The DnaA cell cycle oscillator and its coordination with cell growth and division in Escherichia coli

Ilaria Iuliani

université PARIS-SACLAY

The DnaA cell cycle oscillator and its coordination with cell growth and division in *Escherichia coli*

L'oscillateur du cycle cellulaire DnaA et sa coordination avec la croissance et la division cellulaires chez Escherichia coli

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577 Structure et dynamique des systèmes vivants (SDSV)
Spécialité de doctorat: Sciences de la vie et de la santé
Graduate School : Life Science and Health, Référent : ENS Paris-Saclay

Thèse préparée dans l'unité de recherche LBPA (Université Paris-Saclay, ENS Paris-Saclay, CNRS), sous la direction de Bianca SCLAVI

Thèse soutenue à Paris, le 21 December 2021, par

Ilaria IULIANI

Composition du jury

Stéphanie Bury Moné Professeure, Institut de Biologie Intégrative de la Cellule (Paris-Saclay)	Présidente
Ivan Matic Directeur de recherche, Institut Cochin (Université de Paris)	Rapporteur
Philippe Nghe Maître de conférences, ESPCI (Paris Sciences Lettres)	Rapporteur
Marco Cosentino Lagomarsino Directeur de recherche, IFOM (Milan-Italy)	Examinateur
Petra Levin Professeure, Washington University (St. Louis-USA)	Examinatrice
Gabriele Micali Docteur, EAWAG (Zurich-Switzerland)	Examinateur
Bianca Sclavi Directrice de recherche, LCQB (Sorbonne Université)	Directrice de thèse

THESE DE DOCTORAT
NNT : 2021UPASL104

"The most lasting effect of that time in Buffalo was having the feeling for a while that we were the only people in the world who knew something that no one else knew. I think it was best described by Albert Einstein who, upon gaining the insight of his General Theory of Relativity in 1915, said that 'something snapped inside me'. I too can say that I have felt that 'snap'. Even after 30 years I remember that feeling with joy and excitement. Yet best of all, I remember the wonderful collaboration that I had with my long time friend and associate Chick Helmstetter."

Cooper - 1997 - DNA replication the 30th anniversary of the bacterial model and the 'babymachine'

Acknowledgements

I would like to start by thanking my supervisor Bianca who introduced me to the extraordinary world of bacteria and gave me the possibility to “have fun” with them. In these years we both passed through many different things but she has been always present for me and able to transmit me her calm and her passion for her work.
I am also grateful to Marco, who gave me the opportunity to join his lab for my last year of PhD. Thanks for all the scientific (and not scientific) discussions we had. Thank you also for forcing me to face all the “big decisions” of my life!
You both helped to grow a lot as a researcher and also as person.

A big thanks goes to all the people I had the opportunity to work with in my three labs.
From LBPA lab,
I would like to thanks Jeff who helped me a lot with the microscope and Gladys who helped me with the strain construction. Thanks to “my girls” Manon and Audrey, Adolfo, Sebastian and Huong for being such good friends and colleagues and for all the good times together inside and outside the lab. Thanks to Pamela, for all the chat with “cafe-creme” and all the crazy things we did in Paris in these three years. I hope to meet you very soon in Taiwan!
From LCQB lab,
A special thanks to Alessandra and Gilles for hosting me in the lab in these two years. Thank you also to Nico and Stephane, Allyson, Wenqing, Lexuan, Oana and all the BiG lab. Thanks Sam for considering my beautiful plot a masterpiece to be hung on the wall.
Thank you also to Sara, who has been in Paris just few months but for me it is like we know each other for very long time. I will never forget our swim in Trocadero fountains in the hottest day in the history of Paris!
From IFOM,
I would like to thanks both the Ferrari and Lagomarsino group (the FILARETE peo-

ple), for all the nice things we did together in this year (and for the things we just planned to do!). Thanks to Ludovico, the climbing legend and Cristiano, the Raspberry Pi expert.

A big thanks goes also to all my many flatmates I lived with in these years.
First, I would like to thank Sharon and Vo, who hosted me when I arrived in Paris for the first time and introduced me to “gold hunting”.
Second, thanks to “the Lentilles house”, its artistic walls and Justine, my salsa dancer.
A special thanks to Santosh, my "second brother" and Indian cuisine chef. I am so proud you chose me to be your wedding witness!
Third, thanks to “the Florentine house”, its secret garden, its “lockdown activity board” and Elea&Mushu, for our delicious dinners and time together.
Lastly, thanks to “the Bassini house”, Nico, Cocca and Tob.

Next, I would like to thank all my lifelong “Italian” friends that I am always so happy to meet when I am back in Italy.
Thanks to Sara, “Seth”, Simone and Federica for our mandatory board games evenings.
Thanks to “Furli”, who has been a constant presence since the first year of University and a coffee with him has always been an important moment of the day.
Thanks to Silvia and Valentina, rowing partners and lifelong friends. Be ready to travel all around the world!

I would also like to thanks all my beautiful family.
Thanks to my parents who always believed in me and wherever I go I know that they are always here for me.
A special thanks go to my brother Luca, who illustrated my PhD work with his amazing drawings.

Finally, I would like to thanks Federico just for being my “fixed point” in my chaotic life and for making me happy every single day.

Last but not least, I would like to thanks Prof. Stéphanie Bury Moné, Prof. Ivan Matic, Prof. Philippe Nghe, Prof. Petra Levin and Dr. Gabriele Micali for accepting being part of my PhD Committee thesis and for their useful comments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Despite over 50 years of study, we still need to elucidate important features of the molecular regulatory mechanisms responsible for maintaining a robust cell cycle progression in a large number of growth conditions and in response to specific perturbations[1–7]. In particular, the debate is open concerning the regulation of DNA replication initiation in *E.coli* and its role in the dynamics of cell growth and division [8–13].

In *E.coli*, a key player in cell-cycle progression is the DnaA protein, which is involved in the initiation of DNA replication [14–17]. DnaA is commonly believed to be a cell cycle oscillator and a cell size sensor, but neither of these facts have been firmly established [18–22]. Moreover, experimental evidence of cell cycle-dependent oscillation of DnaA (or more precisely DnaA-ATP) is supported only by indirect population-level data [23].

The aim of this work is to identify the cell cycle oscillator related to DnaA activity, and relate it to the progression of the *E.coli* cell cycle through a cell size sensing mechanism.

To address this, measurements of gene expression and cell growth parameters in real time at the single cell level by using microfluidics coupled to microscopy are helpful to investigate the regulation of gene expression as a function of the cell cycle. I have started by using this approach to study protein expression from a not regulated promoter. Afterwards, I studied the cell cycle dependent changes in gene expression in the special case of the regulation by DnaA and its role in the coordination of DNA replication with cell growth and division.

A deeper understanding in this field could have a broader view if we think that a robust coordination is important to challenge microorganisms' resistance to adversarial attacks such us toxins and drugs [24] or in cancer, where this coordination is lost and cells start growing at a higher growth rate[25]. Therefore, improved un-

derstanding of how bacteria proliferate and which molecular mechanisms are vital for the correct progression of the cell cycle can provide a starting point to develop novel antimicrobials that should be aimed at disturbing normal progression of the cell cycle.

1.1 Outline

In the first two Chapters I review the main works that are the basis of my research with a particular focus on their gaps.

In particular, in **Chapter 2** I focus on cell cycle progression in *E.coli*, and on its role in gene expression variability. Despite the vast literature on this issue, little is known about how gene expression is coupled to the regulation and timing of specific events of the cell cycle, such as initiation of DNA replication or cell division, particularly in fast growing bacteria such as *E.coli* [26]. The most significant models that try to dissect this control mechanisms are reviewed. The goal of these phenomenological models is not to provide as accurate a description of a system as possible, but rather to capture the essence of the phenomena and stimulate further ideas and understanding. This concept is perfectly explained by Box's famous quote "all models are wrong, some are useful".

A vast literature also exists that directly investigates the molecular mechanisms at the basis of a robust cell cycle progression. In **Chapter 3** I review what is already known about DnaA's oscillating activity and its role in the regulation of DNA replication. In particular, I focus on the differential regulation of gene expression by the different DnaA nucleotide-bound forms, including the regulation of its own promoter and on its double role as a transcription factor and an activator of the initiation of the DNA replication process. Lastly, I review other known factors that can contribute to express a gene as a function of the cell cycle, with a focus on SeqA. Some of the questions we still need to answer are:

- How does the amount of DnaA-ATP change as a function of the cell cycle?
- How does the rate of expression of DnaA correlate with cell volume?
- How does positive and negative regulation of the DnaA promoter by DnaA influence its cell cycle dependence?
- Does regulation by SeqA play an important role?

To address these questions, time-lapse single cell microscopy experiments are necessary in order to follow cell cycle progression in real time in *E.coli*. In **Chapter 4** I describe the experimental setup and data-analysis pipeline I have designed for

studying the growth, size, and gene expression of *E.coli* in controlled environmental conditions. This includes an overview of the experimental protocol, microfluidic setup, microscopy techniques, a description of the *E.coli* strains assayed and their culture conditions, the data-analysis pipeline, from segmentation and tracking algorithm to data visualization.

In **Chapter 5**, thanks to a careful analysis of single cell data as a function of the cell cycle, I show that *E.coli* growth is biphasic; deviating from the expected exponential growth. Moreover, I also investigate the gene expression pattern within the cell cycle from promoters placed in different positions in the genome and I show that *E.coli* growth rate follows the expression of a constitutive ori-proximal promoter. Lastly, I explore whether the size of a cell at birth can influence the cell cycle dependence of its gene expression and growth rate to result in a possible cell size correction mechanism.

Chapter 6 investigates the regulation of gene expression by DnaA as a function of the cell cycle. I show how I measured the changes in DnaA-ATP activity *in vivo* in real time. This requires the development of a set of reporters of gene expression using a gene for a fluorescent protein under control of a promoter that is differently regulated by DnaA-ATP and their characterization by single-cell experiments.

In particular, I characterise for the first time the effects of the DnaA-dependent promoter regulatory elements that lead to an oscillatory pattern of gene expression strongly coupled to cell volume. While the standard view of the cell cycle sees it as the result of a single cell-cycle oscillator, here I provide evidence that at least two coupled oscillators are needed to describe the processes that coordinate cell replication, cell growth and cell-cycle progression. Lastly, I show how this approach also makes it possible to detect causality links between these different processes.

Finally, **Chapter 7** focuses on a side project I worked on during my PhD that resulted in a publication [27].

Despite the fact it is not totally connected to the main project, it is a study that focuses to better understand how gene position in the genome can have an effect on gene expression and its variability. To address this question, I present the statistical tendencies for a reference promoter to be inserted and initially maintained in specific chromosomal contexts, as well as to characterize its fate in terms of both gene expression activity and noise of the transcription reporter constructs at different insertion sites on the genome.

1.2 Collaborations

Given the itinerant character of my PhD, where I had to change laboratory almost every year, I think it is useful to quickly review where I have worked and with whom during my PhD. Moreover, I firmly believe this had an important role in defining the main different traits of my PhD. During my PhD I had the opportunity to work on all the different sides of my project: molecular biology, microfluidic, microscopy, image analysis, data analysis and theoretical modelling. Lastly, in these years I had the opportunity to work with different people from very different background that helped me to acquire expertise in all these different fields.

During my first year of PhD I was in the Laboratory of Biology and Applied Pharmacology (LBPA - ENS Paris-Saclay) where I worked mainly on strain construction and plate reader experiments with Gladys Mbemba. In the same year I also started to build the microfluidic experimental setup and the protocol for the experiments and the data analysis was developed. To do this, I have also had the opportunity to visit Pietro Cicuta's laboratory in Cambridge, UK, to be trained in how they were doing microfluidic experiments and image analysis. Back at the LBPA Jeff Audiber helped me for the microscopy while Jean-Pierre Lefevre helped me for the microfluidic side. During my second year my lab moved to Sorbonne Université, at the Pierre and Marie Curie campus, to the Laboratory of Computational and Quantitative Biology (LCQB), headed by Alessandra Carbone. Here I had to build again the microfluidic setup to continue with my experiments, in addition we acquired a new flow control system based on compressed air that needed to be set up and calibrated. On my last year, thanks to a fellowship from ARC, I had the opportunity to join Marco Cosentino Lagomarsino's group at Ifom (in Milan) to mainly work on the data analysis and theoretical part of my project.

Despite all the drawbacks that can arise from a such "dynamic" PhD, especially for the experimental works in the times of COVID lockdowns, I can finally say that it has been an enriching experience that has allowed me to work with people from very different background that helped me in different sides of my PhD project. I hope my work is also a proof of how interdisciplinarity is very fundamental in research. This of course it is not always easy. As Giorgio Parisi said in a past interview: "biologists and physicists use two different glasses to see the world: this diversity can be very useful because it allows one to see features that the other does not see, but there is also the risk of seeing only things that the other is not really interested". Nevertheless, his recent success is the proof the this risk can be overcome and interdisciplinarity is the way to new discovery.

1.3 Publications

1.3.1 Published

- *Early fate of exogenous promoters in E. coli.* Malikmohamed Yousuf, **Ilaria Iuliani**, Reshma T. Veetil, Aswin Sai Narain Seshasayee, Bianca Sclavi and Marco Cosentino Lagomarsino. Nucleic Acids Research, (2020)
- *Threshold accumulation of a constitutive protein explains E. coli cell-division behavior in nutrient upshifts.* Mia Panlilio, Jacopo Grilli, Giorgio Tallarico, **Ilaria Iuliani**, Bianca Sclavi, Pietro Cicuta, and Marco Cosentino Lagomarsino, PNAS (2021)
- *How to study bacteria adaptation to stress and environmental changes by microfluidics?* **Ilaria Iuliani**, Bianca Sclavi, Elveflow-Microfluidic application notes (2021) (see Chapter 4)

1.3.2 In preparation

- *Size-coupled production rate oscillations of a bacterial cell cycle regulator.* **Ilaria Iuliani**, Marco Cosentino Lagomarsino, Bianca Sclavi (see Chapter 6)
- *Biphasic growth in E.coli follows the change in ori-proximal gene dosage.* **Ilaria Iuliani**, Marco Cosentino Lagomarsino, Bianca Sclavi (see Chapter 5)
- *How to express a gene as a function of the cell cycle (Minireview).* **Ilaria Iuliani**, Bianca Sclavi (see Chapter 3)
- *Width and surface-to-volume ratio response to nutrient shifts in E. coli single cells mirrors stochastic fluctuations at equilibrium.* Orso Maria Romano, Mia Panlilio, **Ilaria Iuliani**, Pietro Cicuta, and Marco Cosentino Lagomarsino

Chapter 2

Cell cycle control in *E. coli*

Studies on cell cycle control in *E. coli* have always been marked by close interactions between experiments and modeling since their birth in the mid 20th century [28–30]. In particular, bacterial cell size and cell cycle control has found new interests in modeling with the novelty of microfluidics techniques that allow tracking of thousands of individual cells over hundreds division cycles. In *E. coli* different datasets and models have supported a range of conclusions from one extreme, where these processes are tightly linked to another, to another, where these processes are completely independent of each other.

In this chapter, I will first explain why it is important to take into account the cell cycle progression when we deal with a population of single cells. I will then review the foundations of the bacterial cell cycle and the new results that in the recent years have lead us to re-think about these longstanding results. Several different data sets and models significantly contributed to a deeper understanding of cell growth, division, DNA replication, gene expression and their correlation. It is of fundamental importance to use experimental datasets to propose and to test different models. An example of this "data-driven" approach to understand important mechanisms in *E. coli* growth is reported in the last section where I report the work published in 2020 where our results have allowed us to propose that a protein expressed from a constitutive promoter accumulates to a threshold level and triggers cell division. This publication comes from a collaboration with the Pietro Cicuta's group in Cambridge, Marco Cosentino Lagomarsino in Milan and my group in Paris.

2.1 Gene expression noise mainly comes from variability in cell-cycle progression

At the single cell level, gene expression and cell size can be highly heterogeneous and show phenotypic variability. This can bring advantages, for example, a particular phenotype may be useful for a subpopulation to survive when environmental conditions change or in presence of stresses. Understanding the properties of cellular noise is important for understanding the mechanism that lead to robustness of cell proliferation.

Thanks to new techniques producing single cell data many efforts have been done in these last years to better understand the sources of these variations in their own components and how cells and organisms use and control them to maintain a robust growth in the face of perturbations. It is therefore a key for the understanding of cellular homeostasis, and could help explain heterogeneous phenotypes ubiquitously observed across all domains of life, ranging from persistence to uncontrolled proliferative growth such as cancer.

In 2002, Elowitz and colleagues quantified the variability in the expression from a promoter in *E.coli* by introducing two copies of the same promoter into the genome expressing two different fluorescent proteins of different color. They observed that the amount of protein produced by a particular gene varies from cell to cell in a population and over time in a single cell, suggesting the idea that the expression noise of a gene in a clonal population is determined by intrinsic and extrinsic factors [31]. In their work, intrinsic noise describes the variation at the level of a single gene due to the stochastic nature of the involved biochemical processes, whereas extrinsic noise relates to the variability in expression shared across different genes due to global differences in cellular environment and shared upstream components (Figure 2.1).

Because of gene doubling during DNA replication, when taking a snapshot of a population two cells can differ in protein production rate or concentration because they are in different phases of their cell cycle. Alternatively, two cells at the same cell-cycle phase can differ because of cell-cycle independent effects.

The cell cycle was found to be the major factor of observed noise in the rate of gene expression [32], with the rest coming from other sources such as differences in metabolism [35], transcription factors activity [36, 37], or the activity of the expression machinery including RNA polymerase and ribosomes [38].

In particular, Noreen Walker et al. have found that the cell cycle contributes both with a deterministic part determined by cell cycle phase, and a stochastic con-

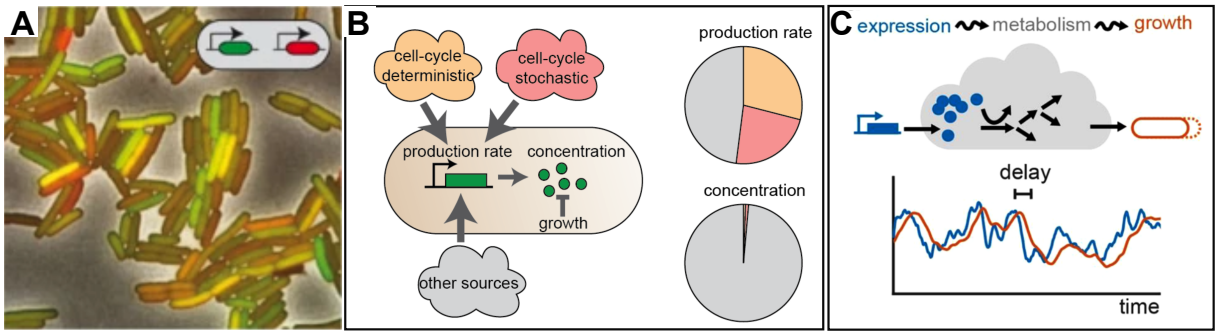


Figure 2.1: The cell cycle as a source of noise A E. coli *expressing two different fluorescent proteins from two identical promoters , in red and green, respectively. Because of noise, the ratio of red to green intensity differs from cell to cell.* ***B*** *The cell cycle progression contributes to the noise in the protein production rate, through deterministic and stochastic contributions. Other not cell cycle-related sources contribute as well. Cell cycle noise has a minor effect on the protein concentration.* ***C*** *Noise in the expression of a single enzyme (blue trace), results in fluctuations in metabolic flux that are transmitted through the metabolic network and affect growth with some time delay (orange trace). Fluctuations in the expression of limiting metabolic enzymes lead to metabolic fluxes that vary in time, which in turn cause delayed growth rate fluctuation. From [32–34]*

tribution caused by variability in the timing of replication [39]. Moreover, they have shown that the cell cycle has a minor effect on the protein concentration because of the balance between protein synthesis rate and dilution rate, both dependent on the overall translation rate. Cells might then achieve concentration homeostasis by coupling reaction rates to cell size via abundant up-stream factors like cell cycle regulators, RNA polymerases or ribosomes.

These results raise the questions of how different fluctuating activities within the cell are coordinated, and which active regulatory mechanisms are implicated in maintaining homeostasis. Moreover, it may be very useful to understand how these fluctuations are transmitted within the cell. For example, Kiviet and coworkers found that stochastic fluctuations in the concentration of a single enzyme can correlate with future cellular growth rates, indicating transmission of noise through metabolic networks (Figure 2.1C) [35].

Such phenomenological relations can be very useful to quantify and predict the behavior of complex biological systems before a full understanding of the processes on the molecular level is available. In the next chapter we will see how particular genetic circuits can be used to control this noise [40].

2.2 The foundations of the bacterial cell cycle

The growth laws

One of the first results in bacterial cell physiology has been the discovery by Schaechter and colleagues that the mass (or size) of individual cells in a growing population is coupled to the growth rate of the population and can be described by an exponential relationship [41]. The composition of the medium is irrelevant to predict cell mass and only the achieved growth rate needs to be considered.

The first definition of cell-cycle (B+C+D)

Another important quantitative principle is the bacterial cell cycle model by Cooper and Helmstetter (C+H model). They developed a new method to synchronize bacteria, the baby machine [42–44]. By using this technique they were able to measure the accumulation of DNA in time and proposed that in balanced growth the duration of DNA replication of the *Escherichia coli* chromosome is constant independently of the growth condition (C period ($\approx 40min$)) and that the cell divides after a constant time has elapsed since replication initiation (C + D period ($\approx 60min$)). The B period is the time between cell birth and initiation of DNA replication (Figure 2.3A) [43–45].

In *E. coli* cell division time can vary from 15-20 mins in fast growth conditions up to days in poor growth conditions. This means that sometimes the division time could be shorter than the time required to copy the genome. This paradox is solved by the presence of overlapping rounds of DNA replication. While at slow growth there are periods without replication activity, at fast growth replication occurs throughout the cell cycle for multiple nested chromosome copies [45]. Replication of the *E. coli* chromosome begins from a single origin, OriC, and divergently oriented replication forks proceed symmetrically along the genome to complete replication at the terminus, near the *dif* site. Since on average a cell divides at a time $C + D$ after replication initiation, an average time lag B before initiation is necessary to make the total replication time $B + C + D$ an integer multiple of the doubling time τ. Thus, defining $n = \mathrm{Int}(C + D/\tau)$ as the integer number of times that τ divides $C+D$ one has that $B+C+D = (n+1)\tau$ Furthermore, we can consider a gene at a chromosomal position defined by its normalized distance from OriC, i.e. $l = 0$ represents a gene at OriC and $l = 1$ a gene at the opposite end at the terminus.

The copy number of this gene, g , changes during the cell cycle following

$$g(t) := \begin{cases} 2^{n'} & \text{if } 0 < t < (n'+1)\tau - (C(1-l)+D) \\ 2^{n'+1} & \text{if } (n'+1)\tau - (C(1-l)+D) < t < \tau \end{cases} \tag{2.1}$$

where $n' = \mathrm{Int}\left[\frac{C(1-l)+D}{\tau}\right]$. By averaging over the cell cycle one obtains the average gene copy number

$$g = \langle g(t) \rangle_{cell\ cycle} = \frac{1}{\tau}\int_0^\tau g(t)dt = 2^n\{1 - n + \mu[C(1-l)+D]\} \tag{2.2}$$

The constant initiation mass

Using observations on Salmonella cell size and growth rate by Schaechter, Maaloe and Kjeldgaard and on the DNA replication pattern in *E. coli* by Cooper and Helmstetter, Donachie proposed that the mass versus growth rate relationship discovered by Schaechter Maaløe and Kjeldgaard results from initiation of DNA replication at multiples of a critical cell mass (mass per replication origin) [46]. This model assumes that the initiation of DNA replication is the molecular event determining the timing of all other cell cycle events such as cell division. During steady-state growth, *E. coli* initiates DNA replication at all origins synchronously once per cell cycle. The cell mass per origin at initiation, the initiation mass, was hypothesized by Donachie to be a constant, independently of growth rate. The constancy of the initiation mass also received strong experimental support under the assumption that the $C + D$ period is constant [47–49].

2.3 Single cell studies challenge the definition of the cell cycle and the invariance of the initiation mass

Deviation from the growth laws

Single cells in a given growth condition with a defined average division rate deviate from the Schaechter-Maaløe-Kjeldgaard (SMK) "growth law" with a stronger deviation at faster growth conditions [50]. In Figure 2.2(A) we can see how single cells do not follow the dependence observed for the population means between size and growth rate. Recently, in 2020, Zheng and coworkers found that this dogma does not hold in either the slow- or fast-growth regime [51].

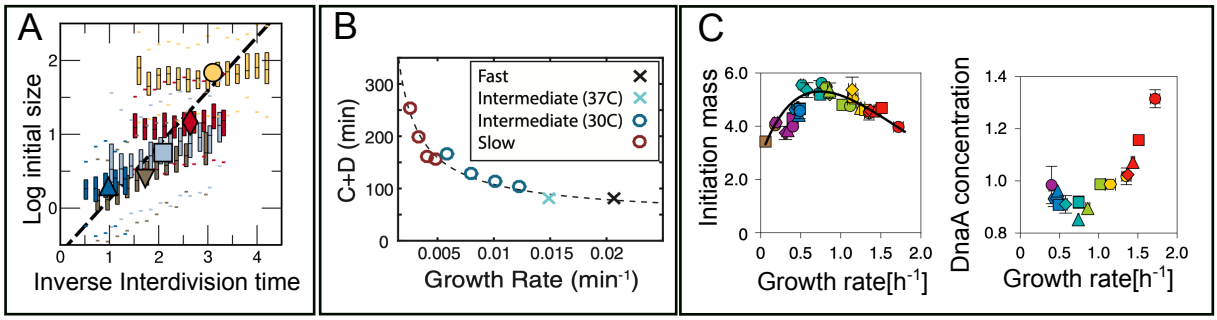


Figure 2.2: Initiation mass is not constant A *Single cells do not follow the dependence observed for the means between cell size and growth rate. From [50]. **B** The time from initiation of DNA replication to cell division is not constant but depends on the individual cell's growth rate. From [48]. **C** In contrast to the Donachie hypothesis, initiation mass has been shown to vary by 50% across different growth rates (left). DnaA concentration is found to mirror this growth rate dependence (right). From [51]*

.

C+D dependence on growth rate

In 2016, Wallden et al formulated a single cell version of the Cooper and Helmstetter (C+H) model [43, 44, 48]. An important step in making a single-cell version of the C+H model was to determine how much time cells spend in DNA replication, segregating their chromosomes, and dividing (the C and the D periods). Cooper and Helmstetter assumed this time to be constant for generation times faster than 60 minutes. Wallden et al found that the average C+D-period is relatively constant for the fast and intermediate growth conditions, but is much longer at slow growth (Figure 2.2B). They concluded that the initiation mass per chromosome is constant, however, they found a growth rate dependent time for the duration of DNA replication and cell division. This growth law, unlike the SMK law, also describes the dependence of cell size/mass on growth rate and the duration of C and D periods under slow growth conditions. In fact, the C and D periods have been shown not to be as constant as initially proposed and they tend to increase with the generation time [52–54]

The initiation mass in not constant

We have seen above that for the average behavior of a population across conditions the traditional consensus was that initiation of DNA replication roughly starts at a constant volume (or mass) per origin [21, 49]. The constancy of the initiation mass received strong experimental support [21, 47–49, 55], albeit studies exists as well that challenged the invariance of the initiation mass [56, 57].

The breakdown of the SMK growth law implies that Donachie's cell mass rela-

tion cannot hold either. Data presented in [51] by Zheng et al and reported in Figure 2.2C(left), clearly show that the initiation mass is not constant, with a non-monotonic dependence on the growth rate, peaking at a growth rate of around $0.7h^{-1}$. Using quantitative proteomics, they also found that the protein concentration of DnaA was negatively correlated with the initiation mass in wild-type cells, and their growth-rate dependencies were mirror images of each other(Figure 2.2C-right). This finding does not necessarily imply a causative relationship, but could be the result of an underlying mechanism that controls both.

2.4 Finding the "perfect" model: propose/validate/falsify

This is an area where data do not speak alone, and quantitative models are needed to understand the data, infer the key mechanisms and guide the experiments. A number of different models have been proposed where either the DNA cycle or the cell division cycle is limiting, or both [58]. The differences between these models reside in their ability to describe all the variability and coupling parameters between cell cycle events that we can measure.

For example, the recent results presented above refute the long-standing idea that cell division is enslaved to DNA-replication initiation control [51], with strong repercussions for the many existing cell cycle models that take the constant initiation mass as a central assumption [21, 49, 59, 60]. Instead, their data and model support the idea that the target of division control is the division process itself, but with inputs from chromosomal events [10, 13, 30, 61].

I will now review the current existing models, with a focus on how a correlation analysis between cell cycle intervals can be used to validate or falsify some of them.

2.4.1 Is the cell division cycle coupled to the DNA replication cycle?

$\ll$ Is the division cycle coupled to the DNA replication cycle? $\gg$. In these recent years many models have been proposed trying to answer to this question. Many of them came up with conflicting conclusions. In particular, three different classes of models can be identified [62].

According to the first class of models, DNA replication and segregation are seen as limiting for cell division, while division has no influence on DNA replication. At the single-cell level, different couplings between DNA replication and cell division

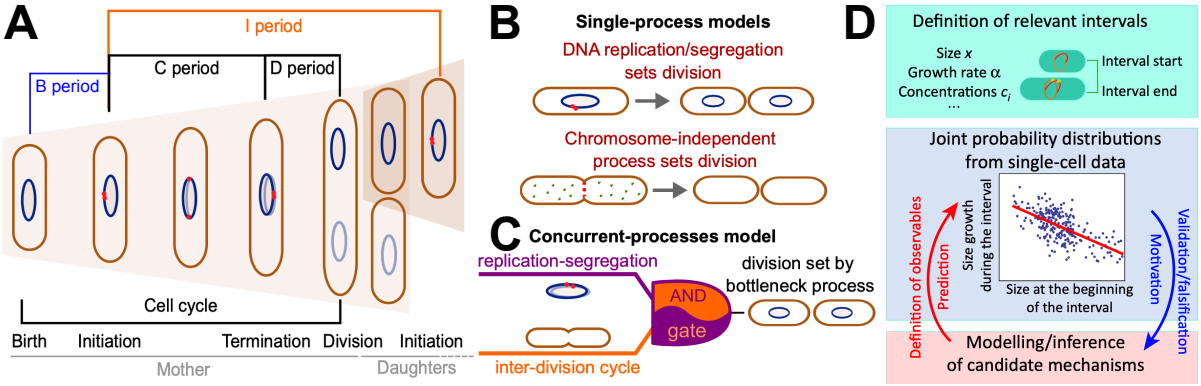


Figure 2.3: Different models have been proposed for cell cycle progression ***A*** *Cartoon of the cell cycle with different sub-intervals: birth-initiation (B period),initiation-termination (C period), termination-division (D period), initiation-initiation (I period).* ***B*** *Models of cell division control based on a single limiting process. Cell division can be controlled by DNA replication and subsequent segregation (top) or by a chromosome-independent inter-division process between birth and division (bottom).* ***C*** *According to the concurrent-processes model cells go through division when both processes (inter-division and replication/segregation) are completed. From [62].* ***D*** *Outlook for correlation analysis. At first, cell cycle intervals and their connection with global observables (e.g. size,growth rate...) are defined (top). Then joint distributions of measured variables are used to motivate a mathematical model (middle). Lastly, mathematical models define key observables and falsifiable predictions to test in the data (bottom). Reciprocal feedback between falsifiable models and data helps identifying mechanisms and molecular players of cell-cycle progression. From [63].*

have been suggested: a "constant" (size-uncoupled) duration since the time of DNA replication initiation (C+D period) [43, 48, 59], or the addition of a "constant" (size-uncoupled) size between replication initiation and division [11].

According to the second set of models, cell division is controlled by a chromosome-independent inter-division process between birth and division [10, 13, 51, 61, 64, 65]. In particular, the accumulation of a molecule or protein is thought to trigger cell division, once the copy number reaches a threshold level. Evidence comes from the observation that the size added by cells between birth and division is independent of their size at birth [66–68] and from experiments that show the independence of this "adder" behavior from perturbations of DNA replication [10]. Different "accumulator" molecules have been suggested such as cell wall precursor molecules [13], components of the divisome or septum [51], or the FtsZ protein [10, 65, 69]. However, whether cells effectively measure a constant size increase, or whether the adder behavior emerges through the accumulation of a single molecule, and/or whether chromosome replication/segregation have a direct influence on cell division remains controversial [10, 11, 51].

Lastly, the third class of models proposes that two processes limit cell division, DNA replication/segregation and a second "inter-division" process that relates cell size at division to cell size at birth, independently of DNA replication or segregation [8]. The inter-division process could be the accumulation of a molecule produced since birth, as summarized above. According to this "concurrent-cycles" model, the slowest process sets the timing of cell division at the single-cell level. Based on recent experimental evidence [10, 11], DNA replication initiation is controlled through an adder-like process between subsequent initiation events, which could also stem from a molecule accumulating during replication events [20, 59].

Sometimes differences between models are minimal and it is easy to fall into misunderstandings [10, 11, 70, 71].

We can conclude that probably cell cycle control involves multiple complementary and redundant mechanisms including multiple check points. Moreover, different mechanisms may be more or less important in different growth conditions.

2.4.2 Correlation analysis between cell-cycle intervals is used to validate or falsify existing models

Using a combination of microfluidics with time-lapse microscopy, in recent years researchers have been exploring to what extent the observed correlation structure of the fluctuations in the sizes and times at birth, initiation of DNA replication and cell division in single cells are consistent with the different models of the cell cycle. The precise mechanism in place is determined by examining the scatter plot of the amount of growth within a time interval versus the cell size at the entrance of the interval(Figure 2.3D) [9, 63].

For example if, as the Cooper-Helmstetter model assumes, initiation of DNA replication is triggered when the cell reaches a critical size, one would expect fluctuations in the size at initiation to be independent of fluctuations in the size at birth. However, recent works have shown that there is a clear positive correlation between size at birth and at initiation [8, 11]. In this way, a particular model of the cell cycle can be falsified by showing that the observed correlations are at odds with predictions of the model. An alternative model of the replication cycle assumes replication initiation is controlled by an inter-initiation adder, i.e. that cells add a fixed volume per origin between consecutive replication events. Indeed, both in [10] and in [11] it was observed that the added volume per origin fluctuates independently from the size at initiation, supporting this model.

Witz and coworkers stressed that, in order to meaningfully compare the evidence that the single-cell data provide for one or another model, one should com-

pare the full correlation structure of the data with the predictions of the different models [11, 71]. Indeed, they found that while the data are consistent with an inter-initiation adder model, the time between replication initiation and division in fact correlates with growth-rate, thereby falsifying a model that assumes division is controlled by a mechanism that acts on the time between initiation and division. Instead, their analysis shows that in the model which is most consistent with the full correlation structure of the data, division is controlled by an adder mechanism that runs from initiation of DNA replication to cell division (and not from birth to division).

One can also think about perturbing the system and looking at correlation variations. For example, Zheng et al. [72] showed that increasing cell width by the titration of the MreB-actin cytoskeleton causes an increase of the period between replication termination and cell division (D period) without affecting the average duration of DNA replication and of the cell-cycle.

Indeed, upon increasing the D period, cell size at division showed continuously decreasing correlations with cell size at initiation of DNA replication [62]. Without any modeling, these findings already suggest that cell division is controlled by a process different from DNA replication but dependent on cell size at birth. On the contrary, in non-perturbed cells, DNA replication appears to have an important limiting role, as supported by the high correlations between cell size at division and cell size at replication initiation, also observed previously by Witz et al. [11]. They also found that only the concurrent-process model is able to describe the experimental data in both perturbed and unperturbed condition, concluding that cell division is controlled by at least two concurrent processes that link cell division to DNA replication and cell birth.

The data presented until now relies on statistical methods such as binning and linear regression that are useful for interpreting data and generating hypotheses. Nonetheless, we have to be very careful, because predicting the relationships between experimentally measured quantities based on these methods might lead to misinterpretations [73].

For example, the existence of two subpopulations with different target sizes or with an overall increase or decrease in average cell size over the course of an experiment will introduce an artefactual positive correlation between birth and division size [74]. This can be referred to as the Simpson's paradox [75].

Moreover, the procedure of binning data in order to eliminate noise may lead to smooth curves that hint at specific functional relations between the two variables plotted that are inconsistent with the true functional relations. This may happen

because of "hidden" noise sources that affect the binning procedure and the phenomenon of "inspection bias" where certain bins have biased contributions [76]. Lastly, the often high intrinsic biological noise can obscure the mathematical relation between variables when not handled properly.

For all these reasons it may be necessary to have an underlying model (or models) that leads to similar data to that observed in the experiments to guide, test and validate data analysis methods [73].

2.4.3 Threshold accumulation of a constitutive protein can explain *E. coli* cell division behavior in nutrient upshifts

In this last section I want to briefly introduce part of a work I contributed to that has been published in 2020 [77]. This has been a side project for me but I think it fits perfectly in this framework. First, in this paper we use the comparison of data with mathematical models to support and / or falsify different existing mechanistic models for cell division. Second, it defines a data driven model based on mother machine data that has been extended to my own datasets (see Chapter 4.2.7). Lastly, this work describes a threshold accumulator model for cell division at the population average level in response to a change in growth medium, an upshift in growth rate. Following this framework, it will be interesting to extend this model to test a similar threshold accumulation model for initiation of DNA replication and cell division at the single cell level, within the cell cycle.

While most available studies try to infer information on cell division from steadily dividing cells in constant nutrient conditions, this study leverages on a high-resolution device to monitor single-cell growth division upon nutrient changes. Comparing these data with different mathematical models, we have been able to discriminate among fundamentally different mechanisms of cell division control, and we have shown that the data support a model where a protein expressed from a constitutive, unregulated promoter accumulates to a threshold level and triggers cell division.

We first tested whether some previously proposed models could reproduce the complex behavior of the division-related variables that we observed across the shift. To this end, we considered different models available in the literature. Specifically, we ran single-cell simulations of the models proposed by Harris and Theriot [13] ("relative-rates" model), Ho and Amir [59] ("incremental" model), and the classic idea of "initiation sizer" (which we implemented here with size-uncoupled C+D period, see [8, 9]]), using both modeled and sampled growth rate distributions across the shift. The relative rates model assumes that chromosome replication-

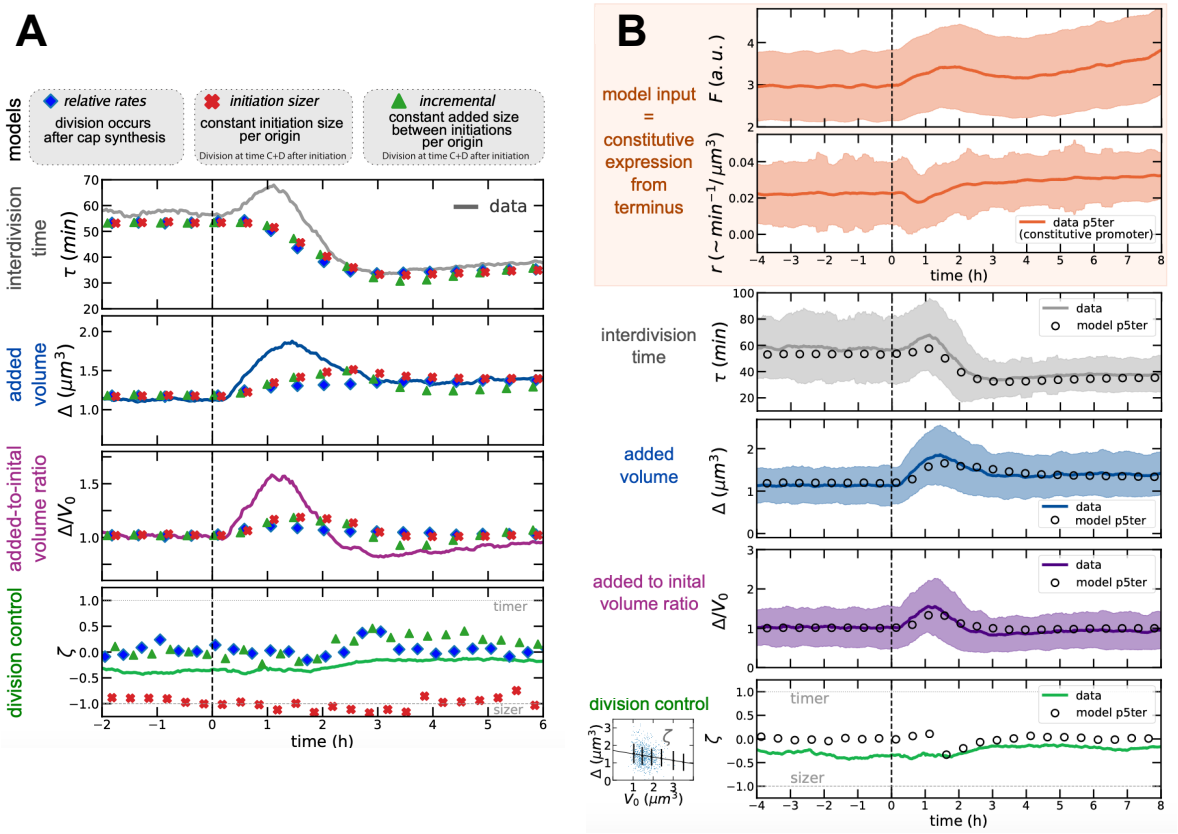


Figure 2.4: Falsification of models for cell-division control A *The nonsteady data across the nutrient shift falsify commonly assumed models for division dynamics. In particular, all the tested models are too slow in reproducing the added-size dynamics and do not reproduce the initial increase in interdivision times.* ***B*** *A putative divisor protein expressed from a consititutive promoter explains the shift data. As model input we used the measured instantaneous growth rates and volume-specific production rate (r) from a constitutive promoter inserted close to the replication terminus. The model predicts faithfully the size dynamic and the observed robustness of near-adder size control. From [78]*

segregation is never limiting for cell division [10], and that completion of synthesis of cap material (produced at the rate of surface synthesis) triggers division. Both the incremental model and the initiation sizer model assume that replication-segregation are limiting for cell division. We did not consider the more complex scenario of concurrent time scales [8] because of the unknown extra parameters used in this framework. The incremental model assumes that the chromosome is always limiting for cell division, and an inter-initiation adder (per origin) based on the cell current growth rate. By contrast, the initiation sizer model assumes a critical size per origin at initiation. None of the considered envelope- or chromosome-limited models for cell division predict correctly the division time dynamics (Figure 2.4A). Specifically, no model predicts that the interdivision time in the data initially

increases, before decreasing to its new steady state. Considering the dynamics of added size and initial size, all models show a delayed dynamics compared to the experimental data, and cannot predict the observed early overshoot in added volume. We figured that, instead, one may need a model where division dynamics is coupled to the mechanistic action of a biological circuit able to sense the physiological state of the cell. We therefore sought to define a mechanistic model where a protein under physiological control could act as the trigger defining cell division. In particular, we focused on a class of "threshold accumulator" models that have been proposed several times in the literature, both for DNA replication and for cell division [10, 20, 65, 79, 80]. We supposed that the fast and complex changes in division rate and added size observed during the shift in growth rate could be due to a coupling between the changes in biosynthetic "sectors" [81]]. The literature offers models that describe proteome sector dynamics and biosynthesis in non-steady regimes [82] and recent attempts were put forward to link these sectors with cell division [69, 83, 84]. However, the descriptions differ, and current data do not allow to select a specific one. Thus, rather than committing to a specific choice, we decided to take an experimentally driven approach to define our model. We believe that, giving our current limited knowledge on the crosstalks between cell growth and cell-cycle physiology, this "empirical" approach to forward models is preferable to more theoretical routes postulating the behavior of different proteome sectors across shifts [69, 83, 84]. Comparing the predictions of the four possible models that can be generated from our data, we show that the data are in agreement with a model where a putative adder protein is under the control of a constitutive promoter located close to the replication terminus of the *E. coli* chromosome. This model predicts efficiently the behavior of interdivision time, initial size and added size of cells across the shift. Conversely, if we reverse-engineer the production rate of an accumulator model in order to reproduce our division data, we find that it matches the expression pattern of the ter-proximate constitutive promoter.

From a molecular point of view, cell division (and its adder correlations) could be set by the accumulation of a putative "divisor" factor, possibly related to the FtsZ division ring [10, 65] or to the synthesis of the septum.

The similar near-adder behavior, also observed between consecutive initiations [10, 11] may derive from accumulation and trigger of an "initiator" protein setting replication initiation [20, 59, 80]. The initiation of DNA replication is known to be effected by a critical accumulation of ATP-bound DnaA. While some molecular mechanisms involved in DNA opening and replisome assembly have been identified, it is still not known how they contribute to setting the timing of initiation of

DNA replication in different growth conditions or how they can contribute to cell-cycle progression in single cells [21, 51, 55].

Chapter 3

The dynamic DnaA regulatory network within the cell cycle

Cells have evolved complex gene regulatory networks to produce appropriate amounts of proteins at appropriate times to adapt to ever-changing environments [1–3, 6, 7]. As we have seen in Chapter 2, efficient coordination of DNA replication, genome segregation and cell division are required for bacteria to adapt to changes in growth rate according to the availability of nutrients and in response to environmental changes and to stresses that can cause delays in DNA replication, such as DNA damage.

In this chapter I review the regulatory mechanisms affecting the expression of the DnaA protein, a key player in the initiation of DNA replication in bacteria that has also been shown to play a specific role in the coordination of DNA replication and gene expression [14–17]. This coordination is necessary for the cell to initiate DNA replication only once per cell-cycle and different models for the initiation of DNA replication have been proposed. DnaA is a regulatory factor essential not only in *E.coli* but also in the vast majority of known bacteria.

In particular, I will focus on the regulation of the *dnaA* gene itself, the special features of the *dnaA* gene's expression, promoter strength, and translation efficiency, as well as the role of DnaA protein, its binding to DnaA-boxes, and its binding of ATP or ADP. Furthermore, I will discuss in general how different factors can contribute to express a gene as a function of the cell cycle and how this could be very useful for the coordination of cell growth and division.

3.1 The roles of auto-regulation in transcription networks

The rate of synthesis of proteins in the cell can be regulated in many ways, at each of the different stages of production. One of the most common method in bacteria is at the level of transcription; modulating the rate at which the RNA polymerase enzymes puts together the long chain mRNA molecules from the DNA sequence of a particular gene. This process is regulated in part by proteins called transcription factors. Transcription factors act to activate or repress a particular gene, making the production of the mRNA from that gene more or less likely. The way in which this is done is that the transcription factor binds to a specific segment of DNA. In the case of repression, the transcription factor will often bind to a sequence near to, or even overlapping, the promoter sequence. In doing this, the transcription factor blocks the RNA polymerase, preventing the enzyme from binding to the promoter and transcribing the gene into mRNA. Most activators work by binding to a sequence upstream of the promoter and interacting with the RNA polymerase to increase its binding affinity to the promoter. This, in turn, leads to an increase in the expression of the gene [85–88].

There is no reason why a protein should not act as a transcription factor on its own gene. This is a means of regulation that is frequently found in cells, and is known as auto-regulation and can come in the form of auto-activation or auto-repression. Auto-regulating transcription factors are typically regulated by other TFs as well. In fact, many auto-regulating TFs in *E. coli* are known to respond to at least two additional regulators [89, 90]. In such cases, the response of the regulated TF to changes in the "input" TF concentrations must reflect an interplay between regulation and auto-regulation.

While simple networks have been studied quantitatively and successfully described by mathematical models, real life natural TF networks are more difficult to describe. The transcriptional network of *E. coli* is probably the best known for any cell, with 4405 ORFs identified [91] and an estimated 8%, or roughly 300 genes, as predicted or known TFs [92]. Despite our broad knowledge of the potential regulatory targets of *E. coli* transcription factors, our ability to predict regulatory behavior on the basis of their expression levels and that of their targets is not better than it would be for random networks[93]. In [94] the authors attribute this partly to the fact that even when a transcription factor is expressed, in many cases it will not bind its targets in the absence of additional signals [95]. In order to model the effect of regulation one needs to take into consideration that the natural timescales

to reach steady-state of gene regulatory systems is set by the expression rate and the decay rate of the protein product. In bacteria such as *E. coli*, where most proteins are stable, the timescale of the decay rate is set by the dilution due to the growth rate of the cell. For DNA-binding proteins one might also need to take into consideration the increase in the number of binding sites on the DNA during DNA replication.

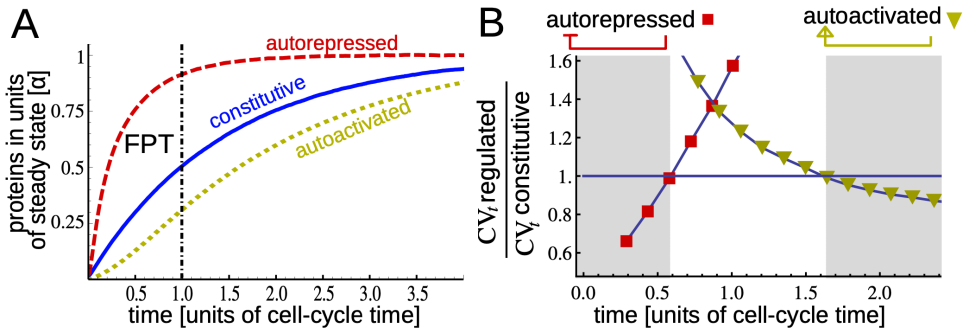


Figure 3.1: Roles of negative and positive auto-regulation within the cell cycle A *Negative auto-regulation allows for a faster response in gene expression. In this case the three regulatory strategies reach the same steady state α in an average time known as FPT (first passage time)* ***B*** *Gene expression noise as a function of time, 0 being the time of gene expression induction. When the average FPT is close to the the cell's doubling time for stable proteins both positive and negative transcriptional auto-regulations increase the timing noise. The two shaded regions show the FPTs for which self-regulation works as a timing noise filter. In particular, auto-repression reduces timing noise for time scales shorter than about 0.5 doubling times, while auto-activation is efficient in timing noise reduction for time scales longer than about 1.5 doubling times. From [95].*

In [96], it is reported that 59% of the TFs in *E. coli* are known to auto-regulate. Negative auto-regulation occurs more frequently than positive auto-regulation, but both are very common: 71 and 34 respectively are found in the current databases of 165 TFs [**sal**, 89]. Moreover, 9 TFs have binding sites for auto-activation as well as for auto-repression in their promoters. This suggests that auto-regulation may play important functions for cell physiology.

3.1.1 Negative auto-regulation

In *E. coli*, negative auto-regulation appears in over 40% of known transcription factors, probably because this is an extremely useful way for a cell to regulate the numbers of proteins in time. Auto-repression will have the effect of switching off the production of a gene when that gene's protein product reaches the concentration necessary to bind to its DNA sequence in its own promoter. This means that the level of that protein can be kept below a particular value. Negative auto-regulation

has also been shown to speed the response times of transcription networks allowing for a higher transcription rate (Figure 3.1A) [97], to reduce gene expression noise [98–100] and to induce oscillatory gene expression [101]. Delays can also help oscillations to arise. A delay in biological circuits can be achieved by adding components in the negative feedback loop to make longer paths in the circuit or by decreasing the translation and protein maturation rates. Despite this, the need to achieve rapid reversible responses in sensory transcription networks may help explain the finding that long cascades are very rare in the gene network of *E .coli* [102].

3.1.2 Positive auto-regulation

Contrary to negative feedback, positive feedback slows down the response kinetics of protein synthesis. In fact, an inherent response delay has long been recognized for positively auto-regulated systems [102–104]. This happens because a certain time interval is required to produce the TF to a level sufficient for activation [103]. The transcription rate will be slow initially, but when the protein number reaches the concentration of its binding affinity for its specific sites, it becomes more likely for the protein to bind to its sequence on the gene and perform its activation role, causing a marked increase in the production rate. While a slow response may not be desirable for many signaling tasks, the frequent occurrence of positive auto-regulation suggests that the cost in response speed can be overcome or tolerated. The introduction of delays is not the only role of positive auto-regulation. It can also increase the sensitivity to signals, produce a switch-like response when in competition with a repressor, and promote bistability [2].

3.1.3 Coupled positive and negative auto-regulation in the control of oscillations of gene expression

Although a single negative feedback loop has the ability to generate oscillations, in some situations a gene regulatory network has both positive and negative feedback loops. This additional positive feedback loop helps generate or amplify oscillation of gene expression created by the negative feedback loop and makes oscillation more robust against noise in gene expression by increasing the expression rate depending on a specific protein concentration threshold [105, 106]. The competition between positive and negative auto-regulation can then result in a step like change in gene expression. It has also been shown that positive feedback loops can provide a tunable period of oscillation [107]. The period can be changed by orders

of magnitude while keeping the amplitude almost constant by varying a reaction parameter in the positive feedback loop. This can be very important for cell cycle dependent oscillations, especially when cell cycle length varies, due for example to changes in growth rate. In contrast, a negative feedback loop without any positive feedback cannot change the period without compromising the amplitude.

Furthermore, it has been postulated that negative auto-regulation coupled to auto-activation can attenuate the inherent response delay caused by positive feedback [104]. For a system with only positive feedback, there is a trade-off between the steady-state expression level and response speed. A coupled negative feedback allows cells to maintain an optimal expression level as well as fast response kinetics. Coupled positive and negative negative auto-regulation has been shown to allow for a strong promoter for fast response without incurring cost of increasing protein expression levels [108].

3.1.4 Auto-regulation contributes to the regulation of gene expression as a function of the cell-cycle by decreasing the noise in the fluctuations

Sometimes an activated gene may be required to reach in a precise time a threshold level of expression that triggers a specific downstream process. However, gene expression is subject to stochastic fluctuations, naturally inducing an uncertainty in this threshold-crossing time with potential consequences on biological functions and phenotypes. These fluctuations might need to be controlled [95].

In particular,it is found that the role of auto-regulation in controlling the timing of fluctuation noise is context dependent. Positive auto-regulation reduces the timing fluctuations at short times compared to the cell cycle, while negative auto-regulation reduces timing noise at longer times (Figure 3.1B).

On the other hand, gene expression noise can result in bacterial population heterogeneity, offering specific advantages for fitness and survival in various environments. In fact stress-response genes have been shown to have a higher noise of expression than the core genes [109]. This trait is therefore selected during the evolution of the species, and is consequently regulated by a specific genetic network architecture [110, 111].

In the next section, I will describe the specific example of a gene, *dnaA*, whose expression is under both positive and negative auto-regulation.

3.2 DnaA is both a transcription factor and an initiator of DNA replication

DnaA as replication initiator

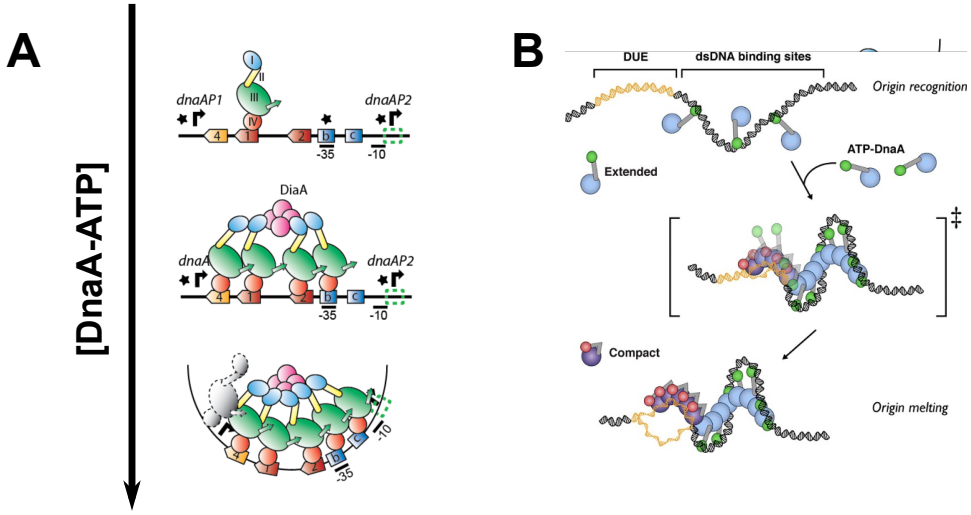


Figure 3.2: DnaA is a transcription factor and an initiator of DNA replication A *Steps leading to DnaA-ATP oligomerization at the* dnaA *promoter region. box1 is the highest affinity site and therefore the first site being occupied by a DnaA monomer, subsequently the other sites are bound with a stabilization of monomer-monomer interactions through the DiaA protein (pink tetramer); finally the entire promoter region becomes protected by the DnaA-ATP oligomer. The DnaA protein is represented by a ball and stick configuration where each domain has a different color.From [112]* ***B*** *At the origin of DNA replication the DnaA-ATP bound form (extended conformation, sky blue and green ball-stick) remains bound to high affinity dsDNA boxes; then in the presence of increasing DnaA-ATP it oligomerizes and binds to the low affinity binding sites. As the DUE (DNA unwinding region) melts, the DnaA molecules evolve from an extended to a compact state (violet and pink ball-stick). Finally the origin opens and the DnaA molecules in the compact state stabilize the ssDNA leading to open complex formation. From [113]*

E. coli has a relatively small circular genome of about 4.6 million base pairs and a single replication origin. The minimal oriC region (245 bp) has two functionally different regions: a duplex-unwinding region (DUE) with three repeats of the AT-rich 13-mer sequence where unwinding occurs during initiation, and a DnaA-oligomerization region (DOR) that comprises five DnaA protein binding 9-mer sequences (the DnaA box) [114]. In addition, in this region, the palindromic sequence GATC appears 11 times. As we will see later, these sequences are recognized by the

enzyme dam methylase and a protein called SeqA [115].

In *E.coli* the DnaA protein is a key factor in the initiation of DNA replication, but it is also an important transcription factor, including for its own expression (Figure 3.2) [15, 16, 116].

Both activities of DnaA can be modulated using an ATP-dependent molecular switch. DnaA can be found inside the cell tightly bound to ATP (Kd 30 nM) or bound to ADP (Kd 100 nM) [117]. Although DnaA has similar and extremely high affinity for ATP and ADP, DnaA is expected to bind preferentially to ATP in the cell because ATP is about six to seven times more abundant than ADP. Furthermore, although DnaA–ATP and DnaA–ADP can bind to the high affinity sites at the origin, only DnaA–ATP can lead to DnaA oligomerization and the opening of the AT-rich sequence (DUE) allowing for initiation of DNA replication [117]. As a transcription factor, the nucleotide switch has been shown to turn DnaA into a stronger repressor or stronger activator when it is bound to ATP [118].

An average *E. coli* cell contains 400–2000 DnaA molecules [119], depending on the growth rate, and the levels of the two nucleotide forms, ATP- and ADP-bound DnaA, are inverted during the replication cycle[15, 17, 120].

About 20–40 monomers of DnaA protein appear to be present in the oriC–DnaA complex (the initiation complex). The amount of the initiator protein DnaA in its ATP-bound, active form is important to trigger initiation of DNA replication. When it has bound ATP, DnaA has a relatively high affinity for the replication origin. When enough DnaA-ATP is present, the small protein HU binds along with DnaA, and together these proteins coil the DNA in such a way that the AT-rich initiation region, DUE, unwinds. Then the DnaB complex, which is the helicase, binds to the single-stranded DNA, and then the other proteins important for replication, such as primase, sliding clamp, and DNA polymerase, bind to form a complex with the helicase. Stimulated by the activity of the other proteins of the replication fork, DnaA hydrolyzes its ATP to ADP via an interaction with the Hda protein (Regulatory inactivation of DnaA, RIDA), and the DnaA protein dissociates from its binding sites . In this ADP-bound form, DnaA is unable to initiate another round of replication.

The binding of ATP in preference to ADP is then essential for the activation of DnaA and a DNA replication block has been recently observed at the level of initiation following ATP starvation [121]. In this case, replication forks were not arrested but new initiations were. Recent studies have shown that the initiation process is quite robust to changes in DnaA concentration [122, 123], suggesting that changes in (total) DnaA concentration may not change significantly the number of DnaA-ATP molecules per origin or the DnaA-ATP to DnaA-ADP ratio.

In Hda deletion strains, DnaA-ATP increases from 30 to 70% and a decreased initiation mass is observed [124]. Moreover, DDAH (datA-dependent DnaA-ATP hydrolysis) stimulates DnaA-ATP hydrolysis at the datA locus [125]. This DnaA inactivation mechanism is switched on once replication has started, thus decreasing the cellular re-initiation potential. RIDA is turned off upon termination of replication, allowing DnaA-ATP to accumulate in the cell and initiate a new replication cycle. Rejuvenation of DnaA-ATP from DnaA-ADP takes place at two DnaA Reactivating Sequences (DARS1 and DARS2) on the genome. The chromosomal locations of the DARS sites are optimized for the strict regulation for timely replication initiation. Deletion of both DARS sequences increases initiation mass to 150% [126]. The observed constant ratio of origin to cell mass at the time of initiation is therefore hypothesized to reflect the DnaA-ATP to origin ratio [127]. Another possible contribution can come from phospholipid mediated nucleotide exchange [128, 129].

Hemimethylated oriC DNA is a direct and immediate consequence of the duplication of oriC. Thus, oriC function is rapidly inactivated by the sequestration system after initiation. As the hemi-methylated form of oriC is maintained for about 10 min by the binding of SeqA to the GATC sites, another system must repress initiations once SeqA dissociates and the origin becomes remethylated by Dam. As the high level of DnaA-ATP before initiation is reduced to the basal level within 20 min after initiation, the RIDA system could take over after the lapse of the 10 min hemimethylated window to ensure that no extra initiations take place before the next scheduled initiation event. These two systems operate independently from one another [23] and presumably compensate for each other during the interinitiation time.

The datA site is a region of the genome where several DnaA proteins can bind. When the datA site is translocated from its original position on the chromosome to a site near the replication termination site, extra initiations are induced in rapidly growing cells[130]. This suggests that the timing of duplication of the datA site is important to titrate DnaA on the genome when the growth rate of cells is high. As the datA site is replicated after 8 minutes from the time of initiation at oriC (at 37°C), the datA function could reinforce the decrease in DnaA activity during the non-sequestrated oriC period. As such, a common and very important feature of these three systems seems to be their functional coupling with the chromosomal replication process.

Sequestration requires the hemi-methylated form of oriC, a direct consequence of origin replication. RIDA depends on the sliding clamps formed upon DNA loading of the replicase. The DnaA titration system can be linked to the timely duplication

of the datA site. Thus, as a principle, the negative regulators of initiation appear to form a feedback circuit in the replication processes[120].

To resume, knowing that about 20 DnaA molecules must bind to oriC for initiation of replication to occur, changes in the abundance and activity of DnaA molecules are necessary to achieve precise cell cycle–coupled initiations. DnaA is normally extremely stable with a half-life exceeding 24 h during steady-state growth [131]; thus, most control is achieved by cell cycle fluctuation in *dnaA* gene expression and DnaA protein activity via its nucleotide bound state.

In Figure 3.3 we can see how several different processes are coordinated in time (left) and/or in space (right) to regulate DnaA's activity.

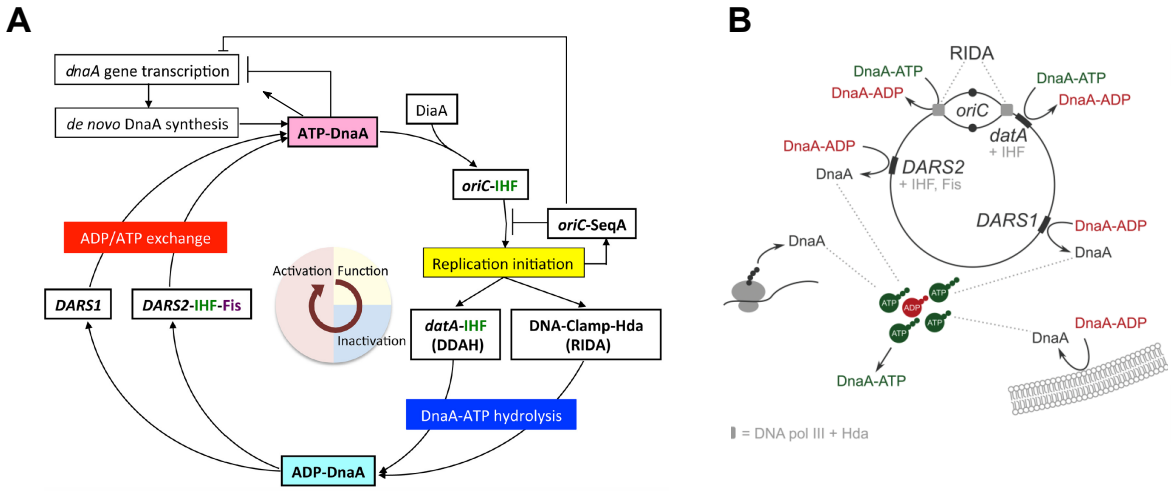


Figure 3.3: The regulatory cycle of DnaA: activation function and inactivation (A) *ATP-DnaA forms oligomers on oriC with the aid of integration host factor (IHF) and DnaA interacting protein DiaA, and initiates replication (period in light yellow). After initiation, ATP-DnaA is converted to ADP-DnaA (RIDA) and datA-dependent DnaA-ATP hydrolysis (DDAH) systems (period in light blue). RIDA involves Hda protein and the clamp subunit of DNA polymerase III holoenzyme, and DDAH involves the datA locus and IHF. DnaA-reactivating sequence (DARS1 and DARS2) loci regenerate ATP-DnaA (period in light red). IHF binds to specific sites in oriC, datA, and DARS2 in a cell-cycle-coordinated manner. Fis binds to DARS2 in log-phase cells. From [16]* ***(B)****Same processes of (A) with a focus on where these processes act in the genome, with differences pre- and post- initiation of DNA replication. From[132]*

DnaA as a transcription factor

DnaA was found to bind the promoter regions of several genes and to regulate their transcription. Depending on the location of DnaA boxes relative to the promoters, the DnaA protein can function either as a transcriptional repressor an activator or both [14, 114, 133–137]. In particular, DnaA has been shown to act as

a repressor or activator for a set of genes involved in regulation of DNA replication depending on the concentration of DnaA-ATP. Other functions of the products of the genes regulated by DnaA includes nucleotide biosynthesis, carbohydrate metabolism, iron homeostasis, amino acid biosynthesis and ribosome biogenesis [138]. Finally, DnaA also regulates the expression of DNA repair enzymes so that it may be proportional to the amount of DNA in the cell to efficiently clear the way for the replication forks to reach completion of the genome synthesis[139–141].

As an activator, DnaA regulates the expression of the *fliC* gene, encoding flagellin, a subunit of *E. coli* flagella [142], *polA*, coding for DNA polymerase I, involved in DNA replication and repair and the *nrdAB* operon, coding for the ribonucleotide reductase enzyme [14, 143]. DnaA overproduction was found to increase *polA* expression in stationary-phase cultures in a RpoS independent manner [133]. In this case the mechanism of transcription activation by DnaA is unknown, however it has been suggested that activating proteins with binding sites near the promoter can cause activation by direct contact with RNA polymerase [87].

As a repressor, DnaA regulates the transcription from the *mioC* promoter, which has been proposed to inhibit initiation of DNA replication [144]. Binding of DnaA upstream of the *rpoH* promoters reduces the transcription of this gene that codes for $\sigma 32$, the primary sigma factor controlling the heat shock response during log-phase growth [137]. A similar situation is found in *uvrB*, which is a subunit of the nucleotide excision repair (NER) complex [141, 145]. Lastly, DnaA represses nrdAB expression [146]. All these genes have DnaA boxes within or upstream of their promoter region.

More importantly, DnaA can act both as a repressor and an activator of its own promoter, depending on the binding of DnaA-ATP on the promoter [112]. We will look at this in detail in the following section.

3.3 Activation and inactivation of the DnaA protein

To better understand the regulation of DnaA activity and that of its promoter we have to introduce briefly the structure and function of the DnaA protein.

E. coli DnaA is a 52 kDa basic protein that has four distinct functional domains and belongs to the AAA+ protein family (ATPases Associated with various cellular Activities)(Figure 3.4)[147, 148]. Different domains have specific roles in DnaA function. Domain I is used primarily for protein-protein interactions. A hydrophobic face of this domain allows DnaA protomers to oligomerize, while the opposite side of the domain interacts with the replicative DNA helicase (DnaB) and with DiaA, a

modulator of the pre-replication complex (pre-RC) assembly. Domain II forms a flexible linker that may accommodate the need to position Domain I over the wide range of distances that separate DnaA recognition sites within OriC, or allow some flexibility in positioning DnaB on single-stranded DNA. Domain III carries the motifs necessary for ATP binding/hydrolysis and oligomerization. Domain IV contains a helix-turn-helix motif for double-stranded DNA binding, as well as an amphipathic region used for membrane interaction. Domains III and IV are the most highly conserved among eubacteria.

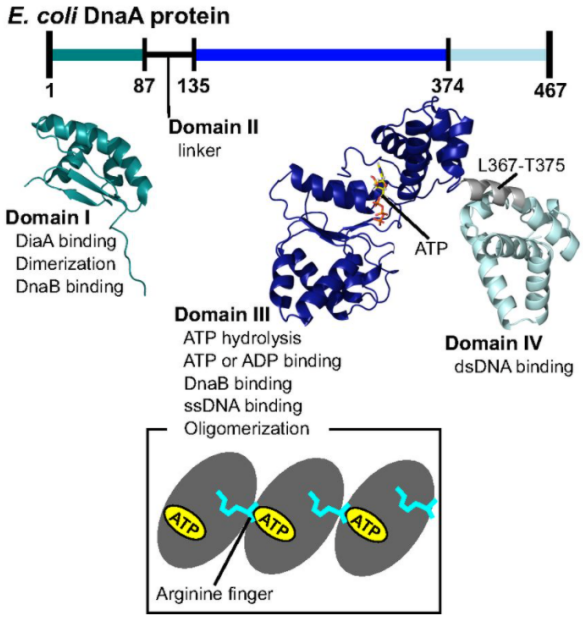


Figure 3.4: Domain architecture of the DnaA protein. *DnaA contains four domains. The structure of Domain I was obtained by NMR (Protein Data Bank 2E0G), and the structure of domain III-IV is a homology model. (Bottom) DnaA domain III oligomerizes in a head-to-tail manner. The ATP-bound side of one molecule contacts the arginine-finger side of another molecule. From [149]*

.

Like other members of AAA+ protein family, the binding of ATP and ADP specifies the conformation and multimerization status of the protein.

3.4 The *dnaA* promoter region and the *dnaA* gene

The *dnaA* promoter region was first identified and characterized in 1979-1980 [150]. The entire promoter region (Figure 3.5) is 606 bp long and it regulates the expression of several proteins involved in the initiation of DNA replication. This region has two promoters, dnaAP1 and dnaAP2 (-317bp and +152bp from the transcription start site), separated by a 53-bp region that contains five binding sites for DnaA

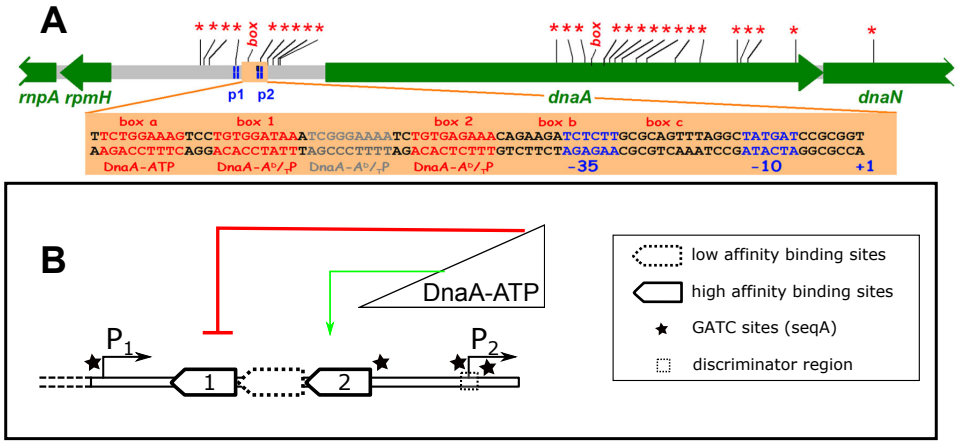


Figure 3.5: Schematic of the promoter region of the* dnaA *operon in* Escherichia coli *A *The GATC sites bound by SeqA are shown by an asterisk (*) and the consensus DnaA boxes are shown in a box. The position of the* dnaA *promoters P1 and P2 are indicated. The light orange rectangle shows the sequence of the P2 promoter and potential DnaA boxes upstream of and overlapping the promoter (adapted from [15]).* ***B*** *Schema of the main players in the regulation of the DnaA promoter: DnaA-ATP binds to the high affinity binding sites, GATC sites for SeqA binding and the P2 GC-rich discriminator region.*

[150]. Box1, box2 and the region in between are protected from nuclease cleavage both by DnaA-ADP and DnaA-ATP, the boxes a, b, and c require DnaA-ATP.

The DnaAP2 promoter region presents several similarities with stable RNA and ribosomal protein promoters. It presents a -35 region that deviates significantly from the consensus sequence while the -10 region is very similar to the consensus one [151]. There is also a GC-rich sequence overlapping with the transcription start site, called a “discriminator” sequence, which confers sensitivity to the effector of the stringent response, the small metabolite ppGpp and suggests that the expression of DnaA could be influenced by the levels of DNA negative supercoiling, which vary following changes in the growth rate and in response to stresses to the replication machinery [22, 151]. DnaAP1 provides a basal level of transcription and it is not subjected to specific regulation, while dnaAP2 is a threefold stronger promoter contributing 60-80% of the dnaA transcripts in exponential phase[22, 134, 135, 152]. The dnaAP1 promoter, instead, is induced mainly upon entry into stationary phase. This could be explained by the previously observed increase in the amount of ppGpp in the transition from exponential to stationary phase, resulting in the inhibition of expression of stable RNA-like promoters, such as dnaAP2, and thus in the increase in the amount of RNA polymerase available in the cell for transcription of weaker promoters.

Moreover, the *dnaA* promoter region presents a divergent promoter configura-

tion where the promoter regions of the *rpmH* and the *dnaA* genes are positioned back to back. Past works in our lab by Bianca Sclavi and the former PhD student Chiara Saggioro have shown that the divergent promoter configuration dnaA-rpmH contributes to a decrease in the expression from the *dnaA* promoter. The *rpmH* gene codes for a ribosomal protein and contains the characteristic promoter organization of the stable rRNA promoters [151, 153].

To conclude, the *dnaA* gene presents a complex regulatory motif characterized by several levels of regulation. In Chapter 6 I will investigate the role of the auto-regulation by DnaA and repression by SeqA in the robust progression of *E.coli* cell-cycle by considering *dnaAP* mutants with mutation in the binding sites necessary for DnaA and SeqA regulation (Figure 3.5B).

3.5 Positive and Negative Auto-regulation by DnaA

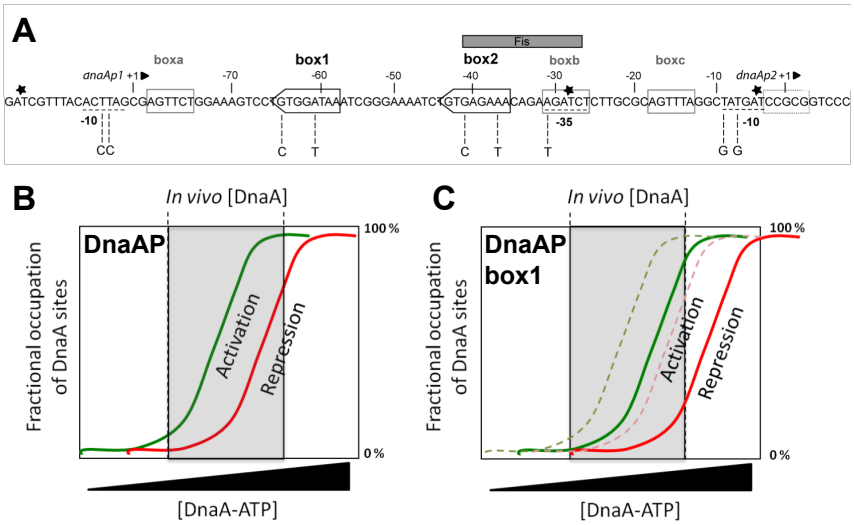


Figure 3.6: DnaA boxes have different affinities for DnaA and play a role in auto-regulation A *The site-directed mutations in the DnaA-binding sites used to study auto-regulation are shown. Mbox1, A-85G and A-86G+T-60A and C-64G; Mbox2, A-85G and A-86G+T-37A and C-41G.* ***B*** *Fractional occupation of DnaA sites in the* dnaA *wild type promoter region at increasing concentrations of DnaA-ATP. The green curve represents the occupation of the high affinity sites, leading to activation of DnaA expression. After DnaA-ATP oligomerization on the entire promoter region at higher concentrations, the repression can take place (red curve). The gray box represents the window of the proposed* in vivo *available DnaA concentration.* ***C*** *After mutation of the box1 DnaA site, a shift of the binding curves is observed, thus occupation of the DnaA sites occurs at higher protein concentration, where activation is still possible in the range of* in vivo *DnaA concentration, but repression can no longer occur. From [112]*

The ability of DnaA to autoregulate its own transcription was discovered at a time when the direct role of DnaA in chromosome replication was not yet clear. Early characterizations of temperature-sensitive *dnaA* mutants of *Escherichia coli* revealed that the copy number of ori decreased, while DnaA levels increased simultaneously at non permissive temperatures, thus accurately predicting that DnaA had a positive involvement in the onset of chromosome replication and a negative involvement in regulating its own synthesis [154]. DnaA's ability to repress its own transcription was later confirmed using *in vitro* and *in vivo* analyses. *In vitro*, DnaA was shown to directly bind its own promoter region using gel shift assays and DNaseI footprinting methods [112, 135, 136, 148]. *In vivo*, two strategies were used to show that DnaA represses its own transcription. First, increased cellular levels of DnaA (expressed from inducible promoters) were shown to decrease the activity of the *dnaA* promoter, and second, a decrease in the levels of DnaA able to bind the *dnaA* promoter (by adding DnaA binding sites on a plasmid that titrated DnaA levels away) was shown to increase the activity of the *dnaA* promoter [134, 135, 152]. In Figure 3.5A we see a representation of the *dnaA* promoter region. Between the two promoters there are different binding sites for the protein itself. The first characterization by DNase I fooptrinting established two DnaA boxes, specific for both ADP and ATP forms, and three DnaA-ATP specific boxes [118]. Afterwards, Hansen et al. proposed that the number of DnaA boxes is probably underestimated and that two additional 9-mer sites, with two misfits in the consensus sequence, are present in the promoter region. One box located between box 1 and box 2 and one located upstream of the box 1, all boxes in the same orientation and properly spaced [155]. Also previous works in my laboratory support this second arrangement [112]. In this study, my supervisor Bianca Sclavi and the former PhD student Chiara Saggioro carried out systematic mutagenesis of the different DnaA binding sites present on the *dnaA* promoter and measured their effects both *in vitro* and *in vivo* in order to gain an improved understanding of the role of the complex regulation of DnaA expression. Mutations are summarized in Figure 3.6A. By the use of a quantitative *in vitro* DNase I footprinting assays, they measured how mutations in the *dnaA* promoter change the affinity of DnaA to its binding sites. In parallel, they used a quantitative approach *in vivo*, using a GFP reporter fluorescence based assay, where the *dnaA* promoter controlled the expression of the *gfp* gene on a medium-to-low copy plasmid.

Figure 3.6B-C summarizes the results obtained *in vitro* in the context of autoregulation of DnaA expression. The binding curves for activation and repression of transcription are drawn in green and red respectively and the dotted grey box in

the graphs represents the *in vivo* concentration of available DnaA in the cell. In this window of concentration DnaA binds to its sites and activates transcription from its own promoter; at higher concentrations it oligomerizes thus repressing transcription (red line). When box1 is mutated, the binding curve is shifted to higher concentrations, DnaA can still bind to box2, thus activation of transcription *in vivo* is still possible, but at higher protein concentrations than for the wild type promoter. The binding curve resulting in repression is also shifted to higher concentrations that are however outside the window of *in vivo* DnaA concentrations. For this reason an increased expression from the dnaAP2 promoter is observed when mutating box 1 *in vivo*.

The presence of activation by DnaA at the dnaAP2 promoter can counteract the negative effects of inhibitors such as ppGpp or Fis [151, 156]. For example, auto activation has been shown to allowed continued transcription in the presence of rifampicin, an inhibitor of transcription initiation [157]. Positive auto-regulation may thus allow the *dnaA* promoter to be expressed differently compared to stable RNA promoters, notably stabilizing RNAP binding in the presence of ppGpp.

Positive and negative auto-regulation can take place within the same cell cycle due to the oscillations in DnaA-ATP. Auto-repression is thought to take place when the DnaA-ATP levels have reached their maximum after DnaA-ATP has filled the low affinity sites at the origin at the time of initiation of DNA replication.

As described above, negative auto-regulation not only contributes to the more stable maintenance of a maximum concentration of a given protein, but it also allows for a faster expression rate compared to an unregulated promoter to achieve the same amount of protein [37, 97].

Thus while the average rate of GFP expression is similar in the presence and absence of auto-regulation [112] we would expect that the cell cycle dependent kinetics of expression should differ.

Auto-regulation of DnaA expression is then important to limit the range of available protein that must be always present in the cell and to respond quickly to changes in the growth rate.

All these results have been obtained from a plasmid. In Chapter 6 we will confirm these results placing the GFP gene under control of the same promoter mutants in the chromosome and in particular we will look at the cell cycle dependent kinetics of expression.

3.6 DnaA and oriC regulation by SeqA sequestration

The average density of GATC sites in a DNA sequence containing equal amounts of the four bases is one in 256 bp. The density of GATC sites in the *dnaA* promoter region is approximately ten times higher. Nine GATC sites are present in the *rpmH-dnaA* promoter region, three of them downstream the transcription start site of dnaAP2, two overlapping with the -10 and -35 regions of dnaAP2 and four upstream of dnaAP1. Those at the -35 and -10 regions are 100% conserved among closely related bacteria[158]. However, mutants of the GATC sites on the *dnaA* promoter have so far failed to show a specific phenotype [158]. Several GATC sites are also found within the *dnaA* gene reading frame.

A protein called SeqA binds specifically to hemi-methylated GATC sites. Hemi-methylated DNA is produced by the replication of the DNA and it is thought to remain so for about 10-15 minutes until the activity of Dam methylase methylates the newly synthesized DNA strand (adenine methylation). SeqA has an N-terminal domain important for oligomerization connected with a flexible linker to the C-terminal domain responsible for the interaction with the GATC sequences.

SeqA binding sites are abundantly present in OriC and several of them overlap with weak DnaA boxes[159]. The binding of SeqA can then contribute to the proper timing of initiation of DNA replication by inhibiting both binding of DnaA to the origin and production of DnaA right after initiation of replication has taken place.

Eventually, SeqA spontaneously dissociates from oriC. Free SeqA binding sites are then methylated by Dam methylase, which blocks further SeqA binding [160, 161]. It has been shown that the GATC sites were sequestered in much shorter periods when the *seqA* gene was mutated or absent [162, 163]. These fully methylated origins are once again able to initiate a new round of replication when sufficient DnaA-ATP has accumulated (Figure 3.7).

The period where reinitiation is not possible is the so called “eclipse” period and it typically last 1/3 of the cell cycle allowing for the levels of DnaA-ATP to decrease below the critical level for initiation.

In addition to preventing duplex DNA at the origin from opening through inhibition of DnaA binding, SeqA can also affect DNA topology directly [165]. SeqA dimers were shown to introduce positive supercoils that could decrease the tendency of the DUE to unwind and thereby also influence replication initiation frequency [165]. SeqA filaments, in addition, were shown to prevent the formation of negative supercoils [165]. In *dam* mutant cells, there is no sequestration by SeqA; consequently, DnaA can immediately rebind to the origins after the first initiation event, and ini-

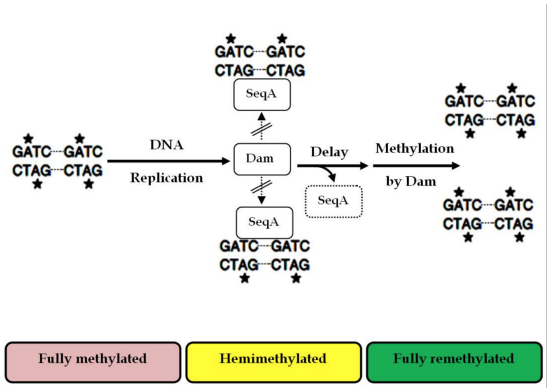


Figure 3.7: DnaA promoter sequestration by SeqA *Two or more GATC sites can be bound by SeqA when they are in the hemi-methylated state. Binding of SeqA inhibits DnaA binding, RNAP binding, transcription of the promoter and Dam methylation, maintaining the hemi-methylated state for a portion of the cell cycle. Dissociation of SeqA allows Dam to methylate the hemi-methylated DNAs, generating fully methylated DNA. From [164]*

tiate a second time when the concentration of the active form of DnaA-ATP is high enough. Transcription from the *dnaA* gene continues throughout the cell cycle although at a reduced level [166]. Dam methylation, therefore, is not essential for replication initiation; rather, the cell uses methylation to discriminate between old and new origins. Also for this reason, SeqA activity becomes important when initiation of DNA replication is induced after DNA damage by UV light [167].

In Chapter 6 I will present the results obtained on the expression of the DnaA promoter where the 3 downstream GATC sites have been mutated.

Chapter 4

Robust and long-term single-cell tracking in steady growth condition

4.1 Introduction

As we have seen in Chapter 2, in the last decades the technological development in high-throughput microscopy, imaging, image analysis, and microfluidics has been critical for the study of single-cell bacterial physiology [63, 168, 169].

An important part of my PhD has been dedicated to the design and testing of a new experimental setup and data analysis pipeline for studying the growth, size, and gene expression of *E. coli* in controlled environmental conditions. In this Chapter, I will introduce the integrated microfluidics and time-lapse microscopy approach I have used to quantify growth, division, and protein expression at the single cell level within the cell cycle and across generations.

The novelty of our experimental setup resides in the use of a new pressure-driven flow control system that improves the quality and the duration of the experiments. At the moment we are one of the first laboratory to use this system and we also had the opportunity to write an application note in the website of ElveFlow describing the main advantages of this experimental setup [170].

4.2 Results

4.2.1 A two-ended microfluidic device

Different techniques have been developed in order to study organisms at the single-cell level. The most popular technique to perform single cell microscopy is using agar pads. This is mainly because of its simplicity and relatively inexpensive

set up. These devices however are not suitable if we need a strict environmental control setup that can stabilize the system for longer time scales and, more importantly, if we need to change growth medium quickly and easily without perturbing the system. Balanced growth is difficult to achieve in bacteria growing in flasks or on an agar pad, as the nutrients at a certain point run out and metabolites can accumulate decreasing the pH and slowing down growth. The solution was met with the introduction of microfluidics. Even in microfluidic devices, different designs have been suggested and manufactured depending on the nature of study.

Thanks to our collaboration with the laboratory of Kevin Dorfman (University of Minneapolis, Minnesota, USA), we have developed different kinds of microfluidic devices that allow to attain balanced cell growth in constant environmental conditions for several hours and then switch the growth medium to observe the changes taking place as the cells adapt [171, 172]. Devices with different geometries can present advantages or disadvantages depending on our experimental needs (Supplementary Figure F.3). We developed these PDMS(Polydimethylsiloxane) devices from the mould by standard procedure using a plasma cleaner. After the devices have been prepared, they are stored and are good to use for about a month, following which collapse of structure is often noticed on the devices. The protocol is reported in Appendix C.

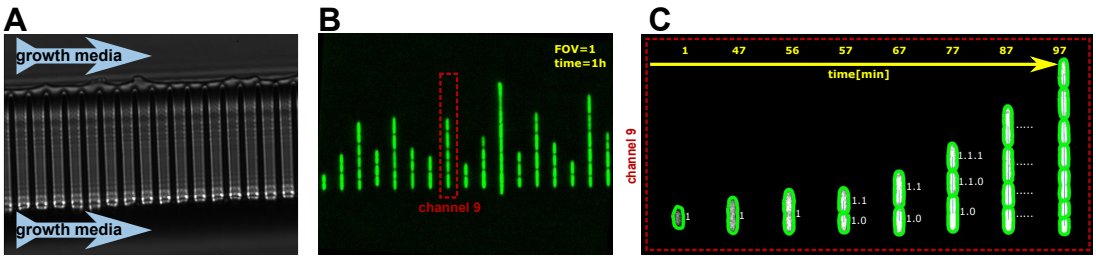


Figure 4.1: A two-ended mother machine A *A two-ended mother machine presents 2 large feeding channels and 2000 narrow channel to trap bacteria* ***B****. The narrow channels are 1 micrometer wide and are open at both ends (thus the name of the device). The lower end is tapered to avoid bacteria escaping but allowing for the growth medium to flow through the channel. This can help with cell loading and maintaining a constant environment* ***C*** *The progeny of the 'mother' cells trapped at the bottom is segmented and tracked.*

The mother machine device was introduced by the Jun lab in 2010 [173, 174], and in the last years has been extended to other organisms such as yeast or mammalian cells [175–177]. In our mother machine there are 2000 parallel channels that are 25 µm long, 1 µm wide and 1.1 µm deep; channels are tapered at one end [172], (Figure D.1A). Channels need to be in a rather narrow range of dimensions for not to display growth-limiting effects[178]. These small channels are flanked by

2 larger channels (50 µm wide and 20 µm deep) where the growth medium flows and brings nutrients to the bacteria at the same time as washing away waste and metabolites. Following loading, some cells enter a narrow channels and become trapped at its narrower base. The progeny of these 'mother' cells then proliferate up the narrow channels until they protrude into the wide channel and flow away with the medium (Figure D.1B). This allows for the bacteria to grow in a constant environment to achieve what is called balanced growth. Such proliferation continues for hundreds of generations while cells are imaged, and thus steady-state high-throughput data are acquired (Figure D.1C).

4.2.2 Pressure-driven flow controlled microfluidics

Usually syringe pumps are used to push the growth media inside the device [78, 172]. I used syringes for the first months and a drawback is that it is difficult to maintain a constant flow rate inside the device. Microfluidic chip resistance can be very different between one chip to another. Moreover, during the experiment the resistance can increase due to the formation of clogs (dust, pdms' fragments..). For this reason, it is very important to have a pressure-driven flow to have fresh growth medium flowing in the device continuously at constant speed. Elveflow® flow control system provides an innovative microfluidic flow control system that uses patented piezoelectric regulators to enable a fast and precise flow control. In this way it is easier to have good reproducibility and long lasting experiments. Moreover, sometimes it can be useful to change growth medium quickly to measure the cell's changes during an adaptation process.

The entire microfluidic setup (Figure 4.2) is composed of a pressure-driven flow controller (0 – 2 bar pressure range), two rotary valves (11-port/10 way Mux Distributor), two flow sensors and a microfluidic chip, ensuring a constant environment for long time experiment and allows to quickly change between growth medium or temperature. Additionally, the PID feedback loop permits an effective control over the flow rate while keeping the stability and responsiveness of pressure-driven flows. These experiments can last up to 72 hours in order to allow us to observe some of the slower adaptation mechanisms and in order to include several switches between different concentrations of nutrients. Thanks to the efficient interface, we can easily check the experiments remotely and this can be very useful particularly for experiments that can last more than 1 day.

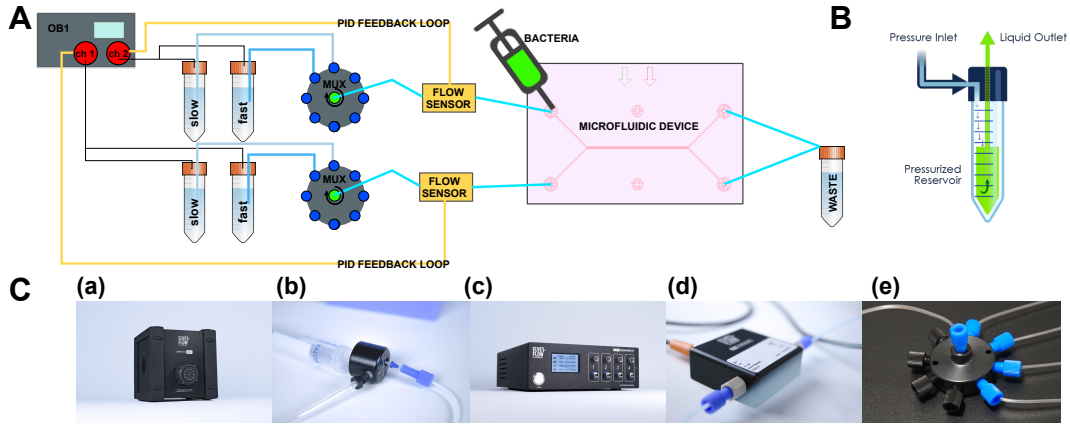


Figure 4.2: Schematic representation of the microfluidic setup A *The experimental setup used for upshift experiments.* ***B*** *Pressurized reservoir: the fluid flows through the outlet if gas pushes on the fluid surface.* ***C****, List of components: Mux Distribution (a) , Tubing, fittings and reservoirs (b), Flow controller OB1 Mk3+ (c), flow sensor (d), Manifold 9 ports (e).*

4.2.3 Reference strains and culture conditions

Our experiments have been carried out on the wild-type *E. coli* strains BW25113. BW25113 is the parent strain of the Keio collection and has been fully sequenced [179].

All the promoters analysed in this work express GFPmut2, a fast folding fluorescent protein [180, 181]. This protein exhibits bright green fluorescence when exposed to light in the blue to ultraviolet range, so was used as biological marker: the "fluorescent strains" absorb light at $\approx 395nm$, then emit it at $\approx 509nm$. The fluorescent protein GFPmut2 folds rapidly: half of all proteins mature in roughly 5 minutes and 90% in just less than half an hour. A kanamycin resistance cassette (KanR) is found upstream of each promoter-GFP construct as the selectable marker.

As a reference, we also considered reporter constructs obtained with two promoters used in a previous study [77], both placed at the same origin-proximate locus (4413507 bp) and at a terminus-proximate locus (1395706 bp). The phage-derived constitutive promoter P5 has a high affinity for RNA polymerase and lacks any transcription factor binding sites, thus the GFP production rate from this construct is largely proportional to the gene copy number in the cell. The expression from the ribosomal reporter P1, derived from the promoter of the ribosomal RNA operon rrnB, is also not regulated by transcription factors, but depends on ppGpp and supercoiling levels, due to the presence of a G-C rich discriminator region overlapping the transcription start site. In particular, we use a full-length ribosomal RNA promoter, rrnBP1, here called P1long, and an upstream-shortened version of

rrnBP1, P1short, where the binding sites for FIS and the high affinity-binding site for H-NS have been excluded. Differently form ribosomal promoters, the growth rate dependence of a constitutive promoter is not proportional to the number of ribosomes, it is limited by the transcription rate, and the gene product is diluted at faster growth [182].

We used two growth media based on the M9 minimal medium as the base and glucose as the carbon source. More precisely, the slow medium is M9 with 0.4% glucose. The fast medium in addition has 0.4% casamino acids. BSA can be added to the media for bacteria growing in a microfluidic environment because it prevents the adhesion and subsequent aggregation of cells along the walls of the device. More details about growth media composition are reported in Appendix D.

A home built temperature control system is used to maintain the entire setup at 30°C but it supports temperatures in the range 20-40°C. We choose to perform experiment at 30°C because in literature there are many other studies at this temperature and it would be then possible to compare them with our results.

4.2.4 Optical microscopy, image acquisition and image analysis

We use a Nikon Inverted Microscope ECLIPSE Ti-E with 100X oil objective high NA (Numerical Aperture) lens, coupled with a Nikon Perfect Focus System (PFS) to rectify drift in focus. A x-y motion plate is used to memorize and loop over different ROIs (Region Of Interests) at specified intervals of time. To avoid shifts of the device during data acquisition a support has been designed and printed with a 3D printer provided by a research group from Sorbonne Université.

Thanks to the automated x-y stage and the perfect focus system we can acquire data on thousands of cells within the same experiment. In this experimental setup, fluorescence images are taken with light from a blue LED passed through a GFP filter. Different strains of bacteria may express GFP at different levels. Henceforth it becomes necessary to adjust the exposure time and acquisition parameters for different strains in order to obtain enough fluorescence and simultaneously not saturate the pixels. Moreover it's very important not to expose the bacteria to blue light during the first few hours when the bacteria are adapting to the microfluidic device.

Movies obtained from the microscope are in the .nd2 format and can be opened and post-processed using the Fiji software. I wrote ImageJ macros to process the .nd2 videos and to convert them in tiff files necessary to proceed with the segmentation and tracking steps. In particular, only frames with a good focus are selected and device drift or rotation are corrected. Background subtraction is then

performed using a 50-pixel rolling ball protocol and different positions (FOV) are stored separately as sets of tiff image files. A folder is created for each FOV. For each FOV, channels are selected manually as region of interest (ROI) and stored in different folders as a sequence of tiff files. The final result is a folder for each experiment containing a folder for each FOV in turn containing a folder for each channel, containing a sequence of tiff files (one for each time frame). This is the input for the next steps: segmentation and tracking.

4.2.5 Segmentation and tracking

Automated image segmentation and tracking are important steps in the analysis process. This process will result in a set of segmented contour cells over time whose label give us information about their genealogy (Figure 4.3). In order to perform segmentation and tracking on mother machine data, we use the codes developed by Mia Panlilio, Cambridge University [78]. Necessary modifications were made based on our experimental setup. An effort has been made to create an easy-to-use pipeline that goes from raw videos to a list of segmented and tracked cells.

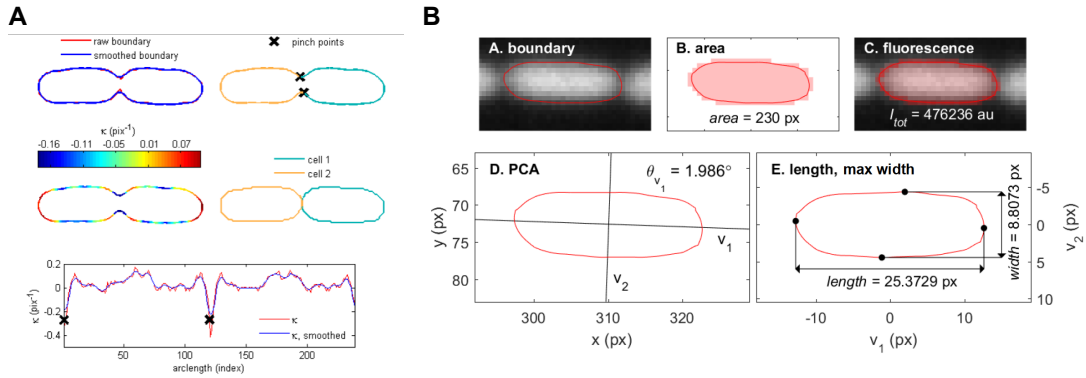


Figure 4.3: Segmentation and tracking algorithm for images obtained with the mother machine A *Cells are boundary coloured according to local curvature with regions of strongly negative curvature (dark blue) indicating pinch points (black cross) at which to further segment cells.* ***B*** *Boundary coordinates overlaid on the original grayscale image (a). Total cross sectional area calculated as the area enclosed by the boundary, including edge pixels (b). Total fluorescence intensity associated with the enclosed area (c). Principal component analysis on boundary to determine axis of orientation (d). Cell presented in principal axis space, where length is the maximum distance between boundary points along the first axis. Along the second axis, maximum width is also shown(e). From [78]*

I will now report briefly how the segmentation and tracking algorithm works. More details can be found in [78]. Functions marked by "†" refers to built-in Matlab

Notation	Name	Unit	Description
l	Length	μm	Greatest pairwise distance between boundary points along the first principal axis (found by pca†)
$Area$	Area	μm^2	area enclosed by boundary points, including edge pixels
w	Width	μm	width assuming the projected area is a rectangle with semicircular caps of radius $\frac{w}{2}$
V	Volume	μm^3	$V = \pi(l - < w >)(\frac{w}{2})^2 + \frac{4}{3}\pi(\frac{w}{2})^3$
F	Fluorescence	$\mu a.u.$	Total fluorescence intensity of the segmented pixels corresponding to A
τ	Division Time	min	time elapsed since birth of a cell

Table 4.1: *Summary of key single cell measurements.*

functions.

First, the cropped image of the k-th microchannel is passed through a 2D median filter to remove hot pixels (medfilt2†) and then resized using bilinear interpolation (imresize†, 3x magnification) to further smooth intensity values before thresholding. The cropped greyscale image is thresholded using Otsu's method with few modification. The thresholded image is then morphologically dilated using a 3 pixel radius disc (imdilate†) to compensate for the slight oversegmentation during thresholding and further smooth object edges. The boundaries of the resulting objects are then found with bwboundaries†.

Each object was then assessed according to its ordered (x, y) boundary points, with each coordinate pair representing an edge pixel. For a single bacterium this set typically consisted of about 120 points on the magnified image. Curvature is then computed (Figure 4.3). The local minima of negative curvature regions are each designated as a "pinch point" and, owing to the configuration of the cells, generally one pair of pinch points can be connected to separate two adjacent cells.

For generic object tracking, the most common algorithmic approach calculates pairwise distances between objects in consecutive frames looking for the assignment that minimizes all pairwise distances.The disadvantage is that this calculation has a huge computational cost . Instead, the tracking procedure can take advantage of the geometric constraints imposed by the mother machine's geometry. In particular,in the mother machine, in the absence of cell divisions a cell's rank within a channel is conserved between frames. The task can be reduced to sorting cells in-channel based on distance from the dead-end and determining markers for cell division to adjust the rank-based pairing as necessary. In particular, division has occurred if two adjacent cells show a significant area reduction.

Fluorescence images and the cell boundaries retrieved from segmentation are used to measure cell size, shape, and GFP expression. Volume is computed considering a cell as a cylinder with two hemispherical caps. Similarly, width is computed assuming the projected area is a rectangle with semicircular caps of radius $\frac{w}{2}$.

The final result is two .txt files for each channel, one with observables reported

for each cell at each time frame, and a second one with observables for each cell at birth and division. In Table 4.1 a list of the main observables obtained by segmentation are reported. Lastly, outputs from different channels are joined with a simple bash script.

Another microfluidic device was developed in our collaborator Kevin Dorfman's lab [171, 172]. This device consists of chambers that are 1 micron high but of different widths (30 x 60, 30 x 120, 120 x 30 etc). If we utilize chambers instead of channels we can't take advantage of the device geometry to proceed with the tracking. In this case we can use the Matlab-based software Supersegger [183]. The main principles are similar to the ones describe before for the mother machine but SuperSegger incorporates machine-learning algorithms to optimize cellular boundaries and automated error resolution to identify and link cells from one frame to the next. The two dimensional displacement of the cells within the chamber however renders more difficult the tacking, unless a high acquisition rate is used, which can result in photodamage in some cases. I have used supersegger to track fluorescent foci inside the cells.

4.2.6 Data handling

Image acquisition for one experiment consist of over 20 fields of view, each one with 10 channels, and each channel with about 10 bacteria. Each 3 minutes a new set of images is acquired. This can result in up to 10 000 cells in a 7 hour experiment. The high throughput and high spatio-temporal resolution nature of acquisition necessitates the development of a fast and flexible pipeline of data-analysis.

In particular, I wrote a set of R Markdown notebooks to easily go from raw data to preliminary results. The first, fundamental step, is data cleaning. Single cell measurements reveal indeed that within a population there is a wide distribution of sizes, division times and gene expression levels. Sometimes these distributions appears bi-modal. Errors in segmentation and tracking process can be a main cause. Other causes can be due to an abnormal growth (or no-growth) of bacteria themselves, such as filamentation. A residual inner stochasticity is expected and it is due to the intrinsic noise [184, 185]. To discriminate between these two kinds of fluctuations we implement a set of filters to our datasets. In particular, the time intervals where bacteria have reached their steady state in a given growth medium are considered: a symmetry filter is applied considering only the daughters with a volume between 0.4 and 0.6 of the mother and only cells with doubling times longer than 15 mins and positive growth rate are considered.

Thousands of single cells for each experiment pass the filtering step and lin-

eages of 3-15 generations are generated. Generation plots in Figure 4.4A were obtained using Tunacell, developed by Joachim Rambeau, a former postdoc in our lab. Tunacell is a Python package that provides tools to analyze data from time-lapse movies of dividing micro-organisms (https://tunacell.readthedocs.io). This package is very useful to analyze time-series defined over many cell-cycles, to compute average values, variance, autocovariance for a single observable, and cross-covariance for a couple of observables over time. I contributed to the debugging of this library and I optimized it to be used with our dataset and it has been very useful in particular for the cross-correlation analysis presented in Chapter 6.

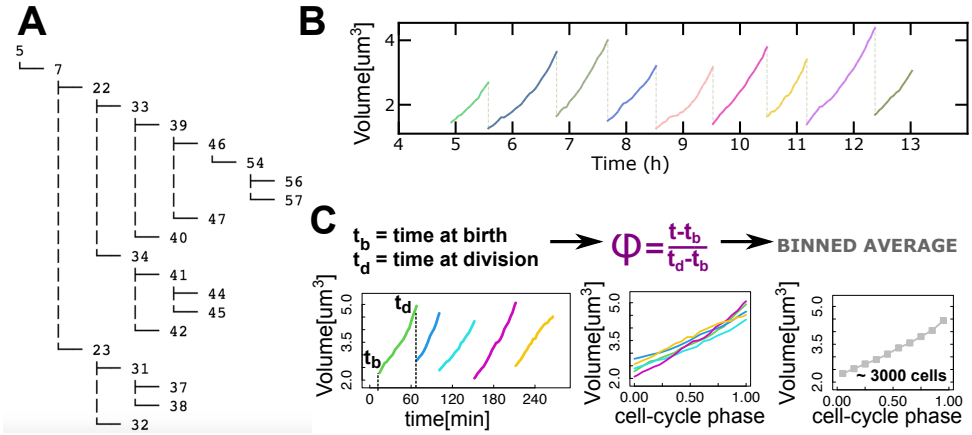


Figure 4.4: Cell cycle analysis through lineages A *Generation plots.* ***B*** *Volume versus experimental time for a 9 generations lineage.* ***C*** *Volume versus cell-cycle phase for the same lineage. Cell-cycle phase is 0 at birth and 1 at division and all the single tracks are averaged together.*

In Figure 4.4 we can see the dynamics of fluorescence and volume change versus experimental time for a 9 generations long lineage.

To examine the effects of cell-cycle progression on gene expression, we can align growth and expression data with respect to a cell-cycle phase defined as the re-scaled time between two consecutive divisions. The cell-cycle phase will be 0 at birth and 1 at division. This procedure makes it possible to average together cells with all doubling times, hence increasing the statistical power (3-8000 cells in our case). Moreover, looking at the fraction of the cell cycle, instead of looking at absolute time, may make a difference in the growth rate and gene expression pattern if, for example, changes in growth rate and/or gene expression reflect the progression of the cell cycle as measured by the occurrence of specific events such as the beginning and end of the DNA replication and cell division processes.

Based on our working hypothesis, which aims to define the gene expression parameters correlated with cell growth and division, alignment is also useful with

Name	Unit	Description
Length	μm	instantaneous cell length
Volume	μm^3	instantaneous cell volume
time from birth	min	time elapsed from cell birth
time to division	min	time missing to cell division
time to daughter's division	min	time missing to cell'daughter division
cell-cycle phase t/τ	min	$(time - time_{min})/(time_{max} - time_{min})$

Table 4.2: *Summary of cell cycle normalizations*

Notation	Name	Unit	Description
$\frac{dF}{dt}$	production rate	$a.u.min^{-1}$	discrete derivative of fluorescence over time
$\frac{dV}{dt}$	growth speed	$\mu m^3.min^{-1}$	discrete derivative of volume over time
$\frac{1}{V}\frac{dF}{dt}$	specific production rate	$a.u.min^{-1}\mu m^{-2}$	production rate normalized by volume
$\frac{1}{V}\frac{dV}{dt}$	specific growth rate	min^{-1}	growth rate normalized by volume

Table 4.3: *Summary of derived observables*

respect to cell volume or time from/to division. In fast growth conditions, when multiple DNA replication forks exists, a normalization with respect to cell's daughter division can be useful to measure possible mother/daughter correlations (Table 4.2).

All these different normalizations allow us to average data from different cells, grouping data based on the value of the x-axis variable, and finding the mean of the fluctuating y-axis variable for this group. Sometimes this binning analysis can be misleading because of "hidden" noise sources that may affect the binning procedure; this leads to the phenomenon of "inspection bias" where certain bins have biased contributions. Having an underlying model (or models) to guide/test/validate data analysis methods may be very important [73].

We assume that the changes in the expression rate of GFP are due to changes in promoter activity and not in changes in translation rate. To estimate the promoter activity in single cells, we define the GFP production rate as the discrete time-derivative of fluorescence dF/dt from the time series of total fluorescence $F(t)$. Similarly, growth rate was defined as the discrete derivative of volume dV/dt, and volume-specific growth rate as $1/V(dV/dt)$. A complete list of these computed variables is provided in Table 4.3.

4.2.7 Design of different microfluidic experiments

Once we have a robust protocol for data acquisition and data analysis we are ready to do experiments. Based on our biological question we can design different experiments, switching for example between different growth media or adding antibiotics. The former is what I did during my PhD, the latter is what the new PhD student Lexuan Liu is working on. Lastly, this experimental setup can also be used to track fluorescent loci inside the bacteria.

Upshift and downshift experiments

Fluctuation in nutrient availability is likely the most common stress faced by single-cell microorganisms in their natural environments. Thanks to the experimental setup described above we are able to quickly change between different growth media. We measured the time necessary to completely fill the device with the new growth medium and determined that it is negligible in the timescale of our experiments and that all the channels are exposed to the change in growth medium at the same time and that therefore we can consider that all the channels get the growth media at the same time (Appendix F).

In our experiments we switch from the slow growth medium containing only glucose to the fast one containing also casamino acids. We can also switch between the fast and slow growth media; this is what is called a downshift experiment. Both steady state data and the adaptation process can give us important information, but exploring the dynamics of different observables in response to a perturbation has the potential to allow one to explore causal relationships between these variables.

Figure 4.5 shows how the mean cell growth, cell division and cell size parameters follow a very complex dynamics across the nutrient upshift, characterized by multiple time scales and trends whereby the same quantity can both increase and decrease in different time windows. In the first growth medium, the cells spend a lot of energy synthesizing their own amino acids and therefore grow more slowly. Once the amino acids are added, the glucose is used mostly for energy production and the growth rate can increase. One can see that the cell division time becomes shorter and cell volume increases as the bacteria adapt to a faster growth rate. In addition, in this case, the fluorescence also increases soon after the change in growth medium, indicating that gene expression has been induced as part of the cellular response.

The two well-defined steady states, respectively in the slow and fast growth

medium, obtained with strains expressing GFP from different promoters chosen to address the questions specific to my project will be described in the following chapters. Thanks to our pressure-driven flow control system, bacteria can reach the exponential growth in a given growth medium and grow for several hours in a constant environment.

Nevertheless, the upshift data are full of information and at the moment the data I obtained have been already used for two side projects I've been working on.

The first set of results contributed to the publication "Threshold accumulation of a constitutive protein explains *E. coli* cell division behavior in nutrient upshifts" [78]. In this work, we used the information obtained about single-cell growth division upon nutrient changes to test different models and to discriminate among fundamentally different mechanisms of cell division control. In particular, we show that the data support a model where a protein expressed from a constitutive promoter accumulates to a threshold and triggers division.

The second set of results is part of a still ongoing project. It is based on an already published theoretical work [186] and consists in the characterization of the width response to nutrient shifts in *E. coli* single cells. In [186] they describe a stochastic relative rates mathematical model for width, or for surface to volume ratio fluctuations at equilibrium. Analysing my data we want to test if the fluctuations in response to a nutritional upshift mirror fluctuations at equilibrium (fluctuation-dissipation theorem).

Labelling of chromosomal loci

If we want to know when and where the initiation of DNA replication happens in the single cell we can label chromosomal loci near the origin of DNA replication and track them within the cell cycle.

We obtained E.coli strains with parB-parS systems expressing ParB-GFP from the Boccard lab [187]. In this strain, a ParB binding site, parS, has been inserted in the chromosome near the replication origin. ParB-GFP proteins proteins production was induced for 30 min with $30\mu M$ IPTG [188] and this allows the easy detection of fluorescent foci that colocalize with replication origins (Figure 4.6. This system allows for the measurement of foci numbers and, if we follow this number over time, we can extrapolate the moment in the cell cycle when origins double, when the initiation of DNA replication happens.

The software SuperSegger includes tools for the quantitation of cellular fluorescence, including the identification of fluorescent foci. To identify and score foci, SuperSegger uses an image-curvature method which is specifically engineered to

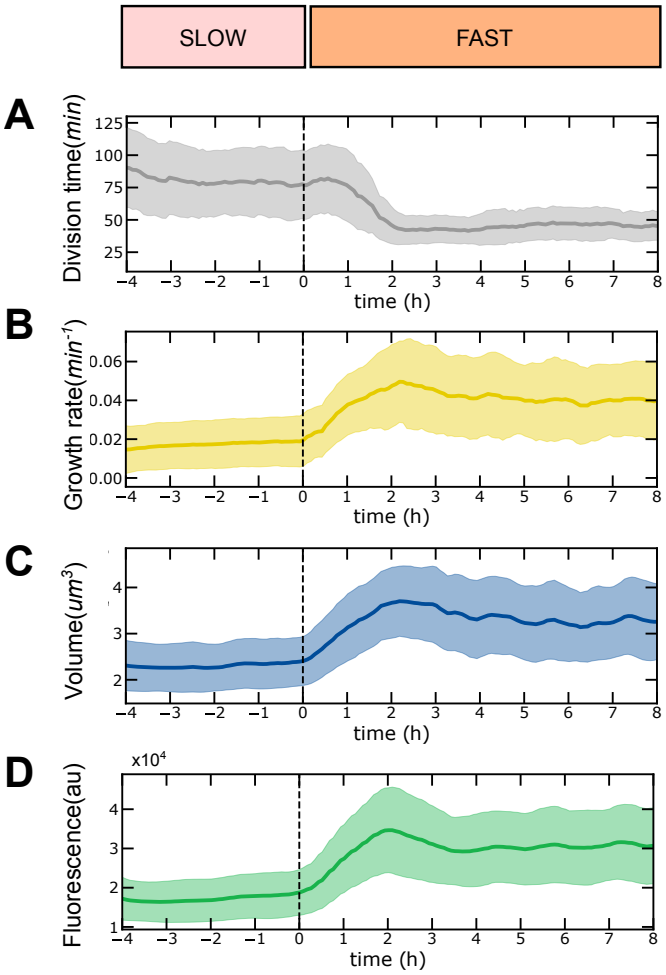


Figure 4.5: Preliminary results for upshift experiments *Population average changes in division time (A), growth rate (B), volume (C) and fluorescence (D) as the bacteria adapt from a poor to a rich growth medium over a 12 hours experiment obtained with the microfluidic setup. Growth medium change happens at time=0 (vertical line)*

avoid the identification of false positive foci due to background intensity from cytoplasmic fluorescence. Foci detection is performed in the union of all cell regions and foci are then allocated to cells to circumvent the double counting of foci close to the membrane at the interface between cells.

The drawback of this protocol is that bacteria are not very fluorescent by themselves and segmentation and tracking results complicated. An important step would be to analyze strains expressing GFP (green) in the cytoplasm with mcherry (red) spots. In this way in the same cell we will be able to have information of growth, DNA replication and promoter activity. The first trials have been unsuccessful but we already have these strains. In Chapter 6 we used this experiment to obtain an estimation of the number of origin and the timing of replication initiation in the two different growth media.

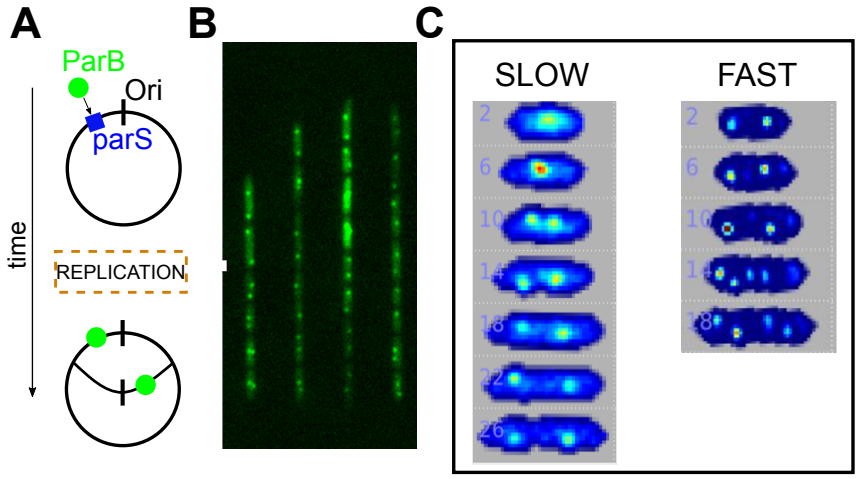


Figure 4.6: Preliminary results for foci experiments (A) *ParB binding site, parS, has been inserted in the chromosome near the replication origin and GFP-ParB proteins form fluorescent foci that localize near the replication origins protein–DNA complex. After replication the number of foci doubles.* ***(B)*** *Fluorescent image of cells with fluorescent loci in microfluidic channels.* ***(C)*** *Localisation of the foci in E. coli bacteria growing in different growth media as a function of the cell cycle.*

4.3 Discussion

To resume, in this chapter I presented a new microfluidic experimental setup that I set up in our laboratory and a new implementation of a pipeline for image analysis and data visualization.

Thank to the versatility of this setup and the straightforward and user friendly pipeline of image acquisition and analysis I have developed, I have already trained bachelor students, master students and a PhD student to use this setup to do their own experiments.

The data I obtained are full of information and, beyond my main project, I have been using them to contribute to different side projects, such has one published article on the upshift dynamics [78], and one in progress on cell width fluctuations.

In the future, a new temperature control system could be useful to keep the temperature constant in a more robust way and to be able to change dynamically the temperature of the device.

We have recently build new bacterial strains expressing more than one different fluorescent reporter protein under control of different promoters. This will allow us to directly measure the activity of multiple factors involved in regulation of gene expression within the same cell and as a function of genome position.

Finally, new results are always followed by new questions and thanks to the versatility of this experimental setup we can always think about new experiments

to do.

Chapter 5

E. coli growth is biphasic and it follows the change in ori-proximal gene dosage

5.1 Introduction

A classic question in bacterial cell biology is how *E. coli* grows. Analysis of growth often involves measurements made on cell populations, which give information on the average properties of the cells in the population. Single cell studies of the evolution of cell growth parameters as a function of the cell cycle can provide unique information that is not available from studies of cell populations [189].

A direct coupling between the metabolic state, the balance between amino acid synthesis and consumption by protein translation, and the cell's growth rate has been measured at the population level, when the growth rate was varied by changing the composition of the growth medium. One of the outstanding questions is whether this coupling could also lead to the observed strong correlation between the growth rate and the protein production rate of a cell at the single cell level[190]. The aim of the work presented in this Chapter is to gain more information on the mode of growth of single *E.coli* cells and on how this is coupled to gene expression.

To address this, we followed cell growth and gene expression at the single cell level from strains expressing GFP from both constitutive and ribosomal promoter placed in the chromosome in 2 different positions, ori-proximal (Ori3) and ter-proximal (Ter3) (see Appendix 4.2.3). All these strains were grown in two growth conditions: a "slow" growth condition in M9-glucose with an average doubling time of 75 ± 5min and in a "fast" growth medium, M9-glucose-casamino acids, with an average doubling time of 45 ± 5min. These growth conditions correspond to the

two steady states from the upshift experiment presented in Chapter 4.2.7.

5.2 Results

5.2.1 E. coli growth is biphasic

One of the earliest attempts to measure the size of individual bacterial cells is seen in Henrici's 1928 book, where he reported the observed changes in cell size in a growth-phase-specific manner and measured cell size using bright-field microscopy observations [191].

More recently, the growth and division of single cells was studied by imaging hundreds of cells growing on agar pads [32]. With the advent of microfluidics techniques it has been easier to follow the growth of thousands of single cells in a constant environment, providing increased statistics necessary for the measurement of what can be small fluctuations [63, 168, 169, 171, 172].

In the literature, there are single cell studies supporting exponential growth within a cell cycle [66, 168, 173, 192–194], as well as linear [189], bilinear [195] or even trilinear growth [196]. This debate now seems considered settled by modern single cell data [67, 173], and an exponential fit of cell size as a function of time or cell cycle phase seems to be a good description for the growth of *E. coli*. Nevertheless, we know how it could be very tricky to discern between different mode of growth in single cells' trajectories [168, 189, 197].

In a recent work on *B. subtilis*, a new method to probe the mode of growth within a cell cycle has been proposed. The authors looked at the variations within the cell cycle of the specific elongation rate $\frac{1}{L}\frac{dL}{dt}$, where L represents the cell length. For cells that are in balanced growth, in a constant environment, the specific growth rate is expected to be constant, if the cell grows exponentially. Thanks to this analysis, they have been able to show that even though the average growth at the population level appears to be exponential, closer inspection of the cell cycle of thousands of single *Bacillus subtilis* cells reveals systematic deviations from a fixed exponential growth rate. This method has been recently proven to be a consistent method to determine the changes in growth rate within a cell cycle and probe the mode of cell growth [73].

Similarly, to investigate *E. coli* growth, I have looked at the growth speed $\frac{dV}{dt}$ and the specific growth rate $\frac{1}{V}\frac{dV}{dt}$ as a function of the cell cycle (Figure 5.1). Here I use the cell volume V, computed considering a cell as a cylinder with two hemispherical caps, as a proxy of cell size (see Chapter 4).

It has been recently found that current methods using only cell length as a proxy for size can lead to inappropriate conclusions [29, 66, 198] and that improved methods accounting for variations in cell width can facilitate reassessment of cell size control mechanisms and other aspects of cell growth [199, 200]. Despite this, we found a biphasic behaviour no matters which proxy for size is considered (Supplementary Figure G.3).

As we have seen in the previous Chapter, the cell cycle phase is obtained by dividing absolute cell age (in minutes) by the cell age at division, resulting in scaling of normalized age from 0 to 1. This normalization allows to average together cells with different doubling times, hence increasing the statistical power of the analysis (5000-10000 cells in our case) [32], avoiding that longer cell cycles contribute more (with more time points) to the population average.

For cells assumed to be growing exponentially, the specific growth rate is expected to be constant throughout the cell cycle. By averaging over multiple cell cycles, the trend of binned data is supposed to be an horizontal line with a value equal to the mean growth rate. However, the data is not consistent with constant exponential growth (Figure 5.1).

The deviations from a simple exponential trend occur in all cells similarly and seem to be associated with specific phases of cell-cycle progression. If they occurred randomly along the cell cycle in individual cells, they would cancel out when the single-cell data are averaged, and one would not observe systematic deviations from exponential growth.

The change in slope in the specific growth rate is present in both growth conditions but we can see a difference in the phase at which this happens. We can then compare the phase at which the change in slope happen with the average initiation time for DNA replication. The latter can be estimated quantitatively by a standard model [45] using experimental parameters for the length of the C and D periods, which also takes into account the case where multiple replication rounds are active in the same cell cycle. We have also measured the dynamics of DNA replication *in vivo* and verified that they are consistent with these estimations (Figure 5.2A). In Figure 5.1 we observe that the average initiation of DNA replication timing precedes the increase in specific growth rate. It is possible that replication of ribosomal operons (5 out of 7 are ori-proximal) may play a role in the change in growth rate. Another possibility could be that this has to do with replication of the genes coding for some limiting cell-wall synthesis components since an operon for peptidoglycan synthesis is present near the origin. This operon contains genes from *murE* (93,166 bp) to *ddlB* (103,153 bp) that code for proteins involved in different

aspects of peptidoglycan synthesis [201].

Similarly, Bruggeman et al. proposed that this change in growth rate within the cell cycle could be coupled to the initiation of DNA replication in *B. subtilis* [202] while other studies have pointed to changes in growth rate at the start of cell constriction [196, 203]. Lastly, in a recent preprint it has been proposed that in *E. coli* the increase in growth rate in the second phase follows the initiation of DNA replication in slow-growing cells and follows cell wall constriction in fast-growing cells, possibly hinting at a replication-related cell cycle control in slow-growth conditions and a divisome-related control in fast-growth conditions [12]. In the latter study slow growth refers to doubling times longer than 100 minutes and this means that the two growth conditions analysed in my work are both considered fast growth and thus in the scenario where the change in growth rate should follow the start of cell wall constriction.

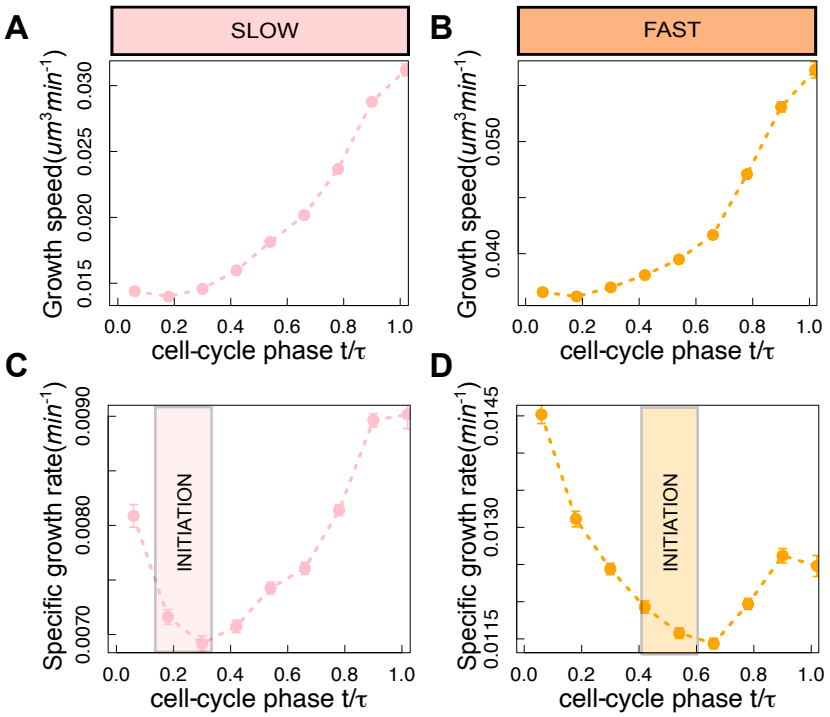


Figure 5.1: E.coli growth is super-exponential in both fast and slow growth conditions *The plots show binned averages of the growth speed (**A**,**B**) and specific growth rate (**C**,**D**), computed with respect to cell-cycle phase. **A**-**B** At both slow and fast growth, growth speed appears biphasic. **C**-**D** At both fast and slow growth, the specific growth rate shows a dip, and the change in slope corresponds to the expected phase of initiation of DNA replication. Around 3000 cell traces are averaged together in each condition. Error bars are standard errors of the mean from a re-sampled distribution of the signal, obtained by bootstrapping from the experimental data for each bin. The average $\pm SD$ initiation time is the shaded box.*

5.2.2 Single-cell growth rate is linked to gene copy number at the origin

Cell growth is a complex process in which cells synthesize cellular components while they increase in size. It is generally assumed that the rate of biosynthesis must somehow be coordinated with the rate of surface growth in order to maintain constant intracellular concentrations [204]. The observed biphasic, non-exponential growth behavior of a single cell along its cell cycle raises the question of how its protein synthesis rate behaves and how it correlates with cell growth.

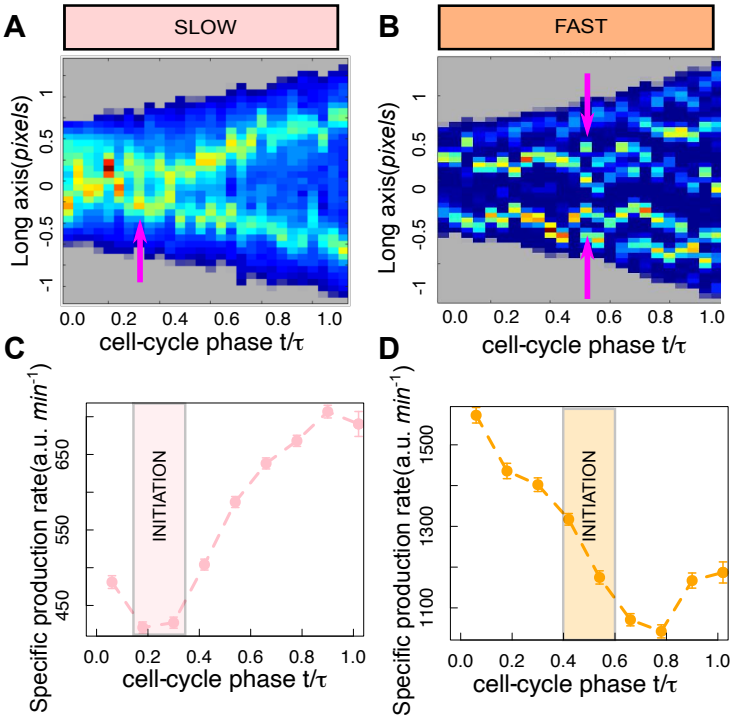


Figure 5.2: The estimated time of initiation of DNA replication is confirmed by* in vivo *measurement of fluorescent loci and the specific GFP production rate from a ribosomal promoter inserted close to the origin shows a pattern similar to the specific growth rate (Figure 5.1). *(**A**, **B**) Foci intensity position and along the long axis of the cell. An increase in intensity is observed before the separation of the origin at slow and fast growth. (**C**, **D**) Binned averages of the specific production rate of GFP, computed with respect to cell cycle phase. At both fast and slow growth, a dip is present at a phase that corresponds to the expected phase of initiation of DNA replication. Around 3000 cell traces are averaged together in each condition. Error bars are standard errors of the mean from a re-sampled distribution of the signal, obtained by bootstrapping from the experimental data for each bin. The average $\pm SD$ initiation time is the shaded box.*

To address this we start by looking at cell cycle-dependent changes in protein expression from a ribosomal promoter placed close to the origin. More precisely, from an upstream-shortened version of rrnBP1, where the binding sites for Fis and

the high affinity-binding site for H-NS have been excluded (see Chapter4). This promoter is expected to have a similar affinity for RNAP as the constitutive promoter P5 [205]. In addition, the GC-rich discriminator region at the transcription start site makes transcription initiation sensitive to the changes in DNA supercoiling and the concentration of ppGpp in the cell.

Figure 5.2A-B shows the timing of initiation of DNA replication in these growth conditions obtained by the observation of parB-GFP fluorescent foci as described in Chapter 4.2.7. We can see how the cell cycle phase at which we observe the doubling of the origin is consistent with the initiation time estimated from the C+H model with C+D = 60 mins. GFP production rate is expected to be dependent on the cell cycle because the replication of the promoter-GFP construct at a specific moment in time increase the probability that it will be transcribed. Consistently, in Figure 5.2C-D we see that the cell cycle phase where GFP production rate start to increase corresponds more or less with the initiation of DNA replication in both growth conditions.

If we compare the pattern of the specific growth rate (Figure 5.1) with the volume-specific expression rate of the ribosomal promoter inserted close to the origin (Figure 5.2) we see that they both exhibit the same biphasic pattern, changing in slope at the same cell cycle phase. This is in agreement with the proposal that the change in growth rate could be coupled with the doubling of a gene close to the origin of replication. In order to test this directly we compared the expression of a gene in two different genomic positions.

5.2.3 A gene's position in the genome can play a role in modulation of gene expression as a function of the cell cycle in addition to the timing of change in gene copy number.

Since gene replication occurs at a time in the cell cycle that depends on its genomic position, more precisely on its distance from the origin, the timing of the change in production rate is expected to be sensitive to the gene's location. To address this, we looked at the cell cycle expression of a constitutive promoter inserted in two different position in the chromosome, one close to the origin, and one to the terminus. The expression of a constitutive promoter only depends on gene copy number and this makes sure that the pattern we see it is only due to gene copy number fluctuations.

It has already been observed that genes located at opposite sides but at the same distance from oriC duplicate at the same time and show the same cell cy-

cle dependence of protein production and concentration, while, if one gene is located closer to the origin compared to the other, the increase in protein production occurs earlier [32]. This happens in the simple case of non-overlapping replication rounds, however, in the presence of multiple rounds of replication a gene farther from the origin can be copied earlier in the cell cycle because it is part of a replication round that has begun in the previous generation. As expected, the ori-proximal specific GFP production rate is higher than the ter-proximal one, while the growth rate is very similar between these two strains (Figure 5.3 A-B-C). All the strains present a biphasic pattern in GFP expression, as observed for the ribosomal promoter. Also, the position of the inflection point is consistent with the different replication timing of the site of insertion (Figure 5.3D). The specific promoter activity for the promoter at the terminus increases earlier than the one near the origin because of the timing of DNA replication. The fact that gene duplication happens earlier (in cell cycle phase) for the promoter near the terminus is due to the fact that we are in the presence of overlapping rounds of DNA replication, as described above.

The cell cycle dependence of the ori-proximal constitutive promoter is similar to the one we found for the ori-proximal ribosomal promoter (Figure 5.2 C-D). This is interesting because in bulk the growth rate dependence of these two promoters is quite different [205]. While they have a similar affinity for RNAP, the ribosomal promoter in addition is regulated by changes in ppGpp concentration and levels of negative DNA supercoiling. In these growth conditions these factors do not contribute to the cell cycle dependence of gene expression.

The phase-dependent GFP production rate is different for the ori- and ter-proximal P5 promoters (Figure 5.3D). While for both promoters, near ori and ter, the increase in gene expression in the second phase can be associated with the timing of the increase in gene copy number, only the specific promoter activity for the ori-proximal promoter decreases until the gene is copied, mirroring the pattern seen with the change in specific growth rate (Figure 5.1). Promoter activity near ori also shows a stronger cell cycle dependence and a larger fold-change in expression.

The difference in the shape of the curve of constitutive gene expression between the ori-proximal and ter-proximal strains could be explained by the high density of native ribosomal promoters in the ori region of the genome that are able to out-compete nearby promoters for local transcriptional resources, especially under faster growth conditions. It is interesting that the decrease in specific promoter activity ends with the duplication of ribosomal operons. As if upon dupli-

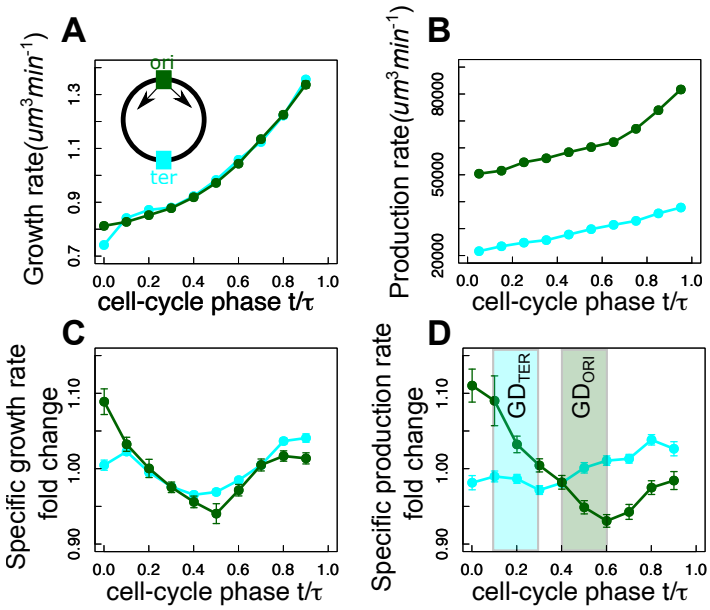


Figure 5.3: Specific production rate of a constitutive promoter inserted close to the terminus is lower than the same promoter inserted near the origin and is not correlated with the specific growth rate*. The expression of GFP from constructs inserted near ori (darkgreen) or ter (cyan) are compared.* ***A*** *GFP production from constructs either near ori or ter does not have an effect on the growth rate, showing the same cell-cycle phase dependence.* ***B*** *GFP production rate from a ori-proximal promoter is higher than for a ter-proximal promoter.* ***C*** *Fold change in the specific growth rate (normalized by the mean) for strains containing the promoter-GFP constructs near the ori or ter show a similar biphasic pattern.* ***D*** *The fold change for the specific GFP production rate (normalized by mean) for a ter-proximal promoter shows less dependence on the cell-cycle phase than the ori-proximal one. Error bars are standard errors of the mean from a re-sampled distribution of the signal, obtained by bootstrapping from the experimental data for each bin. The average $\pm SD$ time of gene duplication is the shaded box.*

cation, each copy of ribosomal operons used less RNA polymerase, allowing an increase of transcriptional resources for nearby genes. The correlation of the specific growth rate with the specific expression rate of a constitutive promoter near the origin suggests that genes in this region of the genome may be limiting for growth rate and not just ribosomal operons. Furthermore, the difference in specific GFP production rate between the beginning and end of the cell cycle suggests that the specific production rate increases after cell division, before starting its descent. At the terminus-proximal construct the opposite trend is observed.

Previous work has proposed that there is differential RNAP availability along the genome due to the presence of highly transcribed ribosomal operons near the origin [206–208]. Our results seem to be in contradiction with Scholz's observa-

tion that there is a higher availability of RNAP close to the ribosomal operons [208]. However, the distance between our promoter insertion site (*aidB* gene) and the closest ribosomal operon (rrnE) is too large to contribute to an higher local concentration of RNAP. The distance is indeed 0.2Mb, while Scholz shows that higher availability of RNAP is negligible already at a distance of 0.15Mb. It should also be noted that Scholtz et al. results were obtained in bulk and thus reflect the average level of expression.

Our results show that expression from genes placed in different positions on the chromosome may differ not only because of the different timing in gene duplication but also because of their change in specific promoter activity as a function of the cell cycle phase. In eukaryotes it is well known that the position of a gene on the chromosome can have an effect on its expression for different reasons [209–212], but such effect in bacteria is still less understood [213]. Several factors can contribute to a position dependence of gene expression, such as proximity to ribosomal operons, as mentioned above [208], the presence of regions with high levels of H-NS binding [214, 215], transcriptional interference [216] and DNA supercoiling coupled to changes in sequence-dependent DNA stability [217, 218].

We next analysed the expression of two versions of a ribosomal promoter (rrnBP1) containing or not the Fis and H-NS binding sites upstream of the core promoter region (Figure 5.4A). The former will be referred to as P1long, the latter as P1short. Both P1short and P1long are expressed in a growth rate dependent manner thanks to the presence of the GC-rich discriminator region overlapping the transcription initiation site, and thus by regulation by ppGpp and DNA supercoiling. In the P1long promoter, activation by Fis at fast growth and repression by H-NS at slow growth increases the growth rate dependent regulation [219, 220]. In principle, the P1long ori-proximal reporter is most representative of the actual ribosomal gene expression: it is the full length promoter, containing all the binding sites for Fis and H-NS and it is near the origin of DNA replication.

These unpublished data have been obtained by Mia Panlilio in our collaborator Pietro Cicuta's laboratory at Cambridge University. The strains used are the same we use in our lab, as well as the growth media (M9-glucose for slow growth and M9-glucose CAA for fast growth) but her experiments are performed at 37°C. At 37°C in the slow growth medium the doubling time ($47 \pm 7min$) is similar to the doubling time of cells growing at 30°C in the fast growth medium ($48 \pm 4min$). As expected, variations in temperature have a larger effect on growth rate than just cell size (Supplementary Figure G.1) [41]. The data obtained for the expression of GFP from the constitutive promoter in these condition are consistent with the

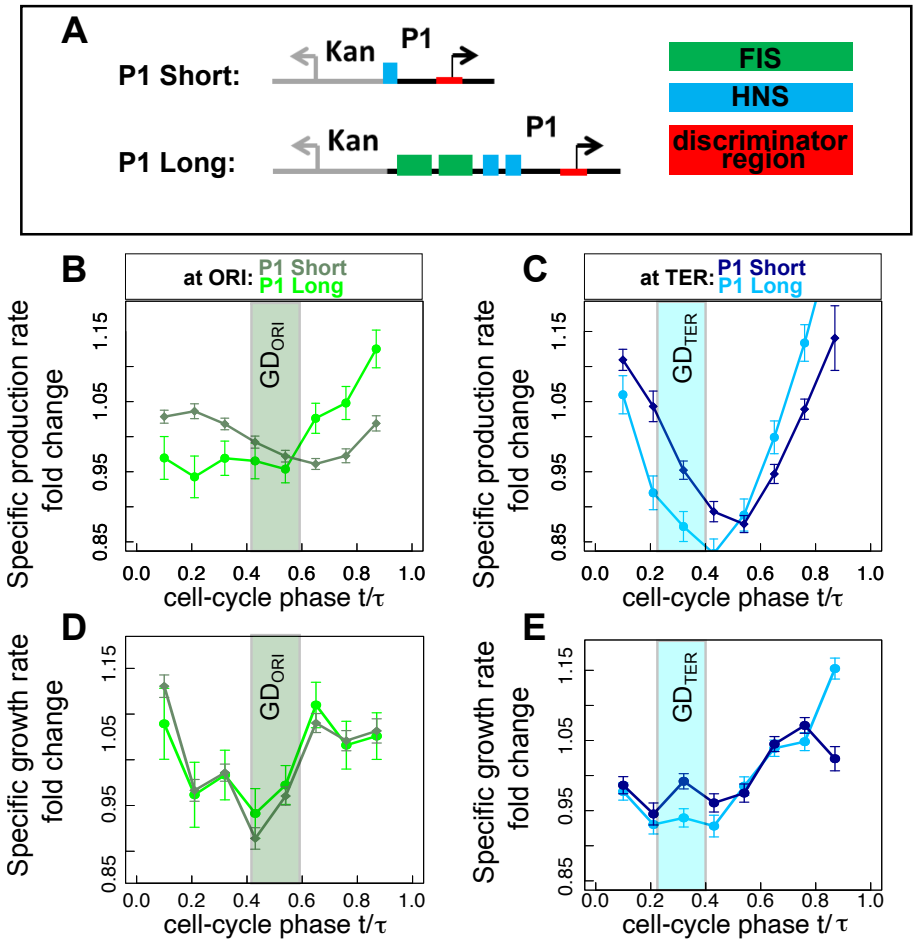


Figure 5.4: The specific production rate of the full length ribosomal promoter at ori does not decrease with the specific growth rate in the first half of the cell cycle ***A*** *Promoter schematics indicating the NAP-binding sites upstream of the promoter. The P1long reporter construct contains all NAP-binding sites since it is the wild type sequence of rrnBP1.* ***B*** *Near ori, P1long (light green) expression does not decrease in the first half of the cell cycle as P1short does (dark green).* ***C*** *Near ter, P1long (light blue) and P1short (dark blue) behave similarly. This data was obtained in M9-glu at 37°C. The data in the previous figures was obtained at 30°C in the same growth medium.* ***(D-E)*** *The specific growth rate changes similarly between P1long and P1short containing strains, however it differs between the strains with the insertions that are either ori-proximal or ter-proximal. This small difference at the beginning of the cell cycle might be due to GFP expression burden. After that, for all strains growth rate increases at the same cell-cycle phase, corresponding to the timing of initiation of DNA replication. Error bars are standard errors of the mean from a re-sampled distribution of the signal, obtained by bootstrapping from the experimental data for each bin. The estimated average* $\pm SD$ *gene duplication time is shown by the shaded box.*

results obtained at 30°C (Supplementary Figure G.2).

In Figure 5.4C we see that at the ter-proximal position P1short and P1long are very similar to each other. In the first half of the cell cycle there is a decrease in the

specific production rate, followed by a sharp increase at the time of gene duplication. At the ori-proximal position, the pattern of the P1short promoter construct looks similar to the one observed for the origin-proximal P1short or P5 at 30°C (Figure 5.3). The construct with the full length ribosomal promoter, on the other hand, does not decrease at the beginning of the cell cycle. In the P1long promoter the presence of Fis helps recruit RNAP, which might result it to better compete with neighboring genes, notably the ribosomal operons, as described above. Lastly, in both chromosomal positions P1long increases earlier than P1short.

P1long at ori is the only construct that does does not show a decrease in specific GFP production rate before gene replication. This suggests that in the first half of the cell cycle, before initiation of DNA replication, the ribosomal production rate is proportional to volume while the expression of other genes that may be rate-limiting for cell growth is slower than volume, as we see for both P1short and P5 at both 30 and 37°C.

5.2.4 Growth rate homeostasis

To better understand the origins of the systematic deviations from exponential growth, we partitioned the single cell data into several bins according to the cell's birth volume. We want to see if these deviations for example happen only for relatively large or small cells and whether they are part of a mechanism for cell size correction.

Figure 5.5A-B shows that cells display systematic deviations from exponential growth regardless of their initial size. Moreover, if we examine each of the birth-volume classes separately, the specific growth rate depends on cell size at the start of the cell cycle, then converges to nearly the same value at the end of the cell cycle. Although cell size and specific growth rate are correlated at birth, these two parameters are nearly independent right before division (Figure 5.5 B-D). Immediately after division, cells born smaller grow faster than larger cells, this results in small cells "catching up" the size difference, and all cells, regardless of their initial size, reach a similar specific growth rate by the end of the cycle. This correction looks stronger at slow growth.

The fact that the specific growth rate is higher in smaller cells may come from two sisters that divide asymmetrically and after that smaller cells that also give smaller cells. Smaller cells have the same DNA content as the larger cells, and the same gene copy number of ribosomal operons, and thus the same absolute potential translation rate. Smaller cells thus produce the same number, but therefore a higher concentration, of ribosomes than larger cells.

Figure 5.6 shows the GFP production rates and specific production rates as a function of cell size at birth. These results confirm the idea that the small cells in the slow growth medium have a faster translation rate per unit volume, which then converges with the other size classes as a function of the cell cycle (Figure 5.6C).

In the fast growth condition one can see that the GFP production rate at birth increases with cell size in a way that is almost proportional to cell volume (Figure 5.6B). During the cell cycle the production rates converge.

The GFP production rate is proportional to cell size at birth at fast growth, to then collapse before cell division, while at slow growth it is the same independently of cell size. These results suggest that at fast growth cell volume (translation?) determines the GFP production rate at birth, while at slow growth the gene copy number (transcription?) seems to play a determining role.

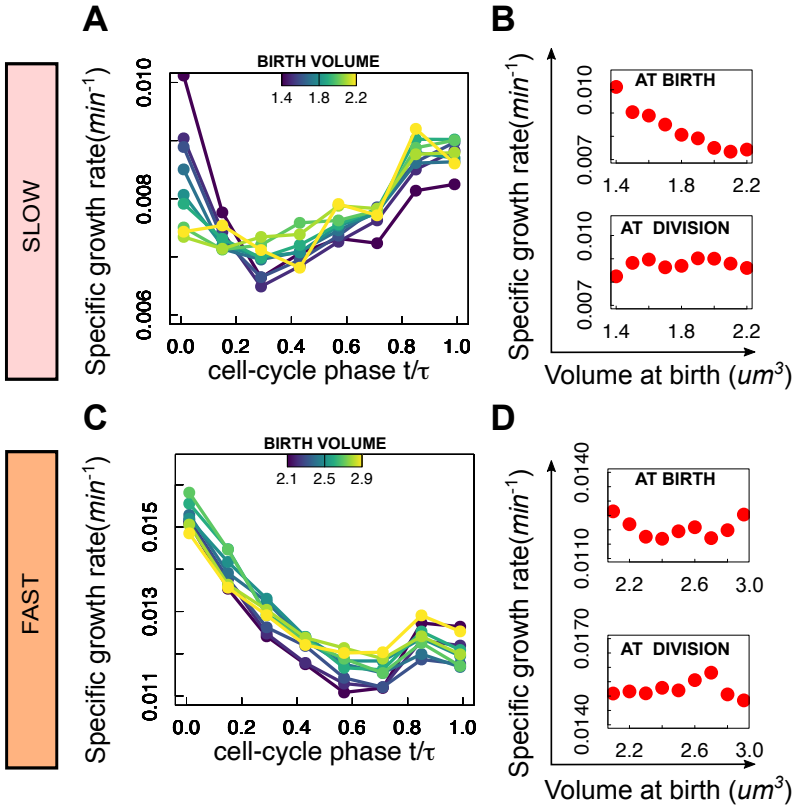


Figure 5.5: Specific growth rate fluctuations are compensated over the cell cycle. *Specific growth rate of different birth volume classes as a function of the cell-cycle phase for slow **A** and fast growth **C**. The legend indicates the mean birth length of the respective birth-volume class. In both conditions we can observe a dependence on birth volume **B, D**. Specific growth rate at birth/division of different birth volume classes as a function of volume at birth. **B** At slow growth, cells with a different initial volume show differences in growth rates at birth that are compensated by the time they arrive at division. **D** At fast growth growth rate fluctuations are weaker and almost constant within the cell cycle.*

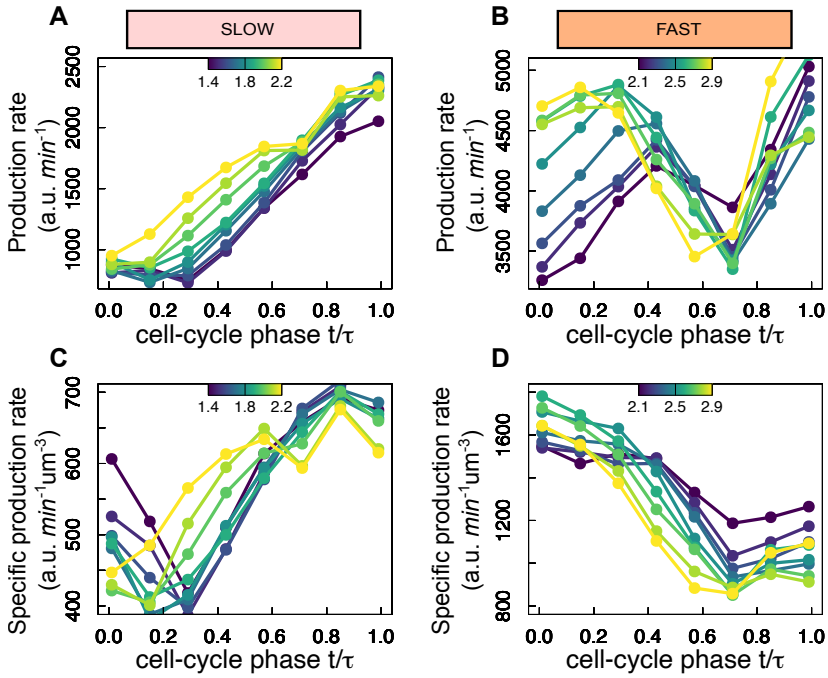


Figure 5.6: Smaller cells have higher volume specific translation rate at slow growth. *GFP production rate (**A,B**) and specific GFP production rate (**C,D**) of different birth volume classes as a function of the cell-cycle phase for slow and fast growth. The legend indicates the mean birth length of the respective birth-volume class. In both conditions we can observe a dependence on birth volume.*

While in Figure 5.5 we analyze the behavior of the specific growth rate, in Figure 5.7A-B we look at the absolute growth rate, or growth speed, as a function of cell cycle. While the specific growth rate gives us information of how growth is coupled to cell volume, which is expected to be constant under exponential growth, growth speed tells us how the absolute growth speed varies as a function of the cell cycle and informs us on the metabolic activity of the cell. Especially at fast growth a rate change point can be easily detected for all the size classes. This inflection point define two different growth phases, with a first phase with a nearly constant growth speed, and a second phase where the growth speed increases.

To understand whether the inflection points are related to cell volume, we plot the growth speed as a function of cell volume (Figure 5.7 C-D). All the curves deviate from the expected exponential curve, and the effect of birth size on growth speed is greatest in the fast growth condition and at the beginning of the cell cycle.

Lastly, we evaluated the average growth speed of cells in the different birth-size classes as a function of the time to division (Figure 5.7 E-F). Interestingly, from this perspective, we found an alignment of phase transitions for different birth-size classes, with rates increasing at a specific time before cell division, independently

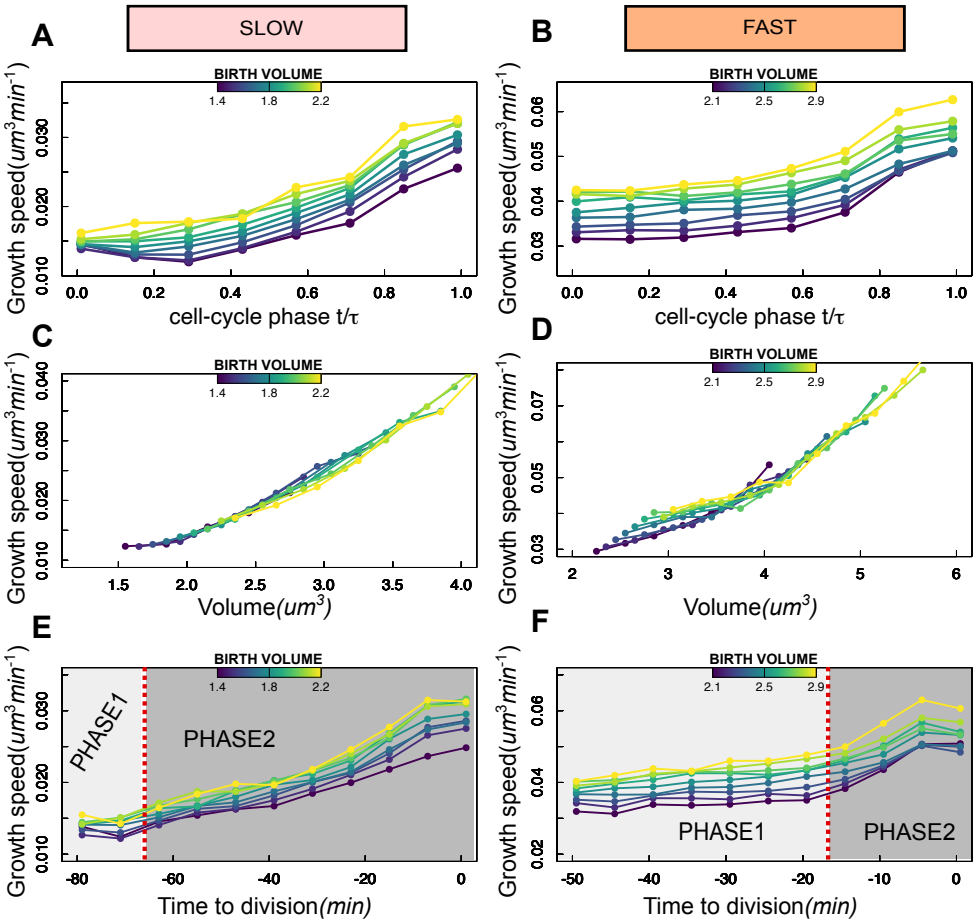


Figure 5.7: Cells start a new phase of growth at a similar time before the cell division, regardless of their birth size (A,B) *The growth speed as a function of the cell-cycle phase. At fast growth most cells start with a fairly constant growth speed that increases as time progresses in a birth-size-dependent manner.* ***(C,D)*** *The growth speed depends on cell volume, which is birth size dependent when cells are in the fast growth condition.* ***(E,F)*** *The growth speed as a function of the time to division reveals that cells start growing faster at an approximately fixed time before they divide. The red line represents the time before cell division at which growth speed changes slope and defines the 2 phases. At slow growth also a single phase of monotonic growth is compatible with the data.*

of birth volume. At slow growth the first phase is really short and the detection of the inflection point can be tricky. We can then define a first phase (phase1) characterized by an approximately constant growth speed and a second phase (phase2) where growth speed increases nearly exponentially (Figure 5.7 A-B).

5.3 Discussion

In this Chapter we have obtained very interesting results on the mode of *E.coli* growth within the cell cycle and its relation with gene expression.

Thanks to robust and long term single cell experiments I have been able to track cell growth and gene expression for thousands of cell for several hours with a time frame of only 3 minutes. A careful analysis of these single cell data as a function of the cell cycle has highlighted not negligible deviations from the classical exponential growth [73, 202, 204]. The observation that the absolute growth rate of an individual cell is not at all times proportional to its size (it is not exponential) implies that it is adjusted continuously in response to cell cycle-dependent events.

Cell cycle dependence of specific growth rate

The observation that in our data growth rate increases soon after the initiation of DNA replication suggests that the replication of the ribosomal operon close to the origin can have a role in this biphasic growth pattern. This may happen because after the replication of ribosomal operons close to the origin there is an increase in the ribosome production rate and an increase in the global translation rate. This implies that after the initiation of DNA replication, cells can start growing more rapidly because they have the internal resources of a larger cell. Whether the cellular materials that promote growth are some subset of the cytoplasmic constituents, or the entire collection, remains to be seen. We propose that ribosomes have a main role in this because we know that ribosomes are one of the dominant constituents in cells, and more specifically in rapidly dividing cells, they begin to take up a significant fraction of the cellular interior and of cellular resources for their production. For example, for a doubling time of 60 mins 14000 ribosomes per cell represents the 25% of dry mass while at a faster doubling time of 24 minutes the 72000 ribosomes per cell represent over 1/3 of the dry mass of the cell [221–223].

Of course, this in not the only possible explanation. We know for example that also an operon for peptidoglycan synthesis is present near the origin and its replication might have a role in the increase of growth rate[201]. Moreover, other studies exist in the literature that lead us to think that this phenomenon is likely to be multi-factorial, dependent on processes and components involved in DNA replication initiation, gene duplication, cell wall synthesis, divisome assembly, and septum formation, all of which in turn will be critically influenced by biosynthesis and growth-rate dynamics.

Cell cycle dependence of specific GFP production rate depends on genome position

Another important result in this Chapter concerns the relation between cell growth and protein expression. To address this, we have followed the expression of constitutive, ribosomal and shortened ribosomal (lacking HNS and Fis regulation) promoters placed in the chromosome both close to the origin and the terminus. This is because a gene's transcription rate could be affected differently depending on its position on the genome. Several factors can contribute to position dependence of gene expression such as H-NS binding [215], transcriptional interference [216] and DNA supercoiling [217, 224].

In particular, we have seen that a strong correlation between the growth rate of the cell and the GFP synthesis rate is only present for the constitutive promoter and the ribosomal promoter that lacks HNS and Fis regulation when they are inserted at the origin-proximal position. This suggests that these promoters are found in different environments. In general, these results confirm previous studies showing that spatial organization of transcription may play a crucial role in coordinating genomic transcription and growth rate in *E. coli* [225], but for the first time we were able to see this at the level of the cell cycle.

Specific GFP expression from non-ribosomal promoters that are ori-proximal match the changes in specific growth rate

The comparison of the results obtained with the full-length and short version of the ribosomal promoter and with the constitutive promoter suggest that the specific growth rate is dictated by the expression of a non-ribosomal promoter whose activity is limited by the availability of RNA polymerase. And this is true particularly before initiation of DNA replication suggesting that the duplication of ribosomal operons might increase the availability of RNAP, by decreasing the demand for RNAP for each copy of ribosomal operons. Moreover, the presence of Fis may help recruit RNAP to ribosomal operons to better compete with neighboring genes before the initiation of DNA replication.

Cell size at birth affects specific growth rate and gene expression in a growth medium dependent manner

Lastly, we have found a growth rate homeostatic mechanisms, where specific growth rate in the first phase of the cell cycle is dependent on cell size at birth. In general, these findings suggests an homeostatic mechanism, by which differences

in cell size at birth are compensated during the cell cycle by differences in their specific growth rate.

From the GFP expression data we can propose that a possible interpretation of this effect could be that in a cell born smaller the concentration of ribosomes and of RNAP will be higher than in a larger cell because their gene copy number will be the same as in a larger cell. During the cell cycle the faster specific growth rate of the smaller cells will dilute these machineries so that by the end of the cell cycle the specific growth rate will be the same as the others.

The difference in the distribution in specific growth rate between birth and division and between slow and fast growth rate suggests that below a specific cell size the same number of ribosomal operons can give rise to different specific growth rates because of the resulting difference in ribosome concentration.

At fast growth, cells are born with a number of ribosomes that depends on cell volume, while at slow growth the number of ribosomes depends on their on gene copy number.

In general, this work represents a step forward in the understanding the coordination between cell growth and gene expression in *E.coli*. Some of these results, such as biphasic growth and its strong coordination with gene expression were already been found in *B.Subtilis* [202] . Here, we confirmed these results in *E.coli* and we also added other interested findings. First, the biphasic growth pattern and its homeostatic mechanism depend on the growth condition and different regimes might arise. Second, the coordination between growth and gene expression is not universal but it depends on the genomic position and regulatory elements of the single promoter.

5.3.1 Future perspectives

What factor is limiting for growth rate? RNAP concentration? RNAP to DNA ratio? Ribosome concentration? Both? Are specific machineries, such as cell wall synthesis enzymes limiting for cell growth? Several studies have tried to address these issues, and the answer appears to depend on the growth conditions such as in the recent studies by the Männik group [12]

Testing new growth conditions would be helpful to determine whether different growth regimes exist, and in which conditions the correlation between growth rate and protein production rate is stronger. For this, it would also be necessary to use strains with fluorescent loci to track initiation of DNA replication at the single cell level in the same strains that express GFP. We already have these strains.

Moreover, we already know that chromosome structure and organisation can

provide an additional level of feedback regulation to obtain a response that is proportional to the environmental change[78]. In the latter example, expression from a constitutive promoter at the ter-proximal site could best model the expression of a protein setting the threshold for cell division across the shift. Preliminary results from our and Pietro Cicuta's lab have already observed different patterns by looking at the adaptation of gene expression for these different promoters during upshift experiments (Figure G.4). The next step would be to understand how this is linked to different patterns of cell growth and gene expression at the cell cycle level.

Lastly, it would also be interesting to stress the system with different antibiotics inhibiting growth, cell division or DNA replication and to characterize the robustness of this biphasic behaviour. For example, inhibition of ribosomes by sublethal concentrations of antibiotics such as chloramphenicol leads to a decrease in the growth rate despite an increase in ribosome content. The limitation of ribosomal activity results in an increase in the level of expression from ribosomal promoters; this can deplete the pool of RNA polymerase (RNAP) that is available for the expression of non-ribosomal genes [18]. It would be interesting to quantify the magnitude of this effect within the cell cycle.

To conclude, a better understanding of gene position effects can also be useful to understand genome assembly and evolution. In Chapter 7 we will see how this might also be important to understand horizontal gene transfer.

Chapter 6

The DnaA cell-cycle oscillator and its coordination with cell growth and division

6.1 Introduction

By its contribution to setting the timing and synchrony of initiation of DNA replication, DnaA plays a key role in the cell cycle, but its contributions to sensing cell size and indirectly setting cell division are not well understood.

The consensus is that DnaA plays a role in coupling cell size to chromosome amount. For the average behavior of a population across conditions, the traditional paradigm was that initiation of DNA replication roughly starts at a constant volume (or mass) per origin [21, 49]. However, it has recently been shown that coupling of cell division with the chromosome cycle likely differs between slow- and fast-growth conditions, two growth regimes roughly separated by the onset of overlapping replication rounds [12, 30, 51].

Recent data show that initiation mass increases and then decreases as a function of growth rate, with a maximum at around $0.7h^{-1}$ at 37°C. The concentration of DnaA follows a mirror pattern decreasing and then increasing again, with a minimum at 0.7 $0.7h^{-1}$ [51].

Furthermore, when studying cell size control in single cells rather than in bulk, as in the studies above, there is some agreement on the idea that the added volume per origin between consecutive initiations may be constant and likely set by the contribution of the DnaA-dependent regulatory circuit [8–12, 59, 62]. Therefore, the activity of DnaA plays an important role in the regulation of DNA replication as a function of growth rate and of cell size as a function of the cell cycle. However,

despite a large number of recent studies, the debate is still heated regarding the question of whether and to what extent replication initiation sets the division time in single cells [8–13].

In this Chapter, we want to gain insight into the DnaA regulatory system, in particular in relation with different phases of the cell cycle.

The challenge is that even if one could quantify the amount of DnaA-ATP in the cell, as some have done in bulk using radioactive ATP, the important parameter might not only be its concentration but the ratio of DnaA-ATP to DNA at any given time during the cell cycle, given the titration effect that the other genomic DnaA binding sites can have. This ratio can indeed act on the DnaA activity by changing the "free concentration" in respect to the "total concentration".

Here, single cell time-resolved data of GFP reporter constructs using promoters differently regulated by DnaA-ATP have been used to assess the relative changes in the activity of DnaA-ATP, and therefore the changes in the amount of DnaA-ATP available to bind to its high affinity sites. An oscillation in DnaA-ATP activity is indeed assumed by most, but so far supported only by indirect population-level data [23].

The main question we want to answer are then:

- Can we establish if DnaA activity oscillates in single cells?
- If so, is this a size sensor?
- How is it linked to the cell-cycle progression and cell growth and division?

6.2 Results

6.2.1 DnaAP mutants: A measure of the change in DnaA's activity

An experimental approach to study gene expression in bacteria *in vivo* consists on carefully choosing gene promoter sequences and their variants regulating the expression of a fluorescent reporter proteins (see for example [14]). The promoter sequence is determined from *in vitro* experiments of DNA-protein interactions and from a careful literature review. In Chapter 3 I have introduced the structure and regulatory elements of the native DnaA promoter (dnaAP). These experiments have been carried out with the wild-type *E. coli* strain BW25113, the parent strain of the Keio collection [226] which has been fully sequenced [179]. We have constructed a reporter cassette, where the fast-folding *mut2gfp* gene is under control of the

dnaA promoter sequence. This construct includes a Kanamycin resistance cassette expressed divergently upstream from dnaAP. I have placed the dnaAP promoter-reporter construct in the genome at the Ori3 position, downstream of the *aidB* gene (4413507 bp). This is the same ori-proximal position used for the constructs we looked at in the previous chapter. Moreover, it is at a similar distance from the origin as the native*dnaA* promoter(3883729 bp) but on the other replicon. The only differences with the native promoter is that in my mutants the divergent promoters for the *rpmH* gene are absent and that the dnaAP1 promoter has been inactivated by site-directed mutagenesis, therefore GFP expression is under control of the dnaAP2 promoter that, as we have seen in Chapter 3 is similar to a ribosomal promoter and is thought to provide the growth rate dependent regulation of DnaA expression. The chromosomal insertion protocol is reported in Appendix A.

The dnaAP mutants used in this work consist of different basepair modification of the high affinity binding sites of DnaA. These mutants result in promoters that are either only positively regulated or not regulated at all by DnaA-ATP, while the wild type sequence is both positively and negatively regulated by DnaA-ATP. The effect of the same mutations on gene expression regulation by DnaA (negative or positive) was first obtained from a previous work in my lab [112] although they used plasmids carrying the promoter-GFP constructs. Hence - crucially for our experiments - measurement of GFP expression rate from chromosomal insertions of these different promoter variants can be used to estimate the amount of regulation that is due to changes in DnaA activity in the cell. DnaA activity depends on the concentration of DnaA-ATP that is available to bind to its own promoter (see Appendix H).

Lastly, the SeqA mutant promoter still has the the dnaAP1 promoter but the three GATC sites just downstream of dnaAP2 have been removed by site-directed mutagenesis, decreasing the effect of repression by SeqA. Three SeqA sites about 170 basepairs upstream of the dnaAP2 promoter and two SeqA sites overlapping with the dnaAP2 promoter remain. The latter could not have been removed without changing the sequence of the -10 and -35 regions.

To summarize, the native dnaAP is positively and negatively autoregulated and DnaA regulates the reporters constructs that have the same promoter sequence as dnaAP but present different mutation in the DnaA and SeqA binding sites (Figure 6.1).

As a reference, we considered a reporter for a constitutive promoter used in a previous study [77]. This phage-derived constitutive promoter “P5” lacks any kind of regulation by specific transcription factors, therefore its GFP production rate

can be considered to be largely representative of the gene copy number in the cell. To determine when the gene copy number changes for different chromosomal sites we have inserted the P5-GFP construct at the same ori3 origin-proximate locus as the dnaAP2-GFP construct as well as at the terminus-proximate locus "ter3" (1395706 bp). In addition, we also considered the ribosomal promoter reporter "P1", derived from the rrnBP1 promoter for the *rrnB* ribosomal RNA operon whose expression depends on the concentration of ppGpp and on the level of negative DNA supercoiling due to the presence of a GC-rich discriminator region, such as the one found at the dnaAP2 promoter. This is a shortened version of the rrnBP1 promoter, lacking the upstream binding sites for Fis and H-NS [219].

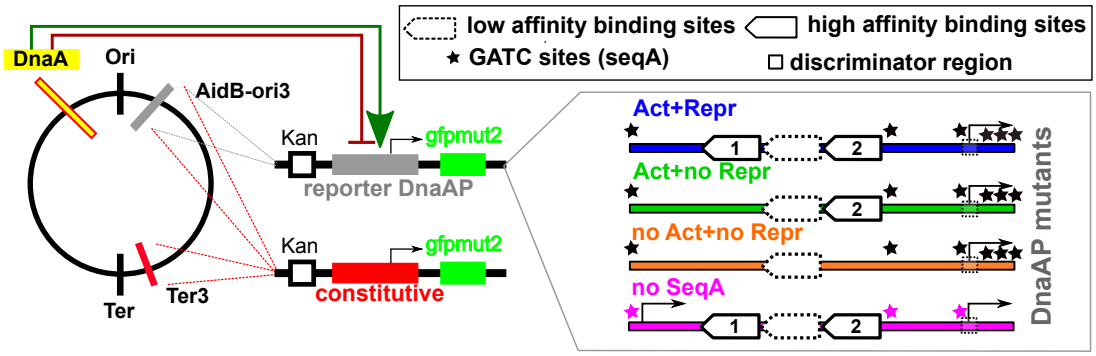


Figure 6.1: The set of DnaA promoter mutants *A construct of the* mut2gfp *gene regulated by the native* dnaA *promoter region and including a Kanamycin resistance cassette was inserted close to the origin of DNA replication. Different promoter mutants with different levels of regulation by DnaA and SeqA were considered. As a reference for baseline gene expression we used a constitutive promoter. The GATC sites bound by SeqA are shown by stars, the DnaA high affinity binding sites are shown in solid line boxes and low affinity binding sites in dashed lines boxes. The GC-rich discriminator region is shown by a small square.*

6.2.2 Activation and repression at the population level for fast and slow growing cells

Measurement of the fluorescence of the GFP protein expressed under the control of the dnaAP or its mutants is used to determine promoter activity (dGFP/dt / OD). Promoter activity is first measured at the population level in a plate reader that allows to rapidly screen several growth media and promoter mutants. We can thus identify the most interesting strains and growth parameters for the subsequent study of gene expression in greater detail, at the single cell level, by time-lapse microscopy. All these strains were grown in two growth conditions: a "slow" growth condition in M9-glucose with an average doubling time of 75 ± 5min and in a "fast"

growth medium, M9-glucose-casamino acids, with an average doubling time of 45 ± 5min.

The rate of GFP production, divided by the cell density as measured by the optical density at 600 nm (OD600), provides a real time measurement of the change in promoter activity as a function of time. A MatLab program is used in order to automate the analysis of the data obtained from the measurement of bacteria growing in 96-well plates in the fluorimeter. An R-Shiny program for the analysis of the data from the plate reader has also been recently developed. See Appendix E for more details.

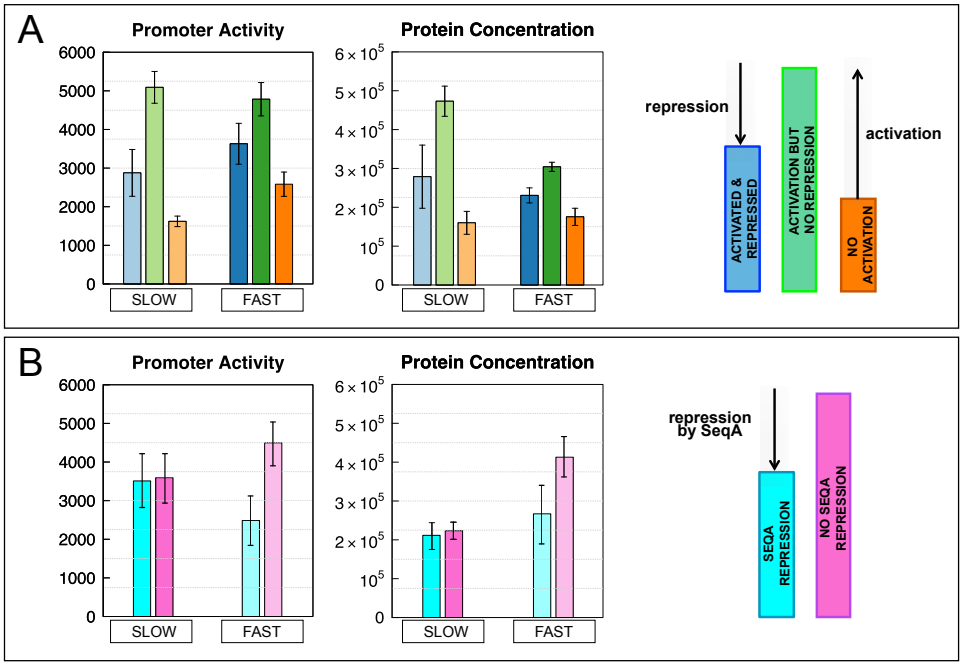


Figure 6.2: Extent of regulation by DnaA and SeqA by comparison of differently regulated promoter mutants *Change in average promoter activity and GFP concentration for the different promoters in exponential phase at fast growth (M9-glucose casamino acids) and slow growth (M9-glucose).(**A**) Extent of activation and repression by DnaA of its own promoter. **B** Extent of repression by SeqA of the dnaA promoter. The values shown are the average of three independent experiments. The error bar is the SEM.*

The comparison between these different mutants can be used to estimate the extent of activation and repression by DnaA-ATP of its own promoter at different growth rates and therefore obtain a measure of the change in DnaA's activity (Figure 6.2A). All the values have been measured in exponential phase after checking that the growth phase dependence of dnaAP2 activity is independent of autoregulation (Supplementary Figure H.4). The expression rate from the promoter that is not regulated by DnaA increases with growth rate, as expected from a ribosomal-like promoter sequence, resulting in a constant DnaA concentration as a function

of growth rate. The activity of the promoter that is only activated is not growth rate dependent because of the increased activation at slow growth when compared with the unregulated promoter. Positive autoregulation of dnaAP2 transcription may play an important role to counteract the effect of possible repressors, such as Fis [227], or in the presence of negative effectors of transcription, such as ppGpp, known to decrease the activity at slow growth of promoters such as dnaAP2 containing a GC-rich discriminator region [22]. At the same time, the promoter that is both activated and repressed changes less than the non regulated one with the change in growth medium due to the increased repression at slow growth. These results therefore suggest that there is a higher activity of DnaA-ATP at slow growth compared to fast growth. This agrees with previously published study by the Skarstad group where they show that DnaA is not limiting for DNA replication initiation at slow growth, since DnaA-ATP seems to be in excess. It could be that at slow growth another factor is more important for setting the initiation timing of DNA replication [123].

Binding of SeqA to the DnaA promoter inhibits novel synthesis of the protein following the passage of the replication fork while the DNA is still hemi-methylated (see Chapter 3). SeqA does not have a strong effect in the slow growth medium.

6.2.3 In absence of any regulation, GFP production rate is proportional to single-cell size and instantaneous gene copy number

To establish a solid reference for monitoring the cell cycle dependence of gene expression from the dnaAP promoter, our first goal was to characterize the “null” relationships between cell cycle progression and gene expression, i.e., the cell-cycle variability of an unregulated promoter.

This is exactly what I have already done in the previous Chapter and I will try to summarise the main results that we are interested in for this work (Chapter 5 and Supplementary Figure H.6). From now on, we will consider experiments at fast growth.

As we already know, protein production rate can vary along the cell cycle because the replication of a gene at a specific moment in time doubles the probability that it will be expressed. Since gene replication occurs at a time in the cell cycle that depends on the gene’s distance from the origin of DNA replication, the cell cycle dependence of its expression rate will depend on the gene’s location along the genome. The copy number of a gene changes during the cell cycle, and can be

estimated quantitatively by a standard model [43], which also takes into account the case of "overlapping DNA replication rounds" where replication forks from different initiation events are active in the same cell (see Appendix G). Additionally, a protein's production rate has been observed to be proportional to cell size, probably in connection with the fact that cell size tends to be proportional to ribosome amounts [228]. To quantify the dependencies of GFP production rate from both gene dosage and cell volume, we monitored the expression of a constitutive promoter (P5, described above) integrated into two genomic loci, one next to the replication origin and the other close to the replication terminus.

We first tested how the GFP production rate depends on cell size. Figure H.6A reports scatter plots of GFP production rate versus volume for origin- and terminus-proximate P5 constructs. These single cell data show an average linear proportionality between the GFP production rate and single cell volume. As expected from the estimation of average gene copy number, the same unregulated promoter shows an increased production rate when it is placed close to the replication origin compared to the one near the terminus. Normalizing the data by the estimated mean gene copy number removes most of this offset. In order to quantify the residual dependencies of gene expression due to the increase in gene copy number as the cell cycle progresses, we averaged the same data conditioning by cell cycle phase [32]. This procedure makes it possible to average together cells with all doubling times, hence increasing the statistical power (3-8000 cells in our case). We found that the GFP production rate from the constitutive promoter was biphasic, which appears more clearly when the promoter is inserted close to the replication origin. Rescaling by both cell volume and by mean gene copy number left some residual oscillations (Figure H.6C), consistent with the timing of cell-cycle dependent increase in gene copy number of the constructs inserted in two different chromosomal positions. Interestingly, while the volume-specific expression rate of GFP from the insertion near the terminus is constant until the gene copy number increases, the one of the GFP from the construct near the origin decreases until the gene is copied by the passage of the DNA replication forks. In the first half of the cell cycle, the increase in volume as the cell cycle progresses is greater than the increase in GFP production rate by transcription and translation. Similar results were obtained with the reporter cassette using the shortened version of the rrnBP1 promoter (see Chapter 5).

In summary, the production rate of an unregulated promoter can be described as linearly proportional to volume, likely through the dependence of protein production rate on ribosome amounts, and to dosage through the increase in the gene

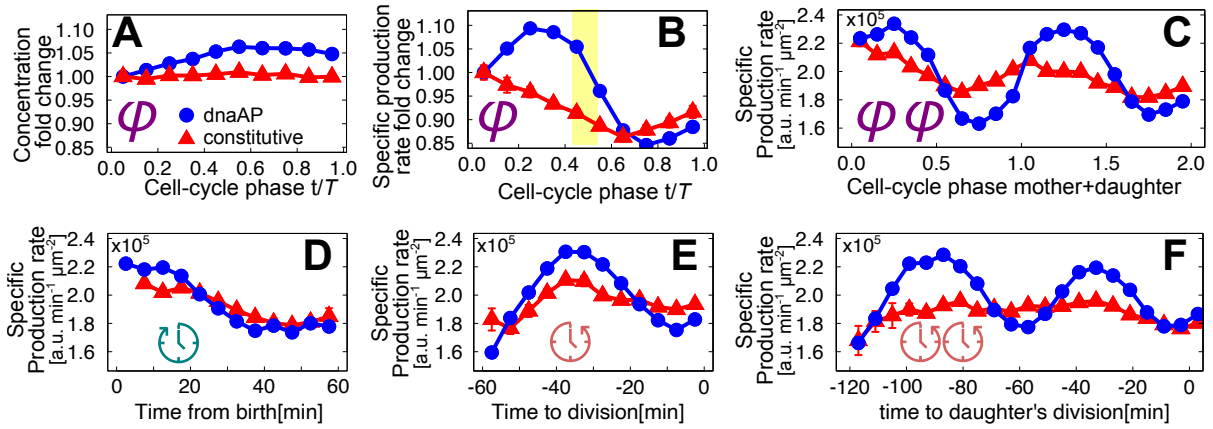


Figure 6.3: The activity of the dnaAP2 promoter oscillates as a function of the cell cycle. ***A*** *Oscillations in GFP concentration as a function of cell cycle phase are weak for both the constitutive (red triangles) and the dnaAP2 promoter (blue circles).* ***B*** *The fold change in volume-specific GFP production rate from dnaAP2 shows a clear peak which is not present for a constitutive promoter (red triangles), which instead changes with the increase in gene dosage.* ***C*** *A sinusoidal oscillation for GFP expression from dnaAP is observed when taking into account the cell cycle phase of two consecutive generations.* ***(D,E)*** *Volume-specific GFP production rate averaged as a function of time from birth is very similar for the two promoters, while oscillations are enhanced for the dnaAP2 promoter when the same data are averaged as a function of the time to division.* ***F*** *Oscillations in GFP expression from dnaAP2 are also coupled with time to daughter division, consistent with DNA replication initiating in the mother cell to terminate in the daughter cell. Error bars (often smaller than symbol size) are standard errors of the mean from a re-sampled distribution obtained by bootstrapping from the experimental data for each bin.*

copy number. Correlating GFP expression rate with the estimated gene replication times along the cell cycle, instantaneous gene dosage may also affect protein production in a cell cycle phase dependent way.

6.2.4 Volume-specific dnaAP2 activity oscillates symmetrically with cell cycle phase and cell volume

Having defined the null features of promoter activity, we set out to ask whether along the cell cycle changes specific to dnaAP2 activity differed from those of a constitutively expressed gene. To this end, we performed (conditional) averages of many cells, by fixing cell cycle phase and cell cycle timing.

Our previous analysis (Figure H.6) tells us that a constitutive promoter shows changes in GFP production rate that depend on cell size and gene dosage. The change in GFP concentration as a function of cell cycle phase shows a small but reproducible difference between the two promoters inserted at the same ori-

proximal position (Figure 6.3A). While the GFP concentration from the constitutive promoter remains constant, the one the dnaAP2 promoter increases slightly (roughly by 5%) in the first half of the cell cycle. The average volume-specific GFP production rate from dnaAP2 clearly shows a peak that is not present in the data for the constitutive promoter, and varies by roughly 20% in a cycle (Figure 6.3B). Moreover, if we consider two consecutive cell-cycle we observe a strong periodic oscillation (Figure 6.3C).

We can gain more insight on these oscillations by performing averages of the volume-specific production rate that are conditioned on other cell cycle variables. For example, by averaging the data as a function of the time from cell birth the specific GFP production rate from dnaAP2 becomes indistinguishable from that of the constitutive promoter (Figure 6.3D), however the oscillation is visibly greater (about 5-fold) than that of the constitutive promoter once we average the data as a function of the time to division (Figure 6.3E), suggesting that the oscillation timing is somewhat agnostic of birth, and synchronized with cell division. Figure 6.3E shows that the difference between dnaAP2 and P5 is even greater when time to daughter's division is used to bin the data (Figure 6.3F). These results show more clearly how oscillations in volume-specific GFP production rate are symmetric and sinusoidal when under control of the dnaAP2 promoter.

6.2.5 Oscillations in dnaAP2 activity are strongly coupled to cell size

We proceeded to test the hypothesis that DnaA activity is a cell-size sensor. Figure 6.4AB shows that volume-binned averages in GFP volume-specific production rate from dnaAP2 follow strong oscillations reaching maxima and minima, roughly placed at multiples of a characteristic volume. These oscillations are thus strongly coupled to the cell volume. The fact that maxima and minima are equally spaced suggests that they may be placed at a fixed volume per origin [63, 72].

In order to shed more light into the size-sensing properties of the dnaAP2 promoter, we performed different conditional averages of volume-specific GFP production rate (the oscillating variable) considering different cell-cycle variables (Figure 6.4C). We performed these averages further grouping cells based on their size at birth, as we figured that the variables that are more strongly coupled to the oscillations should be insensitive to variations of any extrinsic variable, and in particular if the oscillator is a true volume sensor, it should have no memory of size at birth. We divided cells into 11 different birth-size classes, and considered binned averages of specific dnaAP-dependent GFP production rate oscillations as a function of

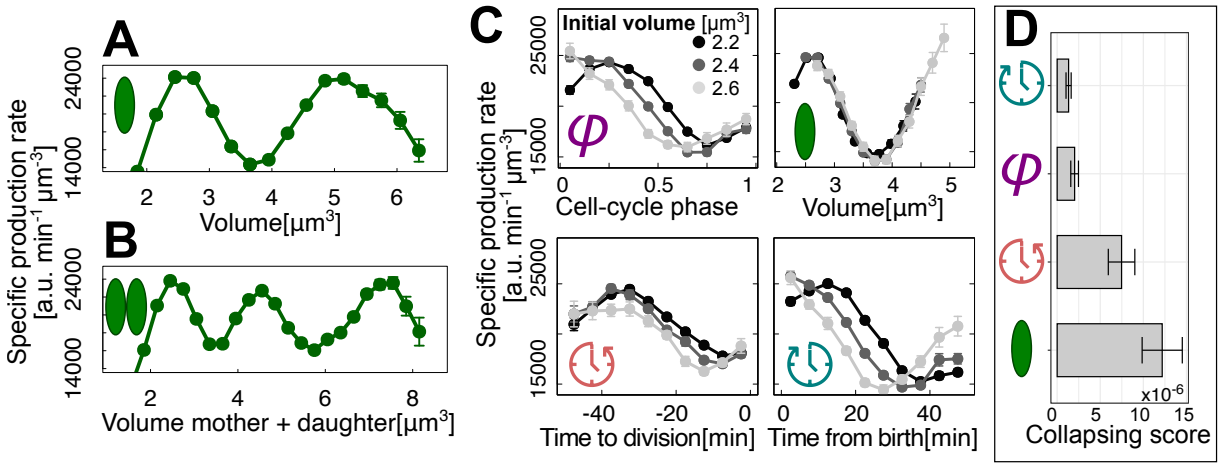


Figure 6.4: DnaAP2 oscillations strongly couple to cell volume. ***A*** *DnaAP2 oscillations are strongly coupled to cell volume. The plot is a conditional average of GFP volume-specific production rate from the dnaAP2 promoter as a function of cell volume.* ***B*** *The same plot on a mother+daughter lineage shows two minima at multiples of a characteristic volume.* ***C*** *Volume-specific dnaAP2 activity oscillations for cells of different initial size show different degrees of overlap when conditionally averaged as a function of cell cycle phase, cell volume, time to division or time from birth. The differently shaded curves result from different bins of cell size (volume) at birth ($2.2 \pm 1\mu m^2$, light-grey , $2.4 \pm 1\mu m^2$, dark-grey and $2.6 \pm 1\mu m^2$, black). Data for cell volume show the best collapse, indicating that the dnaAP2 oscillator is strongly coupled to volume.* ***D*** *Quantification of the collapse of the curves of panel C relative to 11 bins of cell sizes at birth. Binning the data by cell size shows the best collapse and as a function of the time to division shows a good collapse. Error bars are standard errors of the mean obtained by bootstrapping from the experimental data for each bin.*

cell-cycle phase and cell volume. Figure 6.4C shows three birth size bins relative to an average birth size of $2.2 \pm 0.1 \mu m^3$, $2.4 \pm 0.1 \mu m^3$ and $2.6 \pm 0.1 \mu m^3$. The data show that volume-binned oscillations are strongly insensitive to birth size, while binning by time from birth or cell-cycle phase does not result in a good collapse. Binning by time to division determines an intermediate level of collapse. We interpret these two findings as indications that the dnaAP2 oscillator is a potent sensor of cell size, and that the timing of initiation of DNA replication that it contributes to determine has an effect on cell division. Notably, the same oscillations are not coupled to time from birth, despite exponential growth imply a proportionality between cell size and time from birth.

To summarize this behavior, we defined a quantitative score of the collapse of different as inverse of the sum of SE-normalized distances between dnaAP oscillations for all 11 birth-size bins (Figure 6.4D). The higher the score, the higher the collapse of the oscillations for cells with a different birth size. Figure 6.4D shows that cell volume and time to division give the highest scores, which are respectively five and two times higher than those for cycle phase and time from birth.

Until now we talked about volume sensing but its length or its surface could also have an important role in sensing. Supplementary Figure H.10 compares different proxies of cell size (length, surface, volume) as candidates to couple with the dnaAP2 oscillator, showing that volume is the best candidate [72].

In summary, the dnaAP2 oscillator is strongly coupled to cell volume and shows stronger coupling with time to division than it does with time from birth or cell-cycle phase.

6.2.6 dnaAP2 oscillations are due to the combined activity of DnaA-ATP transcription regulation and SeqA binding to the dnaA promoter.

The next step is to understand what leads to these oscillations. In Chapter 3 we have seen that the dnaA promoter activity is regulated by several factors. Here, we focus on how autoregulation and SeqA affect DnaA expression as a function of the cell cycle.

In Figure 6.5 we observe that removing the binding sites for repression and activation causes the disappearance of the oscillating cell-cycle dependence. Only in negatively and positively auto-regulated strains the specific production rate presents strong oscillations within the cell cycle, while the promoters not regulated by DnaA present a phase dependence similar to the one of a constitutive promoter.

We also verified that a ribosomal promoter, exactly as the DnaAP promoter itself stripped form any regulation, does not show this oscillation (Supplementary Figure H.7), suggesting that ppGpp levels and supercoiling, sensed by the discriminator region of this promoter, do not play an important role, but the oscillation is triggered specifically by the ATP-DnaA autoregulatory sites.

On the other hand, removing the downstream binding sites for SeqA, does not cause the disappearance of the oscillating cell-cycle dependence but an alteration of its timing with respect to cell cycle phase (Figure 6.6 and Supplementary Figure H.8).

When we look at the oscillation as a function of cell volume we observe a basal oscillatory pattern for all the promoters (Figure 6.7 and Supplementary Figure H.9). The difference here is in the timing of the peaks, that result altered in the mutant not regulated by SeqA.

Lastly, in Figure 6.8 we observe that the specific production rate for cell classes with different size at birth shows a higher collapsing score as a function of volume and time to division in presence of negative regulation (cfr Figure 6.4).

The data obtained with these strains shows that these oscillations and their coupling to cell volume depend on both activation and repression by DnaA-ATP and that the binding of SeqA following gene duplication by the replication fork is required to regulate the precise timing of the minimum and the maximum. GFP expression from dnaAP2 is then a cell-cycle dependent oscillator, driven by ATP-DnaA autoregulation much beyond the passive effects of replication-related gene dosage changes. A simple model of dnaAP regulation suggests that a maximum of dnaAP oscillations corresponds to a minimum in DnaA-ATP concentration, and a minimum in the oscillation to a maximum ATP-DnaA concentration (see Supplementary methods H). Since we know that the average initiation of DNA replication occurs between a maximum and a minimum (Figure 6.3B), we conclude that both minima and maxima are a candidate proxy for the link of the DnaA circuit to replication initiation. It is possible to interpret the minima as a consequence of the beginning of de-repression of dnaAP due to regulatory inactivation of DnaA and loss of SeqA dependent repression [21].

6.2.7 At the single cell level, dnaAP2 oscillation minima define an adder consistent with the inter-initiation adder.

While all the above analyses firmly establish the coupling of the dnaAP2 oscillator with cell size and cell division using conditional means, they are insufficient to characterize the behavior of single cells [63]. In particular, the correlation signatures

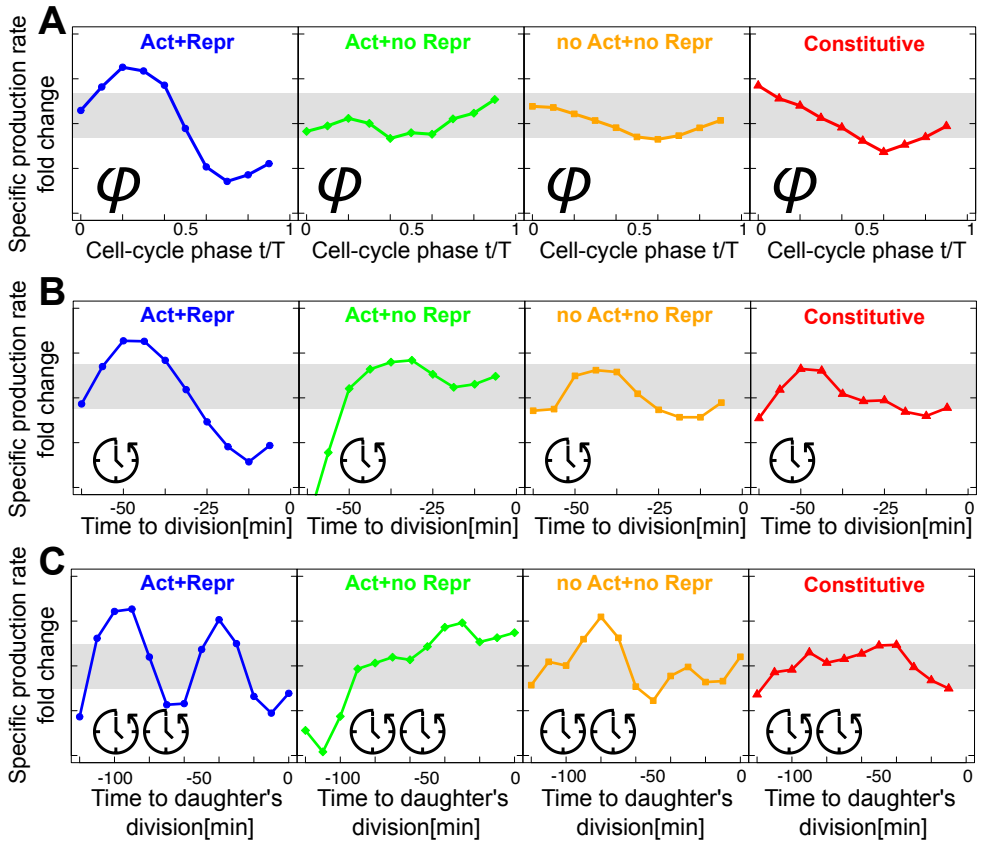


Figure 6.5: Removing the binding sites for repression and activation causes the disappearance of the oscillating cell cycle dependence. *Specific production rate fold change as a function of cell cycle phase(**B**), time to division(**B**) and time to daughter division(**D**). Only in negatively auto-regulated strains the specific production rate presents strong oscillations within the cell cycle (blue lines). The promoters not negatively regulated by DnaA (green and orange lines) present a phase dependence similar to the one of a constitutive promoter(red line).*

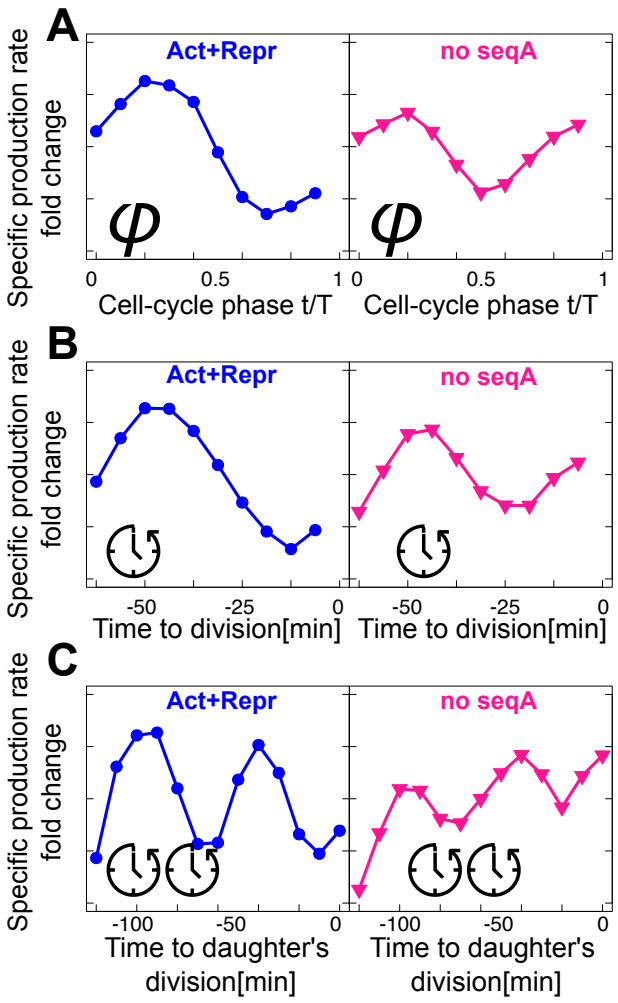


Figure 6.6: Removing the binding sites for SeqA does not cause the disappearance of the oscillating cell-cycle dependence but an alteration of it. *Specific production rate fold change as a function of the cell cycle phase(**A**), time to division(**B**) and time to daughter division(**C**). The negatively auto-regulated strains (blue line) present strong oscillations within the cell cycle also when it is not regulated by SeqA (pink lines) although the peaks' timing is altered.*

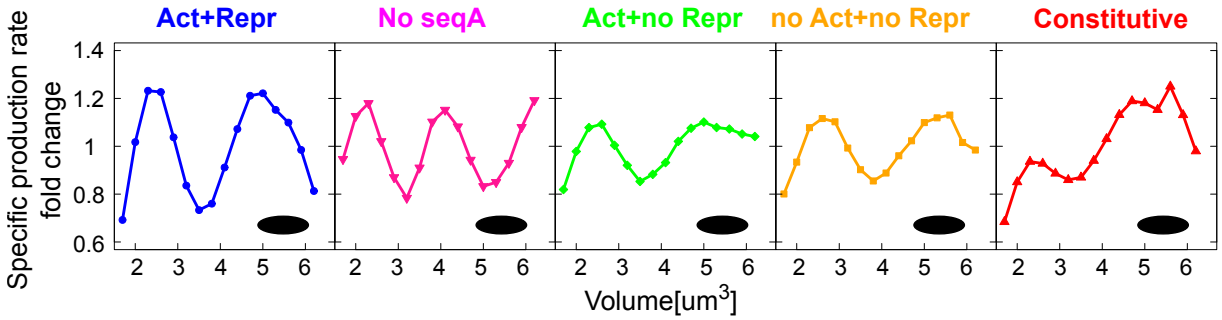


Figure 6.7: Oscillations as a function of volume are robust to mutations although the peaks' timing is altered in the SeqA mutant promoter *Specific production rate fold change as a function of cell volume. The oscillation is robust to mutations although the peaks' timing is altered in the SeqA mutant promoter.*

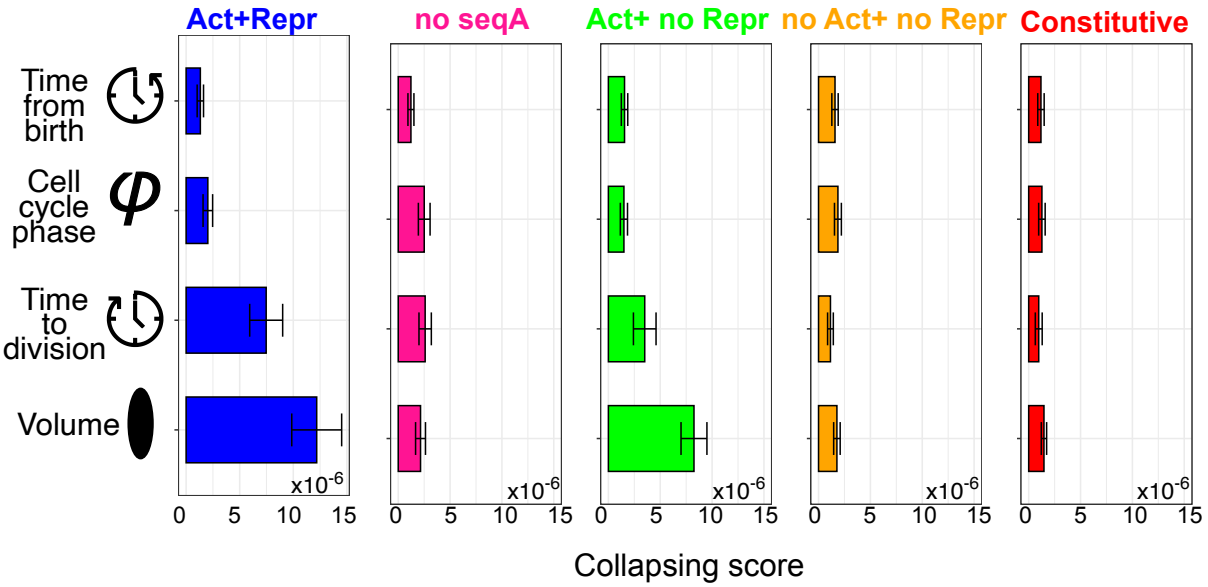


Figure 6.8: Specific production rate for cell classes with different size at birth shows a higher collapsing score as a function of volume and time to division in the presence of negative regulation. *Collapsing score, defined as the inverse mean L_2 distance, is higher as a function of volume and time to division in the presence of negative regulation and with SeqA regulation (compare blue and pink data). Also the promoter that is only activated (green) presents a good collapse with cell volume.*

linking the key events of DNA replication and cell division are well known [8, 10, 11], and we can ask how single cell dnaAP2 oscillations relate to them. Figure 6.9A shows that dnaAP2 oscillations are detectable at the level of a single cell track (Supplementary Figure H.11). We defined an automated algorithm that extracted for every cell and lineage the local maxima and minima of the dnaAP2 oscillations. This gives us two key cell cycle events related to dnaAP2, plus cell division, for each cell and lineage, where we know instantaneous cell size, growth rate, and interdivision time of each cell.

Because of the DnaA-ATP specific auto-regulation and the known features of the DnaA-initiation circuit [21], we expect that minima of the oscillator in single cells are related to initiations triggered by DnaA. Our data lack a proxy for the initiation of DNA replication, but we can ask whether the dnaAP2 oscillations maxima and minima in single cells follow similar patterns to the ones recently observed for initiation of DNA replication [10, 11, 62]. We focus in particular on the minima, which could be interpreted as occurring downstream of initiation of DNA replication, the moment where DnaA-ATP is near its highest activity (the equivalent results for the maxima are shown in Supplementary Figure H.13). Furthermore, the timing of initiation of DNA replication can be estimated for a given growth conditions

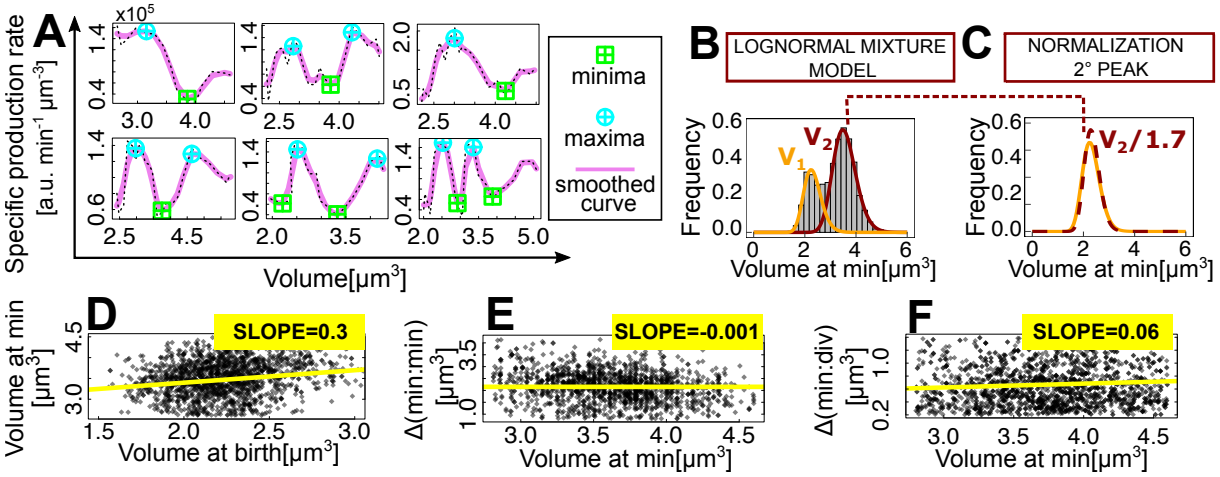


Figure 6.9: In single cells, minima and maxima of dnaAP2 oscillations appear at multiples of a characteristic volume and follow the known correlation patterns of initiation of DNA replication. ***A*** *Oscillations are detectable at the level of single cells. The plots show some single cells dnaAP2 tracks (dashed lines) as well as the smoothed tracks and the local maxima and minima found by our automated algorithm (see Methods).* ***B*** *The distribution of cell size at minima is bimodal with peaks at discrete cell volumes. The bar plots show histograms of cell size at dnaAP2 minima, and the solid lines are a fit with a lognormal mixture model, which results in two lognormal distributions.* ***C*** *If the minima are separated by an adder, the two distributions should collapse if we divide the mean of the second peak by 2. The best collapse is obtained for a factor of 1.7.* ***D*** *Volume at minima is not constant as a function of volume at birth, i.e. minima are not a sizer.* ***E*** *Volume added between two consecutive minima is independent on volume at the first minima (adder between minima).****F*** *Volume added between minima and the next division is independent on volume at the first minima (adder between minima and division).*

thanks to the known growth rate dependence of the C and D periods [49, 72]. This has been already shown in the previous chapter (Figure 5.2 and Appendix G).

In the the slower-growth condition the distribution of cell sizes at dnaAP2 minima is unimodal (Supplementary Figure H.12). Instead, in the faster growth condition we found that the distribution of cell sizes at the minima is bimodal (Figure 6.9B). This is in agreement with the fact that the distribution of cell sizes at initiation of DNA replication in the presence of overlapping rounds of DNA replication should be approximately the sum of two lognormal distributions with means that are one the double of the other [59]. This is because in some cases, at fast growth, initiation of DNA replication takes place before cell division has occurred, and thus at twice the number of origins and at a cell volume that should be proportional. Applying a lognormal mixture model to separate the two distributions we tested whether they would collapse dividing the mean of the second distribution by two. In our data, the best collapse is achieved by dividing by a factor of 1.7, which is very close to two. The small discrepancy could be due to correlations between the volumes at initiation and the probability of extra rounds of replication, previously observed computationally [59].

Next, we decided to investigate the correlation patterns that link DNA replication and cell division. As expected, our cells show adder correlation patterns between cell birth and division independently of cell size at birth (Figure 6.9D) [66, 174]. To test whether the minima are triggered at a critical size (a "sizer") we examined the dependency of the cell size at the minima on birth size (Figure 6.9D). For a sizer pattern (a circuit that triggers at a critical size) there should be no dependency. Instead, the data clearly show some trend, ruling out this hypothesis. On the other hand, Figure 6.9E shows that the added volume between two minima (analogous to the added volume between consecutive initiations) is constant as a function of the cell volume at the first minimum. This recalls the (per origin) inter-initiation adder patterns found by labeling origins or replication-fork proteins in single cells [8, 10, 11, 59, 62]. Lastly, we considered the change in cell size between a minimum and cell size of the daughter at division (analog of the C+D period). The cell size correlations in this period also follow a near-adder pattern as observed previously for the actual C+D period [8, 11, 62] (Figure 6.9F). As anticipated above, similar properties of the distribution and similar correlation patterns are also shared by maxima of the dnaAP2 oscillations (Supplementary Figure H.13).

To summarize, the key timings (maxima and minima) extracted from dnaAP2 oscillations in single cells follow the same correlation patterns followed by replication initiation, corroborating our interpretation that the dnaAP2 cell cycle oscillator

is intimately linked to the single-cell replication cycle.

6.2.8 Volume oscillations are causally coupled to dnaAP2 expression

Beyond cell cycle staging, single cell-level dnaAP2 oscillations can be used to gain further insight into the coupling of this DnaA-dependent cell cycle circuit with cell size and cell division. The standard view of the cell cycle sees it as the result of a single "master-clock" cell cycle oscillator. Having shown that dnaAP2 activity is an oscillator that is strongly coupled to cell size, we figured that the signal we revealed could be a candidate proxy for the oscillator of the cell cycle. These considerations lead us to test the property of the "attractors" (in the language of dynamical systems theory) linking cell size and cell division timing to dnaAP2 oscillations in single cell trajectories. As we detail below, contrary to the expectation of a single master clock, our results suggest that at least two coupled (but independent) oscillators are needed to describe the processes that coordinate DNA replication, cell growth and cell cycle progression.

The term coupled is generally understood to mean “linked” for example through some kind of checkpoint bu here what we mean by “coupled” is that two independent variables such us initiation and division are correlated because they are both dependent on size and on growth dependent accumulation of essential proteins to threshold levels.

The specific activity of dnaAP2, the cell’s volume growth and division events for lineages of variable length are shown in (Figure6.9A). Each of these parameters display oscillations throughout several generations. To test the presence of coupling between specific dnaAP2 activity and volume oscillations, we considered the cross-correlation between these two time series, computed along lineages (Figure 6.9B) [35]. Figure 6.9B shows that this function is markedly periodic, suggesting a strong coupling, with higher-amplitude peaks for positive time delays. This asymmetry of the cross-correlation function, which shows higher-amplitude peaks for positive time delays, could support a "volume to dnaAP2" causality, whereby changes in volume are prognostic to future changes in dnaAP2 oscillations. Importantly, in mutants with deleted DnaA binding sites or in constitutive promoters, these cross-correlations are strongly reduced (Supplementary Figure H.15). Deletion of SeqA binding sites does not suppress the cross-correlations, but alters the pattern and makes it more symmetric (Supplementary Fig H.15).

Finally, while cross-correlations firmly establish a coupling, one has to be careful when inferring causal relations between two observables because of the exis-

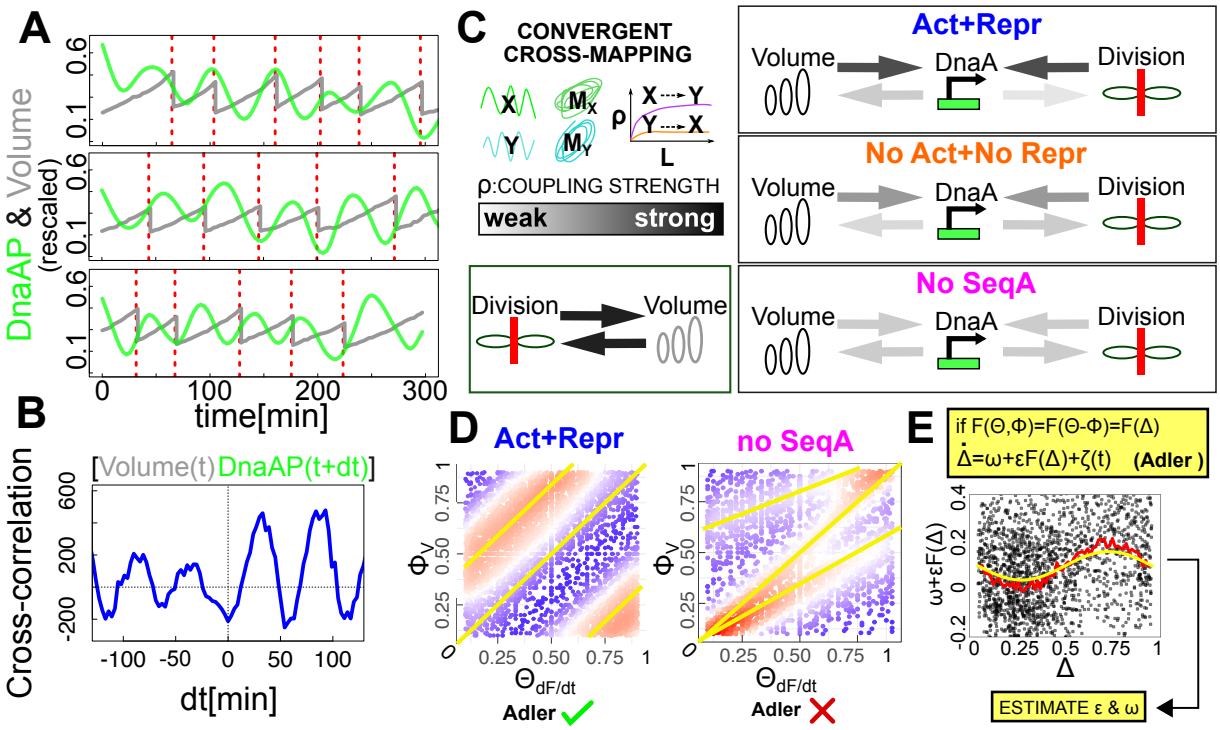


Figure 6.10: DnaAP activity and growth-division are coupled oscillators. A *Lineages of single cells show a clear oscillation of DnaAP activity (green) compared here to volume growth (grey) and division events (red). Volume and DnaAP signal are normalized by their average value to fit them in the same plot.* ***B*** *Cross-correlation between dnaAP activity and volume growth is periodic, supporting synchronization between the two oscillators.* ***C*** *Convergent Cross-Mapping (CCM) identifies causal relationships between oscillators with a common attractor. Volume and division show symmetric causality and are effectively interchangeable, while dnaAP activity presents different causal links that depends on the mutations. In the wild-type promoter, volume and division are strongly causal of dnaAP changes, but the strength of this causal link decreases in the mutants of DnaA-ATP binding sites and the causality becomes symmetric in the mutant without SeqA binding site.* ***D*** *Phase space of the phase of volume (y axis) and of dnaAP activity (x axis) oscillators. For the wild-type promoter, it depends only on phase differences, while this property is disrupted in the mutant without SeqA binding site. As the coupling depends on phase differences, the wild-type promoter follows the Adler equation for coupled oscillators.* ***E*** *The Adler equation can be solved to infer the coupling strength (ϵ) and the detuning (ω). This is tested in simulated data with a sinusoidal coupling function. In red is the rolling average while in yellow the simulated coupling function.*

tence of a non-zero correlation between two signals does not necessarily imply a causal link [229]. To investigate the directionality of the coupling, we used the Convergent Cross Mapping (CCM) technique considering *conditional* correlations of one variable with the a second one which is constrained in a point of its attractor manifold, reconstructed by Takens theorem[230][231] (see Supplementary Figure H.16). These conditional correlations are not symmetric and reveal causal links. Figure 6.10C summarizes the results of our analysis. We detect a volume-to-dnaAP2 causality, and this causal link is weakened in mutants without negative and positive DnaA auto-regulation, as well as in mutants without SeqA regulation. The same analysis shows that volume and cell division are always in a symmetric causal relationship. Consequently, cell division is causally coupled to dnaAP2 oscillations in a much stronger fashion than dnaAP2 causes division. This unintuitive result stems from the strong symmetric coupling between cell division and cell size, and from the fact that the dnaAP2 promoter is a strong size sensor. Importantly, these asymmetric relationships are not detectable by looking at cross-correlations alone (Supplementary figure H.15).

6.2.9 Two phase-locked oscillators govern cell-cycle progression

Having established that dnaP2 and cell division are coupled oscillators, we proceeded to characterize their synchronization properties. Since Huygens' discovery of phase locking [232], this phenomenon has been reported in a variety of contexts [233–235]. These findings show that the synchronized states show common generic properties that can be described by simple predictive mathematical models. Generally, the behavior can be characterized by a coupling strength and a "detuning", describing the difference between the intrinsic frequency of the two oscillators.

In order to capture phase locking between dnaAP2 and cell division, we characterized each oscillator with a phase, a linearly increasing variable reset at each cycle, which advances by 1 between successive cycles. Specifically, we defined a dnaAP2 phase variable Θ, where $\Theta = 0$ corresponds to the minima of dnaAP2 promoter activity and a cell cycle phase Φ as above (Φ = 0 represents birth). These two variables define a periodic square $[0, 2\pi)\ \ [0, 2\pi)$ (this square can be regarded as a torus because opposite sides are equivalent). Subsequently, as time t increases, the combined phase (Θ,Φ) from each single cell lineage traces a trajectory in this phase space. Following single lineages along this space shows that cells follow a diagonal trajectory (i.e. they follow 1:1 "phase-locked" orbits, Figure 6.10D and

Supplementary Figure H.18). Consequently, the phase difference, $\Delta = \Theta - \Phi$ fluctuates around a constant value. Occasionally, one of the two oscillators accumulates an extra oscillation and the phase difference "slips" by one unit (Supplementary Figure H.18). The phase slips are more frequent for the seqA mutant and this suggest that mutation of the SeqA binding site disrupts this phase locked state (Figure 6.10D). These results can be captured by the following general equations describing dnaAP2 production rate(θ) and cell-cycle phase(ϕ) as two coupled oscillators

$$\begin{cases} \frac{d\theta}{dt} = \omega_\theta + \epsilon_\theta F_{\theta\phi}(\theta,\phi) + \zeta_\theta(t) \\ \frac{d\phi}{dt} = \omega_\phi + \epsilon_\phi F_{\phi\theta}(\theta,\phi) + \zeta_\phi(t) \end{cases} , \tag{6.1}$$

where ω_θ and ω_ϕ are the intrinsic frequencies of the two oscillators, ϵ_θ and ϵ_ϕ are the coupling coefficient and ζ_ϕ,ζ_θ represent the noise terms.

Figure 6.10 shows that the wild-type promoter oscillations are concentrated on two juxtaposed diagonal stripes whose slope is one, hence the oscillations only depend on the phase difference. This means that a minimal model of the couplings $F_{\theta\phi}$ and $F_{\phi\theta}$ is a function of phase difference $\Theta - \phi$. In this case, the above equations reduce to a single equation for the phase difference

$$\frac{d\Delta}{dt} = \omega + \epsilon F(\Delta) + \zeta(t) \, , \tag{6.2}$$

where $\omega = \omega_\theta - \omega_\phi$, the detuning, is the difference between the intrinsic frequencies of the two oscillators, $\epsilon = \epsilon_\theta - \epsilon_\phi$ quantifies the overall strength of the interaction, $F(\Delta) = F_{\theta\phi}(-\Delta) - F_{\phi\theta}(\Delta)$ is a coupling function and $\zeta(t)$ noise term. This equation is known in the literature as Adler equation [229], and holds for a wide class of synchronized oscillators [236, 237]. For the SeqA mutant promoter the dependence of the coupling function from the two oscillators is more complicated and the Adler equation is not valid. More details on these theoretical approaches can be found in Appendix H. Since the data support this model, it is possible to infer from data the coupling strength (ϵ) and difference between the intrinsic frequency of the two oscillators (ω) [238]). The inference procedure is simple, and amounts into taking the conditional average of the empirical time derivative of the phase difference, at fixed phase difference. We tested this approach on simulated data where the two oscillators follow Equation 6.1 with a sinusoidal coupling function that depends on the phase difference and we were able the estimate the correct parameters (Figure 6.10E).

6.3 Discussion

The DnaA protein is an important factor regulating the initiation of DNA replication [16]. Its activity depends on its nucleotide bound state, the ATP bound form being the active one for origin recognition and activation of initiation of DNA replication. An oscillation in DnaA-ATP activity is assumed by most, but so far supported only by indirect population-level data [23]. A set of different processes are thought to contribute to the oscillation and help avoid re-initiation of DNA replication within the same cell cycle [15, 17]. The regulated hydrolysis of ATP to ADP coupled to ongoing DNA replication should decrease its activity after initiation has taken place; the binding of SeqA to hemi-methylated DNA at both the origin and the *dnaA* promoter should inhibit access to these sites by DnaA-ATP and RNA polymerase respectively, and DNA replication should result in an increase in the number of DnaA-binding sites. Some of these binding sites can change the nucleotide bound state of the protein and play a different role depending on the cell's growth rate [239]. These results have been the basis for different models that have proposed that the activity of DnaA-ATP oscillates within the cell cycle to trigger initiation [19, 51].

However, one of the major challenges in this field has been to quantify the changes in DnaA-ATP activity *in vivo* in real time, and provide direct evidence for these hypothesized activity oscillations. To address this problem, we have developed a set of reporters of gene expression using a gene for a fluorescent protein under control of a promoter that can be differentially regulated by DnaA-ATP. In previous studies, we have shown that the autoregulation of the expression of the *dnaA* gene takes place via both positive and negative autoregulation [112]. We have used this information to design reporter constructs whose gene expression activity depends on the *in vivo* concentration of DnaA-ATP. Using a microfluidic device coupled to microscopy imaging these strains can be used to follow the changes in gene expression rate during the cell cycle by measuring fluorescence in real time at the single cell level. Thanks to all these efforts, we have been able to identify oscillations dependent on DnaA activity and quantify the effect of the DnaA-dependent promoter regulatory elements that lead to this oscillatory pattern. The study of promoter mutants where the binding of either DnaA or SeqA have been compromised and the comparison with the expression from a constitutive or a growth-rate dependent ribosomal promoter shows that the promoter's regulation by DnaA-ATP causes an oscillation in GFP production rate beyond the effect of dosage. While these experiments have been carried out on a gene expression reporter system, these results can be applied to the autoregulation of the expression of the *dnaA*

gene itself. These results are consistent with DnaA-ATP activity being a cell cycle oscillator, and we have shown that these oscillations strongly depend cell volume and are correlated with the time to daughter cell's division. Both negative autoregulation from DnaA-ATP and SeqA affect the coupling of dnaAP2 oscillations with cell division. Negative regulation is essential to have these oscillations, and SeqA is necessary to couple these oscillations to cell volume and cell division. Hence, the dnaAP2 oscillator is coupled to both the DNA replication and cell division programs.

Remarkably, the signal to noise in our data is high enough that these oscillations are also detectable at the level of the single cell. The minima of these oscillations show the same size-correlation patterns of the initiation timing, detected in other studies by the tracking of fluorescent foci coupled to initiation of DNA replication [48, 240]. While the standard view of the cell cycle sees it as the result of a single oscillator, our single cell data lead us to suggest that at least two coupled oscillators are needed to describe the processes that coordinate DNA replication, cell growth and cell cycle progression. This approach also makes it possible to detect causality links between these different processes. The volume to dnaAP2 causality supports models where chromosome is not always limiting for division [8, 9, 62]. In general, two coupled oscillators linked with specific causality are needed to describe the processes that coordinate cell replication, cell growth and cell-cycle progression.

In eukaryotes, numerous checkpoints exist to ensure that each step of the cell cycle has been completed before proceeding to the next one [241, 242]. Few *bona fide* checkpoints have been identified in bacteria and the mechanisms that ensure orderly progression through the bacterial cell cycle are not completely understood. Our findings combine the use of mathematical models and single-cell dynamic data to pose firmer quantitative bases for a characterization of the mechanisms determining robust cell cycle progression in bacteria.

6.3.1 Future perspectives

Growth regimes play an important role in mechanisms involved in the coordination between DNA replication and cell growth and division. In this work, we tested two growth conditions, and we chose one condition in multiple replication round regime and at slower growth. It would be interesting, although quite time-consuming, to test the results presented in this work also in additional growth conditions.

Moreover, it would be interesting to test the hypothesis that have emerged from my results by making mutations on the endogenous promoters for the DnaA gene as well as its target genes.

We do not know what would happen if we create the same mutations in the native *dnaA* promoter. Probably, because of its central, essential, role in the cell, altering the production of this essential protein would affect too many processes to become difficult to distinguish specific effects, or one could expect that compensatory mechanisms exist in the cell to counteract the problem. In the absence of auto-regulation of *dnaA* expression one would probably observe a perturbation of the timing and synchrony of initiation. People have already done experiments where *dnaA* is expressed from an inducible promoter where cell survive but it is not know what happens to the robustness of cell growth and division control. The chromosomal construction would be useful to detect and measure this effect in a single cell, in particular, under growth conditions slow enough to prevent overlapping DNA replication rounds, where multiple factors, such as supercoiling and increased titration, come into play, and where we have observed the strongest effect of autoregulation.

Oana Ilioaia, engineer in our team, has recently set up a CRISPR protocol to obtain the first set of mutant strains. These and future mutant strains will be useful to characterize the DNA replication properties in the absence of specific regulatory elements. This could be done at first at the population level, by measuring the number and the synchrony of active replication origins by flow cytometry, and then by single cell fluorescence where different sites along the genome have been tagged by fluorescent proteins.

Moreover, in Chapter 5 we have seen how gene position in the chromosome has a large impact on gene regulation within the cell cycle. The native DnaA promoter is placed closed to the origin of DNA replication and it would be interesting to see if the observed oscillatory behaviour is still present for the same promoter placed close to the terminus. This will tell us if there is an evolutionary reason for which the native DnaA promoter is placed close to the origin.

Chapter 7

Early fate of exogenous promoters

Gene gain by horizontal gene transfer is a major pathway of genome innovation in bacteria. The current view posits that acquired genes initially need to be silenced and that a bacterial chromatin protein, H-NS, plays a role in this silencing. However, we lack direct observation of the early fate of a horizontally transferred gene to prove this theory. We combine sequencing, flow cytometry and cell sorting, followed by microscopy to monitor gene expression and its variability after large-scale random insertions of a reporter gene in a population of *E. coli* bacteria. We find that inserted promoters have a wide range of gene-expression variability related to their location. We find that high-expression clones carry insertions that are not correlated with H- NS binding. Conversely, binding of H-NS correlates with silencing. Finally, while most promoters show a common level of extrinsic noise, some insertions show higher noise levels. Analysis of these high-noise clones supports a scenario of gene expression switching due to transcriptional interference from divergent ribosomal promoters. Altogether, our findings point to evolutionary pathways where newly-acquired genes are not necessarily silenced, but may immediately explore a wide range of expression levels to probe the optimal ones.

The work of this chapter is part of a collaboration funded by CEFIPRA between the groups of Bianca Sclavi, Marco Cosentino Lagomarsino and Aswin Sai Narain Seshasayee, at the NCBS, Bangalore. Dr Malikmohamed Yousuf, at the time was a postdoc funded by this collaborative project and he is currently at the Centre for Clinical Brain Sciences at the University of Edinburgh. The experimental part was all performed by Malik, the DNA sequencing was carried out at the NCBS, while I developed the entire data analysis framework. In particular, I have analysed data from flow-cytometry, whole genome sequencing, nanopore sequencing, plate reader experiments and single cell microcolony growth assays by epifluorescence microscopy. Moreover, I have set the statistical analysis protocol and mathematical

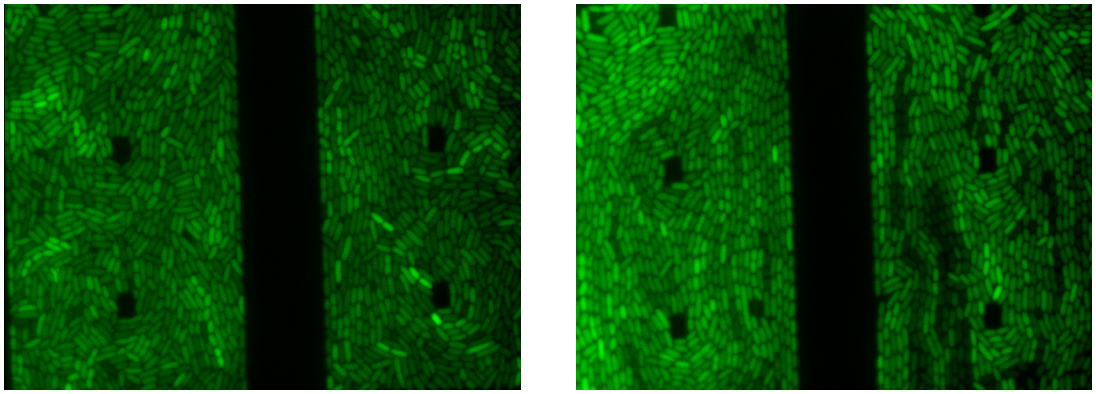


Figure 7.1: Noisy clones in a microfluidic chamber *Two field of view of a microfluidic chamber. The population present a high gene expression variability.*

model to investigate statistical tendencies for transposon insertions to be associated to specific chromosomal contexts. In addition, I have also performed a few experiments in microfluidic chambers to see how this noisy clones behave when growth conditions are changed (Chapter **??** and AppendixC). In Figure 7.1 preliminary results are reported. It will be very interesting to investigate more deeply how this noise propagates in a population of single cells by changing environmental conditions.

This work has been published in 2020 and I am co-fist author [27] and the published paper is reported in this chapter.

7.1 Introduction

The high fraction of mobile genes in bacterial genomes is a source of a great diversity of phenotypes. This large diversity challenges the very concept of species, and has enormous importance for understanding pathogenicity and antibiotic resistance [243]. At the genetic level, *E. coli* genomes vary dramatically in their sizes – ranging from 4.5Mb to 6Mb. Comparative genomic surveys of *E. coli* have shown that there is a core set of genes which is highly conserved across the species and coexists with a large pangenome, the set of genes that can be gained by horizontal gene acquisition from other species [244]. Indeed, bacteria acquire exogenous DNA by transformation (of naked DNA from the environment), transduction (of DNA from bacteriophages) or conjugation (from fellow bacteria through molecular pipes such as pili) [245]. In order to be functional, exogenous acquired genes often need for the metabolic and the regulatory circuitry of the cell to be rewired [246, 247]. Furthermore, expression of a foreign gene can interfere with the resources allocated for endogenous gene expression. Therefore, horizontally acquired genes

must be regulated [248].

A primary mode by which the expression of horizontally-acquired genes is regulated is believed to be transcriptional repression, which is achieved by proteins such as H-NS in enterobacteria, including *E. coli* [215, 249–251](reviewed in ref. [252]). Many previous studies support both the need of initial repression of acquired genes, and the view that H-NS repression is relevant for the successful establishment of these genes [247, 248, 253, 254]. H-NS is among several "global" transcriptional regulators that affect the expression of hundreds of genes in *E. coli*. It binds to AT-rich or intrinsically bent DNA sequences and forms structures such as stiff rods or DNA-protein-DNA bridges, which might act as geometrical motifs for transcriptional silencing [255, 256]. Many H-NS binding regions are up to a few kilobases long. The length of these binding regions correlates with the degree of transcriptional repression imposed on the target gene [214, 215] and genes regulated by H-NS are very highly expressed in the absence of this repressive control [257, 258]. Since the levels and activity of H-NS depend on environmental conditions and growth rate, this level of regulation allows for a coordinated gene-expression change needed for cellular adaptation. Studies of gene expression of inserted reporter cassettes at different genomic locations [213, 214] have demonstrated that gene expression of an identical regulatory system can vary greatly, beyond the effects of gene dosage, for three main reasons, supercoiling, activity of neighbour promoters and H-NS regulatory activity. A gene's local environment can thus provide a fitness advantage, associated to the selection of the gene's position over evolution.

Additionally, at least through the action of H-NS, which is a notorious nucleoid-shaping protein [259], the dynamics of horizontal transfers is related to the physical organization of the chromosome. An important question to be addressed is whether and how the organizational features of the *E. coli* chromosome (such as the "macrodomain" architecture [260–262] are correlated with gene acquisition and control of gene expression, particularly of acquired genes [263].

Horizontally transferred genes are often clustered along the genome [245, 261, 264–266]. In part, this reflects joint transfer of functionally co-dependent genes that would provide no benefit if transferred independently. In part, however, this reflects the existence of "permissive" zones along the chromosome, which experience recurrent integration and high turnover. Permissive zones can originate or be

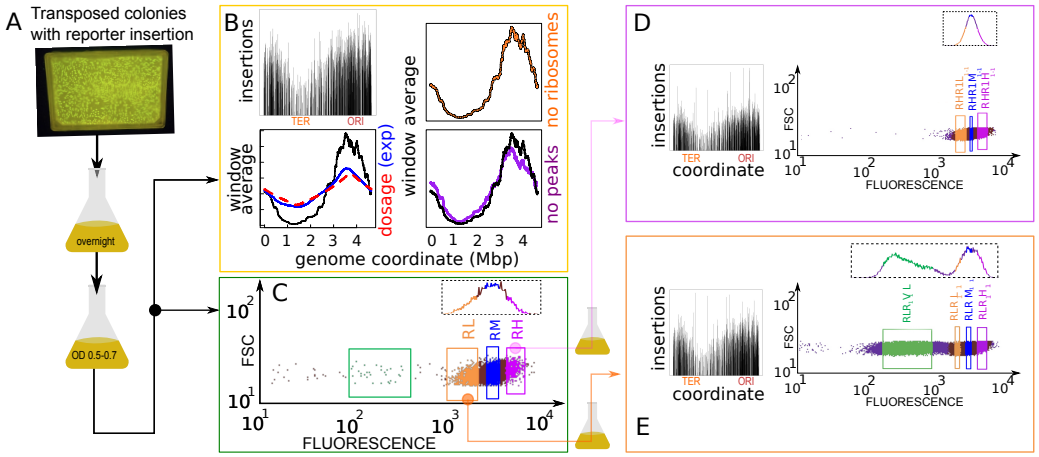


***Figure 7.2: Insertion localization and sorting by gene expression**. A: Experimental pipeline. Massive transposon insertion of a GFP reporter gene cassette in around 100000 founder strains was tested by plating on kanamycin-selective agar and PCR. Surviving colonies were mixed, grown overnight in LB, resuspended and grown to a fixed OD. B: Sequencing of resulting parental populations yields the locations of the insertions, shown in the top-left panel (y axis are counts in logarithmic scale). The bottom panel compares a 3kb sliding average of the coverage (black line, y axis rescaled for comparison) with the prediction from gene dosage, and the experimental dosage (red dashed line) measured by whole-genome sequencing (blue line) the right panels are controls that the trend of insertions copy number is not due to ribosomal genes (orange line) and to the insertions with top 10% coverage (> 3000 reads/bin, purple line). C,D,E: forward scatter vs GFP expression measured by flow-cytometry. FACS Sorting by the level of fluorescence was performed on a total of four rounds (see Supplementary Fig. 1). Selecting for high expression (RH) from the parental population (C) yielded a population with a similar distribution of gene expression (D), while selecting for low expression yielded a population with a bimodal distribution of gene expression (E). Insets in panels D and E show insertions found by population sequencing (y axis are counts in logarithmic scale), with overall similarity but local differences.*

reinforced through physical integration biases, where the presence of integrases and/or recombinogenic sites facilitates acquisition of genetic material [243]. Additionally, in many species including *E. coli*, horizontally acquired genes preferentially accumulate near the (AT-rich) terminus region [243, 244, 261], possibly to avoid deleteriously high expression near the origin due to gene copy number effects.

The genome sequence organization of a given species is a result of selection pressure and architectural constraints. Some of these have been clearly identified [245, 248, 257, 266, 267]. For example, highly expressed, newly acquired genes must be kept from interfering with the expression of essential genes. However, comparatively little is known about the early dynamics of acquired genes. Do clear physical insertion biases emerge? What are the phenotypic impacts of inserted genes and how are they linked with expression levels? How, in turn, is gene ex-

pression a consequence of the locus of insertion? Are they immediately silenced and do H-NS and nucleoid organization play a role?

To access some of the above questions, we devised an experimental assay (Fig. 7.2) where a cassette including an antibiotic resistance gene and a GFP reporter under the control of a highly expressed ribosomal promoter is inserted systematically in the genome, and the resulting mixed and clonal populations are analyzed by sequencing and single-cell biology methods. This methodology allows us to describe statistical tendencies for a reference promoter to be inserted and initially maintained in specific chromosomal contexts, as well as to characterize its fate in terms of both gene expression activity and noise of the transcription reporter constructs at different insertion sites on the genome.

7.2 Results

Detailed Methods are available as Supplementary Materials (Appendix I).

7.2.1 Efficient protocol for production and characterization of systematic exogenous reporter insertions.

The promoter chosen to control GFP expression in the randomly inserted cassette is the rrnBP1 promoter of the rrnB ribosomal operon. We chose this well-characterized highly expressed promoter because it is regulated by changes in DNA supercoiling and by the abundant nucleoid proteins Fis and H-NS [52]. We also considered, as a control, a shortened version of the promoter lacking regulation by the nucleoid proteins Fis and H-NS but still regulated by DNA topology.

Fig. 7.2A describes our pipeline (see Methods and Supplementary Fig. S1). On the order of 10^5 transposed colonies (this estimate was based on manual counting as we knew the number of transposed colonies in each plate) were mixed and grown overnight in minimal medium. This population was regrown to a fixed OD in the same medium, and initially assayed by population sequencing (Fig. 7.2B and flow cytometry Fig. 7.2C,D,E). We then used a cell sorter to select sub-populations based on gene expression levels (Fig. 7.2C,D,E). Each of these subpopulations was grown overnight, regrown to exponential phase and then sorted again as a function of GFP content, for a total of four rounds (Supplementary Fig. S1). Finally, 658 randomly hand-picked clonal populations were individually characterized by flow cytometry. A subset of 96 from the 658 clonal populations representing different sorted populations were randomly selected and sequenced, and 90 of these were

used to measure gene expression and growth rate in a plate-reader assay (see Supplementary File SF1). A smaller selected subset of clones was used to measure the dynamics of gene expression in single-cell microcolony growth assays by epifluorescence microscopy.

7.2.2 Bimodal distribution of gene expression in parental populations and low-expression sub-populations.

Comparison of the fluorescence distribution in the sorted populations obtained from the high (RH) and low-expressiing (RL) fractions of the parental population (Fig 7.2C,D,E) shows that the low-expressiing (RL) population have a sub-population of clones with very low expression. In the parental population (Fig 7.2C), some of these low-expression clones are already visible (green box). Sorting them from the RL population gave rise to the population RLR1V1L. Outside of this low-expression peak, the distribution of gene expression has little variation in the sorted populations compared to the parental one. This variability is the combination of the variability of promoter expression across single cells that are clonal (i.e. where the insertion is in the same exact position) and the variability of mean expression between clones with different insertion locations. Thus, the clonal variability should be considerably high in order to account for the fact that the overall pattern of variability is robust in the sorted sub-populations (which contain less clonal variants). There are, however, some important differences in the distributions, mirrored by differences in the location and frequency of the insertions in the sorted populations, which turn out to be significant (see below).

7.2.3 Insertions are non-uniform and sparse and are more biased towards the replication origin than justified by gene dosage.

TraDIS Sequencing of FACS-sorted populations based on GFP expression shows the presence of transposon insertions at different chromosomal positions. Coverage of the insertions is uneven and sparse (Fig. 7.2B,D,E). In addition, there is a bias with respect to genome coordinate, with a higher insertion frequency close to the replication origin and a lower insertion frequency close to the terminus. The distributions of insertion frequencies in the parental populations of the P1-short and P1-long promoter insertions showed the same qualitative features as a function of genome coordinate (Supplementary Figure S2). Populations derived from the high

or medium GFP expression populations show a bias for insertions closer to the origin of replication (Supplementary Figure S2). We also noted that populations derived from the low GFP expression populations, particularly those filtered for very low expression levels, showed high-frequency insertions in ribosomal regions (Supplementary Figure S2).

We tested a possible role of gene dosage in the origin-to-terminus bias of insertion frequency. The samples are in early log phase in LB medium at 37^{o}C when they are exposed to the transposon. There is therefore a higher number of copies of the chromosome close to the origin than to the terminus. Estimating the dosage from the Cooper-Helmstetter model [43], and assuming an insertion rate proportional to the dosage, we computed the expected insertion bias, keeping into account the population age-structure (see SI text).

Fig. 7.2B shows that the dosage estimated theoretically agrees very well with whole-genome sequencing of genome copy number, but is not sufficient to explain the stronger origin-terminus bias of the insertions. We also verified that this bias was not due to the insertions with top 10% coverage and to the insertions on ribosomal genes. The additional bias may be due to additional factors such as DNA supercoiling or biased binding of nucleoid proteins and differences in nucleoid compaction [151, 268–270]. Additionally, the density of insertions shows a slight left-right asymmetry with respect to the origin, which is visible when the sliding average of insertions is compared with the prediction from dosage (Fig. 7.2B). We verified that a model with time-dependent insertion rate, i.e. where the insertion rate r increases with time t, $r(x,t)$, can fit the data, using insertion rate growing as a power law in time (see SI text). However, there is no empirical motivation to assume such cooperative behavior in insertions that occur *in different cells*. Alternatively, since the exponent linking expected dosage and measured insertions is close to three, one can also hypothesize a cooperative effect of technical or biological origin, but we could not produce a technical or biological explanation for such a simple cooperativity.

7.2.4 H-NS binding sites are enriched at insertions positions

In order to better characterize the genomic positions of the insertions, we investigated the statistical tendencies for localization of insertions using the gene lists from the NuST database [271]. This database contains a large panel of published gene sets measuring several genomic properties such as binding of nucleoid-associated proteins, including several H-NS data sets (see Supplementary Table 1 for a detailed description of each data set). To score for significance, we compared

the co-occurrences of insertions and genes with 5000 realizations of a shuffling null model (see Supplementary Methods for details). Note that the null model subtracts the empirical sliding average of insertions, and not the dosage, thus the results are net of the overall enrichment around the origin. The analysis was applied to the population-sequencing data for the insertion sites in both the parental populations as well as in the ones that were sorted for gene-expression levels.

This analysis, summarized in Fig. 7.3A, shows that H-NS binding is the main property associated with any insertions (even before any sorting by gene expression is performed). The light blue circles in this figure refer to different genome-wide H-NS occupancy (ChIP-ChIP and ChIP-seq) data sets, obtained in different conditions (see Supplementary Table 1). The dark blue circle refers to genes that are sensitive to H-NS knockout under perturbations that make supercoiling more positive [218]. The crossed square refers to generic transcriptionally silenced extended protein occupancy domains, which are largely made of H-NS bound regions. All these sets are correlated but not identical, and essentially contain different categories of H-NS bound regions.

Importantly, a strong enrichment is shared with putative horizontal transfers detected from sequence properties (among which AT-richness [272], red diamond in Fig. 7.3A). Indeed, the full list of insertion sites is enriched in H-NS target genes regardless of the expression level (Supplementary Table 2). We also found an enrichment on H-NS binding sites in the surroundings (10kb regions) of the insertions compared to random sites (Supplementary Fig. S2).

The positive local association of H-NS binding sites with insertion sites is in agreement with a common preference for AT-rich regions. As previously mentioned, genomic insertions generally have a reported bias for AT-rich regions [245, 273, 274], and AT-rich regions are also the preferred binding targets for H-NS [215, 251, 259, 261]. We looked for a correlation between all insertions (regardless of their association with H-NS) and AT-rich regions. In order to do this, we compared the distribution of AT-bias in the sequences surrounding insertions with a random sample of the background sequences of the genome. A one-tailed Kolmogorov-Smirnov test (p-value $< 10^{-16}$) suggests that there is a significant difference between the two distributions, with the sequences surrounding the insertions being richer in AT than the background ones (Supplementary Fig. S2). The role of H-NS has been proposed to inhibiting insertions, in addition to repressing events of spurious transcription [253, 258, 275, 276]. Our results lead us to conclude that, in the tested conditions, at these fast growth rates, H-NS does not appear to inhibit physical events of transposon insertion efficiently. The correlation analysis performed

here does not allow us to conclude that AT-richness causes the enrichment of both insertions and H-NS binding. Another causal chain is possible, but appears less likely, where H-NS binding facilitates insertions, thereby driving them towards AT-rich regions. We also note that the Tn5 transposon has been reported to be biased towards *GC-rich* regions [277], by a similar analysis than that performed in Supplementary Fig. S2, but in different conditions, and in a different organism. This previous result makes the positive association that we find between insertions and AT-rich regions more intriguing. Additionally, Fig. 7.3B shows that the significant enrichments of the different gene sets for insertions are consistent across the two promoters used here (P1-short and P1-long) used here (see also Supplementary Table 5), as expected from a lack of a role of the donor sequence on insertion bias.

7.2.5 Other global regulators are enriched at insertions positions

We now proceed to discuss other gene sets that share enrichment for any insertions (visible in Fig 7.3A). Of notable significance are targets of global regulators Fis (which alters the nucleoid state to aid transcription in exponential growth) and FNR (which alters the distribution of RNA polymerase in response to oxygen starvation). This could be related to AT-richness bias of the binding site of these proteins or to high transcriptional activity (and thus accessibility for insertions) of these genes [275, 276]. It is reasonable to expect that Fis targets are more active in LB medium. However, we found that transcriptionally active RNAP binding regions (measured by ChIP-chip during rapid growth [278]) are *under-represented* for insertions, which suggests a negative interaction between RNAP binding or transcriptional activity and insertion frequency (Supplementary Table 3). Finally, a milder but significant over-representation for insertions was found for CRP and IHF targets, genes that are sensitive to supercoiling perturbations in an H-NS knockout background, and genes with trans-membrane domains (Supplementary Table 4).

Conversely, essential genes (filled green squres in Fig. 7.3A) are the most under-represented set for insertions, as expected (Supplementary Table 3). The other under-represented gene sets for insertions comprise RNAP targets in rapid growth (crossed dark-green circle in Fig. 7.3A), genes whose promoters are sensitive to supercoiling changes and highly transcribed occupancy domains, suggesting a negative correlation between insertion probability and transcriptional activity.

7.2.6 H-NS binding is the sole over-represented signal in low-expression clonal populations

Finally, we compare the parental population with the ones sorted for gene expression levels (Fig. 7.3C and Supplementary Table 6). The comparison of the insertion sites of the high and low expressing populations from the P1-long promoter strains shows that the low expressing population is found preferentially within H-NS binding regions, while the high expressing population is not. Indeed, the low-expression (RL) population maintains a similar association as the parental population with H-NS, and no other binding protein. Conversely, the RH (high expression) populations show no association with H-NS, and only maintain some enrichment with FNR (Fig. 7.3C and Supplementary Table 6). Hence, despite the lack of an effect in inhibiting transposon insertion, H-NS does appear to regulate the level of gene expression of the inserted sequences.

Overall, these results point to a more complex role than is expected for H-NS in modulating genome accessibility and gene expression of recently acquired genes.

7.2.7 Flow-cytometry analysis of clonal populations shows variable noise and gene-expression properties.

Each of the sorted populations was plated separately to yield individual isogenic (clonal) colonies. To gain further insight into these differences in gene expression, 658 individual clones were hand-picked from the different populations and grown in 96-well plates to measure the average fluorescence and its standard deviation by flow cytometry. From these, 90 clones where chosen to measure both the fluorescence and the growth rate in a plate reader in different growth media.

This analysis yields the following main results (see Supplementary Figures S3,4,5)

- There is agreement between a clone's level of fluorescence and the average level of fluorescence of its original population (Supplementary Fig. S3), as measured by both flow cytometry and the fluorimeter. Specifically, the clones from the low-expressing populations have a significantly lower average level of fluorescence.

- The magnitude of the difference between high- and low-expressing clones depends on the growth medium. The difference is greater in the faster growth medium (Supplementary Fig. S4).

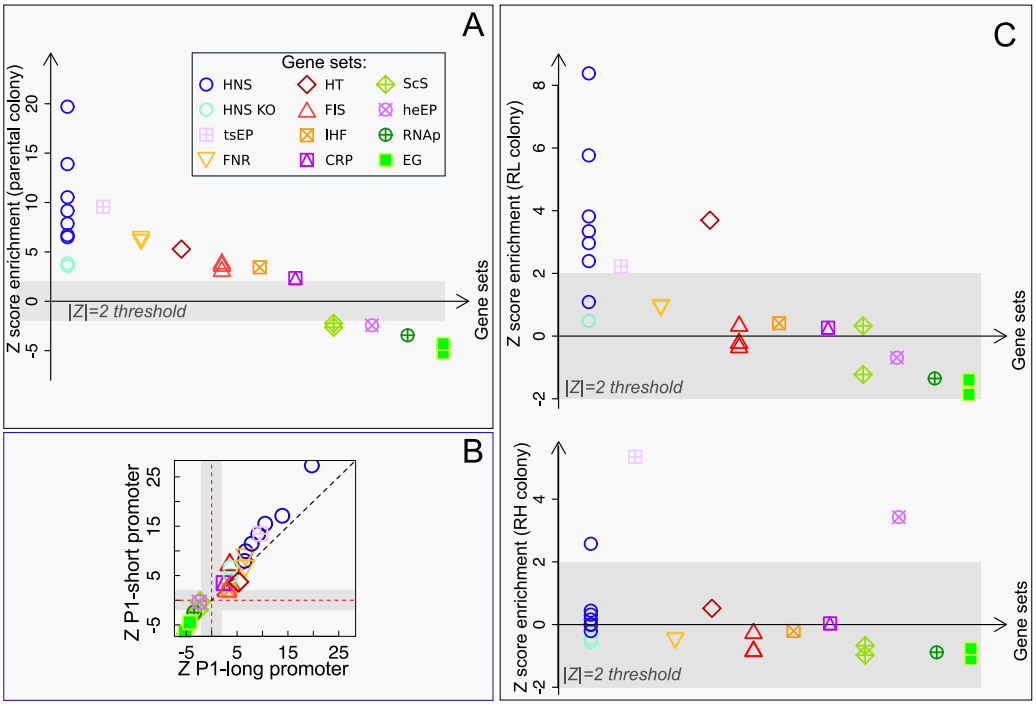


***Figure 7.3: Enrichment of insertions for H-NS and other global regulators**. A. Z-score of enrichment tests for different gene lists (See Supplementary Table 1 for a full legend). H-NS binding sites (from ChIP-seq and ChIp-ChIP data) and H-NS perturbations experiments (from ref. [218]) are highly enriched (circles), indicating a strong positive association of insertions to H-NS binding regions starting from the parental colony. Other global nucleoid regulators (FNR, Fis, IHF, CRP, see legend), and a list of horizontal transfer genes (HT, see legend) also show positive association, lists of essential genes (filled squares) show strong negative enrichment. B. Comparison of the two different promoter tested (with and without Fis and H-NS binding sites) shows a similar behavior. C. Comparison of parental and sorted populations (see Fig. 7.2CDE) shows that H-NS association maintains a strong significance in the low-expression population, and loses significance in the high-expression population, where FNR sites remain highly enriched.*

- The very low-expression strains (Fig. 7.2E) show low expression regardless of the growth media, except for a few outliers showing out-of-trend expression (higher than expected) in the poorer medium (Supplementary Fig. S4). The last two observations are consistent with the known growth-rate dependence of H-NS activity and ribosomal operons transcription rate [214, 279, 280]. This result shows that different regions of the genome can have different growth-rate dependent properties.

- Each clonal population (with a single insertion site, as verified by sequencing) shows a distribution of fluorescence whose spread does not depend solely on the average expression level (Fig. 7.4A and Supplementary Fig. S5).

7.2.8 A set of "noisy" insertion sites.

We found that in all the clones the standard deviation of gene expression is proportional to the mean, meaning that most of the noise shown by the promoter is extrinsic, regardless of expression level and insertion location (Fig. 7.4A). This is expected from the high level of expression of the rrnBP1 promoter [281]. However, a subset of clones show a higher level of noise where the scaling of the standard deviation with the mean follows a steeper slope. In other words, these clones show much larger gene expression variations than expected. All of these high-noise clones were very-low expression bacteria obtained from the RL (low-expression filtered) population. Hence, these clones likely explain the very-low expression sub-population (in green in Fig. 7.2E) in the distribution of gene expression of the RL (low-expression filtered) population. When tested for their distribution in single-cell gene expression by flow cytometry, these low-expression high-noise clones did not show bimodal distributions. Rather, they showed disperse and skewed distributions, whose range overlaps with the expression level of other low-expression clones (Supplementary Fig. S5A,B). Supplementary Fig. S5 C-G recapitulates the noise properties of all the picked clones from different rounds of selection. It is clear that high-noise clones become more frequent in successive sorting rounds where low or very-low expression cells are selected. The following two sections deal with a more detailed characterization of the properties of these low-expression, high-noise insertions.

7.2.9 Noisy sites are associated with the insertions within ribosomal operons.

To characterize the set of clonal colonies from different populations, we performed whole-genome sequencing on 90 selected clones.

The locations of all the insertions of these clones are listed in Supplementary File SF1. 32 clones out of the 90 selected samples show the presence of insertion within a rRNA operon from whole-genome sequencing data. These clones were very-low expressing clones derived from RL filtered populations. For example, in the clones from the RLR1V1L sorted populations we tested a total of 16 insertions, of which 9 were in rRNA regions (Supplementary File SF1). In order to verify these short-read assignments of insertions sites, long-read nanopore sequencing was used on 5 of these 32 clonal samples. This analysis confirmed the presence of the transposon insertion within the 23S ribosomal RNA (rrlE) of the rrnE operon. To recapitulate these results, Fig. S6A shows the association of clones with high-noise promoters

with rrlE insertions. Some of the non-rrlE high-noise clones revealed the presence of multiple (up to three) insertion sites, which can explain the large variance in gene expression of these clones (see column D of Supplementary File SF1 for a list of the multiple insertions and theor coordinates). Indeed, we found that normalization of gene expression by the copy number of the promoter removed these outliers (Supplementary Fig. S6BC). Hence, we believe that the true high-noise phenotype should almost exclusively be associated with rrlE insertions.

Insertion within ribosomal operons is tolerated because *E. coli* has 7 copies of ribosomal operons. To discover the orientation of the cloned sequences with respect to the rrlE promoter position, we used the blast results of *de novo* assembled contigs with the flanks of cloned insert sequence and we compared them with the reference genome. Most of the insertions showed an opposite orientation of the inserted promoter with respect to the rrlE promoter sequence. Separately, we checked the orientation of the inserted GFP cassette in few selected Illumina samples for which the blast results showed reasonable overlap with the genome locus. This confirmed the opposite orientation of GFP with respect to the rrlE promoter sequence in most samples. Strong promoter competition may explain the very low levels of GFP transcript production by which these clones were isolated, as well as the high variability.

We found that the trends of gene expression with growth rate were consistent with the hypothesis of competition with a strong promoter: while most of the other clones increase their expression with growth rate (in agreement with the known regulation of the rrnBP1 ribosomal promoter) the noisy clones decrease in expression with increasing growth rate, in agreement with the idea that their expression is repressed by an interference with transcription of the increasingly transcribed ribosomal operon (Supplementary Fig. S3 and S4). Additionally, the mild reduction in growth rate for insertions giving different mean GFP expression suggests that the cost associated to GFP expression and the possible interference with the ribosomal operon are not dominant for these insertions (Supplementary Fig. S3 and S4).

7.2.10 Noisy promoters may perform switching

The previous analyses strongly indicate an association of the high-noise insertions with an interference of the insertions with ribosomal operons. To gain more insight on the temporal dynamics of these high-noise inserted promoters, we measured gene expression noise in time-lapse microscopy data on growing microcolonies (Fig. 7.4CD). We compared the change in GFP gene expression over time of sin-

gle cells from clones carrying noisy and non-noisy promoters for 3-4 generations and quantified the differences between gene expression in different lineages. The divergence between lineages was quantified as the time average of the absolute value of the gene expression difference between sister cells.

Bacteria were grown on an agar pad to form a microcolony. The time-lapse data in the formation of the microcolony was segmented to obtain the change in the average cell fluorescence as a function of time Fig. 7.4C. An example of gene expression of two lineages, one from a noisy clone, as measured by flow cytometry, and one from a control clone (where the cassette was inserted specifically between two converging genes, AidB and yjfN) is shown in the left panel of Fig. 7.4D. The right panel of Fig. 7.4D quantifies the divergence of gene expression along lineages for different clonal microcolonies, corresponding to clones where the promoter is inserted in different positions. Figure 7.4D shows that in microcolonies from high-noise clones different lineages emanating from the same single cell tend to diverge more in gene expression as time progresses than in the control or low-noise clones. This result points to the possible presence of switching behavior in the high-noise clones.

7.3 Discussion

Our results directly show that the probability of DNA insertion in the *E. coli* genome by a transposon is biased, before any long-term selection may act, other than that related to overnight growth. First, there is a stronger origin-to-terminus bias than explained by gene dosage imbalance, second, insertion probability is higher in AT-rich regions of the genome. This may seem surprising, because these are regions that are preferentially bound by the H-NS protein, which has been proposed to act as a barrier to horizontal gene transfer [245, 248, 250, 251, 266, 273]. However, physical components such as differences in DNA supercoiling [151, 263, 282] and the biophysical properties of AT-rich DNA (lower melting barriers, different stacking energy, etc.) may play a role in establishing these biases. In particular, the measured origin-to-terminus gradients of supercoling [283] and gyrase binding [284] are compatible with the hypothesis that the probability of insertion is increased in regions with higher negative supercoiling, which may explain the origin-to-terminus positive bias for insertions (not justified by origin-to-terminus differences in gene dosage). Note that this insertion bias (only transposon type) should not be confused to the transcriptional consequences of supercoling on the donor promoter. A different promoter sequence is not expected to affect the insertion

probability. On the other hand, changes in gene expression levels depend on promoter sensitivity to supercoiling and local context.

An important technical point to address is the role of kanamycin selection in these experiments. If the donor sequence including the kanR cassette is inserted in a locus where it is completely silenced by H-NS, one might not be able to see the insertion. However, our data show that selection itself does not preclude the identification of an insertion site, since insertions are not excluded from H-NS occupancy-rich regions, as it would be expected if complete silencing of KanR expression had taken place. We do observe that promoters inserted in H-NS rich regions are on average expressed less than the others. We also note that we carried out the transposon reaction in mid exponential phase cells growing in a rich medium, LB. In these conditions the concentration of H-NS is lower due to a high dilution rate [214]. However, at the faster growth rates, H-NS is known to still play a role in the repression of ribosomal promoters by binding to higher affinity sites [224, 269]. Our results suggest that there is probably not enough protein to also cover the lower affinity (non-specific) binding to AT-rich regions, in order to inhibit transposon insertion [252]. This is in contrast to a previous study in *Vibrio cholerae* that has shown that only in the absence of H-NS there was a higher probability of insertion in AT-rich regions of the genome [253]. These results lead us to further suggest that the role of H-NS in regulating the probability of genomic insertion of horizontally acquired genes may depend on the growth conditions and on the specific strain.

Our results also indicate that the dilution of H-NS in rapidly growing cells does not prevent the establishment of a low-expressing population biased for the insertions associated with H-NS binding. Hence, they are consistent with the role of H-NS as a silencer of newly acquired genes. Indeed, we observe that once the full length rrnBP1 promoter cassette is inserted in the genome, its level of expression is lower if it is found near H-NS rich regions. This cassette includes a higher affinity H-NS binding site within the full length rrnBP1 promoter and an AT-rich *gfpmut2* gene sequence stabilizing the formation of H-NS dependent repressing complex. The shorter version of the promoter (P1-short), lacking the high affinity H-NS binding site, does not appear show a stable sub-population of very-low fluorescence clones (data not shown), showing that the high affinity site of the promoter is important in nucleating the repressing oligomeric structure, a question that we are still exploring. In summary, the level of expression of the clones can thus be very heterogeneous, depending on local properties of the site of insertion and the sequence of the fragment.

The cell-to-cell variability of gene expression within a given isogenic clone can

vary significantly, but it typically scales with the mean level of expression as expected from extrinsic noise [185, 279, 281]. This is expected from the promoter used here, rrnBP1, which is a strong promoter, resulting in a high level of expression. The change in the CV as a function of mean expression therefore remains for the most part relatively flat, corresponding to the extrinsic noise regime.

However, in some of the very low expression clones the noise varies in a way that is not expected from the known pattern of gene expression noise correlations that have been described previously [279, 281]. We therefore characterized those clones that have a higher level of gene expression noise and found that in these cases the insertion has taken place within a ribosomal operon. *E. coli* has 7 copies of ribosomal operons, therefore insertion inside one of them does not have a high cost and is not selected against, at least in the short term of this experiment. This results in interference between two transcription processes driven by very similar promoters, of similar strength. Furthermore, the initiation frequency of ribosomal promoters is high enough at fast growth rates that most of the time the operon sequence can be assumed to be covered by transcribing RNA polymerase "trains" [285]. This creates a block for RNA polymerase to bind to the promoter that is found within the operon, creating a stable "off state". However, from time to time RNA polymerase manages to bind to the newly inserted promoter, perhaps after the DNA replication forks have erased the memory from the competing process, starting its own "train" of GFP production. Such transcriptional interference is a well-known phenomenon [286]. Our result on promoter noise also suggests that in the tightly packed bacterial genomes, transcription interference with newly inserted genes might be a natural source of innovations in terms of gene expression noise on evolutionary time scales, as previously speculated for eukaryotes [287].

Altogether, the main novelty of our study is the support for a high (initial) tolerance for insertions with a wide range of expression levels, which challenges the standard view that H-NS not only silences horizontally acquired genes (because they are more AT-rich) but also inhibits insertion in AT-rich regions of the genome. We find that in the tested growth conditions, H-NS does not inhibit insertions in AT-rich regions. However, it can still decrease expression after the insertion has taken place. These findings support the following evolutionary scenario. When a novel gene enters the genome, it is more likely found in a region that is controlled by H-NS, for reasons that most likely have nothing to do with fitness, but have to do with the physico-chemical properties of the DNA in AT-rich regions. However, the wide range of expression levels that we find show that the gene is not necessarily immediately silenced. Rather, the different insertion positions allow it to sample

a wide range of expression levels (including silencing), at (initially) equal promoter strength, while interacting from the start with the cell's housekeeping physiology. We believe that this inherent bet-hedging and exploratory stage may be a key ingredient of genome plasticity, and is underestimated in our current narrative of the process of horizontal transfer, which is centered on the average outcome, and establishes a strict time hierarchy between stages where an exogenous gene is first silenced and then reactivated.

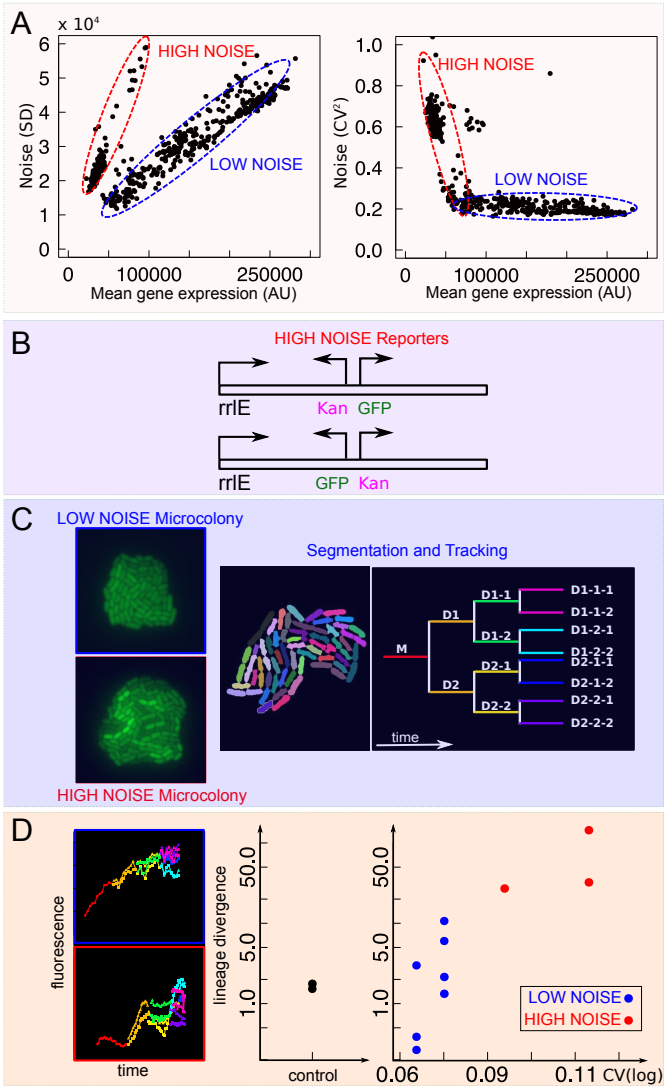


Figure 7.4: Noisy promoters emerge from transcriptional interference within the rrlE ribosomal operon. *A. Characterization of noise of 658 clonal populations from flow cytometry. Left: Standard deviation vs mean GFP expression. Right: Noise (CV^2) vs mean GFP expression B. All the tested "noisy" clones by sequencing typically show association with the* rrlE *ribosomal RNA operon (see Supplementary File SF1). C: Characterization of promoter noise by microcolony time-lapse growth assay. Microcolonies were grown, imaged, segmented and tracked for 3 generations. D. An example of the resulting lineage-specific gene expression data is shown on the left. The plot on the right quantifies the lineage divergence of gene expression (time average of the absolute gene expression difference) of the different clones, comparing it with the CV measured from flow cytometry (see text). High-noise clones show a large lineage divergence - indicating a possible switching behavior.*

Chapter 8

Conclusions

In this final Chapter I will summarise the key results of my work and propose ideas for future research.

8.1 Summary

Despite the great effort scientists put into it, the debate is still open concerning the regulation of DNA replication initiation in *E. coli* and its role in the dynamics of cell growth and division. This work aims to contribute to this debated topic focusing on the regulation of a key player in these processes, the DnaA protein [8–13]. DnaA is commonly believed to be a cell cycle oscillator and a cell size sensor, but neither of these facts have been firmly established [19]. In particular, we were interested in identifying the cell cycle oscillator related to DnaA activity, and relating it to the progression of the *E. coli* cell cycle, focusing in particular on its coordination with cell growth and division.

To address this, I have first designed a dedicated experimental setup and data-analysis pipeline for studying the growth, size, and gene expression of *E. coli* in controlled environmental conditions. Next, to establish a solid reference for monitoring the cell cycle dependence of gene expression from the dnaAP promoter, I characterized the "null" relationships between cell cycle progression and gene expression of an unregulated promoter (P5). Here we have found that the GFP production rate under control of this constitutive promoter can be described as linearly proportional to volume, likely through the dependence of protein production rate on ribosome amounts, and to gene dosage through the increase in the gene copy number by DNA replication. Correlating the GFP expression rate with the estimated gene replication times along the cell cycle, one can see that instantaneous gene dosage also affects protein production in a cell cycle phase dependent

way. Consistently, we have found that the specific GFP expression rate of a constitutive promoter appears to be biphasic: the expression rate increases after the initiation of DNA replication suggesting an increase in gene copy number.

Interesting things happen upon cell division: the volume-specific growth rate suddenly changes, in a way that is dependent on the cell size at birth and reflected in the volume-specific GFP expression rate, this in addition depends on the growth condition, whether it is at fast or slow growth.

Surprisingly, we found that *E. coli* specific growth rate follows the same pattern as that of the specific GFP expression rate from the ori-proximal P5 promoter. We still do not have a clear explanation for this, but one possibility might be that the replication of the ribosomal operons close to the origin can result in increased GFP translation rate and increased volume-specific growth rate after initiation of DNA replication has taken place.

The decrease in the volume-specific expression rate in the first phase does not happen if we look at the expression from a full length ribosomal promoter where Fis may help recruit RNAP to better compete with neighboring genes.

The differences we observed between the expression of GFP from the origin and terminus-proximal strains means that the gene's position in the genome can play a role in modulation of gene expression as a function of the cell cycle, in addition to the timing of change in gene copy number.

In addition to possible effects due to the distance from the origin, differences in the local chromosome structure and organisation can influence gene expression and regulation. In Chapter 7 we find that insertions of exogenous promoters are non-uniform and sparse and are more biased towards the replication origin than justified by gene dosage. Moreover, we found that H-NS binding sites are enriched at insertions positions and noisy phenotype are associated with the insertions within ribosomal operons.

Next, we investigated how DnaA activity correlates with the cell cycle. One of the major novelty of my work has been to quantify the changes in DnaA-ATP activity *in vivo* in real time in order to relate DnaA levels and activity to key cell cycle events. To do this, I have developed a set of reporters of gene expression using a gene for a fluorescent protein under control of a promoter that is differently regulated (activated, activated+repressed, no regulated) by DnaA-ATP. Lastly, it is know that other factors can contribute to express a gene has a function of the cell cycle. We know for example that SeqA might have an important role by shutting off gene expression right after gene duplication. A mutant without GATC binding sites downstream of the promoter is also considered.

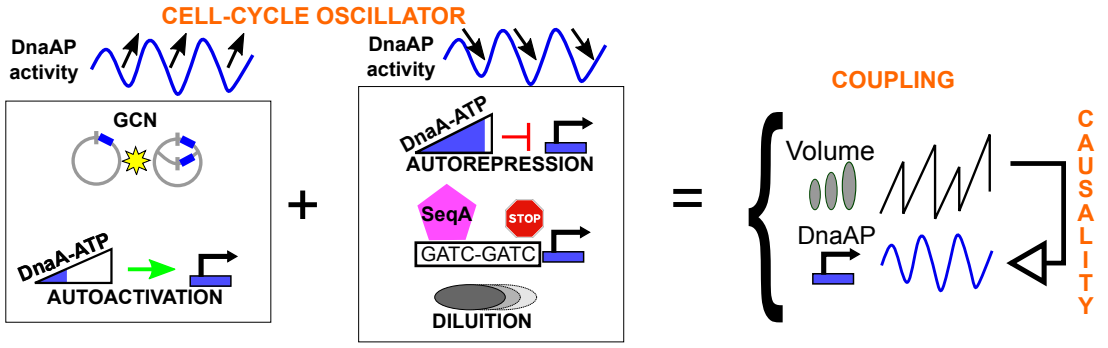


Figure 8.1: Final sketch *The main findings are graphically summarized. First, an oscillatory pattern of DnaA expression within the cell cycle was observed. Thanks to differentially regulated DnaA mutants I was able to dissect the contributions of different factors that boost these oscillation. Gene copy number variation and activation by DnaA-ATP contribute to an increase in volume-specific gene expression. The latter can help recruit RNAP to the promoter in the presence of competition from neighboring genes, notably ribosomal operons, as it is observed for the specific GFP production rate from the constitutive promoter. Transcription repression by DnaA-ATP, repression by SeqA binding and dilution by increased growth rate contribute to its decrease after initiation of DNA replication and* dnaA *gene duplication has taken place. This concert of regulators leads to an oscillation that is strongly coupled with cell volume, indeed a cell volume to DnaA activity causality was observed. This leads us to think of cell growth and DnaA activity as two coupled oscillators with cell division.*

Thanks to the analysis of different reporters, I was able to quantify for the first time the effect of the DnaA-dependent promoter regulatory elements that lead to an oscillatory pattern of gene expression that is coupled to both the DNA replication and cell division programs. For the first time we "observe" negative and positive regulation over the cell cycle, with SeqA contributing to this regulation. Specifically, I found that the volume-specific production rate of GFP from the DnaA promoter is a well-defined cell cycle oscillator and that the signal from this oscillator can be related to key cell cycle processes such as initiation of DNA replication and cell division. However, while the standard view of the cell cycle sees it as the result of a single oscillator, our data lead me to suggest that at least two coupled oscillators are needed to describe the processes that coordinate DNA replication, cell growth and cell-cycle progression. My approach also makes it possible to detect causality links between these different processes. These findings combine the use of mathematical models and single-cell dynamic data to pose firmer quantitative bases for a characterization of the mechanisms determining robust cell cycle progression in bacteria. A lot of previous data consider average idealized cells and when you do this you get cells that do not actually exist and you lose a lot of information. My results highlight the importance of look at how the single cells behaves.

8.2 Future perspectives

8.2.1 Multi-growth conditions switch

An advice I would give to the people that will come to the lab in the next year would be: "play with the bacteria and the experimental setup!". During these years I have used the microfluidic experimental setup at less than half of its potentiality. The 11-port/10 way Mux Distributor from ElveFlow allows a fast switch between up to 10 different growth media. Moreover, thanks to the OB1 pressure controller we can expose the bacteria to oscillating environment with pulses of different frequency and shape. All this would be very useful to study the adaptation process during upshift or downshift experiments. For example, something very interesting that we have seen "by chance" is that following a downshift cells stop growing and for several hours they don't resume growth. Then, as soon as we put back the fast growth medium, they start growing again. It would interesting for me to understand in which condition bacteria are able to grow again after a downshift, for example gradually changing the environment.

8.2.2 Mutations on the native DnaA promoter

As I have already pointed out in Chapter 6, the most direct continuation of my PhD project would be to use the same approach I have presented in this thesis to test the hypothesis that have emerged from my results by making mutations on the endogenous promoters for the DnaA gene as well as its target genes.

8.2.3 A data driven model

Another important step is to build a theoretical model that has DnaA-ATP as main player and it is able to predict the coordination between DNA replication and the cell cycle.

As we have seen, DnaA has an highly regulated promoter, it is difficult to formulate a minimal quantitative model of the initiator circuit that includes the key ingredients known to regulate its activity[18, 19]. Moreover, cell cycle effects that we observed, both for cell growth and DnaA activity, might have an important role in this coordination. A possible approach to this problem would be to use these quantitative data, and especially the ones from the strain with mutations on the native DnaA promoter, to develop a data driven model that will allow us to test possible hypothesis on the regulatory links between gene expression and cell growth

and predict the effect of novel mutations.

More specifically, the changes in gene expression from the different promoter variants should all be consistent with a cell cycle dependent change in DnaA activity. These models of the transcription regulation by DnaA would be useful to obtain an estimate how the activity of DnaA changes to result in the experimentally observed patterns.

Simulations could answer the following questions: can any promoter activated and repressed by the same protein behave as DnaAP? Which form needs to have the concentration of the regulating protein as a function of time (or volume)? Which are other necessary parameters (Transcription initiation rate, RNAP affinity....)?

8.2.4 Beyond *E.coli*

Bacteria like *Escherichia coli* are an ideal model organism for the study of cells that can change their growth rate depending on environmental conditions. When the conditions are optimal they become rapidly proliferating cells requiring a specific set of regulatory mechanisms involved in a robust cell-cycle progression. I find extraordinary the fact that we still need to understand a lot about this simple organism and that new discoveries made on this organisms can have a wider effect on other organisms.

In particular, my work deals with three topics that have a widespread interest and applicability.

First, gene regulation. Although from the seminal work of Jacob and Monod [288] our understanding of gene regulation has increased enormously, there are still many open questions. Gene regulation is a quite simple mechanism that allows living cells to optimize their gene expression based on the external environment. Evolution has selected for regulatory mechanisms of gene expression that result in a coordinated control of different key cellular processes such as DNA replication, DNA repair, cell growth and division. Difference among organisms might reside in different ways their genes are regulated, more than in having different genes [289]. Therefore, while this work sheds more light on the DnaA-dependent regulation in *E. coli* it would me interesting to study how this regulation has changed in other organisms and why.

Second, oscillations. Oscillations - such as a pendulum's swing or a ball's bouncing on the end of a spring - are among the simplest and most common phenomena in physics, but researchers have come to appreciate their ubiquity in the biological world, too [290]. Concentrations of molecules rise and fall, genes alternate between on and off, and circadian clocks keep time almost as well as human-made

machinery. Together, these biochemical fluctuations are crucial for many biological needs: timing daily activities, orchestrating cell division and movement, even mapping out parts of an embryo as it grows. Cells would be unable to function without them [290]. For example, in this work we suggest that in *E.coli* at least two coupled oscillators are needed to describe the processes that coordinate DNA replication, cell growth and cell-cycle progression and probably is can be applied to many others oscillating circuits present in other organisms.

Third, cell-cycle control. The 2011 Nobel Laureates in Physiology or Medicine together with thousands of other researchers have made seminal discoveries concerning the control of the cell cycle. They have identified key molecules that regulate the cell cycle in all eukaryotic organisms, including yeasts, plants, animals and human[291]. Since then, also thanks to new experimental techniques that allows a single cell resolution, many new discoveries has done but at the same time new question have arisen.

What is important to stress is that coordinating cell growth is critical for cell survival across all kingdoms.

In microorganisms, growth control is essential for fitness and resistance to adversarial attacks such us toxins and drugs. On April 15 2021, the World Health Organization once again sounded the alarm on the emergence and spread of drug-resistant strains against existing treatments and the need to develop new strategies to address the problem of the lack of new antibiotics [292]. Identification of novel antibiotics remains a major challenge for drug discovery. DNA replication is essential for cell viability and is therefore an attractive target for antimicrobials. Although several antimicrobials targeting DNA replication proteins have been developed to date, gyrase/topoisomerase inhibitors are the only class widely used in the clinic. Given the numerous essential proteins in the bacterial replisome that may serve as a potential target for inhibitors, it is evident that antimicrobials targeting the replisome are underdeveloped so far and efforts has to be done in this direction [24]. Moreover, some components are well conserved between most bacteria, such as the DNA initiator protein DnaA the bacterial replisome can be used to develop both broad-spectrum antimicrobials.

Lastly, in humans, uncontrolled growth is implicated in several disorders, including cancer, obesity, and diabetes. For example defects in cell cycle control may lead to the type of chromosome alterations seen in cancer cells. As we have seen for antimicrobials, similarly, most cancer therapies work by inhibiting cell-cycle progression. This kind of study and its interdisciplinary approach presented here may in the long term open new possibilities for cancer treatment.

Appendix A

Insertion protocol

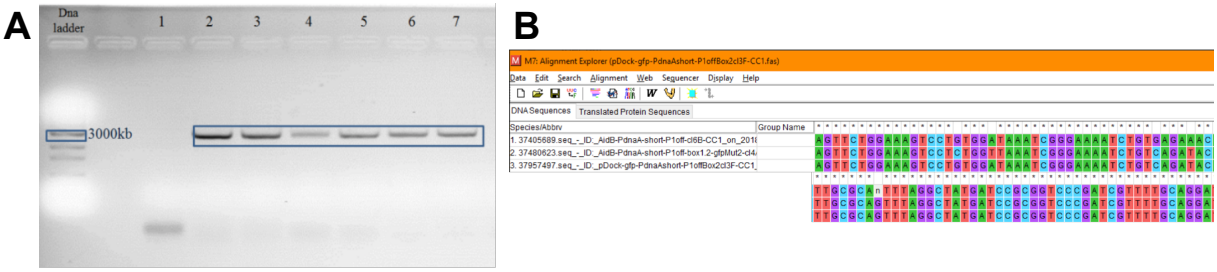


Figure A.1: PCR verification A *Gel electrophoresis verification: compared to the control bacteria BW all the DnaA fragments have bands corresponding to 3000KB so they are positive.* ***B*** *Mega sequencing*

Production of the "promoter + GFP + kan" amplification bordered by 50 bp extensions homologous to the target chromosomal site

This amplification is made from the pDOC-K plasmid with long oligos that hybridize by 20 bp to the pDOC-K, and having 50 bp extensions homologous to the desired chromosomal insertion site. The PCR protocol is reported in Figure A.2 A lot of DNA is needed (at least 200 µl of PCR product = 4x50). Expected cassette size ("promoter + GFP + kan") is between 2000 and 3000 bp depending on the promoter.

Digestion by DpnI of the PCR product in order to eliminate the plasmid pDOC-K

- By promoter, pool the PCR tubes (VT total volume = 200μl)
- Add 1 µl of DpnI
- Incubate at 37°C for 1h minimum

Purification of the digested PCR product on agarose gel 0.8%

- Prepare Sybr Safe
- Cut the band at 3000 bp and purify with the Nucleospin DNA PCR clean up kit (Macherey Nagel)
- Resospend in 50 μl of water
- Measure the concentration of DNA (should be around 100 ng/μl of DNA)

Electrophoresis of the PCR product in the strain BW25113 which carries the plasmid pKD46

- (Red recombinase allowing linear DNA recombination in the cell).
- Incubate the cells for 15 min in ice
- Add 100 ng of PCR product, pipette gently (a small volume of DNA is recommended)
- Transfer the mixture of cells + PCR product into a cooled electroporation vessel (0.1 cm gap) (caution, avoid bubbles in the vessel)
- Electroporate at 1800V
- Immediately add 1 ml of cooled SOC medium and resuspend the suspension with the pipette
- Transfer to a 2 ml tube and incubate at 37°C for 3 hours.
- Spread on LB plates + Kanamycin (50 μg/μl) and incubate overnight at 37°C

Select clones only resistant to Kanamycin

- KanR transformants are selected and put in preculture
- Each clone is then subcultured on one LB + Kan (100μg/ml) and one LB+ Amp plate
- KanR plates are incubated 1 night at 37°C and AmpR plates at 30°C
- The selected clones are those which are only resistant to Kanamycin: KanR After this first selection, the chromosomal insertion of the KanR clones is verified by PCR

- Collect 100μl of transformants
- Centrifuge the cells at 11000 rpm for 5 min then discard the supernatant
- Resuspend the pellet in 100μl of water, vortex and heat for 5 min at 95°C
- Centrifuge the cells at 11000 rpm for 5 min
- Recover the supernatant, it contains the chromosomal DNA
- Perform the verification PCR with short oligos

We use 2 locus-specific primers on each side to verify that we have inserted the "promoter + GFP + kan" cassette of about 2000-3000 bp.

Sequencing of insertion products

- On the correct clones, repeat the insertion verification PCR in a final 50 or 100 μl volume (Figure A.3
- Purification of digested PCR product on 0.8% agarose gel prepare with Sybr Safe
- Cut the band at 3000 bp and purify it with the Nucleospin DNA PCR clean up kit (Macherey Nagel)
- Resospend in 50 μl of water
- Measure the DNA concentration (should be around 100 ng/μl of DNA)
- Send the PCR product to be sequenced with the oligo CC1 (CC1: GAATCTTAAGTTCTACGTGTTCCGC)

The absence of a mutation at the promoter level must be verified. CC1 verifies the insertion in the chromosome (D57785), hybridizes on the kan cassette towards the outside. In reverse direction

	A	B	C
1	Dénaturation	95°C	5 min
2		**95°C**	**30 sec**
3	**30 cycles**	**60°C**	**30 sec**
4		**72°C**	**2'30''**
5	Extension	72°C	5 min

Figure A.2: PCR program 1

	A	B	C	D	E	F	G	H	I
1					782	783	784	785	786
2	AidBF	(100μM)	1 μL	4 μl	4	4	4	4	4
3	AidBR	(100μM)	1 μL	4 μl	4	4	4	4	4
4	dNTP	(10 mM)	1 μL	4 μl	4	4	4	4	4
5	Tampon	(10x)	5 μL	20 μl	20	20	20	20	20
6	Taq polymerase		1 μL	4 μl	4	4	4	4	4
7	ADN	(20 ng/μl)	5 μL	20 μl	5.3	25	20	6.7	10
8	H2O	qsp 50 μl	36 μL	144 μl	158.7	139	144	157.3	154
9	Volume totale		50 μl	200 μl	200	200	200	200	200

Figure A.3: PCR program 2

Appendix B

96 well plate protocol

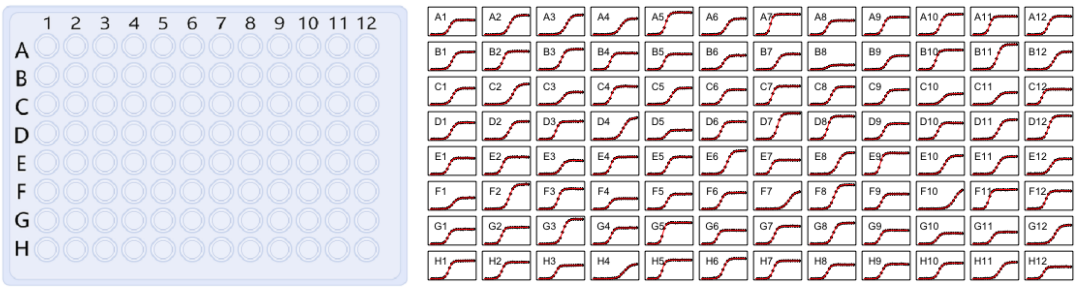


Figure B.1: Plate reader *Plate reader experiments allows to rapidly screen several growth media and promoter mutants at population level*

B.1 Main steps

- The mutant bacteria strains were pre-cultured over night at 37°C in the shaking incubator.
- The next day, growth media were prepared.
- Pre-cultured bacteria strains were diluted by the program 8sample-4media equal dilution set up in Beckman Coulter Biomek 3000 to dilute the bacterial solution from the preculture 1000 times before going to the step of data analysis.
- The mineral oil was put on all the wells to avoid the evaporation.
- 96 well plates were inserted in the Tecan Plate reader machine at 37°C.
- Every 5 minutes for 23 hours, OD600 and fluorescence were measured using the Magellan software and saved in an excel file.

B.2 Data analysis

Data was analyzed by Matlab. A MatLab program was written by a former PhD in the lab in order to automate the analysis of the data. Recently a R-Shiny program was also written in our lab to make analysis easier to all the students in the lab.

Here are reported the main steps:

- The raw data of the measurement of the absorbance and fluorescence as a function of time were smoothed using spline interpolation.
- The background absorbance and fluorescence was obtained from wells containing only the growth medium.
- The curves from several control wells were averaged and subtracted from the data of the different reporter strains.
- The curves of the absorbance measurement were shifted in the time (x) axis so that all the curves reached 20% of their maximal absorbance at 600 nm (A_{600}) at the same time point.
- The curves of the fluorescence measurements were consequently shifted by the same amount.
- The promoter activity was calculated by taking the time derivative of GFP divided by the A_{600}: [(dGFP/dt) /A_{600}] and the growth rate was calculated by taking the time derivative of the absorbance at 600 nm (dA_{600}/dt).
- Graphs of growth rate and promoter activity were used to calculate the average values in exponential phase.

Appendix C

Protocol for chip preparation

Figure C.1: Plasma cleaner *A standard procedure with plasma cleaner(left) is used to increase adhesion potential through a process called surface activation.*

C.1 Main steps

- Coverslips were cleaned by isopropanol, then rinsed in MilliQ water. Slips were dried using pressurized N2 gas and stored in sterile Parafilm-wrapped polystyrene Petri dishes until use.
- PDMS was cast over the master template to yield the negative relief mother machine pattern to be fixed to the coverslip.
- Microfluidic devices are constructed from polydimethylsiloxane (PDMS)). A recipe of 10:1 elastomer to curing agent ratio by weight was mixed thoroughly by hand with a stainless steel microspatula.
- The mixture placed under vacuum in a glass dessicator for at least 20 min to remove gas bubbles.
- The mixture was then gently poured over the master to a height of roughly 7 mm in a alluminium foil before being baked at 80°C for 3-4 h.

- The master was gently removed and individual chips were cut using a scalpel.
- A 0.75 mm biopsy punch was used to make inlet and outlet injection sites.
- Scotch tape was gently applied to the features side of each chip to remove any dust.
- Both coverslips and PDMS chips, features-side up, were plasma cleaned together for 1 minute at 30% power.
- Chips were placed features-side down onto each coverslip with one chip per slip.
- Bonding between the glass and PDMS was usually immediate, however devices were left at 80°C for an additional 10-15 min for increased bonding.
- Heat the assembled device at 80-100 degrees Celsius for 60 seconds in an oven or hot plate. The high temperature provides activation energy for additional bond formation.

C.2 Important advises

Cleanliness

The presence of particulates or oil can block bonds from forming. Avoid touching the surfaces to be bonded when removing PDMS from the plasma chamber.

Air vs Oxygen

Oxygen is more efficient than air due to the higher concentration of reactive oxygen species. Additionally, air from an environment prone to daily fluctuations in humidity or particulates can adversly affect PDMS bonding.

Plasma treatment duration

Plasma treatment should not exceed 2 min, as prolonged plasma exposure causes cracking in PDMS and migration of low molecular mass molecules from bulk to surface, decreasing the number of hydrophilic SiOH groups and resulting in weak or incomplete bonding. Oxidized surfaces should be brought into contact immediately after plasma treatment to achieve strongest bond possible. PDMS surface recovers hydrophobic properties (aging) with time after plasma treatment (1 hour).

Bonding

After Plasma Treatment, press and hold PDMS components together lightly for 30 seconds. Do not pull apart and adjust alignment as this will disrupt bond formation. Pressing with too much force may collapse microfluidic channels.

Conservation

Scotch tape was placed over the injection sites until use to seal the device from dust. When properly stored in Parafilm wrapped Petri dishes, the devices were shown to remain sterile and functional for up to two weeks.

Appendix D

Microfluidic protocol

Part of the protocol can be found at:
www.elveflow.com/microfluidic-applications/microfluidic-cell-culture/\protect\@normalcr\relaxhow-to-study-bacteria-by-microfluidics [170]

D.1 List of components

D.1.1 Pressure control system

- OB1 pressure controller
- 2x Mux-distributor
- 2x Flow sensors
- manifold 9 ports
- 5x Tubing, fittings and reservoirs

D.1.2 Computer interface

Window 1: [ORANGE] Main window

You have the list of all the connected devices. If any device is missed, you can press the "recharge" button on the top/left Press on the play button to open the windows 2 and 3 Press the 3 dots symbol on top/right to open window 4

Window 2: [RED] MUX top and bottom window

Press on the channel relative to the growth medium you want to flow in your circuit

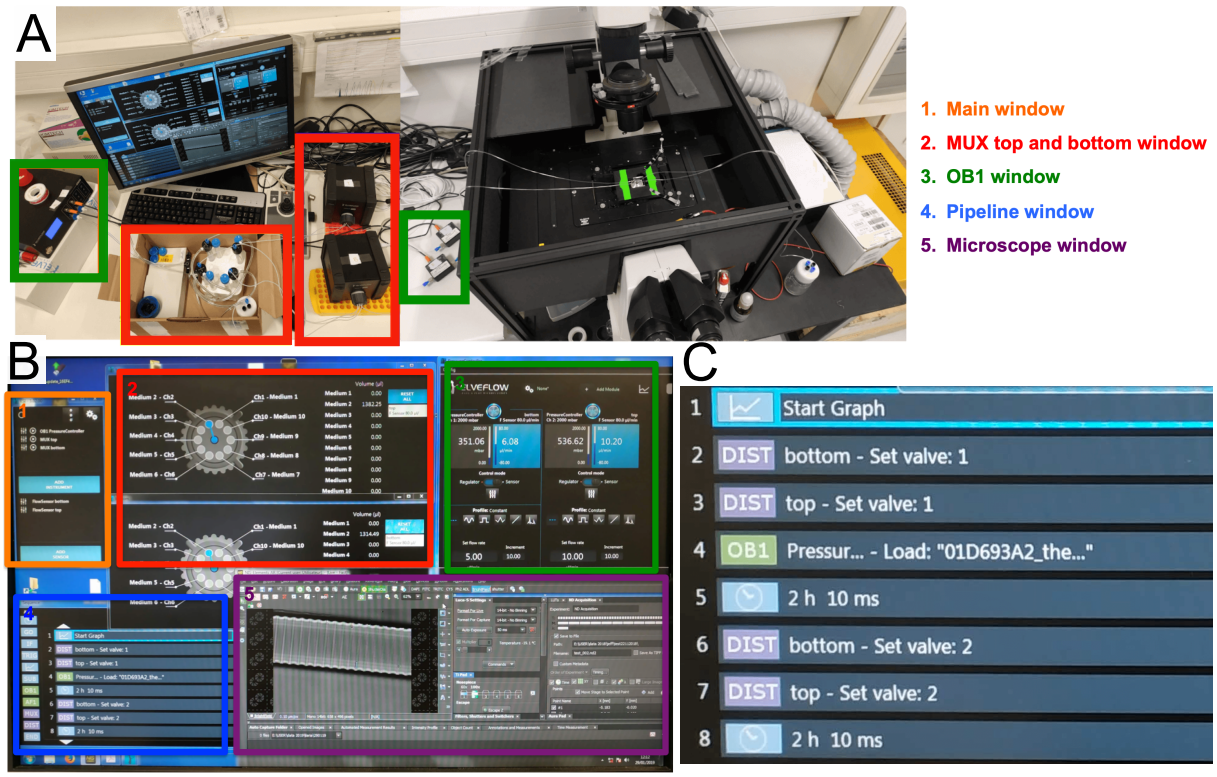


Figure D.1: Microfluidic experimental setup A *Picture of the microfluidic setup employed for the study of bacteria adaptation to stress and environmental changes comprising an OB1 pressure-driven flow controller, flow sensor, Mux Distribution and microfluidic chip.* ***B*** *Picture of the ESI software interface focusing on the Sequence* ***C*** *Scheduler employed for the study of bacteria adaptation to stress and environmental changes comprising an OB1 pressure-driven flow controller, flow sensor, Mux Distribution and microfluidic chip*

Window 3: [GREEN] OB1 window

Press the power bottom to turn on the relative channel If you select Regulator mode you can set a pressure to be maintained constant (flow rate will change based on the circuit's resistance) If you select Sensor mode you can set a flow rate to maintain constant (pressure will change based on the circuit's resistance)

Window 4: [BLUE] ESI Sequence Scheduler window

Press the folder symbol with a plus to load an already existing file. For example, for an upshift experiments open the upshift.sq file.

Window 5: [PURPLE] Microscope window (Ti-Eclipse Nikon inverted microscope)

D.2 Main steps

Bacteria preparation

- Grow the bacteria overnight in filtered growth medium. NOTE: Always use a growth medium filtered at 0.2 µm in order to avoid clogging the microfluidic device.
- In the morning, dilute the bacteria (ex 20 µl in 1.5ml). Put them at 30°C for around 3 hours (we want them to be in exponential phase when they go into the device).

Microfluidic Setup preparation

- Using a 20 μl pipette put around 150 μl of BSA in the inlets (top and bottom) until you see BSA going out from the outlets. This minimizes bacterial interactions and binding to the glass or PDMS components.
- Put the device at 30°C and wait 30 mins
- Fill the reservoirs with the growth medium
- Fill the tubes with a growth medium.
- In Sensor mode with the flow rate at 80 μl /min let the growth mediums flow inside the tubes and fill them completely by switching the MUX distributors between different falcons to avoid bubbles in the entire circuit.
- Set the bottom flow rate to 20 μl min and the top to 10μl/min (dust will be washed out from the channels). Let the medium flow through the device for 30 mins
- Set the bottom flow rate to 10 μl/min and the top to 20μl /min (the bubbles left will go out from the channels).
- In the meantime, start an acquisition in brigth field by selecting around 5 positions to check if the device is stable and there are no problems with the autofocus.
- Once there are no more bubbles in the channels we can inject bacteria.
- Put 0.5 ml of preculture in a 1 ml syringe. Stop pumping. Unplug the 2 input tubes from the device, set them in a tube and start pumping again (to avoid bubbles formation).
- Inject the bacteria by hand (by slowly moving the syringe plunger, not by pushing on it) and check each 0.1 ml if you have enough bacteria in the channels (without unplugging the syringe).

- Set the bottom flow rate to 7μll/min and the top to 15μl/min (to help bacteria to go inside the channels).

Schedule the ESI sequence

The ESI sequence scheduler can be used to automate your microfluidic experiment. In this protocol, the following sequence was created and performed:
STEP 1: Start the acquisition of pressure and or/ flow rate measurements

STEP 2-3: Set both mux distributors (DIST) to valve 1 (slow growth medium)

STEP 4: Start pumping (OB1) with a constant flow rate in both channels but with different flow rates (we can save our OB1 personalized settings in a text file to be loaded for each kind of experiment)

STEP 5: Maintain these conditions for 2h 10ms

STEP 6-7: Switch both the mux distributors to valve 2 (fast growth medium)

STEP 8: Maintain these conditions for 2h 10ms

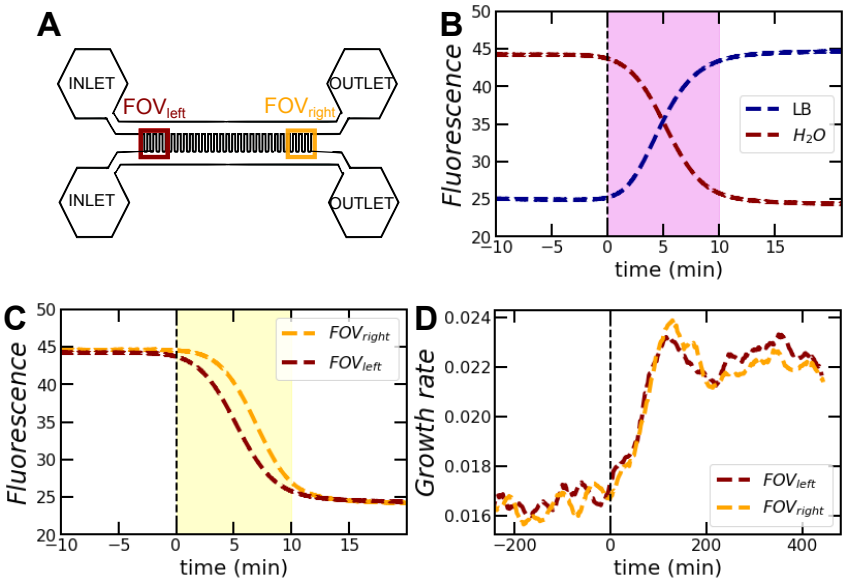


Figure D.2: Test of filling time and delays in nutrient availability within the microfluidic device. ***A*** *Scheme of the device with the outermost fields of view labelled in dark red and orange.* ***B*** *Experiment loading the device with fluorescent LB medium and monitoring mean fluorescence within the channels. Side channels start seeing the new medum immediately, and completely fill in about 4 minutes (purple shaded area) once the medium enters the main channel (t = 0).* ***C*** *The delay between growth rate changes in the experiment averaged in the outermost fields of view in the device is small.* ***D*** *Zoom of panel C showing that the delay is about 5 minute*

Appendix E

Data analysis pipeline

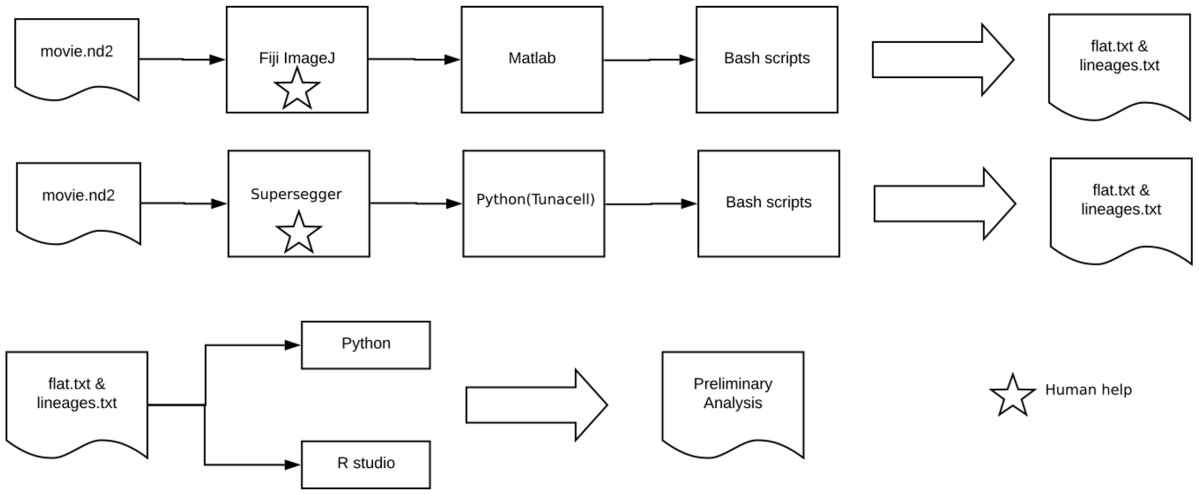


Figure E.1: Data analysis pipeline

1. Open Fiji.
2. Open the movie.nd2 file (Bio-Format Import).
3. Select the following options [view stack with Hyperstack, Color mode Default, Autoscale,Split channels,Split focal panels].
4. Select all the fields of view that you want to analyse and continue (it can take few minutes).
5. Many windows open, one for each FOV.
6. Open Macro “Backgroundrotation.ijm”, you’ll need to modify it based on your dataset.
7. In line 5 change field NFOV : number of fields of view that you want to analyse.
8. In line 12 (“Select window”) you should write the name (and path) of your file.

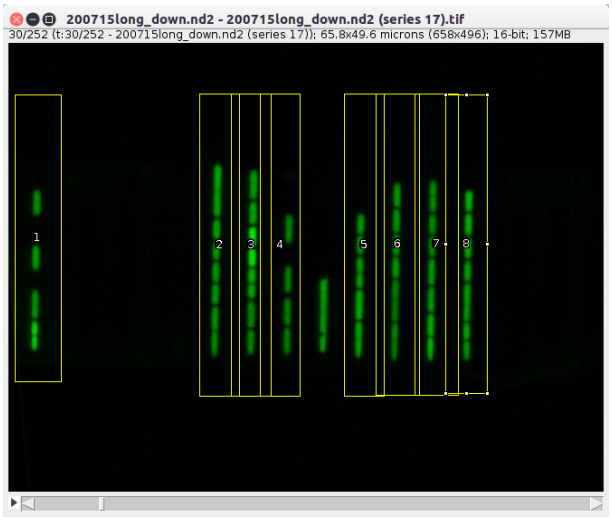


Figure E.2: Background subtraction and Microchannel detection with ImageJ

9. You can use this file to: subtract the background (line18), rotate the movie (line 16) and select the frames you want to analyse (line 17). Please uncomment if you want to use the function.
10. Press run and select the folder where you want to save the output.
11. One folder for each FOV is created.
12. In the folder a .tif file is saved.
13. Open Macro “select-channels.ijm”.
14. In file-FOV write the name of the file you want to analyse.
15. Open the .tif file related to the FOV you want to analyse.
16. Select all the channels manually by pressing “t”. In the window ROI manager you’ll have a list of all the position of all the channels.
17. Check if channels move. If yes you can use the Template Matching (align slices in a stack) plugin.
18. You can subtract the background again in the single channel.
19. Press run and select the folder relative to the FOV you are analysing.

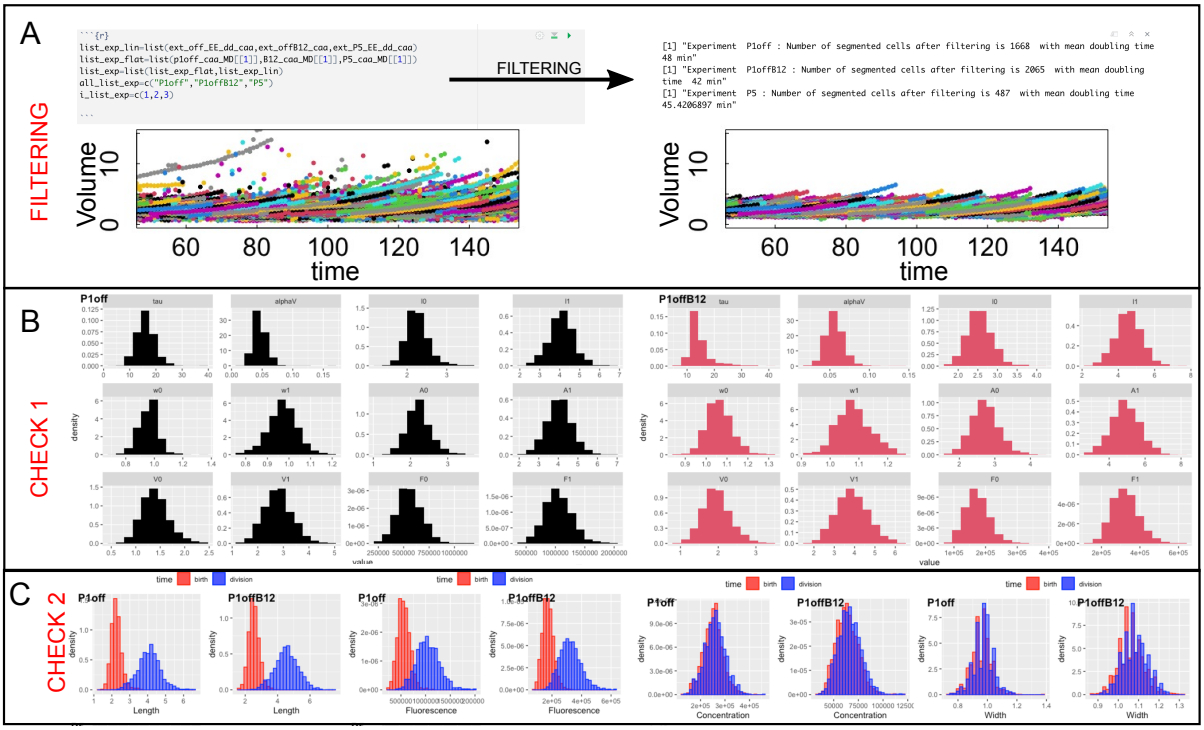


Figure E.3: Filtering and population distribution check *Pipeline to filter experimental data.* ***A*** *Comparison of dataset pre- and post- filtering. Distributions of main observables within all the experiment (****(B)****) or at birth and at division (****(C)****)*

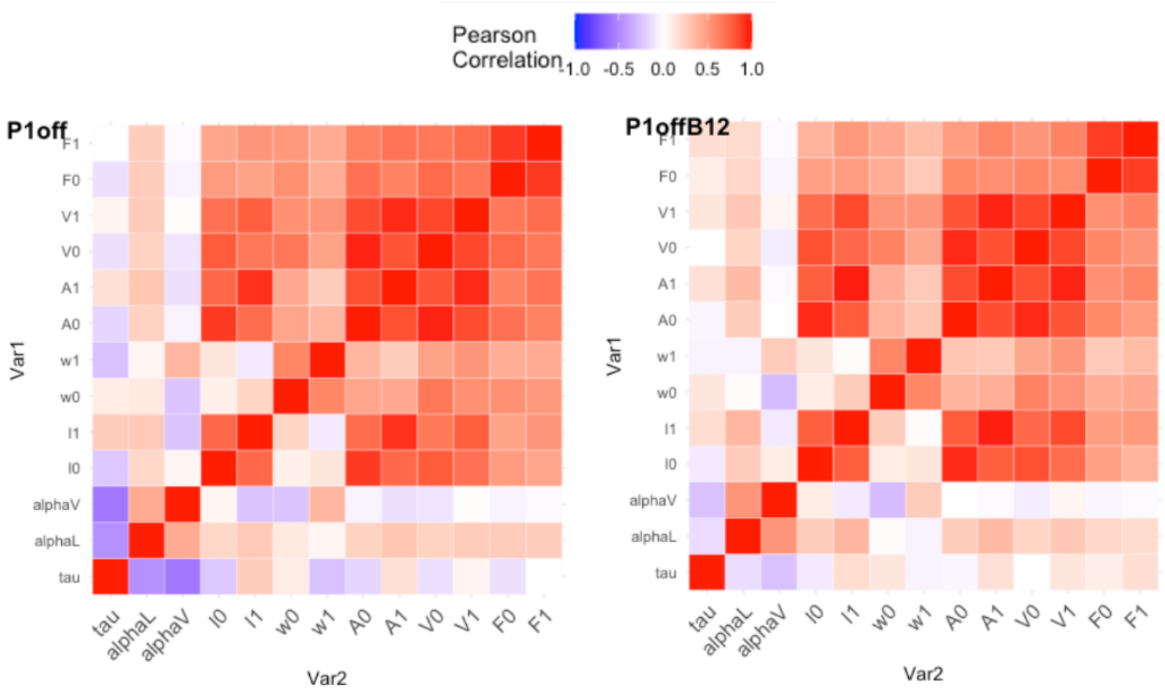


Figure E.4: Correlation matrix for main observables *Correlation matrix for main observables for two different strains. Correlation matrix does not change significantly between different strains.*

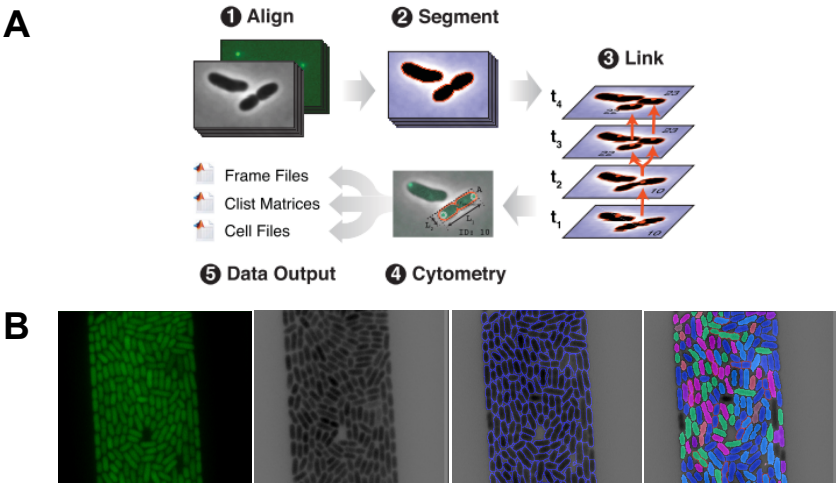


Figure E.5: Supersegger library for segmentation and tracking A *Supersegger workflow.The fluorescence and phase images are processed and aligned. During segmentation the cell regions are identified from the background. Then each cell region is linked to a corresponding cell region in the next frame and the cells receive unique ID numbers. Next, the properties and fluorescence characteristics of each cell are calculated. Finally, the program outputs three different types of outputs: Frame files, Clist matrices and Cell files. From [183]* ***B*** *Example of how segmentation by Supersegger works in chambers.*

Appendix F

Supplementary material Chapter 4

F.1 List of growth media and their composition

Chemical compounds	Composition (1L)
LB broth (Luria-Bertani medium)	10g Tryptone, 5g Yeast Extract, 5g NaCl
M9 5X Complementary Salts	10ml 1M MgSO4, 12.5ml Tymidine, 5ml Tryptophan, 500µL 1M CaCl2
M9 5X Salts	44.5g Na2HPO4.2H2O, 15g KH2PO4, 2.5g NaCl, 5g NH4Cl
SOC media	20g Tryptone, 5g Yeast Extract, 20ml 1M Glucose, 0.5g NaCl, , 10ml 1M MgCl2, 2,5ml 1M KCl.

Figure F.1: List of chemical compounds List of chemical compounds used in growth media

SLOW GROWTH MEDIA: M9 GLUCOSE

- 5X salts
- 5X complementary salts
- 20% glucose 0.4%
- Milli-Q water

FAST GROWTH MEDIA: M9 GLUCOSE CASAMINOACIDS

- 5X salts
- 5X complementary salts
- 20% glucose(0.4%)
- 20% casamino acids(0.5%)
- Milli-Q water

F.2 List of control and mutant strains

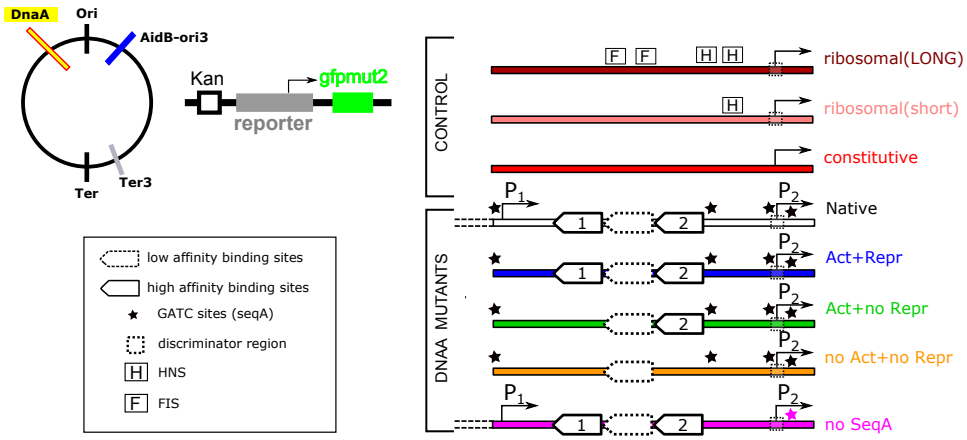
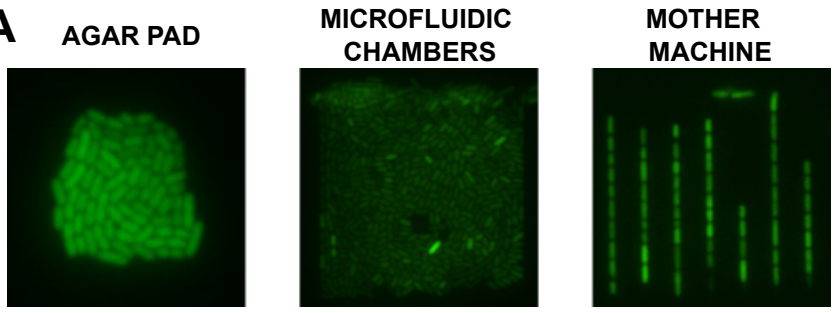


Figure F.2: Control and mutant strains

These experiments were carried out with the wild-type *E. coli* strain BW25113, the parent strain of the Keio collection [226] which has been fully sequenced [179].

F.3 Pro and cons of different experimental setups

A



B

	Agar Pad	Microfluidic Chamber	Mother Machine
Time-Lapse length	<6 hours	~3 days	~3 days
Number Of Cells in an Experiment	Start ~ 50 cells End ~	1800 Cells	500 Cells
Adaptation time	Bad	Good	Good
Single Cell Analysis	Moderate	Bad	Good
Difficulty in Segmentation	Moderate	Bad	Good
Statistics	Moderate	Good	Moderate
Time between recording and Upshift	Bad	Good	Good
Difficulty in capturing cell	Good	Average	Average
Control of growth medium	Bad	Good	Good

Figure F.3: Pro and cons of different experimental setup A *Agar pad, microfluidic chambers and mother machine can be use to have a single-cell resolution* ***A*** *In the mother machine we will have a better statistic and a better estimation of observables like area and width. On the other hand segmentation and tracking will take longer and be less accurate.*

Appendix G

Supplementary materials Chapter 5

G.1 Supplementary methods

G.1.1 Estimation of gene dosage

Since replication occurs in a time in the cell cycle that depends form a gene position, we expect that the production rate is sensitive to gene location. While at slow growth there are periods without replication activity, at fast growth replication occurs throughout the cell cycle for multiple nested chromosome copies [45]. Replication of the *E. coli* chromosome begins from a single origin and oppositely oriented replication forks proceed symmetrically along the genome to complete replication. Since on average a cell divides at a time $C + D$ ($\approx 60min$) after replication initiation, an average time lag B before initiation is necessary to make the total replication time $B + C + D$ an integer multiple of the doubling time τ. Thus, defining $n = \mathrm{Int}(C + D/\tau)$ as the integer number of times that τ divides $C + D$m one has that $B + C + D = (n + 1)\tau$ More generally, we can consider a gene at a chromosomal position defined by its normalized distance from Ori, i.e. $l = 0$ represents a gene in Ori and $l = 1$ a gene in Ter. The copy number of this gene, g , changes during the cell cycle following

$$g(t) := \begin{cases} 2^{n'} & \text{if } 0 < t < (n'+1)\tau - (c(1-l)+D) \\ 2^{n'+1} & \text{if } (n'+1)\tau - (c(1-l)+D) < t < \tau \end{cases} \tag{G.1}$$

where $n' = \mathrm{Int}\left[\frac{C(1-l)+D}{\tau}\right]$. By averaging over the cell cycle one gets the expected gene

$$g = \langle g(t)\rangle_{cell\ cycle} = \frac{1}{\tau}\int_0^\tau g(t)dt = 2^n\{1 - n + \mu[C(1-l)+D]\} \tag{G.2}$$

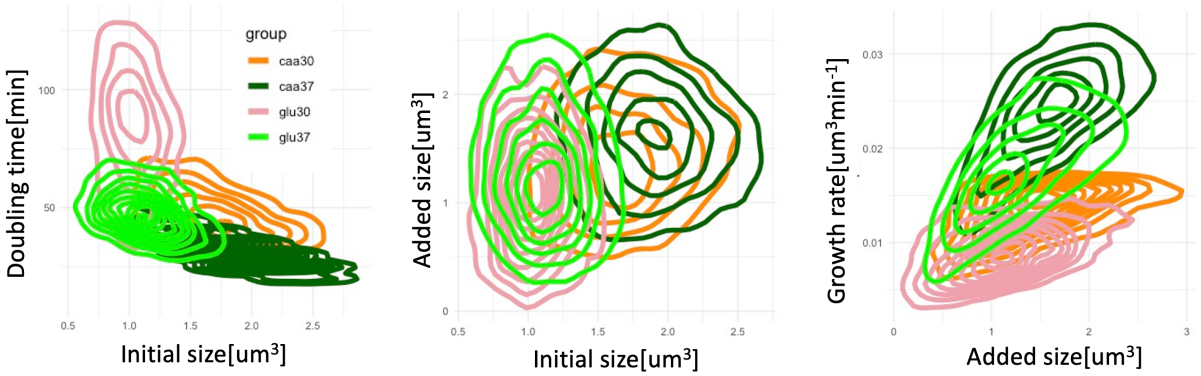


Figure G.1: Strains growing at 30° in glucose-CAA and in glucose at 37°C presents similar doubling times *Doubling time,size and growth rate are differently impacted by an increase in temperature. Temperature does not have a strong effect on cell size.*

G.2 Supplementary figures

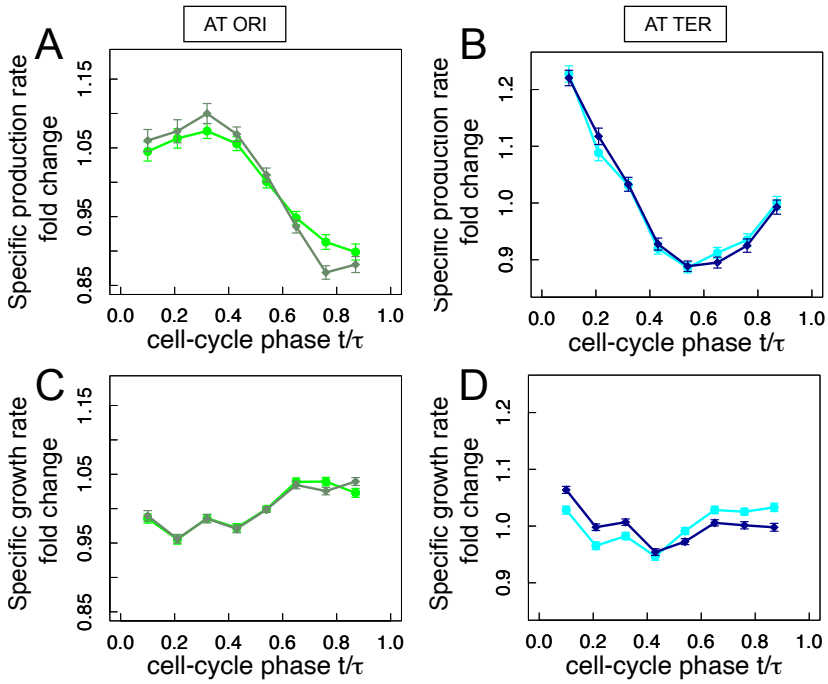


Figure G.2: Specific growth rate and production rate for a constitutive promoter at 30°C are biphasic *Specific production rate(**A-B**) and specific growth rate(**D-D**)) for a constitutive promoter at ori (left) or ter(right) as a function of cell-cycle present a biphasic pattern at 30°C. In both positions gene expression decreases in the first phase.*

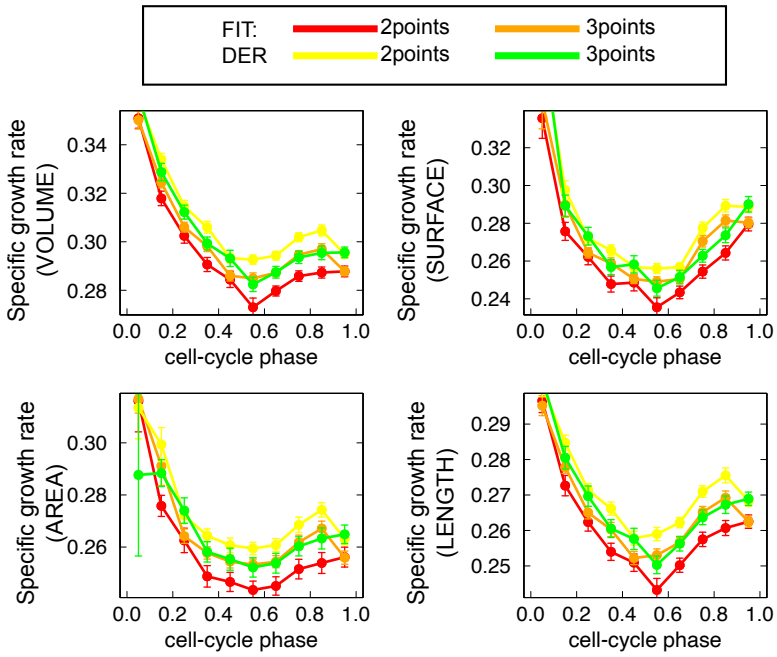


Figure G.3: Biphasic pattern in specific growth rate holds for different proxy of cell size *Specific growth rate can be computed by taking the slope of a 2-3 points fit or by taking the discrete derivative on 2-3 points. Volume, surface, projected area and length as a proxy of cell size. Biphasic growth pattern does not depend on these choices.*

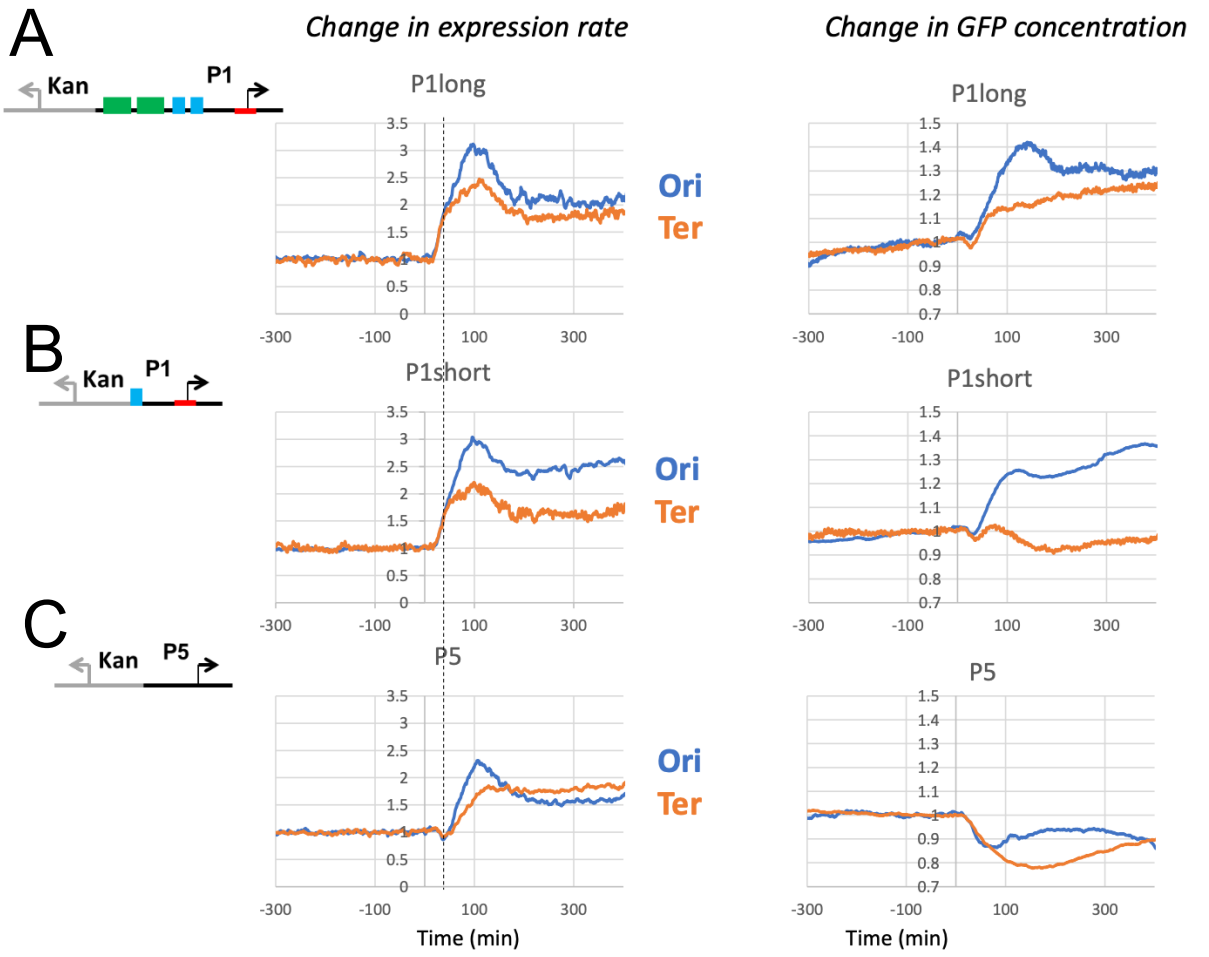


Figure G.4: Gene position affects Change the gene expression dynamics upon up-shift A *P1 promoter is induced at the same time in Ori and Ter* ***B*** *Activation by FIS acts mainly at Ter,probably because at ori there is already enough supercoiling from the other rrn* ***C*** *GFP from P5 is diluted during the overshoot and When P1 slows down P5 activity increases*

Appendix H

Supplementary materials Chapter 6

H.1 Supplementary methods

Considerations on DnaA-ATP concentration

We can consider a simple model of repression by DnaA-ATP and estimate Dna-ATP levels from the DnaA regulated promoter, assuming that the most relevant regulation is selective repression by DnaA-ATP. In such case, we can assume that the specific production rate is the product a basal expression rate r_0 and the probability that RNAP binds to the promoter, which is a decreasing function $P(D)$ of the DnaA-ATP concentration D. The synthesis rate of a regulated protein is often well described by a Hill function[293, 294] that in this case leads to:

$$\frac{1}{V}\frac{dF}{dt} = r_0 P_{bound} \qquad \text{where} \qquad P_{bound} = \frac{1}{1+\frac{D(t)}{k_D}} \tag{H.1}$$

In Equation H.1 any degradation of the protein has been ignored and the parameter r_0 denotes the rate at which the gene is transcribed in the absence of any regulation of the cell. Factors that contribute to r_0 include the number of available RNA polymerase molecules, and how well RNA polymerase binds to the promoter (its binding affinity). K_D reflects the strength of the repression (effective repression parameter). Physically, different strengths occur through differences in the binding affinity of the repression protein to its site on the promoter, as well as the strength of the interaction between the transcription factor and the RNA polymerase. Usually an exponent can be added to the $\frac{D(t)}{k_D}$ term (the Hill coefficient). It represents the way in which transcription factors can cooperate, with, for example, the binding of an initial transcription factor facilitating the binding of a second one.

We can invert Eq. H.1 and get an estimate for the DnaA-ATP concentration $D(t)$

If we consider also activation by DnaA-ADP, since the total DnaA concentration D_{tot} is roughly constant (Figure 6.3), P remains a monotonous decreasing function of D, since $P(D, D_{\text{tot}} - D)$ is a decreasing function of its first argument and an increasing function of

its second argument (which decreases when the first argument increases). Additionally, since Eq. H.1 represents transcription and the experimental time series are relative to the translation of our GFP reporters one can also con consider a delay of 5-10 mins [295].

By these considerations, we can conclude that the maximum of the specific production rate measured by our dnaAP reporter corresponds to a minimum in DnaA-ATP concentration, and vice versa, the minimum production rate to a maximum in DnaA-ATP concentration, with an estimated delay of 5-10 mins. We know that replication starts when DnaA-ATP concentration reaches its maxima (or right before, there are likely delays in the regulatory loop) and this in consistent with our observations that volumes at minima shows the same distribution and correlation patterns that we should see by looking at initiation volumes.

H.1.1 Slope estimation in correlation plots

To evaluate correlations between size fluctuations at different cell-cycle stages in presence of constraints we used the Bayesian method described in [9]. Shortly, the observed volume at maxima for a cell is constrained to be larger than the volume at birth. The correlation between volume at maxima et at birth is therefore masked by the fact that no maxima can be observed below the birth volume. Since the constraint is known, we can exploit the knowledge of the data that are visible in order to estimate the correlations.

H.1.2 Frequency locking for synchronized oscillators

For a fixed value of the forcing amplitude ϵ the frequency of the driven oscillator depends on the detuning $\omega_0 - \omega$. For sufficiently small detuning the external action entrains the oscillator, so that the frequency of the driven oscillator becomes equal to ω. For detuning exceeding a certain critical value, this equality breaks down.The identity of the frequencies that holds within a finite range of the detuning is the designated the synchronization region or the Arnold tongue(Figure H.1) . It is important to note that the synchronization region touches the ω-axis. This means that for vanishing detuning the oscillator can be synchronized by an infinitesimal force. This property is widely used to describe different systems and it is called frequency locking [229].

H.1.3 Causal relationships between two oscillators

The cross-correlation functions between pairs of observables reveal their temporal hierarchy,thus suggesting possible causality links between them.

Interpretation of cross-correlations gets more difficult the less is known about the underlying network which connects the two signals.

The Granger Causality measure determines how well a quantity Z can be predicted from its past values and tests whether the prediction is improved by taking the past of

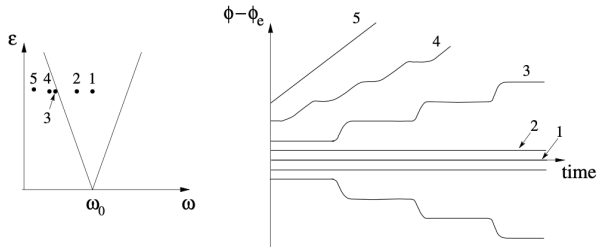


Figure H.1: Dynamics of the phase at the synchronization transition *The phase difference is shown in (right) for different values of the frequency of the driving force; these values are indicated in (left) by points 1–5, within and outside the synchronization region. In the synchronous state (points 1 and 2) the phase difference is constant (lines 1 and 2 in (right)); it is zero in the very center of the tongue and nonzero otherwise. Just outside the tongue the dynamics of the phase are intermittent: the phase difference appears as a sequence of rapid jumps (slips) intermingled with epochs of almost synchronous behavior (point and curve 3). As one moves away from the border of the tongue the dynamics of the phase tend towards uniform growth (points and curves 4 and 5). The transition at the right border of the tongue occurs in a similar way, only the phase difference now decreases. From [229]*

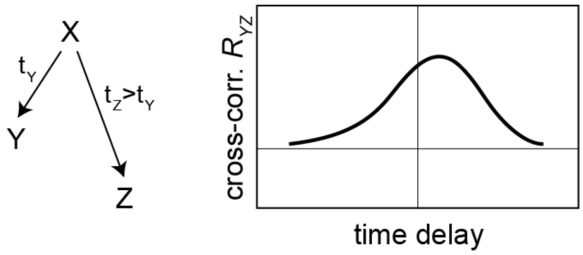


Figure H.2: Asymmetriccross-correlationsdo not imply causality *Fluctuations in an unknown or unmeasured component X propagate to the observables Y and Z with different time de- lays $t_Z > t_Y$. Then, the cross-correlation R_{YZ} is asymmetric, suggesting a noise propagation from Y to Z . However, in reality this link is ab- sent (see network structure on the left). From [39]*

the suspected cause Y into account. If the prediction is significantly improved by Y , then Y is said to Granger-cause Z .

Granger Causality is not equivalent to cross-correlations and can be more powerful. A prerequisite for performing the Granger Causality test is that the data need to be stationary i.e it should have a constant mean, constant variance, and no seasonal component. A crucial feature in this method is that it is model-free and it is able to disentangle the two directions of influence between the two variables, unlike the cross-correlations.

Neverteless, a key requirement of Granger causality is separability, meaning the casual variable is independent of the variable that it influences. This separability tends to be characteristic of stochastic and linear systems. For systems in which separability is not satisfied or which there are shared driving variables, Granger causality is not applicable.

A new approach has been suggested to try to understand causality in dynamical systems where effects of casual variables cannot be separated or uncoupled from the variables that they influence.

This approach, called Convergent Cross Mapping (CCM), was first tested in 1991 but was later evolved by Sugihara et al.for identifying causation in ecological time series [231]. CCM is a method for causality inference based on nonlinear state space reconstruction, a mathematical model commonly used in the theory of dynamical systems, and which can be applied to systems where causal variables have synergistic effects (unlike Granger causality tests).

CCM relies on Takens theorem and on the fundamental assumption that dynamics in the world are not purely stochastic and that in some applications, there are governing dynamics that can be represented by some underlying manifold.

Takens (1980) formally proved that time-delay embedding provides a 1–1 mapping of system dynamics from the original phase space (constructed with all system variables) to the reconstructed shadow phase space so long as the latter has sufficient dimensions to contain the original attractor [230].

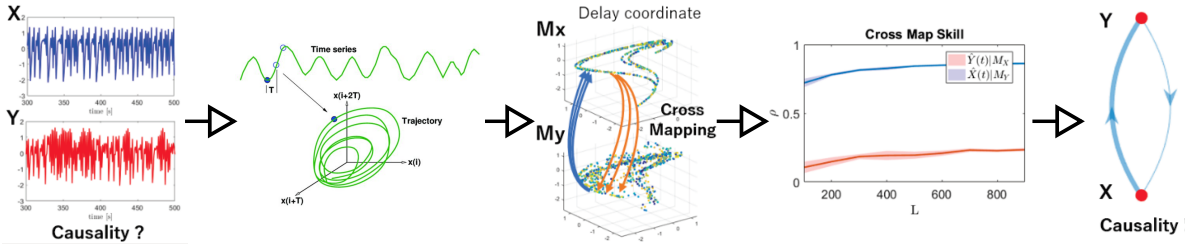


Figure H.3: Convergent cross-mapping *CCM looks for the signature of X in Y's time series by seeing whether there is a correspondence between the "library" of points in the attractor manifold built from Y, M_Y , and points in the X manifold, M_X, where these two manifolds are constructed from lagged coordinates of the time-series variables Y and X, respectively. L is the time-series length (sample size used to construct a library). The accuracy of predictions varies as a function of embedding dimension, which describes the size of the time windows that are used for prediction. One can estimate an optimal embedding dimension by using simplex projection to test the ability of a process to predict its own dynamics through leave-one-out cross-validation.Phase space attractor from the two time series*

The main steps of these method are summarised in Figure H.3. Shortly,we reconstruct a phase space attractor from the two time series. Next, "Shadow" attractor in reconstructed from only one time series Thanks to the delay-coordinate embedding theorem by Takens. Finally, conditional correlations with these "shadow attractors" are computed.Not symmetric conditional correlations reveal causal links. An R package called multispatialCCM is available to implement CCM.

H.2 Supplementary figures

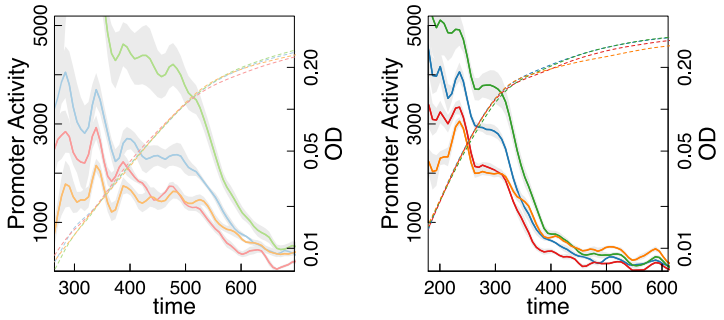


Figure H.4: The growth phase dependence of dnaAP2 activity is independent of autoregulation *Change in promoter activity and concentration for the different promoters in M9-glucose (left) and M9-glucose and casaminoacids(right). The promoter activity is more or less constant in exponential phase. This is followed by a rapid decrease in the entry into stationary phase, as expected for a promoter negatively regulated by ppGpp. Regulation by DnaA does not change the growth phase dependence of promoter activity, except for the box1.2 mutations promoter.The bacteria were grown in a 96-well plate at 37°C. OD and fluorescence were measured every 7 minutes. The error bar is from the average of 3-6 wells.*

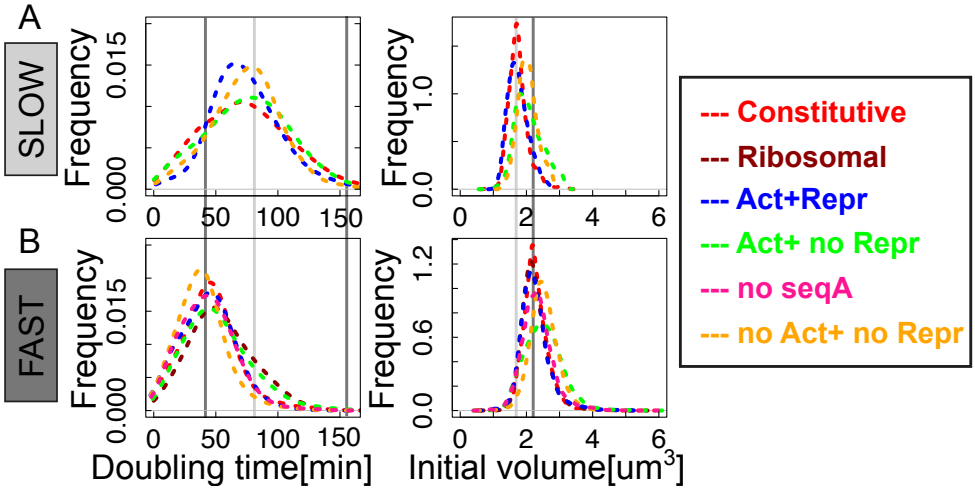


Figure H.5: Doubling time and initial size in consistent in all promoters in both fast and growth conditions A *At slow growth average doubling time(left) is centered around 75 mins for all strain while initial size (right) is centered around 1.9* μm^3 ***B*** *At fast growth average doubling time(left) is centered around 45 mins for all strain while initial size (right) is centered around 2.1* μm^3*.*

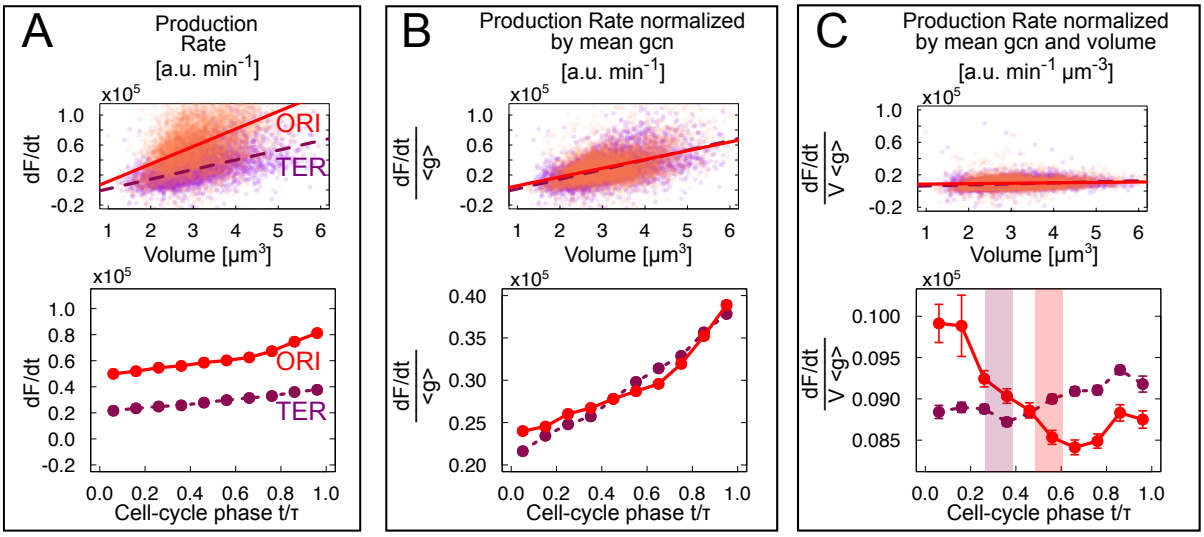


Figure H.6: The production rate of an unregulated promoter is proportional to size and gene dosage. ***A*** *Protein production rate is proportional to volume and gene dosage. Top panel: Scatter plot and trend lines (linear fits) of production rate vs cell size for origin proximate (red, solid line) and terminus proximate (purple, dashed line). Middle panel: normalization by average gene copy number corrects the trend due to gene dosage. Bottom panel, further normalization by volume removes the main trends.* ***B*** *Normalization by volume and average gene copy number leaves a small residual position-dependent cell cycle dependency in the production rate. The plots show binned averages of the quantities shown in panel A, computed with respect to cell-cycle phase (x axis). Normalizations by volume (bottom panel) shows small but significant cell-cycle dependent differences between the production rates of origin-proximate and and terminus-proximate unregulated promoters. Error bars are standard errors of the mean from a re-sampled distribution of the signal, obtained by bootstrapping from the experimental data for each bin.*

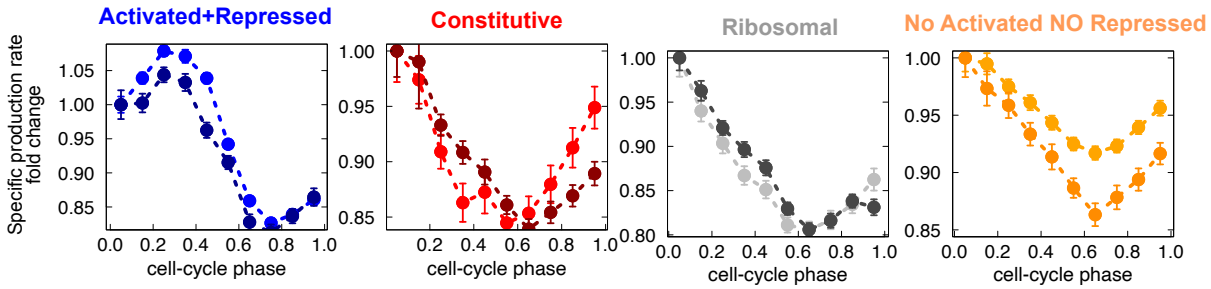


Figure H.7: Experiments present a good reproducibility. *Two replicates for the DnaA and constitutive promoter are shown. They present the same cell-cycle dependence.*

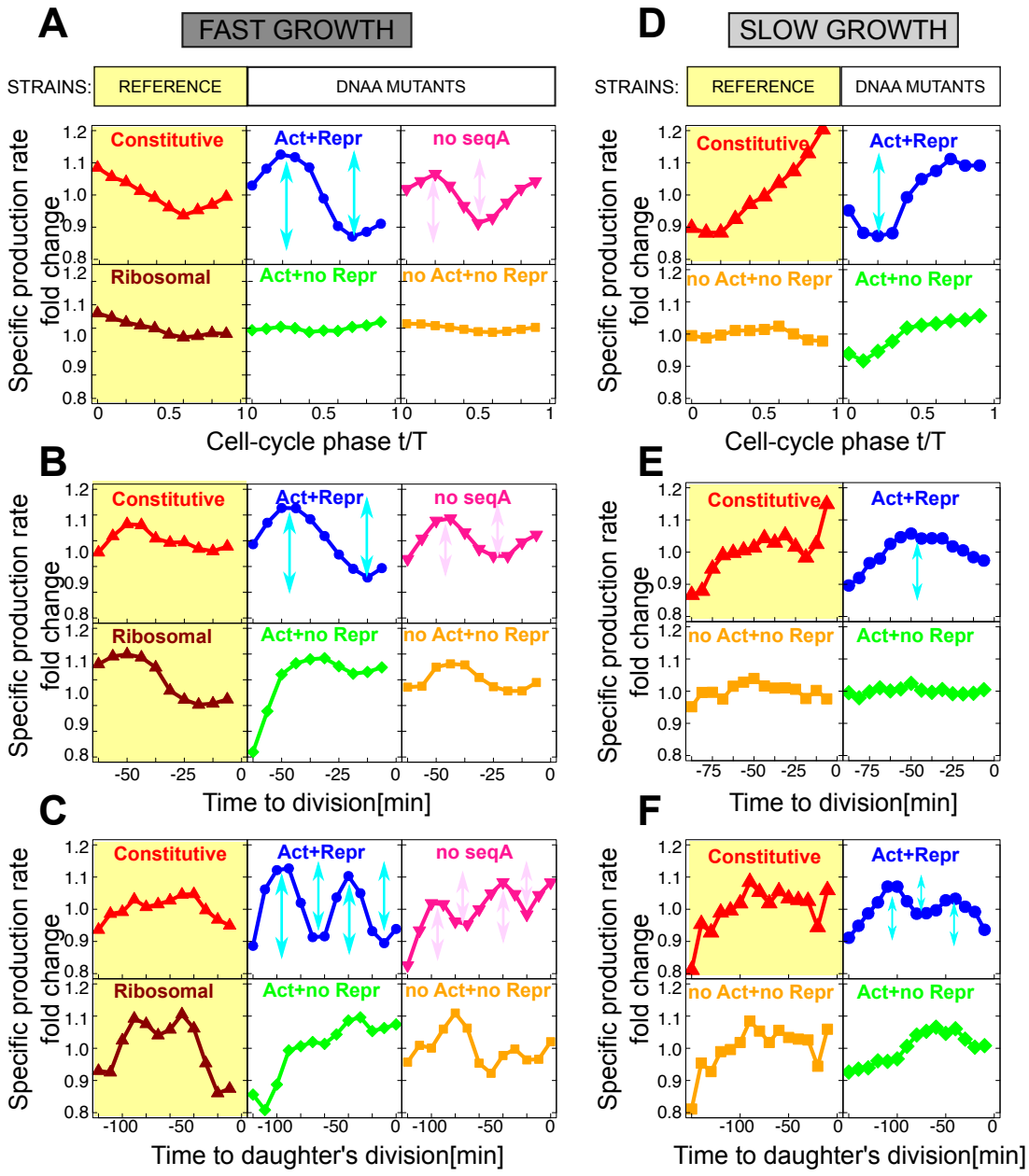


Figure H.8: Oscillation (in time) at fast and slow growth *Oscillation at fast(left) and slow growth(right) as a function of cell-cycle phase(**(A-D)**), time to division(**(B-E)**), time from birth(**(C-F)**). The promoter both activated and repressed (with or without seqA) presents stronger oscillations.*

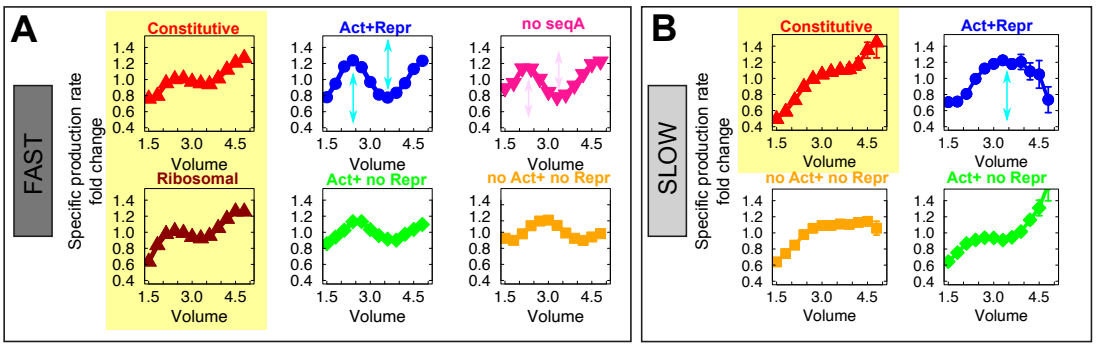


Figure H.9: Oscillation (in volume) at fast and slow growth *Oscillation at fast* ***(A)*** *and slow growth* ***(A)*** *as a function of cell volume.*

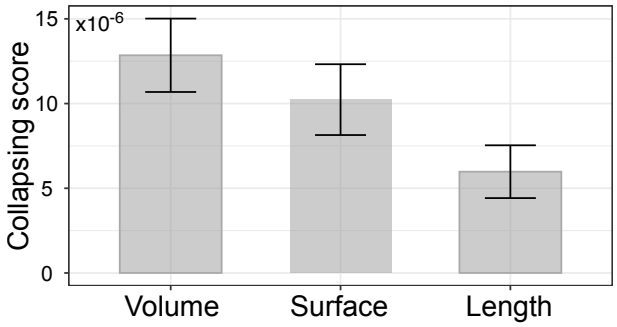


Figure H.10: Collapsing work better when we consider cell volume as a proxy of cell size compared to surface area or length *Collapsing score is computed considering volume,surface area and length as a proxy of cell size. The best collapsing is found by considering cell volume.*

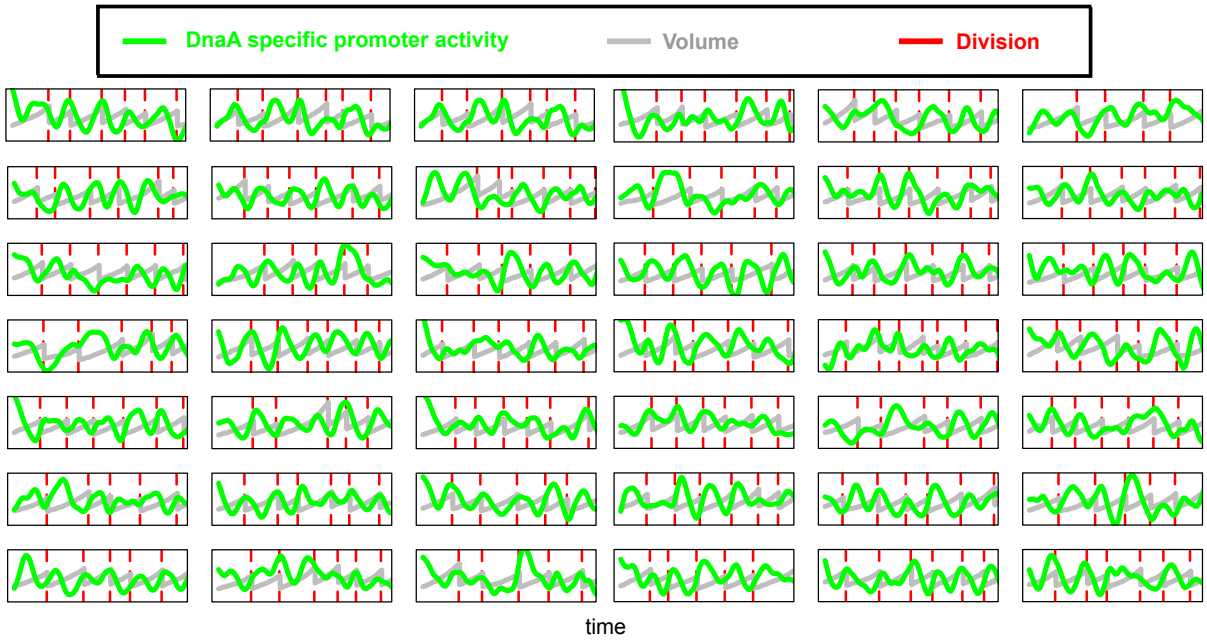


Figure H.11: Oscillations are observable also at the single-cell level *Set of lineages where volume(grey), DnaA promoter activity(green) and division(red) are tracked for several generations.DnaA promoter activity presents strong oscillation with sometimes 2 minima in the same cell cycle.*

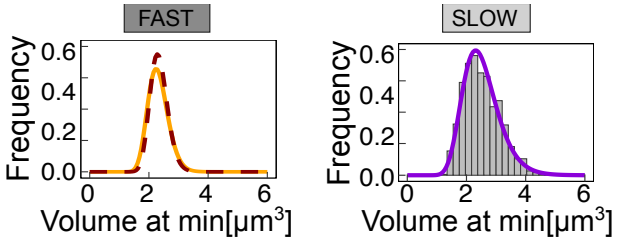


Figure H.12: Minima distribution is bimodal only at fast growth *As we have seen in the main text at fast growth minima present a bimodal distribution and the second peak collapse to the first when it is normalized by a factor of 1.7. At slow growth the minima distribution is unimodal and is consistent with the normalized distribution at fast growth.*

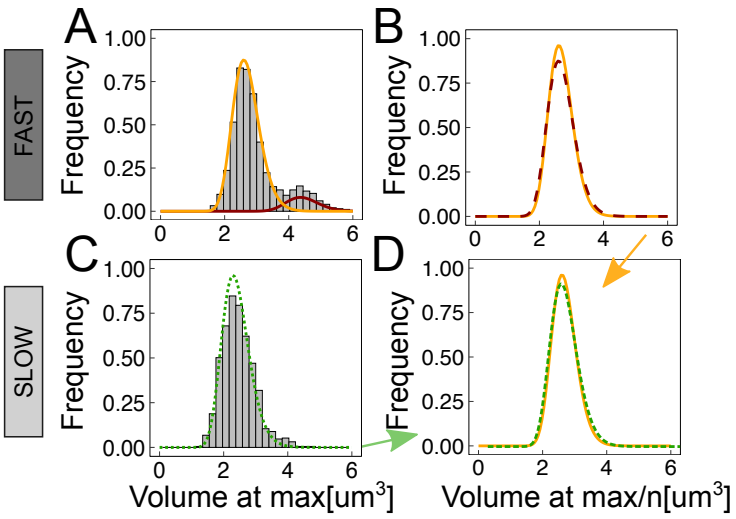


Figure H.13: Maxima,as minima, present a bimodal distribution ***A*** *At fast growth maxima distribution is bimodal and we can fit it with two lognormal distributions* ***B*** *Like for minima we can normalize the second peak by 2 to get an unimodal distribution.* ***C*** *At slow growth the distribution of maxima is unimodal and* ***C*** *it is consistent with the normalized distribution at fast growth.*

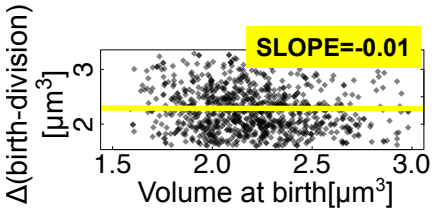


Figure H.14: Data support an adder at division *Scatterplot of volume added between birth and division versus volume at birth has a slope compatible with zero. This is compatible with an adder at division.*

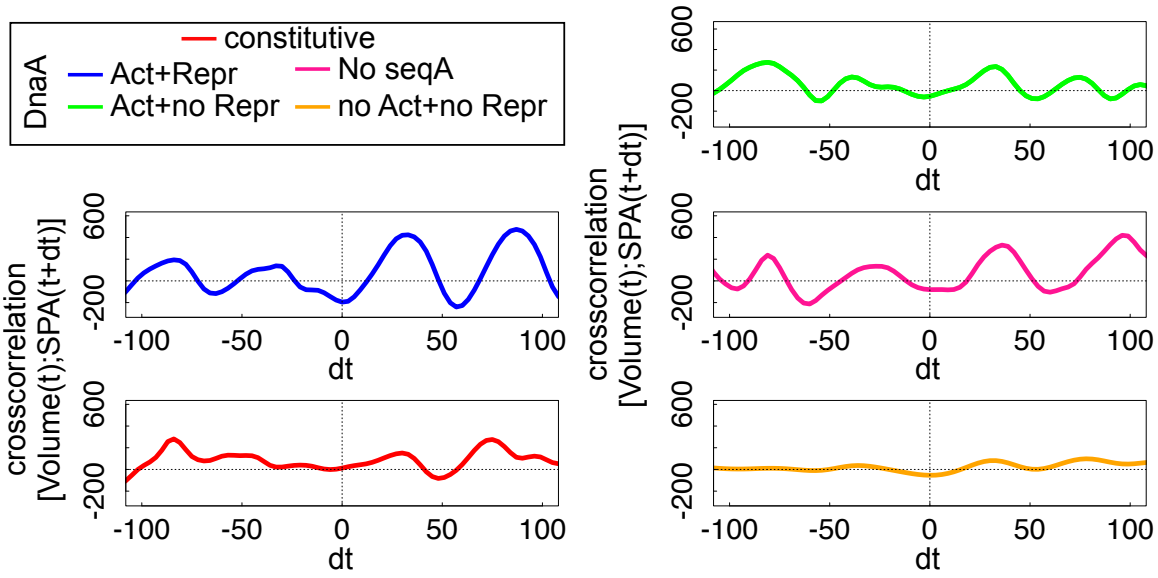


Figure H.15: Cross-correlation for different mutants*Cross-correlation between volume and specific promoter activity. Oscillation are present for each promoter. Promoter not regulated by DnaA show a weaker coupling from volume to specific promoter activity.*

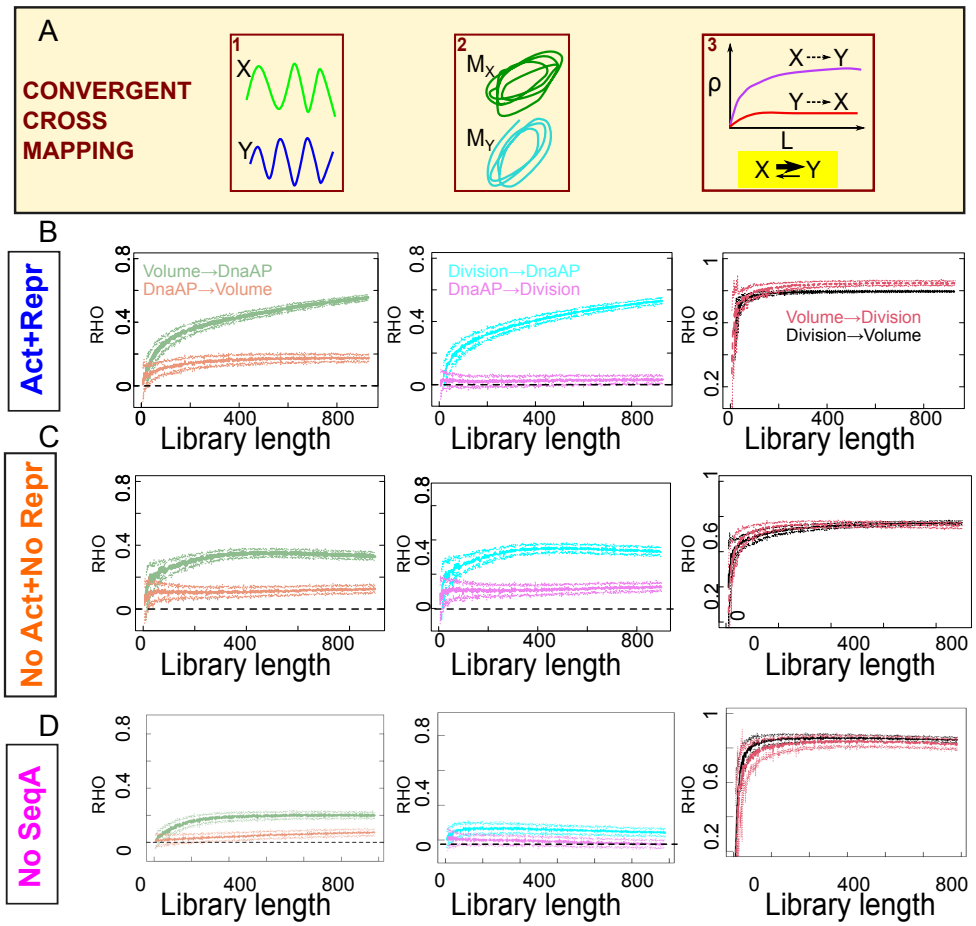


Figure H.16: Convergent Cross-mapping predict a "volume to DnaA" causality whose strength decreases without DnaA or seqA regulation ***A*** *Activated and repressed promoter present a strong "volume to DnaA" and "division to DnaA" causality* ***C-D*** *This causality is weaker for mutants not regulated by dnaA or by seqA. For all promoters volume and division are mutually coupled and crossmap in both directions(each variable can be estimated from the other).*

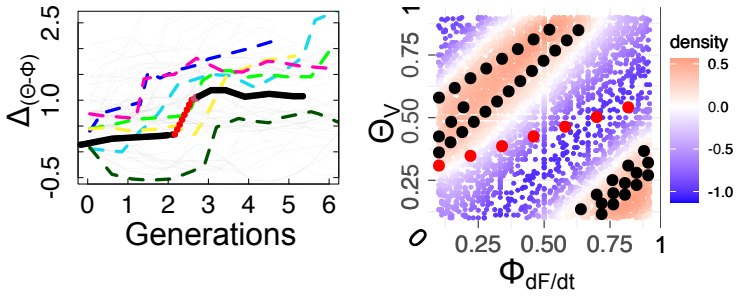


Figure H.17: Phase difference dynamic present phase slips A *Phase difference dynamic as a function of generations. Different lineages (colored dashed lines) present phase slips where phase difference increase by 1 unit. In the background lineage this phase slip is shown in red.* ***B*** *The same trajectory is reported in the phase space. The coupling function depends on the phase difference.*

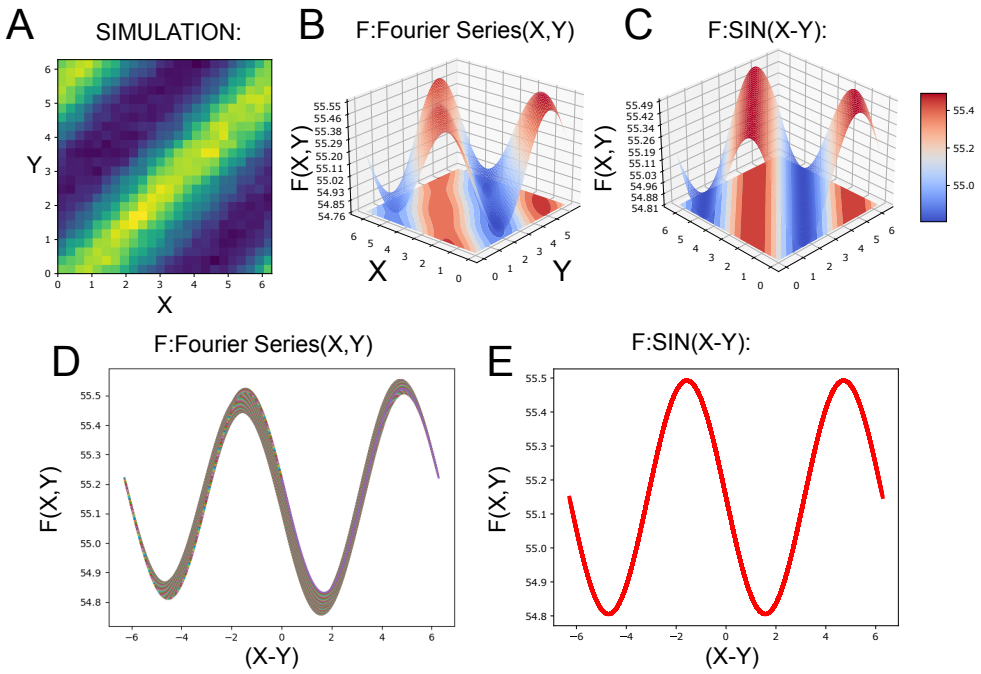


Figure H.18: Simulation of two coupled oscillators A *Phase space (X,Y) present same patter of experimental data* ***B-C*** *3D representation of coupling function F(Y,Y) obtained from sinusoidal fit or Fourier Series fit.* ***D-E*** *Same coupling function as a function of phase difference.*

Appendix I

Supplementary materials Chapter 7

I.1 Supplementary methods

Estimate of the expected gene-dosage effect in transposition.

This section discusses the model used to understand the null trend in the insertion frequency coming from gene dosage [43, 296]. The samples are grown in rich medium and when they are exposed to the transposon, there are a higher number of copies of the chromosome close to the origin than close to the terminus (Figure 1 in the main text).

We proceed to estimate the dosage theoretically. We assume that the $C+D$ periods last 40 + 20 minutes, since we are dealing with fast growth in LB medium, and that the doubling time τ is around 25 minutes (which is the most conservative estimate in terms of dosage bias). Since at time $C + D$ (≈ 60 minutes) the cell divides, a time lag B before initiation is necessary to make the total replication time $B + C + D$ an integer multiple of the doubling time τ, "synchronizing" DNA replication and cell division. Thus, defining $n = \mathrm{Int}\frac{C+D}{\tau}$ as the integer number of times that τ divides $C + D$, the following relation has to be satisfied,

$$B + C + D = (n + 1)\tau \ . \tag{I.1}$$

More generally [296], we can consider a gene at a chromosomal position defined by its normalized distance from the replicaiton origin (Ori), i.e. $x = 0$ represents a gene in the Ori locus and $x = 1$ a gene in the replication terminus (Ter). The copy number of this gene, $g(x, t)$, depends on its distance from Ori, and changes with time t from the beginning of the cell cycle following

$$g(x,t) := \begin{cases} 2^{n'} & \text{if } 0 < t < (n'+1)\tau - (c(1-x)+D) \\ 2^{n'+1} & \text{if } (n'+1)\tau - (c(1-x)+D) < t < \tau \end{cases} , \tag{I.2}$$

where $n' = \mathrm{Int}\left[\frac{C(1-x)+D}{\tau}\right]$ is the local number of active replication processes, which de-

pends on the locus position x.

To evaluate the average gene dosage in a cell population of asynchronous cells with an exponential growth we have to consider the distribution of the time into the cell cycle across the population [297]. For ideal balanced exponential growth with growth rate μ this distribution is given by $a(t,\mu) = 2\ln(2)\mu 2^{-\mu t}$. Averaging $g(x,t)$ over this distribution, the population-average gene copy number per cell becomes

$$g(x) = \langle g(x,t)\rangle_{population} = \int_0^\tau a(t,\mu)g(x,t)dt = 2^{\mu[C(1-x)+D]}. \tag{I.3}$$

In order to estimate the mean dosage of inserted promoters, we also have to introduce a model for insertion kinetics. The simplest model relies on the assumption of constant insertion rate. We hypothesize that during a time interval T a cell population that grows exponentially with division time $\langle\tau\rangle$ is exposed to a transposon which inserts its sequence at a constant rate r per time and genome coordinate unit. The expected number of insertions under this assumption can be obtained by the cumulative distribution function. The mean copy number of each coordinate in a cell population is $g(x)$ and we can assume that there is a Poisson process with the same rate (S) in each genome position, so that the dependency of the rate from position x is only due to dosage, $r(x) = g(x)S$. In this case, the estimate for the insertion probability is

$$P(x) = 1 - e^{r(x)T} \simeq r(x)T \simeq g(x)ST = Ag(x)\ , \tag{I.4}$$

where we have linearized the expression for small T and $A = ST$ is the only free parameter, which depends on the parameters average gene dosage $g(x)$, time interval T and growth rate μ, all derived from experimental data. The trend of normalized position and the dosage estimate for insertions is shown in Fig. 1B in the main text. This plot shows that this estimate does not correspond well quantitatively to the experimental coverage data.

We verified that a model with time-dependent insertion rate, i.e. where the insertion rate r *increases* with time t, $r(x,t)$, can fit the data, using insertion rate growing as a power law in time. However, there is no empirical motivation to assume such cooperative behavior in insertions that occur *in different cells*.

We now address the simple point that the frequency of an insertion is not biased by population growth, as long as it does not confer a sufficiently strong positive or negative fitness effect. Let f_i be the fraction of insertions in site i and f_j the fraction of insertions in site j. If N is the total number of insertions we will have, at time τ, $N * f_i e^{\alpha\tau}$ insertions in site i and $N * f_j e^{\alpha\tau}$ insertions in site j. Then, the total number of insertions will be $Ne^{\alpha\tau}$.

Let $f_{i,j}(\tau)$ be the ratio of the numbers of insertions in site i and j. Then,

$$f_{i,j}(\tau) = \frac{N * f_{i,j}e^{\alpha\tau}}{Ne^{\alpha\tau}} = f_{i,j}\ . \tag{I.5}$$

Since $f_{i,j}(\tau) = f_{i,j}$, for all those insertions not affecting the growth rate, exponential propagation of insertions in progeny does not create a bias in the population. If instead an insertion significantly affects the growth rate of the corresponding clone, its presence will be biased in the population.

Null model for insertion enrichment.

This section describes the null model used to score enrichment of insertions in particular regions defined by gene lists. Since the theoretical prediction scores poorly (and even a fit with a time-dependent insertion rate model does not capture quantitatively the shape of the dosage peak around the origin), in the analysis for scoring overlap between insertions and gene lists we decided to use a more conservative way to subtract the dosage, i.e. directly to consider a sliding-window average of the data as background insertion probability. Specifically, we used a window size of 3000 bp, which is the smallest window size that gives a smooth curve with few local features.

In order to investigate statistical tendencies for transposon insertions to be associated to specific chromosomal contexts, we need a suitable null model. This is simple to define using gene lists, as randomizations of the empirical lists having fixed number of genes. In particular, for each tested gene list we simulate 5000 different stochastic realizations.

The definition of enrichment needs a quantity representative of the tendency of finding the insertion fixed in specific chromosomal contexts. We evaluate a discrete integral considering the coverage corresponding to a portion of genome defined by the genes of the list. We call this quantity "coverage integral". It corresponds to the total overlap of the coverage with the genes of a specific list.

We have evaluated the Z-score as significance score for over/under-representation of the overlap between coverage and target gene lists. The Z-score is defined as $Z = \frac{X-\mu}{\sigma}$ where X is the empirical value and μ and σ are the mean and the variance relative to a particular sample.

DNA primers used:

To verify the insert on the Tn5 construction vector:
Forward Primer : AATTCTACGAGCGTCGTCGCAGACATGATC
Reverse Primer : GGTCTCAGATGGTAGACGCAGGAAGAGACGAGACAG
Primers used for transposon sequencing experiments (* indicates a phosphorothioate group:
Index primer: AAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC
3' Tnspecific primer:
AATGATACGGCGACCACCGAGATCTACACCCAATATGCGAGAACACCCGAGAAAATTCATCG
3' Sequencing primer: CCCGAGAAAATTCATCGATGATGGTTGAGATGTGTA

Illumina Read 1: ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Illumina Read 2: CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
QPCR2.1 : CAAGCAGAAGACGGCATACGA
qPCR2.2 : AATGATACGGCGACCACCGAG
Adapters:
SplA5_top G*AGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
SplA5_bottom /5Phos/G*ATCGGAAGAGCGGTTCAGCAGGttttttttttcaaaaaaa*a
SplAP5.2 C*AAGCAGAAGACGGCATACGAGATAAACATCGGAGATCGGTCTCGGCATTC*C
SplAP5.5 C*AAGCAGAAGACGGCATACGAGATACCACTGTGAGATCGGTCTCGGCATTC*C
SplAP5.6 C*AAGCAGAAGACGGCATACGAGATACATTGGCGAGATCGGTCTCGGCATTC*C
SplAP5.12 C*AAGCAGAAGACGGCATACGAGATAGTACAAGGAGATCGGTCTCGGCATTC*C

Supplementary Figures

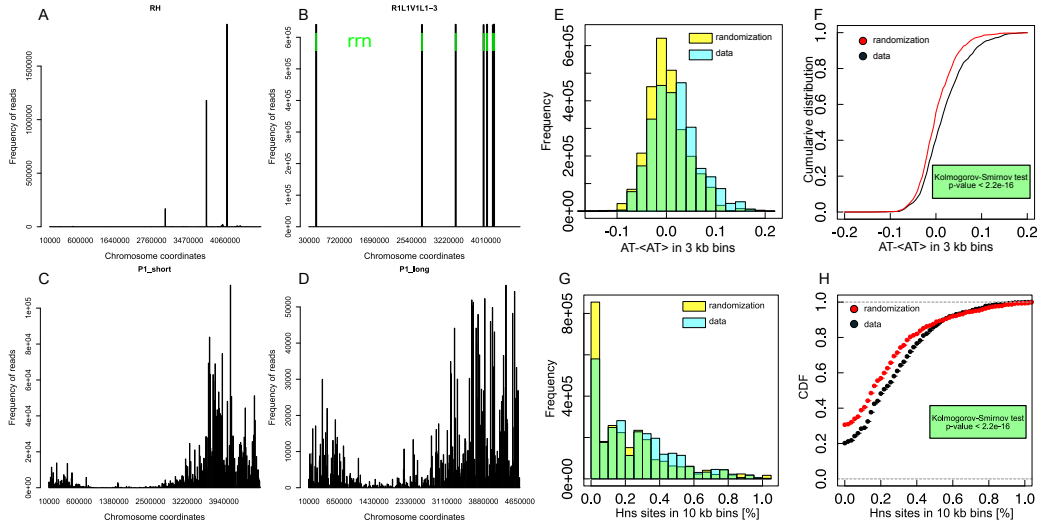


Figure I.1: **Insertion frequency plots and AT bias.** *A): Highly expressing sorted (RH) population shown in Figure 1D shown in linear scale. B): Insertions in very low expressing sorted population (RLR1V1L), shown in linear scale, transposon insertions in rRNA operons are visible close to the origin. C) and D): Insertion frequency plots for parental populations of P1-long and P1-short promoters respectively. The plots were generated from TraDis Data by binning the mapped position of the each read into 10kb non overlapping bins, and the frequency of insertions of each bin was plotted against the chromosome coordinate. E) and F) Insertions have a significant positive bias in AT-richness compared to the background. E): Histograms of deviations from average in AT-richness ($\%AT - <\%AT>$) in the sequences surrounding insertions from the experimental data and from a randomized sample of the background sequences of the genome. $\%AT$ corresponds to a local average in 3Kb intervals around each insertion; $<\%AT>$ refers to the average on the entire genome. F): comparison of the cumulative distributions and results of the Kolmogorov-Smirnov test (in both cases the p-value is smaller than the smallest represented by the ks.test R script). G) and H) Insertions have a positive bias for H-NS binding. The plots report the distribution of the fraction of the 10kb region around an insertion covered by H-NS binding sites (ChIPseq data from ref. [215]), compared to the same quantity for randomized insertion sites.*

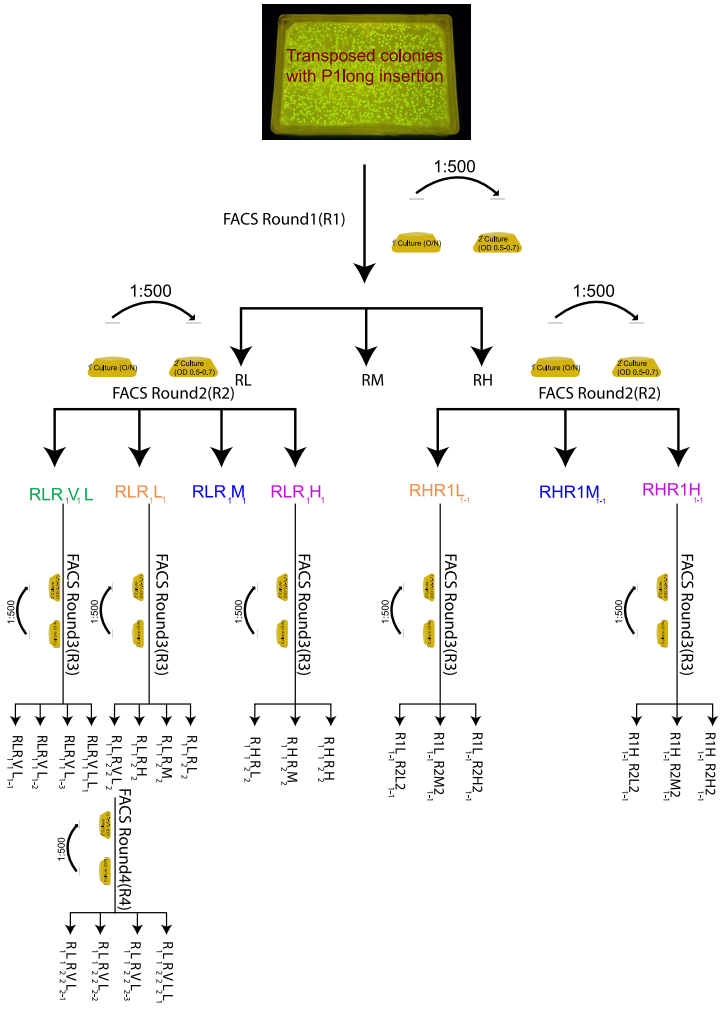


Figure I.2: **Full illustration of the sorting pipeline starting from the parental population and the nomenclature used in our experiment.** *The parental population (the cells that underwent transposition) was grown overnight in M9 medium with 0.5% Glucose + 0.2% CAA ("fast growth" medium) and was diluted to 1:500 to grow in secondary culture with the same medium the following day until the OD reached 0.5 – 0.6. These cultures underwent first round of FACS to yield three population by the level of GFP expression (Low (RL), Medium (RM) and High (RH)). Subsequently the Low (RL) and High (RH) underwent the second round of FACS, the former (RL) yielded 4 population, RLR1V1L1 (Very low), RLR1L1 (Low), RLR1M1 (Medium) and RLR1H1 (High) while the latter (RH) yielded three population, RHR1L1-1 (Low), RHR1M1-1 (Medium) and RHR1H1-1 (High). From the second round of FACS, three populations, RLR1V1L1 (Very low), RLR1L1 (Low), and RLR1H1 (High) were selected from the low expressing population while 2 populations, RHR1L1-1 (Low) and RHR1H1-1 (High) from the high expressing population were selected for the third round of FACS. RLR1V1L1 (Very low) gave rise to 4 different populations, RLR1V1L1L1, RLR1V1L1-1, RLR1V1L1-2, RLR1V1L1-3 and RLR1L1 (Low) also gave rise to 4 different population, R1L1R2V2L2, R1L1R2L2, R1L1R2M2, R1L1R2H2 while RLR1H1 (High) gave rise to 3 different populations, R1H1R2L2, R1H1R2M2 and R1H1R2H2. From the high expressing populations, RHR1L1-1 and RHR1H1-1 gave rise to R1L1-1R2L21-1, R1L1-1R2M21-1, R1L1-1R2H21-1, and R1H1-1R2L21-1, R1H1-1R2M21-1, R1H1-1R2H21-1, respectively. Only the very low expressing population, R1L1R2V2L2 from the third round of FACS gave rise to 4 different very low expressing populations, R1L1R2V2L2-1, R1L1R2V2L2-2, R1L1R2V2L2-3, R1L1R2V2L2-4.*

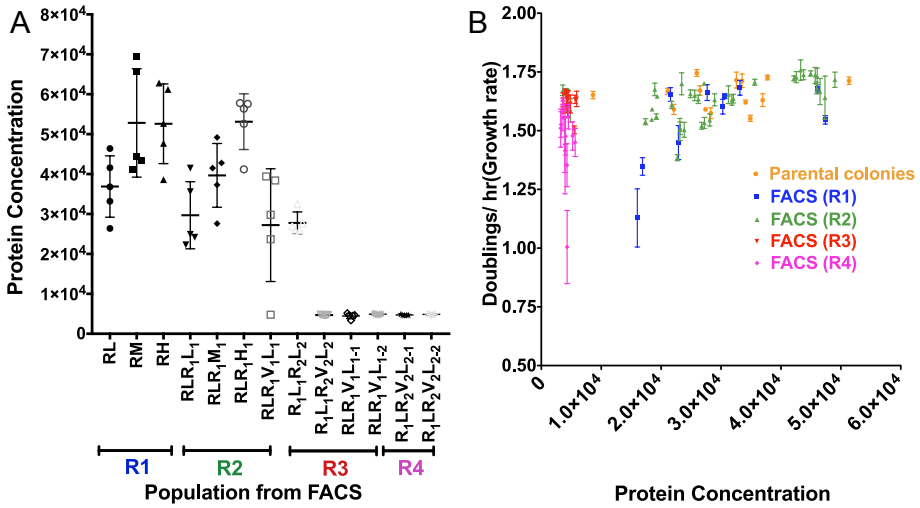


Figure I.3: **Expression levels of clones from sorted populations measured by fluorimetry, and growth rate.** *(A) The fluorimeter expression data are in agreement with the FACS population histograms. GFP expression (total fluorescence over OD) of 5 individual strains randomly chosen from each of the different rounds of FACS sorted population. All the strains were grown in rich medium (0.5% Glucose + 0.2% CAA). Error bars represent standard error of the mean (SEM). R1, R2, R3, R4 denote the first, second, third and fourth round of FACS. Expression is low for very low expressing population (VL in R3 and in R4) and significantly different from the other sorted populations ($P < 0.001$, One-way ANOVA, Tukey's multiple comparison test comparing R1, R2 with R3 and R4). The fluorimeter expression data are in agreement with the FACS population histograms. (B) Scatter plot of expression levels and population growth rates for clones from the different rounds.*

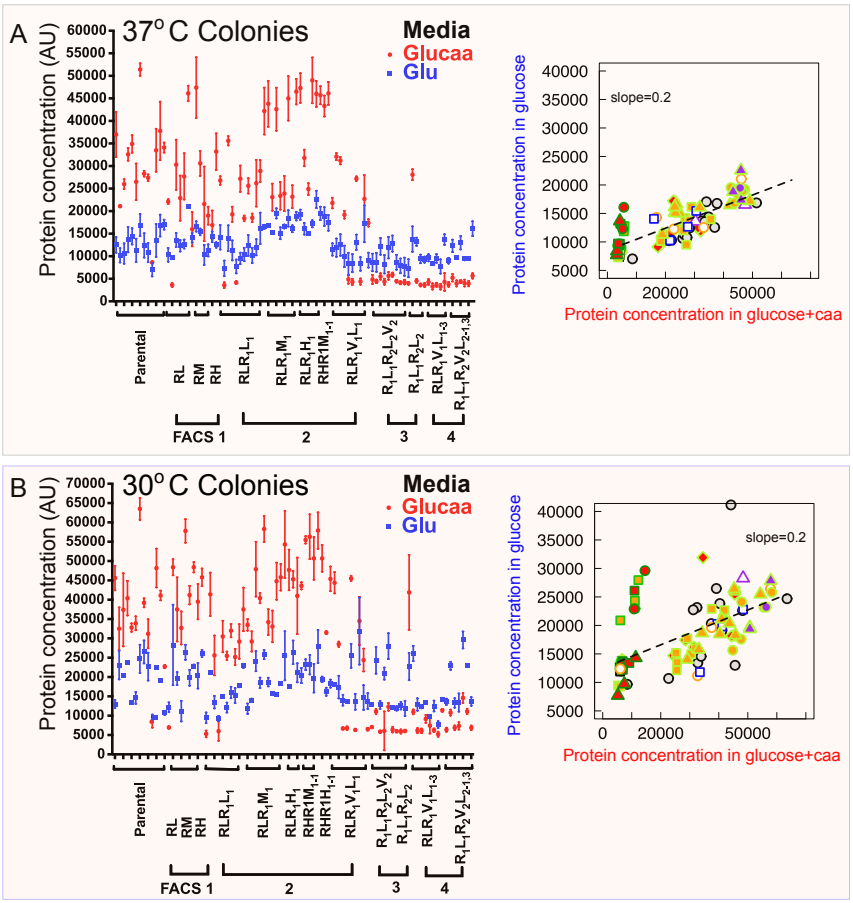


Figure I.4: **Insertions at the ribosomal regions are highly repressed in different growth conditions.** *The expression levels of these colonies are lower when they are grown in a fast-growth medium (M9 + 0.5% Glucose + 0.2% CAA) compared to the slower-growth medium (M9 + 0.5% Glucose) irrespective of the temperature grown (37°C and 30°C). Panel A) reports the gene expression of 90 clones from different rounds of FACS grown in different media (fast growth, Glucose + CAA, compared to the slow growth medium, Glucose at 37°C). Error bars represents standard deviation (SD), n = 3. Panel B) represents the comparison of colonies grown in different media (fast vs slow growth medium) at 30°C, error bars represent SD, n = 2. The scatter plots on the right side of each panel compare gene expression of the same clone in the two growth conditions (color/symbol codes defined in Supplementary Fig. 5). Note that while most of the other clones increase in expression in the faster growth medium (in agreement with a ribosomal promoter) the very-low expression (VL) ones decrease with the faster growth rate, in agreement with the idea that their expression is decreased by the activity of a ribosomal promoter. The comparison between the two temperatures shows that there is a subset of clones that increase in expression in the slower growth Glucose medium at low temperature, spanning different sorted populations. Note that thereare two growth rate dependence results, the first is the fact that we grew the cells in two growth media and we observed an increase in gene expression in the richer medium, as expected from the rrnBP1 promoter, the second is that within a given growth medium different clones grew at different rates (Fig. S3B).*

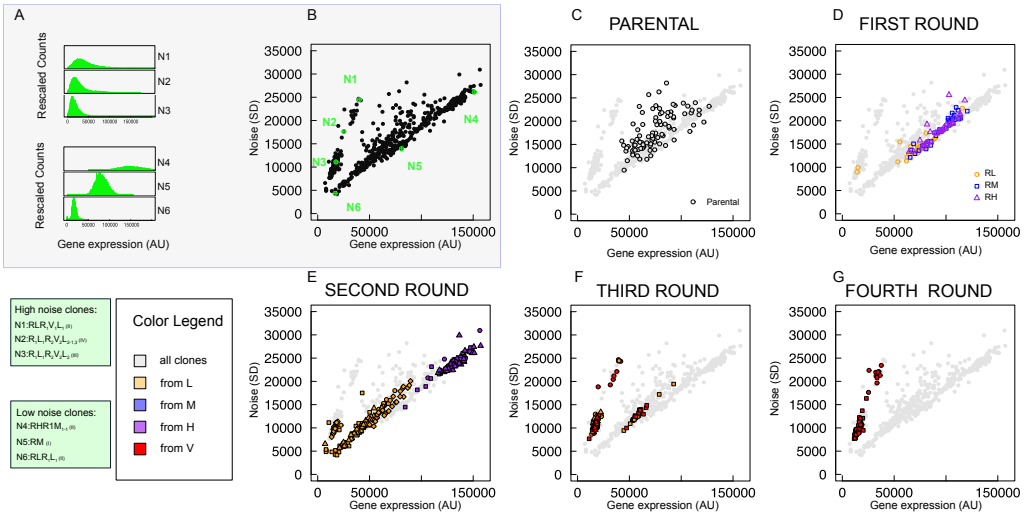


Figure I.5: **Gene expression levels and noise of clonal populations from FACS.** *The figure recapitulates the main results of the selected clonal populations (hence carrying the same insertions). A) and B) N1-N6 are six representative clones chosen from all the measured ones. Panel B (the same plot shown in Figure 3A) shows their gene expression mean level and noise (SD), while panel A compares the histograms of their gene expresion levels. Panels C-G use the same plot to show the noise properties of the clones coming from different rounds of FACS sorting.*

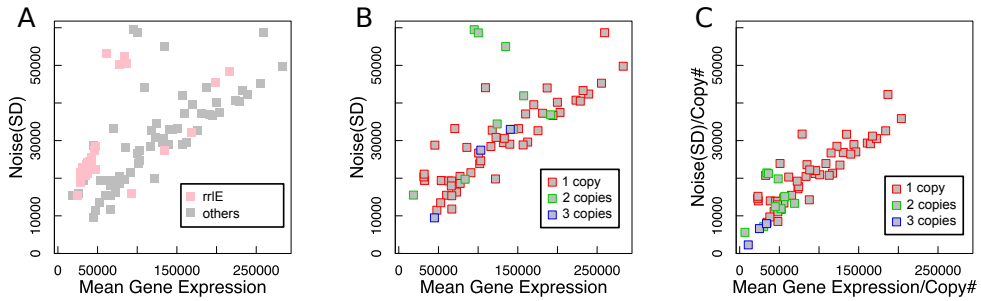


Figure I.6: **Gene expression noise categories of sequenced clones with varying insertion copy number.** *A) Mean vs SD plots of rrlE vs non-rrlE insertions. Noisy promoters mostly associate with rrlE insertions in sequenced clones. B) Mean vs SD plots of non-rrlE sequenced clonal populations (a subset of the data in Supplementary Fig. S5), where the points are colored by number of insertions. C) Different plot of the same data, where gene expression normalized per copy number. Removal of the effect of copy number fold change in the expression levels reconducts many non-rrlE outliers to the cluster of low-noise promoters.*

Supplementary Tables

Genes linked to specific phenotypes (essential genes and genes responsive to external conditions)									Mutated genes		
Multi stress responsive genes (from Nichols et al 2011)	Conditionally essential genes due to auxotrophy (from Nichols et al 2011)	Conditionally essential genes in rich media (from Nichols et al 2011)	Genes whose expression is affected by growth conditions (from Regulon DB)	Genes whose expression is affected by the carbon source (from RegulonDB)	List of E. coli essential genes (from Baba et al 2006)	List of E. coli essential genes (from Gerdes et al 2003)			List of mutated genes in E. coli long-term evolution experiment through 20,000 generations (Barrick et al 2009)	List of mutated genes in E. coli long-term evolution experiment through 40,000 generations (Barrick et al 2009)	
Genes next to binding sites of a regulator from binding data (ChIP-chip, ChIP-seq, etc)											
Putative RNA-polymerase target genes during rapid growth from ChIP-chip experiments (from Grainger et al 2005)	Genes overlapping with heEPODs (highly expressed extended protein occupancy domains) (from Vora et al 2009)	Genes overlapping with lsEPODs (transcriptionally silent extended protein occupancy domains) (from Vora et al 2009)	Putative CRP target genes from ChIP-chip experiments (from Grainger et al 2005)	Putative FIS target genes from ChIP-chip experiments (from Cho et al 2006)	Putative FIS target genes from ChIP-seq experiments (Kahramanoglou et al 2010)	Putative FIS target genes in early-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	Putative FIS target genes in mid-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	Putative FIS target genes in mid-log phase from ChIP-chip experiments (from Grainger et al 2006)	Putative FNR target genes in stationary phase from ChIP chip experiments (from Grainger et al 2007)	Putative FNR target genes in midlog phase in the presence of oxygen from ChIP-chip experiments (from Grainger et al 2007)	Putative FNR target genes in midlog phase from ChIP-chip experiments (Grainger et al. 2007)
Putative H-NS target genes from ChIP-chip experiments (from Oshima et al 2006)	Putative H-NS target genes from ChIP-seq experiments (Kahramanoglou et al 2010)	Putative H-NS target genes in early-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	Putative H-NS target genes in mid-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	Putative H-NS target genes in mid-log phase from ChIP-chip experiments (from Grainger et al 2006)	Putative H-NS target genes in stationary phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	Putative H-NS target genes in transition-to-stationary phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	Putative IHF target genes in mid-log phase from ChIP-chip experiments (from Grainger et al 2006)				
Genes sensitive to nucleoid perturbations from transcriptomics experiments											
FIS knockout sensitive genes in all stages of growth (from Bradley et al 2007)	FIS knockout sensitive genes in the early-exponential growth phase (from Bradley et al 2007)	FIS knockout sensitive genes in the late-exponential growth phase (from Bradley et al 2007)	FIS knockout sensitive genes in the mid-exponential growth phase (from Bradley et al 2007)	FIS knockout sensitive genes in the stationary phase (from Bradley et al 2007)	FIS-knockout sensitive genes under negative supercoiling (from Blot et al 2006)	FIS-knockout sensitive genes under positive supercoiling (from Blot et al 2006)	H-NS knockout sensitive genes under negative supercoiling (from Blot et al 2006)	H-NS knockout sensitive genes under positive supercoiling (from Blot et al 2006)	Supercoiling sensitive genes under FIS knockout (from Blot et al 2006)	Supercoiling sensitive genes (from Blot et al 2006)	Supercoiling sensitive genes under H-NS knockout (from Blot et al 2006)
Genes with specific annotations				Target genes of a regulator (according to external databases)							
Putative horizontally transferred genes (from HGT-DB)	Genes related to small RNAs (from Regulon DB)	Genes whose protein products have transmembrane domains (from Ensembl Bacteria)	Genes with paralogs (from Ensembl Bacteria)	Genes regulated by small RNAs (from Regulon DB)	Genes regulated by the sigma factor 24 (from Regulon DB)	Genes regulated by the sigma factor 28 (from Regulon DB)	Genes regulated by the sigma factor 32 (from Regulon DB)	Genes regulated by the sigma factor 38 (from Regulon DB)	Genes regulated by the sigma factor 54 (from Regulon DB)	Genes regulated by the sigma factor 70 (from Regulon DB)	

Table I.1: **Gene lists used in the enrichment analysis, grouped by category.** *Lists were divided according to the types of biological data and the experimental techniques. See ref. [271] for details. The symbols are the ones used in Fig. 2 in the main text.*

GENE SET	Z-SCORE (p1 long)	Z-SCORE (p1 short)	Z-SCORE (p1 short)
Putative H-NS target genes from ChIP-seq experiments (Kahramanoglou et al 2010)	19,7	27,3	27,3
Putative H-NS target genes from ChIP-chip experiments (from Oshima et al 2006)	13,9	17,1	17,1
Putative H-NS target genes in early-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	10,5	15,5	15,7
Genes overlapping with tsEPODs (transcriptionally silent extended protein occupancy domains) (from Vora et al 2009)	9,5	13,5	13,6
Putative H-NS target genes in mid-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	9,2	13,4	13
Putative H-NS target genes in transition-to-stationary phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	7,9	11,4	11,5
FNR target genes in midlog phase in the presence of oxygen from ChIP-chip experiments (from Grainger et al 2007)	6,7	7,2	7,7
Putative H-NS target genes in stationary phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	6,6	10,8	10
Putative FNR target genes in midlog phase from ChIP-chip experiments (Grainger et al. 2007)	6,5	8,9	9
Putative FNR target genes in midlog phase in the presence of oxygen from ChIP-chip experiments (from Grainger et al 2007)	6,4	7,2	7,2
Putative horizontally transferred genes (from HGT DB)	5,3	4,1	4,1
Putative FIS target genes from ChIP-seq experiments (Kahramanoglou et al 2010)	3,8	7,1	7
H-NS knockout sensitive genes under negative-supercoiling (from Blot et al 2006)	3,7	6,6	6,7

Table I.2: **Groups of genes regarding HNS or horizontally transferred genes result over-represented.** *The table reports the gene sets that result over-represented (Z-score $> +5$ in at least one experiment).*

GENE SET	Z-SCORE (p1 long)	Z-SCORE (p1 short)	Z-SCORE (p1 short)
List of E. coli essential genes (from Gerdes et al 2003)	-5,2	-6,3	-6,2
List of E. coli essential genes (from Baba et al 2006)	-4,4	-4,5	-4,4
Putative RNA-polymerase target genes during rapid growth from ChIP-chip experiments (from Grainger et al 2005	-3,4	-2,5	-2,7
Supercoiling sensitive genes (from Blot et al 2006)	-2,6	-2	-1,8
Supercoiling sensitive genes under FIS knockout (from Blot et al 2006)	-2,4	-0,3	-0,2
Genes overlapping with heEPODs (highly expressed extended protein occupancy domains) (from Vora et al 2009)	-2,2	-0,4	-0,5

Table I.3: **Groups of essential genes result under-represented.** *The table reports the gene sets that result under-represented (Z score < -3 in at least one insertion experiment).*

GENE SET	Z-SCORE (p1 long)	Z-SCORE (p1 short)	Z-SCORE (p1 short)
Putative FIS target genes in mid-log phase from ChIP-chip experiments (from Grainger et al 2006)	3,8	3,2	3
H-NS knockout sensitive genes under positive-supercoiling(from Blot et al 2006)	3,7	4,1	4,9
Putative IHF target genes in mid-log phase from ChIP-chip experiments (from Grainger et al 2006)	3,6	2	2,1
Putative FIS target genes from ChIP-chip experiments (from Cho et al 2006)	3,5	3,1	2,2
Putative CRP target genes from ChIP-chip experiments (from Grainger et al 2005)	2,4	3,5	3,2

Table I.4: **Groups of Fis,IHF,CRP target genes result over-represented considering a lower limit** *The table reports the gene sets that result over-represented (Z score > 3 and < 5 in at least one insertion experiment).*

	P1 long RL	P1 long RH	P1 short 1	P1 short 2
P1 long	0.86	0.50	0.97	0.97
P1 long RL		0.60	0.83	0.84
P1 long RH			0.51	0.50
P1 short 1				0.99

Table I.5: **A correlation exists between Z scores from different strains** *Correlation between Z scores from different stains is represented by the Pearson correlation coefficient. P1 long is the parental colony, then we have the first round high expression (RH) and low-expression (RL) sorted populations. Different colors correspond to different correlation levels.*

POPULATION	GENE SET	Z-SCORE
RL	Putative H-NS target genes from ChIP-seq experiments (Kahramanoglou et al 2010)	8,4
	Putative H-NS target genes from ChIP chip experiments (from Oshima et al 2006)	5,8
	Putative H-NS target genes in early-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	3,8
	Putative horizontally transferred genes (from HGT-DB)	3,7
	Putative H-NS target genes in mid-exponential phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	3,4
	Putative H-NS target genes in transition-to-stationary phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	3,2
	Putative H-NS target genes in stationary phase from ChIP-seq experiments (from Kahramanoglou et al 2010)	2,4
	Genes overlapping with tsEPODs (transcriptionally silent extended protein occupancy domains) (from Vora et al 2009)	2,2
RH	Genes overlapping with heEPODs (highly expressed extended protein occupancy domains) (from Vora et al 2009)	5,4
	Genes overlapping with tsEPODs (transcriptionally silent extended protein occupancy domains) (from Vora et al 2009)	3,4

Table I.6: **Only the subpopulation made of low-expression clones is consistent with the result of the whole population.** *The data in this table refer to Z-scores for the first round high expression (RH) and low-expression (RL) FACS-sorted populations. The table reports the gene sets that result over-represented with Z-score* $> +3$.

Bibliography

1. Wang, J. D. & Levin, P. A. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology* **7,** 822–827 (2009).
2. Alon, U. Alon, U.: Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8, 450. *Nature reviews. Genetics* **8,** 450–61 (2007).
3. MacNeil, L. T. & Walhout, A. J. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research* **21,** 645–657 (2011).
4. Levin, P. A. & Grossman, A. D. Cell cycle: The bacterial approach to coordination. *Current Biology* **8,** R28–R31 (1998).
5. Levin, P. A. & Taheri-Araghi, S. One is Nothing without the Other: Theoretical and Empirical Analysis of Cell Growth and Cell Cycle Progression. *Journal of Molecular Biology* **431,** 2061–2067 (2019).
6. Wolf, D. M. & Arkin, A. P. Motifs, modules and games in bacteria. *Current Opinion in Microbiology* **6,** 125–134 (2003).
7. Uriu, K. Genetic oscillators in development. *Development, Growth & Differentiation* **58,** 16–30 (2016).
8. Micali, G., Grilli, J., Osella, M. & Lagomarsino, M. C. Concurrent processes set E. coli cell division. *Science Advances* **4,** eaau3324 (2018).
9. Micali, G., Grilli, J., Marchi, J., Osella, M. & Cosentino Lagomarsino, M. Dissecting the Control Mechanisms for DNA Replication and Cell Division in E. coli. *Cell Reports* **25,** 761–771.e4 (2018).
10. Si, F. *et al.* Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria. *Current Biology* **29,** 1760–1770.e7 (2019).
11. Witz, G., van Nimwegen, E. & Julou, T. Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism. *eLife* **8** (2019).

12. Tiruvadi-Krishnan, S. *et al.* Replication-related control over cell division in Escherichia coli is growth-rate dependent (2021).

13. Harris, L. K. & Theriot, J. A. Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size. *Cell* **165,** 1479–1492 (2016).

14. Olliver, A., Saggioro, C., Herrick, J. & Sclavi, B. DnaA-ATP acts as a molecular switch to control levels of ribonucleotide reductase expression in Escherichia coli. *Molecular Microbiology* **76,** 1555–1571 (2010).

15. Hansen, F. G. & Atlung, T. The DnaA Tale. *Frontiers in Microbiology* **9,** 319 (2018).

16. Katayama, T., Kasho, K. & Kawakami, H. The DnaA Cycle in Escherichia coli: Activation, Function and Inactivation of the Initiator Protein. *Frontiers in Microbiology* **8** (2017).

17. Messer, W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews* **26,** 355–374 (2002).

18. Zhang, Q., Zhang, Z. & Shi, H. Cell Size Is Coordinated with Cell Cycle by Regulating Initiator Protein DnaA in E. coli. *Biophysical Journal* **119,** 2537–2557 (2020).

19. Grant, M. A. *et al.* DnaA and the timing of chromosome replication in Escherichia coli as a function of growth rate. *BMC Systems Biology* **5,** 201 (2011).

20. Sompayrac, L. & Maaløe, O. Autorepressor model for control of DNA replication. *Nature: New biology* (1973).

21. Donachie, W. D. & Blakely, G. W. Coupling the initiation of chromosome replication to cell size in Escherichia coli. *Current Opinion in Microbiology* **6,** 146–150 (2003).

22. Chiaramello, A. E. & Zyskind, J. W. Coupling of DNA replication to growth rate in Escherichia coli: a possible role for guanosine tetraphosphate. *Journal of Bacteriology* **172,** 2013–2019 (1990).

23. Kurokawa, K., Nishida, S., Emoto, A., Sekimizu, K. & Katayama, T. Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in Escherichia coli. *The EMBO Journal* **18,** 6642–6652 (1999).

24. Van Eijk, E., Wittekoek, B., Kuijper, E. J. & Smits, W. K. DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. *Journal of Antimicrobial Chemotherapy* **72,** 1275–1284 (2017).

25. Cadart, C. *et al.* Size control in mammalian cells involves modulation of both growth rate and cell cycle duration. *Nature Communications* **9,** 3275 (2018).

26. Haeusser, D. P. & Levin, P. A. The great divide: coordinating cell cycle events during bacterial growth and division. *Current Opinion in Microbiology. Cell Regulation* **11,** 94–99 (2008).

27. Yousuf, M. *et al.* Early fate of exogenous promoters in E. coli. *Nucleic Acids Research* **48,** 2348–2356 (2020).

28. Schaechter, M. A brief history of bacterial growth physiology. *Frontiers in Microbiology* **6,** 289 (2015).

29. Jun, S., Si, F., Pugatch, R. & Scott, M. Fundamental Principles in Bacterial Physiology - History, Recent progress, and the Future with Focus on Cell Size Control: A Review. *Reports on progress in physics. Physical Society (Great Britain)* **81,** 056601 (2018).

30. Kleckner, N. E., Chatzi, K., White, M. A., Fisher, J. K. & Stouf, M. Coordination of Growth, Chromosome Replication/Segregation, and Cell Division in E. coli. *Frontiers in Microbiology* **9,** 1469 (2018).

31. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297,** 1183–1186 (2002).

32. Walker, N., Nghe, P. & Tans, S. J. Generation and filtering of gene expression noise by the bacterial cell cycle. *BMC Biology* **14,** 11 (2016).

33. Wehrens, M., Büke, F., Nghe, P. & Tans, S. J. Stochasticity in cellular metabolism and growth: Approaches and consequences. *Current Opinion in Systems Biology* **8,** 131–136 (2018).

34. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467,** 167–173 (2010).

35. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* **514,** 376–379 (2014).

36. Pedraza, J. M. & Oudenaarden, A. v. Noise Propagation in Gene Networks. *Science* **307,** 1965–1969 (2005).

37. Rosenfeld, N. Gene Regulation at the Single-Cell Level. *Science* **307,** 1962–1965 (2005).

38. Yan, J. & Goldbeter, A. Robust synchronization of the cell cycle and the circadian clock through bidirectional coupling. *Journal of the Royal Society, Interface* **16,** 20190376 (2019).

39. Walker, N. E. A single-cell study on stochasticity growth and gene expression (2016).

40. Zopf, C. J., Quinn, K., Zeidman, J. & Maheshri, N. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. *PLoS Computational Biology* **9** (2013).

41. Schaechter, M., MaalØe, O. & Kjeldgaard, N. O. 1. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. *Microbiology* **19,** 592–606.

42. Helmstetter, C. E. A ten-year search for synchronous cells: obstacles, solutions, and practical applications. *Frontiers in Microbiology* **6** (2015).

43. Cooper, S. & Helmstetter, C. E. Chromosome replication and the division cycle of Escherichia coli B/r. *Journal of Molecular Biology* **31,** 519–540 (1968).

44. Helmstetter, C., Cooper, S., Pierucci, O. & Revelas, E. On the bacterial life sequence. *Cold Spring Harbor Symposia on Quantitative Biology* **33,** 809–822 (1968).

45. Cooper, S. Cell division and DNA replication following a shift to a richer medium. *Journal of Molecular Biology* **43,** 1–11 (1969).

46. Donachie, W. D. Relationship between Cell Size and Time of Initiation of DNA Replication. *Nature* **219,** 1077–1079 (1968).

47. Herrick, J., Kohiyama, M., Atlung, T. & Hansen, F. G. The initiation mess? *Molecular Microbiology* **19,** 659–666.

48. Wallden, M., Fange, D., Lundius, E. G., Baltekin, Ö. & Elf, J. The Synchronization of Replication and Division Cycles in Individual E. coli Cells. *Cell* **166,** 729–739 (2016).

49. Si, F. *et al.* Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Current biology: CB* **27,** 1278–1287 (2017).

50. Kennard, A. S. *et al.* Individuality and universality in the growth-division laws of single E. coli cells. *Physical Review E* **93,** 012408 (2016).

51. Zheng, H. *et al.* General quantitative relations linking cell growth and the cell cycle in Escherichia coli. *Nature Microbiology* (2020).

52. Dennis, P. P., Ehrenberg, M. & Bremer, H. Control of rRNA synthesis in Escherichia coli: a systems biology approach. *Microbiology and molecular biology reviews : MMBR* **68,** 639–668 (2004).

53. Allman, R., Schjerven, T. & Boye, E. Cell cycle parameters of Escherichia coli K-12. *Journal of Bacteriology* **173,** 7970–7974 (1991).

54. Michelsen, O., Teixeira de Mattos, M. J., Jensen, P. R. & Hansen, F. G. Precise determinations of C and D periods by flow cytometry in Escherichia coli K-12 and B/r. *Microbiology (Reading, England)* **149,** 1001–1010 (Pt 4 2003).

55. Hill, N. S., Kadoya, R., Chattoraj, D. K. & Levin, P. A. Cell size and the initiation of DNA replication in bacteria. *PLoS genetics* **8,** e1002549 (2012).

56. Churchward, G., Estiva, E. & Bremer, H. Growth rate-dependent control of chromosome replication initiation in Escherichia coli. *Journal of Bacteriology* **145,** 1232–1238 (1981).

57. Wold, S., Skarstad, K., Steen, H. B., Stokke, T. & Boye, E. The initiation mass for DNA replication in Escherichia coli K-12 is dependent on growth rate. *The EMBO Journal* **13,** 2097–2102 (1994).

58. Treut, G. L., Si, F., Li, D. & Jun, S. *Quantitative examination of five stochastic cell-cycle and cell-size control models for Escherichia coli and Bacillus subtilis* preprint (Systems Biology, 2021).

59. Ho, P.-Y. & Amir, A. Simultaneous regulation of cell size and chromosome replication in bacteria. *Frontiers in Microbiology* **6** (2015).

60. Donachie, W. D. Co-ordinate regulation of the Escherichia coli cell cycle or The cloud of unknowing. *Molecular Microbiology* **40,** 779–785 (2001).

61. Ghusinga, K. R., Vargas-Garcia, C. A. & Singh, A. A mechanistic stochastic framework for regulating bacterial cell division. *Scientific Reports* **6,** 30229 (2016).

62. Colin, A., Micali, G., Faure, L., Lagomarsino, M. & van Teeffelen, S. Two different cell-cycle processes determine the timing of cell division in Escherichia coli. *eLife* **10** (2021).

63. Osella, M., Tans, S. J. & Cosentino Lagomarsino, M. Step by Step, Cell by Cell: Quantification of the Bacterial Cell Cycle. *Trends in Microbiology* **25,** 250–256 (2017).

64. Bernander, R. & Nordström, K. Chromosome replication does not trigger cell division in E. coli. *Cell* **60,** 365–374 (1990).

65. Ojkic, N., Serbanescu, D. & Banerjee, S. Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. *eLife* **8** (eds Goldstein, R. E., Barkai, N. & Wolgemuth, C. W.) e47033 (2019).

66. Campos, M. *et al.* A Constant Size Extension Drives Bacterial Cell Size Homeostasis. *Cell* **159,** 1433–1446 (2014).

67. Taheri-Araghi, S. *et al.* Cell-Size Control and Homeostasis in Bacteria. *Current Biology* **25,** 385–391 (2015).

68. Amir, A. Cell Size Regulation in Bacteria. *Physical Review Letters* **112,** 208102 (2014).

69. Serbanescu, D., Ojkic, N. & Banerjee, S. Nutrient-Dependent Trade-Offs between Ribosomes and Division Protein Synthesis Control Bacterial Cell Size and Growth. *Cell Reports* **32,** 108183 (2020).

70. Le Treut, G., Si, F., Li, D. & Jun, S. *Comment on 'Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism'* preprint (Systems Biology, 2020).

71. Witz, G., Julou, T. & Nimwegen, E. v. Response to comment on 'Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism'. *bioRxiv,* 2020.08.04.227694 (2020).

72. Zheng, H. *et al.* Interrogating the Escherichia coli cell cycle by cell dimension perturbations. *Proceedings of the National Academy of Sciences* **113,** 15000–15005 (2016).

73. Kar, P., Tiruvadi-Krishnan, S., Männik, J., Männik, J. & Amir, A. *To bin or not to bin: analyzing single-cell growth data* preprint (Microbiology, 2021).

74. Willis, L. & Huang, K. C. Sizing up the bacterial cell cycle. *Nature Reviews. Microbiology* **15,** 606–620 (2017).

75. Simpson, E. H. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **13,** 238–241 (1951).

76. Stein, W. E. & Dattero, R. Sampling Bias and the Inspection Paradox. *Mathematics Magazine* **58,** 96–99 (1985).

77. Panlilio, M. *et al. Threshold accumulation of a constitutive protein explains E. coli cell division behavior in nutrient upshifts* preprint (Cell Biology, 2020).

78. Panlilio, M. *et al.* Threshold accumulation of a constitutive protein explains E. coli cell-division behavior in nutrient upshifts. *Proceedings of the National Academy of Sciences* **118** (2021).

79. Cadart, C., Venkova, L., Recho, P., Lagomarsino, M. C. & Piel, M. The physics of cell-size regulation across timescales. *Nature Physics* **15,** 993–1004 (2019).

80. Barber, F., Ho, P.-Y., Murray, A. W. & Amir, A. Details Matter: noise and model structure set the relationship between cell size and cell cycle timing. *Front. Cell Dev. Biol.* **5,** 92 (2017).

81. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **330,** 1099–1102 (2010).

82. Erickson, D. W. *et al.* A global resource allocation strategy governs growth transition kinetics of Escherichia coli. *Nature* **551,** 119–123 (2017).

83. Bertaux, F., von Kügelgen, J., Marguerat, S. & Shahrezaei, V. A bacterial size law revealed by a coarse-grained model of cell physiology. *PLoS Computational Biology* **16** (2020).

84. Pandey, P. P., Singh, H. & Jain, S. Exponential trajectories, cell size fluctuations and the adder property in bacteria follow from simple chemical dynamics and division control. *bioRxiv,* 487504 (2018).

85. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* 320 pp. (Chapman and Hall/CRC, New York, 2006).

86. Snyder, L., Peters, J., Champness, W. & Henkin, T. *Molecular Genetics of Bacteria* (2012).

87. Browning, D. & Busby, S. The regulation of bacterial transcription initiation. *Nature reviews. Microbiology* **2,** 57–65 (2004).

88. Martınez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology* **6,** 482–489 (2003).

89. Keseler, I. M. *et al.* EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Research* **33,** D334–D337 (Database Issue 2005).

90. Gama-Castro, S. *et al.* RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Research* **44,** D133–143 (D1 2016).

91. Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science (New York, N.Y.)* **277,** 1453–1462 (1997).

92. Pérez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Research* **28,** 1838–1847 (2000).

93. Freddolino, P. L., Amemiya, H. M., Goss, T. J. & Tavazoie, S. Dynamic landscape of protein occupancy across the Escherichia coli chromosome. *PLOS Biology* **19,** e3001306 (2021).

94. Larsen, S. J., Röttger, R., Schmidt, H. H. H. W. & Baumbach, J. E. coli gene regulatory networks are inconsistent with gene expression data. *Nucleic Acids Research* **47,** 85–92 (2019).

95. Co, A. D., Lagomarsino, M. C., Caselle, M. & Osella, M. Stochastic timing in gene expression for simple regulatory strategies. *Nucleic Acids Research* **45,** 1069–1078 (2017).

96. Hermsen, R., Ursem, B. & ten Wolde, P. R. Combinatorial Gene Regulation Using Auto-Regulation. *PLoS Computational Biology* **6** (2010).

97. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology* **323,** 785–793 (2002).

98. Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439,** 608–611 (2006).

99. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405,** 590–593 (2000).

100. *Intrinsic noise in gene regulatory networks | PNAS* https://www.pnas.org/content/98/15/8614 (2021).

101. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456,** 516–519 (2008).

102. Rosenfeld, N. & Alon, U. Response delays and the structure of transcription networks. *Journal of Molecular Biology* **329,** 645–654 (2003).

103. Maeda, Y. T. & Sano, M. Regulatory Dynamics of Synthetic Gene Networks with Positive Feedback. *Journal of Molecular Biology* **359,** 1107–1124 (2006).

104. Hermsen, R., Erickson, D. W. & Hwa, T. Speed, Sensitivity, and Bistability in Auto-activating Signaling Circuits. *PLOS Computational Biology* **7,** e1002265 (2011).

105. Novak, B. & Tyson, J. J. Design Principles of Biochemical Oscillators. *Nature reviews. Molecular cell biology* **9,** 981–991 (2008).

106. Gérard, C., Gonze, D. & Goldbeter, A. Effect of positive feedback loops on the robustness of oscillations in the network of cyclin-dependent kinases driving the mammalian cell cycle. *The FEBS Journal* **279,** 3411–3431 (2012).

107. Tsai, T. Y.-C. *et al.* Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops. *Science* **321,** 126–129 (2008).

108. Gao, R. & Stock, A. M. Overcoming the Cost of Positive Autoregulation by Accelerating the Response with a Coupled Negative Feedback. *Cell Reports* **24,** 3061–3071.e6 (2018).

109. Silander, O. K. *et al.* A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. *PLOS Genetics* **8,** e1002443 (2012).

110. Munsky, B., Neuert, G. & Oudenaarden, A. v. Using Gene Expression Noise to Understand Gene Regulation. *Science* **336,** 183–187 (2012).

111. Bury-Moné, S. & Sclavi, B. Stochasticity of gene expression as a motor of epigenetics in bacteria: from individual to collective behaviors. *Research in Microbiology* **168,** 503–514 (2017).

112. Saggioro, C., Olliver, A. & Sclavi, B. Temperature-dependence of the DnaA–DNA interaction and its effect on the autoregulation of dnaA expression. *Biochemical Journal* **449,** 333–341 (2013).

113. *Origin Remodeling and Opening in Bacteria Rely on Distinct Assembly States of the DnaA Initiator* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2934688/ (2021).

114. Messer, W. & Weigel, C. DnaA initiator—also a transcription factor. *Molecular Microbiology* **24,** 1–6 (1997).

115. Zyskind, J. W. & Smith, D. W. The bacterial origin of replication, oriC. *Cell* **46,** 489–490 (1986).

116. Ozaki, S. & Katayama, T. DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* **62,** 71–82 (2009).

117. Sekimizu, K., Bramhill, D. & Kornberg, A. ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. *Cell* **50,** 259–265 (1987).

118. Speck, C., Weigel, C. & Messer, W. ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *The EMBO journal* **18,** 6169–6176 (1999).

119. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157,** 624–635 (2014).

120. Katayama, T. Feedback controls restrain the initiation of Escherichia coli chromosomal replication. *Molecular Microbiology* **41,** 9–17 (2001).

121. Charbon, G. *et al.* Energy Starvation Induces a Cell Cycle Arrest in Escherichia coli by Triggering Degradation of the DnaA Initiator Protein. *Frontiers in Molecular Biosciences* **0** (2021).

122. Zhang, Q., Bassetti, F., Gherardi, M. & Lagomarsino, M. C. Cell-to-cell variability and robustness in S-phase duration from genome replication kinetics. *Nucleic Acids Research* **45,** 8190–8198 (2017).

123. Flåtten, I., Fossum-Raunehaug, S., Taipale, R., Martinsen, S. & Skarstad, K. The DnaA Protein Is Not the Limiting Factor for Initiation of Replication in Escherichia coli. *PLoS Genetics* **11** (2015).

124. Kato, J.-i. & Katayama, T. Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli. *The EMBO Journal* **20,** 4253–4262 (2001).

125. Kasho, K. & Katayama, T. DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 936–941 (2013).

126. Fujimitsu, K., Senriuchi, T. & Katayama, T. Specific genomic sequences of E. coli promote replicational initiation by directly reactivating ADP-DnaA. *Genes & Development* **23,** 1221–1233 (2009).

127. Skarstad, K. & Katayama, T. Regulating DNA replication in bacteria. *Cold Spring Harbor Perspectives in Biology* **5,** a012922 (2013).

128. Aranovich, A., Gdalevsky, G. Y., Cohen-Luria, R., Fishov, I. & Parola, A. H. Membrane-catalyzed Nucleotide Exchange on DnaA: EFFECT OF SURFACE MOLECULAR CROWDING*. *Journal of Biological Chemistry* **281,** 12526–12534 (2006).

129. Aranovich, A. *et al.* N-terminal-mediated oligomerization of DnaA drives the occupancy-dependent rejuvenation of the protein on the membrane. *Bioscience Reports* **35,** e00250 (2015).

130. Kitagawa, R., Ozaki, T., Moriya, S. & Ogawa, T. Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein. *Genes & development* **12,** 3032–43 (1998).

131. Atlung, T. & Hansen, F. G. Low-Temperature-Induced DnaA Protein Synthesis Does Not Change Initiation Mass in Escherichia coli K-12. *Journal of Bacteriology* **181,** 5557–5562 (1999).

132. Dewachter, L., Verstraeten, N., Fauvart, M. & Michiels, J. An integrative view of cell cycle control in Escherichia coli. *FEMS Microbiology Reviews* **42,** 116–136 (2018).

133. Quiñones, A., Wandt, G., Kleinstäuber, S. & Messer, W. DnaA protein stimulates polA gene expression in Escherichia coli. *Molecular Microbiology* **23,** 1193–1202 (1997).

134. Kücherer, C., Lother, H., Kölling, R., Schauzu, M. A. & Messer, W. Regulation of transcription of the chromosomal dnaA gene of Escherichia coli. *Molecular & general genetics: MGG* **205,** 115–121 (1986).

135. Braun, R. E., O'Day, K. & Wright, A. Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell* **40,** 159–169 (1985).

136. Wang, Q. & Kaguni, J. M. Transcriptional repression of the dnaA gene of Escherichia coli by dnaA protein. *Molecular and General Genetics MGG* **209,** 518–525 (1987).

137. Wang, Q. P. & Kaguni, J. M. dnaA protein regulates transcriptions of the rpoH gene of Escherichia coli. *The Journal of Biological Chemistry* **264,** 7338–7344 (1989).

138. Goranov, A. I., Katz, L., Breier, A. M., Burge, C. B. & Grossman, A. D. A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 12932–12937 (2005).

139. Tuggle, C. K. & Fuchs, J. A. Regulation of the operon encoding ribonucleotide reductase in Escherichia coli: evidence for both positive and negative control. *The EMBO journal* **5,** 1077–1085 (1986).

140. Truglio, J. J., Croteau, D. L., Van Houten, B. & Kisker, C. Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical Reviews* **106,** 233–252 (2006).

141. Wurihan *et al.* DnaA and LexA Proteins Regulate Transcription of the uvrB Gene in Escherichia coli: The Role of DnaA in the Control of the SOS Regulon. *Frontiers in Microbiology* **9** (2018).

142. Mizushima, T., Tomura, A., Shinpuku, T., Miki, T. & Sekimizu, K. Loss of flagellation in dnaA mutants of Escherichia coli. *Journal of Bacteriology* **176,** 5544–5546 (1994).

143. Sun, L. & Fuchs, J. A. Escherichia coli ribonucleotide reductase expression is cell cycle regulated. *Molecular Biology of the Cell* **3,** 1095–1105 (1992).

144. Ogawa, T. & Okazaki, T. Cell cycle-dependent transcription from the gid and mioC promoters of Escherichia coli. *Journal of Bacteriology* **176,** 1609–1615 (1994).

145. Van den Berg, E. A. *et al.* Analysis of regulatory sequences upstream of the E. coli uvrB gene; involvement of the DnaA protein. *Nucleic Acids Research* **13,** 1829–1840 (1985).

146. Gon, S. *et al.* A novel regulatory mechanism couples deoxyribonucleotide synthesis and DNA replication in Escherichia coli. *The EMBO Journal* **25,** 1137–1147 (2006).

147. Messer, W. *et al.* Functional domains of DnaA proteins. *Biochimie* **81,** 819–825 (1999).

148. Kaguni, J. M. DnaA: controlling the initiation of bacterial DNA replication and more. *Annual Review of Microbiology* **60,** 351–375 (2006).

149. Shimizu, M. *et al.* Near-atomic structural model for bacterial DNA replication initiation complex and its functional insights. *Proceedings of the National Academy of Sciences* **113,** 201609649 (2016).

150. Hansen, E. B., Hansen, F. G. & von Meyenburg, K. The nucleotide sequence of the dnaA gene and the first part of the dnaN gene of Escherichia coli K-12. *Nucleic Acids Research* **10,** 7373–7385 (1982).

151. Travers, A. & Muskhelishvili, G. DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nature Reviews. Microbiology* **3,** 157–169 (2005).

152. Atlung, T., Clausen, E. S. & Hansen, F. G. Autoregulation of the dnaA gene of Escherichia coli K12. *Molecular & general genetics: MGG* **200,** 442–450 (1985).

153. Pemberton, I. K., Muskhelishvili, G., Travers, A. A. & Buckle, M. The G+C-rich discriminator region of the tyrT promoter antagonises the formation of stable preinitiation complexes. *Journal of Molecular Biology* **299,** 859–864 (2000).

154. Hansen, F. G. & Rasmussen, K. V. Regulation of the dnaA product in Escherichia coli. *Molecular & general genetics: MGG* **155,** 219–225 (1977).

155. Hansen, F. G., Christensen, B. B. & Atlung, T. Sequence characteristics required for cooperative binding and efficient in vivo titration of the replication initiator protein DnaA in E. coli. *Journal of Molecular Biology* **367,** 942–952 (2007).

156. Grimwade, J. E. & Leonard, A. C. Blocking, Bending, and Binding: Regulation of Initiation of Chromosome Replication During the Escherichia coli Cell Cycle by Transcriptional Modulators That Interact With Origin DNA. *Frontiers in Microbiology* **12,** 732270 (2021).

157. Flåtten, I., Morigen & Skarstad, K. DnaA protein interacts with RNA polymerase and partially protects it from the effect of rifampicin. *Molecular Microbiology* **71,** 1018–1030 (2009).

158. Wilkinson, T. G. *et al.* The Synchrony Phenotype Persists after Elimination of Multiple GATC Sites from the dnaA Promoter of Escherichia coli. *Journal of Bacteriology* **188,** 4573–4576 (2006).

159. Sánchez-Romero, M. A. *et al.* Dynamic Distribution of SeqA Protein across the Chromosome of Escherichia coli K-12. *mBio* **1,** e00012–10.

160. Waldminghaus, T., Weigel, C. & Skarstad, K. Replication fork movement and methylation govern SeqA binding to the Escherichia coli chromosome. *Nucleic Acids Research* **40,** 5465–5476 (2012).

161. Campbell, J. L. & Kleckner, N. E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* **62,** 967–979 (1990).

162. Lu, M., Campbell, J. L., Boye, E. & Kleckner, N. SeqA: a negative modulator of replication initiation in E. coli. *Cell* **77,** 413–426 (1994).

163. Von Freiesleben, U., Rasmussen, K. V. & Schaechter, M. SeqA limits DnaA activity in replication from oriC in Escherichia coli. *Molecular Microbiology* **14,** 763–772 (1994).

164. Aloui, A., May, A. E., Sahbani, S. K. & Landoulsi, A. *Roles of Methylation and Sequestration in the Mechanisms of DNA Replication in some Members of the Enterobacteriaceae Family* (IntechOpen, 2013).

165. Odsbu, I., Klungsøyr, H. K., Fossum, S. & Skarstad, K. Specific N-terminal interactions of the Escherichia coli SeqA protein are required to form multimers that restrain negative supercoils and form foci. *Genes to Cells* **10,** 1039–1049 (2005).

166. Kedar, G. C. *et al.* Role of DNA methylation at GATC sites in the dnaA promoter, dnaAp2. *Journal of Molecular Microbiology and Biotechnology* **2,** 301–310 (2000).

167. Sutera, V. A. & Lovett, S. T. The role of replication initiation control in promoting survival of replication fork damage. *Molecular Microbiology* **60,** 229–239 (2006).

168. Godin, M. *et al.* Using buoyant mass to measure the growth of single cells. *Nature Methods* **7,** 387–390 (2010).

169. Duncombe, T. A., Tentori, A. M. & Herr, A. E. Microfluidics: reframing biological enquiry. *Nature reviews. Molecular cell biology* **16,** 554–567 (2015).

170. Sclavi, B., Iuliani, I. & Team, T. E. How to study bacteria by microfluidics. *Elveflow* (2021).

171. Long, Z. *et al.* Microfluidic chemostat for measuring single cell dynamics in bacteria. *Lab on a Chip* **13,** 947 (2013).

172. Long, Z. *et al.* Measuring bacterial adaptation dynamics at the single-cell level using a microfluidic chemostat and time-lapse fluorescence microscopy. *The Analyst* **139,** 5254–5262 (2014).

173. Wang, P. *et al.* Robust Growth of Escherichia coli. *Current Biology* **20,** 1099–1103 (2010).

174. Taheri-Araghi, S., Brown, S. D., Sauls, J. T., McIntosh, D. B. & Jun, S. Single-Cell Physiology. *Annual Review of Biophysics* **44,** 123–142 (2015).

175. Potvin-Trottier, L., Lord, N. D., Vinnicombe, G. & Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **538,** 514–517 (2016).

176. Varsano, G., Wang, Y. & Wu, M. Probing Mammalian Cell Size Homeostasis by Channel-Assisted Cell Reshaping. *Cell Reports* **20,** 397–410 (2017).

177. Ryley, J. & Pereira-Smith, O. M. Microfluidics device for single cell gene expression analysis in Saccharomyces cerevisiae. *Yeast (Chichester, England)* **23,** 1065–1073 (2006).

178. Yang, D., Jennings, A. D., Borrego, E., Retterer, S. T. & Männik, J. Analysis of Factors Limiting Bacterial Growth in PDMS Mother Machine Devices. *Frontiers in Microbiology* **9** (2018).

179. Grenier, F., Matteau, D., Baby, V. & Rodrigue, S. Complete Genome Sequence of Escherichia coli BW25113. *Genome Announcements* **2** (2014).

180. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173,** 33–38 (1996).

181. Lissemore, J., Jankowski, J., Thomas, C., Mascotti, D. & deHaseth, P. Green Fluorescent Protein as a Quantitative Reporter of Relative Promoter Activity in *E. coli*. *BioTechniques* **28,** 82–89 (2000).

182. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139,** 1366–1375 (2009).

183. Stylianidou, S., Brennan, C., Nissen, S. B., Kuwada, N. & Wiggins, P. SuperSegger: Robust image segmentation, analysis and lineage tracking of bacterial cells. *Molecular Microbiology* **102** (2016).

184. Thomas, P. Intrinsic and extrinsic noise of gene expression in lineage trees. *Scientific Reports* **9,** 474 (2019).

185. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 12795–12800 (2002).

186. Romano, O. M. & Cosentino Lagomarsino, M. Single rod-shaped cell fluctuations from stochastic surface and volume growth rates. *Physical Review E* **101,** 042403 (2020).

187. Espéli, O. *et al.* A MatP–divisome interaction coordinates chromosome segregation with cell division in E. coli. *The EMBO Journal* **31,** 3198–3211 (2012).

188. Crozat, E. *et al.* Post-replicative pairing of sister ter regions in Escherichia coli involves multiple activities of MatP. *Nature Communications* **11,** 3796 (2020).

189. Mitchison, J. M. Single cell studies of the cell cycle and some models. *Theoretical Biology & Medical Modelling* **2,** 4 (2005).

190. Magnusson, L. U., Farewell, A. & Nyström, T. ppGpp: a global regulator in Escherichia coli. *Trends in Microbiology* **13,** 236–242 (2005).

191. Henrici, A. T. *Morphologic Variation and the Rate of Growth of Bacteria* 224 pp. (C.C. Thomas, 1928).

192. Soifer, I., Robert, L. & Amir, A. Single-Cell Analysis of Growth in Budding Yeast and Bacteria Reveals a Common Size Regulation Strategy. *Current biology: CB* **26,** 356–361 (2016).

193. Cermak, N. *et al.* High-throughput measurement of single-cell growth rates using serial microfluidic mass sensor arrays. *Nature Biotechnology* **34,** 1052–1059 (2016).

194. Iyer-Biswas, S. *et al.* Scaling laws governing stochastic growth and division of single bacterial cells. *Proceedings of the National Academy of Sciences* **111,** 15912–15917 (2014).

195. Kubitschek, H. E. Bilinear cell growth of Escherichia coli. *Journal of Bacteriology* **148,** 730–733 (1981).

196. Reshes, G., Vanounou, S., Fishov, I. & Feingold, M. Cell Shape Dynamics in Escherichia coli. *Biophysical Journal* **94,** 251–264 (2008).

197. Cooper, S. Distinguishing between linear and exponential cell growth during the division cycle: Single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical Biology & Medical Modelling* **3,** 10 (2006).

198. Wood, E. & Nurse, P. Pom1 and cell size homeostasis in fission yeast. *Cell Cycle* **12,** 3417–3425 (2013).

199. Facchetti, G., Chang, F. & Howard, M. Controlling cell size through sizer mechanisms. *Current Opinion in Systems Biology* **5,** 86–92 (2017).

200. Facchetti, G., Knapp, B., Chang, F. & Howard, M. Reassessment of the Basis of Cell Size Control Based on Analysis of Cell-to-Cell Variability. *Biophysical Journal* **117,** 1728–1738 (2019).

201. Typas, A., Banzhaf, M., Gross, C. A. & Vollmer, W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nature reviews. Microbiology* **10,** 123–136 (2011).

202. Nordholt, N., van Heerden, J. H. & Bruggeman, F. J. Biphasic Cell-Size and Growth-Rate Homeostasis by Single Bacillus subtilis Cells. *Current Biology* **30,** 2238–2247.e5 (2020).

203. Banerjee, S. *et al.* Biphasic growth dynamics control cell division in Caulobacter crescentus. *Nature Microbiology* **2,** 1–6 (2017).

204. Knapp, B. D. *et al.* Decoupling of Rates of Protein Synthesis from Cell Expansion Leads to Supergrowth. *Cell Systems* **9,** 434–445.e6 (2019).

205. Zhang, Q. *et al.* A Decrease in Transcription Capacity Limits Growth Rate upon Translation Inhibition. *mSystems* **5** (2020).

206. Rao, L. *et al.* Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *Journal of Molecular Biology* **235,** 1421–1435 (1994).

207. Gaal, T. *et al.* Colocalization of distant chromosomal loci in space in E. coli: a bacterial nucleolus. *Genes & Development* **30,** 2272–2285 (2016).

208. Scholz, S. A. *et al.* High-Resolution Mapping of the Escherichia coli Chromosome Reveals Positions of High and Low Transcription. *Cell Systems* **8,** 212–225.e9 (2019).

209. Dobzhansky, T. Position Effects on Genes. *Biological Reviews* **11,** 364–384 (1936).

210. Wilson, C., Bellen, H. J. & Gehring, W. J. Position Effects on Eukaryotic Gene Expression. *Annual Review of Cell Biology* **6,** 679–714 (1990).

211. Girton, J. R. & Johansen, K. M. in *Advances in Genetics* 1–43 (Academic Press, 2008).

212. Collas, P., Lund, E. G. & Oldenburg, A. R. Closing the (nuclear) envelope on the genome: How nuclear lamins interact with promoters and modulate gene expression. *BioEssays* **36,** 75–83 (2014).

213. Bryant, J. A., Sellars, L. E., Busby, S. J. W. & Lee, D. J. Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Research* **42,** 11383–11392 (2014).

214. Brambilla, E. & Sclavi, B. Gene regulation by H-NS as a function of growth conditions depends on chromosomal position in Escherichia coli. *G3 (Bethesda, Md.)* **5,** 605–614 (2015).

215. Kahramanoglou, C. *et al.* Direct and indirect effects of H-NS and Fis on global gene expression control in Escherichia coli. *Nucleic Acids Research* **39,** 2073–2091 (2011).

216. Conway, T. *et al.* Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. *mBio* **5,** e01442–14.

217. Sobetzko, P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. *Nucleic Acids Research* **44,** 1514–1524 (2016).

218. Blot, N., Mavathur, R., Geertz, M., Travers, A. & Muskhelishvili, G. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO reports* **7,** 710–715 (2006).

219. Zhang, X. & Bremer, H. Control of the Escherichia coli rrnB P1 Promoter Strength by ppGpp. *Journal of Biological Chemistry* **270,** 11181–11189 (1995).

220. Ross, W., Thompson, J. F., Newlands, J. T. & Gourse, R. L. E.coli Fis protein activates ribosomal RNA transcription in vitro and in vivo. *The EMBO journal* **9,** 3733–3742 (1990).

221. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* **85,** 21–38 (2012).

222. Philips, R. M. R. *» How many ribosomes are in a cell?* http://book.bionumbers.org/how-many-ribosomes-are-in-a-cell/ (2021).

223. Ortiz, J. *et al.* Structure of hibernating ribosomes studied by cryoelectron tomography in vitro and in situ. *The Journal of cell biology* **190,** 613–21 (2010).

224. Sobetzko, P., Travers, A. & Muskhelishvili, G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proceedings of the National Academy of Sciences of the United States of America* **109,** E42–50 (2012).

225. Muskhelishvili, G., Forquet, R., Reverchon, S., Meyer, S. & Nasser, W. Coherent Domains of Transcription Coordinate Gene Expression During Bacterial Growth and Adaptation. *Microorganisms* **7,** 694 (2019).

226. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2,** 2006.0008 (2006).

227. Froelich, J. M., Phuong, T. K. & Zyskind, J. W. Fis binding in the dnaA operon promoter region. *Journal of Bacteriology* **178,** 6006–6012 (1996).

228. Brenner, N. *et al.* Single-cell protein dynamics reproduce universal fluctuations in cell populations. *Eur. Phys. J. E,* 9 (2015).

229. Pikovsky, A., Rosenblum, M. & Kurths, J. Synchronization: A universal concept in nonlinear sciences, 433.

230. Takens, F. *Detecting strange attractors in turbulence* in *Dynamical Systems and Turbulence, Warwick 1980* (eds Rand, D. & Young, L.-S.) (Springer, Berlin, Heidelberg, 1981), 366–381.

231. Sugihara, G. *et al.* Detecting Causality in Complex Ecosystems. *Science* **338,** 496–500 (2012).

232. Huygens. *C. Letters to de Sluse* 1665.

233. Droin, C., Paquet, E. R. & Naef, F. Low-dimensional dynamics of two coupled biological oscillators. *Nature Physics* **15,** 1086–1094 (2019).

234. Feillet, C. *et al.* Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. *Proceedings of the National Academy of Sciences* **111,** 9828–9833 (2014).

235. Bieler, J. *et al.* Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular Systems Biology* **10,** 739 (2014).

236. Goldstein, R. E., Polin, M. & Tuval, I. Noise and Synchronization in Pairs of Beating Eukaryotic Flagella. *Physical Review Letters* **103,** 168103 (2009).

237. Kotar, J., Leoni, M., Bassetti, B., Lagomarsino, M. C. & Cicuta, P. Hydrodynamic synchronization of colloidal oscillators. *Proceedings of the National Academy of Sciences* **107,** 7669–7673 (2010).

238. Rosenblum, M. G. & Pikovsky, A. S. Detecting direction of coupling in interacting oscillators. *Physical Review E* **64.**

239. Løbner-Olesen, A., Skarstad, K., Hansen, F. G., von Meyenburg, K. & Boye, E. The DnaA protein determines the initiation mass of Escherichia coli K-12. *Cell* **57,** 881–889 (1989).

240. Adiciptaningrum, A., Osella, M., Moolman, M. C., Cosentino Lagomarsino, M. & Tans, S. J. Stochasticity and homeostasis in the E. coli replication and division cycle. *Scientific Reports* **5,** 18261 (2016).

241. Hartwell, L. H. & Weinert, T. A. Checkpoints: controls that ensure the order of cell cycle events. *Science (New York, N.Y.)* **246,** 629–634 (1989).

242. Murray, A. W. Creative blocks: cell-cycle checkpoints and feedback controls. *Nature* **359,** 599–601 (1992).

243. Touchon, M., Bobay, L.-M. & Rocha, E. P. C. The chromosomal accommodation and domestication of mobile genetic elements. *Current opinion in microbiology* **22,** 22–29 (2014).

244. Lawrence, J. G. & Ochman, H. Molecular archaeology of the Escherichia coli genome. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 9413–9417 (1998).

245. Oliveira, P. H., Touchon, M., Cury, J. & Rocha, E. P. C. The chromosomal organization of horizontal gene transfer in bacteria. *Nature communications* **8,** 841 (2017).

246. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).

247. Lercher, M. J. & Pál, C. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Molecular biology and evolution* **25,** 559–567 (2008).

248. Park, C. & Zhang, J. High expression hampers horizontal gene transfer. *Genome biology and evolution* **4,** 523–532 (2012).

249. Lucchini, S. *et al.* H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathogens* **2,** e81 (2006).

250. Navarre, W. W., McClelland, M., Libby, S. J. & Fang, F. C. Silencing of xenogeneic DNA by H-NS—facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes & Development* **21,** 1456–1471 (2007).

251. Dorman, C. J. H-NS, the genome sentinel. *Nat Rev Microbiol* **5,** 157–161 (2007).

252. Singh, K., Milstein, J. N. & Navarre, W. W. Xenogeneic Silencing and Its Impact on Bacterial Genomes. *Annual review of microbiology* **70,** 199–213 (2016).

253. Kimura, S., Hubbard, T. P., Davis, B. M. & Waldor, M. K. The Nucleoid Binding Protein H-NS Biases Genome-Wide Transposon Insertion Landscapes. *mBio* **7** (2016).

254. Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer. *Science (New York, N.Y.)* **318,** 1449–1452 (2007).

255. Dame, R. T., Noom, M. C. & Wuite, G. J. L. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* **444,** 387–390 (2006).

256. Liu, Y., Chen, H., Kenney, L. J. & Yan, J. A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes & development* **24,** 339–344 (2010).

257. Srinivasan, R., Scolari, V. F., Cosentino Lagomarsino, M. & Seshasayee, A. S. N. The genome-scale interplay amongst xenogene silencing, stress response and chromosome architecture in Escherichia coli. *Nucleic acids research* **43,** 295–308 (2015).

258. Singh, S. S. *et al.* Widespread suppression of intragenic transcription initiation by H-NS. *Genes & development* **28,** 214–219 (2014).

259. Dorman, C. J. Nucleoid-associated proteins and bacterial physiology. *Adv Appl Microbiol* **67,** 47–64 (2009).

260. Espéli, O. & Boccard, F. Organization of the Escherichia coli chromosome into macrodomains and its possible functional implications. *J Struct Biol* **156,** 304–310 (2006).

261. Zarei, M., Sclavi, B. & Cosentino Lagomarsino, M. Gene silencing and large-scale domain structure of the E. coli genome. *Molecular BioSystems* **9,** 758–767 (2013).

262. Dame, R. T., Kalmykowa, O. J. & Grainger, D. C. Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria. *PLoS Genet* **7,** e1002123 (2011).

263. Dorman, C. J. Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription. *Biochem Soc Trans* **41,** 542–547 (2013).

264. Touchon, M. *et al.* Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS genetics* **5,** e1000344 (2009).

265. Junier, I., Martin, O. & Képès, F. Spatial and Topological Organization of DNA Chains Induced by Gene Co-localization. *PLoS Comput Biol* **6,** e1000678 (2010).

266. Dilthey, A. & Lercher, M. J. Horizontally transferred genes cluster spatially and metabolically. *Biology direct* **10,** 72 (2015).

267. Bobay, L.-M. & Ochman, H. The Evolution of Bacterial Genome Architecture. *Frontiers in genetics* **8,** 72 (2017).

268. Lang, B. *et al.* High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic acids research* **35,** 6330–6337 (2007).

269. Berger, M. *et al.* Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Rep* **11,** 59–64 (2010).

270. Lioy, V. S. *et al.* Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172,** 771–783.e18 (2018).

271. Scolari, V. F., Zarei, M., Osella, M. & Cosentino Lagomarsino, M. NuST: analysis of the interplay between nucleoid organization and gene expression. *Bioinformatics* **28,** 1643–1644 (2012).

272. Garcia-Vallve, S., Guzman, E., Montero, M. A. & Romeu, A. HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic acids research* **31,** 187–189 (2003).

273. Touchon, M. & Rocha, E. P. C. Causes of insertion sequences abundance in prokaryotic genomes. *Molecular biology and evolution* **24,** 969–981 (2007).

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

275. Browning, D. F., Grainger, D. C. & Busby, S. J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol* **13,** 773–780 (2010).

276. Grainger, D. C., Goldberg, M. D., Lee, D. J. & Busby, S. J. Selective repression by Fis and H-NS at the Escherichia coli dps promoter. *Mol Microbiol* **68,** 1366–77 (2008).

277. Green, B., Bouchier, C., Fairhead, C., Craig, N. L. & Cormack, B. P. Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mobile DNA* **3,** 3 (2012).

278. Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S. J. W. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 17693–17698 (2005).

279. Zaslaver, A. *et al.* Invariant distribution of promoter activities in Escherichia coli. *PLoS Comput Biol* **5,** e1000545 (2009).

280. Zhang, Q. *et al.* A decrease in transcription capacity limits growth rate upon translation inhibition. *bioRxiv* (2019).

281. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (New York, N.Y.)* **329,** 533–538 (2010).

282. Ouafa, Z.-A., Reverchon, S., Lautier, T., Muskhelishvili, G. & Nasser, W. The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the pel genes, the major virulence factors in the plant pathogen bacterium Dickeya dadantii. *Nucleic acids research* **40,** 4306–4319 (2012).

283. Lal, A. *et al.* Genome scale patterns of supercoiling in a bacterial chromosome. *Nature communications* **7,** 11055 (2016).

284. Jeong, K. S., Ahn, J. & Khodursky, A. B. Spatial patterns of transcriptional activity in the chromosome of Escherichia coli. *Genome biology* **5,** R86 (2004).

285. Klumpp, S. & Hwa, T. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 18159–18164 (2008).

286. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference–a crash course. *Trends in genetics : TIG* **21,** 339–345 (2005).

287. Wang, W., Li, G.-W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333,** 1445–1449 (2011).

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

289. Choubey, S., Kondev, J. & Sanchez, A. Distribution of Initiation Times Reveals Mechanisms of Transcriptional Regulation in Single Cells. *Biophysical Journal* **114,** 2072–2082 (2018).

290. Mann, A. Life's little oscillations. *Knowable Magazine | Annual Reviews* (2020).

291. Watts, G. Three cell cycle scientists win Nobel prize. *BMJ : British Medical Journal* **323,** 823 (2001).

292. World Health Organization. *2020 antibacterial agents in clinical and preclinical development: an overview and analysis* (World Health Organization, Geneva, 2021).

293. Bintu, L. *et al.* Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development* **15,** 125–135 (2005).

294. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development* **15,** 116–124 (2005).

295. Gedeon, T. & Bokes, P. Delayed Protein Synthesis Reduces the Correlation between mRNA and Protein Fluctuations. *Biophysical Journal* **103,** 377–385 (2012).

296. Osella, M. & Lagomarsino, M. C. Growth-rate-dependent dynamics of a bacterial genetic oscillator. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* **87,** 012726 (2013).

297. Marathe, R., Bierbaum, V., Gomez, D. & Klumpp, S. Deterministic and Stochastic Descriptions of Gene Expression Dynamics. *Journal of Statistical Physics* **148,** 608–627 (2012).

université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Title: The DnaA cell-cycle oscillator and its coordination with cell growth and division in *Escherichia coli*

Keywords: Single-cell biology, *E. coli*, DnaA, DNA replication, cell cycle, cell growth, transcriptional regulation, microfluidics, microscopy, time series analysis, coupled oscillators

Abstract: Despite over 50 years of study, key questions on the bacterial cell cycle remain unanswered. In particular, the debate is open concerning the regulation of DNA replication initiation in *E. coli* and its role in the dynamics of cell growth and division. A key player in these processes is the DnaA protein, which is involved the initiation of DNA replication. DnaA is commonly believed to be a cell cycle oscillator and a cell size sensor, but neither of these facts have been firmly established. DnaA activity depends on its nucleotide bound state, the ATP bound form being the active one for origin recognition and activation. DnaA is also a transcription factor, a highly connected node in the network of genes coding for proteins required for DNA replication and the repair of DNA damage. The differential regulation of gene expression by the different DnaA nucleotide bound forms, including the regulation of its own promoter, is believed to be central in its role of cell-cycle oscillator and regulator. Indeed, because of its double role as a transcription factor and an activator of the initiation of the DNA replication process, the DnaA protein can act as the regulatory link between the timing and level of gene expression and the different phases of the bacterial cell cycle.

This thesis addresses the problem of identifying the cell cycle oscillator related to DnaA activity, and relating it to the progression of the *E. coli* cell cycle, focusing in particular on cell-size sensing, individual cell growth rate, and cell division. One of the major challenges in this area has been to quantify the changes in DnaA-ATP activity *in vivo* in real time. This problem requires single-cell dynamic resolution, in order to relate DnaA levels and activity to key cell-cycle transitions. To address this problem, I have developed a set of reporters of gene expression using a gene for a fluorescent protein under control of a promoter that is regulated by DnaA-ATP, and deployed them in single-cell experiments coupling quantitative microscopy and microfluidics. To obtain robust and long-term single-cell tracking in steady growth conditions I have designed a dedicated experimental setup and data-analysis pipeline for studying the growth, size, and gene expression of *E. coli* in controlled environmental conditions.

A careful analysis of these single cell data as a function of the cell cycle shows that *E. coli* growth is biphasic and it follows the expression of a constitutive ori-proximal promoter.

Moreover, thanks to the analysis of different reporters, I was able to quantify for the first time the effect of the DnaA-dependent promoter regulatory elements. Specifically, I found that the volume-specific production rate of GFP from the DnaA promoter is a well-defined cell-cycle oscillator and that the signal from this oscillator can be related to key cell-cycle processes such as DNA replication and cell division. However, while the standard view of the cell cycle sees it as the result of a single oscillator, our data lead me to suggest that at least two coupled oscillators are needed to describe the processes that coordinate DNA replication, cell growth and cell-cycle progression. My approach also makes it possible to detect causality links between these different processes.

These findings combine the use of mathematical models and single-cell dynamic data to pose firmer quantitative bases for a characterization of the mechanisms determining robust cell cycle progression in bacteria.

Titre: L'oscillateur du cycle cellulaire DnaA et sa coordination avec la croissance et la division cellulaires chez *Escherichia coli*.

Mots clés: Biologie de la cellule unique, *E. coli*, DnaA, réplication de l'ADN, cycle cellulaire, croissance cellulaire, régulation transcriptionnelle, microfluidique, microscopie, analyse des séries temporelles, oscillateurs couplés

Résumé: Malgré plus de 50 ans d'études, des questions clés sur le cycle cellulaire bactérien restent sans réponse. En particulier, le débat est ouvert concernant la régulation de l'initiation de la réplication de l'ADN chez *E. coli* et son rôle dans la dynamique de la croissance et de la division cellulaire. Un acteur clé de ces processus est la protéine DnaA, qui est impliquée dans l'initiation de la réplication de l'ADN. On pense généralement que DnaA est un oscillateur du cycle cellulaire et un senseur de taille cellulaire, mais aucun de ces faits n'a été fermement établi. L'activité de DnaA dépend de son état de liaison aux nucléotides, la forme liée à l'ATP étant celle qui est active pour la reconnaissance et l'activation de l'origine. DnaA est également un facteur de transcription, un nœud hautement connecté dans le réseau des gènes codant pour les protéines nécessaires à la réplication de l'ADN et à la réparation des lésions de l'ADN. La régulation différentielle de l'expression des gènes par les différentes formes liées aux nucléotides de DnaA, y compris la régulation de son propre promoteur, est considérée comme centrale dans son rôle d'oscillateur et de régulateur du cycle cellulaire. En effet, en raison de son double rôle de facteur de transcription et d'activateur de l'initiation du processus de réplication de l'ADN, la protéine DnaA peut agir comme le lien régulateur entre le timing et le niveau d'expression des gènes et les différentes phases du cycle cellulaire bactérien.

Cette thèse aborde le problème de l'identification de l'oscillateur du cycle cellulaire lié à l'activité de DnaA, et de sa mise en relation avec la progression du cycle cellulaire de *E. coli*, en se concentrant en particulier sur la détection de la taille des cellules, le taux de croissance des cellules individuelles et la division cellulaire. L'un des principaux défis dans ce domaine a été de quantifier les changements dans l'activité DnaA-ATP *in vivo* en temps réel. Ce problème nécessite une résolution dynamique de cellule unique, afin de relier les niveaux et l'activité DnaA aux transitions clés du cycle cellulaire. Pour résoudre ce problème, j'ai développé un ensemble de rapporteurs d'expression génétique utilisant un gène pour une protéine fluorescente sous le contrôle d'un promoteur qui est régulé par la DnaA-ATP, et je les ai déployés dans des expériences unicellulaires couplant microscopie quantitative et microfluidique. Afin d'obtenir un suivi de la cellule unique robuste et à long terme dans des conditions de croissance stable, j'ai conçu un dispositif expérimental et un pipeline d'analyse de données dédiés à l'étude de la croissance, de la taille et de l'expression génétique des E. coli dans des conditions environnementales contrôlées.

Une analyse minutieuse de ces données de la cellule unique en fonction du cycle cellulaire montre que la croissance de *E. coli* est biphasique et qu'elle suit l'expression d'un promoteur constitutif ori-proximal.

De plus, grâce à l'analyse de différents reporters, j'ai pu quantifier pour la première fois l'effet des éléments régulateurs du promoteur dépendant de DnaA. Plus précisément, j'ai découvert que le taux de production de GFP spécifique du volume à partir du promoteur DnaA est un oscillateur du cycle cellulaire bien défini et que le signal de cet oscillateur peut être lié aux processus clés du cycle cellulaire tels que la réplication de l'ADN et la division cellulaire. Cependant, alors que la vision standard du cycle cellulaire le considère comme le résultat d'un seul oscillateur, nos données me conduisent à suggérer qu'au moins deux oscillateurs couplés sont nécessaires pour décrire les processus qui coordonnent la réplication de l'ADN, la croissance cellulaire et la progression du cycle cellulaire. Mon approche permet également de détecter des liens de causalité entre ces différents processus.

Ces résultats combinent l'utilisation de modèles mathématiques et de données dynamiques du cellule unique afin de poser des bases quantitatives plus solides pour une caractérisation des mécanismes déterminant une progression robuste du cycle cellulaire chez les bactéries.