Jerome Geoffrey Chandraseelan

Synthetic Genetic Circuits: Plasticity and Robustness

Tampere 2015
Jerome Geoffrey Chandraseelan

**Synthetic Genetic Circuits: Plasticity and Robustness**

Thesis for the degree of Doctor of Philosophy to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 17th of September 2015, at 12 noon.
Abstract

Living organisms have evolved to survive in a multitude of environmental conditions. This plasticity, and the robustness to environmental fluctuations, is achieved by altering gene expression levels in response to perturbations while maintaining the basic processes essential for survival. Genetic circuits, which are networks of interacting genes, are responsible for carrying out both the basic processes to sustain life, such as counting time, as well as the specific processes that provide them with adaptability to various environmental conditions while maintaining homeostasis.

Gene expression and its regulation are dynamic, stochastic processes. Namely, although gene expression was previously considered to be identical in cells arising from common ancestors, the observation of multiple, single cells expressing fluorescent proteins has shown that gene expression is noisy, which allows genetically identical cells in a homogeneous environment to behave differently, a phenomenon known as cell-to-cell phenotypic variability. Noise in gene expression arises from the fact that most of the underlying biochemical reactions involve small molecular numbers, which leads to infrequent, to some extent random in time, interactions and processes. While initially noise in gene expression was considered to be disadvantageous to the organisms, a number of recent studies suggest significant functional roles for noise in intracellular processes.

The complexity and size of natural genetic circuits hampers their detailed study at present. One approach to overcome this problem is based on design of small, and thus more tractable, artificial circuits. Aside from its small size, in such circuits, all components are known and there is less chance that they interact with unknown cellular components. This strategy offers additional advantages, such as testability of whether a certain architecture is able to generate a desired trait or function without affecting natural cellular processes e.g., by inducing other downstream effectors that affect cell functioning. Moreover, it provides an opportunity to compare different circuit designs and determine which circuit architecture is preferable. Finally, due to detailed knowledge of

Abstract

Living organisms have evolved to survive in a multitude of environmental conditions. This plasticity, and the robustness to environmental fluctuations, is achieved by altering gene expression levels in response to perturbations while maintaining the basic processes essential for survival. Genetic circuits, which are networks of interacting genes, are responsible for carrying out both the basic processes to sustain life, such as counting time, as well as the specific processes that provide them with adaptability to various environmental conditions while maintaining homeostasis.

Gene expression and its regulation are dynamic, stochastic processes. Namely, although gene expression was previously considered to be identical in cells arising from common ancestors, the observation of multiple, single cells expressing fluorescent proteins has shown that gene expression is noisy, which allows genetically identical cells in a homogeneous environment to behave differently, a phenomenon known as cell-to-cell phenotypic variability. Noise in gene expression arises from the fact that most of the underlying biochemical reactions involve small molecular numbers, which leads to infrequent, to some extent random in time, interactions and processes. While initially noise in gene expression was considered to be disadvantageous to the organisms, a number of recent studies suggest significant functional roles for noise in intracellular processes.

The complexity and size of natural genetic circuits hampers their detailed study at present. One approach to overcome this problem is based on design of small, and thus more tractable, artificial circuits. Aside from its small size, in such circuits, all components are known and there is less chance that they interact with unknown cellular components. This strategy offers additional advantages, such as testability of whether a certain architecture is able to generate a desired trait or function without affecting natural cellular processes e.g., by inducing other downstream effectors that affect cell functioning. Moreover, it provides an opportunity to compare different circuit designs and determine which circuit architecture is preferable. Finally, due to detailed knowledge of
their structure, the behavior of these circuits can be computationally simulated to assist, e.g., the study of their long-term behaviors, among other. Previous studies based on synthetic circuits have already provided key insights into the design principles and architecture of genetic circuits, such as how these organize genes so as to gain the ability to make decisions or track time. These circuits are also expected to become of great use in therapeutic and industrial applications.

In this thesis, we focused on the study of the phenotypic plasticity and robustness of synthetic genetic clocks. We focused on the effects of temperature, copy number and the role of components; proteins and promoters, of the circuit. For this, we made use of the well-known genetic Repressilator, a synthetic genetic clock that is also one of the simplest genetic circuits known to be functional in a living system. To assist the studies, aside from techniques from cell and molecular biology, we made use of state of the art techniques in microscopy and image analysis. From the data, we characterized, first, the phenotypic plasticity of individual genes in Escherichia coli, as these are the main components of the Repressilator. Next, we performed a study of the effects of temperature on the dynamics of the genetic Repressilator. Subsequently, we studied the degree of synchrony between sister cells, each containing the Repressilator, in order to evaluate the extent to which cell division affects the dynamics of this circuit. Finally, we inserted the Repressilator genetic code into a single-copy vector so as to, by comparison with the original construct, study the effects of copy numbers on the dynamics of the circuit.
Preface

First and Foremost, I express my sincere thanks to my supervisor, Andre Ribeiro, for providing me this opportunity to carry out research in the Laboratory of Biosystem Dynamics. I thank him for his constant guidance, support and motivation while carrying out the research work.

I thank Dr. Andrey Kan and Asst. Prof. Ilda Santos Sanches for reviewing the thesis and for their valuable comments and suggestions.

I thank Assistant Professor Meenakshisundaram Kandhavelu, for his support and discussions during the initial stage of this work. My sincere thanks to all the co-authors and the members of the Laboratory of Biosystem Dynamics for their support in completing this thesis.

I thank all the members of the Faculty for providing immense support. My special thanks to the Virve Larmila, Juha Peltonen and the coordinators Elina Orava and Ulla Siltaloppi, for their kind support during my studies.

Special thanks to my beloved wife Shema Angelina and my daughter Stacy Gabriella for their immense love and support.

I sincerely thank my parents, my grandmother and my brother for their everlasting love and support.

Tampere, July 2015

Jerome Geoffrey Chandraseelan

Preface

First and Foremost, I express my sincere thanks to my supervisor, Andre Ribeiro, for providing me this opportunity to carry out research in the Laboratory of Biosystem Dynamics. I thank him for his constant guidance, support and motivation while carrying out the research work.

I thank Dr. Andrey Kan and Asst. Prof. Ilda Santos Sanches for reviewing the thesis and for their valuable comments and suggestions.

I thank Assistant Professor Meenakshisundaram Kandhavelu, for his support and discussions during the initial stage of this work. My sincere thanks to all the co-authors and the members of the Laboratory of Biosystem Dynamics for their support in completing this thesis.

I thank all the members of the Faculty for providing immense support. My special thanks to the Virve Larmila, Juha Peltonen and the coordinators Elina Orava and Ulla Siltaloppi, for their kind support during my studies.

Special thanks to my beloved wife Shema Angelina and my daughter Stacy Gabriella for their immense love and support.

I sincerely thank my parents, my grandmother and my brother for their everlasting love and support.

Tampere, July 2015

Jerome Geoffrey Chandraseelan
## Contents

**ABSTRACT** ....................................................................................................................... 1

**PREFACE** .......................................................................................................................... iii

**CONTENTS** ......................................................................................................................... iv

**LIST OF PUBLICATIONS** .................................................................................................... vii

1 INTRODUCTION ........................................................................................................ 1
  1.1 Background and motivation .................................................................................. 1
  1.2 Thesis objectives .................................................................................................. 3
  1.3 Thesis outline ....................................................................................................... 5

2 GENE EXPRESSION AND REGULATION .................................................................. 7
  2.1 Operon .................................................................................................................. 8
  2.2 Transcription ....................................................................................................... 9
  2.3 Translation ......................................................................................................... 11
  2.4 Regulation of transcription ............................................................................... 14
     2.4.1 Activation .................................................................................................. 14
     2.4.2 Repression .............................................................................................. 16

3 SYNTHETIC BIOLOGY ............................................................................................. 17
  3.1 Design principles of synthetic genetic circuits ................................................. 18
     3.1.1 Logic gates .............................................................................................. 19
  3.2 Components of synthetic genetic circuits ...................................................... 21
     3.2.1 Transcription unit .................................................................................. 21
List of publications

This thesis is based on the following publications. The publications are reproduced with permission from the publishers.


In this thesis we studied the phenotypic plasticity and robustness of synthetic genetic clocks. We focused on the effects of temperature, copy number, role of component proteins, and promoters of the circuit. For that we made use of the genetic Repressilator, which is one of the simplest genetic circuits known to be functional in a living system. In the following section, we describe the contribution of the scientific papers to this effort and the contribution of the author of this thesis in each of these scientific papers.
In Publication 1 (published in Nucleic Acids Research), we studied transcription dynamics from a promoter. Specifically, we studied the influence of the inducer intake kinetics and its transient effect on the kinetics of transcription. The author of the thesis contributed to this study as follows. First, the author, along with M. Kandhavelu, conducted preliminary microscopy measurements to study transcription dynamics from the promoter. Next, the author designed and conducted the qPCR measurements that validated the results obtained from the microscopy measurements. Finally, the author participated in writing of the manuscript, particularly in sections reporting the results in which the author was involved.

In Publication 2 (published in Molecular Biosystems), we studied the effects of temperature on the dynamics of the Repressilator. For that, we subjected cells containing the Repressilator to various temperatures and measured the oscillatory dynamics, in terms of fluorescence levels, exhibited by the cells over time. Specifically, we focused on the effects of temperature on the mean period and robustness of the oscillations. From the results, it is evident that temperature has a profound effect on the dynamics of the Repressilator. Further studies on the dynamics of the CI-Cro Switch, using the MS2-GFP RNA detection method, revealed that the loss of functionality of CI, which is a component protein of both the Repressilator and the Switch, with increasing temperature leads to the observed failure of oscillations in the Repressilator at higher than optimal temperatures. The author of this thesis contributed to Publication II by, first, developing a protocol that allows the cells containing the Repressilator to produce oscillations under the microscope, and then by conducting microscopy measurements at all the temperature conditions considered in the study. The author also contributed in the preparation of the manuscript by suggesting ideas such as that CI loses functionality with increasing temperature based on previous studies.

Publication 3 (published in International Science Index, World Academy of Science, Engineering and Technology) is a follow up of the Publication 2. Here, we studied the effect of temperature on the synchrony of oscillations is sister cells following division of their mother cell. For that, we selected functional sister cells and estimated the loss of synchrony by calculating correlation coefficients between sister cells. The outcomes of
this study suggest that increases in temperature increase the loss of correlation between sister cells over time. The author of this thesis contributed to the study as follows: firstly, by conducting confocal microscopy measurements and then by performing the analysis of the images of sister cells that allowed the subsequent calculation of the degree of loss of synchrony. Finally, the author contributed in the writing of the manuscript.

In publication 4 (published in Molecular Biosystems), we studied the effects of changing the copy-number of the vector on the dynamics of the Repressilator. For that, we inserted the Repressilator in to a single-copy vector and studied the oscillatory dynamics. The study reveals that changing the copy-number of the Repressilator has strong effects on the dynamics of the Repressilator in individual cells. Namely, the results suggest that implementing the Repressilator on a single-copy vector provides this circuit with greater functionality than the original low copy Repressilator, although it causes increased mean and variability of the period of oscillations. From the analysis of the effects of temperature on the dynamics of the Repressilator, we observed also that the functionality of the single-copy Repressilator is higher at optimum and lower-than-optimal temperature conditions. Also, the single-copy Repressilator was found to be more robust than the low-copy Repressilator to external perturbations. The author of this thesis contributed to this study as follows. Firstly, the author designed and constructed the single-copy Repressilator. Next, the author performed most of the microscopy measurements and the qPCR analysis to validate the functionality of single-copy Repressilators in the cells. The author also contributed in the writing of the manuscript.

Publication 1 will be used by Jarno Mäkelä in his doctoral thesis.

Publication 2, 3 and 4 will be used by Samuel Matos Dias Oliveira in his doctoral thesis.
1 Introduction

1.1 Background and motivation

During the course of evolution, living systems have evolved adaptability to survive in a wide range of environmental conditions and compensate for the perturbations caused by changes in the environmental conditions (Cagliero et al., 2013; Stoebel et al., 2009) or by interactions with other beings. Genetic circuits, formed by a network of interacting genes, are responsible for carrying out the complex processes that provide this adaptability at a cellular level, such as counting time, responding to signals from the environment, and regulating cellular processes that perform the basic functions needed for maintaining homeostasis (Becskei et al., 2000).

Some of the genes in the gene regulatory networks (GRN) form ‘genetic clocks’ which control the timing of activity of other genes. For example, circadian oscillators, found in many living systems, control the expression of other genes in the GRN with high temporal stability. These circuits are robust to fluctuations and resilient to perturbations caused by intrinsic factors and extrinsic factors (Mihalcescu et al., 2004). Perturbations in the functioning of these genetic clocks has been linked to many diseases, including cancer, neural disorders, etc. (Fussenegger, 2010). Hence, understanding the regulation of these genetic circuits and their architecture is of importance for understanding several diseases and, thus, for therapeutic purposes.
Many of the elements involved in the regulation of gene regulatory networks have been identified using biochemical and molecular biological approaches. Though scientists were able to identify several of the molecules involved in gene expression regulation and, from there, produce gene interaction maps, some fundamental questions such as why a given architecture is preferred over another and how genetic circuits respond to fluctuations, still remain unanswered (Nandagopal et al., 2011).

Synthetic biology is an alternative approach to answer these questions. Its aim is to engineer synthetic gene networks and study how the architecture of the network influences the regulation of gene expression. In general, well-characterized genetic components are used. Also, the topologies used are derived from the design principles from electrical engineering (Guido et al., 2006; Miyamoto et al., 2013). One of the major advantages of using this approach is expected to be the enhanced predictability of the behaviour of the circuits, via the use of techniques from computational biology. A number of synthetic genetic circuits have been designed and proven to exhibit the predicted behaviour (Elowitz et al., 2000; Stricker et al., 2008). The basic components used for the construction of a synthetic genetic circuit are: the Transcription unit, which comprises of a promoter followed by a gene that encodes a protein and a downstream terminator; the Input signal, which may arise from the environment or within the cell; and the Regulators of transcription, namely activators and repressors which can be produced by the network or be external in origin (Savageau, 2001).

In this thesis, we aimed to study the phenotypic robustness and plasticity of single genes and small synthetic genetic clocks. For that, we used the Repressilator, which is one of the first built, and simplest, synthetic genetic circuits known to be functional in a living organism (Elowitz et al., 2000). It consists of three promoters connected together in a negative feedback loop such that transcriptional activity of one of the promoters in the loop represses the activity of the subsequent promoter (Elowitz et al., 2000), leading to periodic fluctuations in protein levels. These fluctuations can be visualized as oscillations in the level of Green Fluorescence Proteins (GFP) produced from a reporter plasmid, which encodes GFP under the control of one of the promoters of the Repressilator. The thesis, a compilation of works we recently published on this topic, begins with a study on...
the responsiveness and phenotypic plasticity of an individual promoter at the RNA level, followed by a study on the dynamics of the Repressilator at different temperatures. Next, we present a study on the effects of temperature on the degree of synchrony of the Repressilator between sister cells. Finally, we studied the effect of the intra-cellular copy-number of the Repressilator on its dynamics.

### 1.2 Thesis objectives

This thesis is aimed at studying the phenotypic plasticity and robustness of a genetic circuit, namely the Repressilator. The findings of this study are expected to aid in the development of new synthetic genetic circuits. The objectives are as follows:

In objective 1, we aimed to study transcription initiation dynamics when based on the introduction in the media of an external inducer and the transcription dynamics following activation. Transcription initiation is of particular interest as most of the gene expression regulation occurs at this stage (Latchman, 2007). Moreover, it has been reported that bacterial adaptation to stress involves modifications in transcription patterns (Stoebel et al., 2009). Transcription initiation is a multi-stepped sequential process that comprises several rate-limiting steps. The number of rate-limiting steps differ between conditions and between different promoters (Mchule, 1980). Relevantly, objective 1 also included the development of methods of studying transcription dynamics from time-lapse, single RNA detection.

Temperature affects gene expression regulation on a global scale, both by altering the thermodynamic rate of the underlying reactions (Gadgil et al., 2005; Muthukrishnan et al., 2012; Stricker et al., 2008) as well as by influencing the physical and chemical properties of the effector molecules (Hansen et al., 2014). The model organism used in this study, *E. coli*, is known for its ability to survive at different temperatures. Previous studies have shown that changes in temperature affect transcription dynamics in *E. coli* (Bertrand-Burggraf et al., 1984; Muthukrishnan et al., 2013). To study the effect of the temperature induced changes on the transcription dynamics on genetic circuits, in...
objective 2, we studied the effect of temperature on the dynamics of the Repressilator, which exhibits oscillatory dynamics at 30°C (Elowitz et al., 2000). Specifically, we characterized the role of one of the components, the lambda repressor cI, on the temperature-dependence of the dynamics of the Repressilator.

Synchrony of oscillations in different cells is one of the crucial traits observed in biological oscillators, which is important for performing complex, dynamic processes involving multiple cells (Kruse et al., 2005). Natural biological oscillators, such as circadian oscillators, exhibit synchrony with the neighboring cells as well as with the extracellular world (Dunlap., 1999). Since this is one of the desirable traits in the synthetic genetic circuit, in objective 3 we studied how temperature affects the degree of synchrony of the Repressilator between sister cells following their birth.

Copy-number of the constituent genes influences the dynamics of a gene regulatory network. For example, increasing the copy-number of circuit can lead to the relative reduction of the fluctuations in protein numbers, which, in the case of periodic oscillators, ought to affect the periodicity of the oscillations (Ribeiro et al., 2007; Ribeiro, 2007). The Repressilator was originally constructed on a low copy number plasmid, pZS1 (Lutz et al., 1997), which maintains about 3-4 copies of the plasmid per cell. All these copies dynamics are, necessarily, coupled since the proteins they code for are indistinguishable and, as such, any gene product of one of the plasmids can influence the expression of the subsequent gene in the loop in any of the other plasmids in the same cell. In objective 4, we studied the effect of copy-number on the dynamics of the Repressilator. Specifically, we inserted the Repressilator on another plasmid, BAC-2 (Bacterial Artificial Chromosome), which is known to exist as a single-copy in the cells (Gordon et al., 1997; Ogura et al., 1983), and studied how placing the Repressilator in single-copies in the cells affects its mean period and robustness. Also, we examined the effect of external perturbations in this circuit’s dynamics.

objective 2, we studied the effect of temperature on the dynamics of the Repressilator, which exhibits oscillatory dynamics at 30°C (Elowitz et al., 2000). Specifically, we characterized the role of one of the components, the lambda repressor cI, on the temperature-dependence of the dynamics of the Repressilator.

Synchrony of oscillations in different cells is one of the crucial traits observed in biological oscillators, which is important for performing complex, dynamic processes involving multiple cells (Kruse et al., 2005). Natural biological oscillators, such as circadian oscillators, exhibit synchrony with the neighboring cells as well as with the extracellular world (Dunlap., 1999). Since this is one of the desirable traits in the synthetic genetic circuit, in objective 3 we studied how temperature affects the degree of synchrony of the Repressilator between sister cells following their birth.

Copy-number of the constituent genes influences the dynamics of a gene regulatory network. For example, increasing the copy-number of circuit can lead to the relative reduction of the fluctuations in protein numbers, which, in the case of periodic oscillators, ought to affect the periodicity of the oscillations (Ribeiro et al., 2007; Ribeiro, 2007). The Repressilator was originally constructed on a low copy number plasmid, pZS1 (Lutz et al., 1997), which maintains about 3-4 copies of the plasmid per cell. All these copies dynamics are, necessarily, coupled since the proteins they code for are indistinguishable and, as such, any gene product of one of the plasmids can influence the expression of the subsequent gene in the loop in any of the other plasmids in the same cell. In objective 4, we studied the effect of copy-number on the dynamics of the Repressilator. Specifically, we inserted the Repressilator on another plasmid, BAC-2 (Bacterial Artificial Chromosome), which is known to exist as a single-copy in the cells (Gordon et al., 1997; Ogura et al., 1983), and studied how placing the Repressilator in single-copies in the cells affects its mean period and robustness. Also, we examined the effect of external perturbations in this circuit’s dynamics.
1.3 Thesis outline

The structure of the thesis is as follows: chapter 2 provides an overview of the process of gene expression and its regulation, followed by a description of methods used in synthetic biology in chapter 3. In particular, we focused on the design principles and components required for construction of synthetic genetic circuits. In chapter 4, in vivo single-cell methods used to study transcription dynamics are described along with the computational and statistical methods used for segmentation of cells, quantification of single RNA molecules and inference of the number of rate-limiting steps in transcription initiation. Chapter 5 begins with an introduction to the Repressilator circuit, followed by a description of the experimental; computational and statistical methods used in this study. Finally, in chapter 6 we describe and discuss the outcomes of the study.
Gene expression and regulation

Gene expression can be defined as the process of conversion of the genetic information contained in the DNA (Deoxyribonucleic acid) to functional form called ‘proteins’. It begins when information encoded in the DNA is transcribed to RNA (Ribonucleic acid), by the RNA polymerases, in a process called 'Transcription'. This is subsequently followed by protein synthesis by ribosomes, in a process called translation, in which information from the RNA sequence is translated into proteins (Fig. 1) (Crick, 1970).

In prokaryotes, the process of transcription and translation are coupled (Miller, 1970; Brown et al., 1997). Though proteins were considered to be functional components of cells, a number of recent studies suggest that significant functional roles are played by the RNA in the regulation of many processes in gene expression (Liu et al., 2012; Raghavan et al., 2011).

The structure and function of cells are determined by their gene expression pattern. Hence, it is pivotal for a cell to regulate the expression of its genes. Commonly, the decision to initiate transcription of a particular gene is the major point of controlling its expression. In this chapter, an overview about the processes in gene expression regulation is given.
2.1 Operon

In *E. coli*, several protein coding genes, which perform related functions, are organized into functional regions in the genome called operons (Fig. 2) (Jacob *et al*., 1961). The expression of the genes in an operon is under the control of a single promoter; hence activation and repression of the constituent genes of the operon are affected to the same extent, and at the same time, by the regulatory molecules. The genes encoded in the operon are transcribed together as a single poly-cistronic mRNA and either are translated together or, instead, the RNA undergoes trans-splicing to create multiple mono-cistronic mRNAs that undergo translation separately.

Typically, the promoter sequences contain both a “-10 box” and “-35 box” which, as the name indicates, are located, respectively, at about 10 and 35 base pairs (bp) upstream of the transcription start site (TSS) (Pribnow *et al*., 1975; Maniatis *et al*., 1975). The most probable sequences of these two boxes are, respectively, TATAAT and TTGACA (Siebenlist, 1979).

In general, promoters also contain operator sites, which are sites of binding of the regulatory molecules that regulate transcription, positively or negatively. In case of the Lac operon, the classic example of an operon, there are three operator sites: O1, O2 and O3. Repression is achieved when the repressor, LacI, binds to the operator O1 and prevents transcription by the RNA polymerase. Other operators are also involved in repression by binding to the repressor, though with lesser strength (Oehler *et al*., 1990).
and Mueller et al., 1996), leading to less efficient repression. It is also reported that maximal repression is achieved by the formation of a DNA loop between the operators (Reznikoff et al., 1974).

The structural genes of the operon encode for proteins that perform related functions. For example, the lac operon codes for three structural genes \( \text{lacZ}, \text{lacY}, \text{and lacA} \), which are all involved in the metabolism of lactose, a disaccharide composed of glucose and galactose. The \( \text{lacZ} \) encodes the production of an enzyme, namely beta-galactosidase, which breaks lactose into glucose and galactose. \( \text{lacY} \) encodes a transporter protein, beta-galactoside permease, which transports lactose from the environment to the cytoplasm of the cells. Finally, \( \text{lacA} \) codes for an enzyme which converts galactose, produced from the breakdown of lactose to glucose, which in turn is converted into energy (Jacob et al., 1961).

![Figure 2: Schematic representation of the structural organisation of the genes of the lac operon. (Image reproduced from LIFE: THE SCIENCE OF BIOLOGY, seventh edition.)](image)

2.2 Transcription

Transcription is the first step in the process of gene expression. Most of gene expression regulation occurs at this stage, ultimately controlling the rate of synthesis of the proteins (Latchman, 2007). A gene is said to be in the ‘repressed’ or ‘off’ state when the rate of
In transcription initiation, the RNA polymerase binds to the promoter and, once finding a transcription start site, it forms a closed complex (CC). This is followed by a series of intermediate isomerization steps which, eventually, lead to the formation of the open complex (OC). Upon successful open complex formation, the promoter becomes catalytically active allowing the RNA polymerase to synthesize new RNA. During elongation, the RNA polymerase (RNAP) traverses through the template strand until the termination site, synthesizing the RNA. For successful entry into the elongation phase, the RNA polymerase has to undergo a crucial step called ‘promoter clearance’: typically, a minimum of 10 base pairs synthesis should occur for the RNA polymerase to continue synthesizing the RNA. Failing to achieve this usually leads to abortive transcription (Hsu et al., 2003). Following promoter clearance, the polymerase starts synthesis of RNA by adding nucleotides complementary to the template strand from the 5’ to 3’. This step is characterized by the release of the sigma factor from the holoenzyme, in a process called ‘sigma factor recycling’, while the core enzyme proceeds with the synthesis of RNA until the termination site is reached. In the final step, transcription termination, the newly formed mRNA and the RNAP are released from the elongation complex (Browning et al., 2004). In E. coli, transcription termination occurs by one of two mechanisms namely, the Rho-dependent and Rho-independent mechanisms. In the rho-dependent mechanism, a protein factor, ‘Rho’, destabilizes the elongation complex which releases the RNA (Ciampi, 2006), whereas in Rho-independent mechanisms, which are also called intrinsic termination mechanisms, transcription termination occurs typically by the formation of a G-C rich hairpin loop followed by a poly U-tail, which leads to the falling off of the synthesis of the mRNA and the corresponding proteins is very low. Conversely, transcription of a gene is said to be in the ‘active’ or ‘on’ state when the mRNA and the corresponding proteins are produced in abundance, i.e., at high rates. Transcription begins when an RNA polymerase binds to the promoter region of a gene and initiates the synthesis of RNA (Fig. 3). The process of recognition of the promoter occurs by three-dimensional diffusion of the RNA polymerase on the DNA strand (Wang et al., 2013). Following promoter recognition, RNA synthesis occurs in three stages namely ‘Initiation’, ‘Elongation’ and ‘Termination’.

In transcription initiation, the RNA polymerase binds to the promoter and, once finding a transcription start site, it forms a closed complex (CC). This is followed by a series of intermediate isomerization steps which, eventually, lead to the formation of the open complex (OC). Upon successful open complex formation, the promoter becomes catalytically active allowing the RNA polymerase to synthesize new RNA. During elongation, the RNA polymerase (RNAP) traverses through the template strand until the termination site, synthesizing the RNA. For successful entry into the elongation phase, the RNA polymerase has to undergo a crucial step called ‘promoter clearance’: typically, a minimum of 10 base pairs synthesis should occur for the RNA polymerase to continue synthesizing the RNA. Failing to achieve this usually leads to abortive transcription (Hsu et al., 2003). Following promoter clearance, the polymerase starts synthesis of RNA by adding nucleotides complementary to the template strand from the 5’ to 3’. This step is characterized by the release of the sigma factor from the holoenzyme, in a process called ‘sigma factor recycling’, while the core enzyme proceeds with the synthesis of RNA until the termination site is reached. In the final step, transcription termination, the newly formed mRNA and the RNAP are released from the elongation complex (Browning et al., 2004). In E. coli, transcription termination occurs by one of two mechanisms namely, the Rho-dependent and Rho-independent mechanisms. In the rho-dependent mechanism, a protein factor, ‘Rho’, destabilizes the elongation complex which releases the RNA (Ciampi, 2006), whereas in Rho-independent mechanisms, which are also called intrinsic termination mechanisms, transcription termination occurs typically by the formation of a G-C rich hairpin loop followed by a poly U-tail, which leads to the falling off of the
newly synthesized strand from the template strand, and hence the release of the RNA and RNAp from the elongation complex (Nudler et al., 2002).

Figure 3: Steps in transcription (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology, Grieve et al., 2005)

2.3 Translation

In translation, information contained in the mRNA sequence is translated into proteins. This process is carried out by the ribosomes (Fig. 4). In prokaryotes, due to lack of intracellular compartmentalization of the cellular components and of a nuclear membrane, transcription and translation are coupled (Miller, 1970) (Brown et al., 1997); hence translation of proteins can start as soon as the ribosome binding site sequence of the RNA (RBS) is transcribed (Laursen et al., 2005). Like transcription, translation has three main steps (initiation, elongation and termination), with translation initiation being the most rate-limiting and regulated step. Translation initiation occurs when IF-3 (initiation factor) binds to the 30S subunit of the ribosome, leading to the dissociation of the ribosomes into subunits (Petrelli, 2001). This is followed by binding of IF-1 to the A-site (Amino-acyl site) of the ribosome, thus directing the initiator tRNA to the P-site (Peptidyl site) of the ribosome (Carter et al., 2001; Dahlquist et al., 2000). IF-1 also plays a crucial role in stimulating the activities of IF-3, hence aiding in the dissociation of the ribosomes newly synthesized strand from the template strand, and hence the release of the RNA and RNAp from the elongation complex (Nudler et al., 2002).

Figure 3: Steps in transcription (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology, Grieve et al., 2005)

2.3 Translation

In translation, information contained in the mRNA sequence is translated into proteins. This process is carried out by the ribosomes (Fig. 4). In prokaryotes, due to lack of intracellular compartmentalization of the cellular components and of a nuclear membrane, transcription and translation are coupled (Miller, 1970) (Brown et al., 1997); hence translation of proteins can start as soon as the ribosome binding site sequence of the RNA (RBS) is transcribed (Laursen et al., 2005). Like transcription, translation has three main steps (initiation, elongation and termination), with translation initiation being the most rate-limiting and regulated step. Translation initiation occurs when IF-3 (initiation factor) binds to the 30S subunit of the ribosome, leading to the dissociation of the ribosomes into subunits (Petrelli, 2001). This is followed by binding of IF-1 to the A-site (Amino-acyl site) of the ribosome, thus directing the initiator tRNA to the P-site (Peptidyl site) of the ribosome (Carter et al., 2001; Dahlquist et al., 2000). IF-1 also plays a crucial role in stimulating the activities of IF-3, hence aiding in the dissociation of the ribosomes newly synthesized strand from the template strand, and hence the release of the RNA and RNAp from the elongation complex (Nudler et al., 2002).
Following the dissociation of the ribosomal subunits IF2, mRNA, and N-Formylmethionine (fMet-tRNA) associated with the 30S ribosomal subunit, it occurs the interaction of the Shine-Dalgarno (SD) sequence of mRNAs with the anti-SD sequence of the 16S rRNA, which leads to the positioning of the initiation codon to the P-site of the ribosome. At this point, the 30S pre-initiation complex, consisting of the three initiation factors, mRNA and the fMet-tRNA, undergoes a rate-limiting conformational change that promotes the interaction between codon and anti-codon, which leads to the formation of a stable 30S initiation complex (Pon et al., 1984; Gualerzi et al., 1977). Following the formation of a stable 30s initiation complex, IF-3 and IF-1 are ejected from the subunit, while IF-2 promotes the binding of the 50S ribosomal subunit to the 30S initiation complex, by undergoing hydrolysis of the GTP, leading to the formation of 70S initiation complex, which enters the elongation phase.

In elongation, a poly-peptide chain is assembled by the ribosomes from the start codon until a stop codon of the mRNA. Briefly, an aa-tRNA is first bound to the A-site and, following the establishment of proper base-pairing between the codon of the mRNA and the anticodon of the tRNA at the A-site, a peptide bond is formed with the amino acid and/or peptide, attached to the tRNA in the P-site. This is accompanied by the transfer of the peptide to the tRNA in the A-site. The peptidyl tRNA moves from the A-site to the P-site, and the deacylated-tRNA moves from the P-site to the E-site (Exit site), displacing the tRNA deacetylated in the previous cycle. The mRNA is coordinately translocated by one codon, hence retaining the correct reading frame and tRNA-mRNA base-pairing (Merrick et al., 2000; Ramakrishnan, 2002). In optimal conditions, elongation proceeds at a rate of 10–15 amino acids per second and with a very low error rate (Rodnina et al., 2001).
In termination, the nascent polypeptide chain is released and the ribosomal complexes are

Figure 4: Steps involved in translation
(Adapted from: http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-22/CB22.html).

Figure 4: Steps involved in translation
(Adapted from: http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-22/CB22.html).
In bacteria, the recognition of a stop codon in the mRNA in the A-site is performed by
two different release factors, which belong to the ‘class I’ release factors, namely RF1 and RF2. RF1 recognizes the UAA and UAG stop codons while RF2 recognizes UAA and UGA. Dissociation of the nascent peptide chain occurs when a class II release factor, RF3 GDP, binds to the ribosome in the presence of a class I factor. Upon release of the nascent peptide, the GDP bound to RF3 is exchanged to GTP, accompanied by conformational changes and dissociation of the class I factor. GTP hydrolysis, in turn, causes dissociation of RF3. After termination, the ribosome is recycled by a ribosome-recycling factor, RRF, which, together with EF2, completes the process of dissociation of the two subunits (Merrick et al., 2000; Ramakrishnan et al., 2002; Kisselev et al., 2003).

2.4 Regulation of transcription

Gene expression in E. coli, aside from the genes that maintain the basic processes in the cells, occurs, in most cases, in response to different signals that may arise from internal processes or external factors, such as changes in temperature, pH, availability of nutrients etc. It is pivotal for the organism to regulate the expression of the number and the type of proteins needed to survive in a given environment, and prevent unwanted spurious production of proteins or RNA that is not required at a given time. Given the fact that transcription and translation are coupled in E. coli and the fact that transcription initiation is subject to stringent regulation, spurious protein productions are relatively rare, although, sporadic production events are known to occur (Cai et al., 2006). In this section, a brief account on the positive and negative regulation mechanisms of gene expression is provided.

2.4.1 Activation

A promoter is transcriptionally active following the successful binding of the RNA polymerase, leading to the open complex formation and hence production of RNA. Activation of gene expression from a promoter occurs in response to signals that arise from the environment or based on the intracellular needs. For example activation of the...
sugar metabolizing genes occur in the presence of the respective sugars, whereas activation of genes involved in the synthesis of amino acids occur when amino acids are depleted. The strength of activation from a promoter is dependent on the sequences, which play a crucial role in the recruitment of the RNA polymerase to initiate transcription. (VonHippels et al., 1984). However, this provides only static regulation, as sequence level regulation cannot be tuned to the changing conditions.

The promoter sequence also determines the selection of the sigma factor involved in transcription. Sigma factors, in addition to recognizing the promoter, are involved in the melting of the double stranded DNA to form the open complex. There are seven types of sigma factors, which are involved in transcription of specific sets of promoters based on their functions. The intracellular levels of sigma factors vary with differences in growth and stress conditions (Ishihama, 2000; Ishihama, 2010). The recognition of the promoters by sigma factors is highly sequence specific (Campbell et al., 2008) and different sigma factors are involved in transcription of distinct sets of genes. For example, sigma70 is involved in the transcription of the majority of the genes active during active growth, whereas other sigma factors regulate transcription of discrete sets of genes that are required for various processes such as adaptation to different conditions. For example, sigma32 regulates the expression of genes involved in heat shock response, and sigma38 regulates genes involved in stress responses (Gruber et al., 2003).

Based on their activation properties, promoters can be classified as ‘constitutive’ and ‘inducible’ promoters. In case of constitutive promoters, RNA production is always ‘ON’ or active in all circumstances. These promoters control the expression of, in general, the essential genes, such as genes that code for rRNA, RNA polymerase, etc. (Shimada et al., 2014). However, overexpression of the unnecessary and harmful genes under the control of the constitutive promoter is regulated by factors like the H-NS family proteins (Oshima et al., 2006). Inducible promoters, on the other hand, produce RNA upon induction with specific inducers and transcription factors. These include promoters that control the expression of genes involved in metabolizing sugars, shock response genes, etc. (Jacob et al., 1961; Guisbert et al., 2004). In this class of promoters, activation is achieved by derepression of the promoters by binding of the inducer molecules to the repressor, which

Based on their activation properties, promoters can be classified as ‘constitutive’ and ‘inducible’ promoters. In case of constitutive promoters, RNA production is always ‘ON’ or active in all circumstances. These promoters control the expression of, in general, the essential genes, such as genes that code for rRNA, RNA polymerase, etc. (Shimada et al., 2014). However, overexpression of the unnecessary and harmful genes under the control of the constitutive promoter is regulated by factors like the H-NS family proteins (Oshima et al., 2006). Inducible promoters, on the other hand, produce RNA upon induction with specific inducers and transcription factors. These include promoters that control the expression of genes involved in metabolizing sugars, shock response genes, etc. (Jacob et al., 1961; Guisbert et al., 2004). In this class of promoters, activation is achieved by derepression of the promoters by binding of the inducer molecules to the repressor, which
leads to a conformational change in the structure of the repressor, effectively inactivating the binding of the repressor to the DNA, thus allowing the activation of transcription from the promoter. Presence of other transcription factors, such as the Catabolite Activator Protein (CAP), which binds to the CAP binding sites, further, enhance the activation of the promoters by actively recruiting RNA polymerases (Lichstein et al., 1987; Jacob et al., 1961).

### 2.4.2 Repression

Repression is the process of negative regulation or blocking of transcription. This is achieved by binding of the repressor to a repressor binding site, the operator site, of the promoter. The repressors specific for regulation of a particular promoter usually are encoded by the operon that promoter belongs to, although its expression is under the control of a different promoter. For example, the production of the lac repressor is under the control of lacIq, a constitutive promoter (Jacob et al., 1961). Repression usually occurs when there is no inducer present, such as in case of the sugar metabolizing operons, or when the final product is present in excess, leading to repression by negative feedback as in case of the tryptophan operon (Santillan et al., 2000).

Repression of a promoter can be achieved in several ways. The simplest mechanism is by steric hindrance, in which the repressor binds to the DNA, physically blocking the formation of the open complex at the start site by the RNA polymerase. In other cases, repressors form DNA loops by forming oligomers and binding to multiple operators. For example, AraC forms a dimer, which in the absence of arabinose, forms a DNA loop by binding to the araI1 and araO2 (Schleif, 2010). The Lambda phage repressor CI is known to form higher order oligomers up to octamers for effective repression (Dodd et al., 2001). ‘Carbon catabolite repression’ is another mechanism, and allows the bacteria to metabolize the preferred sugar when growing in an environment which contains more than one sugar (Duetscher, 2008). For example, the expression of the Lac operon is blocked when glucose is available (Lichstein et al., 1987; Jacob et al., 1961).
3 Synthetic biology

Synthetic biology is an alternative, bottom-up approach to the more traditional top-down approaches of Molecular Biology. In this approach, well characterized modular genetic components are assembled using principles derived from electrical engineering to create small synthetic genetic circuits that are capable of emulating the characteristics of the naturally occurring genetic circuits, such as oscillatory behavior and switching phenotypes, among other. Fig. 5 shows an example of how principles from electrical engineering can be used to engineer synthetic genetic circuits that emulate a natural system. Use of this strategy to construct genetic circuits offers several advantages. For example, it allows the development of an arbitrary circuit to test its efficiency in performing a desired function, without impairing or altering the basic cellular mechanisms. Also, it can be used to assess the advantages and disadvantages of different architectures that perform similar functions and hence determine the preferable architecture. Finally, this approach provides an opportunity to test the behaviour of circuits developed \textit{in silico}. A number of synthetic genetic circuits have been developed already, which have shed light on some of the key design principles of genetic circuits (Sprinzak \textit{et al.}, 2005). In this chapter, the fundamental aspects of design and development of synthetic genetic circuits are described.
3.1 Design principles of synthetic genetic circuits

The design principles of synthetic genetic circuits are obtained by drawing analogies between naturally occurring genetic circuits and the established techniques in electrical engineering (Miyamoto et al., 2013). This approach involves development of a blue print of the circuit design, which can be analyzed and tested using mathematical models and computational tools, while experimental techniques are used to construct networks according to the blue print. Recent advances in sequencing and genetic engineering aid in the design and development of these circuits, as they help in the characterization and assembly of the modular components, respectively. Logic gates, associated with their truth tables, form the basis for designing a synthetic genetic circuit, which aids in understanding the schematics of the input and output characteristics of the circuits (Weiss et al., 2003).
3.1.1 Logic gates

Biological networks are highly sophisticated devices, but nevertheless resemble electrical devices in the following way: they sense input signals from the environment, then perform decision making by processing the input information and, then, produce an output, similar to electrical circuits. However, biological systems are far more complicated, in that they are able to handle and benefit from the stochasticity of the underlying processes, they can receive and process multiple input signals that arise from various environmental factors such as nutrients, pH, osmotic stress, temperature etc. and, they process the input signals in parallel and respond to these inputs with little or no delays (Eldar et al., 2010).

Synthetic biology's design principles are drawn from electrical engineering and implemented in cells using experimental approaches. Typically, in the process of designing of an electrical device containing a circuit that produces a desired output, requires designing of a logic gate that consists of an arithmetic logic unit, the control unit, memory, and the input and output devices. The input and output characteristics are represented as Boolean logic gates. For example, in case of a binary code, the basic unit information is represented as 0 and 1 which indicate the two possible states of the circuit. A threshold is set to define the input and output range so that, depending on the value, either lower or higher than the threshold, the state of the logic circuit is denoted as 0 or 1, respectively. In digital circuits, a number of such logic elements are interconnected to create logic gates which are capable of executing Boolean logic functions. Fig. 6, shows examples of logic gates that can be used for the development of synthetic genetic circuits. A number of such modular logic gates can be networked together to make complex circuits thus making it possible to build complex circuits that can process versatile inputs (Miyamoto et al., 2013).

In Synthetic Biology, Boolean logic gates can also be used to design synthetic genetic circuits. In this case, chemical concentrations of proteins or inducers act as inputs and outputs for the logic gates (Lim, 2010; Morris et al., 2010; Slusarczyk et al., 2012). One crucial aspect to be considered while applying these concepts to living cells is that the
values of the thresholds of the input and output that define if these equal ‘0’ or ‘1’ must be well defined (Weiss et al., 2003). These thresholds are, usually, defined based on the concentration of the effector molecules; such as proteins, metabolites, and inducers, or on the rate of the chemical reactions, or on the localization of the effector biomolecules. Such characterization allows, e.g. the combinatorial synthesis of a wide variety of synthetic genetic networks with different architectures (Guet et al., 2002; Khalil et al., 2010).

Figure 6: Boolean logic gates and the associated truth tables "Reprinted with permission from (Miyamoto et al., 2013). "American Chemical Society".
3.2 Components of synthetic genetic circuits

Synthetic genetic circuits are built using well characterized, modular genetic components that can act independently of other cellular processes (Sprinzak et al., 2005). In the following section, it is provided a brief description of the basic components required for construction of synthetic genetic circuits.

3.2.1 Transcription unit

Transcription units are the principal components of genetic circuits. They consist of DNA sequences coding for a set of coordinately regulated structural genes that encode proteins. The organization of regulators and genes of a transcription unit of a synthetic genetic circuit typically follows the organization observed in the natural systems (Nandagopal et al., 2011). These contain an up-stream promoter site where regulation of transcription occurs, and a down-stream terminator site where transcription ceases. Modulator sites associated with the promoter can be bound by regulatory proteins that influence the rate of transcription initiation; repressors bind operator sites that down-regulate high-level promoters, or activators bind initiator sites to up-regulate low-level promoters. These units are thus responsible for containing binding sites for regulatory molecules and for detecting such binding events. The regulatory molecules determine the production rate from a gene, whose products can be other regulatory molecules, among other (Jacob et al., 1961).

3.2.2 Input signal

The input signals for transcription units can arise from the environment or from within the cell. When the signal is from the environment, cells make use of cell surface receptors to detect the incoming signal. The signals from membrane-bound receptors are communicated to the inside via short signal transduction pathways (Hoch, 2000). In other systems, molecules are transported across the membrane and modified metabolically, to become signal molecules that influence transcription. Signals that arise from within the

3.2 Components of synthetic genetic circuits

Synthetic genetic circuits are built using well characterized, modular genetic components that can act independently of other cellular processes (Sprinzak et al., 2005). In the following section, it is provided a brief description of the basic components required for construction of synthetic genetic circuits.

3.2.1 Transcription unit

Transcription units are the principal components of genetic circuits. They consist of DNA sequences coding for a set of coordinately regulated structural genes that encode proteins. The organization of regulators and genes of a transcription unit of a synthetic genetic circuit typically follows the organization observed in the natural systems (Nandagopal et al., 2011). These contain an up-stream promoter site where regulation of transcription occurs, and a down-stream terminator site where transcription ceases. Modulator sites associated with the promoter can be bound by regulatory proteins that influence the rate of transcription initiation; repressors bind operator sites that down-regulate high-level promoters, or activators bind initiator sites to up-regulate low-level promoters. These units are thus responsible for containing binding sites for regulatory molecules and for detecting such binding events. The regulatory molecules determine the production rate from a gene, whose products can be other regulatory molecules, among other (Jacob et al., 1961).

3.2.2 Input signal

The input signals for transcription units can arise from the environment or from within the cell. When the signal is from the environment, cells make use of cell surface receptors to detect the incoming signal. The signals from membrane-bound receptors are communicated to the inside via short signal transduction pathways (Hoch, 2000). In other systems, molecules are transported across the membrane and modified metabolically, to become signal molecules that influence transcription. Signals that arise from within the
cells are from other transcription units. These are, in most cases, regulatory proteins or regulatory non-coding RNAs (Friedland et al., 2009; Jacob et al., 1961).

3.2.3 Regulators

Regulation is achieved by factors that control gene expression by activating or by blocking the expression of a gene. Each of these two modes requires the transcription unit to have a promoter with sites for binding of the regulatory molecules. The behaviour of the circuits can be regulated by expressing the regulators constitutively and altering the functional forms of the regulators. In other cases, the regulator is always functional, and the variability in expression is achieved by regulating the rate of synthesis and degradation of the regulators (Savageau, 2001). Small non-coding RNAs have also been widely used to regulate synthetic genetic circuits (Lucks et al., 2011; Qi et al., 2014; Yuan et al., 2015).

3.2.4 Connectivity

Connectivity between genes allows cells to perform complex functions (Hlavacek et al., 1996) and it is determined by how the outputs of the transcription units are connected as inputs of other transcription units. Since these interactions can vary in function (repression or activation) and differ in strength, the number of different networks that can be formed is virtually infinite, assuming a large number of transcription units. A large number of the connections are ensured by proteins modulating the expression of the transcription unit that encoded them (Hlavacek et al., 1995). A number of synthetic genetic circuits that can perform a wide range of functions have been developed by connecting a wide variety of outputs to the input signal. Also, a number of oscillators have been engineered by combining promoters and transcriptional regulators (Elowitz et al., 2000; Gardner et al., 2000; Stricker et al., 2008).

3.2.5 Expression cascades

Expression cascades correspond to the production of output signals from transcription units. The initial signal coming from a transcription unit is an mRNA molecule that has a
sequence complementary to the DNA read. The mRNA is translated into a protein. Finally, the protein, once folded and activated, becomes an enzyme. There can be more stages in such cascades and each of these can be used for regulation. For example, in retroviruses and retrotransposons the cascade might include a stage in which RNA is used to transcribe a complementary DNA. Several mRNA molecules can be produced from the same transcription unit. Also, several proteins can be produced from the same mRNA (e.g. in bacteria) (Hlavacek et al., 1995). It has also been demonstrated that complex functions can be performed by integrating the transcriptional units with the output signals. In ‘metabolator’, oscillations are achieved by integrating transcriptional and metabolic activities (Nem et al., 2005). Other complex functions, exhibited by living systems, such as cell synchronization and communication between cells, have been achieved by connecting the output signals with the input signals (Bacchus et al., 2012; Danino et al., 2010).
4 Single cell, single molecule method to study dynamics of transcription

This section describes the method we used to study transcription initiation dynamics. This method is used in the publications 1 and 2. In publication 1, we used this method to study transcription dynamics from a promoter and the influence of inducer intake on the kinetic of transcription. For that, we made use of P_{BAD} promoter which controls the expression of the arabinose operon. In publication 2, this method was used to confirm that loss of functionality of the Repressilator with increasing temperature is due to the altered properties of cI. For this, we used the cI-cro switch, which is under the control of promoters, P_{R} and P_{RM}, respectively.

4.1 Transcription initiation dynamics

Transcription initiation is a sequential, multi-stepped process (Mcclure, 1980), which begins when an RNA polymerase (R) recognizes and binds to the promoter (P) region of the DNA, forming a closed complex (RP_{c}), which subsequently undergoes a series of isomerization steps (I_{n}) leading to the formation of the open complex (RP_{o}) (Fig. 7). Some of the steps in this process are rate-limiting (“slow” in Fig. 7) and the number of rate-limiting steps and their duration vary between the promoters and with conditions, such as temperature.
4.2 Measuring time intervals between consecutive RNA production events in a single cell

A number of previous studies have been done to determine the dynamics of transcription initiation in different promoters (Bertrand-Burggraf et al., 1984; Hawley et al., 1983; McClure, 1980). However, these studies were based on in vitro methods and fixed cells (Larson et al., 2011).

The MS2-GFP tagging of RNA described in (Golding et al., 2004), offers an opportunity to study the transcription events in vivo (a detailed description is given in the following section). Using this technique, our group has developed a method to infer the dynamics of transcription initiation, specifically the number and the duration of the rate-limiting steps (Kandhavelu et al., 2011; Muthukrishnan et al., 2012).

4.2.1 MS2-GFP tagging method

The MS2-GFP tagging method is one of the currently used single-molecule techniques to detect RNA production in live cells with sensitivity and, thus, measure the real-time kinetics of transcription. The method was originally developed to study eukaryotic mRNA (Bertrand et al., 1998), and was later adapted to prokaryotes with some modifications (Golding et al., 2004; Golding et al., 2005). This method is based on the fact that MS2, the coat protein of MS2 bacteriophage, recognizes and binds to a specific RNA sequence. For RNA tagging using this technique, two plasmids are used, namely the target coding plasmid and the reporter coding plasmid. The target plasmid contains the
The promoter of interest that controls the production of monomeric red fluorescent protein (mRFP-1) followed by tandem 96 binding sites for MS2. The reporter plasmid encodes the production of MS2 tagged with GFP, and is under the control of a different promoter than the target plasmid. Induction of the reporter plasmid floods the cells with MS2-GFP. Subsequently, when the promoter of interest is activated, the RNA produced can be visualized as spots under a fluorescence microscope (Fig. 8), due to the concentration of MS2-GFP molecules at the tail of the RNA produced. Time-lapse imaging of this system in multiple cells allows the extraction of a distribution of intervals between consecutive RNA productions in individual cells, from which one can extract the mean rate of RNA production and variability of the intervals between productions. From this, one can thus also assess the noise in RNA production. Further, assuming a sequential model of transcription initiation (as proposed in (McClure, 1980)), it is possible to estimate from this distribution the number of rate-limiting steps in transcription as well as their duration.

Figure 8: A) Schematic representation of MS2-GFP RNA tagging B) Example image of cells containing fluorescent RNA spots.
4.2.2 Time-lapse microscopy and Image analysis

Cells containing the target and the reporter genes were grown overnight in Lysogeny broth (LB) medium with appropriate antibiotics, after which cells were sub-cultured into fresh LB medium, to an optical density (O.D.) of 0.01, incubated at 37 °C with shaking at 250rpm, and then grown to an O.D. of 0.3. At this point, the reporter gene, which is P$_{\text{TetO1}}$ in both studies, was activated using 100ng/ml anhydro-Tetracycline (aTc) for at least 45 minutes followed by the induction of the target promoter. This induction was achieved differently in each case; full induction of the P$_{\text{BAD}}$ promoter was achieved by addition of 1% arabinose to the medium, whereas no external induction is required for production of RNA by pRM as the production is constitutive in this promoter (Berthoumieux et al., 2013) although its rate of production is temperature dependent (Little et al., 2010). After induction, the cells were seeded on the imaging gel, made of LB media containing the respective inducers, and placed in a temperature controlled imaging chamber, which maintains the desired temperature during the measurements.

Next, the fluorescence images of the induced cells were acquired using a Nikon Eclipse (TE2000-U, Nikon, Japan) inverted microscope with C1 confocal laser-scanning with a 100x Apo TIRF objective. Images were acquired every minute for 2 hours. GFP fluorescence was measured using 488 nm argon ion laser (Melles-Griot) and a 515/30 nm emission filter. Images were acquired using Nikon EZ-C1 software. Independent replicates were produced to ensure reproducibility of the results.

The images were analyzed using custom programs written in MATLAB 2011b (MathWorks). The cells were detected from the fluorescence images obtained from confocal images using a semi-automatic method described in (Kandhavelu et al., 2012). Briefly, in time series, the area occupied by each cell was masked manually at each time point. Principal component analysis was then used to obtain the dimensions and orientation of a cell within each mask. The segmentation of the fluorescent spots in the cells was done using density estimation with a Gaussian kernel (Chen et al., 2008). From this, background-corrected spot intensities were calculated and summed to produce the total spot intensity in each cell, at each time moment.
4.2.3 Data analysis

The moments of appearance of new RNA molecules in each cell were obtained from the fluorescence images acquired by time-lapse microscopy by fitting the corrected total spot intensity over time in each cell to a monotone piecewise-constant function by least squares. The number of terms was selected using F-test with a P-value of 0.01. Each jump corresponds to the production of a single RNA molecule (Kandhavelu et al., 2011). An example of this procedure is shown in Fig. 9. Note that, in cells that do not contain target RNA molecules at the start of the measurements, the number of novel RNA molecules detected since the start of the measurements until a given moment equals the total number of RNA molecules in the cell at that moment. This is because tagged RNA molecules have a negligible degradation rate during measurements a few hours long. Nevertheless, because some cells already contained target RNA molecules at the start of the measurement, the total RNA numbers within cells at a given moment in time was obtained using a different method. Specifically, when comparing measurements using MS2d-GFP tagging and using plate reader, the total number of MS2d-GFP–tagged RNA molecules was extracted from the total spot intensity distribution, obtained from all cells in an image at a given moment after induction. For this, the first peak of the obtained distribution is set to correspond to the intensity of a single-RNA molecule. The number of tagged RNAs in each spot can be estimated by dividing its intensity by that of the first peak (Golding et al., 2005).
4.3 Dynamics of the $P_{BAD}$ promoter

The RNA production dynamics from the $P_{BAD}$ promoter was studied using the method described above except that the target promoter was induced under the microscope so as to observe the first RNA production in each cell. This work was conducted using the strategies previously developed to study transcription dynamics of promoters in vivo (Kandhavelu et al., 2011; Muthukrishnan et al., 2012). Further, here, a method was
Briefly, the waiting time for the appearance of the first RNA ($t_0$) from the $P_{BAD}$ promoter, following induction with arabinose, and the intervals between consecutive productions ($\Delta t$) were measured as described above, namely, for each cell following induction. To verify whether the appearances of target RNA molecules are due to arabinose uptake, another promoter $P_{lac/ara-1}$, which can be induced with two inducers namely, arabinose and Isopropyl Thio-Galactosidase (IPTG) was observed. From the data, it was observed that production from both promoters is sub-poissionian ($CV^2 = 0.37$), similar to production from $P_{lac/ara-1}$ (Kandhavelu et al., 2011). Moreover, the two promoters exhibit statistically indistinguishable waiting times for RNA production with mean $t_0$ of around 2800s, when induced with Arabinose alone. However, when induced with IPTG alone, the mean $t_0$ of $P_{lac/ara-1}$ is close to 2700 and less noisier (Mäkelä et al., 2013). This distribution of RNA production following induction of the $P_{lac/ara-1}$ with IPTG is statistically distinguishable from the others, indicating that a different activation system is being used.

To study the kinetics of the intake process, we defined, $t_{diff}$, the intake time of an inducer, which differs from $t_0$ in that it does not include the time it takes for transcription initiation to be completed, once initiated. To obtain this quantity, the mean and variance of the $\Delta t$ distribution were subtracted to the mean and variance of the $t_0$ distribution. This method was used due to the fact that a correlation could not be drawn between the $t_0$ and $\Delta t$ distributions. The estimated mean and standard deviation of $t_{diff}$ for both $P_{lac/ara-1}$ and $P_{BAD}$ were used due to the fact that a correlation could not be drawn between the $t_0$ and $\Delta t$ distributions. The estimated mean and standard deviation of $t_{diff}$ for both $P_{lac/ara-1}$ and $P_{BAD}$ developed to study the inducer intake kinetics and its effect on the transient dynamics of transcription following the appearance of the inducers in the media.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Inducer</th>
<th>$\mu_0$ (s)</th>
<th>$\mu_{\Delta t}$ (s)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{BAD}$</td>
<td>1% arabinose</td>
<td>2885</td>
<td>532</td>
<td>$2.83 \times 10^{-15}$</td>
</tr>
<tr>
<td>$P_{BAD}$</td>
<td>0.1% arabinose</td>
<td>3519</td>
<td>481</td>
<td>$4.06 \times 10^{-15}$</td>
</tr>
<tr>
<td>$P_{lac/ara-1}$</td>
<td>1% arabinose</td>
<td>2832</td>
<td>516</td>
<td>$2.48 \times 10^{-8}$</td>
</tr>
<tr>
<td>$P_{lac/ara-1}$</td>
<td>1 mM IPTG</td>
<td>2697</td>
<td>576</td>
<td>$3.32 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Table 1: $P$-values of the Kolmogorov-Smirnov test between $t_0$ and $\Delta t$ distributions for each promoter and induction condition

To study the kinetics of the intake process, we defined, $t_{diff}$, the intake time of an inducer, which differs from $t_0$ in that it does not include the time it takes for transcription initiation to be completed, once initiated. To obtain this quantity, the mean and variance of the $\Delta t$ distribution were subtracted to the mean and variance of the $t_0$ distribution. This method was used due to the fact that a correlation could not be drawn between the $t_0$ and $\Delta t$ distributions. The estimated mean and standard deviation of $t_{diff}$ for both $P_{lac/ara-1}$ and $P_{BAD}$ developed to study the inducer intake kinetics and its effect on the transient dynamics of transcription following the appearance of the inducers in the media.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Inducer</th>
<th>$\mu_0$ (s)</th>
<th>$\mu_{\Delta t}$ (s)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{BAD}$</td>
<td>1% arabinose</td>
<td>2885</td>
<td>532</td>
<td>$2.83 \times 10^{-15}$</td>
</tr>
<tr>
<td>$P_{BAD}$</td>
<td>0.1% arabinose</td>
<td>3519</td>
<td>481</td>
<td>$4.06 \times 10^{-15}$</td>
</tr>
<tr>
<td>$P_{lac/ara-1}$</td>
<td>1% arabinose</td>
<td>2832</td>
<td>516</td>
<td>$2.48 \times 10^{-8}$</td>
</tr>
<tr>
<td>$P_{lac/ara-1}$</td>
<td>1 mM IPTG</td>
<td>2697</td>
<td>576</td>
<td>$3.32 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Table 1: $P$-values of the Kolmogorov-Smirnov test between $t_0$ and $\Delta t$ distributions for each promoter and induction condition

To study the kinetics of the intake process, we defined, $t_{diff}$, the intake time of an inducer, which differs from $t_0$ in that it does not include the time it takes for transcription initiation to be completed, once initiated. To obtain this quantity, the mean and variance of the $\Delta t$ distribution were subtracted to the mean and variance of the $t_0$ distribution. This method was used due to the fact that a correlation could not be drawn between the $t_0$ and $\Delta t$ distributions. The estimated mean and standard deviation of $t_{diff}$ for both $P_{lac/ara-1}$ and $P_{BAD}$ developed to study the inducer intake kinetics and its effect on the transient dynamics of transcription following the appearance of the inducers in the media.
were found to be similar (mean and standard deviation of around 1400 and 1100 s, respectively) when both are induced with 1% Arabinose, which is expected since the same intake system is used. Importantly, lac/ara-1, when induced with IPTG, exhibited a smaller standard deviation (700s), although it had a similar mean (approximately 1400s), which suggests that the intake process of arabinose is noisier than the intake process of IPTG. We also found that the mean $t_{\text{diff}}$ is significantly affected by the concentration of arabinose in the medium (Mäkelä et al., 2013). Using a stochastic model of the intake process that accounts for the inducer concentration and the empirical distribution of $t_0$, and that assumes that transcription initiation is a 3-stepped process, we further found that the intake step likely adds variability to the RNA numbers in individual cells during the transient stage to reach the equilibrium (i.e. maximum activation possible given the inducer concentration in the medium).

4.4 Dynamics of the CI-Cro switch

The CI-Cro bistable switch, controlled by $P_{\text{RM}}$ and $P_R$, present in the λ bacteriophage, controls the switching between the lytic and lysogenic states of the virus. In the lysogenic state, the CI repressor is produced in abundance, repressing $P_R$ from producing Cro, which is required to switch from lysogenic to lytic state. In the lytic state, the production of CI from $P_R$ is repressed by Cro (Schubert et al., 2007).

Here, we study the RNA production dynamics of the CI–Cro genetic switch as a means to indirectly validate that the observed changes with temperature in the dynamics of the genetic Repressilator are due to the temperature-dependence of the functionality of CI, a repressor of the promoter $P_R$.

Previous studies suggest that the activity of CI is temperature dependent (Jana et al., 1999). In the switch system used here, the promoter $P_{\text{RM}}$, which controls the expression of CI, also controls the transcription of a 96 MS2d binding array. The other promoter of the switch, $P_R$, controls the expression of Cro. The activity of this promoter is not followed. From previous studies, it was observed that Cro–DNA interactions do not vary.
significantly between temperatures from 24–37°C (Takeda et al., 1992), thus, any behavioural changes observed in the switch with changing temperature in this range should arise from the changes in CI–DNA interactions. For this, time lapse microscopy was performed to obtain the RNA production intervals from the promoter for 2 hours, with images taken every minute. The measurements were conducted at 24 °C, 27 °C, 30 °C, 33 °C and 37 °C.

Table 2 shows the number of samples (i.e. intervals), mean and standard deviation of the intervals’ duration in each condition. It is observed that as temperature increases, the kinetics of production of the target RNA shifts from sub-poissonian ($CV^2 < 1$) to super-poissonian ($CV^2 > 1$). From Table 2 it is also observed that as the temperature increases, the mean interval between consecutive transcription events decreases. The significance of this change was verified by comparing the distributions of intervals in consecutive temperatures with the K–S test. Table 3 shows that the distributions at 24 °C and 27 °C cannot be distinguished from one another. A similar trend was observed in case of the distributions at 33 °C and 37 °C. Meanwhile, the distributions obtained from 27 °C and 30 °C, as well as distributions from 30 °C and 33 °C differs significantly from one another. This suggests that there is a change in the dynamics of transcript production around 30 °C, which is similar to the change observed in the behaviour of the Repressilator (Results in the next chapter).

<table>
<thead>
<tr>
<th>T (°C)</th>
<th>No. of intervals</th>
<th>μ (s)</th>
<th>σ (s)</th>
<th>$CV^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>157</td>
<td>1242</td>
<td>1166</td>
<td>0.88</td>
</tr>
<tr>
<td>27</td>
<td>229</td>
<td>1152</td>
<td>1191</td>
<td>0.67</td>
</tr>
<tr>
<td>30</td>
<td>88</td>
<td>1130</td>
<td>1040</td>
<td>0.85</td>
</tr>
<tr>
<td>33</td>
<td>539</td>
<td>788</td>
<td>807</td>
<td>1.05</td>
</tr>
<tr>
<td>37</td>
<td>324</td>
<td>714</td>
<td>785</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Table 2: Intervals between appearances of new RNA molecules in individual cells. Table shows per condition, number of intervals, $μ$ (s) is mean, $σ$ (s) is standard deviation and $CV^2$ is coefficient of variation.
<table>
<thead>
<tr>
<th>$T$ (°C)</th>
<th>24</th>
<th>27</th>
<th>30</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>0.149</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.06</td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td>0.478</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td>0.478</td>
</tr>
</tbody>
</table>

**Table 3:** P-values of the Kolmogorov-Smirnov test between distributions of intervals between consecutive RNA production events from $P_{RM}$. For p-values <0.01, the hypothesis that the distributions are the same is rejected.
5 Single cell study of the dynamics of Repressilator

This chapter contains a brief introduction to the Repressilator, followed by a detailed description of the experimental and statistical methods used in the study. The Repressilator was chosen as a model for study as the dynamics of this genetic circuit is dependent on the activity of the promoters. The transcription kinetics of one of the promoters of the Repressilator has been described in Chapter 4 and it is found to be affected by temperature. In this chapter, we describe the effect of changes in temperature and copy-number on the dynamics of the Repressilator.

5.1 Repressilator

The Repressilator is one of the first built synthetic genetic circuits, which is known to be functional in a living organism. It is a negative feedback loop consisting of transcriptional regulators that are not a part of any natural biological clocks (Weiss et al., 2003). This is a ring-type circuit, which contains 3 promoters, each controlling the expression of a repressor that represses the activity of the subsequent promoter in the loop. As such, the expression of one of the promoters represses the activity of the next promoter leading to the expression of the third promoter, which in turn suppresses the activity of first promoter (Fig. 10). The resultant fluctuations in the level of the repressors is read-out as periodic oscillations in the green fluorescence, GFP, produced by a reporter plasmid, which is under the control of one of the promoters of the Repressilator (Elowitz et al., 2003).
The promoter and repressor pairs used in the construction of the repressilator are: $P_L$ LacO1 – LacI; $P_L$ TetO1 – TetR; and $P_R$ – cl, and the expression of GFP from the reporter plasmid is kept under the control of $P_L$ TetO1, which is one of the promoters of the repressilator. The proteins produced from the repressilator contain a destruction tag, ssrA, which makes the half-life of the proteins closer to the half-life of mRNA (Keiler et al., 1996).

**Figure 10**: a) Schematic representation of the Repressilator and of the reporter (Reprinted by permission from Macmillan Publishers Ltd, Nature (Elowitz et al., 2000)); b) A logic circuit based representation of the repressilator. (Reprinted with kind permission from Springer Science and Business Media (Weiss et al., 2003).)
5.2 Time lapse microscopy

*E. coli* cells containing the Repressilator and the reporter plasmids were grown in minimal media for up to 10 hours to an O.D. of 0.1, at 28 °C, 30 °C, 33 °C and 37 °C. The cells were then diluted in fresh media and seeded on imaging gel, made of minimal media and 0.2% low melting agarose. Fluorescent images were captured using a confocal microscope every 15 minutes for 10 hours, after which the cells enter the stationary phase and produce no oscillations (Elowitz *et al.*, 2000). The temperature, during the measurement, was maintained constant using a thermal control chamber. Independent replicates were produced to ensure reproducibility of the results.

5.3 Image analysis

Fluorescent images of cells containing the Repressilator were first observed and cells that appear to produce oscillations in the levels of fluorescence were selected by inspection for further analysis. Next, the selected cells were segmented by manually masking the area the cells occupy in each frame. Using custom MATLAB based programs developed for image analysis, the total fluorescence intensities under each mask were extracted and the mean pixel intensity of cells was calculated at each time point (Fig. 11).

![Figure 11](image1.png)

*Figure 11*: Example images of cells with repressilators in five different time points (top), and the corresponding intensity plots (bottom).
corresponding mean intensities extracted from these cells (bottom). The vertical dashed lines indicate the time points corresponding to the images.

5.4 Statistical analysis

The effect of temperature on the dynamics of the Repressilator was studied by analysing the functionality, robustness and period of oscillations from the time-series images of oscillating cells, in different temperatures. The statistical methods used in the study are as follows.

5.4.1 Functionality and estimation of periods

The functionality of the masked cells containing the Repressilator was determined using the criterion proposed in (Elowitz et al., 2000). To do so, a fast Fourier transform was applied to the temporal fluorescence signal from each cell and divided by the transform of a decaying exponential with a time constant of 90 min, which is the measured lifetime of the green fluorescent protein used (GFPαav) as the reporter. Cells were classified as oscillatory when the power spectra produce peaks 4.5 times higher than the background at frequencies of 0.15–0.5 per hour. A larger bandwidth, compared to the bandwidth used in (Elowitz et al., 2000), was used here, so as to include failed oscillations, which seemed to increase the period of apparent oscillations by nearly double. The apparent period, for cells that were considered functional, was estimated as follows. First, a quadratic curve in the least-squares sense was fitted to the intensity of the time series to estimate the general trend, since the measured intensity is known to be affected by different factors such as, e.g., photo bleaching. The estimated trend was subtracted from the time series and the residual was scaled to unit power, followed by computation of the autocorrelation function. From this function, the period of oscillation was estimated by locating the first and the third zeros of the autocorrelation function, and computing the distance between them (Fig. 12).
Figure 12: Estimation of the period of the Repressilator from the fluorescence intensity signal. The top panel shows the raw signal extracted from the confocal images, the dashed line shows the estimated trend. In the middle panel, the trend from the raw signal was subtracted and the residual was scaled to unit power. The bottom panel shows the autocorrelation function of the treated signal. The distance between the first and the third zeros corresponds to the period of oscillation.

5.4.2 Estimation of robustness and true period of oscillations

In time series, at temperatures over 30 °C, it was observed that a certain fraction of cells, in each condition, produced oscillations followed by a brief period of no activity, in terms of fluorescence levels, which was then followed by resumption of oscillations. This can be due to, either, the failure of the reporter to produce GFP to report an oscillation or expression of reporter is repressed. However, in most cases, the signal from the reporter recovered, following the period of inactivity, thus suggesting failure of oscillations followed by recovery. The method, for estimation of the period of the Repressilator, mentioned above, relies on robust periodic behaviour and it cannot detect if a Repressilator halts its activity for a certain time and then resumes its activity. Instead, it assumes a period length that combines the periods of the true and the failed oscillations.
in which cases the measured time became double compared to other oscillations. To extract the true period, a method based on the fact that distribution of period lengths resemble a bimodal distribution when failure of oscillations occur was used (Fig. 13). Here, the mean and standard deviation of the true period in the population was estimated followed by estimation of the fraction of failed oscillations from the measured periods from each cell. For that, the maximum likelihood estimates for a single Gaussian and a mixture of two Gaussians was determined (from mean and standard deviation of the measured periods), such that the mean and the variance of the second are double than that of the first, found using an iterative expectation maximization algorithm (Dempster et al., 1977). Appropriate models were selected using a likelihood ratio test with significance level of 0.01 between the two models, that is the model is selected only when the p-value of a 2-gaussian model is smaller than 0.01. Finally, the fitting was performed, using a leave-one-out-technique, with each subset of data lacking one of the measured periods, which results in N estimates, each using N-1 measure periods from which the variance of the estimates was calculated.

Figure 13: Distribution of periods (magnitude scaled to represent probability density) for each temperature. Solid lines represent the probability densities of the fitted model with one or two Gaussians. Dashed lines represent the densities of individual components in the case of two Gaussians. For 28 °C and 30 °C, the p-values of the likelihood ratio tests are 0.08 and 1, respectively, indicating a lack of evidence for the two-Gaussian model, whereas for 33 °C and 37 °C, the p-values of the likelihood ratio tests are 0.08 and 1, respectively, indicating a lack of evidence for the two-Gaussian model, whereas for 33 °C and 37 °C,
°C, the p-values are 0.0065 and 0.0015, respectively, indicating that the two-Gaussian model should be favoured over the one-Gaussian model.

5.4.3 Loss of synchrony

The effect of temperature on the loss of synchrony between sister cells was studied by detecting pairs of sister cells in which both of the sister cells were functional, and then manually following them by the criterion mentioned above. The loss of synchrony was estimated by calculating the correlation coefficients between functional sister cells. First, the likeliness of a robust cell to produce a robust sister cell, at 37 °C, was obtained. About 112, 40, and 35 pairs were found where none, one, or both of the cells were robust, respectively, suggesting that the numbers of pairs where either none or both sisters are robust are overrepresented. More specifically, there is about a 0.64 chance for a robust cell to have a robust sister, and about a 0.85 chance for a non-robust cell to have a non-robust sister (cf. 0.37). The significance of the correlations was confirmed by calculating the p-value of one-tailed Fisher’s exact test with the null hypothesis. It suggests that being robust or not is independent of the sister cells, resulting in a p-value smaller than 1.29x10^{-10}. Next, we computed the correlation between the intensity signals of each pair of sister cells. This correlation results from loss of synchrony caused both by division and variations in the behaviour of the cells over their lifetime (i.e. variations/drift in the period and noise in the intensity signal). The distributions of correlation coefficient extracted from each pair of cells are shown in Fig. 14, and the mean and standard deviation of the coefficients are shown in Table 4.

It was found that, on average, as temperature increases, sister cells lose correlation (p-values of 3.17x10^{-3} and 9.90x10^{-4} for 28 vs. 30 °C and 30 vs. 37 °C in one-tailed Welch’s t-test with the null hypothesis that the means are equal). This loss of correlation may be due to the increase in the noise of the period as a function of the temperature, which appears to follow a similar pattern. Accordingly, since the non-robust cells contribute much of the variation in the 37 °C condition, the correlation is restored to a level comparable to the 30 °C condition when only the robust cells are considered. Interestingly, Fig. 14 reveals that, in each condition, most cells are very highly correlated.
However, increases in the temperature results in pairs of cells with wider range of correlation coefficients, including a sizable number of pairs whose series are strongly anticorrelated.

**Figure 14:** Distributions of correlation coefficients between functional sister cells in 28 (top), 30 (middle), and 37 °C (bottom). In the 37 °C condition, the pairs where both cells are robust are represented in dark gray bars, while the others are represented in light gray.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>28 °C</th>
<th>30 °C</th>
<th>37 °C</th>
<th>37 °C, robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation mean</td>
<td>0.87</td>
<td>0.66</td>
<td>0.42</td>
<td>0.63</td>
</tr>
<tr>
<td>Correlation s.d</td>
<td>0.16</td>
<td>0.58</td>
<td>0.66</td>
<td>0.58</td>
</tr>
</tbody>
</table>

**Table 4:** Correlation between sister cells in various temperatures.

However, increases in the temperature results in pairs of cells with wider range of correlation coefficients, including a sizable number of pairs whose series are strongly anticorrelated.

**Figure 14:** Distributions of correlation coefficients between functional sister cells in 28 (top), 30 (middle), and 37 °C (bottom). In the 37 °C condition, the pairs where both cells are robust are represented in dark gray bars, while the others are represented in light gray.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>28 °C</th>
<th>30 °C</th>
<th>37 °C</th>
<th>37 °C, robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation mean</td>
<td>0.87</td>
<td>0.66</td>
<td>0.42</td>
<td>0.63</td>
</tr>
<tr>
<td>Correlation s.d</td>
<td>0.16</td>
<td>0.58</td>
<td>0.66</td>
<td>0.58</td>
</tr>
</tbody>
</table>

**Table 4:** Correlation between sister cells in various temperatures.
5.5 Single copy Repressilator

5.5.1 Construction of single copy Repressilator

To study the effect of copy-number of genes on the dynamics of the Repressilator, we inserted the Repressilator cassette into a single copy F-plasmid, to create the single copy Repressilator (SCR) pBAC2-ITIrLLtCL. This was done by amplifying the functional repressilator cassette, flanked with SmaI restriction enzyme recognition sites, using polymerase chain reaction (PCR), from the original plasmid (generously provided by M. Elowitz, Princeton University, USA) and inserting it into an F-plasmid based single copy plasmid, pTRUEBLUEScript, which was also amplified along with SmaI restriction sites during PCR. Fig. 15 gives a schematic representation of the construction of the single copy Repressilator. The construct was validated by sequencing, and the functionality of the newly constructed Repressilator inside cells was confirmed using qPCR, in addition to the oscillations observed under the microscope. The functionality and robustness of oscillations were determined according to the above mentioned criteria.

5.5.2 Effects of plasmid copy number under optimal conditions

To study the effect of copy number on the dynamics of the Repressilator, time-lapse confocal microscopy of cells containing the SCR and the low-copy Repressilator (LCR) was performed. The experiments were carried out at 30 °C, at which the LCR exhibited shorter period of oscillations with higher functionality and robustness (Chandraseelan et al., 2013). The division rate of the cells under the microscope was relatively slow (division time of ~60 min), thus, it is reasonable to assume that the cells maintain one copy of the plasmid most of the time (Churchward et al., 1982).

Also, from the data, it is evident that the SCR cells had fewer copies of the Repressilator than the LCR, as a Kolmogorov-Smirnov (K-S) test of statistical significance, to determine whether the two sets of time lengths of oscillations could be obtained from equal distributions, show that the dynamics of the circuits differ statistically, p-value of 0.006 (p-values of <0.01 suggest a significant difference between two distributions).
Interestingly, the difference (particularly the higher variance) was not reflected in the robustness of the oscillations of the SCR, which was comparable to the robustness of the LCR. Surprisingly, the SCR exhibited higher functionality in cells at 30 °C than the LCR, though the mean period of oscillations from the SCR was longer and noisier (higher CV²).

5.5.3 Dynamics at different temperatures

Following the study on the effect of temperature on the dynamics of the LCR, we studied the effect of temperature on the SCR. The measurements were conducted in a similar fashion like it was previously done in LCR (Chandraseelan et al., 2013). Briefly, the behaviour of the SCR was measured, using confocal microscopy, by subjecting the cells with SCR to temperatures, 28 °C, 30 °C, 33 °C, and 37 °C. Similar to LCR, the measurement period was limited to 10 hours as the cells enter stationary phase after this point, which is evident from the disappearance of oscillatory behaviour in the cells (Chandraseelan et al., 2013; Elowitz et al., 2000).

The data and image analysis were done as described above. Briefly, from the images, obtained from each condition, extracted the fraction of functional cells (F), the number of cells exhibiting oscillatory fluorescent signal, the robustness (R) of the oscillations in ‘functional’ cells, and the mean and standard deviation of the period (m, s) were extracted.

Results are shown in Table 5. K-S tests of statistical significance were performed to determine whether the sets of oscillations’ time length from the LCR and the SCR could be obtained from equal distributions, for each temperature. From Table 6 it is evident that the two circuits differ significantly in their oscillatory dynamics in all temperatures. However, Table 5 shows that the response of both circuits to changes in temperature is similar, although not identical, in the range tested. As expected, the circuits were similar in the loss of robustness at temperatures beyond 30 °C, due to the loss of functionality of the CI repressor (Jana et al., 1999; Koblan et al., 1991a; Koblan et al., 1991b). One major difference observed in the response of the two circuits to temperature changes is that the functionality of the SCR suffered a steeper decrease with increasing temperature.

Interestingly, the difference (particularly the higher variance) was not reflected in the robustness of the oscillations of the SCR, which was comparable to the robustness of the LCR. Surprisingly, the SCR exhibited higher functionality in cells at 30 °C than the LCR, though the mean period of oscillations from the SCR was longer and noisier (higher CV²).

5.5.3 Dynamics at different temperatures

Following the study on the effect of temperature on the dynamics of the LCR, we studied the effect of temperature on the SCR. The measurements were conducted in a similar fashion like it was previously done in LCR (Chandraseelan et al., 2013). Briefly, the behaviour of the SCR was measured, using confocal microscopy, by subjecting the cells with SCR to temperatures, 28 °C, 30 °C, 33 °C, and 37 °C. Similar to LCR, the measurement period was limited to 10 hours as the cells enter stationary phase after this point, which is evident from the disappearance of oscillatory behaviour in the cells (Chandraseelan et al., 2013; Elowitz et al., 2000).

The data and image analysis were done as described above. Briefly, from the images, obtained from each condition, extracted the fraction of functional cells (F), the number of cells exhibiting oscillatory fluorescent signal, the robustness (R) of the oscillations in ‘functional’ cells, and the mean and standard deviation of the period (m, s) were extracted.

Results are shown in Table 5. K-S tests of statistical significance were performed to determine whether the sets of oscillations’ time length from the LCR and the SCR could be obtained from equal distributions, for each temperature. From Table 6 it is evident that the two circuits differ significantly in their oscillatory dynamics in all temperatures. However, Table 5 shows that the response of both circuits to changes in temperature is similar, although not identical, in the range tested. As expected, the circuits were similar in the loss of robustness at temperatures beyond 30 °C, due to the loss of functionality of the CI repressor (Jana et al., 1999; Koblan et al., 1991a; Koblan et al., 1991b). One major difference observed in the response of the two circuits to temperature changes is that the functionality of the SCR suffered a steeper decrease with increasing temperature.
Figure 15: Construction of single copy Repressilator Plasmids used for the construction of the SCR plasmid. The pBAC2-ITiLlCL plasmid was engineered by linking the Repressilator system amplified region to part of the pBAC2 (Plac/ara-1-mRFP1-MS2-96x) vector (generously provided by Ido Golding of the University of Illinois, USA), containing the single-copy origin of replication (the construction history was generated and adapted using SnapGene® 1.5.2).

This is particularly interesting because it shows that the SCR’s functionality is higher at lower and at optimum temperatures (28 °C and, particularly, 30 °C) but lower at higher
temperatures (33 °C and 37 °C). The other significant difference is that in the LCR the steepest change in mean period length was observed when the temperature was increased from 28 °C to 30 °C (a decrease of 142 min in length), in the SCR it was observed when temperature was increased from 30 °C to 33 °C (an increase by 51 min in length was seen). This was verified by K-S tests of statistical significance to determine whether the sets of oscillation lengths at different temperatures could be obtained from equal distributions, for the LCR and the SCR (Table 7). A \( p \)-value smaller than 0.01 in the LCR when comparing the data from 28 °C and 30 °C was seen, while a similar \( p \)-value was obtained in SCR when comparing the data from 30 °C and 33 °C, which is in agreement with the changes observed in the oscillations’ mean period length with the changes in temperature.

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

Table 5: Kinetics of LCR and SCR at different temperatures. Oscillatory signals were classified by their power spectra exhibiting peaks of more than 3 times the background at frequencies of 0.2-0.5 per hour. Temperature (T), fraction of functional cells (F), number of cells exhibiting oscillations, fraction of robust oscillations (R), mean (m) and standard deviation (s) of the period are shown.
Table 6: \( p \)-values of the K-S test between distributions of periods from different copy number plasmids. For \( p \)-values < 0.01, the hypothesis that the two distributions are the same is rejected.

<table>
<thead>
<tr>
<th>Copy no.</th>
<th>28°C vs 30°C</th>
<th>30°C vs 33°C</th>
<th>33°C vs 37°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCR</td>
<td>3.09e-14</td>
<td>0.13</td>
<td>0.33</td>
</tr>
<tr>
<td>SCR</td>
<td>0.44</td>
<td>6.70e-10</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Table 7: \( p \)-values of the Kolmogorov-Smirnov test between distributions of the LCR and SCR periods from different temperatures. For \( p \)-values < 0.01, the hypothesis that the two distributions are the same is rejected.

Table 6: \( p \)-values of the K-S test between distributions of periods from different copy number plasmids. For \( p \)-values < 0.01, the hypothesis that the two distributions are the same is rejected.

Table 7: \( p \)-values of the Kolmogorov-Smirnov test between distributions of the LCR and SCR periods from different temperatures. For \( p \)-values < 0.01, the hypothesis that the two distributions are the same is rejected.

5.5.4 Effect of perturbations

The effect of perturbations with IPTG of the LCR dynamics has been described in (Elowitz et al., 2000). Here, to assess and compare the effect of such perturbations on the dynamics of the SCR and LCR, cells containing the SCR and LCR were subjected to external perturbations by addition of IPTG continuously into the medium. The perturbation was initiated 180 minutes after the start of the microscopy measurements, such that the cells were allowed to produce at least one cycle of oscillations. Perturbation is achieved when IPTG binds to LacI molecules (Sanchez et al., 2011), hence inducing constant expression of \( P_L \) LacO1, which effectively disrupts the functioning of the Repressilator, as it starts to constantly producing TetR, leading to indefinite repression of the \( P_L \) TetO1 promoter. Since \( P_L \) TetO1 is the promoter, which in addition to controlling the expression of the repressor for the next promoter also controls the expression of GFP, this leads to a loss of any fluorescence signal from the cells.

Three such experiments were performed. In the first experiment, cells with the SCR and cells with the LCR were perturbed by adding 50 \( \mu \)M of IPTG in the medium at 30 °C and...
the robustness of the two circuits to external perturbations was compared. Next, cells with the SCR were perturbed by adding 1 mM of IPTG in the medium and the effects of ‘weak’ (50 µM) and ‘strong’ (1 mM) perturbations were compared. In all these tests, the M63 medium was pre-warmed and the desired concentration of IPTG was added, which was then supplied to the imaging chamber by a peristaltic pump at the rate of 0.3 mL/min. Images were acquired every 15 minutes for 6 hours (3 hours before perturbation and 3 hours after perturbation).

The effects of the ‘weak’ perturbation (50 µM of IPTG) on the dynamics of the LCR and of the SCR were compared. For that, functionality of the cells was assessed in the first 3 hours prior to the perturbation and in the subsequent 3 hours following the perturbation. The functionality of cells with the LCR was found to be equal to 93.3% during the first 3 hours, and 1.64% in the next 3 hours after perturbation (61 cells imaged) (i.e. 98% of the cells were perturbed). Interestingly, the functionality of cells containing the SCR equaled 100.0% in the first 3 hours, and 8.96% in the subsequent 3 hours after perturbation (145 cells imaged) (i.e. 91% of the cells were perturbed). This outcome suggests that the robustness of LCR is less affected by this perturbation than that of the SCR.

Next, ‘strong’ perturbation of the cells containing the SCR was performed by adding 1 mM of IPTG to the medium, at the end of the third hour of the measurements (112 cells imaged). Here, we observed a functionality of 100.0% in the first 3 hours as earlier, though a lower proportion of cells, 0.89%, were found to be functional in the latter 3 hours (99.1% of the cells were perturbed). From this, it can be concluded, based on predictions from simulations of models of this and similar circuits (Ribeiro et al, 2007), that the robustness of the SCR to external perturbations by IPTG decreases with the increase in the strength of the perturbation.

The agreement between the stochastic models and the empirical data is an indication that such models are capable of predicting the behaviour of the real circuits. As such, it is reasonable to assume that most assumptions made by such models are accurate. Therefore, it will be of interest to test, in the near future, other predictions made by such models on the long-term behaviour of the circuits, among other. Nevertheless, it is worth noted that the ability of the models to predict the behaviour of the circuits is limited. For
example, the models are unable to predict the changes in the circuit’s behaviour when
cells change phase (following the 10 hours under the microscope). This suggests that the
causes for these changes are external to the circuit.
6 Discussion and Conclusion

In this thesis, the factors that affect the dynamics, particularly the phenotypic plasticity and robustness, of a single gene and of a small synthetic genetic circuit were studied using different methods. The study uses the Repressilator as a model circuit (Elowitz et al., 2000). This genetic circuit is of particular interest because it emulates a natural circuit. Moreover, it performs the desired function, i.e., the production of oscillations. This circuit, comprised of known genetic components, has been used in various studies, although, its plasticity and robustness had not been previously studied using empirical methods.

In publication 1, we developed a method to, from the moments when RNA molecules are produced in live cells, infer the number and duration of steps involved in transcription, and also to infer how long it takes for the first RNA to appear in a cell following the appearance of the inducer in the media. Given that transcription initiation is a multi-stepped process characterized by a number of rate-limiting steps, whose duration far exceeds the duration of subsequent steps, promoter dynamics is a critical factor in the dynamics of any genetic circuit.

In this publication, the \(\text{P}_{\text{BAD}}\) and MS2-GFP tagging of RNA were used to obtain data from which one can determine the rate limiting steps in transcription and also the effect of inducer uptake kinetics on transcription initiation dynamics. The study suggests that inducer uptake kinetics plays a major role in the time it takes for the appearance of the first RNA and in the subsequent dynamics of RNA production. The results also suggest
that this waiting time is very noisy, even in a population of genetically identical cells and, as such, it plays a key role in generating cell to cell diversity in RNA numbers following changes in the environment.

Currently, this method is being used to study the dynamics of the constituent promoters that comprise the Repressilator in order to assess the importance of the dynamics of transcription initiation on the dynamics of the Repressilator, per se. Previous studies suggest that the rate of transcription initiation increases with rise in temperature, when temperature approaches 37 °C (Buc et al., 1985; Muthukrishnan et al., 2012). It was also shown that an additional rate-limiting step emerges, for lacUV5 promoter, when temperature falls below 20 °C. This finding is in agreement with the dynamics of the Repressilator with respect to temperature. At temperatures above 30 °C the circuit exhibited faster dynamics, whereas, at temperatures below 30 °C, the observed dynamics was slower (longer periods of oscillation). Future studies on the dynamics of the constituent promoters should give deeper insights on the temperature dependence of the dynamics of the Repressilator.

In publication 2, the effects of temperature on the dynamics of the Repressilator was studied (Chandraseelan et al., 2013). We observed that the repressor cl loses its functionality with increasing temperature. This finding is in conformity with the biology of the lambda phage. For a lambda phage to switch from lysogenic to lytic cycle, effective repression of PBM, which controls the production of the lambda repressor cl, is required. This is achieved by binding of a cl octamer to the operator 3, effectively repressing the PBM and causing a switch to lysogeny (Dodd et al., 2001). Another study indicated that oligomerization of cl is temperature dependent and production of higher order oligomers, up to octamers, occurs when the temperature approaches 37 °C (Burz et al., 1996). Finally, the RNA polymerase is shown to bind preferentially to Pβ, rather than PBM, at 37 °C (Owens, 1983). In our study, we observed that as the temperature shifts from 27 to 37 °C RNA productions by the PBM promoter changes from being a sub-poissonian process to being a super-poissonian process. Further, the reduced time intervals between RNA production events suggests a brief period of increased production, followed by no production at the end of the measurements (data not shown). This has
consequences in the dynamics of the repressilator. At 30 °C, the repressilator’s functionality was unaffected, but with increase in temperature the functionality was eventually lost; as it is evident from the bimodal distribution of the periods and failed oscillations, due to the altered properties that lead to deficient repression by cl and preferable binding of RNA polymerase to P_R (Fig 4, Chandraseelan et al., 2013). Meanwhile, the loss of functionality at lower temperatures are due to slower rate of the reactions (Buc et al., 1985; Muthukrishnan et al., 2012; Stricker et al., 2008). The outcomes of this study suggest that replacing the lambda repressor and the promoter, cl and P_R respectively; by less a temperature-dependent protein and promoter can make this circuit more functional at temperatures above 30 °C.

The study of effect of temperature on the synchrony of oscillations between sister cells, in publication 3, reveals that temperature has a profound influence in enhancing loss of synchrony between sister cells containing the Repressilator. Analysis of whether a robust cell is likely to have a robust sister cell suggested that it is more likely to have a robust sister cell than not, but the chance for a non-robust cell to have a non-robust sister cell is much higher. We also found that sister cells lose correlation as temperature increases. This effect can be explained with the increase in the noise of the period as a function of the temperature. Since non-robust cells are responsible for much of the behaviour variability at 37 °C, elimination of non-robust cells from the analysis restored correlation to a level comparable to that of 30 °C. Finally, we found that increasing temperature leads to pairs of cells with wider range of degrees of correlation.

Copy number is expected to be one of the key factors controlling the expression dynamics of simple genetic networks. Changes in the number of copies of the genes can result in behaviour switching of the network to different equilibrium points (Mileyko et al., 2007). In study 4, we studied the effect of copy number on the dynamics of the Repressilator by comparing the SCR with the LCR. It was observed that the SCR and the LCR are similar in terms of functionality and robustness to internal fluctuations. However, the dynamics of the SCR differed from that of the LCR in that it exhibited stronger fluctuations in the duration of each oscillation and a longer mean period length. It was also observed that, while the response of the two systems to temperature changes is not identical, as the consequences in the dynamics of the repressilator. At 30 °C, the repressilator’s functionality was unaffected, but with increase in temperature the functionality was eventually lost; as it is evident from the bimodal distribution of the periods and failed oscillations, due to the altered properties that lead to deficient repression by cl and preferable binding of RNA polymerase to P_R (Fig 4, Chandraseelan et al., 2013). Meanwhile, the loss of functionality at lower temperatures are due to slower rate of the reactions (Buc et al., 1985; Muthukrishnan et al., 2012; Stricker et al., 2008). The outcomes of this study suggest that replacing the lambda repressor and the promoter, cl and P_R respectively; by less a temperature-dependent protein and promoter can make this circuit more functional at temperatures above 30 °C.

The study of effect of temperature on the synchrony of oscillations between sister cells, in publication 3, reveals that temperature has a profound influence in enhancing loss of synchrony between sister cells containing the Repressilator. Analysis of whether a robust cell is likely to have a robust sister cell suggested that it is more likely to have a robust sister cell than not, but the chance for a non-robust cell to have a non-robust sister cell is much higher. We also found that sister cells lose correlation as temperature increases. This effect can be explained with the increase in the noise of the period as a function of the temperature. Since non-robust cells are responsible for much of the behaviour variability at 37 °C, elimination of non-robust cells from the analysis restored correlation to a level comparable to that of 30 °C. Finally, we found that increasing temperature leads to pairs of cells with wider range of degrees of correlation.

Copy number is expected to be one of the key factors controlling the expression dynamics of simple genetic networks. Changes in the number of copies of the genes can result in behaviour switching of the network to different equilibrium points (Mileyko et al., 2007). In study 4, we studied the effect of copy number on the dynamics of the Repressilator by comparing the SCR with the LCR. It was observed that the SCR and the LCR are similar in terms of functionality and robustness to internal fluctuations. However, the dynamics of the SCR differed from that of the LCR in that it exhibited stronger fluctuations in the duration of each oscillation and a longer mean period length. It was also observed that, while the response of the two systems to temperature changes is not identical, as the
dynamics of the SCR varies more widely, in both the cases, the minimum period length
was observed when setting temperature to 30 °C.

The higher mean period length of the SCR (30 °C or higher) is explained by the fact that,
in general, in cells with LCR, when gene `A` is in the `ON` state, the system will evolve
to gene `B` being in the ON state, following the production of proteins from the several
copies of gene `C`. This behaviour is expected from the models, provided that a protein
produced by a single copy of a gene is sufficient to repress all the promoter copies of the
subsequent gene in the loop (Golding et al., 2005).

Meanwhile, the shorter mean period length of the SCR at 28 °C can be explained as the
opposite effect occurring at low temperatures, as each gene produces less proteins when
`ON`, also the binding rate of the proteins to promoters is reduced, so a single gene being
`ON` is no longer sufficient to turn ‘OFF’ all the subsequent promoters in the loop. This
possibly explains the longer mean period in length in cells at 28°C, so presence of
multiple copies of a gene may be needed to activate the progression to the next stage of the
cycle (Golding et al., 2005). These hypotheses can be tested in the future, by
implementing the Repressilator on, e.g., a medium copy plasmid (MCR). If the
hypotheses are correct, the differences in behaviour between MCR and LCR, at
temperatures between 30 °C to 37 °C, should be similar to the differences between the
SCR and LCR, and it is also expected that there should be no significant differences
between the LCR and MCR at 28 °C in period length.

Interestingly, the period of the SCR, at 30 °C, was observed to be longer and noisier
(higher CV²) than the period of the LCR. This could be due to the fact that in the LCR the
variability in period lengths is reduced by expression of one copy of a gene going `ON`
producing enough repressors to turn ‘OFF’ all the copies of the subsequent gene in the
loop, consistent with the findings, based on simulations of (Mileiko et al., 2007), which
suggest that copy-number acts as a proxy in coupling of the system and, hence, maintain
oscillations effectively. It was also noted that the LCR and SCR robustness to internal
fluctuations do not differ. Similar to the LCR, loss of robustness was observed with
increasing temperature, which has been shown (Chandraseelan et al., 2013) to be due to
the loss of ability of CI to repress beyond 30 °C (Jana et al., 1999; Koblan et al., 1991a;

dynamics of the SCR varies more widely, in both the cases, the minimum period length
was observed when setting temperature to 30 °C.

The higher mean period length of the SCR (30 °C or higher) is explained by the fact that,
in general, in cells with LCR, when gene `A` is in the `ON` state, the system will evolve
to gene `B` being in the ON state, following the production of proteins from the several
copies of gene `C`. This behaviour is expected from the models, provided that a protein
produced by a single copy of a gene is sufficient to repress all the promoter copies of the
subsequent gene in the loop (Golding et al., 2005).

Meanwhile, the shorter mean period length of the SCR at 28 °C can be explained as the
opposite effect occurring at low temperatures, as each gene produces less proteins when
`ON`, also the binding rate of the proteins to promoters is reduced, so a single gene being
`ON` is no longer sufficient to turn ‘OFF’ all the subsequent promoters in the loop. This
possibly explains the longer mean period in length in cells at 28°C, so presence of
multiple copies of a gene may be needed to activate the progression to the next stage of the
cycle (Golding et al., 2005). These hypotheses can be tested in the future, by
implementing the Repressilator on, e.g., a medium copy plasmid (MCR). If the
hypotheses are correct, the differences in behaviour between MCR and LCR, at
temperatures between 30 °C to 37 °C, should be similar to the differences between the
SCR and LCR, and it is also expected that there should be no significant differences
between the LCR and MCR at 28 °C in period length.

Interestingly, the period of the SCR, at 30 °C, was observed to be longer and noisier
(higher CV²) than the period of the LCR. This could be due to the fact that in the LCR the
variability in period lengths is reduced by expression of one copy of a gene going `ON`
producing enough repressors to turn ‘OFF’ all the copies of the subsequent gene in the
loop, consistent with the findings, based on simulations of (Mileiko et al., 2007), which
suggest that copy-number acts as a proxy in coupling of the system and, hence, maintain
oscillations effectively. It was also noted that the LCR and SCR robustness to internal
fluctuations do not differ. Similar to the LCR, loss of robustness was observed with
increasing temperature, which has been shown (Chandraseelan et al., 2013) to be due to
the loss of ability of CI to repress beyond 30 °C (Jana et al., 1999; Koblan et al., 1991a;
Koblan et al., 1991b). This can be the effect of similar fraction of hampered CI proteins with increasing temperature in LCR and SCR cells are similar.

The functionality of both SCR and LCR were found to be maximized at 30 °C, although the effects of changing temperature on the functionality appeared to differ to a greater extent between the two systems. The functionality was higher in the SCR than in the LCR at 28 °C and 30 °C, but lower at 33 °C and 37 °C. The reasons as to why not all Repressilators (LCR and SCR) are functional in a cell are not known and, at this stage, it is not possible to explain the changes in functionality with temperature, however, these differences may offer clues to understand the mechanisms that underlie the functionality of the circuit.

Another interesting observation in the SCR worth discussing is the robustness of the SCR to external perturbations. The robustness of LCR to a constant perturbation (50 μM of IPTG in the medium) was found to be less than the SCR. This is perhaps surprising but can be explained by the dynamic coupling of all the copies of Repressilator in the LCR system; perturbing one of the copies of the repressor (e.g. causing the operator sites of one of the promoters to become vacant, thus allowing its expression) may be sufficient to disrupt the activity of all other repressilators in the cells due to coupling. Since the cells with LCR have more copies of repressors, for a perturbation of identical ‘strength’, it is more likely that at least one copy will be perturbed in cells with LCR than in cells with SCR.

The outcomes of this study suggest significant differences in robustness to internal and external perturbations as well as the differences in the dynamics of the two circuits. Also, it emphasizes the need for a single-copy version of the synthetic Repressilator which is, for example highly robust to external perturbations and highly functional at lower temperatures that will be a new building block in Synthetic Biology to engineer single copy based artificial genetic circuits. Our findings might also be initial clues to explain why the variation in copy-number of genes has been linked to a number of altered phenotypes, including many diseases (Stranger et al., 2007). Further studies on the regulation of the dynamics of the SCR and LCR and its effects on the phenotypic

Koblan et al., 1991b). This can be the effect of similar fraction of hampered CI proteins with increasing temperature in LCR and SCR cells are similar.

The functionality of both SCR and LCR were found to be maximized at 30 °C, although the effects of changing temperature on the functionality appeared to differ to a greater extent between the two systems. The functionality was higher in the SCR than in the LCR at 28 °C and 30 °C, but lower at 33 °C and 37 °C. The reasons as to why not all Repressilators (LCR and SCR) are functional in a cell are not known and, at this stage, it is not possible to explain the changes in functionality with temperature, however, these differences may offer clues to understand the mechanisms that underlie the functionality of the circuit.

Another interesting observation in the SCR worth discussing is the robustness of the SCR to external perturbations. The robustness of LCR to a constant perturbation (50 μM of IPTG in the medium) was found to be less than the SCR. This is perhaps surprising but can be explained by the dynamic coupling of all the copies of Repressilator in the LCR system; perturbing one of the copies of the repressor (e.g. causing the operator sites of one of the promoters to become vacant, thus allowing its expression) may be sufficient to disrupt the activity of all other repressilators in the cells due to coupling. Since the cells with LCR have more copies of repressors, for a perturbation of identical ‘strength’, it is more likely that at least one copy will be perturbed in cells with LCR than in cells with SCR.

The outcomes of this study suggest significant differences in robustness to internal and external perturbations as well as the differences in the dynamics of the two circuits. Also, it emphasizes the need for a single-copy version of the synthetic Repressilator which is, for example highly robust to external perturbations and highly functional at lower temperatures that will be a new building block in Synthetic Biology to engineer single copy based artificial genetic circuits. Our findings might also be initial clues to explain why the variation in copy-number of genes has been linked to a number of altered phenotypes, including many diseases (Stranger et al., 2007). Further studies on the regulation of the dynamics of the SCR and LCR and its effects on the phenotypic
differences in cells will give deeper insights about how cells alter phenotypes with respect to changing copy numbers of genes.

Although a number of synthetic genetic circuits have been developed, development of a genetic circuit with significant complexity that resemble the natural genetic circuits is still a complex task owing to the complexity of the biological systems and the stochasticity of genetic processes. For the development of complex synthetic circuits, it is of importance to characterise the components and their functioning in the system. In this work, we quantified the effect of different factors on the dynamics of the Repressilator. The outcomes of the study reemphasise the need for characterisation of genetic components that can be used to build genetic circuits with significant functions and complexity. It was also observed, in agreement with previous studies, cell to cell diversity in the populations observed which again supports the notion that noise is a key component of genetic circuits. Finally, we observed that copy-number of genes is a significant factor on the dynamics of the circuit.

Overall, from this study, it is evident that the phenotypic plasticity and robustness of simple genetic circuits can be affected by temperature, copy-number, etc. The methods used in this study, can be used in the future to study other genetic circuits, synthetic and natural, which will enable the development of new synthetic genetic circuits capable of performing complex tasks in cells.
References


promoter in *Escherichia coli*. Evidence for a sequential mechanism with two rate-limiting steps. *BMC Systems Biology*, 5, 149.


Thieffry D, H Salgado, A M Huerta, and J Collado-Vides (1998), Prediction of transcriptional regulatory sites in the complete genome sequence of Escherichia coli K-12, Bioinformatics, 14, 391–400.


In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter
Jarno Mäkelä1, Meenakshisundaram Kandhavelu1, Samuel M. D. Oliveira1, Jerome G. Chandraseelan1, Jason Lloyd-Price1, Juha Peltonen1, Olli Yli-Harja1,2 and Andre S. Ribeiro1,3

1Laboratory of Biosystem Dynamics, Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, FI-33101 Tampere, Finland and 2Institute for Systems Biology, 1441 N 34th Street, Seattle, WA 98103-8904, USA

Received March 21, 2013; Revised April 11, 2013; Accepted April 13, 2013

ABSTRACT
Using a single-RNA detection technique in live Escherichia coli cells, we measure, for each cell, the waiting time for the production of the first RNA under the control of P\textsubscript{BAD} promoter after induction by arabinose, and subsequent intervals between transcription events. We find that the kinetics of the arabinose intake system affect mean and diversity in RNA numbers, long after induction. We observed the same effect on P\textsubscript{BAD} promoter, which is inducible by arabinose or by IPTG. Importantly, the distribution of waiting times of P\textsubscript{BAD} promoter is indistinguishable from that of P\textsubscript{BAD} if and only if induced by arabinose alone. Finally, RNA production under the control of P\textsubscript{BAD} is found to be a sub-Poissonian process. We conclude that inducer-dependent waiting times affect mean and cell-to-cell diversity in RNA numbers long after induction, suggesting that inducible mechanisms have non-negligible effects on the phenotypic diversity of cell populations in natural, fluctuating environments.

INTRODUCTION
Transcription in E. coli is, at a genome-wide scale, a relatively rare stochastic event (I–3). Further, many genes only become active in response to external stimuli (4–7), via processes that are also stochastic (7). Although much is known on the noise in gene expression at the single-cell level (1–7,9–10), most of our present knowledge concerns the average behaviour of cell populations alone (11). However, to characterize the dynamics and the underlying steps of intake processes, it is necessary to observe their effects in individual live cells (12). This observation should inform also on the robustness of cellular response mechanisms by informing us on the degree of change in the responses of a single cell to multiple occurrences of the same stimulus, as well as the difference in responses to different stimuli.

One of the most well-known gene activation mechanisms in the arabinose utilization system of E. coli is the arabinose utilization system of E. coli. This system imports arabinose into the cell by AraFGH, an arabinose-specific high-affinity ABC transporter (11,13–15), and by a low-affinity transporter, AraE, which binds to the inner membrane and makes use of electrochemical potential to intake the arabinose (11,16,17). This system exhibits wide variability in the timing of activation and in the rates of accumulation of inducer molecules (18). It has been hypothesized that this is due to the cell-to-cell variability in the numbers of proteins responsible for the intake of arabinose (18). Interestingly, if the gene araE is placed under a constitutive promoter the inducer rates become more homogenous (19–21), suggesting that the diversity in the number of protein intake is a non-negligible source of cell-to-cell variability in the kinetics of the arabinose utilization system (12).

Evidence suggests that when the intracellular concentration of arabinose exceeds a threshold, the dimeric AraC protein activates the genes that code for the proteins responsible for the intake (AraE and AraFGH) and for the catabolism of arabinose (araBAD) (11,22). In the absence of arabinose, AraC binds two half-sites on the DNA (I\textsubscript{1} and O\textsubscript{2}) and promotes the formation of a DNA loop that prevents access of RNA polymerases to the promoters in that region (P\textsubscript{BAD} and P\textsubscript{BAD}). When bound by arabinose, AraC binds instead to the adjacent I\textsubscript{1} and I\textsubscript{2} half-sites. The resulting configuration promotes transcription initiation at P\textsubscript{BAD} (11).

Transcription initiation is a complex, multi-stepped process (23,24). In vitro measurements suggest that this process has at least two to three rate limiting steps (25,26). It starts when the RNA polymerase binds to the DNA and begins transcription from the promoter.

ABSTRACT
Using a single-RNA detection technique in live Escherichia coli cells, we measure, for each cell, the waiting time for the production of the first RNA under the control of P\textsubscript{BAD} promoter after induction by arabinose, and subsequent intervals between transcription events. We find that the kinetics of the arabinose intake system affect mean and diversity in RNA numbers, long after induction. We observed the same effect on P\textsubscript{BAD} promoter, which is inducible by arabinose or by IPTG. Importantly, the distribution of waiting times of P\textsubscript{BAD} promoter is indistinguishable from that of P\textsubscript{BAD} if and only if induced by arabinose alone. Finally, RNA production under the control of P\textsubscript{BAD} is found to be a sub-Poissonian process. We conclude that inducer-dependent waiting times affect mean and cell-to-cell diversity in RNA numbers long after induction, suggesting that inducible mechanisms have non-negligible effects on the phenotypic diversity of cell populations in natural, fluctuating environments.

INTRODUCTION
Transcription in E. coli is, at a genome-wide scale, a relatively rare stochastic event (I–3). Further, many genes only become active in response to external stimuli (4–7), via processes that are also stochastic (7). Although much is known on the noise in gene expression at the single-cell level (1–7,9–10), most of our present knowledge concerns the average behaviour of cell populations alone (11). However, to characterize the dynamics and the underlying steps of intake processes, it is necessary to observe their effects in individual live cells (12). This observation should inform also on the robustness of cellular response mechanisms by informing us on the degree of change in the responses of a single cell to multiple occurrences of the same stimulus, as well as the difference in responses to different stimuli.

One of the most well-known gene activation mechanisms in the arabinose utilization system of E. coli is the arabinose utilization system of E. coli. This system imports arabinose into the cell by AraFGH, an arabinose-specific high-affinity ABC transporter (11,13–15), and by a low-affinity transporter, AraE, which binds to the inner membrane and makes use of electrochemical potential to intake the arabinose (11,16,17). This system exhibits wide variability in the timing of activation and in the rates of accumulation of inducer molecules (18). It has been hypothesized that this is due to the cell-to-cell variability in the numbers of proteins responsible for the intake of arabinose (18). Interestingly, if the gene araE is placed under a constitutive promoter the inducer rates become more homogenous (19–21), suggesting that the diversity in the number of protein intake is a non-negligible source of cell-to-cell variability in the kinetics of the arabinose utilization system (12).

Evidence suggests that when the intracellular concentration of arabinose exceeds a threshold, the dimeric AraC protein activates the genes that code for the proteins responsible for the intake (AraE and AraFGH) and for the catabolism of arabinose (araBAD) (11,22). In the absence of arabinose, AraC binds two half-sites on the DNA (I\textsubscript{1} and O\textsubscript{2}) and promotes the formation of a DNA loop that prevents access of RNA polymerases to the promoters in that region (P\textsubscript{BAD} and P\textsubscript{BAD}). When bound by arabinose, AraC binds instead to the adjacent I\textsubscript{1} and I\textsubscript{2} half-sites. The resulting configuration promotes transcription initiation at P\textsubscript{BAD} (11).

Transcription initiation is a complex, multi-stepped process (23,24). In vitro measurements suggest that this process has at least two to three rate limiting steps (25,26). It starts when the RNA polymerase binds to the DNA and begins transcription from the promoter.
promoter region of the DNA molecule, forming the closed complex, which is followed by the open complex formation and promoter escape (27,28). The RNA polymerase then elonates these RNA molecules. Evidence suggests that, in general, initiation is much longer in duration than elongation (26,29). Recent in vivo measurements of the kinetics of initiation of P_{BAD} and P_{lac}-promoters have shown that RNA production under the control of these promoters is a sub-Poissonian process (8–10). These studies also support the existence of multiple steps at the stage of initiation, significantly limiting the rate of RNA production, as suggested in vitro measurements (30).

Here, we extend these investigations to include the process of intake of arabinose and of the process of transcription under the control of P_{BAD} to the cell-to-cell diversity in RNA production. Namely, we report measurements of the in vivo kinetics of initiation and transcription production of P_{BAD} with single-molecule sensitivity, making use of the MS2d-GFP tagging of RNA in E. coli (31).

For that, in each cell, we measure the waiting time until the first RNA is produced after induction and the subsequent intervals between consecutive transcription productions. For comparison, we conduct the same measurements for P_{lac/[a]racy} which is produced by either of its two inducers, arabinose and IPTG.

**MATERIALS AND METHODS**

**Strains and plasmids**

*Escherichia coli* strain DH5*-* PRO was generously provided by L. Golding, University of Illinois and contains the construct PROTEC-K13, carrying P_{BAD}/mRFP1-MS2-96bs (31), along with a new construct, pMK-BAC (P_{BAD}/rMP1-MS2d-GFP and pIG-BAC (P_{BAD}/ltetO-1-mRFP1-MS2-96bs), which is a single-copy F-based vector carrying a sequence coding for a monomeric red fluorescent protein (mRFP1) followed by a 96 binding site array under the control of P_{BAD} (cloning information provided in Supplementary Methods) (see also Supplementary Figures S1 and S2). The strain with plasmids P_{lac/[a]racy}-MS2d-GFP and pIG-BAC (P_{lac/[a]racy}-mRFP1-MS2d-96bs) (32) was used as well. The DH5*-* PRO strain [identical to Z1 (31)] is a genuine producer of AraC (33). No modifications were made to the chromosomal DNA of this strain or in our experiments.

**Microscopy**

For microscopy measurements, cells were pelleted and resuspended in ~50 μl of fresh M63 medium. Afterwards, few microcells of cells were placed between a 3% agarose gel pad made with medium and a glass coverslip before assembling the imaging chamber (FCS2, Bioptechs). After finishing the experiment, the chamber was heated to 37°C.

Cells were visualized in a Nikon Eclipse (TE2000-U, Nikon) equipped with a BX51WI camera. Emitted light and laser-scanning system using a ×100 Apo TIRF objective.

A flow of fresh, pre-warmed M63 medium contains the inducer arabinose at a concentration of 1 mM. Images were taken once per minute for 2 h, and the laser shutter was open only during the exposure time to monitor the in situ fluorescence. Emission of ingrown RNA was initialized at the same time as the collection of the time series. For image acquisition, we used Nikon EZ-C1 software. GFP fluorescence was measured using a 488 nm argon ion laser (Melles-Griot), 515/30 nm emission filter and a pixel dwell time of 3.56 μs (total image acquisition time of 3.5 s).

An interacting multiple model filter-based autofocus strategy (34) was used to correct focus drift in time series acquisitions. The method estimates the focal drift using an interacting multiple model filter algorithm to predict the focal drift at time t based on the measurement at t-1. It allows reducing the number of required images at different z-planes for drift correction, thus minimizing photobleaching.

**Data and image analysis**

Data and images were analysed using custom software written in MATLAB 2011b (MathWorks). Cells were detected from fluorescence images by a semi-automatic method described previously (8). In time series, the area occupied by each cell was manually masked. Principal component analysis was used to obtain the dimensions and orientation of the cells within each mask. Fluorescent spots in the cells were automatically segmented using density estimation with a Gaussian kernel (35) and Otus’s thresholding (36). Finally, background-corrected spot intensities were calculated and summed to produce a spot intensity time series for each cell.

Moments of appearance of novel target RNA molecules in each cell were calculated from fluorescence images by fitting the corrected total spots intensity over time in each cell to a monotone polynomial function by least squares (37). An example of this procedure is shown in Figure 1D. For more details on the image analysis see (8). Note that, in cells that do not contain target RNA molecules at the start of the measurements, the number of total RNA molecules detected since the start of the measurements until a given moment equals the total number of RNA molecules in the cell at that moment.

Because some cells already contained target RNA molecules at the start of the measurements, the number of novel RNA numbers within cells at a given moment in time is obtained using a different method. Specifically, when promoter region of the DNA molecule, forming the closed complex, which is followed by the open complex formation and promoter escape (27,28). The RNA polymerase then elonates these RNA molecules. Evidence suggests that, in general, initiation is much longer in duration than elongation (26,29). Recent in vivo measurements of the kinetics of initiation of P_{BAD} and P_{lac}-promoters have shown that RNA production under the control of these promoters is a sub-Poissonian process (8–10). These studies also support the existence of multiple steps at the stage of initiation, significantly limiting the rate of RNA production, as suggested in vitro measurements (30). Here, we extend these investigations to include the process of intake of arabinose and of the process of transcription under the control of P_{BAD} to the cell-to-cell diversity in RNA production. Namely, we report measurements of the in vivo kinetics of initiation and transcription production of P_{BAD} with single-molecule sensitivity, making use of the MS2d-GFP tagging of RNA in E. coli (31). For that, in each cell, we measure the waiting time until the first RNA is produced after induction and the subsequent intervals between consecutive transcription productions. For comparison, we conduct the same measurements for P_{lac/[a]racy} when induced by either of its two inducers, arabinose and IPTG.

**MATERIALS AND METHODS**

**Strains and plasmids**

*Escherichia coli* strain DH5*-* PRO was generously provided by L. Golding, University of Illinois and contains the construct PROTEC-K13, carrying P_{BAD}/mRFP1-MS2d-GFP (31), along with a new construct, pMK-BAC (P_{BAD}/rMP1-MS2d-GFP and pIG-BAC (P_{BAD}/ltetO-1-mRFP1-MS2-96bs), which is a single-copy F-based vector carrying a sequence coding for a monomeric red fluorescent protein (mRFP1) followed by a 96 binding site array under the control of P_{BAD} (cloning information provided in Supplementary Methods) (see also Supplementary Figures S1 and S2). The strain with plasmids P_{lac/[a]racy}-MS2d-GFP and pIG-BAC (P_{lac/[a]racy}-mRFP1-MS2d-96bs) (32) was used as well. The DH5*-* PRO strain [identical to Z1 (31)] is a genuine producer of AraC (33). No modifications were made to the chromosomal DNA of this strain or in our experiments.

**Microscopy**

For microscopy measurements, cells were pelleted and resuspended in ~50 μl of fresh M63 medium. Afterwards, few microcells of cells were placed between a 3% agarose gel pad made with medium and a glass coverslip before assembling the imaging chamber (FCS2, Bioptechs). After finishing the experiment, the chamber was heated to 37°C.

Cells were visualized in a Nikon Eclipse (TE2000-U, Nikon) equipped with a BX51WI camera. Emitted light and laser-scanning system using a ×100 Apo TIRF objective.

A flow of fresh, pre-warmed M63 medium contains the inducer arabinose at a concentration of 1 mM. Images were taken once per minute for 2 h, and the laser shutter was open only during the exposure time to monitor the in situ fluorescence. Emission of ingrown RNA was initialized at the same time as the collection of the time series. For image acquisition, we used Nikon EZ-C1 software. GFP fluorescence was measured using a 488 nm argon ion laser (Melles-Griot), 515/30 nm emission filter and a pixel dwell time of 3.56 μs (total image acquisition time of 3.5 s).

An interacting multiple model filter-based autofocus strategy (34) was used to correct focus drift in time series acquisitions. The method estimates the focal drift using an interacting multiple model filter algorithm to predict the focal drift at time t based on the measurement at t-1. It allows reducing the number of required images at different z-planes for drift correction, thus minimizing photobleaching.
comparing measurements using MS2d-GFP tagging and using plate reader (Supplementary Figure S4), the total number of MS2d-GFP–tagged RNA molecules was extracted from the total spot intensity distribution, obtained from all cells in an image obtained at a given moment after induction. For this, the first peak of the obtained distribution set to correspond to the intensity of a single-RNA molecule. The number of tagged RNAs in each spot can be estimated by dividing its intensity by that of the first peak (32).

RESULTS

Experimental design

To study the kinetics of expression of $P_{BAD}$, we detect individual RNA molecules, as these are produced in live cells and register when these events occur. For this, we placed the $P_{BAD}$ promoter on a single-copy F-plasmid, followed by a coding region for mRFP1 and an array of 96 binding sites for MS2d–GFP–tagging proteins (32) (Figure 1A). The expression of MS2d–GFP is controlled by $P_{BAD}$, which is activated before the gene of interest, $L_{tetO}$ (33) and is inducible by anhydrotetracycline (aTc). The target RNA contains an mRFP1 coding region, followed by an array of 96 MS2d-binding sites. Expression of the target RNA is controlled by $P_{BAD}$ whose activity is regulated by AraC and the inducer aTc. The target construct is on a single-copy F-plasmid. The tagging construct is on a medium-copy vector. (A) Example of E. coli cells expressing MS2d–GFP and target RNA. GFP–tagged RNA molecules are marked by circles. (B) Time course of total intensity of spots in a cell (circles) and monotonie piecewise-constant fit (line).

To study the kinetics of expression of $P_{BAD}$, we detect individual RNA molecules, as these are produced in live cells and register when these events occur. For this, we placed the $P_{BAD}$ promoter on a single-copy F-plasmid, followed by a coding region for mRFP1 and an array of 96 binding sites for MS2d–GFP–tagging proteins (32) (Figure 1A). The expression of MS2d–GFP is controlled by $P_{BAD}$, which is activated before the gene of interest, $L_{tetO}$ (33) and is inducible by anhydrotetracycline (aTc). The target RNA contains an mRFP1 coding region, followed by an array of 96 MS2d-binding sites. Expression of the target RNA is controlled by $P_{BAD}$ whose activity is regulated by AraC and the inducer aTc. The target construct is on a single-copy F-plasmid. The tagging construct is on a medium-copy vector. (A) Example of E. coli cells expressing MS2d–GFP and target RNA. GFP–tagged RNA molecules are marked by circles. (B) Time course of total intensity of spots in a cell (circles) and monotonie piecewise-constant fit (line).

Figure 1. Measurement system. (A) Components of the detection system. The expression of the tagging protein, MS2d–GFP, is controlled by $P_{BAD}$ (33) and is inducible by anhydrotetracycline (aTc). The target RNA contains an mRFP1 coding region, followed by an array of 96 MS2d-binding sites. Expression of the target RNA is controlled by $P_{BAD}$ whose activity is regulated by AraC and the inducer aTc. The target construct is on a single-copy F-plasmid. The tagging construct is on a medium-copy vector. (B) Figureative description of the waiting time for the first RNA production ($t_0$) and intervals between subsequent productions ($\Delta t$). Images are taken once per minute for 2h. (C) Example of E. coli cells expressing MS2d–GFP and target RNA. GFP–tagged RNA molecules are marked by circles. (D) Time course of total intensity of spots in a cell (circles) and monotonie piecewise-constant fit (line).

RESULTS

Experimental design

To study the kinetics of expression of $P_{BAD}$, we detect individual RNA molecules, as these are produced in live cells and register when these events occur. For this, we placed the $P_{BAD}$ promoter on a single-copy F-plasmid, followed by a coding region for mRFP1 and an array of 96 binding sites for MS2d–GFP–tagging proteins (32) (Figure 1A). The expression of MS2d–GFP is controlled by $P_{BAD}$, which is activated before the gene of interest, $L_{tetO}$ (33) and is inducible by anhydrotetracycline (aTc). The target RNA contains an mRFP1 coding region, followed by an array of 96 MS2d-binding sites. Expression of the target RNA is controlled by $P_{BAD}$ whose activity is regulated by AraC and the inducer aTc. The target construct is on a single-copy F-plasmid. The tagging construct is on a medium-copy vector. (B) Figureative description of the waiting time for the first RNA production ($t_0$) and intervals between subsequent productions ($\Delta t$). Images are taken once per minute for 2h. (C) Example of E. coli cells expressing MS2d–GFP and target RNA. GFP–tagged RNA molecules are marked by circles. (D) Time course of total intensity of spots in a cell (circles) and monotonie piecewise-constant fit (line).

Figure 1. Measurement system. (A) Components of the detection system. The expression of the tagging protein, MS2d–GFP, is controlled by $P_{BAD}$ (33) and is inducible by anhydrotetracycline (aTc). The target RNA contains an mRFP1 coding region, followed by an array of 96 MS2d-binding sites. Expression of the target RNA is controlled by $P_{BAD}$ whose activity is regulated by AraC and the inducer aTc. The target construct is on a single-copy F-plasmid. The tagging construct is on a medium-copy vector. (B) Figureative description of the waiting time for the first RNA production ($t_0$) and intervals between subsequent productions ($\Delta t$). Images are taken once per minute for 2h. (C) Example of E. coli cells expressing MS2d–GFP and target RNA. GFP–tagged RNA molecules are marked by circles. (D) Time course of total intensity of spots in a cell (circles) and monotonie piecewise-constant fit (line).
gene, we observe similar trend in activity over time in the native promoter and in the one on the F-plasmid (Supplementary Figure S3).

We next compared expression levels of the target gene, when assessed by independent methods, for two induction levels, namely, 0.1 and 1% L-arabinose (Supplementary Methods). In Figure 2A and B, we show the temporal variation after induction in mean numbers of MS2d-GFP-tagged RNA molecules in cell populations and in the fluorescence intensity of RFP measured by plate reader, respectively.

The plate reader measurements of mRFP1 levels, 2h after induction, show a fold change of 1.67 times when L-arabinose is increased from 0.1 to 1%. The MS2d-GFP in vitro detection method shows a fold change of 1.74 between these same conditions, showing that the results from the two methods are in accordance. From this and the previous experiment, we also conclude that the MS2-GFP method accurately detects RNA production of the target gene, and that the target gene behaves similarly to the natural system.

We also assessed for what range of inducer concentrations is the target gene under full induction. We measured with the plate reader its expression for varying inducer concentration after induction. From Supplementary Figure S5, maximum induction is achieved for 1% arabinose concentration. Here onwards, unless stated otherwise, we use this concentration to assess kinetics of RNA production under the control of PBAD.

First RNA and intervals between consecutive RNA molecules in individual cells. From the time-lapse images acquired with confocal microscopy, after induction, we measure in each cell both $t_0$ and subsequent values of $\Delta t$. $t_0$ is expected to include the time for arabinose to enter the cell via the intake mechanism, the time to find the promoter and release the repressor and also the time for the recruitment of the RNA polymerase and subsequent production of the first target RNA. The latter process includes events such as the closed complex formation. This study thus predicts a faster mean induction time than what our direct measurements (n=1000–3000) show. Two main factors affect this difference. First, in the mutant used previously (12), the chromosomal araBAD operon is deleted, avoiding the negative feedback mechanism, which likely fastens the response time significantly. Additionally, gene expression was assessed from a medium-copy vector, which should respond much faster than the single-copy vector system used here, as its response time depends on the fastest of the response times of several promoter copies. Thus, we find that the results reported previously (12) and ours are in agreement. For example, while observing mean waiting times one order of magnitude longer, we do observe RNA molecules appearing in some of the cells within a time scale of 200–400 s after induction. When provided with a copy number vector instead of the single-copy vector used here, we expect mean waiting times one order of magnitude smaller and thus in agreement with the measurements described previously (12).

Correlations between consecutive processes

To study whether the durations of the processes of intake and of transcription initiation are correlated, we first assessed whether consecutive intervals of $\Delta t$ in individual cells are correlated. Figure 3A shows a distribution of $\Delta t$ for 101 pairs of consecutive intervals, and found it to be 0.16. We obtained a $p$-value of 0.11, assuming no correlation model, which implies that we cannot reject the null hypothesis of similarity. We conclude that the time of intake of inducers affects significantly both mean and standard deviation of RNA numbers in individual cells, locally. Finally, note that the difference between the distributions of $t_0$ and $\Delta t$ provides evidence that the activity of PBAD changes significantly at high inducer concentrations. Otherwise, these two distributions should not differ significantly, as they would both result, e.g. from spurious transcription events alone.

One recent study (12) also focuses on the in vitro induc- tion kinetics of PBAD. This study uses measurements of GFP levels in cell populations, whose expression is controlled by PBAD (inserted into a medium-copy vector) and a model to extrapolate the mean activation time of the promoter, after induction. Assuming a threshold for GFP levels to consider the promoter as active, the mean activation time of GFP after induction was ~960 s. By considering several features of the measurement system, including the mean maturation time of GFP, a value was then extrapolated for the expected activation time of the promoter, namely, ~250 s. This does not include the time for transcription to be completed, once the closed complex is formed. This study thus predicts a faster mean induction time than what our direct measurements indicate (~300–1000 s). Two main factors are responsible for this difference. First, in the mutant used previously (12), the chromosomal araBAD operon is deleted, avoiding the negative feedback mechanism, which likely fastens the response time significantly. Additionally, gene expression was assessed from a medium-copy vector, which should respond much faster than the single-copy vector system used here, as its response time depends on the fastest of the response times of several promoter copies. Thus, we find that the results reported previously (12) and ours are in agreement. For example, while observing mean waiting times one order of magnitude longer, we do observe RNA molecules appearing in some of the cells within a time scale of 200-400 s after induction. Therefore, provided the usage of a multi-copy vector instead of the single-copy vector used here, we expect mean waiting times one order of magnitude smaller and this in agreement with the measurements described previously (12).

Correlations between consecutive processes

To study whether the durations of the processes of intake and of transcription initiation are correlated, we first assessed whether consecutive intervals of $\Delta t$ in individual cells are correlated. Figure 3A shows a distribution of $\Delta t$ for 101 pairs of consecutive intervals, and found it to be 0.16. We obtained a $p$-value of 0.11, assuming no correlation model, which implies that we cannot reject the null hypothesis of similarity. We conclude that the time of intake of inducers affects significantly both mean and standard deviation of RNA numbers in individual cells, locally. Finally, note that the difference between the distributions of $t_0$ and $\Delta t$ provides evidence that the activity of PBAD changes significantly at high inducer concentrations. Otherwise, these two distributions should not differ significantly, as they would both result, e.g. from spurious transcription events alone.

One recent study (12) also focuses on the in vitro induc- tion kinetics of PBAD. This study uses measurements of GFP levels in cell populations, whose expression is controlled by PBAD (inserted into a medium-copy vector) and a model to extrapolate the mean activation time of the promoter, after induction. Assuming a threshold for GFP levels to consider the promoter as active, the mean activation time of GFP after induction was ~960 s. By considering several features of the measurement system, including the mean maturation time of GFP, a value was then extrapolated for the expected activation time of the promoter, namely, ~250 s. This does not include the time for transcription to be completed, once the closed complex is formed. This study thus predicts a faster mean induction time than what our direct measurements indicate (~300–1000 s). Two main factors are responsible for this difference. First, in the mutant used previously (12), the chromosomal araBAD operon is deleted, avoiding the negative feedback mechanism, which likely fastens the response time significantly. Additionally, gene expression was assessed from a medium-copy vector, which should respond much faster than the single-copy vector system used here, as its response time depends on the fastest of the response times of several promoter copies. Thus, we find that the results reported previously (12) and ours are in agreement. For example, while observing mean waiting times one order of magnitude longer, we do observe RNA molecules appearing in some of the cells within a time scale of 200-400 s after induction. Therefore, provided the usage of a multi-copy vector instead of the single-copy vector used here, we expect mean waiting times one order of magnitude smaller and this in agreement with the measurements described previously (12).

Correlations between consecutive processes

To study whether the durations of the processes of intake and of transcription initiation are correlated, we first assessed whether consecutive intervals of $\Delta t$ in individual cells are correlated. Figure 3A shows a distribution of $\Delta t$ for 101 pairs of consecutive intervals, and found it to be 0.16. We obtained a $p$-value of 0.11, assuming no correlation model, which implies that we cannot reject the null hypothesis of similarity. We conclude that the time of intake of inducers affects significantly both mean and standard deviation of RNA numbers in individual cells, locally. Finally, note that the difference between the distributions of $t_0$ and $\Delta t$ provides evidence that the activity of PBAD changes significantly at high inducer concentrations. Otherwise, these two distributions should not differ significantly, as they would both result, e.g. from spurious transcription events alone.

One recent study (12) also focuses on the in vitro induc- tion kinetics of PBAD. This study uses measurements of GFP levels in cell populations, whose expression is controlled by PBAD (inserted into a medium-copy vector) and a model to extrapolate the mean activation time of the promoter, after induction. Assuming a threshold for GFP levels to consider the promoter as active, the mean activation time of GFP after induction was ~960 s. By considering several features of the measurement system, including the mean maturation time of GFP, a value was then extrapolated for the expected activation time of the promoter, namely, ~250 s. This does not include the time for transcription to be completed, once the closed complex is formed. This study thus predicts a faster mean induction time than what our direct measurements indicate (~300–1000 s). Two main factors are responsible for this difference. First, in the mutant used previously (12), the chromosomal araBAD operon is deleted, avoiding the negative feedback mechanism, which likely fastens the response time significantly. Additionally, gene expression was assessed from a medium-copy vector, which should respond much faster than the single-copy vector system used here, as its response time depends on the fastest of the response times of several promoter copies. Thus, we find that the results reported previously (12) and ours are in agreement. For example, while observing mean waiting times one order of magnitude longer, we do observe RNA molecules appearing in some of the cells within a time scale of 200-400 s after induction. Therefore, provided the usage of a multi-copy vector instead of the single-copy vector used here, we expect mean waiting times one order of magnitude smaller and this in agreement with the measurements described previously (12).
indicate inexistence of correlation between durations of consecutive intervals between RNA productions (8). We next assessed whether the distributions of \( t_0 \) and values of \( \Delta t \) (Figure 2A and B) are correlated. Note that \( t_0 \) and the \( \Delta t \) are of similar order of magnitude as the measurement period. This introduces artificial correlations between \( t_0 \) and \( \Delta t \) of individual cells, as, e.g. a cell with a large \( t_0 \) is expected to exhibit smaller than average \( \Delta t \) values, as larger intervals would not be detected during the measurement period as likely as in cells with smaller values of \( t_0 \). To remove these artificial correlations between \( t_0 \) and \( \Delta t \) of individual cells, in this assessment, we only considered RNA productions for a certain window size (Supplementary Methods and Supplementary Figure S6). This window is set so as to maximize the number of data points that can be extracted from the measurements.

From the windowed data, we calculated the Pearson correlation between \( t_0 \) and \( \Delta t \) values in individual cells to be \(-0.15\). We calculated a \( P \)-value of 0.18 assuming no correlation as the null hypothesis, which implies that we cannot prove that the correlation is significant. This result is in line with (12), which reports a lack of correlation between initiation of protein expression and subsequent rate of protein synthesis in individual cells.

**Dynamics of induction and of transcription initiation under different induction schemes**

The distinctiveness of the distributions of \( t_0 \) and \( \Delta t \) of \( P_{BAD} \), as assessed by the \( K \)-S test, suggests that they are, respectively, the partial result of different processes. Although \( t_0 \) ought to depend on the kinetics of intake of arabinose and on the transcription initiation event, \( \Delta t \) values ought to depend mostly on the kinetics of transcription initiation events alone. These assumptions arise from the following. First, in *in vitro* and *in vivo* measurements (26,38) suggest that transcription initiation (including closed and open complex formation) is a long-duration, multi-step process, usually taking \( 10^4\)–\( 10^5 \) s in bacterial promoters (10,25,26,37,38). Other events that need to occur before the appearance of a target RNA because of the tagging of the MS2d-GFP are not expected to affect \( \Delta t \) significantly. These are transcription elongation and the tagging by multiple MS2d-GFP. Elongation of the target RNA was measured to take only tens of seconds (31). Also, the tagging occurs at a rate that makes the RNA visible during elongation or shortly after (31).

To test the two assumptions, we measured the distributions of \( t_0 \) and \( \Delta t \) for another promoter, \( P_{BAD} \), in two conditions. \( P_{BAD} \) can be induced either by IPTG or by arabinose (as \( P_{BAD} \), or by both inducers simultaneously (9). According to our assumption, the distribution of \( t_0 \) of \( P_{BAD} \) is expected to be similar to that of \( P_{BAD} \) when the latter is induced by arabinose, because of depending on the same intake mechanism, whereas it should differ significantly when \( P_{BAD} \) is induced by IPTG, given the different intake mechanisms of IPTG.

We measured the distributions of \( t_0 \) and \( \Delta t \) for \( P_{BAD} \) when induced by IPTG alone and when induced by arabinose alone (Table 1). We used the same concentration of arabinose as when inducing \( P_{BAD} \). The IPTG concentration used is the one required for maximum induction of \( P_{BAD} \) (33). Results in Table 1 follow the windowing procedure described earlier in the text. The table shows mean, standard deviation and square of the coefficient of variation (\( \mu^2/\sigma^2 \)) of \( t_0 \) and \( \Delta t \) for the two promoters, each of which in two induction schemes.

We first assessed the distinctiveness of the distributions of \( t_0 \) and \( \Delta t \) by the \( K \)-S test, for each promoter in each condition (Table 2). In all cases, these two distributions differ in a statistical sense. This is in agreement with the assumption that although both \( \Delta t \) and \( t_0 \) depend on the kinetics of initiation at the promoter, only \( t_0 \) depends on the kinetics of intake of inducers.

We next performed statistical tests to assess the distinctiveness between the induction kinetics \( (t_0) \) of the two promoters (Table 3), when subject to the same inducer and when subject to different inducers. Also, we compared the effects of a different inducer concentration in the case of \( P_{BAD} \). From Table 3, when \( P_{BAD} \) and \( P_{BAD} \) are induced with 1% arabinose, they exhibit distributions of \( t_0 \) that cannot be distinguished. However, when \( P_{BAD} \) is induced in the presence of arabinose, we cannot prove that the correlation is significant. This result is in line with (12), which reports a lack of correlation between initiation of protein expression and subsequent rate of protein synthesis in individual cells.

**Dynamics of induction and of transcription initiation under different induction schemes**

The distinctiveness of the distributions of \( t_0 \) and \( \Delta t \) of \( P_{BAD} \), as assessed by the \( K \)-S test, suggests that they are, respectively, the result of different processes. Although \( t_0 \) ought to depend on the kinetics of intake of arabinose and on the first transcription initiation event, \( \Delta t \) values ought to depend mostly on the kinetics of transcription initiation events alone. These assumptions arise from the following. First, in *in vitro* and *in vivo* measurements (26,38) suggest that transcription initiation (including closed and open complex formation) is a long-duration, multi-step process, usually taking \( 10^4\)–\( 10^5 \) s in bacterial promoters (10,25,26,37,38). Other events that need to occur before the appearance of a target RNA because of the tagging of the MS2d-GFP are not expected to affect \( \Delta t \) significantly. These are transcription elongation and the tagging by multiple MS2d-GFP. Elongation of the target RNA was measured to take only tens of seconds (31). Also, the tagging occurs at a rate that makes the RNA visible during elongation or shortly after (31).

To test the two assumptions, we measured the distributions of \( t_0 \) and \( \Delta t \) for another promoter, \( P_{BAD} \), in two conditions. \( P_{BAD} \) can be induced either by IPTG or by arabinose (as \( P_{BAD} \), or by both inducers simultaneously (9). According to our assumption, the distribution of \( t_0 \) of \( P_{BAD} \) is expected to be similar to that of \( P_{BAD} \) when the latter is induced by arabinose, because of depending on the same intake mechanism, whereas it should differ significantly when \( P_{BAD} \) is induced by IPTG, given the different intake mechanisms of IPTG.

We measured the distributions of \( t_0 \) and \( \Delta t \) for \( P_{BAD} \) when induced by IPTG alone and when induced by arabinose alone (Table 1). We used the same concentration of arabinose as when inducing \( P_{BAD} \). The IPTG concentration used is the one required for maximum induction of \( P_{BAD} \) (33). Results in Table 1 follow the windowing procedure described earlier in the text. The table shows mean, standard deviation and square of the coefficient of variation (\( \mu^2/\sigma^2 \)) of \( t_0 \) and \( \Delta t \) for the two promoters, each of which in two induction schemes.

We first assessed the distinctiveness of the distributions of \( t_0 \) and \( \Delta t \) by the \( K \)-S test, for each promoter in each condition (Table 2). In all cases, these two distributions differ in a statistical sense. This is in agreement with the assumption that although both \( \Delta t \) and \( t_0 \) depend on the kinetics of initiation at the promoter, only \( t_0 \) depends on the kinetics of intake of inducers.

We next performed statistical tests to assess the distinctiveness between the induction kinetics \( (t_0) \) of the two promoters (Table 3), when subject to the same inducer and when subject to different inducers. Also, we compared the effects of a different inducer concentration in the case of \( P_{BAD} \). From Table 3, when \( P_{BAD} \) and \( P_{BAD} \) are induced with 1% arabinose, they exhibit distributions of \( t_0 \) that cannot be distinguished. However, when \( P_{BAD} \) is...
induced with IPTG, the resulting t0 distribution is statistically distinguishable from that of PBAD, when induced ... be rejected.

Table 1. Measurements of t0 and Δt

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Indicator</th>
<th>No. of samples (Δt)</th>
<th>μ0 (s)</th>
<th>σ0 (s)</th>
<th>n02</th>
<th>μ (s)</th>
<th>σ (s)</th>
<th>n2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBAD</td>
<td>1% arab</td>
<td>102</td>
<td>1448.6</td>
<td>522.8</td>
<td>0.14</td>
<td>84</td>
<td>2858.0</td>
<td>1159.8</td>
</tr>
<tr>
<td>PBAD</td>
<td>0.1% arab</td>
<td>78</td>
<td>1474.5</td>
<td>481.2</td>
<td>0.11</td>
<td>70</td>
<td>3519.4</td>
<td>1236.2</td>
</tr>
<tr>
<td>PBAD</td>
<td>1% arabinose</td>
<td>149</td>
<td>1556.5</td>
<td>516.0</td>
<td>0.11</td>
<td>125</td>
<td>2832.5</td>
<td>1236.2</td>
</tr>
<tr>
<td>PBAD</td>
<td>1% arabinose</td>
<td>1100</td>
<td>1100</td>
<td>1100</td>
<td>1100</td>
<td>1100</td>
<td>1100</td>
<td>1100</td>
</tr>
<tr>
<td>PBAD</td>
<td>1mM IPTG</td>
<td>485</td>
<td>1314.4</td>
<td>570.0</td>
<td>0.19</td>
<td>286</td>
<td>2670.7</td>
<td>913.6</td>
</tr>
</tbody>
</table>

The table shows the mean (μ), the standard deviation (σ) and the normalized variance (n(μ)/σ) of the measured distributions of t0 and Δt.

Table 2. P-values of the Kolmogorov-Smirnov test between t0 and Δt distributions for each promoter and induction condition

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Indicator</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBAD</td>
<td>1% arab</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBAD</td>
<td>0.1% arab</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBAD</td>
<td>1% arabinose</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBAD</td>
<td>1% arabinose</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBAD</td>
<td>1mM IPTG</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

For P > 0.05, it is generally accepted that the hypothesis that the two distributions are the same should be rejected.

Table 3. P-values of the Kolmogorov-Smirnov test between t0 distributions for each promoter and induction condition

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Indicator</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBBAD</td>
<td>1% arab</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBBAD</td>
<td>0.1% arab</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBBAD</td>
<td>1% arabinose</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBBAD</td>
<td>1% arabinose</td>
<td>n.s.</td>
</tr>
<tr>
<td>PBAD</td>
<td>1mM IPTG</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

For P > 0.05, it is generally accepted that the hypothesis that the two distributions are the same should be rejected.

induced with IPTG, the resulting t0 distribution is statistically distinguishable from that of PBAD, when induced by either 0.1% or 1% arabinoose. This is also distinct from its own Δt distribution when induced by 1% arabinoose. This statistically significant difference supports the hypothesis that the distributions of t0 are dependent on the kinetics of the intake system of the inducers, and that these differ for IPTG and arabinose.

Finally, we observed that the distributions of t0 of PBAD, when induced by 0.1% and 1% arabinoose, are distinct. This is expected as the time for inducers to ‘first reach the promoter’ ought to depend on the inducer’s concentration.

Kinetics of the intake process

The intake time of an inducer, here named ‘tdiff’, differs from t0 in that it does not include the time for the first transcription initiation event to occur. Because of this, tdiff cannot be measured directly with the MS2-GFP-tagging method. We thus estimate the mean and variance of the distribution of tdiff by subtracting the mean and variances of the Δt distribution from the t0 distribution. This method is based on the fact that we were unable to establish the existence of a correlation between the values of t0 and Δt. Given this, and as tdiff is weakly correlated (Pearson correlation of ~0.15), we assume that they are independent so as to be able to estimate the standard deviation of the duration of the intake process alone (note that the mean of this quantity can be estimated as described later in the text, regardless of the existence of dependence).

The estimated mean and a standard deviation of tdiff are similar for PBAD and for PBBAD, when induced with 1% arabinoose. Namely, in both cases, we obtained a mean of ~1400 s and a standard deviation of ~1100 s. This is expected, given the usage of the same intake mechanism and inducer concentration. Importantly, when PBBAD is induced by IPTG, the standard deviation of tdiff is much smaller (~700 s), whereas the mean is similar to when induced by arabinose (~1400 s). This suggests that the intake of arabinose is a noisier process (concerning the uncertainty of the intake time) than the intake of IPTG. Finally, we find that as the decrease of tdiff, the concentration of arabinose affects the mean of tdiff significantly, as it equals ~2000 s for 0.1% arabinoose.

Effect of the intake process on the temporal cell-to-cell diversity in RNA numbers

Because of being stochastic and thus variable in duration from one event to the next (i.e. it differs from one cell to the next), the intake process impacts on the diversity in RNA numbers of a cell population. This impact should decrease with time, after induction. We estimated the time during which the effect is tangible for each measurement condition. For this, we assume that values of tdiff depend mostly on the intake of arabinose and on the first transcription initiation event at the start site of PBAD. Meanwhile, the distribution of intervals between consecutive RNAs is assumed to depend solely on the kinetics of transcription initiation (8,10,37).

The events determining Δt as well as t0 are modelled as d-step processes, each step with an exponentially distributed duration (Supplementary Methods) (37). From this assumption, it is possible, for a given number of steps, to find the duration of each step that best fits the measurements. We assume transcription initiation to be a three-step process, namely, the open complex formation and promoter escape (27,38), as evidence suggests that these are the most rate-limiting steps in the distribution of values of tdiff, i.e. the ones most contributing to the intervals between production of consecutive RNA molecules (26). This assumption also relies on recent induced with IPTG, the resulting t0 distribution is statistically distinguishable from that of PBBAD, when induced by either 0.1% or 1% arabinose. This is also distinct from its own Δt distribution when induced by 1% arabinose. This statistically significant difference supports the hypothesis that the distributions of t0 are dependent on the kinetics of the intake system of the inducers, and that these differ for IPTG and arabinose.

Finally, we observed that the distributions of t0 of PBBAD, when induced by 0.1% and 1% arabinose, are distinct. This is expected as the time for inducers to ‘first reach the promoter’ ought to depend on the inducer’s concentration.

Kinetics of the intake process

The intake time of an inducer, here named ‘tdiff’, differs from t0 in that it does not include the time for the first transcription initiation event to occur. Because of this, tdiff cannot be measured directly with the MS2-GFP-tagging method. We thus estimate the mean and variance of the distribution of tdiff by subtracting the mean and variances of the Δt distribution from the t0 distribution. This method is based on the fact that we were unable to establish the existence of a correlation between the values of t0 and Δt. Given this, and as tdiff is weakly correlated (Pearson correlation of ~0.15), we assume that they are independent so as to be able to estimate the standard deviation of the duration of the intake process alone (note that the mean of this quantity can be estimated as described later in the text, regardless of the existence of dependence).

The estimated mean and a standard deviation of tdiff are similar for PBBAD and for PBBAD, when induced with 1% arabinose. Namely, in both cases, we obtained a mean of ~1400 s and a standard deviation of ~1100 s. This is expected, given the usage of the same intake mechanism and inducer concentration. Importantly, when PBBAD is induced by IPTG, the standard deviation of tdiff is much smaller (~700 s), whereas the mean is similar to when induced by arabinose (~1400 s). This suggests that the intake of arabinose is a noisier process (concerning the uncertainty of the intake time) than the intake of IPTG. Finally, we find that as the decrease of tdiff, the concentration of arabinose affects the mean of tdiff significantly, as it equals ~2000 s for 0.1% arabinose.

Effect of the intake process on the temporal cell-to-cell diversity in RNA numbers

Because of being stochastic and thus variable in duration from one event to the next (i.e. it differs from one cell to the next), the intake process impacts on the diversity in RNA numbers of a cell population. This impact should decrease with time, after induction. We estimated the time during which the effect is tangible for each measurement condition. For this, we assume that values of tdiff depend mostly on the intake of arabinose and on the first transcription initiation event at the start site of PBBAD. Meanwhile, the distribution of intervals between consecutive RNAs is assumed to depend solely on the kinetics of transcription initiation (8,10,37).

The events determining Δt as well as t0 are modelled as d-step processes, each step with an exponentially distributed duration (Supplementary Methods) (37). From this assumption, it is possible, for a given number of steps, to find the duration of each step that best fits the measurements. We assume transcription initiation to be a three-step process, namely, the open complex formation, the open complex formation and promoter escape (27,38), as evidence suggests that these are the most rate-limiting steps in the distribution of values of tdiff, i.e. the ones most contributing to the intervals between production of consecutive RNA molecules (26). This assumption also relies on recent
studies (37) that indicate that assuming this number of steps suffices to generate distributions that cannot be discriminated, in a statistical sense, from measurements with accuracy and quantity of data similar to the measurements reported here. Finally, we assume the intake to be a two-step process, namely, the binding of extracellular arabinose to an uptake protein and, once bound, its translocation to the cytoplasm (12). The combination of the two processes (intake followed by transcription initiation) is, consequently, assumed to be a five-step process. Assuming these numbers of steps and stable conditions (e.g. induction level), we searched for models that fit the distributions accurately enough so that the K-S test does not find differences between model predictions and measurements. The P-values of these tests are shown in Supplementary Table S1 and show that in all but one case, it is possible to find a model that can be distinguished from the empirical distribution, in a statistical sense.

The case for which we could not find a model that fits the measurements is that of P_{BAD} with 0.1% arabinose induction. This may be due to lack of sufficient data or because the model is unsuitable. Future studies are required to assert this. One explanation may be that, in this case, the distribution of intake times results from two distinct kinetics, one being the productions under induction and the other being spurious productions by promoters in the ‘non-induced’ state.

Given the models aforementioned and provided a rate of RNA degradation, it is possible to estimate the time it takes for the mean RNA numbers of a model cell population to reach equilibrium, as this time depends solely on the rate of degradation of RNAs and t₀. We do not have measurements of the degradation rate of the target RNA, as the tagging with MSD2-GFP ‘immortalizes’ it for the duration of the measurements (32). Instead, the models in Figure 3 assume an RNA degradation rate of 5 min⁻¹, which is within realistic intervals for E. coli (1).

From all of the aforementioned data, we estimated the mean times for RNA numbers to reach near-equilibrium, as well as the Fano factor of this quantity since the start of the simulations. Results are shown in Figure 3, as estimated for each of the models. Also shown is an estimation that assumes the model of transcription initiation of P_{BAD} when induced by 1% arabinose, coupled with an infinitely fast intake.

In all cases, reaching equilibrium in mean RNA numbers takes >1 h, except when assuming infinitely fast intake, in which case the time to reach equilibrium is ~0.5 h. Thus, for a time length as long as 1–2 h, the intake process has a non-negligible contribution on the mean and on the cell-to-cell diversity in RNA numbers of the cell populations. From Figure 3A, one also observes different shapes in the curves of P_{BAD} when induced by IPTG (dashed line) and when induced by arabinose (dotted line), because of different intake kinetics.

From Figure 3B, the contribution of the intake kinetics on the cell-to-cell variability in RNA numbers is also significant. For example, the kinetics of intake causes an increase in the Fano factor in the initial moments not observable in the case of infinitely fast intake. We also tested models of P_{BAD} induced by 1% arabinose (normal and infinitely fast intake) with other RNA degradation rates (Supplementary Figure S7), within realistic intervals (1). Aside from assessing the degree of dependency on the intake time and degradation rate, one also observes from the figure that although the latter determines the rate at which the system reaches equilibrium, the former acts as a delay towards reaching the numbers at equilibrium. Further, one can see that the intake step adds diversity to the RNA numbers in the cells, during the transient to reach equilibrium.

DISCUSSION

We measured, at the single-cell level, how long it takes for the first RNA under the control of P_{BAD} to be produced, followed the introduction of the inducer in the media. Also, we measured the subsequent intervals between studies (37) that indicate that assuming this number of steps suffices to generate distributions that cannot be discriminated, in a statistical sense, from measurements with accuracy and quantity of data similar to the measurements reported here. Finally, we assume the intake to be a two-step process, namely, the binding of extracellular arabinose to an uptake protein and, once bound, its translocation to the cytoplasm (12). The combination of the two processes (intake followed by transcription initiation) is, consequently, assumed to be a five-step process. Assuming these numbers of steps and stable conditions (e.g. induction level), we searched for models that fit the distributions accurately enough so that the K-S test does not find differences between model predictions and measurements. The P-values of these tests are shown in Supplementary Table S1 and show that in all but one case, it is possible to find a model that can be distinguished from the empirical distribution, in a statistical sense.

In all cases, reaching equilibrium in mean RNA numbers takes >1 h, except when assuming infinitely fast intake, in which case the time to reach equilibrium is ~0.5 h. Thus, for a time length as long as 1–2 h, the intake process has a non-negligible contribution on the mean and on the cell-to-cell diversity in RNA numbers of the cell populations. From Figure 3A, one also observes different shapes in the curves of P_{BAD} when induced by IPTG (dashed line) and when induced by arabinose (dotted line), because of different intake kinetics.

From Figure 3B, the contribution of the intake kinetics on the cell-to-cell variability in RNA numbers is also significant. For example, the kinetics of intake causes an increase in the Fano factor in the initial moments not observable in the case of infinitely fast intake. We also tested models of P_{BAD} induced by 1% arabinose (normal and infinitely fast intake) with other RNA degradation rates (Supplementary Figure S7), within realistic intervals (1). Aside from assessing the degree of dependency on the intake time and degradation rate, one also observes from the figure that although the latter determines the rate at which the system reaches equilibrium, the former acts as a delay towards reaching the numbers at equilibrium. Further, one can see that the intake step adds diversity to the RNA numbers in the cells, during the transient to reach equilibrium.
consecutive RNA productions. From the intervals between transcription events, we determined that RNA production under the control of $P_{BAD}$ is a sub- Poissonian process. Two recent studies reached similar conclusions for $P_{lac}$ and $P_{ara}$, for all induction conditions tested (9,10). We hypothesize that this may be a common phenomenon because of the kinetic properties of the process of transcription initiation in bacteria, in particular, because of its multi-stepped nature.

From the different mean times, it takes for the appearance of the first RNA in each cell when under the control of $P_{BAD}$ and of $P_{ara}$, for different induction conditions, we assessed the effect of the kinetics of the intake process on the mean and cell-to-cell diversity in RNA numbers of cell populations. Relevantly, this effect was found to be tangible for a long period after induction. Also, we verified that different intake mechanisms differ significantly not only in mean but also in the degree of variability of the intake time, and that this has a non-negligible effect on RNA population statistics.

Given the aforementioned data, and considering that natural environments are fluctuating, we expect the kinetics of cellular intake mechanisms to have a significant effect on the degree of phenotypic diversity of cell populations. Finally, we expect the methodology used here to assess the in vivo kinetics of intake of arabinose and of IPTG to be applicable to any gene of interest. Such studies should provide valuable insight into the adaptability of prokaryotic organisms to environmental changes and stress. They should also provide a better understanding of the observed cell-to-cell phenotypic diversity in $E. coli$ when in fluctuating environments.

**SUPPLEMENTARY DATA**

Data are available at NAR Online: Supplementary Table 1, Supplementary Figures 1–7, Supplementary Methods and Supplementary References [39–43].

**FUNDING**

The Academy of Finland [257603 to A.S.R.]; the Finnish Funding Agency for Technology and Innovation [40226/12 to O.-Y.H.]; the Foundation for Science and Technology, Portugal [PTDC/BBD-MET/1084/2012 to A.S.R.]; funding for open access charge: Academy of Finland [257605, 2012 to A.S.R.].

Conflict of interest statement. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**REFERENCES**


UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry, 24, 2712–2723.


**Supplementary Information**

**In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter**

Jarno Mäkelä, Meenakshisundaram Kandhavelu, Samuel M.D. Oliveira, Jerome G. Chandraseelan, Jason Lloyd-Price, Juha Peltonen, Olli Yli-Harja, and Andre S. Ribeiro

1 Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland.

2 Institute for Systems Biology, 1441N 34th St, Seattle, WA, 98103-8904, USA

* Corresponding author: andre.ribeiro@tut.fi (Andre S. Ribeiro)

**Supplementary Methods**

**Chemicals**

Bacterial cell cultures were grown in two media, namely Luria-Bertani (LB) and M63. The chemical components of LB broth (Tryptone, Yeast extract and NaCl) were purchased from LabM (UK). For M63 media, the following components were used: 2 mM MgSO4, 7H2O (Sigma-Aldrich, USA), 7.6 mM (NH4)2SO4 (Sigma Life Science, USA), 30 mM FeSO4, 7H2O (Sigma Life Science, USA), 1 mM EDTA (Sigma Life Science, USA), 60 mM KH2PO4 (Sigma Life Science, USA), Glycerol 0.5 % (Sigma Life Science, USA), and Casaminoacids 0.1 % (Fluka Analytical, USA). Isopropyl β-D-thiogalactopyranoside (IPTG), l-(-)-Arabinose and anhydrotetracycline (aTc) used for induction of the cells and the antibiotics (100 mg/ml kanamycin and 35 mg/ml chloramphenicol) were purchased from Sigma-Aldrich (USA). Agarose (Sigma Life Science, USA) was used for the microscopic slide gel preparation.

**Bacterial Strain**

Cloning and expression experiments were performed in *E. coli* DH5α-PRO strain (Clontech; identical to DH5α-Z1 (31). The strain information is: deoR, endA1, gyrA96, hsdR17(rk,mk), recA1, relA1, supE44, thi-1, Δ(lacZYA-argF)U169, F' Δ(lacZΔM15) F-, λ-, P<sub>N<sub>lac</sub></sub> tetR, P<sub>lac</sub>/lacI, and Sp<sup>R</sup>. Frag1A: F-, rha-, thi, gal, lacZ<sup>am</sup>, ΔacrAB::kan<sup>R</sup>, P<sub>N<sub>lac</sub></sub>/tetR, P<sub>lac</sub>/lacI, and Sp<sup>R</sup>. Frag1B: F-, rha-, thi, gal, lacZ<sup>am</sup>, P<sub>N<sub>lac</sub></sub>/tetR, P<sub>lac</sub>/lacI, and Sp<sup>R</sup>. The P<sub>N<sub>lac</sub></sub>/tetR, P<sub>lac</sub>/lacI, Sp<sup>R</sup> cassette was transferred from DH5α<sub>PRO</sub> to Frag1 to generate Frag1B by P1 transduction. The ΔacrAB::kan<sup>R</sup> cassette was transferred from KZM120 to Frag1B, so as to generate Frag1A.
**Construction of the pMK-BAC vector**

To construct the pMK-BAC (PBAD-mRFP1-96 binding site (96 BS) array, the following plasmids were used: a plasmid with mRFP1 plus 96bs array region in the BAC vector, originally designed and generously provided by Prof. Ido Golding (P_{lacz}^+−mrFP1-96 bs) (32). To amplify the construct containing the AraC and pBAD promoter region from the pGLO vector (Biorad), a primer set was designed as follows:

Ara_AatII-Fw-5’CTTAAGACGTACATCGATGATAATAA7’
Ara_AatII-Rv-5’CCCTGATGACGTCATGTATATCTTCTTAAAGTTA3’

The target BAD promoter region along with AraC coding region from the pGLO vector was amplified and inserted into the pIG-BAC vector by standard molecular biology techniques. The construct was verified by sequencing with the appropriate primers and transformed into the *E. coli* DH5α-PRO strain carrying the bacterial expression vector pPROTET.E (Clontech) coding for MS2d-GFP. For more details see Supplementary Figures 1 and 2.

**Plate reader experiment**

The mean fluorescence of RFP under the control of P_{BAD} was measured with a microplate fluorometer (Fluoroscan Ascent, Thermo Scientific). 200 ml of cells at OD600 = 0.5 were induced with 0.1 % or 1 % L-arabinose and placed on 96 well microplate. From this, cells were measured for 2 hours for relative fluorescence levels of mRFP1 protein (excitation and emission wavelengths were 584 nm and 607 nm, respectively). The cell density was kept identical in all wells of the plate for all conditions.

**Quantitative PCR for mean mRNA quantification**

The change in the rate of transcription of genes araB and mRFP was studied using qPCR. *E. coli* DH5α-PRO cells containing the constructs were grown as described in the section describing the microcopy measurements. Cells were grown overnight at 30°C with aeration, diluted into fresh medium and allowed to grow at the appropriate temperature of the experiment until an optical density of OD600 = 0.3-0.5 was reached. For the experiment, 5 ml of cells were pre-incubated with 100 ng/ml of aTc to induce the expression of MS2d-GFP. 1 % L-arabinose was used for induction of the BAD promoter, 30 minutes after induction, the first sample was taken. From then onwards, samples were taken at an interval of 60 minutes. Rifampicin was added to the samples immediately so as to prevent further transcription and the cells were fixed with RNA protect reagent immediately followed by enzymatic lysis using Tris-EDTA lysozyme buffer (pH 8.3). RNA was purified from each sample by RNeasy mini-kit (Qiagen). The total RNA was separated by electrophoresis through 1 % agarose gel and stained with SYBR Safe DNA Gel Stain. The RNA was found intact with discreet bands for 16 S and 23 S ribosomal RNAs. To ensure purity of the RNA samples, they were subject to treatment with DNase free of RNase, to remove residual DNA. The yield of RNA obtained was 0.4 – 0.6 mg/ml. Approximately 40 ng of RNA was used for cDNA synthesis using iSCRIPT reverse transcription super mix (Biorad) according to the manufacturer’s instructions.

**Construction of the pMK-BAC vector**

To construct the pMK-BAC (PBAD-mRFP1-96 binding site (96 BS) array, the following plasmids were used: a plasmid with mRFP1 plus 96bs array region in the BAC vector, originally designed and generously provided by Prof. Ido Golding (P_{lacz}^+−mrFP1-96 bs) (32). To amplify the construct containing the AraC and pBAD promoter region from the pGLO vector (Biorad), a primer set was designed as follows:

Ara_AatII-Fw-5’CTTAAGACGTACATCGATGATAATAA7’
Ara_AatII-Rv-5’CCCTGATGACGTCATGTATATCTTCTTAAAGTTA3’

The target BAD promoter region along with AraC coding region from the pGLO vector was amplified and inserted into the pIG-BAC vector by standard molecular biology techniques. The construct was verified by sequencing with the appropriate primers and transformed into the *E. coli* DH5α-PRO strain carrying the bacterial expression vector pPROTET.E (Clontech) coding for MS2d-GFP. For more details see Supplementary Figures 1 and 2.

**Plate reader experiment**

The mean fluorescence of RFP under the control of P_{BAD} was measured with a microplate fluorometer (Fluoroscan Ascent, Thermo Scientific). 200 ml of cells at OD600 = 0.5 were induced with 0.1 % or 1 % L-arabinose and placed on 96 well microplate. From this, cells were measured for 2 hours for relative fluorescence levels of mRFP1 protein (excitation and emission wavelengths were 584 nm and 607 nm, respectively). The cell density was kept identical in all wells of the plate for all conditions.

**Quantitative PCR for mean mRNA quantification**

The change in the rate of transcription of genes araB and mRFP was studied using qPCR. *E. coli* DH5α-PRO cells containing the constructs were grown as described in the section describing the microcopy measurements. Cells were grown overnight at 30°C with aeration, diluted into fresh medium and allowed to grow at the appropriate temperature of the experiment until an optical density of OD600 = 0.3-0.5 was reached. For the experiment, 5 ml of cells were pre-incubated with 100 ng/ml of aTc to induce the expression of MS2d-GFP. 1 % L-arabinose was used for induction of the BAD promoter, 30 minutes after induction, the first sample was taken. From then onwards, samples were taken at an interval of 60 minutes. Rifampicin was added to the samples immediately so as to prevent further transcription and the cells were fixed with RNA protect reagent immediately followed by enzymatic lysis using Tris-EDTA lysozyme buffer (pH 8.3). RNA was purified from each sample by RNeasy mini-kit (Qiagen). The total RNA was separated by electrophoresis through a 1 % agarose gel and stained with SYBR Safe DNA Gel Stain. The RNA was found intact with discreet bands for 16 S and 23 S ribosomal RNAs. To ensure purity of the RNA samples, they were subject to treatment with DNase free of RNase, to remove residual DNA. The yield of RNA obtained was 0.4 – 0.6 mg/ml. Approximately 40 ng of RNA was used for cDNA synthesis using iSCRIPT reverse transcription super mix (Biorad) according to the manufacturer’s instructions.
Quantification of cDNA was performed by real-time PCR using SYBR-green supermix with primers for the amplification of target and reference genes at a concentration of 200nM. Primers specific to AraB (Forward: 5’ GGTACTTCCACCTGCGACAT 3’, Reverse: 5’ CAACCTGACGCGCAAAATACCT 3’) and mRFP genes (Forward: 5’ TACGAC GCCGAGGTCAAG 3’ and Reverse: 5’ TTGTGGAGGTGATGTCCA 3’) were designed using PRIMER3 (39), the length of the amplicon for the target and reference were maintained at 90bp. The sequence of the primers for the reference gene 16S rRNA (EcoCyc Accession Number: EG30090) (Forward: 5’ CGTCAGCTGTTTGAAA 3’ and Reverse: 5’ GGACCCTGGCAACAAAG 3’) and the primers were obtained from Thermo Scientific. The level of 16s rRNA was used to normalize the expression data of each target gene. 10 ng of cDNA was used as a template. The cycling protocol used was 94 °C for 15 s, 51 °C for 30 s, and 72 °C for 30 s, up to 39 cycles. The amplification was monitored in real time by measuring the fluorescence intensities at the end of each cycle. The experiment was performed in triplicates along with the No-RT and no template controls. The volume used for each reaction was 25 µl in low-profile tube strips in a MiniOpticon Real time PCR system (Biorad). The Cq values were obtained from the CFX ManagerTM Software and the fold change of expression of the target gene was analysed by normalizing against the reference gene according to the Livak method (40). See Supplementary Figure 3 for the results.

Normalization between samples of the distributions of time intervals

The observation time for the production of RNAs is two hours. In some cells, the intervals between transcription events (Δt) are of this order of magnitude. This causes shorter intervals to be ‘favoured’. This is more likely to occur in cells where the waiting time for the first RNA to be produced (t0) is longer, since the remaining observation time is shorter. This introduces an artificial anti-correlation between t0 and Δt in individual cells. Similar correlations are introduced by different division times as well, i.e., shorter division times hamper the collection of longer Δt samples.

Thus, prior to determine if any real correlation exists between t0 and Δt in individual cells, it is necessary to remove these artificial sources of anti-correlation due to the limits in the measurement period. For this, in all cells, all intervals between consecutive RNAs were collected only for a time window of size t, after the previous production. The value of t is identical in all cells. This causes the probability of appearance of the next RNA molecule during that period to be uniform for all cells, if the underlying process is in fact identical in all cells.

This restriction in the collection of values of Δt is made when assessing correlations between t0 and Δt and when comparing these two distributions between conditions. When imposing the restriction, we thus consider only cells that produce at least 2 RNA molecules during their life time and measurement period. The value of t was selected so as to maximize the number of data points collectable from the data sets. Here, t was set to 39 minutes (see Supplementary Figure S6).
Fitting the empirical distributions to a sum of $d$-exponential variates

The arabinose intake mechanism can be described by a single Michaelis-Menten function (41). Since the backward reaction of the intake process is slower than the forward reaction (12), the intake process is modeled, roughly, by a sequence of non-reversible reactions. Interestingly, we found from the measurements and the inference procedure, evidence of two steps at this stage (exponential in duration), which is in agreement with the number of forward steps assumed in other studies for this process (12). Finally, transcription initiation, which follows the intake process, can also be modeled by a 3-step exponential model according recent in vivo measurements (9, 10). Thus, we fit the measured distributions of $t_i$ to a 5-step exponential model.

To fit the empirical distribution with a sum of $d$-exponential variates (of possibly unequal rates), we select the exponential rate parameters $\lambda_1, \ldots, \lambda_d$ such that the Kolmogorov-Smirnov (K-S) statistic is minimized. That is, parameters are selected as $\hat{\Theta} = \arg\max_{\Theta} \sup_{x} |F_\Theta(x) - G(x)|$, where $F_\Theta(x)$ is the cumulative distribution function (CDF) of a sum of $d$ exponentials with parameters $\Theta = (\lambda_1, \ldots, \lambda_d)$, and $G(x)$ is the CDF of the empirical distribution.

$$F_{0-t_1, \ldots, t_d}(X) := \sum_{i=1}^d \left((1 - e^{-\lambda_i X}) \prod_{j=1}^d \frac{I_j}{1 - I_j} \right)$$

The parameter values $\Theta$ are found using a nonlinear numerical optimizer. This method is convenient, since if the K-S test was rejected for the parameters $\Theta$, such a test would also be rejected for any other set of parameters $\Theta$ in this family of fitted distributions, indicating that these distributions are inappropriate models of the data. The results of the fitting are shown in Table S1.

As a final note, the model assumed above can be considered as the simplest possible, i.e., each step is an elementary reaction of the form $A \rightarrow B$, with a constant probability of occurring per unit time. This entails that the distributions of intervals between steps are exponential (42). Notably, the inferred distributions and the experimental data are statistically indistinguishable by the K-S test, which implies that there is no evidence to assume that the model is wrong (see Table S1).

CME solution

To estimate the effect of the intake on the cell-to-cell diversity in RNA numbers we made use of direct integration of the Chemical Master Equation (CME) of the model described in the previous section, using the Finite State Projection algorithm (43). This method truncates the infinite state space of the CME such that the amount of probability outside the truncated region is negligible. In all cases, we truncated the state space at 20 RNA molecules. This number sufficed for this space to contain virtually all of the total probability in the system. The probability mass vector at each time moment is then solved by numerically integrating the truncated CME. From this distribution over time, we calculate mean, variance, and Fano factor of RNA molecules of a model at each moment.

Fitting the empirical distributions to a sum of $d$-exponential variates

The arabinose intake mechanism can be described by a single Michaelis-Menten function (41). Since the backward reaction of the intake process is slower than the forward reaction (12), the intake process is modeled, roughly, by a sequence of non-reversible reactions. Interestingly, we found from the measurements and the inference procedure, evidence of two steps at this stage (exponential in duration), which is in agreement with the number of forward steps assumed in other studies for this process (12). Finally, transcription initiation, which follows the intake process, can also be modeled by a 3-step exponential model according recent in vivo measurements (9, 10). Thus, we fit the measured distributions of $t_i$ to a 5-step exponential model.

To fit the empirical distribution with a sum of $d$-exponential variates (of possibly unequal rates), we select the exponential rate parameters $\lambda_1, \ldots, \lambda_d$ such that the Kolmogorov-Smirnov (K-S) statistic is minimized. That is, parameters are selected as $\hat{\Theta} = \arg\max_{\Theta} \sup_{x} |F_\Theta(x) - G(x)|$, where $F_\Theta(x)$ is the cumulative distribution function (CDF) of a sum of $d$ exponentials with parameters $\Theta = (\lambda_1, \ldots, \lambda_d)$, and $G(x)$ is the CDF of the empirical distribution.

$$F_{0-t_1, \ldots, t_d}(X) := \sum_{i=1}^d \left((1 - e^{-\lambda_i X}) \prod_{j=1}^d \frac{I_j}{1 - I_j} \right)$$

The parameter values $\Theta$ are found using a nonlinear numerical optimizer. This method is convenient, since if the K-S test was rejected for the parameters $\Theta$, such a test would also be rejected for any other set of parameters $\Theta$ in this family of fitted distributions, indicating that these distributions are inappropriate models of the data. The results of the fitting are shown in Table S1.

As a final note, the model assumed above can be considered as the simplest possible, i.e., each step is an elementary reaction of the form $A \rightarrow B$, with a constant probability of occurring per unit time. This entails that the distributions of intervals between steps are exponential (42). Notably, the inferred distributions and the experimental data are statistically indistinguishable by the K-S test, which implies that there is no evidence to assume that the model is wrong (see Table S1).

CME solution

To estimate the effect of the intake on the cell-to-cell diversity in RNA numbers we made use of direct integration of the Chemical Master Equation (CME) of the model described in the previous section, using the Finite State Projection algorithm (43). This method truncates the infinite state space of the CME such that the amount of probability outside the truncated region is negligible. In all cases, we truncated the state space at 20 RNA molecules. This number sufficed for this space to contain virtually all of the total probability in the system. The probability mass vector at each time moment is then solved by numerically integrating the truncated CME. From this distribution over time, we calculate mean, variance, and Fano factor of RNA molecules of a model at each moment.
Figure S1. Plasmids used for the pMK-BAC construction. The pMK-BAC(P_BAD-mRFP1-96bs) plasmid was engineered by linking the amplified region, containing the P_BAD promoter and the araC gene, obtained from pGLO, to the pIG-BAC expression vector, without the lac/ara-1 promoter, obtained from pIG-BAC(P_lac/ara-1-mRFP1-96 bs)-V.

Figure S2. Split gels of the plasmid construction. (A) The PCR fragment of 1347bp amplified from pGLO with the appropriate primers. (B) Lanes containing pIG-BAC-V without the P_lac/ara-1 promoter region (10849bp), and the pIG-BAC-V expression plasmid (11502bp). (C) The pMK-BAC plasmid (12196bp) containing the araC-P_BAD amplified fragment inserted to the BAC expression vector, and the pIG-BAC-V (11502bp). Note the black arrows indicating the bands.
Figure S3. Q-PCR of the native and of the target gene. Q-PCR of RNA expression of the native, integrated AraB gene and of the mRFP1 probe in the F-plasmid, as a function of time, when subject to induction by 1% L-arabinose in liquid culture. The standard deviation bars are from three independent experiments.
Figure S4. MS2-GFP measurement of RNA numbers compared with Plate reader results. (A) RNA numbers over time measured in vivo with the MS2-GFP method for 0.1 % and 1 % L-arabinose. Mean and standard deviation of RNA numbers in individual cells were calculated for each sample separately. Error bars show the standard error of the mean from independent measurements (3 measurements) (B) Fluorescent intensity of RFP over time for 0.1 % and 1 % L-arabinose as measured by Plate reader. Error bars show the standard error of the mean obtained from 8 wells.

Figure S5. Gene expression as measured by Plate Reader. Comparison of different inducer concentrations by plate reader measurements, 2 hours following induction. Maximum induction is achieved with 1 % L-arabinose. Error bars show the standard error of the mean obtained from 8 wells.
Figure S6. Normalization of the data. The values of $t_0$ and the corresponding values of the first $\Delta t$ in each cell. The diagonal line is the total observation time (120 min). Vertical and horizontal ($t_c = 39$ min) lines define the intervals that meet the requirements for un-biasedness.
Figure S7. Models with different degradation rates. The degradation rate was set to the following values: 0.1 min\(^{-1}\), 0.111 min\(^{-1}\), 0.125 min\(^{-1}\), 0.143 min\(^{-1}\), 0.167 min\(^{-1}\), 0.2 min\(^{-1}\), 0.25 min\(^{-1}\), 0.333 min\(^{-1}\), 0.5 min\(^{-1}\). In the figures, only the highest and the lowest values are marked. Mean RNA numbers shown for (A) P_{BAD} with 1% Arabinose and for (C) P_{BAD} with 1% Arabinose and infinitely fast intake. Fano factors of RNA numbers are shown for (B) P_{BAD} with 1% Arabinose and, (D) P_{BAD} with 1% Arabinose and infinitely fast intake.
### Supplementary Table

<table>
<thead>
<tr>
<th></th>
<th>p-value for $t_0$</th>
<th>p-value for $\Delta t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{BAD}$ 1% arabinose</td>
<td>0.2613</td>
<td>0.8930</td>
</tr>
<tr>
<td>$P_{BAD}$ 0.1% arabinose</td>
<td>0.0020</td>
<td>0.5728</td>
</tr>
<tr>
<td>$P_{lac/ara-1}$ 1% arabinose</td>
<td>0.1759</td>
<td>0.3826</td>
</tr>
<tr>
<td>$P_{lac/ara-1}$ 1 mM IPTG</td>
<td>0.1155</td>
<td>0.2413</td>
</tr>
</tbody>
</table>

**Table S1.** Results of the K-S fitting. Asymptotic p-values of the Kolmogorov-Smirnov goodness-of-fit test when fitting the empirical distribution with a sum of 5-exponential variates in the case of $t_0$ and of 3-exponential variates in the case of $\Delta t$. We compare these p-values with a standard value of 0.05.

Available online on the Royal Society of Chemistry’s website, DOI: 10.1039/C3MB70203K
Introduction

Natural genetic circuits can efficiently perform various tasks, including gene expression. Evidence suggests that natural, time-keeping circuits, such as circadian oscillators, evolved to maintain robustness to environmental changes. This is necessary for them to be able to regulate complex cellular processes under various conditions or to efficiently determine cell's response to environmental shifts and signals. Much effort has been made to reproduce their behaviour in synthetic circuits. Once proven reliable, these synthetic circuits should have a wide range of applications. For example, synthetic genetic clocks are desired in synthetic circuits designed for time keeping. Temperature affects the dynamics of most cellular processes, including gene expression. Evidence suggests that natural, time-keeping circuits, such as circadian oscillators, evolved to maintain robustness to temperature fluctuations. Similarly, it is maximized at ~30 °C and is gradually lost as temperature increases, becoming ~50% weaker at 42 °C. This decrease may arise from the fact that the ability of CI to discriminate between operator sites depends on ion binding/release reactions and/or from the temperature-dependence of the CI's dimerization process.

Here, we investigate how temperature affects the dynamics of the repressilator. Afterwards, we search for causes. Motivated by previous evidence that CI’s functionality is temperature-dependent, we also study the temperature-dependence of another circuit, the CI-Cro switch. After comparing the effects of temperature on the kinetics of the two circuits, we propose of a green fluorescent protein (GFP) reporter that is under the control of a promoter that is also present in the 3-gene circuit. Temperature affects the dynamics of most cellular processes, including gene expression. Evidence suggests that natural, time-keeping circuits, such as circadian oscillators, evolved to maintain robustness to temperature fluctuations. Similarly, it is maximized at ~30 °C and is gradually lost as temperature increases, becoming ~50% weaker at 42 °C. This decrease may arise from the fact that the ability of CI to discriminate between operator sites depends on ion binding/release reactions and/or from the temperature-dependence of the CI’s dimerization process.

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

Methods
Repressor: strain, plasmid, and microscopy

Cells of E. coli lac strain MC 4100 with the repressor (pZS1-THL) and the repressilator (pZS1-THL-PRM) were generously provided by I. Golding (University of Illinois, USA). The target plasmid is a single-copy F plasmid with a genetic switch coding for C2H5, under the control of P2319, and C2010 under the control of P2010. Further, the plasmid contains the immunity region of wild-type λ, where the res and reB genes were replaced by a 48 binding site array for MS2 proteins, so as to detect individual RNA whose production is controlled by PMS2. Depending on the occupation of the sites OR1, OR2 and OR3, one of the two promoters will be in a repressed state.30,31 Note that OR1 is absent in the repressilator. Nevertheless, the existence of oscillations shows that the C2 still achieves repression of P5.

Cells were grown in Luria-Bertani (LB) medium with the following components: 10 g l⁻¹ of tryptone (Scharlach, USA), 5 g l⁻¹ of yeast extract (LabM, UK) and 10 g l⁻¹ of NaCl (LabM, UK), with addition of 34 mg ml⁻¹ of Kanamycin and 34 mg ml⁻¹ of Chloramphenicol (both antibiotics from Sigma Aldrich, USA). Cells were grown overnight with shaking at 260 rpm, in an orbital shaker (Labnet), at 30 °C for 12-16 h to an optical density (OD) of 0.1 at 600 nm. Thereafter, cells were transferred to a microscope chamber, for image acquisition.

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

Cells were kept at 24 °C, 27 °C, 30 °C, 33 °C or 37 °C during microscopic observations in a thermal chamber (Heploris, FCNL, Pennsylvannia, USA). We poured 100 μl of melted agarose-medium with 1% agarose (Sigma Life Science, USA), LB medium, and aTc (10 ng ml⁻¹), into a microscope slide with a glass coverslip on top. For time-lapse microscopy, the temperature of the samples was kept stable by a control chamber (Bioptechs, FCS2, Pennsylvania, USA). Time lapse microscopy, the temperature of the samples was kept stable by a control chamber (Bioptechs, FCS2, Pennsylvania, USA). Images were obtained every 15 minutes for 10 hours by a Nikon Eclipse (TE2000-U, Nikon, Tokyo, Japan) inverted C1 confocal laser-scanning system with a 100X oil immersion objective. Images were taken for every 2 minute for 2 hours. GFP fluorescence was measured using 488 nm argon ion laser (Melles-Griot) and a 515/30 nm emission filter. Images were acquired with Nikon EZ-C1 software and were analysed by custom software written in MATLAB 2011b (MathWorks).

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

Paper Molecular BioSystems

Assessing functionality and apparent period of oscillations

To determine if a repressilator is 'functional' during a time series, we use the criterion used in Elowitz et al. A fast Fourier transform is applied to the temporal fluorescence signal from the cell to determine the period of oscillations.32,33 Note that OR1 is absent in the repressilator. Nevertheless, the existence of oscillations shows that C2 still achieves repression of P5.

Cells were grown in Luria-Bertani (LB) medium with the following components: 10 g l⁻¹ of tryptone (Scharlach, USA), 5 g l⁻¹ of yeast extract (LabM, UK) and 10 g l⁻¹ of NaCl (LabM, UK), with addition of 34 mg ml⁻¹ of Kanamycin and 34 mg ml⁻¹ of Chloramphenicol (both antibiotics from Sigma Aldrich, USA). Cells were grown overnight with shaking at 260 rpm, in an orbital shaker (Labnet), at 30 °C for 12-16 h to an optical density (OD) of 0.1 at 600 nm. Thereafter, cells were transferred to a microscope chamber, for image acquisition.

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

Cells were kept at 24 °C, 27 °C, 30 °C, 33 °C or 37 °C during microscopic observations in a thermal chamber (Heploris, FCNL, Pennsylvania, USA). We poured 100 μl of melted agarose-medium with 1% agarose (Sigma Life Science, USA), LB medium, and aTc (10 ng ml⁻¹), into a microscope slide with a glass coverslip on top. For time-lapse microscopy, the temperature of the samples was kept stable by a control chamber (Bioptechs, FCS2, Pennsylvania, USA). Time lapse microscopy, the temperature of the samples was kept stable by a control chamber (Bioptechs, FCS2, Pennsylvania, USA). Images were obtained every 15 minutes for 10 hours by a Nikon Eclipse (TE2000-U, Nikon, Tokyo, Japan) inverted C1 confocal laser-scanning system with a 100X oil immersion objective. Images were taken for every 2 minute for 2 hours. GFP fluorescence was measured using 488 nm argon ion laser (Melles-Griot) and a 515/30 nm emission filter. Images were acquired with Nikon EZ-C1 software and were analysed by custom software written in MATLAB 2011b (MathWorks).

Cell exhibiting oscillatory fluorescence. 5 frames are shown, along with time stamps in minutes. In this case, the images were taken at 30 °C.
For cells considered functional, we estimated the ‘apparent period’ as follows. First, we fit a quadratic curve in the least-squares sense to the intensity time series, to estimate the general trend (Fig. 3, top panel) since the measured intensity is affected by, e.g., photo-bleaching. After subtracting the estimated trend, the residual is scaled to unit power (Fig. 3, middle panel), and then the autocorrelation function is computed (Fig. 3, bottom panel). From this function, we estimate the period by locating the first and the third zeros of the autocorrelation function and computing their distance (Fig. 3, bottom panel, black circles).

Detecting failed oscillations and estimating the true period

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

To extract the ‘true period’, we employed a method that relies on the fact that the distributions of period lengths, when failures occur, resemble bimodal distributions. Namely, we estimate the mean and standard deviation of the true period in the population and the fraction of failed oscillations from each cell and divided by the transform of a decaying exponential with a time constant of 90 min, the measured lifetime of the fluorescent protein used (GFPαav). Power spectra with peaks 4.5 times higher than the background at frequencies of 0.15–0.5 h−1 per hour were classified as oscillatory. The bandwidth used here is larger than in ref. 7 so as to include failed oscillations that should cause apparent oscillations with close to double period.

Detecting failed oscillations and estimating the true period

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

To extract the ‘true period’, we employed a method that relies on the fact that the distributions of period lengths, when failures occur, resemble bimodal distributions. Namely, we estimate the mean and standard deviation of the true period in the population and the fraction of failed oscillations from each cell and divided by the transform of a decaying exponential with a time constant of 90 min, the measured lifetime of the fluorescent protein used (GFPαav). Power spectra with peaks 4.5 times higher than the background at frequencies of 0.15–0.5 h−1 per hour were classified as oscillatory. The bandwidth used here is larger than in ref. 7 so as to include failed oscillations that should cause apparent oscillations with close to double period.
the measured periods from each cell. For that, we find the maximum likelihood estimates for a single Gaussian (given by the mean and standard deviation of the measured periods) and for a mixture of two Gaussians, such that the mean and variance of the second are double that of the first (found using an iterative expectation maximization algorithm). The appropriate model is selected by a likelihood ratio test with significance level of 0.01 between the two models. That is, we only select the 2-Gaussian model if the p-value of this test is smaller than 0.01. Finally, we performed the fitting with each subset of data lacking one of the measured periods (leave-one-out technique). This procedure results in N estimates each using N − 1 measured periods from which the variance of the estimates is estimated.

Results

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

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

Under all conditions, as visible from the distributions in Fig. 4, the period lengths vary widely. Given their mean and variability, a number of short-lasting periods (<100 min) are expected (visible in Fig. 4). To verify that these did not occur in a higher than expected frequency, for the condition 30 °C (the one with most samples), we computed the probability of having such or a more extreme number of periods smaller than 100 min (i.e. a p-value) assuming the fitted model (see below and the Methods section). From the model, 2.93 ‘short periods’ are expected while 3 were detected, which results in a p-value of 0.01 (i.e., the number of events observed is not unlikely.

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

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

Using the fitting, we estimated the number of failed oscillations in each cell, under each condition (see Methods). The fraction of successful oscillations (R) is shown in Table 1, for each condition. For 28 °C and 30 °C, the repressilator loses much of its robustness, as several expected oscillations were not detected. This agrees with the observed decrease in functionality (P) for temperatures above 30 °C (Table 1).

Paper Molecular BioSystems

Paper Molecular BioSystems

Under all conditions, as visible from the distributions in Fig. 4, the period lengths vary widely. Given their mean and variability, a number of short-lasting periods (<100 min) are expected (visible in Fig. 4). To verify that these did not occur in a higher than expected frequency, for the condition 30 °C (the one with most samples), we computed the probability of having such or a more extreme number of periods smaller than 100 min (i.e. a p-value) assuming the fitted model (see below and the Methods section). From the model, 2.93 ‘short periods’ are expected while 3 were detected, which results in a p-value of 0.01 (i.e., the number of events observed is not unlikely.

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

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

Using the fitting, we estimated the number of failed oscillations in each cell, under each condition (see Methods). The fraction of successful oscillations (R) is shown in Table 1, for each condition. For 28 °C and 30 °C, the repressilator loses much of its robustness, as several expected oscillations were not detected. This agrees with the observed decrease in functionality (P) for temperatures above 30 °C (Table 1).
Also in Table 1, we show the mean and standard deviation of both the apparent period and the true, estimated period. The mean true period, \( \mu \), always decreases with increasing temperature. On the other hand, the mean apparent period, \( m \), is minimal at 30 °C.

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

Next, we investigated the causes for the decrease in robustness with temperature. In particular, we investigated how temperature affects the functionality of the three component proteins of the repressilator, namely, CI, LacI, and TetR. First, studies suggest that temperature increases from 30 °C to 42 °C, CI loses approximately half of its DNA-binding stability.27 On the other hand, the DNA-binding affinity of LacI does not vary significantly between 28 °C and 37 °C.28 Similarly, TetR’s functionality is unaltered from 30 °C to 40 °C.29 We thus hypothesized that a possible cause for the loss of robustness of the repressilator with increasing T was the weakening effectiveness of CI as a repressor.

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

conclusions and discussion

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

Molecular Biosystems

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

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

Next, we investigated the causes for the decrease in robustness with temperature. In particular, we investigated how temperature affects the functionality of the three component proteins of the repressilator, namely, CI, LacI, and TetR. First, studies suggest that temperature increases from 30 °C to 42 °C, CI loses approximately half of its DNA-binding stability.27 On the other hand, the DNA-binding affinity of LacI does not vary significantly between 28 °C and 37 °C.28 Similarly, TetR’s functionality is unaltered from 30 °C to 40 °C.29 We thus hypothesized that a possible cause for the loss of robustness of the repressilator with increasing T was the weakening effectiveness of CI as a repressor.

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

conclusions and discussion

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

Molecular Biosystems

Table 1 Kinetics of the repressilator at different temperatures. Temperature (°C), fraction of functional cells (%), true period, and the standard deviation (% of the apparent period) at 30 °C. For p-values < 0.05, the hypothesis that the two distributions are the same is rejected

| Temperature (°C) | No. of intervals | \( \mu \) (s) | \( \sigma \) (s) | CV
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>157</td>
<td>1242</td>
<td>1166</td>
<td>0.88</td>
</tr>
<tr>
<td>27</td>
<td>229</td>
<td>1452</td>
<td>1191</td>
<td>0.67</td>
</tr>
<tr>
<td>30</td>
<td>33</td>
<td>1530</td>
<td>1046</td>
<td>0.85</td>
</tr>
<tr>
<td>33</td>
<td>539</td>
<td>788</td>
<td>807</td>
<td>1.05</td>
</tr>
<tr>
<td>37</td>
<td>324</td>
<td>714</td>
<td>785</td>
<td>1.21</td>
</tr>
</tbody>
</table>

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

conclusions and discussion

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

Molecular Biosystems

Table 2 Kinetics of the repressilator at different temperatures. Temperature (°C), fraction of functional cells (%), true period, and the standard deviation (% of the apparent period) at 30 °C. For p-values < 0.05, the hypothesis that the two distributions are the same is rejected

| Temperature (°C) | No. of intervals | \( \mu \) (s) | \( \sigma \) (s) | CV
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>157</td>
<td>1242</td>
<td>1166</td>
<td>0.88</td>
</tr>
<tr>
<td>27</td>
<td>229</td>
<td>1452</td>
<td>1191</td>
<td>0.67</td>
</tr>
<tr>
<td>30</td>
<td>33</td>
<td>1530</td>
<td>1046</td>
<td>0.85</td>
</tr>
<tr>
<td>33</td>
<td>539</td>
<td>788</td>
<td>807</td>
<td>1.05</td>
</tr>
<tr>
<td>37</td>
<td>324</td>
<td>714</td>
<td>785</td>
<td>1.21</td>
</tr>
</tbody>
</table>

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

conclusions and discussion

We studied the behaviour of the repressilator at different temperatures. We observed that the fraction of functional cells
RNA production is a sub-Poissonian process, within the range studied.27–29 Given this and all of the above, it is useful to discuss the different effects of temperature.31,32 For that, the CI–DNA interaction should be replaced by a less temperature-dependent repression mechanism. This modification is not expected to compromise the sensitivity of the system since CI activity is already temperature-dependent (see ref. 13). In particular, we expect the decay rates of the proteins to increase, which decreases the period length.35 The increased protein decay rates are expected from both increased rates of degradation and increased doubling rate of the cells. This allows the repressor to be sensitive to temperature changes in the range tested.

We hypothesize that the design of genetic clocks that are insensitive to temperature will have to be able to compensate for increased speed of processes such as cell division, open complex formation,31 among others.

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

Recent studies in E. coli suggest that, provided no repression, RNA production is a sub-Poissonian process, within the range of temperature tested here.1244–1248. For that, the CI–DNA interaction should be replaced by a less temperature-dependent repression mechanism. This modification is not expected to compromise the sensitivity of the system since CI activity is already temperature-dependent (see ref. 13). In particular, we expect the decay rates of the proteins to increase, which decreases the period length.35 The increased protein decay rates are expected from both increased rates of degradation and increased doubling rate of the cells. This allows the repressor to be sensitive to temperature changes in the range tested.

We hypothesize that the design of genetic clocks that are insensitive to temperature will have to be able to compensate for increased speed of processes such as cell division, open complex formation,31 among others.

Acknowledgements

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

Notes and references


Synchro ny between Genetic Repressilators in Sister Cells in Different Temperatures
Jerome G. Chandraseelan, Samuel M. D. Oliveira, Antti Häkkinen, Sofia Startceva, and Andre S. Ribeiro

Abstract—We used live E. coli containing synthetic genetic oscillators to study how the degree of synchrony between the genetic circuits of sister cells changes with temperature. We found that both the mean and the variability of the degree of synchrony between the fluorescence signals from sister cells are affected by temperature. Also, while most pairs of sister cells were found to be highly synchronous in each condition, the number of asynchronous pairs increased with increasing temperature, which was found to be due to disruptions in the oscillations. Finally we provide evidence that these disruptions tend to affect multiple generations as opposed to individual cells. These findings provide insight in how to design more robust synthetic circuits and in how cell division can affect their dynamics.

Keywords—repressor, robustness, synchrony, synthetic biology

I. INTRODUCTION

GENETIC circuits are capable of performing tasks such as time keeping [1], state holding [2], and signal modulation and multiplexing [3]. Naturally occurring circuits responsible for these critical tasks have evolved to be sensitive to specific inputs but robust to external fluctuations such as transient environmental changes [4]-[6]. Such behavior is necessary to regulate periodic cellular processes operating under a wide range of conditions, while maintaining efficiency to respond to environmental signals.

Synthetic versions of some important naturally occurring circuits have been engineered [7],[8]. These synthetic constructs aim to allow the programming of novel biological functions but also allow a better understanding of the behavior of naturally occurring circuits, which may allow enhancing their performance as well. To match the performance of the natural circuits, the components of the synthetic circuits must be carefully selected such that both the desired behavior and level of robustness are attained [9]. For this purpose, synthetic circuits utilize chemical components whose physical and chemical properties are well characterized [10].

One example of a synthetic oscillator is the repressilator engineered by Elowitz et al. [7]. This circuit consists of three genes organized in a ring topology, each inhibiting the expression of a neighboring gene. These interactions form a negative feedback loop, which causes the protein levels of the component genes to oscillate over time. Additionally, one of the component genes is used to control a reporter gene producing green fluorescent proteins, which allows visualizing the system’s behavior using fluorescence microscopy. Such oscillator could be used e.g. for time keeping, synchronization via phase-locking, or signal modulation and multiplexing [11].

Temperature is one environmental factor that is known to affect most cellular processes, e.g. by modulating the gene expression dynamics. Evidence suggests that natural time keeping circuits, such as circadian oscillators, have evolved to be robust against temperature fluctuations [4],[5],[12]. In contrast, previous studies of synthetic oscillators have found that the constructed circuits are not immune to temperature changes. In one study, the period of the oscillator was found to decrease monotonically with increasing temperature between 25 and 37 °C, caused by a two-fold change in protein stability [13]. Our previous study on the dynamics of the Elowitz repressilator provides evidence of a similar pattern, but also notes that the most cells exhibit disrupted oscillations for temperatures over 37 °C. Evidence was then provided that this is due to loss of functionality of one of the component proteins [14].

In this work, we use live E. coli containing a synthetic genetic repressilator [7] to study how the synchrony between sister cells changes as a function of temperature and the resulting changes in the robustness of these circuits. In our cells, the loss of synchrony is inevitable since, following cell division, it is not maintained by any process between the sister cells [7]. This would lead to tangible phenotypic differences between them, provided that the clocks were used to regulate some key cellular process. Here we quantify such degree of asynchrony as a function of temperature. The findings will be important in providing insight in designing more robust synthetic circuits, and in understanding the behavior of naturally occurring circuits.
II. MATERIALS AND METHODS
A. Cell Culturing and Microscopy
Cells of E. coli lac strain MC4100 containing the repressor and the reporter plasmids were generously provided by M. B. Elowitz, Princeton University, NJ, USA. The cells were grown in minimal media with 2 mM MgSO\textsubscript{4}/7H\textsubscript{2}O (Sigma-Aldrich, USA), 7.6 mM (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} (Sigma Life Science, USA), 30 mM FeSO\textsubscript{4}/7H\textsubscript{2}O (Sigma Life Science, USA), 1 mM EDTA (Sigma Life Science, USA), and 60 mM KPH (Sigma Life Science, USA) (pH 6.8) supplemented with 0.5% glycerol (Sigma Life Science, USA) and 0.1% casamino acids (Fluka Analytical, USA) overnight at 28, 30, or 37 °C with shaking at 300 rpm to an optical density of 0.1 at 600 nm. Next, cells were diluted into fresh media and a few ml of the culture was placed between a coverslip and a slab of 2% low melting agarose in minimal media. During time lapse microscopy, the temperature of the samples was kept stable by a control chamber (FCS2, Bioptechs, PA, USA). Images were obtained every 15 min for 10 h using a Nikon Eclipse (TE2000-U, Nikon, Tokyo, Japan) inverted C2 confocal laser-scanning system with a 100 Apo TIRF (1.49 NA, oil) objective. GFP fluorescence was measured using a 488 nm laser (Melles-Griot) and a 515/30 nm detection filter. For image acquisition, Nikon EZ-C1 software was used.

B. Image Analysis
The cells were manually segmented in the fluorescence images in each frame of the time series (automatic segmentation is problematic due to the oscillatory signal). Afterwards, cell lineages were established such that a cell (segment) is associated to the cell with largest overlapping cell in the previous frame, after correcting for global translation between the two frames. A cell division was recorded in the case where two cells were associated with the same cell in the previous frame. Finally, the average fluorescence intensity was extracted from each frame and each cell for further analysis.

C. Functionality and Estimation of the Periods
Since a large fraction of the cells do not appear to exhibit oscillations [7], we used the criterion proposed by Elowitz et al. [7] to categorize the cells as functional or dysfunctional, and only included the functional cells in the further part of our analysis. In this method, the power spectral density (estimated using discrete Fourier transform) of the signal is compared with that of a decaying exponential with a time constant of 90 min, the measured lifetime of the fluorescent protein [7]. Cells with spectra exhibiting peaks higher than 3.1 times the background spectral density at frequencies of 0.2 to 0.5 h\textsuperscript{-1} were considered functional. This method was applied to each branch of the lineage trees to determine the functionality of the youngest cells, whereas the other cells were considered to be functional if they had at least one functional child.

The period of oscillations were estimated using the zeros of the autocorrelation sequence of background-corrected intensity signals [14]. In this method, the raw intensity signal is fit with a quadratic polynomial of time, in least-squares sense, to estimate the background trend (caused e.g. by accumulation of GFP and photobleaching). Next, the background trend is subtracted, the residual is scaled to unit power, and the autocorrelation sequence is computed. The period can be estimated by locating the first and third zero of this sequence, as they are expected to occur at lags of 1/4 and 3/4 times the period.

D. Estimating Robustness and True Period Distributions
Particularly at higher temperatures, the period distributions were observed to exhibit bimodality [14]. This might be caused by either the repressor or the reporter failing, causing an apparent doubling of the period [14]. Higher-order harmonics are not expected to be present due to the finite measurement time.

For this, we find the maximum likelihood estimates using a model of a single normal distribution (given by the mean and standard deviation of the data) and a mixture model of two normal distributions, with the mean and variance of the second equal to twice that of the first (found using an iterative expectation maximization algorithm [15]), using the measured periods in each condition. The appropriate model is selected using a likelihood ratio test with a significance level of 0.01, that is, the bimodal model is only selected if it fits significantly better than the unimodal one. Finally, in the bimodal case, the robustness of the population is determined from the total probability mass in the first normal distribution, and the robustness of the individual cells are determined by a maximum-a-posteriori classifier given the estimated parameters.

III. RESULTS
We analyzed time series of cells with reporters imaged in three different temperatures: 28, 30, and 37 °C. The time series were sampled every 15 min and were 10h in duration. The image analysis process produced a total of 172, 186, and 683 cells in 28, 30, and 37 °C, respectively. Fig. 1 shows examples of confocal microscope images of a few cells and the corresponding extracted intensities.

D. Estimating Robustness and True Period Distributions
Particularly at higher temperatures, the period distributions were observed to exhibit bimodality [14]. This might be caused by either the repressor or the reporter failing, causing an apparent doubling of the period [14]. Higher-order harmonics are not expected to be present due to the finite measurement time.

For this, we find the maximum likelihood estimates using a model of a single normal distribution (given by the mean and standard deviation of the data) and a mixture model of two normal distributions, with the mean and variance of the second equal to twice that of the first (found using an iterative expectation maximization algorithm [15]), using the measured periods in each condition. The appropriate model is selected using a likelihood ratio test with a significance level of 0.01, that is, the bimodal model is only selected if it fits significantly better than the unimodal one. Finally, in the bimodal case, the robustness of the population is determined from the total probability mass in the first normal distribution, and the robustness of the individual cells are determined by a maximum-a-posteriori classifier given the estimated parameters.

III. RESULTS
We analyzed time series of cells with reporters imaged in three different temperatures: 28, 30, and 37 °C. The time series were sampled every 15 min and were 10h in duration. The image analysis process produced a total of 172, 186, and 683 cells in 28, 30, and 37 °C, respectively. Fig. 1 shows examples of confocal microscope images of a few cells and the corresponding extracted intensities.
First, we estimated the period of each functional cell, and finally estimated the distribution of periods and robustness in each condition (see methods). Here functional cells are those that exhibit oscillations of "strong enough" power, and robust cells are those that exhibit oscillations at the fundamental frequency (and not some harmonic). Summary of the statistics is shown in Table I.

Next, we computed the correlation between the intensity signals of each pair of sister cells. This correlation results from loss of synchrony caused both by division and variations in the behavior of the cells over their lifetime (i.e. variations/drift in the period and noise in the intensity signal). The distributions of correlation coefficient extracted from each pair of cells are shown in Fig. 2, and the mean and standard deviation of the coefficients is shown in Table II.

The results indicate that the functionality is lower in the 28 °C condition (p-value of 1.56×10⁻⁴), suggesting that an increase in one-tailed binomial test with null hypothesis that the means are different in 28 °C and 30 °C conditions (p-value of 1.84×10⁻¹⁷). Since only around 40% of the cells imaged under 37 °C were found to be robust, we further computed the statistics using only the robust cells in this condition. The mean (standard deviation) period of the robust cells is around 150 (51) min, resulting in a coefficient of variation of 0.34. In comparison to the whole population of the cells in the 37 °C case, the robust cells have significantly lower mean period compared to the 30 °C (p-value of 2.50×10⁻¹⁰), suggesting that an increase in temperature results in a decrease in the period of oscillations of "properly" operating repressilators throughout the whole region. In addition, the stochasticity in the period duration appears to be similar in the 30 and 37 °C conditions and lower in the 28 °C condition.

Next, in these data, we located each pair of sister cells in which both of the sister cells were functional. We found 32, 53, and 187 such pairs in 28, 30, and 37 °C conditions, respectively, and 35 such pairs in the robust cells of 37 °C condition.

We first tested if a robust cell is more (or less) likely to have a robust sister cell than one would expect by chance, in the 37 °C condition. In our data, we found 112, 40, and 35 pairs where none, one, or both of the cells were robust, respectively, suggesting that the number of pairs where either none or both sisters are robust are overrepresented. More specifically, there is about a 0.64 chance for a robust cell to have a robust sister, and about a 0.85 chance for a non-robust cell to have a non-robust sister (cf. 0.37 in Table I). The significance of this correlation was confirmed by computing the p-value of one-tailed Fisher’s exact test with the null hypothesis that being robust or not is independent in the sister cells, resulting in a p-value smaller than 1.29×10⁻⁴.

Next, we computed the correlation between the intensity signals of each pair of sister cells. This correlation results from loss of synchrony caused both by division and variations in the behavior of the cells over their lifetime (i.e. variations/drift in the period and noise in the intensity signal). The distributions of correlation coefficient extracted from each pair of cells are shown in Fig. 2, and the mean and standard deviation of the coefficients is shown in Table II.

The results indicate that the functionality is lower in the 28 °C condition (p-value of 1.56×10⁻⁴) in one-tailed binomial test with null hypothesis that the means are different in 28 °C and 30 °C conditions (p-value of 0.09). We also found that the mean period decreases with a temperature increase from 28 °C to 30 °C (p-value of 1.84×10⁻¹⁵) in one-tailed Welch’s t’s tests with null hypothesis of equal means), and is similar in 30 °C and 37 °C conditions (p-value of 0.28). The coefficients of variation (standard deviation over the mean) extracted from 28, 30, 37 °C, and 40 for 28, 30, and 37, respectively, which suggests that the relative variations in the periods increase with temperature. These results are in agreement with previous findings [14].

Next, in these data, we located each pair of sister cells in which both of the sister cells were functional. We found 32, 53, and 187 such pairs in 28, 30, and 37 °C conditions, respectively, and 35 such pairs in the robust cells of 37 °C condition. We first tested if a robust cell is more (or less) likely to have a robust sister cell than one would expect by chance, in the 37 °C condition. In our data, we found 112, 40, and 35 pairs where none, one, or both of the cells were robust, respectively, suggesting that the number of pairs where either none or both sisters are robust are overrepresented. More specifically, there is about a 0.64 chance for a robust cell to have a robust sister, and about a 0.85 chance for a non-robust cell to have a non-robust sister (cf. 0.37 in Table I). The significance of this correlation was confirmed by computing the p-value of one-tailed Fisher’s exact test with the null hypothesis that being robust or not is independent in the sister cells, resulting in a p-value smaller than 1.29×10⁻⁴.

Next, we computed the correlation between the intensity signals of each pair of sister cells. This correlation results from loss of synchrony caused both by division and variations in the behavior of the cells over their lifetime (i.e. variations/drift in the period and noise in the intensity signal). The distributions of correlation coefficient extracted from each pair of cells are shown in Fig. 2, and the mean and standard deviation of the coefficients is shown in Table II.
equal). This loss of correlation could be due to the increase in the noise of the period as a function of the temperature, which appears to follow a similar pattern. Accordingly, since the non-robust cells contribute much of the variation in the 37 °C condition, the correlation is restored to a level comparable to the 30 °C condition when only the robust cells are considered. Interestingly, Fig. 2 reveals that in each condition, most of the cells are very highly correlated. However, increases in the temperature result in pairs of cells with wider range of correlation coefficients, including a sizable number of pairs whose series are strongly anticorrelated.

IV. CONCLUSION

We used live *E. coli* cells containing genetic Repressilators to study how the synchrony between synthetic genetic repressors contained in sister cells changes as a function of both the temperature and the affected robustness of the cells. We found that the temperature affects both the mean and the variability of the synchrony between sister cells, as measured by the correlation coefficient between their intensity time series. While in each condition most pairs of sister cells are highly correlated, the number of uncorrelated and anticorrelated pairs grows with increasing temperature. These values result in an apparent reduction in the synchrony between the sister cells for the population as a whole.

However, in the 37 °C condition, the non-robust pairs (i.e., cells whose oscillations become disrupted) were found to be responsible for these unlikely pairs, and the synchrony of the cells that remain robust is comparable to the 30 °C case, as predicted by the changes in the stochasticity of the period. Finally, we found that a robust/non-robust cell is more likely to have a sister with similar than the opposite behavior, suggesting that the disruptions in the oscillators propagate to successive generations.

To explain these results, we provide evidence that the changes in the synchrony between the sister cells are reflected in changes of the stochasticity of the period of oscillations. Such stochasticity is expected to result in the sister cells randomly drifting to different behaviors over time. This hypothesis would explain the changes in synchrony both as a function of changes in the temperature and as a function of the changes in the robustness of the cells. However, we note that it remains unclear if other noise sources, such as the stochasticity of partitioning the repressilator or GFP plasmids, differ in the different conditions, and if they have a significant effect on the synchrony of cells of common ancestry. These results further support the hypothesis that in higher temperatures the repressilators become disrupted, which might be due to the component protein CI losing functionality in these temperatures [14]. Furthermore, we have provided evidence that such failure would not only cause a disruption in the oscillatory signals, but also the loss of synchronization between similar the clock signals, which might be important if independent clocks are used to drive downstream circuits.

**REFERENCES**


Available online on the Royal Society of Chemistry's website, DOI: 10.1039/C5MB00012B
Introduction

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

One of the first reported functional synthetic circuits is the repressilator, a ring-oscillator with three genes, each expressing a protein that represses the next gene in the loop. From the study of the signal from GFP reporters in cells at 30 °C, it was shown that it oscillates (stochastically) at a slower rhythm than the cell cycle. Interestingly, for unknown reasons, only approximately 40% of the cells exhibit oscillations, i.e. are ‘functional’. Further, even though the circuit’s behaviour is uncoupled from the cell cycle in the previous phase, these functional cells become non-functional in the stationary phase, suggesting that this synthetic network is not fully uncoupled from the regulatory mechanisms of cell growth. Finally, the oscillatory behaviour can be halted by external signals. E.g. most cells lose the functionality following the addition of 50 μM of isopropyl-β-D-1-thio-galactopranoside (IPTG) to the medium.

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

Originally, the repressilator was implemented on a low-copy plasmid (p251-TiL1LC1). Because of this (as the origin of replication is pC818), each cell has, on average, 3–4 copies of the repressilator, which are functionally coupled in that the proteins coded by a gene in one of the copies can act as repressors of the next gene in the loop in all other copies of the plasmid. However, the plasmid copy number can affect the functionality of the circuit, with the lowest copy number (pSC101) resulting in the highest functionality.

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

This synthetic network is not fully uncoupled from the regulatory mechanisms of cell growth. Finally, the oscillatory behaviour can be halted by external signals. E.g. most cells lose the functionality following the addition of 50 μM of isopropyl-β-D-1-thio-galactopranoside (IPTG) to the medium.

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

This synthetic network is not fully uncoupled from the regulatory mechanisms of cell growth. Finally, the oscillatory behaviour can be halted by external signals. E.g. most cells lose the functionality following the addition of 50 μM of isopropyl-β-D-1-thio-galactopranoside (IPTG) to the medium. However, the plasmid copy number can affect the functionality of the circuit, with the lowest copy number (pSC101) resulting in the highest functionality.
the plasmid in the cell. This coupling, according to simulations of stochastic models, is expected to reduce the fluctuations in period lengths that arise from the stochasticity in gene expression and in RNA and protein degradation. So far, it is unknown whether the repressilator would function if implemented on a single-copy plasmid.

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

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

After verifying the functionality of our SCR, we compared its dynamics with the original LCR at optimal temperatures. Next, we compared their responses to changing temperatures. Finally, we studied their robustness to external perturbations. The results provide new insights into the effects of coupling on genetic circuits in general, and clocks in particular. Understanding the functioning of natural, as well as synthetic clocks, such as the repressilator, will assist in the understanding of how cells regulate the timing of several processes and contribute to ongoing efforts in synthetic biology to produce circuits useful in assisting medicine and biotechnology, particularly given the important role that synthetic clocks are expected to play as sensors and regulators in future synthetic circuits.

Methods

Design, validation, and functioning of single-copy repressilators

The repressilator consists of a three-gene network in a loop formation, with each gene repressing the next gene in the loop. Each network consists of three genes of the repressilator. The signals of the GFP reporter and the actual protein numbers of the component genes are nearly superimposed, as expected.

Finally, we conducted live cell microscopy to determine whether cells with the SCR exhibited a fluorescent signal whose intensity oscillates (for example Fig. 2 and Fig. S7, ESI†). To build the SCR, we transferred the sequence coding for the repressor from the original low-copy plasmid into a single-copy F-plasmid (pBAC2-lTlrLLtCL). The original GFP reporter system was left unchanged. The construction history of the SCR plasmid and PCR are shown in Fig. S4 and S5 (ESI†), respectively. To further confirm the proper construction of the SCR plasmid we performed sequencing and qPCR (Fig. S6 and ESI†, respectively).

Strains, genetic circuit assembly, and growth conditions

Cells of E. coli host strain lac– MC 4100, containing the low-copy repressilator p2SH-lTlrLLtCL and the reporter pG2E1-GFPav, were grown overnight and 1 μL of each was used for the microinjection of a Drosophila melanogaster embryo. The injections were performed using a microinjection apparatus (Eppendorf Micropipette, Hamburg, Germany) and were conducted into the posterior polar discs to induce the repressilator expression in the eye. The resulting flies were raised on standard laboratory food and the RNA and protein degradation. So far, it is unknown whether the repressilator would function if implemented on a single-copy plasmid.

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

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

After verifying the functionality of our SCR, we compared its dynamics with the original LCR at optimal temperatures. Next, we compared their responses to changing temperatures. Finally, we studied their robustness to external perturbations. The results provide new insights into the effects of coupling on genetic circuits in general, and clocks in particular. Understanding the functioning of natural, as well as synthetic clocks, such as the repressilator, will assist in the understanding of how cells regulate the timing of several processes and contribute to ongoing efforts in synthetic biology to produce circuits useful in assisting medicine and biotechnology, particularly given the important role that synthetic clocks are expected to play as sensors and regulators in future synthetic circuits.

Methods

Design, validation, and functioning of single-copy repressilators

The repressilator consists of a three-gene network in a loop formation, with each gene repressing the next gene in the loop. Each network consists of three genes of the repressilator. The signals of the GFP reporter and the actual protein numbers of the component genes are nearly superimposed, as expected.

Finally, we conducted live cell microscopy to determine whether cells with the SCR exhibited a fluorescent signal whose intensity oscillates (for example Fig. 2 and Fig. S7, ESI†). To build the SCR, we transferred the sequence coding for the repressor from the original low-copy plasmid into a single-copy F-plasmid (pBAC2-lTlrLLtCL). The original GFP reporter system was left unchanged. The construction history of the SCR plasmid and PCR are shown in Fig. S4 and S5 (ESI†), respectively. To further confirm the proper construction of the SCR plasmid we performed sequencing and qPCR (Fig. S6 and ESI†, respectively).

Strains, genetic circuit assembly, and growth conditions

Cells of E. coli host strain lac– MC 4100, containing the low-copy repressilator p2SH-lTlrLLtCL and the reporter pG2E1-GFPav, were grown overnight and 1 μL of each was used for the microinjection of a Drosophila melanogaster embryo. The injections were performed using a microinjection apparatus (Eppendorf Micropipette, Hamburg, Germany) and were conducted into the posterior polar discs to induce the repressilator expression in the eye. The resulting flies were raised on standard laboratory food and the RNA and protein degradation. So far, it is unknown whether the repressilator would function if implemented on a single-copy plasmid.

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

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

After verifying the functionality of our SCR, we compared its dynamics with the original LCR at optimal temperatures. Next, we compared their responses to changing temperatures. Finally, we studied their robustness to external perturbations. The results provide new insights into the effects of coupling on genetic circuits in general, and clocks in particular. Understanding the functioning of natural, as well as synthetic clocks, such as the repressilator, will assist in the understanding of how cells regulate the timing of several processes and contribute to ongoing efforts in synthetic biology to produce circuits useful in assisting medicine and biotechnology, particularly given the important role that synthetic clocks are expected to play as sensors and regulators in future synthetic circuits.

Methods

Design, validation, and functioning of single-copy repressilators

The repressilator consists of a three-gene network in a loop formation, with each gene repressing the next gene in the loop. Each network consists of three genes of the repressilator. The signals of the GFP reporter and the actual protein numbers of the component genes are nearly superimposed, as expected.

Finally, we conducted live cell microscopy to determine whether cells with the SCR exhibited a fluorescent signal whose intensity oscillates (for example Fig. 2 and Fig. S7, ESI†). To build the SCR, we transferred the sequence coding for the repressor from the original low-copy plasmid into a single-copy F-plasmid (pBAC2-lTlrLLtCL). The original GFP reporter system was left unchanged. The construction history of the SCR plasmid and PCR are shown in Fig. S4 and S5 (ESI†), respectively. To further confirm the proper construction of the SCR plasmid we performed sequencing and qPCR (Fig. S6 and ESI†, respectively).

Strains, genetic circuit assembly, and growth conditions

Cells of E. coli host strain lac– MC 4100, containing the low-copy repressilator p2SH-lTlrLLtCL and the reporter pG2E1-GFPav, were grown overnight and 1 μL of each was used for the microinjection of a Drosophila melanogaster embryo. The injections were performed using a microinjection apparatus (Eppendorf Micropipette, Hamburg, Germany) and were conducted into the posterior polar discs to induce the repressilator expression in the eye. The resulting flies were raised on standard laboratory food and the RNA and protein degradation. So far, it is unknown whether the repressilator would function if implemented on a single-copy plasmid.

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

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

After verifying the functionality of our SCR, we compared its dynamics with the original LCR at optimal temperatures. Next, we compared their responses to changing temperatures. Finally, we studied their robustness to external perturbations. The results provide new insights into the effects of coupling on genetic circuits in general, and clocks in particular. Understanding the functioning of natural, as well as synthetic clocks, such as the repressilator, will assist in the understanding of how cells regulate the timing of several processes and contribute to ongoing efforts in synthetic biology to produce circuits useful in assisting medicine and biotechnology, particularly given the important role that synthetic clocks are expected to play as sensors and regulators in future synthetic circuits.
plasmids, were generously provided by M. Elozou (Princeton University, NJ, USA).
In cells of the same strain, MC1400, containing only the reporter system (also generously provided by M. Elozou), the LCR was controlled by the tetacycline repressor (TetR) and the promoter Ptet0.2, we inserted the engineered single-copy F-plasmid containing the repressilator system (pBAC2-TflR-LCLC) from p251-TflR-LCLC (Fig. S1, ESIF).

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

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

Cell culturing optimization
To avoid plasmid instability15–19 and to optimize culturing protocols,2–4 we proceeded as follows: (i) as in the original protocol,2–4 cells were taken from a stock (at –80 °C) and streaked onto an LB agar medium with appropriate antibiotics for 14–16 hours; (ii) at this stage, we added to the original protocol an extra step of 8 hours of cell growth from single colonies in LB agar medium;2–4 (iii) next, as in the original protocol, a few colonies from the second plate were inoculated in M63 liquid minimal medium with antibiotics for 10 hours. Finally, we placed cells at 28 °C, 30 °C, 33 °C, or 37 °C for 4 hours, after which we measured the optical density (OD). After these 4 hours of culturing OD₉₀₀ = 0.1–0.2 was reached under all conditions, as reported in ref. 3. No differences in behaviour were found between cells with the LCR and with the SCR during this process.

Microscopy
During time lapse microscopy, cells were kept at a stable temperature in a thermal chamber (Biotepar, PECOS, PA, USA). Images of both LCR and SCR cells were obtained every 15 minutes for 10 hours using a Nikon Eclipse (Ti-E, Nikon, Japan) inverted microscope equipped with a ×100 oil immersion objective lens, a 100× Apo TIRF (1.49 NA, oil) objective, images were taken from multiple locations at each moment. GFP fluorescence was excited using a 488 nm argon ion laser (Melles-Griot) and measured using a 515/30 nm emission filter. The pixel dwell time was set to 2.4 μs, so that the total image acquisition time per frame was 6.0 μs. The laser shutter was open only during exposure to minimize photo-blanching. We used NIS-Elements software (Nikon) for image acquisition.

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

We used the following criterion to determine the function-ality of the repressilator:2 for cells presenting the fluorescence signal from start to end, a discrete Fourier transform was applied and divided by the transform of a decaying exponential of the measured lifetime of the fluorescent protein used (GFPαav), with a time constant of 90 min.6 From these, cells with power spectra with peaks 3 times higher than the background, at frequencies of 0.2–0.5 per hour, were classified as oscillatory. As discussed in ref. 3, the bandwidth was larger than that reported in ref. 2 to detect failed oscillations, which create apparent periods close to the double mean and standard deviation.6

The same method as reported in ref. 1 was used to estimate the period lengths of oscillations for each cell. One important assumption in the detected trend from a raw signal, followed by scaling the residual to unit power and computing an autocorrelation function. Afterwards, periods were estimated by locating the first and third zero of the autocorrelation function, and computing their distance.

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

As reported in ref. 3, these distributions were fitted to either a single Gaussian or to a mixture of two Gaussians. In the distributions where bimodality was observed, the ‘apparent’ and ‘true’ periods were estimated from the fitting of a 2 Gaussian model such that the mean and the variance of the second period were twice that of the first (found using an iterative expectation model).
maximization algorithm). The appropriate model was selected by a likelihood ratio test with a significance level of 0.01 between the two models. That is, we only select the 2-Gaussian model if the p-value of this test is less than 0.01. This methodology was used here solely to quantify the robustness of the repressilator-reporter system to internal fluctuations.

Results

Effects of the plasmid copy number under optimal conditions

We imaged cells with the SCR and LCR at 30 °C, the temperature at which the LCR exhibits shorter period and higher functionality and robustness, due to, among others, the temperature dependence of C21,22 (see results in Table 1). It is noted that 30 °C is the temperature at which cells grow at a relatively slow rate under the microscope (division time of ~60 min). Thus, it is reasonable to assume that, in the case of SCR cells, most of the time only one copy of the repressilator is present in the cells. Nevertheless, in all cases, the SCR cells contain significantly fewer copies of repressilators than the LCR cells (see below).

From these data, we assessed if the dynamics of the LCR and SCR differed significantly by performing a Kolmogorov-Smirnov (K-S) test of statistical significance to determine whether the two sets of time lengths of oscillations could be obtained from equal distributions. We obtained a p-value of 0.006, from which we concluded that the dynamics of the circuits differs statistically (typically, the null hypothesis is rejected at a significance level of 0.01). From this difference, and since the repressilator circuits in the SCR and LCR implementation do not differ, it is possible to conclude that more than one copy of the 3-4 copies of the LCR present in each cell is active.

Interestingly, this difference in the period distributions (particularly the higher variance) is not reflected in the robustness of the oscillations of the SCR, which does not differ from the robustness of the LCR. Also, the SCR exhibits higher functionality at 30 °C than the LCR (similar values to those previously reported in ref. 2). Finally, the period of the SCR is longer and noisier (higher C2).

Dynamics at different temperatures

In Table 3, we show the results from K-S tests of statistical significance to determine whether the distributions of periods from the SCR and LCR could be obtained from equal distributions, at each temperature. This table indicates that the two circuits exhibit different dynamics at all temperatures. Nevertheless, Table 2 indicates that both circuits respond similarly to temperature changes, in the range tested. Specifically, one similarity is that in both circuits

<table>
<thead>
<tr>
<th>Copy no.</th>
<th>F (%)</th>
<th>No. of oscillating cells R (%)</th>
<th>s (min)</th>
<th>m (min)</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCR 42</td>
<td>27</td>
<td>100</td>
<td>251</td>
<td>89</td>
<td>0.128</td>
</tr>
<tr>
<td>SCR 48</td>
<td>59</td>
<td>100</td>
<td>313</td>
<td>122</td>
<td>0.152</td>
</tr>
<tr>
<td>LCR 28</td>
<td>30</td>
<td>100</td>
<td>231</td>
<td>103</td>
<td>0.126</td>
</tr>
<tr>
<td>LCR 30</td>
<td>42</td>
<td>100</td>
<td>251</td>
<td>89</td>
<td>0.126</td>
</tr>
<tr>
<td>LCR 33</td>
<td>35</td>
<td>100</td>
<td>275</td>
<td>100</td>
<td>0.126</td>
</tr>
<tr>
<td>LCR 37</td>
<td>30</td>
<td>100</td>
<td>295</td>
<td>111</td>
<td>0.126</td>
</tr>
<tr>
<td>SCR 28</td>
<td>32</td>
<td>100</td>
<td>342</td>
<td>124</td>
<td>0.126</td>
</tr>
<tr>
<td>SCR 30</td>
<td>40</td>
<td>100</td>
<td>331</td>
<td>112</td>
<td>0.126</td>
</tr>
<tr>
<td>SCR 33</td>
<td>24</td>
<td>98</td>
<td>364</td>
<td>161</td>
<td>0.126</td>
</tr>
<tr>
<td>SCR 37</td>
<td>21</td>
<td>95</td>
<td>404</td>
<td>145</td>
<td>0.126</td>
</tr>
</tbody>
</table>

Table 3. p-values of the K-S test between distributions of periods from different copy number plasmids. For p-values < 0.01, the null hypothesis that the two distributions are equal is rejected.
the mean period is minimized at 30 °C. Another is that the robustness of both circuits is hampered at temperatures beyond 30 °C due to the loss of effectiveness of CI as a repressor.24–25 Finally, in both systems, the functionality is maximized at 30 °C.

On the other hand, as indicated in Table 2, one main difference is that all cells respond to temperature changes is that the functionality of the SCR has a steeper decrease with increasing temperature. This causes the SCR’s functionality to be lower at higher temperatures (53 °C and 57 °C). The other significant difference is in how the oscillations change with temperature. While in the LCR the steepest change in the mean period length occurs when raising the temperature from 28 °C to 30 °C (decrease of 142 min), in the SCR it occurs when raising the temperature from 30 °C to 33 °C (increase by 51 min). We verified this by K-S tests of statistical significance to determine whether the sets of oscillation lengths at different temperatures could be obtained from equal distributions, for both the LCR and the SCR (Table 4). From these, one observes a p-value smaller than 0.01 in the SCR only when comparing data from 30 °C and 33 °C, while in the SCR such observation only occurs when comparing data from 30 °C and 33 °C, in agreement with the observed changes in the oscillations’ mean time length with temperature.

Perturbing the functioning by IPTG induction

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

We compared the robustness of the LCR and the SCR to a ‘weak’ perturbation, by addition of a small amount of IPTG to the medium. Also, we compared the robustness of the SCR to a ‘weak’ and to a ‘strong’ perturbation. For this, after starting measurements as before, we introduced IPTG into the medium at the end of the third hour of the measurements, as this is approximately the time length of one oscillation (see the Methods section). The expected effect of this permanent perturbation is the continuous induction of the P lacO1 promoter (i.e. up-regulation of the repressilator). Consequently, P lacO1 ατ01 might be expected to become permanently repressed. Since this promoter also drives the reporter, the reporter signal should become negligible, following the perturbation, if the perturbation succeeds in disrupting the oscillations.

First, we compared the effects of perturbation (50 µM of IPTG) on the dynamic of the SCR and the LCR. As shown in Table 2, we assessed the functionality (see the Methods section) in the first 3 hours prior to perturbation and in the subsequent 3 hours after the perturbation. We found that the functionality of cells with the LCR equalled 98.3% in the first 3 hours, and 98.6% in the subsequent 3 hours (61 cells imaged [i.e. 98% of the cells were perturbed]). Meanwhile, the functionality of cells containing the SCR equaled 100.0% in the first 3 hours, and 88.6% in the subsequent 3 hours (145 cells imaged [i.e. 91% of the cells were perturbed]). Thus, surprisingly, we conclude that the LCR is less robust to this perturbation than the SCR. As a side note, the reason why the functionality values are measured lower than 100% is explained in the Methods section.

Next, we compared the effects of a ‘strong’ versus a ‘weak’ perturbation in cells containing the SCR (by adding 1 mM or 50 µM of IPTG to the medium, respectively), at the end of the third hour of the measurements (112 cells imaged). We measured a functionality of 100.0% in the first 3 hours as before, but only 0.89% in the later 3 hours (99.1% of the cells were perturbed). We conclude that, as predicted in simulations of models of this and similar circuits,24–25 the robustness of the SCR’s functionality to external perturbations decreases with the strength of the perturbation, in this case defined by the concentration of IPTG in the medium.

Assessing the robustness to perturbations

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

Given the above and the observation that cells with the LCR exhibit weaker robustness to the external perturbations than compared to cells with the SCR, it is possible that cells with the LCR, not all copies of the repressilator need to be perturbed in order to disrupt the periodic signal. This is expected, given that all circuits of the LCR are necessarily dynamically coupled in a cell (as demonstrated by the existence of a periodic signal prior to perturbation), since they produce and are affected by identical proteins, which are equally available to interact with any of the circuits. To exemplify this, we implemented stochastic models of the SCR and the LCR (based on a model proposed by Zhu et al.44).

The methods are described in the ESI†. We simulated the two circuits in the presence of a small amount of IPTG (i.e. up-regulation of the repressilator). Consequently, P lacO1 ατ01 ought to become permanently repressed. Since this promoter also drives the reporter, the reporter signal should become negligible, following the perturbation, if the perturbation succeeds in disrupting the oscillations.

First, we compared the effects of perturbation (50 µM of IPTG) for the dynamic of the SCR and the LCR. As shown in Table 2, we assessed the functionality (see the Methods section) in the first 3 hours prior to perturbation and in the subsequent 3 hours after the perturbation. We found that the functionality of cells with the LCR equalled 98.3% in the first 3 hours, and 98.6% in the subsequent 3 hours (61 cells imaged [i.e. 98% of the cells were perturbed]). Meanwhile, the functionality of cells containing the SCR equaled 100.0% in the first 3 hours, and 88.6% in the subsequent 3 hours (145 cells imaged [i.e. 91% of the cells were perturbed]). Thus, surprisingly, we conclude that the LCR is less robust to this perturbation than the SCR. As a side note, the reason why the functionality values are measured lower than 100% is explained in the Methods section.

Next, we compared the effects of a ‘strong’ versus a ‘weak’ perturbation in cells containing the SCR (by adding 1 mM or 50 µM of IPTG to the medium, respectively), at the end of the third hour of the measurements (112 cells imaged). We measured a functionality of 100.0% in the first 3 hours as before, but only 0.89% in the later 3 hours (99.1% of the cells were perturbed). We conclude that, as predicted in simulations of models of this and similar circuits,24–25 the robustness of the SCR’s functionality to external perturbations decreases with the strength of the perturbation, in this case defined by the concentration of IPTG in the medium.

Assessing the robustness to perturbations

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

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

To exemplify this, we implemented stochastic models of the SCR and the LCR (based on a model proposed by Zhu et al.44).

The methods are described in the ESI†. We simulated the two circuits in the presence of a small amount of IPTG (i.e. up-regulation of the repressilator). Consequently, P lacO1 ατ01 ought to become permanently repressed. Since this promoter also drives the reporter, the reporter signal should become negligible, following the perturbation, if the perturbation succeeds in disrupting the oscillations.

First, we compared the effects of perturbation (50 µM of IPTG) for the dynamic of the SCR and the LCR. As shown in Table 2, we assessed the functionality (see the Methods section) in the first 3 hours prior to perturbation and in the subsequent 3 hours after the perturbation. We found that the functionality of cells with the LCR equalled 98.3% in the first 3 hours, and 98.6% in the subsequent 3 hours (61 cells imaged [i.e. 98% of the cells were perturbed]). Meanwhile, the functionality of cells containing the SCR equaled 100.0% in the first 3 hours, and 88.6% in the subsequent 3 hours (145 cells imaged [i.e. 91% of the cells were perturbed]). Thus, surprisingly, we conclude that the LCR is less robust to this perturbation than the SCR. As a side note, the reason why the functionality values are measured lower than 100% is explained in the Methods section.

Next, we compared the effects of a ‘strong’ versus a ‘weak’ perturbation in cells containing the SCR (by adding 1 mM or 50 µM of IPTG to the medium, respectively), at the end of the third hour of the measurements (112 cells imaged). We measured a functionality of 100.0% in the first 3 hours as before, but only 0.89% in the later 3 hours (99.1% of the cells were perturbed). We conclude that, as predicted in simulations of models of this and similar circuits,24–25 the robustness of the SCR’s functionality to external perturbations decreases with the strength of the perturbation, in this case defined by the concentration of IPTG in the medium.
Conclusions and discussion

We inserted the genetic repressor of Elowitz and Leibler\(^2\) into a single-copy F plasmid to obtain, to our knowledge, the first functional, single-copy, ring-type genetic clock. The SCR is sensitive to the strength of the perturbation, which is particularly relevant in that it increases the number of possible future applications for this circuit.

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

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

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

Acknowledgements

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

Notes and references


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

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

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

Acknowledgements

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

Notes and references

Electronic Supplementary Material (ESI) for Molecular Biosystems.
This journal is © The Royal Society of Chemistry 2015

Electronic Supplementary Material (ESI) for Molecular Biosystems.
This journal is © The Royal Society of Chemistry 2014

Electronic supplementary information

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


*Laboratory of Biosystem Dynamics, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland
**Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland.

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

Supplementary Methods

Media and chemicals

We used Lysogeny Broth (LB) and minimal nutrient (M63) media with the following components (i) and (ii), respectively: (i) 10g/L of Tryptone (Sigma-Aldrich, USA), 5g/L of yeast extract (LabM, UK) and 10g/L of NaCl (LabM, UK); (ii) 2 mM MgSO₄·7H₂O (Sigma-Aldrich, USA), 7.6 mM (NH₄)₂SO₄ (Sigma Life Science, USA), 30 µM FeSO₄·7H₂O (Sigma Life Science, USA), 2.0 mM KH₂PO₄ (Sigma Life Science, USA) pH 6.8 with Glycerol 0.5% (Sigma Life Science, USA) and Casaminoacids 0.1% (Fluka Analytical, USA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used for studying the effects of external perturbations on the Repressilator. All antibiotics used for SCR and LCR strain culturing were purchased from Sigma-Aldrich (USA): (i) 35 mg/mL kanamycin and 35 mg/mL chloramphenicol; (ii) 35 mg/mL kanamycin and 20 µg/mL Ampicillin. Agarose (Sigma Life Science, USA) was used for the microscopic slide gel preparation.

Bacterial strains and single-copy repressilator plasmid construction and validation

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

To construct the single-copy F-plasmid repressilator (SCR) system pBAC2-ITrLlICL, we amplified the functional repressilator cassette from the original plasmid (de novo validation).
Electronic Supplementary Material (ESI) for Molecular Biosystems. This journal is © The Royal Society of Chemistry 2014

Electronic Supplementary Material (ESI) for Molecular Biosystems. This journal is © The Royal Society of Chemistry 2014

Smal restriction sites were added to the end of the cassette during this procedure. The primers used were:

1-Rep.Smal-Fw: 5' CCCGGGTTCGAGAATTGTGACCG 3'
2-Rep.Smal-Rev: 5' CCCGGGTCAAGCCTGCTAAAAACGTAG 3'

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

3-Sc.ori.Cam-Smal-Fw: 5' CCCGGGTTCGAACGCGTATGCATGAG 3'
4-Sc.ori.Cam-Smal-Rev: 5' CCCGGGTAGGGGCCTGACCCAA 3'

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

We validated the SCR construction by performing gel electrophoresis (Fig. S4, for the construction; Fig. S5 for the final product) to confirm the presence of the Repressilator circuit in the single-copy plasmid.

**Sequencing the plasmid for confirmation**

A fraction (covering the vector and insert) of the new plasmid was amplified using PCR, from the Chloramphenicol resistance gene in the vector to the tetR region of the Repressilator, and sequenced using appropriate primers. The primers used for the amplification were:

5-CmR-1-F: 5' CGCGTGAGGATTCGATTCG 3'
6-tetR-3-R: 5' AGCAAGGCCTGTTATTTTATG 3'

The alignment of the sequence obtained from sequencing against the expected original Repressilator sequence using NCBI BLAST (S. F. Altschul et al., J. Mol. Biol. 1990, 215: 403-410) is shown in Fig. S6B. Further, the complete sequence of the single-copy repressilator plasmid is shown in Fig. S6C.

**qPCR verification of RNA expression**

qPCR was used to further validate the presence of each gene in the SCR plasmid in the host cell. For that, lacI E. coli MC 4100 cells, containing the SCR with the reporter system, were grown following the culturing protocols described in the methods section of the manuscript. After 10 hours of culturing in 5 mL liquid M63 medium at shaking 250 rpm, one sample was taken and rifampicin was immediately added to prevent further transcription. RNA protect reagent was used to fix the cells before their enzymatic lysis with Tris-EDTA lysozyme buffer (pH 8.3). The RNA was isolated from cells using RNeasy mini-kit (Qiagen) following manufacturer’s instructions. Total of 1 µg of RNA was used as a starting material. To ensure purity of the RNA, the RNA samples were treated with DNase free of RNase to remove residual DNA. Next, the RNA was reverse transcribed into cDNA using iSCRIPT reverse transcription super mix (Biorad). Finally, qPCR was performed using Power SYBR-green master mix (Life Technologies) with primers for the amplification of the target and the reference genes at a concentration of 200 nM. Reactions were carried out in 20 µL triplicates with 500 nM per primer. The following primers were used for quantification:

1-Rep.Smal-Fw: 5' CCCGGGTTCGAGAATTGTGACCG 3'
2-Rep.Smal-Rev: 5' CCCGGGTCAAGCCTGCTAAAAACGTAG 3'

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

We validated the SCR construction by performing gel electrophoresis (Fig. S4, for the construction; Fig. S5 for the final product) to confirm the presence of the Repressilator circuit in the single-copy plasmid.

**Sequencing the plasmid for confirmation**

A fraction (covering the vector and insert) of the new plasmid was amplified using PCR, from the Chloramphenicol resistance gene in the vector to the tetR region of the Repressilator, and sequenced using appropriate primers. The primers used for the amplification were:

5-CmR-1-F: 5' CGCGTGAGGATTCGATTCG 3'
6-tetR-3-R: 5' AGCAAGGCCTGTTATTTTATG 3'

The alignment of the sequence obtained from sequencing against the expected original Repressilator sequence using NCBI BLAST (S. F. Altschul et al., J. Mol. Biol. 1990, 215: 403-410) is shown in Fig. S6B. Further, the complete sequence of the single-copy repressilator plasmid is shown in Fig. S6C.
Electronic Supplementary Material (ESI) for Molecular Biosystems.
This journal is © The Royal Society of Chemistry 2014

For lacI gene:
7-lacI.pro-Fw: 5’ GTGGTGTGATGGAAGC 3’
8-lacI.pro-Rev: 5’ CTGGTGATGGGTCTGTC 3’

- For tetR gene:
9-tetR.pro-Fw: 5’ CGCTGTCGGGGCATTTTAC 3’
10-tetR.pro-Rev: 5’ AAGAAAGCTGGCTCAGCAC 3’

- For cl gene:
11-cl.pro-Fw: 5’ GATGCGGAGAGATGGGTAAG 3’
12-cl.pro-Rev: 5’ ACTCTATCCCCAAGTCTG 3’

The length of the amplicons was kept at 90 bp. The sequences of the primers of the reference gene 16S rRNA (EcoCyc Accession Number: EG30090) were obtained from Thermo Scientific:
13-Fw: 5’ CGTACAGCTGTTGTGAA 3’
14-Rev: 5’ GGACCCTGGGCAAAAAG 3’

The level of each target gene was normalized with the level of the 16S rRNA for all samples. The PCR cycling protocol used was 94 °C for 15 s, 51 °C for 30 s, and 72 °C for 30 s, up to 39 cycles. We used no-RT enzyme and no-Template as controls. The Ct values were obtained from the CFx Manager 2.0 Software and the fold change of the genes were analysed using the Livak method (K.J. Livak, and T.D. Schmittgen, Methods, 2001, 25, 402-408).

For the SCR, we obtained the following cycle threshold (Ct) values: 18.7 (lacI), 17.3 (tetR), and 21.6 (cl). Meanwhile, in the no-Control template control these numbers equalled 26.6, 33.2, and 31.4, respectively. These result in fold changes no smaller than 200, which indicates that the RNAs are being expressed in the SCR.

Stochastic model of the Repressilator and of coupled Repressilators

We implemented stochastic models of the LCR and of the SCR based on a model proposed by Zhu and others 1. The models are implemented using the stochastic simulation algorithm, which is a Monte Carlo simulation of the stochastic chemical kinetics governed by the chemical master equation. Gene expression is modelled by the following reactions:

\[
\emptyset \xrightarrow{N_i \cdot k_i} P_i \quad k_i = \frac{a_i}{1 + \left( \frac{p^{(i)}}{H_i} \right)^{b_i}}
\]

where the first reaction represents the production of the proteins \( P_i \) (for \( i = 1,2,3,4 \)), and the second their degradation. Here, \( N_i \) represents the copy number of the gene, \( k_i \) is the effective rate of protein production for a single gene (accounting for e.g. transcription and translation rates and the messenger RNA degradation of that gene), and \( d_i \) is the protein lifetime. In our model \( P_1 \), \( P_2 \), \( P_3 \) and \( P_4 \) correspond to TetR, \( \lambda \) cl, LacI and the GFP reporter, respectively, and

\[
\emptyset \xrightarrow{N_i \cdot k_i} P_i \quad k_i = \frac{a_i}{1 + \left( \frac{p^{(i)}}{H_i} \right)^{b_i}}
\]

where the first reaction represents the production of the proteins \( P_i \) (for \( i = 1,2,3,4 \)), and the second their degradation. Here, \( N_i \) represents the copy number of the gene, \( k_i \) is the effective rate of protein production for a single gene (accounting for e.g. transcription and translation rates and the messenger RNA degradation of that gene), and \( d_i \) is the protein lifetime. In our model \( P_1 \), \( P_2 \), \( P_3 \) and \( P_4 \) correspond to TetR, \( \lambda \) cl, LacI and the GFP reporter, respectively, and
the repressor indices are $\sigma(1)=3$, $\sigma(2)=1$, $\sigma(3)=2$ and $\sigma(4)=1$. The production rate $k$ of each gene is modulated by the concentration of the corresponding repressor proteins. In the expression, $a_i$ represents the maximum expression rate, $b_i$ is the Hill coefficient, and $H_i$ is the repressor level that results in half the repression. This model of regulation is appropriate when the repressor binding/unbinding events occur at much faster rates than gene expression.

We model the perturbation caused by the introduction of IPTG in the system at a certain point in time. In our model, IPTG allows the lac promoter to express regardless of the presence of the LacI repressors. This is modelled by the following reaction:

$$ \emptyset \xrightarrow{N_i \cdot a_0} P_i $$

where $I$ is a binary variable denoting the presence of IPTG and $a_0$ is the leak expression rate of the lac promoter when IPTG is present (i.e. when $I = 1$). This rate also accounts for the IPTG concentration, which is not modelled explicitly.

In the above model, we used copy numbers $N_i = 1$ for the SCR and $N_i = 3$ for the LCR. The following parameters were used: $a_i = 1000 \text{ min}^{-1}$, $b_i = 3$, $H_i = 1$, $d^i = 10 \text{ min}$ and $a_0 = 3.5 \text{ min}^{-1}$. The protein lifetimes were set in accordance with measurements (Taniguchi et al., Science 2010, 329, 533-538), and a Hill coefficient of $b_i = 3$ was used, since many proteins function in a multimeric form (Xia et al., Proc. Natl. Sci. U.S.A. 2007, 104, 17329–17334). The parameters $a_i$ and $a_0$ were tuned for the model dynamics to agree with the empirical data.

We simulated the above model for a duration of 10 h, such that the system state in the measurement window is largely unaffected by the initial protein concentrations (set to $P_i = 0$) (in the measurements, a similar procedure occurs, as cells prior to imaging are cultured for ~10 hours). After this, we simulate the model for further 6 h, sampling the state every 15 min. As in the measurements, the perturbation is performed 3 h after the sampling was started. In each simulation (cell), the two 3 h series prior to and after the perturbation are analysed separately to determine the functionality using the same methods as for the measurement data. A total of 1000 instances of both the SCR and the LCR were simulated to compute the statistics.

The protein lifetimes were set in accordance with measurements (Taniguchi et al., Science 2010, 329, 533-538), and a Hill coefficient of $b_i = 3$ was used, since many proteins function in a multimeric form (Xia et al., Proc. Natl. Sci. U.S.A. 2007, 104, 17329–17334). The parameters $a_i$ and $a_0$ were tuned for the model dynamics to agree with the empirical data.

We simulated the above model for a duration of 10 h, such that the system state in the measurement window is largely unaffected by the initial protein concentrations (set to $P_i = 0$) (in the measurements, a similar procedure occurs, as cells prior to imaging are cultured for ~10 hours). After this, we simulate the model for further 6 h, sampling the state every 15 min. As in the measurements, the perturbation is performed 3 h after the sampling was started. In each simulation (cell), the two 3 h series prior to and after the perturbation are analysed separately to determine the functionality using the same methods as for the measurement data. A total of 1000 instances of both the SCR and the LCR were simulated to compute the statistics.
Supplementary Figures

Fig. S1 Illustration of the construction of the SCR plasmid. The pBAC2-lTlrLLtCL plasmid was engineered by inserting the repressilator cassette into pBAC2 (Plac/ara-1-mRFP1-MS2-96x) vector (generously provided by Ido Golding, University of Illinois, USA), containing the single-copy origin of replication. The construction history was generated and adapted using SnapGene® 1.5.2.
Electronic Supplementary Material (ESI) for Molecular Biosystems. This journal is © The Royal Society of Chemistry 2014

Fig. S2 LCR period distributions at different temperatures. Solid lines represent the probability densities of the fitted model with one or two Gaussians as determined by the likelihood ratio tests. Dashed lines represent the densities of the individual components in the case of two Gaussians. Magnitudes were scaled to represent a probability densities.

Fig. S3 SCR period distributions at different temperatures. Solid lines represent the probability densities of the fitted model with one or two Gaussians as determined by the likelihood ratio tests. Dashed lines represent the densities of the individual components in the case of two Gaussians. Magnitudes were scaled to represent a probability densities.
in the case of two Gaussians. Magnitudes were scaled to represent a probability densities.

**Fig. S4** Split gels for the intermediate steps of the SCR plasmid construction. (A) PCR fragment of 3114 bp amplified from the original pZS1-lTlrLLtCL with appropriate primers, in triplicate (bracket). (B) Lane containing pTB-BAC2 backbone amplified region with the single-copy origin of replication (6961 bp) (bracket). (C) Lane 1: plasmid profile of the strain containing only the reporter plasmid (white arrow). Lanes 2 and 3: two replicates of the plasmid profile of the strain with the reporter and the final construct pBAC2-lTlrLLtCL (SCR plasmid, 10069 bp) (bracket). The numbers of the DNA ladders on the left side of Figures A, B and C are shown on an identical ladder on the right side of the figure.
Fig. S5 Split gel of the final product of the SCR plasmid construction. Lane 1: unused. Lanes 2 and 3: PCR amplification of the Repressilator circuit (3114 bp) from the SCR plasmid, pBAC2-iTrLSitCL, with appropriate primers (bracket). The ladder is identical to those in Fig. S4.
Fig. S6A. Complete map of the single-copy plasmid pBAC2-ITLrLLtCL containing the Repressilator. The green region ('sequenced_Hel_10') corresponds to the sequence in Fig. S6B.
Fig. S6B. Alignment of sequence obtained from sequencing, using the primer CmR-1-F, against the expected sequence of the original Repressilator. The alignment shows that the single-copy vector is present and the Repressilator has been inserted into it.
Fig S6C. The complete sequence of the single-copy plasmid containing the Repressilator, as determined by sequencing.
**Fig S7.** Example images of cells at 30 °C exhibiting oscillatory fluorescence levels. For each cell, 8 frames are shown along with the time after starting the imaging in minutes. From top to bottom, the first 5 cells contain a LCR while the last 5 cells contain a SCR.