Controlling DNA replication initiation: from living to synthetic cells

Berger, Mareike Sophie

2023

DOI (link to publisher)
10.5463/thesis.125

document version
Publisher's PDF, also known as Version of record

CONTROLLING DNA REPLICATION INITIATION: FROM LIVING TO SYNTHETIC CELLS

MAREIKE BERGER

This thesis was reviewed by:

dr. C.P. Broedersz	Vrije Universiteit Amsterdam
prof.dr. F.J. Bruggeman	Vrije Universiteit Amsterdam
prof.dr. J. Elf	Uppsala University
prof.dr. R.J. Allen	University of Edinburgh
prof.dr. S.J. Tans	Technische Universiteit Delft
dr. T. den Blaauwen	Universiteit van Amsterdam

The work described in this thesis was performed at AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands, part of the Dutch Research Council (NWO). This work was part of the NWO Zwartekracht programme "Building a Synthetic Cell (BaSyC)". Financial support from the institute AMOLF for the printing of this thesis is gratefully acknowledged.

Cover image: Mareike Berger
Printed by: Ipskamp Printing, Enschede

ISBN 978-94-92323-66-8

A digital version of this thesis is available online at `www.ub.vu.nl` and `ir.amolf.nl`. The DOI of this thesis is `http://doi.org/10.5463/thesis.125`. Printed copies can be obtained by request via `library@amolf.nl`.

VRIJE UNIVERSITEIT

CONTROLLING DNA REPLICATION INITIATION: FROM LIVING TO SYNTHETIC CELLS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op woensdag 3 mei 2023 om 11.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Mareike Sophie Berger

geboren te München, Duitsland

promotor:	prof.dr. P.R. ten Wolde
promotiecommissie:	dr. C.P. Broedersz
	prof.dr. F.J. Bruggeman
	prof.dr. J. Elf
	prof.dr. R.J. Allen
	prof.dr. S.J. Tans
	dr. T. den Blaauwen

Don't let them get you down. Be cheeky, wild and wonderful.

Astrid Lindgren

CONTENTS

1 Introduction **1**

2 The initiator titration model **11**
2.1 Growing cell model of gene expression 14
2.1.1 Basal gene expression 14
2.1.2 Constitutively expressed proteins 15
2.1.3 Negatively autoregulated proteins 16
2.2 The AIT model ensures stable cell cycles at low growth rates 18
2.3 A titration-based mechanism fails at high growth rates 21
2.4 Blocking DnaA synthesis via SeqA cannot fully prevent reinitiation events . 24
2.5 Growth-rate dependence of the cell cycle in the AIT model 25
2.6 Adder correlations in the AIT model 27
2.7 Discussion 28
2.A Stability of the initiator accumulation model 30
2.B Biological parameters of the AIT model 33
2.C Quasi-equilibrium assumption for DnaA-titration site binding 34
2.D Loosening the coupling between replication initiation and division 35

3 An origin-density sensor can generate adder correlations **43**
3.1 An ultra-sensitive switch gives rise to an origin-density sensor 46
3.2 Large amplitude oscillations at high growth rates in LDDR model 47
3.3 Switch requires different volume-dependence of (de)activators. 52
3.4 Growth-rate dependence of the initiation volume 53
3.5 A stochastic model recovers adder correlations 55
3.6 Effect of fluctuations in the total initiator concentration 58
3.7 Negative autoregulation of DnaA generates sizer correlations. 60
3.8 Discussion 61
3.A Experimental findings and parameters 64
3.B The effect of protein synthesis in the LD model 68
3.C The LD model in the ultra-sensitivity regime 68
3.D Derivation of adder correlations from size sensor 69
3.E Sizer and adder correlations in the LDDR model 71
3.F Fluctuations in other switch components. 72

4 Full switch-titration model: model validation and predictions **79**
4.1 Coupling titration with DnaA activation enhances robustness 81
4.2 Titration can generically enhance an activation switch 84
4.3 Switch-titration-SeqA model: large oscillations at all growth rates 88
4.4 Transition between titration, switch or SeqA dominated regime 88

4.5 Discussion . . . 93
4.A Model validation . . . 96
4.A.1 Effect of mutations in the activators and deactivators of DnaA . . . 96
4.A.2 Effect of varying the number of titration sites per origin on the initiation volume per origin . . . 101
4.A.3 Effect of varying the total DnaA concentration on the initiation volume per origin . . . 102
4.A.4 Externally driven oscillation in the DnaA concentration can turn an initiation-adder into an initiation-sizer . . . 103
4.B Novel predictions from the full switch-titration model . . . 108
4.C The role of the lipids . . . 109
4.C.1 The importance of lipid-mediated DnaA rejuvenation . . . 110
4.C.2 The role of *oriC* in lipid-mediated DnaA reactivation . . . 112

5 Synchronous replication initiation of multiple origins 121
5.1 Licensing period must be non-zero and shorter than blocking period . . . 125
5.2 Steep rise in the origin opening probability is essential . . . 128
5.3 Initiation synchrony in molecular model for E. coli . . . 133
5.4 Discussion . . . 143
5.A coarse-grained model for origin opening . . . 145
5.B Derivation of approximation for opening probability . . . 145
5.C Parameter choice for maximal firing rate . . . 148
5.D Derivation of theoretical prediction for degree of synchrony . . . 148
5.E Derivation of average time interval between firing events and CV . . . 151
5.F Model for the initiation cascade . . . 151

6 Towards a synthetic cell cycle 161
6.1 DNA replication: simple yet controlled . . . 164
6.2 DNA segregation: using a biological or a physical approach . . . 167
6.3 Cell division: usual suspects and out-of-the-box alternatives . . . 169
6.4 Cell-cycle control . . . 173
6.5 Integration and compatibility of a synthetic cell cycle . . . 176
6.A Models for cell size and DNA density control in bacteria . . . 179

References 181

Summary 200

Samenvatting 203

List of Publications 206

Acknowledgements 207

About the author 217

Artist impression of a dividing *E. coli* cell with its circular chromosome
Watercolor, 2022
Mareike Berger

1

INTRODUCTION

1

Living systems perform a fascinating range of complex tasks: They grow and divide, process information and react to their environment. Every organism is however made of a large variety of biomolecules, which are themselves well described by the laws of physics: Their trajectories follow simple equations of motion and none of the complex properties of living systems can be found in the dynamics of the individual particles. It is the large number, variety, and complex interaction of these cellular components that give rise to the emerging properties of living systems. While we have extensive knowledge about the individual molecular building blocks that form the basis of living systems, we do not understand how these components collectively work together to form life [1].

One of the key features of living beings is the ability to autonomously reproduce themselves. Contrary to humans and animals, unicellular organisms like bacteria, archea, and certain types of eukaryotes consist of a single cell that can autonomously reproduce itself. A cell is the smallest structural and functional unit of life and consists of a cytoplasm enclosed within a membrane, which contains many biomolecules such as proteins and nucleic acids [2]. Cells make a copy of themselves by growing and dividing into two daughter cells. During each cycle of growth and division, the so-called cell cycle, the cell needs to copy all cellular components. Most importantly, the genetic code that contains all the information about the organism, must be copied exactly once per cell cycle.

Making exactly one copy of the genetic material once per cell cycle presents a major challenge to the cell. The genetic code is distributed over one or several chromosomes in the form of long DNA molecules. In the following, we consider for simplicity organisms like bacteria and archea where all the genetic information is stored on a single chromosome. From a chemical point of view, DNA replication can be seen as the reaction $A \xrightarrow{\kappa} 2\,A$, where A represents the chromosome and κ is the rate at which this reaction happens. After a single copying event, the number of chromosomes A has doubled. Given that there are now two chromosomes that can undergo the duplication reaction, the overall rate to make more chromosomes becomes 2κ. The more chromosomes there are, the more likely it is to make even more of them. Such a reaction is called a run-away reaction (Fig. 1.1). If the DNA replication reaction happens with a fixed rate κ that is faster than the rate at which the cell increases its volume, the DNA content in the cell increases exponentially. Such uncontrolled DNA replication would hamper cell growth and ultimately lead to growth arrest and cell death. Therefore, the cell has developed mechanisms that control the chromosome copy number.

A constant replication rate κ that matches exactly the growth rate of the cell is also not sufficient to control the number of chromosomes in the cell. Bacteria typically grow exponentially [3, 4], which means that they double their volume with a certain doubling rate. If the rate at which DNA replication is fired matches exactly the doubling rate of the cell, then the chromosome is replicated on average exactly once per cell cycle. There are however two complications: Firstly, the rate at which cells double their volume typically varies over several orders of magnitude depending on how nutrient-rich their environment is. A constant DNA replication rate would therefore also have to vary with the growth condition to match this changing doubling time of the cell. A second, even more important problem is that biological systems are noisy. Due to low copy numbers of in-

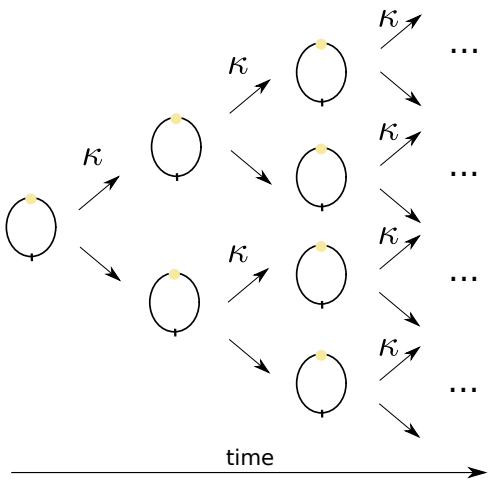


Figure 1.1: DNA replication is a runaway reaction. The circular chromosome is replicated with a constant duplication rate κ. As the number of chromosomes increases with each replication event, the total number of chromosomes increases exponentially in time.

dividual proteins, reactions happen stochastically. This means that even if the rate at which new DNA is made matches the doubling rate of the cell perfectly, DNA replication may be initiated too late or not at all during one division cycle. This would result in the fatal event of a daughter cell without a fully replicated chromosome that cannot reproduce itself. In living cells, this scenario is never observed. Cells therefore have evolved mechanisms to control the rate at which DNA replication starts very precisely to ensure that the genetic code is passed on with high fidelity.

The ability of cells to maintain a stable number of chromosomes over many generations is an example of homeostasis, a key feature of every living being. Homeostasis is the ability of living systems to maintain certain key parameters, such as the number of chromosomes per cell, within a well-defined range of parameters. A key mechanism to maintain homeostasis is to employ negative feedback loops. In a negative feedback system, the output of a reaction has a negative impact on the reaction itself and reduces the probability of the reaction to occur. In the case of DNA replication, negative feedback makes the replication initiation rate κ dependent on how many chromosomes there are in the cell: when there are too many chromosomes, the reaction is inhibited and when there are too few chromosomes then the negative feedback becomes weak and more chromosomes can be produced. Like this, the number of chromosomes in a cell could be kept within a well-defined range of parameters independent of the growth rate of the cell. As the cell typically requires only a single DNA replication event per cell cycle, this negative feedback would have to be extremely sensitive to the small number of chromosomes per cell. It remains however a great puzzle how such negative feedback loops are implemented in different organisms to control the DNA content of a cell so precisely.

Maintaining cell size homeostasis is another fundamental requirement for the survival of cells. Different organisms typically have a well-defined size which can depend on

the growth condition. Again, maintaining a constant size is not automatically ensured: Cells must divide when they reach exactly twice the birth volume, such that after division, the daughter cells start again with the same birth volume as the generation before. This challenge becomes even more difficult in the presence of noise: during cell division, cells do not always divide perfectly in the middle. This gives rise to a smaller and a larger daughter cell. Given that both daughter cells grow at approximately the same rate as the mother, the smaller cell has to grow longer than the larger cell to reach the same division size. Without a homeostatic mechanism, small cells would remain small or become even smaller and large cells would become too large. Indeed, recent single-cell measurements have revealed that many unicellular organisms achieve cell size homeostasis by adding an on average constant volume from birth to division [4–7]. Such a mechanism brings an initially too large or too small cell back to an average size over the course of several generations [8]. This finding inspired researchers to think about molecular mechanisms that could give rise to this phenomenological observation [8–14]. Nevertheless, it remains an open question how bacteria and other unicellular organisms control their size on a molecular level.

To study how life-like features like chromosome copy number and cell size homeostasis emerge from molecular components, mainly two approaches are used today: the bottom up or the top down approach. Bottom up studies start with the individual components and try to reassemble them. As we will discuss later, this often turns out to be extremely challenging. The top down approach on the other hand starts with a living organism and aims at understanding its working principles. This approach is challenging because living organisms are highly complex, involving many biochemical reactions between many often still unknown proteins. Compared to eukaryotic cells, bacteria typically have fewer genes and involve fewer regulatory elements. Due to their relative simplicity, bacteria are thus a good candidate for gaining a better understanding of the basic working principles of life.

Escherichia coli's hardiness, ease of handling, and rapid growth dynamics have made it the most intensively studied and best understood bacterium (and even organism) on the planet [15]. The rod-shaped bacterium *E. coli* is commonly found in the lower intestine of warm-blooded organisms [16]. The bacterium has a single, circular chromosome and replication is initiated at the so-called origin region on the chromosome once per cell cycle [17, 18]. Interestingly, the size of the *E. coli* bacterium depends on how fast it grows: The average cell volume increases exponentially as a function of the growth rate λ [19]. This simple relationship between the volume and the growth rate of the cell inspired researchers early on to search for more quantitative laws that explain how the cell cycle of *E. coli* is organized. In the 60s, Cooper and Helmstetter observed that *E. coli* divides an approximately constant cell cycling time τ_{cc} after initiation of DNA replication [20]. Donachie then combined these two observations and predicted that DNA replication is initiated when the cell reaches a critical volume per number of origins [21]. This early prediction of a constant initiation volume per origin was recently confirmed by single-cell microscopy [22, 23]. The finding that the time from the beginning of DNA replication until cell division is constant led to a replication-centric view of the cell cycle: The main control event in *E. coli* is the moment when DNA replication starts and cell division is then triggered a constant time later. This is an elegant solution to both ensure

that every daughter cell obtains exactly one copy of the chromosome and to maintain cell size homeostasis. Recent single-cell experiments have however challenged this idea and it appears that DNA replication and cell division are more loosely coupled than hitherto believed [23–29]. While the cell cycle of *E. coli* has been studied both theoretically and experimentally in great detail during the last decades [20, 21, 30–33], it is still an open question how fundamental mechanisms such as cell division and DNA replication are regulated on a molecular level.

Using a top-down approach, the Craig Venter institute instead set out to find the minimal set of genes required for an organism to survive. Out of the 4,401 genes of *E. coli*, 1600 genes lack experimental evidence of function (34.6%) and 111 have absolutely no evidence of function [34]. In contrast, the genome of the free-living bacterium called *Mycoplasma mycoides* has one of the smallest genomes known for an autonomously replicating cell found in nature with only 985 putative genes [35]. The researchers replaced the native chromosome of the bacterium with an artificially designed genome [36]. Successively removing more and more genes from the artificial genome led to a minimal cell JCVI-syn3.0 1 that is still able to grow and duplicate itself. Surprisingly, its genome could be reduced to about half of its size [37]. This work demonstrates that one of the most fundamental properties of living systems, namely the ability to reproduce, remains possible in an organism with a much smaller genome as compared to the wild type. Recently, adding some genes related to cell division could restore the wild-type cell morphology and lead to the creation of JCVI-syn3.0 + 126, consisting of 481 genes. Strikingly, approximately 20% of these genes still have unknown functions [38]. Furthermore, it remains to be elucidated how the proteins encoded by the remaining 80% of the genes work together mechanistically to generate a living system. This demonstrates the downside of the top down approach: It remains difficult to understand how the proteins interact in order to give rise to the characteristic features of living systems.

Conversely, reassembling life-like properties from the bottom up has the potential to deepen our understanding of the basic working principles of living systems. This however often turns out to be extremely challenging. Nevertheless, inspired by the famous statement of Richard Feynman “What I cannot create, I do not understand”– the Building a synthetic cell (BaSyC) project set out to create a synthetic cell from bottom up.

As there is no clear definition of what a “living” cell is, it is hard to say what a synthetic cell should be. Yet, arguably the defining feature of a living cell is its ability to make a copy of itself. To achieve this ambitious goal, the cell will need three core functionalities (Fig. 1.2): Firstly, the synthetic cell must be able to use the energy delivered in the form of nutrients from the environment in order to produce its own molecular building blocks. To do so, the cell requires DNA on which the genetic information of the cell is encoded. This DNA could be designed synthetically and distributed over one or several chromosomes. Furthermore, the synthetic cell will need the machinery to transcribe, translate and replicate this genetic code. Finally, the cell should divide autonomously. While each of these modules is already extremely complex and hard to achieve on its own, one main challenge and maybe the key question is how these different functions could be coupled to each other in a minimal cell cycle.

In this thesis, we take inspiration from a living organism to propose modules that can be implemented in a synthetic cell. Specifically, using a top-down approach, we study

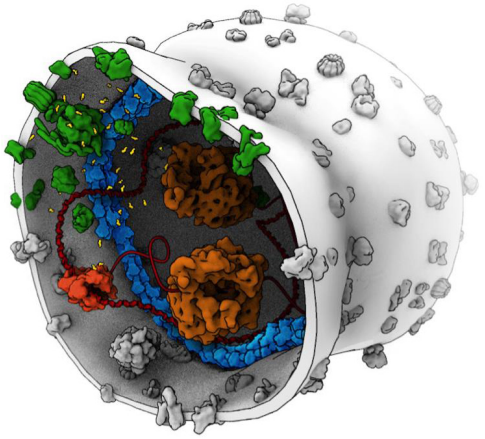


Figure 1.2: Artist's impression of a synthetic cell. The different modules growth (green), DNA processing (red/orange) and Cell division (blue) must be coupled to each other in order to give rise to a stable cell cycle. This Figure was adapted from [1].

how the bacterium *E. coli* controls replication initiation on a molecular level. Based on a wealth of experimental data, we propose a detailed mathematical model for the regulation of DNA replication. We find that the bacterium *E. coli* couples two different mechanisms to sense the number of chromsomes per cell with a high precision. Our model can explain the phenomenological observation that DNA replication is initiated at a critical volume per number of origins and that an on average constant volume is added between consecutive initiation events. Importantly, our model reveals general principles for how a biochemical network must be designed in order to ensure stable DNA replication cycles. As the cell cycle regulation of *E. coli* is probably too complex for implementation in a synthetic cell, we suggest that only one of the two mechanisms employed by *E. coli* could be sufficient for regulating DNA replication in a minimal cell.

In chapters 2, 3 and 4, we present a mathematical model for the regulation of DNA replication initiation in *E. coli*. Experiments indicate that *E. coli* controls replication initiation via two mechanisms: titration and activation of the initiator protein DnaA. In chapter 2, we first show that a titration-based mechanism is not sufficient to generate robust replication initiation cycles at all growth rates. At high growth rates, where multiple chromosomes are replicated simultaneously, the titration model inevitably causes premature reinitiation events which are not observed experimentally. We therefore in chapter 3 show that a system based on activation and deactivation of DnaA can generate stable replication initiation cycles at all growth rates if the total activation and deactivation rate depend differently on the origin density. This requirement is fulfilled by the fact that the lipid-mediated DnaA activation rate is independent of the origin density, while the other (de)activation rates depend on the origin density. Interestingly, even though the switch is at heart an origin-density sensor, it can give rise to the experimentally observed adder correlations in the initiation volume. The finding that titration fails

at high growth rates and the switch is in principle stable at all growth rates raises the question what the role of titration is. In chapter 4, we show that titration can make the switch more robust against fluctuations in the switch components. We therefore predict that both mechanisms together drive robust replication cycles at all growth rates. In chapters 2, 3 and 4 we have assumed that DNA replication is initiated deterministically when a certain threshold in the cell is reached. In chapter 5, we investigate the effect of stochastic replication initiation at the origin. Especially at high growth rates, where replication is initiated at multiple origins synchronously, *E. coli* faces a challenge: Origins that have not yet fired must be allowed to still fire after the first origin has initiated. Simultaneously, every origin should only fire *once* per cell cycle and therefore origins that have already fired must be prevented from re-firing. We find that a global mechanism, namely a delay in the decrease of the initiation potential of the cell in combination with a local mechanism, namely the blocking of already initiated origins, can give rise to synchronous replication initiation at multiple origins. We show that the previously presented theoretical model for replication initiation can ensure a high degree of synchronous replication initiation for biologically realistic parameters. Finally, in chapter 6, we take inspiration from the cell cycle of *E. coli* and other organisms to propose how the cell cycle of a synthetic cell could be regulated. We discuss different mechanisms for DNA replication, DNA segregation, and cell division and propose different routes of how these mechanisms could be coordinated with each other in a minimal cell cycle.

The cascade of life
Clay, 2019
Mareike Berger
Part of the series “Molding Nano”

2

THE INITIATOR TITRATION MODEL

The bacterium Escherichia coli initiates replication once per cell cycle at a precise volume per origin and adds an on average constant volume between successive initiation events, independent of the initiation size. Yet, a molecular model that can explain these observations has been lacking. Experiments indicate that to control replication initiation E. coli employs a mechanism based on the accumulation of an initiator protein up to a threshold number, which is set by the fixed number of titration sites per chromosome. Here we show by mathematical modelling that a mechanism solely based on titration generates stable replication cycles at low growth rates, but inevitably causes premature reinitiation events at higher growth rates. Adding SeqA, which transiently blocks DnaA synthesis after replication initiation, prevents reinitiation events at high but not at intermediate growth rates. In this regime, other mechanisms become essential for stable replication initiation.

The contents of this chapter have been published in Nature Communications **13**, 6556 (2022) [39].

To maintain stable cell cycles over many generations, living cells must coordinate DNA replication with cell growth and cell division. Intriguingly, in nutrient-rich environments, the model organism *Escherichia coli* can even divide faster than the time it takes to replicate its entire chromosome [22, 30, 31, 40]. This apparent paradox was resolved by the model of Cooper and Helmstetter in which new rounds of replication are initiated before the previous round has finished [20] (Fig. 2.1a). Donachie then predicted that replication is initiated at a constant volume per origin v^* [21]. Initiating replication at a constant origin density ensures that DNA replication is initiated once per cell cycle per origin, which is a necessary condition for maintaining stable cell cycles at all growth rates (Fig. 2.1a). Recent experiments at the population level showed that the average initiation volume per origin v^* varies within a $\sim 50\%$ range over a tenfold change in the growth rate [41]. Moreover, single-cell measurements revealed that the initiation volume is one of the most tightly controlled cell-cycle parameters, varying by about 10% for any measured growth rate [22, 23]. Yet, how the initiation volume is controlled so precisely, and what molecular mechanism gives rise to robust cell cycles over many generations remains despite extensive studies poorly understood [32, 33, 42–44].

To obtain insight into the mechanisms that control DNA replication and cell division, fluctuations in cell size have been studied [5, 6]. These experiments revealed that cells obey an adder principle, which states that cells add an on average constant volume independent of the birth volume during each cell cycle. It has been proposed that cell division control is tightly coupled to the control over replication initiation [8, 10, 22], via a sizer on replication initiation and a timer for cell division. Yet, recent experiments revealed the existence of two adders, one on cell division and the other on replication initiation, and that these two processes are more loosely coupled than hitherto believed [23, 24, 26–29, 45]. While these phenomenological observations are vital because they constrain any model on the molecular mechanism for initiation and cell division control, no such molecular model has yet been presented that is consistent with the experimental data.

It has been proposed that replication initiation control in *E. coli* is based on the accumulation of an initiator protein [11, 12]. In the here called initiator accumulation models [8–14], an initiator protein accumulates during the cell cycle proportional to the cell volume, and replication is initiated when a threshold amount per origin has accumulated. As a fixed amount of initiators per origin needs to be accumulated per replication cycle, models of this class are often seen as a mechanistic implementation of an adder [6, 8, 10, 13]. Many variations of this idea with different degrees of detail have been proposed [8, 11–13]. Hansen et al. [12, 14] identified the initiator protein as the protein DnaA, which can be titrated away from the origin by DnaA boxes, high-affinity binding sites on the chromosome [43, 46]. This constant number of titration sites per chromosome sets the critical threshold number of initiator proteins required for initiating replication.

In this chapter, we consider a mechanistic implementation of the initiator accumulation model (Fig. 2.1b). In *E. coli*, the initiator protein DnaA is negatively autoregulated and can be bound to titration sites on the chromosome. Following Hansen et al. [12, 14], we therefore consider a model in which the initiator can bind titration sites and is also autregulated, the Autoregulated Initiator-Titration model (AIT) model. We then study

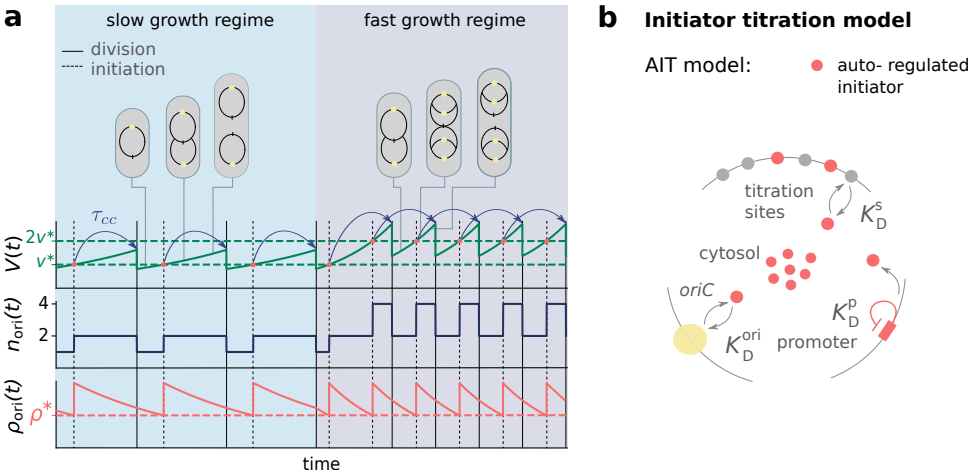


Figure 2.1: We present the Autoregulated Initiator-Titration (AIT) model to elucidate the molecular mechanism by which *E. coli* initiates replication at an on average constant volume per origin. (a) The volume $V(t)$, the number of origins $n_{ori}(t)$ and the origin density $\rho_{ori}(t) = n_{ori}(t)/V(t)$ as a function of time. Initiating replication at a constant origin density ρ^* (dashed red line) or respectively a constant volume per origin $v^* = 1/\rho^*$ (dashed green line) and division a constant time τ_{cc} later (blue arrows) ensures that the cell initiates replication once per division cycle and that it maintains cell size homeostasis at slow (light blue regime) and fast (dark blue regime) growth rates. (b) Scheme of the AIT model: In *E. coli*, the initiator DnaA (red circles) is negatively autoregulated with the dissociation constant K_D^p, and can bind both to the *oriC* and the titration sites with dissociation constants K_D^{ori} and K_D^s, respectively.

the robustness of replication initiation in this model as a function of the growth rate.

Surprisingly, while the AIT model gives rise to stable cell cycles at low growth rates, it exhibits reinitiation events at high growth rates. The effect of the protein SeqA to transiently block synthesis of new DnaA proteins right after replication initiation can prevent reinitiation events at high, but not at intermediate growth rates. We thus argue that the initiator titration model is not sufficient to explain the experimental data on replication initiation in *E. coli* and that a second mechanism based on protein activation is essential to ensure robust replication cylce (see chapter 3).

2.1. GROWING CELL MODEL OF GENE EXPRESSION

In this section we first present the gene expression model, which underlies all our models. In the recently developed growing-cell model by Lin et al. [47], transcription is limited by the availability of RNAPs while translation is limited by the ribosomes. In this model, the mRNA and protein copy numbers are proportional to the cell volume, as recent experiments indicate [47–53]. Concomitantly, the protein synthesis rate is, as observed very recently [48], proportional to the volume, which is a crucial requirement for the stability of the initiator accumulation model (see Appendix 2.A). We start this section by deriving the basal protein synthesis rate in the growing-cell model (section 2.1.1). In section 2.1.2, we show how the synthesis rate of a constitutively expressed protein is proportional to the volume, such that its concentration increases exponentially in time over the course of the cell cycle. In section 2.1.3 we then describe how gene regulation can be included in the growing cell model.

2.1.1. BASAL GENE EXPRESSION

In the gene expression model of Lin et al. [47], the genes and the mRNAs compete for the limiting pool of RNAPs and ribosomes, respectively [47]. Therefore, the transcription rate of a gene i is directly proportional to the total number RNAPs n times the fraction of RNAPs ϕ_i that are transcribing gene i. To quantify the gene allocation fraction ϕ_i, Lin et al. define an effective gene copy number g_i that accounts for its copy number and the binding strength of its promoter [47]. The gene allocation fraction of gene i is then given by the effective gene copy number g_i divided by the sum over all effective gene copy numbers in the cell $\phi_i = g_i / \sum_j g_j$. As the number of ribosomes is assumed to limit translation, the protein synthesis rate of gene i is proportional to the number of ribosomes N_R times the fraction of ribosomes translating the mRNA of gene i. Assuming that the affinity of ribosomes binding to mRNA is equal for all types of mRNA m_i, the ribosome allocation fraction f_i of gene i is given by the number of mRNAs m_i of gene i divided by the total amount of mRNAs, thus $f_i = m_i / \sum_j m_j$. The growing cell model then gives rise to the following set of equations for the change in the number of mRNAs m_i and the number of proteins p_i of gene i:

$$\frac{dm_i}{dt} = k_m \phi_i n - \frac{m_i}{\tau_m} \tag{2.1}$$

$$\frac{dp_i}{dt} = k_R f_i f_a N_R \tag{2.2}$$

where k_m is the transcription rate of a single RNAP, τ_m is the degradation time of the mRNA (taken to be equal and constant for all mRNAs), k_R is the translation rate of a ribosome, f_a is the fraction of actively translating ribosomes and N_R is the number of ribosomes. Due to the fast production and degradation rate of the mRNA compared to the growth rate of the cell, we can approximate the mRNA number to be at a steady state such that

$$\langle m_i \rangle = k_m \phi_i \langle n \rangle \tau_m \tag{2.3}$$

Plugging equation 2.3 into equation 2.2 and using that $\sum_j \phi_j = 1$ gives the following general expression for the change in the number of proteins:

$$\frac{dp_i}{dt} = k_R \phi_i f_a N_R \tag{2.4}$$

The protein production rate of any gene i is therefore proportional to the number of ribosomes N_R times the gene allocation fraction ϕ_i of gene i. The gene allocation fraction ϕ_i is a measure of the relative affinity and amount of gene i with respect to all other genes in the cell. In the simplified scenario of an instantaneous replication of the entire DNA after replication initiation, replication of the DNA does not affect the gene allocation fraction. If the gene i is not regulated, the affinity of gene i is constant in time. If at a given growth rate the total affinity of all genes remains approximately constant in time, the gene allocation fraction ϕ_i is constant in time too.

2.1.2. CONSTITUTIVELY EXPRESSED PROTEINS

In this section, we will first demonstrate that in the growing cell model, the protein production rate is directly proportional to the volume of the cell, which, as we will see in section 2.2, ensures the stability of the AIT model. The total number of proteins N in the cell is given by the sum over all proteins p_j

$$N = \sum_j p_j \tag{2.5}$$

and the fraction of proteins that are ribosomes is

$$\Phi_R = \frac{N_R}{N}. \tag{2.6}$$

From equations 2.4, 2.5 and 2.6, and using that $\sum_j \Phi_j = 1$, we find that the change in the total number of proteins in time is

$$\frac{dN}{dt} = \sum_j \frac{dp_j}{dt} = k_R f_a N_R = k_R f_a \Phi_R N \tag{2.7}$$

while, defining the total number density $\rho \equiv N/V$, the change in the volume is

$$\frac{dV}{dt} = \frac{1}{\rho}\frac{dN}{dt} = k_R f_a \Phi_R V \tag{2.8}$$

Hence, the cell grows exponentially with a growth rate

$$\lambda = \frac{1}{N}\frac{dN}{dt} = \frac{1}{V}\frac{dV}{dt} = k_R f_a \Phi_R \tag{2.9}$$

Using equation 2.9 we can then derive the change in the number of a protein of gene i:

$$\frac{dp_i}{dt} = \phi_i k_R f_a N_R = \phi_i k_R f_a \Phi_R N = \phi_i \lambda N = \phi_i \lambda \rho V \tag{2.10}$$

Therefore, while in the standard model of gene expression the copy number of a constitutively expressed protein i increases bi-linearly in time, in the growing cell model it increases exponentially over the course of the cell cycle. The change in the protein concentration of gene i is then given by

$$\frac{d[p_i]}{dt} = \frac{dp_i}{dt}\frac{1}{V} - p_i \frac{1}{V^2}\frac{dV}{dt} = \phi_i \lambda \rho - \lambda [p_i] \tag{2.11}$$

At steady state, we find that the growth rate drops out and the steady state protein concentration is given by:

$$[p_i]^* = \phi_i \rho \tag{2.12}$$

In order to investigate how the protein number and concentration of an unregulated protein changes over the course of the cell cycle, we evolve the volume of a cell according to to $dV/dt = \lambda V$ (see 2.8 and 2.9) and the protein number according to equation 2.10. Replication is initiated at a fixed volume per origin v^* and the cell divides a fixed time τ_{cc} after replication initiation. The exponential increase in the number of proteins over the course of the cell cycle can be seen in Fig. 2.2a. In the scenario where the entire chromosome is replicated instantaneously and the gene is not regulated, the gene allocation fraction ϕ_i remains constant (Fig. 2.2a, yellow line). While the number of a protein p increases proportional to the volume of the cell (Fig. 2.2a, blue line), the concentration remains perfectly constant in time (Fig. 2.2a, red line).

In reality the chromosome is not replicated instantly. This means that when the part that houses gene i is replicated, the gene allocation fraction ϕ_i rises transiently, as illustrated in the second panel of Fig. 2.2b. The transiently higher gene allocation fraction results in a temporal increase of the production rate (Fig. 2.2b, third panel), which gives rise to weak oscillations in the protein concentration over the course of the cell cycle.

2.1.3. NEGATIVELY AUTOREGULATED PROTEINS

Regulation of gene i can be included by modifying the gene affinity g_i. If gene i is for example negatively autoregulated, the gene affinity becomes

$$g_i = g_i^0 \frac{1}{1 + \left(\frac{[p_i]}{K_D^p}\right)^n} \tag{2.13}$$

where g_i^0 is the basal gene affinity if the promoter is not repressed at all, $[p_i]$ the free initiator concentration, K_D^p is the dissociation constant of the promoter and n is the Hill coefficient. The protein production rate then becomes dependent on the protein con-

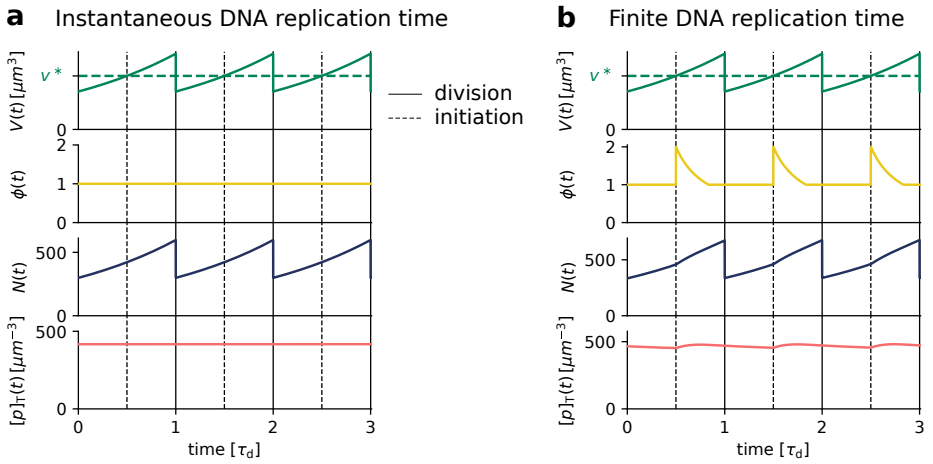


Figure 2.2: The concentration of differently regulated proteins in the growing cell model of gene expression (a, b) The volume $V(t)$, the gene allocation fraction $\phi(t)$, the number of proteins $N(t)$ and the total concentration $[p]_T$ of a constitutively expressed protein within the growing cell model. The volume and the protein number are evolved according to equations 2.8 and 2.10, respectively. (a) While the protein number increases exponentially in time, the total concentration remains perfectly constant. (b) The change of the number and concentration of a constitutively expressed protein when the gene allocation fraction changes in time due to a finite time to replicate the entire chromosome. The gene is assumed to be located at the origin which causes a doubling of the allocation fraction at the moment of replication initiation. When the entire chromosome has been replicated, the gene allocation fraction is again constant. As a consequence, the concentration of a constitutively expressed gene exhibits weak oscillations due to the changes in the gene allocation fraction. The parameters in all simulations are $v^* = 1\,\mu\text{m}^3$, $\tau_{cc} = 1$ h, $T_C = 2/3$ h, $\rho = 10^6\,\mu\text{m}^{-1}$, $\tau_d = 2$ h and $\phi_i = 2 \times 10^{-3}$.

centration via the modified gene allocation fraction ϕ_i:

$$\frac{dp_i}{dt} = \phi_i \lambda \rho V = \frac{g_i}{\sum_j g_j} \lambda \rho V \tag{2.14}$$

$$= \phi_i^0 \frac{1}{1 + \left(\frac{[p_i]}{K_D^p}\right)^n} \lambda \rho V \tag{2.15}$$

where we defined the *basal* gene allocation fraction $\phi_i^0 \equiv g_i^0 / \sum_j g_j$. By defining the gene allocation density as $\tilde{\phi}_i^0 = \phi_i^0 \rho$, we obtain the production rate of a negatively autoregulated protein p (with $i = p$):

$$\frac{dN_p}{dt} = \frac{\tilde{\phi}_p^0 \lambda V}{1 + \left(\frac{[p]}{K_D^p}\right)^n} \tag{2.16}$$

2.2. THE AIT MODEL ENSURES STABLE CELL CYCLES AT LOW GROWTH RATES

Figure 2.1b shows the key ingredients of the AIT model. It consists of a negatively autoregulated initiator protein p, such that the change in copy number N_p is given by

$$\frac{dN_p}{dt} = \frac{\tilde{\phi}_p^0 \lambda V}{1 + \left(\frac{[p]}{K_D^p}\right)^n} \tag{2.17}$$

following the growing cell model of gene expression of Lin et al. [47] (see section 2.1) with gene allocation density $\tilde{\phi}_p^0$, dissociation constant of the promoter K_D^p, Hill coefficient n and concentration of the initiator protein $[p] = N^f/V$ in the cytoplasm. As the DnaA gene is located very closely to the origin, we assume that at the moment of replication initiation the gene number doubles instantaneously. The volume $V(t)$ of the cell grows exponentially,

$$\frac{dV}{dt} = \lambda V \tag{2.18}$$

where the growth rate $\lambda = \ln(2)/\tau_d$, with cell-doubling time τ_d, is a model parameter. The total initiator concentration is then obtained by dividing the total number of initiators N_p by the volume V of the cell.

The model also includes a number N_s of high-affinity titration sites that are distributed randomly on the chromosome [14, 54]. In our model, each chromosome contains a constant number of homogeneously distributed titration sites and the initiator proteins can be either freely diffusing in the cytoplasm or bound to these titration sites. As binding and unbinding dynamics of the initiator protein to the titration sites are relatively fast [55–57], we assume for simplicity a quasi-equilibrium state and use a quadratic equation to calculate at every given total titration site concentration and total initiator concentration in a cell the concentration of initiators freely diffusing in the cytoplasm (see Appendix 2.C). A new round of replication is initiated when the free initiator concentration $[p]$ reaches the dissociation constant for binding to the origin, K_D^{ori}. When

replication is initiated, the number of titration sites $N_s(t)$ increases linearly from the moment of initiation of replication t_i until the end of replication at $t_i + T_C$:

$$N_s(t) = \begin{cases} N_0 & \text{for } t < t_i \\ N_0 + N_0 \frac{t-t_i}{T_C} & \text{for } t_i \le t < t_i + T_C \\ 2\,N_0 & \text{for } t \ge t_i + T_C \end{cases} \tag{2.19}$$

with the C-period $T_C \approx 40$ min being the time to replicate the entire chromosome and $N_0 = n_s\, n_{ori}$ is the total number of titration sites before replication initiation, given by the number n_s of titration sites per chromosomes times the number n_{ori} of origins before replication initiation. Based on the general growth law, the cell divides a constant cycling time τ_{cc} after initiation of replication [22, 40]. The cycling time can be split into two times $\tau_{cc} = T_C + T_D$, the C-period and the D-period: During the C-period, the DNA is being replicated and during the D-period the chromosomes are being separated and the cell divides [20, 22, 40]. The choice of a constant cycling time τ_{cc} is convenient, as it directly couples cell division to replication, thus eliminating the need for implementing an additional mechanism for cell division, yet does not affect our results, as we discuss in Appendix 2.D.

The AIT model generates stable cell cycles at low growth rates (Fig. 2.3a). Evolving the number of initiator proteins and the volume according to equations 2.17 and 2.18, respectively, we find that the total DnaA concentration remains approximately constant in time (Fig. 2.3a). The weak oscillations in the total concentration arise from the effect of a finite replication time of the chromosome as explained in section 2.1.2. Because the dissociation constant of the initiator protein for the titration sites K_D^s is smaller than that for the origin $K_D^{ori} > K_D^s$, the cytoplasmic initiator concentration $[p]$ (see Appendix 2.C) remains below the critical initiation threshold K_D^{ori} as long as there are still unoccupied titration sites (Fig. 2.3a, lowest panel). Yet, when the total number of proteins N_p exceeds the total number of titration sites N_s, the free concentration $[p]$ rapidly rises. When the free initiator concentration $[p]$ reaches the threshold K_D^{ori}, a new round of replication is initiated. New titration sites are now being synthesized faster than new proteins are being produced and therefore the free initiator concentration $[p]$ drops rapidly far below K_D^{ori} (Fig. 2.3a, lowest graph). The cell then divides a constant time τ_{cc} after replication initiation, during which the volume, the number of initiator proteins, and the number of titration sites are halved. In fact, in this mean-field description cell division does not change the concentrations of the components and it therefore does not affect the replication cycle. Only when again enough initiator proteins have been accumulated is a new round of replication initiated.

An open question remains what role negative autoregulation plays in the AIT model. In order to attain the critical initiation concentration K_D^{ori} at the origin, the dissociation constant of the promoter of DnaA K_D^p must be higher than K_D^{ori}. At the same, the mechanism of titration requires that the affinity of the titration sites is higher than that of the origin: $K_D^s < K_D^{ori}$. Combining these two requirements yields: $K_D^s < K_D^{ori} < K_D^p$. The free protein concentration $[p]$ thus remains (far) below the promoter dissociation constant K_D^p, which means the latter is repressed only weakly and proteins are produced approxi-

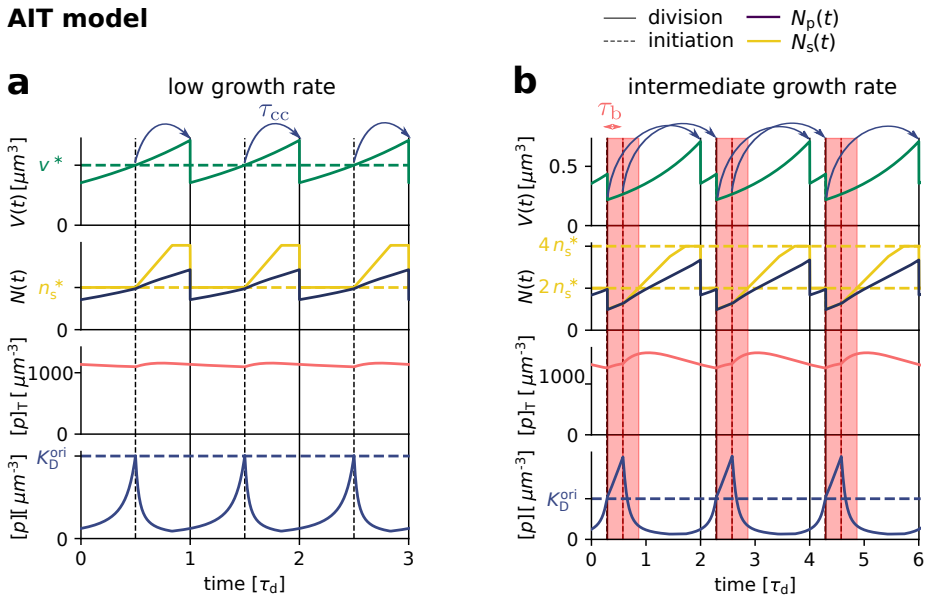


Figure 2.3: While the AIT model ensures stable cell cycles at low growth rates (a), it gives rise to premature reinitiation events at high growth rates (b). (a, b) The volume $V(t)$, the number of initiator proteins $N_p(t)$ and titration sites $N_s(t)$, the total concentration of initiator proteins $[p]_T(t)$, and the concentration of initiator proteins in the cytoplasm $[p](t)$ as a function of time (in units of the doubling time of the cell τ_d) for $\tau_d = 2$ h (a) and $\tau_d = 35$ min (b), respectively. (a) When the number of initiator proteins per origin $n_p(t)$ exceeds the number of titration sites per origin n_s (yellow dashed line), the free concentration $[p](t)$ rapidly rises to reach the threshold concentration K_D^{ori} (blue dashed line), initiating a new round of replication. Due to the homogeneous distribution of titration sites on the chromosome of *E. coli* and the constant DNA replication rate, the number of titration sites then increases linearly in time. At low growth rates, new titration sites are synthesized faster than new initiator proteins and the free concentration $[p](t)$ rapidly drops after initiation. After a fixed cycling time τ_{cc} (blue arrows) the cell divides. The initiation volume per origin v^* (green dashed line) at low growth rates is constant in time. (b) When the doubling time is however smaller than the time to replicate the entire chromosome, $\tau_d < T_C$, new proteins are synthesized faster than new titration sites are formed. After a short period $\tau_b = 10$ min (shaded red area) during which initiation at *oriC* is blocked via the protein SeqA, replication is reinitiated prematurely, dramatically raising the variation in the initiation volume (see Fig. 2.4a and b). (See Table 2.1 for all parameters.)

mately at the maximal rate. Therefore, equation 2.17 can be approximated by

$$\frac{dN_\mathrm{p}}{dt} \approx \tilde{\phi}_\mathrm{p}^0 \lambda V \tag{2.20}$$

The stability of the AIT model arises from the volume dependence in the initiator production rate in equation 2.17 as explained in Appendix 2.A.

2.3. A TITRATION-BASED MECHANISM FAILS AT HIGH GROWTH RATES

At low growth rates, the synthesis of new titration sites after replication initiation prevents reinitiation. Because the titration sites are homogeneously distributed over the chromosome [14, 54], the rate at which new titration sites are formed after replication initiation is given by the DNA duplication rate, which is, to a good approximation, independent of the growth rate [40]. In contrast, the protein synthesis rate increases with the growth rate λ, see Eq. 2.17. In the low growth regime, the time to replicate the entire chromosome T_C is shorter than the time to double the volume of the cell τ_d. The time it takes to double the number of titration sites upon replication initiation is therefore shorter than the time to double the number of initiation proteins. This results in a gradual decrease of the free initiator concentration upon replication initiation (Fig. 2.3a, lowest panel) and prevents reinitiation.

At higher growth rates, the titration mechanism, however, breaks down. In favorable growth conditions, the doubling time of *E. coli* can be shorter than the time it takes to replicate the entire chromosome T_C. As a result, the rate at which new titration sites are formed upon the first replication initiation event (marked by the dashed vertical lines) is therefore lower than the rate at which initiator proteins are produced (Fig. 2.3b); the number of titration sites (yellow line) rises slower than the number of initiators (black line). This means that after the first replication initiation event, the free initiator concentration continues to *rise* (Fig. 2.3b, lower panel). To prevent immediate reinitiation, we introduce a refractory or 'eclipse' period of $\tau_\mathrm{b} \approx 10$ min after replication initiation during which replication initiation is blocked (red shaded area), mimicking the effect of SeqA [43, 58–60]. When this eclipse period is however over, a new round of replication is initiated, which triples the rate at which new titration sites are formed. Now the rate of titration-site formation is higher than the rate at which new initiator proteins are produced, causing the concentration of free initiator to go down. At some point, the first round of replication is finished, causing a small decrease in the rate at which new titration sites are formed and some time later also the next round is finished, causing the number of titration sites to become constant. Then the time τ_cc after the first initiation event is reached and the cell divides. After this division event the cell grows briefly and then divides again, a time τ_cc after the second initiation event in the previous cycle. A given cell cycle thus consists of a long and a short cycle, such that the average division time (time from birth to death) equals the doubling time $\tau_\mathrm{d} = \ln(2)/\lambda$. These unnatural time traces of the volume, namely the oscillation between a short and a long (sub) cycle, have not been observed experimentally and can be prevented by decoupling cell division from replication initiation as described in Appendix 2.D. The reinitiation events,

which are caused by the excess of initiators after the first initiation event are however not affected by the choice of how the replication and the division cycle are coupled. Because in the long (sub) cycle two initiation events are triggered in rapid succession, the initiation volume *per origin* flip-flops between a high and a low initiation volume per origin. This causes a dramatic rise of the Coefficient of Variation (CV) in the initiation volume at higher growth rates (see Fig. 2.4a and b). Importantly, the CV becomes much larger than that observed experimentally, *even though the system is deterministic and no biochemical noise is present; adding noise would only make the CV even higher.*

We emphasize that the breakdown of the titration mechanism arises from the different scaling of two timescales with the growth rate: The rate at which the initiator DnaA is synthesized scales with the growth rate, see Eq. 2.17. In contrast, the titration-site formation rate is nearly independent of the growth rate: when the titration sites are homogeneously distributed, as experiments show [54], then the titration-site formation rate per origin is set by the DNA duplication rate, which indeed varies only little with the growth rate [40]. The protein synthesis rate thus increases faster with the growth rate than the titration-site formation rate, which means that at sufficiently high growth rates the mechanism fails to sequester DnaA proteins after replication initiation; to a good approximation, this breakdown happens when the system enters the overlapping replication-fork regime with $\tau_d \lesssim T_C$, because the rate at which titration sites are formed is given by n_s/T_C, while the rate at which proteins are produced right after replication initiation is given by $dN_p/dt = \phi_p^0 \lambda \rho V = \lambda N_p = \ln(2) N_p/\tau_d \simeq \ln(2) n_s/\tau_d$, where we have used that the fraction of initiator equals the gene (ribosome) allocation fraction ϕ_p^0 (assuming all proteins are made with the same rate) and right after replication initiation $N_p \simeq n_s$. In fact, our modelling predicts that in mutants in which T_C is increased [40], the titration mechanism continues to fail in the overlapping fork regime, even though in these mutants this regime starts at lower growth rates (Fig. 2.4a). This supports the idea that the breakdown of the titration mechanism is due to the different scaling of the titration-site formation rate and the DnaA synthesis rate with the growth rate.

Since this prediction follows from a scaling of two timescales, it is robust, i.e. insensitive to the details of the model. Indeed, this prediction is insensitive to how the other key parameters in the AIT model are varied: the number of titration sites per origin n_s and their affinity K_D^s. Figure 2.4b shows that exactly the same rise in the coefficient of variation (CV) of the initiation volume is observed for different values of n_s and K_D^s, which are varied together to keep the average initiation volume constant and within the range observed experimentally [40]. The fact that the curves nearly fully overlap is because a new replication round is initiated as soon as the eclipse period is over.

Placing the titration sites near the origin would prevent premature reinitiation at high growth rates (see Fig. 2.10), but this is not consistent with experiments [14, 54]. Naturally, if the affinity of the titration sites located at the origin is higher than the affinity of titration sites at the rest of the chromosome, we can recover the behavior of the inhomogeneous titration site distribution. Interestingly, it had been proposed that the site *datA* which is located close to the origin has a very high affinity and can titrate large numbers of proteins, of up to 60-370 [61, 62]. These numbers had been inferred indirectly, from experiments that analyzed the de-repression of dnaA or mioC transcription upon introduction of plasmids containing *datA* sequences [63]. It remained however unclear

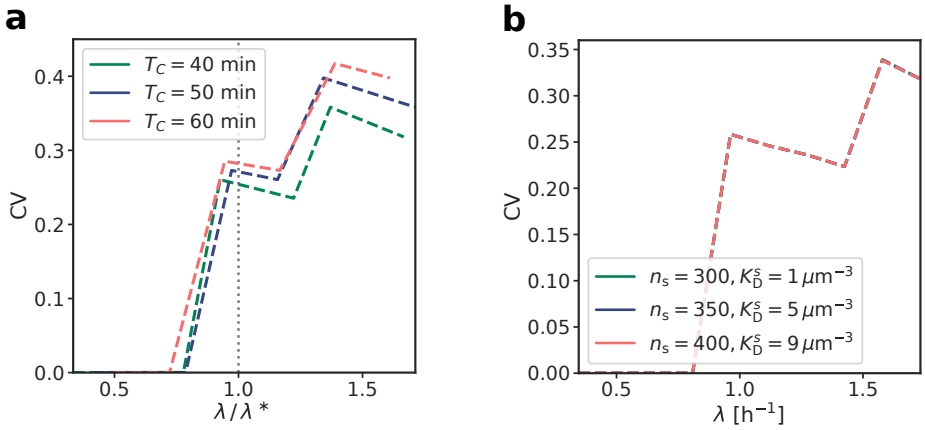


Figure 2.4: A homogeneous titration site distribution on the chromosome in the AIT model causes reinitiation events at high growth rates (e, f) The coefficient of variation CV = σ/μ with the standard deviation σ and the average initiation volume $\mu = \langle v^* \rangle$ as a function of the growth rate for the AIT model with homogeneous titration site distribution. Due to the rapid reinitiation events shown in (d), the coefficient of variation increases strongly in the overlapping fork regime at high growth rates. (e) The onset of the overlapping fork regime can be shifted to lower growth rates by increasing the time to replicate the chromosome T_C. When the growth rate is rescaled with the critical growth rate $\lambda^* = \ln(2)/T_C$ at which the cell-doubling time τ_d equals the C-period all curves essentially overlap, showing that the transition to the overlapping fork regime triggers the re-initiation events. (f) Varying the total number of titration sites in concert with the dissociation constant of the titration sites K_D^S such that the initiation volume remains constant and equal to the experimentally observed initiation volume [40] (by solving equation 2.30 for v^*) cannot prevent these reinitiation events. This demonstrates that the failure of the titration model at high growth rates is independent of the precise parameter choice; it is thus a robust result.

by which mechanism *datA* would be able to absorb so many DnaA molecules. The discovery that the site *datA* can deactivate the initiator protein ATP-DnaA by promoting ATP hydrolysis provides a more likely explanation for this indirect observation [63] (see chapter 3). In the original initiator titration paper by Hansen et al. [64], a bias of titrating DnaA boxes towards the *oriC* region was assumed. Roth and Messer [54] find however that while boxes of the R1 type indeed show such a bias, the high-affinity DnaA boxes show a random distribution on the chromosome.

To conclude, when the titration sites are distributed homogeneously on the chromosome, the titration mechanism inevitably fails at high growth rates in the overlapping fork regime. As the breakdown of the titration mechanism arises from the different scaling of two timescales with the growth rate, this finding is robust and does not depend on the parameters of the model. These observations show that the *E. coli* replication cycle is not regulated via titration only.

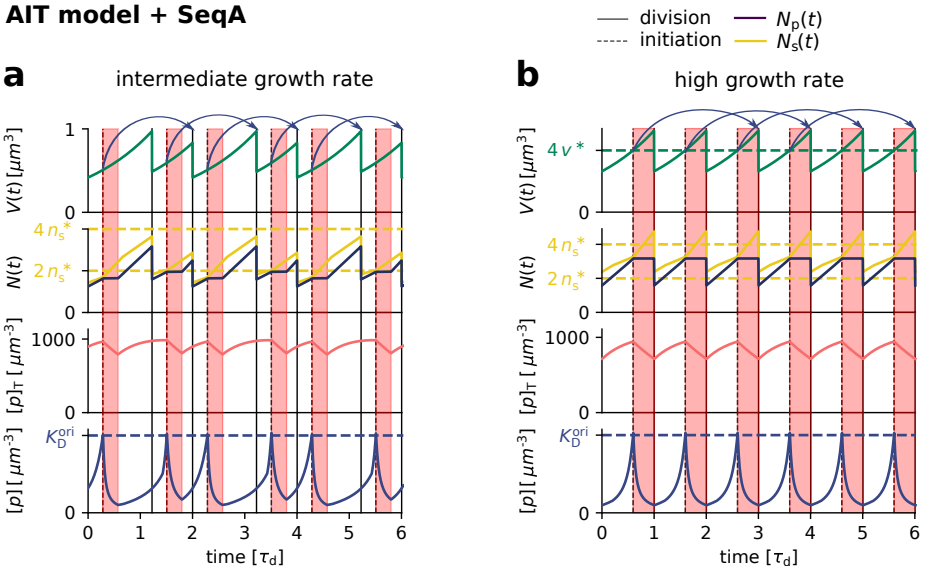


Figure 2.5: Adding SeqA, which transiently blocks DnaA synthesis after replication initiation, to the AIT model prevents reinitiation events at high (a) but not at intermediate growth rates (b). (a, b) The volume $V(t)$, the number of initiator proteins $N_p(t)$ and titration sites $N_s(t)$, the total concentration of initiator proteins $[p]_T(t)$, and the concentration of initiator proteins in the cytoplasm $[p](t)$ as a function of time (in units of the doubling time of the cell τ_d) for $\tau_d = 35$ min (a) and $\tau_d = 25$ min (b), respectively. (c, d) Blocking also transiently DnaA synthesis via SeqA during $\tau_b = 10$ min (shaded red area) can prevent reinitiation at high (d), but not at intermediate growth rates (c). (See Table 2.1 for all parameters.)

2.4. BLOCKING DNAA SYNTHESIS VIA SEQA CANNOT FULLY PREVENT REINITIATION EVENTS

Interestingly, experiments indicate that after replication initiation SeqA not only blocks the origin, preventing immediate reinitiation, but also transiently lowers the DnaA synthesis rate [58–60]. The combination of periodic suppression of DnaA synthesis with DnaA titration enables robust DnaA rhythms at sufficiently high growth rates ($\lambda > 1.5\,h^{-1}$) (Fig. 2.5b). But at lower growth rates, corresponding to longer doubling times, the effect of SeqA becomes weaker. As a result, at intermediate growth rates ($1 < \lambda < 1.5\,h^{-1}$) this combination cannot prevent premature reinitiation events; the time between successive initiation events alternates between a time that is shorter than τ_d and one that is longer than τ_d, giving rise to variations in the initiation volume, even though the system is deterministic (Fig. 2.5a). At intermediate growth rates, the CV of a system based on only titration and SeqA therefore still rises strongly (Fig. 2.6a).

Experiments indicate that the duration of the refractory period depends on the growth rate [58]. Figure 2.6b shows the CV of the initiation volume for a broad range of growth rates and durations of the blocked period in the deterministic system with no biochemical noise and a homogeneous distribution of titration sites. The CV is zero in the non-

overlapping fork regime, in which the cell doubling time τ_d is longer than the time to replicate the entire chromosome T_C, and rises sharply in the overlapping fork regime, where the growth rate $\lambda^* > \ln(2)/T_C$ (marked by the red vertical line). With increasing duration of the blocked period the CV decreases, especially at high growth rates. Importantly however, for all blocked periods shown, there is a range of growth rates in which the CV is larger than zero. Moreover, for the experimentally reported duration of the blocked period of $\tau_b = 4$ minutes at a growth rate of $\lambda = 1.59\ \mathrm{h}^{-1}$ (marked by the white dot on the right), our model predicts strong re-initiation events [58]. Making the duration of the blocking of the origin twice as long as the duration of blocked DnaA synthesis, as reported in [58], did not alter these results. Therefore, also taking into account that the duration of the blocked period varies with the growth rate [58] does not prevent premature reinitiation events at intermediate growth rates (Fig. 2.6b). In this regime, another mechanism is needed. In the next chapter 3, we will present such a mechanism.

2.5. GROWTH-RATE DEPENDENCE OF THE CELL CYCLE IN THE AIT MODEL

Experiments show that while the cell volume increases strongly with the growth rate, the initiation volume per origin v^* varies only little, i.e. within a ~ 50% range over a tenfold change in the growth rate [22, 40, 41]. Here, we discuss how the key cell cycle parameters—the initiation time and volume, and the volume at birth and division—vary with the growth rate λ in the AIT model.

The initiation time is given by $t^* = \tau_d - \tau_{cc}$, where $\tau_d = \ln(2)/\lambda$ is the cell division time and τ_{cc} is the constant time between initiation and division. The volume at birth, V_b, and the initiation volume V^* are related via $V^* = V_b e^{\lambda t^*}$, and the volume at division V_d is simply twice the birth volume V_b. The central quantity is thus the total initiation volume V^*, or its value per origin $v^* = V^*/n_{ori}$, with n^*_{ori} the number of origins at initiation: from this and the initiation time, V_b and V_d follow.

To obtain the initiation volume, we exploit that at the moment of replication initiation the free initiator concentration $[p]$ equals the dissociation constant for binding the origin: $[p] = K_D^{ori}$. For a given total intiator concentration $[p]_T$, we can then combine $[p] = K_D^{ori}$ with Eq. 2.30 to obtain the total titration site concentration $[s]_T$. The latter is given by $[s]_T = n_s/v^*$, where n_s is the known number of titration sites per origin. Hence, for a given $[p]_T$ we can obtain the initiation volume v^* from Eq. 2.30.

To understand how the initiation volume v^* depends on the total initiator concentration $[p]_T$ it is illuminating to consider the limit in which the binding of the initiator proteins to the titration sites is very strong. We connect the critical number of initiators per origin n^* to the initiation volume per origin v^* via the total concentration of initiators at the moment of initiation:

$$[p]^*_T = \frac{N^*_p}{V^*} = \frac{N^*_p/n^*_{ori}}{V^*/n^*_{ori}} = \frac{n^*}{v^*} \tag{2.21}$$

where N^*_p is the total number of initiators at initiation. Hence,

$$v^* = \frac{n^*}{[p]^*_T} \tag{2.22}$$

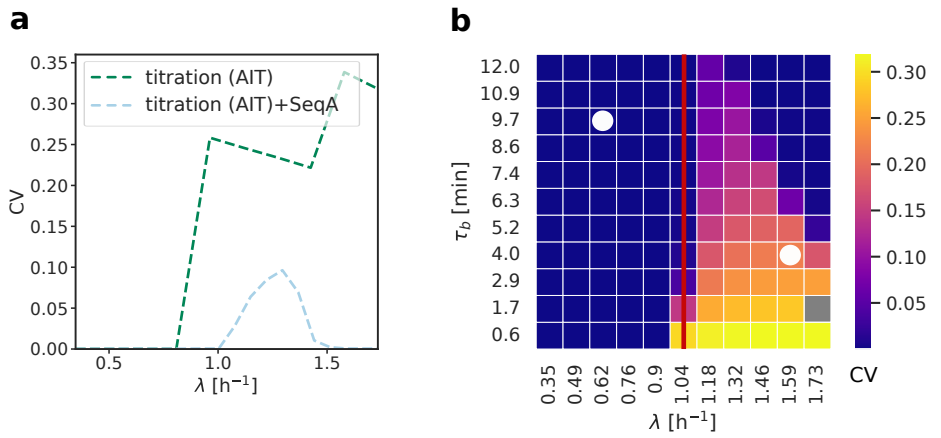


Figure 2.6: Transiently blocking DnaA synthesis via SeqA can prevent reinitiation events in the deterministic AIT model at high, but not at intermediate growth rates. (a) The coefficient of variation CV = σ/μ with the standard deviation σ and the average initiation volume $\mu = \langle v^* \rangle$ as a function of the growth rate for different models. Even in the absence of biochemical noise in DnaA synthesis, the titration model gives rise to a very high CV at high growth rates, due to premature reinitiation (see Fig. 2.3b for time traces). Adding SeqA to the titration model can reduce the CV at high, but not at intermediate growth rates (see Fig. 2.5b for time traces). All models include an eclipse period of about 10 minutes following replication initiation to prevent immediate reinitiation [58–60]. In the titration+SeqA model, DnaA synthesis is additionally blocked for 10 minutes. (b) As the duration of the blocked DnaA synthesis period was reported to vary with the growth rate [58], we here show the CV of the initiation volume for a broad range of growth rates λ and blocked periods τ_b, during which DnaA synthesis is suppressed by SeqA. The system is deterministic and the titration sites are homogenously distributed. The white dots indicate the experimentally reported durations of blocked DnaA synthesis at the respective growth rates [58] and the red line marks the crossover from the regime of non-overlapping replication forks to the regime of overlapping forks, i.e. the growth rate $\lambda^* = \ln(2)/T_C \approx 1.04\text{h}^{-1}$ at which the doubling time τ_d becomes shorter than the C-period of $T_C = 40$ min. Reinitiation events occur only in the overlapping fork regime where the growth rate is larger than λ^*. The longer the blocked period, the smaller becomes the CV, especially at high growth rates. However, for none of the blocked periods shown, reinitiation does not occur at any growth rate. In particular, for the experimentally measured duration of the blocked period at high growth rates (white dot on the right), premature reinitiation events happen, raising the CV above that measured experimentally, even though the model is deterministic. Hence, in the biologically relevant regime premature reinitiation still occurs even when the effect of SeqA to block DnaA synthesis is added to the titration mechanism.

In the limit of very tight binding, the critical number of initiators per origin n^* is set by the fixed number of titration sites per origin, $n^* \approx n_s$, and thus is constant when a new round of replication is initiated in the non-overlapping replication fork regime. Now we see that if the total concentration is maintained approximately constant in time, the initiation volume is also constant and the replication cycle becomes stable. Furthermore, the total concentration could be maintained approximately constant in time for a given growth rate, but vary as a function of the growth rate λ. Then, the growth-rate dependence of the total concentration directly translates into a growth-rate dependence of the initiation volume:

$$v^*(\lambda) = \frac{n_s}{[p]_T(\lambda)} \tag{2.23}$$

Experiments indicate that both the initiation volume and the total DnaA concentration vary by about 50% over a tenfold change in the growth rate, yet in an opposite, anti-correlated fashion [41]. This is consistent with Eq. 2.23. Yet, how the total concentration varies by only 50% over this range in growth rates remains unclear. As we have seen in section 2.2, in the AIT model titration interferes with negative autoregulation such that the initiator is effectively constitutively expressed. Experiments indicate that while the concentration of negatively autoregulated proteins is relatively independent of the growth rate, the concentration of constitutively expressed proteins typically decreases linearly with the growth rate [65, 66]; Eq. 2.23 would then predict that the initiation volume increases with the growth rate. How the total DnaA concentration and hence the initiation volume are kept within a 50% range over the tenfold change in the growth rate thus remains an open question.

2.6. ADDER CORRELATIONS IN THE AIT MODEL

We have shown that a volume-dependent production rate of DnaA in combination with homogeneously distributed titration sites yields stable cell cycles at low growth rates. But does the AIT model also give rise to the experimentally observed adder correlations in the initiation volume?

To test this, we again model the change in the number of DnaA proteins according to equation 2.17 plus a noise term $\xi_D(t)$ accounting for noise in gene expression

$$\frac{dN_D^T}{dt} = \frac{\tilde{\phi}_p^0 \lambda V}{1 + \left(\frac{[D]_T}{K_D^p}\right)^n} + \xi_D(t). \tag{2.24}$$

The noise is modelled as Gaussian white noise, $\langle \xi_D(t)\xi_D(t')\rangle = 2D_D\delta(t-t')$, with the noise strength D_D chosen to match the measured variance in the initiation volume of $CV = 0.1$ [22] (see Table 2.1).

Figure 2.7 shows that the AIT model maintains a stable initiation volume in the presence of gene expression noise. The fluctuations in the total concentration result in fluctuations in the initiation volume. The volume dependence of the production rate in equation 2.24 ensures however that the total concentration regresses back to a stable fixed point as derived in section 2.1.2. Therefore, also the initiation volume regresses back to the initiation volume derived in section 2.5.

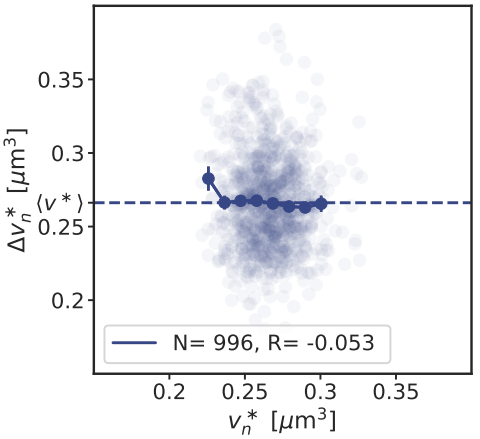


Figure 2.7: Gene expression noise in the DnaA concentration gives rise to adder correlations in the initiation volume in the AIT model The added volume per origin between successive initiation events, $\Delta v_{\mathrm{n}}^* = 2\, v_{\mathrm{n+1}}^* - v_{\mathrm{n}}^*$, is independent of the initiation volume v_{n}^* per origin and on average equal to the average initiation volume, $\langle \Delta v^* \rangle = \langle v^* \rangle$, as expected for an initiation volume adder. The dark blue line shows the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The doubling time is $\tau_{\mathrm{d}} = 2$ h and the number of data points N and the Pearson correlation coefficient R are indicated.

Plotting the volume that is added between two consecutive initiation events, $\Delta v_{\mathrm{n}}^* = 2\, v_{\mathrm{n+1}}^* - v_{\mathrm{n}}^*$, as a function of the initiation volume v_{n}^*, reveals that the AIT model exhibits adder correlations in the initiation volume. This observation can be understood intuitively. The key feature of an adder is that the added volume is on average constant, independent of the last initiation volume [5, 6, 23]. As explained in Appendix 2.A, in the initiator accumulation models replication is initiated when a critical number of initiators per origin have been accumulated. In the AIT model, this critical number is set by the fixed number of titration sites per chromosome. Replication is thus initiated when the number of proteins that have been produced since the last initiation event equals the number of titration sites, irrespective of the magnitude of the last initiation volume. As the number of initiators is accumulated proportionally to the volume of the cell, a fixed amount of accumulated proteins maps directly to a fixed volume that needs to be accumulated (Eq. 2.24). The added volume since the last initiation event is thus, on average, always the same, irrespective of the last initiation volume—the hallmark of an adder.

2.7. DISCUSSION

A titration-based mechanism ensures stable cell cycles at low growth rates, but breaks down at sufficiently high growth rates: Premature reinitiation events cause a dramatic rise of the coefficient of variation (CV) in the initiation volume even in the absence of any biochemical noise. This finding is robust, because it relies on the different scaling of the protein synthesis and the titration-site formation rate with the growth rate. The transient suppression of DnaA synthesis by SeqA after replication initiation can pre-

vent these premature reinitiation events, but only at high growth rates: at intermediate growth rates, we still observed re-initiation events that are not observed experimentally (Fig. 2.5a). From these findings we conclude that a titration-based mechanism is not sufficient to explain how replication initiation is regulated on a molecular level. In the next chapter, we therefore study how protein activation can ensure stable replication cycles.

Our model makes the simplifying assumption that the cell divides a constant time τ_{cc} after replication initiation, independent of the growth rate. Experiments indicate, however, that this is an oversimplification [22–24, 27–29, 67] and that cell division is more loosely coupled to replication initiation [23, 24]. Importantly, our results on replication initiation control are robust to the assumption of a constant τ_{cc}, because on average cell division does not change the densities of the components. Indeed, while this assumption will affect the correlations between the cell volume at birth and the initiation volume, it does not change the correlations between the initiation volume and the volume added until the next initiation event (Fig. 2.9).

Our modelling predicts that negative autoregulation does not play a direct role in replication initiation. This is supported by recent experiments, which show that the average initiation volume and precision of replication initiation are only weakly affected in strains with constitutive *dnaA* expression [68]. Following Hansen et al. [64], we believe that negative autoregulation only plays an indirect role, by setting the growth-rate dependence of the DnaA concentration. Experiments have revealed that the total DnaA concentration varies with the growth rate, anticorrelating with the initiation volume [41]. However, the variation of both the total DnaA concentration and the initiation volume is rather weak, i.e. about 50% over a tenfold change of the growth rate [41]. It seems likely that negative autoregulation is crucial for constraining the growth-rate dependence of the total DnaA concentration [65, 66] and hence the initiation volume [22, 40]. How negative autoregulation with a differential sensitivity of the DnaA promoter to DnaA-ATP and DnaA-ADP [69, 70] and titration conspire to shape the growth-rate dependence of the DnaA concentration and the initiation volume, we leave for future work.

APPENDIX

2.A. STABILITY OF THE INITIATOR ACCUMULATION MODEL

In the initiator accumulation model, an initiator protein accumulates over the course of the cell cycle and replication is initiated when a threshold amount per origin is attained. Here, we study a very general implementation of this model to demonstrate that a volume-dependent protein production rate is essential to obtain stable cell cycles with the initiator accumulation model.

The bacterium *E. coli* must initiate replication once per division cycle in order to be able to distribute two copies of the chromosome in the two daughter cells. In good nutrient conditions, *E. coli* grows exponentially with a growth rate λ such that the volume is given by

$$V(t) = V_b\, e^{\lambda t} \tag{2.25}$$

The growth rate λ can fluctuate due to noise, but on average cells double their entire volume after the cell-doubling time $\langle \tau_d \rangle = \ln(2)/\langle \lambda \rangle$. As in *E. coli* replication is initiated synchronously at all origins also in the overlapping fork regime at high growth rates, we can define the inter-initiation time τ_{ii} as the time between two consecutive initiation events. Any molecular mechanism for replication initiation must ensure that the average inter-initiation time $\langle \tau_{ii} \rangle$ equals the average cell-doubling time $\langle \tau_d \rangle$. If that is not the case, the average origin density, $\langle \rho \rangle = \langle n_{ori} \rangle / \langle V \rangle$, does not remain constant over the course of several generations.

In the initiator accumulation models, an initiator protein is accumulated up to a fixed threshold per origin at which replication is initiated. In the AIT model in section 2.2 we will show that a constant number of high-affinity binding sites for the initiator on the chromosome can ensure such a constant number threshold per origin. Given that this threshold per origin is fixed, the time from one initiation event to the next is determined by how fast the initiator proteins are synthesized. In contrast to the recently proposed growing cell model presented in section 2.1, in an arguably more traditional model of gene expression, the protein production rate of a constitutively expressed gene is given by a constant basal α rate times the gene copy number g [65, 71–74]:

$$\frac{dN}{dt} = \alpha\, g \tag{2.26}$$

Assuming again that the gene is located at the origin, the number of genes g equals the number of origins n_{ori}. Thus, a constant number of initiators per origin $\Delta n = \Delta N / n_{ori}$ is accumulated in a time interval Δt:

$$\Delta n = \alpha\, \Delta t \tag{2.27}$$

As in the initiator accumulation model replication is initiated after a constant amount of proteins per origin Δn^* has been accumulated, we find that the inter-initiation time τ_{ii} in this model is given by

$$\tau_{ii} = \frac{\Delta n^*}{\alpha} \tag{2.28}$$

As the number of initiators that need to be accumulated per origin Δn^* is constant and the basal rate does not explicitly depend on the volume in the traditional model of gene

expression, the inter-initiation time thus is constant. If the basal production rate is not set such that the average replication period exactly equals the doubling time of the cell, $\tau_{ii} = \tau_d$, this system gives rise to an instability in the chromosome density.

We verify this prediction by performing simulations. The cell volume and the number of initiators are evolved according to equations 2.25 and 2.26 and replication is initiated when the number of initiators per origin $n(t) = N(t)/n_{ori}(t)$ equals the critical number per origin n^*. At initiation, the number of origins doubles and the number of initiators per origin in generation i right after initiation thus becomes $n_i = n^*/2$. The number of initiators per origin that needs to be accumulated until the next initiation event is therefore $\Delta n_i = n^* - n_i = n^*/2$. Following the Cooper-Helmstetter model [20], the cell divides a constant cycling time τ_{cc} after replication initiation. In Fig. 2.8a, the replication period τ_{ii} is chosen to be shorter than the doubling time τ_d of the cell. As every replication initiation event triggers a cell division event, the division period τ_{div} equals the replication period $\tau_{div} = \tau_{ii} < \tau_d$. As the replication period and thus the division period is smaller than the doubling time of the cell, the volume of the cell decreases over several generations while the gene density increases. We emphasise that even when τ_{ii} is chosen to be equal to τ_d, any noise, even that coming from the finite machine-precision, will cause the gene density to eventually become unstable. To show that this instability does not depend on the choice of the division control, we also study another model in which cell division is triggered at a fixed division volume V_d instead of a fixed time τ_{cc} after replication initiation. Because in this model the division cycle is independent of the replication cycle and division is triggered at a fixed division volume V_d, the division cycle naturally remains stable (Fig. 2.8b). The replication cycle is however not coupled to this division cycle, because the synthesis rate of the accumulator and the replication threshold are constant, i.e. do not depend on the volume. Replication is therefore initiated at a period that is again shorter than the doubling time of the cell $\tau_{ii} < \tau_d$. Also in this scenario, the gene density increases over the course of several generations.

The initiator accumulation model becomes stable by introducing a volume-dependent production rate, which couples the replication cycle to the cell division cycle. Taking the production rate to be

$$\frac{dN}{dt} = \alpha V^{\gamma} \tag{2.29}$$

where γ is an exponent quantifying the strength of the volume dependence of the production rate. For $\gamma = 0$ the production rate becomes independent of the volume. We show that for the exponents $\gamma = 1$ (Fig. 2.8c) and $\gamma = 0.5$ (Fig. 2.8d) the system recovers from an initial perturbation and becomes stable. The relaxation time increases with decreasing volume dependence.

We have demonstrated that the initiator accumulation model requires a volume dependent production rate. In the traditional model of gene expression, the production rate of an unregulated protein is proportional to the gene copy number times a constant production rate [65, 71–74] and thus cannot fulfill this requirement (it corresponds to $\gamma = 0$). In the previous section, we showed that in the growing cell model, which we use throughout this work, the production rate is directly proportional to the volume of the cell, thus corresponding to the scenario $\gamma = 1$. Therefore, even a constitutively expressed initiator protein can give stable replication cycles in the initiator accumulation model,

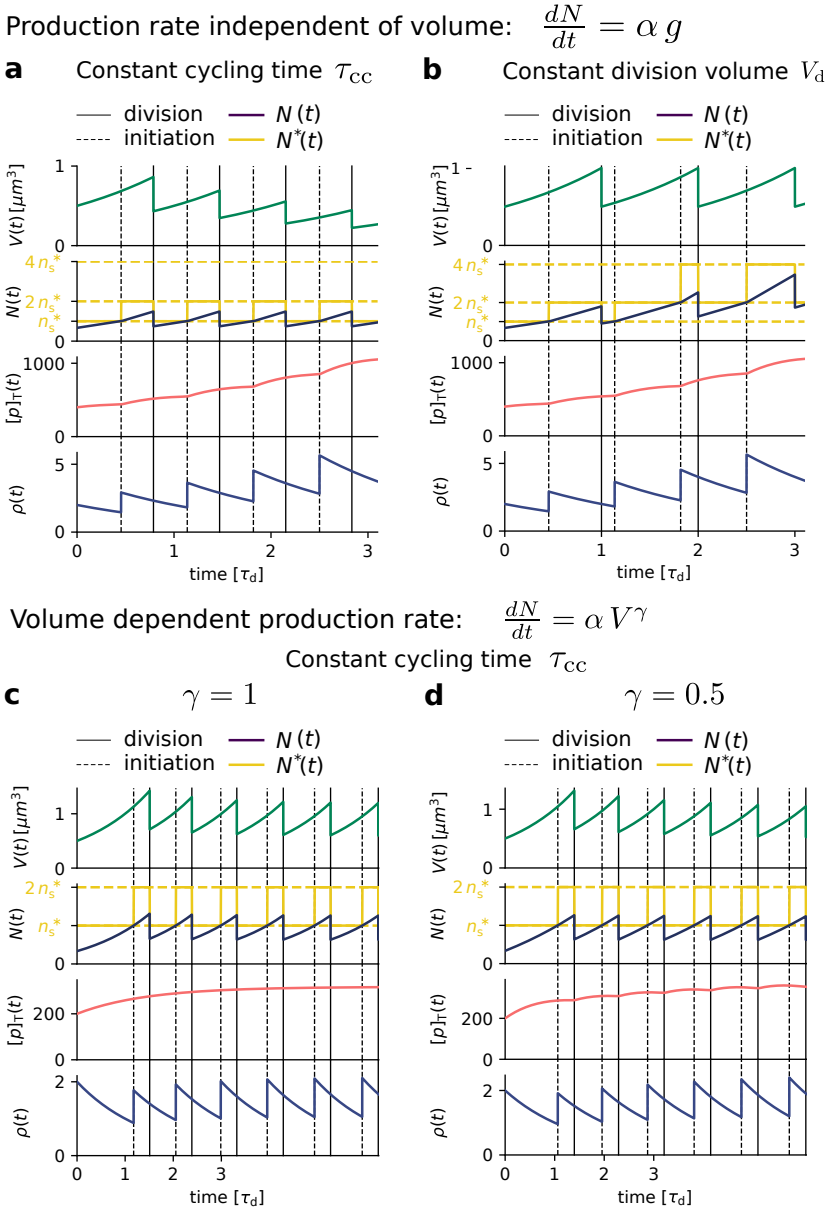


Figure 2.8: For the initiator accumulation model to become stable, the production rate needs to depend on the growth rate of the cell (a, b, c, d) The volume $V(t)$ (according to equation 2.25), the number of proteins $N(t)$ together with the critical threshold $N^* = n^* n_{\mathrm{ori}}$, the total concentration $[p]_T = N(t)/V(t)$, and the origin density $\rho(t) = n_{\mathrm{ori}}(t)/V(t)$ as a function of time. (caption continues on the next page)

provided the ribosomes are limiting gene expression and the synthesis rate is proportional to the cell volume as in the growing cell model.

2.B. BIOLOGICAL PARAMETERS OF THE AIT MODEL

Here, we discuss the experimentally found parameters and compare them to the ones used in the simulations of the AIT model. All parameters of the AIT model can be found in Table 2.1.

The protein DnaA in *E. coli* is generally referred to as the initiator protein, as its ATP-bound form is required to bind to the origin for initiating replication [43]. Both forms of the protein DnaA, ATP-DnaA and ADP-DnaA, have strong affinity for an asymmetric 9 bp consensus sequence on the DNA, the DnaA box [43]. In the replication origin region of *E. coli* several DnaA boxes are present, including R1-R4 and M. [54]. In total, 308 DnaA boxes of the stringent definition (5'- TT $^{A}/_{T}$ TNCACA) have been found on the *E. coli* genome [54]. The dissociation constant of DnaA binding to the DnaA boxes on the DNA lies in the range of $K_D^s = 1 - 50$ nM, depending on the flanking sequences [46]. While for some DnaA boxes, the binding was non-specific $K_D^s \geq 200$ nM, the highest affinity was found for the DnaA boxes R1 and R4 in the origin with $K_D^s = 1$ nM. In *E. coli*, the approximately three hundred 9-mer DnaA boxes are randomly distributed on the *E. coli* chromosome [14, 54]. The *dnaA* gene is regulated by two promoters, *dnaAp1* and *dnaAp2*, with a DnaA box located between them. *dnaAp2* is the stronger promoter and contributes 60–80 % of the dnaA transcripts [69]. Both ATP-DnaA and ADP-DnaA bind cooperatively to these two promoters, but the repression via ATP-DnaA is more efficient [69]. As there are five binding sites for DnaA in the promoter region [69], we choose a Hill coefficient of $n = 5$ in the simulations.

In the AIT model we used $n_s = 300$ titration sites per chromosome with a dissociation constant of $K_D^s = 1$ nM (Table 2.1). We approximate the experimentally reported random distribution of titration sites on the chromosome [14, 54] by a homogeneous distribution. At a concentration of ATP-DnaA of approximately $[D]_{ATP} = 100$ nM, the expression of DnaA was reduced by 50 % [69]. Therefore, we used in the AIT model for the promoter a dissociation constant of $K_D^p = 100$ nM. The dissociation constant of DnaA for the

(a, b) The protein is produced at a constant rate times the number of genes (according to equation 2.26). This gives rise to an unstable chromosome density independent of the division mechanism. (a) Cell division is triggered a constant cycling time τ_{cc} after replication initiation. The time between consecutive replication events τ_{ii} is given by equation 2.28 and is shorter than the doubling time of the cell. Thus, the origin density increases in time. (b) Cell division is triggered at a fixed division volume $V_d = 1\ \mu m^3$ and is thus independent of the replication cycle. Again, the replication period τ_{ii} is shorter than the doubling time of the cell and the origin density increases in time. (c, d) Now, the initiator protein is produced proportional to the volume of the cell according to equation 2.29 with an exponent γ. Cell division is triggered a fixed time τ_{cc} after replication initiation. For any positive exponent that is larger $\gamma > 0$, the gene density stabilizes after an initial perturbation. (c) For an exponent of $\gamma = 1$, the gene density relaxes to a constant average density after an initial perturbation. The total initiator concentration becomes perfectly constant in time. (d) For an exponent of $\gamma = 0.5$ the relaxation time increases and the total concentration oscillates weakly over the course of the cell cycle. The system relaxes to a stable gene density and initiator concentration. The parameters of all simulations are $\tau_{cc} = 1$ h, $\alpha = 110\ h^{-1}$, $\tau_d = 2$ h, $n^* = 300$.

Parameter	Name	Value	Motivation
ϕ_0	gene allocation fraction	10^{-3}	set to match initiation volume reported in [40]
K_D^p [μm^{-3}]	dissociation constant initiator promoter	200	[69]
n	Hill coefficient initiator	5	[69]
n_s	number of titration sites per chromosome	300	[14, 54]
K_D^{ori} [μm^{-3}]	dissociation constant origin	20	[46]
K_D^s [μm^{-3}]	dissociation constant titration sites	1	[46]
ρ [μm^{-3}]	number density	10^6	[75]
D_D	noise strength DnaA	100	set to match CV from [22]
T_C [h]	C-period	2/3	[20]
T_D [h]	D-period	1/3	[20]
D_D	noise strength DnaA	100	set to match CV from [22]
λ [h^{-1}]	growth rate	0.35-1.73	[22, 40]

Table 2.1: Parameters used in the AIT model One molecule per cubic micrometer corresponds to approximately one nM (1 μm^{-3} = 1.67 nM).

origin was chosen to be K_D^{ori} = 20 nM, reflecting the combination of high and intermediate affinity of the titration sites required to be filled by ATP-DnaA in order to initiate replication.

2.C. QUASI-EQUILIBRIUM ASSUMPTION FOR DNAA-TITRATION SITE BINDING

In the main text, we assumed that the binding and unbinding of the initiator proteins to the titration sites is well described by a quasi-steady-state. Here we show that the binding and unbinding dynamics are relatively fast compared to the doubling time of the cell, such that this assumption is well justified. It seems reasonable to assume that DnaA finds its target sites in a way that is similar to that of other transcription factors, such as the lac repressor whose binding dynamics has been well characterized [55]. These transcription factors move by facilitated diffusion, i.e. combining 3D with 1D diffusion along the DNA. Elf et al. [55] have measured that the effective diffusion constant of transcription factors in *E. coli* is of the order of $D_{eff} = 0.4\,\mu$m^2/s. Assuming the binding rate is diffusion-limited, the binding rate is given by $k_{on} = 4\pi\sigma D_{eff}$. For an estimated cross section in the order of $\sigma \approx 10^{-2}\,\mu$m [57], the binding rate therefore becomes $k_{on} \approx 0.05\,\mu$m^3/s. The time

for a transcription factor to bind to its target site is given by one over the concentration of the transcription factor $[c]$ times the binding rate: $\tau_{on} = ([c] \times k_{on})^{-1}$. With a typical volume of an *E. coli* cell of $V = 1\ \mu m^3$, the search time of one transcription factor for finding its target site on the DNA should then be $\tau_{on} = k_{on}^{-1} \times V = 20$ s. This estimate compares well to the measured value of $\tau_{on} = 65 - 360$ s by Elf et al. [55]. The dissociation constant of DnaA binding to the DnaA boxes on the DNA is in the range of $K_D^s = 1 - 50$ nM [46]. Using $K_D^s = k_{off}/k_{on}$ allows us to estimate $k_{off} = K_D^s \times k_{on} \approx 0.015 - 0.8\ s^{-1}$. With an average concentration of the initiator protein DnaA in *E. coli* of $[D]_T \approx 400\ \mu m^{-3}$ [64], the correlation time for binding and unbinding then becomes $\tau = 1/(k_{on}[D]_T + k_{off}) \approx 0.16$ s. This is much faster than the timescale at which the volume changes, set by the growth rate. Recent FRAP experiments combined with single molecule tracking experiments show that DnaA rapidly moves between chromosomal binding sites and has a residence time of less than a second [56]. Thus, the quasi-equilibrium approximation of the initiator binding to the titration sites we make is well justified.

As binding and unbinding dynamics of the initiator protein to the titration sites are relatively fast, we can assume for simplicity a quasi-equilibrium state of the concentration of free initiator proteins $[p] = K_D^s [sp]/[s]$ with the dissociation constant K_D^s. At every given total titration site concentration $[s]_T = [s] + [sp]$ and total initiator protein concentration $[p]_T = [p] + [ps]$, the average free initiator protein concentration $[p]$ is given by the quadratic equation

$$[p]([s]_T, [p]_T) = [p]_T - \frac{K_D^s + [s]_T + [p]_T}{2} + \frac{\sqrt{(K_D^s + [s]_T + [p]_T)^2 - 4[s]_T [p]_T}}{2} \tag{2.30}$$

We use this expression in the main text to calculate at every given total titration site concentration and total initiator concentration in a cell the concentration of initiators freely diffusing in the cytoplasm. As can be seen in Figure 2.3a, as long as there are more titration sites than proteins in the cell, the free DnaA concentration remains low. When the total number of DnaA proteins exceeds the total number of titration sites, the free concentration quickly rises and replication is initiated when the critical free initiator concentration K_D^{ori} is attained. The fixed number of titration sites per chromosome therefore sets the critical number of initiators that need to be accumulated in order to reach the critical free initiator concentration K_D^{ori} in the cytoplasm.

2.D. LOOSENING THE COUPLING BETWEEN REPLICATION INITIATION AND DIVISION

According to experiments at the population level, the time from the initiation of replication until cell division, the cycling time τ_{cc}, is approximately constant [40]. In the main text, we therefore assumed that τ_{cc} is constant. This allowed us to study the cell cycle entirely from the perspective of the replication cycle. Experiments show, however, that this is an oversimplification [22–24, 27–29, 67] and that cell division is more loosely coupled to the replication cycle [23, 24, 26–28, 45]. Of particular interest are two recent single-cell studies, by Si et al. [23] and Witz et al. [24], respectively. Both studies indicate that

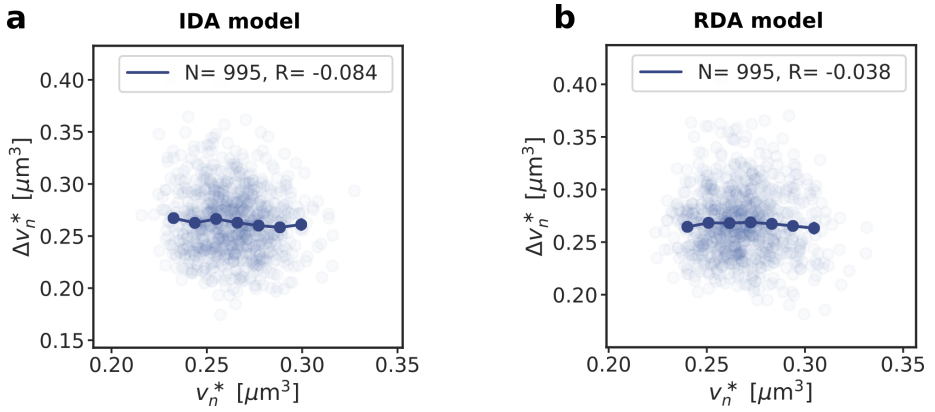


Figure 2.9: The adder correlations in the initiation volume in the AIT model are robust to a more loose coupling between the division and the replication cycle. (a, b) The added volume per origin between consecutive replication initiation events, $\Delta v_{\mathrm{n}}^* = 2\, v_{\mathrm{n+1}}^* - v_{\mathrm{n}}^*$, as a function of the initiation volume v_{n}^*. The dark blue lines show the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The number of data points N and the Pearson correlation coefficient R are indicated. The doubling time in all plots is $\tau_{\mathrm{d}} = 2$ h. In the AIT model, gene expression noise in the initiator protein gives rise to adder correlations in the initiation volume even if the division cycle is coupled more loosely to replication initiation: While in the IDA model, cell division is triggered completely independently via a separate division adder (a), in the RDA model division is triggered when an on average constant volume has been added from replication initiation to cell division (b) (compare to Fig. 2.7).

the cell cycle consists of two adders, a DNA replication adder and a cell-division adder. Both studies also agree on the nature of the replication adder: the data of both studies unequivocally show that the added volume between successive initiation events is independent of the initiation volume, as our model also predicts (Fig. 2.7). However, the authors of these two studies come to different conclusions concerning the nature of the division adder [23, 24, 26, 45]. By employing a statistical framework with stochastic simulations, Witz et al. conclude that the second adder concerns the added volume between replication initiation and cell division [24]. Si et al. showed that by inducing oscillatory perturbations in the concentration of DnaA, the adder correlations in the replication initiation volume can be destroyed, while the adder on the level of cell division remains intact; they conclude that the division adder concerns the added volume from birth to division and suggest that cell division is controlled by a separate molecular mechanism [23].

We emphasise that the central question of our manuscript is how replication initiation is regulated—not how cell division is controlled, nor how this is coupled to replication initiation. Naturally, our assumption that the time τ_{cc} between cell division and replication initiation is constant will affect the correlations between the initiation volume and cell division, since this directly couples division to replication initiation. The pertinent question is, however, whether the adder correlations in the initiation volume remain robust to this assumption.

To address these questions, we compare the results of our models of the main text in which replication is coupled to cell division via a constant time τ_{cc} between these two events, to the predictions of two other models in which replication is coupled to cell division either via the model of Si et al. or that of Witz et al. These two alternative models contain the same molecular, mechanistic description of replication initiation as our models presented in the main text. And like our models, they describe cell division and its coupling to replication initiation phenomenologically. The models differ, however, in the nature and strength of this coupling between cell division and replication. While in our model cell division is tightly coupled to replication initiation, with a constant τ_{cc} between these two events, in the other two models the coupling is more loose. The first of these two models is based on that of Si *et al.* [23, 45], which, following [26], we call the Independent Double Adder (IDA) model. In this model, the cell division cycle is completely independent of the replication cycle. Cells divide when a Gaussian distributed volume Δ_{IDA} with mean $\mu_{IDA} = \langle V_b \rangle$ and a standard deviation σ_{LDA} (with coefficient of variation $CV_{IDA} = \sigma_{IDA}/\mu_{IDA} = 0.1$) has been added to the birth volume, independent of the cell size at birth. As the replication and the division cycle are not coupled in this model, it could happen that a cell attempts to divide before replication has finished. To prevent this biologically unrealistic scenario from happening, we impose in the simulations that replication must be finished before a cell can divide. This scenario however only happens extremely rarely. In the second model, based on that of Witz et al. and called the Replication Double Adder (RDA) model [24, 26], cells divide when a Gaussian distributed volume Δ_{RDA} with mean $\mu_{RDA} = \langle V_d \rangle - \langle v^* \rangle = \langle v^* \rangle \left(\exp(\lambda \tau_{cc}) - 1\right)$ and a standard deviation σ (with coefficient of variation $CV_{RDA} = \sigma_{RDA}/\mu_{RDA} = 0.1$) has been added since replication initiation, independent of the initiation volume. In this model, the coupling between replication and division is thus of intermediate strength.

At the mean-field level, all results on the initiation volume should be independent of the type of division control, as the initiation volume in the accumulation and in the switch model is determined by concentrations of proteins which do not change upon cell division. We show below that, in addition to these mean-field observations, the adder correlations of the initiation volume obtained in the AIT model (and the LD model, see next chapter and Fig. 3.12) remain unchanged when cell division is controlled by either an independent adder running from cell birth to division, as in the IDA model, or by an adder running from replication initiation to cell division, as in the RDA model.

We re-evaluate the correlations between consecutive cell cycles in the AIT model (for LD model see Fig. 3.12) in the case where cell division is controlled according to the IDA or RDA model, as described above. We find that while correlations between the initiation volume and the birth volume are different, as expected, the previously obtained correlations between consecutive initiation volumes per origin remain unchanged, for both models. Specifically, in the AIT model, gene expression noise in the initiator protein production rate still gives rise to adder correlations in the initiation volume per origin, both in the IDA and RDA model (Fig. 2.9a and b).

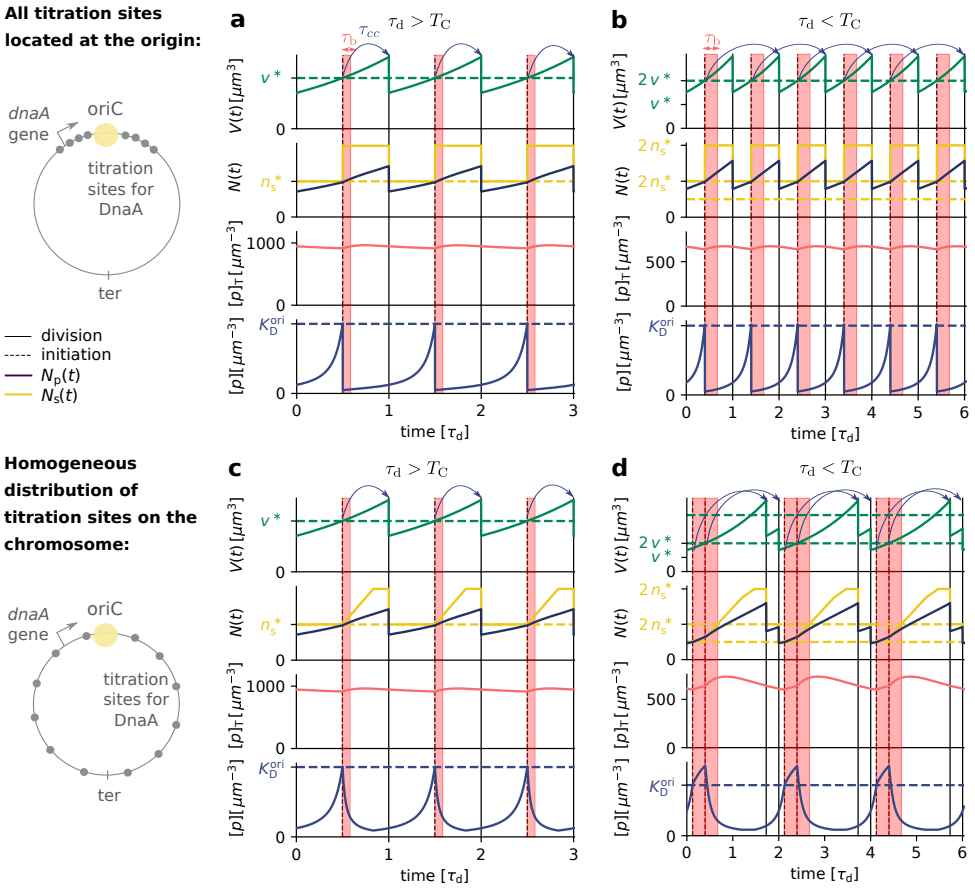
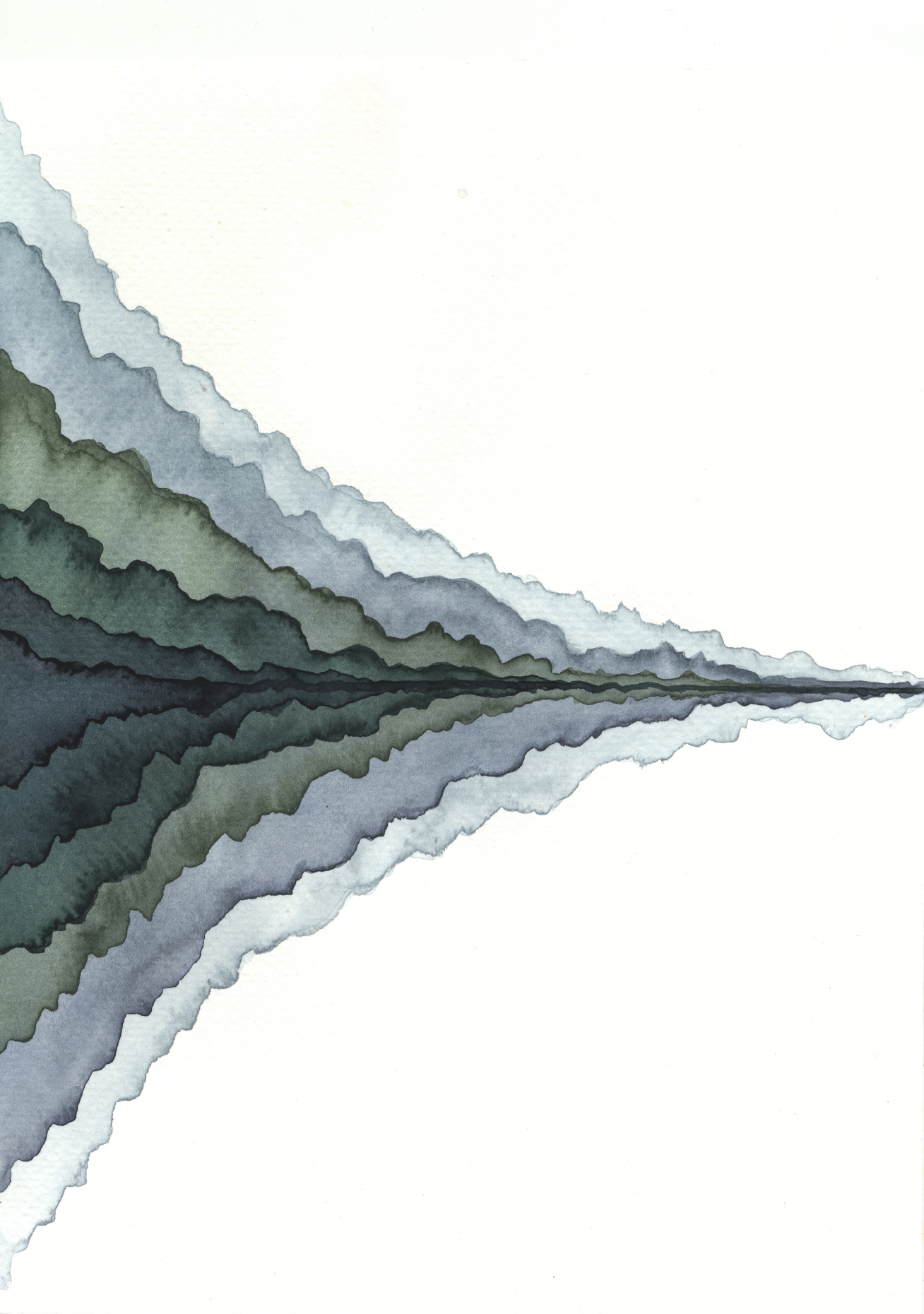


Figure 2.10: A homogeneous titration site distribution on the chromosome in the AIT model causes reinitiation events at high growth rates (a, b, c, d): The volume $V(t)$, the number of initiator proteins $N_p(t)$ (black line) and titration sites $N_s(t)$ (yellow line), the total concentration of initiator proteins $[p]_T(t)$ together with the dissociation constant of the regulator K_D^r (dotted red line), and the concentration of initiator proteins in the cytoplasm $[p](t)$ as a function of time (in units of the doubling time of the cell τ_d) for $\tau_d = 2$ h (a, c) and $\tau_d = 35$ min (b, d), respectively. When the number of initiator proteins per origin $n_p(t)$ exceeds the number of titration sites per origin n_s (yellow dashed line), the free concentration $[p](t)$ rapidly rises to reach the threshold concentration K_D^{ori} (blue dashed line) for initiating a new round of replication. The blue arrows indicate that the cell divides a constant cycling time τ_{cc} after replication initiation. During the blocked period τ_b(red shaded area), no new round of replication can be initiated. (caption continues on the next page)

(a, b) If all titration sites are located at the origin, the free initiator concentration $[p](t)$ decreases immediately after replication is initiated, independent of whether the doubling time of the cell τ_d is smaller (a) or larger (b) than the time T_C to replicate the entire chromosome. (c) When the titration sites are distributed homogeneously along the chromosome, the free initiator concentration decreases during the entire replication time T_C at low growth rates. As the time to produce new titration sites is still faster than the time to synthesize new initiator proteins, we obtain regular stable cell cycles in this regime. (d) When the doubling time is however smaller than the time to replicate the entire chromosome, $\tau_d < T_C$, newly replicated titration sites are being filled faster with new proteins than they are replicated. After a short blocked period τ_b, replication is reinitiated. As a result, each long (sub)cycle is followed by a very short one, together forming the cell cycle. Moreover, replication is not initiated at a constant volume per origin anymore, but oscillates over time. The appearance of premature reinitiation events suggests that replication initiation in *E. coli* can not fully be explained by a titration-based mechanism.

Artist impression of an initial perturbation that decays in time
Watercolor, 2022
Mareike Berger



3

AN ORIGIN-DENSITY SENSOR CAN GENERATE ADDER CORRELATIONS

Experiments indicate that E. coli controls replication initiation not only via titration, but also by switching between an inactive form of DnaA and an active form that induces replication initiation. We here present a model that is based on such an ultra-sensitive switch. Our model shows that at low growth rates the switch is predominantly controlled by activation of DnaA via lipids and deactivation via the chromosomal site datA, while at high growth rates DARS2 and RIDA become essential. Crucially, in our mean-field model DNA replication is initiated at a constant volume per origin, qualifying our model as a sizer. Yet, we show that in a stochastic version of the same model the inevitable fluctuations in the components that control the DnaA activation switch naturally give rise to the experimentally observed adder correlations.

The contents of this chapter have been published in Nature Communications **13**, 6556 (2022) [39].

Cell homeostasis requires that during the cell cycle, all components need to be copied on average once. This is particularly true for the most important component, which is present in extremely small copy numbers: the DNA. It needs to be copied exactly once during each cell cycle. *E. coli* initiates replication from a well-defined site on the origin, called the origin of replication, and experiments have revealed that *E. coli* starts a new round of replication initiation when its volume per origins reaches a well-defined threshold. Yet, how *E. coli* measures this critical volume is poorly understood.

Experiments indicate that *E. coli* controls replication initiation via two mechanisms: titration and activation. In the initiator titration models, the initiator protein DnaA can bind to titration sites located on the chromosome with a high affinity [43, 46]. Only when these sites have been filled up can the initiator protein bind to the origin and initiate replication [12, 14]. In the previous chapter we have shown that a titration-based mechanism is however insufficient to ensure stable cell cycles at all growth rates in *E. coli*. This shows that another mechanism is vital.

Interestingly, experiments have revealed that *E. coli* employs an additional mechanism, based on a switch of the protein DnaA between an active and an inactive form (Fig. 3.1b) [33, 43, 63, 76–78]. In *E. coli*, the initiator protein DnaA forms a tight complex with ATP or ADP, but only ATP–DnaA can initiate replication by forming a complex with the chromosomal replication origin (*oriC*) [17, 18, 79, 80]. While the total DnaA concentration is approximately constant at different growth rates [41, 64], the cellular level of ATP–DnaA oscillates over the course of the cell cycle, with a peak at the time of replication initiation [78, 81, 82]. It has been suggested that the oscillations in the fraction of ATP-DnaA in the cell are the key to understanding how replication is regulated in *E. coli*, but a quantitative description that is consistent with experiments is currently lacking [43, 44, 77, 83, 83–85].

Intriguingly, the level of ATP-DnaA is strictly regulated by multiple systems in the cell. DnaA is activated via acidic phospholipids in the cell membrane [86] and via two chromosomal regions called DnaA-Reactivation Sequence 1 (*DARS1*) and *DARS2* [77, 81], and deactivated via the chromosomal site *datA* in a process called *datA*-dependent DnaA-ATP Hydrolysis (DDAH) [63] and via a mechanism coupled to active DNA replication, called Regulatory Inactivation of DnaA (RIDA) [17, 78, 87, 88] (Fig. 3.1a). Deleting or modifying any of these systems can lead to untimely initiation, asynchrony of initiation, and changes in the initiation volume [44, 63, 89–93]. All systems thus appear important, raising the question not only why *E. coli* has evolved such an intricate set of mechanisms, but also how these conspire to generate robust replication cycles at all growth rates.

To dissect how these multiple mechanisms give rise to a stable cell cycle, we first study the Lipid-*DatA* (LD) model, which consists of only the acidic lipids and *datA* (Fig. 3.1b). This model reveals that the interplay between a constant rate of activation and a rate of deactivation that depends on the origin density gives rise to stable cell cycles. Yet, at higher growth rates these two reactions alone, based on the experimentally estimated rates of activation and deactivation, respectively, are not sufficient to generate large amplitude oscillations in the fraction of ATP-DnaA. Simulations of our Lipid-*DatA*-*DARS1/2*-RIDA (LDDR) model (Fig. 3.1b) show that in this regime, activation via *DARS2* and deactivation via RIDA become essential.

Importantly, in our mean-field switch models, DNA replication is initiated at a thresh-

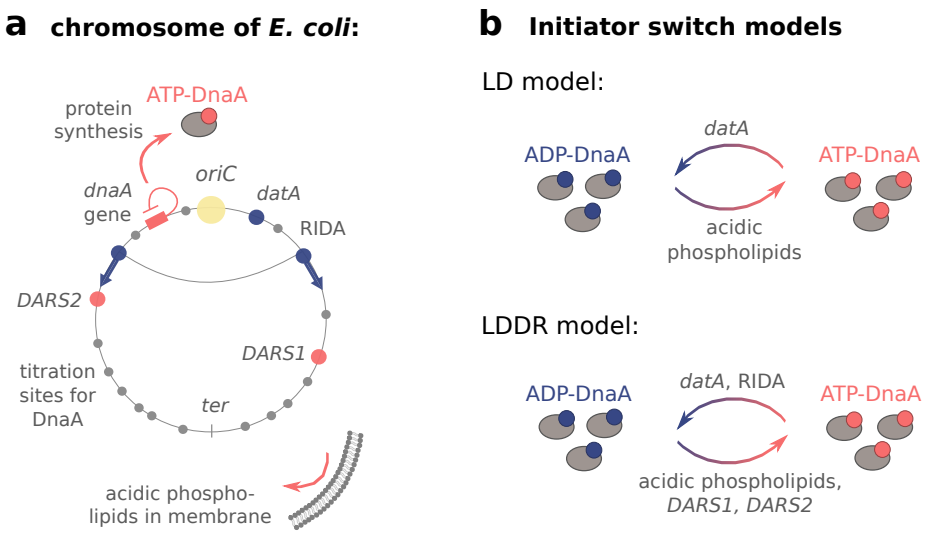


Figure 3.1: We present two initiator activation switch models for the regulation of DNA replication initiation in *E. coli*. (a) Schematic representation of an *E. coli* chromosome: Replication starts at the origin (*oriC*, yellow circle) and proceeds via two replication forks to the terminus (ter, grey bar). Replication is initiated by the ATP-bound form of the initiator protein DnaA. DnaA is activated via the acidic phospholipids in the cell membrane and via the two chromosomal sites *DARS1* and *DARS2*, and deactivated via the chromosomal site *datA* and via regulatory inactivation of DnaA (RIDA), a process coupled to active DNA replication. DnaA also has a high affinity for titration sites (grey circles) located on the DNA. (c) Scheme of the initiator switch models: In the LD model, ATP-DnaA is mainly activated via the acidic phospholipids and deactivated via the site *datA*. In the LDDR model, replication forks overlap and RIDA is the main deactivator in combination with the activators *DARS1* and *DARS2*.

old origin density and mechanistically they should arguably be qualified as a sizer. Yet, experiments have revealed that *E. coli* is an initiation adder, adding an on average constant volume between successive initiation event, independent of the magnitude of the initiation volume [23, 24]. Here, we show that a stochastic version of the switch model naturally gives rise to the experimentally observed adder correlations in the initiation volume [23, 24]. Fluctuations in the components that control the DnaA activation switch (lipids, HdA, Fis, IHF) are transmitted from mother to daughter cells and this generates mother-daughter correlations in the initiation volume that can explain the observed adder correlations [23].

3.1. AN ULTRA-SENSITIVE SWITCH GIVES RISE TO AN ORIGIN-DENSITY SENSOR

In the activation switch models, not the total number of DnaA is the key variable that controls replication initiation, but the concentration or fraction of DnaA that is bound to ATP [76, 85]. While DnaA has a high affinity for both ATP and ADP [43, 94], only ATP-DnaA can initiate replication at the origin [17, 18, 79]. The switch between these two states is controlled by several mechanisms, which, we will argue, play distinct roles in different growth-rate regimes.

We first focus on the regime of slow growth in which the replication forks are non-overlapping. RIDA, a mechanism promoting ATP hydrolysis in a replication-coupled manner, becomes active upon replication initiation, but, since there are no overlapping forks, is inactive *before* replication initiation [17]. The chromosomal locus *datA* can hydrolyze ATP-DnaA via DDAH and is crucial for repressing untimely initiation events (Fig. 3.1a) [63]. The two chromosomal DNA regions *DARS1* and *DARS2* can regenerate ATP-DnaA from ADP-DnaA [17, 44, 77]. The activating site *DARS2* is reported to be only active at high growth rates [77, 95, 96], and the activity of *DARS1* was reported to be ten times weaker than *DARS2* in vitro [77]. In addition to *DARS1/2*, both in vitro [86, 97–101] and in vivo [92, 102–104] experiments indicate that acidic phospholipids can rejuvenate DnaA by promoting the exchange of ADP for ATP. Moreover, as we show in section 3.3, for a switch-based system, activation by *DARS1/2* is not sufficient, while lipid-mediated activation of DnaA is vital to generate stable cell cycles. In summary, our modelling in combination with experiments indicates that at slow growth, the dominant DnaA cycle of the switch setting the initiation volume consists of activation by the phospholipids and deactivation via DDAH. This cycle forms the basis of the Lipid-*DatA* (LD) model.

Since the growing cell model [47] predicts that the total DnaA concentration is nearly constant in time while experiments show that it is nearly independent of the growth rate [41], we here make the simplifying assumption that the total DnaA concentration is strictly constant as a function of time and the growth rate. Later, we also study the effect of variations in the total DnaA concentration (section 3.6). Assuming a constant total DnaA concentration allows us to focus on the fraction $f = [D_{\mathrm{ATP}}]/[D]_{\mathrm{T}}$ of DnaA that is bound to ATP [69]. Exploiting that DnaA is predominantly bound to either ATP or ADP

[17], the change of the active fraction f in the LD model is given by

$$\frac{df}{dt} = \frac{d[D]_{\mathrm{ATP}}}{dt}\frac{1}{[D]_{\mathrm{T}}} \tag{3.1}$$

$$= \tilde{\alpha}_{\mathrm{l}}[l]\frac{1-f}{\tilde{K}_{\mathrm{D}}^{\mathrm{l}}+1-f} - \tilde{\beta}_{\mathrm{datA}}[n_{\mathrm{ori}}]\frac{f}{\tilde{K}_{\mathrm{D}}^{\mathrm{datA}}+f} + \lambda(1-f) \tag{3.2}$$

with the constant, re-normalized activation and deactivation rates $\tilde{\alpha}_{\mathrm{l}} = \alpha_{\mathrm{l}}/[D]_{\mathrm{T}}$ and $\tilde{\beta}_{\mathrm{datA}} = \beta_{\mathrm{datA}}/[D]_{\mathrm{T}}$ and the Michaelis-Menten constants $\tilde{K}_{\mathrm{D}}^{\mathrm{l}} = K_{\mathrm{D}}^{\mathrm{l}}/[D]_{\mathrm{T}}$ and $\tilde{K}_{\mathrm{D}}^{\mathrm{datA}} = K_{\mathrm{D}}^{\mathrm{datA}}/[D]_{\mathrm{T}}$. Note that because *datA* is located close to the origin, we have used here that their concentrations are equal. We further assume that the concentration of the acidic phospholipids $[l]$ is constant. The last term describes the effect of protein synthesis (Appendix 3.B and Fig. 3.8). Since ATP is tenfold more abundant than ADP, new DnaA will predominantly bind ATP [17]. This term is however small at low growth rates ($\lambda \ll \tilde{\alpha}_{\mathrm{l}}, \tilde{\beta}_{\mathrm{datA}}$).

To simulate the LD model, we propagate the fraction of active, ATP-bound DnaA according to equation 3.2. The volume grows exponentially according to $dV/dt = \lambda V$ and the origin density is given by $[n_{\mathrm{ori}}] = n_{\mathrm{ori}}/V$. When the fraction f equals the critical fraction f^*, replication is initiated, and the number of origins n_{ori} doubles. The cell then divides a constant time τ_{cc} later. During cell division, the volume and the number of origins are halved.

Our switch model gives rise to stable cell cycles. The crux of the model is that while the activation rate is independent of the volume of the cell, the deactivation rate decreases with the volume because it is proportional to the density of *oriC* (Fig. 3.2a). The ATP-DnaA fraction $f(t)$ therefore increases with increasing volume $V(t)$ as the origin density decreases (Fig. 3.2b). When the critical initiator fraction $f^* = [D]^*_{\mathrm{ATP}}/[D]_{\mathrm{T}}$ is reached, replication is initiated. As soon as the origin and thus the site *datA* have been replicated, the maximum of the deactivation rate doubles and the active initiator fraction f decreases strongly, preventing reinitiation. As the cell continues to grow, the active initiator fraction rises again. This simple mechanism directly senses the origin density and ensures stable cell cycles (Fig. 3.2b).

At high (de)activation rates, the amplitude of the oscillations $\Delta f = f^* - f_{\mathrm{min}}$ is very large (Fig. 3.2c). At smaller and more biologically realistic rates ($\beta_{\mathrm{datA}} \approx 10\ \mathrm{min}^{-1}$) [63] (Appendix 3.A), the amplitude of the oscillations becomes very small especially at high growth rates (Fig. 3.2c); this continues to hold, even when the activation-deactivation system is deeper in the zero-order regime (Appendix 3.C). Such small amplitudes do not agree with the experiments [78] and are likely to be harmful, as even small fluctuations in the active fraction could result in untimely initiation of replication.

3.2. LARGE AMPLITUDE OSCILLATIONS AT HIGH GROWTH RATES IN LDDR MODEL

Because at biologically realistic (de)activation rates the LD model fails to generate large amplitude oscillations in the active DnaA fraction at high growth rates, the question arises how the cell cycle is regulated in this regime. Interestingly, in the fast growth regime $\lambda > \ln(2)/\mathrm{T_C} \approx 1.04/\mathrm{h}$, where the doubling time τ_{d} is shorter than the time to

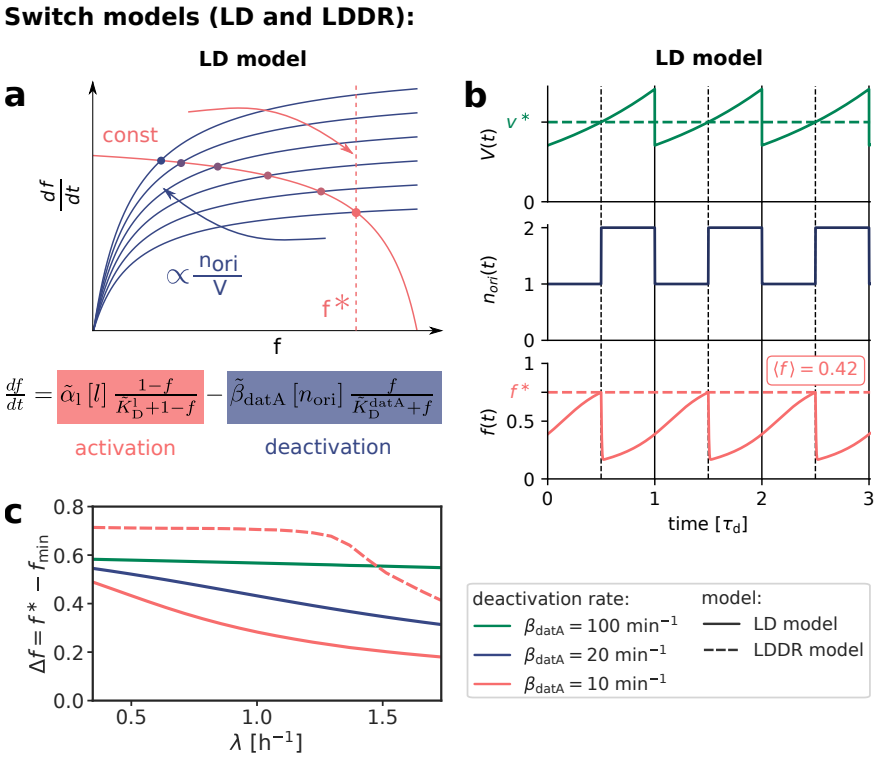


Figure 3.2: An ultra-sensitive switch between ATP-DnaA and ADP-DnaA gives rise to stable cell cycles. (a) LD model: The constant activation rate (red curve) and the origin density-dependent deactivation rate (blue curve) as a function of the active fraction of the initiator protein f at different moments of the cell cycle. The steady-state active fractions are given by the intersection of the activation and deactivation rates (colored dots) and when f equals the critical initiator fraction f^*, replication is initiated. A doubling of the number of origins leads to a decrease of the active fraction f. (b) LD model: The volume of the cell $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$ and the fraction of ATP-DnaA $f(t)$ from equation 3.2 as a function of time (in units of the doubling time $\tau_{\mathrm{d}} = 2$ h). The average active fraction over one cell cycle $\langle f \rangle$ is indicated in red in the third panel. Replication is initiated at a critical initiator fraction f^* (red dashed line) and the system gives rise to a constant initiation volume per origin v^* over time (green dashed line). (c) The amplitude Δf of the oscillations in the active fraction f as a function of the growth rate for different magnitudes of the (de)activation rates ($\alpha_{\mathrm{l}} = 4.6 \times \beta_{\mathrm{datA}}$). The amplitude of the oscillations Δf becomes small for biologically realistic values of the (de)activation rates in the LD model (red solid curve), but not in the LDDR model (red dashed curve). (See Table 3.1 for all parameters and Fig. 3.4 for time traces of LDDR model.)

replicate the entire chromosome T_C, replication is still proceeding when a new round of replication is initiated. This means that at the moment of replication initiation, the deactivation mechanism RIDA, which is associated with active replication forks, is active [88, 105]. Importantly, since RIDA is a potent deactivator [90], its activity must be balanced by another activation mechanism to maintain a roughly constant initiation volume independent of the growth rate [22, 40, 41]. We argue that this is the principal role of *DARS2*.

We therefore included the effects of RIDA and DARS1/2 in our full Lipid-*DatA*-*DARS1/2*-RIDA (LDDR) model (Fig. 3.1b). In the LDDR model, the number of catalytic RIDA complexes is proportional to the number of origins with a rate β_{rida} that is only non-zero during the period of active replication T_C (Fig. 3.3b). The chromosomal sites *DARS1* and *DARS2* are located near the middle of the chromosome and are replicated at constant times τ_{d1} and τ_{d2}, respectively, after the origin (Fig. 3.3a). The activities of DDAH and *DARS2* are temporally regulated during the cell cycle via binding of the integrating host factor (IHF) [43, 44, 63, 77]. IHF binding to *datA* increases within about 5-10 minutes, peaks at about 15 minutes, and decreases again about 20-30 minutes after initiation of replication [43, 63]. The binding of IHF to *DARS2* increases after 10 minutes, peaks after 20 minutes and decreases again 30-40 minutes after initiation of replication [44, 77]. We model these observations via step functions $\alpha_{d2}(t-t_i)$ and $\beta_{\text{datA}}(t-t_i)$ with a high and a low rate for *DARS2* and DDAH, respectively, that vary as a function of the time since initiation of replication $t-t_i$ (Fig. 3.3b). As *DARS2* remains highly active until replication termination, it can counteract the strong deactivator RIDA in the overlapping replication fork regime. *DARS2* is additionally regulated in a growth-rate dependent manner via the protein Fis [77, 96]. As there is no evidence for temporal regulation via Fis, we do not model Fis explicitly. *DARS1* activation is modeled via a constant activation rate α_{d1}. Like in the LD model, we assume that every newly synthesized DnaA binds ATP rather than ADP right after synthesis (Appendix 3.B). Summing up, we thus propose the following expression for the change in the ATP-DnaA fraction (see also Fig. 3.3a and b):

$$\frac{df}{dt} = \left(\tilde{\alpha}_1[l] + \tilde{\alpha}_{d1}[n_{\text{ori}}(t-\tau_{d1})] + \tilde{\alpha}_{d2}(t)[n_{\text{ori}}(t-\tau_{d2})]\right)\frac{1-f}{\tilde{K}_D+1-f} \tag{3.3}$$

$$-\left(\tilde{\beta}_{\text{datA}}(t)+\tilde{\beta}_{\text{rida}}(t)\right)[n_{\text{ori}}]\frac{f}{\tilde{K}_D+f}+\lambda(1-f) \tag{3.4}$$

with the re-normalized activation and deactivation rates $\tilde{\alpha}_1 = \alpha_1/[D]_T$, $\tilde{\alpha}_{d1} = \alpha_{d1}/[D]_T$, $\tilde{\alpha}_{d2} = \alpha_{d2}/[D]_T$, $\tilde{\beta}_{\text{datA}} = \beta_{\text{datA}}/[D]_T$ and $\tilde{\beta}_{\text{rida}} = \beta_{\text{rida}}/[D]_T$ and the Michaelis-Menten constant $\tilde{K}_D = K_D/[D]_T$. The parameters are described in Appendix 3.A and its values are listed in Table 3.1.

The LDDR model gives rise to stable cell cycles at all growth rates. Moreover, in contrast to the LD model, the LDDR model gives rise to large amplitude oscillations at all growth rates, even for realistic parameter values (Fig. 3.2c and Fig. 3.4 for time traces). This is because after a new round of replication, the RIDA deactivation rate is raised immediately while the activation rates of *DARS1/2* are increased only later, after the loci have been duplicated. This differential temporal dependence of the activation and deactivation rates is key to establishing large-amplitude oscillations at all growth rates.

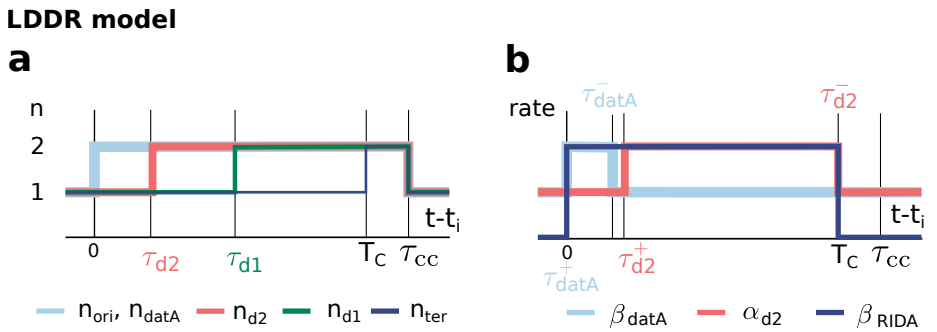


Figure 3.3: The (de)activation rates in the LDDR model are time-dependent. (a) The number of origins $n_{ori}(t-t_i)$, *datA* sites $n_{datA}(t-t_i) = n_{ori}(t-t_i)$, *DARS2* sites $n_{d2}(t-t_i) = n_{ori}(t-t_i-\tau_{d2})$, *DARS1* sites $n_{d1}(t-t_i) = n_{ori}(t-t_i-\tau_{d1})$ and termini $n_{ter}(t-t_i) = n_{ori}(t-t_i-T_C)$ per cell as a function of the time after initiation of replication at t_i. The time to replicate the entire chromosome is T_C and the time from the beginning of replication to cell division is τ_{cc}. Shown is the scenario for the low growth-rate regime of non-overlapping replication forks. (b) The cell cycle time dependent rates of *datA*, *DARS2* and RIDA as a function of the same cell cycle as in (a).

While the temporal separation of the activities of RIDA and *DARS2* can drastically enhance the amplitude of the oscillations, their fixed delay, together with the fact that they are both coupled to the origin density, can also impede robustness, as shown in Fig. 3.4c. More specifically, at low growth rates, the fixed time between replication initiation and replication of the site *DARS2* is very short ($\tau_{d2} = 0.2$ h) compared to the long doubling time of the cell ($\tau_d = 2$ h). Right after replication initiation, the deactivators *datA* and RIDA lower the active fraction of DnaA and prevent reinitiation. A short time τ_{d2} later however, the activity of *DARS2* increases and the active fraction rises. In this low growth rate regime, the active fraction is therefore high for a large fraction of the cell cycle, reducing the robustness of the LDDR model at low growth rates (see Fig. 4.2 of chapter 4). Interestingly, there is experimental evidence that the activity of *DARS2* decreases with decreasing growth rate [77, 95, 96]. Taking this into account does indeed positively affect the shape of the oscillations, yielding a slower rise in the active fraction when *DARS2*

(a, b) While in the LD model at realistic activation and deactivation rates the activating lipids and the deactivating site *datA* generate high-amplitude oscillations at low growth rates, the amplitude of the oscillations becomes very small at high growth rates. (c, d) In the LDDR model, due to the additional temporal modulation of the activation and deactivation rates of *datA*, RIDA and *DARS2*, the amplitude of the oscillations is high both at low and at high growth rates. At low growth rates however, as the activity of the site DARS2 increases a short time after replication initiation ($\tau_{d2} = 0.2$ h) compared to the cell doubling time of $\tau_d = 2$ h, the active fraction is high during a large fraction of the cell cycle, reducing the robustness of the LDDR model at low growth rates. As experiments indicate that the activity of *DARS2* decreases with decreasing growth rate [77, 95, 96], we show in (c) also the time trace of the active fraction when at low growth rates the high activation rate α^+_{d2} of *DARS2* (Fig. 3.3) is lower than that at high growth rate; here, we have taken α^+_{d2} to be equal to the low activation rate α^-_{d2} of *DARS2*, such that $\alpha^+_{d2} = \alpha^-_{d2} = 50$ (dashed red line). This improves the shape of the oscillations as it leads to a weaker increase in the active fraction upon doubling of *DARS2*.

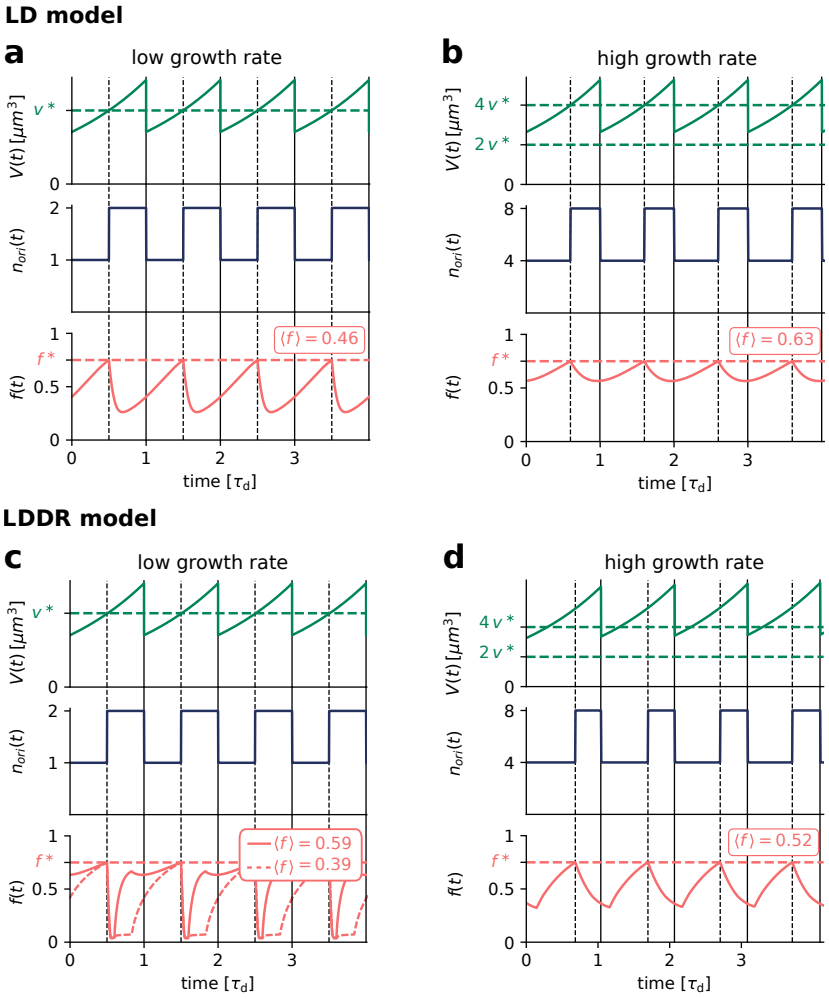


Figure 3.4: The LDDR model ensures high amplitude oscillations in the active fraction even at high growth rates and realistic (de)activation rates (a, b, c, d) The volume of the cell $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$ and the fraction of ATP-DnaA $f(t)$ as a function of time (in units of the doubling time τ_{d}) at a low doubling time of $\tau_{\mathrm{d}} = 2$ h ($\lambda = 1.35\ \mathrm{h}^{-1}$) (a, c) and a at high doubling time of $\tau_{\mathrm{d}} = 25$ min ($\lambda = 1.66\ \mathrm{h}^{-1}$) (b, d). The dashed red line is the critical initiator fraction f^* at which replication is initiated. The average active fraction over one cell cycle $\langle f \rangle$ is indicated in red in the third panel. Replication is initiated at a constant volume per origin v^* over time (green dashed line). (caption continues on the next page)

is duplicated (Fig. 3.4c, dashed red line). Importantly, however, titration naturally enhances the robustness of the switch in the low growth rate regime, by sharpening the oscillations in the concentration of *free, active* DnaA (see chapter 4). Moreover, we find that even with a constant, high activity, deleting *DARS2* does not significantly affect the initiation volume at low growth rates (see section 4.A and Fig. 4.6). Its dominant effect is at high growth rates: in this regime, *DARS2* is essential to vigorously counteract the strong deactivator RIDA, enabling a new round of replication while the old round has not finished yet.

3.3. SWITCH REQUIRES DIFFERENT VOLUME-DEPENDENCE OF (DE)ACTIVATORS

The finding that including RIDA and *DARS2* in the model is necessary to ensure the experimentally observed large amplitude oscillations at high growth rates raises the question whether it is also sufficient, meaning the LD model, and more in particular lipid synthesis, is not essential. In this section, we show that eliminating activation and deactivation mechanisms from the full LDDR model only maintains a stable system as long as the lipids with constant activity indepdendent of the origin density remain part of the model (Fig. 3.5). When instead all activation and deactivation mechanisms are connected to the chromosome, as in a system combining *DARS1/2* activation with *datA*/RIDA deactivation, both the activation and deactivation rates have the same functional dependence on the volume such that the system cannot sense the origin density anymore.

For a switch consisting only of *datA* and *DARS1*, we obtain the following expression for the change in the active fraction of DnaA:

$$\frac{df}{dt} = \tilde{\alpha}_{\mathrm{d1}}\left[n_{\mathrm{ori}}(t-\tau_{\mathrm{d1}})\right]\frac{1-f}{\tilde{K}_{\mathrm{D}}+1-f} - \tilde{\beta}_{\mathrm{datA}}\left[n_{\mathrm{ori}}\right]\frac{f}{\tilde{K}_{\mathrm{D}}+f} \tag{3.5}$$

where the site *DARS1* is replicated a time τ_{d1} after the origin. The concentration of *DARS1* is therefore proportional to the origin density at an earlier time $t-\tau_{\mathrm{d1}}$. A model where both the activator and deactivator are proportional to the (time shifted) origin density does not give rise to stable cell cycles (Fig. 3.5b). We can understand this observation by plotting the activation and deactivation rates as a function of the active fraction at different moments of the cell cycle (Fig. 3.5a). At quasi-steady-state, the active fraction is constant (setting equation 3.5 to zero) and the system will therefore settle to a constant fraction f independent of the volume of the cell. If this fraction f lies above the critical initiation fraction f^*, replication can be initiated (Fig. 3.5a, red dot). Because of its vicinity to the origin, the site *datA* is replicated right after initiation and reduces the active fraction below the initiation threshold (Fig. 3.5a and b, step 1). A constant time τ_{d1} after initiation of replication, the site *DARS1* is replicated as well (Fig. 3.5a and b, step 2). The active DnaA fraction rises again rapidly and when it attains the critical initiation fraction f^*, a new round of replication is initiated. The active DnaA fraction thus oscillates between a high and a low ATP-DnaA state at a period given by the time difference in replicating the sites on the chromosome *datA* and *DARS1*, τ_{d1}. This gives rise to a constant initiation period $\tau_{\mathrm{ii}} = \tau_{\mathrm{d1}}$; the system has become a timer. A system with a constant replication initiation period τ_{ii}, which is thus not coupled to the growth

or the volume of the cell, cannot give rise to stable cell cycles; even the smallest deviation of τ_{ii} from τ_d will inevitably grow and make the system unstable (Fig. 3.5b, green dots in upper panel). If division is coupled to replication initiation via a constant time τ_{cc}, the cell volume does therefore not remain stable after a few generations.

In summary, a system in which all activators and deactivators are located on the chromosome does not ensure stable cell cycles. This is because the volume dependence of the activation and deactivation rates is then the same, which means that the system cannot sense the origin density. Indeed, to sense the origin density, it is vital that the volume dependence of the activation and deactivation rates is different.

3.4. GROWTH-RATE DEPENDENCE OF THE INITIATION VOLUME

Experiments indicate that the initiation volume is relatively independent of the growth rate [22, 40], varying non-monotonically only within a 50% bandwidth over a tenfold change in the growth rate [41]. The question arises whether our switch model is consistent with this experimental observation. In the following we discuss the growth-rate dependence of both the LD and the LDDR model and compare it to experiments.

If the activation and deactivation rates are high compared to the growth rate, the system is well characterized by the steady state ($\frac{df}{dt} = 0$). In the LD model, the theoretical prediction of the initiation volume per origin $v^*_{th} = 1/[n_{ori}]$ is obtained by neglecting the protein synthesis term and setting equation 3.2 to zero:

$$v^*_{th} = \frac{\beta_{datA}}{\alpha_l [l]} \frac{f^*}{\tilde{K}_D^{datA} + f^*} \frac{\tilde{K}_D^{l} + 1 - f^*}{1 - f^*} \tag{3.6}$$

with the critical initiator fraction $f^* = [D]^*_{ATP}/[D]_t$.

All parameters on the right side of equation 3.6 could in principle vary with the growth rate λ of the cell. The initiation volume per origin v^*_{th} is constant at all growth rates if all terms on the right side are constant or if the growth rate dependencies of the parameters cancel each other out. While the deactivation rate of DDAH is known to be temporally regulated over the course of the cell cycle, the explicit growth rate dependence has not been studied so far; such a dependence could well be possible, as several proteins like Dia and IHF (whose concentrations could vary with λ) are involved. A growth-rate dependence of the critical initiation fraction f^* has not been reported but could be possible, as two other proteins, Dia and IHF, are involved in the initiation process [43, 80]. The lipid concentration in equation 3.6 stems from a combination of the two types of lipids CL and PG. The cell membrane composition is complex and could depend on the growth rate. Our model predicts however that if all other parameters of equation 3.6 are growth-rate independent, the for replication initiation relevant phospholipid concentration should be approximately constant in order to obtain a constant initiation volume v^*.

The LDDR model yields two different predictions for the initiation volume per origin v^* in the quasi-equilibrium regime where (de)activation is faster than growth, depending on whether RIDA and *DARS2* are active at the moment of initiation or not. From Eq. 3.4 it follows that in the low growth-rate regime ($\lambda < \ln 2/T_C$) the replications forks

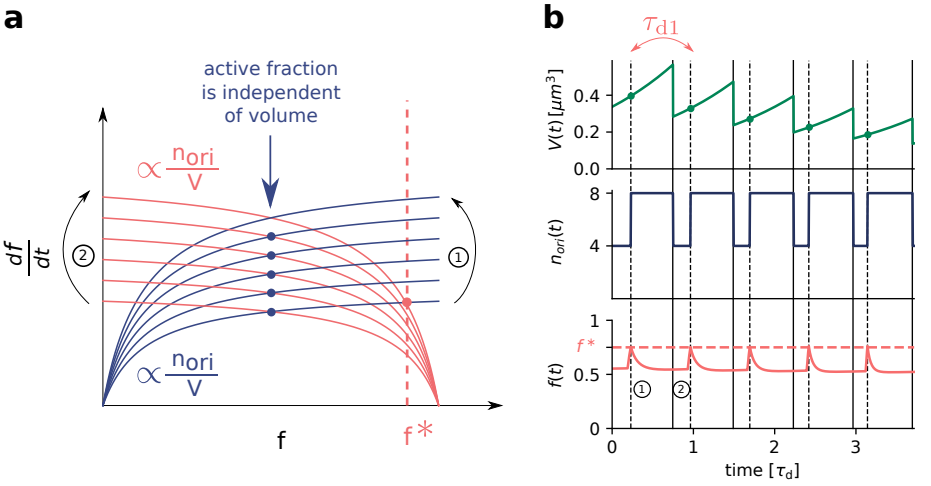


Figure 3.5: A switch model where all activators and deactivators are located on the chromosome does not give rise to stable cell cycles (a) The activation (red curves) and deactivation rates (blue curve) as a function of the active fraction of the initiator protein f at different moments of the cell cycle. The steady-state active fractions are given by the intersection of the activation and deactivation rates (colored dots). As both the activation and the deactivation rate depend on the origin density, the active fraction becomes volume independent. When the active fraction f equals or is larger than the critical initiation fraction f^* (vertical dashed red line), replication is initiated. When replication is initiated (step 1), the number of origin doubles. Due to the vicinity of *datA* to the origin, the deactivation rate doubles right after replication is initiated and the active fraction is reduced to a constant value below the activation threshold. A fixed time $\tau_{d1} = 0.35$ h $= 21$ min after replication was initiated, the site *DARS1* is doubled which causes again an increase of the active fraction beyond the critical active fraction f^* and a new round of replication is initiated again. (b) The volume of the cell $V(t)$, the number of origins $n_{ori}(t)$ and the fraction of ATP-DnaA $f(t)$ as a function of time (in units of the doubling time τ_d) at a doubling time of $\tau_d = 0.5$ h $= 30$ min. The dashed red line is the critical initiator fraction f^* at which replication is initiated. Replication is initiated at a constant time interval $\tau_{ii} = \tau_{d1}$ which, in this example, is smaller than the doubling time of the cell. As a consequence, the volume per origin v^* decreases over time and so does the birth volume of the cell.

are non-overlapping and the initiation volume v_{no}^* is given by

$$v_{\mathrm{no}}^* = \frac{\beta_{\mathrm{datA}}^-}{\alpha_1[l]} \frac{f^*}{\tilde{K}_\beta + f^*} \frac{\tilde{K}_\alpha + 1 - f^*}{1 - f^*} - \frac{\alpha_{\mathrm{d1}} + \alpha_{\mathrm{d2}}^-}{\alpha_1[l]} \tag{3.7}$$

In the high growth regime ($\lambda > \ln 2 / T_{\mathrm{C}}$), the initiation volume per origin is given by

$$v_{\mathrm{o}}^* = \frac{\beta_{\mathrm{datA}}^- + \beta_{\mathrm{rida}}}{\alpha_1[l]} \frac{f^*}{\tilde{K}_\beta + f^*} \frac{\tilde{K}_\alpha + 1 - f^*}{1 - f^*} - \frac{\alpha_{\mathrm{d1}} + \alpha_{\mathrm{d2}}^+}{\alpha_1[l]} \tag{3.8}$$

To obtain equations 3.7 and 3.8, we have again assumed that the (de)activation rates are sufficiently high such that the effect of protein synthesis can be neglected. The initiation volume is only the same at all growth rates, if the rate of RIDA and the high activity rate of *DARS2* are chosen such that they exactly cancel out. By setting $v_{\mathrm{o}}^* = v_{\mathrm{no}}^*$ we obtain the following constraint on the high activity rate of *DARS2* as a function of the rate of RIDA:

$$\alpha_{\mathrm{d2}}^+ = \beta_{\mathrm{rida}} \frac{f^*}{\tilde{K}_\beta + f^*} \frac{\tilde{K}_\alpha + 1 - f^*}{1 - f^*} + \alpha_{\mathrm{d2}}^- \tag{3.9}$$

As emphasized above, experiments indicate that the initiation volume varies non-monotonically within a 50% range over a tenfold change in the growth rate [41]. Clearly, a constant initiation volume independent of the growth rate is not a robust property of the LDDR model, but only emerges if the rates are chosen carefully, matching Equation 3.9. Experiments support, however, this prediction: Specifically, our model agrees with the experimental finding that deleting *datA* or disabling RIDA or *DARS2* not only affects the initiation volume, but, importantly, also makes it dependent on the growth rate [44, 89–91, 96, 106]; moreover, while the effect of deleting *datA* is most pronounced at lower growth rates [89, 106], disabling RIDA or *DARS2* is more severe at higher growth rates [44, 90, 91, 96]. For a more detailed comparison against experimental data, we refer to the model validation section 4.A and Fig. 4.6.

3.5. A STOCHASTIC MODEL RECOVERS ADDER CORRELATIONS

In the titration-based system, a new round of replication is initiated when the number of DnaA proteins that have been accumulated since the last initiation event equals roughly the number of titration sites, irrespective of the previous initiation volume; moreover, DnaA proteins are accumulated proportionally to the volume of the cell. These two elements together naturally give rise to adder correlations (section 2.6). Yet, our switch model is a sizer at the mean-field level: replication is initiated when the origin density reaches a critical threshold. Do the experimentally observed adder correlations [23, 24] rule out our switch model?

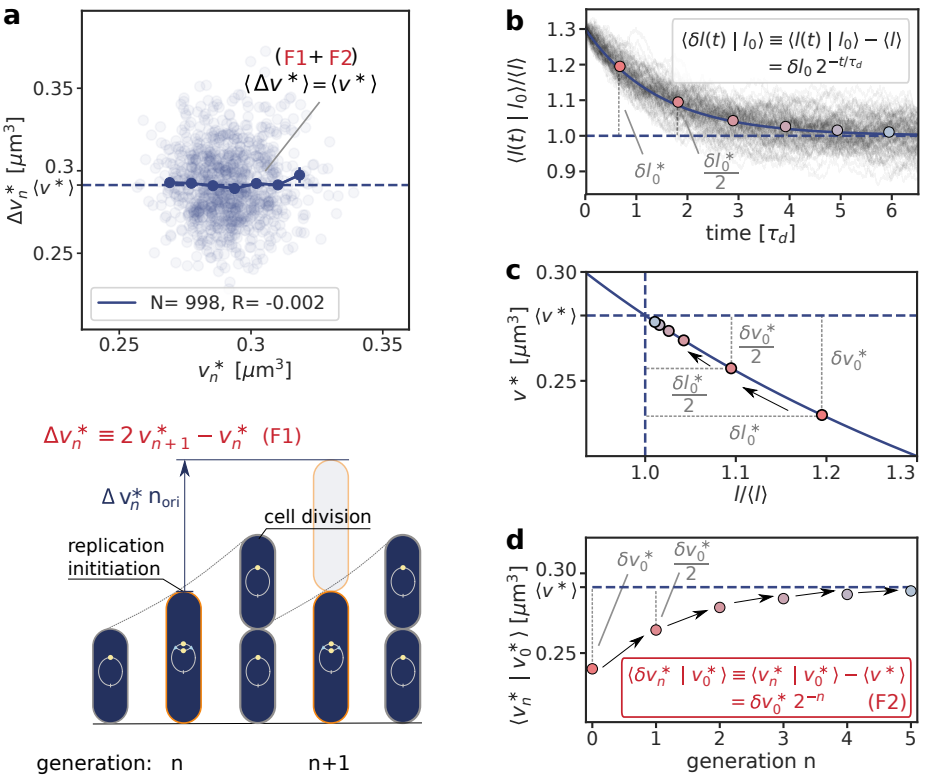


Figure 3.6: Fluctuations in the switch components can give rise to the experimentally observed adder correlations in the initiation volume per origin v^*, illustrated using the LD model with lipid concentration fluctuations (Eq. 3.10). (a) The added volume per origin between successive initiation events, $\Delta v_n^* = 2\,v_{n+1}^* - v_n^*$, is independent of the initiation volume v_n^* per origin and on average equal to the average initiation volume, $\langle \Delta v^* \rangle = \langle v^* \rangle$, as expected for an initiation volume adder. The cartoon below illustrates the volume $\Delta v^*\, n_{\mathrm{ori}}$ that is added between successive initiation events (here, the number of origins before initiation is $n_{\mathrm{ori}} = 1$). (b) Lipid-concentration fluctuations $l(t) \equiv [l](t)$ regress to the mean on a timescale $\tau_{\mathrm{d}} = \ln(2)/\lambda$ set by the growth rate λ, such that an initial perturbation $l_0 - \langle l \rangle$ is halved every subsequent cell cycle. The thin grey lines are time traces from $N = 100$ simulations starting at an initial lipid concentration perturbation δl_0, while the solid line is the analytical prediction for the mean obtained by solving equation 3.10 subject to the same initial condition. The colored dots indicate the average lipid concentration perturbation δl^* at the moment of initiation in generation n and the horizontal dashed line shows the average lipid concentration at steady state. (c) The mapping between the initiation volume v^* and the normalized lipid concentration $l/\langle l \rangle$ (blue line), obtained by solving equation 3.2 in steady state in the limit of high (de)activation rates (see also Fig. 3.7). As in (a), the colored dots show the average initiation volume v^* as a function of $l/\langle l \rangle$ at the successive initiation events. (caption continues on the next page)

To address this question, we systematically studied the effect of fluctuations in the individual components of our switch model [107]. We first consider fluctuations in the lipid concentration, modelled as

$$\frac{d[l]}{dt} = \alpha - \lambda\,[l] + \xi(t), \tag{3.10}$$

where α is the production rate, the second term describes the effect of dilution set by the growth rate λ and $\xi(t)$ models the noise resulting from protein production and partitioning upon cell division. The noise is modelled as Gaussian white noise, $\langle\xi(t)\xi(t')\rangle = 2D_l\delta(t-t')$, with the noise strength D_l chosen to match the measured variance in the initiation volume of $CV = 0.1$ [22] (see Table 3.1). Fig. 3.6 illustrates the effect of noise in the lipid concentration according to equation 3.10 using the LD model, but Fig. 3.10 shows that the principal result also holds for the full LDDR model: the added initiation volume between consecutive initiation events $\Delta v_n^* = 2\,v_{n+1}^* - v_n^*$ is indeed independent of the volume at initiation v_n^*, in agreement with experiments [23, 24].

To understand how fluctuations in the switch components propagate to fluctuations in the initiation volume, it is illuminating to analyze the simpler LD model. The reason why the LD model is more instructive is that in this model the mapping $f(v)$ between the instantaneous fraction $f(t)$ of active DnaA and the current volume $v(t)$ per origin can be obtained and understood straightforwardly. In the regime where (de)activation is faster than growth, this mapping $f(v)$ is obtained by neglecting protein production and solving equation 3.2 in steady state. As we will show, this mapping $f(v)$ depends in an intuitive manner on the concentrations and activities of the switch components (Fig. 3.7), such that it becomes clear how fluctuations in these components propagate to fluctuations in the initiation volume. In contrast, in the LDDR model the mapping between the instantaneous fraction $f(t)$ and the instantaneous volume per origin $v(t)$ is non-trivial because the rates of RIDA and *DARS2* are temporally regulated over the course of the cell cycle. Importantly, however, while the LDDR model is less illuminating, the principle remains: fluctuations in the switch components will propagate to fluctuations in the mapping $f(v)$, and these will propagate to fluctuations in the threshold for replication initiation. The mapping $f(v)$ is shown for different lipid concentrations in panels a and b of Fig. 3.7, for different degrees of ultra-sensitivity, respectively. It shows that the mapping between the active fraction and the volume depends on the lipid concentration. Since replication is initiated when the active fraction f reaches the critical fraction f^* for replication initiation, marked by the horizontal dashed line, fluctuations in the lipid concentration lead to fluctuations in the initiation volume per origin v^* (Fig. 3.7c).

The concentrations of cellular components will fluctuate inevitably, and unless the components are degraded actively or produced with negative feedback control, the fluctuations will persist over several generations, regressing to the mean on a timescale set

(d) The average initiation volume (colored dots) relaxes on the same timescale τ_d as the lipid concentration, such that a perturbation $v_0^* - \langle v^*\rangle$ is halved every cell cycle, giving rise to adder correlations. In (a) the dark blue line shows the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The number of data points N and the Pearson correlation coefficient R are indicated. The model includes an eclipse period of about 10 minutes following replication initiation to prevent immediate reinitiation. (See Table 3.1 for all parameters.)

by the growth rate (Fig. 3.6b). The components that control the threshold of the DnaA activation switch are no exception to this rule. Moreover, their concentration fluctuations will give rise to fluctuations in the initiation volume v^* (Fig. 3.6c) that, to a good approximation, relax on the same timescale because (de)activation is fast compared to the growth rate and the mapping between these components and the initiation volume is roughly linear (Fig. 3.6c). If this timescale is set by the growth rate, then deviations of v^* from its mean are on average halved every cell cycle (Fig. 3.6d), and this gives rise to adder correlations (see Appendix 3.D for derivation) [23]. These ideas are generic and apply not only to the LD model of Fig. 3.6, but also to the full LDDR model (Appendix 3.E). Indeed, in Appendix 3.F we show how fluctuations in the activity of RIDA in the full LDDR model, as induced by e.g. Hda [88], also generate adder correlations in the initiation volume (Fig. 3.11). Fluctuations in switch components that relax with the growth rate, be they lipids or proteins that modulate the activity of *datA*, RIDA, or *DARS1/2* like IHF and Hda [43, 44, 63, 77, 88], thus give rise to adder correlations.

3.6. EFFECT OF FLUCTUATIONS IN THE TOTAL INITIATOR CONCENTRATION

So far, we have assumed that the total DnaA concentration is maintained strictly constant in time and at different growth rates via negative autoregulation. There was thus no difference in initiating replication at a critical concentration $[D]^*_{\mathrm{ATP}}$ or a critical fraction f^*. When the total concentration of DnaA is however fluctuating, the concentration of active proteins $[D]_{\mathrm{ATP}}$ and the fraction of active proteins $f = [D]_{\mathrm{ATP}}/[D]_{\mathrm{T}}$ are not directly proportional anymore. This poses a new question: Is replication initiated at a critical ATP-DnaA concentration $[D]^*_{\mathrm{ATP}}$ or at a critical fraction f^*? Both scenarios could be possible and have been discussed in literature [44, 85]. In the following, we will discuss the effect of fluctuations in the total concentration on the initiation volume in both of these cases in the LD model.

Fluctuations in the total concentration can affect the initiation volume if replication is initiated at a critical ATP-DnaA concentration. We first consider the case where replication is initiated at a critical ATP-DnaA concentration $[D]^*_{\mathrm{ATP}}$. In the LD model for very high (de)activation rates, the active DnaA concentration can be plotted as a function of the volume of the cell for different total concentrations (Fig. 3.7d and e). If the dissociation constants of the activator and deactivator are much smaller than the total concentration, the switch is in the ultra-sensitive regime and becomes very steep (Fig. 3.7d). In this case, the critical initiation concentration is attained at approximately the same volume per origin independent of the total concentration, as shown in Fig. 3.7f. If the dissociation constants of the activator and deactivator are however in the same order of magnitude as the total concentration, the ATP-DnaA concentration rises more gradually and attains the critical initiation concentration at different volumes for different total DnaA concentrations (Fig. 3.7e). Consequently, the initiation volume depends now more strongly on the total DnaA concentration (Fig. 3.7f). To summarize, if replication is triggered at a critical ATP-DnaA concentration $[D]^*_{\mathrm{ATP}}$ and if the switch is not extremely sharp, we predict a dependence of the initiation volume on the total concentration.

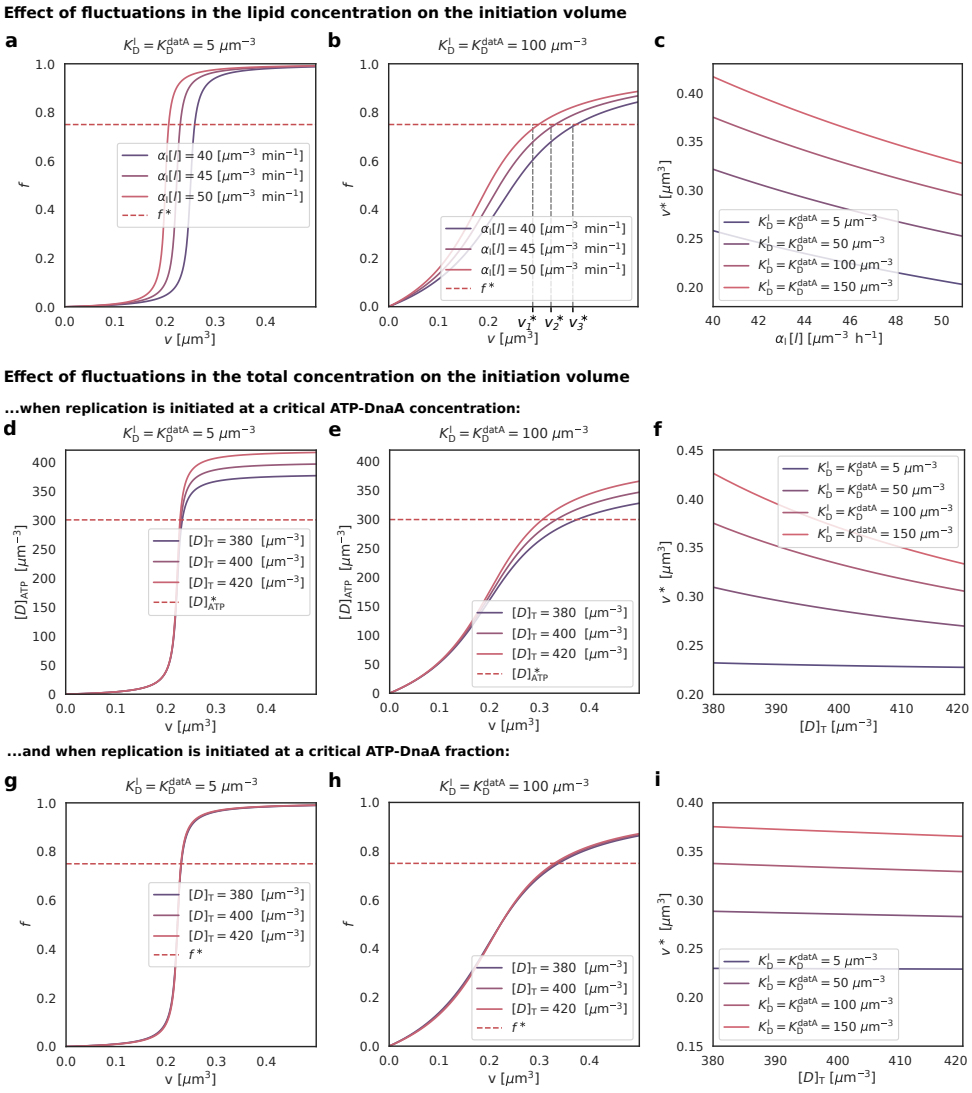


Figure 3.7: Effect of varying lipid and total DnaA concentration on initiation volume in LD model (a, b) The fraction of active ATP-bound DnaA f as a function of the cell volume per origin v for different lipid concentrations $[l]$, and for two different values of the dissociation constants of the activator and the deactivator of $K_D^{\alpha} = K_D^{\beta}$, respectively. The horizontal dashed line shows the critical fraction f^* for replication initiation. Clearly, fluctuations in the lipid concentration lead to fluctuations in the initiation volume v^*, which is the volume per origin v at which the fraction f equals the critical fraction f^*. The initiation volume v^* as a function of the lipid concentration is shown in panel c, scaling as $1/[l]$ both in the regime of strong ultra-sensitivity (a) and weak ultra-sensitivity (b). (caption continues on the next page)

We note here that this implies that care should be taken in inferring molecular mechanisms from experiments in which the expression of DnaA is modulated [108]. In particular, like the initiator accumulation model, also the switch model would predict that the initiation volume decreases as the total DnaA concentration increases.

If replication is however initiated at a critical ATP-DnaA fraction, fluctutations in the total concentration have almost no effect on the initiation volume. Interestingly, again in the limit where the (de)activation rates are higher than the growth rate, the fraction as a function of the cell volume in the LD model is essentially independent of the total DnaA concentration (Fig. 3.7g and h). This finding does also not depend on the steepness of the switch (Fig. 3.7g and h). The critical fraction f^* is attained at an almost perfectly constant volume per origin (Fig. 3.7i) at all dissociation constants. The switch mechanism is thus extremely well protected against variations in the total concentration, if replication is initiated at a critical active fraction of DnaA.

3.7. NEGATIVE AUTOREGULATION OF DNAA GENERATES SIZER CORRELATIONS

In this section, we explicitly model the total concentration $[D]_\mathrm{T}$ and investigate the resulting correlations in the initiation volume. As we have seen in the previous section, the effect of fluctuations in the total concentration is especially high, when replication is initiated at a critical ATP-DnaA concentration and when the system is not too far in the ultra-sensitivity regime. We use a relatively large dissociation constant of $K_\mathrm{D}^\mathrm{datA} = K_\mathrm{D}^\mathrm{l} = 50\,\mu\mathrm{m}^{-3}$ in order to obtain a strong dependence of the initiation volume on the total concentration (Fig. 3.7f). Since the affinities of the two nucleotide binding forms of DnaA to the promoters differ only by a factor of two [69], we here make the simplifying assumption that both forms of DnaA have equal affinity for the promoter. As there are five binding sites for DnaA in the promoter region [69], we choose a Hill coefficient of $n = 5$ in the simulations. The dynamics of the *total* number of DnaA is then given by (the

(d, e) The concentration of active ATP-bound DnaA, $[D]_\mathrm{ATP}$, as a function of the volume per origin v for different values of the total DnaA concentration, both in the regime of strong ultra-sensitivity (d) and weak ultra-sensitivity (e). The horizontal dashed line denotes the critical ATP-DnaA concentration for replication initiation. Fluctuations in the total DnaA concentration generate stronger fluctuations in the initiation volume when the degree of ultra-sensitivity is weaker (panel e). This is highlighted in panel f, which shows the initiation volume as a function of the total DnaA concentration for different values of the dissociation constants. (g, h) The fraction f of active ATP-bound DnaA as a function of the volume per origin v for different total concentrations of DnaA, in the regime of strong (g) and weak (h) ultra-sensitivity. The horizontal dashed line denotes critical fraction f^* for replication initiation. The active fraction f depends only weakly on the total DnaA concentration, almost irrespective of the degree of ultra-sensitivity. As a result, the initiation volume is essentially independent of the total DnaA concentration for nearly all values of the dissociation constant (panel i). Replication initiation is thus well protected against noise in the concentration of DnaA. All curves are obtained by neglecting the protein synthesis term and solving equation 3.2 in steady state. This gives the mapping between the instantaneous concentration $[D]_\mathrm{ATP}(t)$ or fraction $f(t)$ of active ATP-bound DnaA and the instantaneous volume per origin $v(t)$ when the (de)activation rates are higher than the growth rate.

same expression as in the AIT model, see equation 2.24):

$$\frac{dN_{\mathrm{D}}^{\mathrm{T}}}{dt} = \frac{\tilde{\phi}_{\mathrm{p}}^{0}\,\lambda\,V}{1+\left(\frac{[D]_{\mathrm{T}}}{K_{\mathrm{D}}^{\mathrm{p}}}\right)^{n}} + \xi_{\mathrm{D}}(t). \tag{3.11}$$

where we have employed the growing cell model with the basal gene allocation density $\tilde{\phi}_{\mathrm{p}}^{0}$ (see section 2.1) and combined it with a noise term $\xi_{\mathrm{D}}(t)$ accounting for noise in gene expression (see section 2.6). The noise is modelled as Gaussian white noise, $\langle\xi_{\mathrm{D}}(t)\xi_{\mathrm{D}}(t')\rangle = 2D_{\mathrm{D}}\delta(t-t')$, where the noise strength D_{D} is tuned to match the measured variance in the initiation volume (see Table 3.1).

The total DnaA concentration is obtained by dividing the total number of DnaA proteins $N_{\mathrm{D}}^{\mathrm{T}}(t)$ by the explicitly evolved volume $V(t) = V_{\mathrm{b}}\exp(\lambda t)$. As newly produced DnaA proteins are more likely to bind ATP than ADP we add the DnaA production term in equation 3.11 to the change in the number of active DnaA proteins. The change in the number of ATP-DnaA proteins is computed using

$$\frac{dN_{\mathrm{D}}^{\mathrm{ATP}}}{dt} = \frac{dN_{\mathrm{D}}^{\mathrm{T}}}{dt} + \alpha_{\mathrm{l}}\,[l]\,V\,\frac{[D]_{\mathrm{ADP}}}{K_{\mathrm{D}}^{\mathrm{l}}+[D]_{\mathrm{ADP}}} - \beta_{\mathrm{datA}}\,n_{\mathrm{ori}}\,\frac{[D]_{\mathrm{ATP}}}{K_{\mathrm{D}}^{\mathrm{datA}}+[D]_{\mathrm{ATP}}} \tag{3.12}$$

$$= \frac{\tilde{\phi}_{\mathrm{p}}^{0}\,\lambda\,V}{1+\left(\frac{[D]_{\mathrm{T}}}{K_{\mathrm{D}}^{\mathrm{p}}}\right)^{n}} + \xi_{\mathrm{D}}(t) + \alpha_{\mathrm{l}}\,[l]\,V\,\frac{[D]_{\mathrm{ADP}}}{K_{\mathrm{D}}^{\mathrm{l}}+[D]_{\mathrm{ADP}}} - \beta_{\mathrm{datA}}\,n_{\mathrm{ori}}\,\frac{[D]_{\mathrm{ATP}}}{K_{\mathrm{D}}^{\mathrm{datA}}+[D]_{\mathrm{ATP}}} \tag{3.13}$$

and the active initiator concentration $[D]_{\mathrm{ATP}}(t)$ is obtained by dividing the number of ATP-DnaA proteins $N_{\mathrm{D}}^{\mathrm{ATP}}(t)$ by the volume $V(t)$. A new round of replication is initiated when the ATP-DnaA concentration reaches the critical concentration for replication initiation $[D]_{\mathrm{ATP}}^{*}$; the cell then divides a constant time τ_{cc} later. During cell division, the volume and the number of DnaA proteins and the number of origins are halved. The rate constants are the same as in the original LD model (see Table 3.1). To prevent premature reinitiation by stochastic DnaA fluctuations immediately after replication initiation, we also implement a refractory or 'eclipse' period of $\tau_{\mathrm{b}} = 10$ minutes following replication initiation during which replication cannot be reinitiated, mimicking the effect of SeqA [58–60].

Figure 3.10b shows the result. As fluctuations in the total number are reduced via negative autoregulation within less than one generation, we obtain sizer-like correlations in the initiation volume. We emphasise however that it remains to be verified experimentally how strong the effect of negative autoregulation is.

3.8. DISCUSSION

We have shown that a system based on activation and deactivation of DnaA can generate stable replication initiation cycles at all growth rates. Yet, this mechanism requires that the total activation and deactivation rate depend differently on the origin density. This requirement is fulfilled by the fact that the lipid-mediated DnaA activation rate is independent of the origin density, while the other (de)activation rates depend on the origin

density. This differential dependence of the (de)activation rates on the origin density gives rise to replication initiation at a critical volume per origin.

The mechanisms of titration and activation belong to distinct classes of replication initiation control. The titration-based AIT model is an example of an initiator accumulation model, in which an initiator protein needs to accumulate to a threshold number to initiate replication [11, 14, 23, 40, 64]. In contrast, the DnaA activation switch is an example of a push-pull network in which the regulator switches between an inactive and an active state. Conceptually, this switch model is different from the accumulation model because replication is triggered at a critical concentration or fraction and not at a critical number of accumulated initiator proteins. In the switch model, the concentration of ATP-DnaA is set by the balance between DnaA activation and deactivation. Because the (de)activation rates depend differently on the origin density, the critical initiator concentration maps onto a critical origin density for replication initiation. This switch system is thus a bonafide origin-density sensor.

In recent years, single-cell tracking data have revealed that not only *E. coli* but also other evolutionary divergent organisms like *Bacillus subtilis* [6], *Caulobacter crescentus* [5], the archaeon *Halobacterium salinarum* [7], and even budding yeast [4], obey a division adder principle. Our study gives a new perspective on the question whether a cell cycle is controlled via a sizer or adder. While the titration mechanism naturally qualifies as an adder, our switch model should be characterised as a sizer at the mean-field level: the mechanism is based on sensing the origin density. Yet, the inevitable fluctuations in the components that control the density threshold for replication give rise to adder correlations. This idea is general and likely applies to other organisms that obey the adder principle: adder behavior may result from size sensing. Our prediction could be tested by measuring the critical active DnaA concentration for replication initiation and how its fluctuations relax. Since ATP binding induces a conformational switch of DnaA [109], developing a FRET-based ATP-DnaA sensor may be feasible.

For the sake of simplicity we have so far assumed that the cell divides a constant cycling time τ_{cc} after replication initiation. Experiments indicate, however, that cell division is more loosely coupled to replication initiation [22–24, 27–29, 67] and alternative models of division control have been proposed: In the Independent Double Adder (IDA) model, an adder runs from cell birth to division, while in the Replication Double Adder (RDA) model an adder runs from replication initiation to cell division (see section 2.D for more details on IDA/RDA model). Importantly however, fluctuations in the lipid concentration result again in adder correlations for the initiation volume when cell division is controlled by either of these two mechanisms (Fig. 3.12): these correlations depend on the correlation time of the fluctuations in the switch components, but do not depend on the specific type of coupling of the replication cycle to the division cycle.

An open question remains why *E. coli* has evolved two different switch systems, Lipid-*DatA* (LD) and *DARS1/2*-RIDA (DR). In principle, a switch based on activating lipids and deactivating *datA* would be sufficient to control replication initiation at all growth rates. Yet, to ensure high amplitude oscillations in the active DnaA fraction at high growth rates, the (de)activation rates would have to be higher than observed (Fig. 3.2c). This would require higher turnover rates of ATP, which may not be achievable when the growth rate is low. Our model thus suggests that *E. coli* has evolved a slow system to con-

trol the initiation volume at low growth rates, the lipids-*datA* system, and then switches on a faster, more energy-consuming system at higher growth rates, based on RIDA and *DARS2*.

In this chapter, we have assumed that replication initiation is controlled by the DnaA activation switch only. Yet, there is clear evidence that DnaA also binds to chromosomal titration sites. Indeed, as we showed in the previous chapter, a model based on titration can give rise to stable replication cycles, albeit at lower growth rates only. The question that thus arises is how a concentration cycle as generated by titration interacts with an activation cycle as driven by the DnaA activation switch, to control the initiation of DNA replication. This is the topic of the next chapter.

APPENDIX

3.A. EXPERIMENTAL FINDINGS AND PARAMETERS

Here we give an overview of the experimental findings on the regulation of replication initiation in *E. coli* and discuss realistic parameter ranges of the parameters of the switch models. As the protein DnaA binds very strongly to both ATP and ADP with a dissociation constant of $K_D = 10 - 100$ nM [43, 94], we assume that DnaA is always bound to either ATP or ADP. In the following we refer to the total DnaA concentration as $[D]_T$, to the ATP-DnaA as $[D]_{ATP}$ and to the ADP-DnaA as $[D]_{ADP}$. The total DnaA concentration $[D]_T = [D]_{ATP} + [D]_{ADP}$ varies within a 50% range over a tenfold change in the growth rate [41]. Hansen et al. reported a typical number of 330 molecules in an *E. coli* cell with the doubling rate $1/\tau_d = 0.58\ h^{-1}$ [64]. Combining this number with the estimated volume at this doubling rate using the data from Si et al. [40] (for the same *E. coli* strain K-12) of $V(\tau_d^{-1} = 0.58\ h^{-1}) \approx 0.7\,\mu m^3$, we obtain an estimated concentration of DnaA of $[D]_T \approx 471\,\mu m^{-3}$. We use throughout this work a total DnaA concentration of $[D]_T = 400\,\mu m^{-3}$ (See Table 3.1). As discussed in the previous section, the protein DnaA can be bound to the DnaA boxes on the DNA and at the origin or be diffusing in the cytoplasm. In this chapter, we neglect the effect of the titration sites on the chromosome and assume that all DnaA proteins are present in the cytosol. In the full model in chapter 4, we relax this constraint and investigate the effect of a negatively autoregulated DnaA protein that can also bind to titration sites.

In *E. coli*, replication is initiated once per cell cycle at the origin region by the binding of ATP-DnaA to two high-affinity DnaA boxes (R1 and R4) and several low-affinity DnaA boxes together with two other proteins, the DnaA-initiator-associating protein DiaA and the integration host factor (IHF) [43]. While ADP-DnaA can bind to the DnaA boxes on the origin, it does not form the cooperative complex required for the initiation of replication. The fraction of ATP-DnaA is maintained at a low level during most of the cell cycle and increases to approximately 80% at the moment of replication initiation [78, 82]. An interesting and strongly debated question is whether replication is initiated at a critical amount, concentration or fraction of ATP-DnaA in the cell [14, 85, 106]. In our LD and LDDR model, replication is initiated when the ATP-DnaA concentration in the cell attains a critical concentration $[D]^*_{ATP}$. We exploit that the total concentration of DnaA is maintained approximately constant and take $[D]_T$ = constant in our model such that a critical initiation concentration $[D]^*_{ATP}$ corresponds to a critical fraction $f^* = [D]^*_{ATP}/[D]_T$. In section 3.6 we also analyse the implications of the difference between initiating replication at a critical fraction versus a critical concentration of active DnaA. In section 3.7 we investigate the effect of fluctuations in the total concentration of DnaA.

So far, several activators and deactivators of DnaA have been identified and characterized in great detail. Here, we briefly summarize these experimental results, starting with the deactivators.

Deactivation mechanisms: datA and RIDA Regulatory Inactivation of DnaA (RIDA) was

Parameter	name	value (LD)	value (LDDR)	Motivation
$\alpha_1[I]$ [μm^{-3} h^{-1}]	activation rate lipids	2755	750	combined with $\beta_{\text{datA}}, \beta^-_{\text{datA}}, \alpha_{\text{d1}}, \alpha^-_{\text{d2}}$ to match ν^* from [40]
		Fig. 3b: 27550		
		Fig. 4.3b: 2142		
β_{datA} [h^{-1}]	deactivation rate *datA*	600	-	[63]
		Fig. 3b: 6000		
K_{D}^1 [μm^{-3}]	dissociation constant lipids	50	-	taken to be equal to $K_{\text{D}}^{\text{datA}}$
$K_{\text{D}}^{\text{datA}}$ [μm^{-3}]	dissociation constant *datA*	50	-	[46]
K_{D} [μm^{-3}]	dissociation constant in LDDR model	-	50	[46]
τ^+_{datA} [h] after t_{i}	start high deactivation rate *datA*	-	0	[63]
τ^-_{datA} [h] after t_{i}	end high deactivation rate *datA*	-	0.2	[63]
β^+_{datA} [h^{-1}]	high deactivation rate *datA*	-	600	[63]
β^-_{datA} [h^{-1}]	low deactivation rate *datA*	-	300	[63]
$[D]_{\text{T}}$ [μm^{-3}]	total DnaA concentration	400	400	[18, 64]
f^*	critical initiator fraction	0.75	0.75	[78, 82]
		Fig. 4.3b: 0.5		
$[D]^*_{\text{ATP,f}}$ [μm^{-3}]	critical free ATP-DnaA concentration	-	full model: 200	[18, 78, 82]
α_{d1} [h^{-1}]	activation rate *DARS1*	-	100	[43, 77]
			Fig. 3.5: 1200	
τ_{d1} [h] after t_{i}	replication time *DARS1*	-	0.1	[43]
			Fig. 3.5: 0.35	
α^+_{d2} [h^{-1}]	high activation rate *DARS2*	-	643	combined with β_{rida}
			Fig. 3.11: 1930	to match ν^* from [40]
			Fig. 3.4: 50	
α^-_{d2} [h^{-1}]	low activation rate *DARS2*	-	50	set to arbitrary low value
τ_{d2} [h] after t_{i}	replication time *DARS2*	-	0.2	[77]
τ^+_{d2} [h] after t_{i}	start high activation rate *DARS2*	-	0.2	[77]
τ^-_{d2} [h] after t_{i}	end high activation rate *DARS2*	-	2/3	[77]
β_{rida} [h^{-1}]	deactivation rate RIDA	-	500	[63, 88, 110]
			Fig. 3.11: 1500	
τ_{b} [h] after t_{i}	refractory period	0.17	0.17	[58–60]
T_{C} [h]	C-period	2/3	2/3	[20]
T_{D} [h]	D-period	1/3	1/3	[20]
α [h^{-1}]	production rate lipids	955	260	combined with $\beta_{\text{datA}}, \beta^-_{\text{datA}}, \alpha_{\text{d1}}, \alpha^-_{\text{d2}}$ to match ν^* from [40]
D_{l}	noise strength lipids	5000	5000	set to match CV from [22]
			full model: 1000	
			S. Fig. 4.3: 10^6	
ρ [μm^{-3}]	number density	10^6	10^6	[75]
ϕ_0	gene allocation fraction of DnaA	10^{-3}	10^{-3}	to match DnaA conc.
			S. Fig. 4.6: 1.5×10^{-3}	reported in [64]
K_{D}^{p} [μm^{-3}]	dissociation constant DnaA promoter	300	300	[64, 69]
			full model: 400	
n	Hill coefficient DnaA promoter	5	5	[69]
D_{D}	noise strength DnaA	100	100	set to match CV from [22]
D_η	noise strength RIDA	80	80	set to match CV from [22]
λ [h^{-1}]	growth rate	0.35-1.73	0.35-1.73	[22, 40]
			Fig. 4: 0.35	

Table 3.1: Parameters used in the LD/LDDR model One molecule per cubic micrometer corresponds to approximately one nM (1 μm^{-3} = 1.67 nM). section 3.A provides further motivation for chosen parameter values.

the first deactivation mechanism of DnaA that could be identified [105]: The DNA polymerase clamp on newly synthesized DNA forms a complex with ADP and the Hda protein. The resultant ADP-Hda-clamp-DNA complex interacts with ATP-DnaA molecules catalytically and stimulates ATP hydrolysis yielding ADP-DnaA. This system is predominant in the inactivation of DnaA after replication initiation as it strongly represses over-initiation of replication [43]. Importantly, at low growth rates ($\tau_d > T_C$) the replication forks are not overlapping and RIDA is inactive at the moment of replication initiation. RIDA can hydrolyze at least 0.9 molecules of ATP-DnaA per DNA-clamp-Hda complex per minute *in vitro* [63, 88]. Single-cell experiments have shown that the number of DNA-bound sliding clamps increases during the cell cycle, peaking at more than 8 per replication fork [110]. We therefore use a deactivation rate of $\beta_{\text{RIDA}} = 8$ min^{-1} in the LDDR model (See Table 3.1). This RIDA deactivation rate β_{RIDA} is only non-zero during active replication for T_C after initiation and is taken to be constant during this time period. Besides RIDA, a chromosomal site named *datA* can hydrolyze ATP-DnaA via a process called *datA*-dependent DnaA-ATP Hydrolysis (DDAH) [63]. DDAH is temporally regulated over the course of the cell cycle via the protein IHF. The binding of IHF to *datA* increases within about 5-10 minutes, peaks at about 15 minutes, and decreases again about 20-30 minutes after initiation of replication [43, 63]. In the LD model, we neglect this temporal variation in the deactivation rate and take it to be constant for simplicity. In the LDDR model, we have two activity states of DDAH, a high deactivation rate β^+_{datA} from the moment of replication initiation onward ($\tau^+_{\text{datA}} = 0$ h) until 0.2 h after replication initiation ($\tau^-_{\text{datA}} = 0.2$ h) and a low deactivation rate β^-_{datA} during the rest of the cell cycle (See Table 3.1 and Fig. 3.3). *In vitro*, the hydrolysis rate of ATP-DnaA can be at least 1.6 molecules per minute per *datA*, and *in vivo* the deactivation strength of DDAH is about $20-30\%$ of that of RIDA [63]. Therefore, we use a deactivation rate of $\beta_{\text{datA}} = 10$ min^{-1} in the LD and $\beta^+_{\text{datA}} = 10$ min^{-1} in the LDDR model. From the experimental findings on the temporal variation in the IHF binding to *datA* [63], we estimate that the activity of DDAH goes down by a factor of two or three in the low activity state and we use a low deactivation rate of $\beta^-_{\text{datA}} = 5$ min^{-1} (See Table 3.1).

Concerning activation, at least three mechanisms for the production of ATP-DnaA have been characterized: de novo DnaA synthesis; nucleotide dissociation from ADP-DnaA by acidic phospholipids in the cell membrane; and a mechanism involving specific chromosomal DNA sequences termed *DARS* sites [43].

Activation mechanism: acidic phospholipids *In vitro*, acidic lipids such as cardiolipin (CL) and phosphotidylglycerol (PG) can enhance the release of ADP and ATP from DnaA, but Phosphatidylethanolamine (PE), comprising nearly 80% of the phospholipids, is inert [86, 97]. DnaA can bind ATP and ADP in the absence of phospholipids [86]. *In vitro*, CL and PG can restore replication activity of DnaA bound to ADP [43, 86, 93, 101]. *In vivo*, reducing the concentration of acidic lipids leads to growth arrest [93, 102, 103] and inhibited replication initiation [102]; in section 4.A.1, we discuss in more detail the effect of mutations in the acidic lipid synthesis. Based on these experimental observations, we envision the following lipid-mediated activation scenario: DnaA-ADP generated by *datA* and RIDA binds the lipids, causing the ADP to dissociate [86, 97]. After DnaA then

dissociates from the lipids, it rapidly binds ATP in the cytoplasm [86], and subsequently *oriC* [17, 18, 79], initiating replication. This scenario gives rise to the LD/LDDR model, in which the lipid-mediated activation rate is independent of the origin density. We emphasize, however, that while the experiments clearly demonstrate that acidic phospholipids can enhance the release of ADP [86, 97], it remains unclear how important the lipids are for DnaA re-activation *in vivo* [111–113]. Moreover, it is not understood how and where DnaA would be re-activated after it has released ADP. In particular, there is evidence that the rejuvenation of DnaA is contingent on *oriC* [97]. In section 4.C, we discuss these experiments in more detail; here we also analyze a scenario in which lipid-mediated DnaA activation depends on the origin density, and one in which DnaA activation is entirely independent of the lipids. The activation rates of the different types of acidic phospholipids have so far not been characterized experimentally. We combine the experimentally characterized deactivation rates of *datA* and RIDA with the activation rates of *DARS1/2* and the experimentally observed initiation volume per origin v^* to infer reasonable activation rates of the lipids of $\alpha_l\,[l] = 46\ \mathrm{min}^{-1}\,\mu\mathrm{m}^{-3}$ in the LD model and $\alpha_l\,[l] = 12.5\ \mathrm{min}^{-1}\,\mu\mathrm{m}^{-3}$ in the LDDR model.

Activation mechanism: DARS1/2 Experiments have found two activation sites located on the chromosome of *E. coli*: *DARS1* and *DARS2*. ADP-DnaA can form oligomers at *DARS1* and *DARS2*, resulting in the dissociation of ADP and the release of nucleotide-free apo-DnaA, which then binds ATP. *DARS2* requires the binding of the proteins Fis and IHF. The binding of IHF to *DARS2* is cell-cycle regulated: It increases after 10 minutes, peaks after 20 minutes and decreases again 30-40 minutes after initiation of replication [44, 77]. We model this observation via step functions that switch from a low to a high activity state at $\tau^+_{\mathrm{d2}} = 0.2$ h and back to a low activity state at $\tau^-_{\mathrm{d2}} = 2/3\,\mathrm{h} = \mathrm{T_C}$ (See Fig. 3.3). As we could not find an experimental value for the activation rate of *DARS2*, but deleting *DARS2* had a similarly strong effect as deleting RIDA [44], we used a high activation rate of $\alpha^+_{\mathrm{d2}} = 33\ \mathrm{min}^{-1}$ and an arbitrarily low activation rate of $\alpha^-_{\mathrm{d2}} = 0.83\ \mathrm{min}^{-1}$ in the LDDR model (See Table 3.1). Concerning the binding of Fis to *DARS2*, there is no experimental evidence that it is cell-cycle regulated. We therefore do not model Fis explicitly and assume its effect is contained in the values of α^+_{d2} and α^-_{d2}, respectively. Experiments do indicate that the activity of Fis increases with the growth rate [77, 95, 96], but precisely how the binding of Fis to *DARS2* depends on the growth rate of the cell remains to be determined. Since α^+_{d2} contributes to the initiation volume only in the high-growth rate regime of overlapping replication forks, while α^-_{d2} only (weakly) contributes to the initiation volume at low growth rates, see also Eqs. 3.7 and 3.8 below, we assume, for simplicity, that the values of α^+_{d2} and α^-_{d2} are independent of the growth rate. The site *DARS1* was found to be neither cell-cycle regulated nor growth rate-dependent and is approximately ten times weaker than *DARS2 in vitro* [43, 77]. We use a constant activation rate of *DARS1* of $\alpha_{\mathrm{d1}} = 1.67\ \mathrm{min}^{-1}$ in the LDDR model.

As the dissociation constant of the DnaA boxes on the DNA is in the range of $K^{\mathrm{s}}_{\mathrm{D}} = 1-50$ nM and *datA*, *DARS1* and *DARS2* are chromosomal binding sites for DnaA, we use a dissociation constant of $K^{\mathrm{datA}}_{\mathrm{D}} = K^{\mathrm{d1}}_{\mathrm{D}} = K^{\mathrm{d2}}_{\mathrm{D}} = K_{\mathrm{D}} = 50\,\mu\mathrm{m}^{-3}$. Less is known about the dissociation constant of RIDA and the acidic phospholipids and in our model we simply

use the same dissociation constant of $K_D^{\mathrm{RIDA}} = K_D^{\mathrm{l}} = K_D = 50\,\mu\mathrm{m}^{-3}$.

3.B. THE EFFECT OF PROTEIN SYNTHESIS IN THE LD MODEL

Here, we investigate the role of protein synthesis in the LD model and analyse its effect on the amplitude of the oscillations in the active DnaA fraction. As DnaA binds strongly to both ATP and ADP and the concentration of ATP is approximately ten times higher than the concentration of APD in *E. coli* [17], we assume that every newly produced protein binds to ATP right after synthesis. Thus, the change in the total number of DnaA proteins due to protein synthesis equals the change in the ATP-DnaA concentration:

$$\frac{dN_D^{\mathrm{T}}}{dt} = \phi_p^0\,\lambda\,\rho\,V = \frac{dN_{\mathrm{ATP}}^{\mathrm{synth}}}{dt} \tag{3.14}$$

with the gene allocation fraction ϕ_p^0, the growth rate λ, the number density $\rho = N/V$ and the volume V. The change in the total concentration of DnaA is given by

$$\frac{d[D]_{\mathrm{T}}}{dt} = \frac{dN_D^{\mathrm{T}}}{dt}\frac{1}{V} + \frac{d}{dt}\left(\frac{1}{V}\right) N_D^{\mathrm{T}} = \phi_p^0\,\lambda\,\rho - \lambda[D]_{\mathrm{T}} \tag{3.15}$$

As in the previous section, we assume that $[D]_{\mathrm{T}}$ is constant in time, such that $[D]_{\mathrm{T}} = \phi_p^0\,\rho$. Using equation 3.14 and 3.15, and exploiting that $[D]_{\mathrm{T}} = \phi_p^0\,\rho$, we obtain the following expression for the change in the ATP-concentration due to protein synthesis:

$$\frac{d[D]_{\mathrm{ATP}}^{\mathrm{synth}}}{dt} = \frac{dN_{\mathrm{ATP}}^{\mathrm{synth}}}{dt}\frac{1}{V} - \lambda[D]_{\mathrm{ATP}} = \phi_p^0\,\lambda\,\rho - \lambda[D]_{\mathrm{ATP}} = \lambda([D]_{\mathrm{T}} - [D]_{\mathrm{ATP}}) \tag{3.16}$$

By dividing equation 3.16 by the total concentration $[D]_{\mathrm{T}}$ and including the (de)activation terms of *datA* and the lipids, we obtain the change in the active fraction of equation 3.2:

$$\frac{df}{dt} = \tilde{\alpha}_{\mathrm{l}}\,[l]\,\frac{1-f}{\tilde{K}_D^{\mathrm{l}} + 1 - f} - \tilde{\beta}_{\mathrm{datA}}\,[n_{\mathrm{ori}}]\,\frac{f}{\tilde{K}_D^{\mathrm{datA}} + f} + \lambda(1-f) \tag{3.17}$$

The third term on the right-hand side is the additional activation term that comes from protein synthesis, and the fact that newly synthesized proteins rapidly bind ATP; this term is indeed proportional to the growth rate and decreases linearly with the ATP-DnaA fraction.

The effect of protein synthesis on the DnaA oscillations depends strongly on the relative magnitude of the activation and deactivation rates and the growth rate. Fig. 3.8 shows however that the effect of protein synthesis (third term right-hand side Equation 3.2) on the amplitude of the oscillations is small.

3.C. THE LD MODEL IN THE ULTRA-SENSITIVITY REGIME

Figure 3.2c shows that for biological (de)activation rates the amplitude of the oscillations in the active fraction becomes very small. Here we ask whether this effect could be alleviated by bringing the system deeper into the ultra-sensitivity regime. The ultra-sensitivity

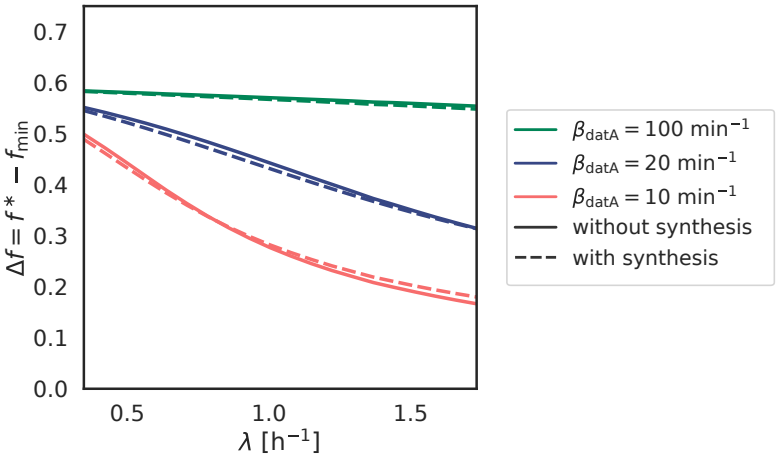


Figure 3.8: Comparison of the LD model with and without synthesis The oscillation amplitude Δf as a function of the growth rate λ for different magnitudes of the activation and deactivation rates ($\alpha_1 = 4.6 \times \beta_{datA}$). The solid lines show the predictions of the LD model that includes the effect of protein synthesis (equation 3.2) while the dashed lines shows the results of the model that does not (equation 3.2 without protein synthesis term). The amplitude of the oscillations decreases with the growth rate, but that this dependence is weaker for higher (de)activation rates.

can be increased by increasing the difference in the dissociation constants K_D^l and K_D^{datA} with respect to the total DnaA concentration $[D]_T$. In the main Figure 3, the dissociation constants of the activator K_D^l and deactivator K_D^{datA} are approximately ten times smaller than the total DnaA concentration $[D]_T = 400\,\mu m^3$ (see Table 3.1). The system is thus already in the ultra-sensitive regime. Here we push the system even deeper in the ultra-sensitivity regime by setting the dissociation constants of both activator and deactivator to $K_D^l = K_D^{datA} = 5\,\mu m^{-3}$, almost two orders of magnitude smaller than the total concentration. Fig. 3.9 shows the amplitude of the oscillations in the active fraction f in this (highly) ultra-sensitive regime. The amplitude of the oscillations at high and intermediate (de)activation rates is slightly higher in this deep ultra-sensitive regime. Importantly, however at low rates ($\beta_{datA} = 10$ min^{-1}) the amplitude of the oscillations drops for high growth rates like in the less ultra-sensitive regime presented in the main section. Therefore, regardless of the degree of ultra-sensitivity, the experimentally reported activation and deactivation rates are too low to explain the experimentally observed high amplitude oscillations in the active initiator fraction [63]. Our modelling predicts that at high growth rates, RIDA and *DARS2* become essential to sustain large amplitude oscillations, as we describe in more detail in section 3.2 on the LDDR model.

3.D. DERIVATION OF ADDER CORRELATIONS FROM SIZE SENSOR

Panels b-d of Fig. 3.6 elucidate how fluctuations in the lipid concentration generate adder correlations in the initiation volume. Panel b shows that lipid concentration fluc-

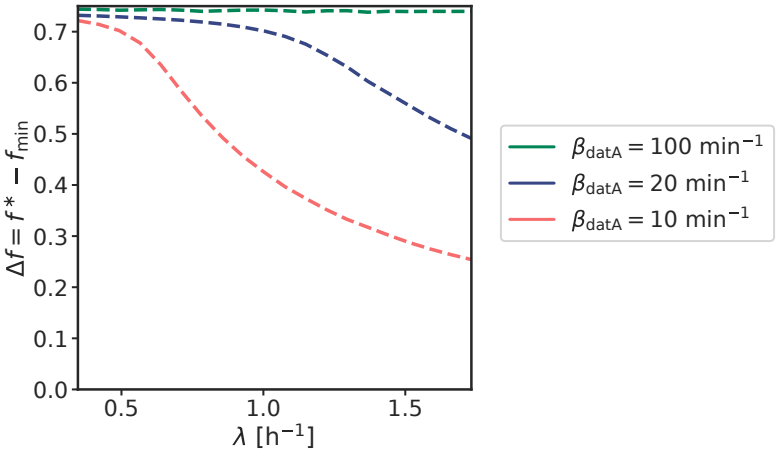


Figure 3.9: The LD model in the ultra-sensitive regime The amplitude Δf of the oscillations in the active fraction f as a function of the growth rate at different magnitudes of the activation and deactivation rates ($\alpha_l = 4.6 \times \beta_{datA}$). The amplitude of the oscillations Δf becomes small for biologically realistic values of the (de)activation rates, even deep in the ultra-sensitive regime. Here, the dissociation constants $K_D^l = K_D^{datA} = 5\,\mu m^{-3}$ for lipid-mediated activation of DnaA and *datA* mediated deactivation, respectively (see Eq. 3.2), is 10 times lower than that used for Figure 3.2.

tuations $l(t) \equiv [l](t)$ regress to the mean on a timescale given by the cell-doubling time $\tau_d = \ln(2)/\lambda$. Here, the thin grey lines are time traces from the simulations, while the solid line is the analytical prediction obtained by solving equation 3.10 subject to an initial concentration fluctuation δl_0:

$$\langle \delta l(t) | l_0 \rangle \equiv \langle l(t) | l_0 \rangle - \langle l \rangle, \tag{3.18}$$

$$= \delta l_0 e^{-\lambda t}, \tag{3.19}$$

$$= \delta l_0 2^{-t/\tau_d}. \tag{3.20}$$

where $\langle l(t) | l_0 \rangle$ is the average lipid concentration at time t given an initial concentration l_0 at time zero and $\langle l \rangle$ is the average lipid concentration; $\langle \delta l(t) | l_0 \rangle$ is thus the average deviation of the lipid concentration from its mean at time t, given an initial concentration fluctuation l_0. Panel c of Fig. 3.6 shows the mapping $v^*([l])$ between the initiation volume v^* and the lipid concentration $[l]$, obtained by neglecting protein synthesis and solving equation 3.2 in steady state; this panel corresponds to panel c of Fig. 3.7. Panel d of Fig. 3.6 demonstrates how the decay of lipid fluctuations shown in panel b (of Fig. 3.6) with the mapping $v^*([l])$ shown in panel c (of Fig. 3.6) causes the initiation volume to regress to the mean on the timescale of the doubling time τ_d:

$$\langle \delta v_n^* | v_0^* \rangle \equiv \langle v_n^* | v_0^* \rangle - \langle v^* \rangle, \tag{3.21}$$

$$= \delta v_0^* 2^{-n}, \tag{3.22}$$

where $\langle v^* \rangle$ is the average initiation volume, v_0^* is the initial initiation volume arising from a spontaneous fluctuation, and $\langle v_n^* | v_0^* \rangle$ is the average initiation volume n cell cycles

later given that initial initiation volume v_0^*. Clearly, fluctuations in the initiation volume relax to the mean via a geometric series, akin to that observed for the volume at birth [6]. Combining $\langle \delta v_n^* | v_0^* \rangle = \delta v_0^* 2^{-n}$ with the definition of the added initiation volume $\Delta v^* \equiv 2v_{n+1}^* - v_n^*$ (see Fig. 3.6) shows that the average added initiation volume $\langle \Delta v^* \rangle$,

$$\langle \Delta v^* \rangle = 2\left(\langle v^* \rangle + \langle \delta v_{n+1}^* | v_0^* \rangle\right) - \left(\langle v^* \rangle + \langle \delta v_n^* | v_0^* \rangle\right) \tag{3.23}$$

$$= 2\left(\langle v^* \rangle + \delta v_0^* 2^{-(n+1)}\right) - \left(\langle v^* \rangle + \delta v_0^* 2^{-n}\right), \tag{3.24}$$

$$= \langle v^* \rangle, \tag{3.25}$$

equals the average initiation volume $\langle v^* \rangle$, independent of the initial initiation volume $v_0^* = \langle v^* \rangle + \delta v_0^*$. Hence, the initiation volume added between successive cell cycles is independent of the initiation volume, and equal to the average initiation volume.

3.E. SIZER AND ADDER CORRELATIONS IN THE LDDR MODEL

The adder or sizer correlations in the initiation volume emerge from the following four ideas: (i) replication is initiated at a critical concentration or critical fraction of active, ATP-bound DnaA; (ii) the mapping between the fraction or concentration of active ATP-bound DnaA and the volume depends on the concentrations and activities of the switch components (Fig. 3.7); (i) and (ii) together imply that fluctuations in the activities and concentrations of the switch components will lead to fluctuations in the initiation volume (Fig. 3.7 c, f, i); (iii) fluctuations in the initiation volume regress on roughly the same timescale as those of the switch components, because the mapping between the initiation volume and the activities or concentrations of the switch components is fairly linear, certainly when the fluctuations are small enough, and the rates of activation and deactivation are faster than the growth rate, which they must be generically in order to generate large-amplitude oscillations in the concentration or fraction of active DnaA; (iv) adder correlations emerge when this timescale is set by the growth rate while sizer correlations emerge when this timescale is significantly faster. These ideas are generic and should apply not only to the LD model of Fig. 3.6 , but also to the full LDDR model. Here, we show that this is indeed the case.

Figure 3.10c demonstrates that also in the full LDDR model, with parameter values estimated from experimental data (see Table 3.1 and section 3.A), adder correlations in the initiation volume emerge from fluctuations in the lipid concentration. The model is described by equation 3.4, but with the lipid fluctuations modelled in the same way as in the stochastic LD model, see equation 3.10. Like the stochastic LD model, the stochastic LDDR model features an eclipse period of $\tau_b = 10$ minutes following replication initiation during which replication cannot be reinitiated [58–60]. In section 3.F below, we argue that this effect is much more generic: any switch component that fluctuates on a timescale set by the growth rate, be it lipids, *datA*, RIDA, or *DARS1/2*, will generate adder correlations in the initiation volume.

Figure 3.10d shows that for a negatively autoregulated initiator protein, fluctuations in the total concentration of DnaA also give rise to sizer correlations in the LDDR model. The stochastic production of DnaA is modelled in exactly the same way as in the LD model, see equation 2.24. Combining this with equation 3.4 yields the following equa-

tion for the dynamics of the number of ATP-bound DnaA molecules:

$$\frac{dN_{\mathrm{D}}^{\mathrm{ATP}}}{dt} = \frac{\tilde{\phi}_{\mathrm{p}}^{0}\,\lambda\,V}{1+\left(\frac{[D]_{\mathrm{T}}}{K_{\mathrm{D}}^{\mathrm{p}}}\right)^{n}} + \xi_{\mathrm{D}}(t) + \left(\alpha_{\mathrm{l}}\,[l]\,V + \alpha_{\mathrm{d1}}\,n_{\mathrm{ori}}(t-\tau_{\mathrm{d1}}) + \alpha_{\mathrm{d2}}(t)\,n_{\mathrm{ori}}(t-\tau_{\mathrm{d2}})\right)\frac{[D]_{\mathrm{ADP}}}{K_{\mathrm{D}}+[D]_{\mathrm{ADP}}}$$
$$-\left(\beta_{\mathrm{datA}}(t)+\beta_{\mathrm{rida}}(t)\right)\,n_{\mathrm{ori}}\,\frac{[D]_{\mathrm{ATP}}}{K_{\mathrm{D}}+[D]_{\mathrm{ATP}}}. \tag{3.26}$$

As in the stochastic LD model of section 3.7, a new round of replication is initiated when the ATP-DnaA concentration reaches the critical concentration for replication initiation. The cell then divides a constant time τ_{cc} later. The volume grows exponentially with growth rate λ and upon cell division the volume and copy numbers of DnaA and the number of origins are halved. And as for the other stochastic switch models, this model features an eclipse period of $\tau_{\mathrm{b}} = 10$ minutes following replication initiation during which replication cannot be reinitiated [58–60].

Fig. 3.10d shows that in the full LDDR model, like in the LD model, DnaA copy number fluctuations give rise to sizer correlations. Negative autoregulation speeds up the regression of the initiation threshold to its mean, turning the system (back) into a sizer. Importantly, it remains to be experimentally verified how strong the effect of negative autoregulation of the protein DnaA is.

3.F. FLUCTUATIONS IN OTHER SWITCH COMPONENTS

The activities of *datA*, *DARS1/2* and RIDA are all influenced by other proteins. IHF affects the activity of *datA* [63] and *DARS2* [77], while Fis modulates the activity of *DARS2* [77]. In addition, the activity of RIDA is influenced by Hda [88]. Fluctuations in these proteins will lead to fluctuations in the respective activation and deactivation rates, just like lipid fluctuations affect the activation rate; in fact, since these proteins are present in (much) lower concentrations than the acidic phospholipids (even though the most potent lipid, cardiolipin, constitutes only a small fraction, 5%, of the total lipid concentration [93]), their fluctuations are likely to be stronger. The fluctuations in the (de)activation rates caused by these proteins will, in turn, generate fluctuations in the concentration or active fraction of ATP-DnaA as a function of the volume, thus causing fluctuations in the initiation size. Because the activation and deactivation rates are typically higher than the growth rate (see section 3.A and Fig. 3.4), fluctuations in the initiation volume regress on the same timescale as that of the fluctuations in the switch components. If the switch components decay on a timescale set by the growth rate, because the proteins are neither degraded actively nor produced via strong feedback control, then their fluctuations will give rise to fluctuations in the initiation volume that relax on the timescale set by the growth rate. These fluctuations will therefore also generate adder correlations in the initiation volume, just like the lipids do. We thus argue that the idea that fluctuations in the switch components can generate adder correlations is general.

To provide support for this idea, we study how fluctuations in the activity of RIDA in the full LDDR model, as induced by e.g. Hda [88], propagate to fluctuations in the initiation volume. The system is modeled in exactly the same way as in the previous section (S33.E), except that now the lipid concentration is constant while the activity of

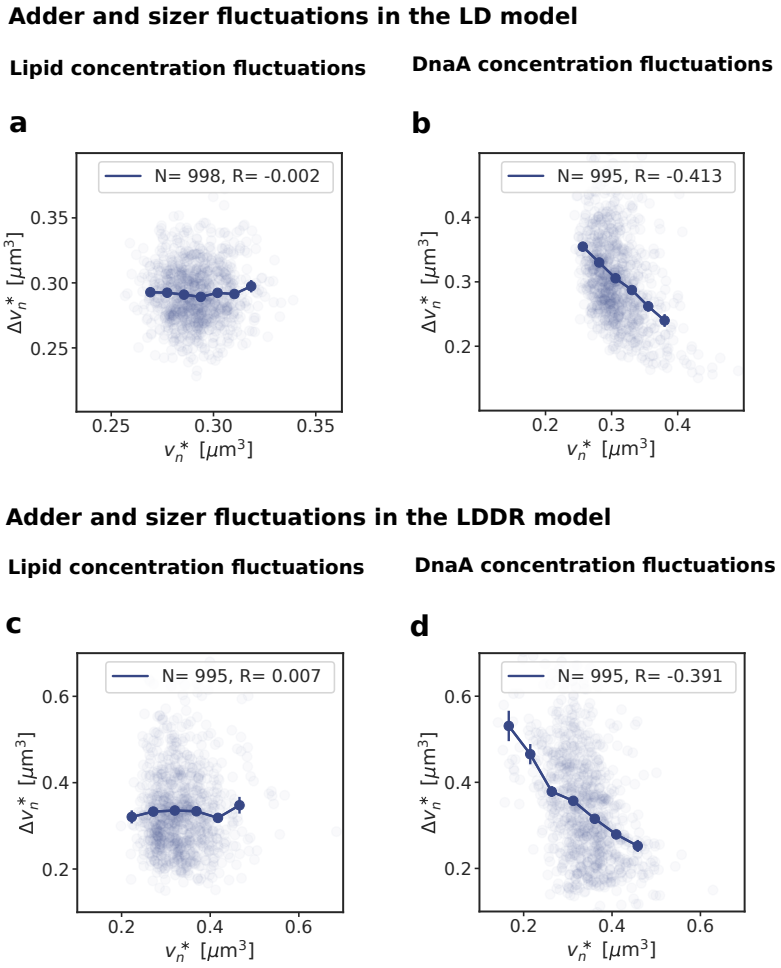


Figure 3.10: Lipid and DnaA concentration fluctuations generate adder and sizer correlations, respectively, in both the LD and LDDR model. (a, b) Scatter plot of the added initiation volume between successive initiation events, $\Delta v_n^* \equiv 2v_{n+1}^* - v_n^*$, and the initiation volume v_n^* in the presence of lipid concentration fluctuations (a) and DnaA concentration fluctuations (b) in the LD model. It is seen that in the presence of lipid fluctuations, Δv_n^* is independent of v_n^* (a), as is characteristic for an adder. In contrast, in the presence of DnaA concentration fluctuations Δv_n^* is anti-correlated with v_n^*; a feature characteristic for a sizer. (c, d) Scatter plot of the same data, but for the LDDR model. Also in the LDDR model, lipid fluctuations generate adder correlations in the initiation volume (c), while DnaA concentration fluctuations yield sizer correlations (d). The dark blue line shows the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The doubling time in both models is $\tau_d = 2$ h, corresponding to non-overlapping replication forks and the number of data points N and the Pearson correlation coefficient R are indicated.

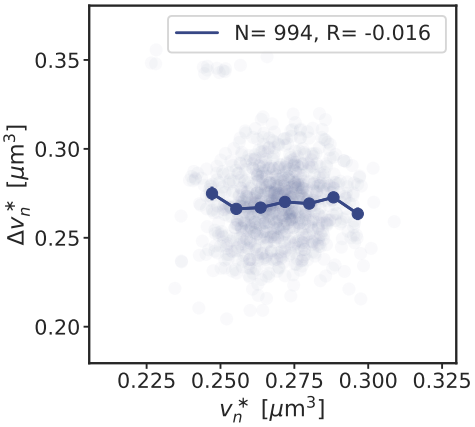


Figure 3.11: RIDA concentration fluctuations generate adder correlations in the LDDR model. Scatter plot of the added initiation volume between successive initiation events, $\Delta v_n^* \equiv 2v_{n+1}^* - v_n^*$, and the initiation volume v_n^*. While the lipids and *datA* control the initiation volume in the low growth-rate regime, *DARS2* and RIDA control the initiation volume in the high growth-rate regime of overlapping replication forks. The Figure shows that in this regime, RIDA fluctuations generate adder correlations in the initiation volume, as observed experimentally [23, 24]. The cell-doubling time is $\tau_d = 0.55$ h ≈ 33 min, corresponding to a growth rate of $\lambda = 1.25$ h^{-1}. The correlations in the initiation volume in the low growth regime are shown in Fig. 3.10. The dark blue line shows the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The number of data points N and the Pearson correlation coefficient R are indicated.

RIDA fluctuates:

$$\tilde{\beta}_{\text{rida}}(t) = \overline{\tilde{\beta}}_{\text{rida}} + \eta(t), \tag{3.27}$$

$$\frac{d\eta}{dt} = -\lambda\eta + \xi_\eta(t). \tag{3.28}$$

Here, $\overline{\tilde{\beta}}_{\text{rida}}$ is the mean RIDA activity, while $\xi_\eta(t)$ models Gaussian white noise with strength $\langle \xi_\eta(t)\xi_\eta(t')\rangle = 2D_\eta\delta(t-t')$, such that the colored noise η describes fluctuations with zero mean that decay on a timescale set by the growth rate λ.

Figure 3.11 shows that, as anticipated, fluctuations in the activity of RIDA generate adder correlations in the initiation volume. Adder correlations will emerge from fluctuations in switch components that relax on a timescale given by the growth rate. This is one of the central findings of our study.

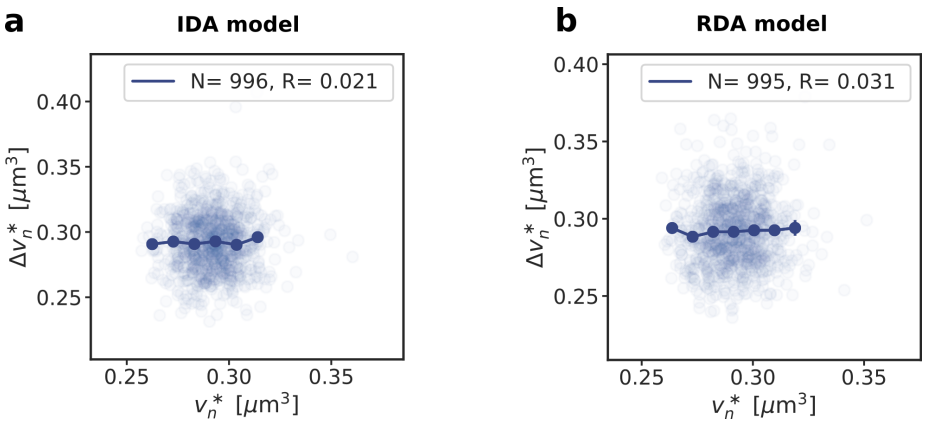


Figure 3.12: The adder correlations in the initiation volume in the LD model are robust to a more loose coupling between the division and the replication cycle. (a, b) The added volume per origin between consecutive replication initiation events, $\Delta v_{\mathrm{n}}^* = 2\, v_{\mathrm{n+1}}^* - v_{\mathrm{n}}^*$, as a function of the initiation volume v_{n}^*. The dark blue lines show the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The number of data points N and the Pearson correlation coefficient R are indicated. The doubling time in all plots is $\tau_{\mathrm{d}} = 2$ h. In the LD model, noise in the lipid concentration (according to equation 3.10) gives rise to adder correlations in the initiation volume even if the division cycle is coupled more loosely to replication initiation: While in the IDA model, cell division is triggered completely independently via a separate division adder (a), in the RDA model division is triggered when an on average constant volume has been added from replication initiation to cell division (b) (compare to Fig. 3.6a or to Fig. 3.10a). See section 2.D for more details on the IDA and RDA model.

Artist impression of the origin of replication of the *E. coli* chromosome
Watercolor, 2023
Mareike Berger

4

FULL SWITCH-TITRATION MODEL: MODEL VALIDATION AND PREDICTIONS

In the bacterium E. coli, both titration and activation of the initiator protein DnaA control replication initiation. Here, we study by mathematical modelling how these two mechanisms interact to generate robust replication-initiation cycles. In the high growth rate regime, where titration alone fails to ensure stable replication initiation, the DnaA activation switch becomes essential for stable replication initiation. Conversely, while the activation switch alone yields robust rhythms at high growth rates, titration can strongly enhance the stability of the switch at low growth rates. Our analysis thus predicts that both mechanisms together drive robust replication cycles at all growth rates. We validate our model by comparing key predictions to experimental observations and make several novel experimentally testable predictions. As the role of acidic phospholipids in activating DnaA is still under debate, we also test the effect of removing the lipids from the model. We find that while an activation switch without the lipids becomes unstable, combining it with a concentration cycle can nevertheless drive robust replication cycles.

The contents of this chapter have been published in Nature Communications **13**, 6556 (2022) [39].

Passing on the genetic information from one generation to the next is essential for every living organism. The bacterium *Escherichia coli* initiates replication at a critical volume per origin and adds an on average constant volume between successive initiation events independent of the initiation size [22, 40]. Initiating DNA replication an on average constant volume per origin ensures that the cell copies its chromosome exactly once per cell cycle independent of the growth rate. Recent single-cell measurements revealed that the initiation volume varies only by about 10% for any measured growth rate, less than most other cell-cycle parameters [22, 23]. A molecular model that can explain how *E. coli* regulates replication initiation with such a high precision has however been lacking.

Experimental evidence indicates that *E. coli* regulates replication initiation via two distinct mechanisms [43, 64, 78]: titration and activation. In *E. coli*, the initiation of DNA replication is controlled by the binding of the initiator protein DnaA to the origin, a special region on the chromosome at which replication can be initiated. In the initiator titration models, the initiator protein first needs to fill up the titration sites that are distributed all over the chromosome before it can initiate replication at the origin [12, 14] (Fig. 2.1b). In chapter 2, we have shown that the titration mechanism can ensure stable cell cycles at low growth rates, but gives rise to re-initiation events at high growth rates where replication forks overlap. The second class of models is based on a switch of the initiator protein DnaA between an active and an inactive form (Fig. 3.1b) [33, 43, 63, 76–78]. The protein DnaA can either be bound to ATP or ADP, but only ATP-DnaA can initiate DNA replication at the origin [17, 18, 79, 80]. A switch of the initiator protein DnaA can ensure stable replication cycles at all growth rates (see chapter 3).

Both the activation switch and the titration model are stable in the presence of biochemical noise. The concentrations do not diverge, also not in the titration-based system at high growth rates as we saw in chapter 2 (Fig. 2.3b). Yet, the precision of replication initiation differs markedly between the respective models (Fig. 4.2). A titration-based mechanism breaks down at sufficiently high growth rates, causing premature reinitiation events and a dramatic rise of the coefficient of variation (CV) in the initiation volume; even in the absence of any biochemical noise, the CV becomes larger than that reported experimentally [22, 23] (Fig. 2.4).

As we discussed in detail in chapter 2, he failure of the titration-based mechanism arises from the different scaling of the protein synthesis and the titration-site formation rate with the growth rate. The synthesis rate of DnaA scales with the growth rate (see Eq. 2.17). In contrast, the titration-site formation rate is set by the DNA replication rate, which varies only very weakly with the growth rate [40]. The different scaling of these two timescales with the growth rate means that a titration-based mechanism must fail inevitably at sufficiently high growth rates — because this prediction is based on a comparison of two timescales, it is insensitive to the details of the model (see also section 2.3). As a titration-based mechanism is not sufficient for generating robust replication-initiation cycles at all growth rates, another mechanism is essential. As we show in chapter 2 (Fig. 2.5a/b), transient suppression of protein synthesis by SeqA after replication initiation [58] can prevent reinitiation events at high but not at intermediate growth rates. Indeed, these reinitiation events still causes the CV of a system based on only titration and SeqA to rise strongly (Fig. 2.6). We thus argue that the switch is essential to generate robust replication cycles at all growth rates.

But could a switch be sufficient? Clearly, the work of the previous chapter shows it could: while a system based on titration alone must fail at high growth rates, one based on a switch alone could generate stable replication cycles at all growth rates. Nonetheless, the experiments indicate that the system combines a switch with titration [14, 46, 54]. What could be the benefit of adding titration and blocking DnaA synthesis via the protein SeqA to the switch?

To answer this question, in this chapter we first include the homogeneously distributed titration sites in the activation switch model. The model shows that at low growth rates the titration sites transiently lower the concentration of ATP-DnaA that is available for initiating replication at the origin. The titration sites therefore help the switch by shaping the oscillations in the *free concentration* of ATP-bound DnaA, such that the precision of replication initiation in the presence of noise is significantly enhanced. We show that a concentration cycle, as induced by titration, can generically enhance an activation cycle, as driven by the switch, by increasing the steepness of the oscillations (the "gain"), which dampens the propagation of fluctuations in the free concentration of active DnaA to the initiation volume. Titration can thus protect a switch from fluctuations in its components.

We then combine the switch-titration model with SeqA and show that in this full model the oscillations in the free ATP-DnaA concentration are large in all growth regimes (section 4.3). We thus argue that *E. coli* has evolved an elaborate set of mechanisms that act synergistically to create robust replication-initiation cycles at all growth rates.

4.1. COUPLING TITRATION WITH DNAA ACTIVATION ENHANCES ROBUSTNESS

To combine the activation switch with titration, we make the following assumptions: 1) Since the affinities of ATP-DnaA and ADP-DnaA for their promoters are fairly similar [69], we assume that active and inactive DnaA have the same affinity for the promoter and only DnaA not bound to titration sites can repress the promoter; 2) the bound and free DnaA are activated and deactivated with the same rate; 3) Since the affinities of ADP-DnaA and ATP-DnaA for the titration sites are fairly similar [54], we assume that the affinities of inactive and active DnaA for the titration sites are the same; 4) replication is initiated when the free concentration of active DnaA reaches a threshold.

Concerning assumption 2, it remains an open question whether all DnaA proteins or only the freely diffusing DnaA can be activated and deactivated via the switch components. Both scenarios could be envisioned: While it might seem more natural to assume that only free DnaA can be activated or deactivated, also DnaA that is bound to titration sites might be in contact with the acidic phospholipids or with the site *datA* via supercoiled DNA. Additionally, as RIDA is moving along the entire chromosome during DNA replication, every titration site will be in the proximity of RIDA once and bound DnaA could be inactivated at that moment. Importantly, however, when the affinities of the two nucleotide bound forms of the DnaA to the titration sites are equal and the binding dynamics are fast (section 2.C), the active fraction in the cytoplasm g equals the total active fraction f, irrespective of whether the activation and deactivation reactions happen only in the cytoplasm or also on the DNA. This question only affects the magni-

tude of the activation and deactivation rates: If only the free DnaA can be (de)activated by the components of the switch, activation and deactivation rates become lower because fewer DnaA proteins are available. If the dissociation constants of the activators and deactivators are however lower than the free DnaA concentration, the system remains in the ultra-sensitivity regime and the titration sites affect the magnitude of the (de)activation rates of the switch only weakly. We therefore here assume out of simplicity that all DnaA, no matter whether bound or unbound to titration sites, can be (de)activated by the switch components (assumption 2).

Using assumption 1, we model the change in the total number of DnaA proteins $N_{\mathrm{T}}(t)$ in the cell via equation 2.17. To obtain the ATP-DnaA fraction, we model the change in the total number of ATP-DnaA proteins $N_{\mathrm{D}}^{\mathrm{ATP}}(t)$ explicitly. As newly produced DnaA proteins are more likely to bind ATP rather than ADP we add the DnaA production term to the change in the total number of ATP-DnaA proteins. Including again all known (de)activators as discussed in the LDDR model (section 3.2) and using assumption 2, we obtain the following expression for the change in the number of ATP-DnaA proteins $N_{\mathrm{D}}^{\mathrm{ATP}}(t)$:

$$\begin{aligned}\frac{dN_{\mathrm{D}}^{\mathrm{ATP}}}{dt} = &\frac{\tilde{\phi}_{\mathrm{p}}^{0}\,\lambda\,V}{1+\left(\frac{[D]_{\mathrm{T,f}}}{K_{\mathrm{D}}^{\mathrm{p}}}\right)^{n}}\\ &+\left(\alpha_{\mathrm{l}}\,[l]\,V+\alpha_{\mathrm{d1}}\,n_{\mathrm{ori}}(t-\tau_{\mathrm{d1}})+\alpha_{\mathrm{d2}}(t)\,n_{\mathrm{ori}}(t-\tau_{\mathrm{d2}})\right)\frac{[D]_{\mathrm{ADP}}}{K_{\mathrm{D}}+[D]_{\mathrm{ADP}}}\\ &-\left(\beta_{\mathrm{datA}}(t)+\beta_{\mathrm{rida}}(t)\right)n_{\mathrm{ori}}\,\frac{[D]_{\mathrm{ATP}}}{K_{\mathrm{D}}+[D]_{\mathrm{ATP}}}.\end{aligned} \tag{4.1}$$

The active initiator concentration $[D]_{\mathrm{ATP}}$ is obtained by dividing the number of ATP-DnaA proteins $N_{\mathrm{D}}^{\mathrm{ATP}}(t)$ by the volume $V(t)$ and the active initiator fraction $f(t)$ is obtained by dividing the number of ATP-DnaA proteins $N_{\mathrm{D}}^{\mathrm{ATP}}(t)$ by the total number of DnaA proteins $N_{\mathrm{T}}(t)$. The total free DnaA concentration $[D]_{\mathrm{T,f}}$ is obtained by using assumption 3) and solving a quadratic equation (section 2.C). Exploiting assumption 3 and fast binding and unbinding dynamics (section 2.C), the fraction $g = [D]_{\mathrm{ATP,f}}/[D]_{\mathrm{T,f}}$ of the concentration of free ATP-DnaA $[D]_{\mathrm{ATP,f}}$ over the concentration of free total DnaA $[D]_{\mathrm{T,f}}$ is equal to the fraction of the total ATP-DnaA concentration over the total DnaA concentration per cell $f = [D]_{\mathrm{ATP}}/[D]_{\mathrm{T}}$. The free ATP-DnaA concentration is therefore given by the concentration of free DnaA $[D]_{\mathrm{T,f}}$ times the active fraction of DnaA f:

$$[D]_{\mathrm{ATP,f}}(t) = [D]_{\mathrm{T,f}}(t) \times f(t) \tag{4.2}$$

Comparing Fig. 4.1a and d shows that including the titration sites in the LDDR model leads to sharper oscillations in the *free concentration* of ATP-DnaA at low growth rates. In the LDDR model, the ATP-DnaA concentration first decreases strongly after replication initiation due to the combined action of the site *datA* and RIDA and then rises again when the activation sites DARS1/2 are being doubled. At low growth rates, the fixed doubling time of e.g. DARS2 τ_{d2} is much shorter than the cell-doubling time τ_{d} ($\tau_{\mathrm{d2}} = 0.2\,\mathrm{h} \ll \tau_{\mathrm{d}} = 2\,\mathrm{h}$), leading to a relatively high ATP-DnaA concentration during most of the cell cycle. The resulting slow rise in the ATP-DnaA concentration towards the initiation

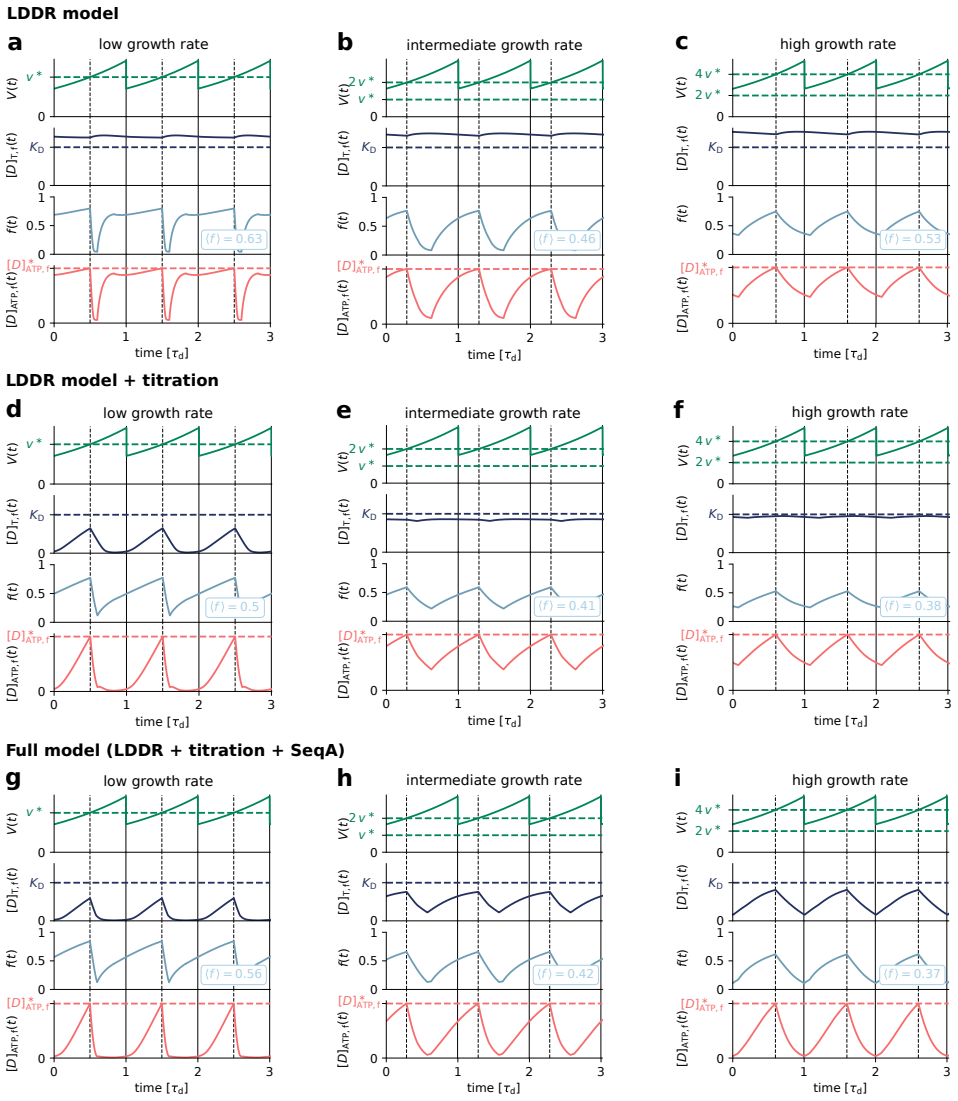


Figure 4.1: Combining the activation switch with tiration sites and SeqA gives rise to large amplitude oscillations in the active free DnaA concentration at all growth rates (a-i) The volume of the cell $V(t)$ (in units of μm^3), the free DnaA concentration $[D]_{\mathrm{T,f}}$ (in units of μm^{-3}), the fraction of DnaA $f(t)$ that is bound to ATP (irrespective of whether the DnaA is in the cytoplasm or on the titration sites) and the concentration of free ATP-DnaA $[D]_{\mathrm{ATP,f}}(t)$ (in units of μm^{-3}) as a function of time (in units of the doubling time τ_{d}) for $\tau_{\mathrm{d}} = 2$ h (a, d, g), $\tau_{\mathrm{d}} = 35$ min (b, e, h) and $\tau_{\mathrm{d}} = 25$ min (c, f, i)). The average active fraction over one cell cycle $\langle f \rangle$ is indicated in light blue in the third panel. While combining the LDDR model with titration helps to shape the oscillations at low growth rates, it does not significantly affect the free DnaA concentration at high growth rates. Adding the effect of SeqA has no strong impact at low growth rates, but strongly increases the oscillations in the free total DnaA concentration $[D]_{\mathrm{T,f}}$ at intermediate and high growth rates.

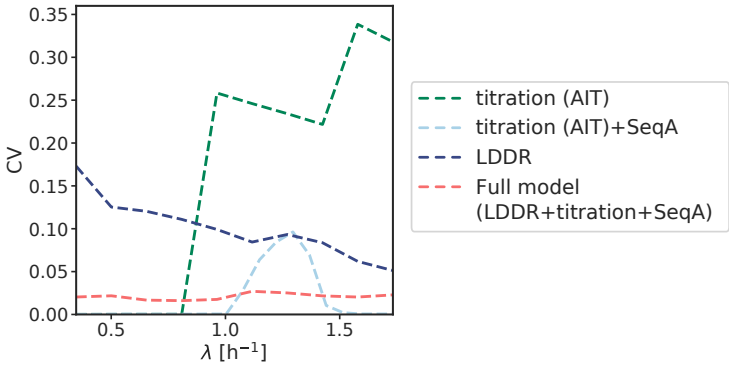


Figure 4.2: Combining the DnaA activation switch with titration and SeqA generates robust replication-initiation cycles over a wide range of growth rates. The coefficient of variation $CV = \sigma/\mu$ with the standard deviation σ and the average initiation volume $\mu = \langle v^* \rangle$ as a function of the growth rate for different models in the presence of noise in the lipid concentration. Even in the absence of biochemical noise in DnaA synthesis, the titration model gives rise to a very high CV at high growth rates, due to premature reinitiation (Fig. 2.3b). Adding SeqA to the titration model can reduce the CV at high, but not at intermediate growth rates (Fig. 2.5b). The large coefficient of variation in the LDDR model at low growth rates is reduced significantly by the titration sites. Conversely, the LDDR model prevents the reinitiation events that inevitably occur at intermediate growth rates in the AIT+SeqA model. Combining DnaA activation with titration thus enhances the robustness of replication initiation at all growth rates, also in the presence of noise in DnaA synthesis (Fig. 4.4). All models include an eclipse period of about 10 minutes following replication initiation to prevent immediate reinitiation [58–60]. (See Table 3.1 for all parameters.)

threshold (Fig. 4.1a) leads to large variations in the initiation volume in the presence of noise in the lipid concentration (Fig. 4.2). Fig. 4.1d shows that including titration sites can significantly sharpen the oscillations in the free ATP-DnaA concentration at low growth rates. As in this growth regime, titration sites are synthesized faster than DnaA proteins, the free concentration drops rapidly after replication initiation and remains low during most of the cell cycle. Only when all titration sites have been filled begins the free concentration to rise and replication can be initiated (Fig. 4.1d). Indeed, in the switch-titration model the precision of replication initiation in the presence of noise is significantly enhanced (Fig. 4.2). At intermediate and high growth rates however, the rate at which new titration sites are being synthesized after initiation is comparable or even lower than the synthesis rate of new DnaA proteins. In this regime, the oscillations in the free DnaA concentration become weaker and the shape of the oscillations in the free ATP-DnaA concentration is dominated by the switch (Fig. 4.1e/f).

4.2. TITRATION CAN GENERICALLY ENHANCE AN ACTIVATION SWITCH

In the previous section, we argued that the combined switch-titration system is more robust to fluctuations than either mechanism alone (Fig. 4.2). The LDDR model contains

however many parameters. Therefore the question arises how general this prediction is. Here, we argue that a concentration cycle, as created by titration, can *generically* enhance an activation cycle, as induced by the switch.

To show this, we consider a minimal model of the switch, namely the LD model. We will optimize this system by minimizing the coefficient of variation in the initiation volume, subject to plausible experimental constraints. We then add to this optimal switch the titration mechanism, *keeping all the parameters of the switch (and also the titration system) the same.* As Fig. 4.3a shows, the combined system is more robust than the optimal system based on only the switch. Clearly, adding titration makes it possible to beat the precision limit of the switch. We then show mathematically how a concentration cycle can, generically, enhance an activation cycle: it can increase the sharpness of the oscillations, the "gain", which means that fluctuations in the cytoplasmic concentration of active DnaA propagate less to fluctuations in the initiation volume (see Fig. 4.3b). This underscores the principal finding of our study: while the switch helps titration by preventing premature reinitiation at high growth rates, titration can help the switch by sharpening the oscillations, increasing the precision of replication initiation.

Concretely, a major source of noise in the switch are the fluctuations in the activation and deactivation components; the switch is inherently fairly robust to fluctuations in the total concentration of DnaA (see Fig. 3.7). We therefore study the coefficient of variation (CV) in the initiation volume arising from lipid fluctuations. In our minimal model of the switch, the LD model, the experimental constraints are: the initiation volume [23, 41, 55] and the maximum (de)activation rates [63] (see Table S2). The optimization parameters are the critical fraction f^* for replication initiation, and the dissociation constants $\tilde{K}_{\mathrm{D}}^{\mathrm{l}} = K_{\mathrm{D}}^{\mathrm{l}}/[D]_{\mathrm{T}}$ and $\tilde{K}_{\mathrm{D}}^{\mathrm{datA}} = K_{\mathrm{D}}^{\mathrm{datA}}/[D]_{\mathrm{T}}$. The noise strength D_{l} only sets the scale for the CV of the initiation volume, and does not affect the outcome of the optimization procedure; it is set such that the CV is comparable to that measured experimentally at high growth rates (see Fig. 4.3a). The parameters of the titration system, the number of titration sites n_{s} and the titration-site affinity $K_{\mathrm{D}}^{\mathrm{s}}$, are taken to be the same as in the main text and in the rest of the SI (see Table 2.1). We then add this titration system to the optimal switch, keeping the parameters of the optimal switch and the titration system the same; also the total DnaA concentration is the same in all three systems. There is only one parameter that remains to be specified in the combined system, which is the threshold for replication initiation, $[D]_{\mathrm{ATP}}^*$; this is set such that the initiation volume matches that observed experimentally.

Fig. 4.3a shows that adding titration to the best switch reduces the CV in the initiation volume. Like any cellular system [107, 114], the robustness of the DnaA activation switch is fundamentally limited by constraints on protein copy numbers and reaction rates and externally induced (extrinsic) fluctuations in these quantities. Our work shows that if these limit the precision of the switch in controlling replication initiation, then adding titration to the switch is a useful strategy to lift these limitations. Titration can help the switch, because a concentration cycle, as generated by titration, can generically enhance the precision of an activity cycle, as driven by the switch. To see this, we consider the mapping between the concentration of cytoplasmic, active DnaA, $[D]_{\mathrm{ATP,f}}$, and the volume per origin v of the cell, see Fig. 4.3b. The former is given by $[D]_{\mathrm{ATP,f}} = [D]_{\mathrm{T,f}} \times f$, where $[D]_{\mathrm{T,f}}$ is the total concentration of cytoplasmic DnaA (i.e, active and inactive) and

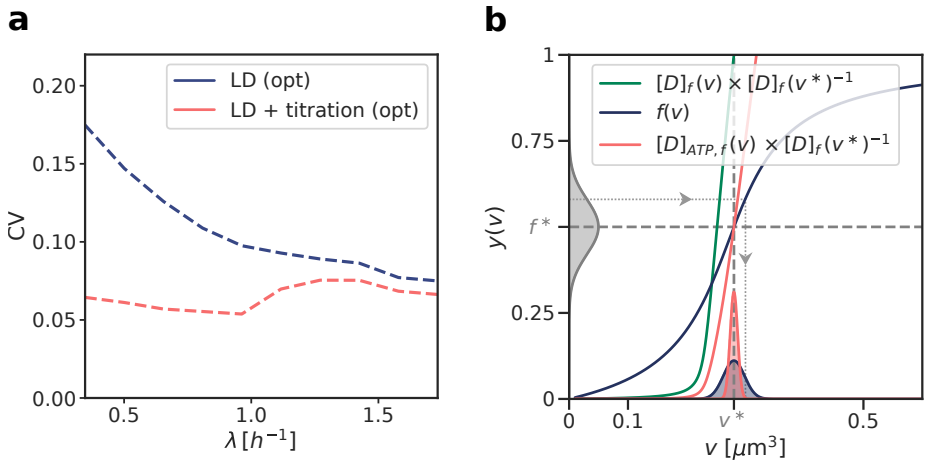


Figure 4.3: Combining the activation switch with titration enhances robustness of replication initiation in the presence of noise in the lipid concentration at all growth rates (a) The system that combines a switch (LD) with titration has a lower Coefficient of Variation (CV) in the initiation volume than the system based on the switch only. This is because a concentration cycle, as induced by titration, can generically enhance an activation cycle, as driven by the switch, as illustrated in the panel on the right. (b) The total concentration of cytoplasmic DnaA, $[D]_{\mathrm{T,f}}$, normalized by its value at the initiation volume v^*, as a function of the cell volume per origin v (green line); the active fraction of DnaA, f, as a function of v (blue line); the cytoplasmic concentration of active DnaA, $[D]_{\mathrm{ATP,f}} = [D]_{\mathrm{T,f}} \times f$, normalized by its value at v^*, as a function of v (red line). It is seen that the gain in the combined system, given by the slope of $[D]_{\mathrm{ATP,f}}$ (red line) at $v = v^*$ is higher than that of the switch-only system, given by the slope of f (blue line) at $v = v^*$: multiplying an activation cycle $f(v)$ with a concentration *cycle* $[D]_{\mathrm{T,f}}(v)$, instead of a constant concentration, leads to sharper oscillations in $[D]_{\mathrm{ATP,f}}(v)$. As a result, fluctuations in f and hence $[D]_{\mathrm{ATP,f}}$ propagate less to fluctuations in the initiation volume: the distribution of v^* in the combined system (red distribution x-axis) is narrower than that of the switch-only system (blue distribution x-axis). The mapping $f(v)$ is obtained by solving Eq. 3.2 in steady state without the protein synthesis term and $[D]_{\mathrm{T,f}}(v)$ is obtained by solving Eq. 2.30. See Table 3.1 for parameters.

f is the fraction of active DnaA, see Eq. 4.2. The general idea is then that oscillations in $[D]_{\mathrm{T,f}}$ (green line of Fig. 4.3b), as induced by titration, can conspire with the oscillations in the active fraction f (blue line of Fig. 4.3b), as driven by the switch, to generate sharper oscillations in $[D]_{\mathrm{ATP,f}}$ (red line of Fig. 4.3b); and these sharper oscillations mean that fluctuations in $[D]_{\mathrm{ATP,f}}$ lead to smaller fluctuations in the initiation volume, as illustrated in Fig. 4.3b.

To make this mathematically concrete, we will exploit that the fluctuations in the switch components cause fluctuations in the active fraction $f(v)$ and not the total cytoplasmic DnaA concentration $[D]_{\mathrm{T,f}}(v)$, which is controlled by titration. The variance of the fluctuations in $[D]_{\mathrm{ATP,f}}$ that arise from fluctuations in the switch components is then given by

$$\sigma^2_{[D]_{\mathrm{ATP,f}}} = [D]^2_{\mathrm{T,f}}\sigma^2_f, \tag{4.3}$$

where σ^2_f is the variance of the fluctuations in the active fraction resulting from the switch. Linearizing the input-output relation $[D]_{\mathrm{ATP,f}}(v)$, and using the rules of error propagation [114], the variance in the initiation volume is given by

$$\sigma^2_{v^*} = \frac{\sigma^2_{[D]_{\mathrm{ATP,f}}}}{g^2_{D\to v}}, \tag{4.4}$$

where $g_{D\to v} = d[D]_{\mathrm{ATP,f}}/dv$ is the "gain", which determines how fluctuations in $[D]_{\mathrm{ATP,f}}$ propagate to variations in the volume per origin v, see Fig. 4.3b; both the numerator and denominator of Eq. 4.4 are evaluated at $v = v^*$. Noting that $[D]_{\mathrm{ATP,f}}(v) = [D]_{\mathrm{T,f}}(v) \times f(v)$ and using Eq. 4.3 we find that

$$\sigma^2_{v^*} = \frac{[D]^2_{\mathrm{T,f}}\sigma^2_f}{[D]^2_{\mathrm{T,f}}(df/dv)^2 + (d[D]_{\mathrm{T,f}}/dv)^2 f^2 + 2[D]_{\mathrm{T,f}} f (d[D]_{\mathrm{T,f}}/dv)(df/dv)} \tag{4.5}$$

$$= \frac{\sigma^2_f}{(df/dV)^2 + (d[D]_{\mathrm{T,f}}/dv)^2 f^2/[D]^2_{\mathrm{T,f}} + 2[D]_{\mathrm{T,f}} f (d[D]_{\mathrm{T,f}}/dV)(df/dv)/[D]^2_{\mathrm{T,f}}} \tag{4.6}$$

$$\leq \frac{\sigma^2_f}{(df/dv)^2} = \sigma^{2,f}_{v^*}. \tag{4.7}$$

Importantly, df/dv and σ^2_f are properties of the switch, and, in comparing the combined to the switch-only system, are evaluated at the same initiation volume $v = v^*$ in the two systems. As a result, $\sigma^{2,f}_{v^*}$ is the variance in the initiation volume of the switch-only system, in which the oscillations in the active, cytoplasmic DnaA concentration, $[D]_{\mathrm{ATP,f}}(v)$, are only driven by the activation cycle $f(v)$ and $d[D]_{\mathrm{T,f}}/dv$ is zero. Eq. 4.6, therefore, shows that by matching the concentration cycle to the activation cycle, such that the total cytoplasmic concentration $[D]_{\mathrm{T,f}}$ rises when the active fraction $f(v)$ rises and both $d[D]_{\mathrm{T,f}}/dv$ and df/dv are non-zero, the concentration cycle of titration can help the activation cycle of the switch by reducing the variance in the initiation volume.

To conclude, we showed that a concentration cycle, as generated by titration, can generically enhance an activation cycle, as driven by the switch, by increasing the steepness of the oscillations; this tames the propagation of fluctuations in the free concentra-

tion of active DnaA to the initiation volume [114] (Fig. 4.3). Combining the switch with titration can thus protect the system against fluctuations in the switch components.

4.3. SWITCH-TITRATION-SEQA MODEL: LARGE OSCILLATIONS AT ALL GROWTH RATES

In the full model, we combine the LDDR model with titration sites and the effect of SeqA to transiently block DnaA production. In section 4.1 we have demonstrated that adding titration to the switch leads to sharper oscillations in the free ATP-DnaA concentration at low, but not at high growth rates, as compared to oscillations driven by the switch alone. Including the effect of SeqA to transiently block DnaA synthesis allows for sharp oscillations in the total free DnaA concentration also at intermediate and high growth rates (Fig. 4.1h and i): By transiently blocking DnaA production after replication initiation, the newly produced titration sites can lower the free DnaA concentration. When the blocked period is over, the free DnaA concentration increases again as new proteins are synthesized faster than titration sites in this regime. These oscillations in the free concentration and the oscillations in the active fraction together lead to large and sharp oscillations in the free ATP-DnaA concentration at high and intermediate growth rates (compare red curve in Fig. 4.1e/f to h/i). Indeed, adding SeqA to the titration-switch model lowers the variance in the initiation volume even more, especially at high growth rates in the presence of noise in the lipid concentration (Fig. 4.4a) or in the DnaA concentration (Fig. 4.4b). The latter result on the containment of DnaA expression noise is particularly interesting: the full model that combines all mechanisms is more precise than all other models that combine only a subset of mechanisms, for all growth rates. This really shows that these three mechanisms - protein activation, titration, periodic suppression of protein synthesis - act synergistically. We thus conclude that the combination of titration, DnaA activation switch and SeqA yields robust oscillations in the concentration of active DnaA over the full range of growth rates. Recent experiments support the idea that *E. coli* combines titration with an activation switch since removing either mechanism alone still yields stable cycles, but with increased variation in the initiation size [68].

4.4. TRANSITION BETWEEN TITRATION, SWITCH OR SEQA DOMINATED REGIME

In the following, we present how in different parameter regimes either titration, the switch or the effect of blocking DnaA synthesis via SeqA can determine the initiation volume. In the full switch-titration-SeqA model, replication is initiated at a critical free, active concentration $[D]^*_{\mathrm{ATP,f}}$ (equation 4.2). Therefore, both the oscillations in the free concentration and the active fraction contribute to reaching the critical initiation threshold. This observation makes it possible to steer the system from a switch-dominated to a titration-dominated regime, by controlling the thresholds of the respective mechanisms. Here, we discuss the effect of modulating the titration threshold, by varying either the basal synthesis rate or the number of titration sites. Specifically, in the full model, the

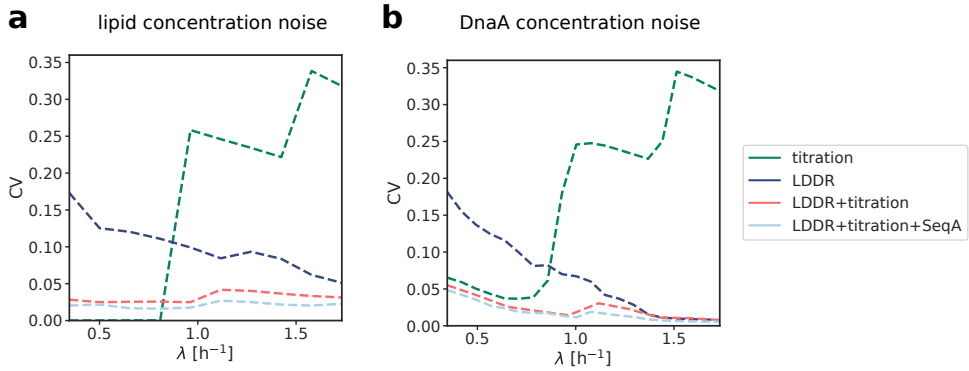


Figure 4.4: Combining the activation switch with tiration sites and SeqA reduces the coefficient of variation of the initiation volume in the presence of noise (a, b) The coefficient of variation $\mathrm{CV} = \sigma/\mu$ with the standard deviation σ and the average initiation volume $\mu = \langle v^* \rangle$ as a function of the growth rate for different models in the presence of noise in the lipid concentration (a) or in the DnaA concentration (b). The large coefficient of variation in the LDDR model at low growth rates is reduced significantly by the titration sites. Conversely, the LDDR model prevents the reinitiation events that inevitably occur at high growth rates in the AIT model. Adding the effect of SeqA to transiently block DnaA synthesis increases the robustness of the system even further. Clearly, the model that combines all mechanisms initiates replication most precisely.

dissociation constant of the promoter K_D^p must again, as in the AIT model, be higher than that of the origin K_D^{ori}, such that the free cytosolic concentration can reach the critical concentration $[D]^*_{ATP,f}$ necessary for replication initiation. Contrary to the AIT model, it is now however possible to attain the (total) free concentration set by the promoter K_D^p without reaching the critical initiation threshold $[D]^*_{ATP,f}$, if the active fraction remains sufficiently low to prevent initiation. Conversely, the titration sites could prevent replication initiation even when the active fraction f is high, by lowering the free concentration for a significant fraction of the cell cycle. Therefore, it depends on the number of titration sites and on the rate at which these sites are filled up, whether the oscillations in the active fraction or the oscillations in the free concentration set the initiation volume.

We first consider the low growth rate regime. In the absence of titration sites, the free DnaA concentration is set by the dissociation constant of the DnaA promoter K_D^p due to negative autoregulation. In this switch-only scenario, replication is initiated at a nearly constant, critical ATP-DnaA fraction $f^* = [D]^*_{ATP,f}/[D]_{T,f}$ because the total free concentration is nearly constant. When we include a finite but small number of titration sites, the production of new titration sites after replication initiation transiently lowers the free DnaA concentration at low growth rates (Fig. 4.5b, panel two). As the number of titration sites per chromosome is however small, they are filled up quickly (especially when the basal production rate as set by ϕ_0 is large) and the free concentration rises quickly until it is again constant (Fig. 4.5b, panel two). As in this regime the free concentration of DnaA is essentially constant in time before the active fraction rises at a volume as set by the balance between the activation and deactivation rates, the change in the free ATP-DnaA is dominated by the change in the ATP-DnaA fraction and the critical volume per origin

v^* is mainly set by the switch (Fig. 4.5b, panel four). In this scenario, the titration sites play a supporting role in preventing premature reinitiation by lowering the free concentration after replication initiation, but they do not set the initiation volume. Indeed, at low numbers of titration sites, the initiation volume remains constant as a function of the number of titration sites and equal to the initiation volume in the absence of titration sites (Fig. 4.5a). However, by increasing the number of titration sites while keeping the basal synthesis rate as set by ϕ_0 constant, the time to fill up the larger number of titration sites per origin increases. Now, the free DnaA concentration remains low for a longer time (Fig. 4.5c). When the cell reaches the critical volume at which the active fraction as set by the switch rises, the free concentration is still very low and prevents replication initiation. Only when the titration sites fill up at a larger volume does the free DnaA concentration rise and is replication initiated. As expected for the AIT model, in this regime the initiation volume increases linearly with the number of titration sites (compare prediction of equation 2.22 to Fig. 4.5a): the initiation volume is now dominated by titration. For a fixed number of titration sites, the system can be brought into the titration-dominated regime by decreasing the basal synthesis rate. Conversely, by increasing the basal synthesis rate, the fixed number of titration sites can be filled up more rapidly and the system is driven into the switch-dominated regime, where the initiation volume is dominated by the rise in the active fraction rather than the rise in the free concentration.

At high grow rates, the replication forks overlap and the time to replicate all titration sites is longer than the inter-initiation time set by the doubling time of the cell. Upon varying the total number of titration sites per origin, we therefore do not observe the same transition as at low growth rates from a switch to a titration-dominated regime (Fig. 4.5d). Instead, the initiation volume rises with the number of titration sites per origin at all basal synthesis rates. As the blocked period by SeqA makes up 1/3 of the cell cycle in this regime, it causes a significant drop in the free DnaA concentration that increases with increasing number of titration sites. This leads to an increasing initiation volume for higher numbers of titration sites. The initiation volume is thus in this regime set by a combined effect of the switch, the titration sites and the relatively long blocked period. It is however important to note that the switch is essential for the stability, as in this regime removing the switch results in reinitiation events as shown in Fig. 2c of the main text.

To summarize, an activation switch is able and hence sufficient to generate stable DnaA oscillations at all growth rates, but titration helps to raise the amplitude of these oscillations. It can do so at all growth rates but most predominantly at low growth rates. Indeed, at low growth rates a titration-based mechanism is sufficient but in the regime of overlapping forks rates it needs another mechanism. At sufficiently high growth rates ($\lambda > 1.4$ h^{-1}), SeqA based repression of DnaA synthesis can play this role, but in the crossover regime ($1.0 < \lambda < 1.4$ h^{-1}) this no longer suffices. In this regime, the switch, according to our model, becomes essential.

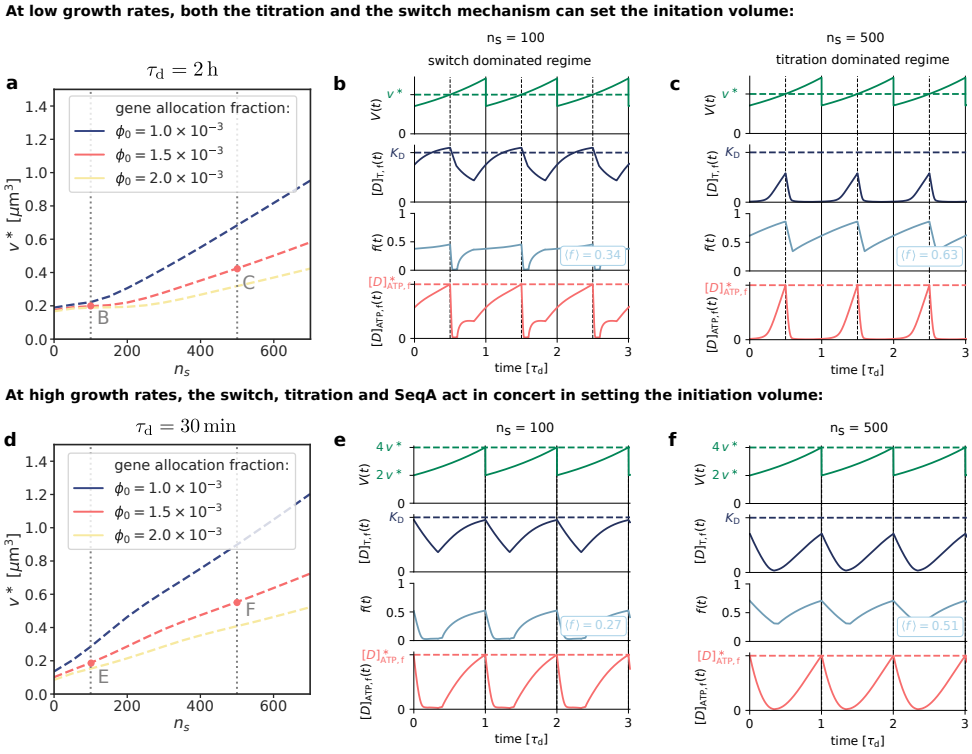


Figure 4.5: While at low growth rates both titration and DnaA activation can set the initiation volume, at high growth rates the switch, titration and SeqA control replication initiation together (b, c, e, f) The volume of the cell $V(t)$ (in units of μm^3), the free DnaA concentration $[D]_{T,f}$ (in units of μm^{-3}), the fraction of DnaA $f(t)$ that is bound to ATP (irrespective of whether the DnaA is in the cytoplasm or on the titration sites) and the concentration of free ATP-DnaA $[D]_{ATP,f}(t)$ (in units of μm^{-3}) as a function of time (in units of the doubling time $\tau_d = 2$ h (b, c) and $\tau_d = 30$ min (e, f)). The blue dashed line is the dissociation constant K_D of the DnaA to its promoter. The average active fraction over one cell cycle $\langle f \rangle$ is indicated in light blue. The dashed red line is the critical free ATP-DnaA concentration $[D]^*_{ATP,f}$ at which replication is initiated. Replication is initiated at a constant volume per origin v^* over time (green dashed line). For simulation details see section 4.1. (caption continues on the next page)

What is the biologically relevant regime? The number of titration sites has been estimated to be around 300, and Fig. 4.1g/h/i shows that in this regime, the switch and titration act in concert to generate DnaA oscillations. This is perhaps not too surprising because in this regime these mechanisms act synergistically to raise the amplitude of the oscillations in the concentration of free, ATP-bound DnaA. However, we emphasize that at this stage we regard this as a model prediction, which can be tested experimentally by varying the number of titration sites. As Fig. 4.5 shows, this makes it possible to drive the system from a titration-based regime to a switch-based one. Interestingly, very recent experiments do support the prediction that the system combines titration with activation because removing either mechanism only still yields stable cell cycles [68]. Moreover, as we show in the next section, our full model agrees, without any additional fitting, quantitatively with experiments in which the effective number of titration sites (4.A.2) or the total DnaA concentration (4.A.3) was increased, further supporting the idea that the system combines titration with activation to drive robust replication cycles.

(a, d) The initiation volume v^* as a function of the number of titration sites per chromosome n_s for different gene allocation fractions ϕ_0. (a) At low growth rates, the initiation volume exhibits two regimes: When the number of titration sites is small compared to the total concentration, the initiation volume is set by the LDDR model in the absence of titration sites. At higher numbers of titration sites the initiation volume increases linearly with the number of titration sites as predicted by the AIT model (equation 2.22). The onset of the titration-dominated regime is shifted to higher numbers of titration sites per chromosome with increasing gene allocation fraction. As the blocked period is short compared to the doubling time of the cell ($\tau_b = 10$ min $\ll \tau_d = 2$ h), SeqA only plays a minor role at low growth rates. (b, c) Two time traces of the parameter regimes indicated by red dots in Figure (a) are shown. At low growth rates ($T_C = 40$ min $< \tau_d = 2$ h), right after replication initiation, new titration sites are synthesized faster than new DnaA proteins and the free concentration of DnaA drops. At a small number of titration sites per origin n_s ($n_s = 100$) the free concentration however only drops weakly and quickly recovers a constant total concentration set by the dissociation constant K_D^p of the DnaA promoter (second panel in b). In this regime, the switch dominates the oscillations in the free, active DnaA concentration (lowest panel in b). When the number of titration sites is however very high ($n_s = 500$), the free DnaA concentration is low during most of the cell cycle (second panel in c). When all titration sites are filled does the free concentration begin to rise. As the active fraction is already high at this large volume (third panel in c), the rise in the free concentration dominates the timing of replication initiation in this regime. (d) At a higher growth rate ($\tau_d = 30$ min), the initiation volume increases strongly with the number of titration sites for all basal rates and the two regimes (switch vs. titration-dominated) that we found at low growth rates have disappeared. Contrary to the low growth rate regime, the blocked-synthesis period takes up a significant fraction of the cell cycle and the blocked period, the switch and the titration sites act together in setting the initiation volume. (e, f) Two time traces of the parameter regimes indicated by red dots in Figure (d) are shown. At high growth rates, blocking DnaA synthesis for 10 minutes in combination with synthesizing new titration sites causes a drop in the free concentration after replication initiation. For a small number of titration sites ($n_s = 100$) the drop in the free concentration is relatively small and the oscillation in the free, active concentration of DnaA (lowest panel in e) are still dominated by the switch. For large numbers of titration sites, the oscillations in the free, active fraction are shaped both by titration in combination with blocked production and the switch (lowest panel in f). Importantly, as titration together with SeqA gives rise to reinitiation events at this growth rate (Fig. 5c of the main text, region of non-zero coefficient of variation), the switch is essential for ensuring stable cycles in this regime.

4.5. DISCUSSION

While the two mechanisms of titration and protein activation have so far been mostly studied independently [14, 23, 63, 64, 77, 78, 81], our manuscript indicates that the robustness arises from the coupling of the two. Interestingly, recent experiments, which show that replication is neither controlled by titration only nor by a DnaA activation switch only, support this prediction from our model [68]. Moreover, the idea that coupling an oscillation in the concentration with an oscillation in the fraction gives rise to more robust rhythms than either oscillation alone, is very generic. Our results are thus expected to apply to any cell-cycle control system that combines titration with protein activation or modification. This finding is of particular interest given the recent observation that also higher organisms employ not only protein modification but also titration for cell-cycle control [115, 116]. In fact, the evidence is accumulating that also other oscillatory systems, most notably circadian clocks in cyanobacteria and higher organisms, derive their robustness to changes in the growth rate by intertwining a protein modification cycle with a protein concentration cycle [117–121].

Our model is supported by many experimental observations. Of particular interest are mutants in which the (de)activation mechanisms are modified or even deleted, because these allow us to test the prediction that replication initiation is controlled by the activation switch (section 4.A.1). Naturally, our model can reproduce the observations on which it is built: deleting *datA* [63, 82, 122] and deactivating RIDA [17, 63, 78, 87, 123] raises the active fraction of DnaA, while deleting *DARS1/2* [77, 81] reduces it (Fig. 4.6). Our model then predicts that impeding activation increases the average volume per origin, while weakening deactivation has the opposite effects. Many experiments support these predictions: deleting *DARS1* and/or *DARS2* increases the initiation volume per origin [81, 124], while deleting *datA* decreases it [106, 124, 125]. Our model cannot only reproduce these observations, but also the effect of combinations of deletions of these chromosomal loci on the initiation volume (Fig. 4.6). Moreover, it can describe how the initiation volume per origin changes when *datA* or *DARS2* is translocated towards the terminus [122, 126, 127] (Fig. 4.6). In addition, our model can reproduce the observation that increasing the number of titration sites via multicopy plasmids increases the initiation volume per origin [128] (section 4.A.2, Fig. 4.5), while increasing the DnaA concentration reduces it [40, 68, 108, 129] (section 4.A.3, Fig. 4.7). Taken together, these experiments support the idea that replication initiation is controlled by both titration and DnaA activation.

Intriguingly, the relative position of *DARS2* with respect to the origin and the terminus is conserved in various genomes of different sizes and strains [124], suggesting it plays an important role. Our modelling provides the following rationale: In the high growth-rate regime of overlapping replication forks, *DARS2* not only serves to balance the strong deactivation by RIDA to yield a roughly constant initiation volume, but also needs to generate oscillations in concert with RIDA. Because the activities of both *DARS2* and RIDA are proportional to the origin density, *DARS2* can only play this dual role if its position meets two constraints: On the one hand, the activity of *DARS2* should rise as late as possible in order to push the active initiator fraction down right after initiation. On the other hand, to achieve a nearly constant initiation volume independent of the growth rate, the activity of *DARS2* must be high to counteract RIDA before the next initi-

ation event; indeed, moving *DARS2* towards the terminus increases the initiation volume [126, 127] (Fig. 4.6i). The shortest period until replication is set by the highest doubling time of *E. coli*, $\tau_d \approx 18$ min. The position of *DARS2* in the middle of the chromosome ($\tau_{d2} \approx 16$ min) therefore naturally results from our model.

As we saw in the previous chapter, perhaps the most non-trivial prediction of our model is that the relaxation timescale of the switch components governs whether the switch generates adder or sizer correlations in the inter-initiation volume. Interestingly, the experiments of Si et al. allow us to test this prediction: they observed that by expressing DnaA in an oscillatory fashion, the adder is turned into a sizer [23]. As we show in Fig. 4.8, our full model can reproduce these experiments, supporting the idea that the adder correlations arise from the regression of the fluctuations in the switch threshold.

Arguably the most enigmatic element of our model is the role of the lipids in rejuvenating DnaA. In vitro experiments have shown that acidic phospholipids in the cell membrane promote dissociation of nucleotides from DnaA very effectively [86], and can restore replication activity of DnaA bound to ADP [97, 98]. Depleting acidic phospholipids in vivo can lead to growth arrest [92] and inhibit initiation at *oriC* [102]. These experiments support the idea that lipids can reactivate DnaA by promoting the exchange of bound ADP for ATP. On the other hand, it has been observed that the lethal effect of a *pgsA* null mutation, which causes a complete lack of the major acidic phospholipids, is alleviated by mutations that change the membrane structure [111, 112]. More recently, it has been reported that while downregulating *pgsA* reduced the growth rate, the initiation volume was not significantly altered [113]. We have therefore also studied models in which lipid-mediated DnaA is absent (section 4.C.1). Our modelling predicts that lipid-mediated DnaA activation is essential for the switch (Fig. 4.9a-d). The capacity of the switch to act as an origin-density sensor hinges on the idea that the activation and deactivation rates scale differently with the origin density. Without the lipids, only protein synthesis remains as an activation mechanism that does not scale with the origin density (Eq. 3.2). Consequently, to obtain a switch-based system that is even deterministically stable, the rates of all other (de)activation mechanisms must be comparable to or smaller than the growth rate. This dramatically lowers the amplitude of the oscillations, to such a degree that the system likely becomes unstable in the presence of biochemical noise. The full model, which combines the switch with titration and SeqA, is, however, surprisingly resilient to the removal of lipids, although the latter does compromise the precision of replication initiation (Fig. 4.9e-g). Indeed, while a system built on only a titration-based concentration cycle fails at high growth rates, and a system based on only an activation cycle driven by DDAH, RIDA and *DARS1/2* is likely unstable at all growth rates, the system that combines these cycles is able to generate strong rhythms at all growth rates (Fig. 4.9e-g). Even an activation cycle without lipids can thus interact with a concentration cycle to drive robust replication cycles. The full model without the lipids also exhibits adder correlations (Fig. 4.9h).

It has also been suggested that lipid-based DnaA rejuvenation is contingent on *oriC* [97] (section 4.C.2). However, a lipid-mediated DnaA activation rate that scales with the origin density effectively reduces the *datA*-mediated deactivation rate; this yields a switch that behaves similarly to that of the lipid-devoid system, because protein synthesis is again the only DnaA activation mechanism that is independent of the origin

density. In summary, lipids enhance replication initiation, but only if their effect is independent of the origin density.

Finally, our model predicts that in the regime of non-overlapping replication forks it should be possible to move the system from a switch-dominated regime to a titration-based one by increasing the number of titration sites or decreasing the basal synthesis rate of DnaA. Our model predicts that the dependence of the initiation volume on the number of titration sites or basal synthesis rate exhibits a marked, characteristic crossover when the system transitions between these two regimes (Fig. 4.5). This is a strong prediction that could be tested experimentally.

ACKNOWLEDGEMENTS

We thank Lorenzo Olivi, Sander Tans, Suckjoon Jun and Erik van Nimwegen for a careful reading of the manuscript. This work is part of the Dutch Research Council (NWO) and was performed at the research institute AMOLF.

APPENDIX

4.A. MODEL VALIDATION

In this section, we compare several key predictions from our full switch-titration-SeqA model against experimental data. First, we discuss several mutants in which activators and deactivators of the DnaA activation switch have been deleted or modified (section 4.A.1). In section 4.A.2, we then show that our model is in quantitative agreement with experiments in which the number of titration sites in the cell was varied via *oriC*-plasmids. We next demonstrate that our full model can reproduce experiments in which the total DnaA concentration was varied (section 4.A.3). Then we show (in section 4.A.4) that we can reproduce the observation of Si et al. [23] that externally driven oscillations in the DnaA concentration can transform an initiation adder into an initiation sizer. Finally, we show that the results of our model on the correlations in the inter-initiation volume are robust to different types of coupling between the replication and the division cycle (section 2.D).

4.A.1. EFFECT OF MUTATIONS IN THE ACTIVATORS AND DEACTIVATORS OF DNAA

In this section, we discuss the experimentally reported effect of mutations in the different (de)activators of DnaA and we compare them to the predictions of our model. We use the full model that combines the LDDR model, with titration sites, and a blocked period via SeqA during which both DnaA synthesis and replication initiation are blocked (see section 4.1).

EFFECT OF MUTATIONS IN THE CHROMOSOMAL DEACTIVATION SITE *datA*

The important role of the chromosomal site *datA* in the regulation of replication initiation has been described in many studies [61, 63, 77, 89, 106, 122, 125, 126]. Deletion of *datA* increases the cellular levels of ATP-DnaA as compared to wild-type cells by 5-10% [63, 82, 122]. While the size and morphology of these mutant cells were indistinguishable from wild-type cells [106, 122], replication is initiated untimely, resulting in a broad distribution of number of chromosomes per cell [122]. The average number of chromosomes per cell mass increases in a growth rate-dependent manner [122]. When *datA* is removed, the initiation volume per origin becomes smaller at low growth rates, while at high growth rates there is almost no effect of deleting *datA* [106]. These findings indicate that *datA* is important to prevent over-initiation and asynchronous initiation [122]. Originally, it was believed that *datA* influences replication initiation negatively by titrating large amounts of DnaA to the chromosome [61, 122]. More recently it was however demonstrated that *datA* can hydrolyze ATP-DnaA and the strong effect of deleting this chromosomal region was attributed to the resulting higher concentration of ATP-DnaA in the cell [63]. Our switch-titration-SeqA model can qualitatively reproduce these results: deleting *datA* results in a lower initiation volume both at high and at low growth rates, although at low growth rates not as pronounced as in the experiments (Fig. 4.6a and b).

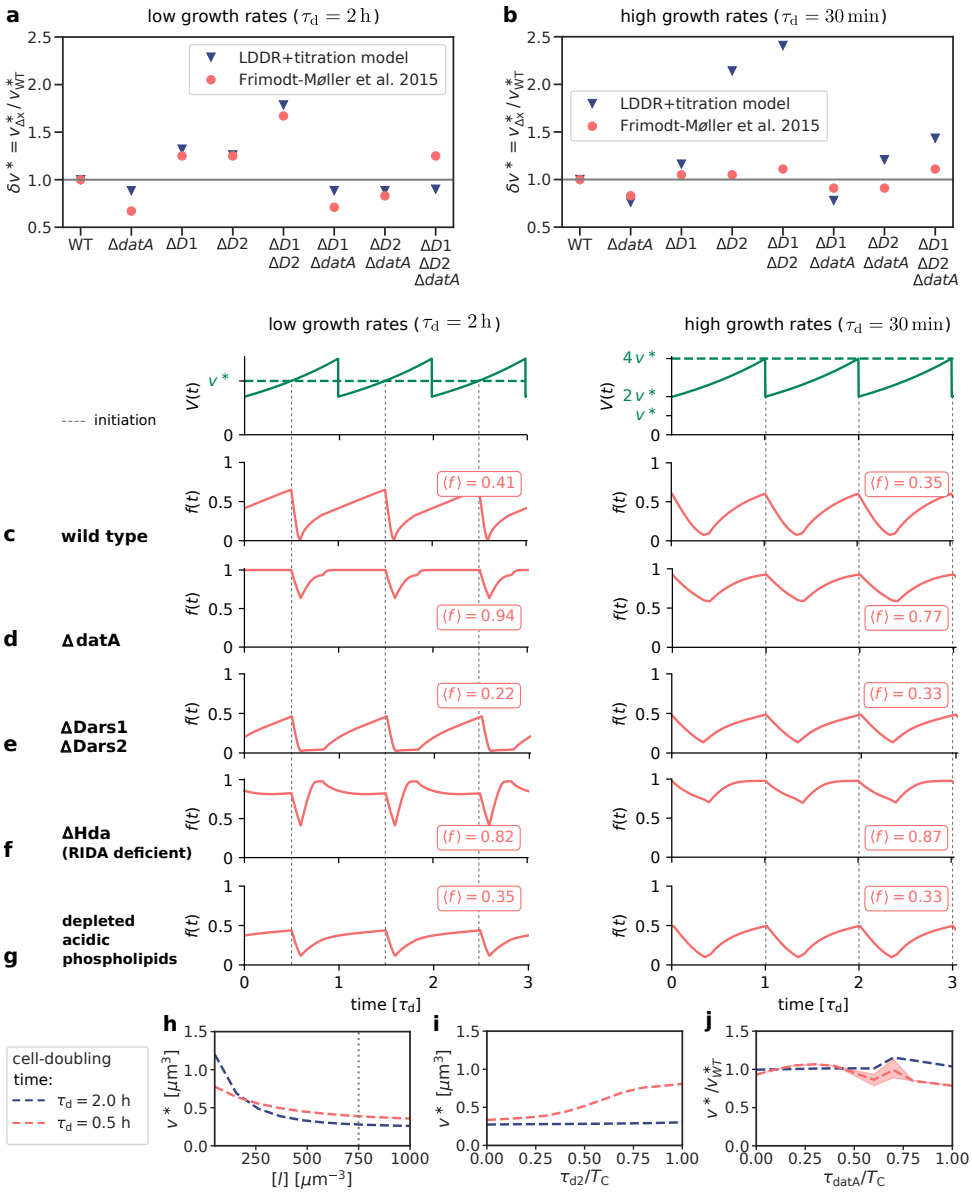


Figure 4.6: The effects of deleting different combinations of the switch (de)activators in the simulations on the initiation volume and the time traces of the active fraction are very similar to experiments, both at high and at low growth rates. All results in this Figure are obtained with the full model containing titration, LDDR switch and SeqA as explained in section 4.1. (caption continues on the next page)

The time traces of the mutant show that the effect of deleting *datA* on the active fraction of DnaA is especially severe at low growth rates, as RIDA is only active for a small fraction of the cell cycle in this regime (Fig. 4.6d, left panel). At high growth rates, the active fraction still exhibits regular oscillations, however at a higher average active fraction (Fig. 4.6d, right panel). This finding agrees well with the experimental observation that deleting *datA* raises the active DnaA fraction and that at high growth rates cells without *datA* were shown to still exhibit regular temporal oscillations in the active fraction [82].

It has been found experimentally that the chromosomal loci *DARS1*, *DARS2* and *datA* are conserved among several *E. coli* strains [124]. This finding suggests that their positions on the chromosome play an important role in the regulation of replication initiation. Focusing on *datA*, it has been observed that moving *datA* towards the terminus can have two effects: 1) change the initiation volume and 2) cause premature reinitiation. Concerning the first observation, it has been observed that placing *datA* near the terminus at a cell doubling time of $\tau_d \approx 40$ min decreases the initiation volume per origin v^* [126]. Also our model predicts that at high growth rates the initiation volume decreases when *datA* is moved towards the terminus, see Fig. 4.6j. This finding can be understood via the change in the effective gene dosage [126]: translocating *datA* to the terminus lowers the effective gene dosage of *datA* and thereby reduces the effective deactivation rate, leading to a lower initiation volume per origin [126]. This effect is especially strong at high growth rates, as in this regime the doubling time is shorter than

(a, b) For various mutants x with initiation volume $v^*_{\Delta x}$ we show the relative change of the initiation volume $\delta v^* = v^*_{\Delta x}/v^*_{WT}$ with respect to the initiation volume of the wild type v^*_{WT} obtained from the simulations of the switch-titration model (as explained in section 4.1) with a blocked production for $\tau_b = 10$ min (blue triangles) and compare it to the experimentally determined relative change in the average volume per number of origin obtained experimentally by Frimodt-Møller et al. [124] (red circles); as shown in the SI of Si et al. [40], there is a direct mapping between the average initiation volume per origin and the average volume per average number of origin. (a) At low growth rates, the effects of deleting different combinations of activators and deactivators on the change in the initiation volume agree very well with the experimentally observed relative change in the average volume per average number of origin. (b) Also at high growth rates, the simulations agree well with experimental observations. (c, d, e, f, g) The volume of the cell $V(t)$ and the active fraction $f(t)$ as a function of time (in units of the doubling time τ_d) for low and high growth rates. Replication is initiated at a critical free ATP-DnaA concentration $[D]^*_{ATP,f}$ (see section 4.1) and the system gives rise to a constant initiation volume per origin v^* over time (green dashed line). For brevity and as it is a quantity that is typically measured in experiments, we here only show the fraction of active, ATP-bound DnaA, and not the concentration. The average active fraction over one cell cycle $\langle f \rangle$ is indicated in red. Deleting DnaA activators (*DARS1/2* and lipids) tends to increase the average active fraction, while deleting DnaA deactivators (*datA* and Hda) tends to lower the active fraction, as observed experimentally. (h) Depleting the lipid concentration $[l]$ leads to an increase in the initiation volume v^* with a stronger effect at low growth rates than at high growth rates. (i) When varying the relative position of the site *DARS2* on the chromosome (where 0 is at the origin and 1 at the terminus), the initiation volume increases when *DARS2* is moved towards the terminus at high growth rates but not at low growth rates. (j) Moving the site *datA* on the chromosome towards the terminus leads to a weak relative increase of the initiation volume per origin v^* compared to the wild type initiation volume v^*_{WT} at low growth rates and to a relative decrease at high growth rates. The red shaded area indicates the maximal and minimal initiation volume per origin occuring at a given *datA* position at high growth rates.

the time to replicate the entire chromosome: changing the position of *datA* has then a large effect on the average *datA* copy number. In contrast, at lower growth rates, it takes longer before replication is initiated during the cell cycle, and the doubling of the copy number thus happens later; moving the position will thus have a smaller effect on the effective copy number. Indeed, our model predicts that at low growth rates the initiation volume does not decrease; in fact, it increases, albeit weakly (Fig. 4.6j). This is because of a second, *spatio-temporal* effect (rather than a change in the *average* copy number), as revealed by our model: moving *datA* towards the terminus means that *datA* will be doubled *later* in the cell cycle and therefore also the activity of *datA* will increase later in the cell cycle. This means that, if a new round of replication has not yet been initiated, the active DnaA concentration will reach the initiation threshold later in the cell cycle, *increasing* the initiation volume. There are thus two competing effects, and which one dominates depends on the growth rate. To our knowledge, it has not been measured how the initiation volume changes when *datA* is translocated towards the terminus at low growth rates; we thus regard this as a novel prediction from our model. Concerning the second observation, the premature reinitiation events: experiments have shown that moving *datA* towards the terminus leads to premature reinitiation at high growth rates (i.e., for $\tau_d \approx 28$ min [122]), but not at low growth rates (i.e., not for $\tau_d \approx 40$ min[126] and $\tau_d \approx 50$ min [122]). Our model can qualitatively reproduce these observations (see Fig. 4.6j). According to our model, these premature reinitiations are due to the second, spatio-temporal, effect associated with moving *datA* towards the terminus: this perturbs the shape of the temporal oscillations as generated by *datA*, *DARS2* and RIDA, such that the active fraction first decreases after initiation due to a high activity via RIDA, then rises due to an increase in *DARS2* activity and then decreases *again* due to a high *datA* activity before the active free concentration rises and reaches the initiation threshold. At higher growth rates, these "double" oscillations can lead to premature reinitiation events (Fig. 4.6j, shaded area). In contrast, at low growth rates, the effect of titration on shaping the oscillations of the active free concentration is much stronger, and this can protect the system from premature reinitiation (Fig. 4.6j, blue line). Our model thus predicts that the effects of moving the position of *datA* are highly non-trivial and growth-rate dependent; to test our predictions, measurements of time traces of the active fraction, albeit experimentally difficult, would be ideal.

EFFECT OF MUTATIONS IN THE CHROMOSOMAL ACTIVATION SITES *DARS1* AND *DARS2*

The two chromosomal sequences *DARS1* and *DARS2* can regenerate ATP-DnaA from ADP-DnaA via nucleotide exchange, resulting in replication initiation both *in vitro* and *in vivo* [77, 81]. Introducing extra copies of *DARS1* and *DARS2* increases the overall ATP-DnaA level and leads to a decrease of the average volume per origin [81]. Deletion of either *DARS1* or *DARS2* reduces the ATP-DnaA concentration in the cell by 14% and 30% [81], respectively, and leads to delayed and in some cases asynchronous initiation as compared to wild type cells [44, 77, 81, 124]. These findings show the important role that *DARS1* and *DARS2* play in replication initiation. While *DARS1* is not known to require an additional protein to activate DnaA, the more potent site *DARS2* requires the proteins IHF and Fis for its functioning [77]. Experiments indicate that Fis is more abundant in cells at high growth rates [77, 96] and the effect of deleting Fis leads to severe

over-initiation at high growth rates but had almost no effect at low growth rates [96]. Combining this observation with the finding that *DARS2* requires Fis for re-activating DnaA, *DARS2* seems to play a more important role at high than at low growth rates. Our model qualitatively reproduces these experimental observations: Both at high and at low growth rates deletion of *DARS1*, *DARS2* or both lead to an increase in the initiation volume as observed experimentally (Fig. 4.6a and b). We also note that the simulations overestimate this effect at higher growth rates, which may indicate that the activation rate of DARS1/2 depends on the growth rate, because, e.g., of the growth-rate dependence of the additional regulators like IHF and Fis. Nonetheless, deleting *DARS1/2* reduces the amplitude of the oscillations, especially at high growth rates, in line with the observation that *DARS2* is particularly important at higher growth rates (Fig. 4.6e, right panel).

Intriguingly, similarly to *datA*, the relative position of *DARS1/2* with respect to the origin and the terminus is conserved in various genomes of different sizes and strains [124]. This suggests that also the position of the site *DARS2* on the chromosome plays a role for the regulation of replication initiation [122, 124, 126]. Indeed, experiments show that relocating the sites *DARS1* and *DARS2* to different positions on the chromosome affects the replication cycle [126, 127]. Translocating *DARS2* to the terminus had the strongest effect and resulted in asynchronous initiation and a lower origin concentration as compared to the wild type. Interestingly, the degree to which the replication cycle was affected by translocating *DARS2* was growth-rate dependent: At high growth rates e.g. for cells growing in LB/glucose medium at 42 °C ($\tau_d = 18 - 10$ min), the origin/mass ratio was 87% of the wild type, almost as low as for cells lacking *DARS2* entirely (with origin/mass ratio of 82% of the wild type) [127]. At low growth rates however, e.g. for cells grown in M9/glucose/caa medium at 30 °C ($\tau_d = 71 - 77$ min), translocating *DARS2* to the terminus only led to a slight decrease of the origin to mass ratio (with origin/mass ratio of 95% of the wild type) [127]. Our model provides a novel explanation for these experimental observations: Within our LDDR model, *DARS2* plays the important role of compensating the strong deactivator RIDA. As described in the main text, putting *DARS2* near the origin would immediately counteract the effect of RIDA after a new round of replication has been initiated; while this would keep the initiation volume constant as a function of the growth rate, it would also nullify the effect of RIDA on raising the amplitude of the oscillations. On the other hand, to keep the initiation volume nearly independent of the growth rate [41], the activity of *DARS2* must be high to counteract RIDA before a new round is started; putting *DARS2* too close to the terminus would make this impossible at sufficiently high growth rates, such that then the initiation volume goes up. This is indeed precisely what our simulations show: shifting the position of *DARS2* towards the terminus increases the initiation volume (Fig. 4.6i) at high growth rates, but not at low growth rates. Our model, therefore, can not only reproduce the observation that moving *DARS2* towards the terminus increases the initiation volume, but also that this effect is stronger for higher growth rates.

EFFECT OF MUTATIONS IN THE DEACTIVATION MECHANISM RIDA

Regulatory Inactivation of DnaA (RIDA) is a mechanism promoting ATP hydrolysis in a replication coupled manner by the formation of a complex with the protein Hda and the

DNA-loaded clamp [17, 78, 87]. It is the predominant mechanism by which reinitiation events in *E. coli* are prevented [17, 90]. RIDA inactivation via the deletion of the Hda gene or inactivation of the clamp increases the cellular ATP-DnaA level to 70–80% of the total number of DnaA molecules [78, 87]. RIDA deficient cells initiate replication asynchronously and earlier than wild type cells at smaller initiation volume [44, 63, 122, 123]. In the simulations we cannot address the effect of asynchronous initiation of replication, because in our mean-field model, by construction, all origins are fired simultaneously when the critical free ATP-DnaA concentration is attained. In combination with the titration mechanism, we obtain stable cell cycles even though the time traces of the active fraction are strongly disturbed. At low growth rates, the active fraction first decreases after replication initiation due to the duplication of *datA* and the blocked protein synthesis, but then rises quickly to its maximal value due to the strong activation via the sites *DARS1* and *DARS2* (Fig. 4.6f, left panel). Also at high growth rates, the active fraction is very high throughout most of the cell cycle (Fig. 4.6e, right panel). Both observations agree well with the experimentally reported strong increase of the ATP-DnaA level in RIDA deficient cells.

EFFECT OF MUTATIONS IN THE ACIDIC PHOSPHOLIPIDS

In section 4.C we discuss the experimental evidence for the role of the lipids in DnaA reactivation. Here, we discuss the predictions from our full model. It predicts that depleting acidic lipids weakens DnaA activation, which increases the initiation volume per origin, see Fig. 3.7 and Fig. 4.6h. Our model also predicts that this effect is stronger at lower growth rates (Fig. 4.6h). From the time-traces of the full model in Fig. 4.6g we furthermore predict that in cells with depleted acidic phospholipids the average fraction of ATP-DnaA is reduced and the amplitude of the oscillations in the active fraction decreases. These are clear predictions that could be tested experimentally, using mutants in which the *pgsA* gene is brought under the control of an inducer [92, 102]. The specific predictions from our model in combination with experiments should make it possible to elucidate the contested role of the acidic lipids in replication initiation, and may clarify the apparent contradictions between the studies of Refs. [111–113] and those of Refs. [93, 102, 103] (see also section 4.C.1.

4.A.2. EFFECT OF VARYING THE NUMBER OF TITRATION SITES PER ORIGIN ON THE INITIATION VOLUME PER ORIGIN

Our full model predicts that varying the number of titration sites per origin affects the initiation volume and that the degree by which the initiation volume changes depends on the relative change in the number of titration sites per origin, the growth-rate regime, and whether the system is dominated by the switch, titration or SeqA (Fig. 4.5 and section 4.4). In the following, we compare this prediction to experiments. Christensen et al. [128] showed that introducing different multicopy plasmids that contain different variants of the *oriC* region changes the initiation volume. While *oriC* does not affect the switch, because it is neither an activator nor a deactivator of DnaA, it does contain several high and low-affinity binding sites for DnaA and as such could affect the titration mechanism. Christensen et al. observed that for all growth rates studied, plasmids with a higher number of DnaA boxes cause a larger decrease in the chromosomal origin

concentration and hence a larger increase in the initiation volume per origin [128]. This finding is consistent with the predictions from our full model: the initiation volume increases with increasing number of titration sites per origin (Fig. 4.5). Remarkably, the agreement is not only qualitative, but even nearly quantitative, without any additional fitting. Specifically, the degree to which the initiation volume changes depends on the relative change in the number of titration sites per origin and the growth rate. The strain carrying plasmid pFHC946, which contained all of the DnaA boxes from the *oriC* region, showed about a 10% increase in the initiation volume per origin at a high growth rate, and a nearly 20% change in the initiation volume per origin at a low growth rate [128]. To compare this observation to the predictions from our model, we note that at the high growth rate, the number of pFHC946 plasmids per origin is about 13 while at the low growth rate it is about 57 (Table 1 of Ref. [128])). To estimate the total number of DnaA titration sites that these plasmid copies carry, we consider that plasmid pFHC946, like the chromosomal *oriC* region, contains two high-affinity DnaA binding sites (R1 and R4) and one intermediate affinity site R2, resulting in 2-3 medium-high affinity titration sites per plasmid copy. The total number of titration sites on the plasmid copies combined is thus $13 \times (2-3) \approx 25-40$ at high growth rates and $57 \times (2-3) \approx 100-150$ at low growth rates. Our model predicts that at high growth rates (Fig. 4.5d, red line), a change in the number of titration sites from $n_s = 300$, corresponding to wild-type cells with no extra plasmids, to $n_s \approx 350$, corresponding to the pFHC946 strain, causes a relative change in the initiation volume of about 10%, in quantitative agreement with the experiments of Christensen et al. At low growth rates, our model predicts (Fig. 4.5a, red line) that a change in the number of titration sites, from 300 in wild-type to $400-450$ in the pFHC946 cells, cause a change in the initiation volume of about $22-35$%, in near quantitative agreement with the reported change of about 20%.

4.A.3. EFFECT OF VARYING THE TOTAL DNAA CONCENTRATION ON THE INITIATION VOLUME PER ORIGIN

In this section, we show that the effect of varying the total DnaA concentration on the initiation volume per origin in our full model is in good agreement with experimental observations. The total DnaA concentration in *E. coli* cells can be varied by introducing plasmids containing inducible *dnaA* promoters into cells [108, 129], by replacing the native *dnaA* promoter with an inducible promoter [68] or by repressing the native *dnaA* promoter further using tunable CRISPR interference [40]. All of these experiments reported a negative dependence of the initiation volume on the total DnaA concentration [40, 68, 108, 129]. Furthermore, Zheng et al. measured the average DnaA concentration and the initiation mass per origin at different growth rates and also reported a negative correlation between these two variables [41]. In a recent paper by Flatten et al. [106] it was however reported that upon increasing the total DnaA concentration two-fold, the average volume per number of origins decreased only very weakly [106]. Upon increasing the total DnaA concentration even further (up to 35 times the wild type concentration), also Flatten et al. observed a decrease in the volume per number of origins in combination with excessive and asynchronous initiations [106]. Finally, Si et al. [23] found that oscillatory perturbation of the total DnaA concentration affected the initiation volume per origin and that the initiation volume was anti-correlated with the total DnaA

concentration. These results clearly show that DnaA synthesis plays a non-negligible role in regulating the initiation volume and that in all experiments, the initiation volume per origin is negatively correlated with the total DnaA concentration.

To test the effect of a varying the total DnaA concentration in our full model, we mimick the plasmid experiments of Hill et al. [108] and of Atlung et al. [129] by adding an external production term of DnaA to our full Switch-titration-SeqA model (as described in sections 4.1 and 4.3)

$$\frac{dN_{\mathrm{D,ATP}}^{\mathrm{ext}}}{dt} = \phi_0^{\mathrm{ext}} \lambda \rho V \tag{4.8}$$

with the external gene allocation fraction ϕ_0^{ext}, the number density ρ and the volume of the cell V. Increasing the external gene allocation fraction ϕ_0^{ext} leads to an increase in the average DnaA concentration in the cell. We find that in the full model the initiation volume decreases with increasing total concentration for a wide range of growth rates, in excellent agreement with experiments (see Fig. 4.7). The initiation volume per origin is approximately inversely proportional to the total DnaA concentration, such that for a doubling of the DnaA concentration the initiation volume is approximately halved. This is consistent with the finding of Zheng et al. who observe that for a decrease of the total concentration of about 20%, the average initiation mass per number of origins increases by about 20% (see Extended Data Fig. 5 of [41]). Also Atlung and Hansen [129] find an almost linear increase in the number of origins per mass (the inverse of υ^*) with increasing total DnaA concentration up to at least two times of the wild type DnaA total concentration; at higher total concentrations, the average number of origins per mass reached a plateau, but this could be explained by a reduced replication speed due to too high replication initiation rates [14, 129]. Therefore, our simulations are in good quantitative agreement with several experimental findings.

4.A.4. EXTERNALLY DRIVEN OSCILLATION IN THE DNAA CONCENTRATION CAN TURN AN INITIATION-ADDER INTO AN INITIATION-SIZER

In this section, we test whether our full switch-titration-SeqA model can explain the experimental result by Si et al. [40] that dynamically perturbing the DnaA concentration can turn an initiation-adder into an initiation-sizer. In section 3.E we found that in the LDDR model noise in a negatively autoregulated DnaA production rate gives rise to sizer correlations in the initiation volume, while noise in the (de)activators of DnaA gives rise to adder correlations. In the initiator titration model, noise in the initiator protein gives however rise to adder correlations (Fig. 2.7b). These findings open two questions: First, what correlations in the initiation volume arise from DnaA gene expression noise when both the switch and the titration sites contribute to setting the initiation volume? We demonstrate that while in the titration-dominated regime (e.g. for high numbers of titration sites or low DnaA synthesis rate, see Fig. 4.5) noise in DnaA gene expression gives rise to adder correlations, in the switch-dominated regime it can give rise to sizer-correlations provided that negative autoregulatinon is strong enough to reduce the relaxation time of DnaA concentration fluctuations. The second question is what correlations arise when we combine a source of noise that gives rise to sizer correlations with one that gives rise to adder correlations? We show that when combining DnaA noise in

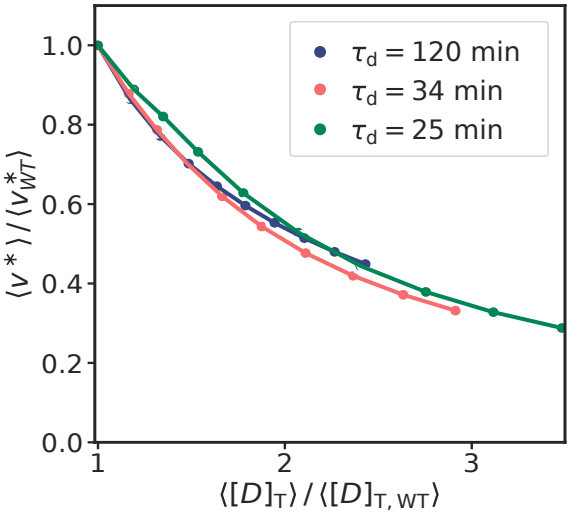


Figure 4.7: The full model can reproduce the experimental observation that the initiation volume per origin is negatively correlated with the total DnaA concentration. The average initiation volume per origin $\langle v^* \rangle$ as a function of the average total concentration $\langle [D]_T \rangle$ for three different cell doubling times τ_d. The total DnaA concentration is varied by adding a constant DnaA production term according to equation 4.8 to the full model (as described in sections 4.1 and 4.3) and varying the external gene allocation fraction $\phi_0^{ext} = 0 - 2.25 \times 10^{-3}$. Both the average initiation volume per origin and the average total concentration have been normalized by their value in the absence of external DnaA expression, $\langle v^*_{WT} \rangle$ and $\langle [D]_{T,WT} \rangle$, respectively. In line with various experiments [14, 40, 41, 68, 108, 129], the average initiation volume per number of origins decreases with increasing average total DnaA concentration. See Table 3.1 for parameters.

the switch-dominated regime, thus giving sizer noise, with adder-noise in the lipid concentration, it depends on the strength of the respective sources of noise whether the system exhibits adder or sizer correlations. Finally, we show that our switch-titration model can reproduce the recent experimental results by Si et al. [23]: By starting from a system in which noise in both the lipids and the DnaA concentration contribute to adder correlations in the initiation volume, dynamically perturbing the DnaA concentration gives rise to sizer noise in the initiation volume.

First, we study the correlations in the initiation volume in the switch-titration model when noise in the DnaA concentration is the only source of noise. We model noise in DnaA gene expression according to equation 2.24 as explained in section 3.7. Whether noise in the DnaA concentration then gives rise to adder or sizer correlations in the initiation volume depends on whether the system is in the switch or in the titration-dominated regime (Fig. 4.8a). In the switch-dominated regime, the basal rate is sufficiently high to fill up the titration sites before the next round of replication is initiated. In this regime, DnaA is negatively autoregulated during most of the cell cycle and fluctuations in the DnaA concentration are reduced rapidly, when, as assumed here, autoregulation is sufficiently strong. Then, the correlations in the initiation volume are sizer-like, as explained in the main text (4.8a, green data points). We can move from the switch to the titration-dominated regime by either increasing the number of titration sites or by decreasing the basal DnaA synthesis rate (see Fig. 4.5). In the titration-dominated regime, the DnaA promoter is almost not repressed during most of the cell cycle. In this regime, fluctuations in the total DnaA concentration are reduced via dilution on a time scale set by the doubling time of the cell, giving rise to adder correlations in the initiation volume (4.8a, red data points).

Now we want to study a system that combines fluctuations that tend to generate sizer-like correlations with fluctuations that tend to generate adder-like correlations. To this end, we will study a system that is in the switch-dominated regime, where DnaA generates sizer-like fluctuations and the lipids generate adder-like correlations. Fig. 4.8b demonstrates that a combination of sizer-noise in the DnaA and adder-noise in the lipid concentration can give rise to both sizer or adder noise, depending on the relative strengths of the respective sources of noise. At a low magnitude of the noise in DnaA as compared to the lipids, the fluctuations in the initiation volume are dominated by the fluctuations in the lipid concentration and we find an initiation adder (Fig. 4.8b, red data points). By increasing the noise strength of the DnaA production rate while keeping the lipid noise strength constant, the observed adder correlations in the initiation volume become sizer-like. From this finding we predict that if replication initiation in *E. coli* is dominated by the switch rather than titration, the dominant source of noise generating adder correlations in the initiation volume is set by fluctuations in the concentration of (de)activators of DnaA such as the lipids, RIDA (via Hda) and *datA* (via IHF). Conversely, if the system is in the titration-dominated regime, the adder correlations can arise from noise *both* the switch (de)activators and DnaA.

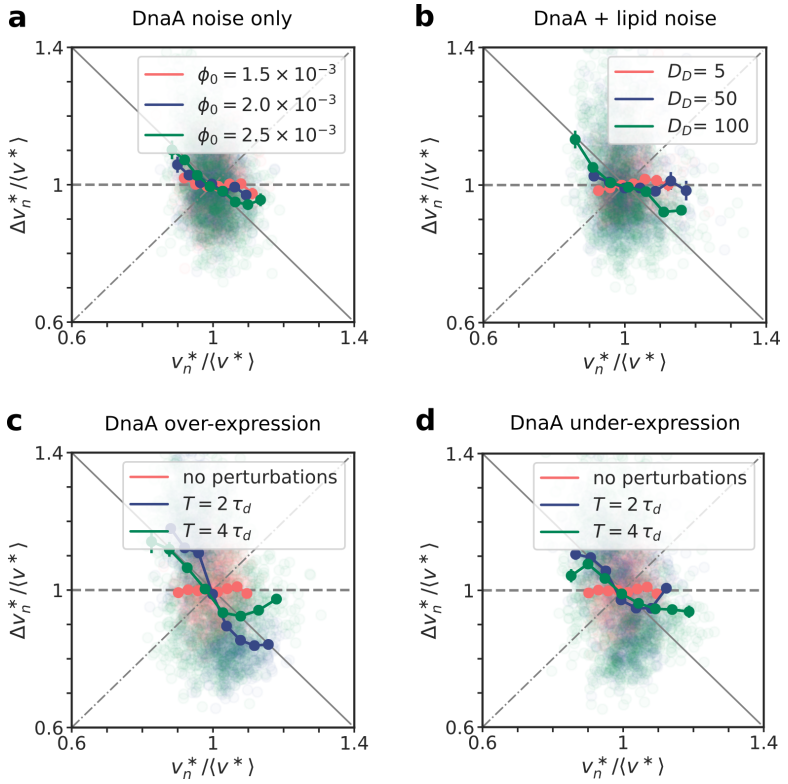


Figure 4.8: Externally driven oscillations in the DnaA concentration can turn an initiation-adder into an initiation-sizer (a, b, c, d) Scatter plot of the added initiation volume between successive initiation events, $\Delta v_n^* \equiv 2v_{n+1}^* - v_n^*$, and the initiation volume v_n^*. Both the x and the y axis are normalized by the average initiation volume $\langle v^* \rangle$ at the respective parameters. For comparison, the dashed line is a perfect adder, the solid line a perfect sizer and the dash-dotted line a perfect timer. The dark blue line shows the mean of the binned data and the error bars represent the standard error of the mean (SEM) per bin. The doubling time in all plots is $\tau_d = 2$ h and the number of data points is $N = 995$. (a) In the switch-titration model it depends on the basal rate of DnaA gene expression whether noise in the DnaA concentration (with noise strength $D_D = 100$) generates an initiation adder or sizer. At a low gene allocation fraction ϕ_0 and thus a low DnaA basal rate, the DnaA promoter is weakly repressed for most of the cell cycle because of the long time it takes to fill all titration sites; consequently, DnaA noise gives rise to adder correlations (red data points). By increasing the gene allocation fraction ϕ_0 of DnaA and thus the basal DnaA expression rate, the system moves from the titration-dominated regime to the switch-dominated regime. In the switch-dominated regime, the titration sites are filled up more quickly and the DnaA promoter is repressed during most of the cell cycle. As explained in the LDDR model, negative autoregulation then generates sizer-correlations in the initiation volume (blue and green data points). (caption continues on the next page)

Now, we want to test whether in the switch-titration model we can turn an initiation adder into an initiation sizer like demonstrated in recent experiments by Si et al. [23]. Starting from the observation that the initiation volume increases with decreasing DnaA concentration, Si et al. dynamically perturbed the DnaA concentration of cells and measured the effect on the initiation volume. Inducing DnaA with a period of $T = 2\tau_d$ causes cells that have initiated at a larger than average volume in generation n to initiate at a smaller than average volume in generation $n+1$. Si et al. therefore proposed that periodically inducing the production of DnaA with a period given by $T = 2\tau_d$ would lead to negative auto-correlations of the initiation volume, thus breaking the for an adder typical mother-daughter auto-correlation of 1/2. In their experiments Si et al. had to use a period of $T = 4\tau_d$ due to the high induction and dilution time of DnaA. They find that the periodic induction of DnaA with a period of $T = 4\tau_d$ can indeed break the initiation adder and the initiation volume showed weak sizer correlations.

Starting from an unperturbed system where both the lipid and the DnaA noise give rise to adder correlations in the initiation volume (Fig. 4.8c, red data points) we now include the effect of strong externally driven oscillations in the DnaA concentration. In the experiments by Si et al. [23], a strain carrying extra *dnaA* under the P_{lac} promoter on plasmids was used to induce DnaA. In the over-expression experiments, DnaA was expressed both by its endogenous promoter on the chromosome and by the inducer controlled extra *dnaA* on the plasmid. In our simulations, we mimic these DnaA over-expression experiments by expressing DnaA as described in section 3.7 from the chromosome and adding the following deterministic oscillation in the expression rate of ATP-DnaA proteins to the system:

$$\frac{dN^{ext}_{D,ATP}}{dt} = a(\cos(2\pi/T\,t) + 1) \tag{4.9}$$

(b) Starting from a system that is in the switch-dominated regime ($\phi_0 = 2.5 \times 10^{-3}$) and in which DnaA noise gives rise to sizer-correlations in the intiation volume, we now add noise in the lipid concentration (see section 3.D). Whether the system is then an initiation-adder or an initiation-sizer depends on the respective strengths of the noise in the DnaA and lipid concentration. For a noise strength of the lipids of $D_l = 1000$ and a noise strength of DnaA of $D_D = 5$, the adder noise from the lipids is dominant and the resulting correlations in the initiation volume are an adder (red data points). With increasing DnaA noise strength, the correlations in the initiation volume become more and more sizer-like (blue and green data points). (c) Now we consider a system where both noise in the DnaA and in the lipid concentration give rise to adder correlations ($\phi_0 = 1.5 \times 10^{-3}$, $D_D = 100$ and $D_l = 1000$, red data points). Similar to the over-expression experiments performed by Si et al. [23], we add external oscillations in DnaA in order to turn the adder into a sizer. We express DnaA both via its endogeneous promoter according to equation 2.24 and via externally driven oscillations according to equation 4.9. For a period of $T = 2\tau_d$ (blue data points), which is the optimal period for inducing negative auto-correlations in the initiation volume, the adder correlations are strongly sizer-like. When DnaA is induced with a period of $T = 4\tau_d$ like in the experiments by Si et al. [23] (green data points), the adder (still) becomes sizer-like as observed experimentally [23]. (d) By removing expression from the endogeneous promoter of DnaA and combining a basal constitutive DnaA expression rate with oscillations according to equation 4.10 with $\phi_0^{ext} = 1.5 \times 10^{-3}$, the switch-titration model can reproduce the under-expression experiments of Si et al. [23]. Again, we can turn an unperturbed system that exhibits adder correlations (red data points, same as in c) into a system that is more sizer-like (green and blue data points).

where a is the amplitude and T is the period of the oscillations. We find that externally driven oscillations according to equation 4.9 with a period of $T = 4\,\tau_d$ lead to strong sizer correlations (Fig. 4.8c, green data points). The amplitude of the oscillations is sufficiently high and thus, the dominant source of noise are not the lipid or the unperturbed DnaA concentration fluctuations anymore, but the driven oscillations in the DnaA concentration. At a period of $T = 2\,\tau_d$, the effect becomes even stronger and we see that the system oscillates as predicted between a high and a low initiation volume, thus creating strong sizer correlations (Fig. 4.8c, blue data points).

Si et al. additionally performed DnaA under-expression experiments by removing the endogeneous expression of DnaA and using the DnaA expressed from the plasmid as the only source of DnaA expression. We model this by removing expression of DnaA via the endogeneous promoter according to equation 2.24. Instead, we add a basal production rate following again the growing cell model of gene expression and combine it with the oscillations in the DnaA number as in equation 4.9:

$$\frac{dN^{\text{ext}}_{\text{D,ATP}}}{dt} = \phi_0^{\text{ext}}\,\lambda\,\rho\,V + a\,(\cos(2\pi/T\,t) + 1) \tag{4.10}$$

Again we find that using this modified DnaA gene expression rate, we obtain sizer-correlations in the initiation volume when the system is driven at the optimal period of $T = 2\,\tau_d$ (Fig. 4.8d, blue data points). Also using the same period $T = 4\,\tau_d$ as in the under-expression experiments Si et al. [23] lead to weak sizer correlations (Fig. 4.8d, green data points). We have shown that our switch-titration model can explain how an unperturbed system that has adder correlations in the initiation volume can become sizer-like by strong dynamic perturbations of DnaA.

4.B. NOVEL PREDICTIONS FROM THE FULL SWITCH-TITRATION MODEL

We here summarize the novel predictions from our full switch-titration model. Some of these have already been discussed in the previous section, when they naturally followed from the model validation.

1. We predict that at low growth rates the activation switch and the titration mechanism act synergistically to raise the amplitude of the oscillations in the concentration of free, ATP-bound DnaA (see section 4.4). Whether the system is in the titration or a switch-dominated regime could be tested by varying the number of titration sites and the basal initiator production rate, as Fig. 4.5 shows. While in the switch-dominated regime the initiation volume should be relatively independent of the number of titration sites, in the titration-dominated regime the initiation volume should increase linearly with the number of titration sites (Fig. 4.5a). This is a strong, robust prediction from our model that could be tested experimentally. Since the titration sites have a characteristic sequence [46, 54] it should be possible to vary the number of titration sites on the chromosome, making it possible to steer the system between a switch-dominated regime to a titration-dominated one.

2. We furthermore predict that due to the random distribution of titration sites on the chromosome, at intermediate and high growth rates the titration mechanism alone cannot ensure stable cell cycles. At high growth rates, the help of SeqA is sufficient to generate robust cell cycles. Yet, at intermediate growth rates the switch becomes essential for preventing premature reinitiation events (Figs. 2 and 5c of main text). The predicted dependence of the importance of the activation switch on the growth rate could be tested using mutants in which the switch is effectively turned off, for example by deleting *datA* [63] and deactivating RIDA [78, 87, 105]. Our model predicts that stable replication cycles can be generated solely by titration and SeqA at low and high growth rates, but not at intermediate growth rates. To further study the interplay between titration, activation, and blocked synthesis by SeqA, it would be of interest to remove the latter mechanism; this may be achieved by removing the GATC site in the promoter region of DnaA, which is necessary for binding SeqA [59]. Our model predicts that removing this mechanism has a much larger effect at high than at low growth rates.

3. We predict that if replication initiation in *E. coli* is dominated by the switch rather than titration, the dominant source of noise generating adder correlations in the initiation volume is set by fluctuations in the concentration of (de)activators of DnaA such as the lipids, RIDA (via Hda) and *datA* (via IHF). Conversely, if the system is in the titration-dominated regime, the adder correlations can arise both from the switch components and DnaA.

4. Our model predicts that the initiation volume is inversely proportional to the total DnaA concentration, see Fig. 4.7. More specifically, it predicts that this relation holds over a wide range of growth rates.

5. According to our switch model, the initiation volume scales inversely proportional to the acidic phospholipid concentration. We predict that in the switch-dominated regime the initiation volume should decrease with increasing lipid concentration and that the effect of depleting the lipids on the initiation volume is stronger at low than at high growth rates (Fig. 4.6h).

4.C. THE ROLE OF THE LIPIDS

The capacity of the switch to act as an origin-density sensor hinges on the idea that the activation and deactivation rates have different functional dependencies on the origin density. In the switch models, all deactivation rates are proportional to the origin density, but not all activation reactions are: while the rate of DARS1/2 activation is proportional to the origin density, the rates of DnaA activation via protein synthesis and the acidic phospholipids are not. However, the importance of the lipids for reactivating DnaA *in vivo* remains unclear [111–113]. Moreover, there is evidence that the effect of the lipids depends on *oriC* [97]. Here, in section 4.C.1, we analyze the importance of the lipids for DnaA reactivation, both for the switch-only models (LD and LDDR) and for the full model. Then, in section 4.C.2, we discuss the role of *oriC* in lipid-mediated DnaA reactivation. The principal findings of these sections are: lipid-mediated DnaA activation is essential to the switch, both in the LD and in the LDDR model: taking the

lipids out entirely, dramatically lowers the amplitude of the oscillations, such that in the presence of biochemical noise they will likely not persist. Interestingly, however, the full model, which includes titration and SeqA, is robust to removing DnaA activation via the lipids. While a model based on titration alone is only stable at low growth rates and a system based on only a lipid-deficient switch produces merely very weak oscillations at any growth rate, the combined system yields robust oscillations at all growth rates.

4.C.1. THE IMPORTANCE OF LIPID-MEDIATED DNAA REJUVENATION

While the *in vitro* evidence that acidic phospholipids enhance the release of nucleotides is compelling [86, 97], the *in vivo* experiments paint a mixed picture [92, 102, 111–113]. In the next section, 4.C.1, we will first discuss these experiments. We will then study the importance of lipid-mediated DnaA activation by taking out the lipids, first in the LD model (4.C.1), then in the LDDR model (4.C.1), and then in the full model (4.C.1). We find that while the lipids are essential for generating high-amplitude oscillations in the switch-only models, the full model is remarkably robust: removing the lipids from the full model does reduce the precision of replication initiation, but not dramatically.

EXPERIMENTS

Acidic phospholipids from the cell membrane promote dissociation of both ADP and ATP from DnaA very effectively [86], and DnaA can be reactivated by exchange of the bound nucleotide *in vitro* in the presence of ATP and *oriC* [97, 98]. Depleting acidic phospholipids *in vivo* can lead to growth arrest [92, 104] and inhibit initiation at *oriC* [102], which supports the idea that lipids can reactivate DnaA by promoting the exchange of bound ADP for ATP. More specifically, Fingland et al. observed that by bringing *pgsA* under an inducible promoter acid-lipid depleted cells slowed down their growth and that as they began to slow down in their growth, the number of origins per cell, as determined via run-out experiments using rifampicin and cephalexin, decreased relative to cells that continue to synthesize acidic phospholipids [102]. Also the DNA content and cell mass decreased. These observations could be explained by the decrease in the growth rate, which tends to reduce the average number of origins per cell, DNA content, and cell mass [40]. Several studies however showed that the reported growth arrest of acidic phospholipid-deficient cells can be suppressed by overexpression of a mutant form of DnaA (L366K) which contains mutations in the membrane-binding domain [93, 102, 103]. Moreover, the growth arrest can be suppressed by mutating rnhA, which allows the cell to bypass normal *oriC*-dependent initiation via a process called recA-dependent constitutive stable DNA replication [92]. Both of these findings strongly indicate that the growth arrest is caused by the inability to initiate replication [92, 93, 102, 103].

In contrast, Shiba et al. found that the lethal effect of a *pgsA* null mutation, which causes a complete lack of the major acidic phospohlipids, phosphatidylglycerol and cardiolipin, is alleviated by mutations that change the membrane structure [111, 112]. Moreover, a recent study by Camsund et al. [113] reported that while downregulating *pgsA* reduced the growth rate of the cell (rather than growth arrest), the initiation volume per origin was not significantly affected. From this finding, the authors concluded that, in contrast to the earlier reports [93, 102, 103], the lipids are not vital for replication ini-

tiation. Motivated by these observations, we therefore analyze in the next sections the behavior of the LD, LDDR, and the full model, when lipid-mediated DnaA activation is taken out completely.

LD MODEL

The change of the active fraction as a function of time for a system where DnaA is activated only via DnaA synthesis and deactivated via the chromosomal site *datA* is given by:

$$\frac{df}{dt} = \lambda(1-f) - \tilde{\beta}_{\mathrm{datA}}[n_{\mathrm{ori}}]\frac{f}{\tilde{K}_{\mathrm{D}}^{\mathrm{datA}} + f} \tag{4.11}$$

As we again combine an origin-density dependent deactivation rate with a constant activation rate (for a given growth rate), initiating replication at a critical active fraction f^* results in stable replication cycles with the initiation volume per origin v^*, see Fig. 4.9a and b. However, the amplitude of the oscillations is very low (Fig. 4.9b). Protein synthesis yields an activation rate that is linear in the active DnaA fraction f (see equation 3.4 and Fig. 4.9a). In contrast, lipid-mediated activation is non-linear in f. This is important, because together with the non-linear deactivation rate, it gives rise to an ultra-sensitive activation mechanism (see Fig. 3a of the main text). Removing lipid-mediated activation from the model thus eliminates the ultra-sensitive activation mechanism, which reduces the amplitude of the oscillations in f (compare Fig. 4.9a with Fig. 3a of the main text).

Yet, there is a second effect, which is best illustrated in the quasi-equilibrium limit, where the critical initiation volume per origin v^*, obtained by setting $df/dt = 0$ in equation 4.11, is given by

$$v^* = \frac{\tilde{\beta}_{\mathrm{datA}}}{\lambda}\frac{f^*}{(\tilde{K}_{\mathrm{D}}^{\mathrm{datA}} + f^*)(1-f^*)} \tag{4.12}$$

The key point to note is that the activation rate is now set by the growth rate λ, which is much lower than the experimentally measured *datA*-mediated deactivation rate $\tilde{\beta}_{\mathrm{datA}}$. Hence, to obtain an initiation volume that is consistent with what is observed experimentally the initiation threshold f^* would have to be lowered drastically. However, this would strongly reduce the amplitude of the oscillations. Alternatively, the deactivation rate $\tilde{\beta}_{\mathrm{datA}}$ itself would have to be much lower. Yet, this would also strongly reduce the amplitude of the oscillations (Fig. 4.9b). In the presence of biochemical noise, these small oscillations would therefore not be sufficient to ensure stable cell cycles. Indeed, to generate large amplitude rhythms, not only the deactivation but also the activation rate needs to be larger than the growth rate. This is what lipid-mediated activation can accomplish.

LDDR MODEL

We have also analyzed the LDDR model without lipid-mediated DnaA activation. This system is described by Eq. 3.4 but without the first lipid term on the right-hand side. Crucially, in order for the switch to be stable, the activation and deactivation rates must have different functional dependencies on the origin density, see section 3.3. Moreover,

the *DARS1/2* activation rate scales with the origin density like the rates of *both* deactivation mechanisms, *datA* and RIDA. Taken together, these two observations mean that, in the absence of origin-independent lipid-mediated DnaA activation, the *DARS1/2* activation rate cannot be too large: it must be smaller than the protein synthesis rate, which is then the only mechanism of DnaA activation or deactivation that is independent of the origin density; otherwise, the net, overall rates of both activation and deactivation would scale with the origin density, making the system unstable (see Fig. 3.5b). Since the protein synthesis rate is set by the growth rate, this means that the activation and deactivation rates must be smaller than or comparable to the growth rate, at all growth rates. The net result is that a LDDR model without the lipids exhibits oscillations of only weak amplitude, for all growth rates (see Fig. 4.9c/d).

FULL MODEL

The full model, including titration, SeqA, and the switch without lipid-mediated activation, is given by Eq. 4.1, but with the lipid term in 4.1 removed. Crucially, in the presence of titration and SeqA, the rates of *DARS1/2* and RIDA do not have to be dialed down to make the system stable, as is the case for the LDDR model without the lipids; only the rate of *datA*, which is no longer balanced by the lipids, needs to be reduced, to keep the initiation volume consistent with experiments (indeed, merely taking out the lipids raises the initiation volume, see Fig. 4.6g/h). This system is surprisingly robust, for the full range of growth rates, see Fig. 4.9e/f. This is particularly interesting, because a system with *only* titration and SeqA or with *only* a (lipid-deficient) switch, cannot generate robust oscillations at all growth rates, while the full system, which combines all three mechanisms, can. The antagonism between *DARS1/2* and RIDA generates large oscillations in the fraction, titration and SeqA generate large oscillations in the concentration, and the interplay between *DARS1/2*-RIDA and titration-SeqA ensures stability. In particular, homeostasis is ensured by titration, making the unstable lipid-devoid switch stable, while, conversely, the lipid-devoid switch prevents the reinitiation events that inevitably happen in the titration-SeqA-only system at intermediate growth rates (Figs. 2 and 5c). We do find that a full model with lipid-mediated DnaA activation is more robust than a full model without it, but the difference is surprisingly small (Fig. 4.9g, compare solid blue to solid red line). Yet, in the full model without SeqA (switch + titration), a switch without lipids can no longer fully prevent reinitiations in the intermediate growth rate regime (Fig. 4.9g, compare dashed blue to dashed red line). Finally, in the presence of gene expression noise in DnaA, the full system without the lipids still exhibits adder correlations in the initiation volume, as shown in experiments [23, 24] (Fig. 4.9h).

4.C.2. THE ROLE OF *oriC* IN LIPID-MEDIATED DNAA REACTIVATION

In all the switch models discussed above and in the main text, we have assumed that lipid-mediated DnaA rejuvenation is independent of the origin density. However, there is evidence that the effect of the lipids depends on *oriC* [97]. In section 4.C.2 below, we first discuss these experiments in more detail. We then discuss the picture that emerges from these experiments, and the different DnaA rejuvenation scenarios that could be envisioned. We then discuss a scenario in which the rate of DnaA reactivation is proportional to the origin density. We show that an activation rate that is proportional to

the origin density effectively lowers the *datA* deactivation rate, thereby reducing the amplitude of the oscillations. Hence, our modelling predicts that if lipids are to play an activating role, the activation rate should not depend on the origin density, or at least not linearly.

EXPERIMENTS

Experimental observations While acidic phospholipids like cardiolipin (CL) and phosphatidylglycerol (PG) enhance the release of ADP and ATP from DnaA [86, 97, 101], it has also been reported that CL blocks the binding of ATP to DnaA [86]. Moreover, while phospholipids can restore replication activity of DnaA bound to ADP [86], it has also been observed that CL can inactivate nucleotide-free DnaA for replication initiation [86]. Furthermore, incubating DnaA-ADP or DnaA-ATP in the presence of phospholipids inhibits the binding of DnaA-ADP/ATP to *oriC* [97], and adding phospholipids speeds up the dissociation of DnaA-ADP/ATP from *oriC*, albeit weakly [97]. Of particular interest is the observation that while incubating DnaA-ADP with the lipids and *oriC* restores replication activity (irrespective of whether the other replication components were added initially or later) [97], first incubating DnaA-ADP with the lipids and then adding *oriC* later does not [97]; the temporal order in which components are added thus appears to matter. In addition, lipids lower the apparent affinity between DnaA and ATP, yet the presence of *oriC* during the incubation of DnaA with lipids restores it [97]. Consistent with this observation, the stimulating effect of the lipids on the release of nucleotides is weakened by the binding of DnaA-ADP/ATP to *oriC* [97].

Emerging picture The picture that emerges from these studies is that acidic phospholipids and nucleotides mutually exclude each other in (and thus compete for) binding DnaA. This could explain the observation that lipids both promote the release and inhibit the binding of nucleotides [86]; in turn, the release-promoting effect could explain why lipids can restore replication-initiation activity [86, 97]. In addition, the lipids also compete with *oriC* for binding DnaA. This may explain why adding lipids impedes the binding of DnaA to *oriC* [97] and hence replication initiation [86, 97], and why it stimulates the dissociation of DnaA-ADP/ATP from *oriC* [97] (the fact that this effect is weak indicates that the complex of *oriC* and DnaA is very stable). At the same time, DnaA retains its high affinity for nucleotides when bound to *oriC* [97]. These observations are consistent with structural studies [93], which show that DnaA binds the nucleotides and DNA (*oriC*) via different domains, while it binds the lipids via the domain (III) that also binds the nucleotides (hence their mutually exclusive binding), yet via a site that is on the border with the domain (IV) that binds *oriC*; the latter could explain why *oriC* and lipid binding are mutually exclusive.

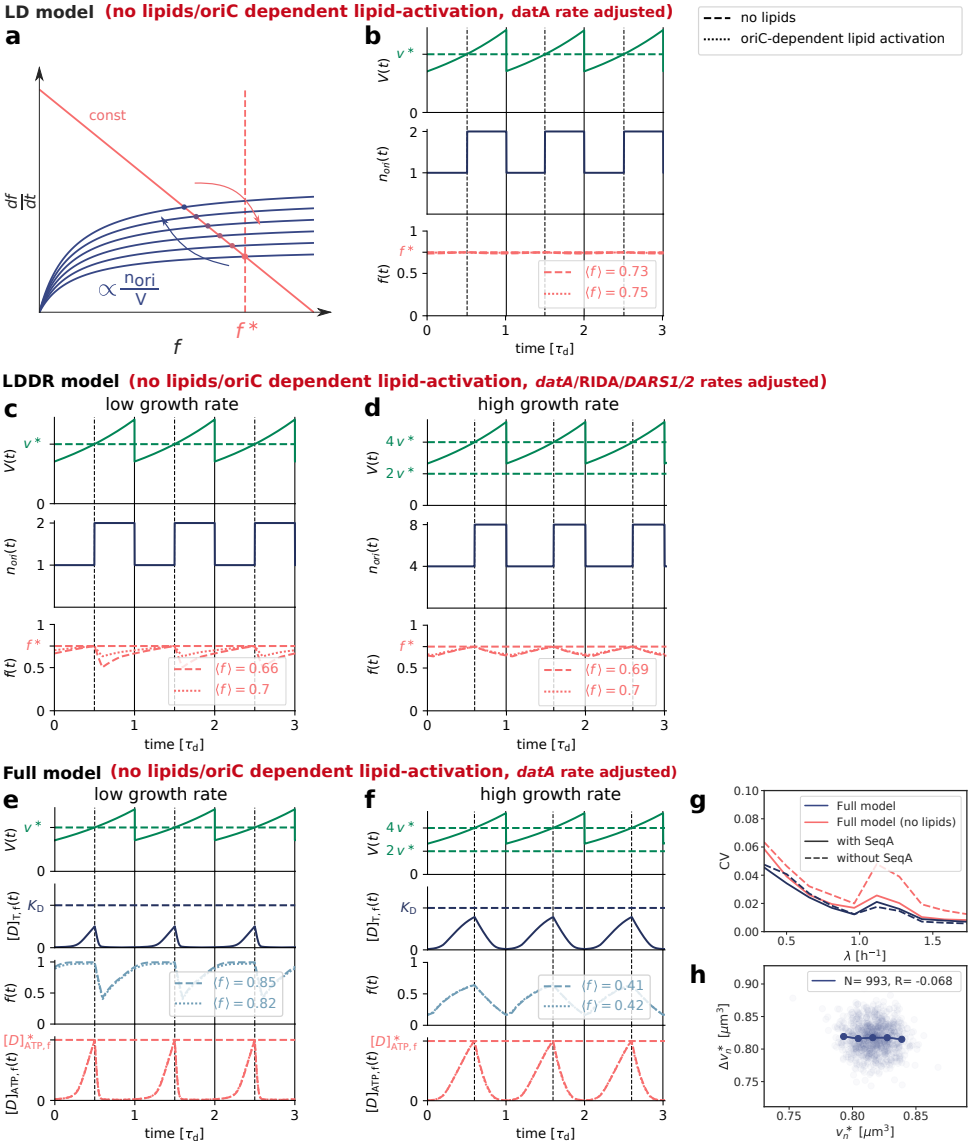


Figure 4.9: Effect of removing the lipids and of *oriC*-dependent lipid-activation in the LD, the LDDR and the full model (a) Combining activation via DnaA synthesis with deactivation via *datA*: The constant activation rate (red line) and the origin density-dependent deactivation rate (blue curve) as a function of the active fraction of the initiator protein f at different moments of the cell cycle (see equation 4.11). Contrary to the LD model, where both activation and deactivation are non-linear (see Fig. 3a of the main text), here the activation rate is linear in f, eliminating the ultra-sensitivity and thus giving rise to a smaller amplitude in the active fraction f. (caption continues on the next page)

Open questions The biggest open question is how to reconcile the observation that lipids in a mixture of DnaA-ADP, *oriC*, ATP and other replication components can initiate DNA synthesis (Table I of [97]), while ADP is much less likely to be released from DnaA by the lipids when the DnaA-ADP is bound to *oriC* (Figs. 3 and 4 of [97]). The observation of Crooke et al. that the temporal order in which *oriC* and lipids are added to DnaA-ADP [97] is particularly confusing, because the reactions involved are association-dissocation reactions (i.e. DnaA-lipid binding, DnaA-*oriC* binding, and DnaA-(*oriC*)-nucleotide binding), which can reach thermodynamic equilibrium; and in the equilibrium state, the temporal order in which the components have been added should be irrelevant. It is conceivable that the reactions do not reach equilibrium, which would make it hard to interpret the results. It is also possible that the results depend in a non-

(b, c, d, e, f) The volume of the cell $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$, the fraction of ATP-DnaA $f(t)$, the free DnaA concentration $[D]_{\mathrm{T,f}}(t)$ and the concentration of free ATP-DnaA $[D]_{\mathrm{ATP,f}}(t)$ as a function of time (in units of the doubling time τ_{d}) with $\tau_{\mathrm{d}} = 2$ h (b, c, e) and $\tau_{\mathrm{d}} = 25$ min (d, f). Replication is initiated at a critical initiator fraction f^* in the LD and LDDR model (red dashed line in b, c, d) and at a critical free, ATP-DnaA concentration $[D]^*_{\mathrm{ATP,f}}$ (red dashed line in e, f) in the full model. The average active fraction over one cell cycle $\langle f \rangle$ is indicated and replication is initiated at a constant volume per origin v^* over time (green dashed line). (b) LD model (no lipids): With DnaA synthesis (set by the growth rate $\lambda = 0.35\ \mathrm{h}^{-1}$) being the only activator, the rate of *datA* ($\beta_{\mathrm{datA}} = 20\ \mathrm{h}^{-1}$) must be small in order to ensure that replication is initiated at the experimentally observed initiation volume per origin v^*. Due to low (de)activation rates and the lack of ultra-sensitivity, the amplitude of the oscillations becomes extremely small and would not be sufficient to generate stable cell cycles in the presence of noise. LD model (*oriC*-lipids): Similarly, when lipid-activation is *oriC*-dependent (see equation 4.14), the amplitude of the oscillations in the active fraction becomes small due to a small effective deactivation rate from *datA* and the lipids (dotted line, $\alpha_{\mathrm{l}} = 750\ \mathrm{h}^{-1}$ and $\beta_{\mathrm{datA}} = 600\ \mathrm{h}^{-1}$). (c, d) LDDR model (no lipids): In the LDDR model, not only the rate of *datA* ($\beta^-_{\mathrm{datA}} = 20\ \mathrm{h}^{-1}$, $\beta^+_{\mathrm{datA}} = 120\ \mathrm{h}^{-1}$), but also the rates of *DARS1/2* and RIDA must be low ($\beta_{\mathrm{RIDA}} = 100\ \mathrm{h}^{-1}$, $\alpha^-_{\mathrm{d2}} = 10\ \mathrm{h}^{-1}$, $\alpha^+_{\mathrm{d2}} = 128\ \mathrm{h}^{-1}$, $\alpha^+_{\mathrm{d1}} = 10\ \mathrm{h}^{-1}$), because otherwise the replication cycles become unstable as described in Fig. 3.5 and section 3.3. Low (de)activation rates thus result in weak oscillations in the active DnaA fraction f both at low (c) and at high growth rates (d). LDDR model (*oriC*-lipids): When lipid-activation is *oriC*-dependent, the amplitude of the oscillations in the active fraction is again small due to the small effective deactivation rate via *datA* and the lipids (dotted line, $\alpha_{\mathrm{l}} = 100\ \mathrm{h}^{-1}$, $\beta^-_{\mathrm{datA}} = 120\ \mathrm{h}^{-1}$ and $\beta^+_{\mathrm{datA}} = 220\ \mathrm{h}^{-1}$). (e, f) Full model (see section 4.1): While the rate of *datA* must again be small (same *datA* rates as in c and d) to ensure the experimentally observed initiation volume, homeostasis (stability) is ensured via titration, allowing for high rates of RIDA and *DARS1/2* (set to same values as in original full model). The full model without lipids (dashed lines) or with *oriC*-dependent lipid activation (dotted lines) thus gives rise to stable cell cycles with large amplitude oscillations both at low (e) and at high growth rates (f). (g) The coefficient of variation $\mathrm{CV} = \sigma/\mu$ with the standard deviation σ and the average initiation volume $\mu = \langle v^* \rangle$ as a function of the growth rate for different models in the presence of noise in the DnaA concentration. Removing the lipids from the full model raises the coefficient of variation, especially at intermediate growth rates in the absence of SeqA (red dashed line). (h) In the full model without lipids, the added volume per origin between successive initiation events, $\Delta v^*_{\mathrm{n}} = 2\, v^*_{\mathrm{n+1}} - v^*_{\mathrm{n}}$, is independent of the initiation volume v^*_{n} per origin, for gene expression noise in DnaA (modelled according to equation 2.24 as explained in section 3.7). The doubling time is $\tau_{\mathrm{d}} = 35$ min and the number of data points N and the Pearson correlation coefficient R are indicated.

trivial manner on the (relative) concentrations of the components, especially because the reactions involve competitive binding. This is particularly pressing, because the concentrations in the *in vitro* experiments are likely to be very different from the effective concentrations *in vivo*; the latter depend not only on the concentrations of the lipids in the membrane, but also on the membrane area and the cytoplasmic volume of the cell.

Alternative activation scenarios. To make progress, we have analyzed alternative lipid-mediated DnaA activation scenarios. Besides the model of the main text (scenario 1), which assumes that DnaA is reactivated in the cytoplasm after it has dissociated from the membrane (see section 3.A), we considered two other scenarios: 2) DnaA-ADP binds to *oriC*, and then DnaA-ADP-*oriC* interacts with the lipids, leading to the exchange of ADP for ATP; 3) DnaA-ADP binds the lipids, ADP is released, *oriC* moves to the DnaA bound to the lipids, causing ATP to bind. Scenario 1 of the main text is consistent with the observation that lipids can stimulate the release of ADP [86], that DnaA can bind ATP [86, 97], and that DnaA-ATP can bind to *oriC* [17, 18, 79]; the only open question is on what timescale DnaA dissociates from the membrane [99, 100]. The second scenario is consistent with the observation that incubating DnaA-ADP with the lipids and *oriC* restores replication activity [97], but, as mentioned above, it seems at odds with the observation that lipids are less likely to induce the release of ADP when DnaA-ADP is bound to *oriC* [97]. Moreover, there is no evidence that *oriC* is associated with the membrane [68]. The same criticism applies to scenario 3. We therefore believe that the scenario of the main text, scenario 1, is the most likely scenario.

Yet, while scenarios 2 and 3 appear less likely, they also cannot be ruled out. Importantly, in scenarios 2 and 3 the rejuvenation of DnaA is contingent on *oriC*, making the lipid-mediated activation reaction dependent on the origin density. Below we therefore discuss a switch in which the activation rate scales with the origin density. Given the ambiguous experimental results discussed above, and the lack of quantitative, time-series data like that obtained for the Kai system [120, 121], we have not developed a detailed mathematical model that includes the competitive DnaA binding between the lipids, nucleotides, and *oriC*, but rather a coarse-grained model similar to that of the main text.

SWITCH WITH LIPID-MEDIATED ACTIVATION RATE THAT DEPENDS ON THE ORIGIN DENSITY

The dynamics of the active fraction in the LD model of the switch, yet with a lipid-mediated activation rate that depends on the origin density, is given by

$$\frac{df}{dt} = \lambda(1-f) + \tilde{\alpha}_{\mathrm{l}}[l][n_{\mathrm{ori}}]\frac{1-f}{\tilde{K}_{\mathrm{D}}^{\mathrm{l}}+1-f} - \tilde{\beta}_{\mathrm{datA}}[n_{\mathrm{ori}}]\frac{f}{\tilde{K}_{\mathrm{D}}^{\mathrm{datA}}+f} \tag{4.13}$$

$$= \lambda(1-f) - [n_{\mathrm{ori}}]\left(\tilde{\beta}_{\mathrm{datA}}\frac{f}{\tilde{K}_{\mathrm{D}}^{\mathrm{datA}}+f} - \tilde{\alpha}_{\mathrm{l}}[l]\frac{1-f}{\tilde{K}_{\mathrm{D}}^{\mathrm{l}}+1-f}\right) \tag{4.14}$$

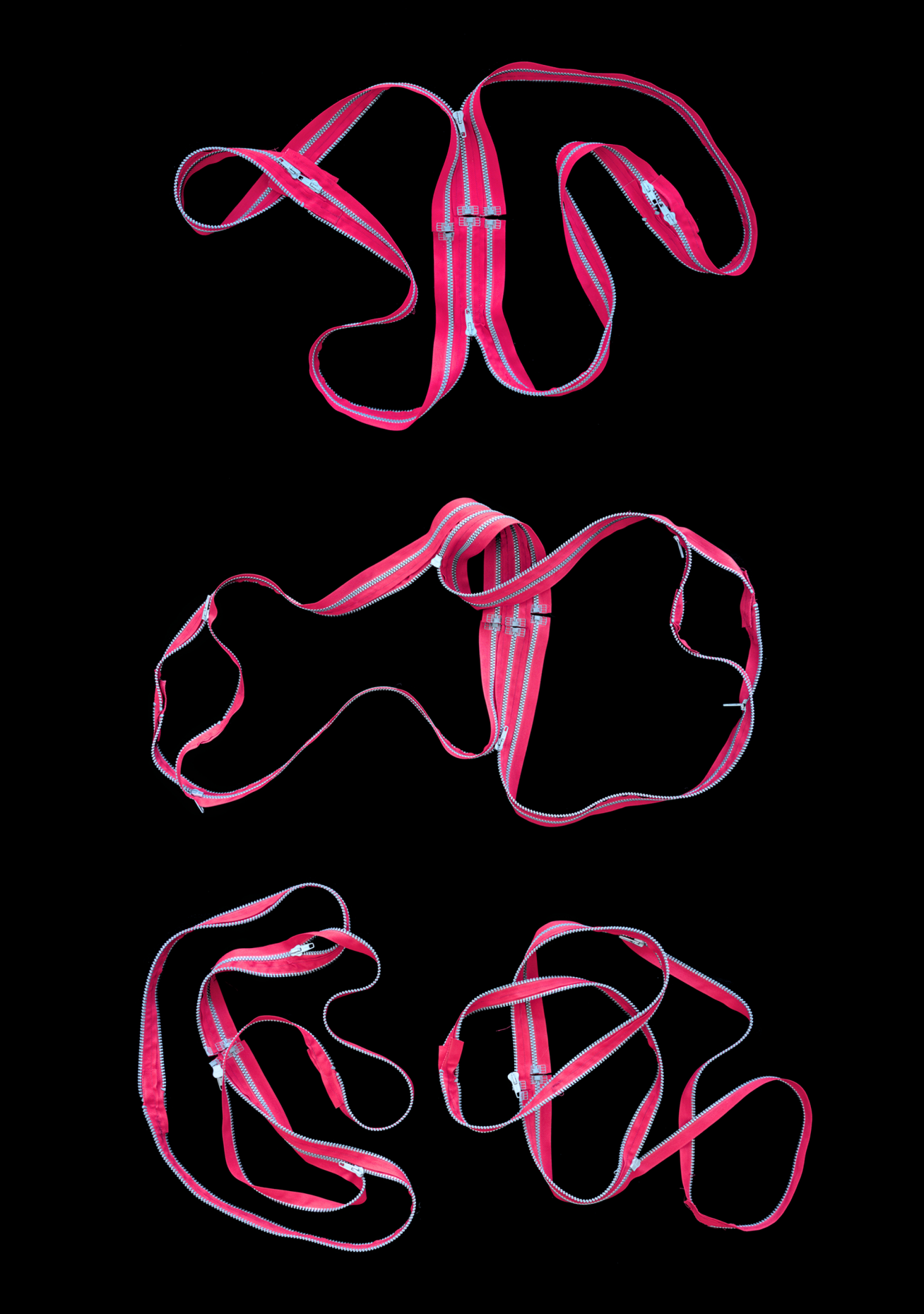
Clearly, a lipid-mediated activation rate that scales with the origin density tends to renormalize the *datA* de-activation rate: it indeed tends to lower the effective deactivation rate. Importantly, at the mathematical level, the system therefore becomes similar to the LD model in which the lipids have been taken out completely, with one activation rate

that is independent of the origin density and one effective deactivation rate that scales with the origin density. This effective deactivation rate, from *datA* and the lipids, needs to balance the activation rate from protein synthesis. And since even at high growth rates, the protein synthesis rate is relatively low as compared to the measured *datA* deactivation rate (see [63] and Table 3.1), the effective deactivation rate must be low as well in order to yield an initiation volume that is consistent with experiments. This yields oscillations of low amplitude, similar to the oscillations of the lipid-independent model (Fig. 4.9b).

We have also considered an LDDR model in which the lipid-mediated activation rate is proportional to the origin density. However, precisely as in the LD model, this merely renormalizes, i.e. lowers, the *datA* deactivation rate. The system therefore becomes mathematically similar to the LDDR model in which lipid-mediated DnaA activation is taken out entirely, with only one term that is independent of the origin density, namely DnaA activation via protein synthesis; and, again, because the protein synthesis rate is low, all other activation and deactivation rates must be low. The system therefore only exhibits very weak oscillations in the fraction of active DnaA, very similar indeed to those of the lipid-independent model (Fig. 4.9c/d, dotted lines in third panel).

For completeness, we have also considered a full model with a lipid-mediated DnaA activation rate that is proportional to the origin-density, but as found for the LD and LDDR model, the results are similar to the full model with the lipids taken out completely (Fig. 4.9e/f, dotted lines in third panel). Indeed, the full model is surprisingly robust to taking out the lipids fully or making its effect scale with the origin density, although the precision of replication initiation is highest when the effect of the lipids is independent of the origin density. Lipids can thus enhance replication initiation by promoting the exchange of DnaA-bound ADP for ATP, but only if the associated activation rate is independent of the origin density.

Crafted model of the *E. coli* chromosome to illustrate overlapping replication cycles
Zippers sewn together, 2019
Mareike Berger



5

SYNCHRONOUS REPLICATION INITIATION OF MULTIPLE ORIGINS

Initiating replication synchronously at multiple origins of replication allows the bacterium Escherichia coli to divide even faster than the time it takes to replicate the entire chromosome in nutrient rich environments. What mechanisms give rise to synchronous replication initiation remains however poorly understood. Via mathematical modelling, we identify four distinct synchronization regimes depending on two quantities: the duration of the so-called licensing period during which the initiation potential in the cell remains high after the first origin has fired and the duration of the blocking period during which already initiated origins remain blocked. For synchronous replication initiation, the licensing period must be long enough such that all origins can be initiated, but shorter than the blocking period to prevent reinitiation of origins that have already fired. We find an analytical expression for the degree of synchrony as a function of the duration of the licensing period, which we confirm by simulations. Our model reveals that the delay between the firing of the first and the last origin scales with the coefficient of variation (CV) of the initiation volume. Matching these to the values measured experimentally shows that the firing rate must rise with the cell volume with an effective Hill coefficient that is at least 20; the probability that all origins fire before the blocking period is over is then at least 94%. Our analysis thus reveals that the low CV of the initiation volume is a consequence of synchronous replication initiation. Finally, we show that the previously presented molecular model for the regulation of replication initiation in E. coli can give rise to synchronous replication initiation for biologically realistic parameters.

Passing on the genetic information from one generation to the next with high fidelity is crucial for the survival of every organism. Many bacteria contain several copies of their chromosome [22, 40, 130–133]. The bacteria *Escherichia coli* and *Bacillus subtilis* initiate DNA replication of several copies of the same chromosome synchronously in nutrient rich environments [22, 40, 131]. Also the mycobacterium *Mycobacterium smegmatis* has been shown to initiate 78% of their chromosomes synchronously before cell division in the mother cell [132]. The genetic code of the bacterium *Vibrio cholerae* is divided between two replicons and both replicons are initiated in synchrony [133]. Conversely, the cyano bacterium *Synechococcus elongatus* contains on average 3.3 chromosomes in exponentially growing cells and initiates replication both asynchronously and independent of the position of the chromosome in the cell [130]. While all these bacteria are evolutionarily divergent and have very different molecular mechanisms to control the initiation of replication, some of them share similar features, like the degree by which replication of several chromosomes is initiated synchronously. Yet, what underlying principles and molecular mechanisms determine the degree of synchrony of replication initiation of multiple chromosomes remains poorly understood.

The bacterium *E. coli* initiates replication synchronously with a very high precision. Already in the 1960s, Cooper and Helmstetter suggested that initiating new rounds of replication synchronously at several origins enables *E. coli* to divide even faster than the fixed time it takes to replicate its entire chromosome [20]: Rounds of replication that started in the mother cell continue to being replicated during cell division and finish only in the following generations (Fig. 5.1a). To ensure that all daughter cells obtain a fully replicated copy of the chromosome at these high division times, replication must be initiated at all chromosomes synchronously. Later, Skarstad et al. confirmed the prediction of Cooper and Helmstetter by counting the numbers of origins in rapidly growing cultures: They found that most cells have 2^n ($n = 1, 2, 3$) copies of their chromosome and only a small fraction of cells (2-7%) contained 3, 5, 6 or 7 chromosomes [134]. Recent single cell measurements indeed show that *E. coli* initiates replication synchronously at up to eight origins with a very high precision in the fast growth regime [22, 135]. It remains however an open question how *E. coli* achieves such a high degree of synchrony.

Replication initiation in *E. coli* is controlled by the initiator protein DnaA. This protein can switch between two nucleotide-binding states, an inactive state in which DnaA is bound to ADP and an active one in which it is bound to ATP. Both the inactive and active form can bind to an origin of replication, but binding of the inactive state is not sufficient: replication initiation requires the binding of ATP-DnaA [17, 43]. The evidence is accumulating that the origin binding of DnaA and hence replication initiation is controlled via two distinct mechanisms, titration and protein activation [43, 64, 78]. Titration of DnaA via sites on the chromosome generates a cycle in the concentration of free DnaA that is available for binding to the origin, while an activation switch induces a cycle in the fraction of active DnaA. These two cycles together conspire to generate robust oscillations in the concentration of free and active DnaA [39]. This concentration of free and active DnaA forms the initiation potential of the cell, which determines the propensity of origin firing.

Initiation synchrony entails that all origins are initiated during each cell cycle, yet also only once per cell cycle. This is a major challenge, because the cell needs to meet

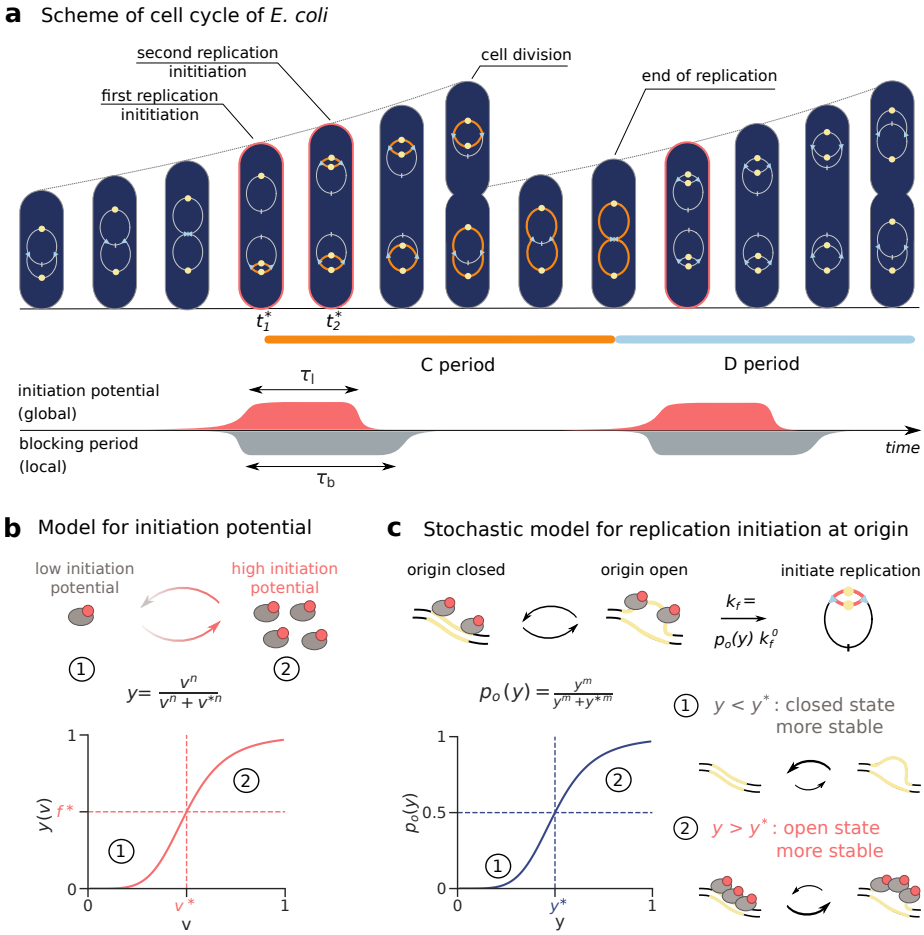


Figure 5.1: Model of stochastic replication initiation at each origin. (a) Scheme of the cell cycle of *E. coli*: The volume of the cell grows exponentially with a growth rate λ. At doubling times $\tau_d = \ln(2)/\lambda$ that are shorter than the time to replicate the entire chromosome and divide (C+D period), cells are typically born with an ongoing round of chromosomal replication. Replication is initiated stochastically at each origin (yellow circles) at times t_1 and t_2, respectively, and the replication forks (blue triangles) advance towards the terminus (grey bar) with a constant replication speed. In *E. coli*, all origins fire within a very short time interval, thus giving rise to synchronous replication initiation. To fire replication synchronously a global and a local mechanism are required: the global mechanism keeps the initiation potential high for a licensing period τ_l (red shaded area), while the local mechanism based on SeqA prevents already initiated origins from refiring for a blocking period $\tau_b > \tau_l$ (grey shaded area). In our model, cell division is triggered a fixed cycling time $\tau_{cc} = T_C + T_D$ after replication has been initiated. (b) We model the initiation potential y in the cell as a function of the volume per origin $v = V/n_{ori}$ via a Hill function with the Hill coefficient n. At the critical volume per origin v^* the initiation potential equals the critical initiation potential $y^* = 0.5$. (c) Stochastic model of replication initiation at the origin as a function of the initiation potential in the cell: The origin can be in an open or in a closed configuration and replication can be initiated with a constant rate k_f^0 if the origin is open. The probability to be in the open state $p_o(y)$ depends on the initiation potential in the cell and is modeled via a Hill function with the Hill coefficient m and the critical active fraction of DnaA f^*.

two potentially conflicting constraints. The requirement that all origins must fire during each cell cycle means that when the first origin fires, the initiation potential cannot go down immediately: it must continue to rise so that also the other origins can fire. On the other hand, the origin that has fired, should not fire again, even though the initiation potential is still rising. It appears that *E. coli* employs two distinct mechanisms to meet these two constraints. The oscillations in the initiation potential, the concentration of free and active DnaA, constitute a global mechanism that induces not only the first origin to fire, but also prompts, and allows, the remaining origins to fire (Fig. 5.1a). To prevent the immediate reinitiation of origins that have already fired, a local mechanism is used.

The local mechanism that prevents the immediate reinitiation of new replicated origins is based on the so-called sequestration of these origins. In *E. coli*, after an origin has initiated replication, the protein SeqA transiently binds to this origin and thus prevents that new rounds of replication start immediately again at the same origin [44, 58]. When either of the two proteins SeqA or Dam that are required for sequestration after replication initiation are deleted, synchrony is lost and replication is initiated throughout the entire cell cycle [44]. While the bacterium *B. subtilis* lacks the protein SeqA, it instead contains the protein Spo0A, which can inhibit replication initiation in the *B. subtilis* phage ϕ29 in vivo and has been shown to bind to specific sites on the origin in vitro [17]. These experiments suggest that Spo0A, similar to SeqA in *E. coli*, represses chromosomal replication by binding directly to the origin region of *B. subtilis*. Blocking of recently initiated origins during a so-called blocking period is therefore an essential mechanism to ensure synchronous replication initiation (Fig. 5.1a).

The combination of global oscillations in the initiation potential, which induce all origins to fire, and local origin sequestration, which prevents the newly replicated origins from reinitiation, appears to be an elegant solution to the problem of initiation synchrony. Yet, many questions remain. Newly replicated origins are only sequestered for a finite amount of time: the blocking period is about 10 minutes long [17, 59, 60]. Hence, while, after the initial origin has fired, the initiation potential must first continue to rise sufficiently long in order to allow all the remaining origins to fire, it must also come down before this blocking period is over, because otherwise the newly replicated origin(s) will fire again after all. The licensing period during which origins can fire must thus be long enough for all origins to fire, yet also shorter than the blocking period (Fig. 5.1a). Given that the blocking period is only 10 minutes, this constraint is likely to pose a major challenge.

The problem of replication synchrony is compounded by the fact that the oscillations in the initiation potential are directly shaped by replication initiation itself. When a new origin is fired, the newly generated replisomes will stimulate the deactivation mechanism called RIDA. Moreover, a few minutes after an origin has initiated DNA replication, the locus *datA* is duplicated, which enhances deactivation by stimulating the hydrolysis of ATP bound to DnaA. Furthermore, the newly duplicated DNA will harbor new titration sites, which also tend to reduce the initiation potential by lowering the concentration of cytoplasmic DnaA. How these molecular mechanisms cause the initiation potential to first continue to rise during the licensing period and then fall before the blocking period is over is far from understood.

To study how replication can be initiated synchronously at several origins, we first

propose a minimal coarse-grained model in which an initiation potential rises when the cell reaches a critical volume per origin. Each origin can initiate stochastically with a firing probability that depends on the initiation potential. We show that for synchronous replication initiation, origins that have already fired must be prevented from reinitiation during a blocking period and that the initiation potential must remain high after the first origin has fired during a licensing period. By varying the duration of the licensing and the blocking period we reveal four synchronization regimes. In order to initiate synchronously, the licensing period must be long enough for all origins to fire, yet shorter than the blocking period. However, given that the measured blocking period is only 10 minutes [17, 59, 60], the licensing period must be shorter than 10 minutes. To fire all origins within this blocking period with a success rate of 94%, the firing rate must rise with the volume with a Hill coefficient of at least 20, such that the average time between the first and last initiation event is less than 4 minutes, as measured experimentally by Skarstad et al. [134]. Our modelling thus provides a rationale for the question of why DNA replication initiation in *E. coli* is so tightly controlled.

We then investigate how these general synchronization requirements could be realized in the bacterium *E. coli*, by replacing the coarse-grained initiation potential with the previously proposed molecular model, in which the free ATP-DnaA concentration oscillates over the course of the cell cycle. We find that if replication is initiated dominantly by an activation switch, initiation synchrony is only achieved if the (de)activation rates of the switch components are very high, the switch is in the ultra-sensitivity regime and for parameter combinations that do not agree with the experimentally measured values. Adding titration and bringing the system in a regime where the DnaA concentration in the cytoplasm is low during most of the cell cycle significantly improves the degree of synchrony by sharpening the rise of the initiation potential at a critical volume per origin. This suggests that combining titration with protein activation is crucial for initiating replication synchronously at multiple origins in the bacterium *E. coli*.

5.1. LICENSING PERIOD MUST BE NON-ZERO AND SHORTER THAN BLOCKING PERIOD

Our coarse-grained model to study the effect of stochastic replication initiation on the cell cycle consists of two parts: Firstly, we model the available amount of initiator proteins in the cell as an initiation potential y that depends on the volume per origin $v(t) = V(t)/n_{\mathrm{ori}}(t)$ according to

$$y(v) = \frac{v^n}{v^n + v^{*n}} \tag{5.1}$$

with the Hill coefficient n and the critical volume per origin v^* (Fig. 5.1b). Secondly, each origin is modelled as a two-state system that can switch stochastically between an open and a closed configuration (see Appendix 5.A for details). Exploiting that the origin is more likely to be in the open state when the initiation potential y in the cell is high, we assume that the probability to be in the open state p_{o} increases with the activation potential y as

$$p_{\mathrm{o}}(y) = \frac{y^m}{y^m + y^{*m}} \tag{5.2}$$

with the Hill coefficient m and the critical initiation potential y^* (Fig. 5.1c). Molecularly, this non-linear opening probability $p_o(y)$ could arise via cooperative binding of the initiator to the origin or via a Monod-Wyman-Changeux model, where the open configuration becomes more energetically favorable the more initiators bind to the origin. Assuming rapid opening and closing dynamics of the origin, the origin firing rate is given by the probability to be in the open state p_o times the maximal firing rate k_f^0:

$$k_f = k_f^0 \, p_o \tag{5.3}$$

To investigate the effect of stochastic replication initiation on the cell cycle of *E. coli*, we model the volume $V(t)$ of the cell as an exponential function, $V(t) = V_b \, e^{\lambda t}$, where the growth rate $\lambda = \ln(2)/\tau_d$, with cell-doubling time τ_d, is a model parameter. We track the number of chromosomes together with their state of replication (e.g. fully replicated or replication ongoing) and whenever an origin fires a new round of replication at time t^*, a new division time a constant cycling time τ_{cc} after replication initiation is set at $\tau_{div} = t^* + \tau_{cc}$. The constant cycling time τ_{cc} is given by the sum of the time to replicate the entire chromosome T_C and the time from the end of replication until cell division T_D (Fig. 5.1). Every set division time τ_{div} is linked to the chromosome that just initiated replication. This choice ensures that a cell never divides before the chromosome has been replicated. When the next division time is reached, the cell volume is divided by two and one of the two daughter chromosomes is kept at random for the next cell cycle. When the cell inherits a chromosome that is already being replicated but has not yet reached its division time, it also inherits the next division time (see also Fig. 2.1a).

For synchronous replication initiation at several origins, the initiation potential must remain high after the first initiation event during a licensing time τ_l and already initiated origins must be prevented from reinitiation during a blocking time $\tau_b > \tau_l$. To study the effect of stochastic replication initiation on the degree of synchrony, we consider the fast growth regime ($\tau_d < \tau_{cc}$), where there are typically two or more origins in the cell at the moment of replication initiation. At the critical volume per origin v^*, the activation potential $y(t)$ rises and the probability to initiate replication $p_o(t)$ increases strongly (Fig. 5.2, lowest panel). When the first origin fires, the number of origins in the cell increases stepwise and the volume per origin $v(t)$ drops instantaneously (Fig. 5.2, second and third panel). If the initiation potential $y(t)$ (and therefore also the opening probability $p_o(y)$) followed the change in the volume per origin instantaneously, it would become very unlikely for the second origin to initiate replication as well, resulting in asynchronous replication initiation. We therefore introduce a licensing time τ_l, during which the initiation potential does not yet sense the change in the volume per origin $v(t)$ and continues to rise (Fig. 5.2). The opening probability $p_o(t)$ therefore rises sharply during this licensing time and the second origin also initiates replication stochastically. In order to prevent that already initiated origins fire again, we additionally introduce a blocking period τ_b, during which replication cannot be initiated again at the same origin. At the end of the licensing time, the activation potential is updated according to the current number of origins in the cell and it thus drops instantaneously (Fig. 5.2). For a sufficiently long licensing time and a block period that is longer than the licensing time, $\tau_b > \tau_l$, we indeed obtain stable cell cycles with synchronous replication initiation events (Fig. 5.2).

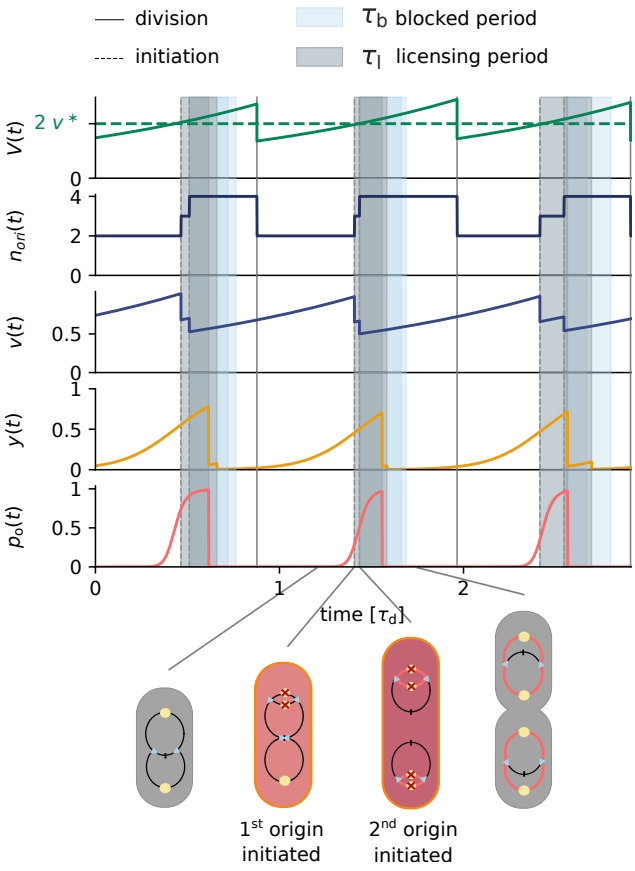


Figure 5.2: Replication is initiated synchronously at several origins by introducing a blocking and a licensing period. The volume $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$, the volume per number of origins $v(t) = V(t)/n_{\mathrm{ori}}(t)$, the initiation potential $y(t)$ and the opening probability $p_{\mathrm{o}}(t)$ as a function of time (in units of the doubling time of the cell τ_{d}). Every origin is initiated stochastically (dashed vertical grey lines) and during the blocking period τ_{b} (light blue shaded area), the newly replicated origins cannot be re-initiated. The initiation potential $y(t)$ and the opening probability $p_{\mathrm{o}}(t)$ continue to increase during the licensing period τ_{l} (grey shaded area), such that the remaining origins that have not yet initiated replication are also initiated. At the end of the licensing period, the initiation potential $y(t)$ and therefore also the opening probability $p_{\mathrm{o}}(t)$ instantaneously decreases to a lower value, making re-initiation highly unlikely. At cell division (vertical solid grey lines), the cell volume is divided by two and one of the two chromosomes is chosen at random for the next cell cycle. The cartoon below shows cells and their circular chromosome at four different moments of the cell cycle. Replication is initiated at the origin (yellow circle) and advances to the terminus (black bar) via two replication forks (light blue triangles). Blocked origins are marked by red crosses and the shaded color of the cell indicates whether the initiation potential in the cell is low (grey color) or high (red color). (See Table 5.1 for all parameters.)

To quantify the degree of synchrony of replication initiation for a given parameter set, we define the degree of synchrony s as the change of the number of origins Δn_{ori} from the beginning of an initiation cascade t_{i} to the end of an initiation cascade t_{f}, relative to the number of origins $n_{\mathrm{ori}}(t_{\mathrm{i}})$ at the beginning of the cascade (Fig. 5.3a):

$$s = \frac{\Delta n_{\mathrm{ori}}}{n_{\mathrm{ori}}(t_{\mathrm{i}})} = \frac{n_{\mathrm{ori}}(t_{\mathrm{f}}) - n_{\mathrm{ori}}(t_{\mathrm{i}})}{n_{\mathrm{ori}}(t_{\mathrm{i}})} \tag{5.4}$$

The beginning of the initiation cascade t_{i} is given by the time at which the first origin fires and the initiation cascade ends at $t_{\mathrm{f}} = t_{\mathrm{i}} + \tau_{\mathrm{l}}$, when the licensing period of the first origin that has fired is over. When the degree of synchrony s is one, replication is initiated synchronously, as all origins that were present at the beginning of the initiation cascade have fired (Fig. 5.1a). For $s < 1$ or $s > 1$, replication is under- or over-initiated, respectively (Fig. 5.3a). We measure the degree of synchrony s for many cell cycles to obtain the average degree of synchrony $\langle s \rangle$ for any parameter set.

By varying the duration of the licensing and the blocking period, we find four different synchronization regimes (Fig. 5.3). All simulations start with a single, fully replicated chromosome and if replication is initiated in perfect synchrony, the system settles to a state, where the number of origins oscillates between two and four. However only in regime four is replication initiated synchronously (Fig. 5.3b, regime 4). When the blocking period τ_{b} is zero but the licensing period τ_{l} is larger than zero, replication is severely over-initiated, such that no stable cell cycles can be obtained (Fig. 5.3b, regime 1, grey fields). If on the other hand the licensing period τ_{l} is zero or very short and the blocking period τ_{b} is larger than zero, we obtain a highly under-synchronized cell-cycle: After each initiation event, the initiation potential drops rapidly, preventing further initiation events. This results in periodic, individual initiation events throughout the entire cell cycle (Fig. 5.3b, regime 2). When both the licensing and the blocking period are non-zero and the licensing period is longer than the blocking period, $\tau_{\mathrm{l}} > \tau_{\mathrm{b}}$, origins that have already fired can fire again after the end of the licensing period. This results in initiation cascades where all origins fire synchronously, but with too many initiation events. As can be seen in Fig. 5.3b, in this third regime the number of origins goes from one to four in one initiation cascade instead of oscillating between two and four. We therefore call this regime "over-synchronized". Replication is only initiated synchronously once per cell cycle, when the licensing period τ_{l} is sufficiently large and the blocking period is even larger $\tau_{\mathrm{b}} > \tau_{\mathrm{l}}$ (Fig. 5.3b, regime 4).

5.2. STEEP RISE IN THE ORIGIN OPENING PROBABILITY IS ESSENTIAL

A key question is what controls the size of the synchronization regime 2. While the transition from the over-initiation (Fig. 5.3b, regime 3) to the perfect synchronization regime (Fig. 5.3b, regime 4) is sharp and clearly defined by the requirement that $\tau_{\mathrm{b}} > \tau_{\mathrm{l}}$, the transition from the under-synchronization (Fig. 5.3b, regime 2) to the perfect synchronization regime (Fig. 5.3b, regime 4) is smooth and there is no clear separation between these two regimes. In the following, we show that the average degree of synchrony $\langle s \rangle$ at different delay periods τ_{l} and Hill coefficients n and m can be derived from the proba-

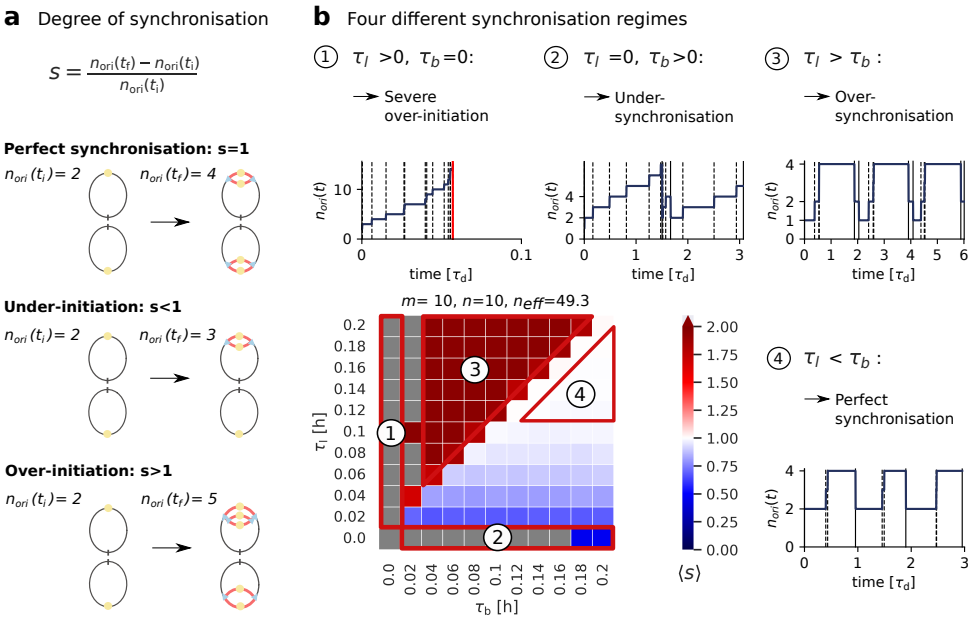


Figure 5.3: Replication is only initiated synchronously when the licensing period is sufficiently long, yet shorter than the blocking period. (a) The degree of synchronization s of an initiation cascade is given by the number of origins at the end of the initiation cascade $n_{\mathrm{ori}}(t_{\mathrm{f}})$ minus the number of origins at the beginning of the initiation cascade $n_{\mathrm{ori}}(t_{\mathrm{i}})$ relative to the number of origins at the beginning of the initiation cascade $n_{\mathrm{ori}}(t_{\mathrm{i}})$. An initiation cascade begins with the moment where the first origin fires and ends after the licensing time τ_{l}. Replication is initiated synchronously when all origins that were present at the beginning of the cascade have fired exactly once during the cascade ($s = 1$) and replication was under- or overinitiated when less or more origins have initiated, respectively. (b) The average degree of synchronization $\langle s \rangle$ as a function of the licensing period τ_{l} and the blocking period τ_{b}. The effective Hill coefficient n_{eff} is obtained by fitting the opening probability $p_{\mathrm{o}}(f(\nu))$ to a Hill function $p_{\mathrm{o}}(\nu)$ (Appendix 5.B). For each parameter set, the average degree of synchronization was obtained from $N = 5000$ consecutive cell cycles. We show example time traces of the number of origins as a function of time (in units of the doubling time of the cell τ_{d}) for four different synchronization regimes as indicated in the heatmap. When no cell cycle could be obtained, the field in the heatmap is marked in grey. (See Table 5.1 for all parameters.)

bility that two independent origins fire within a time interval $\Delta t < \tau_l$.

To calculate the probability that two independent firing events happen within a time interval Δt, we first derive an analytical expression for the instantaneous firing probability $k_f(t) = k_f^0\, p_o(t)$. In our model, the opening probability $p_o(y)$ depends indirectly on the time-dependent volume per origin $v(t)$ via the activation potential $y(v)$. The opening probability as a function of the volume per origin v can however be approximated by a Hill function (see Appendix 5.B for derivation)

$$p_o(v) \approx \frac{v^{n_{\text{eff}}}}{v^{n_{\text{eff}}} + v^{*n_{\text{eff}}}} \tag{5.5}$$

with the effective Hill coefficient

$$n_{\text{eff}} = \frac{n\,m}{2}. \tag{5.6}$$

This is a good approximation for the opening probability $p_o(y(v))$, when both the Hill coefficient of the activation potential and that of the opening probability, n and m, respectively, are relatively high (see Fig. 5.11c and d). The firing rate is then given by equation 5.3 with the approximate opening probability $p_o(v)$ from equation 5.5. In the following, we use the the procedure proposed in Ref. [22] where the maximal firing rate k_f^0 is chosen such that the average initiation volume $\langle v^* \rangle$ equals the theoretical initiation volume v^* in equation 5.5 (See Appendix 5.C). Using this analytical approximation for the opening probability in the regime of sufficiently high Hill coefficients n and m, we can now calculate the probability that two independent initiation events at times t_1 and $t_2 > t_1$ happen within a time interval $\Delta t = t_2 - t_1 \leq \tau_l$ (Appendix 5.D). In order to compare this probability $\langle P(\Delta t < \tau_l) \rangle$ to the degree of synchronization obtained from the simulations in the growth regime where two origins are present at the beginning of an initiation cascade, we re-scale the probability to range from $s_{\min} = 0.5$ to $s_{\max} = 1$:

$$s_{\text{th}} = 0.5 + \langle P(\Delta t < \tau_l) \rangle \times 0.5 \tag{5.7}$$

The average degree of synchrony $\langle s \rangle$ as a function of the licensing period τ_l for different Hill coefficients n and m is indeed very well approximated by s_{th} (Fig. 5.4a). The transition from the under-synchronized to perfect synchronization regime in Figure 5.3b is therefore given by the probability that two independent origin firing events happen within a short time window given by the licensing time τ_l. The higher the effective Hill coefficient n_{eff}, the higher the degree of synchrony for a give delay period τ_l (Fig. 5.4a). The degree of synchrony $\langle s \rangle$ increases with the effective Hill coefficient because that raises the firing rate more steeply, making it more likely that the two origins fire within the licensing time τ_l.

While the degree of synchronisation $\langle s \rangle$ increases with τ_l, τ_l cannot be longer than the blocking period τ_b, because otherwise origins that have already fired will fire again. The blocking period thus bounds τ_l. In the bacterium *E. coli*, the origin of the blocking period τ_b is well understood: The protein SeqA can bind to specific sites on the newly replicated origins which overlap with the binding sites for the initiation protein and thus prevent it from reinitiating. After about ten minutes, the protein SeqA unbinds and new rounds of replication can start again [17, 59, 60]. As the licensing time must be shorter than the blocking period in order to prevent over-initiation (Fig. 5.3b, regime 3), the

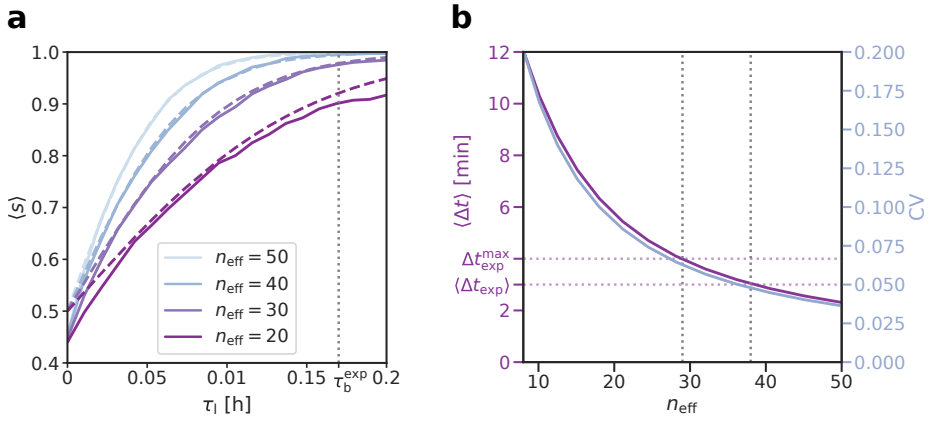


Figure 5.4: The experimentally observed high precision of replication initiation is required to ensure synchronous replication initiation at multiple origins. (a) The average degree of synchrony $\langle s \rangle$ as a function of the licensing time τ_l for varying effective Hill coefficients n_{eff} (with $n = m = \sqrt{2\, n_{\text{eff}}}$) from the simulations (solid lines) agrees well with the theoretical prediction derived in the Appendix 5.D (dashed lines). The small difference between the simulations and theory at very low delay periods arises from the fact that while in the theory for two synchronous firing events the minimal degree of synchrony is $s_{\min} = 0.5$, in the simulations there can be more origins at the beginning of an initiation cascade, leading to a lower degree of synchrony $s_{\min} < 0.5$. In these simulations, the blocking period τ_b is set larger than all tested licensing τ_l periods ($\tau_b = 0.25\,$h), such that over-initiation events cannot occur. The maximal firing rate k_f^0 is set such that the average initiation volume $\langle v^* \rangle$ equals the theoretical initiation volume v^* in equation 5.5 as explained in Appendix 5.C. The experimentally measured blocking period τ_b^{exp} is marked as a grey vertical dotted line. For each parameter set, the average degree of synchrony was obtained from $N = 5000$ consecutive cell cycles. (b) The theoretical average time interval between two consecutive firing events $\langle \Delta t \rangle$ (pink line and axes) and the coefficient of variation of the initiation volume (CV, blue line and axes) as a function of the effective Hill coefficient n_{eff} (see Appendix 5.E for derivation). Skarstad et al. [134] found experimentally that the average time interval to fire all origins in the B/r A *E. coli* strain is about $\langle \Delta t_{\text{exp}} \rangle = 3\,$min with an upper estimate of $\Delta t_{\text{exp}}^{\max} = 4\,$min (horizontal dotted pink lines). The effective Hill coefficient lies therefore in the range $n_{\text{eff}} = 29-38$ (vertical dotted grey lines), corresponding to a coefficient of variation of CV=0.05-0.07. This agrees well with the experimentally measured precision of the initiation volume of CV≤0.1 [22, 23, 135]. Interestingly, the average degree of synchrony at $n_{\text{eff}} = 30$ and $n_{\text{eff}} = 40$, respectively, is given by $\langle s \rangle(n_{\text{eff}} = 30) = 0.975$ and $\langle s \rangle(n_{\text{eff}} = 40) = 0.996$, corresponding to the probabilities to fire all origins synchronously of $\langle P(\Delta t < \tau_l) \rangle(n_{\text{eff}} = 30) \equiv P_s(n_{\text{eff}} = 30) = 95.5\%$ and $P_s(n_{\text{eff}} = 40) = 98.9\%$. This prediction of the degree of synchrony agrees well with the qualitative experimental observation that in *E. coli* DNA replication is typically initiated synchronously at multiple origins. Our model therefore provides a rational for the experimentally observed high precision of replication initiation. (See Table 5.1 for all parameters.)

licensing time in *E. coli* must be less than τ_b^{exp} =10 min. Fig. 5.4a shows this puts a major constraint on the Hill coefficient: to get a degree of synchronisation $\langle s \rangle$ that is above 95%, the effective Hill coefficient must be at least $n_{eff} = 30$.

The question remains what the effective Hill coefficient n_{eff} is. Interestingly, Skarstad et al. has measured the time between the first and last firing event, which we can compare against our theoretical prediction [134]. However, to do so, we must first examine the dependence on the growth rate, because Skarstad performed their measurements at a higher growth rate. Fig. 5.13 shows that while the average degree of synchrony $\langle s \rangle$ as a function of the licensing time τ_l varies strongly with the effective Hill coefficient n_{eff}, it is fairly independent of the doubling time of the cell τ_d.

Given that $\langle s \rangle$ as a function of τ_l is fairly independent of the growth rate, we now examine the data of Skarstad et al. [134]. To obtain an experimental estimate for the effective Hill coefficient and thus for the average degree of synchrony, we calculate the average time interval between the first and last initiation event $\langle \Delta t \rangle$ (see Appedix 5.E) and compare it to the experiments. Skarstad et al. find that this time is on average $\langle \Delta t_{exp} \rangle = 3$ min with an upper limit of $\Delta t_{exp}^{max} = 4$ min [134]. Our theory predicts that to fire two initiation events within an average time interval of $\langle \Delta t \rangle = 3-4$ min, the effective Hill coefficient must be in the range $n_{eff} = 29-38$ (Fig. 5.4b, vertical grey dotted lines), corresponding to a coefficient of variation (CV) of the initiation volume CV$\approx 0.05-0.06$. This agrees well with the experimental finding that the initiation volume is one of the most tightly controlled cell cycle parameters with CV=0.08-0.1 [22, 23]. A CV of 0.1 as measured by Ref. [22] corresponds to $n_{eff} \approx 20$ (Fig. 5.4b, see also calculation in Ref. [22]) and would thus only result in a relatively low degree of synchrony of less than $\langle s \rangle = 0.92$ corresponding to a probability of initiating synchronously of $\langle P(\Delta t < \tau_l) \rangle \equiv P_s = 84\%$. Recent experiments show however that the contribution from the intrinsic noise in replication initiation to the CV is only about CV_{int}=0.04 - 0.05 [135]. Our model, which only concerns the effect of intrinsic noise, then predicts that n_{eff} must be at least 40 (Fig. 5.4b), which then means that at least 98% of the initiation events happen synchronously within a period of 10 min (Fig. 5.4a).

Before we conclude, we must discuss one key parameter, which is the firing rate k_f^0. In our theoretical model, we covaried k_f^0 with n_{eff} to keep the average initiation volume per origin $\langle v^* \rangle$ constant and equal to v^* of Eq. 5.5, akin to the procedure of Wallden et al. [22]. Fig. 5.14 shows however the degree of synchronisation $\langle s \rangle$, the average initiation volume $\langle V^* \rangle$, and the CV of the initiation volume as a function of k_f^0, keeping the effective Hill constant at a large value $n_{eff} = 50$. It is seen that $\langle V^* \rangle$ is not very sensitive to the firing rate. Moreover, while the precision of initiation (higher $\langle s \rangle$, lower CV) increases for small k_f^0, it plateaus for $k_f^0 > 100\ h^{-1}$. Importantly, for the limiting k_f^0 (or $k_f^0 \to \infty$), the Hill coefficient necessary to match the Skarstad data is lower than the theoretical model above predicts, but it is still very high: it should be at least $n_{eff} = 20$ (Fig. 5.15). Our model of stochastic replication initiation thus provides a rationale for the experimentally observed high precision of DNA replication initiation in *E. coli*: Given the constraint set by the duration of the blocking period, the system requires a very high Hill coefficient in order to initiate replication synchronously. This indicates that synchronous replication initiation is a major challenge for the cell.

To summarize, we have shown that the smooth transition from the under-synchronized

regime to perfect synchronization in Figures 5.3b is well described by the probability that all the required initiation events happen independently during the licensing period τ_l. To fire all origins with a success rate of 94% before the blocking period is, the firing rate must rise with the cell volume with a Hill coefficient that is at least 20. Since increasing the Hill coefficient beyond this already large value becomes progressively harder, it seems that the system operates close to what is theoretically possible given the duration of the blocking period.

5.3. INITIATION SYNCHRONY IN MOLECULAR MODEL FOR E. COLI

Our coarse-grained model of replication initiation revealed general requirements for initiating replication synchronously at several origins. It remains however an open question how these requirements are implemented on a molecular level in different organisms. In *E. coli*, both a protein activation cycle and a concentration cycle are required for robust replication initiation at all growth rates (see Chapters 2, 3 and 4). In the following, we investigate how the well studied biochemical network for replication initiation of *E. coli* achieves initiation synchrony. For simplicity, we here first study the effect of stochastic replication initiation on the degree of synchrony in an activation switch driven only by the deactivator *datA* and the activating lipids in the simple Lipid-*DatA* (LD) model. In a second step, we also include all known (de)activators and the titration sites.

An optimal delay in replicating the deactivation site *datA* after the origin is required for synchronous replication initiation in the LD model. To test the effect of stochastic origin firing in a more molecular model, we replace the abstract initiation potential $y(v)$ we used in the coarse-grained model with the LD model for the ATP-DnaA fraction $f(t)$ in the cell. The opening probability $p_o(f)$ is again modeled as a simple Hill function according to equation 5.2 with Hill coefficient $m = 10$ and $f^* = 0.5$ and the initiation potential y is now given by the ATP-DnaA fraction $f(t)$ in the cell. Like in the coarse-grained model, already initiated origins are blocked transiently during a blocking period of τ_b =10 min. While the deterministic LD model (see section 3.1) ensures stable replication initiation cycles, combining it with stochastic replication initiation leads to asynchronous replication initiation (Fig. 5.5b). Importantly, in the deterministic LD model we have assumed that due to the small distance between the chromosomal site *datA* and the origin, the *datA* site is replicated instantaneously after replication initiation. This leads to a rapid decrease in the ATP-DnaA fraction $f(t)$ whenever an origin initiates replication and makes it less likely that replication is initiated at the remaining origins as well (Fig. 5.5b). The site *datA* (at the genomemap position of 94.6 min) is however located about one fifths between the origin (84.6 min) and the terminus (36 min) [42] (Fig. 5.5a). Given that it takes about 40 minutes to replicate the entire chromosome [20, 40], the time from initiation until replication of the *datA* site τ_{datA} takes about 8 minutes. Placing the site *datA* at the experimentally observed chromosomal position leads however to over-initiation (Fig. 5.5b). After the first origin has been initiated, the active fraction first continues to rise. During this time, the remaining origins can initiate replication. When the first *datA* site is replicated after about 8 minutes, deactivation becomes stronger and thus the ATP-DnaA fraction begins to decrease. However, contrary to

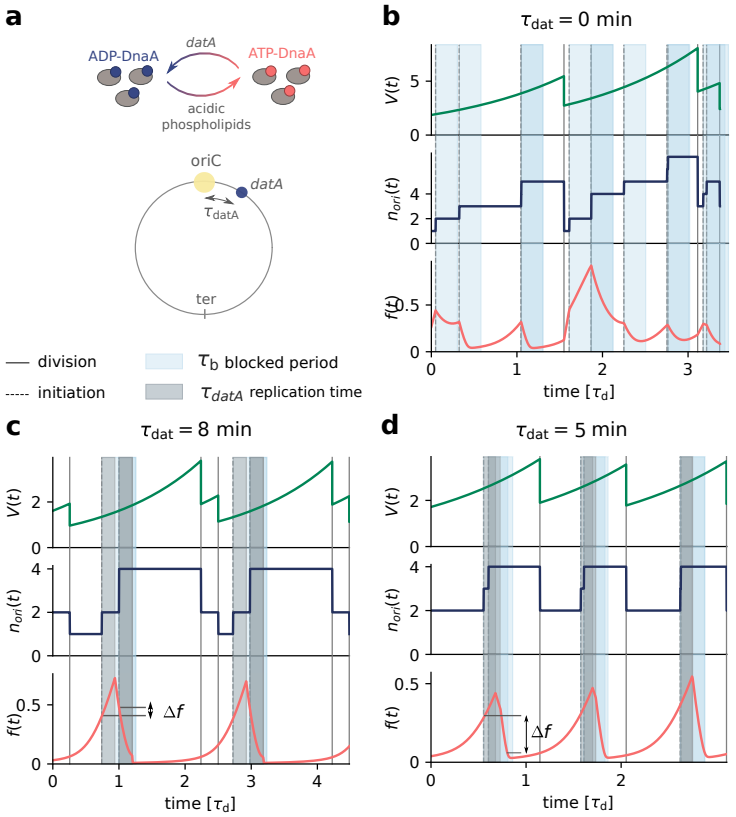


Figure 5.5: Whether replication is initiated synchronously in the LD model depends on the position of the deactivation site *datA* on the chromosome. (a) In the Lipid-*DatA*-model (LD model), the initiator protein DnaA can switch between ADP-DnaA and ATP-DnaA, but only ATP-DnaA can initiate replication at the origin. The protein DnaA can be deactivated via the deactivation site *datA* which is replicated a time τ_{datA} after the origin and activated by acidic phospholipids in the cell membrane. (b-d) The volume $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$ and the ATP-DnaA fraction $f(t)$ as a function of time (in units of the doubling time of the cell τ_{d}). The blocking period τ_{b} is marked as a light blue shaded area and the replication time of *datA* is marked as a grey shaded area. The difference of the ATP-DnaA fraction $\Delta f = f(t_{i,1}+\tau_{\mathrm{b}}) - f(t_{i,1})$ from the first initiation event at $t_{i,1}$ to the end of the blocking period at $t_{i,1}+\tau_{\mathrm{b}}$ is marked by two horizontal lines for the first cell cycle. The (de)activation rates are sufficiently high compared to the growth rate of the cell ($\beta_{\mathrm{datA}} = 6000\,\mathrm{h}^{-1}$, $\alpha_{\mathrm{l}} = 5000\,\mathrm{h}^{-1}$) to ensure large amplitude oscillations and the Michaelis Menten constants are low compared to the total DnaA concentration ($K_{\mathrm{D}} = 5\,\mu\mathrm{m}^{-3}$) such that the system is far in the ultra-sensitivity regime. (b) If the site *datA* is replicated instantaneously with the origin, replication is rarely initiated synchronously at several origins. (c) At the experimentally measured replication time of *datA* of $\tau_{\mathrm{datA}} = 8$ min, replication is over-initiated because at the end of the blocking period the active fraction f is still higher than at the moment of the first replication initiation ($\Delta f < 0$). (d) When the replication time of *datA* is shorter ($\tau_{\mathrm{datA}} = 5$ min) the ATP-DnaA fraction has sufficient time to decrease after the doubling of *datA* and at the end of the blocking period, the active fraction f is much lower than at the first initiation event ($\Delta f > 0$). (See Table 5.1 for all parameters.)

the coarse-grained model, where the initiation potential decreases instantaneously after the end of the licensing period, the finite (de)activation rates in the LD model lead to a slower decrease of the active fraction. At the end of the blocking period of $\tau_b = 10$ min, the active fraction is therefore still higher than it was at the moment where the first origin has fired (see Δf in Fig. 5.5b). Effectively, the licensing period is longer than the blocking period. Therefore, a new round of replication starts, thus leading to over-initiation. When the *datA* site is however placed at an optimal position relative to the origin such that it is replicated 5 minutes after the origin, replication is initiated synchronously (Fig. 5.5c). The active fraction begins to decrease early enough such that now, at the end of the blocking period it is much lower than at the moment of initiation of the first origin (see Δf in Fig. 5.5c). At the same time, the active fraction remains high long enough such that all origins can initiate replication.

While a synchronization parameter cannot be defined uniquely, we will define one to quantify the degree to which replication is initiated synchronously, and then show that the result is fairly robust to the precise definition. Specifically, the degree of synchrony is obtained by counting the number of initiation events per initiation cascade, where the duration of the cascade τ_{cas} is a parameter that we will choose carefully (Fig. 5.3a). In the coarse-grained model, an initiation cascade starts when the first origin initiates and ends after the licensing period is over. As after the end of the licensing period the initiation potential drops instantaneously to a very low value, re-initiation events after the end of the licensing period are very unlikely in the coarse-grained model. In the LD model, the active fraction f does however not decrease instantaneously after the site *datA* has been doubled (Fig. 5.5c and d). Therefore, it is less clear what the duration of a cascade event should be. We test the effect of varying the cascade duration τ_{cas} on the average degree of synchrony $\langle s \rangle$ for different replication times of *datA* τ_{datA} (Fig. 5.6). The average degree of synchrony $\langle s \rangle$ varies strongly with the cascade duration in parameter regimes where replication is initiated asynchronously: At very low replication times of *datA* τ_{datA} replication is under-initiated (see also Fig. 5.5b), but the degree of synchrony nevertheless becomes larger than one at high cascade durations $\tau_{cas} > 0.6\,\tau_d$ (Fig. 5.6). The larger the cascade duration τ_{cas} the more initiation events are counted per cascade event, leading to an average degree of synchrony that is larger than one. Conversely, when the site *datA* is located too far away from the origin and replication is over-initiated (Fig. 5.5c), the degree of synchrony can nevertheless be smaller than one, if the cascade duration is chosen too short. At a replication time of *datA* of $\tau_{datA} = 0.08\,\text{h} \approx 5$ min, where replication is initiated synchronously, the choice of the cascade duration becomes however less relevant: Because all initiation events happen within a relatively small time window, increasing the cascade duration further does not change the degree of synchrony significantly. Only if the cascade duration is chosen too small ($\tau_{cas} = 0.1\,\tau_d \approx 4$ min) becomes the average degree of synchronization smaller than one. We therefore in the following choose an intermediate cascade duration of $\tau_{cas} = 0.4\,\tau_d \approx 16$ min.

Replication is only initiated synchronously in the LD model when the system is far in the ultra-sensitivity regime. As we have seen in the coarse-grained model, the initiation potential must rise sharply at a critical volume per origin in order to initiate replication at all origins within a short time window. In the LD model, the sharpness of the rise

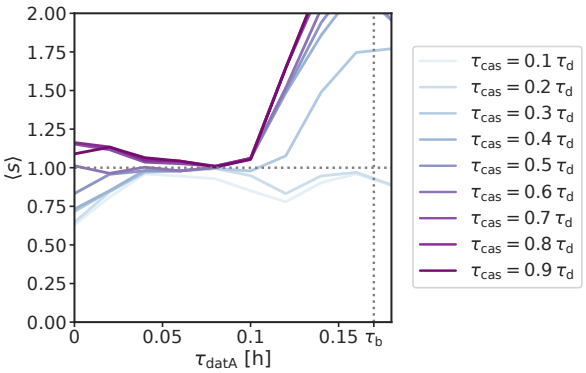


Figure 5.6: The average degree of synchrony $\langle s \rangle$ depends strongly on the cascade duration when replication is initiated asynchronously, but not when replication is initiated synchronously. The average degree of synchrony $\langle s \rangle$ as a function of the replication time of the site *datA* τ_{datA} for varying cascade durations τ_{cas} (in units of the doubling time τ_{d}). The degree of synchrony s is obtained by counting the number of initiation events from the first origin firing until the end of the cascade durations τ_{cas}. When replication is initiated synchronously at all origins within a short time interval (at $\tau_{\mathrm{datA}} \approx 0.08$ h), the average degree of synchrony does not depend strongly on the maximal cascade event duration τ_{cas}. When origins are however initiated asynchronously over the course of the cell cycle, the average degree of synchrony can either be smaller or larger than one, depending on the duration τ_{cas}. In the rest of this manuscript, we use an intermediate cascade duration of $\tau_{\mathrm{cas}} = 0.4\,\tau_{\mathrm{d}}$. (See Table 5.1 for all parameters.)

5

of the active fraction at a critical volume can be tuned by varying the degree of ultra-sensitivity. When the difference in the dissociation constants of the (de)activators K_{D} with respect to the total DnaA concentration $[D]_{\mathrm{T}}$ is high, the system is deep in the ultra-sensitivity regime and replication is initiated synchronously (Fig. 5.6d). For lower dissociation constants of the (de)activators K_{D}, not all origins initiate replication during an initiation cascade and replication is initiated asynchronously (Fig. 5.7b). Also increasing the replication time of *datA* τ_{datA} cannot recover synchronous replication initiation, because when τ_{datA} becomes too long, replication is over-initiated (Fig. 5.7a). It seems however reasonable to assume that the switch is highly ultra-sensitive, given that the binding of DnaA to the site *datA* was reported to be very strong [61, 89, 122].

For synchronous replication initiation in the LD model, the (de)activation rates must be higher than observed experimentally. To obtain large amplitude oscillations in the active fraction f, we have so far chosen relatively high (de)activation rates. When the (de)activation rates are however lower and closer to the experimentally measured rates [63], replication is over-initiated independent of the position of the site *datA* on the chromosome (Fig. 5.7a and b). Due to the low rates, the active fraction f decreases slowly after replication initiation, such that the active fraction is still high after the end of the blocking period τ_{b} and more origins initiate replication (Fig. 5.7b). The longer the delay in replicating the site *datA* after the origin, the higher the degree of over-initiation (Fig. 5.7a).

As we have shown in section 3.2, the deterministic LDDR model that includes all

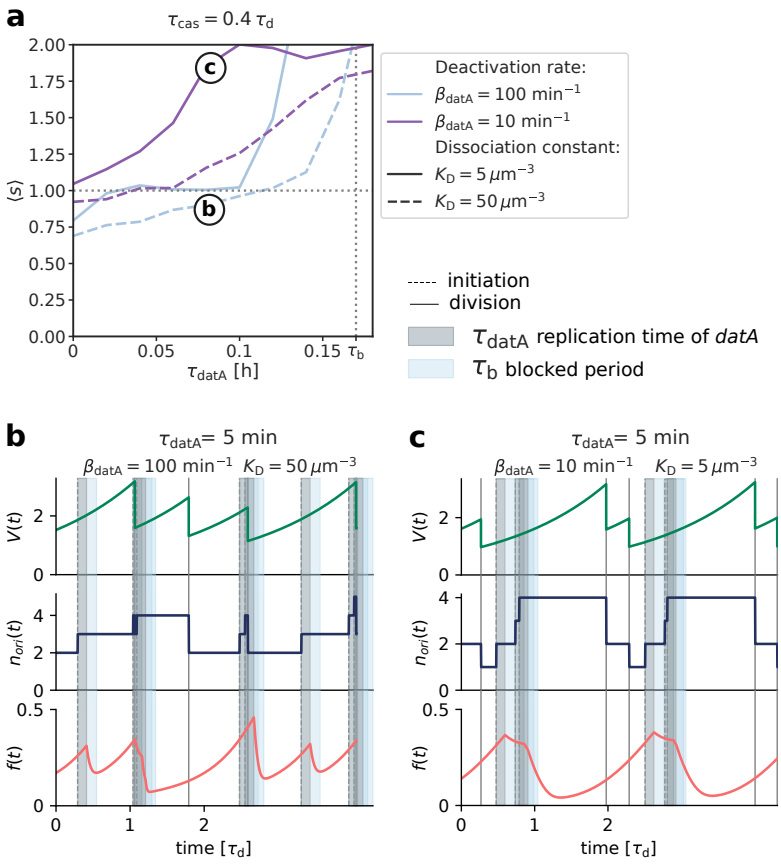


Figure 5.7: Synchronous replication initiation in the LD model only for high ultra-sensitivity and relatively high (de)activation rates. (a) The average degree of synchrony $\langle s \rangle$ as a function of the replication time of the deactivation site τ_{datA} for different magnitudes of the (de)activation rates ($\alpha_l = 0.83 \times \beta_{datA}$) and for different dissociation constants $K_D^l = K_D^{datA} = K_D$. The duration of the blocking period τ_b is indicated by the vertical, dotted line and perfect synchronization at $\langle s \rangle = 1$ is marked by the horizontal, dotted line. (b, c) The volume $V(t)$, the number of origins $n_{ori}(t)$ and the ATP-DnaA fraction $f(t)$ as a function of time (in units of the doubling time of the cell τ_d). (b) When the system is less ultra-sensitive than in Fig. 5.6, not all origins are initiating before the end of the licensing period τ_{datA}, resulting in under-initiation. (c) At too low (de)activation rates ($\beta = 10\,min^{-1}$), the active fraction does not decrease fast enough to prevent re-initiation events after the end of the blocking period, resulting in over-initiation. (See Table 5.1 for all parameters.)

known (de)activators of DnaA in *E. coli* does give sharp oscillations in the active fraction at realistic (de)activation rates also at high growth rate. In this growth regime, replication forks overlap and Regulatory Inactivation of DnaA (RIDA) is the main deactivator in combination with the two chromosomal regions called DnaA-Reactivation Sequence 1 (*DARS1*) and *DARS2* (Fig. 5.8a). We therefore test whether the LDDR model can ensure sufficiently large amplitude oscillations to drive synchronous replication initiation.

The LDDR model can indeed give rise to synchronous replication initiation at multiple origins, but only for a small range of parameters: When the (de)activators *DARS1*, *DARS2* and *datA* are located at the experimentally measured positions on the chromosome, replication is initiated asynchronously when RIDA starts directly after an origin has fired (Fig. 5.8b at $\tau_{\mathrm{rida}} = 0$ h). As RIDA is a strong deactivator, it causes the active fraction to drop rapidly after the first origin has been initiated and thus prevents other origins from firing as well. By varying both the position of *datA* on the chromosome and the time at which RIDA starts after an origin is initiated, we find that replication can be initiated synchronously in the LDDR model for a small range of parameters: At the experimentally measured replication time of *datA* of $\tau_{\mathrm{datA}} = 0.13\,\mathrm{h} \approx 8$ min, replication is initiated with a high degree of synchrony when the deactivation rate of RIDA becomes high with a delay of $\tau_{\mathrm{rida}} = 0.1$ h=6 min after the origin is fired (Fig. 5.8b and c). The closer the site *datA* is to the origin, the later RIDA should start for synchronous replication initiation (Fig. 5.8a).

It remains however unclear what molecular mechanism could cause a delay in the onset of RIDA of about 6 minutes. In RIDA, the DNA polymerase clamp on newly synthesized DNA forms a complex with ADP and the Hda protein. The resultant ADP-Hda-clamp-DNA can bind ATP-DnaA and stimulates ATP hydrolysis yielding ADP-DnaA. It is conceivable that Hda binding is slow, but whether it would yield a delay of about 6 minutes is far from clear. For experimentally realistic parameters, the LDDR model appears therefore not sufficient to explain how replication is initiated synchronously in *E. coli*.

In *E. coli*, DNA replication initiation is not only controlled via an activation switch, but also via titration [12, 14]. To study the effect of titration on the degree of synchrony, we add homogeneously distributed titration sites on the chromosome to the LDDR model (see chapter 4 for details). In the LDDR-titration model, the initiation potential is given by the free ATP-DnaA concentration $[D]_{\mathrm{ATP,f}}$ in the cell and both oscillations in the active fraction f and in the free DnaA concentration $[D]_{\mathrm{T,f}}$ contribute to regulating replication initiation. We again model the stochastic opening probability of the origin as a Hill function (equation 5.2) with Hill coefficient $m = 10$. The critical initiation potential y^* is now given by a critical free ATP-DnaA concentration $[D]^*_{\mathrm{ATP,f}}$ at which ATP-DnaA binds cooperatively to the origin. We here neglect the effect of the relatively small number of about 10-20 DnaA proteins that are bound to the origin on the free DnaA concentration. As explained in section 4.4, we set the parameters (by varying the lipid activation rate α_{l}) such that the initiation volume of the switch v^*_{s} and the initiation volume of the titration mechanism v^*_{t} are approximately the same. This optimal choice ensures that both the free concentration and the active fraction rise at the same critical volume per origin, thus increasing the amplitude of the oscillations in the free ATP-DnaA concentration.

Fig. 5.9a/c show the time traces of the model that combines titration with the activation switch. The small jump in the total free DnaA concentration upon cell division

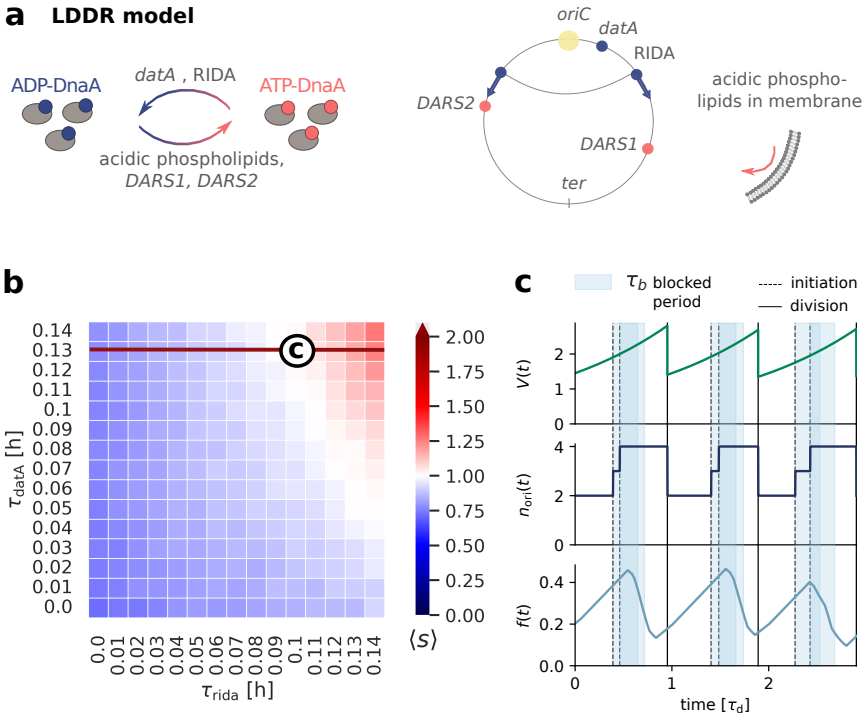


Figure 5.8: The LDDR model can ensure a high degree of synchronous replication initiation for a narrow range of parameters. (a) In the Lipid-*DatA*-*DARS1/2*-RIDA (LDDR) model, replication forks overlap and RIDA is the main deactivator in combination with the activators *DARS1* and *DARS2*. (b) The average degree of synchronization $\langle s \rangle$ as a function of the replication time of the site *datA* τ_{datA} and the onset time of RIDA τ_{rida}. The sites *DARS1* and *DARS2* are replicated at the experimentally measured times $\tau_{\mathrm{d1}} = 0.25\,\mathrm{h} = 15$ min and $\tau_{\mathrm{d1}} = 0.4\,\mathrm{h} = 24$ min, respectively. Replication is only initiated synchronously for a small range of parameters: When the site *datA* is replicated after the experimentally measured time of $\tau_{\mathrm{datA}} = 0.13\,\mathrm{h} \approx 8$ min (red horizontal line), replication in the LDDR model is only initiated synchronously if RIDA starts only about 6 minutes after the origin is initiated. It is however not clear what could cause a delay of 6 minutes in the onset of RIDA. (c) The volume $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$ and the ATP-DnaA fraction $f(t)$ as a function of time (in units of the doubling time of the cell τ_{d}) for the parameter combination marked in b. The large amplitude oscillations in the active fraction in combination with a long delay in the onset of deactivation via RIDA and *datA* can give rise to a high degree of synchrony for a small range of parameters. For each parameter set in b, the average degree of synchronization was obtained from $N = 1000$ consecutive cell cycles. (See Table 5.1 for all parameters.)

results from the fact that only one out of two chromosomes is selected per daughter cell (Fig. 5.9a/c, second panel). The stochastic firing of the origins causes a temporal delay between the initiation of replication at the respective origins. Moreover, in the growth-rate regime of overlapping replication forks considered here, not all chromosomes have been fully replicated at the moment of cell division. Taken together, this means that at the moment of cell division not all chromosomes have the same number of titration sites (the sites are distributed uniformly). The difference in the number of titration sites per chromosome causes a slight change in the free concentration upon cell division.

Adding titration sites to the LDDR model affects the degree of synchrony only little when the critical free ATP-DnaA concentration at which replication is initiated is high. When a new round of replication is initiated, new titration sites are generated and the free DnaA concentration drops. As discussed in section 2.3, at high growth rates, where multiple chromosomes are present in the cell, new titration sites are however replicated at a similar rate as new DnaA proteins. Titration therefore introduces only weak oscillations in the free total DnaA concentration (Fig. 5.9a). If the critical free DnaA concentrationat which DNA replication is initiated is relatively high, the oscillations in the free DnaA concentration contribute only little to the oscillations in the initiation potential (Fig. 5.9a). In this scenario, adding titration to the LDDR model does not significantly change the degree of synchrony, the optimal position of *datA* on the chromosome or the optimal onset time of RIDA (compare Fig. 5.9b to Fig. 5.8b).

When the free DnaA concentration $[D]_{\mathrm{T,f}}$ is however low, titration can significantly enhance the degree of synchrony of the LDDR model. Setting the critical free ATP-DnaA $[D]^{*}_{\mathrm{f,ATP}}$ to a value that is comparable to the affinity of the titration sites increases the oscillations in the free DnaA concentration (Fig. 5.9c). The resulting sharper rise of the free ATP-DnaA concentration gives rise to a higher degree of synchrony at all positions of *datA* and onset times of RIDA (Fig. 5.9c). The regime of parameters in which replication is initiated with a high degree of synchrony now extends also to shorter and more realistic onset times of RIDA than in the LDDR model (Fig. 5.9d).

Transiently blocking DnaA synthesis via SeqA after replication initiation leads to significant under-initiation. Blocking DnaA synthesis for about $\tau_{\mathrm{b,s}}$ =10 min after an origin has been initiated leads to a strong decrease of the free DnaA concentration after replication initiation. Even though the ATP-DnaA fraction continues to rise after the first origin has been initiated, the strong decrease in the free concentration prevents the remaining origins from firing. This leads to under-initiation for a broad range of experimentally realistic parameters (Fig. 5.10a). We therefore predict that the effect of SeqA to block DnaA synthesis must be rather weak at high growth rates. Indeed, experiments indicate that the duration of DnaA synthesis blocking via SeqA depends on the growth rates: at high growth rates (λ = 1.6 h^{-1}), SeqA only binds to the DnaA promoter for about 4 minutes [58]. Importantly, for synchronous replication initiation the effect of blocking DnaA synthesis needs to be delayed in order to allow the remaining origins to fire. This delay could be arising from the fact that the *dnaA* gene is replicated only about one minute after the origin. Furthermore, a delay in the effect of blocking DnaA synthesis could arise from the finite time it takes to transcribe the *dnaA* gene and translate the mRNA into a functioning protein [65, 136]. Including a shorter blocked synthesis period of $\tau_{\mathrm{b,s}}$ = 3.6 min with a delay of $\tau^{+}_{\mathrm{b,s}}$ = 6.6 min indeed alleviates the effect of blocking DnaA synthesis and

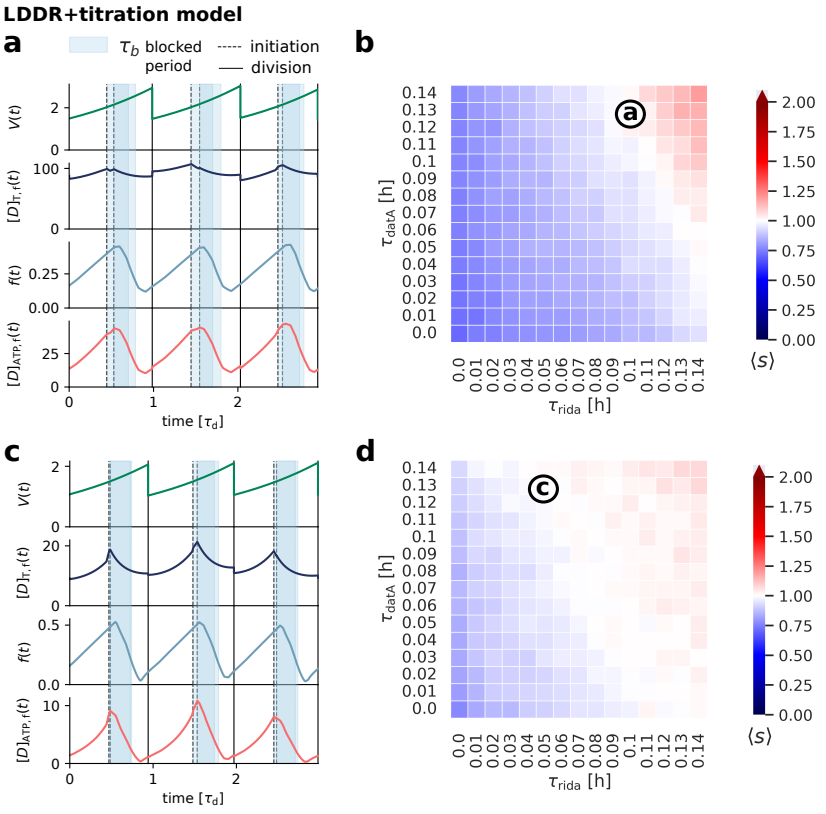


Figure 5.9: Adding titration sites to the LDDR model enhances initiation synchrony for low critical free DnaA concentrations. (a, c) The volume $V(t)$, free DnaA concentration (independent of whether DnaA is bound to ATP or ADP) $[D]_{T,f}(t)$, the ATP-DnaA fraction $f(t)$, and the free ATP-DnaA concentration $[D]_{ATP,f}(t)$ as a function of time (in units of the doubling time of the cell $\tau_d = 0.67$ h=40 min) for a critical free ATP-DnaA concentration of $[D]^*_{f,ATP} = 50\,\mu m^{-3}$ (a) and $[D]^*_{f,ATP} = 10\,\mu m^{-3}$ (c). During the blocking period τ_b (light blue shaded area), the newly replicated origins cannot be re-initiated. (a) When the critical free ATP-DnaA concentration is relatively high, the free DnaA concentration $[D]_{T,f}(t)$ oscillates only weakly and decreases slightly after new rounds of replication are initiated due to the synthesis of new sites. The shape of the oscillations in the free ATP-DnaA concentration $[D]_{ATP,f}(t)$ is therefore mainly determined by the oscillations in the ATP-DnaA fraction $f(t)$. (b, d) The average degree of synchronization $\langle s \rangle$ as a function of the replication time of the site *datA* τ_{datA} and the onset time of RIDA τ_{rida} for $[D]^*_{f,ATP} = 50\,\mu m^{-3}$ (b) and $[D]^*_{f,ATP} = 10\,\mu m^{-3}$ (d). The sites *DARS1* and *DARS2* are replicated at the experimentally measured times $\tau_{d1} = 0.25$ h = 15 min and $\tau_{d1} = 0.4$ h = 24 min, respectively. (b) When the critical free ATP-DnaA concentration is high, the effect of the titration sites on the degree of synchrony is small and almost undistinguishable from the scenario without titration sites (compare to Fig. 5.8b). (c) At a lower critical free ATP-DnaA concentration, the oscillations in the free concentration are larger and lead to sharper oscillations of the free ATP-DnaA concentration. This causes a broader range of parameters for which replication is initiated synchronously (d). For each parameter pair in b and d, the average degree of synchronization was obtained from $N = 1000$ consecutive cell cycles. (See Table 5.1 for all parameters.)

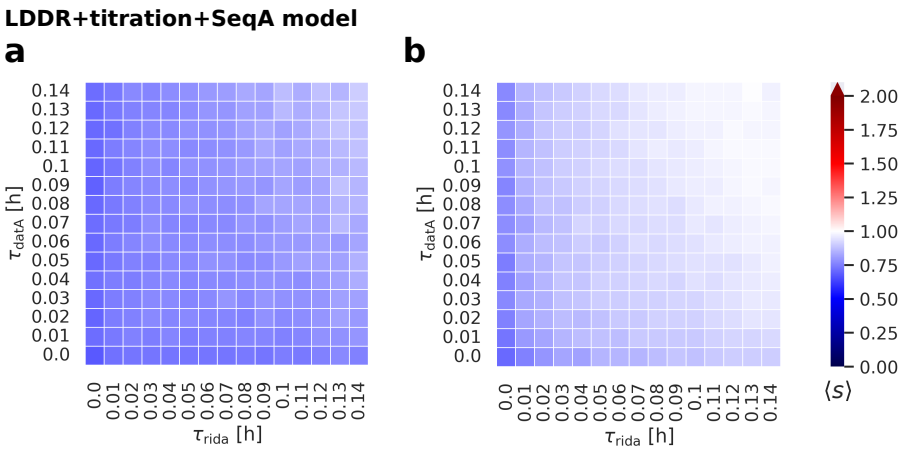


Figure 5.10: Transiently blocking DnaA synthesis via SeqA reduces the degree of synchrony in the LDDR model (a, b) The average degree of synchronization $\langle s \rangle$ as a function of the replication time of the site *datA* τ_{datA} and the onset time of RIDA τ_{rida}. The sites *DARS1* and *DARS2* are replicated at the experimentally measured times $\tau_{\mathrm{d1}} = 0.25\,\mathrm{h} = 15$ min and $\tau_{\mathrm{d1}} = 0.4\,\mathrm{h} = 24$ min, respectively. (a) Adding the effect of SeqA to transiently block DnaA synthesis for $\tau_{\mathrm{b,s}} = 10$ min right after replication initiation leads to under-initiation over the full range of tested parameters. (b) When the duration of the blocked synthesis period is shorter ($\tau_{\mathrm{b,s}} = 3.6$ min) and delayed by $\tau^{+}_{\mathrm{b,s}} = 6.6$ min, replication synchrony can be partially recovered for long delay times in the onset of RIDA. In all simulations, the origin is prevented for a blocking period of $\tau_{\mathrm{b}} = 10$ min immediately after replication initiation. To obtain the experimentally observed total DnaA concentration, the gene allocation fraction was set slightly higher than in the other Figures to a value of $\phi_0 = 6 \times 10^{-4}$ to compensate for the duration of blocked DnaA synthesis. For each parameter set in a and b, the average degree of synchronization was obtained from $N = 1000$ consecutive cell cycles. (See Table 5.1 for all parameters.)

recovers initiation synchrony for relatively high delay times in the onset of RIDA (Fig. 5.10b).

To conclude, we found that the LD model can ensure synchronous replication initiation at an optimal position of the deactivator *datA* on the chromosome, but only for higher (de)activation rates than observed experimentally. The LDDR model can ensure a high degree of synchronization also at low and biologically more realistic (de)activation rates, but only for a small range of parameters and a long delay in the start of RIDA after replication initiation. Slow binding of the protein hda to form the RIDA complex might lead to a delay in the onset of RIDA. It is however unclear how a delay of approximately 6 minutes could be achieved, given the strong binding of Hda to the replication clamps that form right after a new round of replication starts [88]. Including titration sites at a low critical free ATP-DnaA concentration can enhance initiation synchrony and extend the parameter regime in which replication is initiated synchronously to more realistic parameters. Transiently blocking DnaA synthesis right after replication initiation for about ten minutes on the other hand reduces the degree of synchronization for all tested parameters. We therefore predict that the effect of blocking DnaA synthesis should be

weaker at high growth rates and delayed in order to nevertheless initiate synchronously for reasonable parameters.

5.4. DISCUSSION

The bacterium *E. coli* initiates replication at several origins synchronously with high precision. How it achieves this high degree of synchrony remained however unknown. In this work, we have revealed several general principles that govern whether replication is initiated synchronously at several origins: (1) the initiation potential must remain high after the first origin has fired, so that the remaining origins can fire; (2) origins that have already fired must be prevented from reinitiating immediately as long as the initiation potential remains high; and (3) the initiation potential must come down before the blocking period is over to prevent reinitiation of (the first) newly replicated origins. The licensing period, during which the origins can fire, must thus be shorter than the blocking period. To ensure that all origins fire during the licensing period, the initiation potential must rise sharply, and to guarantee that the initiation potential is low again before the blocking period is over, it also must fall sharply. Synchronous replication initiation thus requires sharp oscillations in the initiation potential. Such oscillations will also give rise to small variations in the initiation volume. Our results therefore predict that the experimentally observed small variations in the initiation volume is a result of the robust replication initiation that is necessary for synchronous replication initiation.

We showed that the previously presented model for the regulation of replication initiation in the bacterium *E. coli* can ensure a high degree of initiation synchrony for a range of parameters that agree with the experimentally measured ones. We find that if replication initiation is governed by a protein activation switch only, the optimal onset time of the RIDA mechanism would have to be about 6 min in order to ensure synchronous replication initiation. As RIDA is coupled to active replication [105], protein diffusion in cells is typically in the order of seconds rather than minutes [55] and binding of HdA to the replication clamps is rather strong [87, 88], it seems natural to assume that the deactivation via RIDA becomes strong directly after a new round of replication starts. It is however conceivable that HdA concentration rises slowly, that HdA binding is slow or that several RIDA complexes are required for a strong deactivation rate of RIDA. Adding titration sites to the activation switch and bringing the system to a regime where the free DnaA concentration is low during the entire cell cycle enhances the degree of synchronization significantly for a broad range of parameters. Importantly, in the combined model replication is initiated with a high degree of synchrony also for shorter onset times of RIDA. Combining protein activation with titration could therefore be a key ingredient to initiate replication synchronously in *E. coli*.

So far, we have not modelled the binding of about 10-20 ATP-DnaA proteins to the origin explicitly. It has however been proposed in the so-called 'initiation cascade model' that initiating replication at the first origin could cause other origins to fire as well by releasing the bound initiator proteins into the cytoplasm. The resulting higher concentration of free DnaA proteins could lead to a redistribution of the free DnaA proteins to the remaining origins, making the next replication initiation event more likely [137]. We tested this idea by introducing weak, cooperative origin binding sites to which only

ATP-DnaA can bind into our model. When in this extended model the concentration of ATP-DnaA in the cytoplasm rises, the weak binding sites at the respective origins begin to fill up and then trigger the initiation of replication at a randomly selected origin (see Appendix 5.F). After replication has been initiated, the binding sites at the origin that fired become unavailable for binding DnaA for the duration of the blocking period, causing a rise in the free DnaA concentration. We find, however, that the ATP-DnaA binding to the origin has two opposing effects: on the one hand, the initiation potential indeed increases right after the first initiation event due to the released ATP-DnaA proteins, making the next initiation event more likely (see Fig. 5.12a and b). On the other hand, binding of ATP-DnaA proteins to the origin leads to a less sharp rise in the free DnaA concentration right before the first origin initiates replication (see Fig. 5.12a and b). A sharp rise of the initiation potential right before replication initiation is however a necessary requirement for synchronous replication initiation. Therefore, the net effect of the initiation cascade on the degree of synchrony is approximately zero and we do not find a significant increase in the degree of synchrony (see Fig. 5.12c).

An important question is how cells can recover from a state with an uneven number of origins. Interestingly, cell division enhances recovery both from over- and under-initiation events. Let us consider the scenario where after an initiation cascade, a cell has three instead of the desired four origins of replication (Fig. 5.16a): At cell division, one daughter cell will obtain the chromosome with an ongoing round of replication (two origins) and the other daughter cell will obtain a single chromosome that is not yet replicating. As the volume $V(t)$ is divided by two upon cell division but only one out of three origins ends up in the latter daughter, the volume per origin $v(t)$ of this daughter cell increases instantaneously to a higher value (Fig. 5.16a, third panel). The volume per origin is now higher than the critical volume per origin v^* at which the active fraction rises sharply, causing the active fraction $f(t)$ to increase strongly. A new round of replication initiation is therefore triggered shortly after cell division, leading to the correct number of origins in this cell. The active fraction first decreases and when it rises again in the same cell cycle, both origins now initiate synchronously. The cell has thus recovered from an asynchronous initiation event. Similarly, a cell with too many origins at birth has a lower volume per origin $v(t)$ and the active fraction thus remains low for a longer time (Fig. 5.16b, third panel). The daughter cell divides before a new round of replication is initiated. This again leads to a sharp rise of the active fraction and the cell recovers via the same mechanism as described in the under-initiation scenario. This leads to an interesting prediction of our model: Cell division should be correlated with triggering new rounds of replication in cells where replication was initiated asynchronously.

APPENDIX

5.A. COARSE-GRAINED MODEL FOR ORIGIN OPENING

We describe the origin region as a two-state system that can switch between an open (O) or a closed (C) configuration with the opening rate k_o and the closing rate k_c. If the origin is open, replication can be initiated (I) with a maximal firing rate k_f^0:

$$C \underset{k_c}{\overset{k_o}{\rightleftharpoons}} O \xrightarrow{k_f^0} I \tag{5.8}$$

In thermal equilibrium, the ratio of the transition rates between the open and closed state is given by the Boltzman distribution of the energy difference of the two states:

$$\frac{k_c}{k_o} = \frac{e^{-\beta E_c}}{e^{-\beta E_o}} = e^{\beta \Delta G} \tag{5.9}$$

with $\beta = k_B T$ and the energy difference

$$\Delta G = E_o - E_c \tag{5.10}$$

where E_o is the energy of the open state and E_c is the energy of the closed state. The probability to be in the open state as a function of the energy difference ΔG is given by

$$p_o = \frac{e^{-\beta E_o}}{e^{-\beta E_o} + e^{-\beta E_c}} = \frac{1}{1 + e^{\beta \Delta G}} \tag{5.11}$$

Assuming rapid opening and closing dynamics of the origin, the origin firing rate is given by equation 5.3. The higher the initiation potential f in the cell, the more likely is it that the origin is open and that replication can be initiated. We model this observation phenomenologically by assuming that the opening probability p_o increases with the activation potential f following a Hill function (see equation 5.2).

5.B. DERIVATION OF APPROXIMATION FOR OPENING PROBABILITY

We want to find an analytical expression for the opening probability p_o and therefore also the instantaneous firing rate k_f (equation 5.3) as a function of time. We therefore insert equation 5.1 into equation 5.2 to obtain:

$$p_o(f(v)) = \frac{v^{nm}}{f^{*m}(v^{*n} + v^n)^m + v^{nm}} \tag{5.12}$$

$$= \frac{v^{nm}}{f^{*m} v^{*nm}(1 + \tilde{v}^n)^m + v^{nm}}, \tag{5.13}$$

Parameter	name	value	Motivation
n	Hill coefficient of initiation potential	5	set to match initiation precision reported in [76]
v^* [μm^3]	initiation volume per origin	1	set to match initiation volume reported in [76]
m	Hill coefficient of opening probability	10	set to match initiation precision reported in [76]
y^*	critical initiation potential	0.5	set to maximal sharpness of opening probability
K_D [μm^{-3}]	dissociation constant of (de)activators	5	[46]
$\alpha_l[l]$ [μm^{-3} h^{-1}]	activation rate lipids	LD:5000	set to match initiation volume reported in [76]
		LDDR:500	
		LDDR+titration:800	
β_{datA} [h^{-1}]	deactivation rate *datA*	LD:6000	
		LDDR:600	[63]
τ_{datA} [h]	replication time *datA*	0.13	[63]
f^*	critical initiator fraction	0.5	[78, 82]
τ_{cas} [h]	initiation cascade duration	0.27	see Fig. 5.6
α_{d1} [h^{-1}]	activation rate *DARS1*	100	[43, 77]
τ_{d1} [h]	replication time *DARS1*	0.4	[43]
α^+_{d2} [h^{-1}]	high activation rate *DARS2*	600	combined with β_{rida}
α^-_{d2} [h^{-1}]	low activation rate *DARS2*	50	set to arbitrary low value
τ_{d2} [h]	replication time *DARS2*	0.25	[77]
τ^+_{d2} [h]	start high activation rate *DARS2*	0.2	[77]
τ^-_{d2} [h]	end high activation rate *DARS2*	2/3	[77]
β_{rida} [h^{-1}]	deactivation rate RIDA	500	[63, 88, 110]
$[D]_T$ [μm^{-3}]	total DnaA concentration	400	[18, 64]
ϕ_0	gene allocation fraction	4×10^{-4}	set to match $[D]_T$
K^s_D [μm^{-3}]	dissociation constant of titration sites	1	[46]
n^s_{ori}	number of origin binding sites	10	[43]
$[D]^*_{ATP,f}$	critical free ATP-DnaA concentration	10	[46]
ρ [μm^{-3}]	number density	10^6	[75]
k^0_f [s^{-1}]	maximal origin firing rate	1000	
τ_b [h]	blocking period	0.17	[58–60]
λ [h^{-1}]	growth rate	1.04	[22, 40]
T_C [h]	C-period	2/3	[20]
T_D [h]	D-period	1/3	[20]

Table 5.1: Parameters used in the simulations One molecule per cubic micrometer corresponds to approximately one nM (1 μm^{-3} = 1.67 nM).

where we used $\tilde{v} = v/v^*$. According to the binomial formula, we can write

$$(1+\tilde{v}^n)^m = \sum_{k=0}^{m} \binom{m}{k} 1^k (\tilde{v}^n)^{m-k} \tag{5.14}$$

$$= \sum_{k=0}^{m} \binom{m}{k} (\tilde{v}^n)^{m-k}. \tag{5.15}$$

with the binomical coefficient

$$\binom{m}{k} := \frac{m!}{k!(m-k)!}. \tag{5.16}$$

We introduce the shifted parameter $k' = k - m/2$, such that equation 5.15 can be rewritten as:

$$(1+\tilde{v}^n)^m = \sum_{k'=-m/2}^{m/2} \binom{m}{k'+\frac{m}{2}} (\tilde{v}^n)^{\frac{m}{2}-k'}. \tag{5.17}$$

By examining the first and the second term of the sum in equation 5.17 separately, we find that the binomial coefficient has a maximum at $k' = 0$ and decays quickly for $k' \neq 0$ (See Fig. 5.11a). Secondly, as can be seen in Figure 5.11b, for small $k' \ll \pm m/2$ and sufficiently large Hill coefficient m, the second term is approximately given by

$$\tilde{v}^{n(\frac{m}{2}-k')} \approx \tilde{v}^{\frac{mn}{2}}. \tag{5.18}$$

Combining these two observations, we can approximate equation 5.17 by

$$(1+\tilde{v}^n)^m \approx \sum_{k'=-m/2}^{m/2} \binom{m}{k'+\frac{m}{2}} \tilde{v}^{\frac{mn}{2}}. \tag{5.19}$$

Finally, using that

$$\sum_{k'=-m/2}^{m/2} \binom{m}{k'+\frac{m}{2}} = 2^m, \tag{5.20}$$

we find

$$(1+\tilde{v}^n)^m \approx 2^m \tilde{v}^{\frac{mn}{2}} \tag{5.21}$$

Plugging this expression into equation 5.13 gives

$$p_o(v) \approx \frac{v^{nm}}{f^{*m} v^{*nm} 2^m \tilde{v}^{\frac{mn}{2}} + v^{nm}} \tag{5.22}$$

$$= \frac{v^{nm}}{f^{*m} 2^m v^{*\frac{nm}{2}} v^{\frac{mn}{2}} + v^{nm}} \tag{5.23}$$

$$= \frac{v^{\frac{nm}{2}}}{f^{*m} 2^m v^{*\frac{nm}{2}} + v^{\frac{nm}{2}}} \tag{5.24}$$

For $f^* = 0.5$ we then find equation 5.5 of the main text with the effective Hill coefficient $n_{\text{eff}} = n\,m/2$. By comparing the approximation of $p_o(v)$ in equation 5.5 to a function

$$p_o^{\text{fit}}(v) = a^{\text{fit}} \frac{v^{n_{\text{eff}}^{\text{fit}}}}{v^{*n_{\text{eff}}^{\text{fit}}} + v^{n_{\text{eff}}^{\text{fit}}}} \tag{5.25}$$

5

that was fitted to $p_o(f(v))$ (equation 5.13), we find that the approximation in equation 5.5 is indeed a good approximation for sufficiently large Hill coefficients n and m, especially at volume close to the critical volume v^* (Fig. 5.11c). Indeed, the fitted Hill coefficient agrees well with the approximated Hill coefficient in equation 5.6 for a broad range of Hill coefficients n and m, respectively (Fig. 5.11d).

5.C. PARAMETER CHOICE FOR MAXIMAL FIRING RATE

Combining the approximation for the opening probability as a function of the volume per origin (equation 5.5), the exponentially growing cell-volume $V(t) = V_b\, e^{\lambda t}$, and the expression for the firing rate (equation 5.3), we find the following time-dependent firing rate of a single origin:

$$k_f(t) = k_f^0 \frac{\left(V_b\, e^{\lambda t}\right)^{n_{eff}}}{v^{*\, n_{eff}} + \left(V_b\, e^{\lambda t}\right)^{n_{eff}}} \tag{5.26}$$

From this rate, we can calculate the survival probability

5

$$S(t) = e^{-\int_{t_0}^{t} dt' k_f(t')} \tag{5.27}$$

$$= e^{-\frac{k_f^0}{n_{eff}\lambda} \ln\left(\frac{(V_b\, e^{\lambda t})^{n_{eff}} + v^{*\, n_{eff}}}{V_b^{n_{eff}} + v^{*\, n_{eff}}}\right)} \tag{5.28}$$

where we solved the integral with the initial condition $S(t_0 = 0) = 1$. We now impose that at the theoretical initiation volume per origin $v(t = t^*) = v^*$, the survival probability is exactly $S(t^*) = 0.5$. Using this constraint, we obtain the following expression for the maximal firing rate as a function of the effective Hill coefficient n_{eff}:

$$k_f^0(n_{eff}) = \frac{n_{eff}\,\lambda \ln(2)}{\ln\left(\frac{2\, v^{*n}}{V_b^n + v^{*n}}\right)} \tag{5.29}$$

This parameter choice ensures that the average initiation volume $\langle v^* \rangle$ is given by v^*.

5.D. DERIVATION OF THEORETICAL PREDICTION FOR DEGREE OF SYNCHRONY

In the following, we derive a theoretical prediction for the probability that two initiation events happen within a time interval τ_l. We assume here that the two firing events are statistically independent, meaning that between the first initiation event at time t_1 and the time $t_1 + \tau_l$, the change in the number of origins induced by the first event has no effect on the firing of the second event. Using the firing rate in equation 5.26, we can now calculate the error probability S_{err} that the second event does *not* happen within a time τ_1 after the first event, given that the first event happened at time t_1:

$$S_{err}(t_2 - t_1 > \tau_1 | t_1) = e^{-\int_{t_1}^{t_1+\tau_1} dt' k_f(t')} \tag{5.30}$$

$$= e^{-\frac{k_f^0}{n_{eff}\lambda} \log\left(\frac{\left(V_b\, e^{\lambda\left(t_1+\tau_1\right)}\right)^{n_{eff}} + v^{*\, n_{eff}}}{\left(V_b\, e^{\lambda t_1}\right)^{n_{eff}} + v^{*\, n_{eff}}}\right)} \tag{5.31}$$

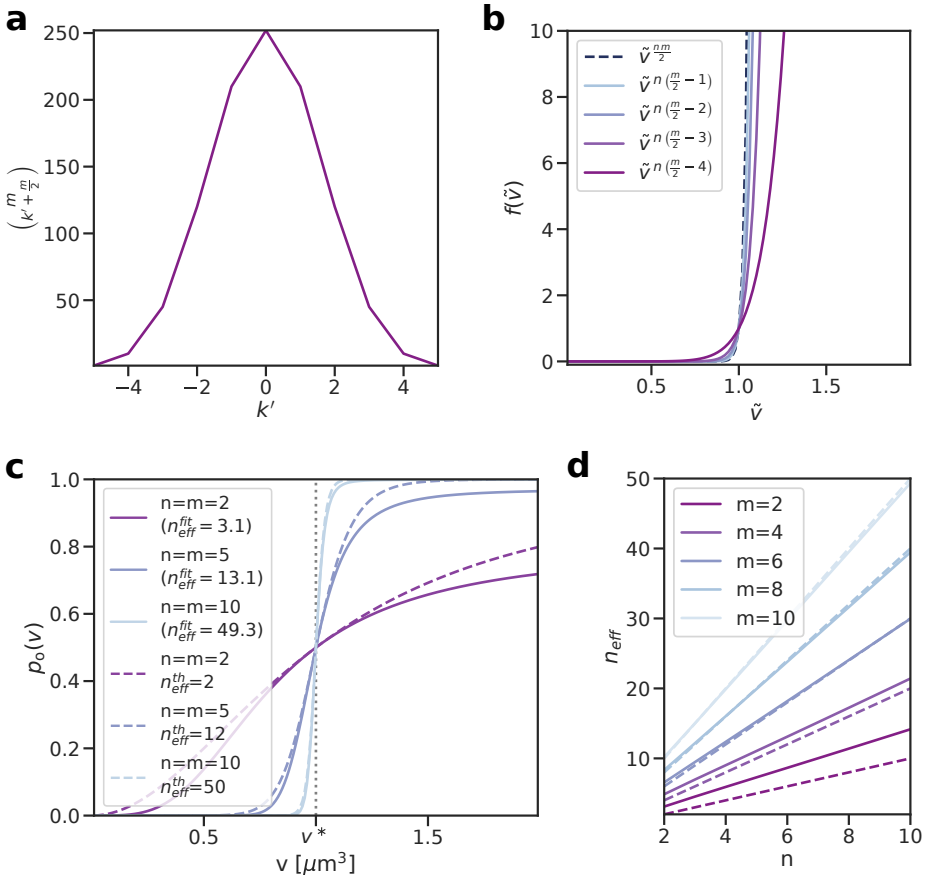


Figure 5.11: The instantaneous opening probability can be approximated by a Hill function with an effective Hill coefficient (a) The binomial coefficient as defined in equation 5.16 as a function of the index k' for $m = 10$. The binomial coefficient is centered around and maximal at $k' = 0$ and becomes small for $k' \gg 0$. (b) The second term in equation 5.17 for different values of the index k' as a function of the rescaled volume $\tilde{v} = v/v^*$. For small values of k', the second term in equation 5.17 is well approximated by the term $\tilde{v}^{\frac{nm}{2}}$ (dashed blue line). (c) The opening probability of the origin $p_o(f(v))$ (equation 5.13) as a function of the volume per origin v for different Hill coefficients n and m (solid lines). The effective Hill coefficient $n_{\text{eff}}^{\text{fit}}$ is obtained from a fit of the function $p_o(f(v))$ to a Hill function (equation 5.25). The dashed lines show the approximated opening probability (equation 5.5) with the effective Hill coefficient as defined in equation 5.6. The vertical dotted line indicates the critical volume per origin v^* at which the opening probability equals 1/2. (d) The fitted (solid line) and the approximated (dashed line, equation 5.6) Hill coefficient as a function of the Hill coefficient n for different values of the Hill coefficient m. Except for very low Hill coefficients n and m, the approximated Hill coefficient agree well. In all graphs $f^* = v^* = 0.5$.(See Table 5.1 for all parameters.)

The average error probability $\langle S_{\text{err}} \rangle$ over all initiation times of the first event, t_1, is then given by

$$\langle S_{\text{err}} \rangle = \int_0^{\tau_{\text{d}}} d\,t_1\, q_1(t_1)\, S_{\text{err}}(t_2 - t_1 > \tau_{\text{l}} | t_1) \tag{5.32}$$

$$= \int_0^{\tau_{\text{d}}} d\,t_1\, q_1(t_1)\, e^{-\int_{t_1}^{t_1+\tau_{\text{l}}} dt' k_{\text{f}}(t')} \tag{5.33}$$

where τ_{d} is the doubling time of the cell. The propensity $q_1(t_1)$ that one out of two origin events happens at time t_1 is given by

$$q_1(t_1) = 2\, k_{\text{f}}(t_1)\, e^{-\int_{t_0}^{t_1} dt' 2\, k_{\text{f}}(t')} \tag{5.34}$$

Therefore, the average error probability $\langle S_{\text{err}} \rangle$ that the second origin fires after a time τ_{l} after the first event is given by plugging expression 5.34 into equation 5.33:

$$\langle S_{\text{err}} \rangle = 2 \int_0^{\tau_{\text{d}}} d\,t_1\, k_{\text{f}}(t_1)\, e^{-2\int_{t_0}^{t_1} dt' k_{\text{f}}(t')}\, e^{-\int_{t_1}^{t_1+\tau_{\text{l}}} dt' k_{\text{f}}(t')} \tag{5.35}$$

$$= 2\, k_{\text{f}}^0 \int_0^{\tau_{\text{d}}} d\,t_1\, \frac{\left(V_{\text{b}}\, e^{\lambda\, t_1}\right)^{n_{\text{eff}}}}{\upsilon^*\, n_{\text{eff}} + \left(V_{\text{b}}\, e^{\lambda\, t_1}\right)^{n_{\text{eff}}}}$$

$$\times\; e^{-\frac{2\, k_{\text{f}}^0}{n_{\text{eff}}\, \lambda} \ln\left(\frac{\left(V_{\text{b}}\, e^{\lambda\, t_1}\right)^{n_{\text{eff}}} + \upsilon^*\, n_{\text{eff}}}{V_{\text{b}}^{n_{\text{eff}}} + \upsilon^*\, n_{\text{eff}}} \right)} \tag{5.36}$$

$$\times\; e^{-\frac{k_{\text{f}}^0}{n_{\text{eff}}\, \lambda} \ln\left(\frac{\left(V_{\text{b}}\, e^{\lambda\, (t_1 + \tau_l)}\right)^{n_{\text{eff}}} + \upsilon^*\, n_{\text{eff}}}{\left(V_{\text{b}}\, e^{\lambda\, t_1}\right)^{n_{\text{eff}}} + \upsilon^*\, n_{\text{eff}}} \right)}, \tag{5.37}$$

where τ_{d} is the average division time of the cell and $\tau_{\text{d}} \gg \tau_{\text{l}}$, such that the probability that both origins have not yet fired at τ_{d} becomes negligible. The average probability that the second origin fires within a time interval $\Delta t = t_2 - t_1 < \tau_{\text{l}}$ after the first has fired at t_1, is then given by:

$$\langle P(\Delta t < \tau_{\text{l}}) \rangle = 1 - \langle S_{\text{err}}(\tau_{\text{l}}) \rangle \tag{5.38}$$

We solve the integral in equation 5.37 numerically and use expression 5.7 to predict the degree of synchronization for two origins (see Fig. 5.4b).

One can also calculate analytically the degree of synchrony at higher growth rates where there are typically four or more origins per cell at the beginning of an initiation cascade. The probability that none of the $n-1$ origins fire within the time τ_{l} after the first origin has fired at t_1 is similar to equation 5.37 and given by

$$\langle S_{\text{err}} \rangle = n \int_0^{\tau_{\text{d}}} d\,t_1\, k_{\text{f}}(t_1)\, e^{-n\int_{t_0}^{t_1} dt' k_{\text{f}}(t')}\, e^{-(n-1)\int_{t_1}^{t_1+\tau_{\text{l}}} dt' k_{\text{f}}(t')} \tag{5.39}$$

This is the probability that given that the first origin fires at t_1, all $n-1$ other origins fire later than $t_1 + \tau_{\text{l}}$. Importantly, one now also needs to take into account the cases where only one or more origins fire $t_1 + \tau_{\text{l}}$ and the other fire before. We here do not derive an expression for the scenario $n > 2$.

5.E. DERIVATION OF AVERAGE TIME INTERVAL BETWEEN FIRING EVENTS AND CV

The average time interval between two independent firing events $\langle \Delta t \rangle$ can be calculated analytically for the approximate opening probability in equation 5.5 via

$$\langle \Delta t \rangle = \int_0^{\tau_d} dt_1 \int_{t_1}^{\tau_d} dt_2 \, 2\, k_f(t_1)\, k_f(t_2)\, e^{-2\int_{t_0}^{t_1} dt' k_f(t')} \times e^{-\int_{t_1}^{t_2} dt'' k_f(t'')} \tag{5.40}$$

Solving this integral numerically gives the pink line in Fig. 5.4b.

The theoretical coefficient of variation of the initiation volume V is given by

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{\langle V^2 \rangle - \langle V \rangle^2}}{\langle V \rangle} \tag{5.41}$$

where we use

$$\langle V \rangle = \int_0^{\tau_d} dt\, k_f(t)\, e^{-\int_{t_0}^{t} dt' k_f(t')}\, V_b\, e^{\lambda t} \tag{5.42}$$

and

$$\langle V^2 \rangle = \int_0^{\tau_d} dt\, k_f(t)\, e^{-\int_{t_0}^{t} dt' k_f(t')} \left(V_b\, e^{\lambda t} \right)^2 \tag{5.43}$$

In this theoretical model, the only source of noise is intrinsic noise and the CV in equation 5.41 therefore also corresponds to the intrinsic noise as defined by Ref. [135] based on the derivation of Elowitz et al. [107].

5.F. MODEL FOR THE INITIATION CASCADE

To initiate DNA replication, 8 ATP-DnaA proteins and 3 DnaA proteins independent of the nucleotide binding state form a cooperative complex at the origin, which induces a conformational change and leads to the opening of the origin [43]. Consequently, the DNA replication machinery binds to the open origin and replication is initiated. Upon replication initiation, the DnaA proteins that were bound to the origin are likely being released into the cytoplasm, leading to a transient rise in the free DnaA concentration. It has been suggested that the release of origin-bound DnaA proteins from one origin triggers replication initiation at the remaining origins in a so-called 'initiation cascade' [137]. Here we propose a model to test whether the rise in the free DnaA concentration upon replication initiation at one origin could be sufficient to trigger replication initiation at the other origins.

So far, we have modelled the origin opening and firing process in a coarse-grained manner using a Hill function as a function of the initiation potential for the opening probability of the origin (see Appendix 5.A). Now, we instead model the binding of ATP-DnaA to the origin explicitly by introducing weak, cooperative binding sites for ATP-DnaA proteins at the origin. Specifically, we neglect the three strong binding sites to which both ATP- and ADP-DnaA can bind and assume that there are n weak binding sites with dissociation constant K_D^{ori} to which only ATP-DnaA can bind cooperatively. The probability that n ATP-DnaA proteins are bound to the origin is given by

$$p_b^n = \frac{Z_b^n}{\sum_{i=0}^{N} Z_i} \tag{5.44}$$

where Z_b^n is the partition function of n proteins bound to the origin and $\sum_{i=0}^{N} Z_i$ is the sum over all possible configurations the origin can be in. Let us first consider the scenario of only two cooperative binding sites. This gives rise to the following probability that two ATP-DnaA proteins are bound to the origin:

$$p_b^2 = \frac{Z_b^2}{Z_b^0 + 2\,Z_b^1 + Z_b^2} \tag{5.45}$$

The statistical weight of zero bound ATP-DnaA proteins is normalized to one $Z_b^0 = 1$ and the weight of one bound ATP-DnaA protein is given by $Z_b^1 = [D]_{f,ATP}/K_D$ with the dissociation constant $K_D = c_0^{-1}\, e^{-\beta \Delta G}$ and the free ATP-DnaA concentration $[D]_{f,ATP}$. The weight of two bound ATP-DnaA proteins is then given by $Z_b^2 = w\,[D]_{f,ATP}^2/K_D^2$ where $w = e^{\beta \Delta E}$ accounts for the additional energy gain from cooperative binding of two ATP-DnaA proteins. When cooperative binding is very strong then $\Delta E \gg \Delta G$ and we can neglect terms with lower powers of w:

$$p_b^2 \approx \frac{Z_b^2}{Z_b^0 + Z_b^2} = \frac{w\,[D]_{f,ATP}^2/K_D^2}{1 + w\,[D]_{f,ATP}^2/K_D^2} = \frac{[D]_{f,ATP}^2}{\left(\frac{K_D}{\sqrt{w}}\right)^2 + [D]_{f,ATP}^2} \tag{5.46}$$

This expression can be generalized to the case of n strongly cooperative ATP-DnaA origin binding sites:

$$p_b^n \approx \frac{[D]_{f,ATP}^n}{\left(\frac{K_D}{\sqrt[n]{w}}\right)^n + [D]_{f,ATP}^n} \tag{5.47}$$

We therefore recover the expression 5.2 for the origin opening probability in the coarse-grained model where now the critical free ATP-DnaA concentration is given by $[D]_{ATP,f}^* = K_D/\sqrt[n]{w}$.

In order to calculate the free DnaA concentration $[D]_f$ in the scenario where both DnaA forms can bind to the 300 homogeneously distributed strong binding sites on the chromosome and ATP-DnaA can additionally bind cooperatively to n weak binding sites on the origin, we write down the following expression

$$[D]_f = [D]_T - [D]_s - [D]_o \tag{5.48}$$

where $[D]_T$ is the total DnaA concentration in the cell, $[D]_s$ is the concentration of titration-site bound DnaA and $[D]_o$ is the origin-bound concentration of ATP-DnaA. An expression for the titration-site bound concentration $[D]_s$ as a function of the free DnaA concentration $[D]_f$ is obtained from the quasi-equilibrium approximation as explained in section 2.C

$$[D]_s = \frac{[s]_T\,[D]_f}{K_D^s + [D]_f} \tag{5.49}$$

with the total titration site concentration $[s]_T$ and the dissociation concentration of the titration sites K_D^s. The origin-bound ATP-DnaA concentration $[D]_o$ is given by the probability p_b^n that n ATP-DnaA proteins are bound to the origin times the total concentration of proteins that can be bound to these origin sites. This total concentration is given

by the concentration of origins that are available for ATP-DnaA binding $[n^{\mathrm{f}}_{\mathrm{ori}}]$ times the number of binding sites per origin n. Therefore, we obtain the following expression for the free DnaA concentration:

$$[D]_{\mathrm{f}} = [D]_{\mathrm{T}} - \frac{[s]_{\mathrm{T}}\,[D]_{\mathrm{f}}}{K^{\mathrm{s}}_{\mathrm{D}} + [D]_{\mathrm{f}}} - \frac{n\,[n^{\mathrm{f}}_{\mathrm{ori}}]\left([D]_{\mathrm{f}}\,f\right)^n}{\left(\frac{K_{\mathrm{D}}}{\sqrt[n]{w}}\right)^n + \left([D]_{\mathrm{f}}\,f\right)^n} \tag{5.50}$$

We here made the simplifying assumption that the free ATP-DnaA concentration $[D]_{\mathrm{f,ATP}}$ is given by the ATP-DnaA fraction f times the free DnaA concentration $[D]_{\mathrm{f}}$. This is a reasonable approximation, because the number of origin binding sites is small compared to the total number of DnaA proteins and the total number of titration sites. The total fraction of ATP-DnaA proteins in the cell f is therefore approximately equal to the fraction of ATP-DnaA proteins in the cytoplasm. Importantly, as explained in section 4.1, we assume that the switch components (de)activate DnaA independent of whether it is bound to the chromosome (either titration sites or origin sites) or freely diffusing in the cytoplasm. We solve equation 5.50 numerically at every time step of the simulations given the total titration site concentration $[s]_{\mathrm{T}}$, the total DnaA concentration $[D]_{\mathrm{T}}$ and the total number of for ATP-DnaA available origin binding sites $[n^{\mathrm{f}}_{\mathrm{ori}}]$. Replication is again initiated stochastically at every origin with a rate $k_{\mathrm{f}} = k^0_{\mathrm{f}}\,p^{\mathrm{n}}_{\mathrm{b}}$. We model the effect that DnaA proteins are released to the cytoplasm upon replication intiation by transiently reducing the number of available origin binding sites for a duration of $\tau_{\mathrm{b}} = 10$ min after an origin has initiated replication.

Modelling the ATP-DnaA binding to the origins explicitly does not significantly increase the degree of synchrony for a broad range of parameters. When the free ATP-DnaA concentration rises and approaches the critical free ATP-DnaA concentration $[D]^*_{\mathrm{ATP,f}}$, ATP-DnaA begins to bind cooperatively to the origin binding sites. This causes a weaker rise in the free DnaA concentration right before replication initiation as compared to a system without these origin binding sites (Fig. 5.12a and b). After an origin has been initiated, the origin binding sites become unavailable, causing an increase in the free DnaA concentration after replication initiation (Fig. 5.12a and b). While this increase should enhance the probability of other origins to fire replication as well, the weaker rise in the free concentration before replication initiation reduces the sharpness of the rise in the free ATP-DnaA concentration and should therefore lead to a decrease in the degree of synchrony. Indeed, comparing the simulations in which the origin-binding is modelled explicitly (Fig. 5.12c) to the previous model in which we simply used a Hill function for the opening probability (Fig. 5.9d) shows that the degree of synchrony is not significantly enhanced by the initiation cascade. The reason likely is that the positive effect of the initiation cascade is counterbalanced by the negative effect of a lower rise in the free concentration before replication initiation.

ACKNOWLEDGEMENTS

I want to thank Vahe Galstyan for the fruitful discussions and his mathematical insights, and Lorenzo Olivi for an inspiring discussion and for sharing his profound biological knowledge of this system with me.

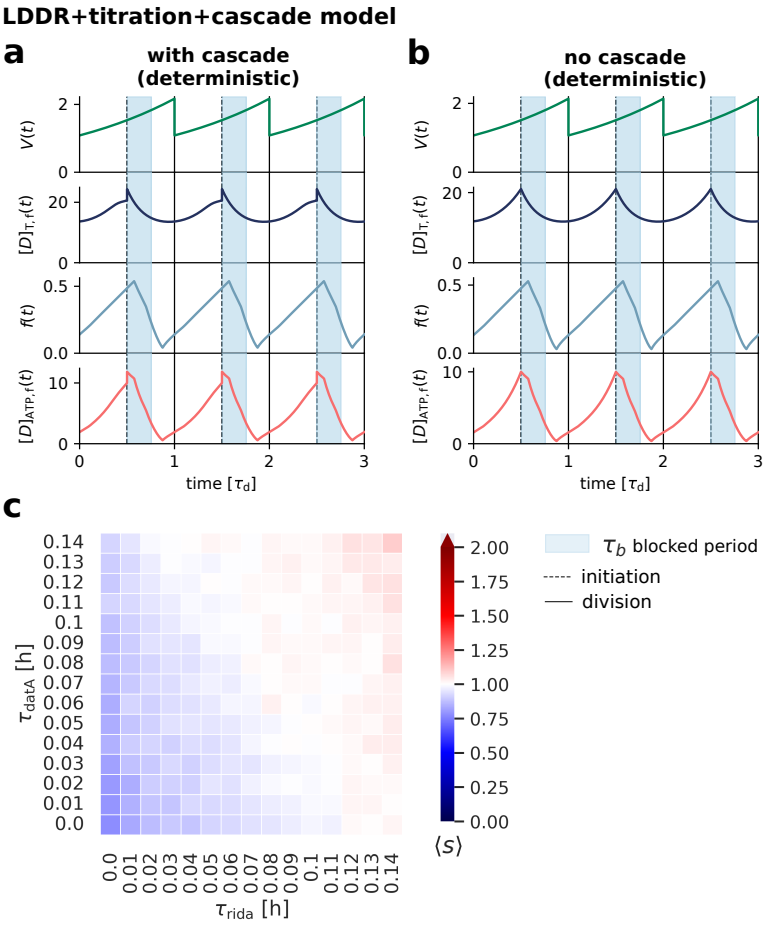


Figure 5.12: The initiation cascade does not significantly enhance the degree of synchrony. (a, b) The volume $V(t)$, free DnaA concentration (independent of whether DnaA is bound to ATP or ADP) $[D]_{\mathrm{T,f}}(t)$, the ATP-DnaA fraction $f(t)$, and the free ATP-DnaA concentration $[D]_{\mathrm{ATP,f}}(t)$ as a function of time (in units of the doubling time of the cell τ_{d} = 0.67 h=40 min) for a critical free ATP-DnaA concentration of $[D]^*_{\mathrm{f,ATP}} = 10\,\mu\mathrm{m}^{-3}$ with (a) and without (b) the initiation cascade. For illustration purposes, replication is here initiated deterministically at all origins as soon as the critical free ATP-DnaA concentration in the cell is reached. (a) When the free ATP-DnaA concentration approaches the critical free ATP-DnaA concentration of $[D]^*_{\mathrm{f,ATP}}$, ATP-DnaA proteins begin to bind to the weak, cooperative origin binding sites. This leads to a decrease in the rise of the free DnaA concentration right before replication initiation. Upon replication initiation, the bound ATP-DnaA proteins become unavailable, leading to a sharp increase in the free DnaA concentration. (b) For comparison, we here also show the time traces of a system in which the origin binding sites are not modelled explicitly and the only binding sites are the strong DnaA boxes distributed homogeneously all over the chromosome. (c) The average degree of synchronization $\langle s \rangle$ as a function of the replication time of the site *datA* τ_{datA} and the onset time of RIDA τ_{rida} for the same parameters as in Fig. 5.9d. Comparing the two panels reveals that the initiation cascade does not significantly enhance the degree of synchrony. For each parameter pair in c, the average degree of synchronization was obtained from $N = 1000$ consecutive cell cycles. (See Table 5.1 for all parameters.)

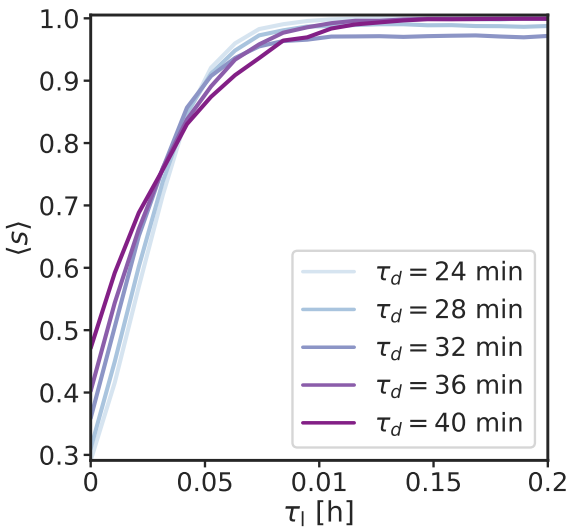


Figure 5.13: The average degree of synchrony does not depend strongly on the growth rate of the cell The average degree of synchrony $\langle s \rangle$ as a function of the duration of the licensing period τ_l for different cell-doubling times τ_d. The average degree of synchrony $\langle s \rangle$ is to a good approximation independent of the doubling time τ_d. The small decrease in the degree of synchrony at high licensing times τ_l at a doubling time of $\tau_\mathrm{d} = 32$ min is because at this doubling time replication initiation happens almost at the same time as cell division. When the cell divides during an initiation cascade, the number of origins decreases and the counted total number of origins at the end of the cascade is smaller than the total number of origins in both daughter cells. This error in counting the total change in the number of origins Δn_ori in equation 5.4 effectively reduces the degree of synchrony especially for long cascade durations (high licensing times τ_l). The decrease in the minimal average degree of synchrony $\langle s \rangle$ (at $\tau_\mathrm{l} = 0$) with increasing growth rate comes from the overall higher number of origins at higher growth rates and thus a higher denominator in equation 5.4. The effective Hill coefficient is set to $n_\mathrm{eff} = 50$. (See Table 5.1 for all parameters.)

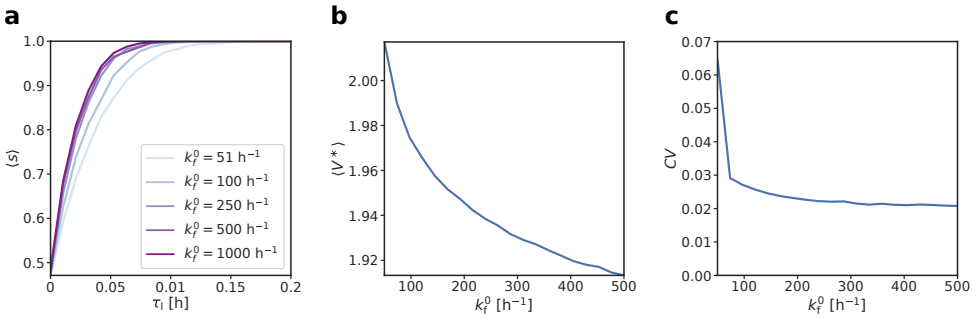


Figure 5.14: Replication cycle parameters depend only weakly on the maximal firing rate for high values of the maximal firing rate. (a) The average degree of synchrony $\langle s \rangle$ as a function of the duration of the licensing period τ_l for different maximal origin firing rates k_f^0 and an effective Hill coefficient of $n_{\text{eff}} = 50$. The maximal firing rate of $k_f^0 = 51\ \text{h}^{-1}$ corresponds to the parameters in the main text where the average initiation volume $\langle v^* \rangle = v^*$ (equation 5.29 in Appendix 5.C). The average degree of synchrony $\langle s \rangle$ first increases as a function of the maximal firing rate k_f^0 and then converges to a constant value for higher maximal firing rates. (b, c) The average initiation volume $\langle V^* \rangle$ (a) and the coefficient of variation (CV) (b) as a function of the maximal firing rate k_f^0 for the effective Hill coefficient $n_{\text{eff}} = 50$, the blocking period $\tau_b = 10$ min and delay period $\tau_b = 9.6$ min. (b) The average volume at which DNA replication is initiated $\langle V^* \rangle$ decreases weakly with the maximal origin firing rate k_f^0. (c) The CV first decreases sharply with the maximal origin firing rate k_f^0 and then settles to a constant value for higher firing rates. The high value of the coefficient of variation is a cell cycle effect that arises when an origin did not initiate replication during an initiation cascade. This error can only be corrected by cell division or by firing the remaining origin when the opening probability is sufficiently high. At a very low maximal origin firing rate, the cell volume strongly increases due to cell growth before the uneven number of origins can be corrected, leading to a high coefficient of variation. For each parameter set, the average degree of synchronization was obtained from $N = 5000$ consecutive cell cycles. (See Table 5.1 for all parameters.)

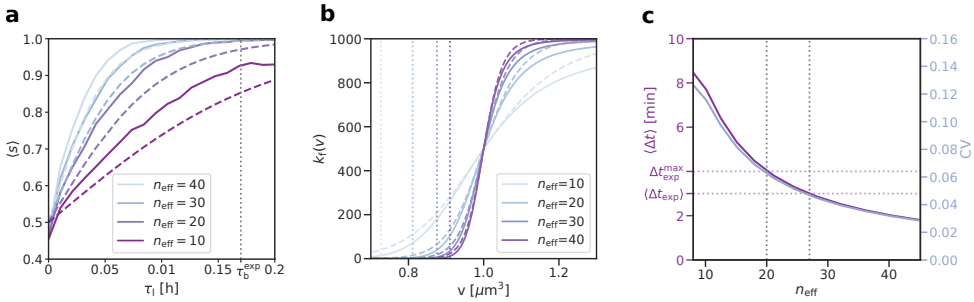


Figure 5.15: At a higher maximal origin firing rate than in Fig. 5.4a the effective Hill coefficient can be lower but must still be very high to obtain a high degree of synchrony at a licensing time shorter than the experimentally observed blocking period. (a) The average degree of synchrony $\langle s \rangle$ as a function of the duration of the licensing period τ_l for varying effective Hill coefficients n_eff (with $n = m = \sqrt{2\,n_\mathrm{eff}}$) and a maximal origin firing rate of $k_\mathrm{f}^0 = 1000\ \mathrm{h}^{-1}$ from the simulations (solid lines) and the theoretical prediction derived in the Appendix 5.D (dashed lines). In the simulations, origin firing is a two-step process (Fig. 5.1) while in the theoretical model this is coarse-grained into a one-step process. In these simulations, the blocking period τ_b is set larger than all tested licensing τ_l periods ($\tau_\mathrm{b} = 0.25\,\mathrm{h}$), such that over-initiation events cannot occur. The experimentally measured blocking period $\tau_\mathrm{b}^\mathrm{exp}$ is marked as a grey vertical dotted line. For each parameter set, the average degree of synchrony was obtained from $N = 5000$ consecutive cell cycles. Both in the simulations and in theory, the degree of synchrony is higher than for the parameter choice of the maximal firing rate following equation 5.29 at any given effective Hill coefficient (compare to Fig. 5.4a). At this higher maximal firing rate, the cell cycle simulations agree less well with the theoretical prediction. This can be understood by looking at the average initiation volume $\langle v^* \rangle$ in this parameter regime in the next panel. (b) The firing rate k_f as a function of the volume per origin v together with the average initiation volume $\langle v^* \rangle$ (vertical dotted lines) at different effective Hill coefficients n_eff, for both the simulations with a two-step initiation process (solid lines) and the theoretical model with a one-step process (dashed lines). Note that the resulting activation functions are highly similar, but do deviate away from the inflection point. This small deviation leads to the differences observed in panel a. In particular, due to the higher maximal firing rate k_f^0 (as compared to Fig. 5.4a), the initiation volume is shifted to lower values and in this regime, the theory of an approximated one-step process (dashed lines, equation 5.5) agrees less well with the cell-cycle simulations of a two step process (solid lines, equations 5.1 and 5.2). (c) The theoretical average time interval between two consecutive firing events $\langle \Delta t \rangle$ (pink line and axes) and the coefficient of variation of the initiation volume (CV, blue line and axes) as a function of the effective Hill coefficient n_eff (see Appendix 5.E for derivation). The experimentally reported average time interval to fire all origins in the B/r A *E. coli* strain of about $\langle \Delta t_\mathrm{exp} \rangle$ =3 min with an upper estimate of $\Delta t_\mathrm{exp}^\mathrm{max}$ =4 min are shown as horizontal dotted pink lines [134]. For this parameter regime ($k_\mathrm{f}^0 = 1000\ \mathrm{h}^{-1}$), the effective Hill coefficient for the experimentally reported firing time interval Δt is lower than in Fig. 5.4b, but with $n_\mathrm{eff} = 21 - 28$ still very high (vertical dotted grey lines). While the required effective Hill coefficient depends on k_f^0, the mapping between the CV of the initiation volume and the average time $\langle \Delta t \rangle$ between the first and last initiation event remains nearly unchanged. Moreover, the probability to fire all origins synchronously at the experimentally measured upper limit of $\Delta t_\mathrm{exp}^\mathrm{max}$ is also nearly unchanged, with $P_\mathrm{s} = 94\%$ in Fig. 5.15c and $P_\mathrm{s} = 95.5\%$ in Fig. 5.4a. This demonstrates that the predicted degree of synchrony does not depend on the precise parameter choice of the maximal origin firing rate and is thus a robust prediction. Also the finding of the main text that the effective Hill coefficient must be high to ensure synchronous replication initiation remains valid in this parameter regime. (See Table 5.1 for all parameters.)

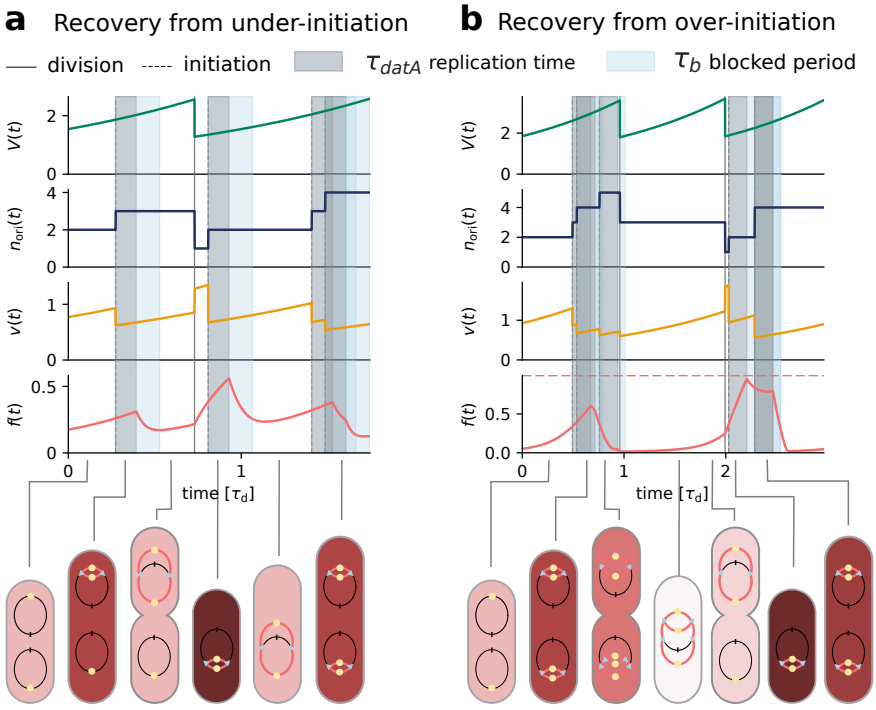
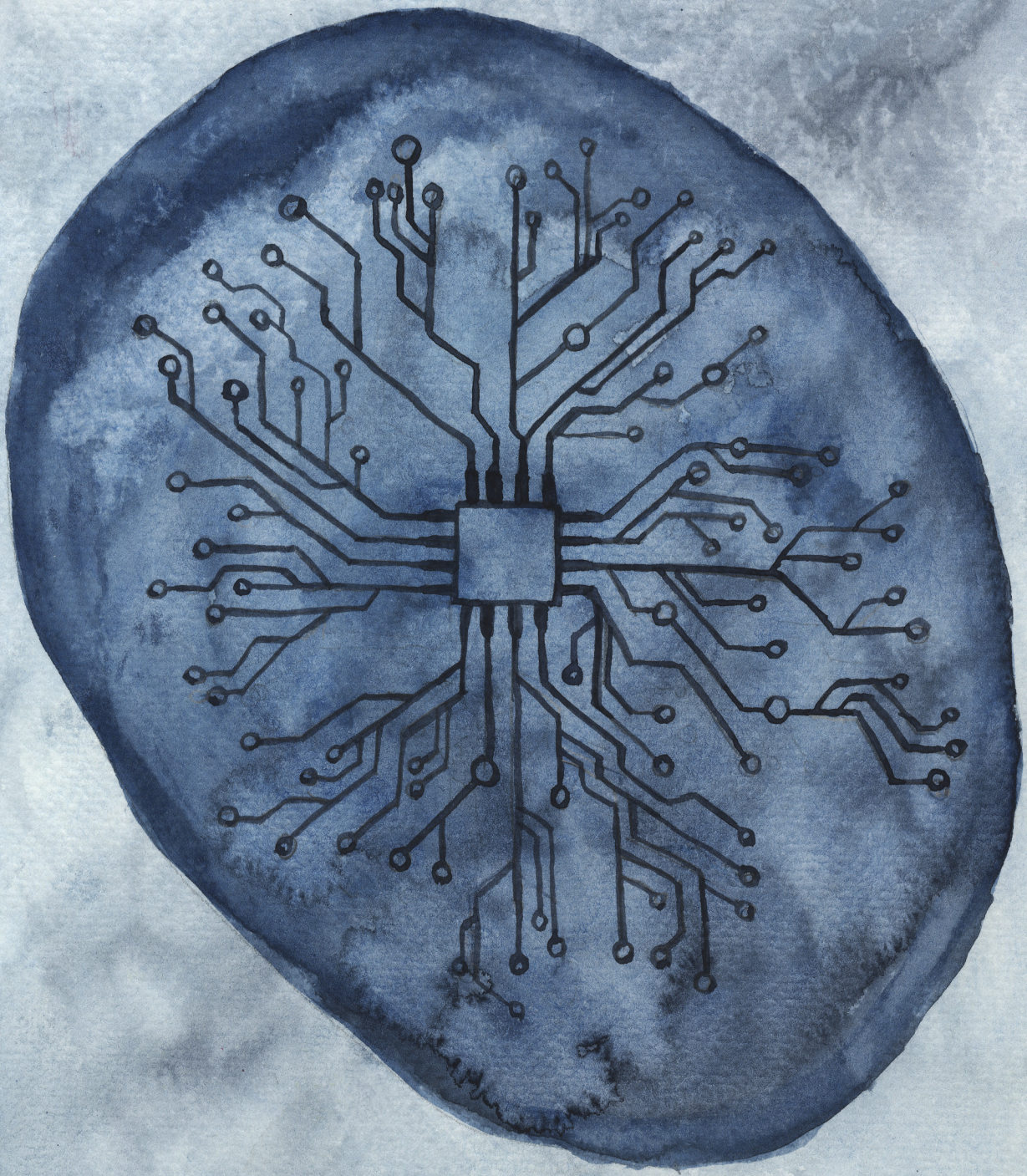


Figure 5.16: Cell division enhances recovery from over- or under-initiation. (a, b) The volume $V(t)$, the number of origins $n_{\mathrm{ori}}(t)$, the volume per number of origins $v(t) = V(t)/n_{\mathrm{ori}}(t)$ and the ATP-DnaA fraction $f(t)$ as a function of time (in units of the doubling time of the cell τ_{d}). During the blocking period τ_{b} (light blue shaded area), the newly replicated origins cannot be re-initiated and during the licensing period τ_{l} (grey shaded area) the initiation potential continues to increase. (a) Due to the stochastic nature of replication initiation, it can happen that only one out of two origins initiates replication during the licensing period. As the ATP-DnaA fraction $f(t)$ decreases after the end of the licensing period, it becomes less likely that the second origin fires. Upon cell division however, the volume per origin $v(t)$ increases abruptly, because of the unequal distribution of the chromosomes over the daughter cells. This causes a rapid increase in the initiation potential in the cell that is born with a single origin, causing replication initiation quickly after cell division. (b) Similarly, if the licensing period is too long, it can happen that an origin that has already initiated replication, initiates a second time. This causes the initiation potential to drop drastically. Upon cell division, the volume per origin $v(t)$ of the cell that has too many origins drops even further, preventing a new round of replication until the next cell division. At cell division, the now high volume per origin leads to a rapid rise in the ATP-DnaA fraction, causing a first and then a second round of replication initiation, until the number of origins is equal to four. (See Table 5.1 for all parameters.)

Artist impression of a synthetic cell
Watercolor, 2023
Mareike Berger



6

TOWARDS A SYNTHETIC CELL CYCLE

LORENZO OLIVI and MAREIKE BERGER

Recent developments in synthetic biology may bring the bottom-up generation of a synthetic cell within reach. A key feature of a living synthetic cell is a functional cell cycle, in which DNA replication and segregation as well as cell growth and division are well integrated. Here, we describe different approaches to recreate these processes in a synthetic cell, based on natural systems and/or synthetic alternatives. Although some individual machineries have recently been established, their integration and control in a synthetic cell cycle remain to be addressed. In this chapter, we discuss potential paths towards an integrated synthetic cell cycle.

This chapter was written by Lorenzo Olivi (DNA replication) and Mareike Berger (Cell-cycle control) as shared first authors, together with Ramon N.P. Creyghton, Nicola De Franceschi, Cees Dekker, Bela M. Mulder, Nico J. Claassens, Pieter Rein ten Wolde and John van der Oost.

Understanding the basic operating principles of life is a major scientific challenge, let alone using this knowledge to reconstruct a living cell from a basic set of essential parts. In recent years, significantly improved biological insights and impressive developments of molecular tools have been achieved. This prompted dreams to synthesise a living cell-like object by combining its molecular components. Attempts to create such 'synthetic life' will certainly contribute to addressing key fundamental questions on how life may have emerged and evolved, while elucidating its basic design principles. Furthermore, the generation of synthetic life forms may, on the longer term, provide novel platforms for a wide range of applications. A major breakthrough towards synthetic cells was the generation of a minimal cell by a top-down approach [138]. Using insights from transposon mutagenesis of a free-living bacterium with a small genome (1 Mb), *Mycoplasma mycoides*, a minimised functional genome was designed (530 kb). An innovative workflow including DNA synthesis, assembly of a complete circular chromosome, and its transplantation to a close relative of *M. mycoides*, eventually led to a minimal cell JCVI-syn3.0 [138]. Recently, addition of a few genes related to cell division has restored the wild type cell morphology and resulted in the creation of JCVI-syn3.0+126, consisting of 481 genes; remarkably, approximately 20% of these genes still have an unknown function[38]. Alternatively, several interdisciplinary teams are currently developing even more challenging bottom-up approaches [138], aiming to combine well-studied molecular components of different origins to eventually construct a living synthetic cell [139, 140] (Table 6.1). These bottom-up approaches require the development of functional modules as well as their stepwise integration [141].

There is no single definition of what constitutes a 'living' cell. Yet, arguably the defining characteristic of a living cell is its ability to make a copy of itself. In other words, living cells can grow and divide autonomously, which implies that during each cell cycle on average all its components are faithfully duplicated and partitioned over the daughter cells (Fig. 6.1). The modules involved in that process are DNA replication, DNA segregation, cell growth and cell division, which together form the cell cycle. In addition, the proteins that drive and control these processes need to be produced, which means that a minimal cell also should include a transcription-translation machinery. An early proposal suggested that this already requires at least 151 genes in a 113kb genome [171]. Another key feature of all forms of life, including a synthetic cell, is that these processes co-occur in some form of 'cellular' compartment or container. Moreover, some basic metabolism is also required for providing the energy and facilitating the reducing power and building blocks for biosynthesis of all the crucial components (e.g. DNA, RNA, proteins, lipids, co-factors). At least to some extent, however, these building blocks may be supplied externally. Hence, we conclude that a minimal synthetic cell should at least contain a cell cycle that combines DNA replication and segregation with cell growth and division.

In this chapter, we discuss major challenges and potential approaches to create a synthetic cell cycle. For comprehensive information about natural cell cycles, metabolic modules as well as general engineering and evolution strategies towards a synthetic cell, we refer to some excellent recent reviews [139–141]. Below, a comparison is made of promising modules for DNA replication, DNA segregation, and cell division. In addition, cell growth is obviously an essential feature of a living cell, and some examples are available of coupling the cell cycle with growth. However, in this chapter we do not discuss in

Module	Potential solution	References
Container	Liposomes, water-in-oil droplets, coacervates	[142–144]
Chromosomal configuration	Linear or circular, single/complete or multiple/partial chromosomes, single/multiple copies	[145–147]
Transcription and translation	*In vitro* transcription-translation systems (IVTT): PURE, TX-TL system	[140]
Energy	Arginine breakdown pathway, decarboxylase pathways, proton-pumping rhodopsins combined with ATP synthase	[139, 148]
Cell growth	Lipid biosynthesis, vesicle fusion	[140, 149–152]
DNA replication	ϕ29, T7, T4, *Escherichia coli* replicative machineries	[140]
DNA segregation	Random or active partitioning, entropy-driven segregation	[153–155]
Cell division		
Symmetry breaking	Reaction-diffusion, lipid phase separation, DNA partitioning	[156–160]
Membrane deformation	FtsZ, Cdv, actin-processing proteins, Min system, microtubules, lipids, DNA origami, mechanical deformation	[161–167]
Membrane abscission	ESCRT-III, dynamin, active droplets, mechanical splitting	[166, 168–170]

Table 6.1: Synthetic Cell modules. Different modules and processes that are considered essential to create a synthetic cell and alternatives proposed in the literature.

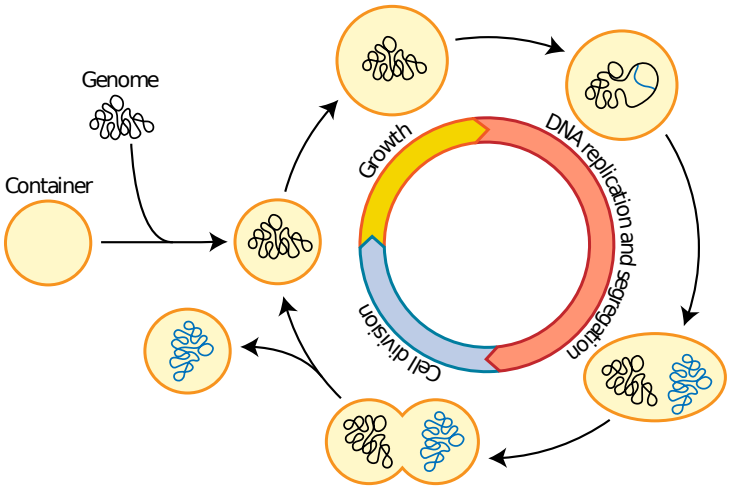


Figure 6.1: The natural and synthetic cell cycle. A natural/synthetic cell consists of a genome expressing all essential components inside a container. After a first phase of growth (yellow), the cell enters a phase in which the genomic content is replicated and segregated (red). Finally, cell division happens in the last phase of the cell cycle (blue), generating two daughter cells. Note that, although the first phase of the cell cycle is one of growth, the cell continuously grows also during DNA replication, DNA segregation and cell division.

detail reconstitution of biosynthesis routes for lipids or other essential building blocks that are also at the basis of cellular growth. For an extensive overview of such processes, we thus refer to a comprehensive recent review [140]. As coordination of DNA replication, DNA segregation and cell division is considered crucial, they should be integrated with mechanisms that monitor cell volume (growth) and DNA content. Altogether, these modules and their control will provide a basis for achieving the grand scientific goal of a growing and stably dividing synthetic cell. After discussing potential promises and pitfalls of the different modules and mechanisms, we will provide an integrated outlook on what a synthetic cell cycle could look like.

6.1. DNA REPLICATION: SIMPLE YET CONTROLLED

Every living cell executes DNA replication to ensure that its daughter cells will inherit a copy of the genomic content. Many efforts have been made to reconstitute this process extra-cellularly. An initial attempt to rebuild the DNA replication machinery of the model bacterium *Escherichia coli* set out to express the 13 core genes of its replication system in an *in vitro* transcription/translation system (IVTT) [172]. However, this minimal system did only result in synthesis of the complementary strand of a single-stranded circular DNA template [172]. For fully functional replication, including lagging strand synthesis and dissociation of the sister chromosomes, additional genes are required. Next, the full *E. coli* replication system (encoded by 25 genes) was reconstituted from purified proteins, resulting in one round of replication of circular double-stranded DNA

templates up to 200 kb [173]. Production of a full *E. coli* replication system within IVTT reactions has not yet been reported, probably because of its relatively complex nature, the relatively large number of components, and the lack of chaperones in IVTT systems [172].

A promising, simpler alternative to achieve replication in synthetic cells is the single-protein DNA polymerase (DNAP) of bacteriophages. Several of these replication systems have been reconstituted *in vitro* [174, 175], with the machinery of bacteriophage ϕ29 [176] being investigated in most detail. The ϕ29 system consists of a DNAP and three associated proteins (Fig. 6.2a). Unlike other well-known phage DNA polymerases, ϕ29-DNAP exhibits strand-displacement activity that circumvents the need for additional helicases to unwind the upstream duplex DNA [177]. The ϕ29 DNAP has been demonstrated to achieve high replication rates of linear DNA fragments up to 70 kb *in vitro* [178]. Primer-independent initiation of replication from defined origins of replication by the ϕ29 DNAP requires the associated proteins. The use of the phage DNAP and associated proteins allowed for self-replication of linear DNA in IVTT reactions, where ϕ29 origins of replication were placed at both ends of the DNA fragment [145]. This strategy was also demonstrated to result in DNA replication within a liposome container (Fig. 6.2a) [145]. Compartmentalised self-replication is an essential step toward synthetic life, as well as a suitable approach for *in vitro* evolution of other modules [141].

The ϕ29 DNAP has also been reported to mediate rolling-circle amplification of double-stranded circular DNA in IVTT mixtures. In the absence of the associated proteins, this DNA polymerase can use either DNA [146] or RNA [179] oligo-nucleotide primers to generate long linear DNA multimers that need to be processed back to circular chromosomes before cellular division [146]. A follow-up study demonstrated splitting and re-circularisation of these multimers through Cre-Lox recombination [179] (Fig. 6.2b). Moreover, it was shown that ϕ29 DNAP produced in IVTT reactions could amplify a set of 11 plasmids in parallel, covering a total size of 116 kb [147] (Fig. 6.2b).

Even though ϕ29 DNA replication is currently the best developed system for eventually driving replication in a synthetic cell, there are relevant bottlenecks that need to be addressed. Firstly, the DNA amplification rates of the ϕ29 system in IVTT conditions should be optimised [147], through adjusting the standard composition of IVTT reaction mixtures (PURE), and/or through laboratory evolution of the ϕ29 system itself [141, 180]. Secondly, it is important to consider that the natural function of this phage system is to replicate a linear genome. This may imply restrictions for the actual genomic configuration of the synthetic cell, such as to use linear chromosomes, which could be a viable option, although most small genomes in cells (e.g. those in bacteria) are circular. Alternatively, the ϕ29 system could be adjusted to replicate circular chromosomes. For circularisation, the aforementioned Cre-Lox approach could be used, which was achieved by directed evolution of the initial DNA template [146]. Alternatively, initial indications of spontaneous resolution of these multimers into circular chromosomes under IVTT reaction conditions [147] should be investigated in more detail. Another bottleneck of the ϕ29 system may be that the processivity of its DNAP (responsible for replicating its 20 kb genome) is not sufficient to enable replication of a single chromosome of the size needed for a synthetic cell [145, 178]. As splitting the genome into multiple smaller chromosomes is not preferred (see DNA segregation, section 6.2), a potential solution could

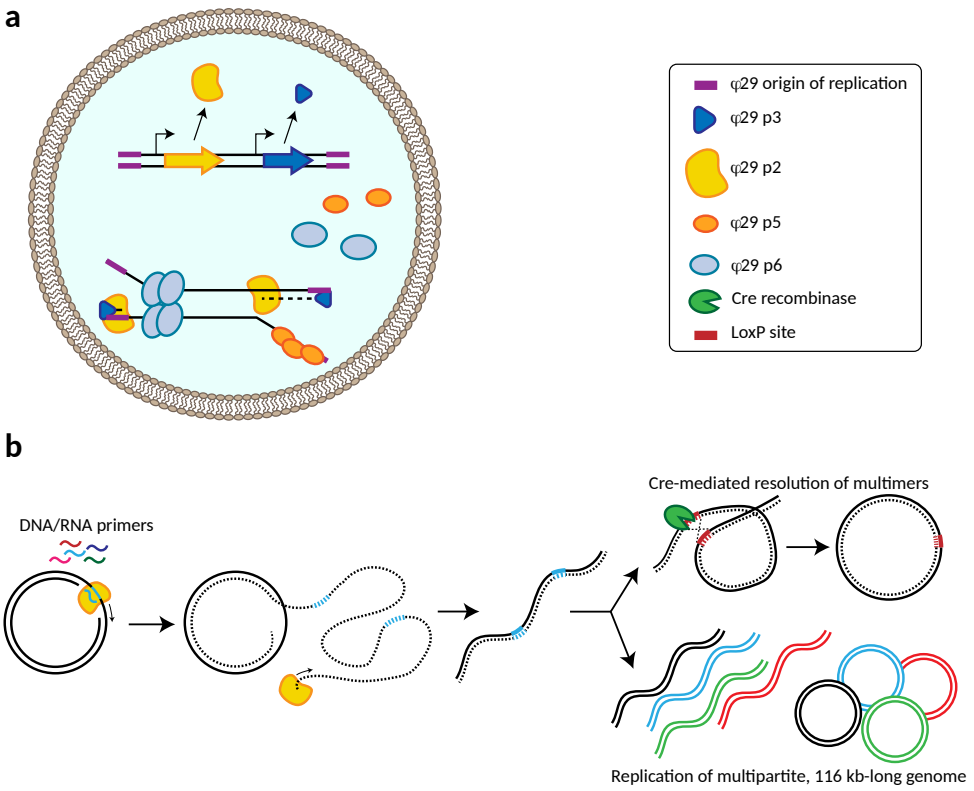


Figure 6.2: Synthetic DNA replication. Implementation of the bacteriophage ϕ29 system for DNA replication in *in vitro* transcription-translation (IVTT) reactions. **(a)** The reconstitution of the complete ϕ29 DNA replication machinery has been demonstrated inside a synthetic liposome. Bi-directional self-replication of a linear DNA has been achieved, with the linear DNA template encoding the DNA-polymerase (DNAP, protein p2) and the Terminal protein (p3), which binds to the origin of replication at the end of the linear fragment and provides an hydroxyl group to prime DNA-polymerisation. The single-stranded DNA binding protein (p5) and the origin-binding protein (p6), promoting the unwinding of the phage origin of replication, were supplemented as purified proteins [145]. **(b)** ϕ29 DNAP-mediated rolling-circle amplification of circular dsDNA molecules, resulting in a linear dsDNA multimer [146]. The synthetic multimer can be re-converted into circular DNA monomers through Cre-Lox recombination [179] (top). In a ϕ29 DNAP-based replication experiment using a multipartite genome [147] (bottom right), spontaneous monomerisation, disentangling and circularisation was observed, notably resulting in a mixture of linear multimers and circular monomers [147].

be to improve the processivity of ϕ29 DNAP by laboratory evolution [141] . Alternatively, replicase systems from known viruses with larger, circular dsDNA genomes (250-500 kb) might be suitable candidates for replicating the genome of a synthetic cell.

In addition, an important limitation of viral replicative systems is the lack of regulation, as these systems evolved to drive uncontrolled amplification of viral DNA fragments. Appropriate timing of genomic replication, however, is essential for the fitness of biological cells [181], and most likely for synthetic cells as well. To obtain a synthetic cell surviving throughout generations, we envision the introduction of control over replication initiation as an essential step, as discussed in more detail below.

6.2. DNA SEGREGATION: USING A BIOLOGICAL OR A PHYSICAL APPROACH

A cell does not only need to duplicate its genome, but it also needs to make sure that the generated sister chromosomes are properly partitioned over the daughter cells. This implies that, seemingly against the odds of entropy, a spatial order should be established through disentangling and separating the two chromosome copies. Well-known natural modules that drive this segregation process are the mitotic spindle apparatus of eukaryotic cells [182] and the bacterial Par system [153]. The DNA segregation module of a synthetic cell could be adopted from the latter systems, but alternatively could also consist of a much more minimal mechanism. In order to be fully functional and controllable, the DNA segregation module should accomplish three main tasks: (i) breaking the symmetry to initiate disaggregation, (ii) achieving complete spatial segregation, and (iii) ensuring correct partitioning over the daughter cells. There are several alternative scenarios to achieve these tasks, each with consequences for the configuration of the chromosome, and for the type of DNA replication machinery of such a cell.

The simplest solution to accomplish appropriate segregation of identical genomes would be through random partitioning (Fig. 6.3a). This model was originally proposed for plasmid partitioning [183], and relies on a large number of relatively small, identical DNA molecules. This sharpens the peak of the binomial distribution of inherited copies over the daughter cells to such an extent that the probability of a daughter cell receiving only a few or no copies of DNA becomes acceptably low. However, mathematical modelling of such a stochastic partitioning for other cytosolic macromolecules (e.g. proteins), shows that this results in major heterogeneity [154]. According to this model, random partitioning requires many copies of the chromosome, possibly in the order of tens. Replication of such a high number of DNA molecules would result in a huge burden, even when a designed synthetic cell may require a smaller genome (250-500 kb) than most bacteria. In addition, this approach would require precise fine-tuning of transcription and translation of all essential coding sequences, to avoid issues with imbalanced gene expression. Overall, we consider this solution inappropriate for DNA segregation in a synthetic cell.

Alternatively, an active molecular machinery can be implemented to achieve DNA segregation (Fig. 6.3b). Although a minimal mitotic spindle has successfully been reconstituted *in vitro* [184], this system might still be too complex to be properly integrated and coordinated in a synthetic cell. Instead, inspiration can be taken from the bacterial

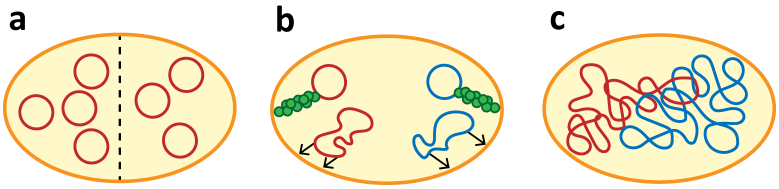


Figure 6.3: Synthetic DNA segregation mechanisms. **(a)** Random partitioning of DNA molecules, generating viable offspring in synthetic cells with sufficiently high genome copy numbers. **(b)** Molecular machineries for active partitioning of DNA molecules, such as the Par system (top) or membrane anchors in elongating cells (bottom). **(c)** Entropy-driven segregation, relying on the physical behaviour that large overlapping DNA macromolecules constrained in a small volume will spontaneously get separated from each other.

world, in the form of plasmid or chromosomal segregation machineries. The bacterial actin-like plasmid partitioning (Par) system has been reported to push coupled plasmids apart [153], or to space them regularly along the nucleoid [185], contributing to equal partitioning. Regarding bacterial chromosomes, there is a long list of proposed mechanisms, ranging from dynamic anchoring to the poles or to other sites at the elongating membranes of growing cells [186, 187], to directional biases in either transcription or replication [188, 189]. Yet, it is unlikely that any of these mechanisms alone would be sufficient to achieve full segregation [190]. Moreover, and especially relevant in the perspective of a synthetic cell, all these mechanisms require coordinated interactions between the involved structures, inevitably leading to an overwhelming complexity.

Because of the downsides of the previous options, it is interesting to explore a third scenario, namely entropy-driven segregation (Fig. 6.3c). This possibility relies on *in silico* observations that long polymer chains tend to segregate spontaneously when sufficiently spatially confined [155]. This effect can be explained by the strong increase in the entropic cost associated with overlapping polymers when they are constrained to volumes much smaller than their natural size [191]. There are at least three requirements for entropic segregation to occur. The primary one is a sufficient spatial confinement, and thus a minimal mass (size and density) of the polymers involved [156]. For instance, it is unlikely that entropic segregation could drive partitioning of a 500 kb genome in cells with a diameter of more than a few micrometres. If the totality of the coding sequences envisioned would not provide sufficient length, non-functional "filler" DNA could be incorporated in the genome. Secondly, the confinement shape plays a major role, which is why entropic segregation has been extensively explored for elongated cylindrical shapes like that of the *E. coli* cell [192]. For a synthetic cell that starts out spherical, the required symmetry breaking may be achieved by the constriction machinery needed for division (see Cell division, section 6.3). Lastly, entropy-driven segregation can be optimised by adapting the DNA topology, for example by introducing circularisation [193], supercoiling or loop-creating cross-links [194], which can be induced by recently established molecular mechanisms [195]. Furthermore, macromolecular crowding can increase the effective degree of confinement through the depletion effect [196].

It is important to consider that the requirement of a relatively large DNA molecule

would influence the choice of the replicative system. Another caution is that entropy-driven DNA segregation has so far been based on *in silico* predictions that are supported by indirect evidence only [197, 198]. Hence, verification of entropy-driven partitioning of 250-500 kb genomes in different containers is required to decide on the suitability of this approach. Nevertheless, the strongest argument for entropy-driven segregation as the mechanism of choice for a synthetic cell is that its success does not hinge on precisely tuned biochemistry, but rather on a generic physical principle. In our perspective, this potentially makes it more robust, and therefore more likely to work independent of the constraints imposed by the other functional modules.

6.3. CELL DIVISION: USUAL SUSPECTS AND OUT-OF-THE-BOX ALTERNATIVES

After chromosome segregation, cells proceed with the division into daughter cells. Conceptually, cell division is a multistep process that can be broken down into three distinct steps: symmetry breaking, membrane deformation, and membrane abscission. Regarding their implementation in molecular systems, one could adopt the well-studied natural cell division machineries, but also several "out-of-the-box" alternatives (Fig. 6.4).

Before considering division strategies, it is important to briefly consider the pros and cons of different types of synthetic containers. Water-in-oil droplets are an attractive option for testing purposes: very easy to mass-produce through microfluidics [142] and offering excellent encapsulation of components. Yet, the large surface tension makes these droplets difficult to deform, and the external hydrophobic environment hardly allows any transport in and out of the synthetic cell. Another interesting container candidate is based on the recently rediscovered coacervates [143]. In this case liquid droplets are formed upon mixing two immiscible liquid phases, which would facilitate easy exchange of material due to the lack of a physical boundary. Coacervates, however, can spontaneously fuse, making it difficult to maintain a clear identity of protocells and hindering their use as containers for a synthetic cell. Finally, liposomes represent an excellent minimal model for a container of a synthetic cell. Their lipid bilayer mimics natural cell membranes, structures that can potentially be equipped with molecular machineries to deform and divide [199, 200]. The non-permeable bilayer will prevent the influx of building blocks, but this can be overcome by the reconstitution of functional channels and transporters [157]. Based on these arguments and on the large amount of available knowledge, liposomes are to date the most obvious candidates as containers for a synthetic cell.

Cell division inevitably calls for symmetry breaking as a first step in the process of generating offspring. Several physical and biochemical mechanisms can be considered to induce asymmetries. Reaction-diffusion at the membrane can generate protein gradients and set up polarity. The best-studied example of reaction-diffusion is the Min system of rod-shaped bacteria like *E. coli* [199, 201], which uses pole-to-pole oscillations to define the division site at the cell midpoint. This Min system can also segregate membrane-bound proteins to spatial gradients along the cell length [202].

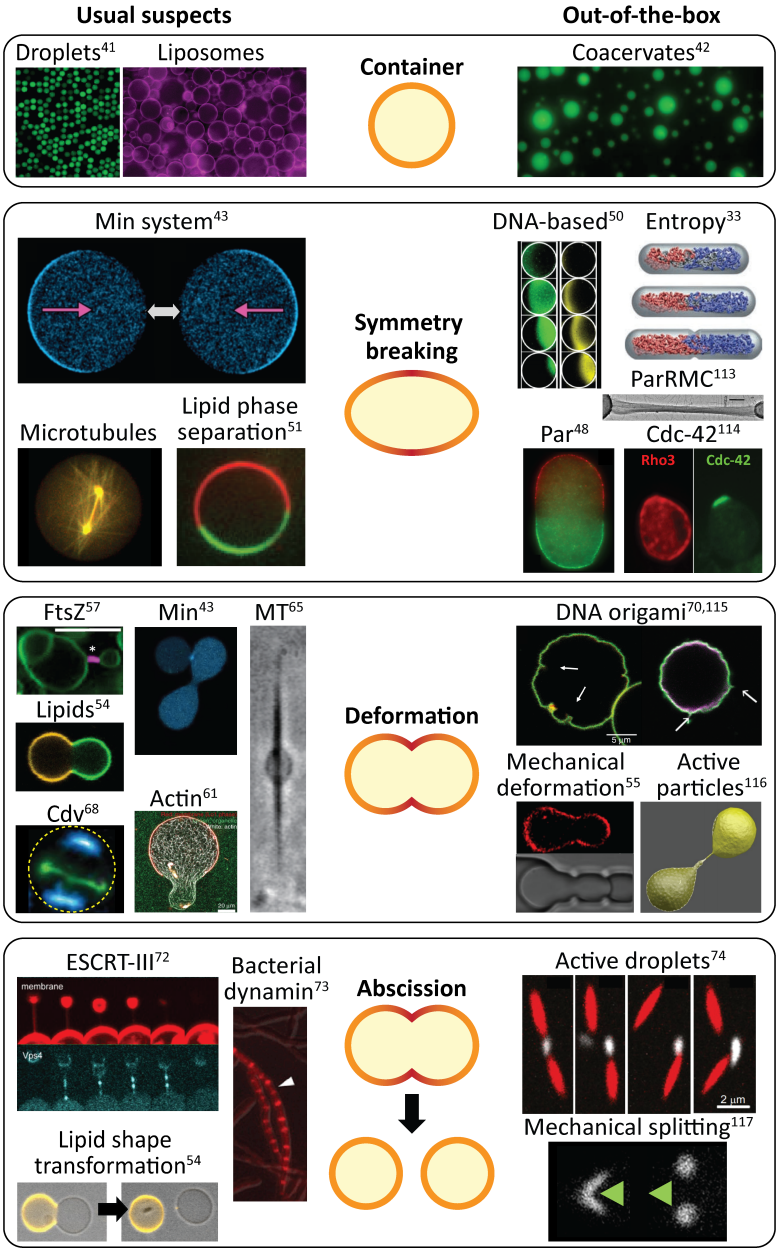


Figure 6.4: Synthetic cell division. Alternative synthetic cell container and the stages of symmetry breaking, deformation and abscission during cell division. For each step, well-characterised (left) and alternative "out-of-the-box" candidates (right) are presented. (caption continues on the next page)

Although more complex, alternative reaction-diffusion systems are the metazoan PAR system [158] (not to be confused with the aforementioned bacterial Par system) that produces a stable concentration gradient along the membrane, and the Cdc-42 system [159] that localises the division site in budding yeast. Interestingly, fully synthetic DNA-based reaction-diffusion systems have also been demonstrated to create chemical waves in solutions in closed reactors [213]. Although still relatively complex, such systems may provide spatial cues for division and can be expanded to other rationally designed modules. Lipid phase separation is another well-characterised and robust process to break symmetry within a liposome [160]. However, here the downside is the requirement for an accurately-controlled lipid composition throughout multiple generations in a synthetic cell. The systems involved in DNA segregation, discussed in the previous section, may also be of interest for synthetic cell division, since they break symmetry by partitioning sister chromosomes to the poles of the cell. For example, entropy-driven segregation[31,33] in rod-shaped vesicles may help to identify a middle plane through nucleoid occlusion [161]. Nonetheless, for now reaction-diffusion systems seem to represent the most promising choice to achieve symmetry breaking as they have been very well characterised. The Min system is considered to be of particular interest, as pole-to-pole oscillations have already been established in liposomes [199].

During the division process significant changes occur in the shape of the cellular container. *In vivo*, cell deformation is accompanied by excess membrane area that is generated by growth. In *in vitro* studies, this has often been achieved by applying an osmotic pressure difference across the membrane [215]. Osmotically deflated vesicles revealed that periodic Min protein binding can result in dumbbell-shaped liposomes [199], which is an important step towards abscission, although completion of the cell division process has not been accomplished yet *in vitro*. In mixed-lipid liposomes, the

Images depicting droplets, coacervates, lipids and lipid shape transformation, mechanical deformation, FtsZ, and DNA origami are reproduced from refs. [143, 168, 197, 203–205], respectively, under the CC BY 4.0 licence. Images depicting Min system, Min are reproduced/adapted from ref. [199] under the CC BY-NC 4.0 licence. The image depicting MT was reproduced from ref. [206] (10.1021/acs.langmuir.6b00799) and further permissions related to the material excerpted should be directed to the ACS. The image depicting mechanical splitting was reproduced from ref. [207] (10.1021/acsnano.7b08411) and further permissions related to the material excerpted should be directed to the ACS. The image depicting DNA origami was reproduced from ref. [208] under the CC BY 3.0 licence. The image depicting Par is reproduced/adapted with permission from Development from ref. [158]. Images depicting entropy symmetry breaking, actin, and active particles were reprinted by permission from refs. [156, 209, 210], respectively. The image depicting ParRMC is reproduced from ref. [211]. Reprinted with permission from AAAS. The image depicting ESCRT-III is reproduced from ref. [169]. Reprinted with permission from AAAS. The image depicting Cdv is reproduced from ref. [212]. Reprinted with permission from AAAS. The image depicting DNA-based symmetry breaking was reprinted with permission from ref. [213]; copyright 2013 American Chemical Society. Images depicting Cdc-42 were reproduced from ref. [214], with permission from American Society for Microbiology. Images depicting bacterial dynamin and active droplets were reproduced from refs. [166, 170], respectively, with permission from Proceedings of the National Academy of Sciences of the United States of America. The image depicting lipid phase separation was reproduced from ref. [160] , with permission from Europhysics Letters. The image depicting microtubules was kindly provided by Prof. Marileen Dogterom (TU Delft).

line tension at the interface between two lipid phases can deform and even split the liposome, albeit typically asymmetrically [203]. External interventions in the form of mechanical deformation of the vesicles using microfluidic traps [168] offer good spatial control, and may be very useful in initial stages of the synthetic cell research. However, for achieving a fully autonomous synthetic cell, this eventually needs to be substituted with a dedicated division machinery.

In natural cells, most deformations are controlled by dedicated protein machineries. By far the best-studied system of this type is the bacterial division protein FtsZ that orchestrates membrane constriction by a contracting Z ring. This system is appealing due to its simplicity. Moreover, FtsZ rings were proven to localise to membrane necks in liposomes [204, 216]. Interestingly, the Min system has been proven to be able to dynamically localise FtsZ near mid cell [217]. However, it is currently unclear whether *in vitro* the Z-ring can exert sufficient force to constrict the membrane of the synthetic cell to the point of division, or if instead its mere role is to recruit other proteins that actually drive this process. Recent studies reported that FtsZ is able to stabilise membrane deformations such as local buds [204, 216, 218, 219], but a dynamic constriction from large to small diameter rings remains to be demonstrated. Eukaryotic actin [209] is another interesting filamentous protein that, in concert with actin-processing motor proteins, is able to induce blebbing [162], protrusions and vesicle elongation [163], as well as reversible spindle-like vesicle deformation [164]. These proteins can exert high forces on a membrane in a spatially controlled way, but their implementation seems less attractive due to the large number of components required. Alternatively, microtubules are a relatively simple cytoskeletal system able to deform membranes, either autonomously through polymerisation [206] or by coupling to the motor protein kinesin [165, 220]. However, microtubules generate a linearly extended rather than a dumbbell-shaped vesicle, hindering division. Another system that recently is receiving increasing attention is the archaeal Cdv system [212]. Cdv employs a constriction mechanism similar to the eukaryotic ESCRTs, its evolutionary homologue. While *in vitro* Cdv reconstruction remains to be established, the relative stability of these archaeal proteins may be an advantage when used in a bottom-up approach.

On a different note, DNA origami nanotechnology [221] holds great potential to design functional components with tailored features such as local bending of the membrane [205, 222]. However, implementation in a synthetic cell poses several challenges, such as achieving a proper folding of the origami or the need of an energy-burning module to achieve active constriction. On top of this, DNA origami cannot be synthesised in a cell. Some of these issues can, however, likely be tackled with the use either of RNA origami or of accessory DNA-binding proteins, but additional research is needed to demonstrate the feasibility and use of this solution in synthetic cells.

In the final stage of division, the membrane is fully pinched off and two separate daughter cells are born, ready to undergo a new cell cycle round. Membrane abscission is so far the least studied phenomenon when it comes to reconstituted systems. The eukaryotic ESCRT-III system has recently been reconstructed and proven to be functional *in vitro* [169]. This work, however, highlighted the challenges involved in controlling this protein complex both spatially and temporally. A simpler system is the bacterial dynamin that is recruited by the Z-ring [170]. Further investigation is needed to assess

whether dynamin is sufficient to efficiently mediate abscission on its own. Due to its natural connection with FtsZ, the bacterial dynamin system is a promising candidate to accomplish membrane abscission and drive cell division for a synthetic cell. Regarding non-natural solutions, "active droplets" of coacervates with chemical reactions occurring inside have been shown to be able to autonomously split using proteins such as actin and myosin [166]. Assessing whether such coacervates can be encapsulated in vesicles and drive their division is another interesting future perspective [167] that needs to be tackled, together with preventing spontaneous fusion between coacervate droplets.

6.4. CELL-CYCLE CONTROL

Obtaining an autonomously replicating cell by bottom-up, modular assembly of lifeless components would be a major scientific breakthrough. However, to obtain a synthetic cell surviving throughout multiple generations, a certain level of homeostasis of cell size and genomic content must be ensured. Disturbing the latter balance is associated with severe fitness-loss in natural [181, 223], and most likely also generates non-viable offspring of synthetic cells (Fig. 6.5a). We argue that the cell cycle of a synthetic cell should fulfil three minimal requirements. Firstly, replication must be coupled to cell growth, such that, independently from the chosen chromosomal configuration, the DNA density remains constant over many generations. Secondly, during DNA segregation, sister chromosomes must be correctly partitioned during the cell division process, such that each daughter cell contains at least one copy of the chromosome. Finally, cell division must be coordinated with growth in such a way that during each cell cycle a doubling in cell size occurs, thus achieving cell size homeostasis over many generations (see Appendix 6.A and Fig. 6.7). In efforts to build a synthetic cell, we can explore and exploit the large variety of solutions that different natural life forms have evolved to couple these core processes.

Cell growth and the cell cycle are coordinated, yet separable processes. In most unicellular species, inhibiting cell growth causes a delay or even an arrest of the cell cycle [224, 225]. Conversely, blocking or perturbing the cell cycle typically has a minor impact on cell growth [226, 227]. While in some eukaryotes (e.g. fission yeast) a dependence of the growth rate on the stage of the cell cycle has been reported [228], bacteria are known to grow exponentially at an approximately constant rate throughout the whole cell cycle [3, 4]. Assuming that the minimal metabolism of a synthetic cell will probably lead to a continuous increase in cell volume (e.g. through lipid biosynthesis), the main challenge of controlling the cell cycle of a synthetic cell is to couple its core processes to growth and to coordinate the timing of the different events.

In most eukaryotes, the cell cycle is strictly regulated and consists of four discrete phases (Fig. 6.5b). An intricate system of molecular checkpoints and feedback control mechanisms leads to sequential activation of different transcriptional master regulators, which in turn trigger the transitions between different phases of the cell cycle [229]. A similar system is also present in the bacterium *Caulobacter crescentus* [230]. However, implementing even the simplest version of such a controlled cell cycle would be a considerable challenge, since it involves a large number of components and non-linear feedback loops [231, 232]. For these reasons, simpler alternatives should be explored.

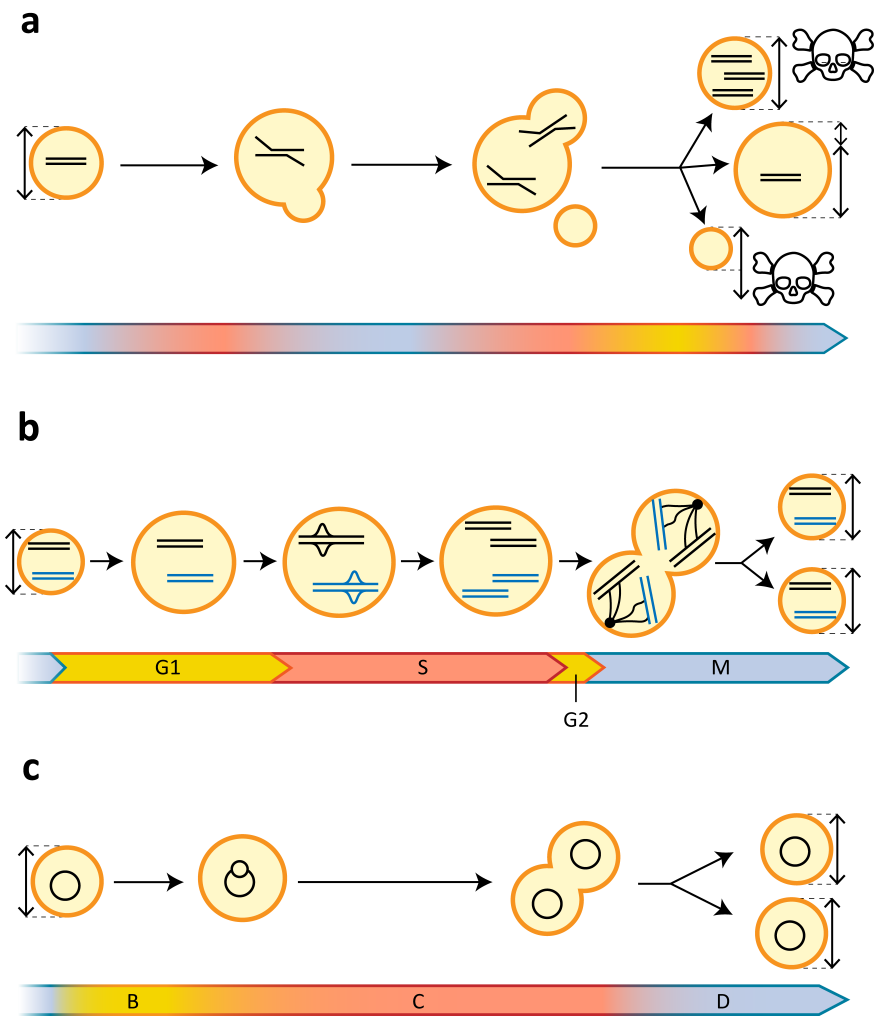


Figure 6.5: Cell cycle in natural and synthetic cells. **(a)** In a synthetic cell that lacks control over DNA replication, DNA segregation and cell division, nothing prevents these processes from happening simultaneously, most likely leading to loss of cell-size homeostasis, hyper-replicative stress, inconsistency in genomic content, and non-viability of daughter cells. **(b)** In eukaryotic cells, the passage through phases of the cell cycle is controlled by molecular checkpoints, depicted as bold lines separating the different phases. **(c)** In the bacterium *Escherichia coli*, the cell cycle lacks a eukaryotic-type checkpoint system. However, the order of the cell cycle phases is maintained through coupling with growth. coupling with growth.

The cell cycle of the model bacterium *E. coli* appears to be less strictly regulated than its eukaryotic counterparts, as the different phases of the bacterial cell cycle are not strictly separated by checkpoints [233] (Fig. 6.5c). Despite the lack of master regulators, *E. coli* cells exhibit a robust cell cycle that even allows them to divide faster than the time it takes to complete DNA replication [20, 22]. This apparent paradox was resolved by early studies reporting that at high growth rates, *E. coli* cells start a new round of chromosomal replication before the previous one is finished, giving rise to chromosomes with multiple replication forks [21]. It was initially proposed that DNA replication is initiated in *E. coli* at a critical volume per origin of replication, independently of the growth rate [21]. Indeed, recent single-cell experiments provide strong experimental support for this prediction [22, 24]. In addition, it was observed that, in total, the time required for the DNA replication and the cell division processes together is roughly constant. Overall, this led to the idea that replication initiation and cell division are co-regulated with growth [40] via a combination of so-called "sizer" and "timer" mechanisms (see Appendix 6.A and Fig. 6.7). In this scenario, *E. coli* initiates chromosomal replication at a constant volume per origin (sizer) and divides a fixed time later (timer), yielding stable cell cycles.

Yet, recent studies have challenged this idea [23, 24]. *E. coli* adds a constant volume each generation, independently from its size at birth, a concept known as "adder" (see Appendix 6.A and Fig. 6.7). At closer inspection, the "adder" principle is also valid for the added volume between replication initiation and cell division and even between replication initiation events of successive cell cycles, leading to "double adder" behaviour [23, 24]. This suggests that DNA replication and cell division, while both individually coupled to growth, might be less coordinated with each other than previously thought.

It is commonly believed that DnaA and FtsZ [23] are the central components that regulate bacterial DNA replication and cell division, respectively. The DnaA system controlling replication initiation in *E. coli* has been fairly well-characterised and could therefore be potentially used to control DNA replication by the aforementioned ϕ29 machinery (Fig. 6.2). Replication by the ϕ29 system starts at the origin of replication through helicase-based unwinding of the linear, double stranded phage DNA, but this is an uncontrolled process that normally leads to a replication burst that is typical for phage proliferation. In a synthetic cell, control mechanisms could be designed to couple expression of the ϕ29 helicase to the cell cycle. Alternatively, substituting the uncontrolled viral helicase by the well-controlled bacterial DnaA system would allow a synthetic cell to exert strict control over the initiation step of chromosomal replication, while synthesis of the new DNA strands would be mediated by the minimal ϕ29 replicative machinery.

How DnaA and FtsZ control chromosome replication initiation and cell division is not yet completely understood. A long-standing idea has been that these processes are timed via protein accumulation up to a threshold level [10, 13, 234]. However, the initiator protein DnaA exists in two states, either active origin-unwinding (ATP-bound) or inactive (ADP-bound), and most likely this activation switch is responsible for the coupling of replication initiation to growth [233]. Despite extensive research [43], however, a mechanistic understanding of this system is still lacking. Therefore, simplified synthetic solutions based on the accumulation of an initiator protein up to a threshold level could be considered for implementation in a synthetic cell.

Plasmid copy-number control systems provide alternative mechanisms for coordinating replication with growth. Plasmids are extrachromosomal genetic elements that, in some instances, replicate independently of the chromosome of the host [133, 235]. Most, if not all, plasmids regulate the initiation of their replication based on their intracellular concentration in order to maintain a constant plasmid density[236–238]. Three mechanisms have been described for control of plasmid replication initiation, all mediated by different plasmid-encoded regulators: a negatively-autoregulated transcription regulator protein [236, 239], inhibitory antisense RNAs [240], or a plasmid-binding protein that inhibits replication through cross-linking plasmids at elevated concentrations ("handcuffing") [238]. In addition to these replication control mechanisms, some plasmids have evolved addiction mechanisms, such as toxin/anti-toxin systems, to enhance the equal partitioning over daughter cells [241]. While plasmid copy-number control systems maintain a constant DNA density, they do not ensure simultaneous replication at a constant cell volume. The lack of coupling DNA replication initiation to cell division, appears to hamper the potential of these plasmid control systems as regulators in a synthetic cell cycle.

6.5. INTEGRATION AND COMPATIBILITY OF A SYNTHETIC CELL CYCLE

To conclude this perspective, we will discuss a few potential routes towards building a synthetic cell cycle (Fig. 6.6). The presented routes have been selected on basis of the aforementioned insights of individual module variants, as well as on considerations on their functional compatibility. First, the presence of a container is essential to define the context of the synthetic cell. Being at the centre of extensive research throughout the years [144], inverted emulsion-generated liposomes of a defined size (1-5 μm diameter), and composed of selected bacterial lipids are considered the best choice to fulfil this role.

As to the chromosome configuration, we propose to include all genetic information on a single, circular or linear dsDNA chromosome (total 250-500 kb). Because of its simplicity and its reported in vitro performance [145, 147], the ϕ29 system appears very attractive for DNA replication, although, as discussed, its processivity and its preference for linear chromosomes may require optimisation to cope with replication of chromosomes up to 500 kb. In addition, implementing a sizer-like modulation of replication activity (see Appendix 6.A and Fig. 6.7), would require including an additional control module, for which the natural *E. coli* DnaA-mediated regulation of replication initiation could serve as inspiration.

Another important design criterium to be considered concerns the copy number of the chromosome. Employing multiple copies of the chromosome means that a plasmid-like copy number control mechanism, together with random partitioning, could be sufficient for dividing the chromosomes over the daughter cells. However, as simultaneous initiation of replication of all the copies at a certain volume could not be easily controlled in this scenario, division cannot be coupled to replication via a timer. Thus, an additional control module should be implemented to ensure cell size homeostasis through generations. Another important consequence of multiple chromosomes would be that symme-

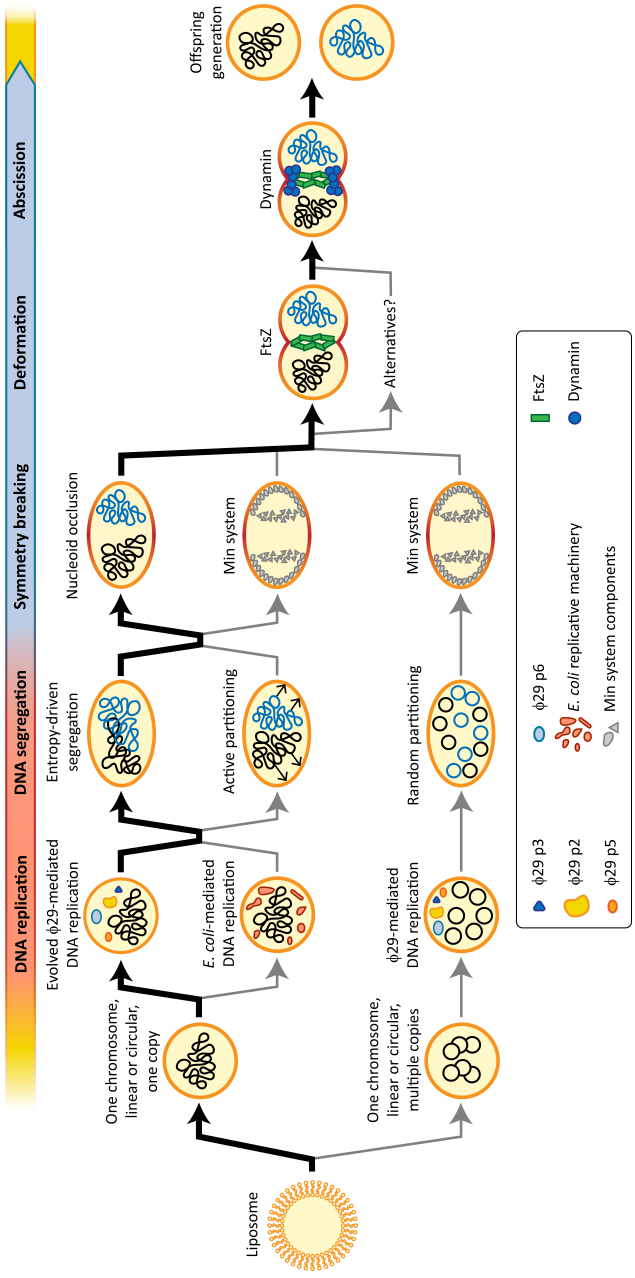


Figure 6.6: Towards a synthetic cell cycle. Various possible routes are indicated to accomplish a sustainable synthetic cell cycle. The pathway that, in our opinion, is most promising based on current insights, is marked in bold black arrows. Alternative paths are depicted in thin grey arrows. Note that not all possible alternative paths are displayed and that we only depicted circular chromosomes here. However, a linear chromosome architecture could also be a viable option, especially in combination with the ϕ29 machinery.

try breaking cannot operate through nucleoid occlusion. Hence, a more appealing alternative is to only use a single copy of the chromosome, and to ensure a sufficient degree of confinement by generating liposomes of appropriate dimensions. Importantly, this genomic configuration would allow for employing entropy-driven segregation. Compared to alternatives, this mechanism would considerably reduce the molecular complexity of the synthetic cell. If entropy-driven chromosome segregation alone proves to be insufficient to break the cell symmetry, implementation of the Min system could be considered.

As nucleoid occlusion prevents assembly of membrane-deforming machineries close to the chromosome [161], the cell division process appears to initiate only after full partitioning occurred, thus achieving an intrinsic timer behaviour in synthetic cells that employ a sufficiently confined chromosome. Subsequently, FtsZ can assemble the Z-ring at the cell midpoint to induce membrane deformation. FtsZ is to be initially preferred over other systems due to the natural connection with other well-known constriction and abscission mechanisms, although it still remains to be confirmed whether the Z-ring alone is able to constrict a membrane. Finally, bacterial dynamin can be recruited at the Z-ring and promote membrane fission to complete the cell division. This process then may be expected to generate two daughter cells, ready to undergo a new round of the synthetic cell cycle. Integrating a sizer for replication initiation and a timer for division is considered a promising combination that might give rise to robust cell cycles in a synthetic cell.

Apart from the modular routes proposed here (Fig. 6.6), several alternative sets of modules could be proposed. Despite individual preferences, we believe that the interconnected nature of the processes taking part in the cell cycle should be considered as a key feature when developing a synthetic cell. Smart design and tuning by evolution eventually should provide a robust and well-integrated synthetic cell cycle that will be a major step towards a synthetic cell that grows and divides autonomously.

ACKNOWLEDGEMENTS

The authors would like to thank Raymond Staals and many members of the BaSyC consortium for inspiring discussions and feedback on the topic of this Perspective. All authors acknowledge financial support from The Netherlands Organization of Scientific Research (NWO/OCW) Gravitation program Building A Synthetic Cell (BaSyC) (024.003.019). N.J.C. acknowledges support from an NWO-Veni Fellowship (VI.Veni.192.156). C.D. acknowledges the ERC Advanced Grant Looping DNA (no. 883684) and the NWO NanoFront program.

APPENDIX

6.A. MODELS FOR CELL SIZE AND DNA DENSITY CONTROL IN BACTERIA

All organisms need to maintain a stable average cell size in a variety of different growth conditions. Symmetrically dividing cells must, on average, double their volume once per cell cycle, such that the average division volume $\langle V_d \rangle$ is twice the average birth volume $\langle V_b \rangle$. In the presence of biological noise (e.g. non-symmetrical cell division), cells can be born bigger or smaller than average ($V_b = \langle V_b \rangle + \delta V$). In order to maintain cell size homeostasis, it is imperative to reduce deviations in birth size over the course of the cell cycle and over several generations.

Fig. 6.7a shows how the size at division, V_d, depends on the size at birth, V_b, in three different phenomenological models of cell size homeostasis in exponentially growing cells. In the "sizer" model, the division volume is independent of the birth volume and equals twice the average birth volume ($V_d = 2\langle V_b \rangle$). Therefore, the "sizer" reduces noise in the birth volume within one generation only, providing the optimal division size control. In the "timer" model, cells divide a constant time after birth. In exponentially growing cells, the birth volume V_b is doubled after a constant time τ_d, set by the growth rate of the cell. Fluctuations in the birth volume are therefore not reduced upon cell division. In fact, in the entire unstable region shown in Fig. 6.7a, cell size homeostasis is not maintained, because initial variations in the birth volume are further amplified instead of reduced. Thus, the "timer" does not ensure a stable cell size. The "adder" behaviour, based on more recent single cell measurements, reduces fluctuations in the birth volume by adding an on average constant volume $\langle V_b \rangle$ from birth to division. An initial deviation in the birth volume δV is reduced by a factor of two in every generation, thus ensuring cell size homeostasis over generations (Fig. 6.7a).

In addition to cell size control, cells must also maintain a constant chromosome density and couple DNA replication to cell division. It has been suggested that *E. coli* achieves this by initiating replication at a constant volume per number of origins of replication, and dividing an approximately constant time later (Fig. 6.7b). This combination of a sizer for DNA replication and a timer for cell division reduces fluctuations in the birth volume on the level of replication initiation and a stable cell size is maintained at all growth conditions. Additionally, this simple mechanism ensures that every replication initiation event is followed by a cell division, thus also coupling replication to division. The biochemical networks regulating the timing of replication initiation and cell division are often very complex. Until today, it remains poorly understood what molecular mechanisms give rise to the phenomenological observations, even in well-studied bacteria such as *E. coli*. These phenomenological models, however, provide a good framework for designing regulation modules that ensure a robust cell cycle of a synthetic cell.

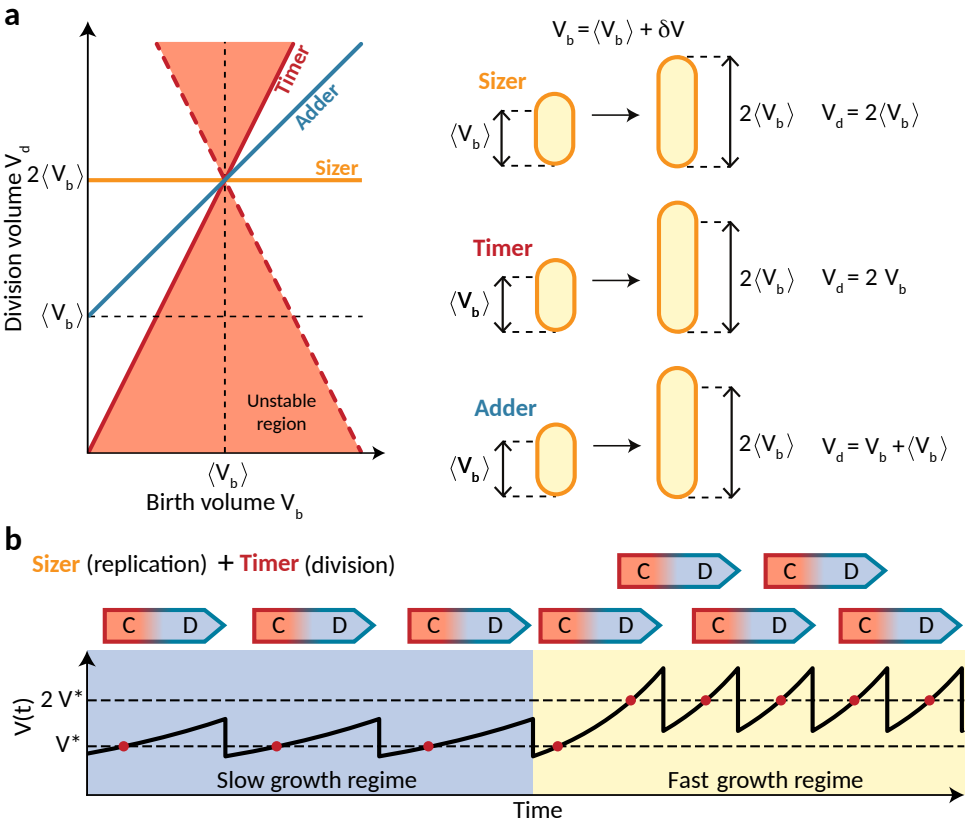


Figure 6.7: Phenomenological models for cell size and DNA density control in bacteria. (a) The plot on the left shows the division volume V_d as a function of the birth volume V_b for different phenomenological models of cell size control. The schemes on the right show cells at birth and at division for the same phenomenological models. In the sizer (yellow line), the division volume is independent of the birth volume. The adder (blue line) adds an on average constant volume from birth to division. In the timer (red line), the cell divides after a fixed time. Contrary to the sizer and the adder, the timer does not reduce fluctuations in the birth size and can therefore not ensure cell size homeostasis. (b) The volume of a cell as a function of time at slow (blue shaded area) and fast growth rates (yellow shaded area). By combining a sizer for replication initiation at a critical volume per number of origins (horizontal dashed lines) and dividing a constant time later (C+D period, marked as arrows), a cell can both ensure that every daughter cell obtains exactly one chromosome and that cell size homeostasis is maintained at all growth rates.

REFERENCES

[1] M. Dogterom, *Gravitation application form: Building a Synthetic Cell,* (2016).

[2] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson, *Molecular Biology of the Cell*, 4th ed. (Garland, 2002).

[3] S. Iyer-Biswas, C. S. Wright, J. T. Henry, K. Lo, S. Burov, Y. Lin, G. E. Crooks, S. Crosson, A. R. Dinner, and N. F. Scherer, *Scaling laws governing stochastic growth and division of single bacterial cells,* Proceedings of the National Academy of Sciences of the United States of America **111**, 15912 (2014).

[4] I. Soifer, L. Robert, and A. Amir, *Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy,* Current Biology **26**, 356 (2016), arXiv:1410.4771 .

[5] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, *A constant size extension drives bacterial cell size homeostasis,* Cell **159**, 1433 (2014), arXiv:NIHMS150003 .

[6] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, *Cell-size control and homeostasis in bacteria,* Current Biology **25**, 385 (2015).

[7] Y.-J. Eun, P.-Y. Ho, M. Kim, S. LaRussa, L. Robert, L. D. Renner, A. Schmid, E. Garner, and A. Amir, *Archaeal cells share common size control with bacteria despite noisier growth and division,* Nature Microbiology **3**, 148 (2018).

[8] A. Amir, *Cell size regulation in bacteria,* Physical Review Letters **112**, 1 (2014), arXiv:1312.6562 .

[9] F. Barber, P. Y. Ho, A. W. Murray, and A. Amir, *Details matter: Noise and model structure set the relationship between cell size and cell cycle timing,* Frontiers in Cell and Developmental Biology **5**, 1 (2017), arXiv:1710.06486 .

[10] P.-Y. Ho and A. Amir, *Simultaneous Regulation of Cell Size and Chromosome Replication in Bacteria,* Frontiers in microbiology **6**, 662 (2015).

[11] L. Sompayrac and O. Maaloe, *Autorepressor Model for Control of DNA Replication,* Nature New Biology **241**, 133 (1973).

[12] F. G. Hansen, B. B. Christensen, and T. Atlung, *The initiator titration model: computer simulation of chromosome and minichromosome control,* Research in Microbiology **142**, 161 (1991).

[13] M. Basan, M. Zhu, X. Dai, M. Warren, D. Sévin, Y. Wang, and T. Hwa, *Inflating bacterial cells by increased protein synthesis,* Molecular Systems Biology **11**, 836 (2015).

[14] F. G. Hansen and T. Atlung, *The DnaA tale,* Frontiers in Microbiology **9**, 1 (2018).

[15] Z. D. Blount, *The natural history of model organisms: The unexhausted potential of E. coli,* eLife **4**, e05826 (2015).

[16] O. Tenaillon, D. Skurnik, B. Picard, and E. Denamur, *The population genetics of commensal escherichia coli,* Nature Reviews Microbiology **8**, 207 (2010).

[17] T. Katayama, S. Ozaki, K. Keyamura, and K. Fujimitsu, *Regulation of the replication cycle: Conserved and diverse regulatory systems for DnaA and oriC,* Nature Reviews Microbiology **8**, 163 (2010).

[18] C. Speck and W. Messer, *Mechanism of origin unwinding: Sequential binding of DnaA to double- and single-stranded DNA,* EMBO Journal **20**, 1469 (2001).

[19] M. Schaechter, O. Maaloe, and N. O. Kjeldgaard, *Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium,* Journal of General Microbiology **19**, 592 (1958).

[20] S. Cooper and C. E. Helmstetter, *Chromosome replication and the division cycle of Escherichia coliBr,* Journal of Molecular Biology **31**, 519 (1968).

[21] W. D. Donachie, *Relationship between Cell Size and Time of Initiation of DNA Replication,* Nature **219**, 1077 (1968).

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

[23] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, *Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria,* Current Biology **29**, 1760 (2019).

[24] G. Witz, E. van Nimwegen, and T. Julou, *Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism,* eLife **8**, e48063 (2019).

[25] G. Le Treut, F. Si, D. Li, and S. Jun, *Comment on 'Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism',* (2020), https://www.biorxiv.org/content/early/2020/05/11/2020.05.08.084376.full.pdf .

[26] G. Witz, T. Julou, and E. van Nimwegen, *Response to comment on 'Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism',* (2020).

[27] G. Micali, J. Grilli, J. Marchi, M. Osella, and M. Cosentino Lagomarsino, *Dissecting the Control Mechanisms for DNA Replication and Cell Division in E. coli,* Cell Reports **25**, 761 (2018).

[28] G. Micali, J. Grilli, M. Osella, and M. C. Lagomarsino, *Concurrent processes set E. coli cell division,* Science Advances **4**, 1 (2018).

[29] A. Adiciptaningrum, M. Osella, M. C. Moolman, M. Cosentino Lagomarsino, and S. J. Tans, *Stochasticity and homeostasis in the E. coli replication and division cycle,* Scientific Reports **5**, 18261 (2015).

[30] O. Maaløe and N. O. Kjeldgaard, *Control of macromolecular synthesis : a study of DNA, RNA, and protein synthesis in bacteria* (New York (N.Y.) : Benjamin, 1966).

[31] C. E. Helmstetter and S. Cooper, *DNA synthesis during the division cycle of rapidly growing Escherichia coli Br,* Journal of Molecular Biology **31**, 507 (1968).

[32] L. Willis and K. C. Huang, *Sizing up the bacterial cell cycle,* Nature Reviews Microbiology **15**, 606 (2017).

[33] L. Dewachter, N. Verstraeten, M. Fauvart, and J. Michiels, *An integrative view of cell cycle control in Escherichia coli,* FEMS Microbiology Reviews **42**, 116 (2018).

[34] S. Ghatak, Z. A. King, A. Sastry, and B. O. Palsson, *The y-ome defines the 35coli genes that lack experimental evidence of function,* Nucleic Acids Research **47**, 2446 (2019).

[35] J. Westberg, A. Persson, A. Holmberg, A. Goesmann, J. Lundeberg, K.-E. Johansson, B. Pettersson, and M. Uhlén, *The genome sequence of mycoplasma mycoides subsp. mycoides SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP),* Genome Res **14**, 221 (2004).

[36] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, and J. C. Venter, *Creation of a bacterial cell controlled by a chemically synthesized genome,* Science **329**, 52 (2010).

[37] C. A. Hutchison, R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, and J. C. Venter, *Design and synthesis of a minimal bacterial genome,* Science **351**, aad6253 (2016), https://www.science.org/doi/pdf/10.1126/science.aad6253 .

[38] J. F. Pelletier, L. Sun, K. S. Wise, N. Assad-Garcia, B. J. Karas, T. J. Deerinck, M. H. Ellisman, A. Mershin, N. Gershenfeld, R.-Y. Chuang, J. I. Glass, and E. A. Strychalski, *Genetic requirements for cell division in a genomically minimal cell,* Cell **0** (2021), 10.1016/j.cell.2021.03.008.

[39] M. Berger and P. R. t. Wolde, *Robust replication initiation from coupled homeostatic mechanisms,* Nature Communications **13**, 6556 (2022).

[40] F. Si, D. Li, S. E. Cox, J. T. Sauls, O. Azizi, C. Sou, A. B. Schwartz, M. J. Erickstad, Y. Jun, X. Li, and S. Jun, *Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli,* Current Biology **27**, 1278 (2017).

[41] H. Zheng, Y. Bai, M. Jiang, T. A. Tokuyasu, X. Huang, F. Zhong, Y. Wu, X. Fu, N. Kleckner, T. Hwa, and C. Liu, *General quantitative relations linking cell growth and the cell cycle in Escherichia coli,* Nature Microbiology **5**, 995 (2020).

[42] K. Skarstad and T. Katayama, *Regulating DNA replication in bacteria.* Cold Spring Harbor perspectives in biology **5**, a012922 (2013).

[43] T. Katayama, K. Kasho, and H. Kawakami, *The DnaA cycle in Escherichia coli: Activation, function and inactivation of the initiator protein,* Frontiers in Microbiology **8**, 1 (2017).

[44] L. Riber, J. Frimodt-Møller, G. Charbon, and A. Løbner-Olesen, *Multiple DNA Binding Proteins Contribute to Timing of Chromosome Replication in E. coli,* Frontiers in Molecular Biosciences **3**, 1 (2016).

[45] G. Le Treut, F. Si, D. Li, and S. Jun, *Quantitative examination of five stochastic cell-cycle and cell-size control models for escherichia coli and bacillus subtilis,* Frontiers in Microbiology **12**, 3278 (2021).

[46] S. Schaper and W. Messer, *Interaction of the Initiator Protein DnaA of Escherichia coli with Its DNA Target,* Journal of Biological Chemistry **270**, 17622 (1995), http://www.jbc.org/content/270/29/17622.full.pdf+html .

[47] J. Lin and A. Amir, *Homeostasis of protein and mRNA concentrations in growing cells,* Nature Communications **9** (2018), 10.1038/s41467-018-06714-z.

[48] M. Panlilio, J. Grilli, G. Tallarico, I. Iuliani, B. Sclavi, P. Cicuta, and M. C. Lagomarsino, *Threshold accumulation of a constitutive protein explains E. coli cell-division behavior in nutrient upshifts,* Proceedings of the National Academy of Sciences of the United States of America **118** (2021), 10.1073/pnas.2016391118.

[49] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, *Single-cell protein dynamics reproduce universal fluctuations in cell populations,* The European Physical Journal E **38**, 102 (2015).

[50] H. Kempe, A. Schwabe, F. Crémazy, P. J. Verschure, and F. J. Bruggeman, *The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise,* Molecular Biology of the Cell **26**, 797 (2015).

[51] O. Padovan-Merhar, G. Nair, A. Biaesch, A. Mayer, S. Scarfone, S. Foley, A. Wu, L. Churchman, A. Singh, and A. Raj, *Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms,* Molecular Cell **58**, 339 (2015).

[52] R. Ietswaart, S. Rosa, Z. Wu, C. Dean, and M. Howard, *Cell-Size-Dependent Transcription of FLC and Its Antisense Long Non-coding RNA COOLAIR Explain Cell-to-Cell Expression Variation,* Cell Systems **4**, 622 (2017).

[53] X.-y. Zheng and E. K. O'Shea, *Cyanobacteria Maintain Constant Protein Concentration despite Genome Copy-Number Variation,* CellReports **19**, 497 (2017).

[54] A. Roth and W. Messer, *High-affinity binding sites for the initiator protein DnaA on the chromosome of Escherichia coli,* Molecular Microbiology **28**, 395 (1998).

[55] J. Elf, G.-W. Li, and X. S. Xie, *Probing transcription factor dynamics at the single-molecule level in a living cell,* Science **316**, 1191 (2007), https://science.sciencemag.org/content/316/5828/1191.full.pdf .

[56] K. Schenk, A. B. Hervás, T. C. Rösch, M. Eisemann, B. A. Schmitt, S. Dahlke, L. Kleine-Borgmann, S. M. Murray, and P. L. Graumann, *Rapid turnover of dnaa at replication origin regions contributes to initiation control of dna replication,* PLOS Genetics **13**, 1 (2017).

[57] J. Paijmans and P. R. Ten Wolde, *Lower bound on the precision of transcriptional regulation and why facilitated diffusion can reduce noise in gene expression,* Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **90**, 1 (2014).

[58] J. L. Campbell and N. Kleckner, *E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork.* Cell **62**, 967 (1990).

[59] M. Lu, J. L. Campbell, E. Boye, and N. Kleckner, *SeqA: A negative modulator of replication initiation in E. coli,* Cell **77**, 413 (1994).

[60] T. Waldminghaus and K. Skarstad, *The Escherichia coli SeqA protein,* Plasmid **61**, 141 (2009).

[61] R. Kitagawa, H. Mitsuki, T. Okazaki, and T. Ogawa, *A novel DnaA protein-binding site at 94.7 min on the Escherichia coll chromosome,* Molecular Microbiology **19**, 1137 (1996).

[62] F. Blaesing, C. Weigel, M. Welzeck, and W. Messer, *Analysis of the DNA-binding domain of Escherichia coli DnaA protein,* Molecular Microbiology **36**, 557 (2000).

[63] K. Kasho and T. Katayama, *DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation,* Proceedings of the National Academy of Sciences **110**, 936 (2013), https://www.pnas.org/content/110/3/936.full.pdf .

[64] F. G. Hansen, T. Atlung, R. E. Braun, A. Wright, P. Hughes, and M. Kohiyama, *Initiator (DnaA) protein concentration as a function of growth rate in Escherichia coli and Salmonella typhimurium,* Journal of Bacteriology **173**, 5194 (1991).

[65] S. Klumpp, Z. Zhang, and T. Hwa, *Growth Rate-Dependent Global Effects on Gene Expression in Bacteria,* Cell **139**, 1366 (2009).

[66] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, *Interdependence of cell growth and gene expression: origins and consequences.* Science (New York, N.Y.) **330**, 1099 (2010).

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

[68] A. Knöppel, O. Broström, K. Gras, D. Fange, and J. Elf, *The coordination of replication initiation with growth rate in escherichia coli,* bioRxiv (2022), 10.1101/2021.10.11.463968, https://www.biorxiv.org/content/early/2022/06/12/2021.10.11.463968.full.pdf .

[69] C. Speck, C. Weigel, and W. Messer, *ATP- and ADP-DnaA protein, a molecular switch in gene regulation,* EMBO Journal **18**, 6169 (1999).

[70] C. Saggioro, A. Olliver, and B. Sclavi, *Temperature-dependence of the DnaA–DNA interaction and its effect on the autoregulation of dnaA expression,* Biochemical Journal **449**, 333 (2012), https://portlandpress.com/biochemj/article-pdf/449/2/333/674788/bj4490333.pdf .

[71] J. Paulsson, *Models of stochastic gene expression,* Physics of Life Reviews **2**, 157 (2005).

[72] M. Thattai and A. Van Oudenaarden, *Intrinsic noise in gene regulatory networks,* Proceedings of the National Academy of Sciences of the United States of America **98**, 8614 (2001).

[73] N. Friedman, L. Cai, and X. S. Xie, *Linking stochastic dynamics to population distribution: An analytical framework of gene expression,* Physical Review Letters **97**, 1 (2006).

[74] V. Shahrezaei and P. S. Swain, *Analytical distributions for stochastic gene expression,* Proceedings of the National Academy of Sciences of the United States of America **105**, 17256 (2008), arXiv:0812.3344 .

[75] R. Milo, *What is the total number of protein molecules per cell volume? A call to rethink some published values,* BioEssays **35**, 1050 (2013).

[76] M. Wallden, D. Fange, Ö. Baltekin, and J. Elf, *Fluctuations in growth rates determine the generation time and size distributions of E. coli cells,* (2015).

[77] K. Kasho, K. Fujimitsu, T. Matoba, T. Oshima, and T. Katayama, *Timely binding of IHF and Fis to DARS2 regulates ATP-DnaA production and replication initiation,* Nucleic acids research **42**, 13134 (2014).

[78] K. Kurokawa, S. Nishida, A. Emoto, K. Sekimizu, and T. Katayama, *Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in Escherichia coli,* The EMBO Journal **18**, 6642 (1999), https://www.embopress.org/doi/pdf/10.1093/emboj/18.23.6642 .

[79] S. Nishida, K. Fujimitsu, K. Sekimizu, T. Ohmura, T. Ueda, and T. Katayama, *A Nucleotide Switch in the Escherichia coli DnaA Protein Initiates Chromosomal Replication,* The Journal of biological chemistry **277**, 14986 (2002).

[80] K. Keyamura, Y. Abe, M. Higashi, T. Ueda, and T. Katayama, *DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading,* Journal of Biological Chemistry **284**, 25038 (2009).

[81] K. Fujimitsu, T. Senriuchi, and T. Katayama, *Specific genomic sequences of E. coli promote replicational initiation by directly reactivating ADP-DnaA,* Genes and Development **23**, 1221 (2009).

[82] T. Katayama, K. Fujimitsu, and T. Ogawa, *Multiple pathways regulating DnaA function in Escherichia coli: Distinct roles for DnaA titration by the datA locus and the regulatory inactivation of DnaA,* Biochimie **83**, 13 (2001).

[83] M. Grant, C. Saggioro, U. Ferrari, B. Bassetti, B. Sclavi, and M. C. Lagomarsino, *DnaA and the timing of chromosome replication in Escherichia coli as a function of growth rate,* BMC Systems Biology **5**, 201 (2011).

[84] A. C. Leonard and J. E. Grimwade, *Regulation of DnaA Assembly and Activity: Taking Directions from the Genome,* Annual Review of Microbiology **65**, 19 (2011), pMID: 21639790, https://doi.org/10.1146/annurev-micro-090110-102934 .

[85] W. D. Donachie and G. W. Blakely, *Coupling the initiation of chromosome replication to cell size in Escherichia coli,* (2003).

[86] K. Sekimizu and A. Kornberg, *Cardiolipin activation of dnaA protein, the initiation protein of replication in Escherichia coli,* Journal of Biological Chemistry **263**, 7131 (1988).

[87] J. I. Kato and T. Katayama, *Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli,* EMBO Journal **20**, 4253 (2001).

[88] K. Nakamura and T. Katayama, *Novel essential residues of Hda for interaction with DnaA in the regulatory inactivation of DnaA: Unique roles for Hda AAA⁺ Box VI and VII motifs,* Molecular Microbiology **76**, 302 (2010).

[89] T. Ogawa, Y. Yamada, T. Kuroda, T. Kishi, and S. Moriya, *The datA locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in Escherichia coli,* Molecular Microbiology **44**, 1367 (2002).

[90] J. E. Camara, A. M. Breier, T. Brendler, S. Austin, N. R. Cozzarelli, and E. Crooke, *Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of Escherichia coli DNA replication,* EMBO Reports **6**, 736 (2005).

[91] L. Riber, J. A. Olsson, R. B. Jensen, O. Skovgaard, S. Dasgupta, M. G. Marinus, and A. Løbner-Olesen, *Hda-mediated inactivation of the DnaA protein and dnaA gene autoregulation act in concert to ensure homeostatic maintenance of the Escherichia coli chromosome,* Genes and Development **20**, 2121 (2006).

[92] W. Xia and W. Dowhan, *In vivo evidence for the involvement of anionic phospholipids in initiation of DNA replication in Escherichia coli,* Proceedings of the National Academy of Sciences of the United States of America **92**, 783 (1995).

[93] R. Saxena, N. Fingland, D. Patil, A. K. Sharma, and E. Crooke, *Crosstalk between DnaA protein, the initiator of Escherichia coli chromosomal replication, and acidic phospholipids present in bacterial membranes,* International Journal of Molecular Sciences **14**, 8517 (2013).

[94] H. Kawakamii, K. Keyamura, and T. Katayama, *Formation of an ATP-DnaA-specific initiation complex requires DnaA arginine 285, a conserved motif in the AAA+ protein family,* Journal of Biological Chemistry **280**, 27420 (2005).

[95] P. Mallik, B. J. Paul, S. T. Rutherford, R. L. Gourse, and R. Osuna, *DksA Is Required for Growth Phase-Dependent Regulation, Growth Rate-Dependent Control, and Stringent Control of fis Expression in Escherichia coli,* Journal of Bacteriology **188**, 5775 (2006).

[96] I. Flåtten and K. Skarstad, *The Fis protein has a stimulating role in initiation of replication in Escherichia coli in vivo,* PLoS ONE **8**, 1 (2013).

[97] E. Crooke, C. E. Castuma, and A. Kornberg, *The chromosome origin of Escherichia coli stabilizes DnaA protein during rejuvenation by phospholipids,* Journal of Biological Chemistry **267**, 16779 (1992).

[98] C. E. Castuma, E. Crooke, and A. Kornberg, *Fluid membranes with acidic domains activate DnaA, the initiator protein of replication in Escherichia coli,* Journal of Biological Chemistry **268**, 24665 (1993).

[99] A. Aranovich, G. Y. Gdalevsky, R. Cohen-Luria, I. Fishov, and A. H. Parola, *Membrane-catalyzed nucleotide exchange on DnaA: Effect of surface molecular crowding,* Journal of Biological Chemistry **281**, 12526 (2006).

[100] J. Garner, P. Durrer, J. Kitchen, J. Brunner, and E. Crooke, *Membrane-mediated release of nucleotide from an initiator of chromosomal replication, Escherichia coli DnaA, occurs with insertion of a distinct region of the protein into the lipid bilayer,* Journal of Biological Chemistry **273**, 5167 (1998).

[101] B. Y. Yung and A. Kornberg, *Membrane attachment activates dnaA protein, the initiation protein of chromosome replication in Escherichia coli,* Proceedings of the National Academy of Sciences of the United States of America **85**, 7202 (1988).

[102] N. Fingland, I. Flåtten, C. D. Downey, S. Fossum-Raunehaug, K. Skarstad, and E. Crooke, *Depletion of acidic phospholipids influences chromosomal replication in Escherichia coli,* MicrobiologyOpen **1**, 450 (2012).

[103] W. Zheng, Z. Li, K. Skarstad, and E. Crooke, *Mutations in DnaA protein suppress the growth arrest of acidic phospholipid-deficient Escherichia coli cells,* EMBO Journal **20**, 1164 (2001).

[104] P. N. Heacock and W. Dowhan, *Alteration of the phospholipid composition of Escherichia coli through genetic manipulation,* Journal of Biological Chemistry **264**, 14972 (1989).

[105] T. Katayama, T. Kubota, K. Kurokawa, E. Crooke, and K. Sekimizu, *The initiator function of DnaA protein is negatively regulated by the sliding clamp of the E. coli Chromosomal replicase,* Cell **94**, 61 (1998).

[106] I. Flåtten, S. Fossum-Raunehaug, R. Taipale, S. Martinsen, and K. Skarstad, *The DnaA Protein Is Not the Limiting Factor for Initiation of Replication in Escherichia coli,* PLoS Genetics **11**, 1 (2015).

[107] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic gene expression in a single cell.* Science **297**, 1183 (2002).

[108] N. S. Hill, R. Kadoya, D. K. Chattoraj, and P. A. Levin, *Cell size and the initiation of DNA replication in bacteria,* PLoS Genetics **8**, 14 (2012).

[109] J. P. Erzberger, M. L. Mott, and J. M. Berger, *Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling.* Nature Structural & Molecular Biology **13**, 676 (2006).

[110] M. C. Moolman, S. T. Krishnan, J. W. Kerssemakers, A. van den Berg, P. Tulinski, M. Depken, R. Reyes-Lamothe, D. J. Sherratt, and N. H. Dekker, *Slow unloading leads to DNA-bound β2-sliding clamp accumulation in live Escherichia coli cells,* Nature communications **5**, 5820 (2014).

[111] Y. Shiba, Y. Yokoyama, Y. Aono, T. Kiuchi, J. Kusaka, K. Matsumoto, and H. Hara, *Activation of the Rcs signal transduction system is responsible for the thermosensitive growth defect of an Escherichia coli mutant lacking phosphatidylglycerol and cardiolipin,* Journal of Bacteriology **186**, 6526 (2004).

[112] Y. Shiba, H. Miyagawa, H. Nagahama, K. Matsumoto, D. Kondo, S. Matsuoka, K. Matsumoto, and H. Hara, *Exploring the relationship between lipoprotein mislocalization and activation of the Rcs signal transduction system in Escherichia coli,* Microbiology **158**, 1238 (2012).

[113] D. Camsund, M. J. Lawson, J. Larsson, D. Jones, S. Zikrin, D. Fange, and J. Elf, *Time-resolved imaging-based CRISPRi screening,* Nature Methods **17**, 86 (2020).

[114] C. C. Govern and P. R. ten Wolde, *Optimal resource allocation in cellular sensing systems,* Proceedings of the National Academy of Sciences of the United States of America **111**, 17486 (2014).

[115] M. D'Ario, R. Tavares, K. Schiessl, B. Desvoyes, C. Gutierrez, M. Howard, and R. Sablowski, *Cell size controlled in plants using DNA content as an internal scale,* Science **372**, 1176 (2021).

[116] N. Rhind, *Cell-size control,* Current Biology **31**, R1414 (2021).

[117] D. Zwicker, D. K. Lubensky, and P. R. t. Wolde, *Robust circadian clocks from coupled protein-modification and transcription–translation cycles,* Proceedings of the National Academy of Sciences **107**, 22540 (2010).

[118] S. W. Teng, S. Mukherji, J. R. Moffitt, S. de Buyl, and E. K. O'Shea, *Robust Circadian Oscillations in Growing Cyanobacteria Require Transcriptional Feedback,* Science **340**, 737 (2013).

[119] L. F. Larrondo, C. Olivares-Yanez, C. L. Baker, J. J. Loros, and J. C. Dunlap, *Decoupling circadian clock protein turnover from circadian period determination,* Science **347**, 1257277 (2015).

[120] T. Nishiwaki-Ohkawa, Y. Kitayama, E. Ochiai, and T. Kondo, *Exchange of ADP with ATP in the CII ATPase domain promotes autophosphorylation of cyanobacterial clock protein KaiC.* Proceedings of the National Academy of Sciences of the United States of America **111**, 4455 (2014).

[121] J. Paijmans, D. K. Lubensky, and P. R. ten Wolde, *A thermodynamically consistent model of the post-translational Kai circadian clock,* PLoS Computational Biology **13**, e1005415 (2017).

[122] R. Kitagawa, T. Ozaki, S. Moriya, and T. Ogawa, *Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein,* Genes and Development **12**, 3032 (1998).

[123] K. Fujimitsu, M. Su'etsugu, Y. Yamaguchi, K. Mazda, N. Fu, H. Kawakami, and T. Katayama, *Modes of overinitiation, dnaA gene expression, and inhibition of cell division in a novel cold-sensitive hda mutant of Escherichia coli,* Journal of Bacteriology **190**, 5368 (2008).

[124] J. Frimodt-Møller, G. Charbon, K. A. Krogfelt, and A. Løbner-Olesen, *Control regions for chromosome replication are conserved with respect to sequence and location among Escherichia coli strains,* Frontiers in Microbiology **6**, 1 (2015).

[125] S. Nozaki, Y. Yamada, and T. Ogawa, *Initiator titration complex formed at datA with the aid of IHF regulates replication timing in Escherichia coli,* Genes to Cells **14**, 329 (2009).

[126] J. Frimodt-Møller, G. Charbon, K. A. Krogfelt, and A. Løbner-Olesen, *DNA Replication Control Is Linked to Genomic Positioning of Control Regions in Escherichia coli,* PLoS Genetics **12**, 1 (2016).

[127] Y. Inoue, H. Tanaka, K. Kasho, K. Fujimitsu, T. Oshima, and T. Katayama, *Chromosomal location of the DnaA-reactivating sequence DARS2 is important to regulate timely initiation of DNA replication in Escherichia coli,* Genes to Cells **21**, 1015 (2016).

[128] B. B. Christensen, T. Atlung, and F. G. Hansen, *DnaA boxes are important elements in setting the initiation mass of Escherichia coli,* Journal of Bacteriology **181**, 2683 (1999).

[129] T. Atlung and F. G. Hansen, *Three distinct chromosome replication states are induced by increasing concentrations of DnaA protein in Escherichia coli,* Journal of Bacteriology **175**, 6537 (1993).

[130] I. H. Jain, V. Vijayan, and E. K. O'Shea, *Spatial ordering of chromosomes enhances the fidelity of chromosome partitioning in cyanobacteria,* Proceedings of the National Academy of Sciences of the United States of America **109**, 13638 (2012).

[131] J. T. Sauls, S. E. Cox, Q. Do, V. Castillo, Z. Ghulam-Jelani, and S. Jun, *Control of bacillus subtilis replication initiation during physiological transitions and perturbations,* mBio **10**, e02205 (2019), https://journals.asm.org/doi/pdf/10.1128/mBio.02205-19 .

[132] I. Santi, N. Dhar, D. Bousbaine, Y. Wakamoto, and J. D. McKinney, *Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria,* Nature Communications **4**, 3 (2013).

[133] E. S. Egan, A. Løbner-Olesen, and M. K. Waldor, *Synchronous replication initiation of the two Vibrio cholerae chromosomes,* Current Biology **14**, 501 (2004).

[134] K. Skarstad, E. Boye, and H. B. Steen, *Timing of initiation of chromosome replication in individual Escherichia coli cells.* The EMBO journal **5**, 1711 (1986).

[135] T. Boesen, G. Charbon, H. Fu, C. Jensen, D. Li, S. Jun, and A. Lobner-Olesen, *Robust control of replication initiation in the absence of dnaa-atp/dnaa-adp regulatory elements in escherichia coli,* bioRxiv (2022), 10.1101/2022.09.08.507175, https://www.biorxiv.org/content/early/2022/09/08/2022.09.08.507175.full.pdf .

[136] S. T. Liang, Y. C. Xu, P. Dennis, and H. Bremer, *mRNA composition and control of bacterial gene expression,* J Bacteriol **182**, 3037 (2000).

[137] A. Løbner-Olesen, F. G. Hansen, K. V. Rasmussen, B. Martin, and P. L. Kuempel, *The initiation cascade for chromosome replication in wild-type and Dam methyltransferase deficient Escherichia coli cells,* EMBO Journal **13**, 1856 (1994).

[138] C. A. Hutchison, R. Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z. Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, and J. C. Venter, *Design and synthesis of a minimal bacterial genome,* Science **351**, 6280 (2016).

[139] H. R. Sikkema, B. F. Gaastra, T. Pols, and B. Poolman, *Cell Fuelling and Metabolic Energy Conservation in Synthetic Cells,* (2019).

[140] N. J. Gaut and K. P. Adamala, *Reconstituting Natural Cell Elements in Synthetic Cells,* Advanced Biology , 2000188 (2021).

[141] Z. Abil and C. Danelon, *Roadmap to Building a Cell: An Evolutionary Approach,* Frontiers in Bioengineering and Biotechnology **8**, 927 (2020).

[142] G. Zubaite, K. Simutis, R. Galinis, V. Milkus, V. Kiseliovas, and L. Mazutis, *Droplet Microfluidics Approach for Single-DNA Molecule Amplification and Condensation into DNA-Magnesium-Pyrophosphate Particles,* Micromachines **8**, 62 (2017).

[143] B. Drobot, J. M. Iglesias-Artola, K. L. Vay, V. Mayr, M. Kar, M. Kreysing, H. Mutschler, and T.-Y. D. Tang, *Compartmentalised RNA catalysis in membrane-free coacervate protocells,* Nature Communications **9**, 3643 (2018).

[144] E. Rideau, R. Dimova, P. Schwille, F. R. Wurm, and K. Landfester, *Liposomes and polymersomes: a comparative review towards cell mimicking,* Chemical Society reviews **47**, 8572 (2018).

[145] P. van Nies, I. Westerlaken, D. Blanken, M. Salas, M. Mencía, and C. Danelon, *Self-replication of dna by its encoded proteins in liposome-based synthetic cells,* Nature Communications **9**, 1583 (2018).

[146] Y. Sakatani, N. Ichihashi, Y. Kazuta, and T. Yomo, *A transcription and translation-coupled dna replication system using rolling-circle replication,* Scientific Reports **5**, 10404 (2015).

[147] K. Libicher, R. Hornberger, M. Heymann, and H. Mutschler, *In vitro self-replication and multicistronic expression of large synthetic genomes,* Nature Communications **11**, 904 (2020).

[148] T. Pols, H. R. Sikkema, B. F. Gaastra, J. Frallicciardi, W. M. Śmigiel, S. Singh, and B. Poolman, *A synthetic metabolic network for physicochemical homeostasis,* Nature Communications **10**, 1 (2019).

[149] C. Xu, S. Hu, and X. Chen, *Artificial cells: from basic science to applications,* (2016).

[150] D. Blanken, D. Foschepoth, A. C. Serrão, and C. Danelon, *Genetically controlled membrane synthesis in liposomes,* Nature Communications **11**, 1 (2020).

[151] A. Bhattacharya, R. J. Brea, H. Niederholtmeyer, and N. K. Devaraj, *A minimal biochemical route towards de novo formation of synthetic phospholipid membranes,* Nature Communications **10**, 1 (2019).

[152] Y. Kuruma, P. Stano, T. Ueda, and P. L. Luisi, *A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells,* Biochimica et Biophysica Acta - Biomembranes **1788**, 567 (2009).

[153] K. Gerdes, M. Howard, and F. Szardenings, *Pushing and Pulling in Prokaryotic DNA Segregation,* Cell **141**, 927 (2010).

[154] D. Huh and J. Paulsson, *Random partitioning of molecules at cell division,* Proceedings of the National Academy of Sciences **108**, 15004 (2011).

[155] S. Jun and B. Mulder, *Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome,* Proceedings of the National Academy of Sciences **103**, 12388 (2006).

[156] S. Jun and A. Wright, *Entropy as the driver of chromosome segregation,* Nat Rev Microbiol **8**, 600 (2010).

[157] M. Garten, S. Aimon, P. Bassereau, and G. E. Toombes, *Reconstitution of a transmembrane protein, the voltage-gated ion channel, KvAP, into giant unilamellar vesicles for microscopy and patch clamp studies,* Journal of Visualized Experiments , 52281 (2015).

[158] J. Nance and J. A. Zallen, *Elaborating polarity: PAR proteins and the cytoskeleton,* Development **138**, 799 (2011).

[159] K. J. A. Vendel, S. Tschirpke, F. Shamsi, M. Dogterom, and L. Laan, *Minimal in vitro systems shed light on cell polarity,* Journal of Cell Science **132**, jcs217554 (2019).

[160] S. F. Shimobayashi, M. Ichikawa, and T. Taniguchi, *Direct observations of transition dynamics from macro- to micro-phase separation in asymmetric lipid bilayers induced by externally added glycolipids,* EPL **113**, 56005 (2016).

[161] L. J. Wu and J. Errington, *Nucleoid occlusion and bacterial cell division,* Nature Reviews Microbiology **10**, 8 (2012).

[162] E. Loiseau, J. A. M. Schneider, F. C. Keber, C. Pelzl, G. Massiera, G. Salbreux, and A. R. Bausch, *Shape remodeling and blebbing of active cytoskeletal vesicles,* Science Advances **2**, e1500465 (2016).

[163] F.-C. Tsai and G. H. Koenderink, *Shape control of lipid bilayer membranes by confined actin bundles,* Soft Matter **11**, 8834 (2015).

[164] S. Tanaka, K. Takiguchi, and M. Hayashi, *Repetitive stretching of giant liposomes utilizing the nematic alignment of confined actin,* Communications Physics **1**, 1 (2018).

[165] Y. Sato, Y. Hiratsuka, I. Kawamata, S. Murata, and S.-i. M. Nomura, *Micrometer-sized molecular robot changes its shape in response to signal molecules,* Science Robotics **2**, eaal3735 (2017).

[166] K. L. Weirich, K. Dasbiswas, T. A. Witten, S. Vaikuntanathan, and M. L. Gardel, *Self-organizing motors divide active liquid droplets,* Proceedings of the National Academy of Sciences **116**, 11125 (2018).

[167] M. G. F. Last, S. Deshpande, and C. Dekker, *pH-Controlled Coacervate-Membrane Interactions within Liposomes,* ACS Nano **14**, 4487 (2020).

[168] K. A. Ganzinger, A. Merino-Salomon, D. A. Garcia-Soriano, A. N. Butterfield, T. Litschel, F. Siedler, and P. Schwille, *Ftsz reorganization facilitates deformation of giant vesicles in microfluidic traps,* Angewandte Chemie International Edition **59**, 21372 (2020), https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202001928 .

[169] J. Schöneberg, M. R. Pavlin, S. Yan, M. Righini, I.-H. Lee, L.-A. Carlson, A. H. Bahrami, D. H. Goldman, X. Ren, G. Hummer, C. Bustamante, and J. H. Hurley, *ATP-dependent force generation and membrane scission by ESCRT-III and Vps4,* Science **362**, 1423 (2018).

[170] S. Schlimpert, S. Wasserstrom, G. Chandra, M. J. Bibb, K. C. Findlay, K. Flärdh, and M. J. Buttner, *Two dynamin-like proteins stabilize FtsZ rings during Streptomyces sporulation,* Proceedings of the National Academy of Sciences **114**, E6176 (2017).

[171] A. C. Forster and G. M. Church, *Towards synthesis of a minimal cell,* (2006).

[172] K. Fujiwara, T. Katayama, and S. I. M. Nomura, *Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system,* Nucleic Acids Research **41**, 7176 (2013).

[173] M. Su'etsugu, H. Takada, T. Katayama, and H. Tsujimoto, *Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle,* Nucleic Acids Research **45**, 11525 (2017).

[174] Y. Schaerli, V. Stein, M. M. Spiering, S. J. Benkovic, C. Abell, and F. Hollfelder, *Isothermal DNA amplification using the T4 replisome: Circular nicking endonuclease-dependent amplification and primase-based whole-genome amplification,* Nucleic Acids Research **38**, e201 (2010).

[175] D. Hürtgen, J. Mascarenhas, M. Heymann, S. M. Murray, P. Schwille, and V. Sourjik, *Reconstitution and Coupling of DNA Replication and Segregation in a Biomimetic System,* ChemBioChem **20**, 2633 (2019).

[176] M. Salas, I. Holguera, M. Redrejo-Rodríguez, and M. de Vega, *DNA-binding proteins essential for protein-primed bacteriophage Φ29 DNA replication,* (2016).

[177] J. G. Paez, M. Lin, R. Beroukhim, J. C. Lee, X. Zhao, D. J. Richter, S. Gabriel, P. Herman, H. Sasaki, D. Altshuler, C. Li, M. Meyerson, and W. R. Sellers, *Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification.* Nucleic acids research **32**, e71 (2004).

[178] L. Blanco, A. Bernad, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas, *Highly efficient DNA synthesis by the phage Φ29 DNA polymerase. Symmetrical mode of DNA replication,* Journal of Biological Chemistry **264**, 8935 (1989).

[179] Y. Sakatani, T. Yomo, and N. Ichihashi, *Self-replication of circular dna by a self-encoded dna polymerase through rolling-circle replication and recombination,* Scientific Reports **8**, 13089 (2018).

[180] Y. Sakatani, R. Mizuuchi, and N. Ichihashi, *In vitro evolution of phi29 DNA polymerases through compartmentalized gene expression and rolling-circle replication,* Protein Eng Des Sel **32**, 481 (2019).

[181] L. A. Simmons, A. M. Breier, N. R. Cozzarelli, and J. M. Kaguni, *Hyperinitiation of DNA replication in Escherichia coli leads to replication fork collapse and inviability,* Molecular Microbiology **51**, 349 (2004).

[182] D. C. Bouck, A. P. Joglekar, and K. S. Bloom, *Annual Review of Genetics*, Vol. 42 (NIH Public Access, 2008) pp. 335–359.

[183] K. Nordstrom and S. J. Austin, *Mechanisms that Contribute to the Stable Segregation of Plasmids,* Annual Review of Genetics **23**, 37 (1989).

[184] M. Vleugel, S. Roth, C. F. Groenendijk, and M. Dogterom, *Reconstitution of Basic Mitotic Spindles in Spherical Emulsion Droplets,* Journal of Visualized Experiments **114** (2016), 10.3791/54278.

[185] R. Ietswaart, F. Szardenings, K. Gerdes, and M. Howard, *Competing ParA Structures Space Bacterial Plasmids Equally over the Nucleoid,* PLoS Computational Biology **10**, e1004009 (2014).

[186] C. L. Woldringh, *The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation,* Molecular Microbiology **45**, 17 (2002).

[187] C. L. Woldringh, F. G. Hansen, N. O. E. Vischer, and T. Atlung, *Segregation of chromosome arms in growing and non-growing Escherichia coli cells,* Frontiers in Microbiology **6**, 448 (2015).

[188] K. Lemon and A. Grossman, *The extrusion-capture model for chromosome partitioning in bacteria,* Genes & development **15**, 2031 (2001).

[189] E. P. C. Rocha, *Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes?* Trends in Microbiology **10**, 393 (2002).

[190] Z. Gitai, M. Thanbichler, and L. Shapiro, *The choreographed dynamics of bacterial chromosomes,* Trends in Microbiology **13**, 221 (2005).

[191] P. G. De Gennes, *Physics Today*, Vol. 33 (1979) p. 324.

[192] B.-Y. Ha and Y. Jung, *Polymers under confinement: single polymers, how they interact, and as model chromosomes,* Soft Matter **11**, 2333 (2015).

[193] E. Minina and A. Arnold, *Entropic Segregation of Ring Polymers in Cylindrical Confinement,* Macromolecules **48**, 4998 (2015).

[194] M. Bohn and D. W. Heermann, *Topological interactions between ring polymers: Implications for chromatin loops,* The Journal of Chemical Physics **132**, 44904 (2010).

[195] T. J. Krogh, J. Møller-Jensen, and C. Kaleta, *Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes,* Frontiers in Microbiology **9**, 2019 (2018).

[196] J. Shin, A. G. Cherstvy, and R. Metzler, *Mixing and segregation of ring polymers: spatial confinement and molecular crowding effects,* New Journal of Physics **16**, 53047 (2014).

[197] A. Japaridze, C. Gogou, J. W. J. Kerssemakers, H. M. Nguyen, and C. Dekker, *Direct observation of independently moving replisomes in escherichia coli,* Nature Communications **11**, 3109 (2020).

[198] L. J. Wu, S. Lee, S. Park, L. E. Eland, A. Wipat, S. Holden, and J. Errington, *Geometric principles underlying the proliferation of a model cell system,* Nature Communications **11**, 4149 (2020).

[199] T. Litschel, B. Ramm, R. Maas, M. Heymann, and P. Schwille, *Beating Vesicles: Encapsulated Protein Oscillations Cause Dynamic Membrane Deformations,* Angewandte Chemie International Edition **57**, 16286 (2018).

[200] E. Godino, J. N. López, D. Foschepoth, C. Cleij, A. Doerr, C. F. Castellà, and C. Danelon, *De novo synthesized min proteins drive oscillatory liposome deformation and regulate ftsa-ftsz cytoskeletal patterns,* Nature Communications **10**, 4969 (2019).

[201] B. Ramm, T. Heermann, and P. Schwille, *The E. coli MinCDE system in the regulation of protein patterns and gradients, Cellular and Molecular Life Sciences,* **76**, 4245 (2019).

[202] B. Ramm, P. Glock, J. Mucksch, P. Blumhardt, D. A. Garcia-Soriano, M. Heymann, and P. Schwille, *The MinDE system is a generic spatial cue for membrane protein distribution in vitro,* Nature Communications **9**, 3942 (2018).

[203] Y. Dreher, K. Jahnke, E. Bobkova, J. P. Spatz, and K. Göpfrich, *Division and Regrowth of Phase-Separated Giant Unilamellar Vesicles,* Angewandte Chemie - International Edition **60**, 10661 (2021).

[204] E. Godino, J. N. López, I. Zarguit, A. Doerr, M. Jimenez, G. Rivas, and C. Danelon, *Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes,* Communications Biology **3**, 539 (2020).

[205] H. G. Franquelim, A. Khmelinskaia, J.-P. Sobczak, H. Dietz, and P. Schwille, *Membrane sculpting by curved DNA origami scaffolds,* Nature Communications **9**, 811 (2018).

[206] M. Hayashi, M. Nishiyama, Y. Kazayama, T. Toyota, Y. Harada, and K. Takiguchi, *Reversible Morphological Control of Tubulin-Encapsulating Giant Liposomes by Hydrostatic Pressure,* Langmuir **32**, 3794 (2016).

[207] S. Deshpande, W. K. Spoelstra, M. van Doorn, J. Kerssemakers, and C. Dekker, *Mechanical Division of Cell-Sized Liposomes,* ACS Nano **12**, 2560 (2018).

[208] H. G. Franquelim, H. Dietz, and P. Schwille, *Reversible membrane deformations by straight DNA origami filaments,* Soft Matter (2020), 10.1039/d0sm00150c.

[209] K. Y. Lee, S. J. Park, K. A. Lee, S. H. Kim, H. Kim, Y. Meroz, L. Mahadevan, K. H. Jung, T. K. Ahn, K. K. Parker, and K. Shin, *Photosynthetic artificial organelles sustain and control ATP-dependent reactions in a protocellular system,* Nature Biotechnology **36**, 530 (2018).

[210] H. R. Vutukuri, M. Hoore, C. Abaurrea-Velasco, L. van Buren, A. Dutto, T. Auth, D. A. Fedosov, G. Gompper, and J. Vermant, *Active particles induce large shape deformations in giant lipid vesicles,* Nature **586**, 52 (2020).

[211] E. C. Garner, C. S. Campbell, D. B. Weibel, and R. D. Mullins, *Reconstitution of DNA Segregation Driven by Assembly of a Prokaryotic Actin Homolog,* Science **315**, 1270 (2007).

[212] G. Tarrason Risa, F. Hurtig, S. Bray, A. E. Hafner, L. Harker-Kirschneck, P. Faull, C. Davis, D. Papatziamou, D. R. Mutavchiev, C. Fan, L. Meneguello, A. Arashiro Pulschen, G. Dey, S. Culley, M. Kilkenny, D. P. Souza, L. Pellegrini, R. A. de Bruin, R. Henriques, A. P. Snijders, A. Šarić, A. C. Lindås, N. P. Robinson, and B. Baum, *The proteasome controls ESCRT-III-mediated cell division in an archaeon,* Science (New York, N.Y.) **369** (2020), 10.1126/science.aaz2532.

[213] A. Padirac, T. Fujii, A. Estévez-Torres, and Y. Rondelez, *Spatial Waves in Synthetic Biochemical Networks,* Journal of the American Chemical Society **135**, 14586 (2013).

[214] H. Wu and P. Brennwald, *The Function of Two Rho Family GTPases Is Determined by Distinct Patterns of Cell Surface Localization,* Molecular and Cellular Biology **30**, 5207 (2010).

[215] Y. Caspi and C. Dekker, *Divided we stand: splitting synthetic cells for their proliferation,* Systems and Synthetic Biology **8**, 249 (2014).

[216] D. A. Ramirez-Diaz, A. Merino-Salomón, F. Meyer, M. Heymann, G. Rivas, M. Bramkamp, and P. Schwille, *Ftsz induces membrane deformations via torsional stress upon gtp hydrolysis,* Nature Communications **12**, 3310 (2021).

[217] K. Zieske and P. Schwille, *Reconstitution of self-organizing protein gradients as spatial cues in cell-free systems,* eLife **3**, e03949 (2014).

[218] T. Furusato, F. Horie, H. T. Matsubayashi, K. Amikura, Y. Kuruma, and T. Ueda, *De Novo Synthesis of Basal Bacterial Cell Division Proteins FtsZ, FtsA, and ZipA Inside Giant Vesicles,* ACS Synthetic Biology **7**, 953 (2018).

[219] E. J. Cabré, A. Sánchez-Gorostiaga, P. Carrara, N. Ropero, M. Casanova, P. Palacios, P. Stano, M. Jiménez, G. Rivas, and M. Vicente, *Bacterial division proteins FtsZ and ZipA induce vesicle shrinkage and cell membrane invagination,* Journal of Biological Chemistry **288**, 26625 (2013).

[220] F. C. Keber, E. Loiseau, T. Sanchez, S. J. DeCamp, L. Giomi, M. J. Bowick, M. C. Marchetti, Z. Dogic, and A. R. Bausch, *Topology and dynamics of active nematic vesicles,* Science **345**, 1135 (2014).

[221] H. Ramezani and H. Dietz, *Building machines with DNA molecules,* Nature Reviews Genetics **21**, 5 (2019).

[222] M. W. Grome, Z. Zhang, and C. Lin, *Stiffness and membrane anchor density modulate dna-nanospring-induced vesicle tubulation,* ACS Applied Materials & Interfaces **11**, 22987 (2019).

[223] C. Fernandez-Fernandez, D. Gonzalez, and J. Collier, *Regulation of the Activity of the Dual-Function DnaA Protein in Caulobacter crescentus,* PLoS ONE **6**, e26028 (2011).

[224] G. C. Johnston, J. R. Pringle, and L. H. Hartwell, *Coordination of growth with cell division in the yeast Saccharomyces cerevisiae,* Experimental Cell Research **105**, 79 (1977).

[225] D. J. Ferullo and S. T. Lovett, *The stringent response and cell cycle arrest in Escherichia coli,* PLoS Genetics **4**, 21 (2008).

[226] P. Jorgensen and M. Tyers, *How cells coordinate growth and division,* Current Biology **14**, 1014 (2004).

[227] M. Wehrens, D. Ershov, R. Rozendaal, N. Walker, D. Schultz, R. Kishony, P. A. Levin, and S. J. Tans, *Size Laws and Division Ring Dynamics in Filamentous Escherichia coli cells,* Current Biology **28**, 972 (2018).

[228] M. Pickering, L. N. Hollis, E. D'Souza, and N. Rhind, *Fission yeast cells grow approximately exponentially,* Cell Cycle **18**, 869 (2019), pMID: 30957637, https://doi.org/10.1080/15384101.2019.1595874 .

[229] *The Cell: A Molecular Approach. 2nd edition* (Sinauer Associates, Inc., 2000).

[230] J. M. Skerker and M. T. Laub, *Cell-cycle progression and the generation of asymmetry in Caulobacter crescentus,* Nature Reviews Microbiology **2**, 325 (2004).

[231] S. Tanaka, T. Umemori, K. Hirai, S. Muramatsu, Y. Kamimura, and H. Araki, *CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast,* Nature **445**, 328 (2007).

[232] P. Kraikivski, K. C. Chen, T. Laomettachit, T. M. Murali, and J. J. Tyson, *From START to FINISH: Computational analysis of cell cycle control in budding yeast,* npj Systems Biology and Applications **1**, 1 (2015).

[233] R. Reyes-Lamothe and D. J. Sherratt, *The bacterial cell cycle, chromosome inheritance and cell growth,* Nature Reviews Microbiology **17**, 467 (2019).

[234] L. Somoayrac and O. Maaloe, *Autorepressor Model for Control of DNA Replication,* Nature New Biology **241**, 133 (1973).

[235] D. K. Chattoraj, R. J. Mason, and S. H. Wickner, *Mini-P1 plasmid replication: The autoregulation-sequestration paradox,* Cell **52**, 551 (1988).

[236] S. B. Lee and J. E. Bailey, *A mathematical model for λdv plasmid replication: Analysis of wild-type plasmid,* Plasmid **11**, 151 (1984).

[237] J. Paulsson and M. Ehrenberg, *Quarterly Reviews of Biophysics*, Vol. 34 (Cambridge University Press (CUP), 2001) pp. 1–59.

[238] N. Das, M. Valjavec-Gratian, A. N. Basuray, R. A. Fekete, P. P. Papp, J. Paulsson, and D. K. Chattoraj, *Multiple homeostatic mechanisms in the control of P1 plasmid replication,* Proceedings of the National Academy of Sciences of the United States of America **102**, 2856 (2005).

[239] K. Nordström, *Control of plasmid replication-How do DNA iterons set the replication frequency?* Cell **63**, 1121 (1990).

[240] J. Paulsson, K. Nordström, and M. Ehrenberg, *Requirements for rapid plasmid ColE1 copy number adjustments: A mathematical model of inhibition modes and RNA turnover rates,* Plasmid **39**, 215 (1998).

[241] A. Harms, D. E. Brodersen, N. Mitarai, and K. Gerdes, *Molecular Cell*, Vol. 70 (Cell Press, 2018) pp. 768–784.

SUMMARY

The ability to reproduce autonomously is a key feature of living beings. Cells reproduce via growth and cell division. During each cycle of growth and division, the so-called cell cycle, the cell needs to copy all cellular components, including its genetic information, exactly once. Controlling the number of chromosomes per cell is not an easy task and requires precise control over the moment at which DNA replication starts. How the cell regulates DNA replication on a molecular level remains however poorly understood. In this thesis, we aim to understand via mathematical modeling how the cell orchestrates the duplication of its genetic material with growth and cell division.

There are two approaches to study how life-like features emerge from individual components. While the top-down approach starts with a living organism and tries to understand its working principles, bottom-up studies start from individual components and reassemble them. In this thesis, we first use a top-down approach to study the replication cycle of the bacterium *E. coli*. We then take inspiration from living organisms to suggest routes towards building an autonomously dividing synthetic cell from the bottom up.

The bacterium *Escherichia coli* has a single, circular chromosome that is replicated once per cell cycle. A new round of replication always starts at one specific site on the chromosome, the so-called origin of replication, when the cell reaches a critical volume per number of origins. It is clear from biological experiments that opening the chromosome requires the activator protein DnaA that binds to the origin. In **chapter 2**, we present the so-called Autoregulated Initiator-Titration (AIT) model, in which DnaA first binds strong titration sites located on the chromosome, and only when all of these sites have been saturated, does it bind the origin to start a new round of replication. We test this titration model in different growth conditions and find that it only produces stable cell cycles at low growth rates. *E. coli* can however grow even faster than the time it takes to replicate the entire chromosome by having overlapping replication forks at high growth rates. In this growth regime, new initiator proteins are synthesized faster than new titration sites per origin, causing rapid re-initiation events that are not observed experimentally. Including the effect of SeqA to transiently block DnaA synthesis right after replication initiation can prevent these re-initiation events at high, but not at intermediate growth rates. We therefore conclude that there must be another mechanism that ensures stable replication cycles in *E. coli*.

Many experiments have shown that the activator protein DnaA can be in an active ATP-bound and an inactive ADP-bound form and that it switches from one state to the other over the course of the cell cycle via several mechanisms. In **chapter 3**, we therefore develop a new mathematical model for the regulation of replication initiation, which takes this protein activation switch into account. A model based on protein activation exhibits stable cell cycles if the activation and deactivation rates depend differently on the origin density. We argue that this requirement is fulfilled by the fact that lipid-

mediated activation of DnaA is independent of the origin while the other (de)activators in the cell are located on the chromosome, making their (de)activation rates proportional to the origin density. We show that including all known (de)activators of DnaA leads to high amplitude oscillations in the active fraction of DnaA. Importantly, the fact that in our mean-field model DNA replication is initiated at a constant volume per origin qualifies it as a sizer. Yet, experiments observe that *E. coli* adds an on average constant volume from one DNA replication initiation event to the next. We show that in a stochastic version of the same model the inevitable fluctuations in the components that control the threshold of the DnaA activation switch naturally give rise to the experimentally observed adder correlations.

The experimental evidence for the existence of both mechanisms is strong, raising the question why the system employs both mechanisms. While the switch ensures stable cell cycles at all growth rates, titration produces stable cell cycles only at low growth rates. Transiently blocking DnaA synthesis via SeqA stabilizes the titration mechanism at high, but not at intermediate growth rates. To generate robust oscillations over the full range of growth rates the switch is therefore essential. The converse question is however less obvious: the switch is not only necessary but could also be sufficient. What is therefore the benefit of combining the titration system with the switch? To answer this question, in **chapter 4**, we show that a concentration cycle, as induced by titration, can generically enhance an activation cycle, as driven by the switch, by sharpening the oscillations in the free concentration of active DnaA. Combining the switch-titration model with SeqA in the full model results in large oscillations in the free ATP-DnaA concentration in all growth regimes. We therefore conclude that *E. coli* has evolved an elaborate set of mechanisms that act synergistically to create robust replication-initiation cycles at all growth rates. We show that our full model agrees well with the experiments by comparing it to a large body of experimental data and we make new, experimentally testable predictions. Finally, in view of the enigmatic role of the lipids in activating DnaA, we remove the lipids from the full model and demonstrate that the full model is surprisingly resilient to the removal of lipids.

At high growth rates, DNA replication is initiated synchronously at multiple origins in *E. coli*. In **chapter 5**, we investigate how such a high degree of replication synchrony can be achieved given that replication is initiated stochastically at each origin. Initiating DNA replication synchronously is challenging because origins that have not yet fired must be allowed to still fire after the first origin has initiated. Simultaneously, every origin should only fire *once* per cell cycle and therefore origins that have already fired must be prevented from re-firing. We show that initiation synchrony at multiple origins requires both a local and a global mechanism: Combining a delay in the decrease of the initiation potential of the cell (global mechanism) with the local mechanism of blocking already initiated origins can give rise to synchronous replication initiation at multiple origins. While the combination of a global mechanism for allowing all origins to fire with a local mechanism for preventing the refiring of origins that have already fired appears to be an elegant solution, hard constraints need to be met. Given the fixed duration of the blocking period, the initiation potential must first rise, then remain high for a licensing period and eventually drop sufficiently fast before the end of the blocking period. We show both analytically and via simulations that a sharp rise in the initiation potential is

required for synchronous replication initiation within the finite blocking period. Finally, we test whether the previously presented theoretical model for replication initiation can ensure synchronous replication. We find that it can, even for biologically realistic parameters.

In **chapter 6**, we give a perspective on how a synthetic cell that can autonomously grow, replicate its genetic material and divide might be achieved. First, we describe different approaches to recreate DNA replication, segregation as well as cell growth and division in a synthetic cell. As inspiration, we take both natural systems and synthetic alternatives. A key question is how these modules can be coupled to each other to fulfill the fundamental requirements of chromosome number and cell size homeostasis. We therefore discuss how the different modules could be coupled to each other in a synthetic cell cycle. Finally, we suggest potential paths towards such an integrated synthetic cell cycle.

SAMENVATTING

Een essentieel kenmerk van levende wezens is het vermogen om zich autonoom voort te planten. Cellen planten zich voort via groei en celdeling. Tijdens elke cyclus van groei en deling, de zogenaamde celcyclus, moet de cel alle cellulaire componenten, inclusief zijn genetische informatie, precies één keer kopiëren. Het beheren van het aantal chromosomen per cel is geen gemakkelijke taak en vereist nauwkeurige controle over het moment waarop DNA-replicatie begint. Hoe de cel DNA replicatie op moleculair niveau regelt blijft echter slecht begrepen. In dit proefschrift proberen we met behulp van wiskundige modellen te begrijpen hoe de cel het dupliceren van zijn genetisch materiaal afstemt op zijn groei en celdeling.

Er zijn twee benaderingen om te bestuderen hoe karakteristieke kenmerken van het leven ontstaan uit individuele componenten. Een top-down benadering begint met een levend organisme en probeert de werkingsprincipes ervan te begrijpen, terwijl bottom-up studies juist uitgaan van individuele componenten en die met elkaar verenigen. In dit proefschrift gebruiken we eerst een top-down benadering om de replicatiecyclus van de bacterie *Escherichia coli* te bestuderen. Vervolgens laten we ons inspireren door levende organismen om routes voor te stellen naar het van onderaf opbouwen van een autonoom delende synthetische cel.

De bacterie *E. coli* heeft een enkel cirkelvormig chromosoom dat één keer per celcyclus wordt gekopieerd tijdens de zogeheten DNA-replicatie. Een nieuwe replicatieronde begint altijd op één specifieke plaats op het chromosoom, de zogenaamde replicatie-oorsprong, wanneer de cel een kritisch volume per aantal oorsprongen bereikt, ofwel de oorsprongsdichtheid onder een kritische waarde zakt. Uit biologische experimenten blijkt dat voor het openen van het chromosoom het activator-eiwit DnaA nodig is, dat zich bindt aan de oorsprong. In **hoofdstuk 2** presenteren we het zogenaamde Autoregulated Initiator-Titration (AIT) model, waarin DnaA eerst bindt aan sterke titratieplaatsen op het chromosoom, en pas als al deze plaatsen verzadigd zijn, aan de oorsprong bindt om een nieuwe replicatieronde te beginnen. Wij testen dit titratiemodel in verschillende groeicondities en stellen vast dat het alleen bij lage groeisnelheden stabiele celcycli produceert. *E. coli* kan echter sneller groeien dan de tijd die nodig is om het hele chromosoom te repliceren, door bij hoge groeisnelheden overlappende replicatievorken te creëren. Bij zulke hoge groeisnelheden worden nieuwe initiator-eiwitten sneller gesynthetiseerd dan er nieuwe titratieplaatsen per oorsprong ontstaan, waardoor snelle re-initiatie ontstaat die experimenteel niet wordt waargenomen. Het meenemen van het effect van SeqA, dat de DnaA-synthese onmiddellijk na replicatie initiatie tijdelijk blokkeert, kan deze re-initiatie voorkomen bij hoge, maar niet bij gemiddelde groeisnelheden. Wij concluderen daarom dat er een ander mechanisme moet zijn dat zorgt voor stabiele replicatierondes in *E. coli*.

Vele experimenten hebben aangetoond dat het activator-eiwit DnaA in een actieve ATP-gebonden en een inactieve ADP-gebonden vorm kan verkeren en dat het gedurende

de celcyclus via verschillende mechanismen van de ene naar de andere toestand overschakelt. In **hoofdstuk 3**, ontwikkelen we daarom een nieuw wiskundig model voor de regulatie van replicatie initiatie, dat rekening houdt met deze schakeling in eiwitactiviteit. Een model gebaseerd op dit soort eiwitactivering vertoont stabiele celcycli als de activerings- en deactiveringssnelheden verschillend afhangen van de oorsprongsdichtheid. Wij stellen dat aan deze eis wordt voldaan doordat de lipide-gemedieerde activering van DnaA onafhankelijk is van de oorsprong, terwijl de andere (de)activatoren in de cel zich op het chromosoom bevinden, waardoor hun (de)activeringssnelheden evenredig zijn met de oorsprongsdichtheid. Wij laten zien dat het meenemen van alle bekende (de)activatoren van DnaA leidt tot hoge amplitude oscillaties in de actieve fractie van DnaA. Een belangrijk punt is dat in ons mean-field model de DNA-replicatie wordt gestart bij een constant volume per oorsprong, wat het als een 'sizer' kwalificeert. Uit experimenten blijkt echter dat *E. coli* een gemiddeld constant volume toevoegt in de tijd tussen twee DNA-replicatie-initiaties, wat betekent dat het een 'adder' zou zijn. Wij tonen aan dat in een stochastische versie van hetzelfde model de onvermijdelijke fluctuaties in de componenten die de drempel van de DnaA-activeringsschakelaar controleren, leiden tot de experimenteel waargenomen 'adder'-correlaties.

Het experimentele bewijs voor het bestaan van beide mechanismen is sterk, wat de vraag doet rijzen waarom het systeem beide mechanismen gebruikt. Terwijl de activiteitsschakelaar zorgt voor stabiele celcycli bij alle groeisnelheden, zorgt titratie alleen bij lage groeisnelheden voor stabiele celcycli. Het tijdelijk blokkeren van DnaA-synthese via SeqA stabiliseert het titratiemechanisme bij hoge, maar niet bij gemiddelde groeisnelheden. Om robuuste oscillaties over het hele bereik van groeisnelheden te genereren is de schakelaar dus essentieel. De omgekeerde vraag is echter minder duidelijk: waarvoor dient het titratiemechanisme als het schakelmechanisme reeds voldoende is? Wat is het voordeel van de combinatie van het titratiesysteem met de schakelaar? Om deze vraag te beantwoorden, laten we in **hoofdstuk 4** zien dat het gecombineerde model robuuster is dan de modellen gebaseerd op één van beide mechanismen. Voral bij lage groeisnelheden beschermt titratie de schakelaar tegen fluctuaties in de schakelcomponenten. Wij laten zien dat een concentratiecyclus, zoals geïnduceerd door titratie, in het algemeen een activeringscyclus, zoals gedreven door de schakelaar, kan versterken door de oscillaties in de vrije concentratie van actief DnaA scherper te maken. De combinatie van het schakel-titratiemodel met SeqA in het volledige model resulteert in sterke oscillaties in de vrije ATP-DnaA-concentratie in alle groeiregimes. Wij concluderen daarom dat *E. coli* een uitgebreide verzameling mechanismen heeft ontwikkeld die synergetisch werken om bij alle groeisnelheden robuuste replicatie-initiatiecycli te creëren. We laten zien dat ons gecombineerde model goed overeenkomst met een grote hoeveelheid experimentele data, en we doen nieuwe, experimenteel toetsbare voorspellingen. Gezien de raadselachtige rol van de lipiden bij het activeren van DnaA, verwijderen we tenslotte de lipiden uit het gecombineerde model en laten we zien dat het volledige model verrassend goed bestand is tegen het verwijderen van lipiden.

Bij hoge groeisnelheden wordt DNA replicatie in *E. coli* bij meerdere oorsprongen synchroon gestart. In **hoofdstuk 5**, onderzoeken we hoe zo'n hoge mate van replicatiesynchronisatie kan worden bereikt terwijl de replicatie bij elke oorsprong stochastisch wordt geïnitieerd. Het synchroon starten van DNA replicatie is een uitdaging omdat

oorsprongen die nog niet gestart zijn nog steeds moeten kunnen starten nadat replicatie bij de eerste oorsprong is geïnitieerd. Tegelijkertijd mag replicatie op elke oorsprong slechts één keer per celcyclus geïnitieerd worden. Daarom moet worden voorkomen dat oorsprongsgebieden waar replicatie al geïnitieerd is, opnieuw starten. Wij tonen aan dat initiatiesynchronisatie bij meerdere oorsprongen zowel een lokaal als een globaal mechanisme vereist: het globale mechanisme initieert de replicatie op alle oorsprongen door afname van het initiatiepotentieel van de cel, terwijl het lokale mechanisme reeds geïnitieerde oorsprongen blokkeert, en zo voorkomt dat oorsprongen die al gestart zijn voor een tweede keer worden gestart. Deze combinatie lijkt een elegante oplossing voor synchrone replicatie-initiatie te zorgen, al moet daarvoor wel aan harde voorwaarden worden voldaan. Gegeven de vaste duur van de blokkeringsperiode moet het initiatiepotentieel eerst stijgen, vervolgens gedurende een 'licensing'-periode hoog blijven en uiteindelijk voldoende snel dalen voor het einde van de blokkeringsperiode. Wij tonen zowel analytisch als via simulaties aan dat een scherpe stijging van de initiatiepotentiaal vereist is voor synchrone replicatie-initiatie binnen de blokkeringsperiode. Ten slotte testen we of het eerder gepresenteerde theoretische model voor replicatie-initiatie synchrone replicatie kan garanderen. Wij laten zien dat dit kan, zelfs voor biologisch realistische parameters.

In **hoofdstuk 6** geven we een perspectief op hoe een synthetische cel tot stand kan komen die autonoom kan groeien, zijn genetisch materiaal kan repliceren en zich kan delen. Eerst worden verschillende benaderingen beschreven om DNA-replicatie, -segregatie, -groei en -deling in een synthetische cel na te bootsen. Als inspiratiebron nemen wij zowel natuurlijke systemen als synthetische alternatieven. Een belangrijke vraag is hoe deze modules aan elkaar gekoppeld kunnen worden om te voldoen aan de fundamentele eisen van chromosoomaantal en celgrootte-homeostase. Daarom bespreken wij hoe de verschillende modules aan elkaar gekoppeld zouden kunnen worden in een synthetische celcyclus. Tenslotte suggereren wij mogelijke wegen naar zo'n geïntegreerde synthetische celcyclus.

LIST OF PUBLICATIONS

Lorenzo Olivi*, **Mareike Berger***, Ramon N. P. Creyghton, Nicola De Franceschi, Cees Dekker, Bela M. Mulder, Nico J. Claassens, Pieter Rein ten Wolde, and John van der Oost, *Towards a synthetic cell cycle*, Nature Communications **12**, 4531 (2021) (* shared first author).

Mareike Berger and Pieter Rein ten Wolde, *Robust replication initiation from coupled homeostatic mechanisms*, Nature Communications **13**, 6556 (2022).

Mareike Berger and Pieter Rein ten Wolde, *Synchronous replication initiation of multiple origins*, in preparation.

ACKNOWLEDGEMENTS

There are so many people who accompanied me during my PhD journey and I cannot thank all of you enough for your support. My PhD time was wonderful mostly because of all the amazing people I met during this time. Without you I would never have made it that far.

First of all I would like to thank my supervisor Pieter Rein for guiding me throughout this challenging project. Your inexhaustible enthusiasm about the ideas and models that we developed together always motivated me to go one step further. You taught me to develop my ideas into solid, quantitative models that could stand up against the reviewers comments. I really appreciate how much time you took to discuss with me and that you were always approachable when I needed help. I also want to thank you for all the fun activities we did as a group, from the adventurous sailing trips to virtual diplomacy sessions during the Corona lock down!

A special thanks goes to my beloved para- and partynymphs, Manuel, Kathi, Lucie and Chris. Manuel, I was very excited when Pieter Rein told me that a new Master student from Munich was going to join our group. And you did not disappoint my expectations: Always eager to go out and explore Amsterdam together, your warm and thoughtful presence always cheers me up. Thank you for your open ear and your advise whenever I needed it. Having lunch and often also dinner almost every day during some of the lock downs together with Alex helped me survive this tough period. I really admire your deep understanding of physics and mathematics, your quick grasp and critical questions during the group meetings. Kathi, what a wonderful moment when you joined AMOLF. Your sunny presence and warm laugh makes AMOLF a nicer place. You are such an attentive person and you make every one around you feel at ease. I enjoyed going to the cinema together and discussing about the movies afterwards for almost as long as the duration of the movie, I loved our cocktail and cooking evenings, chatting while painting together, playing board games... Manuel and Kathi, I wish you a lot of success during the last year of your PhD. I am sure you're gonna rock it! Dear Lucie, you were one of the first people outside of the bio-department bubble that I met. I remember first talking to you in the AMOLF corridors and at the Christmas dinner. Already after a few chats, you invited me and my friends to your birthday party in Paris, which I found extremely generous of you. Your warmth and charmingness, your positive energy and encouraging words were extremely helpful during my PhD time and I want to thank you for this. Chris, I was so happy meeting you on one of my first days in AMOLF and finding someone from home with whom I could explore Amsterdam together. And that we did with all our senses: I remember tasting chalk at the art happening at Frascati (yumyum...), experiencing my first scent-dj party at Mediamatic and dancing to Nederhop at the silent disco in Noorderlicht cafe. These adventures were a great change to the academic life and I loved doing all these crazy things together with you. Thank you four for making my defense ceremony and party a memorable and fun day!

I would like to thank my defense committee for having taken the time to carefully read my manuscript and ask critical questions about it. Thank you for being present at my defense ceremony.

Thank you Bela, for never skipping a single coffee break! Especially during the pandemic, these coffee breaks kept me motivated and made us feel connected as a group despite the physical distance. I also want to thank you for your scientific input during the group and theory meetings and for your guidance during difficult moments of my PhD.

Lorenzo, meeting you at my second BaSyC meeting was what one could call a game changer. When I started my PhD, I felt like nobody in the BaSyC project was interested in the cell cycle ("We don't even have a synthetic cell, why would we care about its cell cycle?"). But then you popped up at my poster and your great enthusiasm blew me away. This is how our collaboration started, and every meeting since then gave me new motivation and cool ideas to work on. I enjoyed our in-depth discussions about the cell cycle regulation in *E. coli* and I am fascinated by how quickly your grasped the theoretical concepts I was developing. Thank you for initializing to write a perspective article on the cell cycle together. This project brought different views and expertise of the members of the BaSyC consortium together, exactly what BaSyC was aiming at. Writing the perspective article together was a great experience and broadened my horizon. I am very proud to hear that now people approach you, because they would like to work on the cell cycle ;)

Thanks to my various office mates for the fun atmosphere you created in our office. Ramses, thanks to your friendly and relaxed attitude I directly felt at ease when I arrived at AMOLF. When Harmen, Luci and Gerard joined the office, the usually quite boring white board full of equations suddenly turned into a crazy, somewhat disturbing art exhibition! Thank you for all the fun office conversations, coffee breaks and borrels we had together. The "Delta-Daltons" only lived for one summer, but they definitely left a mark both on the office (I recently found another flummy!) and on me. Harmen, thank you for welcoming me in the group when I arrived in Amsterdam. Your eagerness to help others and your calm, reassuring way makes you a very special person. I am very glad to having been your paranymph and having gotten to know you even better through this experience. Gerard, thank you so much for your critical questions and disbelieve in my simple models on the cell cycle of *E. coli*. You really helped me progress a lot during my project. I also appreciate that you visited my poster at almost every occasion, be it online or in person! Luci, being Harmen's paranymph together was definitely one of the most fun moments of my PhD. I remember the long, Britney Spears-heavy brainstorm sessions and the beautiful video-clip performance in your garden! I love your sarcastic humor, critical opinions and your passion for science. I wish you a lot of success in your postdoc. Gerard and Luci, I really missed you at AMOLF when your group moved to Delft. Luckily, Alex joined the office not much later and gave it a completely new touch. Your precise working style, regular research on unusual topics such as the amount of features a good office chair must have and your willingness to always interrupt work whenever I started talking to you really made you a wonderful office mate and a good friend. Thank you for being there for me when things got a bit difficult and for fuelling my work with the best croissants of Amsterdam. Thank you also for helping me take pictures of my artwork and editing them afterwards. Vahe, I really enjoy your calm and friendly pres-

ence in the office. Thank you for your spontaneous help with mathematical puzzles I encountered and thanks for taking over the important duty of watering the office plants! Muriel, it was a pleasure to supervise your short Master project and I was so happy to hear that you chose to come back and do your Master thesis in our group. Alex and I had been reluctant to let a fourth person in our office (how should we keep the one and a half meter distance?!), but for you we clearly made an exception: Your positive and fun energy cheers me up every day and the Yoga positions you taught me are really helpful. I am happy to know that it is you who is now sitting on my desk in the office!

A special thanks goes to all my (former) group members, thank you for making the group such a welcoming and great place to work in. Thank you Giulia, for welcoming me so warmly in Amsterdam and for inviting me to join the Medical Datascience Meetup. I profited a lot from your experience as a former post-doc in the group and I was glad that I could share with you some of the difficulties I had during my first year of PhD. Yao, I will never forget your opera songs at the Sinterklaas evening at Pieter Rein's home. Lotte, it was fantastic to have a second female physicist in the group. Your easy way of talking to people and your great interest in others made you a wonderful group member. And not to forget: Your Dutch directness lead to many funny moments! Age, you joined our group during Covid times, which was for sure not easy. Even though we only saw each other via zoom in the beginning, I felt right away connected with you. Being stuck first on a sandbank and then in weeds on our sailing trip definitely brought us closer together and I will never forget these moments! I was really impressed by how well you taught yourself all the physics and information theory backgroud and how confident you are now in your knowledge and understanding of these concepts. Raimon, it was a real pleasure to work together with you and guide you during your Master project. Your enthusiasm and eagerness to learn and understand things challenged me and lead to a deeper understanding of the system we were studying. I wish you lots of luck with your PhD, I'm sure you are going to do great work. Daan, thank you for teaching me so many Dutch sayings and for cheering up our lunches with Dutch poems and songs. Michael, your love for bad movies and fantastic food is legendary and I loved our conversations about food and even more cooking together. Jenny, thank you for introducing me to the British culture and so warmly inviting me and the group to your birthday party in Brighton. I really enjoyed how you integrated every one in the group and directly made us your new family in Amsterdam. Avischek, chatting with you about art and thesis cover design directly created a link between us. I am really happy you joined our group, your friendly and warm presence is wonderful to be around and I found your critical questions during the group meeting often very thought-provoking. Julien, even though we only spent little time together in the group, I really enjoyed getting to know you and discussing the differences of the French, German and Dutch culture and research environment. I hope you can finally leave the administrative post-Brexit obstacles behind you and I wish you a lot of fun and success in Edinburgh!

I also want to thank the members of Bela's group, who at some point really felt like my own group members. Ramon, your thoughtful comments, broad knowledge about societal and political topics and your subtle humour made every tea or coffee break more interesting and entertaining. I was glad to have a BaSyC theory buddy during my PhD and I loved our critical discussions about the BaSyC meetings. Preparing our talk and

presenting at the Pint of Science event was one of the most enjoyable moments of my PhD and it was great to do this together with you! I wish you and your family all the best and especially a lot of health. Marco, you always made me laugh with your italian view on the Dutch culture: From doctor appointments to Dutch cuisine, we had a lot of fun topics to chat about! Thank you for helping me set up the theoretical biophysics course at AMOLF, in which we taught each other interesting methods and new aspects of biophysics. Melle, it was really nice to discuss both physics and philosophical topics with you during lunch and coffee breaks. I found your often very different points of view very refreshing and interesting. Roberto and Kyriakos, it was great to finally meet you at the Christmas dinner after months of daily virtual coffee breaks! I wish you a lot of success with your PhDs.

I want to thank all the lovely people in the bio-department for your support and feedback during the biomeetings and poster sessions. Lori, my Dutch-Bio roomie, whenever I suggested to do something together, be it game evenings or making Käsespätzle together, you were down! I love your calm and reassuring attitude and your sarcastic humor. Xuan, you are such a kind and warm person and it is always great to be around you. Your poster was the first one I visited when I started my PhD and I remeber that I was very impressed by how much you had found out already after less than one year. Timo, thank you so much for the unforgettable dinner at your place. I have rarely eaten so well and laughed so much! I will honor the brownie recipie you gave me and hope that there will be many more occasions to enjoy your amazing cooking skills. Agatha, your defense was the first one I saw in the Netherlands and it was a great experience to see you shine! Yuval, our paths never crossed at AMOLF, but you were the first person I met from AMOLF. Thank you for your super positive energy and the great discussions we had before I started my PhD. Mattijs, thanks for your wonderful humor and energy, you brightened up the beginning of my AMOLF time! Nebojsa and Sofija, thank you so much for the amazing parties and Nebojsa's surprise performance! It is always great to drop into you in the corridors of AMOLF. Kristina, thank you very much for your advise on a career in academia and for your openness in sharing your experiences with me. Chi, I always enjoyed talking to you at the lunch break and at parties and I wish you a lot of success with your PhD. Evan, thank you for the funny and diverse conversations we had towards the end of my PhD. I hope you will enjoy your time at AMOLF as much as I did!

But the AMOLF family is even bigger than that and expands beyond the different departments. Caro, meeting you at the Art of Presenting Science course made this course so much more fun. And I am so happy that we kept in touch after our first romantic candle light dinner! I want to thank you for your emotional support during the tedious submission process. Thank you so much for encouraging me, when I struggled with the review process, and for sharing with me the difficulties you had during your own PhD. Dominique, I was so happy to finally meet a Dutch person with whom I could practice my rudimentary Dutch skills, in exchange for some German conversations. In the process of struggling with Dutch and German, we became friends and having not just espresso, no, fancy cappuccino together was always a lot of fun! Nasim, you charming and positive attitude is an absolute joy. You see something positive and beautiful in everything and everyone and your acknowledgement section is just one proof of this (the statement "It looks like a watermelon that is blown away by the wind." is another

one). Thank you for the wonderful watercolor evening that helped me relax during the stressful thesis writing period. Lukas, I really enjoyed our Duwo dinners, especially during the lock down period. Thank you for your tips for designing and editing my thesis cover. Jenny, I will never forget your laughter ringing through the corridors of AMOLF, even though we did not have a lot of overlap. Thank you for inviting me to your party when we did not know each other for long, it was a wonderful occasion to meet the people from different departments at AMOLF. Julia, thank you for choosing me as one of your paranymphs. Seeing how you managed to start a family, move into a new house and simultaneously did an amazing PhD was a great inspiration for me. I really admired how calmly you planned your defense and the great party afterwards was truly beautiful. Tom, thanks for being such a well-organized co-paranymph of Julia. Susan, thank you so much for helping me planning Julia's defense and making the gift for her. Your creative ideas, hands-on mentality and your humour was absolutely wonderful. I love meeting you at the coffee corner and chatting about this and that and I wish you a great continuation with your PhD. Robin, thank you so much for your surprise poem at Sinterklaas. It really made me laugh a lot! Thank you also for your help in preparing the defense gift for Julia, I really appreciated your creative input and your positivism! Ale, thank you for telling me about the Dutch course at the Institute of Physics at the UVA, which boosted my Dutch! Magda, unfortunately I only met you rather late at one of the sustainability meetings at AMOLF. Luckily, I could get to know you better even after your time at AMOLF at several occasions like for example the Dutch biophysics conference. Your positive energy and hearty laugh make you a wonderful person to spend time with and I hope to get many more occasions to get to know you better! Imme, I greatly enjoyed our regular conversations during coffee break and before and after going to our ice skating class. Learning how to make a real Dutch stamppot from you was great fun. I wish you a lot of success in the last phase of your PhD! Moritz, I am always very happy to meet you in the AMOLF corridors. Your calm and warm energy is contagious.

I want to thank all members of the PV for creating so many occasions to meet new people at AMOLF and to make everyone feel more connected. I think that this is one of the most special things of AMOLF and I appreciate that you spend so much extra time in making AMOLF a fun place to work at. Many thanks also to the Works council members, who represent the employees and try to guide the decision making process in the interests of the employees.

A special thanks goes to all the members of the support teams at AMOLF. You are often acting behind the scenes and one of the main reasons why AMOLF is such a great place to work at. Many thanks to the secretaries, for your help with organizing meetings. Thanks to the reception for creating a cheerful and welcoming atmosphere at AMOLF. A special thanks to the IT support, Gerben, Michiel, Kallinikos and Sean, for helping me with any linux related problems I encountered. Petra and Erny, thank you very much for your support when communicating about outreach events or paper publications. Thank you Wessel, for helping me take and edit pictures for the cover image for one of my paper and thesis. Dion, thank you so much for the surprise plant gift and for helping me making a beautiful, lens-shaped vase for Julia's defense. Thank you Clyde, for dropping into our office from time to time and explaining us how to sit better or how we can better clean our white board. These moments really pulled me back to reality when I was

pondering about an analytical problem. Thank you Hinco, for your positive energy that you spread all over AMOLF and for being always ready to help, whenever help is needed. I also want to thank the finance department, especially Angela, to patiently answer my questions when trying to fill in the more and more complicated travel expenses claims. My special thanks also go to Wouter, Linda and Reshma, for helping me with contract related, housing and unemployment money questions. It is really great that at AMOLF everyone helps eachother.

I also want to thank all the amazing people I met outside of AMOLF. My friends from Lindy Hop rapidly became my second family in Amsterdam. First I want to thank Alix, Pierre, Maria and Marine. Alix, you even found my number on the AMOLF website after I had given you a wrong number, in order to go to a Lindy event in Utrecht with me. I really liked going out with you and doing Lindy and step classes together. I always admired your high moral standards, your critical thinking and your compassion. Pierre, you were my number one dancing partner in Amsterdam for many years and it was great to do so many fun things together. I remember going to the beach, playing board games, going to the plant and parrot farm together ... and making a hearty Cassoulet for Maxime and me. Thank you so much for your emotional support in difficult times and for doing so many fun things together! I hope you are enjoying the scandinavian life, and if not you should know that you can always come back to the Netherlands or France. Maria, thank you so much for the creativity, energy and warmth that you shared with me. Thank you for inviting us to your cute little apartment in Jordan, where we had the best waffles with toppings I ever had. I wish we could see each other more often and I still wanna visit your house in Zaltbommel. Marine, hanging out with you after work, going to bars, parties and restaurants together was a great balance to work. Your focus on art, creativity and meditation really stood out to me and you taught me a lot of new things. I loved visiting you in London just before the lockdown and it was great that we kept in touch during the pandemic by playing games and having video calls. Pierre, Maria and Marine, we really were a wonderful team and it makes me a bit sad to know that I am the only one left in Amsterdam. But it still makes me laugh so much to think of the furniture building day, were you helped me in one of my first weeks to build all my furniture!

Thank you so much, my lovely Lindy and sewing ladies Laura, Miha and Elisa for all the wonderful moments we spent together! It all started with an unofficial Lindy hop course, where I met you, Laura. You told me about the classes you are taking at Swing Street and I directly signed up. Quickly, our pre-course dinner with yummy South Corean food became a tradition. The line dance course we took together was a real bonding moment for the three of us: Neither of us wanted to submit to the imposed dress code and the dance movements did not feel very natural. Our sewing afternoons and evenings really helped me during Covid times and created a balance to work. Moving in with you, Laura, was one of the best things that could happen to me in Amsterdam and I really love our Ehpad/plongeoir! Thank you for inviting Maxime to join our flat share without hesitation.

I want to thank Andy, Gui, Nina and Timon for the many fun moments we had during and after our Dutch classes. You were one of the first people I met in Amsterdam outside of work and I really enjoyed our game evenings and exploring the Dutch language and culture together.

One of the great joys during my time in Amsterdam was to communicate science to a broad audience at Pint of Science. I want to thank you, Laura, for your fantastic energy and for motivating me to speak about my research at Pint of Science. It is great how you involve and encourage everyone around you. Thanks to the entire Amsterdam team for having so much fun together organizing the events and helping each other whenever help was needed!

I also want to thank my Master project supervisor Chase for having suggested doing a PhD in the group of Pieter Rein. Having done my Master thesis in your group was a very good preparation for my PhD and on top of that a lot of fun. Thanks for fuelling my passion for science and physics and encouraging me to do a PhD!

Joris, you helped me realize more than anyone else that publishing a high impact paper is difficult for everyone. You made me feel less alone in my struggle and I really appreciate how you helped me with your light and cheerful way to overcome these difficulties. Our road-trip in the US was unforgettable, thank you so much for spontaneously joining just a week before having to defend your thesis. Your light-heartedness and passion for science was always a great motivation for me! I am looking forward to visiting you in Japan.

Karsten, our many conversations about all kinds of topics helped me a lot during my PhD time. I want to thank you for making me realize and avoid fallacies and cognitive biases. Thank you so much for being such a committed friend and for guiding me in making some very important decisions. I highly value your enthusiasm, querky humor and infinite amount of ideas, thoughts and questions!

Ich kann meinen vielen Freunden in München nicht genug dafür danken, wie ihr mich aus der Ferne unterstützt und ermutigt habt! Es war nicht leicht, euch zurückzulassen, und ich bin so froh, dass viele von euch mich in Amsterdam besucht haben. Es freut mich sehr zu sehen, dass die räumliche Distanz kein Hindernis ist und ich bin wirklich dankbar für unsere Freundschaft! Meine liebe Giada, danke für die vielen lustigen und intensiven Gespräche in Kombination mit Sauna-Entspannung! Liebste Claudia, deine sonnige Art und deine positive Ausstrahlung erfreuen mich jedes Mal, wenn wir uns sehen. Ich fand es so schön, dass ihr beide mich gleich in den ersten Monaten in Amsterdam besucht habt und wir die Stadt zusammen entdeckt haben. Jana, ich liebe deine starke Ausstrahlung und Selbstsicherheit, deinen Humor und deine liebevolle und ermutigende Art. Ich habe mich so gefreut, als ich das Gefühl hatte, nicht genug Zeit für meine Münchener Freunde zu haben und du mir erklärt hast, dass das nichts an unserer Freundschaft ändert und wir uns, wenn wir uns sehen, wie immer super verstehen. Liebe Annalena, ein zweites Mal habe ich München verlassen und es ist mir wieder nicht leicht gefallen. Aber unsere vielen Gespräche, unser regelmäßiger Austausch über unsere PhD Schwierigkeiten und wunderbare Momente wie deine Hochzeit haben uns all die Zeit verbunden. Vielleicht schaffen wir es ja demnächst wieder in derselben Region zu wohnen? Ich fände es so schön dich und Christoph öfter zu sehen und Annika aufwachsen zu sehen. Christoph, dein Humor und deine lockere Art erheitern mich immer, wenn wir uns sehen! Kilian, ich freue mich immer sehr, dich in München zu sehen und bei unseren Spaziergängen viele spannende Themen zu besprechen. Korbinian, auch wenn wir uns seit der Masterarbeit nicht mehr so oft sehen, freue ich mich immer sehr, mit dir einen Schlehenschnapps zu trinken und von deinen spannenden Projekten

zu erfahren. Koushik, I am super happy to have you as a friend and I am very glad to hear that you enjoy your PhD in Helsinki. I hope to come visit you soon! Tobi, deine sonnige, entspannte Art ist total ansteckend und ich denke immer wieder gerne an den unterhaltsamen Kürbis-Abend! Antonia, ich finde es so schön, dass wir trotz der wenigen Male an denen wir uns sehen, noch viel Kontakt miteinander haben. Deinen Besuch in Amsterdam habe ich in wunderbarer Erinnerung! Annele, ich finde es so schön, dass wir uns so lange kennen und uns noch immer so gut verstehen. Behalte deine liebevolle, gütige und humorvolle Art, die finde ich fantastisch! Clari, mein Sonnenschein, unsere lange Freundschaft ist mir sehr kostbar und ich freue mich jedes Mal dich und Brian wiederzusehen, egal ob in Bremen, Berlin oder Amsterdam! Marina, it is always such a pleasure to meet you. You are an inexhaustible source of joy, good humor and fun and I am so happy that you want to come to my PhD defense and party! Meine liebe Veri, wie schön, dass wir uns nach all den Jahren immer noch so gut verstehen. Unsere Freundschaft ist mir sehr viel Wert und ich schätze deine ruhige, überlegte und sonnige Art sehr. Es freut mich riesig, dass du planst bei diesem wichtigen Schritt in meinem Leben dabei zu sein.

Ein großes Dankeschön geht auch an meine Freunde aus dem SommerARTelier! Die lustigen Abende voller Verkleiden und gutem Wein und die mit Kunst und Kreativität gefüllten Tage haben mir immer wieder klargemacht, dass die Doktorarbeit nicht alles ist und die Welt nicht nur mit Logik zu begreifen ist. Eva, vielen Dank für deine Gastfreundschaft und deine vielen lustigen Nachrichten, Postkarten und Anrufe. So schade, dass du bei der Verteidigung nicht persönlich dabei sein kannst. Liebe Kerstin, von dir habe ich nicht nur Nähen gelernt (eine Fähigkeit, die ich sogar in meiner Doktorarbeit brauchen konnte). Du hast mir auch so viel Zuneigung und Wärme geschenkt und ich bin dir dafür sehr dankbar. Hoffentlich sehen wir uns bald wieder.

Vielen Dank, Marga, meine Mentorin, dafür, dass du mich auf dem Weg in meinen ersten Job nach der Doktorarbeit begleitest. Danke, dass du immer erreichbar bist und mir die Angst und Sorgen vor dem Unbekannten nimmst!

Un grand merci à mes amis français. Faire la fête et me défouler avec vous m'a toujours fait tellement du bien et m'a aidé de ne pas prendre ma thèse trop au sérieux. Zaz, merci beaucoup de m'avoir laissé emménager dans votre colloque pendant la pandémie, cela nous a beaucoup aidé Maxime et moi. Alicia et Edouard, merci de toujours m'accueillir chez vous si chaleureusement. Mathieu, merci pour ton énergie fantastique et ces merveilleux moments de fête que nous passons toujours ensemble. Fanny et Hugo, merci pour les soirées pyjama musicales durant la pandémie, c'était un superbe divertissement de notre routine de confinement. Sophie, je te remercie pour ton humour et les super moments qu'on a passé ensemble en montagne. Chère Citlali, je suis super heureuse de t'avoir rencontré au laboratoire des atomes froids à Toulouse, et même si tu n'as pas pu me convaincre de devenir expérimentaliste, on est resté amis depuis ce moment-là. Merci pour ton rire contagieux, ton optimisme et tous tes conseils précieux.

Je veux aussi remercier la famille de Maxime: Didier, Catherine, Claire, Nicolas, Arthur et Julia. Vous m'avez intégré dans votre famille sans hésitation. Les voyages en bateau à voile, de la Rochelle jusqu'au Glénans, m'ont donné l'énergie nécessaire pour survivre aux dernières étapes de ma thèse. Je suis super heureuse de vous voir souvent et j'ai hâte de vous accueillir à Amsterdam pour ma défense de thèse.

Zu guter letzt möchte ich meiner Familie und vorallem meinen Eltern danken. Mapa, deine Faszination für die Natur und deren Wunder hat mich früh stark geprägt und zu meiner Entscheidung beigetragen, Physik zu studieren. Unsere tiefen Gespräche über Gott und die Welt machen mir immer wieder viel Spaß. Du hast mir beigebracht, dass es gut und wichtig is neugierig zu sein und Fragen zu stellen und du hast immer an mich geglaubt. Vielen Dank dafür! Liebe Mama, ich kann nur schwer in Worte fassen wie dankbar ich dir bin für alles, was du während meiner Doktorarbeit für mich getan hast. Danke für deinen Humor, deine Unterstützung in schwierigen Zeiten, deine vielen Besuche in Amsterdam und die unzähligen lustigen Postkarten. Ich bin unendlich froh, dass es dich gibt! Vielen Dank auch an meine Tanten und Onkels, Cousinen und Cousins, für euer Interesse an meiner Arbeit und eure ermutigenden Worte bei unseren regelmäßigen Familientreffen. Klaus, deine Begeisterung für die Naturwissenschaften und unsere tiefen Gespräche bedeuten mir sehr viel. Ina, vielen Dank für deine Beratung bei schwierigen beruflichen Entscheidungen und der Frage, wie man Beruf, Partnerschaft und Familie miteinander vereinbart. Du bist ein großes Vorbild für mich.

Ces dernières lignes sont pour mon amoureux et partenaire Maxime. Je ne peux pas te dire à quel point je suis heureuse que nous nous soyons rencontrés il y a presque huit ans. Merci beaucoup pour tout le soutien que tu m'as apporté tout au long de mon doctorat. Surtout pendant la dernière période d'écriture intense, tu m'as régulièrement proposé de sortir faire une promenade ou du rolleur, tu m'as préparé des plats fantastiques et tu étais à l'écoute de mes problèmes. Cela m'a vraiment empêché de devenir un zombie ! T'avoir à mes côtés était et sera toujours un vrai bonheur.

ABOUT THE AUTHOR

Mareike Berger was born in Munich on September 29, 1993. After completing her high school studies in 2012, she felt a strong curiosity for understanding how the natural world works. She obtained her Bachelor's degree in Physics from the *Ludwig-Maximilians-Universität* in Munich, in 2016. During her Bachelor's program she spent one year at the University *Paul Sabatier* in Toulouse, where she also did her Bachelors thesis.

Desiring to deepen her understanding of Physics, she continued her studies at the *Ludwig-Maximilians-Universität*, in Munich. There, she obtained her Masters in 2018. During this Masters, she completed an experimental research internship in the group of ultra-cold atoms in the Laboratoire Collisions Agrégats Reactivité (LCAR). In her Masters thesis, she studied the rupturing dynamics of biological actin filament networks under the supervision of Chase Broedersz. In March 2018, she obtained a prize for the best poster presentation at the DPG conference.

Besides her studies, she worked for two years as a working student at Siemens, where she developed software and applied machine learning techniques to build a parking place classifier.

In 2018, she started her PhD at AMOLF in Amsterdam in the Biochemical Networks group under the supervision of Pieter Rein ten Wolde. The results of her work on coupling of growth, DNA replication and cell division in bacteria and synthetic cells are presented in this thesis. She obtained the second poster prize at the Dutch Biophysics conference 2021. Next to her thesis, she participated in the Sustainability group of AMOLF.

Mareike also enjoys scientific outreach and is part of the organization team of Pint of Science in Amsterdam. She presented her work at the Pint of Science event on "Synthetic Life" in May 2022.