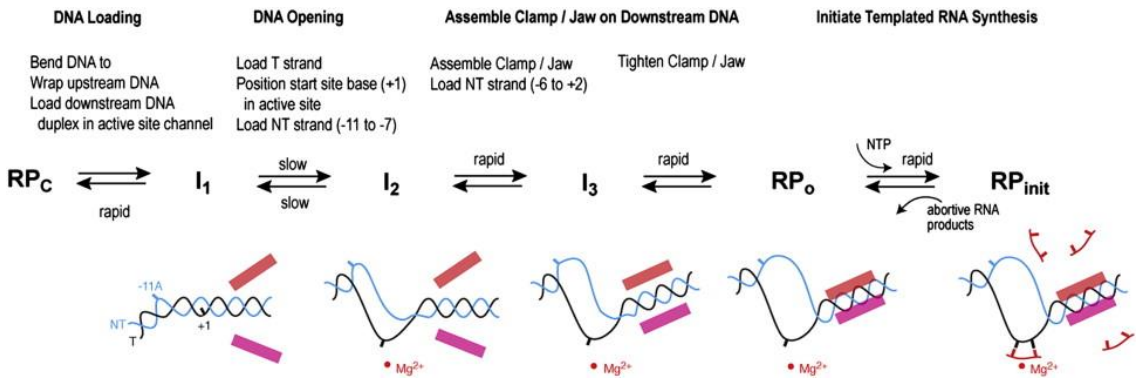


Figure 4. Summary of the steps in transcription initiation, from the formation of the promoter closed complex (RP_c), with the end product being the initiating complex (RP_{init}). The intermediate complexes are isomerizing complexes (I_1, I_2, I_3) and the promoter open complex (RPo). Reused from (Saecker et al., 2011).

Transcription elongation quickly follows the formation of the initiating complex. In this process, RNAP slides in between and unwinds the DNA double helix from 3' to 5' with a speed of ~20-90 base pair per second (Ryals et al., 1982; Veloso et al., 2014; Vogel and Jensen, 1994) and, while doing so, it reads the information on the gene segment and assembles a single-stranded RNA copy of the gene. Upon reaching the end of the gene segment (usually marked by G-C rich inverted repeats which form a hairpin loop (Wilson and von Hippel, 1995)), both RNAP and the template RNA are released from the DNA.

Several studies suggest that the transcription dynamics of several *E. coli* genes, due to fast rates of promoter escape (Hsu, 2002) and transcription elongation (Herbert et al., 2006; Ryals et al., 1982; Vogel and Jensen, 1994), is mainly controlled by the duration of transcription initiation (Saecker et al., 2011). In these studies, both *in vitro* (Lutz et al., 2001; McClure, 1985; Nierman and Chamberlin, 1979) and *in vivo* (Kandhavelu et al., 2012a, 2012b; Muthukrishnan et al., 2012), it was shown that there are two main rate-limiting steps in transcription initiation (associated with the closed and open complex formations) and that other steps only become rate-limiting at temperatures as low as 24^{o}C (Muthukrishnan et al., 2012). Because of this, the intervals between consecutive RNA productions can be fitted well by a model of two to three rate-limiting step, where the steps are sequential and exponentially distributed in time (Kandhavelu et al., 2012a, 2012b; Muthukrishnan et al., 2012) (Figure 5).

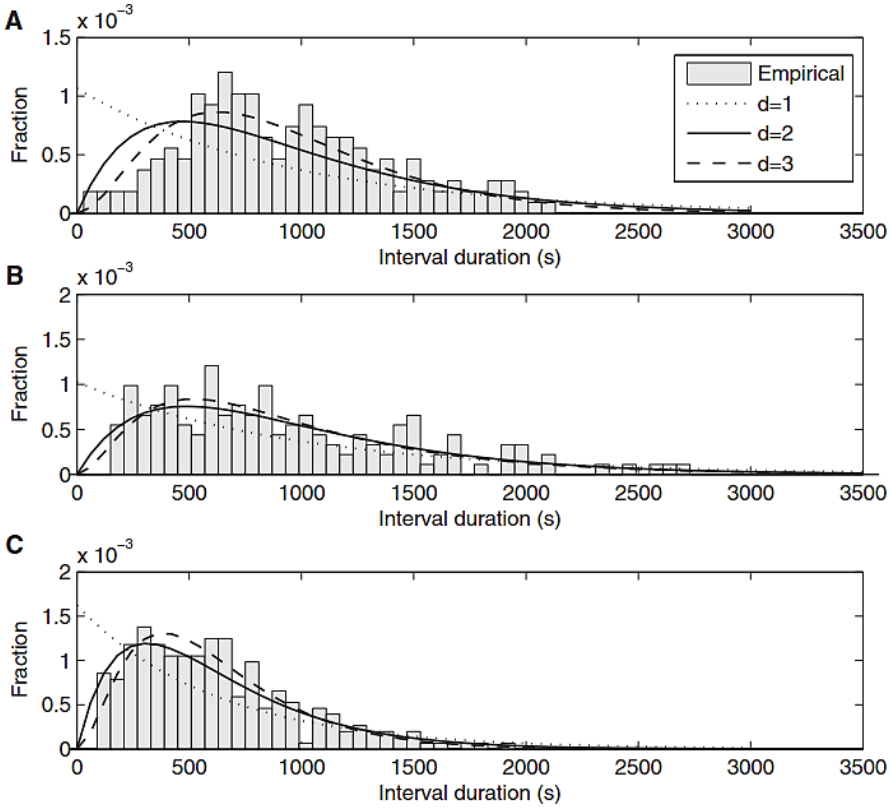


Figure 5. Distributions of time intervals between consecutive RNA production events in individual cells under the control of P_{tetA} in conditions: (A) 0 ng/ml aTc at 37 ^{o}C (B) 15 ng/ml aTc at 24 ^{o}C, and (C) 15 ng/ml aTc at 37 ^{o}C. Also shown are the probability density functions of inferred models of transcription initiation differing in number of rate-limiting steps. Reused from (Muthukrishnan et al., 2012).

As the lengthy initiation process occurs only at the promoter region, the dynamics of transcription and consequently the dynamics of transcripts' abundance are strongly dependent on the sequence of the controlling promoter. In (Lutz et al., 2001), using a panel of promoter sequences derived from lac and araBAD promoters of *E.coli*, the transcription initiation process on these promoters was found to differ widely. In agreement, in (Brewster et al., 2012) it was shown that, by altering a few nucleotides of the promoter region, one can cause significant differences in the stability of the RNAP - promoter complex (Figure 6) and consequently achieve a fold change in the gene expression rate on the order of 10^3, without changing the level of the gene expression regulators (Kinney et al., 2010).

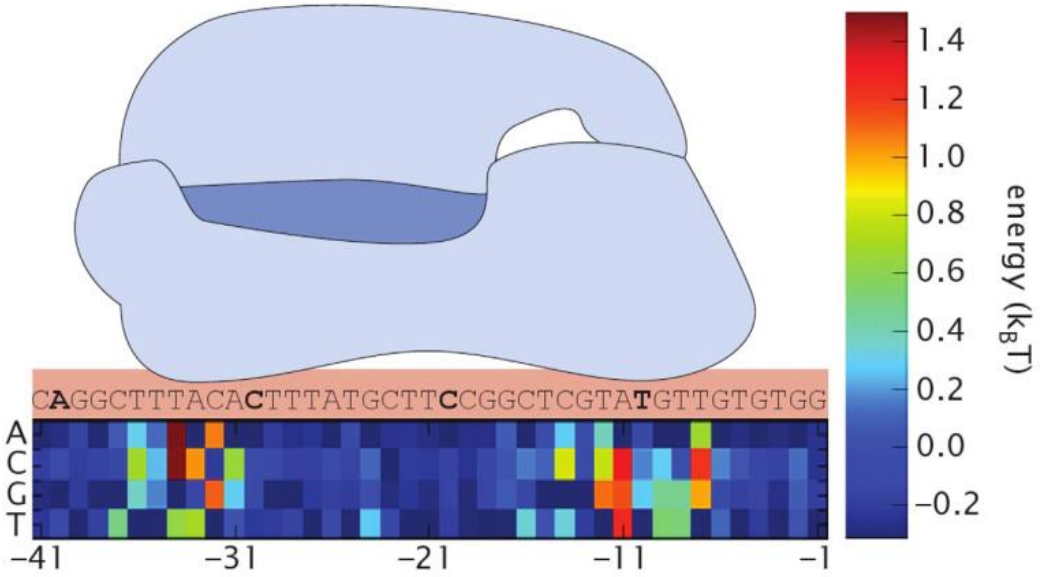


Figure 6. Energy matrix for RNAPσ^{70} binding. The contribution of each base pair to the total binding energy is represented by color. The total binding energy of a particular promoter sequence can be calculated by summing the contribution from each base pair. Positive values indicate unfavorable contributions to binding energy. x-axis coordinates are with respect to TSS. Reused from (Brewster et al., 2012).

The dependence of gene expression on the promoter sequence is a critical knowledge in bioengineering. For example, by modifying the promoter sequence, one can tune the production of recombinant proteins, without the need of altering the gene sequence coding for the protein (which could affect the quality of the desired proteins). This allows achieving high-enough yields of recombinant proteins while avoiding their aggregation which would hamper the proteins' ability to fold properly (Baneyx and Mujacic, 2004).

2.4 RNA polymerase holoenzyme

2.4.1 The functioning of RNA polymerase

RNA polymerase (RNAP) is the main enzyme of the transcription machinery that is responsible to assemble RNA chains from the template genes (Alberts et al., 1994). RNAP is involved in all the three steps of the transcription cycle: initiation, elongation and termination. To initiate transcription (see above), the RNAP binds to the DNA and unwinds a small portion of the DNA double helix located prior to the TSS, leaving the bases of the

two DNA strands exposed. One of the DNA strands is used as a template for the synthesis of a new RNA. The RNAP then escapes from the promoter region and moves stepwise along the DNA, from the 3' end to 5' end of the template strand. While doing so, it reads the exposed DNA template and adds consecutive nucleotides to the RNA chain at the polymerization site, until reaching the template's terminator (Figure 7).

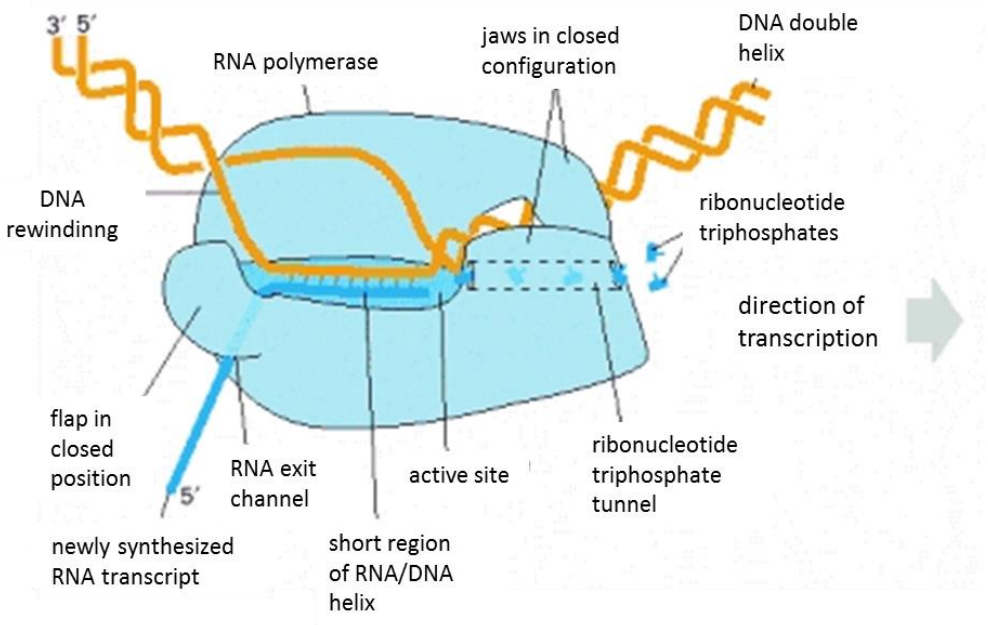


Figure 7. DNA is transcribed by the enzyme RNA polymerase: The RNA polymerase (pale blue) moves stepwise along the DNA (gold), unwinding the DNA helix at its active site. As it progresses, the polymerase adds consecutive nucleotides (here, small "T" shapes) to the RNA chain at the polymerization site using one of the two exposed DNA strands as the template. The polymerase has a flap in the closed position that displaces the newly formed RNA, allowing the two strands of DNA behind the polymerase to rewind. A short region of DNA/RNA helix is therefore formed only transiently, and a "window" of DNA/RNA helix therefore moves along the DNA with the polymerase. Adapted from (Alberts et al., 1994).

In bacteria, RNAP has several subunits: α^{i}, α^{ii}, β, β', ω and σ (Cramer, 2002), with the first five being the 'core' subunits forming the core enzyme with a stable structure. The level of RNAP 'core' enzyme was reported to change with media richness (Cabrera and Jin, 2001; Grigorova et al., 2006; Jishage et al., 1996; Liang et al., 1999), which should result in correlated changes in the expression of all genes in the genome.

In order to initiate transcription, the RNAP core enzyme requires the aid of the σ subunits (Friedman et al., 2013; Ishihama, 2000; Mooney et al., 2005). Upon binding to a σ factor

molecule, the transcription enzyme (called RNAP holoenzyme (RNAPσ)) has increased binding affinity to specific sequences of the promoter region and is therefore able to form the closed complex with the promoter (Friedman et al., 2013). It should be noted that σ factors are not considered 'core' subunits of RNAP as they can leave the RNAP stochastically during transcription elongation (Mooney et al., 2005; Raffaelle et al., 2005). The interaction between RNAP core enzymes and σ factors is illustrated in Figure 8.

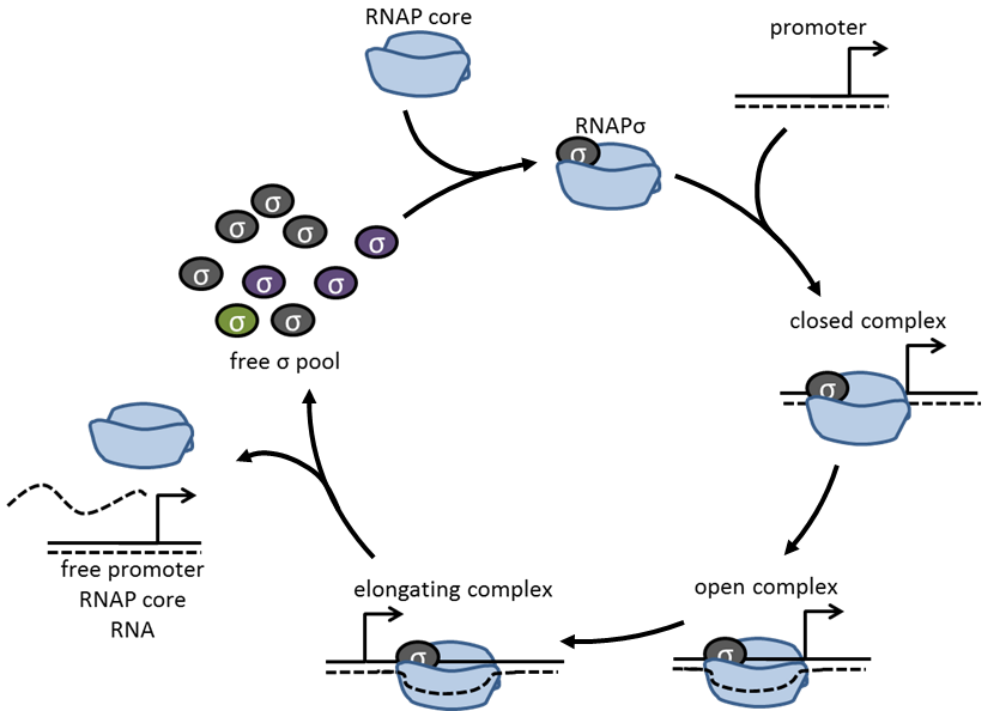


Figure 8. The σ cycle. The pool of free floating σ factors compete for RNAP core enzymes to form holoenzymes (RNAPσ), which in turn bind to the free promoter to form a closed complex. This is followed by the open complex and elongating complex formation. The σ factor is released stochastically from the elongating RNAP – DNA complex and can again participate in the competition for core enzyme binding.

2.4.2 Competition between sigma factors for RNAP holoenzymes

The bacteria *E. coli* contains 7 main sigma factors, which are named after their molecular weight in kDa (Gruber and Gross, 2003). These are σ^{70}, σ^{38}, σ^{54}, σ^{24}, σ^{32}, σ^{19}, σ^{28}. These subunits compete for a limited number of RNAP core enzymes (Grigorova et al., 2006; Maeda et al., 2000) and generate heterogeneity in the pool of RNAP holoenzymes (Figure 9).

The content of the σ pool is highly responsive to environment cues. During normal growth conditions, the σ factor population consists mostly of the house-keeping σ^{70} (RpoD). When cells are put in specific stresses, the expression of other σ factors is either triggered or enhanced. For example, during starvation, both the transcription and the translation of the stress-responding σ^{38} (RpoS) are enhanced, leading to an increase in its protein yield (Altuvia et al., 1997; Sledjeski et al., 1996). Another σ factor, σ^{54}, is found to be activated in nitrogen-limiting conditions (Zimmer et al., 2000). Once produced, these σ factors compete with the existing σ factors for RNAP core enzymes, resulting in perturbations in the RNAP holoenzyme distribution.

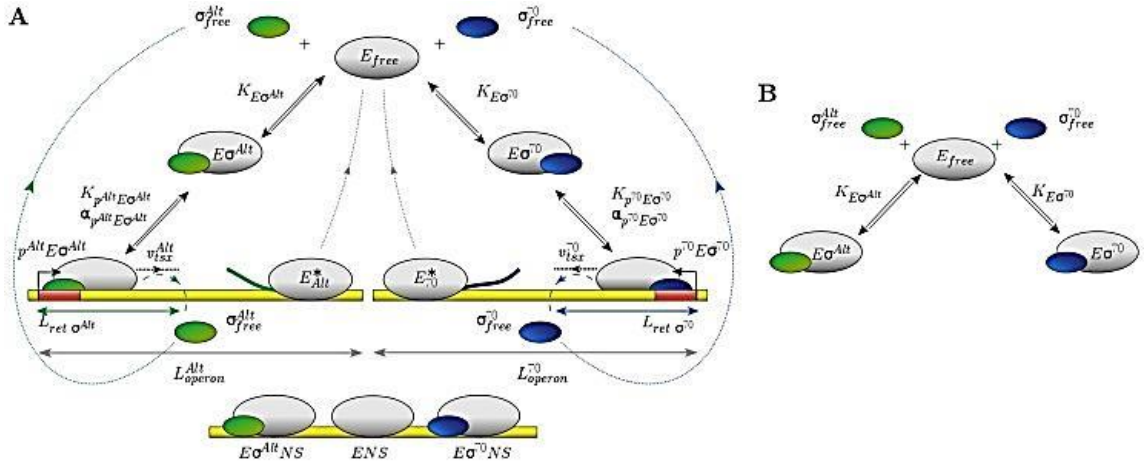


Figure 9. (A) Model for σ factor competition with two types of σ factors, the housekeeping σ factor σ^{70} and σ^{Alt}, a generic alternative σ factor. The model describes binding of σ^{70} and σ^{Alt} to RNAP enzymes (*E*) to form holoenzymes ($E\sigma^{70}$ and $E\sigma^{Alt}$) as well as the transcription process. (B) Core model for holoenzyme formation. Adapted from (Mauri and Klumpp, 2014).

Each promoter, depending on its sequence, can be recognized only by a selection of σ factors (Barrios et al., 1999; Becker and Hengge-Aronis, 2001; Gruber and Gross, 2003; Hengge-Aronis, 2002). It was shown from standard *in vitro* transcription assays (Gaal et al., 2001; Tanaka et al., 1995) that promoters with the conserved sequence (TATAAT) at the -10 position from TSS are recognized by holoenzymes carrying either σ^{70} or σ^{38}. The promoters' preference for RNAPσ^{38} can be enhanced with the presence of nucleotides C and T at the -13 and -14 positions, respectively (Becker and Hengge-Aronis, 2001). This preference for RNAPσ^{38} can be tuned down with the presence of the consensus (TTGACA) at the -35 position from TSS (Hengge-Aronis, 2002). Meanwhile, RNAPσ^{54} are able to bind only to promoters with the consensus sequences at -12 and -24 positions (Barrios et al., 1999).

Due to the diversity in the promoters' selectivity for σ factors, the perturbations in the σ content would propagate to the gene expression level, but not necessarily in a correlated manner. For example, in (Chang et al., 2002; Dong and Schellhorn, 2009a; Farewell et al., 1998) it was observed that, during starvation, the majority of genes whose promoter regions are recognized by σ^{70} have reduced expression while an array of stress responding genes recognized by the stress responding σ^{38} are activated. Interestingly, even among promoters that are dominantly recognized by a single type of RNAPσ, some are more responsive to changes in the σ population than others. In **Study V**, we aim to explain this phenomenon from observations of the kinetic of transcription initiation of these promoters.

2.5 Transcription factors

2.5.1 Repressors, activators, inducers

There are several mechanisms that can regulate the expression of a gene, even when RNAPσ are abundant in cells (Jacob and Monod, 1961; Shimoni et al., 2007). One of such mechanisms is the alteration of the promoter's secondary and ternary structures, signaled by the binding of proteins called transcription factors (Bell and Lewis, 2001; Lewis, 2005; Malan et al., 1984; Miyada et al., 1984). Upon binding to the promoter region, a transcription factor may restrict the promoter's interaction with RNAPσ and hamper transcription (in this case, the factor is called repressor). Transcription factors can also actively recruit RNAPσ to the promoter region and enhance gene expression (and thus named activator). Since the interaction between transcription factors and the promoter depends on both the promoter sequence and the structure of regulatory protein (Dong et al., 1999; Kandhavelu et al., 2012b; Lutz and Bujard, 1997), this mechanism grants cells the ability to diversify the gene expression profile even further than by employing σ factors, due to the limitation in the variety of σ subunits.

The functionality of transcription factors can be altered when bound by specific substances, called inducers. These inducers can render repressors inactive and indirectly promote the expression of the inducible gene. An example of such interaction is the case of the *lac* operon's repressor (LacI), with lactose being its inducer (Bell and Lewis, 2001; Lewis, 2005). The lac repressor is a homo-tetramer, encoded by the *lacI* gene. In the absence of lactose, LacI tetramers slide unspecifically along the DNA (Elf et al., 2007) and bind to

the two operator sites of the lac promoter. The bound tetramer then forms a stable complex with the promoter and prevents RNAPσ from initiating the transcription of the genes in the lac operon. When lactose is introduced to the media and absorbed by cells, lactose molecules bind to LacI tetramers and reduce the binding affinity of the repressors to the operator sites, thus allowing RNAPσ to reach the promoter and initiate the transcription of *lac* genes (Bell and Lewis, 2001; Lewis, 2005). In lab experiments, the lactose mimic, Isopropyl β-D-1-thiogalactopyranoside (IPTG), is frequently used to regulate the activity of lac promoters (Jensen et al., 1993; Kandhavelu et al., 2012b; Lutz and Bujard, 1997; Marbach and Bettenbrock, 2012) instead of lactose due to its immunity to cellular hydrolysis.

Apart from neutralizing repressors, some inducers can promote gene expression by enhancing the functionality of transcription activators. This is the case of cyclic adenosine monophosphate (cAMP) activating the cAMP receptor protein (CRP), the activator of many catabolite genes (Busby and Ebright, 1999). Some inducers can even inverse the repressor's function, turning it into an activator (as in the case of arabinose and the repressor of the L-arabinose operon, AraC (Schleif, 2010)).

2.5.2 Inducer entrance in cells

Many inducers of gene expression are not synthesized by cells, but rather, enter cells from the environment. Upon fluctuations in the concentration of inducers in the media, one expects that the intracellular numbers of inducer and consequently the expression of its target gene will be affected. In (Jensen et al., 1993), it was shown that in wild type *E. coli*, when the concentration of IPTG in the media increases from 1 μM to 10 μM, the activity of lac promoter increases over 10 folds. In (Lutz and Bujard, 1997), in an *E. coli* strain overexpressing the lac repressor (DH5-αZ1), this fold change is reported to be on the order of 10^3.

To enter cells and regulate gene expression, inducer molecules must travel across the cell membrane. In *E. coli*, the membrane consists of two layers (Alberts et al., 1994; Zimmermann and Rosselet, 1977), the inner membrane and the outer membrane, separated by the periplasmic space (Figure 10). The fabric of the membranes is two stacked sheets (referred as the lipid bilayer) of amphipathic molecules, with the hydrophilic heads pointing outward and the hydrophobic tails buried inward. The lipid bilayer is semi-permeable: it only allows lipid-soluble molecules and some small, uncharged molecules to pass through (Finkelstein, 1976; Lodish et al., 2000), while repelling ionic and larger

molecules (Decad and Nikaido, 1976; Ronald Kaback, 1983; Walter et al., 1986). Located on the membrane layers are porins, membrane and transmembrane proteins (Alberts et al., 1994; Benz and Bauer, 1988), which are various in type and allow more specific molecules to travel across the membrane via either facilitated diffusion or active transport (Nikaido, 2003). Along with the lipid bilayers, these proteins play an important role in the dynamics of the intracellular inducer numbers and consequently the response of gene expression to the appearance of these inducers in the media.

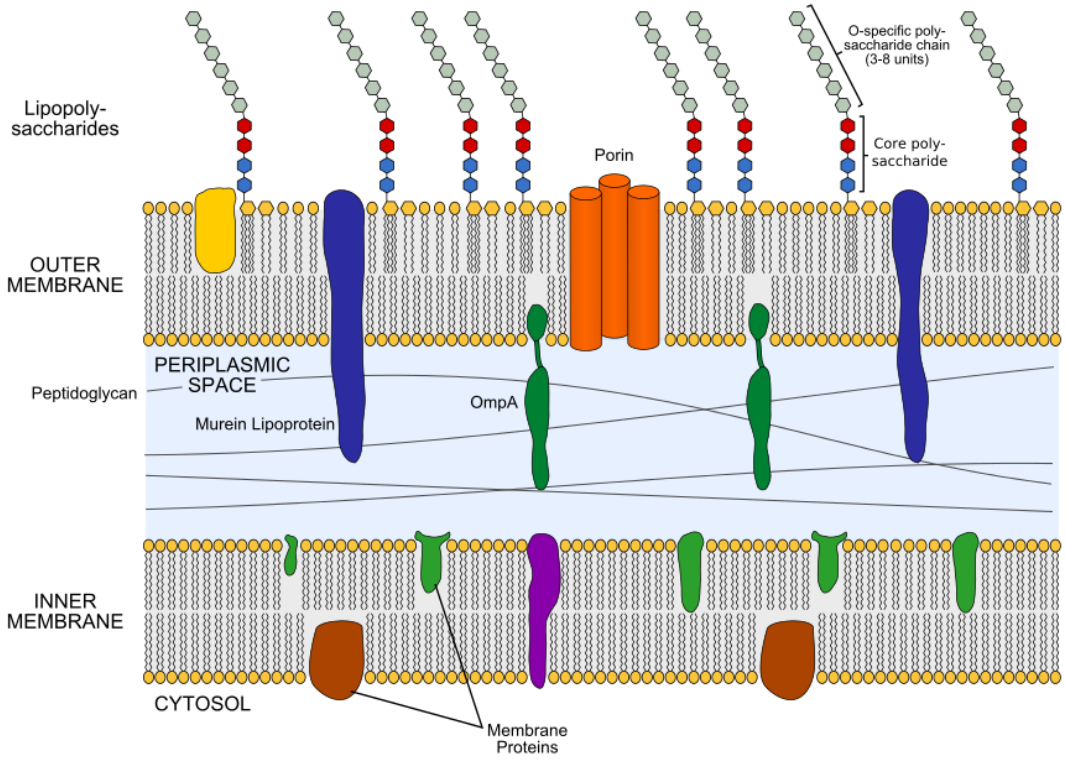


Figure 10. Illustration of *E. coli* membrane. The membrane consists of two lipid bilayers, the outer membrane and the inner membrane, separated by the periplasmic space. The membrane fabric is made up mostly of amphipathic molecules, with hydrophilic end (yellow) exposed to aqueous solution and hydrophobic tails buried inward. Bound to the membrane layers are porins (orange), membrane (brown and green) and transmembrane (dark blue) proteins. Reused from Wikimedia (https://commons.wikimedia.org/wiki/File:Gram_negative_cell_wall.svg).

The lactose intake system is one of the best studied intake mechanisms of inducers in molecular biology (Bentaboulet and Kepes, 1981; Ghazi and Shechter, 1981; Hansen et al., 1998; Jacob and Monod, 1961; Jensen et al., 1993; Marbach and Bettenbrock, 2012; Ozbudak et al., 2004; Ronald Kaback, 1983). Most of these studies were conducted in the regime of low lactose concentration. In this regime, the mean expression rate of the target

gene exhibits a close-to-linear dependence on the intracellular inducer level (Hansen et al., 1998; Jensen et al., 1993), making it possible to study the intake kinetics with assays of the inducible gene's products alone. At low concentrations (less than 0.25 mM), the cellular intake of lactose is found to have a positive feedback, conferred by the lactose permease LacY. The LacY protein is a proton symporter (Ronald Kaback, 1983), which transports protons and lactose molecules simultaneously using the transmembrane proton gradient (Ramos and Kaback, 1977). Following the entrance of the first few lactose molecules in the cell, the genes in the *lac* operon are activated. LacY proteins are then synthesized and enhance the intake of lactose even further (Hansen et al., 1998; Jensen et al., 1993; Ozbudak et al., 2004) (Figure 11). At higher lactose concentrations, the role of LacY is no longer significant (Jensen et al., 1993). Lactose molecules enter cells by means of alternative symporters (Bentaboulet and Kepes, 1981; Ghazi and Shechter, 1981) and potentially through passive diffusion (Decad and Nikaido, 1976).

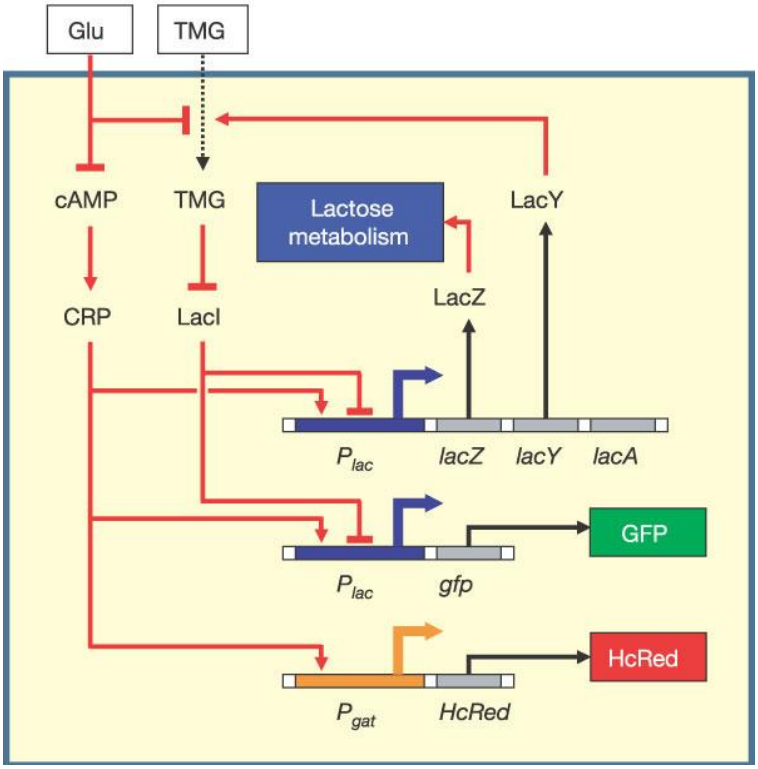


Figure 11. Diagram of the intake system of Trimethylglycine (TMG), used here as a lactose mimic: In the absence of glucose (Glu) in the media, TMG can enter cells and activate the genes in the lac operon (*lacZYA*), among which is *lacY* gene encoding for the lactose permease protein. Upon translation, LacY protein will enhance the intake rate of TMG even further. Reused from (Ozbudak et al., 2004).

Generally, the kinetics of the inducer intake process in the regime of high concentrations is less explored. In this regime, the transient period of gene expression following the addition of inducers to the media is relatively short. Also, once reaching steady states, the target gene's activity is close to full-induction and thus can no longer reflect the changes in the intracellular inducer levels. In **Publication IV**, we addressed this problem with a method to characterize the intake kinetics of lactose in cells in this regime of concentration (from 0.25 mM up to 2 mM) using *in vivo* measurements of transcription at the single cell, single event level (Golding et al., 2005).

2.6 Genetic circuits

2.6.1 Motifs

Individual genes have a limited set of possible dynamical behaviors. For example, due to the time-consuming transcription (Kandhavelu et al., 2012a; Lutz et al., 2001; Muthukrishnan et al., 2012) and degradations (Chen et al., 2015; Taniguchi et al., 2010), most genes are responsive only to slow changes in the regulatory molecules' number and therefore can act as a low pass filter to external signals (Samoilov et al., 2002). Also, their expression rate was found to vary with the dose of inducers (Jensen et al., 1993; Müller-Hill, 1990) and temperature (Muthukrishnan et al., 2012; Rouvière et al., 1995), within a certain range. However, these behaviors are usually monotonic and simple.

In practice, genes rarely function separately. The chromosome of organisms as simple as *E. coli* contains thousands of genes interacting with one another forming a highly connected gene network with a hierarchical organization (Martínez-Antonio et al., 2012; Ravasz et al., 2002). The expression of one gene can regulate directly or indirectly the expression of another gene or even itself via multiple overlapping pathways (Shen-Orr et al., 2002). These feedforward and feedback loops allow cells to achieve more complex and non-linear functions (Wolf and Arkin, 2003).

Usually, each cellular function is performed by a module, a unique cluster of genes interacting with each other via their gene products in repeating patterns called motifs (Shen-Orr et al., 2002). The module's motif determines the flow of information within the small gene network and consequently the copy dynamics of proteins and RNAs of the constituent genes. For example, in Cyanobacteria, the circadian rhythm, which allows cells to adapt better to rhythmic environments, is administrated by the *kaiABC* gene cluster

(Nakahira et al., 2004). In this cluster, the expressions of the three genes *kaiA*, *kaiB* and *kaiC* are wired so that the products of *kaiC* gene indirectly represses its own expression and thus gives rise to oscillations in the number of KaiC protein over time (Ishiura et al., 1998).

Along with the identification of existing gene modules in nature, there have been attempts to engineer new modules that exhibit the desired behaviors (Elowitz and Leibler, 2000; Gardner et al., 2000; Stricker et al., 2008). These modules are designed to mimic the behaviors of their electronic counterparts (Wolf and Arkin, 2003) and aimed to serve as a platform to control cellular functions or alter cellular responses to environmental conditions.

2.6.2 Toggle switch and Repressilator

Among the designed genetic circuits, the Toggle switch and the Repressilator are the most widely studied and implemented. They recapture two complex behaviors that have been observed in nature, namely multistability (Arkin et al., 1998; Shea and Ackers, 1985) and oscillation (Ishiura et al., 1998), respectively.

The Toggle Switch, which was realized for the first time in (Gardner et al., 2000), consists of two genes (named gene A and gene B) with antagonistic interactions (Figure 12A). When gene A is activated, its proteins are produced and down-regulate the expression of gene B. This reduces the protein synthesis rate of gene B and consequently the number of protein B, which in turn allows gene A to express even further. Likewise, when gene B is activated, gene A is repressed, which enhances gene B's expression further. The positive feedback in the expression of both genes causes their protein numbers either to stay low or to stay high. The circuit therefore has two stable states "A on B off" and "B off A on" (the system is thus regarded as being bistable). Due to several sources of noise, both intrinsic and extrinsic (Elowitz et al., 2002), the circuit may "flip" stochastically from one state to the other with certain probability rates. Also, by changing the media conditions, it can be biased to favor the one state over the other (Gardner et al., 2000). The system is therefore said to be capable of making "decisions" based on environment cues.

The Repressilator, realized for the first time in (Elowitz and Leibler, 2000), is composed of three genes (named gene A, gene B and gene C) repressing one another in a loop (Figure 12B). With this structure, the expressions of all genes in the circuit have delayed negative feedbacks. This delayed feedback causes the protein numbers of the genes in the

circuits to oscillate over time, with relatively accurate periods following a sub-Poisson distribution (Chandraseelan et al., 2013) (i.e. the rises of the protein level of a certain gene in the circuit occur in a less noisy fashion than a Poisson process).

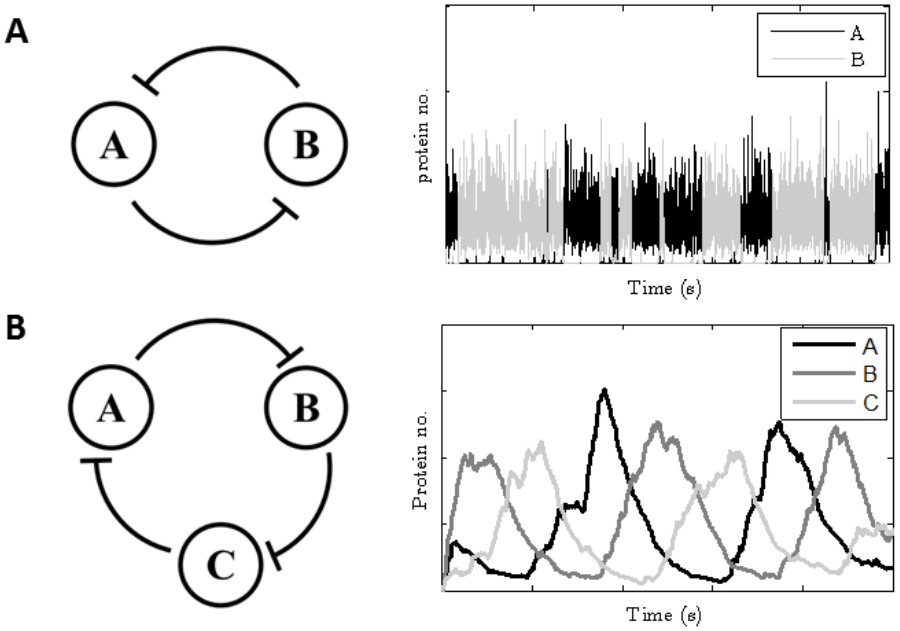


Figure 12. Topology (left) and the corresponding behavior (right) of two genetic motifs. (A) Toggle switch (Gardner et al., 2000): The motif consists of two genes antagonistically repressing (denoted by the "blunt" arrows) each other, resulting in a system whose state is randomly switched between "A on B off" state and "A off B on" state. (B) Repressilator (Elowitz and Leibler, 2000): The motif consists of three genes repressing one another in a loop. The protein level of each gene in this motif oscillates over time with non-exponentially distributed periods.

There are also other genetic constructs that can exhibit multistability (Ozbudak et al., 2004) and oscillation (Atkinson et al., 2003; Stricker et al., 2008). Despite employing different topologies from the classic circuits, they operate following the same principal of positive feedback (for the multistability) and delayed negative feedback (for oscillation).

2.6.3 Signal processing with genetic circuits

Not all circuits function autonomously in cells, but rather some operate according to external signals. These external signals, originated from another cellular process or from the extracellular environment, directly affect the expression of one or more input genes and propagate throughout the small gene network. Thus, one can treat small genetic networks as genetic filters, where the expression of each constituent gene is a filter output.

Naturally, the response of a genetic filter depends on the filter's structure, i.e. the wiring schemes between the promoter-regulator pairs. By following the design of electrical circuits, one can hope to create genetic circuits the desired responses both in the amplitude domain and in the frequency domain (Wolf and Arkin, 2003). Such filters when realized are expected to provide more selective controls over cellular processes. For example, by linking the output of genetic filters to the control mechanism of a process, one can activate or deactivate that process by applying input signals of certain magnitude or frequency to the system.

Several designs of filters in the amplitude domain have been proposed. These include the biphasic amplitude filter (Wolf and Arkin, 2003) and logic gates (Anderson et al., 2007; Moon et al., 2012), which allows the expression of the output gene to be triggered when the input signal is within a certain range of magnitude. Meanwhile, filtering on the frequency domain is more challenging as most of the network components (promoter-regulator pairs) only allow signals of low frequencies to pass by (Samoilov et al., 2002), due to the time consuming transcription initiation (Muthukrishnan et al., 2012) and degradation of proteins and RNA (Chen et al., 2015; Taniguchi et al., 2010).

Aside from the structure, the performance of genetic filters is largely determined by the quality of its components. Due the stochastic nature of gene expression (Elowitz et al., 2002), signals are constantly added with noise when traveling through the gene network, which severely degrades the filters' performance. This problem is assessed in **Publication I**. Also investigated in this publication are the measures to abate the problem by modifying the kinetics of transcription initiation of the motifs' constituent genes.

2.7 Cell division and the partitioning of molecules in cell division

2.7.1 The cell cycle of *E. coli*

The life of *E. coli* switches between feast and famine (Wang et al., 2010). To survive in hazards, cells have to rely on proliferation and diversification to ensure the survival of at least one cell lineage in the population.

The proliferation of *E. coli* is performed through binary fission, where the parent cell is divided into two daughter cells of identical DNA contents. This process is split into three periods (Cooper and Helmstetter, 1968) (Figure 13). The first, also called as period B,

elapses between cell "birth" and the initiation of chromosome replication. In this period, cells contain a single copy of the chromosome. When cells reaches a certain length, they enter the second period, period C, with the trigger of chromosome replication at the replication origin (oriC) (Cooper and Keasling, 1998; Gordon et al., 1997). From the origin site, the chromosome is replicated bidirectionally with the unwinding of the double-stranded DNA into two template strands and the assembly of two new double-helix strands from the template strands. Once replication reaches the opposite site of oriC (terC), the two double helix strands are separated from each other and migrate directly and rapidly to the quarter-cell destinations (Gordon et al., 2004). The last period, period D, is the preparation of cell division, where the parent cell is compartmentalized by the division septum and eventually divided into two new daughter cells, each inheriting a chromosome and approximately half of the cytoplasm's content. Surprisingly, even though the duration of period C, during which the chromosome is being replicated, is non-negligible (on the order of 40 minutes in *E. coli* (Cooper and Helmstetter, 1968)), it was reported that the process of gene expression does not appear to be interrupted throughout the cell cycle (Arends and Weiss, 2004; Liu et al., 1994, 1993).

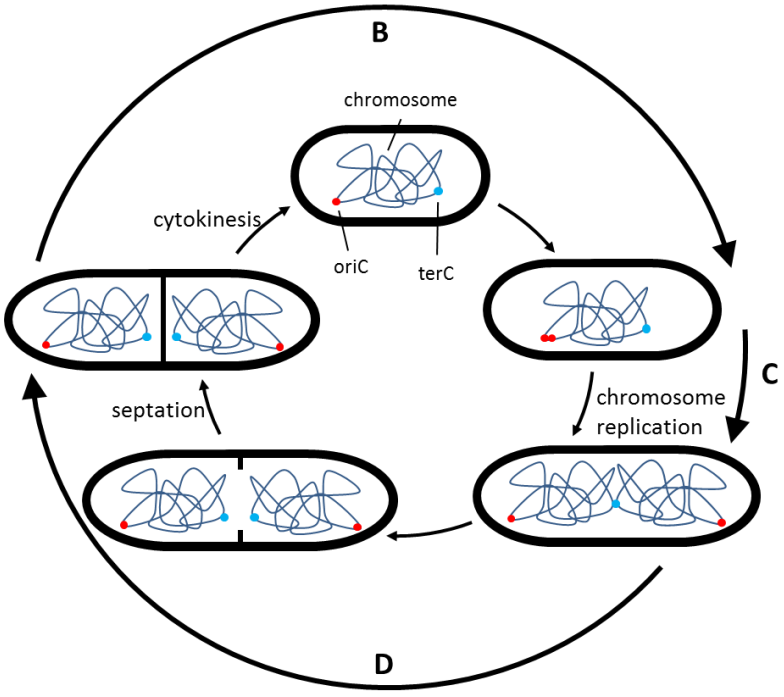


Figure 13. The cell cycle of *E. coli*. There are three periods (B, C and D) in *E. coli*'s cell cycle, the milestones of which are the chromosome replication initiation at the origin site oriC (red dots), the completion of chromosome replication at the termination site terC (blue dots), and cell division.

2.7.2 Partitioning of molecules in cell division

Along with cell proliferation to counter the death rate during hazards, the bacterial population also employs cell differentiation to diversify the death rate, so as to avoid the annihilation of the whole population (Kussell and Leibler, 2005). Cell differentiation can be generated not only from the stochastic nature of the biochemical reactions involved in the process of gene expression (Kaern et al., 2005; McAdams and Arkin, 1999) but also from the partitioning of plasmids, RNAs, proteins and organelles during cell division (Coquel et al., 2013; Lindner et al., 2008).

In (Huh and Paulsson, 2011b), it is defined the level of errors in partitioning $Q_x{}^2$ as the statistical difference between the number of molecules inherited by each daughter cell:

$$Q_x{}^2 = \frac{\left\langle (L-R)^2 \right\rangle}{\left\langle x \right\rangle^2} = \frac{\left\langle \sigma_{L|x}^2 \right\rangle}{\left\langle L \right\rangle^2} = CV_L^2 - CV_x^2 \tag{2.1}$$

where $\langle ... \rangle$ denotes the averages over all dividing cells in the population, x is the number of partitioning molecules before division, L and R are the numbers of molecules inherited by the left daughter and the right daughter, respectively.

In the absence of active segregation mechanisms, most molecules segregate randomly to two daughter cells with the probability ½ after division, resulting in the numbers of molecules inherited by each daughter cell following a binomial distribution. The level of errors in partitioning is therefore given as:

$$Q_x{}^2 = \frac{1}{\left\langle x \right\rangle} \tag{2.2}$$

It is notable that, for species that exist in large copy number ($\langle x \rangle \gg 1$), $Q_x{}^2$ is small. However, considering that many proteins and RNA, especially ones with signaling functions, are low-copied (Taniguchi et al., 2010), the errors in the partitioning of their molecules would add significant variance to their molecular copy numbers at cell division.

In the presence of segregation mechanisms, the level of errors in partitioning in division can deviate from that of random partitioning ($1/\langle x \rangle$). These mechanisms can be either disordered (Figure 14), which increases the level of errors, or ordered (Figure 15), which decreases the level of errors.

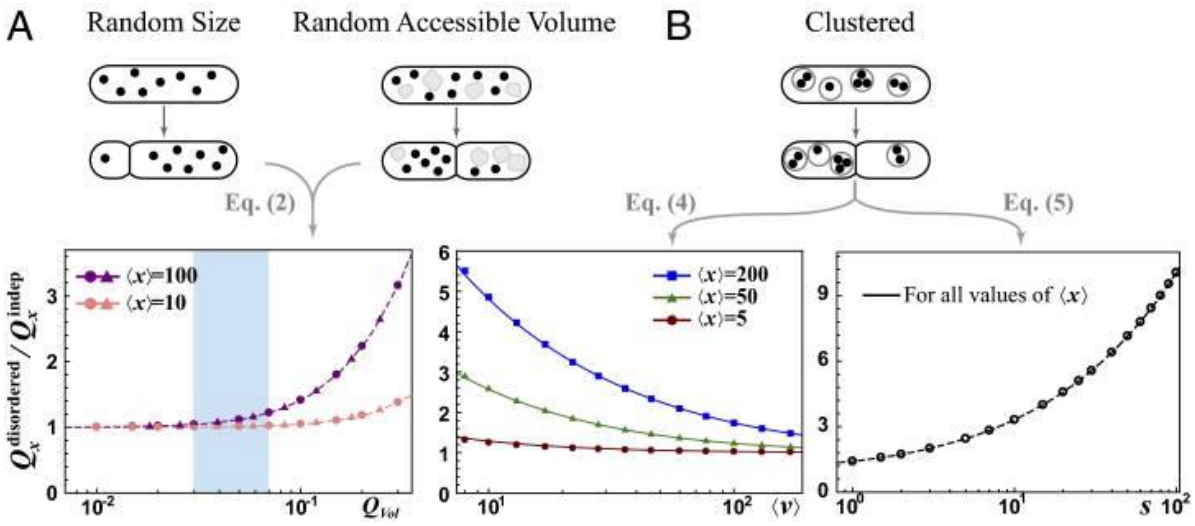


Figure 14. Examples of disordered segregation schemes. (A) Variation in division site or random segregation of other large components differentiates sibling cells' size or the volume accessible to the segregating component. The level of errors in partitioning ($Q_x^{\text{disordered}}$) is increased with the level of errors in the partitioning of the assessable volume Q_{Vol}. (B) Segregating molecules (dots) are randomly grouped into vesicles (circles) and the vesicles are independently partitioned into daughter cells. $Q_x^{\text{disordered}}$ increases with increasing vesicles' size (s). Adapted from (Huh and Paulsson, 2011b).

The disordered schemes can generate significant differences between the numbers of molecules inherited by each daughter cell in division even when the copy number of the partitioning molecules before division is high. For example, due to the imperfect positioning of the FtsZ ring at mid-cell in division (Kerr et al., 2006), the size of daughter cells may not be equal (Gupta et al., 2014; Männik et al., 2012), resulting in the difference in the volumes accessible to each partitioning molecule. These molecules therefore segregate randomly and biasedly to the bigger daughter cell with the probability p>½. The level of errors in partitioning is therefore given by (Huh and Paulsson, 2011b):

$$Q_x^{\,2} = \frac{1 + Q_{vol}^{\,2}(CV_x^{\,2}\langle x\rangle + \langle x\rangle - 1)}{\langle x\rangle} \tag{2.3}$$

Here, $Q_{vol}^{\,2}$ is the level of errors in the partitioning of accessible volume. From (2.2) and provided that $\langle x\rangle \geq 1$, $Q_x^{\,2}$ is always greater than the level of errors achievable with the random partitioning scheme ($1/\langle x\rangle$).

The ordered schemes allow the variance of the molecules inherited by each daughter cell to be below that of binomial distribution. This can be achieved by reducing the number of molecules able to segregate stochastically at division (e.g. using spindles or pair formations to pre-allocate an equal number of molecules to segregate to each daughter cell). Other possibility is to enhance uniform spatial distribution of the partitioning molecules before cell division (e.g. the partitioning molecules are big enough to exclude themselves). In general, the ordered segregation schemes would require extreme parameters to greatly suppress the partitioning errors (Huh and Paulsson, 2011b). However, the phenomenon is not necessarily absent in nature. One prominent example is the highly ordered partitioning of *E. coli* chromosomes in cell division. In (Sheahan et al., 2004), it was reported that *Nicotiana tabacum* uses actin filaments to segregate the chloroplasts in mitosis more evenly than the independent segregation scheme.

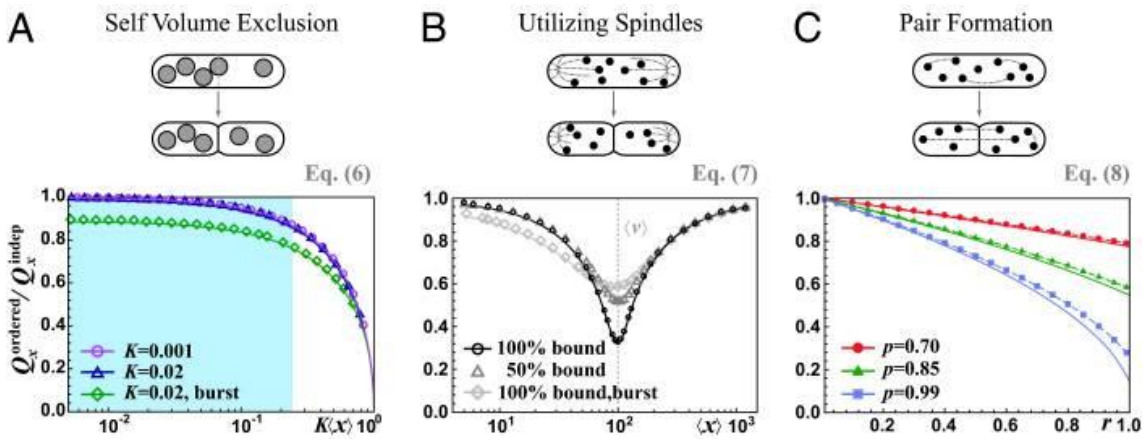


Figure 15. Examples of ordered segregation schemes. (A) Self-volume exclusion: Segregating units with non-negligible size (gray circles) exclude each other and thereby promote more even segregation. The level of errors in partitioning (Q_x^{ordered}) can be further reduced with increasing unit's size ($\sim K\langle x\rangle$). (B) Utilizing spindles: Organelles (dots) compete for available binding sites (ends of astral, gray) and unbound organelles are partitioned independently. (C) Pair formation: The segregating units form, with certain probability *r*, pairs, which at division push each other away to the two daughter cells. Q_x^{ordered} is reduced by increasing the pairing probability *r*. Adapted from (Huh and Paulsson, 2011b).

The variation in the number of the partitioning molecules caused by noisy partitioning is added only at the moment cell division occurs. Therefore, one expects that the effects of stochastic partitioning on the dynamics of single gene expression would be qualitatively different from the effects of noise in gene expression, as the latter is generated continu-

ously and independently throughout the cell cycle. In **Publication II**, by observing the dynamics of two genetic circuits, the Toggle Switch and the Repressilator, with varying levels of errors in partitioning, the effects of errors in partitioning on the network level are investigated.

3 Methods of study

This chapter provides an overview on the methods which have been employed to characterize the dynamics of gene expression in *E. coli* and, based the observed kinetics, to study the dynamics of single gene expression and of small genetic circuits in different scenarios.

3.1 Measurements of gene expression

The constant development of new microscopy techniques has allowed us to peek into the microscopic world with greater and greater details. From the classic bright-field microscope with the resolution of ~0.2 μm, nowadays, an electron microscope can generate images with the resolution as high as ~0.1 nm, which is smaller than the size of a typical protein molecule (Erickson, 2009). Nevertheless, most intracellular processes are not readily visualized under the microscope due to the presence of the cell wall and the myriads of molecules floating behind it. For such direct observation, it is required that the object of study is either isolated (Buc and McClure, 1985; Miller et al., 1970), i.e. *in vitro*, or modified (e.g. using fluorescent fusion proteins (Elf et al., 2007; Hammar et al., 2012; Romantsov et al., 2010)), which may cause the target system's dynamics to deviate from that of the native system.

The discovery of the *lac* operon and its regulation by lactose (Jacob and Monod, 1961) offered alternative possibilities to characterize several cellular processes by observing the expression dynamics of a dependent gene. For example, in (Hansen et al., 1998; Jensen et al., 1993), from measurements of the expression of *lacZ* gene (using Beta-galactosidase assays) with varying IPTG concentrations, the intake process of lactose in

E. coli was revealed to be with positive feedback in regime of low lactose concentrations. In (Farewell et al., 1998), it was observed that the presence of the stress responding σ factor, σ^{38}, in the stationary growth phase down-regulates the expression of many genes that are expressed in the exponential growth phase. This signifies the competition between σ^{38} and the house keeping sigma factor (σ^{70}) for RNAP core enzymes.

Here, we present measurement methods that have been employed to study the dynamics of gene expression. Also discussed are the scenarios where their application is proper.

3.1.1 Relative quantification methods

Relative quantification methods (also referred as assay methods) allow us to quantify the total amount of a certain target entity (the analyte) in a cell population. The target measurement results are usually shown in regard to certain reference entities that are relatively invariant to environmental conditions so that one can obtain the average level of analyte in each cell. The quantification process essentially involves the identification of the analyte from other entities by the use of exogenous analyte-specific reagents (Alberts et al., 1994). The information on the analyte level is then amplified so that it can be distinguished and extracted from the noise background (Livak and Schmittgen, 2001).

Several types of entities in cells, such as DNA, RNA and proteins, can be quantified using different assay methods. However, the information on their levels is shown only as averaged over the cell population. Thus, the assay results do not disclose the cell-to-cell diversity in this amount in the population. Further, since the host cells are killed during the assay process, studying the time evolution of such quantities is challenging.

3.1.2 Fluorescent proteins

Fluorescent proteins, due to their ability to emit photons under proper excitation, can be used as a reporter of gene expression. The first fluorescent protein discovered is the green fluorescent protein (GFP), which was initially isolated from a species of illuminating jellyfish and then modified to become functional when expressed in other organisms (Shanner et al., 2005).

To study the dynamics of a target promoter, one can insert into cells a genetic construct containing the gene encoding for fluorescent proteins under the control of that target promoter (Megerle et al., 2008; Ozbudak et al., 2004; Rosenfeld et al., 2005). These

cells are then imaged with confocal or epifluorescence microscopes. From the brightness of each cell in the fluorescence images, one can extract information on the dynamics of the fluorescence protein numbers and consequently the target promoter's dynamics in each cell (Figure 16).

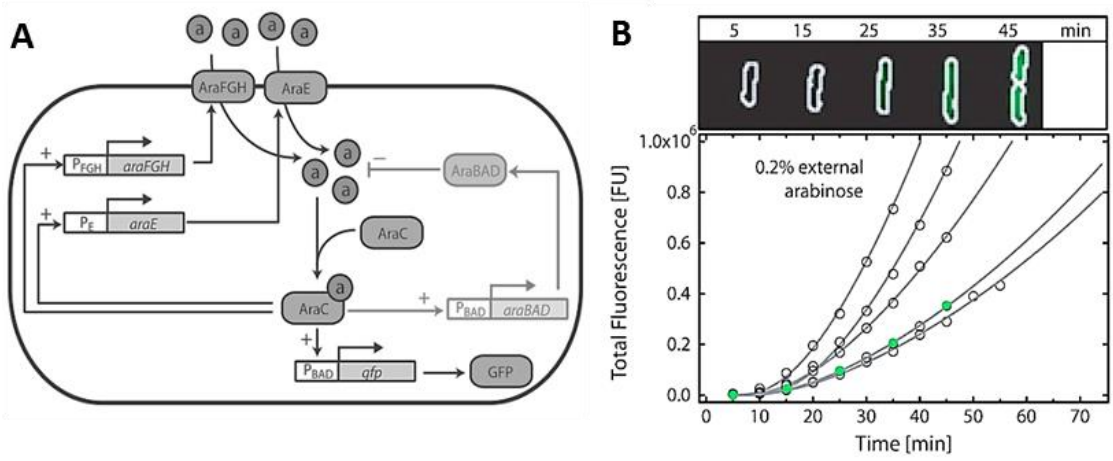


Figure 16. Measurement of P_{BAD} activity upon the introduction of arabinose (a) in the media. (A) Diagram of the arabinose intake system (P_{FGH}-*ara*FGH, P_E-*ara*E) and the reporter system (P_{BAD}-gfp). The arabinose utilization system (P_{BAD}-*ara*BAD blurred) has been deleted in cells. (B) Example of single cell induction kinetics of P_{BAD} promoter with external media of 0.2% arabinose. Traces (open circles) of the GFP total intensity in each cell are shown at 5 minute intervals. The image panel corresponds to the fluorescence traces marked with green circles. Adapted from (Megerle et al., 2008).

Unlike assay methods, the use of fluorescent proteins allows the observation of the cell-to-cell diversity in gene expression and of the gene expression dynamics over time (Elowitz et al., 2002). However, the long protein maturation time (from 5 to 15 minutes (Shanner et al., 2005)) causes a noisy delay in the observation process. Added to this, due to the low fluorescent intensity and the fast diffusion speed of individual fluorescent molecules (Nenninger et al., 2010), a large amount of fluorescent proteins must be produced to be detectable under the microscope. Therefore, the reporter system is usually multi-copied (e.g. the genetic construct is placed on a high copy plasmid), making the observation of single gene expression dynamics unfeasible.

3.1.3 Fluorescence *in situ* hybridization (FISH)

FISH is a cytogenetic technique that uses multiple exogenous fluorescent probes that bind specifically to parts of the DNA or RNA with a high degree of sequence complementarity (Langer-Safer et al., 1982). As the probes co-localize at the target RNAs or

DNA segments, the clusters become visible under fluorescence microscope as bright spots, allowing one to study their abundance, localization and structure (Figure 17). This method has been further developed to be able to quantify the RNA level with single molecule sensitivity (Raj et al., 2006; Skinner et al., 2013). From the distribution of the RNA numbers in each cell in a population, one can infer on the dynamics of transcription of the target gene (So et al., 2011).

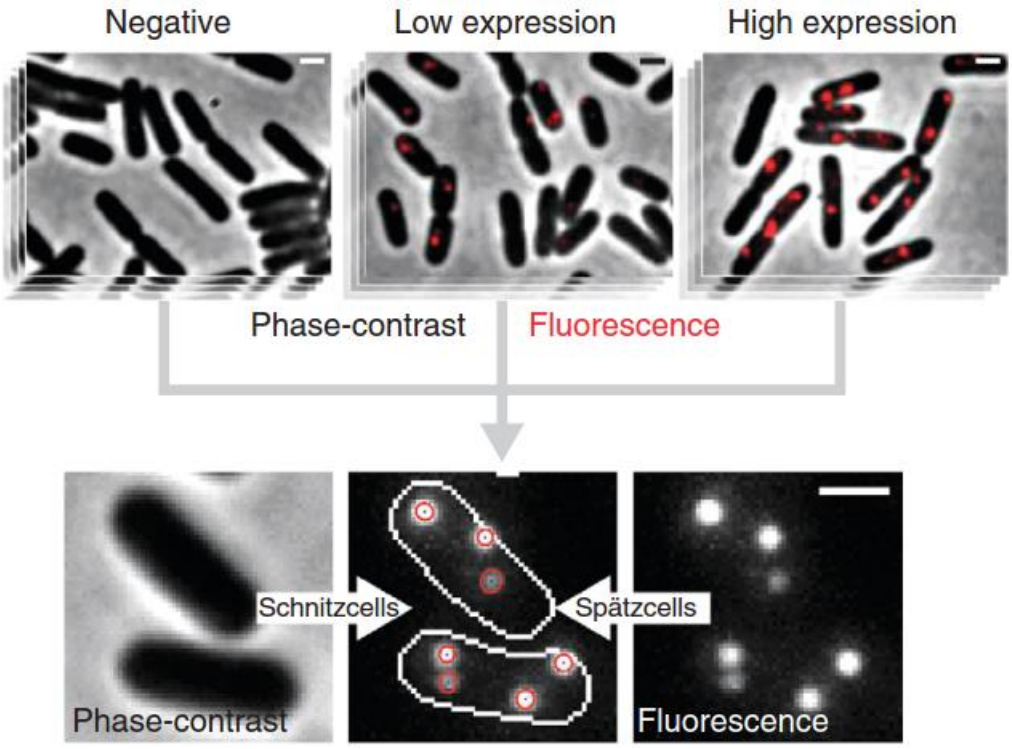


Figure 17. Example of images acquired with the FISH technique. Both phase contrast and fluorescence images of the fixed cells after hybridization are captured for cell segmentation and spot segmentation, respectively. The two sets of images are then aligned to determine which cell each spot belongs to. From the number and intensity of the spots in each cell, further information on the RNA number, localizations can be extracted. Adapted from (Skinner et al., 2013).

Despite the sensitivity and the ease of experiment preparation (e.g. probes can be designed based on the target gene sequence, no genetic modification on the host cells is required), the FISH method requires that host cells are fixed in the hybridization process (the binding of the probes to the RNA segment). Therefore, with this method, it is not possible to observe the transcription dynamics in individual cells in real-time.

3.1.4 RNA-tagging system using MS2-GFP fusion proteins

The RNA-tagging system using MS2-GFP (Golding et al., 2005; Le et al., 2005), like FISH, makes use of fluorescent probing molecules that bind specifically to the target RNA. Here, the probe is a fusion between a green fluorescent protein variant (GFPmut3) and the coating protein of bacteriophage MS2, capable of binding to the specific sites ("hairpin") of the phage DNA (Valegård et al., 1990). The target RNA is a tandem array of 96 MS2 binding sites, separated from one another by randomized sequences to increase the RNA's stability.

The measurement system contains two constructs (Figure 18). The first construct is placed on a single copy plasmid, containing the gene encoding for the RNA containing the 96 MS2 binding sites under the control of the target promoter. For the cross validation of the transcription activity, another tractable gene (e.g. lacZα (Golding and Cox, 2004), mRFP1 (Muthukrishnan et al., 2012)) can also be inserted behind the promoter of interest. The second construct, placed on a multiple copy plasmid, contains the gene encoding MS2-GFP proteins under the control of a tunable promoter.

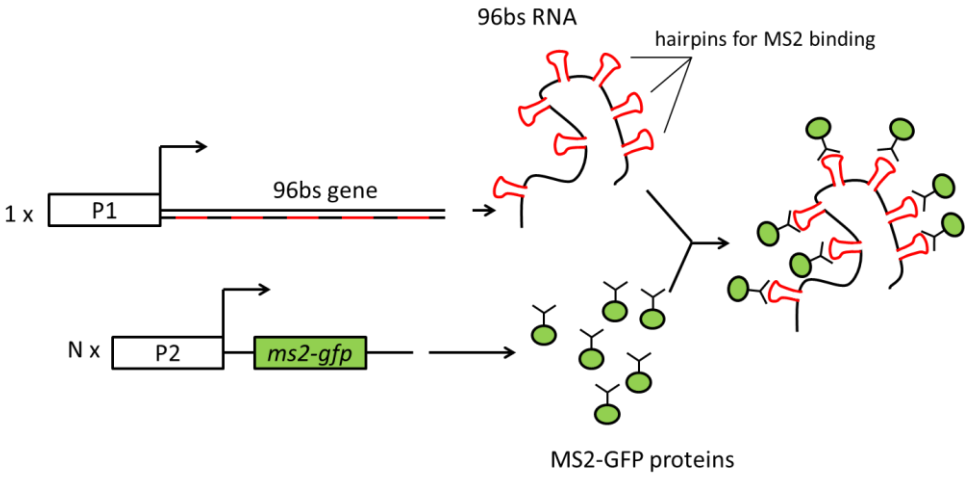


Figure 18. Diagram of the MS2-GFP RNA tagging system. The gene encoding for RNA with 96 binding sites (red) is placed under the control of the single copied target promoter (P1). The reporter MS2-GFP proteins (green circles with hooks) are produced under the control of the multi-copied promoter (P2). When a new target RNA is produced, its binding sites are rapidly bounded by MS2-GFP proteins and become foci of GFP, which are visible above the cell background under the fluorescence microscope.

To measure the transcription dynamics of the target promoter, before activating the target gene, one has to prepare the reporter system by overexpressing MS2-GFP proteins

in cells. The amount of MS2-GFP should be high enough to be able to bind to and report the appearance of novel target RNAs as they are transcribed. After the reporter incubation time, the production of the target RNA can then be triggered and observed under the confocal microscope. As they are transcribed, the binding sites on the target RNA are bound by and become foci of MS2-GFP. These foci, due to the large number of bound MS2-GFP, appear as bright spots under the confocal microscope.

Fluorescence images are captured in time series and processed with cell and spot-segmentations (Chowdhury et al., 2012; Häkkinen et al., 2013). As the target RNAs are wrapped by MS2-GFP proteins and therefore protected from degradation mechanisms (Golding and Cox, 2004), the spots do not lose their intensity over time. Therefore, the total spot intensity over time can be fitted by a monotonically increasing piecewise-constant function, where each step corresponds to the production of a single target RNA (Figure 19). With the function fitted, the moments of RNA appearances in each cell can then be extracted.

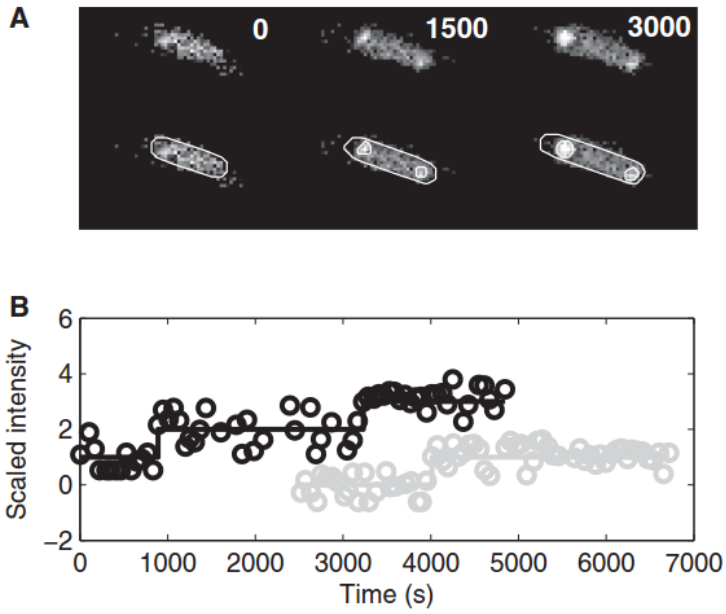


Figure 19. Tagged RNA in *E. coli* DH5α-PRO cells. (A) Examples of unprocessed frames and segmented cells and RNA spots. The moments of image capture are shown for each frame. (B) A time series of scaled spot intensity levels from individual cells (circles) and the corresponding estimated RNA numbers (solid lines). Reused from (Muthukrishnan et al., 2012).

The RNA-tagging system using MS2-GFP allows a lag-free observation of the transcription dynamics at the single cell single event level in real time. This system has

been employed to parameterize the kinetics of transcription initiation of many bacterial promoters in the strain DH5α-PRO (Kandhavelu et al., 2012a, 2012b; Muthukrishnan et al., 2012).

3.2 Stochastic modeling

3.2.1 Goals of modeling

Due to the limitation of measurement techniques, we are not able to observe every biological process directly and in its native form. For example, the interaction between lac repressor and DNA is only visualized with the use of a lac repressor – YFP (venus) fusion protein (Elf et al., 2007; Hammar et al., 2012). Another example is the competition between σ factors for RNAP core enzyme, which is evidenced only through the tradeoff between the expression of genes recognized by each σ factor (Chang et al., 2002; Farewell et al., 1998; Rahman et al., 2006).

To overcome these problems, we employ models, which are simplified representations of biological processes but can still capture the processes' key behaviors. The model may serve two main purposes. The first purpose is validation, i.e. solving the inverse problem, which involves the finding of the model that is likely to produce the observable data. The second purpose is prediction, i.e. solving the forward problem: by apply the model to different contexts, we can also assess how the kinetics of the process would play a role in the dynamics of more complex systems, which are not readily observable or have not been engineered.

3.2.2 Representation of biological processes in *E. coli*

Due to the random movement of both reactive and non-reactive molecules in the cytoplasm (Alberts et al., 1994) and that reactions between multiple reactants are discrete events that occur only when the molecules collide, biochemical processes in cells are inherently stochastic, resulting in stochastic dynamics of substances' copy numbers over time.

Originally, by treating biochemical systems as solvent systems where all the interacting species exist in large copy numbers, one employs the deterministic approach to model biochemical reactions in cells. In this approach, the dynamics of the system is described

by a set of Ordinary Differential Equations (ODE), which deterministically characterizes the change in the quantity of each reactant at any given time. By solving the ODE, we are able to study the time evolution of each species' quantity over time. The deterministic model may give us some information on the dynamics' trends of the molecule numbers in the system, and is particularly useful in population studies.

However, many processes in cells involve species that exist only in small copy number (e.g. genes, RNAs). The role of intrinsic noise (noise arises from the random movement of reactive and non-reactive molecules in cells) on the dynamics of molecule numbers over time therefore cannot be ignored. The use of ODE is no longer appropriate to capture the system dynamics in individual cells (Gillespie, 1992). In order to capture the stochastic nature of chemical reactions in cells, we study the kinetics of molecules in a solution based on the following assumptions: (i) The system is well-stirred and of constant volume V. (ii) The system is in thermal equilibrium at constant temperature T. (iii) Reactions occur only when two or more molecules collide, while most collisions do not lead to reactions.

The first assumption requires that the spatial distribution of one species' molecules is uniform within the cell volume V. Also, the positions of all molecules are independent of one another, no matter whether they are of the same species or not. The second assumption requires that every molecule in the solution has independent, normally distributed velocity, with mean equal k_BT/m, in which k_B is the Boltzmann constant (Kittel and Kroemer, 1980), T is the solution's absolute temperature and m is the molecular weight. The third assumption requires that the reactant molecules' number is low enough to ensure that the system follows Maxwell-Boltzmann statistics (Gillespie, 1976; Kittel and Kroemer, 1980). With these assumptions, we are able to ignore individual molecules' position and velocity, and adopt a probabilistic approach on the evolution of the system's state over time. The general reaction has the stochastic form written as follow:

$$\sum_{i=1}^{N} s_{ij} X_i \xrightarrow{c_j} \sum_{i=1}^{N} r_{ij} X_i \tag{3.1}$$

where X_i is the i^{th} reactant out of N reactants in the system. s_{ij} and r_{ij} indicate how many molecules of substance X_i are consumed/produced via the j^{th} reaction. The reaction constant c_j is "reaction probability per time unit" indicating how likely the reac-

tion j^{th} out of M reactions is to happen given the current reactants' molecule. Its value can be calculated from the reactant collision rate, Maxwell's velocity, and the probability for reactions to occur upon collisions. We define the propensity function as:

$$a_j(x) = c_j.h(x) \tag{3.2}$$

$a_j(x).dt$ is the probability for the j^{th} reaction to occur in the infinitesimal time window $[t, t+dt)$. $h(x)$ is the number of possible reactant combinations of a reaction at a given time, with x the reactants' molecule number vector.

Because the propensity function at a specific time depends only on the current state rather than previous ones, we can consider the system dynamics as a Markov process, where each reaction marks a change in state. Therefore, one can model the system's dynamics as probabilistic switching between the discrete states in its state space using Chemical Master Equation (Gillespie, 1977) and find the time-evolution of system's probability in any given state over time.

The analytical solution to Chemical Master Equation (CME) is usually intractable and computational challenging to obtain especially for complex systems involving either a large number of substances or fast and reversible reactions. Added to these, the generation of the probability densities is calculated on the continuous time scale, which is vulnerable to error when performed on a digital computer. To address these problems, the numerical Stochastic Simulation Algorithm (SSA) (Gillespie, 1977, 1976) is employed.

3.2.3 Stochastic Simulation Algorithm

The Stochastic Simulation Algorithm (SSA) numerically simulates the underlying Markov process described by the CME using random sampling of the reaction times. A single simulation of the system over time yields explicitly a single possible sequence of reactions, corresponding to a single trajectory of the system's state space with the appropriate probability density. Acquiring a single trajectory is generally simple and generally inexpensive with current digital computers. However, to characterize the system dynamics, one is usually interested in the probability density described by the CME. Thus, multiple realizations are generated and analyzed together to approximate the real probability density. This can be both time and data storage consuming. Nevertheless, employing SSA with a finite number of realizations is still a preferred approach as it

allows the user to pre-allocate resource consumption to achieve a certain level of approximation of CME's analytical results.

Another reason that supports the preference of SSA is that, by studying the dynamics of interacting substances in single trajectories, one can observe the trends and the correlations of substances' quantity unobtainable by CME. For example, robust oscillation with fluctuation in its period, when averaged, becomes damped oscillation. The information on the mean and noise of the oscillating period is therefore not observable by the CME's solution. However, one can extract these properties on individual trajectories acquired with SSA.

Unlike CME which tracts the time evolution of the probability density at fixed rates, SSA realizes the time evolution of substance number with a rate chosen from a set of rates randomly with a certain probability mass function. The function, which describes the probability that the next reaction in the system is the j^{th} reaction and will occur in an infinitesimal time interval $[t,t+dt)$, is dependent of the temporal system's state x at time t. We denote $a(x)$ the sum of all reactions' propensity functions as follow:

$$a(x)=\sum_{j=1}^{M} a_j(x) \tag{3.3}$$

Given $a(x)$ the probability for any reaction to occur in the infinitesimal time window $[t,t+dt)$, the time for the first reaction to occur in the system has the following distribution:

$$\tau=-a(x)^{-1}\ln(1-r_1) \tag{3.4}$$

Here, r_1 is a random variable following a uniform distribution $U[0,1)$. As all the reactions compete with each other to be first one to occur given the system state x.

The probability that the next reaction to occur is the j^{th} reaction is proportional to its propensity function $a_j(x)$. Therefore, the next reaction to occur is drawn by:

$$j=j' \text{ such that } \sum_{i=1}^{j'-1} a_i(x)\le r_2.a(x)<\sum_{i=1}^{j'} a_i(x) \tag{3.5}$$

where r_2 is a random variable following a uniform distribution $U[0,1)$.

From (3.4) and (3.5), we have the outline to implement the SSA: From any given state x at time *t*, we calculate τ and draw *j* and update the system with the new state $x+S_j$ at time $t+\tau$. S_j is the stoichiometric vector indicating the changes in molecule numbers after one reaction j^{th} occurs. Therefore, from the initial state x_0 at time *t*, we can generate one trajectory describing the time evolution of substances' molecule numbers. The algorithm is executed in the following steps:

1. Initialize the step $n=0$, time: $t_n=t_o$ and state $x_n=x_o$

2. Calculate $a_j(x_n)$ and $a(x_n)$ given the current state x_n

3. Generate two random values r_1, r_2 from the uniform distribution $U[0,1)$

4. Calculate τ and *j* from equations (3.4) and (3.5)

5. Perform the j^{th} reaction: Update $t_{n+1}=t_n+\tau$ and $x_{n+1}=x_n+S_j$

6. Set $n=n+1$. Return to step 2.

It should be noted that, during the iteration of the algorithm, all values of (t_n, x_n) are recorded, giving a complete realization of the system's state with the probability described by the CME. The trajectory is usually resampled at fixed time intervals to make the analysis of multiple trajectories more convenient.

3.2.4 Delayed Stochastic Simulation Algorithm

Biological processes such as transcription initiation, transcription elongation, translation, etc., generally involve several sub-processes. To model these processes properly, one usually has to break them into or approximate them with sequences of elementary reactions. For example, the complex transformation of species *A* to species *B* can be described by reaction (3.6):

$$A \xrightarrow{c_0} I_1 \xrightarrow{c_1} I_2 \xrightarrow{c_0} \ldots \xrightarrow{c_{n-1}} I_n \xrightarrow{c_n} B \qquad (3.6)$$

Such approximation, however, requires the identification of the rate of each step, making the number of free parameters in the model to increase drastically. Further, processes with little noise require a large number of steps to approximate. This increase in

number of reactions and reactants makes the simulations computationally and storage expensive. As the intermediary products I_1 to I_n do not play any other role in the system, reaction (3.6) can be shortened to the delay form:

$$A \xrightarrow{c_0} B(\tau_B) \tag{3.7}$$

The delay term τ_B represents the time elapsed of the transformation from I_1 to *B*, that is, the molecule *B* is released at τ_B after the reaction consuming A. τ_B can have arbitrary probability densities, predefined by user rather than by convolving the probability densities of elementary reaction times. The system is therefore semi-Markovian: the probability of the system being in the next state within the infinitesimal time window depends not only on the current state but also on the previous ones.

To numerically simulate the system with delayed product releases, we employ the Delayed Stochastic Simulation Algorithm (DSSA) (Bratsun et al., 2005; Roussel and Zhu, 2006). The DSSA, extended from the SSA, is employed. The algorithm is first developed by (Bratsun et al., 2005) and generalized by (Roussel and Zhu, 2006) to account for multiple delayed products. The outline of the DSSA is as follow:

1. Initialize the step $n = 0$, time: $t_n = t_o$ and state $x_n = x_o$

2. Calculate $a_j(x_n)$ and $a(x_n)$ given the current state x_n

3. Generate two random values r_1, r_2 from the uniform distribution $U[0,1)$

4. Calculate τ and j from equations (3.4) and (3.5)

5. If there are any delayed products released from the time interval $[t_n, t_n + \tau)$:

   (a) Release the delayed product $x_{(i)}$ with the closest release time t': Update $t_n = t'$ and $x_{n+1(i)} = x_{n(i)} + 1$

   (b) Return to step 2.

6. Perform the j^{th} reaction: Update $t_{n+1} = t_n + \tau$ and $x_{n+1} = x_n + S_j$. If the reaction has delayed products $x_{(i)}$ with the release times $\tau_{(i)}$:

(a) Generate random values $\tau'_{(i)}$ following the distributions of $\tau_{(i)}$

(b) Put the delayed products $x_{(i)}$ in the release queue with the release times $t' = t_n + \tau'_{(i)}$

7. Set $n = n + 1$. Return to step 2.

In this thesis, the stochastic simulations are performed with Stochastic Genetic Network Simulator (SGNSim) (Ribeiro and Lloyd-Price, 2007) and its updated version Stochastic Genetic Network Simulator 2 (SGNS2) (Lloyd-Price et al., 2012a). SGNSim, which was used in **Publication I** and **Publication III**, is a tool to model stochastic biochemical systems whose dynamics is driven by the DSSA. SGNS2, employed in **Publication II** and **Publication IV**, is also based on the DSSA and allows the modeling of compartmentalized systems, in which the compartments can be created, destroyed and divided at runtime.

3.2.5 Connecting stochastic systems of different time scales

With the stochastic approach, though offering more information on the system dynamics (e.g. fluctuations), an appropriate modeling strategy is required to gain results of relevance without the explosion of the parameter space and the computation time.

In practice, biological systems often take place on vastly different time scales, with ‘‘fast’’ reactions having much greater propensity rate function than ‘‘slow’’ ones. As fast reactions occur with must greater frequency than slow ones, an exact stochastic simulation of such a system will necessarily spend most of its time simulating the more frequent occurrences of fast reactions (Cao et al., 2005a, 2005b).

If fast reactions and slow reactions do not share any common species, the system can be divided into non-interdependent systems, the “fast” system and the “slow” system, the dynamics of which can be concluded through separate simulations differing in approaches, simulation times, numbers of trajectories, etc.

However, when fast and slow reactions involve some common species, separate simulations are not possible. This is frustrating when knowing that a single fast reaction event will be of much less importance to the system’s dynamics than a slow reaction event. To avoid the waste of computational efforts, in (Cao et al., 2005b), the use of the “virtu-

al fast system" was proposed. This system contains species connected to one another only by fast reactions, while assuming that all the slow reactions are switched off. We can impose the solution of the virtual fast system (either a value or a probability density) directly on the number of the common species and with this, study the dynamics of the slow system (which involves only slow reactions) alone. It should be noted that this approximation is only valid when the virtual fast system is stable and when its relaxation time is much less than the expected time for the next slow reaction.

To demonstrate this, we have a model of transcription of a single-copy gene, regulated by some repressor molecules (McClure, 1985; Ribeiro et al., 2006):

$$R + Pr \underset{k_{unbind}}{\overset{k_{bind}}{\rightleftarrows}} PrR \tag{3.8}$$

$$Pr \xrightarrow{k_c} Pr_c \tag{3.9}$$

$$Pr_c \xrightarrow{k_o} Pr_o \xrightarrow{\infty} Pr + M \tag{3.10}$$

Reaction (3.8) models the reversible binding between a repressor (*R*) and a promoter (*Pr*) with the binding rate k_{bind} and unbinding rate k_{unbind} to form the promoter repressor complex (*PrR*). When unbound by repressors, the free promoter can be bound by an RNAP holoenzyme to form the promoter closed complex (Reaction (3.9)) with the rate constant k_c. Here, the number of RNAP holoenzymes is assumed to be very high and thus omitted in the value of k_c. The formation of the closed complex is followed by the formation of the open complex Pr_o, with the rate constant k_o. This is followed by very fast promoter escape, transcription elongation and finally the release of a novel mRNA (*M*) and the free promoter.

Here, we assume that k_{bind} and k_{unbind} are much greater than k_c and k_o. Thus, reaction (3.8) is considered "fast", and reaction (3.9), (3.10) are considered "slow". Here, it is noticeable that when using SSA to simulate the system, most of the reaction occurrences are either the binding or the unbinding of repressors to the promoter region. The firings of reaction (3.9), which greatly determine the dynamics of RNA production, occur with much smaller frequency.

From the system's reactions, it can be seen that the virtual fast system involves *R*, *Pr* and *PrR* interacting with one another through reaction (3.8). The solution of the system when $t \to \infty$ is:

$$\mathrm{P}(Pr=1)=\frac{K}{R+K};\ \mathrm{P}(Pr=0)=\frac{R}{R+K} \tag{3.11}$$

$$\mathrm{P}(PrR=1)=\frac{R}{R+K};\ \mathrm{P}(PrR=0)=\frac{K}{R+K} \tag{3.12}$$

in which K, also called the dissociation constant, is the ratio between k_{unbind} and k_{bind}. By applying this solution of P(Pr) to reaction (3.8), the original system can then be approximated by:

$$Pr \xrightarrow{\frac{K\times k_c}{R+K}} Pr_c \tag{3.13}$$

$$Pr_c \xrightarrow{k_o} Pr_o \xrightarrow{\infty} Pr + M \tag{3.14}$$

This new approximating system contains essentially two reactions on the low time scale and thus can be simulated with much less computational efforts. However, the approximation on the dynamics of RNA production will become worse as the time scale between the fast and slow systems converge (Figure 20).

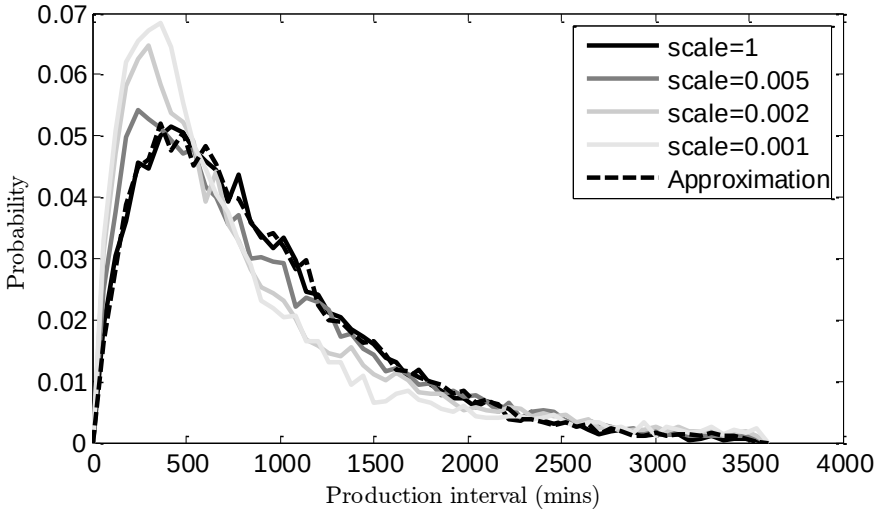


Figure 20. Distributions of RNA production intervals with different models of transcription. The input parameters are: k_{unbind}=1×scale (s^{-1}), k_{bind}=0.1×scale (s^{-1}), R(0)=10, k_c=1/300 (s^{-1}), k_o=1/300 (s^{-1}). Here, the parameter "scale" is tuned to vary the difference in the time scale between fast and slow reactions. The distributions are shown along the approximated distributions (dash) with the dissociation constant unvaried with "scale".

The approximation of multi-timescale systems is employed in all the studies of the thesis, as it allows effective modeling of the transcription factor-promoter interactions (all

publications), the intake process of inducer in cells (**Publication IV**) and the σ factor competition for RNAP core enzymes (**Study V**).

3.2.6 A detailed model of gene expression

This section presents a detailed stochastic model of single gene expression in *E. coli*. It is conceived based on previous studies on the kinetics of transcription (Kandhavelu et al., 2012b; Muthukrishnan et al., 2012; Saecker et al., 2011), translation (Yu et al., 2006) and degradations (Chen et al., 2015; Taniguchi et al., 2010). The model reactions, along with their descriptions, are presented in Table 1. Also shown are some of the previous works where the information on the reactions' rate constants can be found.

Table 1. A stochastic model of single gene expression.

Reaction	Description	References
$R + Pr \underset{k_{unbind}}{\overset{k_{bind}}{\rightleftharpoons}} PrR$	Binding / unbinding between repressors (R) and promoter (Pr) to form the repressed promoter complex.	(Elf et al., 2007) (Rosenfeld et al., 2005) (Hammar et al., 2012)
$RNAP + Pr \xrightarrow{k_c} Pr_c$	Formation of promoter closed complex (Pr_c) with the binding of RNA polymerase (*RNAP*)	(Buc and McClure, 1985) (Lutz et al., 2001) (Muthukrishnan et al., 2012)
$Pr_c \xrightarrow{k_o} Pr_o$	Formation of promoter open complex (Pr_c)	(Buc and McClure, 1985) (Lutz et al., 2001) (Muthukrishnan et al., 2012)
$Pr_o \xrightarrow{\infty} Pr + M + RNAP$	Promoter escape, transcription elongation (very fast), followed by the release of a new mRNA (*M*), and *RNAP*	(Hsu, 2002) (Ryals et al., 1982) (Vogel and Jensen, 1994)
$M \xrightarrow{k_P} M + P$	Translation from mRNA to protein (*P*)	(Yu et al., 2006) (Taniguchi et al., 2010)
$M \xrightarrow{d_M} \phi$	Degradation of mRNA	(Bernstein et al., 2002) (Taniguchi et al., 2010) (Chen et al., 2015)
$P \xrightarrow{d_P} \phi$	Degradation of protein	(Taniguchi et al., 2010) (Shanner et al., 2005)

It should be noted that, as σ factors exist in large copy numbers in cells, the release of a single σ factor molecule from the active RNAP during the transcription elongation (Mooney et al., 2005; Raffaelle et al., 2005) of this gene should not affect the σ factor population significantly and thus is not accounted in this model.

From the model of single gene expression, one can predict the dynamics of gene expression, in both mean and noise (Pedraza and Paulsson, 2008), with different numbers of repressors [*R*] and RNA polymerases [*RNAP*] analytically as follow:

Table 2. Properties of gene expression dynamics as a function of the numbers of repressors [*R*] and RNA polymerases [*RNAP*].

Description	Mean (μ)	Coefficient of Variation Squared (CVS=σ^2/μ^2)
Regulation function (f)	$k_{unbind}k_{bind}^{-1}([R]+k_{unbind}k_{bind}^{-1})^{-1}$	0
Duration of closed complex formation (τ_{cc})	$f^{-1}k_{cc}{}^{-1}[RNAP]^{-1}$	1
Duration of open complex formation (τ_{oc})	$k_{oc}{}^{-1}$	1
Interval between consecutive RNA productions ($\Delta t=\tau_{cc}+\tau_{oc}$)	$\mu_{\tau_{cc}}+\mu_{\tau_{oc}}$	$\frac{(\mu_{\tau_{cc}})^2+(\mu_{\tau_{oc}})^2}{(\mu_{\tau_{cc}}+\mu_{\tau_{oc}})^2}$
RNA number (M)	$(\mu_{\Delta t})^{-1}d_M^{-1}$	$\sim\frac{CVS_{\Delta t}+1}{2}$
Protein number (P)	$\mu_M k_P d_P^{-1}$	$\sim\frac{1}{\mu_P}+\frac{CVS_M}{\mu_M}\times\frac{d_M^{-1}}{d_M^{-1}+d_P^{-1}}$

In each study of the thesis, a different simplified variant of this model was employed depending on the context of the study.

4 Summary of the Results

This thesis focused on the long-term dynamics of genetic circuits and on the cellular mechanisms due to which such dynamics becomes environment dependent. Two mechanisms were investigated: the intake of gene expression inducers from the media and the σ factor competition for RNAP. The first mechanism affects only the expression dynamics of a specific gene, while the second mechanism is expected to have global impacts on gene expression. As the time scale of the dynamics of genetic circuits can span several cell cycles (Elowitz and Leibler, 2000; Stricker et al., 2008; Zeng et al., 2010), this dynamics was also observed in the context of a dividing cell population, where errors in the partitioning of regulatory molecules in cell division occur.

The five studies contributed to the thesis' goal as follows: **Publication I** established a framework to study the dynamics of genetic circuits as a function of the transcription kinetics of the constituent genes. Here, the impact of the choice of promoters, the host cells and the environmental conditions are shown explicitly in the kinetics of transcription. **Publication II** extended this framework by considering the effects of cell growth and cell division, where the partitioning of regulatory molecules is inherently stochastic. **Publication III** and **Publication IV** featured the intake kinetics of a gene expression inducer and its effects on the transcription dynamics of a gene. **Studies V** focused on the σ factor competition and the degree to which promoters with different transcription initiation kinetics are affected by this phenomenon.

In **Publication I**, the dynamics of genetic circuits with varying transcription initiation kinetics was studied. Two genetic circuits are considered: a biphasic amplitude filter and a frequency filter. The input signals here are deterministic and are set constant in the time domain (for the biphasic amplitude filter) and in the frequency domain (for the

frequency filter). According to recent studies, the transcription initiation process in *E. coli*, controlled mostly at the promoter region, is well-modeled by a sequence of multiple rate-limiting steps, the durations of which are exponentially distributed (Kandhavelu et al., 2012a, 2012b, 2011; Muthukrishnan et al., 2012). This results in a distribution of the RNA production intervals having a coefficient of variation (σ/μ) less than 1. Therefore, we investigated whether the multi-step nature of transcription initiation could be utilized to compensate for the noise in filters' response to the input signals.

Using stochastic simulations driven by the multi-scale SSA (Cao et al., 2005a, 2005b; Gillespie, 1976), we found that, for realistic parameter values, the level of noise in the output of genetic filter motifs increases with increasing durations of rate-limiting steps in transcription initiation. As RNAs are produced with slower rates, the filters suffer from low-copy noise and thus have degraded performance. The effects of low copy number noise can be partially compensated by employing a less noisy transcription process (i.e. having more rate-limiting steps in transcription initiation).

It should be noted that the numbers and durations of rate-limiting steps in transcription initiation can vary not only between promoters (Lutz et al., 2001) but also between host cells and between different environmental conditions (Dong et al., 2011; Farewell et al., 1998; Kandhavelu et al., 2012b; Muthukrishnan et al., 2014). Thus, these results will aid in the exploration of the range of behaviors of genetic motifs as a function of the kinetics of transcription initiation and in tuning of synthetic motifs to attain specific characteristics without affecting their protein products.

In **Publication II**, we studied the behavior across cell generations of two genetic circuits, namely the Toggle Switch and the Repressilator, when subject to the random partitioning of their regulatory molecules in cell division. Several schemes of partitioning, both ordered and disordered (Huh and Paulsson, 2011b), have been employed to vary the level of errors in partitioning. The results from stochastic simulations suggest that genetic circuits are not impervious to this source of cell-to-cell variability, although the level of impact is heavily network-dependent.

In the case of the Toggle Switch, it was found that increasing the level of partitioning errors, while increasing the switching rate between the noisy attractors of the Toggle Switch, decreases the variance of the phenotypic distribution of the population to levels lower than that of a binomial distribution. This is because of the anti-correlation between the numbers of molecules inherited by each daughter cell: with high-variance

partitioning, the stability of the inherited state of one cell is enhanced at the cost of the stability of its sister cell. This may be interesting in the critical case when switching to the other state can be lethal to the cells (e.g. the case of the bacteriophage λ-switch's lytic circle (Oppenheim et al., 2005; Shea and Ackers, 1985)).

In the case of the Repressilator, increasing partitioning errors result in more noisy periods of oscillation. However, the rate of desynchronization of the genetic clock in a cell population of the same lineage (the rate at which the phase of oscillation of the genetic clock in each cell deviates from the population's mean phase) was considerably slow for all level of partitioning errors, except for the extreme case (where one daughter cell inherits all the regulatory molecules).

These results show that the effects of noisy partitioning of regulatory molecules, especially low copied ones, on the dynamics of genetic circuits are not trivial to predict at the population level. The effects were also shown to differ widely from those of noise in gene expression. The differences arise from the inherent anti-correlation in the numbers of molecules inherited by sister cells and the periodicity of the division process, in contrast to noise in gene expression, which affects all cells in the population continuously and independently.

In **Publication III**, the expression dynamics of individual genes in response to the addition of inducers to the media with different inducer intake kinetics was studied. Three intake mechanisms were considered: purely diffusive, with positive feedback and with negative feedback. Using detailed stochastic models of gene expression and intake processes known to exist in *E. coli*, it was found that differing intake kinetics result in differences in the transient period of gene expression before reaching near-equilibrium, as well as the steady-state noise in the expression of the inducible gene. Namely, the intake process with negative feedback has the shortest transient time and causes least noise in the RNA and protein numbers of the inducible gene at steady state. The intake process with positive feedback, on the other hand, has the longest transient time and amplifies the noise in gene expression at equilibrium.

This work also showed that different intake kinetics can lead to different behaviors of genetic circuits, exemplified by a biased Toggle Switch. By having the expression of both genes of the Switch triggered by the same inducer, it was observed that the level of feedback in the intake process can alter the probability that the Switch chooses its first noisy attractor.

In **Publication IV**, the author and colleagues characterized the intake kinetics of the lactose analogue IPTG in *E. coli* DH5α-PRO. Like lactose, IPTG is hydrophilic and therefore does not diffuse easily through *E. coli*'s lipid bilayers (Decad and Nikaido, 1976; Zimmermann and Rosselet, 1977). Rather, it enters cells via means of active transport using H^+ symporters (Hansen et al., 1998; Ronald Kaback, 1983). Using the MS2-GFP RNA-tagging system (Golding et al., 2005), the transcription activity of lac-ara1 promoter upon the introduction of IPTG at different concentrations (from 0.25mM up to 1.25mM) in the media was captured. We then constructed a hybrid model of inducer intake through the bilayer membrane coupled with gene expression and fitted the model to observed data.

The results indicate that at high concentrations, each IPTG molecule on average takes on a considerable amount of time (~30 minutes at 37^{o}C) to travel cross the cell membranes. This duration is significantly longer than the reported time for L-arabinose (less than 5 minutes (Fritz et al., 2014; Megerle et al., 2008)), which is likely due to the latter's lower molecular weight (~150 Da, compared to ~238 Da of IPTG) (Benz and Bauer, 1988). The long traveling time for IPTG implies a lower bound for the response time of genetic systems to external excitation using IPTG. Also, in the regime of IPTG concentrations tested, the intake rate at the outer membrane was found to be linearly dependent on the external inducer concentration. This signifies that inducers can cross the outer membrane with a dynamics similar to that of a Michaelis–Menten process, where the amount of pores or transport proteins contributing to the intake process may take up only a small portion of the total amount of pores/proteins capable of intake.

Aside the knowledge on the lactose intake system in the regime of high concentrations, the work also established a method that can potentially be used to study the intake kinetics of other inducers.

In **Study V**, σ factor competition and its effects on the activity of different promoters are studied. Two RpoD-dependent promoters having different transcription kinetics are selected. In particular, the first promoter (P_{BAD}) has a long closed complex formation in comparison to the open complex formation. The second promoter (P_{tetA}) has a short closed complex formation when compared with the other step. Using the MS2-GFP RNA-tagging system (Le et al., 2005) with independent validations, the transcription dynamics of the two promoters is observed with and without the appearance of the stationary stress-induced factor RpoS.

From the measurements, it was found that the two promoters behave differently in the exponential phase (where *rpoS* gene is repressed at the translation process) and in the stationary phase (where *rpoS* gene is unrepressed). This leads to the hypothesis that the ratio between the durations of the closed complex formation, which is dependent on the specific RNAP holoenzyme copy number, and the unaffected open complex formation is responsible for promoter sensitivity to the σ factor competition.

Relevantly, the kinetics of closed and open complex formation are sequence dependent, implying that they are evolvable, either naturally or artificially. This “freedom” in the kinetics of the two rate-limiting steps in transcription initiation is expected to allow cells to diversify the behavior of each gene in response to changes in σ factor numbers.

5 Discussion

In this thesis, we established a theoretical framework to study the long-term dynamics of genetic circuits subject to noise in gene expression and errors in the partitioning of the regulatory molecules in cell division. Subsequently, we investigated the effects of different dynamics of intake of gene expression inducers and the effects of σ factor competition on the dynamics of transcription of individual promoters.

In our first study, we showed that the performance of chromosome-integrated genetic circuits can be improved by employing promoters of constituent genes with less noisy transcription processes (e.g. by having more rate-limiting steps). Therefore, we hypothesize that, when beneficial for the cellular functions to have a close-to-deterministic dynamics, natural promoters will evolve towards accommodating more rate-limiting steps in transcription initiation and, thus, achieve more robust cellular functions. These findings here are potentially applicable to Synthetic Biology: recent works have shown that the promoter sequence can be modified to achieve faster/slower transcription rates (Brewster et al., 2012; Kinney et al., 2010) or to accommodate varying number of rate-limiting steps in transcription initiation (Lutz et al., 2001). By replacing or modifying the promoter regions of the constituent genes, aside from altering their protein products, one can alter the kinetics of a genetic circuit significantly, without altering their gene products.

Next, we showed that the random partitioning of molecules in cell division has effects that depend on the gene network studied, as shown by the differences observed between the effects on the two genetic circuits modeled. In the switch, the increased variability results in a counter-intuitive decrease in the variance of the phenotype distribution. In the clock, no new features appeared aside from increased speed of desynchronization

over generations. This suggests that the effects of partitioning errors, to be correctly predicted, need to be considered from the point of view of the network under scrutiny.

Subsequently, in **Publication III** and **Publication IV**, the intake kinetics of lactose in cells was shown to play a major role in determining the dynamics of intracellular inducer numbers upon media perturbations. Given that the passage of each lactose molecule through the cell membrane was found to be time-consuming, the cell membrane is expected to act as a low-pass filter of chemical signals travelling from the external media to the cell cytoplasm. That is, the intracellular inducer levels will fail to capture the very fast dynamics of the extracellular concentrations. This poses an upper bound for the frequency of lactose concentrations that cellular functions can respond to.

Also in **Publication IV,** we performed the inference on the transport rates of IPTG through the outer membrane, without the aid of LacY permease proteins, as a function of inducer concentrations in the media, and under optimal conditions (37^{o}C, pH~7). Also estimated was the average time for each inducer molecule to travel through the membrane layers. This time proved to be of significance to the measured response time of gene expression, and future studies may be able to make use of this or similar measurements to learn more about the process of intake of other molecules and the membrane features involved in it. Also, in the future, it would be of interest to observe these parameters in cells placed in sub-optimal conditions. The results could inform on how the properties of the periplasm and the membrane pores, symporters change with, e.g., temperature or acidic levels.

A recent study on the localization of membrane proteins (Romantsov et al., 2010) reported that the lactose permease proteins (LacY) locate mostly at the cell poles. Because of this, aside from noise in gene expression, the dynamics in its numbers is inherently subject to partitioning errors in cell division (Huh and Paulsson, 2011a; Lindner et al., 2008). Thus, it is expected that the response of lac promoters to a small dose of lactose in the media would vary not only between wild type and lacY deletion strains, but also between dividing and non-dividing populations of wild-type *E. coli*. In the future, we plan to study whether the errors in the partitioning of LacY proteins in cell division contribute to the multistability of the lactose utilization network reported in (Ozbudak et al., 2004).

Finally, in **Study V**, it was shown that the response of gene expression to changes in the σ factor population varies between promoters of different σ factor selectivity and tran-

scription initiation kinetics. We expect that this phenomenon also occurs in genetic circuits. Namely, the constituent genes of the circuits can also react diversely and in opposite manners to these changes depending on their initiation kinetics, thus altering the circuits' responses and subsequent behavioral properties, such as the degree of bias of the two noisy attractors of a Toggle Switch, or the properties of the pass-band of genetic filters. We hypothesize that one of the features evolved in promoters is their initiation kinetics as means to govern to what extent they react to stress conditions, and that this is one of the variables determining cells epigenetics.

In conclusion, despite the recent technological advances in genetic engineering (Hasty et al., 2002; Wall et al., 2004), the operations performed by the realized genetic circuits remain almost circumstantial (Cardinale and Arkin, 2012; Cardinale et al., 2013). Given the findings reported in the thesis, the author corroborates this by showing that the behaviors of genetic circuits depend not only on the circuits' composition but also on the host cells and the environmental context. By providing information on the hosts' properties and how they communicate with the environment, one future direction for these studies should be directed towards identifying the boundaries that the host cells and the environment impose on the operations performed by the circuits.

References

Alberts, B., Bray, D., Lewis, J., Raff, M., 1994. Molecular biology of the cell, 1994, Garland, New York.

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., Storz, G., 1997. A small, stable RNA induced by oxidative stress: Role as a pleiotropic regulator and antimutator. Cell 90, 43–53. doi:10.1016/S0092-8674(00)80312-8

Anderson, J.C., Voigt, C. a, Arkin, A.P., 2007. Environmental signal integration by a modular AND gate. Mol. Syst. Biol. 3, 133. doi:10.1038/msb4100173

Arends, S.J.R., Weiss, D.S., 2004. Inhibiting cell division in *Escherichia coli* has little if any effect on gene expression. J. Bacteriol. 186, 880. doi:10.1128/JB.186.3.880

Arkin, A., Ross, J., McAdams, H.H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. Genetics 149, 1633–1648.

Atkinson, M.R., Savageau, M. a., Myers, J.T., Ninfa, A.J., 2003. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. Cell 113, 597–607. doi:10.1016/S0092-8674(03)00346-5

Baneyx, F., Mujacic, M., 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nat. Biotechnol. 22, 1399–1408. doi:10.1038/nbt1029

Barrios, H., Valderrama, B., Morett, E., 1999. Compilation and analysis of σ54-dependent promoter sequences. Nucleic Acids Res. 27, 4305–4313. doi:gkc653 [pii]

Becker, G., Hengge-Aronis, R., 2001. What makes an *Escherichia coli* promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S). Mol. Microbiol. 39, 1153–65.

Bell, C.E., Lewis, M., 2001. The Lac repressor: A second generation of structural and functional studies. Curr. Opin. Struct. Biol. 11, 19–25. doi:10.1016/S0959-440X(00)00180-9

Bentaboulet, M., Kepes, A., 1981. Dependence on pH of parameters of lactose transport in *Escherichia coli*: evidence for an essential protonated group of the carrier. Eur. J. Biochem. 117, 233–238.

Benz, R., Bauer, K., 1988. Permeation of hydrophilic molecules through the outer membrane of gram-negative bacteria. Eur. J. Biochem. 176, 1–19. doi:10.1111/j.1432-1033.1988.tb14245.x

Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 99, 9697–702. doi:10.1073/pnas.112318199

Boezi, J.A., Cowie, D.B., 1961. Kinetic studies of β-Galactosidase induction. Biophys. J. 1, 639–647. doi:10.1016/S0006-3495(61)86913-0

Bratsun, D., Volfson, D., Tsimring, L.S., Hasty, J., 2005. Delay-induced stochastic oscillations in gene regulation. Proc. Natl. Acad. Sci. U. S. A. 102, 14593–14598.

Brewster, R.C., Jones, D.L., Phillips, R., 2012. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. PLoS Comput. Biol. 8. doi:10.1371/journal.pcbi.1002811

Buc, H., McClure, W.R., 1985. Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry 24, 2712–2723. doi:10.1021/bi00332a018

Burns, H., Minchin, S., 1994. Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. Nucleic Acids Res. 22, 3840–3845. doi:10.1093/nar/22.19.3840

Busby, S., Ebright, R.H., 1999. Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. 293, 199–213. doi:10.1006/jmbi.1999.3161

Cabrera, J.E., Jin, D.J., 2001. Growth phase and growth rate regulation of the *rapA* gene, encoding the RNA polymerase-associated protein RapA in *Escherichia coli*. J. Bacteriol. 183, 6126–6134. doi:10.1128/JB.183.20.6126-6134.2001

Cao, Y., Gillespie, D., Petzold, L., 2005a. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. J. Comput. Phys. 206, 395–411. doi:10.1016/j.jcp.2004.12.014

Cao, Y., Gillespie, D.T., Petzold, L.R., 2005b. The slow-scale stochastic simulation algorithm. J. Chem. Phys. 122, 14116. doi:10.1063/1.1824902

Cardinale, S., Arkin, A.P., 2012. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. Biotechnol. J. 7, 856–866. doi:10.1002/biot.201200085

Cardinale, S., Joachimiak, M.P., Arkin, A.P., 2013. Effects of genetic variation on the *E. coli* host-circuit interface. Cell Rep. 4, 231–237. doi:10.1016/j.celrep.2013.06.023

Chandraseelan, J.G., Oliveira, S.M.D., Häkkinen, A., Tran, H., Potapov, I., Sala, A., Kandhavelu, M., Ribeiro, A.S., 2013. Effects of temperature on the dynamics of the LacI-TetR-CI repressilator. Mol. Biosyst. 9, 3117–23. doi:10.1039/c3mb70203k

Chang, D., Smalley, D.J., Conway, T., 2002. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. Mol. Microbiol. 45, 289–306.

Chen, H., Shiroguchi, K., Ge, H., Xie, X.S., 2015. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. Mol. Syst. Biol. 11, 781. doi:10.15252/msb.20145794

Chowdhury, S., Kandhavelu, M., Yli-Harja, O., Ribeiro, A.S., 2012. An interacting multiple model filter-based autofocus strategy for confocal time-lapse microscopy. J. Microsc. 245, 265–75. doi:10.1111/j.1365-2818.2011.03568.x

Cohen, S.N., Chang, a C., Hsu, L., 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. U. S. A. 69, 2110–2114. doi:10.1073/pnas.69.8.2110

Cooper, S., Helmstetter, C.E., 1968. Chromosome replication and the division cycle of *Escherichia coli* B/r. J. Bacteriol. 31, 519–540. doi:10.1016/0022-2836(68)90425-7

Cooper, S., Keasling, J., 1998. Cycle specific replication of chromosomal and F • plasmid origins. FEMS Microbiol. Lett. 163, 217–222.

Coquel, A.-S., Jacob, J.-P., Primet, M., Demarez, A., Dimiccoli, M., Julou, T., Moisan, L., Lindner, A.B., Berry, H., 2013. Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. PLoS Comput. Biol. 9, e1003038. doi:10.1371/journal.pcbi.1003038

Cramer, P., 2002. Multisubunit RNA polymerases. Curr. Opin. Struct. Biol. 12, 89–97.

doi:10.1016/S0959-440X(02)00294-4

Decad, G.M., Nikaido, H., 1976. Outer membrane of gram-negative bacteria. XII. Molecular-sieving function of cell wall. J. Bacteriol. 128, 325–36.

DeHaseth, P.L., Helmann, J.D., 1995. Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol. Microbiol. 16, 817–824. doi:10.1111/j.1365-2958.1995.tb02309.x

Dong, F., Spott, S., Zimmermann, O., Kisters-Woike, B., Müller-Hill, B., Barker, a, 1999. Dimerisation mutants of Lac repressor. I. A monomeric mutant, L251A, that binds Lac operator DNA as a dimer. J. Mol. Biol. 290, 653–666. doi:10.1006/jmbi.1999.2902

Dong, T., Schellhorn, H.E., 2009a. Control of RpoS in global gene expression of *Escherichia coli* in minimal media. Mol. Genet. Genomics 281, 19–33. doi:10.1007/s00438-008-0389-3

Dong, T., Schellhorn, H.E., 2009b. Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933. BMC Genomics 10, 349. doi:10.1186/1471-2164-10-349

Dong, T., Yu, R., Schellhorn, H., 2011. Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in *Escherichia coli*. Mol. Microbiol. 79, 375–86. doi:10.1111/j.1365-2958.2010.07449.x

Elf, J., Li, G.-W., Xie, X.S., 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. Science 316, 1191–1194. doi:10.1126/science.1141967

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338. doi:10.1038/35002125

Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. Science 297, 1183–1186. doi:10.1126/science.1070919

Emtage, J.S., Angal, S., Doel, M.T., Harris, T.J., Jenkins, B., Lilley, G., Lowe, P. a, 1983. Synthesis of calf prochymosin (prorennin) in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 80, 3671–3675. doi:10.1073/pnas.80.12.3671

Erickson, H.P., 2009. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biol. Proced. Online 11, 32–51. doi:10.1007/s12575-009-9008-x

Farewell, A., Kvint, K., Nyström, T., 1998. Negative regulation by RpoS: A case of sigma factor competition. Mol. Microbiol. 29, 1039–1051. doi:10.1046/j.1365-2958.1998.00990.x

Finkelstein, A., 1976. Water and nonelectrolyte permeability of lipid bilayer membranes. J. Gen. Physiol. 68, 127–135. doi:10.1085/jgp.68.2.127

Friedman, L.J., Mumm, J.P., Gelles, J., 2013. RNA polymerase approaches its promoter without long-range sliding along DNA. Proc. Natl. Acad. Sci. U. S. A. 110, 9740–5. doi:10.1073/pnas.1300221110

Fritz, G., Megerle, J. a., Westermayer, S. a., Brick, D., Heermann, R., Jung, K., Rädler, J.O., Gerland, U., 2014. Single cell kinetics of phenotypic switching in the arabinose utilization system of *E. coli*. PLoS One 9. doi:10.1371/journal.pone.0089532

Gaal, T., Ross, W., Estrem, S.T., Nguyen, L.H., Burgess, R.R., Gourse, R.L., 2001. Promoter recognition and discrimination by EσS RNA polymerase. Mol. Microbiol. 42, 939–954. doi:2703 [pii]

Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in *Escherichia coli*. Nature 403, 339–342. doi:10.1038/35002131

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Bala, G., Ravasz, E., Daugherty, M.D., Somera, a L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., Bhattacharya, A., Kapatral, V., Souza, M.D., Baev, M. V, Grechkin, Y., Mseeh, F., Fonstein, M.Y., Overbeek, R., Baraba, A., Oltvai, Z.N., 2003. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J Bacteriol. 185, 5673–5684. doi:10.1128/JB.185.19.5673

Ghazi, A., Shechter, E., 1981. Lactose transport in *Escherichia coli* cells. Dependence of kinetic parameters on the transmembrane electrical potential difference. Biochim. Biophys. Acta - Biomembr. 644, 305–315. doi:10.1016/0005-2736(81)90388-6

Gillespie, D.T., 1992. A rigorous derivation of the chemical master equation. Phys. A Stat. Mech. its Appl. 188, 404–425. doi:10.1016/0378-4371(92)90283-V

Gillespie, D.T., 1977. Concerning the validity of the stochastic approach to chemical kinetics. J. Stat. Phys. 16, 311–318. doi:10.1007/BF01020385

Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22, 403–434. doi:10.1016/0021-9991(76)90041-3

Goeddel, D. V, Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, a, Itakura, K., Riggs, a D., 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc. Natl. Acad. Sci. U. S. A. 76, 106–110. doi:10.1073/pnas.76.1.106

Golding, I., Cox, E.C., 2004. RNA dynamics in live *Escherichia coli* cells. Proc. Natl. Acad. Sci. U. S. A. 101, 11310–11315. doi:10.1073/pnas.0404443101

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

Gordon, G.S., Sitnikov, D., Webb, C.D., Teleman, A., Straight, A., Losick, R., Murray, A.W., Wright, A., 1997. Chromosome and low copy plasmid segregation in *E. coli*: Visual evidence for distinct mechanisms. Cell 90, 1113–1121. doi:10.1016/S0092-8674(00)80377-3

Gordon, S., Rech, J., Lane, D., Wright, A., 2004. Kinetics of plasmid segregation in *Escherichia coli*. Mol. Microbiol. 51, 461–469. doi:10.1046/j.1365-2958.2003.03837.x

Gralla, J.D., Carpousis, a J., Stefano, J.E., 1980. Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. Biochemistry 19, 5864–5869. doi:10.1021/bi00566a031

Grigorova, I.L., Phleger, N.J., Mutalik, V.K., Gross, C. a, 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. Proc. Natl. Acad. Sci. U. S. A. 103, 5332–5337. doi:10.1073/pnas.0600828103

Gruber, T.M., Gross, C.A., 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. 57, 441–66. doi:10.1146/annurev.micro.57.030502.090913

Gupta, A., Lloyd-Price, J., Oliveira, S.M.D., Yli-Harja, O., Muthukrishnan, A.-B., Ribeiro, A.S., 2014. Robustness of the division symmetry in *Escherichia coli* and functional consequences of symmetry breaking. Phys. Biol. 11, 066005. doi:10.1088/1478-3975/11/6/066005

Häkkinen, A., Muthukrishnan, A.-B., Mora, A., Fonseca, J.M., Ribeiro, A.S., 2013. CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli*. Bioinformatics 29, 1708–9. doi:10.1093/bioinformatics/btt194

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G., Elf, J., 2012. The lac Repressor displays facilitated diffusion in living cells. Science (80-.). 336, 1595–1598. doi:10.1126/science.1221648

Hansen, L.H., Knudsen, S., Sørensen, S.J., 1998. The effect of the lacY gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and Pseudomonas fluorescens. Curr. Microbiol. 36, 341–347. doi:10.1007/s002849900320

Hasty, J., McMillen, D., Collins, J.J., 2002. Engineered gene circuits. Nature 420, 224–230. doi:10.1038/nature01257

Hengge-Aronis, R., 2002. Stationary phase gene regulation: What makes an *Escherichia coli* promoter σS-selective? Curr. Opin. Microbiol. doi:10.1016/S1369-5274(02)00372-7

Henkin, T.M., 2000. Transcription termination control in bacteria. Curr. Opin. Microbiol. 3, 149–153. doi:10.1016/S1369-5274(00)00067-9

Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., Block, S.M., 2006. Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell 125, 1083–94. doi:10.1016/j.cell.2006.04.032

Hsu, L.M., 2002. Promoter clearance and escape in prokaryotes. Biochim. Biophys. Acta - Gene Struct. Expr. 1577, 191–207. doi:10.1016/S0167-4781(02)00452-9

Huh, D., Paulsson, J., 2011a. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat. Genet. 43, 95–100. doi:10.1038/ng.729

Huh, D., Paulsson, J., 2011b. Random partitioning of molecules at cell division. Proc. Natl. Acad. Sci. U. S. A. 108, 15004–9. doi:10.1073/pnas.1013171108

Ishihama, A., 2000. Functional modulation of *Escherichia coli* RNA polymerase. Annu. Rev. Microbiol. 54, 499–518.

Ishii, S., Sadowsky, M.J., 2008. *Escherichia coli* in the environment: implications for water quality and human health. Microbes Environ. 23, 101–108. doi:10.1264/jsme2.23.101

Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Andersson, C.R., Tanabe, a, Golden, S.S., Johnson, C.H., Kondo, T., 1998. Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria. Science 281, 1519–1523. doi:10.1126/science.281.5382.1519

Jacob, F., Monod, J., 1961. Genetic Regulatory Mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Jensen, P.R., Westerhoff, H. V., Michelsen, O., 1993. The use of lac-type promoters in control analysis. Eur. J. Biochem. 211, 181–191. doi:10.1111/j.1432-1033.1993.tb19885.x

Jishage, M., Ishihama, A., 1995. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular Levels of σ70 and σ38. J. Bacteriol. 177, 6832–6835.

Jishage, M., Iwata, A., Ueda, S., Ishihama, A., 1996. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli* : intracellular levels of four species of sigma subunit under various growth conditions. J. Bacteriol. 178, 5447–5451.

Johnson, C.M., Schleif, R.F., 1995. In vivo induction kinetics of the arabinose promoters in *Escherichia coli*. J. Bacteriol. 177, 3438–3442.

Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464. doi:10.1038/nrg1615

Kandhavelu, M., Häkkinen, A., Yli-Harja, O., Ribeiro, A.S., 2012a. Single-molecule dynamics of transcription of the lar promoter. Phys. Biol. 9, 026004. doi:10.1088/1478-3975/9/2/026004

Kandhavelu, M., Lloyd-Price, J., Gupta, A., Muthukrishnan, A.B., Yli-Harja, O., Ribeiro, A.S., 2012b. Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter. FEBS Lett. 586, 3870–3875. doi:10.1016/j.febslet.2012.09.014

Kandhavelu, M., Mannerström, H., Gupta, A., Häkkinen, A., Lloyd-Price, J., Yli-Harja, O., Ribeiro, A.S., 2011. 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. 5, 149. doi:10.1186/1752-0509-5-149

Kauffman, S.A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437–467. doi:10.1016/0022-5193(69)90015-0

Kerr, R. a, Levine, H., Sejnowski, T.J., Rappel, W.-J., 2006. Division accuracy in a stochastic model of Min oscillations in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 103, 347–352. doi:10.1073/pnas.0505825102

Kinney, J.B., Murugan, A., Callan, C.G., Cox, E.C., 2010. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proc. Natl. Acad. Sci. U. S. A. 107, 9158–9163. doi:10.1073/pnas.1004290107

Kittel, C., Kroemer, H., 1980. Thermal Physics.

Kussell, E., Leibler, S., 2005. Phenotypic diversity, population growth, and information in fluctuating environments. Science 309, 2075–2078. doi:10.1126/science.1114383

Langer-Safer, P.R., Levine, M., Ward, D.C., 1982. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc. Natl. Acad. Sci. U. S. A. 79, 4381–4385. doi:10.1073/pnas.79.14.4381

Laursen, B.S., Sørensen, H.P., Mortensen, K.K., Sperling-Petersen, H.U., 2005. Initiation of protein synthesis in bacteria. Microbiol. Mol. Biol. Rev. 69, 101–123. doi:10.1128/MMBR.69.1.101

Le, T.T., Harlepp, S., Guet, C.C., Dittmar, K., Emonet, T., Pan, T., Cluzel, P., 2005. Real-time RNA profiling within a single bacterium. Proc. Natl. Acad. Sci. U. S. A. 102, 9160–9164. doi:10.1073/pnas.0503311102

Lewis, M., 2005. The lac repressor. Comptes Rendus - Biol. 328, 521–548. doi:10.1016/j.crvi.2005.04.004

Liang, S., Bipatnath, M., Xu, Y., Chen, S., Dennis, P., Ehrenberg, M., Bremer, H., 1999. Activities of constitutive promoters in *Escherichia coli*. J. Mol. Biol. 292, 19–37. doi:10.1006/jmbi.1999.3056

Lindner, A.B., Madden, R., Demarez, A., Stewart, E.J., Taddei, F., 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proc. Natl. Acad. Sci. U. S. A. 105, 3076–81. doi:10.1073/pnas.0708931105

Liu, B., Wong, M.L., Alberts, B., 1994. A transcribing RNA polymerase molecule survives DNA replication without aborting its growing RNA chain. Proc. Natl. Acad. Sci. U. S. A. 91, 10660–10664. doi:10.1073/pnas.91.22.10660

Liu, B., Wong, M.L., Tinker, R.L., Geiduschek, E.P., Alberts, B.M., 1993. The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. Nature 366, 33–39. doi:10.1038/366033a0

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. doi:10.1006/meth.2001.1262

Lloyd-Price, J., Gupta, A., Ribeiro, A.S., 2012a. SGNS2: A compartmentalized stochastic chemical kinetics simulator for dynamic cell populations. Bioinformatics 28, 3004–3005. doi:10.1093/bioinformatics/bts556

Lloyd-Price, J., Lehtivaara, M., Kandhavelu, M., Chowdhury, S., Muthukrishnan, A.-B., Yli-Harja, O., Ribeiro, A.S., 2012b. Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*. Mol. Biosyst. 8, 565. doi:10.1039/c1mb05100h

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J., 2000. Molecular Cell Biology.

Lutz, R., Bujard, H., 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. 25, 1203–1210. doi:10.1093/nar/25.6.1203

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

Maeda, H., Fujita, N., Ishihama, A., 2000. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res. 28, 3497–3503. doi:10.1093/nar/28.18.3497

Mäkelä, J., Kandhavelu, M., Oliveira, S.M.D., Chandraseelan, J.G., Lloyd-Price, J., Peltonen, J., Yli-Harja, O., Ribeiro, A.S., 2013. In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter. Nucleic Acids Res. 41, 6544–6552. doi:10.1093/nar/gkt350

Malan, T.P., Kolb, a, Buc, H., McClure, W.R., 1984. Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. J. Mol. Biol. 180, 881–909. doi:10.1016/0022-2836(84)90262-6

Männik, J., Wu, F., Hol, F.J.H., Bisicchia, P., Sherratt, D.J., Keymer, J.E., Dekker, C., 2012. Robustness and accuracy of cell division in *Escherichia coli* in diverse cell shapes. Proc. Natl. Acad. Sci. U. S. A. 109, 6957–62. doi:10.1073/pnas.1120854109

Marbach, A., 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, 82–88. doi:10.1016/j.jbiotec.2011.10.009

Martínez-Antonio, A., Velázquez-Ramírez, D. a., Sánchez-Mondragón, J., Santillán, M., 2012. Hierarchical dynamics of a transcription factors network in *E. coli*. Mol. Biosyst. 8, 2932. doi:10.1039/c2mb25236h

Mauri, M., Klumpp, S., 2014. A model for sigma factor competition in bacterial cells. PLoS Comput. Biol. 10, e1003845. doi:10.1371/journal.pcbi.1003845

McAdams, H.H., Arkin, A., 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet. 15, 65–69. doi:10.1016/S0168-9525(98)01659-X

McClure, W.R., 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. 54, 171–204. doi:10.1146/annurev.bi.54.070185.001131

Megerle, J.A., Fritz, G., Gerland, U., Jung, K., Rädler, J.O., 2008. Timing and dynamics of single cell gene expression in the arabinose utilization system. Biophys. J. 95, 2103–15. doi:10.1529/biophysj.107.127191

Miller, O.L., Hamkalo, B. a, Thomas, C. a, 1970. Visualization of bacterial genes in action. Science 169, 392–395. doi:10.1126/science.169.3943.392

Miyada, C.G., Stoltzfus, L., Wilcox, G., 1984. Regulation of the *araC* gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on araBAD expression. Proc. Natl. Acad. Sci. U. S. A. 81, 4120–4124. doi:10.1073/pnas.81.13.4120

Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., Voigt, C. a., 2012. Genetic programs constructed from layered logic gates in single cells. Nature 1–5. doi:10.1038/nature11516

Mooney, R.A., Darst, S. a., Landick, R., 2005. Sigma and RNA polymerase: An on-again, off-again relationship? Mol. Cell 20, 335–345. doi:10.1016/j.molcel.2005.10.015

Müller-Hill, B., 1990. The isolation of the lac repressor. Bioessays 12, 41–43. doi:10.1073/pnas.56.6.1891

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

Muthukrishnan, A.-B., Martikainen, A., Neeli-Venkata, R., Ribeiro, A.S., 2014. In Vivo transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells. PLoS One 9, e109005. doi:10.1371/journal.pone.0109005

Nakahira, Y., Katayama, M., Miyashita, H., Kutsuna, S., Iwasaki, H., Oyama, T., Kondo, T., 2004. Global gene repression by KaiC as a master process of prokaryotic circadian system. Proc. Natl. Acad. Sci. U. S. A. 101, 881–885. doi:10.1073/pnas.0307411100

Nenninger, A., Mastroianni, G., Mullineaux, C.W., 2010. Size dependence of protein diffusion in the cytoplasm of *Escherichia coli*. J. Bacteriol. 192, 4535–4540. doi:10.1128/JB.00284-10

Nierman, C., Chamberlin, J., 1979. Studies of RNA chain bound to T7 DNA by *Escherichia coli* RNA polymerase bound to T7 DNA. J. Biol. Chem. 254, 7921–7926.

Nikaido, H., 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol. Mol. Biol. Rev. 67, 593–656. doi:10.1128/MMBR.67.4.593

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., Adhya, S., 2005. Switches in bacteriophage lambda development. Annu. Rev. Genet. 39, 409–429. doi:10.1146/annurev.genet.39.073003.113656

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., Van Oudenaarden, A., 2004. Multistability in the lactose utilization network of *Escherichia coli*. Nature 427, 737–740. doi:10.1038/nature02298

Pedraza, J.M., Paulsson, J., 2008. Effects of molecular memory and bursting on fluctuations in gene expression. Science 319, 339–343. doi:10.1126/science.1144331

Raffaelle, M., Kanin, E.I., Vogt, J., Burgess, R.R., Ansari, A.Z., 2005. Holoenzyme switching and stochastic release of sigma factors from RNA polymerase in vivo. Mol. Cell 20, 357–366. doi:10.1016/j.molcel.2005.10.011

Rahman, M., Hasan, M.R., Oba, T., Shimizu, K., 2006. Effect of *rpoS* gene knockout on the metabolism of *Escherichia coli* during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular metabolite concentrations. Biotechnol. Bioeng. 94, 585–595. doi:10.1002/bit.20858

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S., 2006. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4, 1707–1719. doi:10.1371/journal.pbio.0040309

Ramos, S., Kaback, H.R., 1977. The electrochemical proton gradient in *Escherichia coli* membrane vesicles. Biochemistry 16, 848–54.

Ravasz, E., Somera, a L., Mongru, D. a, Oltvai, Z.N., Barabási, a L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555. doi:10.1126/science.1073374

Ribeiro, A.S., Lloyd-Price, J., 2007. SGN Sim, a Stochastic Genetic Networks Simulator. Bioinformatics 23, 777–779. doi:10.1093/bioinformatics/btm004

Ribeiro, A.S., Zhu, R., Kauffman, S.A., 2006. A general modeling strategy for gene regulatory networks with stochastic dynamics. J. Comput. Biol. 13, 1630–9. doi:10.1089/cmb.2006.13.1630

Romantsov, T., Battle, A.R., Hendel, J.L., Martinac, B., Wood, J.M., 2010. Protein localization in *Escherichia coli* cells: comparison of the cytoplasmic membrane proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. J. Bacteriol. 192, 912–24. doi:10.1128/JB.00967-09

Ronald Kaback, H., 1983. The Lac carrier protein in *Escherichia coli*. J. Membr. Biol. 76, 95–112. doi:10.1007/BF02000610

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., Elowitz, M.B., 2005. Gene regulation at the single-cell level. Science 307, 1962–5. doi:10.1126/science.1106914

Roussel, M.R., Zhu, R., 2006. Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. Phys. Biol. 3, 274–284. doi:10.1088/1478-3975/3/4/005

Rouvière, P.E., De Las Peñas, A., Mecsas, J., Lu, C.Z., Rudd, K.E., Gross, C.A., 1995. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. EMBO J. 14, 1032–1042.

Ryals, J., Little, R., Bremer, H., 1982. Temperature dependence of RNA synthesis parameters in Temperature Dependence of RNA Synthesis Parameters in *Escherichia coli* 151, 879–887.

Saecker, R.M., Record, M.T., Dehaseth, P.L., 2011. Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J. Mol. Biol. 412, 754–771. doi:10.1016/j.jmb.2011.01.018

Samad, H. El, Vecchio, D. Del, Khammash, M., 2005. Repressilators and promotilators: loop dynamics in synthetic gene networks. Proc. 2005, Am. Control Conf. 2005. 4405–4410. doi:10.1109/ACC.2005.1470689

Samoilov, M., Arkin, A., Ross, J., 2002. Signal processing by simple chemical systems. J. Phys. Chem. A 106, 10205–10221.

Schleif, R., 2010. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. FEMS Microbiol. Rev. 34, 779–796. doi:10.1111/j.1574-6976.2010.00226.x

Shanner, N.C., Steinbach, P.A., Tsien, R.Y., 2005. A guide to choosing fluorescent proteins.

Nat. Methods 2, 905–909.

Shea, M.A., Ackers, G.K., 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. J. Mol. Biol. 181, 211–30.

Sheahan, M.B., Rose, R.J., McCurdy, D.W., 2004. Organelle inheritance in plant cell division: The actin cytoskeleton is required for unbiased inheritance of chloroplasts, mitochondria and endoplasmic reticulum in dividing protoplasts. Plant J. 37, 379–390. doi:10.1046/j.1365-313X.2003.01967.x

Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat. Genet. 31, 64–68. doi:10.1038/ng881

Shimizu, K., 2013. Metabolic regulation of a bacterial cell system with emphasis on *Escherichia coli* metabolism. ISRN Biochem. 2013, 1–47. doi:10.1155/2013/645983

Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., Margalit, H., 2007. Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol. Syst. Biol. 3, 138. doi:10.1038/msb4100181

Skinner, S.O., Sepúlveda, L. a, Xu, H., Golding, I., 2013. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. Nat. Protoc. 8, 1100–13. doi:10.1038/nprot.2013.066

Sledjeski, D.D., Gupta, a, Gottesman, S., 1996. The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. EMBO J. 15, 3993–4000.

So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L. a, Segev, R., Golding, I., 2011. General properties of transcriptional time series in *Escherichia coli*. Nat. Genet. 43, 554–560. doi:10.1038/ng.821

Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., Hasty, J., 2008. A fast, robust and tunable synthetic gene oscillator. Nature 456, 516–519. doi:10.1038/nature07389

Tanaka, K., Kusano, S., Fujita, N., Ishihama, a, Takahashi, H., 1995. Promoter determinants for *Escherichia coli* RNA polymerase holoenzyme containing sigma 38 (the rpoS gene product). Nucleic Acids Res. 23, 827–834. doi:4b0506 [pii]

Tani, T.H., Khodursky, A., Blumenthal, R.M., Brown, P.O., Matthews, R.G., 2002. Adaptation to famine: A family of stationary-phase genes revealed by microarray analysis. Proc. Natl. Acad. Sci. U. S. A. 99, 13471–13476.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–8. doi:10.1126/science.1188308

Valegård, K., Liljas, L., Fridborg, K., Unge, T., 1990. The three-dimensional structure of the bacterial virus MS2. Nature 345, 36–41. doi:10.1038/345036a0

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. doi:10.1101/gr.171405.113

Vogel, U., Jensen, K.F., 1994. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. J. Bacteriol. 176, 2807–2813.

Wall, M.E., Hlavacek, W.S., Savageau, M. a, 2004. Design of gene circuits: lessons from

bacteria. Nat. Rev. Genet. 5, 34–42. doi:10.1038/nrg1244

Walter, A., Gutknecht, J., Carolina, N., 1986. Permeability of small nonelectrolytes through Lipid bilayer membranes. J. Membr. Biol. 90, 207–217.

Wang, P., Robert, L., Pelletier, J., Dang, W.L., Taddei, F., Wright, A., Jun, S., 2010. Robust growth of *Escherichia coli*. Curr. Biol. 20, 1099–103. doi:10.1016/j.cub.2010.04.045

Wilson, K.S., von Hippel, P.H., 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc. Natl. Acad. Sci. U. S. A. 92, 8793–8797. doi:10.1073/pnas.92.19.8793

Wolf, D.M., Arkin, A.P., 2003. Motifs, modules and games in bacteria. Curr Opin Microbiol 6, 125–134. doi:10.1016/S1369-5274(03)00033-X

Yu, J., Xiao, J., Ren, X., Lao, K., Xie, X.S., 2006. Probing gene expression in live cells, one protein molecule at a time. Science 311, 1600–3. doi:10.1126/science.1119623

Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M., Golding, I., 2010. Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell 141, 682–91. doi:10.1016/j.cell.2010.03.034

Zhou, Z., Schnake, P., Xiao, L., Lal, A. a., 2004. Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. Protein Expr. Purif. 34, 87–94. doi:10.1016/j.pep.2003.11.006

Zimmer, D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, a B., Peter, B.J., Bender, R. a, Kustu, S., 2000. Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. Proc. Natl. Acad. Sci. U. S. A. 97, 14674–14679. doi:10.1073/pnas.97.26.14674

Zimmermann, W., Rosselet, A., 1977. Function of the outer membrane of *Escherichia coli* as a permeability barrier to beta-lactam antibiotics. Antimicrob. Agents Chemother. 12, 368–72.

PUBLICATION I

A Häkkinen, **H Tran**, O Yli-Harja, and AS Ribeiro (2013) “Effects of rate-limiting steps in transcription initiation on genetic filter motifs”, *PLoS ONE*, 8(8): e70439. doi: 10.1371/journal.pone.0070439.

OPEN ACCESS Freely available online

PLOS ONE

Effects of Rate-Limiting Steps in Transcription Initiation on Genetic Filter Motifs

Antti Häkkinen[1], Huy Tran[1], Olli Yli-Harja[1,2], Andre S. Ribeiro[1]*

1 Department of Signal Processing, Tampere University of Technology, Tampere, Finland, **2** Institute for Systems Biology, Seattle, Washington, United States of America

Abstract

The behavior of genetic motifs is determined not only by the gene-gene interactions, but also by the expression patterns of the constituent genes. Live single-molecule measurements have provided evidence that transcription initiation is a sequential process, whose kinetics plays a key role in the dynamics of mRNA and protein numbers. The extent to which it affects the behavior of cellular motifs is unknown. Here, we examine how the kinetics of transcription initiation affects the behavior of motifs performing filtering in amplitude and frequency domain. We find that the performance of each filter is degraded as transcript levels are lowered. This effect can be reduced by having a transcription process with more steps. In addition, we show that the kinetics of the stepwise transcription initiation process affects features such as filter cutoffs. These results constitute an assessment of the range of behaviors of genetic motifs as a function of the kinetics of transcription initiation, and thus will aid in tuning of synthetic motifs to attain specific characteristics without affecting their protein products.

Citation: Häkkinen A, Tran H, Yli-Harja O, Ribeiro AS (2013) Effects of Rate-Limiting Steps in Transcription Initiation on Genetic Filter Motifs. PLoS ONE 8(8): e70439. doi:10.1371/journal.pone.0070439

Editor: Christophe Herman, Baylor College of Medicine, United States of America

Received March 13, 2013; **Accepted** June 18, 2013; **Published** August 5, 2013

Funding: This work was supported by the Academy of Finland and the Finnish Funding Agency for Technology and Innovation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: andre.ribeiro@tut.fi

Introduction

Genes function in networks, whose building blocks are motifs of few genes. Several motifs have been identified, which perform a specific function in networks [1]. Examples include genetic switches, which can be used as memory circuits or for digital control of processes; oscillators, which can be used for time-keeping and synchronization; and genetic filters, which can be used for noise filtering and computation via genetic logic [1].

In addition to the gene-gene interactions, the behavior of a motif depends on the expression pattern of each constituent gene. Investigating this dependency is of relevance given recent evidence that both mean level and the cell to cell diversity in RNA and protein numbers vary between genes by several orders of magnitude [2]. For that, we need to use models that account for the nature of gene expression, since genes with low expression levels are abundant in bacteria [2,3]. Such low numbers cause the dynamics of motifs to be poised with correlations and low copy number fluctuations.

Much effort has been made to characterize the processes of transcription and translation in bacteria. In vitro studies [4,5] showed that transcription, the process by which RNA molecules are produced, is controlled mostly at the promoter region of the gene. Once the RNA polymerase reaches the transcription start site and forms the closed complex, it remains there until the open complex is complete. Following this, the polymerase can escape the promoter and elongate along the DNA sequence, according to which the RNA sequence will be assembled. Both in vitro and in vivo studies suggest that the closed and open complex formations are the lengthiest (rate-limiting) steps of the process of gene expression, along with protein folding and activation.

Recently, the intervals between transcription events in individual, live cells have been measured for two promoters, lac-ara-1 [6] and tetA [7]. These studies suggest that, under optimal conditions, there are two to three major rate-limiting steps, which occur during initiation, that control both mean rate and noise in RNA production. These steps durations were also shown to vary widely with induction level and environmental conditions [6,7]. In that sense, they are major regulators of the dynamics of mRNA production.

Since the duration of the rate-limiting steps in transcription is both sequence-dependent and regulated by activator and repressor molecules, these steps are both evolvable and adaptive to the environment [6]. Since in prokaryotes translation is coupled with transcription, these steps are likely also key regulators of protein numbers [8]. However, it remains unknown to what extent one can tune the behavior of genetic motifs by selecting specific kinetics of initiation of the constituent genes.

In this work, we study the behavior of stochastic genetic motifs, while varying the kinetics of transcription initiation of the constituent genes. Two motifs are considered: one performs filtering in the amplitude domain, and the other in the frequency domain. The response of the motifs is quantified for a wide range of transcriptional dynamics that are in accordance with measurements.

The results indicate that the dynamics of these two genetic motifs, while dependent of the gene-to-gene interactions, is also affected by the kinetics of transcription initiation of each component gene. This, in turn, suggests that it is possible to engineer synthetic circuits to be more robust or having higher plasticity than the present ones, by selecting for promoters with appropriate initiation kinetics.

Methods

Gene expression

We use the delayed stochastic modeling strategy [9,10], which correctly accounts for the low copy number effects, that is, the fluctuations and correlations, of the interacting components, coupled with non-exponential waiting times. The results are quantified from Monte Carlo simulations of the reaction system, using SGN Sim [11].

To model gene expression we use the following set of reactions. The syntax $A \xrightarrow{k} B + C(\tau)$ denotes a reaction where A is transformed into B and C, with a stochastic rate of k. While B is released in the vessel of reactions instantaneously once the reaction occurs, C is released after a delay of τ [10].

$$S \xrightarrow{\infty} S(\tau) + M(\tau) \tag{1}$$

$$M \xrightarrow{d_M} \varnothing \tag{2}$$

$$\varnothing \xrightarrow{k_P M} P \tag{3}$$

$$P \xrightarrow{d_P} \varnothing \tag{4}$$

where $S=1$ ($S=0$) denotes that the promoter is free (occupied), M is the messenger RNA, and P is the protein. Reaction 1 models transcription, Reaction 2 mRNA degradation (d_M being the mRNA degradation rate), Reaction 3 translation (k_P representing the per-mRNA translation rate), and Reaction 4 protein degradation (d_P denoting the protein degradation rate).

The infinite rate set for Reaction 1 derives from the assumption that there is an inexhaustible pool of polymerases (which is a common assumption for bacteria in optimal growth conditions). The delay τ represents the effects of all rate-limiting steps, including the initiation of transcription up to the production of an mRNA. As mentioned, recent evidence suggests that, in *E. coli* under optimal growth conditions, τ is determined to a great extent by the sum of two to three rate-limiting steps, each following an exponential distribution in duration [6,7]. We use $\tau \sim \Gamma(\alpha, \alpha^{-1}\lambda^{-1})$, which denotes that the delay τ is drawn from gamma distribution with a shape of α and a mean of λ^{-1}. Integer values of α indicate that transcription consists of α sequential steps, each with a rate of $\alpha\lambda$. The gamma distribution has a coefficient of variation (the standard deviation over the mean) of $\alpha^{-1/2}$ regardless of the mean (cf. unity of the exponential distribution, which is a gamma distribution with $\alpha=1$). Consequently, values of $\alpha=1$ will result in a Poisson distributed $M \sim \mathrm{Poi}(\lambda d_M{}^{-1})$, while values of $\alpha<1$ result in a more noisy (super-Poisson), and values of $\alpha>1$ less noisy (sub-Poisson) mRNA number dynamics. We note that even if transcription initiation consists of sequential steps of unequal duration, the gamma distribution is still a good approximation. If the steps are of the same order of magnitude, they can be considered approximately equal, else, fast steps can be neglected.

Finally, we let $\lambda \doteq k_M f(X_1, \ldots, X_n)$, where k_M indicates the maximal expression rate of the promoter, and $f(X_1, \ldots, X_n) : \mathbb{N}_0{}^n \mapsto [0,1]$ is a regulatory function of the promoter, which depends on substances X_1 through X_n. It is generally not known which steps are affected by which transcription factors, so we assume that each step is affected in an equivalent manner. The choice of these functions is discussed in the next section. Moreover, we let $\mu \doteq k_M d_M{}^{-1} k_P d_P{}^{-1}$, which coincides with the expected protein level of a gene under full expression.

Unless otherwise stated, we use the parameters $k_M d_M{}^{-1}=5$, $d_M=(5\ min)^{-1}$, $k_P d_P{}^{-1}=100$, and $d_P=(60\ min)^{-1}$. These values were selected in accordance with measurements in live *E. coli* [2]. In the results presented, each simulation is ran for 10^6 *min*, and the system is sampled uniformly every 1 *min*. To assess the kinetics of initiation within a realistic range of parameter values, we set the number of rate-limiting steps $\alpha \in \{1,2,3,5,10\}$. The first three have been observed in measurements of mRNA production kinetics in live *E. coli* cells [6,7]. In vitro studies of the kinetics of this process (see e.g. [12]) provide evidence for the existence of, at least, five rate-limiting steps, namely, closed complex formation, three isomerization steps, and promoter clearance. We also study the effects of setting α to 10 to observe the behavior of the model in limit conditions and due to the fact that some of the steps might be non-exponential in duration, thus requiring multiple exponentially distributed steps to be well described.

Gene regulation

The genes are coupled by interactions between their promoter regions and the proteins they express. The activation/repression of a gene is achieved by the binding of the protein expressed by another gene. Once bound, this protein can either degrade while bound, or unbind. While bound, the propensity for the gene to express differs from the unbound case. The activation/repression of gene B by gene A could be represented by the following set of reactions:

$$P_A + T_B \xrightarrow{k_{AB}} T'_B \tag{5}$$

$$T'_B \xrightarrow{k_{AB} K_{AB}} P_A + T_B \tag{6}$$

$$T'_B \xrightarrow{d_P} T_B \tag{7}$$

where P_A denotes the protein product of gene A, $T_B=1$ denotes that the binding site of the gene B for that protein is free, and $T'_B=1$ (implying $T_B=0$) that the binding site is occupied. Here, Reaction 5 models the binding of the activator/repressor molecule P_A to the promoter region of gene B, Reaction 6 its unbinding, and Reaction 7 the degradation of a bound protein. The rate of binding is denoted by k_{AB} and the disassociation constant by K_{AB}.

To simplify the model, we take the limit $k_{AB} \to \infty$. In this limit, the binding of the regulatory proteins is assumed to be much faster than the rate of transcription. It can be found that in this limit, the expectation $\mathbf{E}[T'_B] = (1 + K_{AB} P_A{}^{-1})^{-1}$ if P_A is constant. Following this, to implement the regulation, we vary the transcription rate such that:

$$f_{AB}(A) = \left(1 + (K_{AB} P_A^{-1})^{+d}\right)^{-1} \quad \textit{iff gene A activates gene B} \tag{8}$$

$$f_{AB}(A) = \left(1 + (K_{AB} P_A^{-1})^{-d}\right)^{-1} \quad \textit{iff gene A represses gene B} \tag{9}$$

and

$$f_{ABC}(A,B)=f_{AC}(A)f_{BC}(B) \quad \text{iff genes A and B regulate gene C} \tag{10}$$

where d denotes the Hill coefficient, which represents the cooperativity of binding, (e.g. $d=2$ can be taken that there are two binding sites for a same type of protein) determining how steep the transition between on- and off-states (e.g. $\mathbf{E}[T'_B]=0$ and $\mathbf{E}[T'_B]=1$) is. Also, the role of the disassociation constant in this context is now apparent, namely, it follows that $\mathbf{E}[T'_B]=0.5$ iff $K_{AB}=P_A$. In our simulations, we use $d=2$, since many proteins are known to function in a dimeric form [13].

Results

Amplitude filtering

We start by examining how the properties of a genetic motif performing amplitude filtering are affected by the transcriptional dynamics. A genetic motif capable of behaving as a biphasic amplitude filter should allow the output to be active only for a certain range of input levels, which allows a process to be trigged by a narrow range of molecular concentration [1]. The region of inputs where the output is active is called the passband and the non-active regions are referred by stopbands. We model a biphasic amplitude filter consisting of four genes as follows. Gene A activates the expression of genes B and C, and gene B activates the expression of gene D, while gene C represses gene D. We model explicitly the expression of genes B through D, while the relative expression level of gene A acts as an input parameter. This is illustrated in Figure 1. Such a circuit was used to explain the narrow range of induction triggering the expression of Xbra in *Xenopus laevis* [14].

We simulate the model for various values of shape α' and rate k'_M of transcription of genes B and C, while the output gene shape and rate are kept constant ($\alpha=2$, $k_M d_M{}^{-1}=5$). This is due to the fact that the effects of changes in α and k_M in the protein distribution of the output gene are more apparent and not related to the internal behavior of the filter, and because it allows the different cases to be easily compared. We set $K_{BD}=0.25\mu'$ and $K_{CD}=0.1\mu'$, which is expected to produce a biphasic response (see Equations 11 through 13). In this, $\mu'=k'_M d_M{}^{-1} k_P d_P{}^{-1}$ denotes the expression rate of genes B and C under full expression. To vary the mean input level, we vary the quantity $\rho=K_{AB}{}^{-1}$ $P_A=10K_{AC}{}^{-1}P_A\propto P_A$.

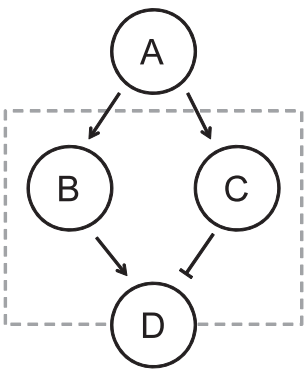


Figure 1. Illustration of the biphasic amplitude filter motif. In the biphasic amplitude filter, gene A acts as input to the filter, while genes C and D compose the filter, represented by the dashed box, along with the regulatory connections between each gene. The protein level of gene D acts as the output.
doi:10.1371/journal.pone.0070439.g001

If all molecule numbers were constant, the response of the filter could be characterized by the following equations:

$$P_B=\mu'\left(1+(K_{AB}P_A{}^{-1})^d\right)^{-1} \tag{11}$$

$$P_C=\mu'\left(1+(K_{AC}P_A{}^{-1})^d\right)^{-1} \tag{12}$$

$$P_D=\mu\left(1+(K_{BD}P_B{}^{-1})^d\right)^{-1}\left(1+(P_C K_{CD}{}^{-1})^d\right)^{-1} \tag{13}$$

which is a good approximation for high expression levels. Note that in Equation 13, P_D is a function of ρ, but invariant to the parameters α' and k'_M, thus the effects of varying them lie beyond this formula. The response of the filter using Equations 11 through 13 is depicted in Figure 2.

The molecular levels will not be constant in our stochastic model. We quantify the noise in molecular levels using Fano factor (the variance over the mean), which is convenient, since Fano factor of Poisson-distributed molecules equals unity regardless of the mean. Even in the limit $\alpha\to\infty$ the protein levels will remain highly noisy (Fano factor $\mathbf{Fano}[P]\geq 1$), since in this case $P_B,P_C\sim\mathrm{Poi}(\mu')$ and their noise further propagates through the probabilistic expression of gene D to the output protein levels P_D.

Next, we present the response of the biphasic amplitude filter using the stochastic model, and study how much it deviates from the expected response when the shape and rate of transcription are varied. The mean output level of the output gene D is presented in Figure 3. As expected, the response resembles the curves in Figure 2. Lower values of α (which imply higher noise) produce slightly degraded performance in terms of the response of the filter. That is, the maximum output protein level will be lower, and the transition between the on- and off-states will be less steep. In addition, the increased noise makes the passband to shift toward a higher input level, since the distributions resulting from the model tend to have right skew.

We also assessed the response for various mean expression levels μ' of the component genes (Figure 4). The results are qualitatively similar to those in Figure 3. Decreasing α' or k'_M (either leading to higher noise) will degrade the filter performance. Moreover, as the expression rate is lowered the shape of the transcription takes greater role in determining the filter behavior. This implies that for rarely expressed genes, it might be important to have sub-Poissonian transcript dynamics, to compensate the increased low copy number noise.

Adding noise in the processes within the filter must shift downwards the value of the maximum output protein level. Generally, adding noise results in a flatter response, which can be interpreted as a degradation in performance, since the filter aims to selectively turn the output on or off. Furthermore, it is possible that adding noise also shifts the input level for which the maximal output is attained or the locations of the transition bands. The results depend on whether the input distributions and the response function of the filter are symmetric or not.

Finally, we assessed quantitatively the effects on the output of having different values of α', for each expression ratio of the input gene shown in Figure 4. For $\mu'\mu^{-1}=0.01$, increasing α' from 1 to 2, causes the output amplitude in the passband to increase by 10.8%. Increasing α' from 1 to 3 causes the output amplitude to

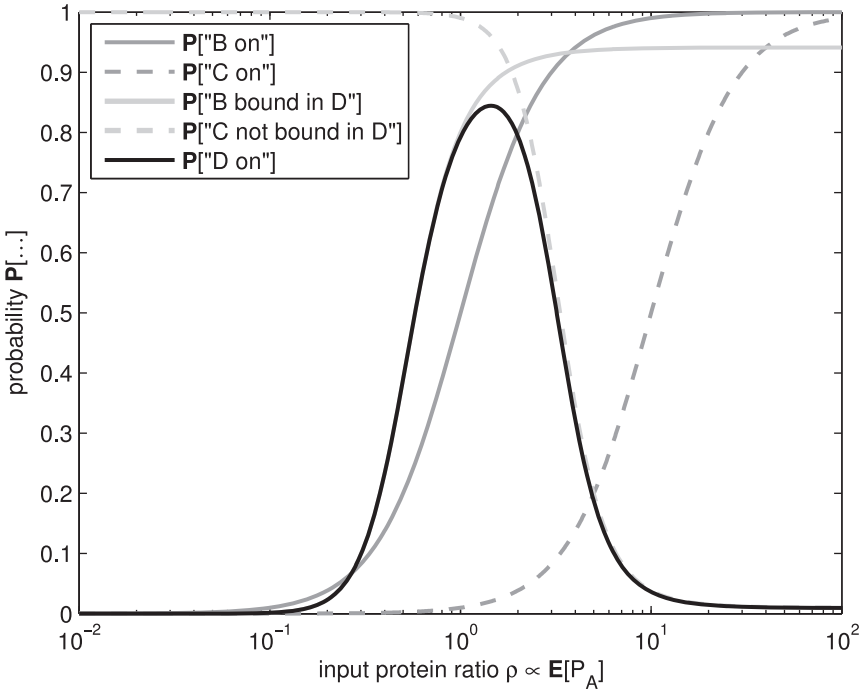


Figure 2. Event probabilities in biphasic amplitude filter. Probabilities of events in the biphasic amplitude filter as a function of the input protein level $\mathbf{E}[P_A]$. The solid black line denotes the probability that the output gene D is expressing, while the dark gray lines denote those of the intermediate genes (solid denoting gene B and dashed gene C). The probabilities that the intermediate genes allow the output gene to express are depicted by the light gray lines (solid denoting gene B and dashed gene C).
doi:10.1371/journal.pone.0070439.g002

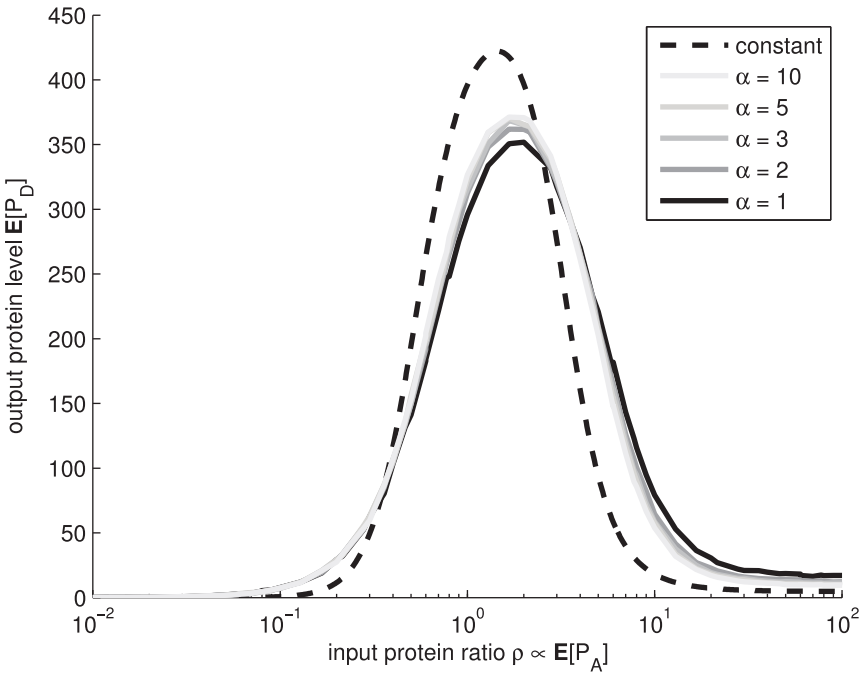


Figure 3. Mean response of biphasic amplitude filter. Mean response $\mathbf{E}[P_D]$ of the biphasic amplitude filter as a function of the protein level $\mathbf{E}[P_A]$ of the input gene, for various shapes α'. Different levels of gray denote different shape parameter α'. The simulations were performed with $\mu'\mu^{-1}$ of 0.05. The dashed black line is an approximation, assuming constant molecular levels.
doi:10.1371/journal.pone.0070439.g003

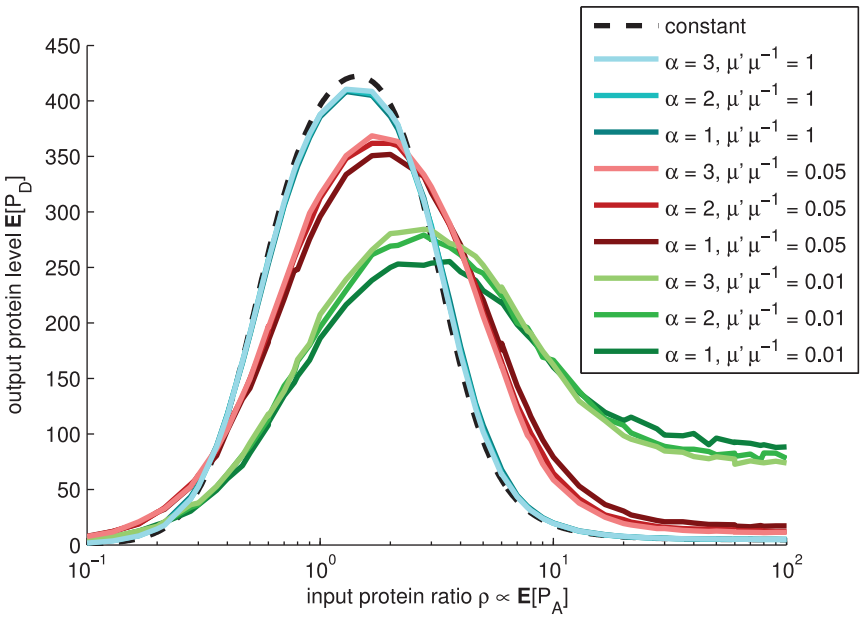


Figure 4. Mean response of biphasic amplitude filter for various transcription rates. Mean response $\mathbf{E}[P_D]$ of the biphasic amplitude filter as a function of the input gene protein level $\mathbf{E}[P_A]$, for various shapes α' and rates k'_M of transcription. Different levels of brightness denote different shape parameter α'. The simulations were performed with $\mu'\mu^{-1}$ of 1 (cyan), 0.05 (red), and 0.01 (green), in the order of decreasing performance. The three cyan lines overlap. We also performed simulations with $\mu'\mu^{-1}$ of 0.5, 0.2, 0.1, and 0.02 (not shown) to assert that the changes are generally nonlinear and more drastic for low mean levels. The dashed black line is an approximation, assuming constant molecular levels.
doi:10.1371/journal.pone.0070439.g004

increase by 12.9%. For other values of $\mu'\mu^{-1}$, the differences are smaller. For example, for $\mu'\mu^{-1}=0.05$, these increases are, respectively, 7.2% and 8.5%, while for $\mu'\mu^{-1}=1$, these differences are of the order of 1.5%.

Since our model dynamics is poised with noise, we study the noise in the output gene protein level, as a function of the input gene level. One might expect the noise to take a shape that is characteristic to the output gene, e.g. constant for Poisson, or some monotonically decreasing curve in our case. In the presence of noisy molecular levels in the circuit, this is generally not true. The noise in the output of this motif is expected to be higher in the transition bands of the biphasic amplitude filter, with the magnitude more characteristic to the output gene in the pass- and stop-bands. An example from stochastic simulations is presented in Figure 5.

From Figure 5 we find that even when the effects of changes in transcription initiation on the response of the biphasic amplitude filter are slight, the change in the fluctuations of the protein numbers of the output gene might be significant. In Figure 6, we present the output noise for various mean levels. For very low expression levels, the low copy number noise in the output becomes dominant.

As a consequence of the amplification of the noise in the transition bands, the output of the filter becomes unpredictable in these regions. Therefore, for this circuit to operate properly in these regions, it is of importance to minimize the noise in the genes composing the filter, for example, by adding rate-limiting steps in initiation. Alternatively, regulation schemes that can provide steeper transition bands are required, which can be accomplished via regulatory schemes of higher-order. We hypothesize that the latter scheme has less effect, since it cannot remove the problem, only reduce its effects. Moreover, it is harder to implement in real genetic circuits, as it requires altering both the protein and the promoter sequences.

Frequency filtering

In this section, we study the effects of changes in the transcription dynamics to a motif that performs filtering in the frequency domain. It is known that changes in the transcriptional dynamics can affect the period and its robustness of genetic oscillators [15], so we expect that these changes affect the response of certain frequency filters as well.

We constructed a motif that can perform low-pass frequency filtering composed of four genes (A through D). This filter suppresses highly transient signals while letting slowly varying signals to pass through as-is. Such a filter would allow a specific set of genes to be subject to only stable signals, by filtering out fast fluctuations in the numbers of the regulatory molecules. Here, gene A acts as an input, required to enable the expression of gene B. Gene B represses gene C, C represses D, and D represses B, that is, genes C through D form a loop (three-gene repressilator). The structure of the motif is illustrated in Figure 7.

When a periodic signal P_A is applied, the behavior of this circuit should vary, depending on the frequency of the signal. When the signal is of high frequency, the feedback loop should be the main responsible for the frequency content of the output. For low frequencies, the input from gene A will disconnect the feedback loop periodically, and lower frequencies, including that of P_A, are introduced in the output. Thus, it is expected that the modulated circuit would have a synchronization point when the input frequency equals that of the repressilator, and that a phase transition would occur in the output frequency response.

For simplicity, we let the Hill coefficient $d'\to\infty$, in the regulatory connection where A activates B. That is, the regulatory connection becomes Boolean, with a threshold of K_{AB}. We denote the Boolean input signal by $X \doteq (1+(K_{AB}P_A{}^{-1})^{d'})^{-1}$. This allows us to omit the explicit modeling of gene A, and consequently this parameter does not need to be determined. Instead, we can apply an arbitrary $X\in\mathbb{B}$. In this case, it does not

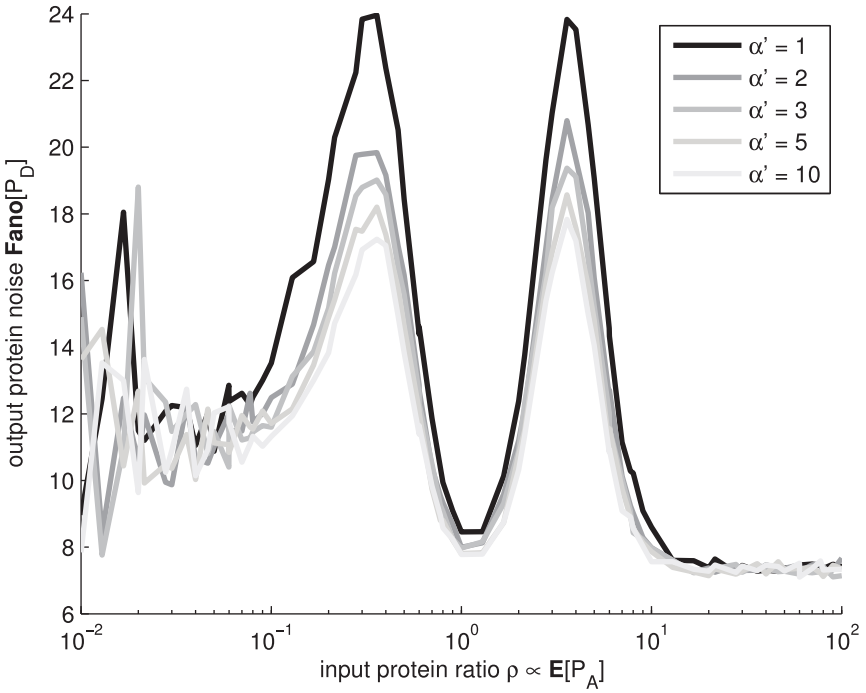


Figure 5. Noise of response of biphasic amplitude filter. Noise of the response $\mathbf{Fano}[P_D]$ of the biphasic amplitude filter as a function of the input gene protein level $\mathbf{E}[P_A]$, for various shapes α'. Different levels of gray denote different shape parameter α'. The simulations were performed with $\mu'\mu^{-1}$ of 1.
doi:10.1371/journal.pone.0070439.g005

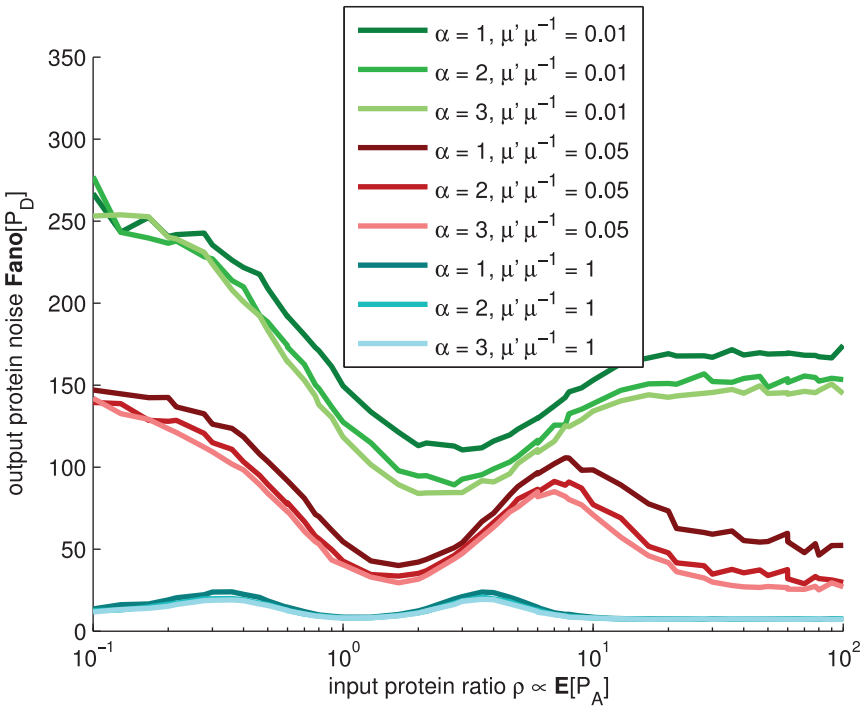


Figure 6. Noise of response of biphasic amplitude filter for various transcription rates. Noise of the response $\mathbf{Fano}[P_D]$ of the biphasic amplitude filter as a function of the input gene protein level $\mathbf{E}[P_A]$, for various shapes α' and rates k'_M of transcription. Different levels of brightness denote different shape parameter α'. The simulations were performed with $\mu'\mu^{-1}$ of 0.01 (green), 0.05 (red), and 1 (cyan), in the order of decreasing noise. We also performed simulations with $\mu'\mu^{-1}$ of 0.5, 0.2, 0.1, and 0.02 (not shown) to assert that the changes are generally nonlinear and more drastic with low mean levels. The dashed black line is an approximation, assuming constant molecular levels.
doi:10.1371/journal.pone.0070439.g006

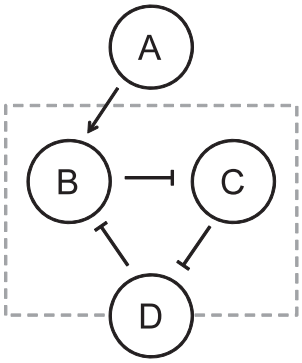


Figure 7. Illustration of the frequency filtering motif. In the frequency filtering motif, gene A acts as an input to the motif, while the filter consists of genes B, C, and D in a feedback loop structure along with the modulation by the input, represented by the dashed box. The protein level of gene D acts as an output of the filter.
doi:10.1371/journal.pone.0070439.g007

matter if the connection is an activating (as in Figure 7) or repressing, since the Boolean input can be flipped.

First, we let the input signal to be constant $X=1$. We analyze the periodic behavior characteristic to the submotif of genes B, C, and D. Since the genes B, C, and D are identical, we can treat them interchangeably and quantify the distribution of periods from any of the protein levels, denoted by T_{BCD}, from the zeros of the autocorrelation function of each time series.

We simulate our model for values of shape α' and rate k'_M of genes B, C, and D, and μ' is defined analogously to the previous subsection. Moreover, the disassociation constants are set to $K_{BC}=K_{CD}=K_{DB}=0.05\mu'$, which were found to produce an oscillatory signal under constant input. The mean period of the protein levels of genes B, C, and D, as a function of the mean expression level μ' of the genes, is shown in Figure 8.

Interestingly, the period changes as a function of the number of steps in transcription initiation. Also, changing the mean transcription level affects the period (note that the disassociation constants are a function of the expected expression level μ', which would make a deterministic model invariant of μ').

We also examined if the robustness of the period is affected. We quantify robustness by the coefficient of variation of the periods of the protein numbers. This measure is convenient, since it equals unity for exponentially distributed periods regardless of the mean. The results are shown in Figure 9. For low mean protein numbers, the period becomes unpredictable (i.e. exponential-like), whereas for moderate levels, the period distribution is Gaussian-like, due to lower noise in transcript production, implying more robust period length. The shapes of the distribution were verified from period histograms (see examples in the insets in Figure 9).

Next, we apply an unbiased Boolean square wave to X, that is, $X(t)=0$ for time t that satisfies $kT\le t<(k+1/2)T$ with any integer k and $X(t)=1$ otherwise, and we denote its frequency by $f_X=T^{-1}$, where T refers to the period. The autocorrelation function of this signal X is a triangular wave of the same frequency, and consequently its spectral power is concentrated to the harmonics of f_X. The spectral power is measured in terms of power spectral density (PSD), which is given by the Fourier transform of the autocorrelation function and measures how much of the signal power per unit frequency is concentrated around certain frequency. Specifically, the PSD of X at frequency f_X is $4\pi^{-2}$ (cf. Figure 10).

We measure the power spectral densities of the input X and the output P_D. An example is shown in Figure 10, with the input PSD plotted for reference. The motif exhibits a low-pass behavior in the frequency domain. Frequencies lower than those corresponding to the mean period of the three-gene submotif when functioning

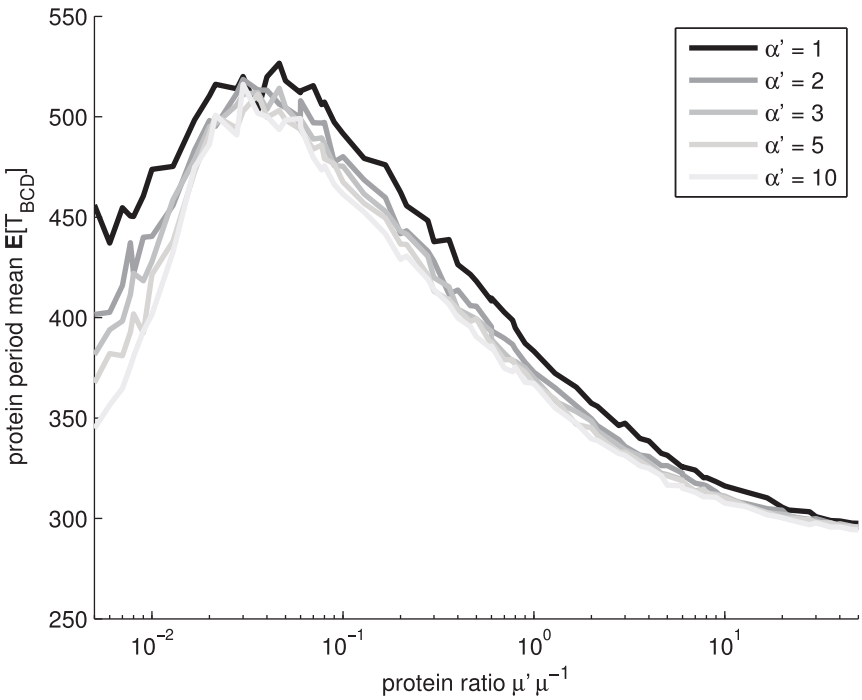


Figure 8. Mean period of frequency filtering motif with constant input. Mean period of the protein levels of genes B, C, and D ($\mathbf{E}[T_{BCD}]$), for constant input $X=1$. Different levels of gray denote different shape parameter α'.
doi:10.1371/journal.pone.0070439.g008

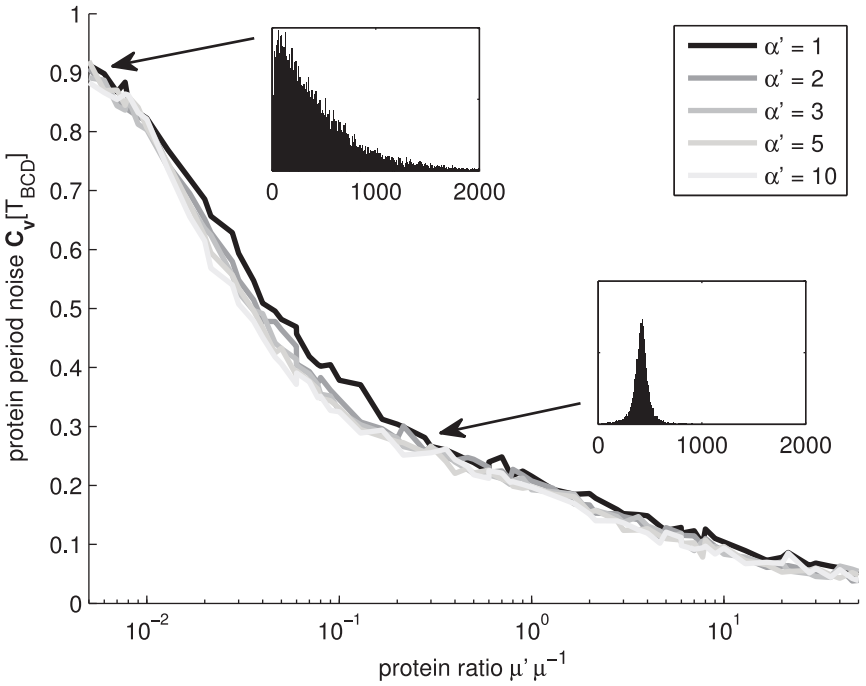


Figure 9. Noise in period of frequency filtering motif with constant input. Noise in the period of the protein levels of genes B, C, and D ($\mathbf{C}_v[T_{BCD}]$), for constant input $X=1$. Different levels of gray denote different shape parameter α'. The insets exemplify the distributions of periods T_{BCD} for shape of $\alpha'=1$ and ratios $\mu'\mu^{-1}$ of 0.005 and 0.5 (units of the x-axis are seconds).
doi:10.1371/journal.pone.0070439.g009

independently (see Figure 8) are only slightly attenuated (amplification factor of $>10^{-1}$). In contrast, higher frequencies are highly attenuated (amplification factor of $<10^{-4}$).

Changing the shape parameter α' of the transcription results in slight variations in the performance of the frequency filter, while the main characteristics are not changed. Namely, the attenuation

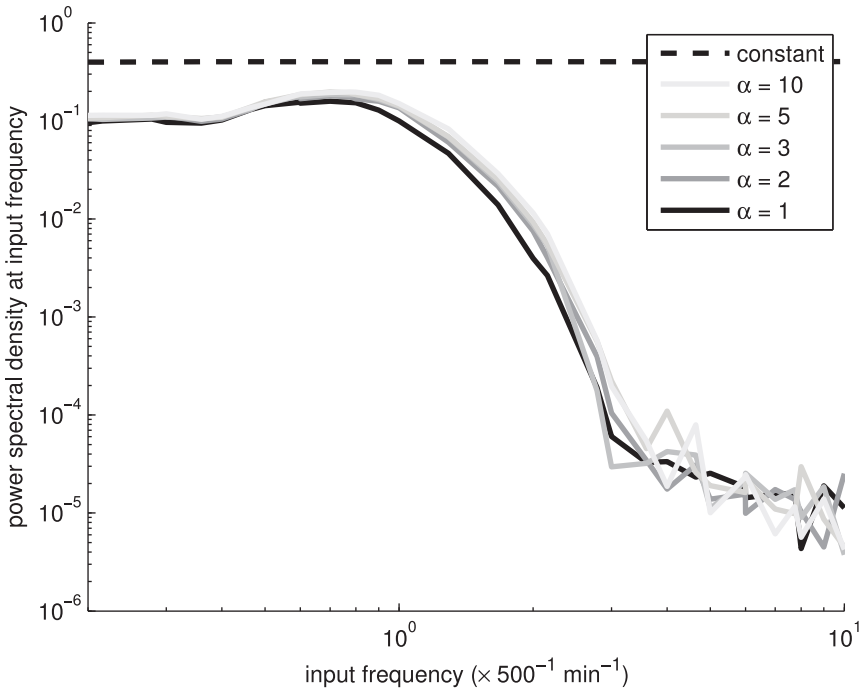


Figure 10. Power spectral density of the frequency filtering motif. Power spectral density of the frequency filter as a function of the input frequency. Different levels of gray denote different shape parameter α'. The simulations were performed with $\mu'\mu^{-1}$ of 0.1. The dashed black line represents the PSD of the input X at the input frequency.
doi:10.1371/journal.pone.0070439.g010

of the frequencies is of the same order of magnitude, more noisy shapes resulting in slightly higher attenuation in the passband. Moreover, the cutoff frequency is affected by changes in the characteristic frequency of the three-gene submotif (Figure 8). We also varied the transcription rate k'_M of the genes in the motif (Figure 11). Again, lower transcription rates, implying more noise in mRNA and protein levels, degrades performance, similarly to when varying α'. The changes in the steepness of the transition band of the filter are more apparent in the former case.

Similarly to the amplitude domain filter, the performance of the frequency domain filter is affected by changes in the transcriptional dynamics of the constituent genes. A transcription process that is less noisy results in a frequency filter with steeper transition bands. Consequently, an efficient frequency domain filter requires limited noise level in transcription, which in the case of low transcript levels can be implemented by a promoter with a sequential initiation process. Interestingly, the cutoff frequency of the filter is also affected by the kinetics of transcription.

As in the case of the amplitude filter, we assessed quantitatively the effects on the output of having different values of α', for each expression ratio of the input gene shown in Figure 11. For $\mu'\mu^{-1}=0.01$, increasing α' from 1 to 2, causes the magnitude of the PSD in the passband to increase by 236.0%. Increasing α' from 1 to 3, causes the PSD to increase by 275.1%. For other values of $\mu'\mu^{-1}$, the differences are smaller as before. In particular, for $\mu'\mu^{-1}=0.05$, these increases are, respectively, 32.5% and 41.9%, while for $\mu'\mu^{-1}=1$, these differences are of the order of 7%.

Discussion

Motivated by recent findings of the relevance of the kinetics of the process of transcription initiation on the dynamics of RNA production in bacteria [6,16], we investigated the functioning of genetic filter motifs as a function of the kinetics of transcription initiation of the constituent genes. We focused on two common filters, namely, an amplitude filter and a frequency filter, as these have several practical applications. One major concern regarding their performance is that most genes in bacteria exhibit very low expression levels. We investigated whether one can utilize the multi-step nature of the process of initiation to compensate for the low copy number noise.

We found that, for realistic parameter values, genetic motifs with stochastic dynamics differ significantly from their deterministic counterparts. Consequently, the latter do not serve as a means to predict the realistic behavior of genetic motifs in live cells. Also, for low expression levels, high noise in the transcripts production significantly degrades the performance of the motifs. The effects of low copy number noise can be compensated by a multi-step (less noisy) transcription process. We suggest that natural motifs with low-expressing constituent genes might employ a multi-step transcription initiation process so as to limit the noise in the mRNA and protein levels, therefore allowing the motif to behave robustly.

The sequence-dependent distribution of transcripts production can have intriguing effects on the behavior of the motifs. These were most prominent in the characteristic frequency of the oscillatory circuit, in which, within a realistic interval of parameter values, it is possible to have a period double that of the one of high

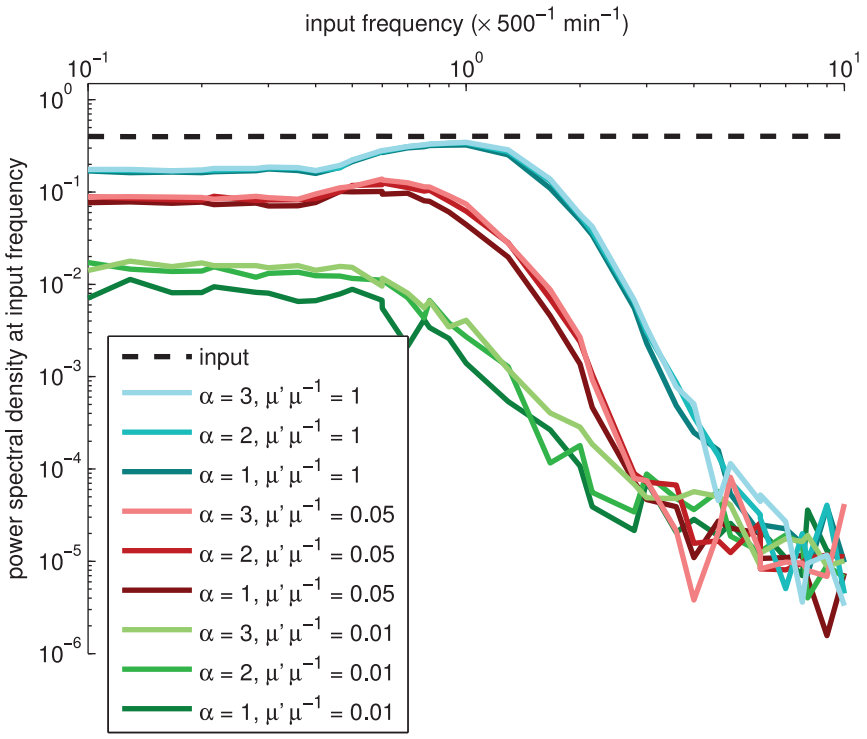


Figure 11. Power spectral density of the frequency filtering motif for various transcription rates. Power spectral density of the frequency filter as a function of the input frequency, for various shapes α' and rates k'_M of transcription. Different levels of brightness denote different shape parameter α'. The simulations were performed with $\mu'\mu^{-1}$ of 1 (cyan), 0.05 (red), and 0.01 (green), in the order of decreasing performance. We also performed simulations with $\mu'\mu^{-1}$ of 0.5, 0.2, 0.1, and 0.02 (not shown) to assert that the changes are generally nonlinear and more drastic with low mean levels. The dashed black line represents the PSD of the input X at the input frequency.
doi:10.1371/journal.pone.0070439.g011

mean levels. Importantly, in both motifs studied, the cutoffs that separate the different regimes of operation of the filters were found to be tunable. The effects of changing the kinetics of transcription initiation were found to be slight, partly masked by the noise, but non-negligible.

It is known that changes in the kinetics of the sequential process of transcription initiation affect the dynamics of mRNA abundances of individual genes [16,17]. Here, we provided tentative evidence that these changes affect the behavior of genetic motifs as well. This is of relevance, since both the number and the kinetics of these steps are dependent of the promoter sequence and transcription factors alone, i.e., are independent of the protein coding region. Due to this, we hypothesize that it is possible to alter the kinetics of a genetic circuit significantly by replacing the promoter region of the constituent genes, without the need of altering the protein under their control. Further, we hypothesize that changes in the promoter sequence of the constituent genes of motifs constitutes a significant degree of freedom in their evolutionary process in natural organisms.

Author Contributions

Conceived and designed the experiments: ASR HT AH. Performed the experiments: ASR HT AH. Analyzed the data: ASR HT AH. Wrote the paper: ASR HT AH OYH.

References

1. Wolf DM, Arkin AP (2003) Motifs, modules and games in bacteria. Curr Opin Microbiol 6: 125–134.
2. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329: 533–538.
3. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci USA 99: 9697–9702.
4. McClure WR (1985) Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem 54: 171–204.
5. Lutz R, Lozinski T, Ellinger T, Bujard H (2001) Dissecting the functional program of *Escherichia coli* promoters: The combined mode of action of lac repressor and arac activator. Nucl Acids Res 29: 3873–3881.
6. Kandhavelu M, Mannerstrom H, Gupta A, Hakkinen A, Lloyd-Price J, et al. (2011) *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 5: 149.
7. Muthukrishnan AB, Kandhavelu M, Lloyd-Price J, Kudasov F, Chowdhury S, et al. (2012) Dynamics of transcription driven by the *tetA* promoter, one event at a time, in live *Escherichia coli* cells. Nucleic Acids Res 40: 8472–8483.
8. Pedraza JM, Paulsson J (2008) Effects of molecular memory and bursting on fluctuations in gene expression. Science 319: 339–343.
9. Ribeiro AS, Zhu R, Kauffman SA (2006) A general modeling strategy for gene regulatory networks with stochastic dynamics. J Comp Biol 13: 1630–1639.
10. Roussel MR, Zhu R (2006) Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression title. Phys Biol 3: 274–284.
11. Ribeiro AS, Lloyd-Price J (2007) Sgn sim, a stochastic genetic networks simulator. Bioinf 23: 777–779.
12. deHaseth PL, Zupancic ML, Record MT Jr (1998) Rna polymerase-promoter interactions: The comings and goings of rna polymerase. J Bacteriol 180: 3019–3025.
13. Xia K, Manning M, Hesham H, Lin Q, Bystroff C, et al. (2007) Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. Proc Natl Acad Sci USA 104: 17329–17334.
14. Dyson S, Gurdon JB (1998) The interpretation of position in a morphogen gradient as revealed by occupancy of activin receptors. Cell 93: 557–568.
15. Loinger A, Biham O (2007) Stochastic simulations of the repressilator circuit. Phys Rev E 76: 051917.
16. Kandhavelu M, Häkkinen A, Yli-Harja O, Ribeiro AS (2012) Single-molecule dynamics of transcription of the lar promoter. Phys Biol 9: 026004.
17. Ribeiro AS, Hakkinen A, Mannerstrom H, Lloyd-Price J, Yli-Harja O (2010) Effects of the promoter open complex formation on gene expression dynamics. Phys Rev E 81: 011912.

PUBLICATION II

J Lloyd-Price, **H Tran**, AS Ribeiro (2014), "Dynamics of small genetic circuits subject to stochastic partitioning in cell division", *Journal of Theoretical Biology*, 356, pg. 11-19. doi: 10.1016/j.jtbi.2014.04.018.

Journal of Theoretical Biology 356 (2014) 11–19

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Dynamics of small genetic circuits subject to stochastic partitioning in cell division

Jason Lloyd-Price, Huy Tran, Andre S. Ribeiro *

Laboratory of Biosystem Dynamics, Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, PO Box 527, FI-33101 Tampere, Finland

HIGHLIGHTS

- We study effects of partitioning errors on the dynamics of genetic circuits.
- Effects of partitioning errors differ widely with network topology and behavior.
- In switches, errors reduce the phenotype distribution's variance across generations.
- The synchrony of a population with clocks is robust to the majority of errors.
- Errors produce qualitatively different effects than noise in gene expression.

ARTICLE INFO

Article history:
Received 21 October 2013
Received in revised form
20 February 2014
Accepted 15 April 2014
Available online 23 April 2014

Keywords:
Partitioning errors
Toggle switch
Repressilator
Cell-to-cell diversity
Synchrony

ABSTRACT

In prokaryotes, partitioning errors during cell division are expected to be a non-negligible source of cell-to-cell diversity in protein numbers. Here, we make use of stochastic simulations to investigate how different degrees of partitioning errors in division affect the cell-to-cell diversity of the dynamics of two genetic circuits, a bistable switch and a clock. First, we find that on average, the stability of the switch decreases with increasing partitioning errors. Despite this, anti-correlations between sister cells, introduced by the partitioning errors, enhance the chances that one of them will remain in the mother cell's state in the next generation, even if the switch is unstable. This reduces the variance of the proportion of phenotypes across generations. In the genetic clock, we find that the robustness of the period decreases with increasing partitioning errors. Nevertheless, the population synchrony is remarkably robust to most errors, only significantly decreasing for the most extreme degree of errors. We conclude that errors in partitioning affect the dynamics of genetic circuits, but the effects are network-dependent and qualitatively different from noise in gene expression.

1. Introduction

Phenotypic diversity is a feature of all cell populations, including monoclonal ones, that significantly affects their survival chances, particularly in fluctuating environments (Kussell and Leibler, 2005; Samoilov et al., 2006). The stochastic nature of the biochemical reactions involved in the dynamics of gene regulatory networks is one well-known contributing source of phenotypic diversity (Kaern et al., 2005; McAdams and Arkin, 1999).

Recently, it has been recognized that the partitioning of plasmids, RNAs, proteins and other macromolecules during cell division can also be a non-negligible source of phenotypic diversity (Huh and Paulsson, 2011a, 2011b; Lloyd-Price et al., 2012). Similar to noise in gene expression, this source generates diversity that can propagate through reaction networks to high-copy number components, even in organisms with a morphologically symmetric division process, such as *Escherichia coli* (Huh and Paulsson, 2011a). After establishing a mathematical framework with which to characterize this source of noise (Huh and Paulsson, 2011b), it was shown that the random errors in partitioning result in cell-to-cell diversity in RNA and protein numbers that is difficult to distinguish from the diversity arising from gene expression noise, when observing cell populations at a single time moment (Huh and Paulsson, 2011a). Nevertheless, while noise in gene expression continuously generates diversity, noise from partitioning only occurs sparsely, when cells divide. Thus, the effects of these two sources should be readily distinguishable from a temporal perspective (Lloyd-Price et al., 2012). So far, it is unknown how these two sources of noise differ in regards to their effects on the dynamics of genetic circuits.

* Corresponding author. Tel.: +358331153928; fax: +358331154989.
E-mail addresses: jason.lloyd-price@tut.fi (J. Lloyd-Price), huy.tran@tut.fi (H. Tran), andre.ribeiro@tut.fi (A.S. Ribeiro).

http://dx.doi.org/10.1016/j.jtbi.2014.04.018
0022-5193/

Most cellular processes are regulated by small genetic networks, named motifs (Wolf and Arkin, 2003; Alon, 2007). It is conceivable that the noise in the process of partitioning of the products of gene expression affects the cell-to-cell diversity of behaviors of these motifs. Here, we study the effects of errors in partitioning on the behavior of two such motifs, the Toggle Switch (Gardner et al., 2000) and the Repressilator (Elowitz and Leibler, 2000). These two circuits differ widely in behavior. While the former is able to switch between two noisy attractors (Ribeiro et al., 2006; Ribeiro and Kauffman, 2007; Zhu et al., 2007), the latter only has one noisy attractor, a limit cycle (Elowitz and Leibler, 2000; Zhu et al., 2007; Loinger and Biham, 2007). Due to their dynamic properties, these circuits are likely candidates to serve as master regulators of future synthetic genetic circuits. Also, similar circuits have evolved in natural cells to perform similar tasks (Wolf and Arkin, 2002; Arkin et al., 1998; Lahav et al., 2004; Nelson et al., 2004). Thus, understanding the effects of random partitioning of RNA and proteins in cell division on the dynamics of these two circuits may aid in understanding how cells maintain robust phenotypes across generations.

The Toggle Switch is a two-gene motif, where each gene expresses a transcription factor that represses the expression of the other gene. As this circuit has two noisy attractors (Gardner et al., 2000; Arkin et al., 1998), it can store one bit of information. It can thus be used to make decisions (Arkin et al., 1998), or to store the results of one (Wolf and Arkin, 2003). The level of gene expression noise determines the frequency at which the Toggle Switch changes between its noisy attractors (Loinger et al., 2007; Potapov et al., 2011). A well-studied Toggle Switch is the "λ-switch", a decision circuit of the λ phage (Arkin et al., 1998), which determines whether an infecting phage will lyse the cell or, instead, integrate itself into the bacterial genome, forming a lysogen. The lytic cycle can be activated in lysogens either stochastically (Neubauer and Calef, 1970), or due to environmental cues such as irradiation by UV light (Baluch and Sussman, 1978). Meanwhile, the Repressilator is a synthetic three-gene motif which exhibits oscillatory behavior (Elowitz and Leibler, 2000), as each gene represses the next gene in the loop. In the Repressilator, gene expression noise determines the robustness of the period of oscillation (Häkkinen et al., 2013).

We study the effects of errors in partitioning on the behavior of these two circuits, focusing on their ability to 'hold state' (i.e. on the stability of their noisy attractors) across cell lineages, when subject to different partitioning schemes. Namely, we explore a wide range of magnitudes of partitioning errors, since in *E. coli* the process of partitioning of gene expression products ranges from highly symmetric (Di Ventura and Sourjik, 2011) to heavily asymmetric, e.g. due to spatially organized protein production (Montero Llopis et al., 2010). For this, we first examine the switching dynamics of the Toggle Switch in cell lineages. In this context, we further consider two biologically motivated scenarios: the phenotypic diversity in a continuous cell culture, and the population dynamics when one state of the switch is lethal to the cells, as in the case of λ lysogens. We then study the effects of errors in partitioning on the behavior of the Repressilator. Specifically, we study the robustness of the period of oscillations, and the rate of desynchronization across cell lineages of an initially synchronous population.

2. Methods

The models used here contain three main components. The first is the genetic circuit within each cell. The second is cell growth and division, and the last is the partitioning scheme of the proteins and RNA molecules in division. For simulations, we used the SGNS2 stochastic simulator (Lloyd-Price et al., 2012), which utilizes the Stochastic Simulation Algorithm (Gillespie, 1977).

2.1. Stochastic model of gene expression

The model of gene expression, illustrated in Fig. 1A, consists of the following set of reactions (Häkkinen et al., 2013):

$$\mathrm{Pr} \xrightarrow{k_c \times f(R,V)} \mathrm{Pr}_c \tag{1}$$

$$\mathrm{Pr}_c \xrightarrow{k_o} \mathrm{Pr} + \mathrm{M} \tag{2}$$

$$\mathrm{M} \xrightarrow{k_P} \mathrm{M} + \mathrm{P} \tag{3}$$

$$\mathrm{M} \xrightarrow{d_M} \varnothing \tag{4}$$

$$\mathrm{P} \xrightarrow{d_P} \varnothing \tag{5}$$

The model includes transcription (Reactions (1) and (2)), translation (reaction (3)), and degradation of mRNA (M, Reaction (4)) and proteins (P, reaction (5)). Transcription initiation is a two-step process, consisting of the closed and open complex formations (Buc and McClure, 1985; Ribeiro et al., 2010). The free promoter is represented by Pr and the promoter-RNAP complex is represented by Pr_c. Here, the closed complex formation can be repressed by a transcription factor produced by another gene by blocking access to the transcription start site. The repression function is a hill function with hill coefficient 2, as in (Zhu et al., 2007). Specifically, it is

$$f(R,V) = \frac{K_d^2}{\left(\frac{R}{V}\right)^2 + K_d^2} \tag{6}$$

where R is the number of repressor molecules, V is the normalized volume of the cell ranging from 0.5 to 1 over the cell cycle, and K_d is the dissociation constant. This repression function arises when the promoter has two operator sites, and there is strongly cooperative binding between the two repressors which bind there.

2.2. Cell growth and division

Cell division in *E. coli* is remarkably stable, with little variance in division time of sister cells when under optimal growth conditions (squared coefficient of variation of division times ≈ 0.02 (Hoffman and Frank, 1965)). We therefore divide cells according to a fixed doubling time T_D, implying that the population doubles in size synchronously. Cell growth is modeled by increasing V linearly from 0.5 to 1 over the lifetime of the cell.

Each cell division is modeled as an instantaneous process which occurs at regular intervals, wherein the DNA (i.e. the promoter region, Pr) is replicated, and the M and P molecules are randomly partitioned into the daughter cells (see next section). After division, we assume that the promoters in the daughter cells are in the initial state (Pr), since any bound molecules are assumed to have been removed from the DNA by the DNA polymerase during replication (Guptasarma, 1995).

To illustrate the dynamics of the single-gene expression model from the previous section with the growth and division model here, we show several time traces in Fig. 1C, as well as the average behavior. Note that the subtle oscillatory behavior observed in the average behavior is due to the effects of linear cell growth and exponential protein degradation.

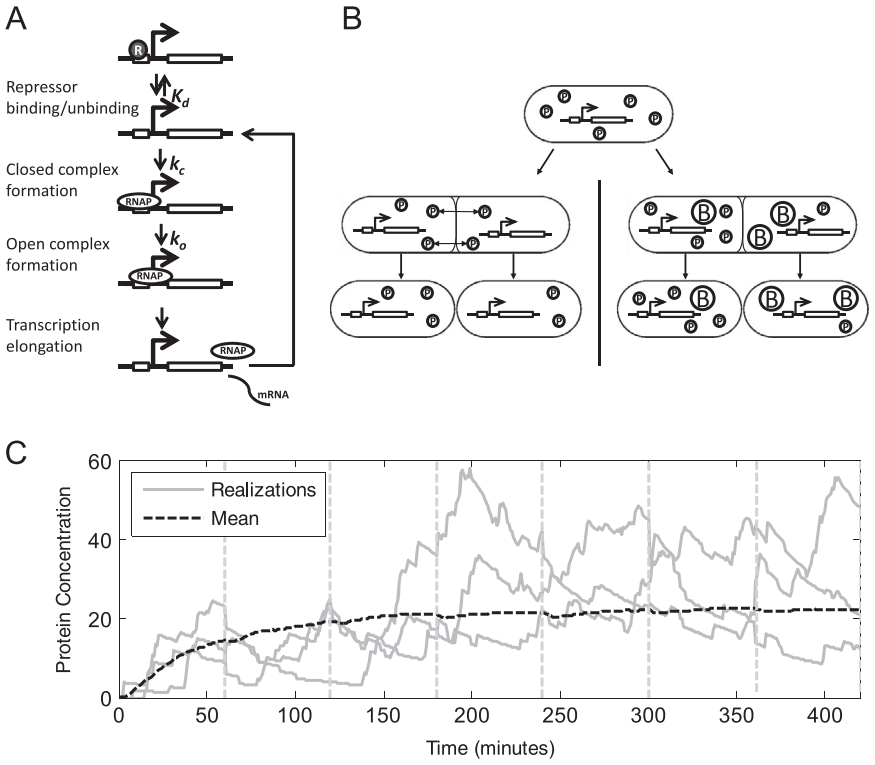


Fig. 1. (A) Illustration of the stochastic model given by reactions (1)–(6). (B) Illustration of the partitioning schemes used, pair formation (left) and random accessible volume (right), which result in lower- and higher-than-binomial variance in partitioning, respectively. (C) Protein concentration (P/V) over time for three independent realizations of a model with a single gene, and the overall mean from 1000 realizations. The vertical dashed lines show division points.

2.3. Molecule partitioning in cell division

The molecule partitioning in cell division is done according to one of three partitioning schemes, which differ in the amount of variance in the molecule numbers that they introduce in division. In (Huh and Paulsson, 2011b), the partitioning error was quantified by $Q_X^2 = CV_L^2 - CV_X^2$, where CV_X^2 and CV_L^2 are the squared coefficients of variation (CV^2, defined as the variance over the squared mean) of the number of molecules in parent cells immediately before division (X), and in daughter cells immediately after division (L), respectively. If the molecules are partitioned independently and randomly, this will result in a binomial distribution in the number of molecules which are inherited by a given daughter, and thus $Q_X^2 = \langle X\rangle^{-1}$(Huh and Paulsson, 2011b). We quantify differences between the variances produced by partitioning schemes by the log of the ratio between the Q_X^2 produced by that scheme and what would be expected from a binomial partitioning, giving $\lg \tilde{Q}$, defined as

$$\lg \tilde{Q} = \lg(\langle X\rangle Q_X^2) \tag{7}$$

If molecules are partitioned independently, i.e. binomially, $\lg \tilde{Q}$ is 0. "Ordered" partitioning schemes resulting in lower variance have $\lg \tilde{Q} < 0$. For this, we use the 'Pair Formation' partitioning scheme (Huh and Paulsson, 2011b), where the segregated molecules first form pairs with probability k. These pairs then are equally divided into the daughter cells while the unpaired molecules are segregated independently (left side of Fig. 1B). It can be shown (see Supplementary Material) that to achieve a given $\lg \tilde{Q} < 0$, k must be set as (from Eq. (8) of Huh and Paulsson, 2011b):

$$k = 1 - 10^{\lg \tilde{Q}} \tag{8}$$

"Disordered" partitioning schemes, resulting in greater variance, will have $\lg \tilde{Q} > 0$. For this, we use the 'Random Accessible Volume' segregation scheme (Huh and Paulsson, 2011b), where large macromolecules in low copy number are independently segregated into the daughter cells. These macromolecules (denoted by B in Fig. 1B) reduce the volume accessible to other molecules, and the error in their partitioning is imparted to the segregated molecules (right side of Fig. 1B). If the same number of B molecules is used to partition all molecules in the cell, this will introduce a correlation in the number of molecules inherited by a given daughter. In all cases, we assessed whether this correlation affected the results by testing both a correlated model and an uncorrelated model, where the B molecules are different for each partitioned molecule. To achieve a given $\lg \tilde{Q} > 0$, we use the following number of B molecules (derived from (Huh and Paulsson, 2011b), see Supplementary Material):

$$B = \frac{\langle X\rangle CV_X^2 + \langle X\rangle - 1}{10^{\lg \tilde{Q}} - 1} \tag{9}$$

The values of CV_X^2 and $\langle X\rangle$ in (9) were calculated by running a simulation of the model with the binomial partitioning scheme, and extracting the CV^2 and mean of the protein distribution prior to divisions. We verified that Eqs. (8) and (9) produce the desired values of $\lg \tilde{Q}$. Supplementary Fig. S1 shows a good correspondence between the input $\lg \tilde{Q}$ and the value determined by simulation.

Finally, to test the behavior of the model in the limit of disordered partitioning, we use an all-or-nothing scheme, where one daughter always receives all molecules, while the other receives none. The $\lg \tilde{Q}$ for this scheme is labeled as "max" in the figures. As might be expected, greater variance in partitioning (higher $\lg \tilde{Q}$) increases the CV^2 of the protein concentration P/V

taken over all time (Supplementary Fig. S2), although the mean concentration is unaffected.

2.4. Stability of the toggle switch

We quantify the stability of the noisy attractors of the Toggle Switch by τ_s, the mean time for this 2-gene network to change from one noisy attractor to the other, similar to (Potapov et al., 2011). Switching points are defined as the moments when the sign of the difference between the two protein concentrations differs from the previous moment. However, a large amount of small switching intervals are generated when the system is at the border between the two noisy attractors. We discount any intervals shorter than 40 min, so as to only consider 'definitive' switches. To get τ_s, we use a maximum likelihood estimator of the conditioned exponential distribution, i.e. the sample mean subtracted by the threshold of 40 min.

2.5. Continuous culture of cells containing a toggle switch

For the study of the Toggle Switch in continuous culture, we use an abstracted model, so as to support the simulation of thousands of cells for many generations. The number of cells in each state is represented by c_1 and c_2, which evolve according to the following reactions:

$$p_i \xrightarrow{1/\tau_s} p_{3-i} \tag{10}$$

$$p_i \xrightarrow{\text{divide}} p_i^+ + p_i^- \tag{11}$$

$$p_i^+ \xrightarrow{k_\infty} p_i(\tau_P) \tag{12}$$

$$p_i^- \xrightarrow{k_\infty} p_i \tag{13}$$

$$p_i^- \xrightarrow{B_P \times k_\infty} p_{3-i} \tag{14}$$

Cells switch stochastically between states with mean time τ_s (reaction (10)). Cells divide synchronously every hour into two daughters, one with more molecules, p_i^+, and one with less, p_i^- (reaction (11)). The sister cell inheriting more molecules is protected from switching for a given amount of time, proportional to the level of bias in partitioning (B_P, ranging from 0 to 1), denoted in reaction (12) by the fixed time delay τ_P. Here, we set τ_P to $B_P \times T_D$. Meanwhile, the sister cells inheriting less molecules become more prone to switch to the other state with increasing B_P (reactions (13) and (14)). In the above reactions, k_∞ is an arbitrarily fast rate. In order to maintain the number of cells stable, exactly half of each type of cells, p_i^+ and p_i^-, are removed from the system after each division event.

3. Results

We study the effects of partitioning errors in the dynamics across cell generations of, first, the Toggle Switch, and second, the Repressilator. In both cases, we explore a wide range of partitioning error rates per division, given the known diversity of the partitioning schemes of proteins and plasmids in bacteria (Huh and Paulsson, 2011a, 2011b). For that, and by using different partitioning schemes, we vary $\lg \tilde{Q}$ from -1 (corresponding to highly symmetric partitioning) to 0 (binomial), up to the maximum allowed by the mean protein number in the mother cell at the moment of division (here labeled 'max').

3.1. Toggle switch

We first examined the effects of errors in the partitioning of regulatory molecules on the dynamics of the Toggle Switch. This circuit is constructed by duplicating reactions (1)–(5), and controlling the expression of each gene with the protein concentration of the other gene (full model presented in Supplementary Material, reactions (1)–(10)). Unless stated otherwise, we set the model parameters as described in Table 1, which result in a mean protein number before division $\langle X \rangle$ of approximately 20, if unrepressed. We set the dissociation constant, K_d, to a value that makes the noisy attractors stable ($K_d = \langle X \rangle / 3$) by maintaining the protein numbers of the repressed gene close to zero at all times. We simulated the system for 2×10^8 s, sampling every 2 min, for each $\lg \tilde{Q}$ tested. After each division, only one daughter of each lineage was simulated (randomly selected), to avoid exponential population growth.

In Fig. 2, we show the stability of the Toggle Switch for different levels of error in partitioning (black lines with crosses). As expected, increasing the error in partitioning from $\lg \tilde{Q} = 0$ to max destabilizes the switch on average, with the mean switching

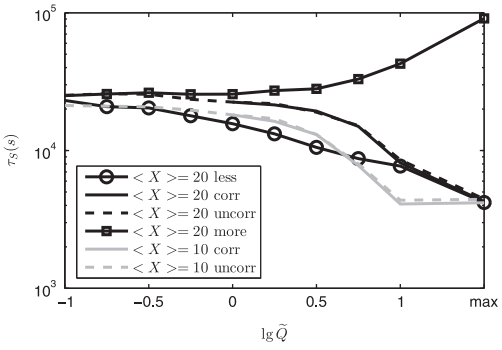


Fig. 2. Mean switching interval (τ_s) of the Toggle Switch for different levels of error in partitioning, two mean protein levels, and correlated/uncorrelated disordered partitioning schemes. For $\langle X \rangle = 20$, switching intervals are also shown for the lineage which always inherits more/less molecules. Data is from one 2×10^8 s simulation for each data point.

Table 1
Parameters used in the model, unless stated otherwise.

Parameter	Description	Value	Source
T_D	Doubling time	3600 s	Yu et al. (2006)
k_c	Closed complex formation (maximum)	1/300 s^{-1}	Kandhavelu et al. (2011)
k_o	Open complex formation	1/300 s^{-1}	Kandhavelu et al. (2011)
d_M	mRNA degradation rate	1/200 s^{-1}	Bernstein et al. (2002)
k_P	Translation rate	3/200 s^{-1}	Yu et al. (2006)
d_P	Protein degradation rate[a]	1/10,000 s^{-1}	Taniguchi et al. (2010)

[a] The degradation rate was set to match a mean protein number of ~20 molecules (Taniguchi et al., 2010), assuming no regulation and given the values of the other parameters.

time decreasing by a factor of ~4. Interestingly, the stabilizing effect from partitioning schemes with less variance than binomial does not seem to be as strong as the destabilizing effect from high-variance partitioning. We suspected that this may be due to the already-low variance introduced due to binomial partitioning with $\langle X\rangle = 20$. We therefore halved the translation rate k_P to 1.5/200 s^{-1} so as to reduce the $\langle X\rangle$ to 10 (gray line). This increases the error in partitioning in the binomial case, and should therefore increase the impact of error correction. Even after this change, however, the stabilizing effects of the lower-variance partitioning schemes appear to be minimal. Note that for the $\langle X\rangle = 10$ case, the maximum value of $\lg \tilde{Q}$ is 1, and thus the mean switching time for $\lg \tilde{Q} = 1$ and $\lg \tilde{Q} = \max$ is the same. Finally, we did not observe any significant difference in the stability of the switch when using correlated and uncorrelated disordered partitioning (Fig. 2).

With high errors in partitioning, we noted one interesting phenomenon. Although the added variance destabilizes the Toggle Switch on average, there are many instances where one daughter cell inherits most of the proteins of the gene that was 'ON' in the mother cell at the moment of division. This generates a transient time during which the probability of switching state in that daughter cell is much smaller than otherwise. As such, high errors in partitioning can be a source of robustness of the states of the circuit in some cells, at the cost of loss of robustness in the sister cells. To show this, we simulated the lineage which inherits more molecules from the parent (black line with squares in Fig. 2). For these cells, an increased stability is observed for the high $\lg \tilde{Q}$ cases. Conversely, the lineage which inherits less molecules exhibits reduced stability (black line with circles in Fig. 2). Thus, high-variance partitioning of the proteins of the Toggle Switch leads to the splitting of the population into sub-populations of cells that differ in the degree of stability in their noisy attractors at birth. This has a far from straightforward effect on the phenotypic distribution of the cell population.

To study this, we constructed an abstracted model of a population of cells, each containing a Toggle Switch (see Methods). High-variance partitioning was modeled by protecting the daughter cells which inherit more molecules from switching, and destabilizing the daughter cells which inherit less, by increasing the probability that they change state after division (see Methods). For simplicity, we assume that the partitioning is biased, i.e. one cell always inherits significantly more than the other. We set the mean switching time when there is no error in partitioning to the measured time in Fig. 2 for $\lg \tilde{Q} = -1$ and $\langle X\rangle = 20$, i.e. $\tau_S = 2.5\times 10^4 s$. We simulated this model with 1000 cells for 10^8 s and recorded the fraction of cells in one of the two states at each time moment. The variance-to-mean ratio (VMR) of this number is shown in Fig. 3 for different levels of bias in partitioning B_P.

If the phenotype of each cell is randomly, independently and unbiasedly chosen, the phenotypes should follow a binomial distribution with $p=0.5$. This distribution has a VMR of $1-p=0.5$, which is observed for the lower biases in partitioning in Fig. 3. Increasing the variance in partitioning (by increasing the bias in partitioning) has the counterintuitive effect of reducing the variance of the phenotype distribution. In other words, while the frequency of the fluctuations of this distribution is faster, the amplitude of the fluctuations is smaller, and thus the distribution is less broad over time. In the limit of fully biased partitioning ($B_P=1$), the VMR decreases to 1/3, independent of the population size and the switching rate (see Supplementary Material). This lower limit on the VMR is due to the combined effect of the randomization of the states of half of the population after each division (p_i^-) and the noise arising from previous events, namely divisions and switches between noisy attractors. Note that these

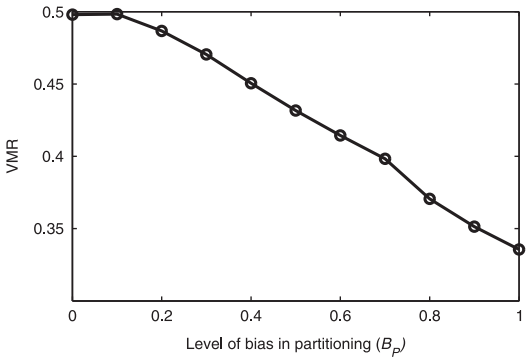


Fig. 3. Variance-to-mean ratio (VMR) of the phenotype distribution in a cell population with Toggle Switches in each cell, as a function of the bias in partitioning, B_P. Data is from populations of 1000 cells simulated for 10^8 s.

levels of VMR would not be possible to reach without anti-correlations in protein numbers between sister cells (which do not arise from noise in gene expression).

It is possible to envision realistic scenarios in which reduced variance is advantageous. Assume that, under certain conditions, one of the two noisy attractors is lethal for example. In this case, high-variance partitioning may protect a population from extermination. To show this, we simulated a population of cells containing a switch that either produces a maintenance protein or a protein that leads to the lethal noisy attractor (we refer to this protein as being, in that sense, 'lethal').

For these simulations, we used the model of the Toggle Switch at the molecular level (built from reactions (1)–(5)), along with one extra condition: if the lethal protein exceeds a threshold, here set to eight molecules, the cell dies (full model presented in Supplementary Material, reactions (1)–(10)). The simulations were initialized with one cell in the nonlethal noisy attractor (25 maintenance proteins, and none of the lethal ones). Finally, in contrast to the switch in Fig. 2, we tripled the rate of the closed complex formation for the gene controlling the lethal state and weakened the repression strength to $K_d=14$ for both genes, so as to mimic a lysogenic cell under stress, e.g. due to UV irradiation (Baluch and Sussman, 1978). From the simulations, we obtained the probability that the resulting population survived the stress (here lasting 9.5 generations, or 30,600 s). The results are shown in Fig. 4 for both correlated and uncorrelated disordered partitioning schemes.

Fig. 4A shows that high-variance partitioning increases the chance of survival of a small population of cells. In particular, the survival chance of each cell increased by ~1.3 fold in the correlated case, despite the increased rate at which the switch changes state on average (Fig. 2), and by ~1.7 fold in the uncorrelated case. This increase in the latter case is due to the reduced chance that the cell inheriting the maintenance protein also inherits most lethal proteins. This strategy comes at a cost: the mean number of cells in the surviving population decreases with increasing variance in partitioning (Fig. 4B). In this case, the mean drops from 3.3 cells to 1.5 cells when increasing $\lg \tilde{Q}$ from 0 to max in both the correlated and uncorrelated schemes. Interestingly, the increased survival probability becomes apparent for intermediate values of $\lg \tilde{Q}$, without incurring a large loss in the mean surviving population size (mean of 2.6 and 2.7 for $\lg \tilde{Q} = 1$ cases with correlated and uncorrelated partitioning, respectively). No large differences were observed in the survival chances or in the mean number of cells in surviving populations when the variance in partitioning was decreased below binomial.

We note that the survival chance (Fig. 4A) is not monotonic with $\lg \tilde{Q}$, as it decreases slightly for small positive $\lg \tilde{Q}$, for both the correlated and uncorrelated cases. We expect that this is due

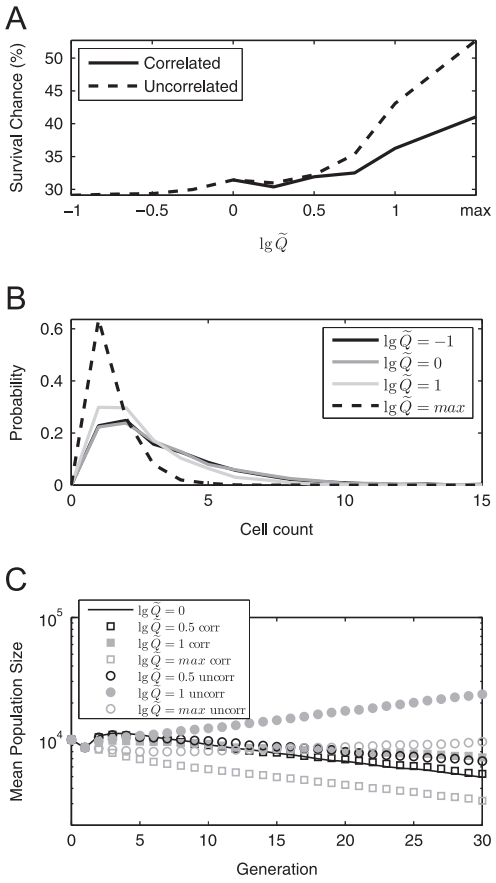


Fig. 4. (A) Survival chance of cell populations for different levels of error in partitioning and correlated/uncorrelated disordered partitioning schemes. (B) Distribution of the number cells in surviving population. (C) Mean population size over time for different partitioning schemes. Corr/uncorr refer to the correlated and uncorrelated disordered partitioning schemes. All data is from 10,000 simulations starting from one initial cell.

to the weakness of the aforementioned protection effect for small lg $\tilde{Q}$. Without it, cells inheriting the majority of the proteins would be less stable, since they also inherit the majority of the lethal proteins. If this is the case, this effect should be exacerbated in the correlated case, where one daughter cell always falls into this category. As expected, Fig. 4A shows a slightly larger drop in survival chance for the correlated case compared to the uncorrelated case.

The above simulations assumed small initial populations. Increasing the initial size of a population increases its survival chances, since the probability that all cells die decreases. Nevertheless, the partitioning scheme affects the mean size of the population over time. To exemplify this, the number of cells in a population starting with 10,000 cells is shown in Fig. 4C for both the correlated and uncorrelated disordered partitioning schemes. After a transient of ~5 generations, all populations enter an exponential phase (linear on the log-linear plot). The slope of this line informs on how quickly the population is growing/shrinking in the different conditions. Table 2 shows the change in population size over 10 generations in this exponential phase. In agreement with the above results, all the disordered partitioning schemes improved populations' numbers, with the uncorrelated disordered schemes providing the best improvements, to the point where the numbers even grow. In addition, the point at which the growth rate is maximized is for an intermediate value of lg $\tilde{Q}$, consistent with the above observation that the benefits of disordered partitioning appear before the drawbacks.

Table 2
Change in population size over 10 generations in exponential phase, for the independent partitioning scheme (lg $\tilde{Q}=0$) and two disordered partitioning schemes with varying lg $\tilde{Q}$. Data is from a least-squares linear fit to generations 15–30 in Fig. 4C.

lg $\tilde{Q}$	Uncorrelated	Correlated
0	−27%	
0.5	−16%	−25%
1	+37%	−12%
Max	+8%	−24%

The increase of survival rates with high-variance partitioning should apply at least so long as the protein lifetime is on the order of, or longer than the cell doubling time, to ensure that the added stability of the state of the cells is not lost during the cell cycle. Nevertheless, we tested several other parameter sets, including protein degradation faster than cell division. Qualitatively, the results hold, except for extreme parameter values. For example, high-variance partitioning decreases the population survival chance when the lethal protein's interaction with the maintenance gene's promoter is extremely cooperative.

We also tested whether resetting the promoter state at division (see Methods) affected the above results. The only significant effect was a reduction by ~1 to 5% in the survival chances in the lethal noisy attractor scenario. Finally, we tested whether, in this context, the use of a hill function (Eq. (6)) is equivalent to using elementary reactions by measuring the stability of Toggle Switches in both cases, as recent studies show that these two modeling strategies can exhibit significant differences (Zhu et al., 2007; Thomas et al., 2012). For this test, we did not allow cells to divide. We found no significant differences between the two models.

3.2. Repressilator

We next studied how errors in partitioning affect the behavior of the Repressilator across cell generations. The circuit is constructed by triplicating reactions (1)–(5), and controlling the production of each gene (labeled A, B and C) with the protein concentration of the previous gene (the full model is presented in the Supplementary Material, reactions (11)–(25)). We set the model parameters to those in Table 1, and K_d to five proteins. We simulated the system for 10^7 s, sampling every minute, for each lg $\tilde{Q}$ tested, and quantified the period by the zeros of the autocorrelation function of the concentration of one gene's product. While the mean of the period (~375 min) does not differ between conditions, the variance does, as it increases with increasing variance in partitioning for both the correlated and uncorrelated disordered partitioning schemes (Fig. 5). The uncorrelated case exhibits lower variance than the correlated case, since there is a higher probability of transmitting some of the phase information to the daughter cells. No significant change in either the mean or the variability of the periods was observed for lower-variance partitioning schemes.

With less robust periods in the higher-variance partitioning case, we predict that an initially synchronized population of cells will desynchronize faster. To test this, we simulated the growth of 500 initially synchronous cells, and measured the mean protein concentration of each protein within the entire population at each moment (an example is shown in Fig. 6A for binomial partitioning). Note that the mean overall protein concentration exhibits a small oscillation, and does not converge to a constant value because of the combined effects of protein degradation and the linear increase of the cell volume over the cell cycle (same as in Fig. 1C). To quantify the loss of synchrony, we measured the

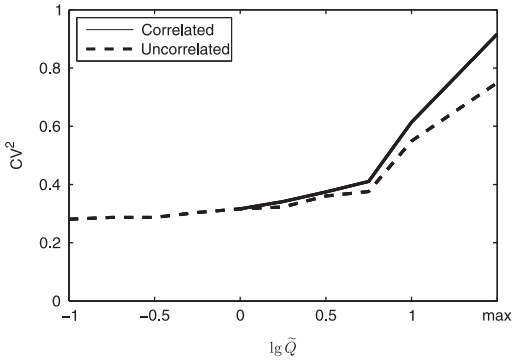


Fig. 5. CV^2 of the period of oscillation of the Repressilator, subject to differing levels of errors in partitioning and correlation in partitioning. Data is from one 10^7 s simulation for each data point.

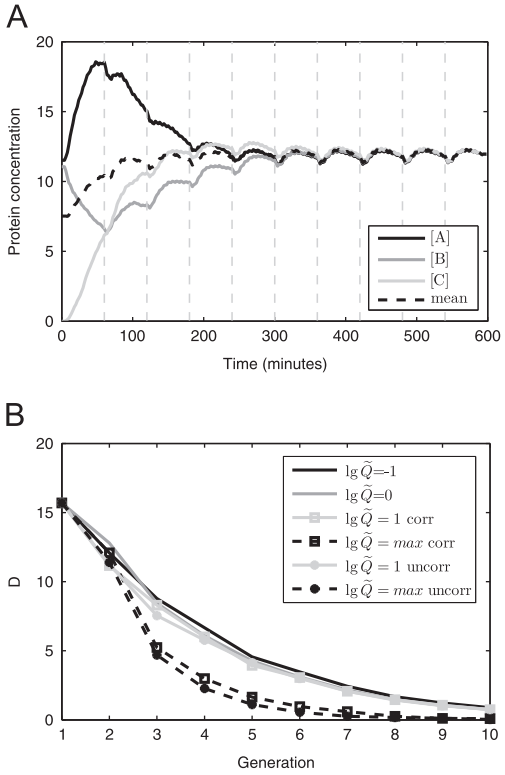


Fig. 6. (A) Mean protein concentrations of each protein in the Repressilator (solid lines), when subject to independent partitioning of molecules at division (i.e. $\lg \tilde{Q} = 0$). The overall mean protein concentration is also shown (dashed line). Vertical dashed lines indicate division points. Data is from a population starting with 500 cells in the same state ([A]=12, [B]=12, [C]=0). (B) Difference between the mean protein concentration of each protein and the mean overall protein concentration, averaged over each generation (D), for differing levels of error in partitioning and different partitioning schemes. Corr/uncorr refer to the correlated and uncorrelated disordered partitioning schemes. Data is from a population of 500 initial cells for each line.

absolute difference between the mean protein concentration of each protein (solid lines in Fig. 6A) and the mean overall protein concentration of the three genes (dashed line in Fig. 6A). We define D as the average of this value over each generation. If cells are in perfect synchrony, as in the beginning of the simulations, D will be large. Random fluctuations and random partitioning will reduce D, since the mean protein concentrations will converge to the overall mean, indicating that the population is less synchronous. The values of D across generations for different levels of error in partitioning are shown in Fig. 6B.

For correlated partitioning, Fig. 6B shows that the synchronization of the Repressilator is remarkably robust to partitioning errors, as there is only a slight change in the rate of desynchronization (i.e. how quickly D approaches 0) for $-1 \leq \lg \tilde{Q} \leq 1$. This is despite the observed increase in the noise in the lengths of the periods at $\lg \tilde{Q} = 1$ (Fig. 5).

When all-or-nothing partitioning is applied, the cells desynchronize rapidly, with its effects being visible already in the third generation. This is explained as follows: in the second generation, the cell receiving all proteins will oscillate more robustly than in the cases with less variance in partitioning, while the cell that received no proteins remains in an undetermined phase for most of its lifetime.

Finally, despite the decreased variance in the periods (Fig. 5), uncorrelated partitioning decreases the synchrony of the population slightly faster than correlated partitioning. This is due to the highly synchronous subpopulation of cells inheriting the majority of molecules in the correlated case. Overall, our results show that the synchrony of the Repressilator is not affected by moderate partitioning errors, but decreases for more extreme errors.

When relaxing the assumption that the promoter state resets during division, we found no significant effect on either the noise in the oscillation period or the rate of desynchronization. Further, we looked for phase-locking effects when the period of the Repressilator was near an integer multiple of the cell cycle. We were, however, not able to observe phase-locking in our simulations, perhaps due to the magnitude of the noise in the period's length, even in the lowest $\lg \tilde{Q}$ case (Fig. 5). This is in agreement with a lack of change in behavior when changing the ratio between the mean period and the mean length of the cell cycle (data not shown).

4. Conclusions and discussion

From stochastic simulations, we studied the effects of errors in partitioning on the behavior across cell generations of two genetic circuits, the Toggle Switch and the Repressilator. Knowledge of the effects is necessary not only to understand their kinetics in long scales across cell lineages but also in the context of synthetic biology, where partitioning schemes may potentially be used as regulatory mechanisms. The results suggest that genetic circuits are far from immune to this source of cell-to-cell variability, although the extent to which they are affected is heavily network-dependent.

We found that increasing partitioning errors not only decreases the stability of the noisy attractors of the Toggle Switch but also decreases the variance of the phenotypic distribution of the population below that of a binomial distribution. Notably, while the former result could be obtained by increasing noise in gene expression, the latter could not. This effect was due to the anti-correlation between the protein numbers inherited by sister cells, which is enhanced with high-variance partitioning and increases the stability of the inherited state of one cell at the cost of the stability of its sister cell. In this context, we considered an extreme case, by assuming that one of the states of a switch led to cell death. We found that finite cell populations, originally heading to extinction when employing binomial partitioning of components, increase their survival chances and may even grow in numbers over time, if they employ disordered partitioning schemes instead. This is due to the increased chances that, following each division, one of the daughter cells will remain in the non-lethal state.

The Repressilator was found to be more robust than the switch to increasing partitioning errors. Though the variance in the periods increased, consistent with an increase in noise in gene expression, the synchrony of a population was remarkably robust

to this increase. Only the strongest errors in partitioning (i.e., the all-or-nothing partitioning scheme) were able to significantly affect the degree of synchrony of the population. This is interesting, in that the function of circuits to track time is commonly the maintenance of synchrony between cells in a population. Further, even in the 'all-or-nothing' scenario, we expect that a simple cell to cell communication system will suffice to quickly resynchronize the clocks of sister cells following division.

We also studied the effects of correlations in the partitioning errors of the different proteins of the two circuits above. Such correlations are expected if the division process is morphologically asymmetric. We found that these correlations have no effect on the stability of the genetic switches and only slightly increase the rate of desynchronization of Repressilators in sister cells.

In all conditions tested, we did not observe any significant effect in the behavior of cells when using ordered partitioning schemes, when compared to binomial partitioning. This, combined with the fact that the implementation of such schemes is likely energy-consuming (due to requiring error correction (Huh and Paulsson, 2011b)), may explain why its use, while not absent (Di Ventura and Sourjik, 2011), is seemingly rare in nature, at least for low-to-medium-copy components such as RNA and regulatory proteins.

Non-binomial partitioning errors can arise in a number of different ways. Here, we have used pair formation to achieve $\lg \tilde{Q} < 0$, and random accessible volume to achieve $\lg \tilde{Q} > 0$. Though we believe that Q_X^2, and thus $\lg \tilde{Q}$, captures the most important aspect of the partitioning schemes (Huh and Paulsson, 2011b), other partitioning schemes could result in similar values of $\lg \tilde{Q}$, but lead to different behaviors. For example, correlated and uncorrelated disordered partitioning schemes produce slightly different effects for the same $\lg \tilde{Q}$. As more complex networks are analyzed in this context in the future, it will likely become necessary to characterize the various possible partitioning schemes more comprehensively.

One sort of error in partitioning not considered here occurs when the circuit is expressed from a multi-copy plasmid (Gardner et al., 2000; Elowitz and Leibler, 2000), whose numbers are also partitioned stochastically in division (Reyes-Lamothe et al., 2013). We expect that errors in plasmid partitioning will affect the dynamics of the circuits they code for in a manner similar to the correlated disordered partitioning schemes employed here, since the same partitioning error eventually affects all proteins. However, the impact of the division event will be delayed and diluted over time by noise in plasmid replication and gene expression. Other extrinsic noise sources not considered here include cell to cell diversity in RNA polymerase and ribosome numbers, among others. Future studies may assist in quantifying the contribution of these sources on the temporal distributions of cellular phenotypes.

The results above show that the effects of errors in partitioning differ widely from those of noise in gene expression. The differences arise primarily from the unavoidable anti-correlation in the numbers of molecules inherited by sister cells, whereas noise in gene expression affects all cells in the population independently. In other words, unlike noise in gene expression, the division process forces sister cells to move in opposite directions in the network's state space, starting from the location of the mother cell the moment prior to division.

The qualitative differences in the effects of errors in partitioning in the Toggle Switch and the Repressilator derive from the differences in their long-term behaviors. In the Toggle Switch, with two noisy attractors, division can move one of the daughter cells close to the border of the basin of attraction that the mother cell lied on, while moving the other daughter cell further into the basin. That creates a strong possibility that the former cell switches into the neighbor attractor while the chances that the latter remains in the present attractor are enhanced. In other words, there are increased chances that sister cells will exhibit opposite behaviors. The Repressilator, on the other hand, has only one attractor, a state cycle. Regardless where the daughter cells lie on the state space following division, they will both travel towards the same attractor, thus becoming closer in the state space with time. Thus, in this network, the effects of partitioning are hardly distinguishable from those of noise in gene expression.

From all of the above, it is possible to infer general consequences of errors in partitioning on the dynamics of small genetic circuits. In circuits with only one noisy attractor, the effects of these errors are not expected to differ qualitatively from those of noise in gene expression. Meanwhile, in circuits with more than one noisy attractor, large errors in partitioning enhance the chances for sister cells to begin their lifetime in different noisy attractors, with one sister cell deeper into the mother's basin of attraction and the other jumping way from it. There is a process in natural organisms that exhibits some similarity. Namely, multicellular organisms have stem cells which, in division, produce both a renewed stem cell (i.e. on the same noisy attractor as the mother cell) and a differentiated cell (i.e. on another noisy attractor of the gene regulatory network). It would be of interest to assess in the future the degree to which the many asymmetries in these division events are deliberate.

Finally, it is interesting to note that the effects of noise in gene expression and of errors in partitioning must, in one or more aspects, differ for all networks. This is because, first, their effects at the single gene level differ, in that while noise in gene expression enhances fluctuations at all time points, errors in partitioning occur only at specific, rare moments. Second, at the network level, increased noise in gene expression decreases the stability of all noisy attractors at all times. Meanwhile, partitioning errors promote transitions between attractors at specific points in time, without affecting their stability otherwise. As such, it is reasonable to hypothesize that both of these 'perturbation mechanisms' will be of use in present efforts in Synthetic Biology and are likely to be used for different aims in natural organisms.

Acknowledgments

Work supported by Finnish Funding Agency for Technology and Innovation (ASR), Academy of Finland (HT, ASR), Portuguese Foundation for Science and Technology (ASR) and TUT President's graduate programme (JLP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2014.04.018.

References

Alon, U., 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8, 450–461.

Arkin, A., Ross, J., Mcadams, H.H., 1998. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in phage. Genetics 149, 1633–1648.

Baluch, J., Sussman, R., 1978. Correlation between UV dose requirement for lambda bacteriophage induction and lambda repressor concentration. J. Virol. 26, 595–602.

Bernstein, J.A., Khodursky, A.B., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 99, 9697–9702.

Buc, H., McClure, W.R., 1985. Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry 24, 2712–2723.
Di Ventura, B., Sourjik, V., 2011. Self-organized partitioning of dynamically localized proteins in bacterial cell division. Mol. Syst. Biol. 7, 457.
Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338.
Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in Escherichia coli. Nature 403, 339–342.
Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.
Guptasarma, P., 1995. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of Escherichia coli? Bioessays 17, 987–997.
Häkkinen, A., Tran, H., Yli-Harja, O., Ribeiro, A.S., 2013. Effects of rate-limiting steps in transcription initiation on genetic filter motifs. PLoS One 8, e70439.
Hoffman, H., Frank, M.E., 1965. Synchrony of division in clonal microcolonies of Escherichia coli. J. Bacteriol. 89, 513–517.
Huh, D., Paulsson, J., 2011a. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat. Genet. 43, 95–100.
Huh, D., Paulsson, J., 2011b. Random partitioning of molecules at cell division. Proc. Natl. Acad. Sci. U. S. A. 108, 15004–15009.
Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464.
Kandhavelu, M., Mannerstrom, H., Gupta, A., Häkkinen, A., Lloyd-Price, J., Yli-Harja, O., et al., 2011. 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. 5, 149.
Kussell, E., Leibler, S., 2005. Phenotypic diversity, population growth, and information in fluctuating environments. Science 309, 2075–2078.
Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A.J., Elowitz, M.B., et al., 2004. Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat. Genet. 36, 147–150.
Lloyd-Price, J., Lehtivaara, M., Kandhavelu, M., Chowdhury, S., Muthukrishnan, A.-B., Yli-Harja, O., et al., 2012. Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of Escherichia coli. Mol. Biosyst. 8, 565–571.
Lloyd-Price, J., Gupta, A., Ribeiro, A.S., 2012. SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations. Bioinformatics 28, 3004–3005.
Loinger, A., Biham, O., 2007. Stochastic simulations of the repressilator circuit. Phys. Rev. E. 76, 051917.
Loinger, A., Lipshtat, A., Balaban, N., Biham, O., 2007. Stochastic simulations of genetic switch systems. Phys. Rev. E. 75, 021904.
McAdams, H.H., Arkin, A., 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet. 15, 65–69.
Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., et al., 2010. Spatial organization of the flow of genetic information in bacteria. Nature 466, 77–81.
Nelson, D.E., Ihekwaba, A.E.C., Elliott, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., et al., 2004. Oscillations in NF-kappaB signaling control the dynamics of gene expression. Science 306, 704–708.
Neubauer, Z., Calef, E., 1970. Immunity phase-shift in defective lysogens: non-mutational hereditary change of early regulation of λ prophage. J. Mol. Biol. 51, 1–13.
Potapov, I., Lloyd-Price, J., Yli-Harja, O., Ribeiro, A.S., 2011. Dynamics of a genetic toggle switch at the nucleotide and codon levels. Phys. Rev. E. 84, 031903.
Reyes-Lamothe, R., Tran, T., Meas, D., Lee, L., Li, A.M., Sherratt, D.J., et al., 2013. High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. Nucleic Acids Res. 42, 1042–1051.
Ribeiro, A.S., Kauffman, S.A., 2007. Noisy attractors and ergodic sets in models of gene regulatory networks. J. Theor. Biol. 247, 743–755.
Ribeiro, A.S., Zhu, R., Kauffman, S.A., 2006. A general modeling strategy for gene regulatory networks with stochastic dynamics. J. Comput. Biol. 13, 1630–1639.
Ribeiro, A.S., Häkkinen, A., Mannerstrom, H., Lloyd-Price, J., Yli-Harja, O., 2010. Effects of the promoter open complex formation on gene expression dynamics. Phys. Rev. E. 81, 011912.
Samoilov, M.S., Price, G., Arkin, A.P., 2006. From fluctuations to phenotypes: the physiology of noise. Sci. STKE 2006, re17.
Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., et al., 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538.
Thomas, P., Straube, A.V., Grima, R., 2012. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. BMC Syst. Biol. 6, 39.
Wolf, D.M., Arkin, A.P., 2002. 15 min of fim: control of phase variation in E. coli. Omi. J. Integr. Biol. 6, 91–114.
Wolf, D.M., Arkin, A.P., 2003. Motifs, modules and games in bacteria. Curr. Opin. Microbiol. 6, 125–134.
Yu, J., Xiao, J., Ren, X., Lao, K., Xie, X.S., 2006. Probing gene expression in live cells, one protein molecule at a time. Science 311, 1600–1603.
Zhu, R., Ribeiro, A.S., Salahub, D., Kauffman, S.A., 2007. Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. J. Theor. Biol. 246, 725–745.

Supplement to "Dynamics of small genetic circuits subject to stochastic partitioning in cell division"

Jason Lloyd-Price, Huy Tran, and Andre S. Ribeiro

Parameters for partitioning schemes

The formulas used to adjust the $lg\tilde{Q}$ of a given partitioning scheme are presented here.

The partitioning error for the Pair Formation scheme is given by equation 8 of ref. 6 in the main manuscript (where the probability of evenly partitioning a pair is $p = 1$):

$$Q_X^2 = \frac{1-k}{\langle X \rangle}$$

where k is the fraction of molecules that form pairs. The value of $lg\tilde{Q}$ is given by:

$$lg\tilde{Q} = lg(\langle X \rangle Q_X^2) = \lg(1-k)$$

To achieve a given $lg\tilde{Q}$, we therefore set $k = 1 - 10^{lg\tilde{Q}}$.

The partitioning error for the Random Accessible Volume segregation scheme, from equation 2 of ref. 6 in the main manuscript, is:

$$Q_X^2 = \frac{1-Q_{vol}^2}{\langle X \rangle} + Q_{vol}^2(CV_X^2 + 1)$$

where Q_{vol}^2 is the partitioning error of the accessible volume. The value of Q_{vol}^2 is determined by the number of macromolecules (denoted by B) that reduce the volume accessible to other molecules:

$$Q_{vol}^2 = \frac{1}{\langle B \rangle}$$

The value of $lg\tilde{Q}$ is given by:

$$lg\tilde{Q} = lg(\langle X \rangle Q_X^2) = \lg(1 - Q_{vol}^2 + Q_{vol}^2(CV_X^2 + 1)\langle X \rangle)$$

$$= \lg\left(1 + \frac{(CV_X^2 + 1)\langle X \rangle - 1}{B}\right)$$

To achieve a given $lg\tilde{Q}$, we therefore set $B = \frac{\langle X \rangle CV_X^2 + \langle X \rangle - 1}{10^{lg\tilde{Q}} - 1}$, where the values of CV_X^2 and $\langle X \rangle$ were calculated by simulating a model with the binomial partitioning scheme, and sampling these values immediately before division events. Figure S1 shows that the above formulas produce the desired values of $lg\tilde{Q}$ when applying the different partitioning schemes.

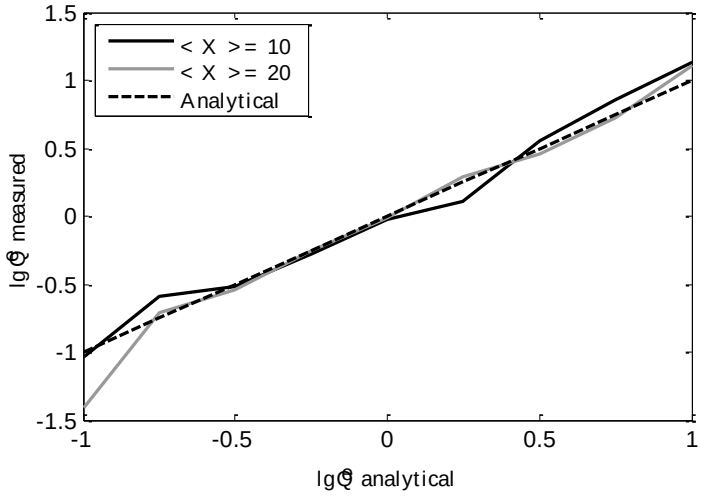


Figure S1: Input lg$\tilde{Q}$ and simulated results after applying the partitioning schemes.

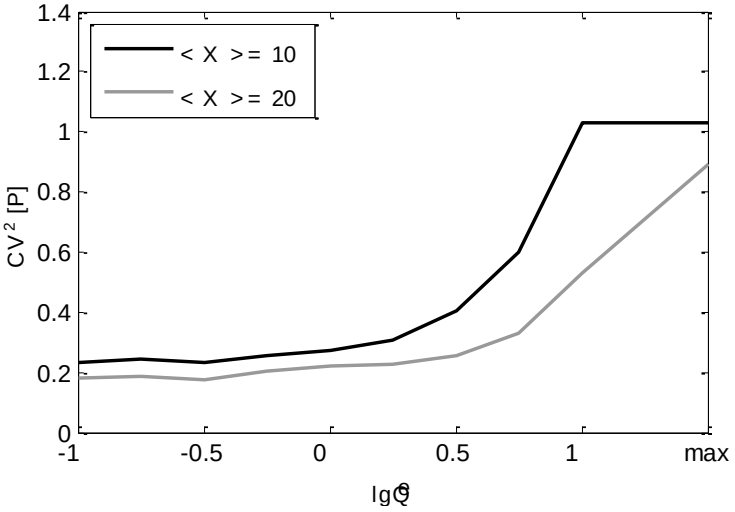


Figure S2: CV^2 of the protein concentration ([P] = P/V), taken over all time with different errors in partitioning in division, for different mean protein levels before division. Data is from a single simulation of length 10^8 s, for each level of partitioning error.

Stochastic model of the Toggle Switch

The Toggle Switch's model comprises two genes A and B, which repress each other via their protein products. The repression function is a hill function, described in equation (6) of the manuscript. Parameters are shown in Table S1.

Parameter	Description	Value
T_D	Doubling Time	3600 s
k_{cA}	Gene A's Closed Complex Formation Rate	1/300 s^{-1}
k_{cB}	Gene B's Closed Complex Formation Rate	1/300 s^{-1}
k_o	Open Complex Formation Rate	1/300 s^{-1}
d_M	mRNA Degradation Rate	1/200 s^{-1}
k_P	Translation Rate	3/200 s^{-1}
d_P	Protein Degradation Rate	1/10000 s^{-1}
K_d	Dissociation Constant	20/3

Table S1: Parameters used in the single lineage simulation of a Toggle Switch. To vary the mean protein level from 20 to 10, the translation rate k_P is halved from 3/200 s^{-1} to 1.5/200 s^{-1}. The dissociation constant K_d, set as <X>/3, is 20/3 or 10/3 respectively. For parameter sources, see Table 1 in the main manuscript.

The model consists of the following set of reactions:

$$\mathrm{Pr}_A \xrightarrow{k_{cA}\times f(P_B,V)} \mathrm{Pr}_{Ac} \quad (1)$$

$$\mathrm{Pr}_{Ac} \xrightarrow{k_o} \mathrm{Pr}_A + \mathrm{M}_A \quad (2)$$

$$\mathrm{M}_A \xrightarrow{k_P} \mathrm{M}_A + \mathrm{P}_A \quad (3)$$

$$\mathrm{M}_A \xrightarrow{d_M} \varnothing \quad (4)$$

$$\mathrm{P}_A \xrightarrow{d_P} \varnothing \quad (5)$$

$$\mathrm{Pr}_B \xrightarrow{k_{cB}\times f(P_A,V)} \mathrm{Pr}_{Bc} \quad (6)$$

$$\mathrm{Pr}_{Bc} \xrightarrow{k_o} \mathrm{Pr}_B + \mathrm{M}_B \quad (7)$$

$$\mathrm{M}_B \xrightarrow{k_P} \mathrm{M}_B + \mathrm{P}_B \quad (8)$$

$$\mathrm{M}_B \xrightarrow{d_M} \varnothing \quad (9)$$

$$\mathrm{P}_B \xrightarrow{d_P} \varnothing \quad (10)$$

In the case where one noisy attractor is 'lethal', the model consists of reactions (1)-(10) with one additional condition: if P_B equals or exceeds 8, the simulation of that cell is immediately ended. Parameters are shown in Table S2.

Parameter	Description	Value
T_D	Doubling Time	3600 s
k_{cA}	Gene A's Closed Complex Formation Rate	1/300 s^{-1}
k_{cB}	Gene B's Closed Complex Formation Rate	3/300 s^{-1}
k_o	Open Complex Formation	1/300 s^{-1}
d_M	mRNA Degradation Rate	1/200 s^{-1}
k_P	Translation Rate	3/200 s^{-1}
d_P	Protein Degradation Rate	1/10000 s^{-1}
K_d	Dissociation Constant	14

Table S2: Parameters used in the Toggle Switch simulation in the case where one noisy attractor is 'lethal'. Gene B, the 'lethal' gene, has the rate of closed complex formation k_{cB} 3 times faster than that of gene A (k_{cA}), the maintenance gene. For parameter sources, see Table 1 in the main manuscript.

Stochastic model of the Repressilator

The model of the Repressilator consists of three genes A, B, and C, which repress each other in a ring. The repression function is a hill function, described in equation (6) of the manuscript. The model parameters are shown in Table S3.

Parameter	Description	Value
T_D	Doubling Time	3600 s
k_c	Closed Complex Formation Rate	1/300 s^{-1}
k_o	Open Complex Formation Rate	1/300 s^{-1}
d_M	mRNA Degradation Rate	1/200 s^{-1}
k_P	Translation Rate	3/200 s^{-1}
d_P	Protein Degradation Rate	1/10000 s^{-1}
K_d	Dissociation Constant	5

Table S3: Parameters used in the Repressilator simulation. For parameter sources, see Table 1 in the main manuscript.

The model consists of the following set of reactions:

$$\mathrm{Pr}_A \xrightarrow{k_c \times f(P_C, V)} \mathrm{Pr}_{Ac} \quad (11)$$

$$\mathrm{Pr}_{Ac} \xrightarrow{k_o} \mathrm{Pr}_A + \mathrm{M}_A \quad (12)$$

$$\mathrm{M}_A \xrightarrow{k_P} \mathrm{M}_A + \mathrm{P}_A \quad (13)$$

$$\mathrm{M}_A \xrightarrow{d_M} \varnothing \quad (14)$$

$$\mathrm{P}_A \xrightarrow{d_P} \varnothing \quad (15)$$

$$\mathrm{Pr}_B \xrightarrow{k_c \times f(P_A, V)} \mathrm{Pr}_{Bc} \quad (16)$$

$$\mathrm{Pr}_{Bc} \xrightarrow{k_o} \mathrm{Pr}_B + \mathrm{M}_B \quad (17)$$

$$\mathrm{M}_B \xrightarrow{k_P} \mathrm{M}_B + \mathrm{P}_B \quad (18)$$

$$\mathrm{M}_B \xrightarrow{d_M} \varnothing \quad (19)$$

$$\mathrm{P}_B \xrightarrow{d_P} \varnothing \quad (20)$$

$$\mathrm{Pr}_C \xrightarrow{k_c \times f(P_B,V)} \mathrm{Pr}_{Cc} \quad (21)$$

$$\mathrm{Pr}_{Cc} \xrightarrow{k_o} \mathrm{Pr}_C + \mathrm{M}_C \quad (22)$$

$$\mathrm{M}_C \xrightarrow{k_P} \mathrm{M}_C + \mathrm{P}_C \quad (23)$$

$$\mathrm{M}_C \xrightarrow{d_M} \varnothing \quad (24)$$

$$\mathrm{P}_C \xrightarrow{d_P} \varnothing \quad (25)$$

Minimum VMR with fully biased partitioning

In the case of fully biased partitioning in the continuous culture model of a population of cells containing Toggle Switches, it is possible to derive the VMR of the phenotype distribution as follows. Let *N* be the number of cells before a division, σ^2 be the variance of the number of cells in state 1 at that time. The variance of the number of cells in state 1 after the division (after reactions (13) and (14) of the manuscript), is then $\sigma^2/4 + Np(1-p)/2$, where p is the probability that one of the cells receiving nothing will end up in state 1 after division, which is 0.5 since the switch is unbiased. We then obtain the variance of the stationary phenotype distribution by setting:

$$\sigma^2 = \frac{\sigma^2}{4} + \frac{Np(1-p)}{2}$$

The VMR is therefore:

$$\frac{\sigma^2}{N/2} = \frac{1}{3}$$

We note that this result does not apply if the population is allowed to grow indefinitely. In this case, the VMR converges to the VMR of a binomial.

PUBLICATION III

H Tran and AS Ribeiro (2013), “Effects of inducer intake kinetics on the dynamics of gene expression”, *Advances in Artificial Life, ECAL 2013: Proceedings of the 12th European Conference on the Synthesis and Simulation of Living Systems*, pg. 1174-1181. Taormina – Sicily, Italy, September 2013. doi: 10.7551/978-0-262-31719-2-ch181.

Effects of inducer intake kinetics on the dynamics of gene expression

Huy Tran and Andre S. Ribeiro
Dept. of Signal Processing, Tampere University of Technology, P.O Box 553, 33101 Tampere, Finland
andre.ribeiro@tut.fi

Abstract

Many genes are only active following the intake of an inducer by the cell, via passive diffusive, positive, or negative feedback intake mechanisms. Based on measurements of the *in vivo* kinetics of intake and subsequent transcription events in *Escherichia coli*, we use stochastic models to investigate how the kinetics of intake affects both transient and near-equilibrium dynamics of gene expression. We find that the intake kinetics affects mean and variability of the transient time to reach the steady state of proteins numbers and also the degree of fluctuations in these numbers. Fluctuations in the extracellular number of inducers affects the variability of protein numbers at steady state in a degree that differs with the intake kinetics. Finally, changing the intake kinetics of an inducer of a genetic switch allows tuning the bias in the choice of noisy attractor. We conclude that the kinetics of inducer intake affects transient and near-equilibrium gene expression dynamics and, consequently, the phenotypic diversity of organisms in fluctuating environments.

Introduction

To survive, cells must adapt to environmental changes, such as in the concentration of nutrients and toxics. Some of these changes can occur at rates faster than, e.g., the cell cycle, and thus require rapid adaptability from the cells. This adaptability may involve modifications in the kinetics of membrane-associated mechanisms (Sajbidor, 1997), metabolic rates (Talwalkar and Kailasapathy, 2003), or gene expression (Yamamoto and Ishihama, 2005; Allen and Tresini, 2000).

Studies suggest that organisms such as *Escherichia coli* can adjust the reception of some external signals. For example, in normal conditions, transcription of the genes of the *lac* operon (Elf et al., 2007; Hansen et al., 1998) is inhibited by the native *lac* repressor. When allolactose is present in the environment, it is absorbed by passive intake transport and triggers the expression of the *lac* genes. One of the proteins expressed, *lacY*, enhances the intake of allolactose further, thus forming a positive feedback. Such feedback mechanisms are particularly useful in saving cellular resources in periods when inducers are not present(Jacob and Monod, 1961). On the other hand, negative feedback mechanisms are appropriate for, e.g., quickly pumping unwanted substances out of the cell (Schnappinger and Hillen, 1996). One example is the tetracycline intake system, as the tetracycline-induced proteins *tetA* actively transport the tetracycline out of cell (Beck et al., 1982).

A recent study (Megerle et al., 2008) showed that the timing of intake of inducers differs widely between cells in monoclonal populations of *E. coli*. Such variability in intake times was visible in the differing timings for the appearance of proteins in cells following the introduction of the inducer in the media even though, after a transient period, all cells exhibited the same rate of protein production. Another recent study (Makela et al., 2013) supported this hypothesis, by showing that there is a wide variability in the timing of activation of transcription following the appearance of the inducer in the media that is not due to noise in gene expression but causes high cell to cell variability in RNA numbers for long periods of time (i.e. longer than several cell cycles). This source of phenotypic diversity is likely to be of particular relevance in fluctuating environments (Acar et al., 2013; Ribeiro, 2008).

Here, using detailed stochastic models of gene expression and intake processes in *E. coli*, we investigate if differing intake kinetics results in differing kinetics of expression of the target gene both in the transient period for protein numbers to reach near-equilibrium, as well as in the subsequent stable phase. Next, we investigate how fluctuations in inducer numbers in the environment affect this variability. Finally, we investigate whether the intake kinetics of inducers can affect the behavior of a small genetic circuit, namely, a toggle switch.

Methods

We compare the kinetics of expression when the inducer of the target gene enters cells via positive feedback mechanisms, passive diffusion, or negative feedback mechanisms. While varying the intake kinetics, the mean number of proteins expressed by the target gene is kept invariant in the stable phase.

The dynamics of the models is driven by the delayed

Stochastic Simulation Algorithm (delayed SSA) (Roussel and Zhu, 2006). This algorithm, unlike the original SSA (Gillespie, 1977), allows delaying the release of products, following a reaction. Furthermore, it differs from previous algorithms that can accommodate delays (e.g. (Bratsun et al., 2005; Barrio et al., 2006)) in that it can handle multiple delayed events in one reacting event, which facilitates the modeling of genetic circuits(Ribeiro, 2010). The delayed SSA uses a wait list to store delayed output events. The wait list is a list of elements (e.g., proteins being produced), each to be released after a time interval has elapsed (also stored in the wait list).

Each model includes an extracellular environment, which contains inducers, a cellular intake mechanism of inducers, and a gene expression mechanism that requires activation. The proteins produced can affect the intake kinetics, so as to model positive or negative feedback mechanisms. The models are simulated by SGNSim (Ribeiro and Lloyd-Price, 2007). All parameter values are extracted from measurements in *E. coli*, unless stated otherwise.

Environment and passive transport of inducers

We assume that the inducers in the environment are inexhaustible. To model passive diffusion of inducers into cells, we set a rate constant of intake (k_{Iin}) and an extracellular amount of inducers (I_e) equal to 1, for simplicity. The intake is modeled by the following reaction:

$$\emptyset \xrightarrow{I_e \times k_{I\,in}} I \tag{1}$$

where I is the number of inducers inside a cell. Note that, in all cases, even when an active mechanism is present, there is always passive intake.

To model fluctuations in inducer numbers in the environment, we assume that these follow a Gaussian distribution with standard deviation σ_e and unity mean. The fluctuations are set to occur at a rate slow enough to allow for one fluctuation to have a visible effect in the protein numbers before the next one occurs. For that, we use a first order autoregressive model to restrict the degree of change in inducer numbers from one moment to the next, with the following update rule:

$$I_e(t) = 1 - \phi + I_e(t - \delta t) * \phi + \epsilon_t \tag{2}$$

where ϕ is constant, ϵ_t is white noise with standard deviation of σ_t, and δt is the update interval. The model generates values for the extracellular inducer numbers according to $I_e \sim N(1, \sigma_e^2)$, where:

$$\sigma_e^2 = \frac{\sigma_t^2}{1 - \phi^2} \tag{3}$$

The inducers' extracellular concentration thus has the autocorrelation function's decay rate of $-ln(\phi)/\delta t$. By tuning ϕ, one can adjust the rate of change in this concentration. Finally, cells can dispose of inducers via diffusion, modeled as a first order reaction event:

$$I \xrightarrow{k_{I\,out}} \emptyset \tag{4}$$

Reaction 4 is assumed to account also for possible degradation events of inducers when inside the cell.

Active transport mechanisms

We assume that the active transport rate is proportional to the number of proteins of the target gene. Let such transport be done by a protein P.

Positive feedback mechanisms are modeled by reaction 5a, where one inducer I is introduced in the cell by a protein P, while negative feedback mechanisms are modeled by reaction 5b, where one inducer is pumped out of the cell by a protein P:

$$P \xrightarrow{I_e \times k_{P\,in}} P + I \tag{5a}$$

$$P + I \xrightarrow{k_{P\,out}} P \tag{5b}$$

These two mechanisms are never present simultaneously.

Gene expression

We assume that the gene only expresses once activated. An inducer I interacts with the operator site O at the promoter region of the gene via the following reactions:

$$O + I \underset{k_d}{\overset{k_a}{\rightleftharpoons}} O.I \tag{6}$$

where k_a and k_d are the association and the disassociation rate constants, respectively. We assume half lives of inducers much longer than the expected time for disassociation to occur. Thus, we do not model degradation of inducers when bound to the operator. Additionally, we assume that leaky expression is negligible and that the operator site is not overlapped by the RNA polymerase for any significant time, so that the interaction between promoter and inducer is independent of the transcription process, particularly initiation.

The model of gene expression used was proposed and validated in (Ribeiro et al., 2006; Zhu et al., 2007), by comparing its kinetics with the real-time production of tsr-venus proteins under the control of a lac promoter in *E. coli* (Yu et al., 2006). The model consists of transcription (7), in this case of the activated gene, and translation of the resulting RNA molecules (8). Also modeled are first order degradation processes of RNA (9) (Bernstein et al., 2002) and proteins (10). Transcription events can occur when the operator-inducer complex $O.I$ is formed. RNA polymerases are not explicitly modeled, as it is assumed that these exist in sufficient amount so that fluctuations in their numbers are not

significant. The transcription start site (TSS) is modeled explicitly so that transcription initiation events do not interfere with operator-activator reactions:

$$O.I + TSS \xrightarrow{\infty} TSS(\tau) + O.I + M(\tau) \quad (7)$$

$$M \xrightarrow{k_P} M + P \quad (8)$$

$$M \xrightarrow{d_M} \emptyset \quad (9)$$

$$P \xrightarrow{d_P} \emptyset \quad (10)$$

Reaction 7 describes the process of transcription. In particular, τ (which follows a Gamma distribution (Kandhavelu et al., 2011)) accounts for the finding of a promoter region by an RNA polymerase, the formation of the closed complex at the transcription start site, the open complex formation, and finally, the promoter escape (DeHaseth et al., 1998) and elongation. Of these, in general, the most rate limiting steps are the isomerization steps and the open complex formation (McClure, 1985; Lutz et al., 2001). To model this multi-step process, we set the reaction rate to infinity, which causes the reaction to occur the moment the reactants become available. Thus, in this model, τ determines the interval between consecutive productions of transcripts when the gene is induced.

Also, given the short duration of the elongation time in comparison to transcription initiation (Kandhavelu et al., 2011), the transcript (M) is released at the same time as the TSS (i.e. the elongation time is assumed negligible). This allows for translation events of the RNA to initiate (reaction 8) as soon as the assembly of that RNA begins (Miller et al., 1970).

Genetic Toggle Switch

The toggle switch consists of a network of two genes (here, B and C), whose proteins (P_B and P_C, respectively) repress the other gene's activity. We assume that these two genes are only active when bound by a protein produced by an operon, 'A', which is itself activated by the extracellular inducer. The transcription and translation processes in each of these genes are modeled as described in the previous section.

In this model, the operon A, once activated by the inducer, expresses two proteins, P_{A1} and P_{A2}. The former is involved in the intake of the inducer via a feedback mechanism (described in a previous section), while the latter activates genes B and C via the following reaction:

$$P_{A2} + OR_i \xrightarrow{k_a} O_i \quad (11)$$

where P_{A2} is the activator, OR_i is an operator site at the promoter of either gene B or C in the inactive state ($i = B, C$), k_a is the association rate constant, and O_i is the operator region with the activator bound to it.

The interactions between genes B and C form a switch. Namely, each of these genes, once activated by P_{A1}, is free

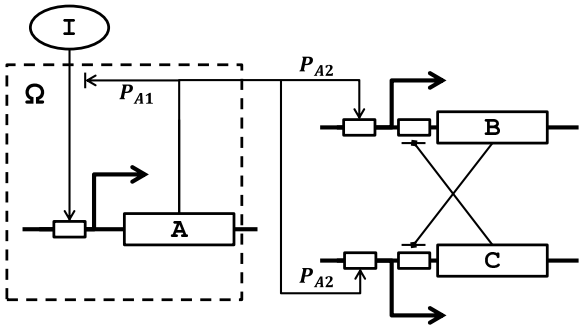


Figure 1: Model switch and activation mechanism. Inducers enter the cell by the intake mechanism, whose kinetics is determined by Ω (dashed box). Protein P_{A1} is responsible for the feedback mechanism, while P_{A2} activates genes B and C, whose mutual interactions form a switch.

to express or to be repressed by the protein of the other gene (P_B represses gene C while P_C represses gene B). Such repressions occur at a second operator site at the promoter regions, via:

$$P_j + O_i \xrightarrow{k_{ai}} O_i P_j \quad (12)$$

where P_j is the repressor, O_i is an active operator site of either gene B or C, $O_i P_j$ is the operator region with a repressor bound to it, and k_{ai} is the association rate constant of that repressor. Importantly, this rate differs in the two genes, being higher for gene C, which biases the choice of noisy attractor made by the switch, when first initialized (Ribeiro and Kauffman, 2007). The noisy attractor favored is "gene C on and gene B off". In Fig. 1 we show a schematic representation of this model.

Characterization of the intake process

To compare the effects of different active intake systems we modeled these such that, for the same extracellular concentration of inducers, one has the same mean protein number of the inducible gene at steady state ($[P]$), for varying active transport kinetics. To achieve this, the mean number of inducers in the cells at steady state ($[I]$) has to be kept constant for varying intake kinetics, which is done by tuning the kinetics of passive intake.

In the stable phase, the influx and outflux rates of inducers are identical. Let f_1 and f_2 be the passive influx and outflux of inducers, respectively. These occur, respectively, via reactions 1 and 4. Given these, $f_1 = I_e \times k_{I\,in}$ while $f_2 = [I]k_{I\,out}$. Let f_3 be the flux due to the active transport. Since the flux (f) from active and passive transports, in the stable phase, must equal zero, then:

$$\begin{aligned} f &= f_1 - f_2 + f_3 \\ &= [I_e]k_{I\,in} - [I]k_{I\,out} + f_3 = 0 \end{aligned} \quad (13)$$

The active transport flux f_3 can thus be calculated as follows, in each condition:

$$f_3 = \begin{cases} [I_e][P]k_{P\,in} > 0 & \text{positive feedback} \\ -[P][I]k_{P\,out} < 0 & \text{negative feedback} \\ 0 & \text{passive diffusion} \end{cases} \quad (14)$$

Finally, we define Ω as the log ratio between the passive disposal flux and the passive intake flux. The value of Ω informs on the rate at which inducers are transported into cell by passive intake when in the stable phase:

$$\Omega = lg(\frac{f_2}{f_1}) = lg(\frac{f_3}{f_1} + 1) = lg(\frac{[I]k_{Iout}}{k_{Iin}}) \quad (15)$$

Ω can take positive values in the presence of positive feedback (i.e. $f_3 > 0$), and negative values in the presence of negative feedback ($f_3 < 0$). Given passive intake alone ($f_3 = 0$), $\Omega = 0$.

Results

The models used to study the effects of the kinetics of the intake of inducers on the dynamics of gene expression are stochastic. Thus, to assert if their dynamics changes as a function of a parameter's value, we perform tests of statistical significance. Also, the models are initialized without any proteins of the target gene, so as to assess the kinetics both at the stable phase, as well as during the transient to reach the stable phase.

Gene expression and intake kinetics

We first study the dynamics of protein numbers of a target gene as a function of the intake kinetics of the inducer, both in the transient phase and in the near-equilibrium or stable phase. In all cases, we use the following parameter values for the model of gene expression: $k_P = 0.005$ s^{-1} (Taniguchi et al., 2010), $d_M = 0.002$ s^{-1}(Bernstein et al., 2002), and $d_P = 0.0005$ s^{-1}(Taniguchi et al., 2010). Also, we let τ be a random variable following a gamma distribution $\Gamma(\alpha, \theta)$, with the shape α equal to 2 (Kandhavelu et al., 2011) and the scale θ equal to 25 s. The value of θ was set so that, given the other parameter values, the mean RNA number in the stable phase is $\sim$ 10, in accordance with *in vivo* measurements in *E. coli* (Taniguchi et al., 2010). The reaction rates of inducer-operator interactions are set to: $k_a = 10^{-5}s^{-1}$ and $k_d = 0.02$, so that the expected time for inducers to bind to the promoter is in accordance with measurements reported in (Elf et al., 2007).

We vary Ω, while maintaining constant the mean protein ($[P] \sim 30$) and mean inducer numbers within the cells ($[I] \sim 1400$), when in the stable phase. For this, the range of variation of Ω was constrained between -1 and 1. This range complies with measurements of the intake kinetics of known inducers (namely, of tet and lacY) in *E. coli* (Brown and Hogg, 1972; Hansen et al., 1998; Beck et al., 1982).

For each value of Ω, we simulate 500 cells, each for 60 000 s, sampling their state every 60 seconds. We define the stable phase as the phase in which the mean protein numbers in the cells do not differ, in a statistical sense, for different values of Ω. We found that all cells reach the stable phase after, at most, $t = 4 \times 10^4$ s. In Table 1, we show the p-values of the Kolmogorov-Smirnov (KS) test comparing the distribution of mean protein numbers when $\Omega = 0$ (passive intake) with each of these distributions for the other values of Ω, in the stable phase.

Table 1: p-values of the KS test comparing the distribution of mean protein numbers when $\Omega = 0$ and when Ω takes other values, in the stable phase ($t > 4 \times 10^4$ s)

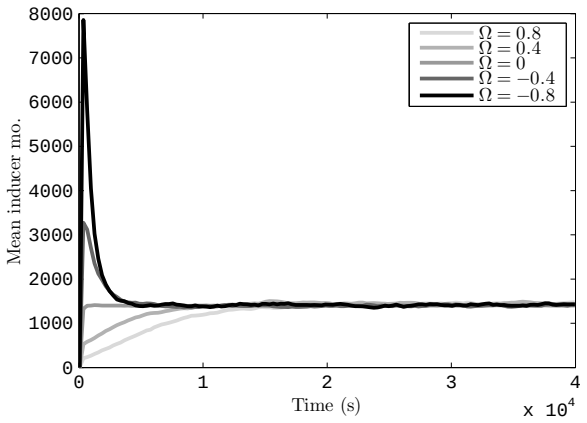
Ω	-1	-0.4	1	0.4	1
p-value	0.03	0.19	1.00	0.26	0.34

The p-values from likelihood ratio tests with the null hypothesis that the distributions are identical are larger than 0.01, thus, we cannot reject that they are identical in a statistical sense. Therefore, in this range of values of Ω, the models have identical mean protein numbers over time, in a statistical sense, when in the stable phase.

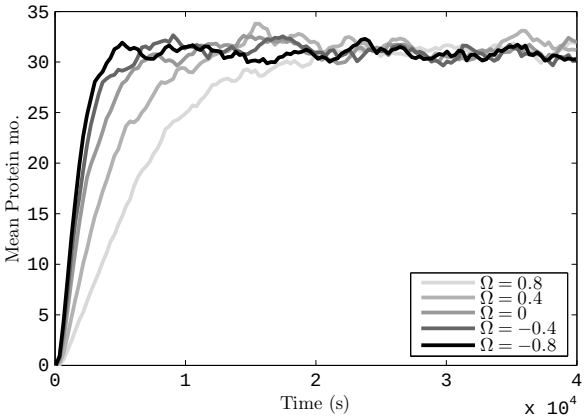
We next study the intracellular dynamics of inducer numbers as a function of Ω. In Fig. 2a we show the mean number of inducers in the cells over time, from the start of the simulations. These vary significantly as a function of Ω, before reaching the stable phase. In particular, for $\Omega < 0$ (negative feedback), there is a rapid influx of inducers, followed by a steady decrease towards the numbers at near-equilibrium. For $\Omega > 0$ (positive feedback), the inducer numbers take longer time to reach near-equilibrium. The passive intake mechanism ($\Omega = 0$) is, of the cases modeled, the one for which the intracellular inducer numbers stabilizes faster. We found by inspection that this mean time is minimized for values of Ω close to, but slightly smaller than 0. Finally, from Fig. 2b we observe that the proteins reach slower the numbers observed in the stable phase the greater is Ω.

In Fig. 3, we show the mean transient time (t_0) for each value of Ω, along with the square of the coefficient of variation ($CV^2(t_0)$), obtained from the multiple simulations in each condition. We find that the mean t_0 is shorter for negative intake mechanisms. However, the $CV^2(t_0)$ does not change significantly with Ω.

Next, we assess the fluctuations in protein and inducer numbers in the stable phase as a function of Ω. Fig. 4 shows the variance over the mean (σ^2/μ) ratio (i.e. the fano factor) of these numbers. This quantity is minimized for $\Omega = 0$ in the case of intracellular inducer numbers. In the absence of feedback, these follow a Poisson distribution, as expected since both the passive intake and disposal are first-order processes. When there is an active feedback mechanism, the noise in protein numbers causes an noise in the intracellular



(a) Mean intracellular inducer numbers over time



(b) Mean protein numbers over time

Figure 2: Mean numbers of inducers (top) and proteins (bottom) over time: $\Omega < 0$ corresponds to negative feedback, $\Omega = 0$ corresponds to passive intake, and $\Omega > 0$ corresponds to positive feedback mechanisms.

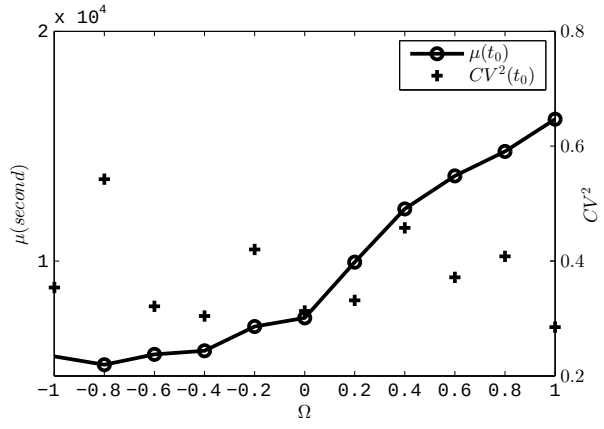


Figure 3: Mean (μ) and CV^2 of the transient time, t_0, for protein numbers to reach near-equilibrium as a function of Ω.

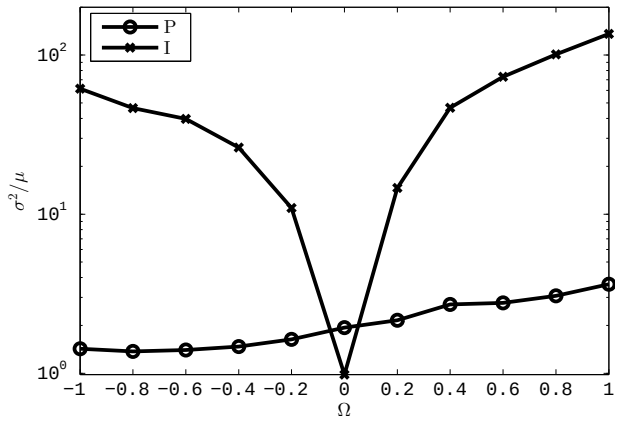


Figure 4: Noise in protein and intracellular inducer numbers in the stable phase as a function of Ω.

inducer numbers to be higher, then when only passive intake is present. The stronger the feedback mechanism (larger deviation from $\Omega = 0$), the stronger is this effect.

On the other hand, the noise in protein numbers increases for increasing Ω, being lower for negative feedback mechanisms and higher for positive feedback mechanisms. To investigate this, we calculated the normalized cross-correlation between protein and intracellular inducer numbers in the stable phase for varying Ω (Fig 5). For positive feedbacks ($\Omega > 0$), the protein and inducer numbers are positively correlated. This means that the noise in intracellular inducer numbers will be propagated to the protein numbers, causing it to be higher than in the passive diffusion case. In the regime of negative feedbacks, the numbers of intracellular inducers and proteins are anti-correlated, and the noise in the numbers of proteins is suppressed, when compared to passive diffusion case. Finally, as expected, in the absence of feedback mechanisms, the noise in intracellular inducer numbers does not affect the protein numbers, which is indicated by zero cross-correlation at $\Omega = 0$.

Finally, we study how the intake mechanisms behave in environments with fluctuating number of inducers. We assume that the extracellular number of inducers follows a Gaussian distribution with variance $\sigma_e^2 = 0.2$ and unity mean, generated by the autoregressive model (see methods). We set δt to 30 s. We set the rate of environmental change, ϕ, from 0.5 to ~ 1. The closer the value of ϕ to 1, the slower the decay rate of the autocorrelation function of the extracellular inducer concentration. For $\phi \sim 1$, the extracellular inducer concentration is constant, corresponding to $\sigma_e = 0$. For each pair of values [ϕ,Ω], we simulate one cell for 5×10^6 s, sampling every 60 s. Fig. 6 shows the changes in fluctuations in protein numbers in the stable phase (as assessed by the CV^2) due to the fluctuations in the inducer numbers, relative to when in stable environments.

From Fig. 6, with values of ϕ larger than 0.8, the noise

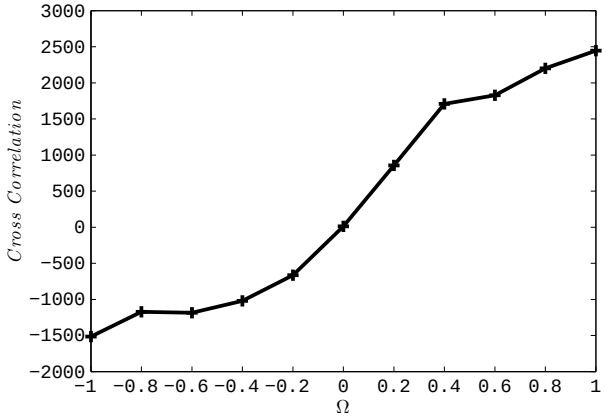


Figure 5: Cross-correlation between protein and inducer numbers in the stable phase.

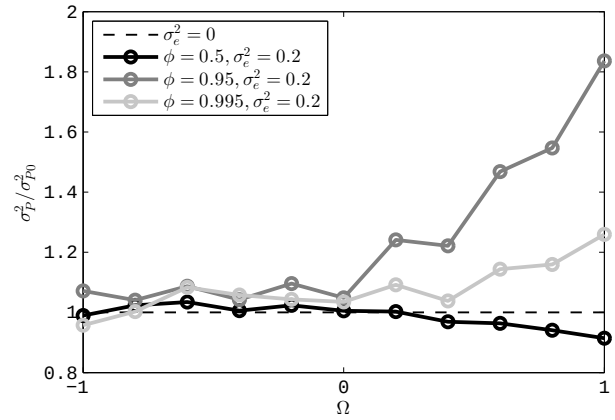


Figure 6: Noise amplification ratio (σ_P^2/σ_{P0}^2) due to fluctuations in the number of inducers in the environment.

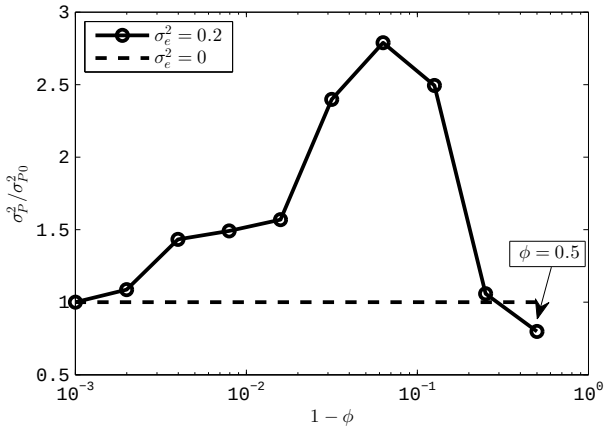


Figure 7: Noise amplification ratio (σ_P^2/σ_{P0}^2) with $\Omega = 1$, $\sigma_e = 0.2$ for different values of ϕ.

amplification in the protein numbers appears to increase with increasing Ω. For each value of $\phi > 0.8$, we performed tests of statistical significance between the protein numbers when Ω is 0, and when is -1 and 1, respectively. In both tests, we found that the distributions are distinct (p-values smaller than 10^{-10}), confirming that the increase in the noise amplification effect with increasing Ω is statistically significant.

For positive feedback mechanisms, the noise amplification ratio for different values of ϕ resembles a band pass filter (Fig.7). As ϕ increases up to 0.9, the fluctuations in the external inducer numbers propagate more efficiently to the protein numbers of the induced gene (as shown in Samoilov et al. (2002)). When ϕ increases beyond 0.9, the noise amplification ratio decreases as the positive feedback, affected by the extracellular inducer numbers, loses the ability to reflect the fluctuations in protein numbers. Interestingly, for $\Omega = 1$, $\phi = 0.5$, $\sigma_e = 0.2$, the protein CV^2 is reduced by $\sim 10\%$ when compared with the noiseless case (i.e. $\sigma_e = 0$). This reduction is significant, namely the resulting distributions of protein numbers in the two cases are statistically distinct (p-value smaller than 10^{-10}).

Inducible genetic switch

We study the behavior of a biased genetic switch as a function of the intake kinetics of an inducer. The model consists of two genes, B and C, which form a switch via mutually repressing interactions, and of a third gene, A, responsible for, once activated by the inducer, activate both genes B and C (figure 1).

In this model, P_{A_2} activates the expression of genes B and C in the same fashion as reaction 6, with the association rates: $k_{A_2B} = 0.001\ s^{-1}$, $k_{A_2C} = 0.002\ s^{-1}$. The other parameters of genes B and C are: $k_{PB} = k_{PC} = 0.05s^{-1}$, $d_{MB} = d_{MC} = 0.002\ s^{-1}$, $d_{PB} = d_{PC} = 0.0005\ s^{-1}$. The formation of the open complex is regulated by: $\tau_B, \tau_C \sim \Gamma(1, 100s)$. Finally, we denote $[t_B]$ and $[t_C]$ as the

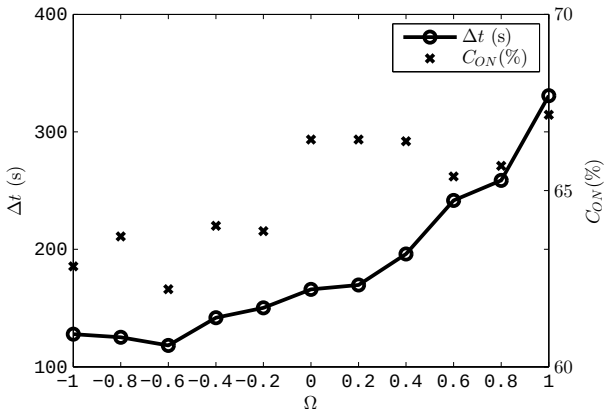


Figure 8: Mean temporal difference between the first activation of genes B and C (Δt), and bias in the choice of noisy attractor towards "gene C on" (C_{ON}).

mean elapsed times from the introduction of inducers to the first closed complex formation in each promoter. The bias in the association rate constants biases the moments of formation of the closed complex. Namely, on average, these occur at moments that differ in time by: $\Delta t = [t_B] - [t_C] > 0$.

Given these parameter values, we observed that both noisy attractors of the switch are stable enough so that, once a noisy attractor is reached, the switch will remain there until the end of the simulation. Also, as noted in the methods section, the association rate of the activator protein P_{A_2} is higher in the case of gene C, which biases the first choice of noisy attractor of the switch. For $\Omega = 0$, the first noisy attractor selected by the switch will be "gene C on" $\sim 65\%$ of the times (see Fig. 8).

We study the bias in the choice of noisy attractor as a function of Ω. For each value of Ω, we simulate 1000 independent cells, each in 30 000 s, sampled every 5 seconds. The results are shown in Fig. 8. From this figure, for increasing values of Ω, the bias in choice of noisy attractor is reduced from $\sim 67\%$ to $\sim 62\%$.

We performed statistical tests of significance comparing the distribution of Δt when $\Omega = 0$ to the same distribution when $\Omega = -1$ and 1 respectively. The test results show that all tested pairs of distributions are distinct (p-values smaller than 10^{-10}).

This result can be explained as follows. As Ω is increased, P_A reaches the stable phase slower. Thus, both $[t_B]$,$[t_C]$ increase, but not by equal amounts (e.g. for half the number of proteins A, each of these times is doubled). Accordingly, Δt increases and the distribution of chosen noisy attractors becomes more biased.

Discussion

Many genes in *E. coli*, as well as other single-celled organisms, only become active in response to an external signal, either individually, or as part of a small network. Additionally, even when active, in the stable phase, most genes exhibit very small mean RNA numbers (from one to a few) (Taniguchi et al., 2010). This implies that differences in the intake time of inducers between sister cells can have significant implications on phenotypic differences between them. Similarly, differences in the kinetics of the intake process of different inducers may lead to significant differences on the mean and variability of response times to those inducers. Relevantly, Recent measurements *in vivo* showed that the intake time of inducers can be of the same order of magnitude as the cell division time and transcription initiation (Kandhavelu et al., 2011). The degree of cell-to-cell variability in these times is also equally high (Megerle et al., 2008; Makela et al., 2013).

Using a stochastic model with parameter values extracted from measurements in *E. coli*, we showed that the nature of the intake mechanism, that is, whether it is based on passive diffusion, positive feedback or negative feedback mechanism, has a significant impact on the dynamics of gene expression, both in the transient phase, as well as in the stable phase. The intake kinetics not only affects mean and variability of the transient time to reach the stable phase but also the degree of fluctuations in these numbers once it that phase. These effects are tangible in the behavior of small genetic circuits.

The results presented here show that the kinetics of the response, in terms of gene expression, of single-celled organisms to external signals, depends to great extent not only on the intake mechanism of the inducer/repressor molecule as well as on the mechanisms of transcription and translation. Also relevant is the observation that the intake mechanism also has an effect on the kinetics of gene expression, long after the transient period. This implies that the kinetics of genes responsive to environmental signals ought to be studied accounting for the effects of the intake mechanism on RNA and protein numbers dynamics. In the future, it would be of interest to further explore how active transport mechanisms, able of positive or negative feedback processes, can be used to tune the behavior and adaptability of small genetic circuits to fluctuating environments.

Acknowledgements

We acknowledge Tekes and Academy of Finland. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Acar, M., Mettatal, J. T., and van Oudernaarden, A. (2013). Stochastic switching as a survival strategy in fluctuating environments. *Nature Gen.*, 40(4):471–475.

Allen, R. and Tresini, M. (2000). Oxidative stress and gene regulation. *Free Radic. Biol. Med.*, 28(3):463–499.

Barrio, M., Burrage, K., Leier, A., and Tian, T. (2006). Oscillatory regulation of hes1: Discrete stochastic delay modelling and simulation. *PLOS Comp. Biol.*, 2(9):1017–1030.

Beck, C., Mutzel, R., Barbe, J., and Muller, W. (1982). A multifunctional gene (tetr) controls tn10-encoded tetracycline resistance. *J. Bacteriol.*, 150(2):633–642.

Bernstein, J., Khodursky, A., Lin, P., Lin-Chao, S., and Cohen, S. (2002). Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. *Proc. Natl. Acad. Sci. USA.*, 99(15):9697–9702.

Bratsun, D., Volfson, D., Tsimring, L., and Hasty, J. (2005). Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. USA.*, 102(41):14593–14598.

Brown, C. and Hogg, R. (1972). A second transport system for l-arabinose in escherichia coli b/r controlled by the arac gene. *J. Bacteriol.*, 111(2):606–613.

DeHaseth, P. L., Zupancic, M., and M. T. Record, J. (1998). Rna polymerase-promoter interactions: the comings and goings of rna polymerase. *J. Bacteriol.*, 180(12):3019–3025.

Elf, J., Li, G., and Xie, X. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316:1191–1194.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361.

Hansen, L., Knudsen, S., and Srensen, S. (1998). The effect of the lacy gene on the induction of iptg inducible promoters, studied in escherichia coli and pseudomonas fluorescens. *Curr. Microbiol.*, 36(6):341–347.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, 3(3):318–356.

Kandhavelu, M., Mannerstrom, H., Gupta, A., Hakkinen, A., Lloyd-Price, J., Yli-Harja, O., and Ribeiro, A. (2011). In vivo kinetics of transcription initiation of the lar promoter in escherichia coli. evidence for a sequential mechanism with two rate-limiting steps. *BMC Sys. Biol.*, 5:149.

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. *Nuc. Acid. Res*, 29(18):3873–3881.

Makela, J., Kandhavelu, M., Oliveira, S., Chandraseelan, J. G., Lloyd-Priced, J., Peltonen, J., Yli-Harja, O., and Ribeiro, A. (2013). In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter. *Nucl. Acids Res.*, *in press*.

McClure (1985). Kinetics of open complex formation between *Escherichia coli* rna polymerase and the lac uv5 promoter. evidence for a sequential mechanism involving three steps. *Biochem.*, 24(11):2712–2723.

Megerle, J., Fritz, G., Gerland, U., Jung, K., and Radler, J. (2008). Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.*, 95:2103–2115.

Miller, O., Hamkalo, B., and Thomas, C. (1970). Visualization of bacterial genes in action. *Science*, 169(3943):392–395.

Ribeiro, A. (2008). Dynamics and evolution of stochastic bistable gene networks with sensing in fluctuating environment. *Phys. Rev. E*, 78(1):061902.

Ribeiro, A. (2010). Stochastic and delayed stochastic models of gene expression and regulation. *Math. Biosci.*, 233:1–11.

Ribeiro, A. and Kauffman, S. (2007). Noisy attractors and ergodic sets in models of gene regulatory networks. *J. Theor. Biol.*, 247:743–755.

Ribeiro, A. and Lloyd-Price, J. (2007). Sgn sim, a stochastic genetic networks simulator. *Bioinf.*, 23(6):777–779.

Ribeiro, A., Zhu, R., and Kauffman, S. (2006). A general modeling strategy for gene regulatory networks with stochastic dynamics. *J. Comp. Biol*, 13(9):1630–1639.

Roussel, M. and Zhu, R. (2006). Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys. Biol.*, 3:274–284.

Sajbidor, J. (1997). Effect of some environmental factors on the content and composition of microbial membrane lipids. *Crit. Rev. Biotechnol.*, 17(2):87–103.

Samoilov, M., Arkkin, A., and Ross, J. (2002). Signal processing by simple chemical systems. *J. Phys. Chem. A*, 106:10205–10221.

Schnappinger, D. and Hillen, W. (1996). Tetracyclines: antibiotic action, uptake, and resistance mechanisms. *Arch. Microbiol.*, 165(6):359–369.

Talwalkar, A. and Kailasapathy, K. (2003). Metabolic and biochemical responses of probiotic bacteria to oxygen. *J. Dairy Sci.*, 86(8):2537–2546.

Taniguchi, Y., Choi, P., Li, G., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. (2010). Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538.

Yamamoto, K. and Ishihama, A. (2005). Transcriptional response of escherichia coli to external zinc. *J. of Bacteriol.*, 187(18):6333–6340.

Yu, J., Xiao, J., Lao, K., and Xie, X. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science*, 311:1600–1603.

Zhu, R., Ribeiro, A. S., Salahub, D., and Kauffman, S. (2007). Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models. *J. Theor. Biol.*, 246(4):725–745.

PUBLICATION IV

H Tran, SMD Oliveira, NM Goncalves, AS Ribeiro (2015), “Kinetics of the cellular intake of a gene expression inducer at high concentrations”, *Molecular BioSystems*, 11, pg. 2579. doi: 10.1039/C5MB00244C.

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).

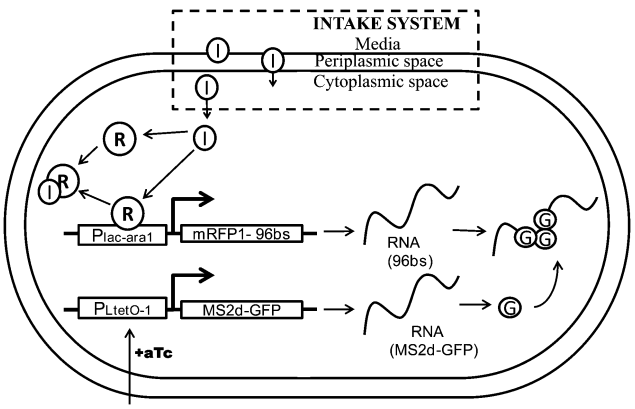


Fig. 1 Diagram of the inducer intake system, target gene and RNA tagging system: IPTG molecules (*I*) in the media enter the cytoplasm by passing through two membrane layers, with a periplasmic space in between. When in the cytoplasm, they neutralize *lacI* repressors (*R*) by forming inducer–repressor 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†).

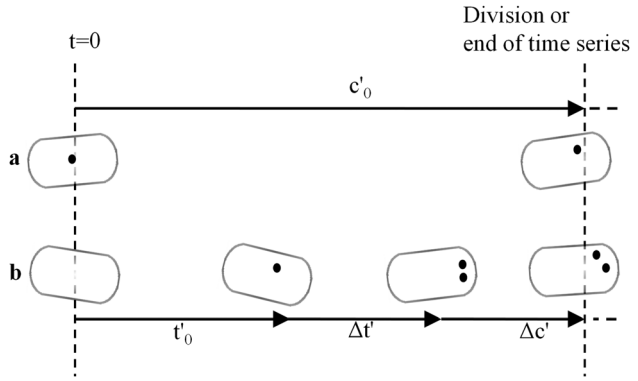


Fig. 2 Collection of t_0 and Δt samples subject to censoring: from cells in the initial population that do not produce any RNA during their lifetime (a), we obtain censored samples of t_0, denoted c_0', whose value equals the cell lifetime. From cells that produce at least one RNA (b), we obtain actual 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.

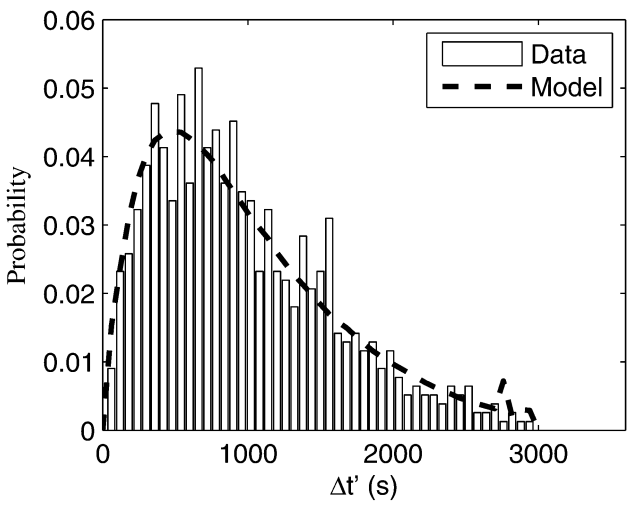


Fig. 3 Distribution of the actual samples of RNA production intervals $\Delta t'$ (bars) and distribution curve (dashed line) inferred from the transcription model of two sequential steps. The samples are obtained from cells subject to 1 mM and 1.25 mM of IPTG in the media.

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

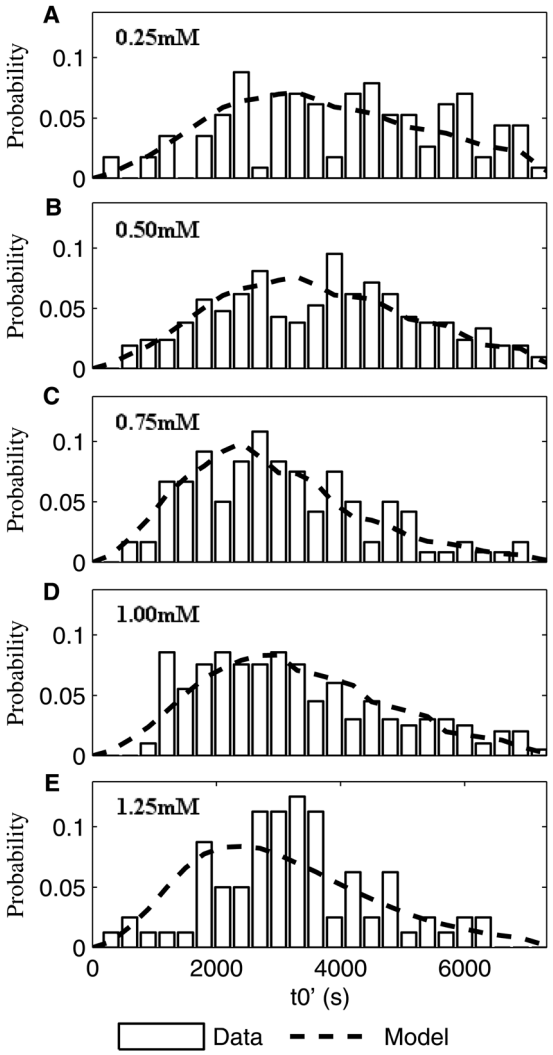


Fig. 4 Distribution of the actual samples t_0' (bars) and distribution curves inferred from the model (dashed lines). Data for (A) 0.25 mM, (B) 0.50 mM, (C) 0.75 mM IPTG, (D) 1.0 mM and (E) 1.25 mM IPTG concentrations.

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

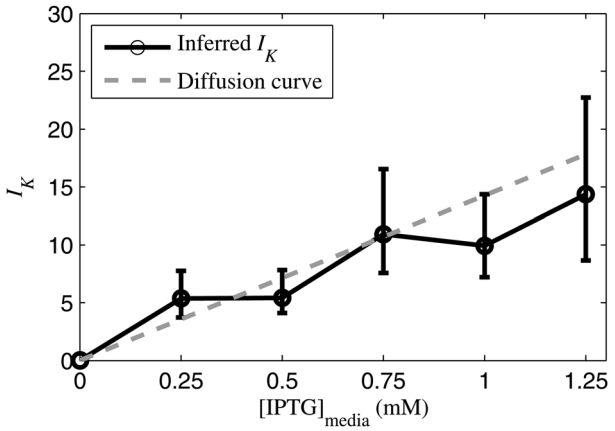


Fig. 5 Intracellular inducer level, I_K, as a function of external inducer concentration $[IPTG]_{media}$. The vertical bar indicates the margins of error, for α-value of 90%. The diffusion curve (grey dashed) is the approximation of I_K using a line through the origin (0, 0). The slope of the diffusion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

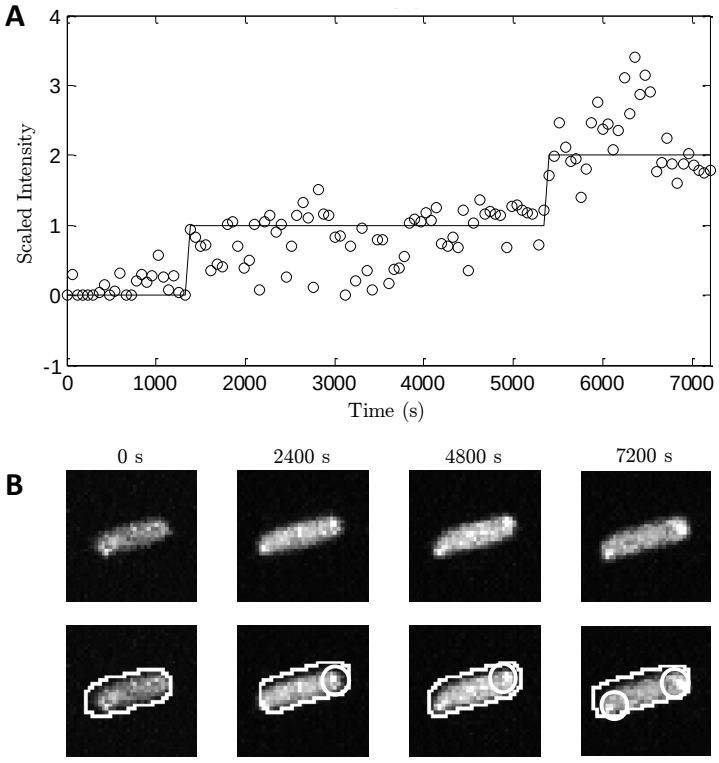


Figure S1. Tagged RNAs in *E. coli* cells. (A) Unprocessed frames and segmented cells and RNA spots. The moments when images were taken are shown for each frame. (B) Examples of time series of scaled spot intensity levels from one cell (circles) and the corresponding estimated RNA numbers (solid lines).

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: $\Lambda(\langle\delta_{1..n}, z_{1..n}\rangle|\theta)$. 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]:

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

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:

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

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:

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

By comparing $P_{T'|\theta}(t|\theta)$ 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:

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

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

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

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:

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

The general solution for this equation is:

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

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:

$$\begin{aligned} I(t) &= \frac{\int e^{d_I t} k_o(1 - e^{-k_i t}) dt + C_2}{e^{d_I t}} \\ &= \frac{k_o(\int e^{d_I t} dt - \int e^{(d_I - k_i)t} dt) + C_2}{e^{d_I t}} = \frac{k_o}{d_I} - \frac{k_o e^{-k_i t}}{d_I - k_i} + \frac{C_2}{e^{d_I t}} \\ &= \frac{k_o(d_I - k_i) - d_I k_o e^{-k_i t} + C_2 d_I (d_I - k_i) e^{-d_I t}}{d_I(d_I - k_i)} \\ &= \frac{k_o(d_I e^{-k_i t} + C_2 d_I (k_i - d_I) e^{-d_I t} + k_i - d_I)}{d_I(k_i - d_I)} \end{aligned} \tag{14}$$

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:

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

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:

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

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

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

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]).

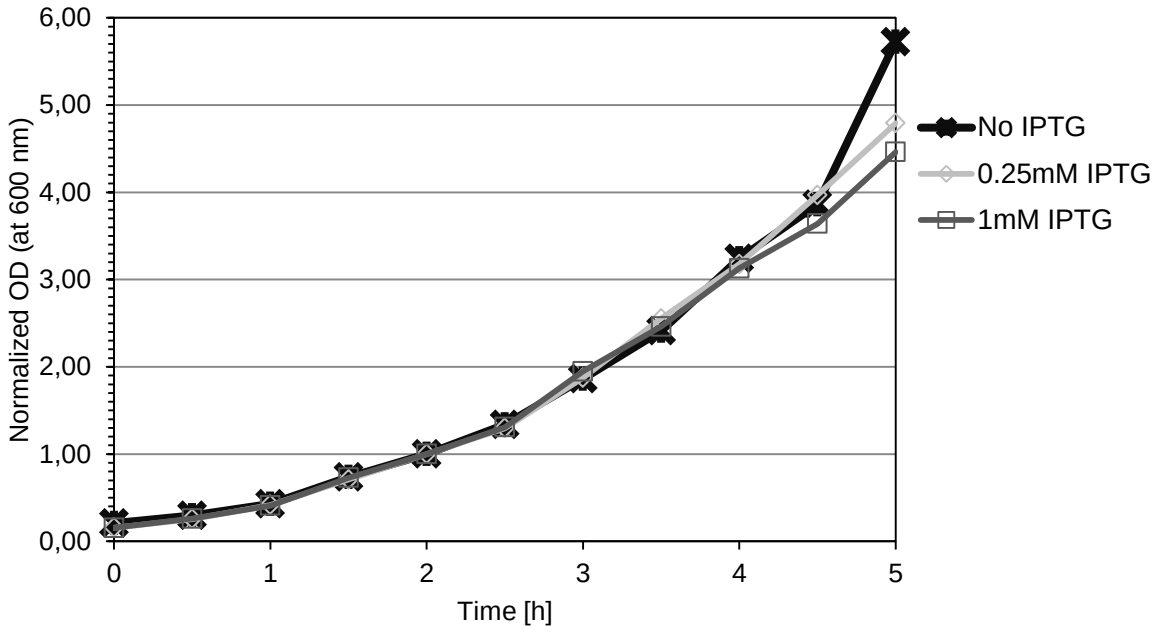
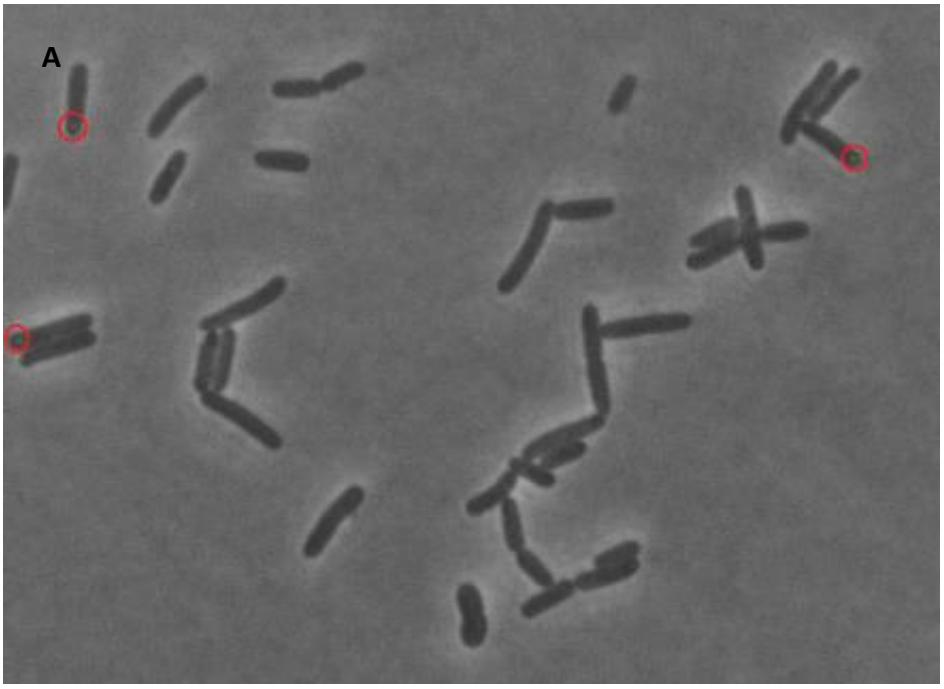
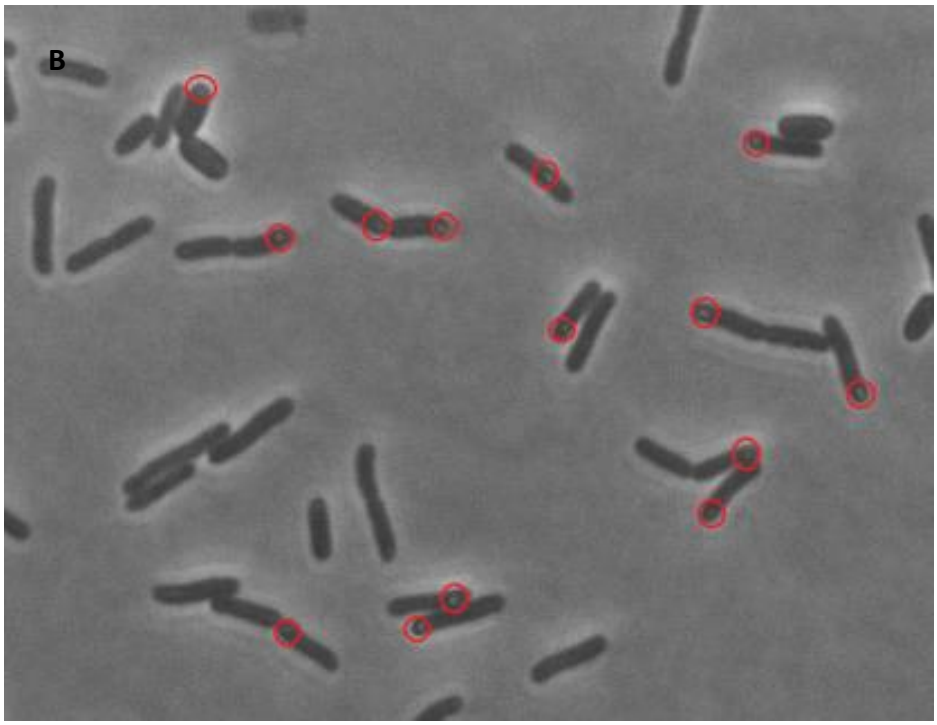


Figure S2. Normalized optical density (OD) curve at 0.25mM (diamond) and 1mM (square) IPTG and without IPTG (cross). Inducers are added at the end of the second hour, where the normalized OD's values equal 1.

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

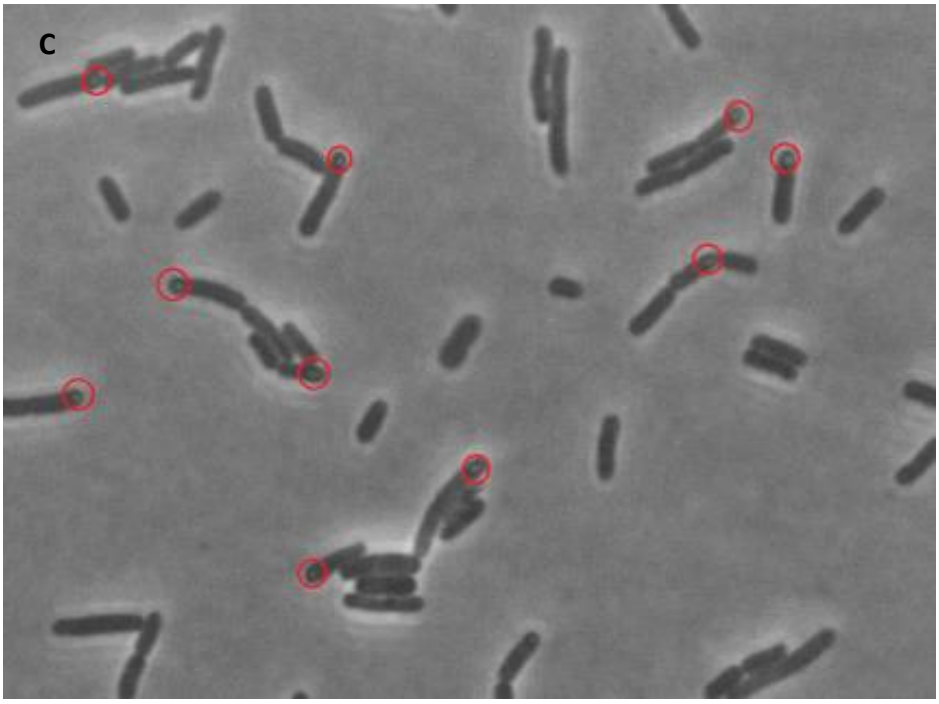


Figure S3. Phase contrast images with cells with marked inclusion bodies (appears as a bright spots), when induced with IPTG concentrations of (Top) 1mM, (Middle) 2mM and (Bottom) 4mM.

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].

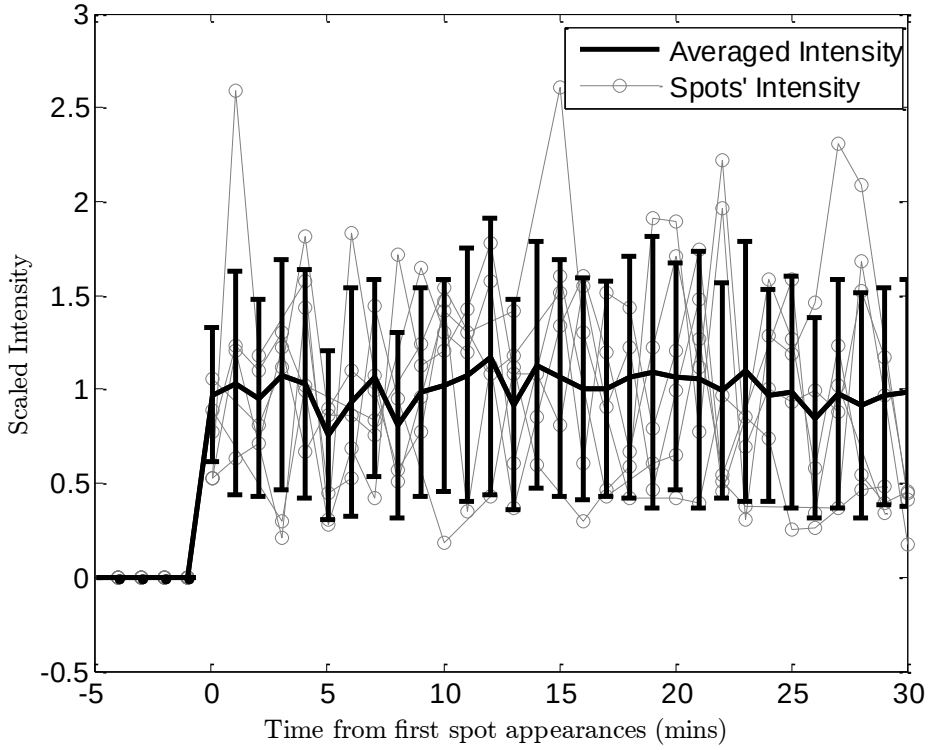


Figure S4. Fluorescence of tagged RNAs in *E. coli* cells over time. Each of the five thin lines shows the fluorescence of a single tagged RNA molecule (randomly selected from the data) since first detected, for a period of 30 minutes. The solid black line shows the mean fluorescence intensity of individual 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 37^{o}C, 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.

STUDY V

VK Kandavalli[+], **H Tran**[+], JG Chandraseelan, AS Ribeiro (2015), "Degree of regulation by σ factor competition in *Escherichia coli* depends on the kinetics of transcription initiation", Submitted. ([+]equal contributions)

Degree of regulation by σ factor competition in *Escherichia coli* depends on the kinetics of transcription initiation

Vinodh K. Kandavalli [1,+], Huy Tran [1,+], Jerome G. Chandraseelan [1] and Andre S Ribeiro[1,] *

[1] Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, Tampere, Finland.

[+]Equal contributions.

Keywords: σ Factors competition; *In vivo* Transcription dynamics; Single RNA detection; Open Complex Formation; Closed Complex Formation

* To whom correspondence should be addressed. Tel: +358408490736; Email: andre.ribeiro@tut.fi

ABSTRACT

Background

In *Escherichia coli*, stress-responding expression of a specific σ factor indirectly downregulates some, but not all genes expressed by other σ factors, due to the σ factor competition for a limited pool of RNA polymerase core enzymes. We studied the unknown mechanisms of selectivity of indirect regulation of promoters whose transcription is primarily initiated by RNAP holoenzymes carrying RpoD.

Results

We performed qPCR and *in vivo* single-RNA measurements of transcription kinetics in various conditions, along with stochastic simulations of σ factor population-dependent models of transcription initiation. We provide empirical and theoretical evidences that the responsiveness of RpoD dependent promoters' kinetics to changes in σ factor numbers other than RpoD, decreases widely as the duration of their closed complex formation decreases and the duration of its open complex formation increases.

Conclusions

We conclude that, in *E. coli*, promoter responsiveness to indirect regulation by σ factor competition is determined by its sequence-dependent and dynamically regulated multi-step initiation kinetics. Finally, we argue that a similar mechanism may exist in eukaryotes.

BACKGROUND

In *Escherichia coli*, cell growth arrests by starvation affect the numbers of several components of the transcription machinery, such as RNA polymerase (RNAP) core enzymes (1, 2) and σ factors (3). It has been

suggested that the active regulation of σ factor numbers is one of the means by which *E. coli* cells implement genome-wide changes in expression rates (4–7) and, consequently, alter growth phase.

Studies (4, 5, 8, 9) reported that, during stationary growth, while most genes have reduced activity compared to that during exponential growth, some have more stable or even enhanced expression. This diversity in responses is, to an extent, made possible by differences in the promoters' selectivity for σ factors (5, 10, 11) and/or changes in transcription factor numbers (5).

Another means by which changes in σ factor numbers cause genome-wide changes in the expression rates is the mechanism of indirect negative regulation (5), which affects rates of transcription initiation (12), where most gene expression regulations occur (13–16). For example, in the stationary growth phase, *rpoS* (σ^{38}) expression is enhanced (11), while in exponential growth it is silenced (3, 17). Since RpoS competes with the house-keeping RpoD (σ^{70}) for a limited pool of RNAP core enzymes (2, 3, 17), increasing its numbers decreases the fraction of RNAP holoenzymes carrying RpoD (5, 18). Consequently, there is a decrease in the transcription rate of most, but not all genes that are expressed by RNAP-RpoD holoenzymes. It is so far unknown why some of these genes are impervious to this change in RpoS numbers (4, 5, 8, 9).

In *E. coli*, transcription initiation is a sequential process (19–22). The first step is the 'closed complex formation', and consists of the binding of the RNAP to the promoter region and the finding of the transcription start site (TSS) (23). It is followed by the 'open complex formation' (19, 20, 24), which consists of the isomerizations of the closed complex, starting with the DNA untwisting at the TSS (22) (this step is usually well approximated as irreversible (25)). Next, the RNAP escapes the promoter and transcription elongation begins (26), ending with the release of a nascent RNA (27).

In vitro measurements of the duration of the events at the promoter region (22, 28) and of elongation and termination (29), showed that in *E. coli* the rate-limiting steps between consecutive RNA productions by active promoters are the closed and open complex formations. Recent *in vivo* measurements using MS2-GFP tagging of RNA (30), which can detect target RNA molecules rapidly and efficiently (28, 31, 32), are in agreement, supporting the existence of 2 to 3 rate-limiting steps between consecutive RNA productions (32–35).

The multi-step nature of transcription initiation and the sequence dependence of the duration of each step allow active promoters to differ widely in dynamics (22, 36). *In vitro* measurements on fully induced variants of the lar promoter, showed that the mean interval between transcription events of these variants

differs by hundreds of seconds (22). Further, promoters also differ widely in range of induction (ratio of production rate between zero and full induction), even when differing only by a couple of nucleotides (21, 22, 37). E.g. while P_{larS17} has an induction range of 500 fold, $P_{larconS17}$ has an induction range of 4.5 fold, even though it only differs by 3 point mutations (37). Genome wide studies suggest that most genes in *E. coli* have an activity such that only less than one to a few RNAs are found in the cell at any given moment (13, 38), in agreement with direct observations that, in several promoters, even when fully induced, it takes hundreds of seconds between transcription events (19, 22, 24, 28, 33, 34), due to the rate-limiting steps, i.e. the closed and open complex formations (19, 39, 40).

The duration of the closed complex formation of any promoter is partially determined by how fast a free promoter is bound by a free floating RNAP holoenzyme carrying the appropriate σ factor (20). Being the number of RNAP core enzymes limited (18), the σ factor numbers are expected to affect the kinetics of transcription initiation of all promoters in a cell (2, 5, 41). For example, increasing the numbers of a given σ factor should increase the number of holoenzymes carrying it. This should not only directly increase the transcription rate of promoters responsive to that σ factor, but also indirectly decrease the transcription rate of all other promoters, by reducing the numbers of RNAP carrying other σ factors (5).

However, as noted above, while some genes are indirectly affected by changes in σ factor numbers, others are not (4, 5, 8, 9). The mechanisms underlying this selectivity are unknown but *in silico* studies using an equilibrium model of σ factor competition suggest that this could be related to promoter strength (2, 41).

Here, we address this issue. First, we measure the effects of changes in RpoS numbers on the *in vivo* transcription dynamics of the promoter P_{BAD}, whose transcription is primarily initiated by RNAP-RpoD holoenzymes (see (38, 42–44) and Supplement). For this, we perform measurements of time intervals between consecutive RNA productions in individual cells of the strain BW25113 in the exponential and stationary growth phases, as the RpoS numbers differ in these two conditions (3, 11, 17), and also observe a deletion mutant for *rpoS*, JW5437-1. We find that P_{BAD} is heavily responsive to changes in RpoS numbers.

Next, we perform both measurements by qPCR of the relative durations of the closed and open complex formation under full induction in each condition as well as an *in silico* analysis of a stochastic model of transcription, and conclude that the high responsiveness of P_{BAD} to changes in RpoS numbers is due to having a closed complex formation with longer duration than its open complex formation.

To support our conclusion, we subject P_{BAD} to lower induction levels and, first, show that the duration of the open complex formation increases more than the duration of the closed complex formation with

decreasing induction. Then, we show that the difference in P_{BAD}'s transcription rates between exponential and stationary phase is smaller when weakly induced, as expected from the smaller duration of the closed complex relative to the open complex formation.

Afterwards, we study another promoter, P_{tetA}, whose transcription is also primarily initiated by RNAP-RpoD (Supplement) but that, unlike P_{BAD}, it has a short-length closed complex formation compared to the open complex formation. We show that, in agreement with our hypothesis, its activity does not differ between exponential and stationary growth phase, neither it is affected by the deletion of the RpoS-coding gene. Finally, we present a stochastic model of transcription initiation with σ factor dependency based on the empirical data, and perform simulations to explore the degree of dependency of transcription dynamics on σ factor numbers as a function of the relative duration of the closed and open complex formation. In end, we discuss the role of the sequence-dependent, multi-step initiation kinetics of promoters in *E.coli* as a determinant factor in their selectivity for regulation by σ factor competition, and the possibility of existence of similar mechanisms in eukaryotes.

RESULTS AND CONCLUSIONS

σ factor numbers in the wild type strain, *rpoS*$^+$, and in the deletion mutant *rpoS*$^-$

We first measured the protein levels of *rpoS*, *rpoD*, and *rpoC* genes in *rpoS*$^+$ (BW25113) and *rpoS*$^-$ (JW5437-1) cells in the exponential and stationary growth phases. From previous studies (3–5, 11, 17, 18), we expect the RpoS numbers to differ between these phases in *rpoS*$^+$ cells, but little differences are expected in the RpoD or RpoC numbers (3–5, 17, 18). Meanwhile, in *rpoS*$^-$ cells no difference is expected due to the deletion of the rpoS gene.

After inducing the specific growth phase (Methods), we performed Western blot of RpoS, RpoD, and RpoC protein levels (Methods). Figure 1 shows that, as previously observed (3, 11, 17), in *rpoS*$^+$ cells, the RpoS levels are higher in the stationary phase. Meanwhile, the protein levels of *rpoC* and *rpoD* genes do not differ greatly between growth phases in either *rpoS*$^+$ or *rpoS*$^-$ cells. Given this limited changes in RpoC numbers, any difference in P_{BAD}'s dynamics between growth phases should not be due to changes in core enzyme numbers. Meanwhile, the lack of changes in RpoD levels implies that, in *rpoS*$^+$ cells, the ratio between RpoS and RpoD numbers differs between the growth phases. Finally, note that, aside from RpoS, the numbers of other σ factors existing in some abundance in *E. coli* (specifically, RpoN, and RpoF) are not expected to differ between these growth phases in this strain (3, 45).

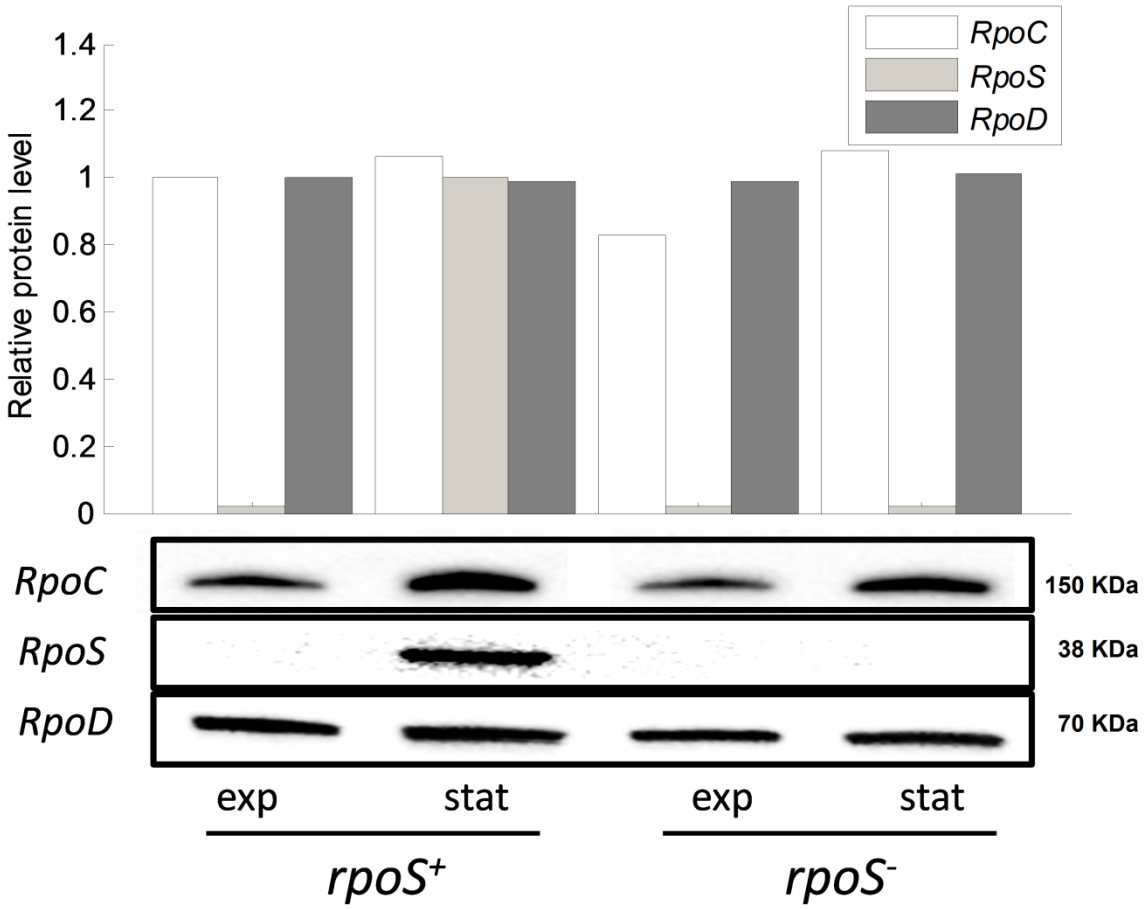


Figure 1. Protein levels of *rpoC*, *rpoS*, and *rpoD* genes in BW25113 (*rpoS*$^+$) and in JW5437-1 (*rpoS*$^-$) cells in the exponential and stationary growth phases, as measured by Western blot. The values are relative

Dynamics of transcription of P_{BAD} as a function of growth phase

Having established the differences in σ factor numbers between conditions, we studied the *in vivo* dynamics of RNA production under the control of P_{BAD} in *rpoS*$^+$ and *rpoS*$^-$ cells in the exponential and stationary growth phases.

For each condition, we performed time-lapse microscopy measurements (2 hours long, with 30 s intervals between consecutive images). The photo-toxicity caused by the confocal microscopy was accessed and found to be insignificant (Supplementary Material). From the time-lapse images, we extracted the duration of the intervals between consecutive RNA appearances in individual cells (Methods) and calculated their mean (μ) and coefficient of variation squared ($CV^2=\sigma^2/\mu^2$). Finally, we performed Kolmogorov-Smirnov (KS) tests of statistical significance to assess whether there is a significant difference between the interval

distributions in the two growth phases (usually it is accepted that, for p-values smaller than 0.01, the null hypothesis that the two distributions are identical is rejected). Results are shown in Table 1.

Table 1

Strain – Phase	No. samples	μ (s)	CV^2	Relative RNA level (qPCR)	p-value (*rpoS*$^+$ vs *rpoS*$^-$)	p-value (Exp. vs Stat.)
rpoS$^+$ – Exp	624	700	0.79	1	0.548	$<10^{-5}$
rpoS$^+$ – Stat	342	1595	0.69	0.11	$<10^{-5}$	
rpoS$^-$ – Exp	368	679	0.76	0.66		$<10^{-5}$
rpoS$^-$ – Stat	244	1053	0.72	0.17		

Table 1. *In vivo* transcription dynamics of P_{BAD} in *rpoS*$^+$ (BW25113) and *rpoS*$^-$ (JW5437-1) cells when in the exponential growth phase and when in the stationary growth phase. Shown are the number of samples of intervals between consecutive transcription events in individual cells, mean intervals duration (μ), coefficient of variation squared ($CV^2=\sigma^2/\mu^2$) of the intervals duration, and p-values of KS tests comparing the interval distributions between growth phases and between *rpoS*$^+$ and *rpoS*$^-$ cells Also shown are the relative RNA levels as measured by qPCR.

From Table 1, both in *rpoS*$^+$ and *rpoS*$^-$ cells, the RNA production dynamics of P_{BAD} differs in the exponential and stationary growth phases (p-value smaller than 10^{-5}), having a much higher mean rate and noise in the former. However, while in *rpoS*$^+$ cells the intervals between consecutive RNA productions are ~2.3 times longer in mean duration in the exponential phase, in *rpoS*$^-$ cells this difference is only of ~1.55 fold, which is significantly smaller than the ~2.3 fold in the *rpoS*$^+$ strain (p-value < 10^{-5}). qPCR measurements of the activity of P_{BAD} in the four conditions are in qualitative agreement with the results from the microscopy measurements (Table 1 and Supplementary Figure S3).

Kinetics of transcription initiation of P_{BAD}

Another important result in Table 1 is that, in all conditions, the CV^2 of the intervals between transcription events is smaller than 1, in agreement with previous reports on the dynamics of this promoter, where no tangible transcriptional bursting was observed (46). The sub-Poissonian nature of the transcription dynamics of this promoter is expected due to the mean length of the intervals between consecutive RNAs (10^2-10^3 seconds), which should allow the buildup of DNA positive supercoiling around the target gene segment to be resolved by DNA gyrases in between events (47, 48).

Given that the transcription dynamics of P_{BAD} is a sub-Poissonian process it can be well-fitted by a model of multiple rate-limiting steps, each exponential in duration (32–35, 46). In Table 2 we show the

results of the statistical inference of the best fit multi-step model of transcription initiation for each condition (Methods). In all conditions, the best fit models have two rate-limiting steps (in agreement with the existence of a closed and an open complex formation, both of which rate-limiting), with one step being significantly longer than the other (Figure S4). As such, transcription initiation of P_{BAD} can be well described by reactions (1), with the last step not being rate limiting (see also reactions (6) and (7) in Supplement).

Table 2

Strain - Phase	No. steps of the best fit model	Duration of steps (s) (95% confidence interval)		p-value	Ratio between longest and shortest inferred step
rpoS$^+$ – Exp	2	109 (±35)	591 (±59)	0.43	5.4
rpoS$^+$ – Stat	2	182 (±112)	1413 (±199)	0.35	7.7
rpoS$^-$ – Exp	2	137 (±51)	542 (±75)	0.37	4.0
rpoS$^-$ – Stat	2	212 (±100)	841 (±147)	0.92	4.0

Table 2. Results of the best fit, for each condition, of the inference of the number and duration of the rate-limiting steps in transcription from the *in vivo* distribution of time intervals between consecutive RNA productions in individual cells under the control of P_{BAD}. For each condition, it is shown the number of steps of the best fit model along with the number and duration of the steps of the model and the 95% confidence interval of the duration of the steps. Also, the results of the KS test comparing the model and the empirical data are shown. As the temporal order of the rate-limiting steps cannot be found by the inference method, we opted for sorting them ascendingly according to their duration. Finally, the ratio between the durations of the longest and shortest inferred steps is shown for each condition.

Finally, from Table 2, the differences between mean transcription intervals (μ (s) in Table 1) in the two growth phases (particularly, in *rpoS*$^+$ cells) appear to be caused by changes in the duration of the longest rate-limiting step alone. Along with the fact that this difference is much smaller in *rpoS*$^-$ cells, we hypothesized that the longest rate-limiting step ought to correspond to the closed complex formation.

Next, we assess whether the longest rate-limiting step in P_{BAD}'s transcription initiation dynamics thus in fact correspond to the closed complex formation.

Relative durations of the closed and open complex formations of P_{BAD}

In order to identify which of the two rate-limiting steps of P_{BAD}, the close or the open complex formation, has longer duration, we make use of the methodology proposed in (49), so as to determine the

duration of the open complex formation (τ_{cc}) relative to the duration of the closed complex formation (τ_{oc}). We performed this test both in *rpoS*$^+$ and *rpoS*$^-$ cells in the exponential growth phase (Methods).

Measurements are performed in cells in exponential phase, in 0.25x, 0.5x and 1x media conditions, so that cells differ in intracellular RNAP concentrations (49). Results are shown in the "τ plot" (Methods) in Figure 2. Also shown is the best fit line, from which we inferred the ratio between τ_{cc} and τ_{oc} for each strain (Methods).

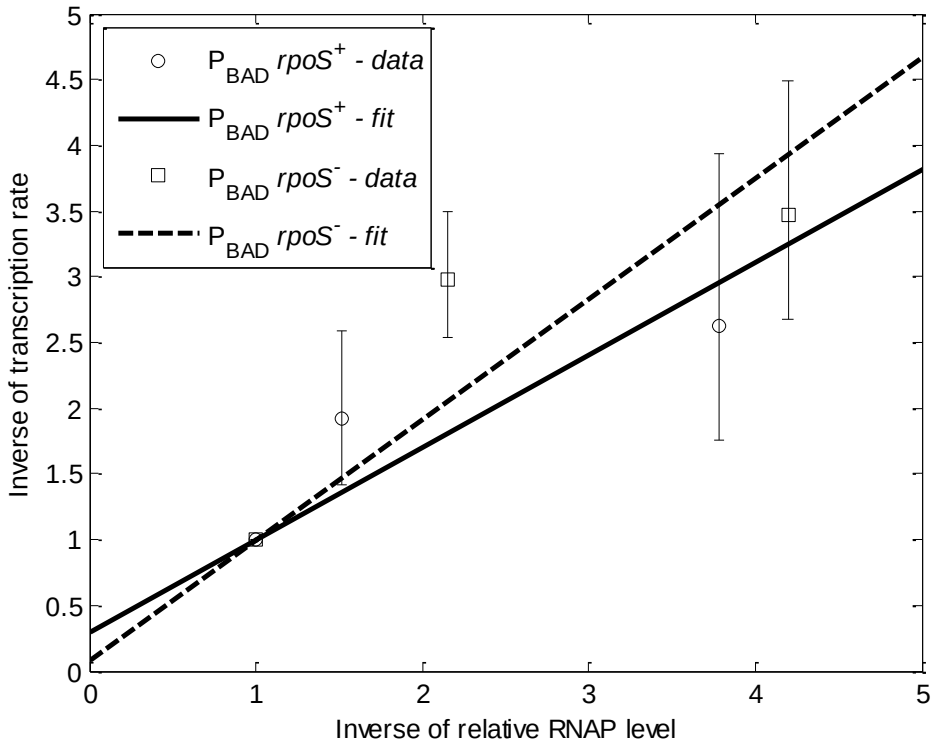


Figure 2. τ plot for P_{BAD} for *rpoS*$^+$ cells and *rpoS*$^-$ cells in the exponential growth phase. All data is shown relative to the RNA and RNAP levels in condition 1x. The lines are Weighted Least Squares fits. The error bars represent the SEM of the estimate of the inverse of the relative level of transcription for the target RNA (vertical bars) in each condition. All errors are calculated including the uncertainty in the first, right-most point in the plot (thus removing the error from that point). The ratio τ_{cc}/τ_{oc} equals ~2.4 for *rpoS*$^+$ cells and ~11.0 for *rpoS*$^-$ cells.

From the best fit lines in Figure 2, the ratio τ_{cc}/τ_{oc} was determined. It equals 2.4 for *rpoS*$^+$ cells and 11.0 for *rpoS*$^-$ cells (Methods). As such, we conclude that τ_{cc} is the rate-limiting step with longest duration.

Note that, qualitatively, the results are in agreement with the microscopy measurements in that this ratio is larger in *rpoS*$^+$ cells than in *rpoS*$^-$ cells (5.4 for *rpoS*$^+$ and 4.0 for *rpoS*$^-$, Table 2).

Relative durations of the closed and open complex formation in P_{BAD} under medium and weak induction in the stationary and the exponential growth phases

From the above, differences in the mean transcription rate of P_{BAD} between growth phases in *rpoS*$^+$ cells (Table 2) are caused by changes in the duration of the longest rate-limiting step, which is the closed complex formation (in agreement with the fact that changes in RpoS numbers should only affect the closed complex formation (reactions (6) and (7) in Supplement)). As such, it is reasonable to hypothesize that the degree of adaptability of the kinetics of transcription of P_{BAD} to changes in RpoS numbers depends on the ratio between τ_{cc} and τ_{oc}. This can be tested, provided that this ratio can be altered, which should be possible by regulating the activation of P_{BAD} by arabinose.

In the absence of arabinose, AraC binds two half-sites on the DNA (I1 and O2), promoting the formation of a DNA loop that blocks the access of RNA polymerases to P_{BAD}. When bound by arabinose, AraC binds instead to the adjacent I1 and I2 half-sites, which promotes transcription initiation of P_{BAD} (25). It is also known that changing the induction strength of P_{BAD} by altering arabinose concentration in the media alters tangibly its RNA production rate (here verified in Figure 7A). This is known to derive from a change in the rate of transcription initiation, but it is yet not well established whether it occurs due to changes in the duration of the open complex formation, of the closed complex formation, or of both closed and open complex formation. Studies using *in vitro* techniques suggest that both steps are altered as they indicate that AraC can enhance both the binding of holoenzymes (which affects the closed complex formation) as well as the transition from the closed to the fully open complex state (22, 25, 35).

Thus, we first assessed whether the closed and open complex formation change by a different amount with changing arabinose concentration. For that, we measured RNA transcription rates by qPCR in cells differing in intracellular RNAP concentrations (as in the previous section), when induced by 0.001%, 0.01% and 0.1% (i.e. full induction) arabinose concentrations. Measurements are performed in *rpoS*$^+$ cells in the exponential phase.

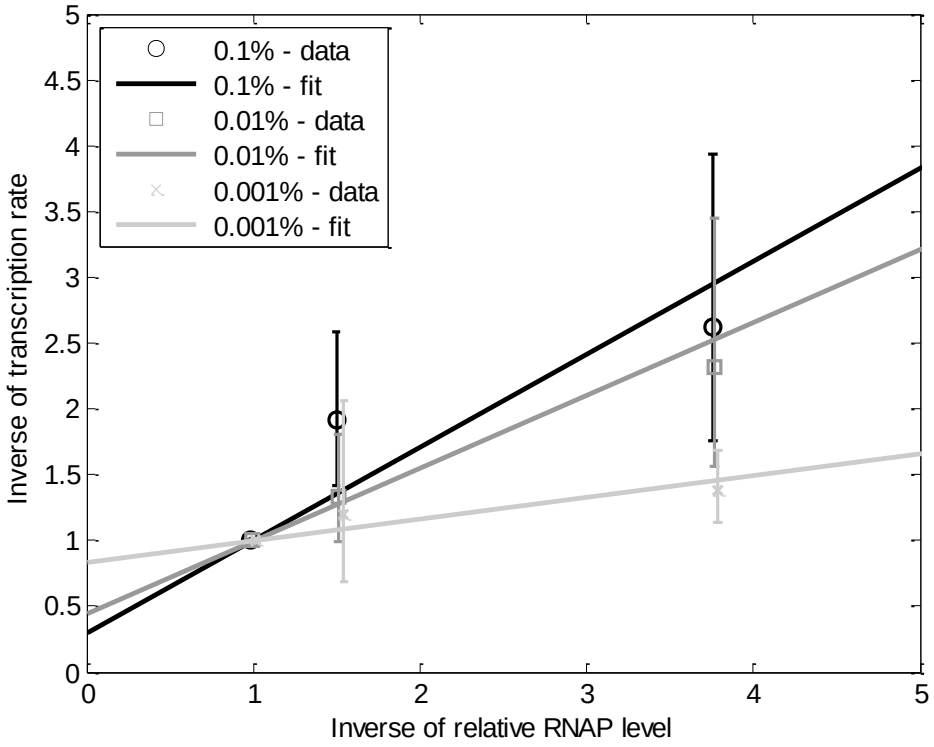


Figure 3. τ plot for P_{BAD} for *rpoS*$^+$ cells in the exponential growth phase in three media conditions (0.001%, 0.01% and 0.1%, i.e. full induction, arabinose concentrations). All data is shown relative to the RNA and RNAP levels in condition 1x. The lines are Weighted Least Squares fits. Error bars represent the SEM of the estimate of the inverse of the relative level of transcription for the target RNA (vertical bars) in each condition. All errors are calculated including the uncertainty in the left-most point in the plot (thus removing the error from that point). The ratio τ_{cc}/τ_{oc} equals ~2.4 at 0.1%, and ~1.2 at 0.01% and ~0.2 at 0.001%.

Results in Figure 3 show that the lower the arabinose concentration, the smaller is the change in RNA production as a function of RNAP concentration (weaker inclination as a function of 1/[*RNAP*]). This indicates that, as the induction strength is decreased, the duration of the open complex formation increases more than the duration of the closed complex formation. Namely, τ_{cc}/τ_{oc} equals 2.4 for 0.1% arabinose, 1.2 for 0.01% arabinose and 0.2 for 0.001% arabinose.

Given this result, it is possible to make the following prediction. If the effects of changing RpoS numbers are stronger in promoters whose closed complex formation is longer than the open complex formation, then the weaker the induction strength of P_{BAD}, the smaller should become the difference in RNA production between exponential and stationary phases.

We tested this in *rpoS*$^{+}$ cells, and the results (Figure 4) verify that our hypothesis is accurate. Visibly, from Figure 4, the weaker the induction by arabinose, the smaller is the difference in expression levels between the exponential and stationary phases. Namely, the fold change in RNA levels when changing from the exponential to the stationary phase equal 9.2 for 0.1% arabinose, 2.5 for 0.01% arabinose and 1.8 for 0.001% arabinose.

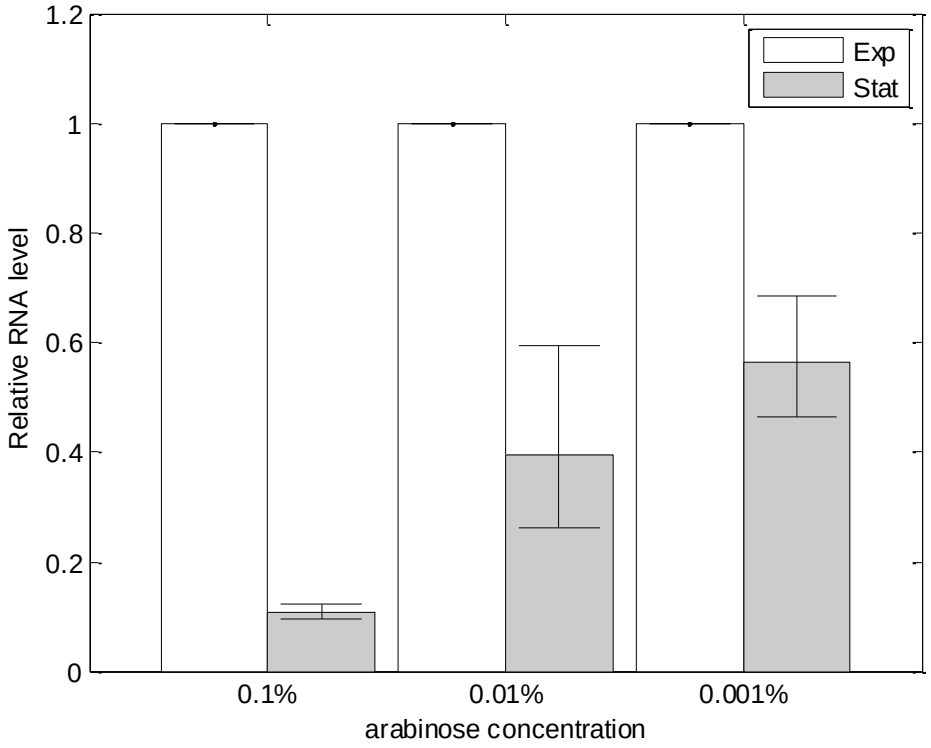


Figure 4: Relative levels of RNA under the control of P_{BAD} in BW25113 (*rpoS*$^{+}$) cells when in the exponential and stationary growth phases for 0.1% arabinose, 0.01% arabinose and 0.001% Arabinose in the media, as measured by qPCR using the 16S RNA housekeeping gene for internal reference. The mean and SEMs (error bars) of relative RNA numbers per cell in each condition were estimated from 3 technical replicates

Dynamics of transcription of P_{tetA} as a function of growth phase

We concluded that the effects of changing RpoS numbers are stronger in promoters whose closed complex formation is longer than the open complex formation. To test this further, we searched for a promoter whose closed complex formation is much shorter in duration than the open complex formation, so as to assess if, in this case, no changes are observed when RpoS numbers, either by changing growth phase, or by deleting the gene coding for RpoS.

In *E. coli*, resistance to tetracycline is conferred by the extra-chromosomal Tn10 transposon encoded class B molecular determinants, forming the tet operon (37, 50–52). This operon consists of two structural genes, *tetA* and *tetR*. *tetA* is essential for tetracycline resistance (53), as it encodes for a membrane-targeted antiporter protein, TetA, responsible for active efflux of tetracycline, whereas *tetR* codes for TetR that regulates the tet operon. In the absence of tetracycline, TetR binds to the operator sites of P_{tetA} and inhibits transcription (54). In the presence of tetracycline, TetR binds as a dimer to the biologically active tetracycline–Mg2+ complex, causing an allosteric conformational change in the repressor protein (55). This releases the repressor from the DNA, allowing the RNAP to bind and initiate transcription of tetA and tetR.

Previous *in vitro* measurements (39) suggest that, for RNAP concentrations similar to those in the *E. coli* cytoplasm (56), P_{tetA} has a much larger τ_{oc} than τ_{cc}. As such, its dynamics should differ neither between growth phases, nor between *rpoS*$^+$ and *rpoS*$^-$ cells.

To test this, we first verified if, *in vivo*, τ_{oc} is indeed longer than τ_{cc}. We conducted these tests in both *rpoS*$^+$ and *rpoS*$^-$ cells in the exponential growth phase (Methods). Results are shown in Figure 5. Also shown are the best fit lines, from which we inferred the ratio τ_{cc}/τ_{oc} (Methods).

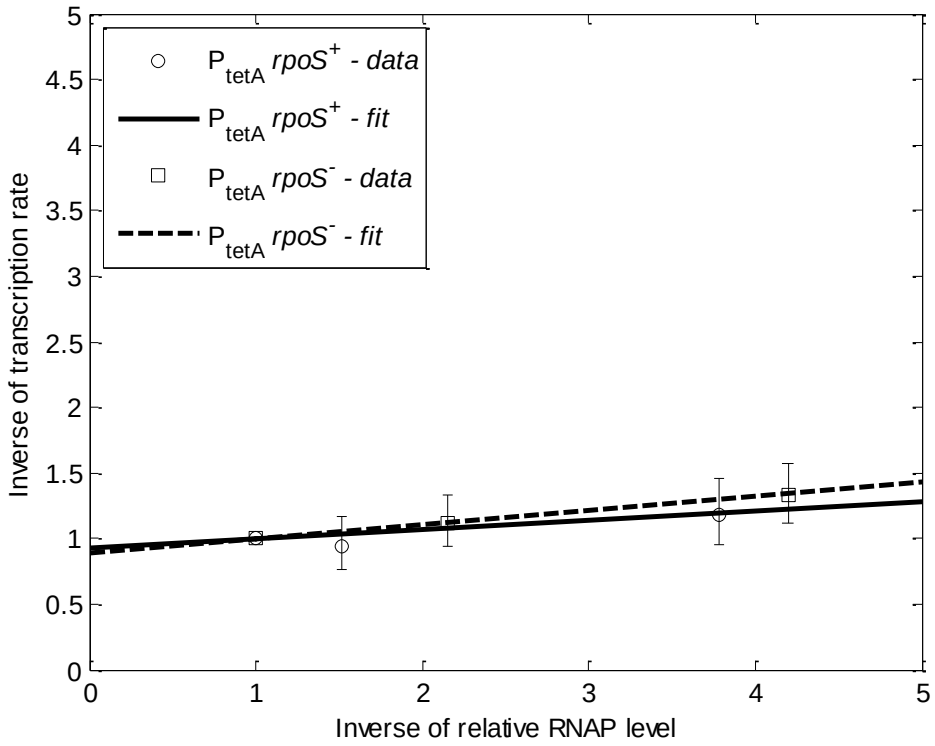


Figure 5. τ plot for P_{tetA} for *rpoS*$^+$ cells and *rpoS*$^-$ cells in the exponential growth phase. All data is shown relative to the RNA and RNAP levels in condition 1x. The lines are Weighted Least Squares fits. The error

bars represent the SEM of the estimate of the inverse of the relative level of transcription for the target RNA (vertical bars) in each condition. All errors are calculated including the uncertainty in the first, right-most point in the plot (thus removing the error from that point). The ratio τ_{cc}/τ_{oc} equals 0.08 for *rpoS*$^+$ cells and 0.12 for *rpoS*$^-$ cells.

First, from Figure 5, the solid and dashed lines of the best fit are nearly superimposed. Also, in both *rpoS*$^+$ and *rpoS*$^-$ cells, for P_{tetA}, the open complex formation is much longer in duration than the closed complex, as τ_{cc}/τ_{oc} equals 0.08 for *rpoS*$^+$ cells and 0.12 for *rpoS*$^-$ cells.

We thus hypothesized that there should be no tangible differences in the dynamics of transcription of P_{tetA} between cells in the exponential and stationary phase neither between *rpoS*$^+$ and *rpoS*$^-$ cells. To test it, we performed measurements of intervals between consecutive transcription events in individual cells (Methods) from time lapse microscopy measurements for P_{tetA} in the two growth phases in *rpoS*$^+$ and *rpoS*$^-$ cells. Results are shown in Table 3, along with the results of tests of statistical significance of the differences between conditions. From the results, we conclude that the dynamics of P_{tetA} does not exhibit tangible differences between the conditions, as predicted.

Table 3:

Strain –Phase	No. samples	μ (s)	CV^2	relative RNA level (qPCR)	p-value (*rpoS*$^+$ vs. *rpoS*$^-$)	p-value (Exp. vs. Stat.)
rpoS$^+$ – Exp	435	982	0.59	1	0.188	0.014
rpoS$^+$ – Stat	447	1157	0.56	0.92	0.994	
rpoS$^-$ – Exp	160	996	0.63	0.90	0.188	0.142
rpoS$^-$ – Stat	192	1156	0.58	0.93	0.994	

Table 3. *In vivo* transcription dynamics of P_{tetA} in *rpoS*$^+$ and *rpoS*$^-$ cells in the exponential and in the stationary growth phase. Shown are the number of samples, mean duration (μ), coefficient of variation squared ($CV^2=\sigma^2/\mu^2$) of the intervals between transcription events. Also shown are the p-values of KS tests comparing the intervals' distributions between growth phases, the relative transcription rates, inferred from the mean production interval, and the relative levels of RNA, as measured by qPCR, in the different growth phases.

Also, from the distribution of intervals between consecutive transcription events obtained from cells in the exponential conditions the ratio τ_{cc}/τ_{oc} equals 0.48 for *rpoS*$^+$ cells and 0.32 for *rpoS*$^-$ cells, which in

qualitative agreement with the qPCR measurements (in that they are much smaller than 1, indicating a longer open complex formation).

Finally, as in the case of P_{BAD}, we performed the statistical inference of the best fit multi-step model of transcription initiation for each condition (Methods). Results in Table 4 show that, in all conditions, the best fit model has two rate-limiting steps (in agreement with the existence of a closed and an open complex formation, both of which rate-limiting) (Figure S4), with one step being significantly longer than the other (in this case, the open complex formation).

Table 4:

Strain - Phase	No. steps of the best fit model	Duration of steps (s) (95% confidence interval)		p-value
rpoS$^+$ – Exp	2	321 (±78)	660 (±88)	0.10
rpoS$^+$ – Stat	2	364 (±93)	793 (±104)	0.60
rpoS$^-$ – Exp	2	240 (±60)	757 (±75)	0.88
rpoS$^-$ – Stat	2	349 (±150)	807 (±173)	0.85

Table 4. Results of the best fit, for each condition, of the inference of the number and duration of the rate-limiting steps in transcription from the *in vivo* distribution of time intervals between consecutive RNA productions in individual cells under the control of P_{tetA}. For each condition, it is shown the number of steps of the best fit model along with the number and duration of the steps of the model and the 95% confidence interval of the duration of the steps. Finally, the results of the KS test comparing model and empirical data are shown. As the temporal order of the rate-limiting steps is unknown, we opted for sorting them according to their duration.

Estimation of the differences in relative expression rates due to changes in σ factor numbers as a function of τ_{cc}/τ_{oc}

Using the model described by reactions (1) to (8) in Supplement, it is possible to predict, for a given value of τ_{cc}/τ_{oc}, the change in transcription rate of a given promoter (whose transcription is primarily initiated by RpoD) has a function of changes in the fraction of RNAP-RpoD holoenzymes. Supplementary Figure S5 shows the predicted results for a wide range of realistic parameter values (i.e. for τ_{cc}/τ_{oc} ranging from 10^{-2} to 10^{2}), along with the predicted results using the measured values of τ_{cc}/τ_{oc} for P_{BAD} and P_{tetA}.

These predictions suggest that altering the kinetics of the rate-limiting steps in initiation (namely, varying τ_{cc}/τ_{oc}) results in a wide range of different behavioral responses to changes in σ factor numbers.

Importantly, examples exist showing that the required changes in τ_{cc}/τ_{oc} are possible either by changing the promoter sequence by a few nucleotides (see e.g. (22) or by changing the induction scheme (e.g. the induction mechanism of P_{lar} allows altering the durations of closed and open complex formation separately by at least 2 orders of magnitude each (37)).

DISCUSSION

One of the global regulatory mechanisms of the transcriptional program of *E. coli* is the population of σ factors (8, 9, 57). Evidence shows that it acts via direct positive regulation as well as by indirect negative regulation. Regarding the former, once a specific σ factor is expressed and forms holoenzymes, a specific array of genes will be activated. Meanwhile, regarding the latter, as different σ factors compete for RNAP core enzymes, increasing the numbers of one σ factor and its corresponding holoenzyme decreases the numbers of other holoenzymes. Consequently, all genes expressed by those other holoenzymes should be, indirectly, downregulated. However, measurements show that not only not all genes expressed by the other holoenzymes are downregulated, but also the degree of down-regulation differs between genes. Given that this selection for down-regulation is not random, there should be a controlling mechanism.

We investigated this by studying the relationship between the dynamics of transcription initiation of promoters whose transcription is primarily initiated by RNAP-RpoD and the degree of indirect regulation caused by changes in the numbers of RpoS. Our findings, based on measurements of P_{BAD} (for three induction regimes), a promoter with a closed complex longer than the open complex, and P_{tetA}, a promoter with a closed complex shorter than the open complex, indicate that the ratio between the durations of the closed and the open complex formation is responsible for promoter selectivity for indirect down-regulation and for the degree of change in transcription rates due to changes in σ factor numbers. Thus, we concluded that the sequence-dependent, multi-step initiation kinetics of promoters in *E.coli* influences their degree of regulation by σ factor competition.

Relevantly, the kinetics of closed and open complex formation are sequence dependent, implying that they are evolvable. Also, they are subject to regulation by, e.g., transcription factors, meaning that the relationship between their time lengths can be readily and widely altered (as shown here for P_{BAD}) to provide, at least to an extent, adaptability to different conditions.

According to our analysis of an *in silico* model, the ‘elasticity’ in the kinetics of the two rate-limiting steps in transcription initiation is expected to provide genes the ability of performing a wide range of different

behaviors in response to changes in σ factor numbers. Nevertheless, we expect this elasticity to vary widely between promoters and to be weaker in genes essential for survival, while being stronger in genes associated to stress or environmental changes responses (as is the case of P_{BAD}).

It is worth noting that the ability to select which genes are responsive to changes in σ factor numbers by indirect negative (or positive) regulation (via the control of the relative durations of closed and open complex formation) may have significant selective advantages when compared to other regulatory mechanisms (e.g. transcription factor regulation). For example, it does not require the production of regulatory molecules and, thus, saves energy, cellular components, and it is not subject to errors in production (a common problem in protein production in *E. coli*).

We find plausible the existence of a similar mechanism in eukaryotic cells. In these, for transcription to be initiated, a TFIID factor must first bind to the TATA box. This factor is composed of both a TBP (TATA binding protein) as well as 1 out of at least 15 different TAFs (TBP associated factors), and has been described as being a 'relative' of σ factors in *E. coli* (58). It has also been proposed that the needed association between TAFs and TBP allows for the coordinated regulation of transcription in eukaryotes, similar to σ factors in bacteria (59), as distinct sets of TAFs likely dictate the type of promoter at which a given TFIID will function. By regulating the intracellular numbers of one TAF, it should be possible to indirectly up/down regulate the speed of transcription initiation of a promoter not transcribed by that TAF, provided that there is competition in the cell for TBP factors. Evidence for such competition exists, in that TBP is not found in an isolated form *in vivo* (60). In that case, promoters whose first rate limiting step is longer in duration than subsequent steps would likely have their transcription activity indirectly affected by changes in the numbers of TAF not necessary for its transcription.

We note that we do not expect the differences in RpoS numbers to be the only cause for the differences in transcription dynamics of P_{BAD} between the conditions observed. An evidence for this is the difference in P_{BAD}'s dynamics between the exponential and stationary growth phases in *rpoS*$^-$ cells. Nevertheless, the differences in RpoS numbers are certainly one of the main causes, as the difference between exponential and stationary growth phases is much weaker in the deletion mutants than in the control strain. In the future, it should be of interest to study what other factor (not necessarily a σ factor) is responsible for the 'remaining' difference in dynamics between growth phases. Whatever it is, it does not act on P_{tetA}, which exhibited no dynamical changes between growth phases in either *rpoS*$^+$ or *rpoS*$^-$ cells.

Finally, as a side issue, previous analysis of time-lapsed, single RNA detection data allowed extracting the distributions of intervals between consecutive RNA productions in individual cells (33–35), which, by being sub-Poissonian, can be modeled by a multi-step process with each step being exponentially distributed in duration and allow extracting from them the number and duration of the rate-limiting steps in transcription initiation of a promoter (33–35). However, it is not possible to extract the order of occurrence of these steps studying the distributions of the wild type strain alone. The present work shows that studying the same dynamics in single deletion mutants of a specific σ factor solves this problem, by allowing the identification of which of the steps is the closed complex formation. Relevantly, we expect this methodology to have significant genome wide applicability (i.e. to any promoter primarily expressed by RNAP-RpoD holoenzymes).

MATERIALS AND METHODS

Strains and plasmids

The *E. coli* strains used are BW25113 (61) and a deletion mutant, JW5437-1 (62), lacking the gene encoding for RpoS (62), were obtained from the Keio single-gene knockout collection. For simplicity, we denote BW25113 as *rpoS*$^+$ and JW5437-1 as *rpoS*$^-$. Note that these strains do not contain the *araB-araD* genes. As such, the negative feedback loop of the arabinose utilization system is inactive (63), which allows attaining maximum RNA production when fully inducing P_{BAD} by arabinose.

We inserted an MS2-GFP RNA tagging system (64) into both strains to monitor the dynamics of transcript production of the promoter of interest. The tagging system consists of a low-copy reporter plasmid (pZS12MS2-GFP) carrying P_{lac}-*ms2-gfp* (generously provided by Philippe Cluzel, Harvard University, MA, USA) and a single-copy target plasmid (pTRUEBLUE-BAC) coding for mRFP1 protein and an RNA sequence with 96 target binding sites (96bs) for MS2-GFP, controlled by the promoter of interest, which is either P_{tetA} (33) or P_{BAD} (46). Given their sequences, both promoters are expected to be transcribed by RNAP holoenzymes carrying RpoD (Supplement).

Media conditions and growth phase induction

The method used here to reach specific growth phases was proposed in (65). Cells were grown in LB media (10 g/l tryptone, 5 g/l yeast extract and 10 g/l NaCl) overnight in an orbital shaker at 30°C with aeration at 250 rpm. Afterwards, to induce exponential growth phase, cells were diluted in fresh LB media to

reach an optical density (OD600) of ~0.05 and then grown at 37°C with aeration at 250 rpm for 1 hour. Meanwhile, to induce the stationary growth phase, cells were diluted in a stationary phase inducing media. Namely, the cells were placed on a media obtained by centrifuging the overnight cultured cells at 10000 rpm for 10 minutes. Next, cells were allowed to growth at 37°C with aeration at 250 rpm for 1 hour. As shown by the growth curve analysis, this halts cell growth (Figure S1). Finally, to assess whether the addition of arabinose alters cell growth rates, we measured the cells' OD600 over time using a spectrophotometer. From the OD curves, we verified that the growth phases of neither the wild-type nor the deletion mutant strain were altered significantly for at least 2 hours after the addition of arabinose to the media (Figure S1).

Intracellular RNA polymerase concentrations

The concentration of RNA polymerases in the cells, which includes both core enzymes and holoenzymes, differs with media richness (49). Since all measurements are carried out in LB medium, to obtain cells with differing RNAP concentrations, we placed them in modified LB media. These media have lower tryptone and yeast extract concentrations in order to reduce the RNAP concentration in the cells. The media used are denoted as 0.25x, 0.5x, and 1x, and their composition per 100 ml is, respectively: (0.25x) 0.25 g tryptone, 0.125 g yeast extract and 1 g NaCl (pH – 7.0); (0.5x) 0.5 g tryptone, 0.25 g yeast extract and 1 g NaCl (pH – 7.0); and (1x) 1 g tryptone, 0.5 g yeast extract and 1 g NaCl (pH – 7.0).

The resulting concentrations of RNAP were assessed by measuring the level of RpoC protein, a core subunit of RNAP (66) using Western blot (see Method). These measurements confirmed that the RNAP levels are reduced in the poor media (Figure S6).

Time-lapse microscopy

One hour after incubation in the respective phase-inducing media, for both cells containing P_{BAD} or P_{tetA}, we induced the MS2-GFP reporter, under the control of P_{lac}, with 1 mM IPTG for 45 minutes in liquid culture. We verified by inspection that, at this stage, cells contained sufficient, uniformly distributed MS2-GFP in the cytoplasm to detect target RNAs (30, 32–35, 46). To activate P_{BAD}, we added 0.1% of arabinose for 5 minutes while in liquid culture. Cells were then placed under the microscope. P_{tetA} does not require induction, as cells from BW25113 and JW5437-1 strains lack the gene coding for TetR, the repressor of P_{tetA} (62).

We used the following microscopy setting to image cells. We used an inverted microscope body Nikon Eclipse (Ti-E, Nikon, Japan) for all experiments. In the left port of the body, we used a confocal C2+

scanner connected to LU3 laser system (Nikon) with a 488 nm argon ion laser. The laser shutter was open only during the exposure time, to minimize photobleaching. For phase-contrast, we used an external phase-contrast setting (Nikon) with a DS-Fi2 CCD-camera. For both phase-contrast and confocal imaging, we used a 100x oil-immersion objective (Apo TIRF, Nikon). The software used to capture both image types and control the microscope was NIS-Elements (Nikon). To maintain, during microscopy, stable growth conditions and induction of the promoters controlling the production of the RNA target for MS2-GFP and of MS2-GFP, a peristaltic pump was used to introduce a constant flow of the appropriate media and inducers.

For imaging, a few µl of culture was placed between a glass coverslip and a slab of 3% agarose containing the respective media and inducers. During image acquisition, the slide was kept in a temperature-controlled chamber (Bioptechs, FCS2) at 37 °C. Cells were imaged by the confocal microscope every 30 seconds for 2 hours, for the purpose of cell segmentation and detecting the MS2-GFP tagged RNA molecules (which appear as a bright spots in the image). Example images of cells along with the results of cell segmentation, spot detection, and extraction of the spots total fluorescence intensity over time are shown in Figure 6.

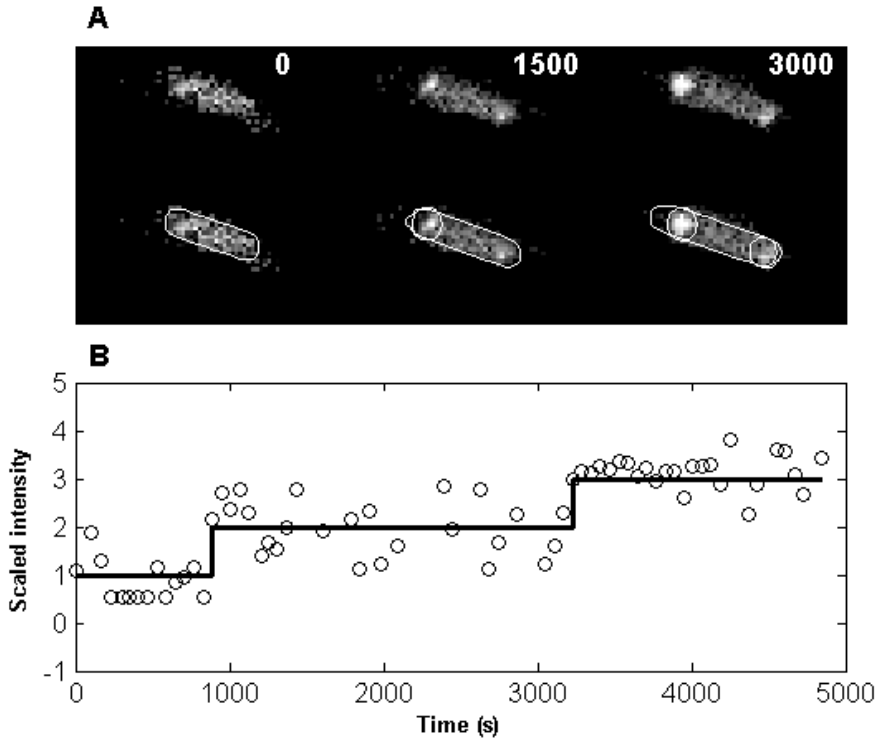


Figure 6. Tagged RNAs in *E. coli* cells over time. (A) Unprocessed frames (top) along with segmented cells and RNA spots (bottom). The moment when the images were captured is shown at the top of each frame.

(B) Example time series of a scaled spot intensity level in a cell (circles) and the corresponding estimated RNA numbers (solid line).

Image processing and data extraction

Image analysis was performed as in (33, 35, 46). We used a semi-automated cell segmentation strategy as in (67), which uses the software MAMLE (31) followed by manual correction. Afterwards, fluorescent spots in each cell, at each time moment, are detected automatically as in (34), by estimating the cell background intensity distribution using its median and median absolute deviation, and then performing thresholding with a given confidence level and assuming that this distribution is Gaussian.

For time-series analysis, since the MS2-GFP tagged RNA molecules do not degrade during the measurement time (33, 68, 69), the moments of appearance of novel target RNAs in each cell were obtained as in (35), by least squares fitting a monotonically increasing piecewise-constant function to the corrected total spot intensity in that cell over time (Supplement, section 7). The number of terms for the fitting 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 (Figure 6) (32–35, 46). Finally, the time intervals between consecutive RNA production events in each cell were extracted (events separated by cell divisions are not considered). This method, first proposed in (70) and subsequently improved in (71), allows estimating the accuracy of the estimation based on the number of cells observed and the level of noise of the 'spot fluorescence' signal from individual cells (71). For a set of 300 cells and a noise level of 1 in the fluorescent signal (as measured by the coefficient of variation and in agreement with the present measurements), we expect an accuracy of 0.8 (71).

An example of the application of this method to the signal from one cell is shown in Figure 6.

Inference of the number and durations of the sequential rate-limiting steps in transcription from the measured intervals between transcription events in individual cells

The number and durations of rate-limiting steps in transcription initiation are inferred by fitting a multi-step model, where the steps are sequential and exponentially-distributed, to the intervals between consecutive transcription events in individual cells using maximum likelihood (72). For a given number of steps, d, the distribution to fit is the sum of d exponential variables with mean durations μ_i ($i=1..d$). The model with smallest d is selected that cannot be rejected at the significance level 0.01 in favor of a higher order model. To assess the error of the inference, we also computed the 95% confidence interval of the step

durations inferred from time intervals. Note that this method does not allow us to determine the temporal order of the sequential steps inferred. Only their number and durations can be assessed.

Quantitative PCR

Cells (5 ml) at different phases were harvested as described previously, followed by an addition of 10 ml of RNA protect bacteria reagent and mixed immediately by vortexing for 5 seconds. The 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 by using RNeasy kit (Qiagen) according to the instructions of the manufacturer. Samples were then quantified using a Nanovue plus spectrophotometer (GE Healthcare life sciences), and the quality of the isolated RNA was checked by measuring the ratio of the 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. cDNA was synthesized (Fermentas, Finland) from 1 µg of RNA with iScript Reverse Transcription Supermix. The cDNA templates with a final concentration of 10 ng/µl were added to the qPCR master mix containing iQ SYBR Green supermix (Fermentas, Finland) with primers for the target and reference genes at a final concentration of 200 nM. We used the 16S RNA housekeeping gene for internal reference. The primer set for the target mRNA (mRPF1) was (Forward: 5’ TACGACGCCGAGGTCAAG 3’ and Reverse: 5’ TTGTGGGAGGTGATGTCCA 3’) and for the reference gene (16S RNA) was (Forward: 5’ CGTCAGCTCGTGTTGTGAA 3’ and Reverse: 5’ GGACCGCTGGCAACAAAG 3’). For *rpoS*, *rpoD* and *rpoC* RNA quantification, we used a specific set of primers, i.e. for the *rpoS* gene (Forward: 5’TATTCGTTTGCCGATTCACA 3’ and Reverse: 5’ CTTGGTTCATGGTCCAGCTT 3), the *rpoD* gene (Forward: 5’GATCTGATCACCGGCTTTGT 3’ and Reverse: 5’ TCTTCCTGGGAAAGCTCAGA 3’) and the *rpoC* gene (Forward: 5’ CGTCAGATGCTGCGTAAAGC 3’ and Reverse: 5’ GCGATCTTGACGCGAGAGTA 3’). The qPCR experiments were performed using a Biorad MiniOpticon Real time PCR system (Biorad, Finland). The following thermal cycling protocol was used: 40 cycles of 95°C for 10s, 52°C for 30 s, and 72°C for 30 s for each cDNA replicate. These reactions were performed in three replicates for each condition with a final reaction volume of 25 µl. We use no-RT controls and no- template controls to crosscheck non-specific signals and contamination. PCR efficiencies of these reactions were greater than 95%. The data from CFX Manager TM Software was used to calculate the relative gene expression and its standard error (73).

Western blotting

E.coli strains were grown as described above and cells were harvested by centrifuging at the OD600 of 0.3. Cells were then lysed with the B-PER Bacterial protein extraction reagent (Thermo Scientific) containing the protease inhibitors. Cell lysate was incubated in the room temperature for 10 mins and then centrifuged at 15000 ×g for 5 mins to remove the debris and collected the supernatant. The samples containing the total protein were diluted with the 4X lamella sample loading buffer containing the β mercaptomethanol and boiled for 5 mins at 95°C. 30 µg of the total protein in each sample was resolved by 4 – 20 % TGX gels (Biorad). Proteins were separated by electrophoresis and then electro-transferred on the PVDF membrane (Biorad). Membranes were incubated with respective primary antibodies for RpoC, RpoS and RpoD (Biolegend) of 1:2000 dilution overnight at 4°C followed by the appropriate HRP-secondary antibodies (Sigma Aldrich) 1:5000 dilution for 1 hour at room temperature. Detection was done by the chemilumiscence reagent (Biorad). Images were generated by the chemidoc XRS system (Biorad). Band quantification was done by using Image Lab software version 5.2.1.

Estimation of the ratio between the closed and open complex formation

To estimate the ratio between the durations of closed and open formation of a target promoter, we followed a strategy similar to the one used in (49). Given the multi-step model of transcription described in (19, 23, 28).

$$RNAP + Pr \xrightarrow{k_{cc}} Pr_c \xrightarrow{k_{oc}} Pr_o \xrightarrow{\infty} Pr + RNAP + M \qquad (1)$$

where *RNAP* is the active RNA polymerase, *Pr* is a free promoter, Pr_c is a closed complex, Pr_o is an open complex (fully formed). Here, promoter escape and transcription elongation are assumed to be infinitely fast due to their duration being at least one order of magnitude smaller than the duration of transcription initiation (26, 29, 74). The duration of the closed complex formation (τ_{cc}) is inversely proportional to the abundance of RNAP ($[RNAP]$):

$$\tau_{cc} = \frac{1}{k_{cc}[RNAP]} \qquad (2)$$

Here, k_{cc} is the rate constant of the closed complex formation in the condition when $[RNAP] = 1$. Provided that the closed and open complex formations are the only main rate-limiting steps in transcription (19, 22, 23, 28) the interval between consecutive RNA production events can be given by:

$$\Delta t = \tau_{cc} + \tau_{oc} = \frac{1}{k_{cc}[RNAP]} + \frac{1}{k_{oc}} \quad (3)$$

where Δt are the intervals between consecutive transcription events, while τ_{oc} is the duration of the open complex formation.

The variables in equation (3) whose values can be obtained from the measurements are [*RNAP*] and Δ*t*. In particular, as Δ*t* is inversely proportional to the target RNA production rate, it can be extracted by qPCR measurements (13, 16). Similarly, relative values of [*RNAP*] can be obtained by protein immunoblot. Then, it is possible to estimate the value of τ_{cc}/τ_{oc}, by weighted least square fit of a line to the measured Δt values when plotted against the 1/[*RNAP*] relative numbers.

In practice, we obtain the value of τ_{cc}/τ_{oc} by measuring both [*RNAP*] and 1/Δt, in three conditions, differing in the abundance of RNAP in cells. To alter the abundance of RNAP, we place cells in different media, with chemical compositions of 1x, 0.5x and 0.25x (see above). The growth curves show that the cells are in the exponential phase in all three conditions (Figure S2). As the RNA degradation rates are not affected by the growth conditions (16), the relative transcription rates (proportional to 1/Δt) at 0.5x and 0.25x when compared to that in 1x media are assessed from the RNA levels, as measured by qPCR.

Validation of promoter activity for P_{BAD} and P_{tetA} in BW25113

To confirm the presence of P_{BAD} in BW25113, a derivative of *E. coli* K-12, and to determine the minimum concentration of arabinose needed to maximize its induction, we measured from microscopy images, for varying inducer concentration, the mean number of RNAs per cell. Results are shown in Figure 7A. We also measured the fold change in the RNA synthesis rate by qPCR, which is in agreement with the microscopy measurements (Figure 7A). From the measurements, 0.1% arabinose suffices to reach maximum induction of P_{BAD}. Also, from both measurement methods, we observe a change of ~10 fold induction ratio in RNA production (similarly to the data reported in (46)).

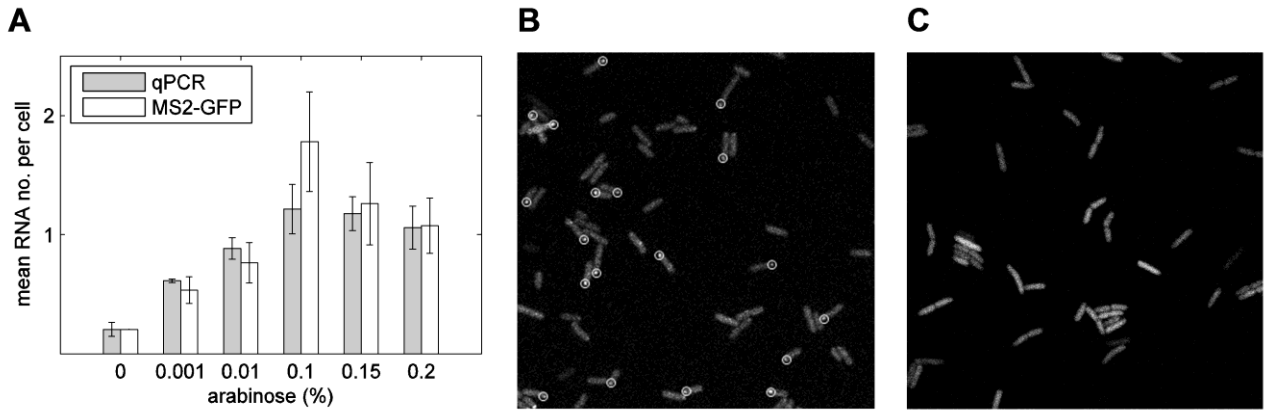


Figure 7. (A) Induction curve of P_{BAD} with varying inducer concentration obtained from images of cells carrying P_{BAD} and the MS2-GFP tagging system (white) and from qPCR (grey). The data from qPCR was normalized so as to match the microscopy measurements at 0% arabinose. In both measurement procedures, the mean and standard errors of the means (SEMs) (error bars) of RNA numbers per cell in each condition were estimated from 3 technical replicates. (B) Cells with the plasmid containing P_{tetA}-mrfp1-96bs under confocal microscope and no target gene inducer. White circles were placed around fluorescent spots identified by the spot detection algorithm. (C) Cells without the plasmid containing target genes under confocal microscope and no target gene inducer. No fluorescent spots were identified

In the case of P_{tetA}, it is only possible to observe its dynamics at full induction, since tetR is absent in BW25113 (61). To verify that the promoter is active, we compared, by inspection, the number of fluorescent spots in cells with and without the target plasmid carrying P_{tetA}-mrfp1-96bs, 1 hour after inducing the reporter plasmid. While for cells with the target plasmid (example Figure 7B), spots appear under the confocal microscope (on average ~0.5 spot per cell), in the absence of the target plasmid (example Figure 7C) we detected no spots (500 cells observed), from which we conclude that P_{tetA} is active. From this, it is also possible to infer that the number of false positives in cells containing the target plasmid is negligible.

Given these results, in experiments on both *rpoS*$^+$ as well as *rpoS*$^-$ we used arabinose concentration of 0.1% to fully induce P_{BAD}, unless stated otherwise. Note that the transcription rate of this promoter for concentrations of 0.1%, 0.01% and 0.001% differs significantly between conditions. As for P_{tetA}, we only induce the MS2-GFP reporter system. Finally, note that the mean RNA numbers as measured by the MS2-GFP tagging system does not differ significantly between replicates.

ACKNOWLEDGMENTS

We thank Jason Lloyd-Price for valuable advice.

FUNDING

Work supported by Academy of Finland [257603 to ASR), Portuguese Foundation for Science and Technology [PTDC/BBB-MET/1084/2012 to ASR] and Centre of International Mobility (13.1.2014/TM-14-91361/CIMO, VK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY DATA

Supplement File.

AUTHORS' CONTRIBUITION

ASR, VK and HT conceived the study. VK and JG performed experiments. ASR, VK and HT analyzed results. All authors wrote the manuscript.

REFERENCES

1. Cabrera,J.E. and Jin,D.J. (2001) Growth phase and growth rate regulation of the rapA gene, encoding the RNA polymerase-associated protein RapA in *Escherichia coli*. *J. Bacteriol.*, **183**, 6126–6134.

2. Grigorova,I.L., Phleger,N.J., Mutalik,V.K. and Gross,C. a (2006) Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 5332–5337.

3. Jishage,M., Iwata,A., Ueda,S. and Ishihama,A. (1996) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli* : intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.*, **178**, 5447–5451.

4. Rahman,M., Hasan,M.R., Oba,T. and Shimizu,K. (2006) Effect of rpoS gene knockout on the metabolism of *Escherichia coli* during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular metabolite concentrations. *Biotechnol. Bioeng.*, **94**, 585–595.

5. Farewell,A., Kvint,K. and Nyström,T. (1998) Negative regulation by RpoS: A case of sigma factor competition. *Mol. Microbiol.*, **29**, 1039–1051.

6. Rouvière,P.E., De Las Peñas,A., Mecsas,J., Lu,C.Z., Rudd,K.E. and Gross,C.A. (1995) *rpoE*, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *EMBO J.*, **14**, 1032–1042.

7. Dong,T. and Schellhorn,H.E. (2009) Control of RpoS in global gene expression of *Escherichia coli* in minimal media. *Mol. Genet. Genomics*, **281**, 19–33.

8. Tani,T.H., Khodursky,A., Blumenthal,R.M., Brown,P.O. and Matthews,R.G. (2002) Adaptation to famine: A family of stationary-phase genes revealed by microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 13471–13476.

9. Chang,D., Smalley,D.J. and Conway,T. (2002) Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol. Microbiol.*, **45**, 289–306.

10. Hengge-Aronis,R. (2002) Recent insights into the general stress response regulatory network in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.*, **4**, 341–346.

11. Lange,R.P. and Hengge-Aronis,R. (1991) Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol. Microbiol.*, **5**, 49–59.

12. Gruber,T.M. and Gross,C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.*, **57**, 441–66.

13. Taniguchi,Y., Choi,P.J., Li,G.-W., Chen,H., Babu,M., Hearn,J., Emili,A. and Xie,X.S. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533–8.

14. Yarchuk,O., Guillerez,J. and Dreyfus,M. (1992) Interdependence of translation, transcription and mRNA degradation in the ZacZ Gene. *J. Mol. Biol.*, **226**, 581–596.

15. Lodish,H., Berk,A., Zipursky,S.L., Matsudaira,P., Baltimore,D. and Darnell,J. (2000) Molecular Cell Biology.

16. Chen,H., Shiroguchi,K., Ge,H. and Xie,X.S. (2015) Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.*, **11**, 781.

17. Jishage,M. and Ishihama,A. (1995) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular Levels of σ70 and σ38. *J. Bacteriol.*, **177**, 6832–6835.

18. Maeda,H., Fujita,N. and Ishihama,A. (2000) Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.*, **28**, 3497–3503.

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

20. Saecker,R.M., Record,M.T. and Dehaseth,P.L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.*, **412**, 754–771.

21. Browning,D.F. and Busby,S.J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, **2**, 57–65.

22. 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**, 3873–3881.

23. McClure,W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.*, **54**, 171–204.

24. DeHaseth,P.L. and Helmann,J.D. (1995) Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol. Microbiol.*, **16**, 817–824.

25. Schleif,R. (2010) AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.*, **34**, 779–796.

26. Hsu,L.M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta - Gene Struct. Expr.*, **1577**, 191–207.

27. Uptain,S.M., Kane,C.M. and Chamberlin,M.J. (1997) Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.*, **66**, 117–172.

28. Buc,H. and McClure,W.R. (1985) Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry*, **24**, 2712–2723.

29. Herbert,K.M., La Porta,A., Wong,B.J., Mooney,R.A., Neuman,K.C., Landick,R. and Block,S.M. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell*, **125**, 1083–94.

30. Golding,I., Paulsson,J., Zawilski,S.M. and Cox,E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, **123**, 1025–1036.

31. Chowdhury,S., Kandhavelu,M., Yli-Harja,O. and Ribeiro,A.S. (2012) An interacting multiple model filter-based autofocus strategy for confocal time-lapse microscopy. *J. Microsc.*, **245**, 265–75.

32. Muthukrishnan,A.-B., Martikainen,A., Neeli-Venkata,R. and Ribeiro,A.S. (2014) *In vivo* transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells. *PLoS One*, **9**, e109005.

33. Muthukrishnan,A.B., Kandhavelu,M., Lloyd-Price,J., Kudasov,F., Chowdhury,S., Yli-Harja,O. and

Ribeiro,A.S. (2012) Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucleic Acids Res.*, **40**, 8472–8483.

34. Kandhavelu,M., Häkkinen,A., Yli-Harja,O. and Ribeiro,A.S. (2012) Single-molecule dynamics of transcription of the lar promoter. *Phys. Biol.*, **9**, 026004.

35. Kandhavelu,M., Lloyd-Price,J., Gupta,A., Muthukrishnan,A.B., Yli-Harja,O. and Ribeiro,A.S. (2012) Regulation of mean and noise of the *in vivo* kinetics of transcription under the control of the lac/ara-1 promoter. *FEBS Lett.*, **586**, 3870–3875.

36. Record,M.T.J., Reznikoff,W.S., Craig,M.L., McQuade,K.L. and Schlax,P.J. (1996) *Escherichia coli* RNA polymerase (Es70), promoters, and the kinetics of the steps of transcription initiation. 2nd ed. American Society for Microbiology, Washington, DC.

37. Lutz,R. and Bujard,H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.*, **25**, 1203–1210.

38. Becker,G. and Hengge-Aronis,R. (2001) What makes an *Escherichia coli* promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S). *Mol. Microbiol.*, **39**, 1153–65.

39. Bertrand-Burggraf,E., Lefèvre,J.F. and Daune,M. (1984) A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322. *Nucleic Acids Res.*, **12**, 1697–1706.

40. Chamberlin,M.J. (1974) The selectivity of preparation. *Psychol. Rev.*, **81**, 442–464.

41. Mauri,M. and Klumpp,S. (2014) A model for sigma factor competition in bacterial cells. *PLoS Comput. Biol.*, **10**, e1003845.

42. Wise,A., Brems,R., Ramakrishnan,V. and Villarejo,M. (1996) Sequences in the -35 region of *Escherichia coli* rpoS-dependent genes promote transcription by E.RpoS. *J. Bacteriol.*, **178**, 2785–2793.

43. Colland,F., Fujita,N., Kotlarz,D., Bown,J.A., Meares,C.F., Ishihama,A. and Kolb,A. (1999) Positioning of σS, the stationary phase σ factor, in *Escherichia coli* RNA polymerase-promoter open complexes. *EMBO J.*, **18**, 4049–4059.

44. Barrios,H., Valderrama,B. and Morett,E. (1999) Compilation and analysis of σ54-dependent promoter sequences. *Nucleic Acids Res.*, **27**, 4305–4313.

45. Dong,T., Yu,R. and Schellhorn,H. (2011) Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in *Escherichia coli*. *Mol. Microbiol.*, **79**, 375–86.

46. Mäkelä,J., Kandhavelu,M., Oliveira,S.M.D., Chandraseelan,J.G., Lloyd-Price,J., Peltonen,J., Yli-Harja,O. and Ribeiro,A.S. (2013) *In vivo* single-molecule kinetics of activation and subsequent activity of the arabinose promoter. *Nucleic Acids Res.*, **41**, 6544–6552.

47. Reece,R.J. and Maxwell, a (1991) DNA gyrase: structure and function. *Crit. Rev. Biochem. Mol. Biol.*, **26**, 335–375.

48. Chong,S., Chen,C., Ge,H. and Xie,X.S. (2014) Mechanism of rranscriptional bursting in bacteria. *Cell*, **158**, 314–326.

49. Liang,S., Bipatnath,M., Xu,Y., Chen,S., Dennis,P., Ehrenberg,M. and Bremer,H. (1999) Activities of constitutive promoters in *Escherichia coli*. *J. Mol. Biol.*, **292**, 19–37.

50. Hillen,W., Gatz,C., Altschmied,L., Schollmeier,K. and Meier,I. (1983) Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. *J. Mol. Biol.*, **169**, 707–721.

51. Gossen,M. and Bujard,H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 5547–5551.

52. Nguyen,T.N., Phan,Q.G., Duong,L.P., Bertrand,K.P. and Lenski,R.E. (1989) Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli* K12. *Mol. Biol. Evol.*, **6**, 213–225.

53. Beck,C.F., Mutzel,R., Barbit,J. and Muller,W. (1982) tetracycline resistance . A multifunctional gene (

tetR) controls TnJO-encoded tetracycline resistance. **150**, 633.

54. Hillen,W. and Berens,C. (1994) Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.*, **48**, 345–369.

55. Orth,P., Schnappinger,D., Hillen,W., Saenger,W. and Hinrichs,W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.*, **7**, 215–219.

56. Bremer,H., Dennis,P. and Ehrenberg,M. (2003) Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie*, **85**, 597–609.

57. Dong,T. and Schellhorn,H.E. (2009) Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933. *BMC Genomics*, **10**, 349.

58. Jaehning,J. a (1991) Sigma factor relatives in eukaryotes. *Science*, **253**, 859.

59. Tansey,W.P. and Herr,W. (1997) TAFs: Guilt by association? *Cell*, **88**, 729–732.

60. Goodrich,J. a. and Tjian,R. (1994) TBP-TAF complexes: Selectivity factors for eukaryotic transcription. *Curr. Opin. Cell Biol.*, **6**, 403–409.

61. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 6640–6645.

62. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.

63. Fritz,G., Megerle,J. a., Westermayer,S. a., Brick,D., Heermann,R., Jung,K., Rädler,J.O. and Gerland,U. (2014) Single cell kinetics of phenotypic switching in the arabinose utilization system of *E. coli*. *PLoS One*, **9**.

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

65. Sezonov,G., Joseleau-Petit,D. and D'Ari,R. (2007) *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.*, **189**, 8746–8749.

66. Cramer,P. (2002) Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.*, **12**, 89–97.

67. Häkkinen,A., Muthukrishnan,A.-B., Mora,A., Fonseca,J.M. and Ribeiro,A.S. (2013) CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli*. *Bioinformatics*, **29**, 1708–9.

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

69. 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**, 2579–2587.

70. Kandhavelu,M., Mannerström,H., Gupta,A., Häkkinen,A., Lloyd-Price,J., Yli-Harja,O. and Ribeiro,A.S. (2011) *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.*, **5**, 149.

71. Häkkinen,A. and Ribeiro,A.S. (2014) Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data. *Bioinformatics*, **31**, 69–75.

72. Mannerstrom,H., Yli-Harja,O. and Ribeiro,A.S. (2011) Inference of kinetic parameters of delayed stochastic models of gene expression using a Markov chain approximation. *Eurasip J. Bioinforma. Syst. Biol.*, **2011**, 11–15.

73. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.

74. Vogel,U. and Jensen,K.F. (1994) The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J. Bacteriol.*, **176**, 2807–2813.

Supplement for "Degree of regulation by σ factor competition in *Escherichia coli* depends on the kinetics of transcription initiation"

1. Promoter sequences

Sequences of P_{tetA} (1, 2) and P_{BAD} (3) with the consensus boxes (in red) at the -10 and -35 elements:

P_{BAD}:

```
ccataagattagcggatcctacctgacgctttttatcgcaactctctactgttctccatA
                       -35                    -10         +1
```

P_{tetA}:

```
ccagatgattaattcctaatttttgttgacactctatcattgatagagttattttaccacT
                          -35                   -10      +1
```

Both promoters have very conserved consensus at position -10, suggesting that they are transcribed by RNAP holoenzymes carrying factors of the σ^{70} family (4–6). Also, their binding affinity to holoenzymes carrying RpoD is expected to be much higher than to those carrying RpoS (4, 7), due to the conserved consensus at position -35 for RpoD recognition (8, 9). Furthermore, both promoters lack the consensuses for recognition by the σ^{54} family at positions -12 and -24 (10). We thus assume that their transcription initiation is triggered only by RNAP holoenzymes carrying RpoD.

2. Bacterial growth rates

We used the method proposed in (11) to induce exponential and stationary growth phases. To validate the method, we measured the cultures' optical density (OD600) every 30 minutes using the spectrophotometer for each growth media (Figure S1). The strains tested are BW25113 (wild type, or WT) and a deletion mutant, JW5437-1, both of which containing the MS2-GFP tagging system (with P_{BAD} as the target promoter and P_{lac} as the reporter promoter). Both strains were obtained from the Keio single-gene knockout collection (12).

All measurements were performed with and without 0.1% arabinose, added to the media after 2 hours of measurements at 37°C, to ensure that cells were adapted to the conditions. This comparison allows testing whether the cells are under stress, by comparing their growth rate in the presence and absence of arabinose.

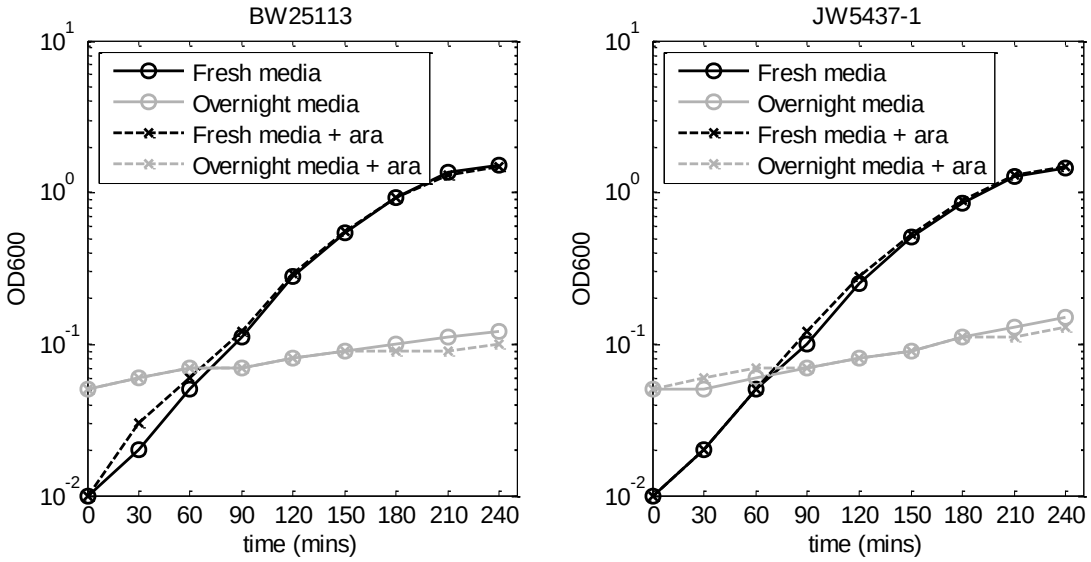


Figure S1. OD curves of bacterial populations in exponential growth-inducing media (black lines) and stationary growth-inducing media (grey lines). The media either contained 0.1% arabinose (solid lines and circles) or not (dashed lines and crosses). The moment t = 0 min. corresponds to the moment when arabinose was added to the media (2 h after the start of the experiment). The *E. coli* strains used are **(A)** BW25113 and **(B)** JW5437-1.

Figure S1 shows that the cell growth rates differ widely between the two different phase-inducing media. Also, it shows that, in both cases, the growth rates are not altered significantly by the addition of 0.1% arabinose in the media, for (at least) 2 hours following its addition.

We also assessed whether the confocal imaging of the cells (for 2 hours with images taken every 30 seconds), caused significant photo-toxicity. For this, we compared the division times of BW25113 cells in the exponential growth phase under the microscope, when and when not exposed to confocal microscopy (cells imaged by phase contrast in both cases).

We observed doubling times of 72 min. and 68 min. with and without confocal microscopy, respectively, from which we conclude that the confocal imaging does not introduce significant photo-toxicity. Media richness in these tests was 1x (see main manuscript).

In addition, the tests were also performed in two other media, differing in richness, as described in the main manuscript and referred to as 0.25x and 0.5x media. From Figure S2, we observe that the mean growth rate decreases slightly with decreasing media richness, from 1x to 0.25x.

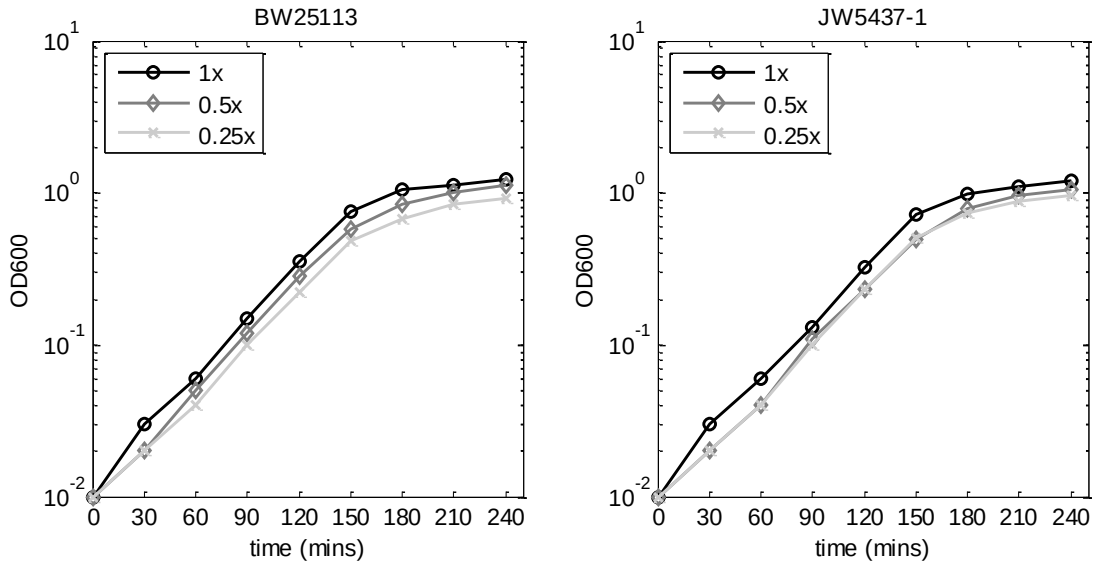


Figure S2. OD curves of bacterial populations in exponential growth-inducing media with 0.25x, 0.5x, and 1x richness. The *E. coli* strains used are **(A)** BW25113 and **(B)** JW5437-1.

3. qPCR measurements of relative RNA levels under the control of P_{tetA} and P_{BAD}

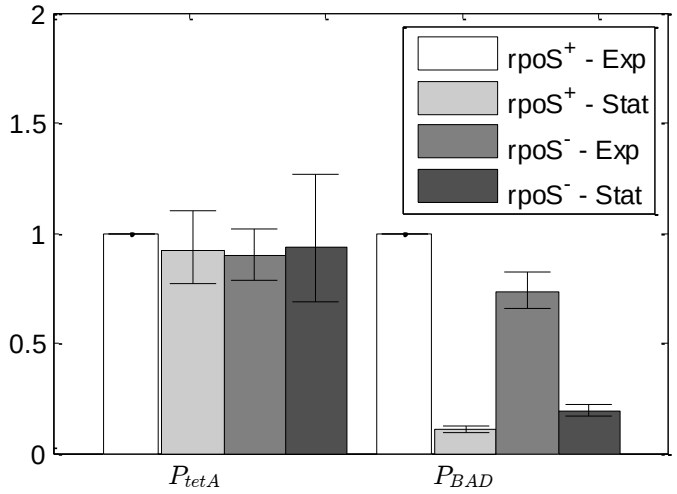


Figure S3. Relative levels of RNA under the control of P_{tetA} and P_{BAD} in BW25113 (*rpoS*$^+$) and JW5437-1 (*rpoS*$^-$) cells when in the exponential and stationary growth phases, as measured by qPCR using the 16S RNA housekeeping gene for internal reference.

4. Empirical distributions of intervals between consecutive RNA productions in individual cells and the models that best fit the data

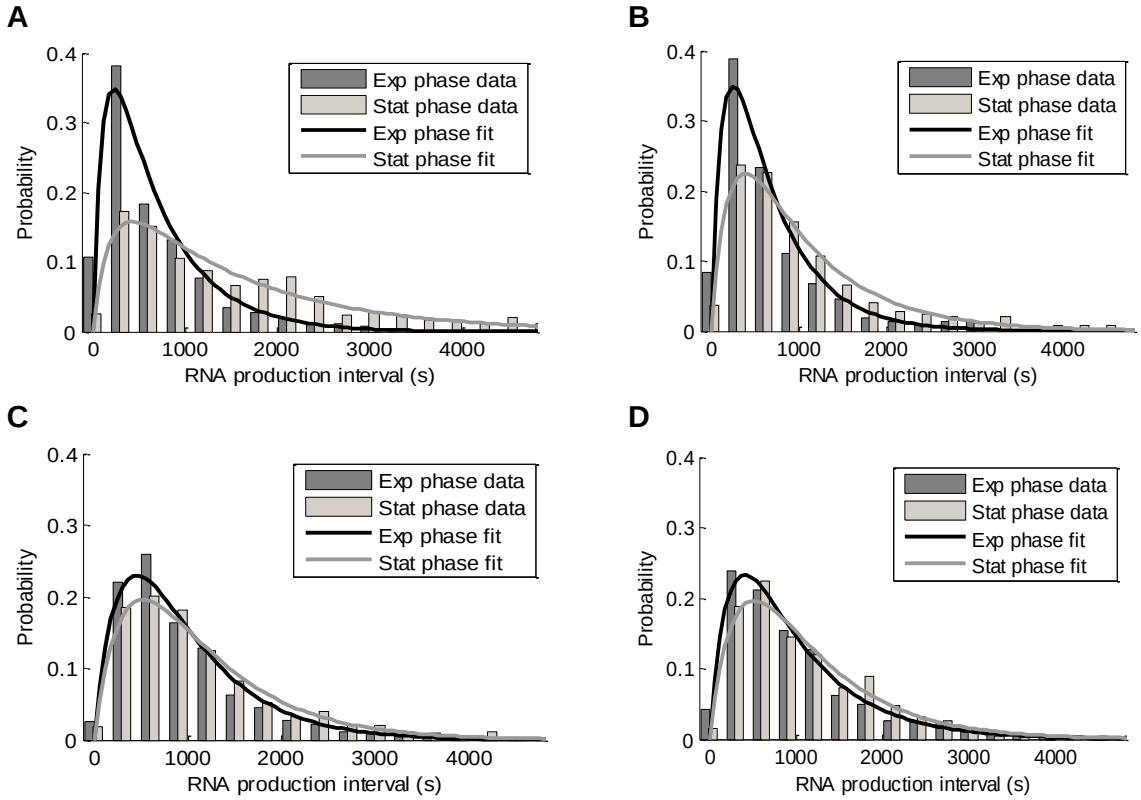


Figure S4. Distributions of intervals between consecutive RNA productions in individual cells from (A) P_{BAD} in *rpoS*$^+$ cells, (B) P_{BAD} in *rpoS*$^-$ cells, (C) P_{tetA} in *rpoS*$^+$ cells, and (D) P_{tetA} in *rpoS*$^-$ cells. Results from both cells in the exponential growth phase (dark grey bars) and the stationary growth phase (grey bars) are shown. Also shown are the fitted models of transcription with two exponentially-distributed rate-limiting steps, the closed and open complex formation (dark grey and grey solid curves).

5. Model of transcription initiation with σ factor dependency

Our model first accounts for the competition between different σ factors for binding to an RNAP core enzyme which exists in limited numbers in *E. coli* (13–16). Since the promoters studied are transcribed by σ^{70}, this reaction is modeled explicitly (reaction 1). Since the mutant strain studied lacks σ^{38}, this reaction is also modelled explicitly (reaction 2). Other σ factors are all referred to as σ^i and their interactions with RNAPs are modeled by a single reaction, for simplicity (reaction 3):

$$\sigma^{70} + RNAP \overset{K_{70}}{\longleftrightarrow} RNAP.\sigma^{70} \quad (1)$$

$$\sigma^{38} + RNAP \overset{K_{38}}{\longleftrightarrow} RNAP.\sigma^{38} \quad (2)$$

$$\sigma^{i} + RNAP \overset{K_{i}}{\longleftrightarrow} RNAP.\sigma^{i} \quad (3)$$

Reactions 1, 2 and 3 describe the binding/unbinding of RpoD (σ^{70}), RpoS (σ^{38}) and other σ factors (σ^i) to RNAP core enzymes (*RNAP*) respectively, in order to form corresponding RNAP

holoenzymes (*RNAP*.σ^{70}, *RNAP*.σ^{38}, and *RNAP*.σ^{i}, respectively). These reactions have a ratio between association and dissociation rate constants that are given by K_{70}, K_{38} and K_i, respectively.

We assume that in wild-type *E. coli*, the numbers of functional σ factors (either free floating or in a holoenzyme form) are [σ^{70}], [σ^{38}], [σ^{i}] for RpoD, RpoS and the other factors, respectively. The number of free floating RNAP core enzymes (i.e not bound to DNA) equals [*RNAP*].

Due to a limited pool of core enzymes when compared to the σ factor numbers, most RNAPs are expected to be in the holoenzyme form (14, 15, 17). Therefore, the number of *RNAP*.σ^{70} in wild-type *E. coli* (here, the r*poS*$^{+}$ strain) is given by:

$$[RNAP.\sigma^{70}]_{rpoS+} \sim [RNAP]\ \frac{[\sigma^{70}] \times K_{70}}{[\sigma^{70}] \times K_{70} + [\sigma^{38}] \times K_{38} + [\sigma^{i}] \times K_i} \quad (4)$$

Meanwhile, in the mutant strain JW5437-1 (rpoS^{-}), the number of RNAP.σ^{70} is given by:

$$[RNAP.\sigma^{70}]_{rpoS-} \sim [RNAP]\ \frac{[\sigma^{70}] \times K_{70}}{[\sigma^{70}] \times K_{70} + [\sigma^{i}] \times K_i} \quad (5)$$

Provided that [σ^{38}] is significantly greater than 0 in the wild type strain (which occurs when, e.g., these cells are in the stationary growth phase), then: $[RNAP.\sigma^{70}]_{rpoS+} < [RNAP.\sigma^{70}]_{rpoS-}$. Consequently, the transcription initiation rate of promoters transcribed by *RNAP*.σ^{70} should be slower in the stationary growth phase.

The model also accounts explicitly for the multi-step transcription process, based on the modelling strategy proposed in (18), which follows from the empirical models proposed in (19–21):

$$RNAP.\sigma^{70} + Pr \xrightarrow{k_{cc}} Pr_{cc} \quad (6)$$

$$Pr_{cc} \xrightarrow{k_{oc}} Pr_{oc} + \sigma^{70} \quad (7)$$

$$Pr_{oc} \xrightarrow{\infty} Pr + M + RNAP \quad (8)$$

Reaction 6 models the closed complex formation, which consists of the binding of RNAP holoenzymes with σ^{70} to the promoter region (*Pr*). Note that only σ^{70} is assumed to be able to bind stably to *Pr*, i.e., it is assumed that other holoenzymes either only bind very rarely to the promoter or, when doing so, do not remain bound for a significant amount of time. As this process is modeled by a first order reaction, the expected time to occur follows an exponential distribution.

Reaction 7 models the formation of the open complex, which, once complete, is followed by the release of σ^{70} from the promoter-RNAP complex (22) and by the escape of this complex from the promoter region (23), which allows initiating the elongation process and, subsequently, the termination and RNA release (24). All steps following the open complex formation are modeled by a single reaction (reaction 8). Their duration is not accounted for since they are expected to be much shorter in duration than the closed and open complex formation in normal conditions (23, 25).

From reactions 6-8, and assuming one and only one *Pr* promoter in each cell, the mean time to produce one RNA molecule is given by:

$$\tau_{cc} + \tau_{oc} = \frac{1}{k_{cc}[RNAP.\sigma^{70}]} + \frac{1}{k_{oc}} \quad (9)$$

From equations 5 and 9 it is deducible that, the absence of a given σ factor (other than σ^{70}) will decrease ($\tau_{cc} + \tau_{oc}$) of a promoter preferentially transcribed by $RNAp.\sigma^{70}$. The magnitude of this decrease depends on the magnitude of the various rate constants (some of which are unknown, such as k_{cc}), and thus it cannot be determined analytically. This was therefore assessed empirically for P_{BAD} and P_{tetA} (main manuscript).

6. Changes in transcription rate due to changes in σ factor numbers as a function of the duration of the closed and the open complex formation.

The model described by reactions 1-3 and to 6-8 can be used to explore, within realistic intervals of parameter values, the degree to which the duration of closed and open complex formation determine the degree of changes in transcription rate as a function of σ factors numbers.

For that, we use equation 9 to estimate the RNA production rate for different values of (τ_{cc}/τ_{oc}) as a function of the fraction of RNA polymerases occupied by σ^{70}, normalized by the expression rate when all holoenzymes are $RNAp.\sigma^{70}$. Results are shown in Figure S5.

Also shown are the expected relative transcription rates of P_{tetA} and P_{BAD} under full induction, based on the empirical values of (τ_{cc}/τ_{oc}). As expected, P_{tetA} is within the region of the parameter space of promoters whose closed complex is longer in duration than the open complex formation, while P_{BAD} is found in the opposite region of the parameter space.

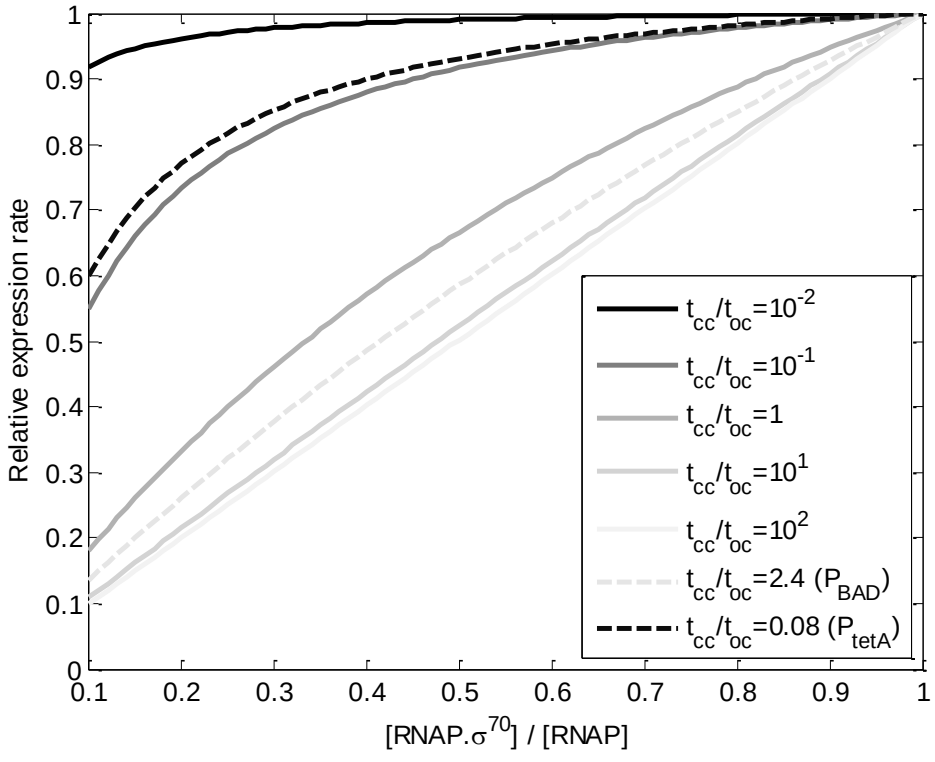


Figure S5. Relative RNA expression rate as a function of the fraction of RNA polymerases occupied by σ^{70}, for model promoters with differing empirical values of τ_{cc}/τ_{oc} and for two model promoters with values of τ_{cc}/τ_{oc} identical to those measured for P_{BAD} and P_{tetA}.

From Figure S5, the larger is τ_{cc}/τ_{oc}, the larger will the change in the relative expression rate of the target gene as a function of the change in σ factor numbers.

On the other hand, this figure suggests that too large changes in [RNAP.σ70] / [RNAp] are needed for the observed changes in transcription rates. Namely, for P_{BAD} we measured a change in transcription rate of ~50%, which would demand that the [RNAP.σ70] / [RNAp] would have changed by ~50% as well (which is not in accordance with the results in Figure 3). As such, and in agreement with differences in transcription dynamics of P_{BAD} between exponential and stationary phases in the mutant strain lacking RpoS, we expect that additional factors are contributing to the changing dynamics of transcription of P_{BAD} between cells these two growth phases.

7. Temporal fluorescence intensity of MS2-GFP tagged RNA molecules

The technique of detecting new RNA molecules in individual cells as these appear in time lapse microscopy images using the MS2-GFP RNA-tagging system consists of fitting the total corrected RNA spot intensity with a step-increasing function (26).

For this method to be valid, new RNA molecules need to appear nearly fully-tagged when first detected, so as to cause a significant "jump" in the total spots fluorescence intensity of the cell. This is possible if the speed of elongation (expected to be ~60 and ~90 base pairs per second, at 37^{o}C (27–29)) and the MS2-GFP binding to the target RNA do not take much longer than the interval between consecutive images, which in our measurement settings is 30 seconds long. Second, it is necessary that an MS2-GFP tagged RNA, once tagged, does not degrade significantly (neither abruptly nor gradually) during the measurement period (to allow using a step-increasing function). These two assumptions were recently tested (30), by observing the fluorescence intensity of individual, tagged RNA molecules for 30 minutes (1 image per minute) in individual cells.

From the data, first, it was found that new RNA molecules are nearly fully-tagged when first detected, as no significant RNA fluorescence increases are observed after the detection of the complex. Second, by fitting the intensity of each tagged RNA over time with a decaying exponential function and inferring its intensity degradation rate, a mean decaying rate of ~8.1×10^{-5} s^{-1} was measured, corresponding to a mean half-life of ~144 mins, which is longer than our observation window (120 mins).

These measurements agree with previous studies of the coat protein of bacteriophage MS2 (31, 32), which showed that most MS2 binding sites are constantly occupied by MS2-GFP proteins, resulting in the 'immortalization' of the target RNA due to isolation from RNA-degrading enzymes (33, 34). We thus assume that the fluorescence of tagged RNAs does not decrease significantly over time (gradually or abruptly) during the measurement period, in agreement with previous reports using the

same RNA detection system (26, 34–36). Finally, it is noted that this method of single RNA counting can tolerate infrequent "blinking" of the tagged RNAs, due to moving out of focus transiently, without loss of information (26).

8. Measurements of RpoC protein levels in the exponential growth phase when in 0.25x, 0.5x, and 1.0x media and in the stationary phase

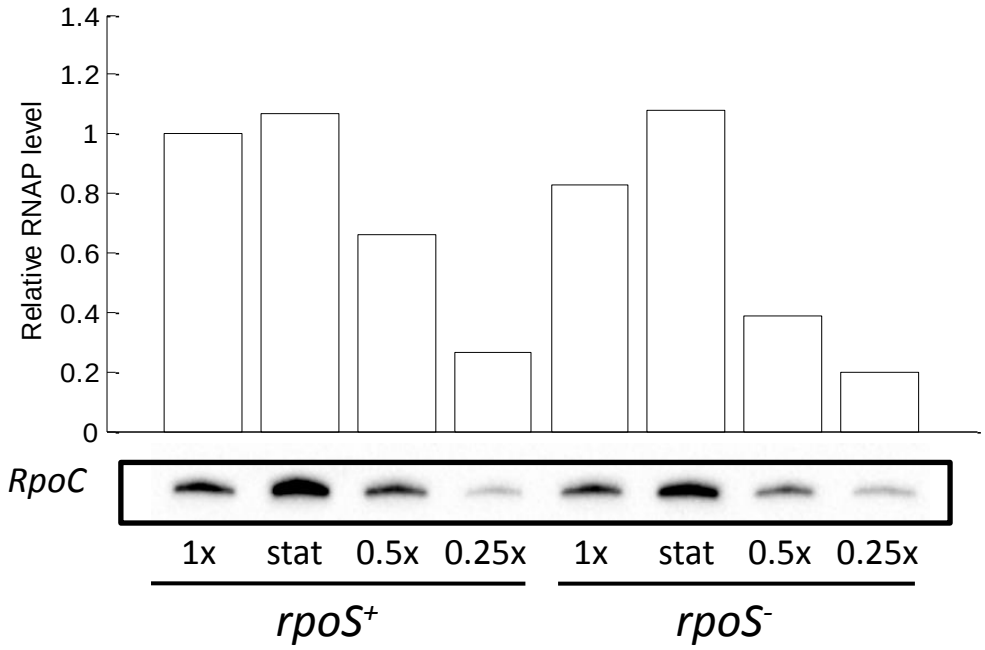


Figure S6. *RpoC levels* in BW25113 (*rpoS*$^{+}$) and JW5437-1 (*rpoS*$^{-}$) cells in the exponential growth phase when in 0.25x, 0.5x, and 1.0x media and in the stationary phase (stat), as measured by protein immunoblot. The protein levels are shown in relative to that in 1x media for *rpoS*$^{+}$ cells.

REFERENCES

1. Korpela,M.T., Kurittu,J.S., Karvinen,J.T. and Karp,M.T. (1998) A recombinant *Escherichia coli* sensor strain for the detection of tetracyclines. *Anal. Chem.*, **70**, 4457–4462.

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

3. Smith,B.R. and Schleif,R. (1978) Nucleotide sequence of the L-arabinose regulatory region of *Escherichia coli* K12. *J. Biol. Chem.*, **253**, 6931–6933.

4. Colland,F., Fujita,N., Kotlarz,D., Bown,J.A., Meares,C.F., Ishihama,A. and Kolb,A. (1999) Positioning of σS, the stationary phase σ factor, in *Escherichia coli* RNA polymerase-promoter open complexes. *EMBO J.*, **18**, 4049–4059.

5. Wise,A., Brems,R., Ramakrishnan,V. and Villarejo,M. (1996) Sequences in the -35 region of

Escherichia coli rpoS-dependent genes promote transcription by E.RpoS. *J. Bacteriol.*, **178**, 2785–2793.

6. Gruber,T.M. and Gross,C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.*, **57**, 441–66.

7. Becker,G. and Hengge-Aronis,R. (2001) What makes an *Escherichia coli* promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S). *Mol. Microbiol.*, **39**, 1153–65.

8. Lisser,S. and Margalit,H. (1993) Compilation of E.coli mRNA promoter sequences. *Nucleic Acids Res.*, **21**, 1507–1516.

9. Hawley,D.K. and Mcclure,W.R. (1983) Nucleic compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.*, **11**, 2237–2255.

10. Barrios,H., Valderrama,B. and Morett,E. (1999) Compilation and analysis of σ54-dependent promoter sequences. *Nucleic Acids Res.*, **27**, 4305–4313.

11. Powell,E.O. (1956) Growth rate and generation time of bacteria, with special reference to continuous culture. *J. Gen. Microbiol.*, **15**, 492–511.

12. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.

13. Farewell,A., Kvint,K. and Nyström,T. (1998) Negative regulation by RpoS: A case of sigma factor competition. *Mol. Microbiol.*, **29**, 1039–1051.

14. Grigorova,I.L., Phleger,N.J., Mutalik,V.K. and Gross,C. a (2006) Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 5332–5337.

15. Maeda,H., Fujita,N. and Ishihama,A. (2000) Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.*, **28**, 3497–3503.

16. Mauri,M. and Klumpp,S. (2014) A model for sigma factor competition in bacterial cells. *PLoS Comput. Biol.*, **10**, e1003845.

17. Shepherd,N. and Bremer,H. (2001) Cytoplasmic RNA polymerase in *Escherichia coli*. *Society*, **183**, 2527–2534.

18. 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.

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

20. Buc,H. and McClure,W.R. (1985) Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry*, **24**, 2712–2723.

21. McClure,W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.*, **54**, 171–204.

22. Raffaelle,M., Kanin,E.I., Vogt,J., Burgess,R.R. and Ansari,A.Z. (2005) Holoenzyme switching and stochastic release of sigma factors from RNA polymerase *in vivo*. *Mol. Cell*, **20**, 357–366.

23. Hsu,L.M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta - Gene Struct. Expr.*, **1577**, 191–207.

24. deHaseth,P.L., Lohman,T.M., Burgess,R.R. and Record,M.T. (1978) Nonspecific interactions of *Escherichia coli* RNA polymerase with native and denatured DNA: differences in the binding behavior of core and holoenzyme. *Biochemistry*, **17**, 1612–1622.

25. Herbert,K.M., La Porta,A., Wong,B.J., Mooney,R.A., Neuman,K.C., Landick,R. and Block,S.M. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell*, **125**, 1083–94.

26. Kandhavelu,M., Häkkinen,A., Yli-Harja,O. and Ribeiro,A.S. (2012) Single-molecule dynamics of transcription of the lar promoter. *Phys. Biol.*, **9**, 026004.

27. Vogel,U. and Jensen,K.F. (1994) The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J. Bacteriol.*, **176**, 2807–2813.

28. Dennis,P.P., Ehrenberg,M., Fange,D. and Bremer,H. (2009) Varying rate of RNA chain elongation during rrn transcription in *Escherichia coli*. *J. Bacteriol.*, **191**, 3740–3746.

29. Ryals,J., Little,R. and Bremer,H. (1982) Control of rRNA and tRNA syntheses in *Escherichia coli* by guanosine tetraphosphate. *J. Bacteriol.*, **151**, 1261–1268.

30. 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**, 2579–2587.

31. Talbot,S.J., Goodman,S., Bates,S.R.E., Fishwick,C.W.G. and Stockley,P.G. (1990) Use of synthetic oligoribonucleotides to probe RNA-protein interactions in the MS2 translational operator complex. *Nucleic Acids Res.*, **18**, 3521–3528.

32. Fusco,D., Accornero,N., Lavoie,B., Shenoy,S.M., Blanchard,J.M., Singer,R.H. and Bertrand,E. (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.*, **13**, 161–167.

33. Golding,I., Paulsson,J., Zawilski,S.M. and Cox,E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, **123**, 1025–1036.

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

35. Barrios,H., Valderrama,B. and Morett,E. (1999) Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res.*, **27**, 4305–4313.

36. Muthukrishnan,A.-B., Martikainen,A., Neeli-Venkata,R. and Ribeiro,A.S. (2014) *In vivo* transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells. *PLoS One*, **9**, e109005.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3678-6
ISSN 1459-2045