UNIVERSITY OF CALIFORNIA SAN DIEGO

DNA replication and cell size control in *Escherichia coli*

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Biology

by

Dongyang Li

Committee in charge:

Professor Suckjoon Jun, Chair
Professor Arshad Desai
Professor Kit Pogliano
Professor Milton Saier
Professor Massimo Vergassola

2018

The dissertation of Dongyang Li is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2018

DEDICATION

To my parents for their unwavering support and love, to Shihan for being there for me and making me want to do better than yesterday.

EPIGRAPH

We are the reckless
We are the wild youth
Chasing visions of our futures
One day we'll reveal the truth
That one will die before he gets there
And if you're still bleeding, you're the lucky ones
'Cause most of our feelings, they are dead and they are gone

— Daughter

from the song "Youth"

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis advisor Suckjoon Jun, for taking me in and teaching me how to think and act as a scientist. More importantly, he taught me the courage to face various challenges and my inner self. I would also like to acknowledge the rest of thesis committee members, in no specific order, Professor Kit Pogliano, Professor Arshad Desai, Professor Massimo Vergassola and Professor Milton Saier. They gave me suggestions, encouragement and feedback to help me push my thesis work forward.

I would also like to thank all Jun lab members, in the past and present, for the good rapport in the lab and their peer support. The dark nights I shared with JT Sauls in the lab or on the road trip, the discussion I had with Steven Brown for the sake of pure curiosity, the minilectures Michael Erickstad gave me to explain optics, circuitry and fluidics in layman's words, the generous help I received from Fangwei Si and Guillaume Le Truet. Xintian Li taught me bacterial genetics and helped me out on numerous occasions when thing just did not work for me for no obvious reasons, Yonggun Jun encouraged and gave me a new perspective on scientific life. Sarah Cox was always chilled and helped me get things done and passing on good vibes. I cannot put them all on record, but they took a fair share of my inner world in grad school life.

I would also like to thank Professor Jens Lykke-Andersen, our graduate chair, for helping me out after my decision of changing lab. Ying Hu gave me a lot of guidance when I was lost and helpless in the transition. My previous advisor, Björn Lillemeier, despite the hardship I felt at the time, taught me perseverance in face of difficulty and how to pursue one's passion. I left his lab on good terms

and I am grateful for his truly caring for me during and after my stay there.

Chapter 3 is a subset of material as it appears in Current Biology 2017. Fangwei Si, Dongyang Li, Sarah E. Cox, John T. Sauls, Omid Azizi, Cindy Sou, Amy. B. Schwartz, Michael J. Erickstad, Yonggun Jun, Xintian Li, Suckjoon Jun, Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. Curr. Biol. 27, 1278–1287 (2017). The dissertation/thesis author was the primary investigator and author of this paper.

VITA

2011-2018	Ph.D. in Biology, University of California, San Diego
2007-2011	B.S in Biology, Peking University

PUBLICATIONS

Si F*, **Li D***, Cox S, Sauls J, Azizi O, Sou C, Schwartz A, Erickstad M, Jun Y, Li X, Jun S. Invariance of Initiation Mass and Predictability of Cell Size in *Escherichia coli*. *Current Biology*. 2017 27(9):1278-1287
equally contributed

* indicates co-authorship

ABSTRACT OF THE DISSERTATION

DNA replication and cell size control in *Escherichia coli*

by

Dongyang Li

Doctor of Philosophy in Biology

University of California San Diego, 2018

Professor Suckjoon Jun, Chair

The defining feature of living organisms is their capacity to reproduce and pass on the genetic information so that their progeny can flourish. For bacteria, reproduction is a feat by itself – *Escherichia coli* cells cultured in optimal conditions grow rapidly and divide about every 20 minutes. In other words, the cell has to replicate all cellular contents, and be ready to divide evenly into two daughter cells within this 20 minutes. Biosynthesis of new cellular materials, e.g. proteins, nucleic acis, lipids and other metabolites accumulate and roughly doubles after every generation. Notably, the deoxyribonucleic acid (DNA) encodes genetic information

and needs to be duplicated in order to faithfully pass on this information to the progeny. This process of DNA replication in the cell needs to dynamically adapt to fluctuation in growth condition and cellular physiology. Such coordination is controlled at the first step of replcation – the initiation of replication. In this thesis, I presented the development of methods for measuring DNA replication duration (replication period), the quantitative relationship between DNA replication and cell size as well as the mechanism of replication initiation. DNA replication measurement laid the foundation of studying the quantitative relationship between cell size and DNA replication. A general growth law was proposed to describe cell size regulation in light of three physiological variables including biosynthesis rate, cell cycle progression and replicaiton initiation. Of the three variables, the mass when cell initiates replication (initiation mass) remains invariant despite a wide spectrum of antibiotics or growth limitation challenge. This invariant initiation mass called into question about the mechanism of initiation to achieve such constancy. We proposed a simple threshold model to explain how cells can maintain a invariant initiation mass by regulating the expression of initiation regulators (initiators). The initiaiton mass is inversely proportional to the initiator levels, which is held constant. Experimental evidence was provided to test our model prediction.

Chapter 1

Introduction

1.1 Introduction

The defining feature of living organisms is their capacity to reproduce and pass on the genetic information so that their progeny can flourish. For bacteria, reproduction is a feat by itself – *Escherichia coli* cells cultured in optimal conditions grow rapidly and divide about every 20 minutes. In other words, the cell has to replicate all cellular contents, and be ready to divide evenly into two daughter cells within 20 minutes. In a typical *E. coli* cell, proteins occupy about half of the mass fraction and there are about 10^6 proteins in each cell [24]. In steady state, this means the cell has to synthesize an equal number of proteins within one generation. The number is large enough that protein synthesis can be seen as a continuous process. The same argument goes for RNA synthesis and other metabolic processes with one exception: DNA replication. Each cell has on the order of one chromosome [24], therefore the cell on average only needs to replicate its genome once per generation. Hence the DNA replication is a discrete process

that can only happen once per generation in steady state. If more DNA is replicated, the average DNA content will increase and deviate from steady-state values. Meanwhile the opposite scenario of under-replicating the genome is equally undesirable, where there is a loss of genetic information passed onto the progeny. How to ensure that DNA is only replicated once and only once before cell division, so that the progeny cells inherit a complete genome? Moreover, the generation time can vary more than 10-fold, depending on the nutrient quality of growth media. Does the cell dynamically adjust the progress of DNA replication and growth to different length of generation time?

These questions are central to bacterial physiology, a field studying the basic principles underlying cell growth and division. For a population of cells to stably expand, DNA replication needs to be coupled to cellular growth to counteract any fluctuations in growth rate over time. It is the goal of this thesis to study how cells couple DNA replication to cell growth that ensures steady-state propagation of cells.

It may be beneficial to introduce some of the basics of bacteria growth and clarify some terms that will be repetedly used throughout this thesis (see Figure 1.1 for schematic explanation.)

Bacteria grown in batch culture (e.g. in a culture flask) go through four distinct stages of growth, depending on its nutrient conditions and its metabolic state. The growth of cells can be measured by turbidity of the liquid culture, usually in a spectrophotometer, reported in optical density units (OD). A liquid culture inoculatd from a single colony starts with **lag phase** where growth is barely noticeable. After certain amount of time, the culture enters into **exponential**

phase when the OD increases exponentially as time progresses. Nutrient in the media eventually runs out and the OD growth plateaus as the culture begins **stationary phase**. The OD of the culture slowly decays as cells start to die off in the **death phase**.

The exponential phase can be sustained if fresh media is supplied, in most cases through serial dilution of the culture (Figure 1.1). The exponential phase can be sustained in theory indefinitey and such a culture is also called an **exponential culture**. At exponential phase, the total cell number and total cell mass increases exponentially and doubles after every **mass doubling time $\boldsymbol{\tau}$**:

$$\begin{aligned} N(t) &= N(0)2^{t/\tau} \\ &= N(0)e^{\lambda t} \end{aligned} \tag{1.1}$$

where N(t) is the cell number at time t and N(0) is the initial cell number. As can be seen from the above equation, the **growth rate $\boldsymbol{\lambda}$** is the inverse of the mass doubling time as in $\lambda = \ln 2/\tau$.

The exponential culture can be characterized by different properties, which fall into two categories: intensive property and extensive property[I]. Intensive properties do not change with the scale of the system (e.g. density, pressure or temperature), whereas extensive properties depend on the size (extent) of the system (e.g. mass, volume, energy or entropy). In the realm of bacterial physiolgy, total cell numbers, total cell mass and total DNA or protein are all extensive properties that increase exponentially over time. Average cell mass, average cell

[I] The term intensive and extensive properties were borrowed from thermodynamics. Intensive properties do not change with the size or mass of the system. Extensive properteies, on the contrary, increases as the system scales up and are additive.

size or average DNA/protein content are intensive properties that stay the same. This is a feature of **steady-state** culture, where intensive properties become time-invariant.

When cell culture is said to be in steady-state, the culture is in both exponential phase and **balanced growth**, which is defined as "every extensive property of the growing system increases by the same factor" [31]. In other words, cell culture in balanced growth is simply scaling up its total amount of DNA, protein, metabolites etc., all in the same rate as the **growth rate**.

From a single-cell perspective, each cell grows in size exponentially (not to be confused with exponential growth of the culture). The rate of growth can be defined as **elongation rate $\boldsymbol{\alpha}$** as $\alpha = \frac{1}{l}\frac{\mathrm{d}l}{\mathrm{d}t}$. Under steady-state growht, the average elongation rate is equivalent to the population growth rate. The time interval between two consecutive divisions (from birth to division) is defined as **interdivision time**, also called **generation time** τ. Note that they share the same symbol with the mass doubling time of the culture, as in steady state the population average generation time is the mass doubling time. As the cell grows in size, DNA replication progresses so that the DNA amount would on average double from birth to division. The time duration of the replication process is the **replication period** also known as the **C period**. The time from termination of replication upto cell division is the **D period**. The C and D period combined is called the **cell cycle period $\boldsymbol{\tau}_{\mathrm{cyc}}$**. Note that cell cycle period is not necessarily limited within one generation. In certain growth conditions, the cell cycle can span several generations and the cell is said to have **overlapping cell cycles**, a phenomenon that will be covered in more detail in the following Chapter 2.

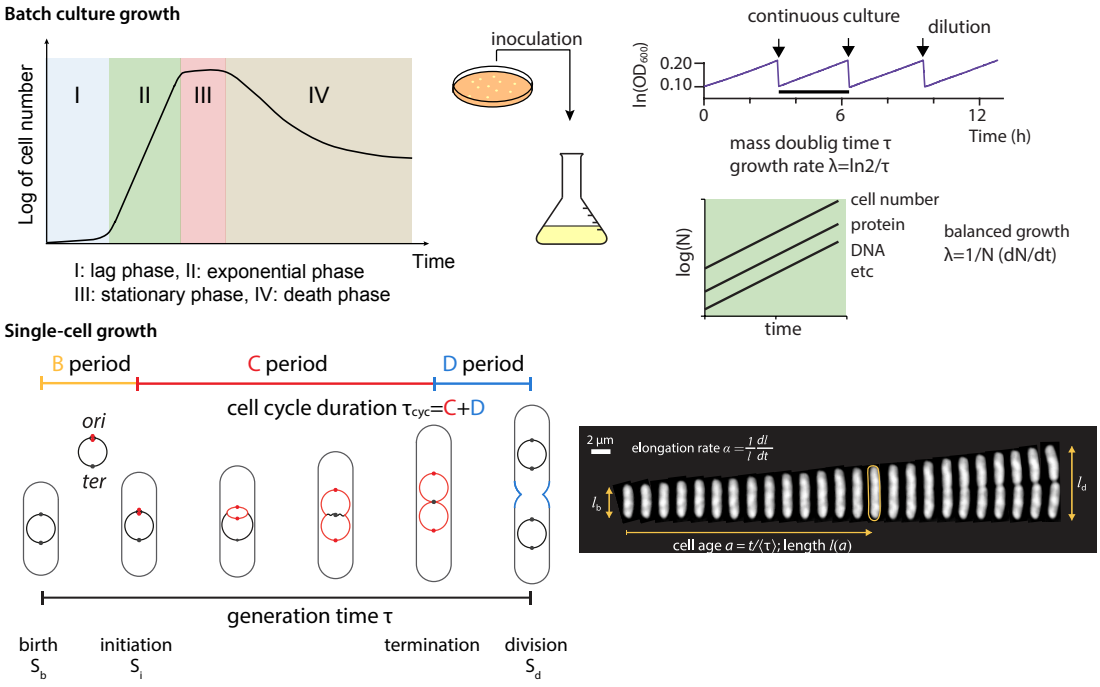

Figure 1.1: **Schematics of basic terms in bacterial physiology**. Figure adapted from [88].

1.2 Background on DNA replication study

Let us go back in time to peruse the classic works in this field and better understand how the field has evolved. In the 50s, it was clear to microbiologists that DNA bears genetic information that has to be passed on [77]. However, the physical structure of condensed DNA, the chromosome, was less well-understood. The bacterial chromosome was observed in different bacterial species as a less-defined structure inside the cell [41]. However, not much was known about the mechanism of DNA replication. One could imagine, that DNA would be replicated given the proper enzymes (DNA polymerase, primase, topoisomerase, etc.), substrates (nucleotides) and energy. Without any regulatory mechanism, replication would start when the proper combinations of enzymes are present locally at an arbitrary chromsome locus. However, there is no guarantee that the entire chromosome would be replicated if this process were solely driven by fluctuations of enzyme concentrations at different loci. This "random" replication would be problematic as it does not ensure the completeness of replication, or timely replication for cell division. In fact, the replication is more regulated in at least the following aspects (Figure 1.2): 1) DNA replcation is semiconservative[II] [131]; 2) DNA replication is an autonomous process that does not require new protein synthesis or any signal from last replication [13, 39, 58, 121]; 3) DNA replication is sequential[III] and starts at a unique and fixed region in the chromosome [137, 138, 192]; 4) Replication

[II] Semiconservative replication refers to the fact that replicated DNA each contains one strand of original DNA and one strand of new DNA. This was confirmed by the Meselson-Stahl experiment as opposed to other modes of replication.

[III] Sequential replication refers to that DNA replication has a fixed directionality, starting from a single fixed replication origin. This is in contrarary to other less regulated modes of replication, e.g. replication starting at random genomic locus/loci.

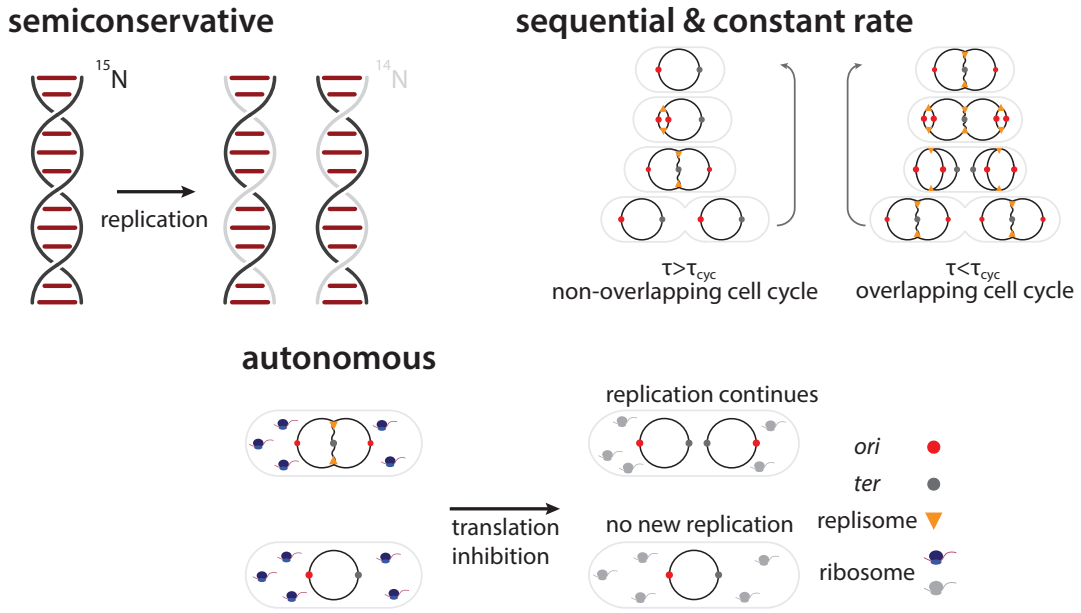

Figure 1.2: **Schematics of features of Bacterial DNA replications**.

rate does not vary with growth rate[IV] but multiple rounds of DNA can happen concurrently [70, 182, 209].

1.2.1 Semiconservative DNA replication

Despite the discovery of DNA double helix structure [210], it was not clear how DNA is replicated until Meselson and Stahl beautifully demonstrated it with isotope labeling [131]. The key to their technique is to use heavey isotope and then differentiate them in caesium chloride (CsCl) gradient centrifugation, the counter-action of centrifugal force and diffusion gives distinct peak for DNA fragments with ^{14}N or ^{15}N. At the time, there were three competing models for how DNA is replicated. Conservative replication assumes a new double-stranded DNA is made during replication; semiconservative replication assumes that the two strands will

[IV] Replication rate is constant when the growth rate $\lambda \geq 0.7$ doubling/hr. Replication slows down at slower growth rates.

be replicated separately to give a new strand, and pair up with each old strand respectively; dispersive replication assumes newly synthesized DNA will disperse throughout old DNA and gives a mixed double strand. In their experiment, an exponential culture grown in ^{15}N medium was transferred to ^{14}N medium, and cell samples were collected at different time points after the shift. They found out that early samples predominantly have heavy bands (both strands with ^{15}N), then hybrid band (one of the two strands with ^{15}N) and eventually light bands (both strands with ^{14}N). The strict order of apperance of heavy, hybrid and light bands compelled Meselson and Stahl to conclude that DNA replication is semiconservative and replication is sequential, i.e. any locus cannot be replicated again before the current round of replication is finished (Figure 1.2).

1.2.2 Ongoing DNA replication is independent from RNA and protein synthesis

One of the big questions was whether DNA replication is an autonomous process independent from RNA/protein production. Maaløe and Hanawalt drew inspiration from the nutrient shift experiment carried out in the same group [104]. In that experiment, Kjeldgaard, Maaløe and Schaechter compared the production of DNA, RNA and protein upon nutrient shift-down, and discovered a delay in the slowdown of DNA synthesis compared to the other two species (Figure 1.3). They reasoned that this differential change in synthesis rate upon nutrient shift suggests that ongoing DNA synthesis does not need RNA or protein synthesis. This idea was also supportd by evidence from amino acid withdrawal [58] and chloramphenicol treatment [13]. In both cases, DNA synthesis continue if the DNA replication

was underway at the time of treatment, but no new round of replication is initiated (Figure 1.2). In contrast to the continued DNA synthesis under translation inhibition, removal of thymine immediately stopped DNA synthesis but not protein synthesis in the thymine auxotroph strain $15T^-$ [39]. Interestingly, if DNA, RNA and protein synthesis are blocked altogether, the viability loss is partially recovered [121]. This cell death is termed thymineless death since the thymine-auxotroph cannot proceed DNA synthesis. It was Hanawalt who combined both

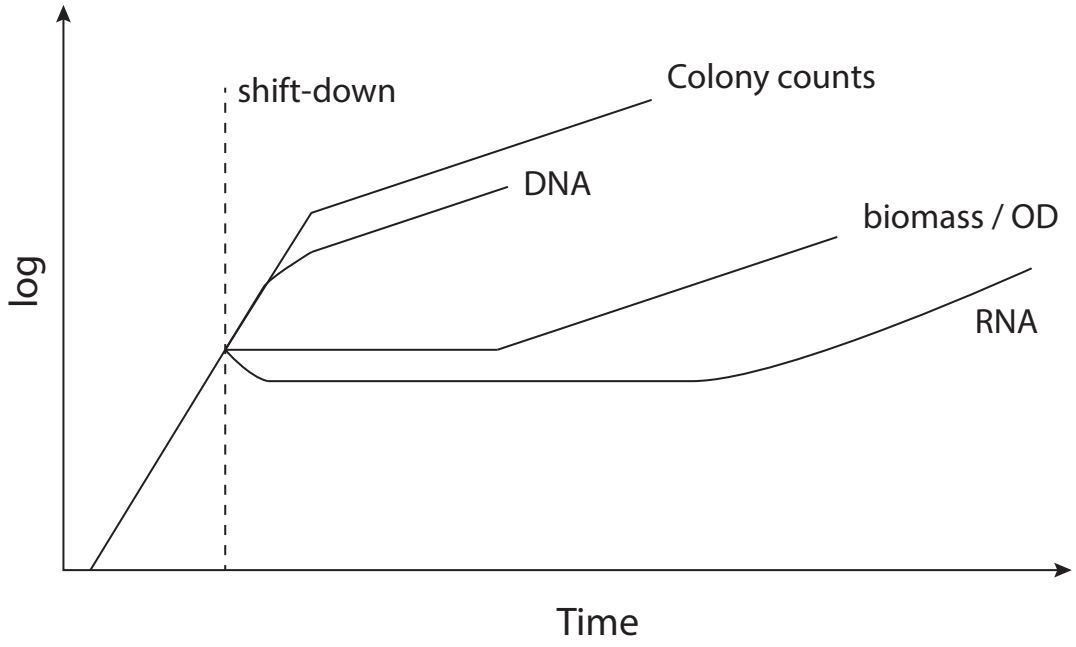

Figure 1.3: **Biosynthesis activities during nutrient shiftdown**. Immediately after nutrient shift-down, RNA and protein (biomass as measured by OD) came to a halt, while DNA continues replication. Adapted from [104], not drawn to the exact scale.

sides of evidence and shed light on the mechanism of DNA replication. He utilized a thymine, arignine and uracil triple auxotroph strain (15TAU) that can block DNA or RNA/protein synthesis separately. Cells incubated with thymine but

blocked for RNA and protein synthesis has a higher survival rate in the ensuing thymine removal. However, the DNA incorporation is capped at ∼40%, unlike the continuous increase in DNA synthesis in the control group where thymine, arginine and uracil are supplied and none of the synthetic activities are blocked. This level of incorporation coincides with the average DNA incorporated to finish all ongoing DNA replication but not any new round of replication. Therefore this analaysis led to the conclusion that ongoing DNA synthesis is independent of RNA or protein synthesis, whereas a new round of replication requires RNA or protein production. This implies that, once initiated, DNA replication can proceed autonomously without *de novo* synthesis of RNA or protein, yet the licensing of a new round of replication requires protein/RNA factors (Figure 1.2). This notion would be further developed in the following section on replication initiation.

1.2.3 Replication start site is unique and fixed

In the 1960s, it was not known whether replication starts at a random locus on the chromosome or at a specific one. Sueoka and Yoshikawa solved this problem using marker frequency anlaysis [143, 192, 215, 216]. In marker frequency analysis one quantifies the relative amount of DNA segments (genetic markers) from a population sample. If replication starts randomly along the chromosome, the relative amount of each fragment should be similar. If, however, replication starts at a specific locus, there would exist systematic difference in marker frequency. Furthermore, the closer the locus is to the replication origin, the higher the average copy number of that locus. This is because loci replicated later will appear in a smaller fraction of older cells. They used both *B. subtilis* spores and

chloramphenicol run-out samples[V] to show that a fixed replication origin would best explain the observed marker frquency pattern. They termed it replication polarity to indicate the uneven distribution of marker frequency and the directionaity of replication. This served as important evidence that there is a unique origin region where all replication starts (also known as sequential replication at the time). Given a steady-state exponential culture and assuming replication can be simplified as a linear progression on the chromosome, the amount of a specific marker (chromomsomal locus) can be expressed as a decreasing exponential function of distance from the replication origin (Figure 1.4). This turned out to be a very useful tool for cell cycle duration measurement (to be detailed in following sections). At the same time, Nagata came to the same conclusion of sequential replication studying the replication of prophage in thymine starvation-synchronized cell culture [137, 138].

Another important implication from Sueoka and Yoshikawa's experiment is that when they compared DNA prepared from spores to those from exponential culture, the ratio of certain markers reached 4. However, one round of replication would only double the marker copy numbers. Therefore, they deduced that there could be multifork replication ("dichotomous replicaiton" as used in their original paper) where a second round of replication starts before the frist round gets finished. It also served as evidence that starting a new round of replication does not rely on finishing the current round of replication.

[V] In run-out experiments, extended administration of rifampicin and cephalexin blocks transcription and cell division respectively. The net result is that cells are allowed to finish current round of replication ("run-out" of current replication) without initiating new round of replication or cell division. The run-out experiment is used to infer the DNA content of the cell in flow/image cytometry measurement [147].

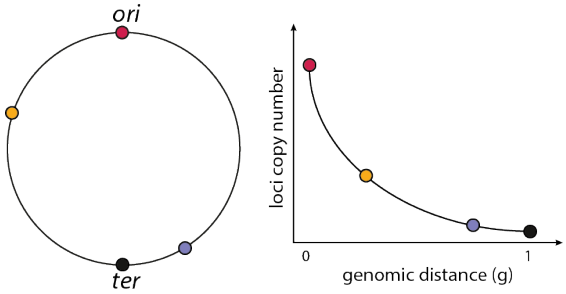

Figure 1.4: **Schematic of genomic locus copy number distribution**. Genomic loci copy number decreases exponentially as it gets further away from the replication origin (measured in genomic distance, normalized to a range between 0 and 1). Assuming the chromosome replication is symmetric, i.e. two loci having the same distance from *ori* on either arm would have the same copy number. The figure illustrates the relative copy number of two arbitrary loci as well as replication origin (*ori*) and terminus (*ter*).

1.2.4 Circular chromsome and bidirectional replication

Cairns took autoradiographs of lysed *E. coli* cells labeled with [^{3}H]thymidine, and estimated the physical length of the chromsome. He was also able to show that replication happens through a fork where the new strand and old strand forms a new helix. Notably, based on the various broken forms of chromosome, he proposed that the chromosome exist as a circle [28, 29]. However, both Cairns and Bonhoeffer [17, 29] assumed there is only one replication fork that starts at the same position on the chromosome based on Sueoka's finding from marker frequency analysis (Figure 1.5). Bonhoeffer used density distribution of 5'-bromouracil labeled DNA and compared the size of BU-labeled DNA segment with increased pulse duration to the size of a reference DNA labeled with [^{3}H]-thymine and concluded there can be a maximum of one replication fork. This miscalculation was

due to a lack of understanding of the genetic map at the time of their publication[VI]. This discrepancy was later resolved by Masters and Broda using P1 phage transduction experiment to show that replication is bidirectional [127]. The experiment is conceptually similar to the marker frequency analysis pioneered by Sueoka and his colleagues – ratio of marker frequency was mapped for different genes on the chromosome based on the number of transductants. With a detailed linkage map of over 300 genes of *E. coli* [200], Masters and Broda were able to map the marker frequency ratio to the chromosome and found that this ratio peaks at 60 minutes position and decrease as one moves away from it in either direction. This was also confirmed by a similar trend when simply looking at transduction efficiency, as the transduction efficiency depends little on integration efficiency and gene expression level [127]. Similar conclusion was reached by Bird *et al.* using marker frequency analysis through DNA-DNA hybridization [14]. Notably, this study also allowed them to map the origin of replication to be near *ilv* at 74 minutes, and they verified that replication speed is roughly the same in two chromosomal arms. The physical evidence of bidirectional replication came from Prescott and Kuempel, who used autoradiograph to show that incorporation of [^{3}H]-thymine is enriched on both ends of the circular chromosome [151].

It was, however, a decade later when Kaguni and Kornberg first showed that replication is bidirectional in a purified enzymatic system *in vitro* as well. They introduced replication terminating substrate 2',3'-Dideoxythymidine 5'-Triphosphate (ddTTP), which was just adopted in Sanger DNA sequencing for chain termination

[VI] The Bohoeffer experiment relied solely on the incorporation of radiolabeled nucleotides measured in minimal media where growth is slow. While it was successful in supporting the semi-conservative replication concept, the replication rate thus calculated would be twice the actual rate.

Figure 1.5: **Autoradiograh of an intact, replicating *E. coli* chromosome**. Cairns used [^{3}H]thymidine-label chromosome to provide physical evidence that bacterial chromosome exhibits as one intact ring and deduced the pattern of DNA replication based on the observed topology of the chromosome. Figure taken from [29].

[171]. By spiking in different amount of ddTTP, they were able to stop replication of an *oriC*-containing plasmid at different stages and examine them using electron micrograph. This method allowed them to directly visualize replication [90]. Even though this was only confirmatory of previous *in vivo* findings, this enzymatic system they had developed would prove to be of much more use in laying the foundation of biochemical characterization of the initiation process (as discussed in the following section 1.3).

1.2.5 Replication period is constant at different growth rates

Cell cycle in bacteria is divided into three stages with reference to chromosome replication: the time from last cell division to the initiation of replication is the B period; the duration for chromosome replication is the C period; and the time from termination of replication to the next cell division is the D period (Fig-

ure 1.6). This replication-centric view of the cell cycle has its root in the seminal work by Helmstetter and Cooper. It was them who first studied the link between cell growth and cell cycle systematically when they applied their new membrane elution technique, the "baby machine", to collect culture in small fractions that only contain newly divided cells [71, 72]. Newborn cells were eluted from the membrane whereas the mother cell remain on the membrane. This way cells can be synchronized to cell division/birth events, and DNA replication duration (or the C period) be derived from heightened level of incorporation of isotope-labeled nucleotides. They were able to show that DNA incorporation is periodic and the period does not change when cells are growing relatively fast (growth rate $\lambda > 0.7$ doubling/hour). Same was found true for the D period (the time between replication termination and cell division)[VII]. This implies that, when generation time gets shorter than the C period, multiple replication forks are present in any given generation to ensure there is at least two copies of entire chromosomal DNA for progeny cells. This phenomenon is termed overlapping cell cycle (Figure 1.6). This is consistent with the first observation of overlapping cell cycle in *B. subtilis* by Sueoka and Yoshikawa using marker frequency analysis [192]. The constancy of replication period was also confirmed by Gudas and Pardee. They synchronized culture by selecting smallest cell fraction from sucrose gradient centrifugation, and measured replication period by quantifying [^{3}H]-thymidine incorporation period [61]. In addition, Chandler, Bird and Caro calculated C period based on the ratio of two chromosomal markers, and confirmed the constancy of C period [34].

Since replication period is relatively constant over different nutrient-imposed

[VII] When cell culture further slows down, the (C+D) period increases proportionally as the generation time [182, 209].

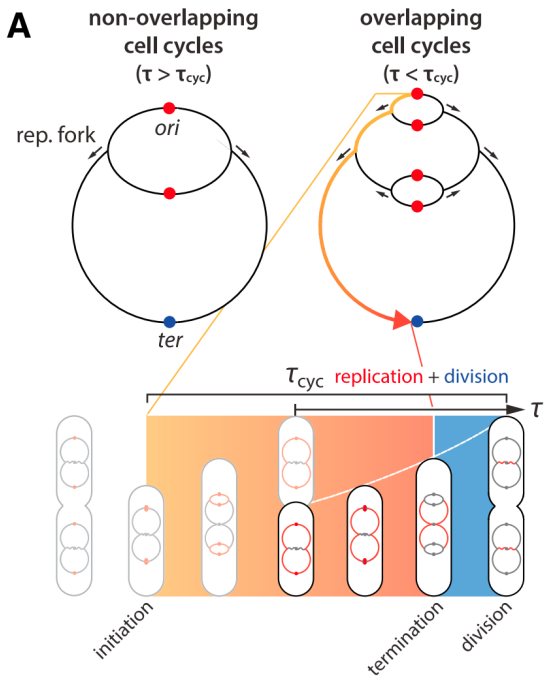

Figure 1.6: **Schematic of overlapping cell cycles and exponential growth** If $\tau_{\text{cyc}} < \tau$, cells initiate and finish replication within the same generation. If $\tau_{\text{cyc}} > \tau$, cells initiate before current generation so that there will be at leat two copies of fully replicated chromosome by the time of entering D period. At any given moment, there could be more than one pair of replication forks on the chromosome. The bottom panel demonstrates the cell cycle periods: the orange region marks the C period, during which a cell division event occured as the generation time τ is shorter than the cell cycle period τ_{cyc}. The blue region between termination of the replication that started before current generation and division is the D period.

growth rates, the only way to adjust DNA synthesis rate is to regulate the frequency of replication initiation to match the growth rate. As a result, replication initiation is a candidate for the key to the coordination of cell growth with DNA replication.

1.3 Background on replication initiation

Based on the aforementioned works and others, Jacob, Brenner and Cuzin proposed the very first model on replication initiation, the replicon model [85].

The motivation for the replicon model was to apply their work on transcription control (or better known as work on the *lac* operon) to replication control. The model can be summarized as the following: 1) the DNA that can be replicated together as a unit is termed a replicon; 2) there is a unique operator of replication (the replicator) on the DNA where replication starts and a structural gene whose product (the initiator) diffuses to the replicator and acts in *trans* to trigger replication (Figure 1.7). Because early evidence showed that chromosome fragments and certain defective phage cannot replicate DNA unless being integrated [86], they also suggested that this regulation has to be positive. It is worth noting that this model predates any molecular studies on repliation initiation yet remains for the most part factually correct [64]. However, the replicon model does not provide any insight into the link between cell growth and DNA replication. The replicon model inspired discovery and characterization of the replicator and initiator, which have laid the groundwork for study of the regulatory mechanism of replication initiation.

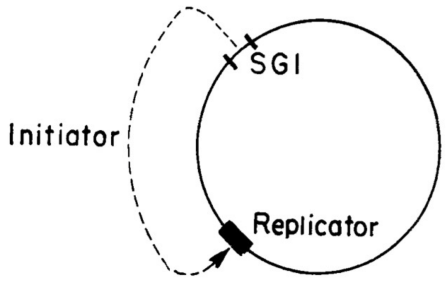

Figure 1.7: **Schematic of the replicon model**. The replicon is a unit of replication and has to be replicated as a whole before the next round of replication of any part. There is a positive signal that initiates replication. The structural gene for the initiator (SGI) encodes the initiator, which diffuses and acts on the replicator sequence. Adpated from [85].

1.3.1 DnaA as the replication initiator

Around the same time when the replicon model was proposed, Masamichi Kohiyama, a student of François Jacobb, carried out a genetic screen for temperature-sensitive mutants of DNA replication. Some of the mutants did not stop DNA replication right after shifting to non-permissive temperature. Instead, these mutants cotinued replication longer than expected. He calculatd the amount of added DNA, which was similar to that reported for the chloramphenicol run-out experiment. This led him to believe that the mutants finished the ongoing replication but could not initiate a new round of replication. They further showed that adding chloramphenicol post shifting the culture back to permissive temperature did not rescue the mutants from replication halt. Hence, they concluded that the initiation requires a protein factor [80, 85, 105]. Some of the mutants Kohiyama isolated were later found to map to *dnaA* gene [62, 64, 80]. More *dnaA* temperature-sensitive mutants became available through systematic genetic screenings [62]. This led to the idea that DnaA is the initiator molecule and its production and availability underlie initiation regulation [64, 94].

DnaA protein can be divided into four domains: a protein-protein interaction domain, a flexlible linker domain, an AAA+ ATPase domain and a DNA binding domain (Figure 1.8). DnaA exist in either ATP-bound form or ADP-bound form.

1.3.2 Characterization of *oriC*

Even though multiple *dnaA* mutants were identified, biochemical characterization of the DnaA protein was not possible since the genomic location of the

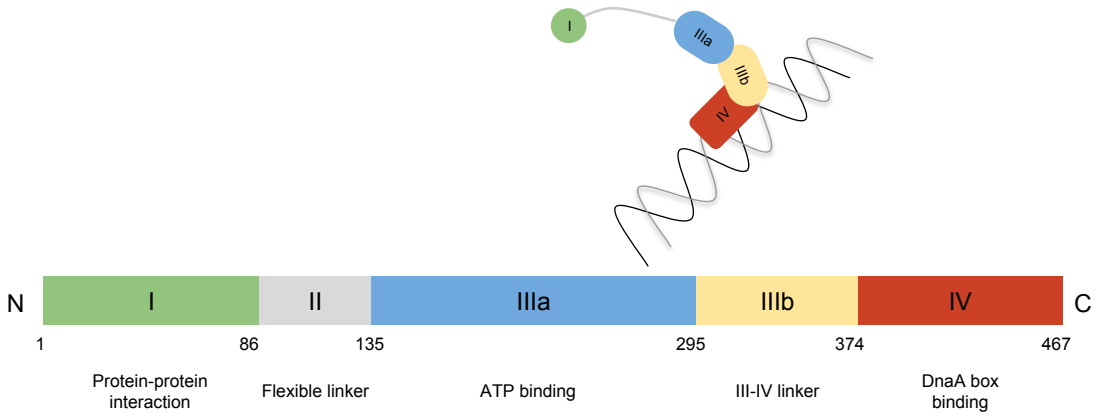

Figure 1.8: **DnaA function domains**. Figure adapted from [64].

replicator (replication origin) was not known. Thus biochemical assays with small purified DNA cannot be carried out. As mentioned above, location of replication origin was first identified by Hirota and colleague by way of marker frequency analysis. They used two strains of phage, Mu-1 and λ to infect cells and make them lysogenic[VIII]. The Mu-1 phage inserts randomly in the host genome whereas the λ inserts in a fixed location on the *E. coli* chromsome. The relative amount of Mu-1 phage over λ phage was detected by DNA hybridization and worked as a measure of relative copy number of Mu-1 integration locus. The gradient of marker frequency was drawn as function of the map distance on the chromosome where *ilv* had the highest frequency, suggesting the replication origin is close to this locus (similar to Figure 1.4) [14] [IX]. Hohlfeld and Vielmetter carried out marker frqeuency analysis using mutagenesis. The mutagen N-methyl-N'-nitro-N-

[VIII] Lysogenic cells contains phage that integrated its genome into the host genome and can replicate with host chromsome normally. In the contrary, the virus excises its own genome from that of the host and replicate independently of host cell cycle in the lytic cycle. The lysogenic cycle is stable whereas the lytic cyclc leads to cell lysis.

[IX] The authors were careful as to limit the study to where integration did not change the generation time. Such physiology alteration could suggest a mutagenic integration, i.e. insertion into the coding sequence of household genes.

nitrosoguanidine (MNG) preferentially induce mutation close to replication forks. Combined with synchronized culture from membrane elution technique pioneered by Helmstetter and Cummings [73], the time-series of mutation frequency acts as a proxy for replication fork progression across the chromosome. They confirmed that the replication origin is near *ilv* [81], also validated by Louarn, Funderburgh and Bird [119].

It was not until Hiraga isolated an F' plasmid capable of replicating extrachromosomally that the replication origin was limited to a shorter genomic region (named as poh^+) [79]. This finding allowed Yasuda and Hirota to refine the region containing replication origin by subcloning the origin-containing region to an extrachromosomal DNA that is otherwise incapable of self-replication [214]. This region was shortened to a 422 bp and sequenced [130, 133, 194].

Marsh and Worcel also mapped the *oriC* to be in a 1kb region by exploiting an initiation temperature sensitive mutant (*dnaC2*) to synchronize the culture. Upshift to non-permissive temperature works like run-out except that downshift back to permissive temperature allows all cells to initiate. The DNA was pulse labeled with small amount of [^{3}H]-thymidine before downshift so that there the amount of radiolabels incorporated into each cell differed. The radioactivity was enriched in a $\sim$40 kb region close to *oriC* since cells ran out of the radiolabel as replication progressed. The replicated DNA were digested with DNA endonuclease HindIII or EcoRI to generate fragments, resolved by molecular weight in agarose gel using electrophoresis. The radioactivity distribution works essentially like a marker frequency histogram to allow determination of *oriC* location [124].

Zyskind and Smith identified the functional replication origin in *Salmonella*

typhimurium as a 296-bp fragment, which contains an unproportionally large number of GATC sites (14 in total). Since the *E. coli oriC* also contains high density of GATC sites, which prompted them to hypothesize sequence conservation of replication origin among bacteria [223]. Further analysis of more bacteria species from the *Enterobacteriacea* family led to the discovery of a conserved 9-bp repated sequence (R-box) among the species sequenced [221, 222], later known as the DnaA box due to its binding to the DnaA (Figure 1.9).

The study of *dnaA* and *oriC* came together under the work of Fuller *et. al.* [53]. Using nuclease protection assay[X], they showed that DnaA bind specifically to 9-bp region in the *oriC*, and the interaction was also visualized with electron microscopy [52]. This work provided solid evidence to support the replicon model, proposed two decaded ago by Jacob and his colleaue [85]. The *oriC* contains multiple of these 9-bp DNA binding sequence motifs (Figure 1.9), later known as DnaA box, and bind about 20-30 DnaA in total [52]. Since the number of DnaA bound to *oriC* outnumbers the number of DnaA box, and the binding curve of *oriC* has a sigmoidal dependence on DnaA amount, it was suggested that the binding of DnaA to *oriC* is copperative [52]. Shortly afterwards, Andrew Wright's group discovered binding sites in *dnaA* promoter region which contains this 9-bp DNA motif sharing the TTAT sequence but differing in the rest [21]. It was later found out that there are many DnaA box motifs in the genome (307 in total estimate) and their distribution is random across the genome [64].

Not all DnaA boxes are the same. Variation on the nucleotide composition to the consensus sequence changes the protein-DNA binding affinity, and the DnaA

[X] Nuclese cannot digest DNA bound by protein and therefore can be used to map sequence bound by proteins.

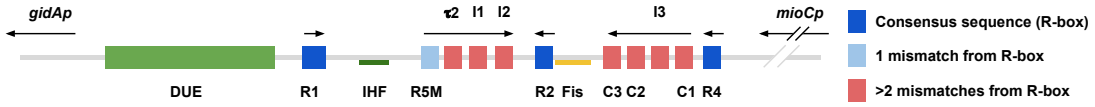

Figure 1.9: **oriC structure**. DnaA boxes (dark blue, light blue and red) are labeled according to the legend. Binding sites for nucleotide-associated protein (NAP) integration host factor (IHF) and Fis are labeled. The duplex unwinding element (DUE) is an AT-rich 70-bp region containing three 13-bp tandem repeats. Figure recreated from [112].

boxes can be divided into high and low affinity groups. Schaper and Messer used electrophoretic mobility assay (EMSA) to show that DnaA binds tightly to the isolated DnaA box from *oriC* or other chromosomal loci (K_D in the range 1-50 nM), they also determined the optimal sequence to be TTA(T)TNCACA [176, 177]. DnaA-ATP can bind to both groups whereas DnaA-ADP can only bind to high-affinity DnaA box. There are 3 high affinity DnaA boxes in *oriC* (R1, R2, R4) and 8 low-affinity DnaA boxes (R5M, τ2, I1-I3, C1-C3) (Figure 1.9).

The exact mechanism by which DnaA initiates replication at *oriC* remains elusive, but it is generally agreed that DnaA binds to *oriC* in a cooperative fashion that forms a filament. Additionally, the binding of nucleotide associated protein (NAP) IHF induces a 180 degree bend in the DNA and *Fis* binding has been shown to work similarly [42, 112, 186]. The binding of DnaA box with DnaA protein also causes a 40 degree bend of the DNA [177]. This discovery confirmed the idea that DnaA binding facilitates the *oriC* for replication: not only by recruiting other proteins to assemble the pre-replication (pre-RC) complex, but also by causing a conformational change that promotes binding of helicase DnaB/DnaC [177].

1.3.3 DnaA-ATP is the active form in replication initiation

Even though genetic studies accumulate valuable information about DnaA and its importance as the replication initiator, little is known about how initiation works on a molecular level. Many of the proteins involved in replication are essential, of which genetic mutations would be most detrimental or even lethal to the cells. Kaguni and Kornberg brought their expertise in biochemistry to the field starting from the 80s. For the first time, Kaguni and Kornberg reconstituted the replication initiation system using purified protein [91]. Despite previous studies with crude protein extract capable of replication [90], this was the first report of a reconstituted system with all components purified. Three classes of proteins were added into the solution of *oriC*-containing plamid: 1) *oriC*-recognizing proteins (DnaA, RNA polymerase, gyrase); 2) replication proteins [DNA polymerase III holoenzyme, single-strand binding (ssb) protein, primosomal proteins] and 3) specificity proteins (topoisomerase I, ribonuclease HI, HU). Kaguni and Kornberg showed biochemically that this reconstituted system was capable of initiating and sustaining replication of the plsamid DNA.

Building on their reconstitute system, Sekimizu, Bramhill and Kornberg systematically studied the functionality of nucleotide binding of DnaA, which has a AAA+ ATPases domain that binds to either ATP or ADP [132]. This led to the discovery that even though both forms can bind to the oriC, only DnaA-ATP can trigger the conformation change in *oriC* prior to replication start [113]. They also showed that both ADP and ATP have strong binding affinity to the DnaA ($K_D^{ATP} = 30$nM, $K_D^{ADP} = 100$nM), and the hydrolysis rate of DnaA-ATP is slow ($\sim$ 30 minutes at 37 °C) [179].

1.3.4 Regulation of initiator DnaA

It has been shown that overinitiation bears deleterious results including collapsed replication fork and even cell death [98, 111, 184]. It is no surprise that initiation is tightly controlled in the cell. As the master regulator for initiation, DnaA is regulated from multiple channels: 1) sequestration of *oriC* to prevent initiator binding; 2) titration of initiator to limit its availability; 3) regulation on initiator expression; 4) regulation on active form of initiator.

Sequestration prevents reinitiation

Fast-growing cells with overlapping cell cycles initiate replication at several origins[XI] simultaneously, a phenomenon known as initiation synchrony. This was first visualzied by Skarstad and her colleagues using the DNA histogram from run-out cell samples. They observed the majority of cells have DNA content corresponds to 2^n of genome equivalent[XII] [185]. If initiation happens asynchronously among origins, certain fraction of cells would have less than the predicted number of origins initiated at the time of shift to run-out conditions and the histogram would deviate from the regular pattern observed (Figure 1.10).

At initiation, the concentration of the DnaA-ATP reaches peak level [139] and all origins initiate repliation simultaneously or within an undetectable short time window[XIII]. A sequestration mechanism is in place to prevent reinitiation

[XI] The number of origins starting replication is given by 2^n_{OC}, where $n_{OC} = \lceil \tau_{cyc} \rceil / \tau$ is the number of overlapping cell cycle as calculated by the ratio of cell cycle period and generation time. Detail derivation is given in Section 2.3.1.

[XII] One genome equivalent is the total DNA, measured in weight or fluorescence intensity of DNA-labeling dyes, that correponds to one complete chromosome.

[XIII] Whether initiation happen all at once or in a cascade is still a matter of controversy [64, 160], but it is of less significance in the current discussion.

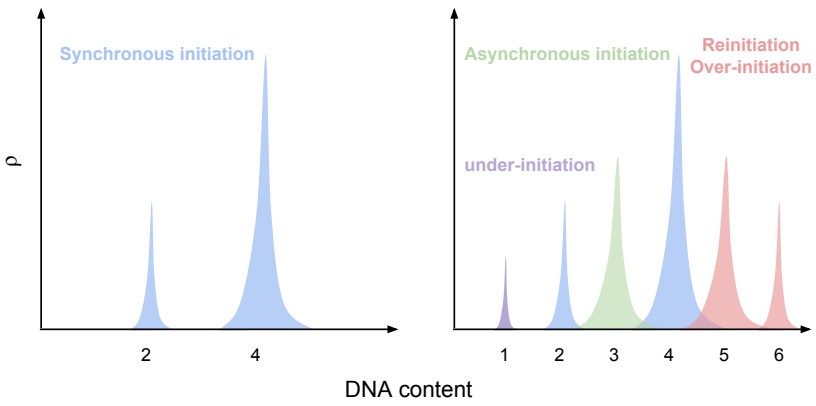

Figure 1.10: **Schematic of DNA histogram from a run-out experiment**. Cells in this Gedankenexperiment have 2 chromosomes and double to 4 after replication. Left: DNA histogram of a normal cell population with two overlapping cell cycles. At the time of shift to run-out, cells have not initiated end up showing 2 genome equivalents while cells that have initiated would eventually finish replication to show 4 genome equivalents. Right: DNA histogram of cells with abnormal initiation. The purple peak corresponding to only 1 genome equivalent is a result of under-initiation. The green peak corresponding to 3 genome equivalents arise mostly due to asynchronous initiation. The red peaks with higher genome contents form as a result of over-initiation or reinitiation (a second initiation event within one generation).

from newly replicated origins. The GATC sequence in *E. coli* is methylated by DNA adenine methyltransferase *dam* [162], and newly replicated DNA remain hemimethylated until the action of *Dam*, with a delay typically on the order of several mintues [32]. These hemimethylated GATC sites recruit *SeqA*, a negative regulator of initiation that physically precludes DNA from occupation of other proteins including DnaA. There are 11 GATC sites in *oriC* and the delay in *Dam* action can last up to one-third of generation time [144, 206]. During this "eclipse" period, the binding of *SeqA* precludes binding of DnaA to the origin [120]. Cells defective in *seqA* or overexpressing *dam* have growth defects due to over-initiation and unresolved replication fork collapse [20, 120, 207].

Initiator titration decreases effective DnaA concentration

The discovery of DnaA boxes in places other than *oriC* [21, 102, 168] led to question whether these other DnaA boxes compete with *oriC* for available initiator. DnaA titration effect was discovered by Hansen using plasmid that contains different regions from *oriC* and showed a derepression of *dnaA* transcription (assayed in a *dnaA-lacZ* fusion expression) as a result of titrating away DnaA that would otherwise bind to and suppress *dnaA* promoters. In addition to confirm that binding affinity variation based on DnaA box sequence [177], they also demonstrated that cooperativity exists in DnaA binding as either R4 or R2+R3 boxes caused little derepression compared to all three combined, as shown previously with *in vitro* binding from Kornberg's group [52, 63]. Of note *datA* (DnaA titration) is a 950-bp region between *glyV* and *amiB-mutL* operons at 94.7 min, has a high titrating capacity with an estimate of up to 300 DnaA molecules [103, 136], although that number was later lowered to 70 by *in vivo* measurements [66].

Similar experiments using DnaA box-bearing plasmid showed that titraiton of DnaA caused change in initiation mass in a growth rate-dependent manner (more dramatic in slow growth condition). The number of extra DnaA boxes introduced to the cell correlates positively with the increase in initiation mass [37]. Reduction of DnaA boxes has the opposite effect: deletion of *datA* did not cause any physiological defects, but did lead to over-initiation and initiation asynchrony [103]. In line with these dicoveries, the titration of DnaA prevents reinitiation after the eclipse period, since regions near *oriC* duplicates (including *datA*) contains multiple DnaA boxes [103, 186]. This idea was corroborated when *datA* was moved close to *ter* hence cannot titrate away DnaA released from *oriC*. Such cells diaplayed

over-initiation. Interestingly, this over-initiation phenotype was reverted when a second copy of *datA* was introduced close to *ter*, arguing for the importance of *datA* copy number in regulating initiation [48].

Autoregulation of *dnaA* transcription

The promoter region of *dnaA* has four DnaA boxes [21] such that its expression is autoregulated and kept at a nearly constant concentration [5, 21]. These data are in line with the autorepressor model proposed by Sompayrac and Maaløe in 1973, which states that the initiator expression is regulated by an autorepressor, whose expression is limited by itself (Figure 1.11). Hansen and Rasmussen demonstrated the autoregulation by using the *dnaA46* temperature sensitive mutant. At nonpermissive temperature, no initiation occured but the mutant DnaA protein continued to accumulate. Upon return to permissive temperature, the cell initiated replication but the protein production was prevented due to an excess of DnaA protein [67]. Similar regulation has been validated for *Bacillus subtilis* as well [142].

Interconversion of two nucleotide forms of DnaA

Newly synthesized DnaA is quickly bound to ATP due to its high affinity to the nucleotides [179] and a high ATP-to-ADP ratio in the cytomplasm (∼3-10) [82, 205]. However, the percentage of DnaA-ATP does not necessarily reflect the ratio of ATP to ADP – DnaA-ATP only comprises about 40% of total DnaA (DnaA-ATP and DnaA-ADP), and undergoes cyclic fluctuation that peaks at 80-90% at replication initiation [139]. This suggests additional regulatory mechanisms

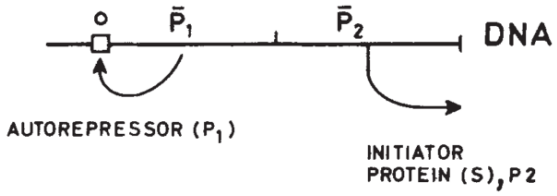

Figure 1.11: **Autorepressor model**. The autorepressor model was proposed as a simple feedback control system to achieve constant protein concentration at different growth rates [191]. The initiator protein P2 is produced in proportion to the autorepressor P1. P1 binds to its own operator to repress its own expression in a concentration-dependent manner. When there is an excess of P1, its own expresssion would be shut down until volume increase reduces its concentration. If the P1 concentration is below the steady-state value, expression would be derepressed and more P1 will be made. The end result is that P1 expression as well as of P2, will fluctuates around the steady-state value. Figure taken from [191].

are in place to actively regulate the level of two nucleotide-bound forms of DnaA (Figure 1.12)

DnaA-ATP hydrolysis is important in maintaining a percentage level of DnaA-ATP lower than that of ATP in the cytoplasm [30, 111]. Higher DnaA-ATP level causes overinitiation, even collapsed fork or cell death [98, 111, 184]. However, Sekimizu and Kornberg have shown that autonomous hydrolysis rate is very slow and cannot explain the dynamic behavior of DnaA-ATP fluctuations [179]. Katayama and colleagues discovered replicative inactivation of *DnaA* (RIDA), which requires the β clamp of pol III to actively decrease DnaA-ATP level during replication [93, 95]. RIDA also involves *Hda* (homologous to *dnaA*), an AAA+ family protein that shares the ATPase motif with DnaA (domain III, Figure 1.8) [97] but binds with ADP specifically [193]. *hda* was found in a screen for DNA segments that rescued temperature-sensitive mutant (*dnaN36*) defective of replication. ADP-*Hda* binds to DNA-loaded β clamp and recruits DnaA-ATP to promotes its hydrolysis [94, 97, 186].

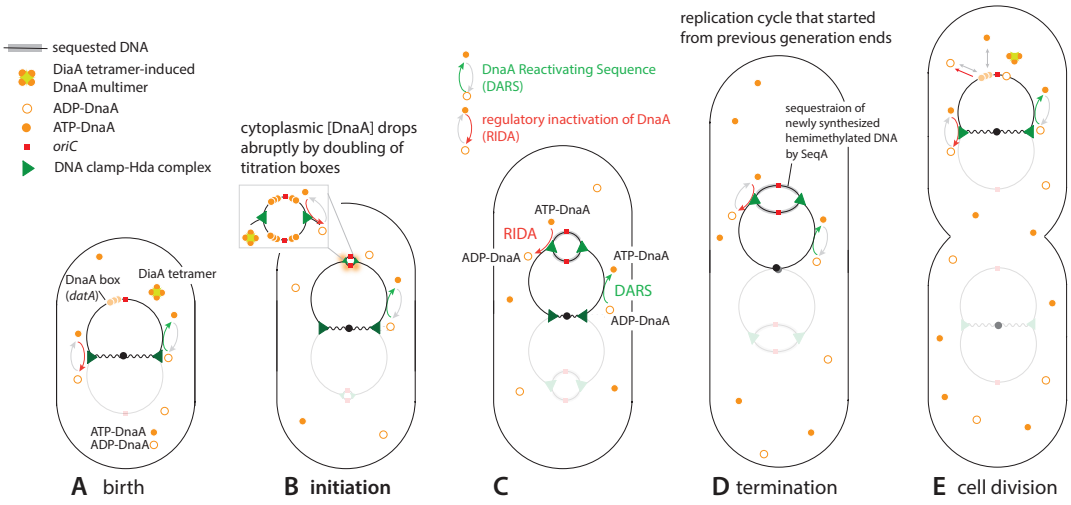

Figure 1.12: **Cell cycle-dependent dynamics of DnaA**. This figure summarizes current understanding of molecular mechanism involved in regulating initiator as discussed in this section. Autoregulation is not shown in this figure but the constant concentration of DnaA, as a result of this, is depicted.

The same group discovered another replication-independent hydrolysis mechanism, which requires the DnaA-titrating locus *datA*, in a process termed *datA dependent ATP hydrolysis* (DDAH) [92]. RIDA was found to be the predominant mechanism for DnaA-ATP hydrolysis and responsible for timely initiation [94, 160].

The reverse process of hydrolysis, also known as DnaA rejuvenation or reactivation, was also discovered. Sekimizu and Kornberg *et al.* reported that DnaA protein bound to the *oriC* DNA can be rejuvenated with ATP, a process mediated by anionic phospholipids *in vitro* [180]. However, the experiment was done *in vitro* and the evidence supporting the role of phospholipid in reactivating DnaA *in vivo* is lacking [64, 213]. Fujimitsu and Katayama reported a novel 70-bp DNA element containing a DnaA box and two DnaA box-like motifs, termed DnaA-reactivating sequence (DARS), that is capable of regenerating DnaA-ATP from DnaA-ADP [50, 51]. DARS1 and DARS2 are responsible for the cyclic fluctuation

of DnaA-ATP level that peaks at replication initiation and deletion of DARS led to diminished DnaA-ATP level. Consequently, replication initiation is inhibited in this mutant strain, and this mutant can suppress overinitiation caused by *seqA* and *datA*, two negative regulators of initiaiton [51].

1.4 Outline

The rest of the thesis following this chapter will be organized as follows (also summarized in Figure 1.13):

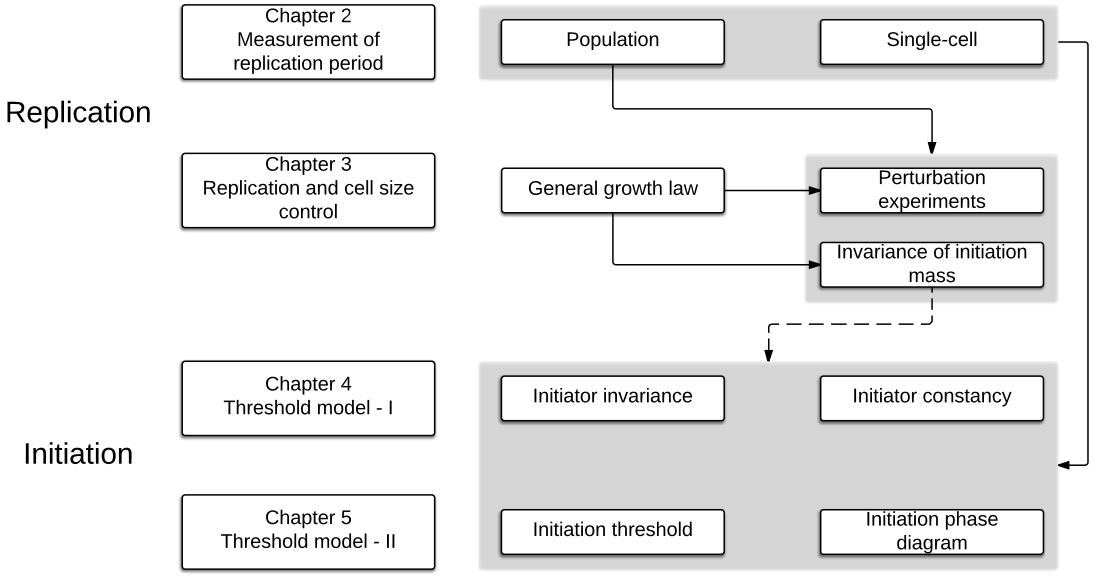

Figure 1.13: **Outline of thesis**.

In Chapter 2, I detail the experimental methods to measure the replication period (C period) both on population level and single-cell level. The quantitative PCR (qPCR) is successfully adpated for measurement of C period from a series exponential culture in different growth media. Multiple attempts at developing single-cell measurement methods are documented with explanation/analysis for

failure. Altenatives are pointed out at the end of the chapter.

In Chapter 3, I apply the C period measurement to a wide range of growth conditions in a collaborative work with others in the lab. We propose a general growth law to project all relevant biological processes to three parameters - unit cell size (S_0)[XIV], growth rate (λ) and cell cycle duration (τ_{cyc}). We apply antibiotics treatment or genetic perturbation to perturb each parameter independently of the other two and verify that cell size did change according to the prediction of the general growth law. We also apply a broad range of inhibition targetting different biosynthetic processes to validate the general growth law. Importantly, we find an invariance of the unit cell size in all perturbation conditions except when replication initiaiton is specifically targeted. This discovery suggests that replication initiaiton has an important role in cell size homeostasis.

In Chapter 4, I delve into the mechanism of replication control to understand the significance of invariant initiation mass in the context of size control. I try to explain the invariance in initiation mass as a result of the control strategy of the replication initiation system. A simple threshold model is proposed: The initiator occupies an *invariant* fraction of the proteome regardless of the growth condition and maintain a *constant* concentration throughout the division cycle. Initiation triggers when *accumulation of initiator reaches a threshold* per replication origin. The first two assumptions of initiator invariance and constancy are systematically tested with a *dnaA* reporter strain.

In Chapter 5, I test the critical threshold of initiator by varying DnaA level while simultaneously measuring initiation mass and DnaA concentration. The

[XIV] Unit cell size is inherenetly linked to the initiation mass S_i, which will be discussed in Chapter 3.

DnaA concentration and initiation indeed followed a simple inverse relationship as predicted. I further explore the effect of ATP-to-ADP ratio on initiation, by overexpressing a nonhydrolyzable ATP-DnaA mutant to change this raito. I manage to explain initiation control with two dimensions (initiator level, ATP-to-ADP ratio) in a phase diagram. An outlook for further work leading to a more complete understanding of the initiation control system is given at the end of the chapter.

Chapter 2

DNA replication measurement

2.1 Abstract

In this chapter, I described the development of replication period (C period) measurement on population level and single-cell level. The population measurement is based on marker frequency analysis of different genomic loci using qPCR. Derivation of expected copy number of genomic locus and its application for C period calculation was provided in detail. qPCR-based measurement was tested and applied to measurement of cells grown with nutrient limitation. Genomic sequencing data was used to show the applicability of marker frequency analysis to a different experimental format. Single-cell measurement was unsuccessful, but multiple methods were described for their comparative advantages and disadvantages, along with analysis for reasons that they failed.

2.2 Introduction

In 1958, Schaechter, Maaløe and Kjeldgaard proposed that "*...at a given temperature, average mass, RNA, DNA and number of nuclei/cell can be described as exponential functions of the growth rate...*" [175]. When a cell culture recovers from lag phase into exponential phase, the cell size becomes larger as generation time gets shorter due to the improvement of nutrient quality. The total DNA content per cell also increases during this transition. It is not clear, however, whether the change in cell size reflects change of growth rate and total biosynthesis activities or the change to DNA content. A long-standing hypothesis in biology is that organism size correlates with its DNA content (C-value enigma, Figure 2.1, [60]) – the more DNA it contains, the larger the individual size. It is therefore tempting to explain the size change in *E. coli* in light of the increased DNA content. An alternative explanation, however, is that cell will continue growth if the DNA replication cannot finish in time and cell gets larger as a result of this delay in cell division.

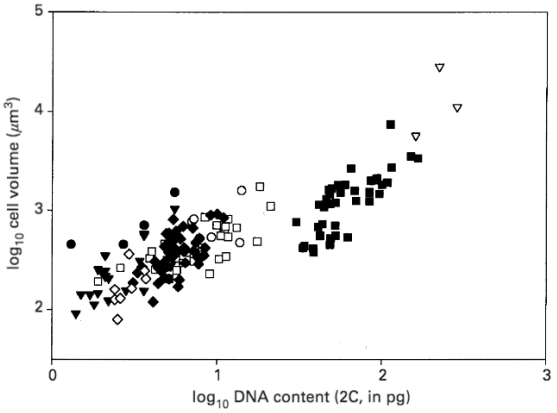

Figure 2.1: **Cell size scale with DNA content**. A cross-species survey of erythrocytes volume and DNA content among 159 species of vertebrates. A positive correlation between cell volume and DNA content is shown. Figure taken from [60].

2.3 Establishing marker frequency analysis to measure C period

In order to systematically test the relationship between cell size and C period, a robust assay system to measure C period is needed. Classic methods include 1) pulse chase assay with isotope-labeled nucleotides in a synchronized culture [72]; 2) pulse chase assay using cells blocked for replication initiation and monitoring for incorporation of labeled nucleotides [38]; 3) flow cytometry to measure population DNA content ditribution and estimate C period [134, 187]; 4) marker frequency analysis by comparing the ratio of origin to terminus using DNA hybridization [6] or microarray [23]; 5) Imaging fluorescent nucleotie incorporation of DNA replication on glass by a receding water-air interphse ("DNA combing") [12, 23, 75, 110]. Isotope labeling is less common nowadays due to its lengthy procedure and inherent risk of radiation exposure, meanwhile flow cytometry requires special instrument setup and further assumptions about C and D period in order to simulate the histogram of DNA distribution within the population [134]. Marker frequency analysis remains to be a simple yet direct method to measure C period.

2.3.1 Calculation of C period from marker frequency ratio

Marker frequency analysis works as each chromosomal locus (marker) has a unique average copy number based on its distance from the origin. This can be seen as the average copy number of any locus is exponentially decreasing as a function of its distance from the replication origin (see Figure 1.4).

We will now derive the expression of average copy number of an arbitrary locus in the genome as it will be important for the replication period measurement in the following section. Two assumptions are made here to derive this: 1) all cells behave identically in a deterministic way, i.e. two cells at the same age are interchangeable; 2) replication forks progress at a constant speed on both sides of the chromosome[I]. In a steady-state culture, the age probability distribution decreases exponentially as cell age (time passed from last cell division, denoted as a), and can be written as

$$\rho(a) = \frac{2\ln 2}{\tau} \cdot 2^{-a/\tau} \tag{2.1}$$

Note the probability of observing newborn cells ($a = 0$) is twice as much as that for dividing cells ($a = \tau$), consistent with the notion that each dividing cell splits into two newborn cells. We then calculate the average copy number of a genomic locus X based on this. Because DNA replication would double the copy number of the locus, the population mean copy number $\langle X \rangle$ is simply the weighted average of cells before and afer replication of this locus. Denote the cell age when locus X is replicated as $t(X)$, then one can write down the average $\langle X \rangle$ as:

$$\langle X \rangle = \int_0^{t(X)} \rho(a)\,\mathrm{d}a + \int_{t(X)}^{\tau} 2 \cdot \rho(a)\,\mathrm{d}a \tag{2.2}$$

$t(X)$ is the sum of time from birth to replication initiation plus from initiation to the duplication of X. Since we have assumed replication fork speed to be constant, the latter time duration can be expressed as the fraction of C period that

[I] There are more evidence [2, 11, 23, 59, 101] arguing against the assumption of a continuous replisome with high processivity based on direct single-molecule imaging or DNA microarray. However, the calculation of C period does not require detailed information about the dynamics of replisome and heterogeneity in different regions of the chromosome.

corresponds to the fraction of genomic distance between *oriC* and X. Therefore $t(X) = t_{\text{init}} + gC$, where t_{init} is the time elapsed from cell birth to replication initiation and g is the normalized genomic distance between *oriC* and X[II]. Putting it altogether we have

$$\langle X \rangle = 2^{\frac{(1-g)C+D}{\tau}} \tag{2.3}$$

The above expression links the average copy number of a genomic locus $\langle X \rangle$ to the duration of generation time τ, replication period (C period) and division period (D period). In other words, if the copy number $\langle X \rangle$ can be measured, one should be able to estimate the cell cycle period, given that generation time τ can be measured. qPCR is widely used in biology to quantify gene dosage and have been applied to measure the copy number of genomic loci in bacteria [78]. It is not trivial, however, to directly measure the absolute copy number of a given genomic locus with qPCR as it requires careful calibration using standards, whereas it is easier to measure the relative copy number of two loci [148]. By measuring the copy number ratio of two loci X_1, X_2, we can solve for C as

$$C = \frac{\tau}{g_2 - g_1} \log_2 \frac{\langle X_1 \rangle}{\langle X_2 \rangle} \tag{2.4}$$

Since generation time can be measured, in theory one only needs to know the relative copy number of any combination of two loci and their relative genomic distance (Δg). The relative copy number can be measured by southern blot (DNA-DNA hybridization) [150] or more recently using qPCR methods [78]. *oriC* and *ter* are ususally chosen because they have the largest possible genomic distance

[II] The normalized distance is calculated as the fraction of the genomic distance relative to half genome size. This way, $g(oriC) = 0$, $g(ter) = 1$

difference Δg and can thus reduce relative measurement error. The ratio of the two is simply

$$\langle ori \rangle / \langle ter \rangle = 2^{C/\tau} \tag{2.5}$$

By measuring ori/ter along with generation time, one can thus calculate C period as

$$C = \tau \log_2(\langle ori \rangle / \langle ter \rangle) \tag{2.6}$$

2.3.2 qPCR experiment setup and data analysis

In a qPCR experiment, the amount of DNA is measured by using fluorescently labeled probes. The DNA amplification process is recorded as a fluorescence time trace over number of amplification cycles and shows a typical sigmoidal shape. The amount of template DNA can be quantified by the threshold cycle (C_t), which denotes the number of cyclces it takes to surpass a threshold value of fluorescence signal. Conceivably, the more templates thera are in the sample the sooner it reaches the threshold cycle. Assuming DNA amount doubles after each round of amplification, the ratio of two DNA fragments in the sample is 2 to the power of their difference in threshold cycle $2^{\Delta C_t}$. However, in reality the amplification efficiency (α) can be less than 2 due to imperfections[III] but can be fit by examining the exponential section of the amplification curve(Figure 2.2). The ratio can thus be adjusted as $\alpha^{\Delta C_t}$. Equate this to the expected ratio based on genomic

[III] PCR reaction can be affected by multiple factors including replication inhibitor or contaminant from DNA sample, false priming of oligos to other location, secondary structure of amplified region etc.

posititions of two loci, one gets

$$C = \frac{\Delta C_t}{\Delta g} \tau \log_2 \alpha \tag{2.7}$$

Thus for a given cell sample, different loci would give different ΔC_t and Δg, but the ratio of these two would be consistent.

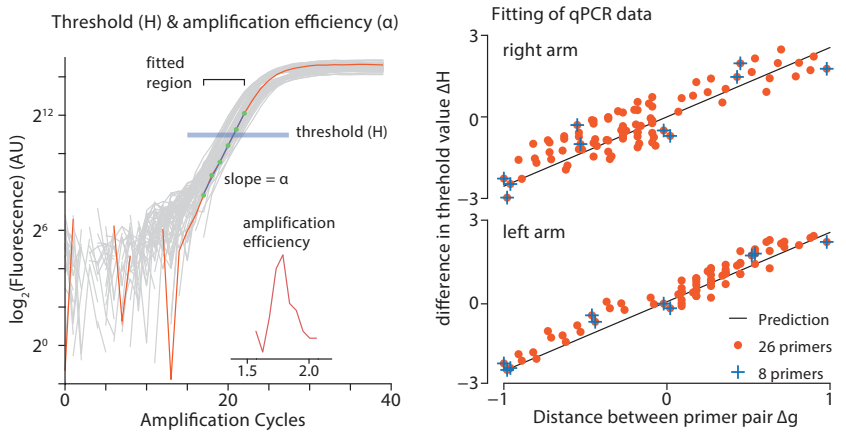

Figure 2.2: **An example of qPCR data analysis from one DNA sample**. Left: Typical fluorescence time trace depicts the process of fitting of amplification efficiency α and threshold cycle C_t. The amplifiication efficiency is fitted with green points in the exponential regime of the trace and the inset shows the distribution of α of all traces. C_t is evaluated as the interpolated cycle number the trace crosses the set threshold (blue bar).Right: ΔC_t vs. Δg for all primers (orange point) are used to estimate the C period by linear fitting (black lines). The estimate from all positions are consistent. A subset of 8 primers (plus signs) give consistent results compared to all 26 primers.

Previously, similar methods have been employed by only using the *ori*/*ter* [78]. To make sure the C period estimate is not biased by the seleciton of primers at specific genomic location, I picked 26 target sites spread over the left and right chromosome arms and measured their relative copy number systematically. All primers were designed such that the size of each amplicon[IV]/genomic locus is about 100 bp

[IV] DNA fragment being amplified in the PCR reaction is referred to as amplicon.

with similar binding propensity (annealing temperature, see Appendix B.1.2). The specificity of PCR amplification was validated by inspecting the melting curves of each reaction. The melting curve showed one major peak across the temperature gradient tested, indicating that the amplification is specific and no significant false-priming or amplification of unintended sequence exists (Figure 2.3). Additionally, the dynamic range of qPCR reaction was probed by serial dilution of samples and testing for corresponding changes in C_t. The system works well with genomic DNA amount ranging 10 - 2,000 ng (Figure 2.3).

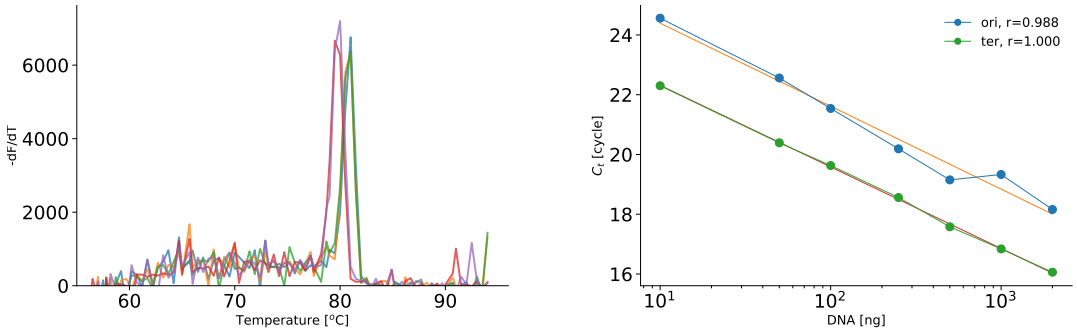

Figure 2.3: **qPCR quality control for amplification specificity and dynamic range**. Left: Melting curve of example qPCR reactions. The change in fluorescence is plotted against temperature. The peak in the curve reflects the annealing temperature of the amplicon. A single peak indicates only one major DNA species being amplified; multiple peaks indicate non-specific amplification due to false priming. Right: Linearity of qPCR reaction for two select primers near *ori* and *ter*. qPCR data were taken from two primers and C_t values were determined for different amount of total genomic DNA input (10, 50, 100, 250, 500, 1000, 2000 ng). The C_t values change exponentially as a function of input DNA (linear in a semi-log plot). The linearity suggests the dynamic range of the system where the readout is prportional to DNA input.

To estimate the threshold cycle C_t, all fluorescence traces were background-subtracted and fit to a single exponential function at the amplification stage, from which the amplification efficiency (the exponential prefactor, $\alpha \in (1, 2]$) was de-

rived (Figure 2.2). Amplification efficiency α is normally distributed with a narrow spread (CV $< 5\%$, Figure 2.2), therefore I decided to use the average as a global parameter for all qPCR amplification . The threshold cycle C_t was determined by interpolating the fluorescence trace to estimate the abscissa (replication cycle number) in two decimal points when it crosses the fluorescence cutoff. The cutoff is set arbitrarily to cross the traces at the exponential section and moving of the actual value within that region does not affect ΔC_t since the traces are all linear in semi-log scale. To make sure all primers are performing equally well, I checked the linearity of ΔC_t over Δg for loci falling onto the left or right arm of the chromosome respectively, and they behaved as expected. Finally, to reduce the workload for each qPCR run, I picked 8 target sites out of the 26. This downsampling does not affect the C period estimate, as they recapitulate the trend of the entire dataset in ΔC_t vs. Δg plot (Figure 2.2).

2.3.3 Extracting C period from whole genome sequencing data

An altnerative to qPCR is to carry out genome sequencing. To test if the same strategy of marker frequency analysis can be used, I took advantage of a set of whole-genome sequencing data genearted by Steven Brown in the lab [26]. In theory, genome sequencing gives the finest detail for all genomic loci which is ideal for marker frequency analysis. However, sequencing depth (number of sequencer reads per region) may show systematic bias and it is incovenient to handle data with single nucleotide resolution as the file size gets too big for efficient computation. In light of this, the sequencing counts were tallied for a bin size of

100 bp using Bowtie 2, which corresponds about 46,000 loci for the entire genome. The histogram of bin counts were then plotted against their genomic location based on the refernce genome sequence (GenBank ID: U00096.3). The histogram can be fit to a decreasing exponential line and gives replication period estimate similarly to qPCR, thus validating the measurement of qPCR (Figure 2.4).

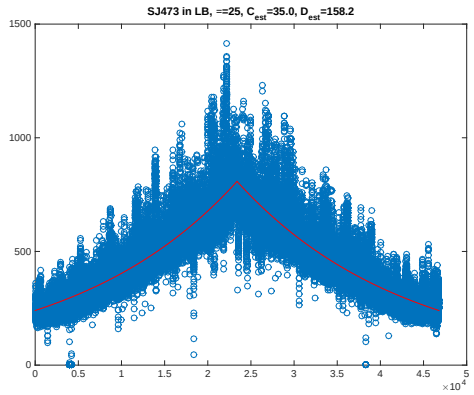

Figure 2.4: **Fitting for C period based on genomic sequencing data**. Exponential culture of a NCM 3722 strain was used for sequencing as reported in [26]. Reads of short pieces were mapped using the reference genome published and counts were binned per 100 bp. Each dot represents a bin count and the overall profile of the count distribution agrees with the idea that the locus copy number as a decreasing exponential function of genomic distance (see Figure 1.4).

2.4 C period as a function of growth rate

Given the qPCR system setup, I set out to measure C period from cell samples grown in different nutrient conditions. Replication period under nutrient limitation have been measured measured previously so my mueasurements can be compared with published results [72, 134]. Indeed the result agreed well with previous report in that C period is about 40 minutes (38.0 ± 4.5 min) when the growth rate is higher than 0.7 doubling/hour and slowly increases as growth slows

down more (Figure 2.5).

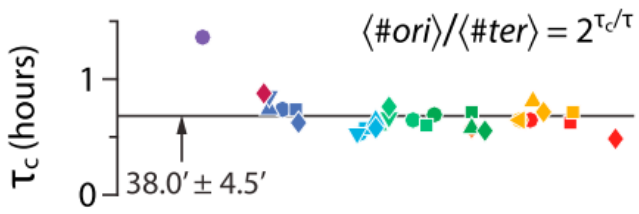

Figure 2.5: **C period at different nutrient conditions**. Exponential culture in different media were taken to prepare genomic DNA samples for qPCR. Each color represents a different nutrient type and different shapes denote biological replicates.

2.5 Attempts to establish single-cell cell cycle measuremnt using gene copy number variation

2.5.1 Motivation for single-cell cell cycle measurement

Even though the qPCR-based C period measurement aligns well with reporter values, there are several drawbacks of the system: 1) the results are sensitive to operation errors, e.g. sampling error, pipetting error, contaminants in the reaction system; 2) the readout is only an average of the sample drawn with no additional information about heterogeneity within the sample; 3) the measurement relies on exponential culture and wildtype gene and chromosome arrangement. This third requirement limits our measurement to normal physiological growth conditions, and cells with rearragned or engineered chromosome have to be treated separtely with different oligo designs, as the relative locations of target

sites may change. This added layer of complexity makes it even more challenging to directly compare the numbers as each oligo design could generate results with a different level of accuracy.

In comparison, imaging-based single-cell measurement can be readily incorporated into single-cell mother machine experiment procedures. Such a single-cell cell cycle measurment will overcome all the abovementioned shortcomings – the results of cell cycle duration of individual cells are available and no assumption about gene or chromosome arrangement is taken. There were several reported attempts on developing imaging-based measurement, mostly relying on tracking replication-related proteins, as discussed in the following sections.

2.5.2 Overview of reported methods

Most single-cell cell cycle measurements are imaging-based. The constraints of imaging-based assays are the temporal resolution of the measurement (imaging frequency) since frequent exposure could cause phototoxicity and cell pathology [196], and decreased level of signal-to-noise ratio due to photobleaching. Santi *et al.* measured C period in mycobacteria by trakcing the replisome (*dnaN-mCherry*). However, the doubling time is in the range of 120-300 minutes, much slower than that in *E. coli* [172]. As such, the imaging frequency can be lowered to avoid the phototoxicity and bleaching, with little compromise on the temporal resolution. Nevertheless, this method cannot be directly applied to my project, especially when cells are growing faster with a doubling time ranging 20-100 minutes. Adiciptaningrum *et al.* used fluorescently labeled *SeqA* to infer cell cycle stage based on the appearance (C period) and disapperance (B or D period) of fluorescent foci

in *E. coli* [1]. However, the work is also limited in slow-growing cells (generation time ranging 60 to 200 minutes). Moreover, there is evidence that *SeqA* does not closely follow the replisome and persists in the B and D period as well [69], arguing against the effectiveness of using *SeqA* for cell cycle measurement. A third and more recent study from Elf group reported cell cycle measurement using a combination of *SeqA*, *oriC* localization and *DnaQ* as the latter photobleaches after a few cycles of imaging and cannot be used to track an entire generation [209]. Therefore the estimation of C period appeared to be indirect and convoluted: they mapped the average replisome number to cell size and deduced initiaiton and termination volume through fitting the replisome-size distribution to a Gaussian distribution. The C period is then calculated as

$$C = \frac{1}{\mu} \ln(n_{ori} \frac{V_T}{V_I}) \tag{2.8}$$

where n_{ori} is the average number of *oriC* at a given growth condition and V_I, V_T are volume at replication initiation and termination, respectively. Note that the second term $\ln(n_{ori} \frac{V_T}{V_I})$ is a fixed number for a given conditions, the estimate assumes that C period is inversely proportional to growth rate with a prefactor, thus attributing all variation in cell cycle to growth rate variation [209].

To summarize, the abovementioned methods either suffered from low signal-to-noise ratio, phototoxicity that makes it unable to do long-term imaging in steady state or indirect measurement based on assumptions unjustified. The poor imaging quality in part stems from the stoichiometry of the replisome-related proteins. Most proteins have a low copy number of 1-6 per replication fork [159] and the

brightness per protein molecule is low such that total fluorescence per focus is low. Another candidate replisome-related protein is *SSB*, single stranded DNA-binding protein, is made into a functional chimera by Andrew Wright [217], but suffers from poor signal-to-noise ratio in imaging from our experience. This is likely due to a wide distribution of stoichiometry (∼30 molecules per replisome) [159] and significant cytoplasmic pool that contributes to high background fluorescence (personal communication with John Sauls in the lab, who pioneered imaging-based cell cycle measurement in our lab).

2.5.3 Cell cycle measurement using chromosomal foci labeling with ParS-ParB system

In light of the problem of imaging replisome-related proteins, we decided to bypass this issue completely by measuring the timing of duplication at specific genomic loci. Duplication of a gene close to *oriC* can be used as a proxy for replication initiation, and duplication of one near *ter* for termination. There have been reports using fluorescence repressor-operator system (FROS) to study chromosome dynamics during cell cycle [217]. Specific locus can be labeled by inserting a *parS* sequence and visualized by expressing *ParB* binding protein fused with fluorescent proteins. Time-lapse fluorescence imaging can therefore provide a time course of the estimated foci number. Furthermore, two foci can be labeled with different *parS* sequences and correspondingly different fluorescence protein. In this way, the time difference of two duplication events can be extracted. Given the genomic distance between the two loci (suppose two loci are on the same side of the chromosome) and assumption that the replication fork speed is constant,

the fork speed and C period can be readily calculated (Figure 2.6).

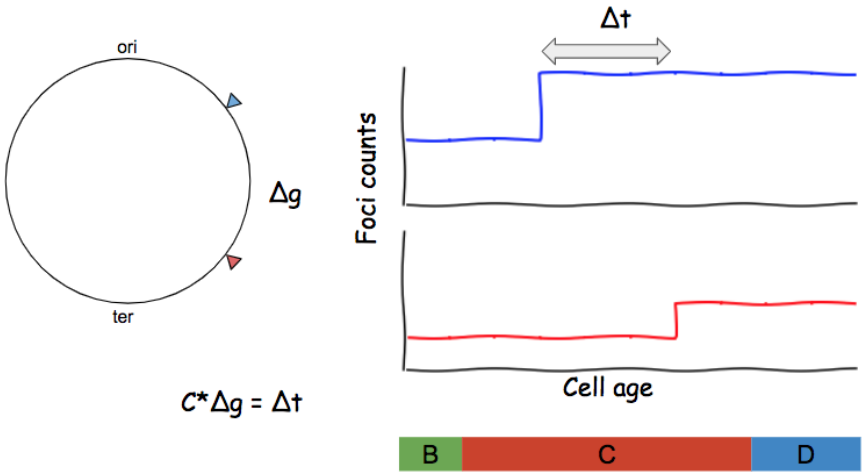

Figure 2.6: **Schematics of C period calculation based on foci number time trace**. In this example, two loci are labeled as blue and red, respectively. The righ panel shows the expected foci number time trace, where the blue focus duplicates earlier than the red. The time difference Δt equals the fraction of C period ($g\Delta C$) corresponding to their genomic distance difference.

A technical problem, though, is the delay between separation of two newly duplicated foci and the catual duplication of the locus due to cohesion of two newly replicated DNA strands. The two foci thus reside in a optically diffraction-limited spot that cannot be resolved. This adhesion would add additinoal uncertainty to the timing accuracy. The problem can be solved if the total fluorescence intensity of the fluorescence spot is proportional to the number of actual genomic loci involved. Unfortunately, the linearity between foci number and fluorescence intensity does not hold because the stoichiometry of the *parS-ParB* system is not clear: even though only a homodimer of *ParB* is bound to *parS* sequence, more *ParB* proteins are recruited to *parS*-neighboring region more than 10 kb to form a large DNA-protein complex in a cooperative manner [19, 22, 170]. Therefore, the exact number of *ParB* is context-dependent and unknown.

In a preliminary analysis, I took data from a two-color *parS-ParB* strain (data from Jia-Wei Yeh, a past lab colleague). Two loci (22' and 84.2') were labeled with GFP- and mCherry-fused *ParB* in this strain. Samples were fixed from a exponential culture in MOPS glycerol and cells were collected on an agar pad after fixation and imaged for cell size, GFP and mChery fluorescence respectively. Since no cell history can be extracted from one time point of fixed sample, cell-to-cell variability was ignored and individual cells are treated as snapshots of one imaginary cell at different cell age/size. Since the size distribution can be expressed as a function of cell age, I assigned each cell its age based on the growth behavior of the imaginary cell. This allowed me to reconstruct a "time series" of foci copy number from an average-behaving cell (Figure 2.7).

The distribution of fluorescence intensity over time is then used to estimate the intensity from a single focus by fitting the data to a sum of two gausssian distributions of the form

$$A_1 e^{-(x-\mu_1)/\sigma_1^2} + A_2 e^{-(x-\mu_2)/\sigma_2^2} \tag{2.9}$$

The two modes of the distribution are two-fold different with $\mu_1 = 1.43$, $\mu_2 = 2.62$. Based on this estimation, the time trace can be thresholded to estimate the ratio of two foci at any given time point in this pseudo-time trace. The time point at which the copy number changes for either foci were estimated. Based on this, the period is estimated to be 45-47 minutes, in accordance to the qPCR measured 44 minutes in the same growth condition. Despite the general agreement, the raw time series trace is noisy and had to be smoothed in order to extract the duplication

timing. The noisiness could come from the limitation of FROS system (undefined stoichiometry) [170]. Another possible source of noise is the cell-to-cell variability in cell cycle timing: this dataset is taken from a population of fixed cells, and their age is mapped to absolute size with the assumption that cells of the same age have the same size and same cell cycle timing. Such assumption may not stand, and this would greatly benefit from real time-series data from the mother machine.

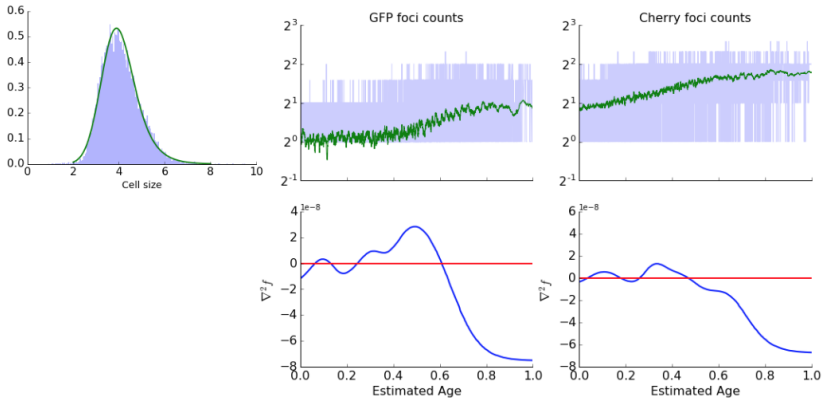

Figure 2.7: **Estimation of C period based on agar pad imaging of foci strain**. A second-order polynomial Savitzky-Golay filter with kernel size 101 was used to smooth the time trace of their foci ratio. To further determine the transition point (or the sharp change in foci), a 1D Laplacian of Gaussian filter was used to identify the time at which their ratio changed (the first one being the duplication of the first of the two foci, the second one being the duplication of the second).

It is worth noting that another FROS system has been used in a recent single-cell study on replication: *MalI* (Maltose inhibitor) is a transcription repressor that binds to *malO*. As discussed earlier, the authors used Equation 2.8 to indirectly infer C period based on the timing of initiation and termination [209].

2.5.4 Cell cycle measurement based on copy number change in reporter gene

Due to the limitation of FROS-based method, we aimed for more direct measurements. We conceived a new method that still depends on the copy number of a specific locus, but does not require direct binding of fluorescent proteins. Instead, a reporter gene can be inserted into the locus of interest. The reporter gene product will likelydouble due to duplication of the reporter gene. The limit of the system is set by the time for gene expression (transcription, translation and protein folding). In the following two sections, two methods based on reporter gene expression will be discussed, one based on bioluminescence and one based on fluorescence.

Bioluminescence-based assay system development

Bioluminescence is based on luminescence produced by chemical reactions catalyzed by luciferase or related enzymes. The substrate D-luciferin (or its analogs) can be supplied in the media and the reaction emits light in ATP-dependent manner. Alternatively the entire luciferae luxCDABE operon can be introduced to act as an autonomous light-emitting system. Since luminescence does not require light activation or excitation, phototoxicity is minimal compared to fluorescence imaging. Therefore, images can be taken as frequently as it's possible for the microscope setup to obtain better temporal resolution. The time delay between gene doubling and increase in luminescence is set by the time to express the luciferase. Given the average rate of transcription (50 nt/s) and translation (15 aa/s), expression of luciferase will take less than two minutes, which is still less than 10% of

the generation time under the fastest growth condition.

However, the brightness of luciferase system is much lower compared to fluorescence imaging. In fact it is mostly used in population-level measurement or biochemical assays whereby the sample (colony, liquid culture, cell lysate) is imaged with a long integration time (up to 30 seconds) with photomultiplier tube (PMT) or charge-coupled device (CCD) camera. The signal was not strong enough to be captured in a typical epifluorescence microscope setup. Only recently has there been reports of single-cell level luminescence imaging [128, 169, 198]. In order to test whether our imaging setup is capable of luminescence imaging, I obtained the luciferase-encoding plasmid pAKlux2 (courtesy of Dr. Susan Golden) and transformed it into *E. coli* MG1655. However, the luminescence can only be seen from a turbid culture or colonies on a plate by eye or CCD camera (Figure 2.8), but cannot be captured by the microscope using either a electron-multiplying CCD (EMCCD) camera (Hamamatsu C9100-13) or a complementary metal-oxide semiconductor (CMOS) camera (Andor Neo). No signal can be detected even with extended exposure time (up to 2 minutes), after which the background signals overwhelms the camera chip. This is probably due to the lower efficiency of light collection compared and orders of magnitude smaller cell number. Even though there is improved version of luciferase or other luciferase species (e.g. Nanolattern), they are not significantly brighter than the other version to warranty successful imaging in our system [169].

This method was also motivated by application of luciferase in yeast cells to measure cell cycle in a microscope setup similar to ours [128]. The difference could be due to a much higher copy number of luciferase protein in yeast cell. The

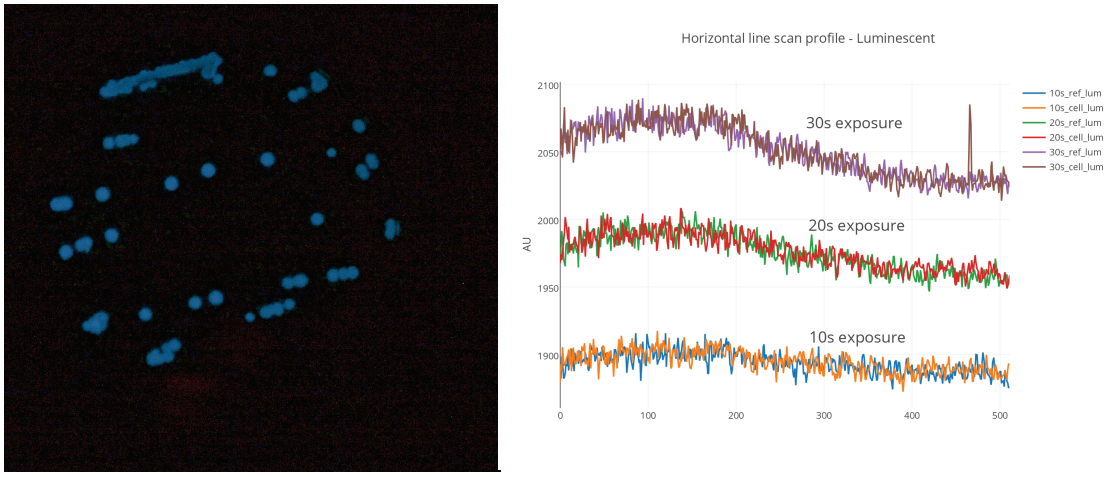

Figure 2.8: **Imaging luciferase strain**. Left: a picture of plate with bacterial colony showing bioluminescence. Photo taken with a commercial DSLR camera. Right: Bioluminescent bacteria were loaded onto glass slide for imaging on a wide-field microscope (Nikon TiE) with no illumination. The plot shows line scans across field of views with or without bacteria. No discernble difference can be observed even with longer exposure, and the rising pixel values are due to increased camera noise due to extended exposure (dark counts) as well as ambient light exposure.

substrate was supplied in excess and not the limiting factor in the imaging process.

Fast-maturing fluorophore-based assay system development

Since luminescence is undetectable in our system, flurescence imaging beomes the only option left. The time delay for a fluorescent protein to report gene dosage change is largely due to the fluorophore maturation process. Besides the typical time of transcription, tranlation and protein folding, the chromophore core needs extra time to go through conformational change before it fluorescence [135]. Depending on the choice of fluorescent protein, the maturation time can range from 5 minutes up to 2 hours. Since the actual maturation time for each protein molecule differs, it is ideal to apply the separation of time scale strategy such that the maturation time is insignificant compared to the typical time range of the process

involved (generation time in this case). The fastest generation time is about 20 minutes, so it is important to have fluorophore with as short a maturation time as possible.

We collaborated with Philippe Cluzel and Enrique Balleza, who have systematically measured maturation time of common fluorescent proteins in live cells. We were able to choose a CFP variant (sCFP3a) and YFP variant (mVenus NB) with short maturation time ($\tau_{1/2}^{\mathrm{CFP}} = 6.4\mathrm{min}, \tau_{1/2}^{\mathrm{YFP}} = 4.1\mathrm{min}$) [9]. The fast-maturing fluorescent proteins greatly shortens the time delay between reporter gene duplcation and the increse in resulting fluorescence.

Once the expression cassette is validated, I introduced both CFP and YFP cassettes separately to a locus near *oriC* as a preliminary test. If this method were successful, one should observe periodic fluctuations of fluorescence intensity and the pixel-average fluorescence peak right after initiation/replication of *oriC*. This is because the output from the reporter gene would almost double after being replicated, and continuous cell growth serves to decrease the concentration afterwards. To test whether this prediction, I performed a mother machine experiment with YFP-*ori* cells grown in MOPS rich glucose. While the total fluorescence intensity tracks the cell growth, the fluctuation of average fluorescence is masked by the generation-to-generation variation (Figure 2.9). It is therefore not straightforward to accurately identify the time point of highest fluorescence concentration corresponding to the duplication in each generation. However, when I looked at the average behavior of the flurescence fluctuation, the periodicity is clear: the autocorrelation of average fluorescence intensity for each cell clearly shows periodic

fluctuation. The autocorrelation is calculated as

$$R(\tau) = \frac{\left\langle \big(I(t) - \mu\big)\big(I(t+\tau) - \mu\big) \right\rangle_t}{\sigma^2}, \tag{2.10}$$

, where μ, σ^2 are the mean and variance of the fluorescence. The autocorrelation function reflects the self-similarity of the fluorescence trace in time. If the protein were produced in a spike, the autocorrelation would decay as the protein is being diluted through cell growth. In contrary, if the level of the protein is fixed, the autocorrelation would be close to 1 as it is always identical regardless of the time-delay. The fluctuation in autocorrelation curve suggests periodic fluctuation of the fluorescence time trace, and the spacing between adjacent peaks correspond to 25 minutes, which is the generation time. This periodicity suggest that the YFP reporter is indeed fluctuating in concentration, as there is inter-initiation time is equivalent to genearation time on average. However, the level of fluctuation within each generation is not big enough to easily extract the initiation timing information directly from this time trace. Similar experiments have been done by Sander Tans group in slow growth condition (0.6 doubling/h), where they showed doubling of the production rate in fluorescence time trace. They also noticed that at higher growth rates the noise is higher [208], potentially prohibiting extraction of cell cycle timing.

2.6 Conclusion

I established population-level C period measurement based on marker frequency analysis using qPCR. C period was found to be constant at growth rate

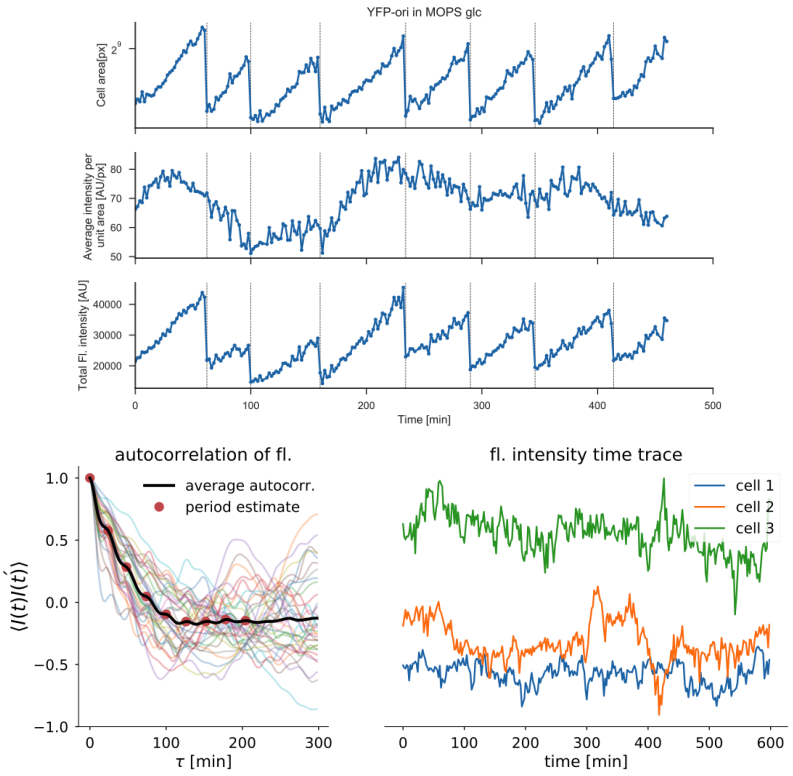

Figure 2.9: **Fast-maturing fluorescent protein time trace and fluctuation**. Top: example time trace of cell growth (top), average YFP fluorescence (middle) and total fluorescence of a YFP-*ori* strain grown in MOPS rich glucose. The generation-to-generation fluctuation in the average fluorescence dominates the Bottom left: Autocorrelation of average fluorescence shows fluctuation with a period correponding to generation time τ. Bottom right: select single-cell time trace of three different cells.

down to 0.7 doubling/hr and increase monotonically afterwards. I had multiple attempts were made to measure single-cell level C period that did not work. All methods are summarized in the table that follows.

Table 2.1: Comparison of different C period measurement methods

Method	Pros	Cons	Feasibility	Comment
		Marker Frequency Analysis		
qPCR	Robust against small errors	1) experiment optimization takes time; 2) relies on good measurement of doubling time; 3) measurements limited to popualtion level; 4) requires exponential culture.	very feasible and currenlty used	
Sequencing	Enormous amount of data	1) not cost-effective; 2) long processing time.	feasible, requires special sample prep	only analyzed data
Northern blot	Tried and true system	obsolete	requires special setup for radioactivity	did not try
		Chromosome locus tracking		
ParS-ParB	No delay due to fluorophore maturation	1) *ParB*:*ParS* highly variable and unknown 2) requires plasmid that encodes ParB-FP; 3) intensity weighting not working well.	doable with minimal genetic engineering	only analyzed data
MalO-MalI	No delay due to fluorophore maturation	requires additional plasmid that encodes MalI-FP	doable with minimal genetic engineering	did not try
Luciferase reporter	No laser excitaiton and phototoxicity	too dim to be captured	not possible with current version of luciferase and imaging system	did not work
FP reporter	Avoids foci counting and correction	1) requiress fast-decaying mRNA; 2) delay due to fluorophore maturation	requires systematic sequence optimization	did not work due to design issue
		Replisome tracking		
SSB-GFP		low SNR due to high copy number	data anlaysis compromised due to low SNR	
YPet-dnaN	Good SNR and works well with intensity weighting due to small stoichiometry.		1) very feasible and currently used; 2) minimal genetic engineering	

Chapter 3

Physiological control of cell size and the growth law of a fundamental unit of cell size

3.1 Introduction

Under different nutrient conditions, *E. coli* cell of the same genetic background can vary in size for almost an order of magnitude (Figure 3.1). This was best characterized by Schaechter, Maaløe and Kjeldgaard in their seminal paper that studied the nutrient-imposed growth rate dependence of cell size and chemical composition. The exponential relationship between nutrient-imposed growth rate (hereinafter referred to as the growth rate) and cell size is widely known as the nutrient growth law, wherein cell size is an exponential function of the nutrient-imposed growth rate regardless of the detailed chemical composition [175]. This

paper laid the foundation of bacterial physiology, of which a big part consists of studying the nutrient-imposed growth rate dependence of physiological parameters.

However, one may wonder if the growth rate alone is sufficient to describe the physiological state of any given cell. In other words, given two cells growing at the same nutrient-imposed growth rate, are they physiologically equivalent? One missing component is the cell cycle, as I would expect cells to seek a balance between the need for DNA replication and biomass accumulation. This would ensure maximl growth in nutrient rich environment without compromising the genetic integrity. As Helmstetter and Cooper discovered, cell cycle duration (both C and D period) are relatively invariant and independent of the growth rate at fast growth conditions (Figure 2.5; [72]). It is not known if or how cell size would change with perturbation aimed at changing the cell cycle duration. Whether there exists some degeneracy of cell cycle or other parameters in the nutrient growth law.

In this chapter, I will describe the experiments lead to our discovery of the growth law of a fundamental unit of cell size (hereinafter referred to as the general growth law for simplicity), which decomposes cell size into three independent physiological parameters, growth rate λ, cell cycle duration τ_{cyc} and unit cell size S_0. These three parameters were perturbed selectively and independently, and the cell size changed predictably according to the general growth law. Furthremore, we also showed that the unit cell size S_0 is remarkably invariant when cells are challenged with a broad spectrum of antibiotics or growth inhibitions. Since the unit cell is inherently linked to the initiation mass, this confirms the notion of constant initiation mass in a much wider range of conditions than when it was first proposed for by Donachie [43]. This general growth law can be used to explain cell

size change in a series of experiments we conducted here and by others.

3.2 The growth law of a fundamental unit of cell size

3.2.1 Revisiting the nutrient growth law

To verify the nutrient growth law using our own measurments, we did a series of turbidostat (TSTAT) experiments in media comprised of different carbon sources and supplements (Table 3.1). Each vial of cell culture was kept in exponential phase for at least 8-10 generations to ensure steady-state growth as monitored with constant optical density (OD) measurements with TSTAT. The slope of the growth curves in semi-log scale represent the growth rate λ, whose values are estimated by linear fitting of log-transformed curves (Appendix A.1.2). Cell samples were fixed with formaldehyde for phase contrast imaging to measure the average cell size (see Appendix for detailed protocol). Our results agree well with the growth law as cell size increases exponentially depending on the growth rate (Figure 3.1).

The nutrient growth law describes cell size as an exponential function of the growth rate, which can be expressed as

$$S(\lambda) = S_0 e^{\gamma\lambda} \tag{3.1}$$

where $S(\lambda)$ is the average cell size (=volume) of the steady-state population at the nutrient-imposed growth rate λ. Our data indeed showed an exponential rela-

Table 3.1: List of growth media and their respective growth and size measurment

Growth Media	**Growth Rate** [doubling/hr]	**Cell Size** μm^3
TSB	1.98 ± 0.15	3.45 ± 0.52
MOPS rich glucose	1.86 ± 0.07	2.93 ± 0.20
MOPS glucose + 12 a.a.	1.52 ± 0.08	1.60 ± 0.01
MOPS glucose + 6 a.a.	1.22 ± 0.06	1.12 ± 0.23
MOPS glucose	1.02 ± 0.09	0.98 ± 0.04
MOPS glycerol	0.71 ± 0.10	0.55 ± 0.04
MOPS sorbitol	0.65 ± 0.08	0.61 ± 0.02

tionship between size and growth rate. We can see that S_0 is the y-intercept and γ is a exponential prefactor. Theoretically, S_0 is the average size of the cells as the growth rate approaches zero. Biologically, this theoretical limit would only be reached if the growth condition becomes so poor that cells can barely start growth and the cell cycle. Importantly, when the work by Schaechter, Kjeldgaard, and Maaløe was published, it was unclear whether the constant γ has any biological underpinnings. The only certainty is that γ should have the unit of time since the unit of the growth rate is inverse time ($\lambda = \ln 2/\tau$, where τ is the average mass doubling time of the steady-state population).

Besides cell size and growth rate measurement, we also measured cell cycle period ($\tau_{\text{cyc}} = C + D$) for these conditions. C period was measured using qPCR and D period using image cytometry respectively (see Appendix B.1.3 for detail). The values for τ_{cyc} agrees well with best-fit numerical values for the exponential prefactor γ (Figure 3.1). Comparintg this with Equation 3.1, we can see that

$$S(\tau) = S_0 2^{\tau_{\text{cyc}}/\tau} \tag{3.2}$$

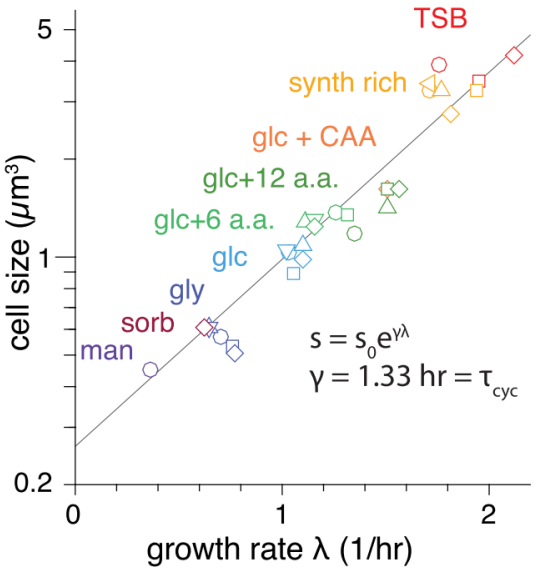

Figure 3.1: **The nutrient growth law**. Cell size is exponentially dependent on growth rate conferred by specific type of media (differentiated by color in this plot). The slope of the curve is γ and can be fitted as shown here. This has been used to derive τ_{cyc} before the application of flow cytometry and qPCR techniques. Different symbol types represent biological replicates.

where τ_{cyc} is the duration of cell cycle that equates the prefactor γ and τ is the generation time that follows $\lambda = \ln 2/\lambda$. This observation supports Donachie's interpretation of the nutrient growth law as an expression of cells initiating at a fixed size and grow exponentially for a combined (C+D) period before division [43]. Nevertheless, the assumption that τ_{cyc} is constant has been taken for granted for any condition without experimental support other than Helmstetter and Cooper's, which focused on nutrient limitation [72]. We thus questioned whether τ_{cyc} is always a constant value and cell size can be expressed as a univarite function of growth rate λ or generation time τ.

3.2.2 Translational inhibition reveals the unexpected size changes

The importance of the nutrient growth law lies not only in its quantitative predictive power, but also that it established the practice of studying the growth rate dependence of physiological parameters. Despite this widely adopted view, one may question if the growth rate is sole variable to proxy the physiological state of any cell [24, 174]. As detailed below, we found that the growth rate alone is not enougth to encapsulate cellular physiology when it comes to cell size.

Since there has not been systematic study of cell size in growth inhibition conditions, we decided to treat cells with sublethal dosage of chloramphenicol, an antibiotics that blocks translation by binding to the ribosome. As the ribosome represents the total biosynthesis capacity and the cell size is the product of all biosynthesis, one would predict cell size to be decreasing under growth limitation. In contrast, when cells were treated with sublethal dosage of chlorampenicol, the change in size was unpredictable. In rich growth media conditions, cell size decreased as the growth rate; in poor growth media conditions, cell size increased even though the growth rate decreased; in intermedidate media growth conditions, cell size did not change as the growth rate decreased, similar to previous findings (Figure 3.2) [10]. This complex behavior can no longer be explained by the nutrient growth law, as cells with the same growth rate could show difference in size depending on its growth media and chloramphenicol concentration.

Remarkably, this cannot be explained solely by the change in ribosome content either, as ribosome fraction increased monotonically following the increase

in chloramphenicol concentration (Figure 3.2). Under chloramphenicol actions, a fraction of the ribosome becomes inactivated. To compensate for this, cells increase their ribosome content until a new equilibrium is reached where the active ribosome fraction can support the biosynthesis activities corresponding to the new growth rate (Figure 3.2). It is nevertheless not possible for us to predict how the active ribosome fraction change in response to different level of inhibition. In order

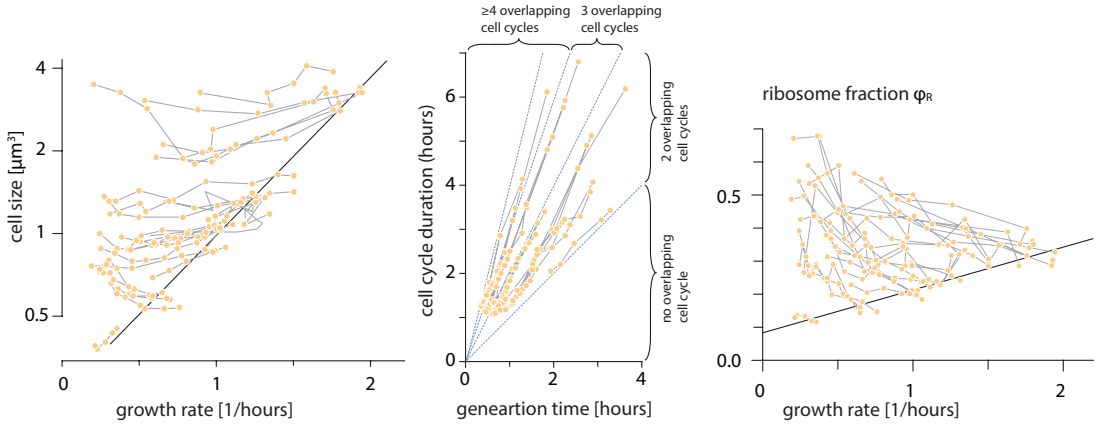

Figure 3.2: **Physiological changes upon chrloramphenicol treatment**. Left: cell size change upon chloramphenicol treatment. Each line series represents a group of experiments with the same nutrient condition and increasing concentration of the antibiotics starting from a control sample without antibiotics (same for the next two panels). Samples without chloramphenicol (mostly rightmost point with highest growth rate of the series) did follow the nutrient growth law as an exponential line in this graph (solid line). Center: changes of cell cycle duration τ_{cyc} as a function of generation time τ, from the same cell samples as the left panel. Depending onthe ratio of τ_{cyc}/τ, the number of overlapping cell cycle n_{OC} can be determined. Right: ribosome fraction as measured by RNA/protein ratio (see Appendix B.1.4 for detail) all increased after chloramphenicol treatment as a compensatory effect of inactivated ribosome.

to explain this unpredicted change, we then examined the cell cycle parameters of all the samples in the above experiments, measuring their cell cycle duration τ_{cyc}. Unlike in the case of nutrient limitation (Figure 3.1), τ_{cyc} all increased when cells were challenged with chloramphenicol. When the τ_{cyc} is plotted against the

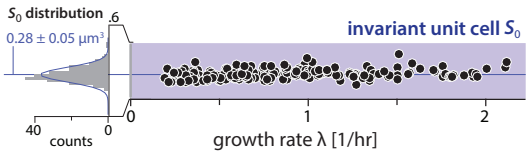

Figure 3.3: **Rescaling of cell size under chloramphenicol treatment**. Cell size is rescaled to calculate S_0 accoring to Equation 3.3, and plotted against the growth rate for each measurement. The distribution of S_0 is shown on the left.

doubling time τ, the two variables are roughly linearly related to each other (Figure 3.2), similar to what was reported for slow growth conditions without growth inhibition [70, 209]. From Equation 2.3.1, we know that $2^{(C+D)/\tau}$ represents the average copy number of *oriC*. If the cell size is rescaled by this number of *oriC*, the normalized cell size $S/oriC$ becomes surprisingly constant irrespective of the growth rate, the type of growth media or the inhibition conditions (Figure 3.3). Rewriting Equation 3.2, we have

$$\frac{S}{2^{\tau_{\rm cyc}/\tau}} = S_0 \tag{3.3}$$

where S_0 remains constant under chloramphenicol treatment.

3.2.3 The general growth law and the unit cell

Even though Equation 3.2 is similar to the nutrient growth law in form, it implies that cell size can change by varying any of the three variables $\tau_{\rm cyc}$, λ and S_0. The former two variables represent cell cycle progression and rate of biosynthesis respectively. S_0 denotes cell size at initiation. This can be understood from the perspective of a single cell going through exponential expansion, which can be

written as

$$S_d = S_i \cdot 2^{\tau_{cyc}/\tau} \tag{3.4}$$

in slow growth conditions where $\tau_{cyc} < \tau$. Under overlapping cell cycle regime ($\tau_{cyc} > \tau$), this relationship becomes

$$S_d = 2^{n_{OC}-1}\, S_i \cdot 2^{\tau_{cyc}/\tau} \tag{3.5}$$

since the cells would have divided $(n_{OC}-1)$ times by the time the last division event corresponding to the current cell cycle happens (Figure 1.6) and $n_{OC} = \lceil \tau_{cyc} \rceil / \tau$ denotes the number of overlapping cell cycles. The average cell size $\langle S \rangle$ can be related to the division size S_d as

$$\begin{aligned} \langle S \rangle &= \int_0^\tau \rho(a) S(a)\, \mathrm{d}a \\ \rho(a) &= \frac{2 \ln 2}{\tau} 2^{-a/\tau} \\ S(a) &= S_b 2^{a/\tau} = S_d 2^{a/\tau - 1} \end{aligned} \tag{3.6}$$

Simplify this one gets $\langle S \rangle = \ln 2 S_d$. Therefore one has the following

$$\langle S \rangle = \ln 2 S_i \cdot 2^{\tau_{cyc}/\tau} \tag{3.7}$$

Comparing Equation 3.7 with Equation 3.2 we can see that S_0 is prportional to the initiation size S_i. S_0 is the unit cell size that represents a minimal cell with a complete set of chromsome. In balanced growth conditions, any given cell would contain a set number of unit cells growing in parallel and the difference in size is

rooted in the difference in the number of unit cells contained. Even though the general growth law is similar in form to the nutrient growth law, it impiles that the cell size is a function of three physiological parameters: the global biosynthesis rate λ, the progression of replication-division cycle τ_{cyc} and replication initiaiton S_0. We further postulated that these three parameters can be changed independently from each other and the cell size should change according to the general growth law.

3.3 Perturbation to nutrient condition

We sought to perturb each of the three physiological parameters. Since nutrient limitation is shown to only vary growth rate but not the other two, we focused on perturbing the cell cycle progression and the initiation process and test if the general growth law can predict the change in size.

A trivial case is to observe size change with different growth media. In this case the general growth law is reduced to the nutrient growth law as only growth rate is changed. We systematically varied carbon source and nutrient additives to achieve a range of growth rate between 0.4 - 2.1 doublings/hr. The resulting cell size vs. growth relationship is consistent with the original growth law where cell size is shown to be an exponential function of the growth rate (see 3.2.1; Figure 3.1). The replication period was also measured for the same set of samples, and C period was also constant as utnil growth rate drops below 0.7 doubling/hr (2.5).

3.4 Perturbation to the cell cycle progression: Changing C period

Next we sought to change cell cycle progression with targeted perturbation aimed at changing the cell cycle duration τ_{cyc}. This section presents work focusing on perturbation to the DNA replication, while there is an accompanying effort in perturbing the division progression (D period) that I did not partake. The reader is referred to the appendix in this chapter 3.7.1 for details.

3.4.1 Thymine limitation

To study the effect of C period on cell size, we started with thymine limitation. It is a classic method for observing replication progression where the thymidylate synthase gene (*thyA*) is inactivated or mutated. This thymidylate synthase is required for the reaction of converting deoxyuridine monophosphate (dUMP) into deoxythymidine monophosphate (dTMP) and eventually thymine. Thymine is one of the four nucleotides that forms the monomer of DNA. Cells with deficiency in this enzyme have to rely on externally supplmented thymine to maintain a nucleotide pool in order to keep DNA replication (Figure 3.4) [164]. External thymine can be radiolabeled to facilitate tracking of replication progression by means of observing radioactivity changes [109, 121]. Replication progression is thus limited by the size of nucleotides pool, which in turn can be modulated by the amount of thymine provided in the growth media [154, 218, 219]. In addition, thymine limitation does not seem to affect growth rate under mild thymine-limiting conditions, whereas more severe limitaiton leads to thymine starvation and even

thymineless death of cells [219]. In the former case, cell size change can be associated with change in C period but not the growth rate.

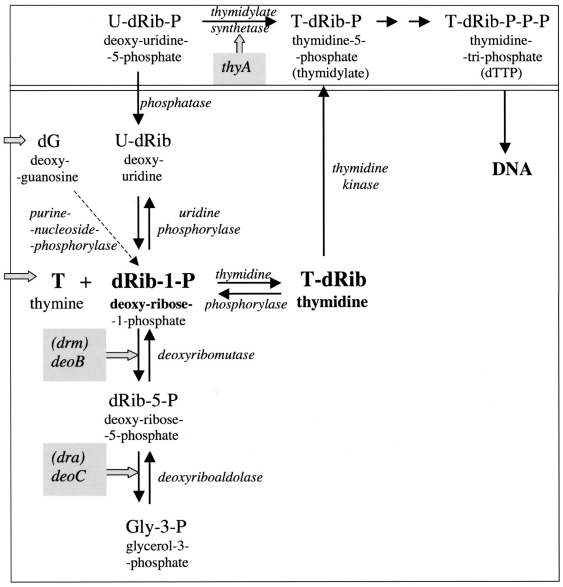

Figure 3.4: **Schematics of thymidylate synthesis** Thymine can rescue the deficiency of thymidylate synthetase (*thyA*) and support the demand for thymidine from DNA replication. Figure reproduced from [219] .

Indeed when we grew $\Delta thyA$ cells in TSTAT, there was a range of thymine concentraiton where the cell culture remains in steady state. Further limitation leads to significant slowdown in growth and deviation from exponential growth (Figure 3.5). Under thymine limitation, C period increases monotonically as the concentration of supplemented thymine decreases. More importanty, cell size also increases as C period and can be fit to single exponential function of C period (Figure 3.5). A logical deduction is that the number of overlapping cell cyclce would increase as $n_{\mathrm{OC}} = \lceil C + D \rceil / \tau$. To directly visualize this, a parS-parB strain was made where parS is inserted close to *oriC* and the binding protien parB is fused with mCherry. Because the number of *oriC* increases as the overlapping cell cycle number ($\langle ori \rangle = 2^{C+D/\tau}$), increase in C period only would lead to more *oriC*

per cell. When imaged, cells with higher level of thymine limitaion (less thymine) did show more foci, confirming the qPCR-based C period measurement.

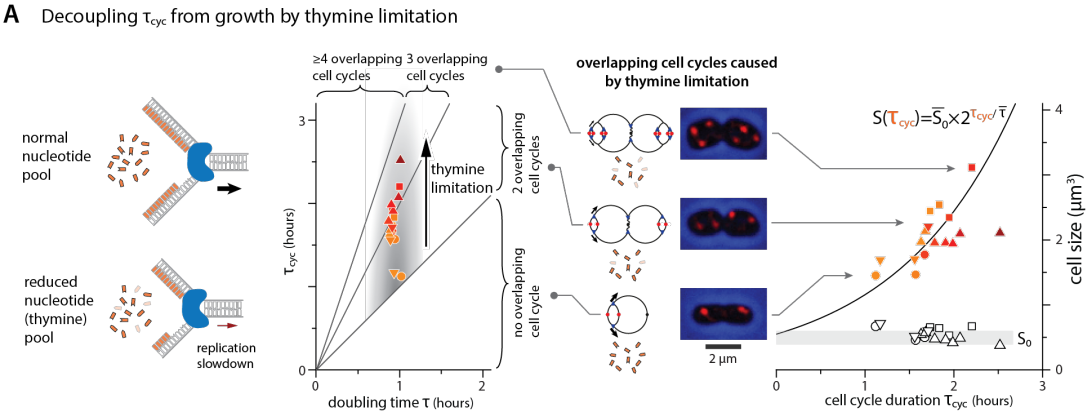

Figure 3.5: **Thymine limitation** Left: thymine limitation reduces the nucleotide pool and replication slows consequently. Middle: τ_{cyc} increases in thymine limitation while t remains unchanged, increasing the number of overlapping cell cycles. Chromosome schematics and cell images with foci qualitatively show increasing number of orisas a result of multifork replication. An odd number of foci in some cell images are possibly due to cohesion and/or stochasticity in replication initiation [217]. Right: cell size increases exponentially with tcyc in thymine limitation, as predicted by Equation 3.2 (solid line, no free parameters). The empty symbols are the cell size per *ori* (S_0), and the thickness of the gray band denotes ±SD. Symbol shapes reflect biological replicates and the symbol colors indicate the level of thymine limitation.

3.4.2 Alternative replication inhibition mimicks thymine limitation

Despite the agreement of cell size chnage accompanying the extension of C period, thymine limitation also caused a change of cell shape. It can be seen from the images that cell rounds up and deviate form the typical rod shape in $\Delta thyA$ cells (Figure 3.5). This is due to the link between thymidylate synthase and cell wall synthesis [219]. To rule out the possibility that this change in cell

wall synthesis rather than increase in replication period causes cell size change, we went on to look for other methods that changes C period similarly but not cell wall synthesis. One of the candidates was hydroxyurea (HU) treatment.

HU acts on ribonucleotide reducatase (RNR) pathway and generally limits the nucleotide pool size for replicatoin [141]. Cells appear normally shaped under HU treatment. However, qPCR experiment revealed that C period increased as the dosage of HU. In addition, cell size also increased as predicted by the general growth law (Figure 3.6).

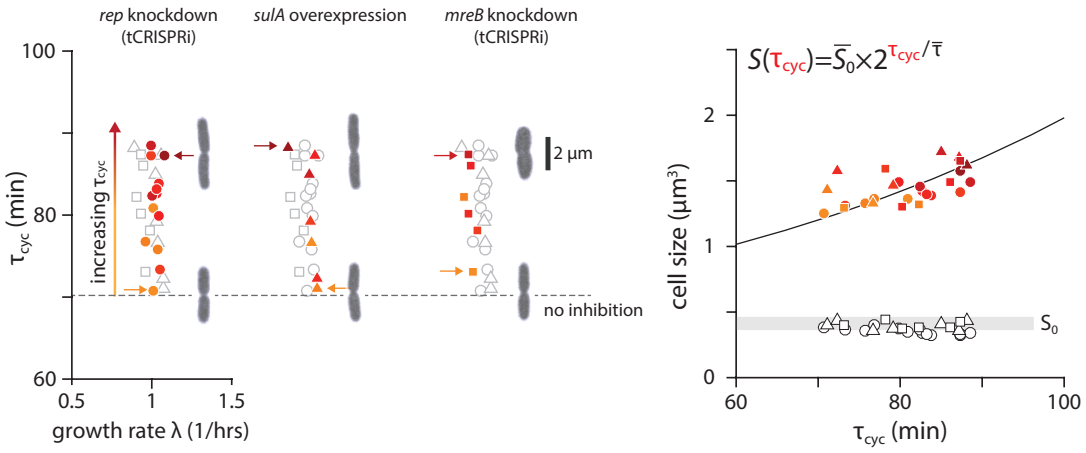

Figure 3.6: **Cell size changes by perturbation to the cell cycle**. Left: cell cycle duration $\tau_{\rm cyc}$ increases as a result of targeted perturbation to replication (circles, *rep* tCRISPRi knockdown), cell division (triangles, *sulA* overexpression) and cell shape change (squares, *mreB* tCRISPRi knockdown). The color reflects the level of perturbation with darker color indicating higher level of perturbation. Horizontal dashed line indicates wildtype cell cycle duration time at ~ 70 minutes. Right: cell size of the same samples as in the left panel using the same symbol. Cell size increases exponentially as a funciton of $\tau_{\rm cyc}$, consistent with the prediction of the general growth law (solid line). Cell size normalized by their number of unit cells $2^{\tau_{cyc}/\tau}$ collapsed onto one horizontal line, showing the constancy of the unit cell size in all cell cycle perturbation conditions.

To preclude the possibility of any drug side effect, we did targeted knock-down by employing the tunable CRISPR interference[I](tCRISRi) developed by Xin-

tian Li (Jun lab, [115]). The system expresses dCas9 protein in a titratable fashion by controlling the inducer arabinose level. dCas9 blocks transcription hence reducing expression level of the gene of interest, as targetd by the short guide RNA (sgRNA). By targeting the *rep* gene, which encodes a helicase in the replisome, the replication is slowed down [108]. Similar to the antibiotics experiment, a dose-dependent slowdown of the replication was observed. In the meantime, cell size also increased in a similar fashion as the other two methods (Figure 3.6).

3.5 Perturbation specific to unit cell

After showing that both growth rate (λ) and cell cycle duration (τ_{cyc}) can be changed independently and causing cell size to change according to the general growth law, we sought ways to alter the unit cell size. Since unit cell is related to the replication initiation, we shortlisted a few candidate genes that could change unit cell size.

The above results all confirm that τ_{cyc} is an independent physiological parameter in cell size control. Even though the molecular mechanisms of delaying replication and division can differ significantly, the effect on cell size is only reflected through change of τ_{cyc} and interchangeable.

This part of work was mostly carried out by my colleague and the reader is referred to the appendix (3.7.2) for details.

[I] Clustered regularly interspaced short palindromic repeats (CRISPR) interference is a RNA-programmable gene expression control technique developed based on the prokaryotic immune system CRISPR-Cas9 [45]. Instead of nicking double-stranded DNA, the inactivated endonuclease dCas9 will bind to the target DNA site located by the single guide RNA (sgRNA) and interfere with the transcription process [155].

3.6 Discussion

The general growth law provides a quantitative framework for understanding cell size control. The classical nutrient law is but a special case of the general growth law, which encompasses a larger physiological space. The cell size can be summarized as the product of three physiologicl variables S_0, λ and τ_{cyc}. Among these threee variables, the unit cell S_0 remains invariant under a broad array of growth inhibition conditions targeting transcription, translation, ribosome content, fatty acid synthesis, cell wall synthesis, replication speed, and cell division.

This universal invariance suggests a deeper level of size control: There exists a critical control point for the replication initiation that ensures cell size homeostasis. Nevertheless, it is not immediately obvious why inhibition to biosynthesis like trancription and translation did not affect unit cell size. One may well point out that these perturbations would affect the synthesis of the initiators hence the initiation process. We think the answer lies in balanced growth, wherein all components are increasing as the same rate as the growth rate λ [31]. When total biosynthesis is inhibited, all cellular components are inhibited to the same degree as is the growth rate λ. Therefore, every component are still growing at the same proportion to each other, albeit at a slower rate. In this way, cells maintain that size threshold for initiation but it takes longer for the cell to reach the threshold (Figure 3.7). It should be noted that this may well be a simplifying picture as there are part of the proteome that changes in regard to growth rate [83]. However, it is reasonable to assume that the initiator belongs to the proteome sector that maintains a constant fraction of the proteome [65, 83].

The invariance of unit cell suggests an important role for replication initi-

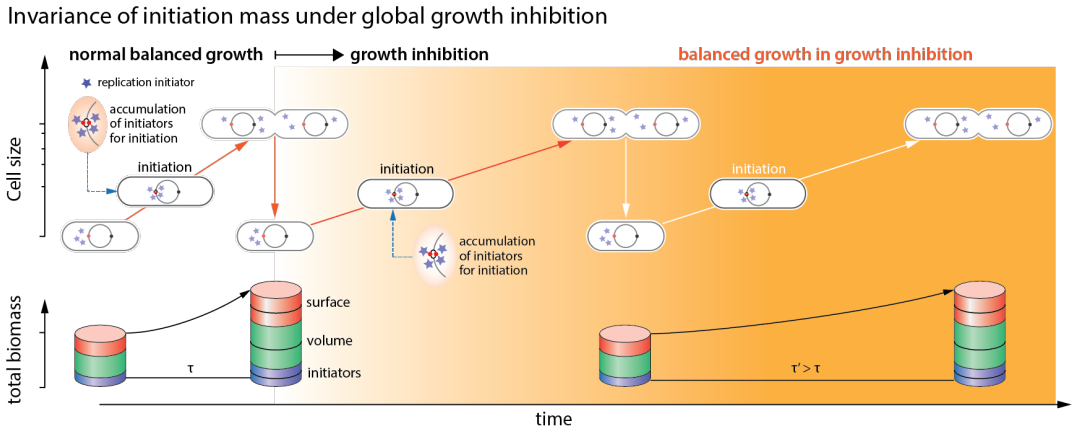

Figure 3.7: **Balanced growth and initiation threshold explains the invariant unit cell**. The initiator ocuipes a fixed fraction of the total proteome so that under limit

ation in maintaining cell size homeostasis. Additionally, single-cell data suggest that initiation mass has a narrow distribution with coefficient of variation (CV) of 0.1 - 0.15, much smaller than other physiological parameters like generation time, cell cycle period [195, 209]. This elicits the question of how this tight distribution is achieved in the cell and ultimatley why it is so, which is the focus of the next two chapters of this thesis.

3.7 Appendix

This appendix contains contents integral to the study described in this chapter, but was carried out by my colleagues.

3.7.1 Cell cycle progression perturbation: Changing D period

Given that delay in replication caused cell size increase, one immediate question that follows is if a delay in the division process would have a similar effect on cell size. An obvious candidate was *FtsZ*, an essential cell division protein that forms a ring structure in midcell to guide division machinery assembly. We constructed an FtsZ tCRISPRi strain to test this. However, the cells were very sensitive to the level of ftsZ knockdown and become filamentous at moderate level of induction. We could not observe intermediate state where cell size increase incrementally in a dose-dependent manner - the cell size distribution was bimodal where each cell is either filamentous or indistinct from normal wildtype cells at intermediate level of knockdown. Higher level of knockdown blocks cell division and cannot be used for steady-state measurement (Figure 3.8)

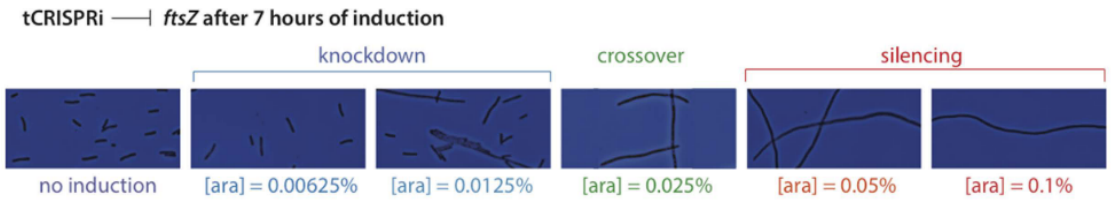

Figure 3.8: ***ftsZ* knockdown to block cell division** Phase contrast images of knocking down *ftsZ* after 7 hours of treatment. Figure adapted from [115].

Therefore we looked for other genes that regulate *FtsZ* activity and *SulA* was chosen as it inhibits *FtsZ* polymerization and Z-ring formation by competitively binding to free FtsZ subunit [35, 40]. Overexpression of *SulA* did delay division in that D period became longer but C period did not (Figure 3.8). The change in cell size follows the same exponential relationship with regard to cell

cycle period change (Figure 3.6).

As an alternative to the genetic method, cephalexin inhibits the synthesis of peptidoglycan in the cell wall and delay cell division. Sublethal dosage of cephalexin indeed extended cell cycle duration τ_{cyc} as well as population average size $\langle S \rangle$. More importantly, the size change can be explained by the general growth law formulation as the size changes expnentially as a function of τ_{cyc} (Figure 3.8).

3.7.2 *dnaA* knockdown changes unit cell size

Since both λ and τ_{cyc} caused expected change in cell size, we wondered if perturbation to the initiation process can lead to change in S_0 and hence cell size. Among all candidates, *dnaA* is the key regulator of replication initiation. Since *dnaA* is an essential gene, a knockout mutation was not possible without extensive modification to the genome to ensure its viability. There is the possibility to activate an alternative mode of replication, constitutive stable DNA replication (cSDR) by removing both RNase H (*rnhA*) and *dnaA*, but this will greatly change the way replication works and cell physiology is very different from normal conditions [33, 126]. As a result, we sought to construct a dnaA-tCRISPRi strain to downregulate the *dnaA* expression level incrementally. We reasoned that a lower level of *dnaA* would cause a delay in the process of initiator buildup hence an increase in cell size at initiation.

In a series of *dnaA* knockdown, unit cell size increased in a dose-dependent manner. The growth rate did slow down at higher knockdown level, which is expected given the essentiality of *dnaA*. Nevertheless, the general growth law can well predict the change in cell size (Figure 3.9).

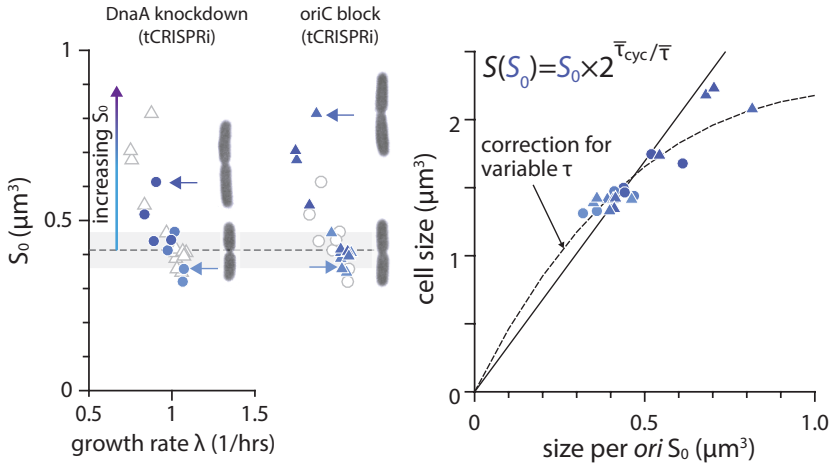

Figure 3.9: **Inhibition to initiation process led to increased** S_0. Left: unit cell size S_0 increases as a result of perturbation to the initiaiton process, either due to tCRISPRi knockdown of the initiator *dnaA* or blockage of the replication origin *oriC*. Dashed line and grey band represent typical range of S_0 values found in wildtype conditions (mean and standard deviation, respectivey). Right: The size changes of samples from th left as a function of unit cell size. Cell size changes almost linearly according to the general growth law (solid black line, assuming constant growth rate regardless of perturbation). The dashed line is the correction taken into account of slightly lower growth rate at higher inhibition level shown on the left.

3.7.3 Alternative methods to perturb the initiation process

To corroborate our observation on *dnaA*, we looked at other ways of changing unit cell size via perturbation to the initiation process. Inspired by work from Cees Dekker and his colleagues [212], we decdided to reinstate their strategy of blocking initiation by physically blocking *oriC* using tCRISPRi. It has been shown that targeting dCas9 to *oriC* will cause the cell to stop new rounds of replication initiation [212]. We applied a very low amount of arabinose to the cell culture so that an exponential culture at steady-state can be maintained. Within such range of possible arabinose, we found a similar negative correlation betweeen arabinose concentration and the unit cell size, which can be fit to the general growth law

(Figure 3.9).

Altogether, the evidence suggests that replicaiton initiation is the key process that underlies the unit cell size. The initiation process is well regulated such that the initiation mass is kept constant in different nutrient growth conditions or translational inhibition.

3.7.4 Wide range of perturbation confirms constancy of unit cell

The notion of constant initiation mass has been held as true without much experimental scrutiny in the context of growth inhibition. Donachie derived the constant initiation mass based on constant τ_{cyc} as measured by Helmstetter and Cooper and the further assumption of exponential cell growth [43, 72]. As shown in the previous section (Section 3.4), cell cycle duration can change at certain growth inhibition conditions. Therefore, it remains to be seen whether the initiation mass stays constant when challenged with different growth inhibition.

To systematically test whether initiation mass is constant, we included perturbations that inhibit major biological processes in the cell including transcription, translation, ribosome content, fatty acid synthesis, cell wall synthesis, replication speed, and cell division as listed in the table below.

We did TSTAT experiments on all the abovementioned conditions and measured their growth and cell cycle parameters accordingly. For many growth inhibition conditions, the $S - \lambda$ curve deviates from the nutrient growth law and do not present similar dependency on growth rate λ. However, the unit cell size (calculated by $S/2^{\tau_{cyc}/\tau}$) collapsed all points in the $S - \lambda$ space to a single line (Figure

Table 3.2: List of extensive peturbation to test constancy of unit cell size

Targeted process	**Treatment**
Transcription	Rifampicin
Translation	Chloramphenicol
Translation	Erythromycin
Ribosome content	Nutrient condition
Fatty acid synthesis	Triclosan
Cell wall synthesis	Fosfomycin
Replication progression	Rep tCRISPRi
Replication progression	Hydroxyurea
Cell Division Control	SulA overexpression
Cell Division Control	Cephalexin

3.10). This is a very stong evidence that the unit cell is invariant over all growth rates under a broad range of perturbation. As a result, we learned that the initiation mass is indeed invariant, not only under different nutrient limitations as it was proposed for, but also under extensive growth inhibition.

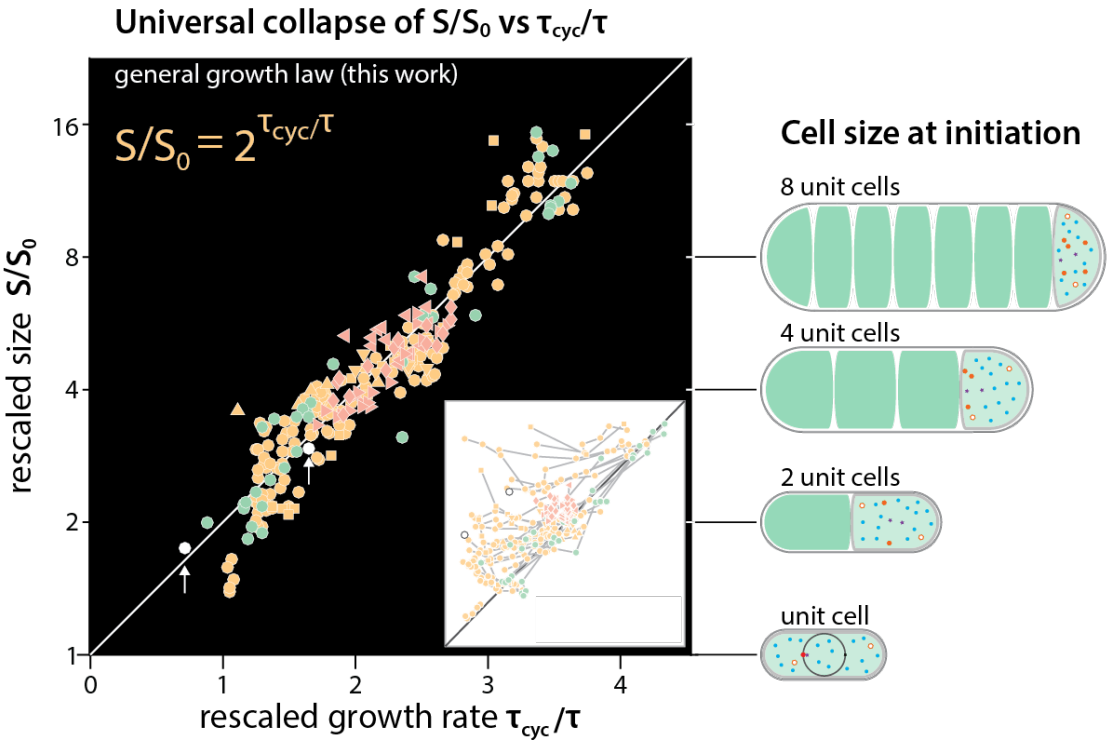

Figure 3.10: **The general growth law unveils the principle of size control** Left: All raw data from the size vs. growth rate plot (inset) collapse onto a single master curve after rescaling, demonstrating the predictive power of the general growth law. Cell size is normalized by unit cell size S_0 and τ_{cyc} normalized by τ. Empty circles (with arrows) represent the average of pooled single-cell data from previously reported data [209]. This confirms that our data are consistent with single-cell data, and the conclusion is not dependent on our experimental system. Right: Illustration of cells at different size can be seen as a multiple of the fundamental unit cells. The observed size change is achieved by varying the number of unit cells.

3.8 Acknowledgment

This chapter is entirely adpated from [182] and represents work from a joint effort among the coauthors. The dissertation/thesis author was the primary investigator and co-first author of this paper.

Chapter 3 is a subset of material as it appears in Current Biology 2017. Fangwei Si, Dongyang Li, Sarah E. Cox, John T. Sauls, Omid Azizi, Cindy Sou,

Amy. B. Schwartz, Michael J. Erickstad, Yonggun Jun, Xintian Li, Suckjoon Jun, Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. Curr. Biol. 27, 12781287 (2017). The dissertation author was the primary researcher an author of this material.

Chapter 4

Replication initiation control: Constancy of initiator

4.1 Introduction

4.1.1 Initiation control models

In the previous chapter, the discovery of constant initiation mass under a broad range of growth inhibition put replication initiation control in the center stage. Conceptually, this constancy connects DNA replication with cellular growth, so that the two processes could work together to reach optimal growth. An obvious question that follows is the origin of this constancy, which likely involves the control of replication initiation. In this chapter, a simple threshold model for initiation control is given to tie initiator constancy to the initiation mass constancy. The requirements for this simple model is examined in this chapter and the next.

There have been several models proposed for replication initiation control.

However, they are either not up-to-date (consistent) with all important observations or lack the quantitative predictive power for us to directly test (Table. 4.1). In 1963, Jacob, Brenner and Cuzin put forward the replicon model to describe the initiation control (1.7). The replicon is a unit of replication that has to be duplicated altogether. The initiation of replication is triggered by an initiator that diffuses and acts on a specific sequence the replicator. The replicator then opens up DNA to allow for one round of DNA replication [85]. One can see that the replicon model remains true to a great degree, even with our current kwnoledge of replication initiation control. The initiator is DnaA and the replicator *oriC*. Despite its remarkable insight, the replicon model does not explain quantitatively the constancy of the initiation mass we observed. In addition, the model does not explain how the cell prevents reinitiation, an important mechanism to balance DNA replication with cell growth.

The earliest quantitative model for initiation contorl came from Sompayrac and Maaløe (the autorepressor model) and Pritchard (the inhibitor titration model) (Table 4.1). Both models employed a critical threshold for initiation, so that initiation occurs when the threshold is reached for a specific regulator (Figure 4.1). The autorepressor model assumes a positive regulator that is produced by a negative feedback loop such that its concentration stays the same and its number increase as cell size. The initiator is immediately destroyed after initiation to prevent reinitiation [191]. Meanwhile the inhibitor titration model requires a inhibitor that prevents initiation when its concentration is higher than a critical threshold. As the cell increases in size, the concentration drops until the next round of replication is allowed to initiate. At initiation, there is a burst of new inhibitor synthesized

to block reinitiation until the next round of replicaiton [15, 152, 165]. Simulation study [123, 191] later showed that the inhibitor titration model to be unlikely as it cannot achieve the narrow distribution of initiation mass experimentally observed. In addition, there is no sound mechanistic explanation for a burst of gene expression at initiation. Hansen proposed the initiator titration model that is similar to the autorepressor model (Figure 4.1). The difference is that the latter does not assume immediate destruction of the initiator DnaA, instead they are titrated after the replication of *datA* locus which contains many *dnaA* boxes [65].

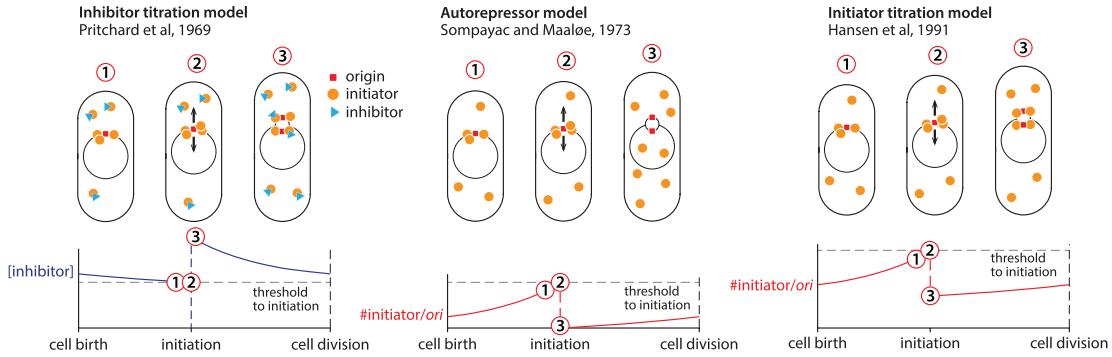

Figure 4.1: **Threshold-based initiation control models**. Schematics of three initiation control models: inhibitor titration, autorepressor and initiator titration model (in the order of left to right). Bottom graphs show expected dynamics of the initiation regulator (inhibitor/initiator) from birth to division. All three models have a threshold for initiation (inhibitor concentration for the first one, and number of initiators per *ori* for the latter two). Adapted from [88].

All the abovementioned models have not taken into account that only the ATP-bound DnaA is the active form capable of initiating replication [64, 179, 186]. Donachie proposed a ATP/ADP competition model that relies on the ratio of the two forms of DnaA as the trigger for initiation (Figure 4.2). While this model incorporates important molecular details, a mechanism that detects the ratio of two DnaA forms remain elusive and not likely. In fact, The author himself instated a

high-order function of the actual ratio as the actual initiation trigger [44], although it remains mostly speculative.

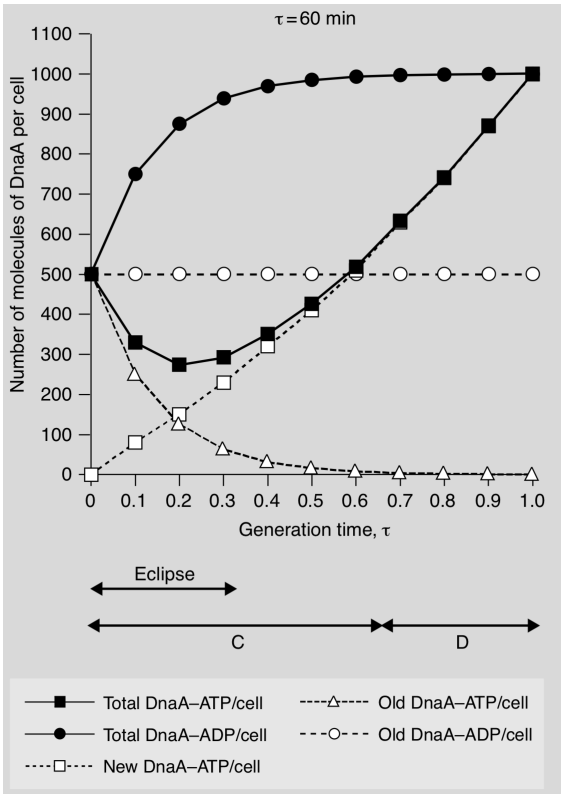

Figure 4.2: **Schematics of the ATP/ADP competition model**. Newly synthesized DnaA primarily becomes the ATP-bound form, thus the ADP-form is generated only through the hydrolysis pathway [44]. The illustrated example assumes a generation time of 60 minutes and (C+D) period of 60 minutes so that initiation happens at cell birth when the ATP/ADP ratio is the highest. The hydrolysis process brings down the ratio while over longer period, this is counteracted by the synthesis of new DnaA which raises the ATP/ADP ratio.

4.1.2 A simple threshold model for replication initiation

It can be seen that none of the models hitherto proposed is satisfactory to quantitatively explain the initiation control that suits ours. Therefore, we prpose a simple threshold mode to address this issue (Figure 4.3): 1) **invariance**: the

initiator constitutes a fixed fraction of the proteome in all growth conditions so that its concentration does not change over different steady-state growth conditions; 2) **constancy**: the initiator concentration remains nearly constant by its autoregulatory expression throughout the division cycle; 3) **accumulation to a critical number of initiators**: initiation occurs a fixed number of initiators per origin has accmulated. Given these three conditions are met, the initiation is tied to cell growth via the modus operandi of initiator accumulation and action.

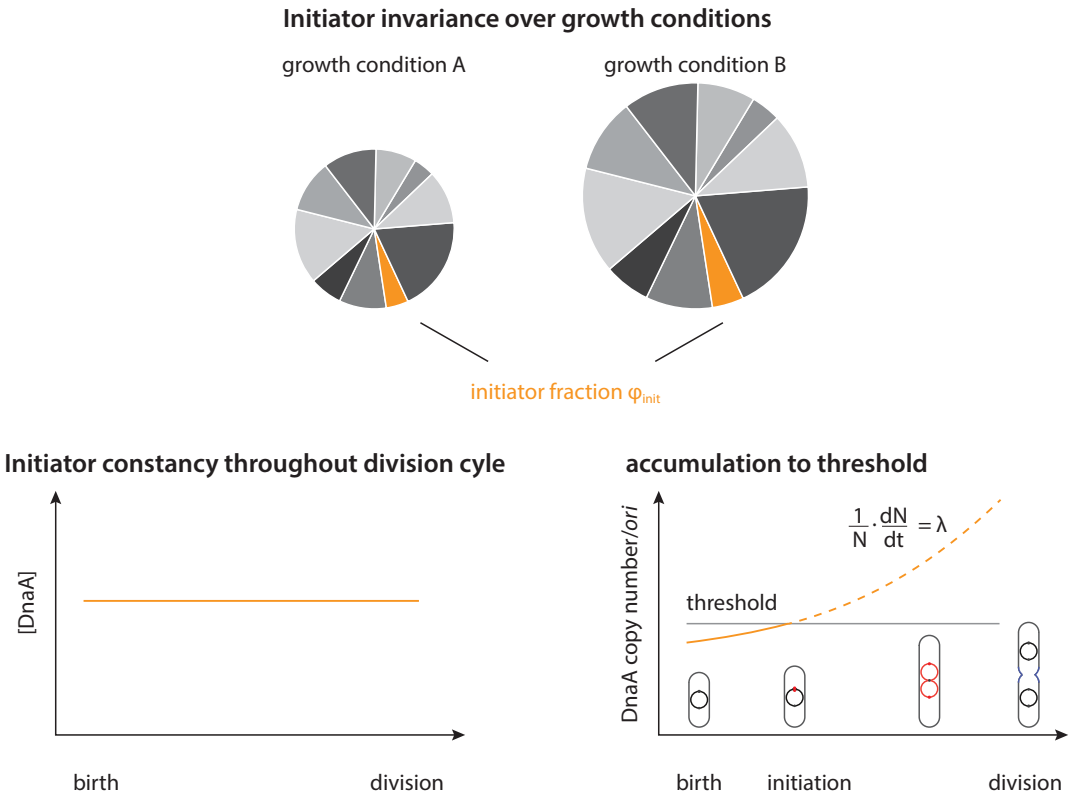

Figure 4.3: **Schematics for the simple initiation model**. The initiator occupies a fixed fraction of the proteome regardless of the growth condition (**invariance**). The initiator concentration remains at the same level throughout division cycle (**constancy**). Initiation is triggered when the number of iniaitors per origin has accumulated to a critical threshold (**accumulation to threshold**).

Before examining the three conditions experimentally, it is worthwhile to

point out some hints from the literature. In the last chapter (**??**), we showed that decreasing the concentration of DnaA causes an increase in the initiation mass [182]. This consonates with earlier work from Pritchard that showed an increase in cell size with the temperature seentive *dnaA46* mutant at increasingy higher temperature before cells became nonviable [153]. Similarly, overexpression of *dnaA* has been shown to cause overinitiation (decrease in initiation mass) [6, 7]. The invariance of initiators is logically consistent with our discovery of the invriant initiation mass in a wide range of steady-state growth conditions, with or without inhibition (Chapter 3). Therefore, it is reasonable to assume that DnaA concentration directs initiation. Here we assume simply an invserse relationship between DnaA concentration and initiation mass. This has been shown tentatively before by pooling together multiple experimental results together [76].

Besides, earlier work has shown independence and invariance of DnaA concentration by quantifying immunoblotting (Figure 4.4) over a large range of growth rate (0.6-2.3 doublings/hr) [6, 180]. Several different polyplonal antibodies have been used to reach the same conclusion [65, 76, 204]. More recently, several proteomic studies have measured DnaA copy number at different growth rates or with growth inhibition, but there is fluctuation in the measurement and no definitive conclusion can be drawn [83, 114, 178]. Both conventional biochemistry and proteomic works support the invariance of DnaA. There is a singular contradicting study, though, that indicated DnaA concentration per *oriC*is higher in slow growth condition [47].

The prediction of DnaA constancy comes from its gene expression regulation. The *dnaA* gene promoter region harbors several DnaA boxes, and bound

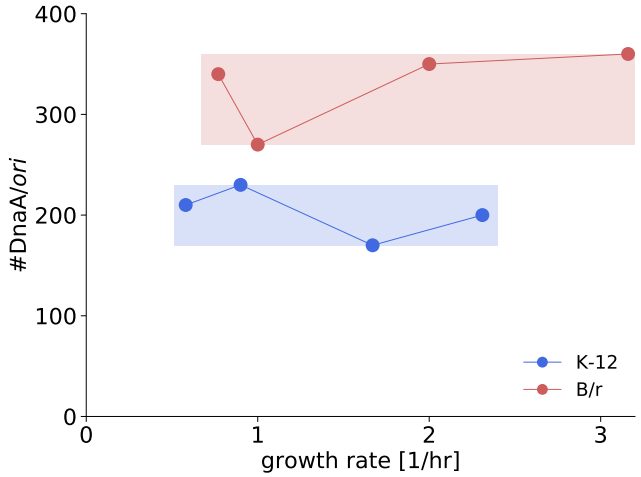

Figure 4.4: **Number of DnaA protein molecules per *oriC***. Two *E. coli* strains, K-12 and B/r were measured for DnaA concentraiton based on immunoblotting, and their protein copy number per origin calculated based on origin quantification (based on flow cytometry). Reproduced based on data from [65].

DnaA suppresses its own expression [4, 5, 21, 106]. This negative feedback of *dnaA* expression ensures that DnaA concentration is close to its equilibrium state [191].

The following section will focus on tests of the invariance and constancy of DnaA using imaging-based methods. Initiation threshold will be addressed in the next chapter.

Table 4.1: Models for replication initiation control

Model	Explanation	Pros	Cons	Key reference
Replicon model	Initiator structure is assembled at replicator to triggers initiation then gets destroyed	conceptually concise	no molecular details	Jacob et al., 1963
Inhibitor titration model	A fixed number of inhibitor molecules synthesized immediately post-initiation, then diluted out as cell grows till it reaches a critical level that derepresses initiation		no molecular mechanism could explain the fixed size of inhibitor expression burst	Pritchard et al., 1969
Autorepressor model	Autorepressor binds to its own promoter to maintain constant concentration of itself and initiator; under balanced growth, cells initiate when a fixed number of initiator/repressor is synthesized since last initiation event	agreed with experimental proof of constant concentration of DnaA; stochastics simulation suggested 250 initiators per origin matches well with biochemical measurement	The assumed protein:mRNA stoichiometry of 50:1 is significantly deviated from current understanding.	Sompayrac & Maaløe, 1973
Mahaffy & Zyskind		considered two nucleotide-bound forms of DnaA	the model was based on the now obsolete idea that DnaA concentration correlates with growth rate	Mahaffy & Zyskind, 1989
Initiator titration model	DnaA has high affinity to DnaA boxes on the chromosome, and lower affinity to initiation site; replication starts when other DnaA boxes are saturated and DnaA are able to bind to origin; replication pushed off bound DnaA from its site and origin cannot be reinitiated right after last round of initiation	took into account of more recent molecular understandings of DnaA and replication initiation	overlooked the two nucleotide-bound forms of DnaA	Hansen et al., 1991
ATP/ADP ratio	Cellular ATP-DnaA/ADP-DnaA ratio triggers initiation. Two species compete antagonistically for binding sites in the replication origin.	took into account of experimental evidence that ATP-DnaA form is required for binding to low-affinity DnaA boxes in replication origin and forms polymer	the model cannot explain well DnaA titration experiments when only the absolute concentration of DnaA but not the chemical equilibrium between two species are altered. Difficult to find a mechanism that sense the ratio of two forms of DnaA and how the ratio triggers initiation instead of a molecular species.	Donachie & Blakely, 2003
Simple threshold model	A critical number of DnaA per oriC is required to trigger initiation so that DnaA concentration and the initiation mass is inversely proportional. Increase in ATP-DnaA/ADP-DnaA ratio will further decrease initiation mass as ATP-DnaA is the limiting factor in opening of DUE in oriC. Additionally, we suggest that there exists a threshold for ATP-DnaA in the cell since ATP-DnaA is the only form that cooperatively binds to oriC to open up the oriC.	Consistent with experimental evidence of dnaA over-/under-expression reduces/increases initiation mass. Compared to the autorepressor model, this model takes into account that ATP-DnaA is the predominant form of DnaA that drives initiation.	The functional form of how ATP/ADP ratio affects initiation mass remains unclear.	our working hypothesis.

4.2 Gene expression reporter

Reliable quantifications are needed to establish the quantitative relationship between DnaA and the initiaton mass. The initiation mass can be measured on population level by combining measurement of cell cycle and cell size (calculating s_i based on Equation 3.3). On the single cell level, initiation mass can be measured by tracking fluorescently labeled replisome marker DnaN (detail in Appendix B.2). As for the measurement of the initiator, DnaA was traditionally measured using quantiative immunoblotting, where the proteins of interests are detected by antibodies and detected via chemical luminescence or fluorescence coupled to the antibody. The quality of blotting is subject to the quality of antibody that recognizes the epitope on the DnaA protein. Reliable quantitation from immunoblotting relies on careful calibration of antibody reactivity as the variability between different batches of antibodies would confound the result [18]. However, protein quantification based on blotting requires extensive and meticulous calibration, which involves expression and purification of DnaA protein and establishing a standard curve based on the purified DnaA (a dilution series of protein blotted to map out the detection dynamic range) [65, 125]. Experimental uncertainties include the variation in efficiency of protein transfer between the polyacrylamide gel and the blotting membrane [54]. This warrants careful interpretation of detection signal and actual protein amount. Additionally, this kind of biochemical methods takes samples from a population of cells, thus limiting any conclusion drawn to population-level correlation.

Instead we chose to measure DnaA concentration by quantitative fluorescence imaging. This can be easily incorporated into our mother machine exper-

imental procedures. Imaging-based measurement can probe protein level of individual cells, thus allowing us to look at not just the mean values of DnaA concentration and initiation mass but also their distribution and correlation [195]. The quantification can be achieved via genetically labeling *dnaA* with a fluorescent protein (FP) tag co-transcriptionally or co-translationally. Each tagging strategy has its own caveat and strength as summarized in Table 4.2. Notably, co-translation reporter is superior in ensuring the stoichiometry of FP to DnaA protein since they are translated as one polypeptide chain and two protein molecules physically linked together. However, fused FP may affect the folding of the tagged protein and render the protein dysfunctional. As a result, successful protein fusion requires careful design of the posititioning of FP relative to the gene of interest and multiple rounds of iteration yet still not possible on all genes [189, 190]. There is however a third different tagging method using cleavable FP tags. It is similar to translaional reporter in that the FP is fused to the protein of interest when translated. However, the protein will be cleaved in the linker region by induction of a protease or inherent autocatalytic activity [74, 220]. In principle, this method combines the strength of both methods as it retains 1:1 stoichiometry while not affecting the protein functionality. However, this method is relatively recent and has not seen many applications in bacteria and requires systematic optimization. In light of this, I opted to use a transcriptional msfGFP reporter, which is a bright FP suitable for quantification. Additionally, msfGFP has a short maturation time, which helps minimize the time delay between dnaA expression and fluorescence readout.

However, it remained unknown whether the transcriptional reporter acts as an accurate proxy of the DnaA due to potential discrepancy between transcrip-

Table 4.2: Comparison of fluorescent protein reporter labeling schemes

Methods	**Advantage**	**Disadvantage**
Transcriptional	minimal side effect on protein function	protein level may differ from transcript level
Translational	exact stoichiometry	protein function may be compromised
Autocleavage	exact stoichiometry; protein function intact after clevage	Incomplete cleavage or delay in cleavage may affect protein function

tion and translation. Fortunately, there has been reports of successful chimeric DnaA-FP fusion protein implemented. The FP is inserted in the flexible linker region of domain II of *dnaA*, and the initiator function remains intact [16, 140, 158]. Based on this translational fusion reporter, I constructed a strain that co-expresses mCherry-*dnaA* and msfGFP udner the inducible promoter P_{BAD}. I measured both GFP and mCherry fluorescence level for each cell at a range of arabinose induction levels. It turned out that DnaA concentration (as measured by mCherry fluorescence level) can be well represented by the transcriptional reporter msfGFP: The population average GFP vs. mCherry fluorescence intensity can be fit with a simple linear relationship with zero intercept (r^2>0.99, Figure 4.5). This suggests DnaA concentration can be directly measured by fluorescence protein, and transcriptional reporters can be used as a good proxy for DnaA concentration.

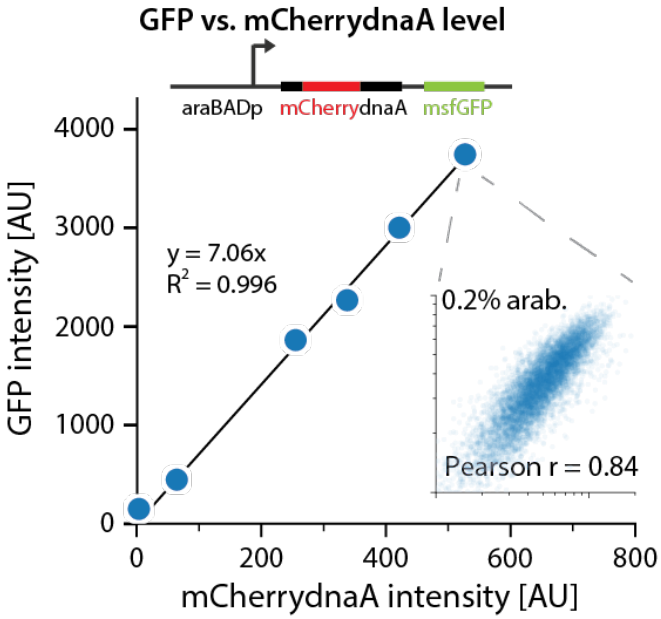

Figure 4.5: ***dnaA* gene expression reporter**. Top: an msfGFP Transriptional reporter is inserted after the chimeric mCherry-*dnaA* gene. The entire cassette is driven by a tritratable P^*_{BAD} promoter. Bottom: this strain is cultured at multiple arabinose induction level and fluorescence levels for both reporters are measured and compared. The average GFP and mCherry fluorescence level can be described with a simple linear relationship and at each condition (one condition shown in inset), the single-cell level GFP and mCherry level is well correlated.

4.3 DnaA invariance over different growth conditions

To test the invariance of DnaA, I first collected batch culture samples of *dnaA*-msfGFP cells growing with different nutrients from TSTAT. The fluorescence level of GFP is quantified for each condition. The average fluorescence level of all 8 conditions (with growth rate spanning from 0.5 to 2 doublings/hr) agreed with the invariance hypothesis as the inter-sample variation is much smaller in comparison to the variation within each sample (Figure 4.6).

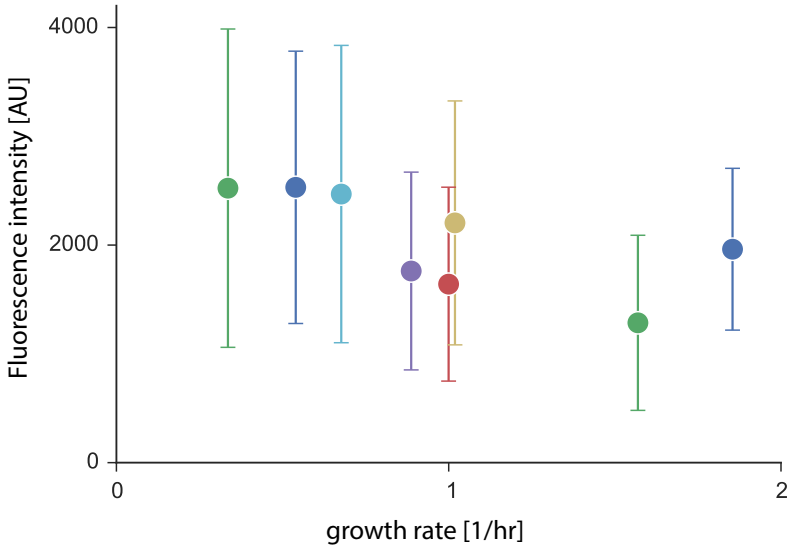

Figure 4.6: **Population level measurement of DnaA.** DnaA concentration is measured by the transcriptional reporter GFP fluorescence level. The samples were taken from 8 different nutrient conditions in a TSTAT experiment that spans a wide range of growth rates.

4.3.1 Protein quantification by binomial partitioning

Even though direct fluorescence imaging of the GFP reporter suggested the initiator invariance, the results are limited as they are in arbitrary fluorescence unit, and cannot be diretly compared with other experiments. It is also impossible to directly compare with classical quantitative blotting experiments, as they were reported in molecular copy numbers or molar concentration [36, 64, 65, 180]. Therefore it is beneficial to use actual numbers rather than fluorescence level to facilitate cross-checking with literature-reported values.

This calls for a calibratin method to convert arbitrary fluorescence units into real numbers. Assuming the fluorescence level increases linearly with fluorophore numbers, the key is to determine the conversion factors. This can be obtained either direclty or indrirectly. It is possible to directly measure the fluorescence

emission from single molecules given purified fluorescence molecules and special optical setup [199]. This kind of experiment typically requires diluting the protein onto a glass slide to separate single molecules, which could differ from *in vivo* situations drastically. On the other hand, a series of indirect measurement is based on the scaling between fluorescence variation and number of fluorescent molecules involved. Binomial partitioning is one of such methods.

Given that the excitation power and detection are not limiting, the fluorescence I is linearly scaled with the number of fluorophores N:

$$I = \eta N \tag{4.1}$$

The key is then to find the value of the conversion factor η. Upon septum formation in a dividing cell, proteins are partitioned in one of the two daughter cells. Binomial partitioning models such outcome with a binomial distribution: Freely diffusing protein molecules are randomly assigned to one of the two daughter cells. The assignemnt of each molecule is a Bernoulli trial (coin flip test), with equal probability of being in either cell. The result is represented by repeated Bernoulli trials.

The binomial distribution $\mathbf{B}(n, p)$ models the number of success in n individual trials with each having probability of p. The mean of the distribution (expected times of success after large number of trials) is given by np, and the variation $np(1-p)$. In the case of random assignment, the probability of one particular protein ended up in one cell is 0.5. Denote the number of protein molecules in two daughter cells as n_1 and n_2 and plug the number in gives the expression

of variaion as and the partitioning error is tied to the average copy number of proteins

$$\begin{aligned}\sigma_N &= np(1-p) = n/4 \\ &= \langle (n_1 - \langle n_1 \rangle)^2 \rangle \\ &= \langle (n_1 - \frac{n_1+n_2}{2})^2 \rangle \\ &= \langle (n_1 - n_2)^2 \rangle / 4\end{aligned} \tag{4.2}$$

Using the linearity between fluorescence intensity and fluorescent protein copy number 4.1, we have $\eta(I_1 + I_2) = \langle (I_1 - I_2)^2 \rangle$ where I_1 and I_2 are measured total fluorescence intensity from each daughter cell. This binomial partitioning method was developed by Elowitz group [166, 167] and also adopted for similar applications in quantifying transcription factors [25, 201] from fluorescence of fusion protein.

To verify the binomial paritioning method, I took a strain that constitutively expresses YFP and performed mother machine experiment. Dividing cells are identified through the Python anlaysis package after cell segmentation. Each division gives rise to two sibling daughter cells, and their total fluorescence intensity were quantified. Based on Equation 4.3.1, the partition error $(I_1 - I_2)^2$ and the sum of fluorescence $(I_1 + I_2)$ are calculated. The partition error is indeed well correlated with the sum of fluorescence (Figure 4.7). Points with similar sum of fluorescence are binned together, as if they represent different outcomes of a repeated cell division event. This way, the data from different cells resembles a repeated Bernoulli trials of an imaginary cell with certain number of proteins. Thus, one can apply the equation derived above to determine η. Afte binning, the average partition error within each bin showed good linear relationship with the mean sum of fluorescence for each bin (Figure 4.7).

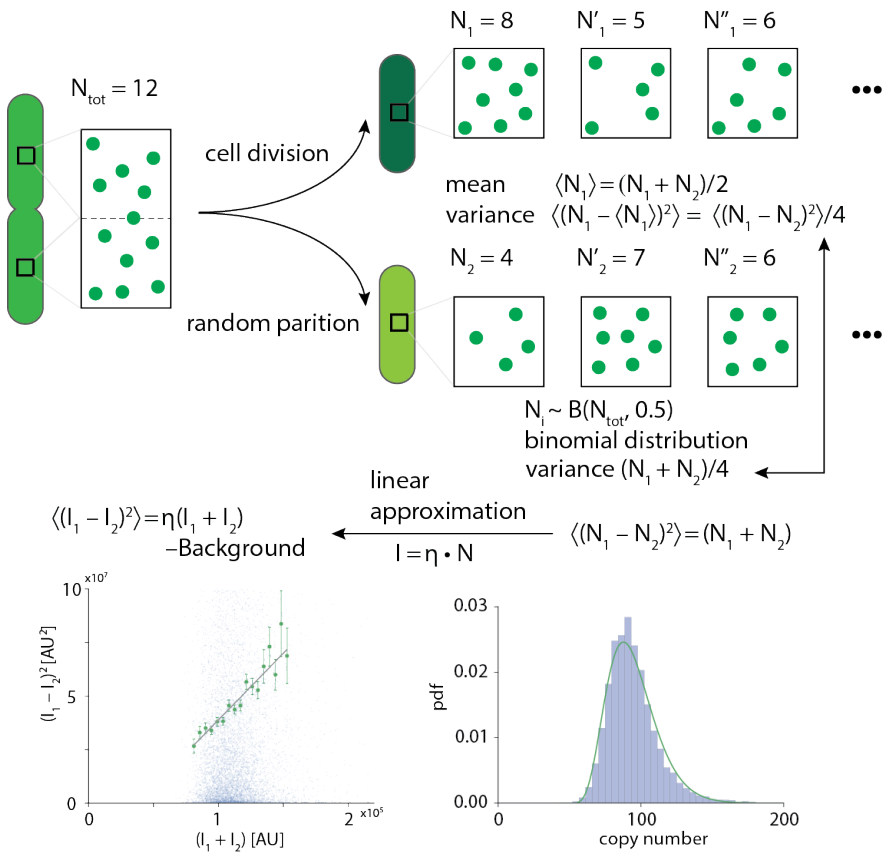

Figure 4.7: **Binomial partitioning applied to a strain expressing YFP constitutively** Top: schematics of binomial partitioning and modeling the partitioning as a bernoulli trial. Bottom: estimation of conversion factor η based on the partition error $(I_1 - I_2)^2$ vs. binned $(I_1 + I_2)$. The distribution of YFP protein numbers is shown to the right. The green line represents a fit to a gamma distribution, commonly used to model protein copy number distribution [181, 199]. The strain used is SJ77 with the genotype *intC*::[P(λR):yfp].

4.3.2 DnaA quantification

To quantify the DnaA protein copy number, I applied the binomial partitioning method to *dnaA*-GFP strain (SJ1171) grown in different nutrient conditions. The DnaA protein copy number estimates agree with previously reported value of 250/*oriC* [6, 64] (Figure 4.3).

Our prediction for DnaA invariance also applies to growth inhibition. To

see if this is the case, I chose translation inhibition since that was widely tested in our study on invariance of initiation mass [182]. However, because the *dnaA*-msfGFP reporter strain has the selection marker gene *cat*, I used erythromycin to block translation and achieve similar effect as chloramphenicol treatment. Despite changes in cell size, DnaA concentration is not changed (Table 4.3). To show the generality of this invariance, a different type of antibiotics, triclosan was used. Triclosan inhibits cell wall synthesis but preserves the initiation mass [182]. Not surprsingly, the DnaA concentration remains unchanged (Figure 4.8, Table 4.3).

In summary, the DnaA concentration is invariant under nutrient conditions and growth inhibitions, similar to the initiation mass. When DnaA concentration is specifically tuned down by tCRISPRi, the initiation mass increases accordingly [182]. These suggest that an initiation control mechanism based on the copy number of DnaA.

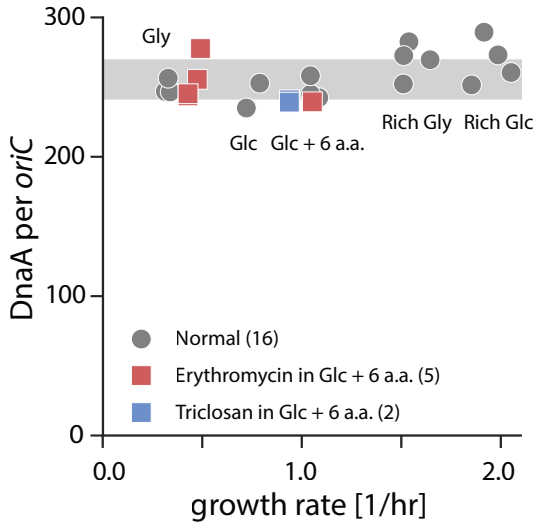

Figure 4.8: **Invariance of DnaA** DnaA concentration is measured from mother machine growth experiment by employing binomial partitioning methods. Grey sybmols are experiments with different media but no growth inhibition. Colored symbols represent experiment of growth inhibition. DnaA copy numbers taken from Table 4.3, and number of origins is interpolated from [182].

4.4 DnaA constancy over cell generation

The second prediction of our initiation model is that the concentration of initiator protein is effectively constant within the cell generation. To test this, I reanalyzed the DnaA quantitation experiment. Each individual cell generation trace is aligned to their birth time and the fluorescence level of the GFP reporter is plotted against cell age. One confounding factor is the delay between gene expression and corresponding change in fluorescence due to fluorophore maturation [9]. To bypass this problem, I chose slow growing conditions (MOPS glycerol) where the generation time is longer than the fluorophore maturation time so that there is a separation of two time scales (fluorophore maturation and cell generation).

The fluorescence level does not change significantly over the entire generation. In sharp contrast, a constitutively expressed YFP showed systematic fluctuation in fluorescence intensity (Figure 4.9). This confirms that DnaA concentration is constant over the cell generation.

However, our current model does not take into account of the effect of eclipse period on *dnaA* trascription. It has been proposed that the eclipse period may negatively impact transcription due to seclusion of SeqA [64]. More experiments are needed to test whether the eclipse period does signficantly alter *dnaA* transcription acitivity.

4.5 Summary

The discovery of invariant initiation mass over a broad range of growth conditions led us to believe that initiation control is important in coupling DNA

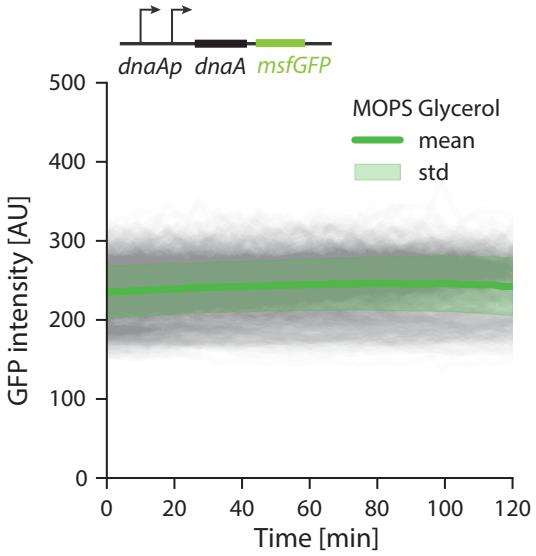

Figure 4.9: **Constancy of DnaA concentration within cell generation** GFP Fluorescence time trace of individual cells are plotted as gray traces in the background. The average at each time point is shown as solid green line with the light green band as the standard deviation. The DnaA concentration as quantified by GFP fluorescence intensity does not fluctuate systematically over the entire generation.

replication and cell growth. Since the speed of replicaiton fork does not change in fast-growing conditions, the rate of initiaiton has to be modulated to match changes in growth rate [85]. The initiation mechanism is central to this coupling between the two proceses.

We proposed a simple threshold model for initiation control with three components: the fraction of proteome occupied by the initiator is *invariant* different growth conditions; the initiator concentration remains *constant* over the cell generation; the initiation is triggered when a *critical threshold* number of initiator per origin has accumulated. I used an msfGFP transcriptional reporter of *dnaA* was used to quantify initiator concentration, and showed that the initiator concentration indeed does not change in differnt growth conditions or fluctuate within cell

generation.

It should be noted that the reasoning for the model assumptions of initiator constancy and invariance could well be a simplification of actual cellular processes. We proposed that the the invariance of initiator concentration as a natural result of DnaA occupying a fixed fraction of the proteome. However, there have been studies that suggest *dnaA* transcription is modulated by factors in a growth-rate dependent fashion: *Fis*, a nucleotide associated protein (NAP) has been shown to bind to the *dnaA* promoter and downregulate its expression [49]. The expression level of *fis* is closely tied to the growth condition[I] [8, 122]. Another exmaple is the stringent response regulator ppGpp, which also negatively regulates *dnaA* expression [64]. Despite this complexity, one could benefit from adopting the Occam's razor here to acknowledge that the initiator constancy and invariance is not due to a lack of regulation but an emergent property of the system.

[I] *fis* is upregulated transiently at nutrition upshift, and it remains to be seen if this is relevant in steady-state growth condition. However, proteomic data seem to suggest that *Fis* expression to be higher in nutrient rich conditions [83, 178].

Table 4.3: Summary of DnaA quantitation results by binomial partitioning

Date	**Condition**	τ	**DnaA copy number**		
		[min]	μ	σ	N*
20170523	Rich Glc	20.3	448.5	62.3	1,428
20170528	Rich Glc	21.7	498.2	161.6	6,523
20170607	Rich Glc	21.0	470.2	94.1	2,327
20170628	Rich Gly	27.0	486.4	97.8	6,520
20170703	Rich Gly	25.3	464.2	95.2	7,826
20170728	Glc 6aa	38.3	417.4	112.0	637
20170729	Glc 6aa	39.8	421.6	200.6	1,913
20170801	Glc	57.6	404.6	97.6	138
20170811	Glc	52.7	435.3	127.9	3,535
20170814	Rich Glc	22.4	433.1	101.7	27,752
20170830	Gly	132.3	425.1	102.1	934
20170907	Gly	122.6	424.5	122.1	1,014
20170928	Gly	126.4	441.4	109.3	212
20170818	Rich Gly, 0.004% arab	22.6	247.4	50.9	2,251
20171202	Rich Gly, 0.005% arab	27.5	434.3	75.3	5,771
20171204	Rich Gly, 0.0075% arab	39.9	444.4	89.5	1,007
20171215	Rich Gly, 0.006% arab	27.5	469.4	80.0	8,209
20170922	Glc 6aa, Erythromycin 5uM	39.5	412.2	98.6	1,098
20171027	Glc 6aa, Erythromycin 80uM	97.1	419.6	105.0	1,134
20171030	Glc 6aa, Erythromycin 60uM	87.5	440.0	150.4	1,714
20171031	Glc 6aa, Erythromycin 80uM	97.0	422.2	55.8	1,151
20171129	Glc 6aa, Erythromycin 40uM	84.7	478.1	77.9	1,604
20171109	Glc 6aa, Triclosan 0.15 μg/ml	44.3	414.7	82.0	5,775
20171209	Glc 6aa, Triclosan 0.05 μg/ml	44.4	411.8	109.8	1,352

* sample size for each experiment.

Chapter 5

Replication initiation control: Initiation threshold

5.1 Introduction

Replication initiaiton is a multi-step process that requires coordination among all molecular players involved (Figure 5.1). While it can be argued that all protein and DNA elements involved in this process are responsible for setting the initiation mass, DnaA remained the focus of initiation study as it is the master regulator. The prereplication complex (pre-RC) is formed after occupation of DnaA boxes in *oriC* and unwinding of DUE region within. A series of events follows till the assembly of replisome: The formation ensures landing of helicase loading protein *DnaC* and replicative helicase *DnaB* to be poised for DNA replication. Loading of the primase *DnaG*, then β-clamp and DNA polymerase follows as bidirectional replicaiton starts [111, 113]. Nevertheless, DnaA loading onto *oriC* is the rate-limiting step [89, 111, 157, 186], and most of the evidence suggest that chang-

ing DnaA level or DnaA activity (point mutation) causes corresponding changes in replication initiation (initiation mass and initaition synchrony) [6, 7, 64, 67, 118].

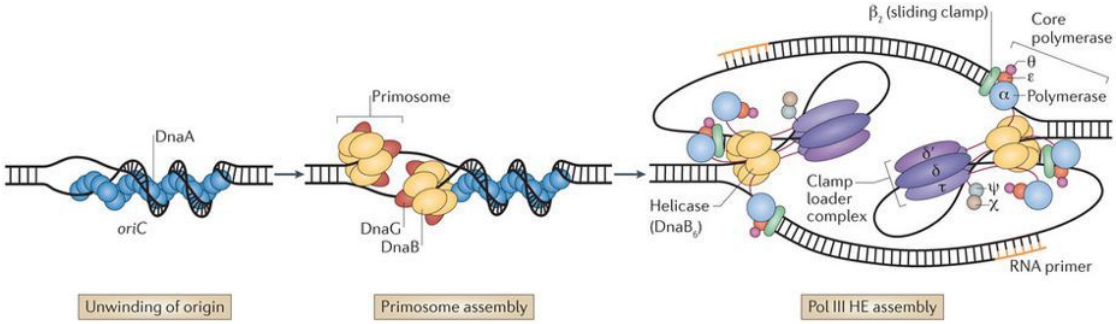

Figure 5.1: **Schematics of three stages of replcation initiation**. DnaA binding and unwinding of *oriC* is the first step and key to replication initiation. Helicae *DnaB* and primase *DnaG* fomrs the primosome complex, joined by the polymerase to assemble the replisome complex. Figure taken from [163].

Despite efforts to measure DnaA level (current work and [65]), it is not clear the "decision-making" mechanism of initiation in regards to DnaA level. In other words, how does a cell process the information of intracellular DnaA level and what is the functional relationship between initiation mass and DnaA level. In this chapter, we presented a simple threshold model that proposes initiation is triggered once a critical level of initiators have been accumulated per *oriC*. This model predicts a simple inverse relationship between DnaA level and the initiation mass. Such a relationsihp is tested this chapter to verify the model. Since endogenous *dnaA* gene is autoregulated and does not show much variation in expression level (as tested in the last chapter), I described the construction of *dnaA*-titratable strain. Using this titratable strain, *dnaA* expression level was varied and its relationship with initiation mass was tested both on population level and on the single-cell level.

However, the simple threshold model did not take into account the fact

that DnaA-ATP is the active form of initiator that drives initiaiton. To this end, the dependency of initiation on DnaA-ATP was examined in the framework of a initiation phase diagram consisting both total DnaA level and ATP-to-ADP ratio. At the end of the chapter, directions for further work to deepen our understanding of replication initiation control is given.

5.2 Construction of *dnaA*-titratable strain

Since the initiator level does not vary in wildtype strains (Figure 4.8), an extra *dnaA* titration system is needed to systematically vary DnaA level. Overexpression of *dnaA* has been done with plasmids bearing *dnaA* gene under inducible promoters [6, 64, 139, 161, 184]. Underexpression of *dnaA* has not been achieved prior to our work using tCRISPRi knockdown [182]. To achieve a wide range of possible DnaA level below and above that of the wildtype, I combined two systems in one strain.

To make such a strain, I started with a plasmid that encodes *dnaA* under a wild-type P_{lac} promoter (pSN306 from Katayama, Kyushu Univ., Japan, [139]). This plasmid was introduced into a tCRISPRi strain with sgRNA targeting *dnaA* [182] so that in principle *dnaA* level can be both upregulated and downregulated depending on the relative strength of overexpression and knockdown. Since the two systems use different inducer (Isopropyl β-D-1-thiogalactopyranoside, IPTG for *dnaA* overexpression and arabinose for *dnaA* underexpression, respectively), DnaA level can be directly modulated by using a combination of two inducers. This strain was used in population measurement on the relationship between DnaA

level and initiation mass.

Additionally, a similar *dnaA* overexpression plasmid (pLR40, Løbner-Olesen, Univ. of Copenhagen, [161]) was used to verify the results. A YPet-*dnaN* replisome marker was introduced to enable single-cell measurement of intiation mass in the mother machine.

The shortcoming of the above two strains is that there is no direct measure of *dnaA* expression level. DnaA level was estimated with separate calibration based on literature reported values of the expression system [6, 139], and any expression variability at the same induction level was ignored. A *dnaA* expression reporter can directly address this issue. In principle this can be done by adding a transcriptional reporter for both the endogenous *dnaA* and the plasmid-borne one. However, having only one source of *dnaA* and transcriptional reporter is preferrable as two sources might need separate calibration. To this end, the endogenous *dnaA* can be knocked out, leaving only the titratable *dnaA*. The following sections will discuss in details the process of constructing such a strain for tunable expression of initiator capable of simultaneously quantifying DnaA level and initiation events.

5.2.1 Deletion of endogenous *dnaA*

The *dnaA* is the first gene of an operon that contains *dnaA*, *dnaN* and *recF* (Figure 5.2). As mentioned in prvious chapters, *dnaN* encodes the β sliding clamp of the DNA replisome III holoenzyme replisome, while *recF* encodes a DNA repair protein involved in *RecA*-mediated recombination. The operon is transcribed from two *dnaA* promoters *dnaAp1* and *dnaAp2*, both of which negatively regulated by binding of DnaA-ATP [64, 67]. There are, however, four additional promoters

dnaNp1-4 residing in the C terminus of *dnaA* gene and two more inside *dnaN* *recFp1* and *recFp2* [3] (Figure 5.2).

To minimize the effect on *dnaN* expression, the C terminus (starting from 313rd amino acid residue) of *dnaA* needs to be preserved (Figure 5.2). Therefore, the N terminus of *dnaA* was replaced by a chloramphenicol resistance gene *cat*, which was fused to *dnaA* to minimize negative effects on *dnaN* expression due to Rho dependent polarity[I] [116, 117, 211]. In addition, since overexpression of the *dnaA* will negatively affect the transcription of the operon, the promoter region of *dnaA* should be replaced to bypass this regulation and decouple DnaA level from the rest of the operon. Therefore, I opted to use a constitutive promoter P_{cat} to drive the expression of downstream genes.

This scheme was carried out using *in vivo* homologous recombination-based engineering (recombineering) with λ Red system [107, 173] (see Appendix C). The strain was first supplied with plasmid expressing *dnaA* under a *lac* promoter (pLR40), then removed of the *dnaA* gene to generate the knockout. The resulting strain relies on IPTG for survival, as *dnaA* expression requires induction.

5.2.2 Construction of titratable *dnaA* expression cassette

dnaA overexpression can be done with aforementioned *dnaA*-bearing plasmids [6, 149, 184]. However, these systems are susceptible to variable expression level due to plasmid copy number variation and the promoters used (P_{lac},

[I] In prokaryotes, transcription and translation are coupled such that ribosome will bind to nascent mRNA transcripts that are still being transcribed. As a result, RNA polymerase(RNAP) and the translational machinery can physically interact [46]. When there is a premature stop codon, the ribosome falls off the transcript, leaving the nascent transcript open to Rho factor binding. Rho factor then interacts with the RNAP and causes it to fall off, resulting in a premature transcription termination.

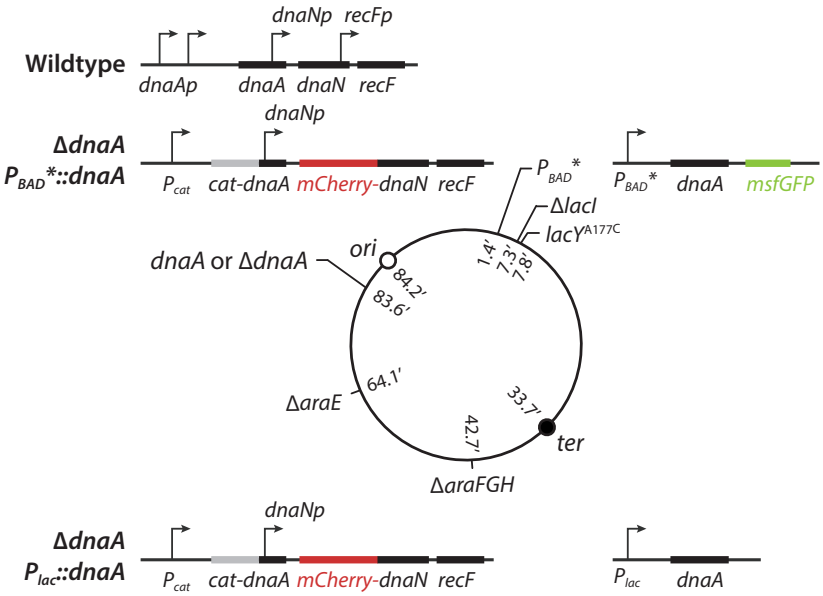

Figure 5.2: **Schematics for the $\mathbf{P^*_{BAD}}$::*dnaA* titration strain**. The *dnaA* titration strain is based on the P^*_{BAD} system. The arabinose transporter system was eliminated by deletion of *araE* and *araFGH*. The $lacY^{A177C}$ mutation allows arabinose to freely diffuse into the cell through the mutated transporter, and $\Delta lacI$ enables constitutive expression of this transpoter. The *araBAD* operon was replaced by *dnaA*-GFP so that *dnaA* expression level can be titrated by controlling the amount of arabinose added to the growth media. The endogenous *dnaA* operon contains three genes: *dnaA*, *dnaN* and *recF*, which share the two *dnaA* promoters *dnaAp1* and *dnaAp2*. Meanwhile, *dnaN* also has its own promoter inside the C-terminus of the *dnaA* sequence. The *dnaA* operon was engineered to remove the N-terminus of *dnaA* and fuse it with *cat* selection marker and the endogenous promoter was swapped for P_{cat}. The *dnaN* was fused to mCherry to enable fluorescence imaging of the replisome. Figure partially adpated from [115].

P_{araBAD}) are known to give switch-like expression pattern (bistable expression) [129, 146, 183]. This would amplify the cell-to-cell variability in response to a given level of initiator. To address this issue, I opted to use the P^*_{BAD} system developed in the lab [115], which allows fine-tuning of *dnaA* expression with minimal leaky expression compared to other available expression system [115].

This P^*_{BAD}::*dnaA* along with the endogenous *dnaA* knockout ($\Delta dnaA$) is capable of achieving a wide range of expression level from near zero to more than

10-fold of the wildtype initiator level [115] (Figure 5.3). To test whether the increase in *dnaA* expression level does lead to change in initiation mass, I tested with a strain that contains P^*_{BAD}::*dnaA* and the replisome marker. Three mother machine experiments were recorded with increasing amount of arabinose induction, and the initiation mass indeed decreased with increasing level of *dnaA* induction (Figure 5.3).

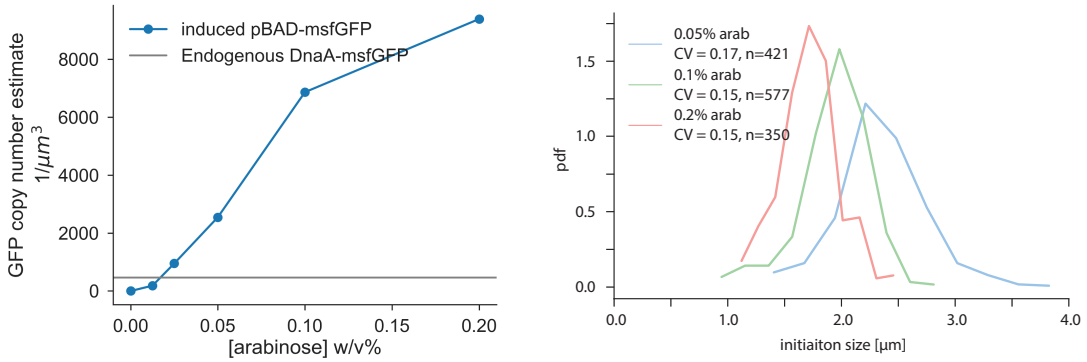

Figure 5.3: **Expression capacity of P^*_{BAD} system**. Left: A P^*_{BAD}::GFP strain (SJ_XTL177) was cultured at different arabinose induction level. The copy number of GFP was estiamted by using the binomial partitioning method (see 4.3.1). The horizontal line rerpresents wildtype DnaA level by imaging GFP transcriptional reporter of a wildtype strain (SJ_XTL571). The P^*_{BAD} system is capable of producing protein at a level below or above the wildtype DnaAlevel depending on the strengt of induction. Right: Probability distribution of initiation size (measured in length) from three different level of *dnaA* expression. Data were taken from mother machine experiment of a P^*_{BAD}::*dnaA* strain.

Since P^*_{BAD}::*dnaA* works as expected, I then introduced a fluorescently tagged *dnaN* into the strain so that one can simultaneously modulate and measure DnaA level as well as measure intiation mass on single-cell level (Figure 5.2). After confirming the functionality of this *dnaA* titration strain, I continued to characterize the relationship between initiation mass and the DnaA level.

5.3 DnaA level sets the intiation mass

If there does exist a critical threshold of DnaA for initiation, i.e. certain number of initiators need to be synthesized per *oriC*, one can simply express the number of DnaA as the product of its concentration and the cell volume. Therefore, rewriting this gives

$$[DnaA] = \frac{\text{const.}}{S_i} \tag{5.1}$$

which indicates that the concentration of DnaA is inversely proportional to the initiation mass S_i. To verify this, I carried out a series of *dnaA* induction experiments, using both the plamid-borne and chromosome-integrated *dnaA*. The results of two steady-state measurement, along with population measurement are pooled together in one graph to show the quantitative relationship between DnaA level and initiation mass. Similar measurements have been done by Atlung and Hansen [6], which are included in the same graph for comparison (Figure 5.4). It can be seen that all points fell on one master curve as initiation mass is inversely proportional to the DnaA level (Figure 5.4).

It can be seen from Figure 5.4 that the initiation mass did change inversely proportionally to the change in DnaA level. However, one caveat of the overexpression strain used above is that it does not allow direct quantitation of the initiator level. Instead, it has to be inferred based on previous reported values and the concentration of inducer used. This result can be improved by using the dual marker strain (*dnaA* transcriptional reporter with replisome marker).

By using this titratable strain, I was able to systematically vary *dnaA* expression level and measure the initiation mass. The result agrees well with the

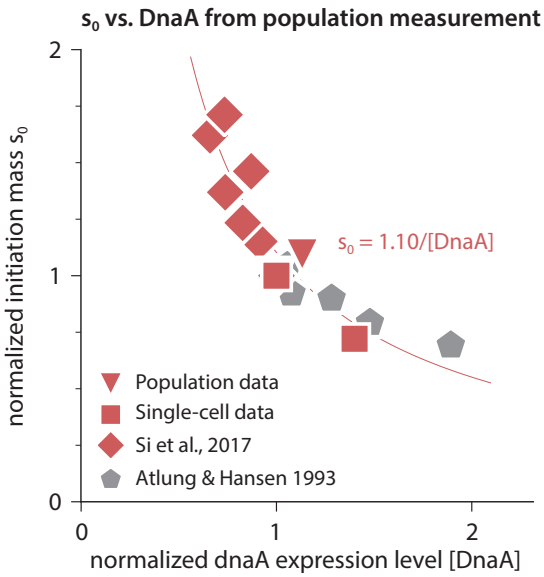

Figure 5.4: **DnaA level and the initiation mass are inversely proportional**. Population measurement of DnaA level and initiation mass are shown with various markers distinguishing different sources of data. Each set of data is normalized independently so that the control group representing wildtype becomes 1. All points are used to fit a simple $y = C/x$ curve and the fitting result is shown as the red line. The single-cell data (square) is generated from mother machien experiments with strains containing pLR40 plasmid and population data (triangle) came from TSTAT experiment of strains harboring pSN305 plasmid. The tCRISPRi data (diamond) from [182] is also included to show that this relationship holds in initiator underexpression situations. A classical study from Hansen (pentagon) was also indicated (grey markers) on this plot to show its consistency with our results [6]. DnaA level cannot be directly quantified in the overexpression strains and are estimated based on the induction curve of published reports where the respective plasmid was used [6, 139]. In tCRISPRi data, DnaA level is estimated based on a *dnaA* transcriptional reporter. Initiation mass is directly measured for single-cell data with the fluorescent replisome marker YPet-*dnaN*, and calculated for population data and tCRISPRi data with cell cycle and cell size measurement.

population data (Figure 5.4) and the inverse relationship still holds for the initiation mass and the DnaA level.

5.4 Initiation control through DnaA level and ATP-to-ADP ratio

The inverse relationship between initiaiton mass and DnaA level does not take into account that only the active form DnaA-ATP can trigger initiation. If in the above titration experiment, all overexpressed *dnaA* were in the ADP form, it would not be as effective in reducing initiation mass as what we observed. Similarly, overexpression of the active form of DnaA-ATP will have a more pronunced effect on triggering initiation. In this section, the role of two forms of DnaA and their ratio (ATP-to-ADP ratio for short) on regulating initiation will be discussed.

As discussed in the previous chapter (Figure 4.2), Donachie proposed that the ATP-to-ADP ratio triggers initiation when it reaches to a critical threshold. However, as we have showed in the above section, an elevated level of total DnaA is sufficient to cause overinitiation. This observation contradicts Donachie's model since overproduced DnaA did not significantly alter the ATP-to-ADP ratio [139]. It is not known if changing ATP-to-ADP ratio but not the total DnaA level can trigger initiation. It is also possible that the critical threshold of DnaA level we observed exists for the active form of *dnaA*.

We can see initiation as a function of two variables - both DnaA level and the ATP-to-ADP ratio of the initiators. More intuitively, one can view this in a phase diagram of these two variables and ask how does initiation mass change in this 2D

plane (Figure 5.5). In this phase diagram, the origin represent wildtype DnaA level and ATP-to-ADP ratio. The four quandrants of the phase space represents high DnaA level and high ATP-to-ADP ratio (top right quadrant), high DnaA level but low ATP-to-ADP ratio (lower right quadrant), low DnaA level but high ATP-to-ADP ratio (upper left quadrant) and low DnaA level and low ATP-to-ADP ratio (lower left quadrant). Previous studies have shown that increasing DnaA level leads to overinitiaiton [6] and vice versa [182]. Similarly, increasing the ATP-to-ADP ratio by perturbing the interconversion system leads to overinitiation [97, 139] and decreasing this ratio leads to underinitiation [48, 51]. However, it is not clear how to map current understandings onto this phase diagram as the interdependency of DnaA level and ATP-to-ADP ratio is not well established. While it was suggested that *dnaA* overexpression has minimal effect on the ATP-to-ADP ratio [139], the case for downregulation is not clear.

In light of this problem of interdependency between DnaA level and ATP-to-ADP ratio, we looked for a direct way to change ATP-to-ADP ratio of *dnaA*, orthogonal to *dnaA* overexpression. *dnaA*R334A is a mutant locked in the active form and cannot be effectively hydrolyzed [139]. Expression of this R334A mutant caused overinitiation and initiation asynchrony [139, 186]. This mutant can be expressed as a way to directly increase the DnaA-ATP level without changing the DnaA-ADP level, thus provides direct control of the ATP-to-ADP ratio. Alternative ways of changing ATP-to-ADP ratio includes perturbing the interconversion of two forms of DnaA including RIDA, DDAH or DARS. However, the latter requires direct biochemical measurement of the change in ATP-to-ADP ratio, which we were unable to carry out at the time. Instead, titration of the R334A mu-

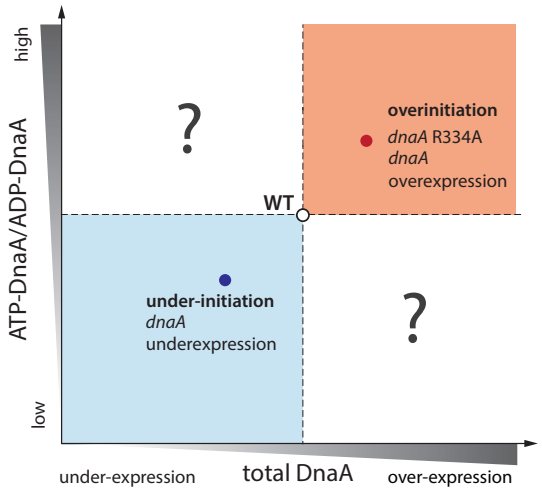

Figure 5.5: **Initiation phase diagram**. The phase diagram illustrates the relationship between initiation mass and both the DnaA level and the ATP-to-ADP ratio. Initiation mass decreases either through wildtype *dnaA* overexpresssion or increasing ATP-to-ADP ratio (by means of expressing R334A mutant or disrupting DnaA hydrolysis [96]). Conversely, knocking down *dnaA* increases initiation mass [182]. However, it is not clear how initiation mass will change in the top-left or bottom right quadrant when DnaA level and the ATP-to-ADP ratio sit on the opposite side relative to the wildtype point.

tant provides an easier way to directly modulate the ATP-to-ADP ratio as all the mutant proteins are considered to be in the ATP form [94].

By combining the expression plasmid of R334A with the *dnaA* tCRISPRi strain, I was able to change the ATP-to-ADP ratio systematically and measure the response in initiation change. To isolate the effect of overexpression of *dnaA* from the effect of changes in ATP-to-ADP ratio, a control group of overexpressing wildtype *dnaA* was added. At comparable DnaA levels, overexpression of the R334A greatly reduced the initiation mass (Figure 5.6), suggesting an independent role of the ATP-to-ADP ratio in triggering initiation. However, more data is needed to draw quantitatitve conclusion on the functional relationship between initiation mass and the two variables in the phase diagram.

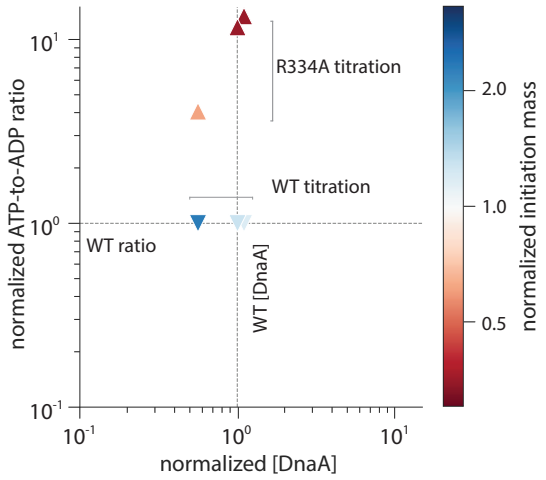

Figure 5.6: **DnaA titration phase diagram**. Two strains overexpressing either wildtype *dnaA* (downward triangle ▼) or R334A mutant (upward triangle ▲) were grown in TSTAT and measured for their initiaiton mass. Each marker is color-coded based on the initiation mass mesasured in reference to the colormap shown on the right. Both ATP-to-ADP ratio (y-axis) and DnaA concentration (x-axis) are normalized based on the wildtype values.

5.5 Future work on initiation control

5.5.1 Differentiating two regulatory mechanisms of intiation

While our initiation threshold model sheds light on the regulatory mechanism of replication initiation, it remains unclear how cells maintain a tight control of the initiation mass (CV $\sim$ 0.1 - 0.15) in many grwoth conditions. There remains a gap of knowledge in the molecular mechanism to achieve such tight regulation. Based on our model, two possible reasons can explain this regulation: 1) the regulation of initiator dnaA expression noise; 2) higher order protein-protein interaction (cooperativity).

These two possibilities can be differentiated if we examine the functional

response of initiation to varying level of initiator input, similar to a dose-reponse curve. One can synthesize such dose-response curve by examining the single-cell initiation trajectory (Figure 5.7). The temporal accumulation of DnaA can be expressed in reference to the total DnaA in the cell, and cells exist either in pre-initiation or post-initiation state. The fraction of cells switched state at any given DnaA level provides estimation of the probability of initiaiton at that corresponding DnaA level. Then the ensemble curve can be constructed that estiamtes the probability function of initiation.

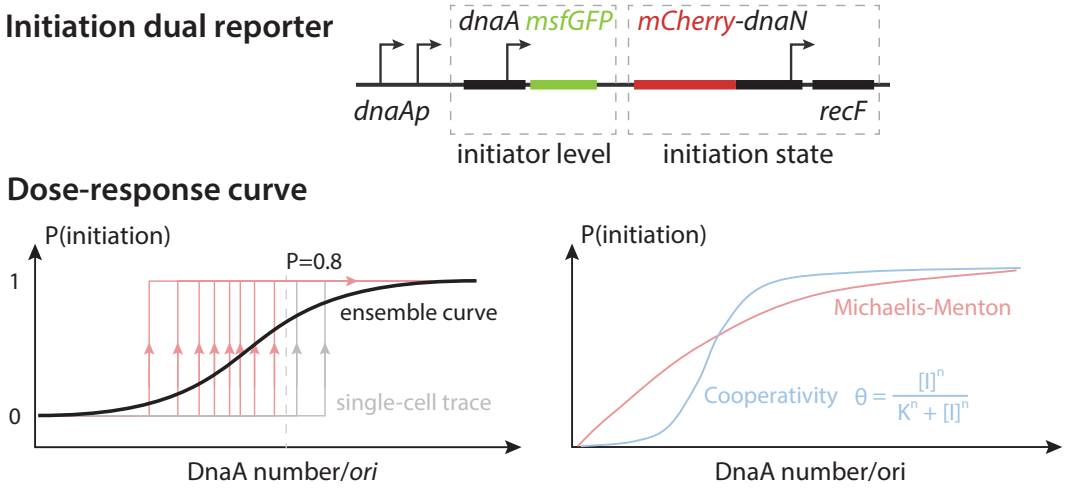

Figure 5.7: **Schematics of initiation dose-response measurement**. Upper: design of the initiation dual reporter strain. The transcriptional reporter msfGFP is used to determine the initiator level and the fused mCherry-dnaN is used to determine the state of initiation. Middle: initiation probability compiled from single-cell initiation curve. Each single-cell trajectory is drawn in grey line, and initiation trigger is visualzed as the cell switch from state 0 to state 1. Variation exists as to the exact DnaA level at which each cell triggers, and the probability of initiation can be calculated as the fraction of cells that have initiated at the given DnaA level. In the illustrated example, eight of the ten cells have initiated (red lines) at the given DnaA level hence the initiation probability is estimated to be 0.8. Bottom: The dose-response curve of initiation regulation. The shape of this curve reflects the regulatory mechanism of initiation. The sigmoidal curve suggests of cooperativity among initiators, quantified by the Hill coefficient n in the equation. The hyperbolic curve suggests simple Michaelis-Menton type of reaction where Hill coefficient equals to 1.

If the cooperativity is the predominant feature of initiation control, one would expect a sigmoidal shape of curve with a high Hill coefficient. Otherwise, it would show up as a hyperbolic curve, typical of Michaelis-Menton type of reaction due to mass action (Hill coefficient approximates 1). This can be directly tested in our system in mother machine with a dual-reporter strain, one for dnaA level, one for replisome/initiation. Specifically, a dnaA with msfGFP transcriptional reporter and mCherry-dnaN replisome marker is sufficient. One can measure the DnaA level at which each cell triggers initiation. The initiation probability at any given DnaA level is calculated as the fraction of cells with that DnaA level that have initiated.

5.5.2 Initiator cooperativity

If the above dose repsonse curve can be confirmed to resemble a sigmoidal shape, we would further look into the roles of DnaA protein-protein interaction. DnaA domain I is involved in protein-protein interaction. Among the binding parterns of DnaA, the DnaA initiator associating protein (*DiaA*) is a DnaA homologue that recruits DnaA in tetrameric form, and brings DnaA to close proximity for cooperative binding to DnaA boxes in *oriC*. Mutation study of *diaA* does not show a change in average behavior in initiation, but *diaA* can rescue dnaA46 mutant phenotype [84, 99, 100]. Our own TSTAT resutls did not show a change in initiation mass with *diaA* knockdown using tCRISPRi (personal communication). However, it is not known whether *diaA* knockdown will change the cooperativity behavior of DnaA in initiation. This can be readily tested by measuring the dose-response curve using our tCRISPRi system.

Additionally, certain residues in this domain are responsible for DnaA dimerization or recruitment of helicase *DnaB* [64]. Other nucleotide associated protein (NAP) including *Fis* and IHF have been shown to bend the *oriC* to regulate DnaA oligomerization and initiation complex assembly [111]. These interactions can also be exploited to test the importance of cooperativity/protein interaction in initiaiton.

5.5.3 Initiator gene expression noise

The other possible source of initiation noise comes from the expression of initiator itself. There are two notable features of *dnaA* gene that potentially limits its expression noise. First it is the autoregulation feature, i.e. the initiator DnaA binds to its own promoter to supress *dnaA* expression. The second feature is the fact that *dnaA* has high transcription rate and low translation rate. Based on Flemming Hansen's estimate, this leads to a roughly one-to-one correspondence of mRNA tranascript to protein [64]. The high transcription rate is an effective way to reduce fluctuation in transcription (as opposed to certain bursty promoter[II]), the low translation rate further reduces bursty behavior of the system[III]. Experimental and theoretical work by van Oudenarrden suggested that such high transcription rate coupled with low translation rate is the key to suppressing expression noise [145, 202]. From a molecularpoint of view, the atypical strat codon GTG instead

[II] Transcriptional bursting has been shown to be a wide-spreadh phenomenon across species and has been experimentally demonstrated in bacteria [57]. Promoters are modeled to be switching between on an off state, and the random switching between two states contributes to the fluctuation of gene expression.

[III] High expression level would result in autoregulation and shutdown of the expression system. The expression cannot resume until volumetric dilution of the initiator concentration. In contrast, a lower level of protein production helps maintain a steady production over time.

of ATG reduces the translation rate by an estimate of 20% [68], combined with a poor ribosome binding site (GGAG) could explain 2 - 3 fold lower translation efficiency [64] the observed low translation rate.

To test this, one can first compare the expression noise of *dnaA* to a consitutively expressed fluorescent protein. Then, engineering of the start codon and ribosome binding site (RBS)[IV] will presumably increases the translation efficiency. More proteins produced per transcript will lead to a bursty expression pattern due to the autoregulation nature of the *dnaA* gene. Therefore, by modifying the sequence in *dnaA* start codon and RBS, we expect to see a change in initiator expression noise, and an increase in initiation control noise accordingly.

A comprehensive census of gene, transcript and protein copy number of the *dnaA* could help quantify the initiator expression noise. Such an approach would also benefit modeling effots to better understand initiator expression dynamics and its fluctuation within cell cycle. The *dnaA* gene copy number can be approximated by that of *oriC*, hence derived from the initiation state. The protein copy number is measurable with the transcriptional reporter. However, the measurement of transcript number is non-trivial. There exists live-cell imaging of transcript using the MS2 system derived from the RNA bacateriophage coat protein. MS2-binding sites are inserted in the gene where the MS2-FP fusion protein can bind once trascribed [56]. However, the binding of fusion protein onto nascent transcript likley affects protein translation due to the steric hindrance (personal communication with Scott Rifkin, UCSD). An alternative is to employ fluorescence

[IV] The RBS is a region in mRNA upstream of the start codon that recruits the ribosome for translation. This sequence, also known as the Shine-Dalgarno (SD) sequence, affects the translation initiation rate.

in situ hybridization (FISH) (Figure 5.8), which has been used for RNA imaging and transcript counting [156, 181, 188]. However, typical FISH experiment requires fixation of cells which can not be directly combined with our mother machine time lapse imagine regime. To remedy this, we plan to fix the cells *in situ* at the end of the mother machine experiment (Figure 5.8). In this way, one can identify individual cells in FISH experiments and map their identities in the mother machine experiment where their life histories of growth, cell cycle, initiator expression have been recorded. This "film-and-fix" approach combines the information collected from two experiments to allow for deciphering the expression dynamics indirectly.

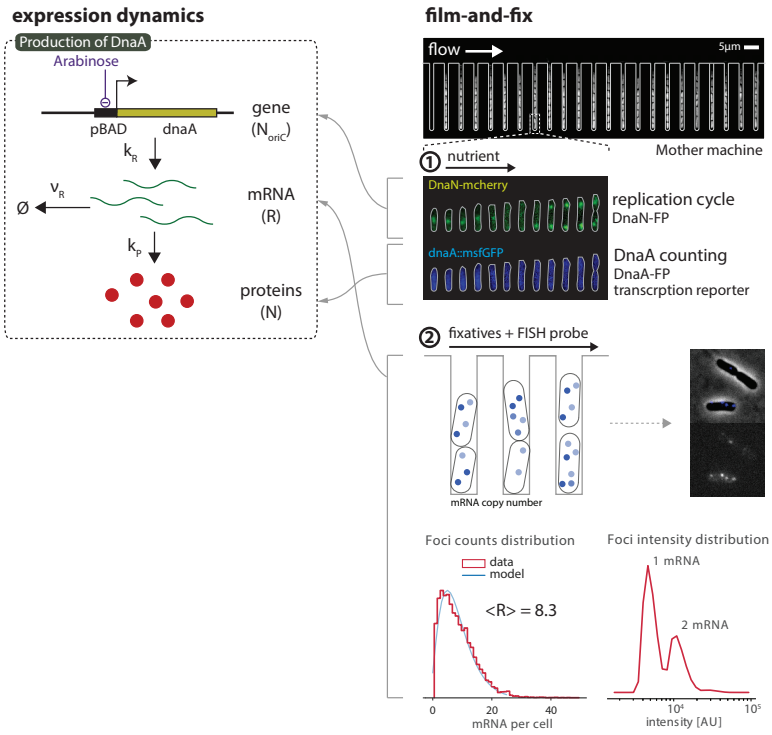

Figure 5.8: **Schematics of film fix experiment**. The expression dynamics of *dnaA* from gene to protein can be measured with the film fix method at single-cell level as described. FISH works through binding of fluorophore-labeled short oligos tiling the RNA transcript. Usually 20-50 different oligos are used to cover the transcript and they act together to amplify the signal and achieve single-molecule sensitivity in detection. Preliminary FISH results (bottom) on batch culture samples shows the applicability of FISH to *dnaA* transcript counting. The number of transcript is counted by detection of fluorescent spot, and weighted by intensity to correct for overlapping spots.

Appendix A

Cell culture

Most of the bacterial physiology works around steady-state culture and it is therefore of great importance to maintain cell culture in steady state. Apart of manually monitoring cell growth and diluting the cutlture, two major automated methods are used throughout the thesis to generate steady-state cell samples for various measurements, turbidostat (TSTAT) and microfluidic single-cell culture device ("mother machine"). They are detailed in the following sections.

A.1 Turbidostat

TSTAT is a continuous culturing device that monitors cell growth by OD measurement and dilutes the culture at preset OD values (Figure A.1). TSTAT is a multiplex device that handles up to eight vials of culture simultaneously, hence increasing the experiment throughput.

A.1.1 TSTAT system setup

Different from the typical turbidostat which maintains a constant turbidity of the cell culture, our TSTATs are set to allow the accumulation of turbidity up to a preset threshold. Upon reaching this upper turbidity threshold each TSTAT dilutes its culture until a defined lower turbidity threshold. After dilution, the culture in each TSTAT grows continuously until the upper turbidity threshold is reached again. This process repeats and the culture maintains steady-state growth. The turbidity of the culture is measured at 15-second intervals. The individual exponential curves between consecutive dilutions can be fit to an exponential for measurement of the doubling times.

Each TSTAT has its own fixed-wavelength spectrophotometer for taking real time turbidity measurements. The spec- trophotometer componentry of each TSTAT is built into a culture vial cup holder with one 600 nm T-1$\frac{3}{4}$ form factor LED fitted into one end facing opposite to a 570 nm (FWHM 350 nm) broad band T-1$\frac{3}{4}$ form factor phototransistor press fitted into the other end. With a culture vial placed in the vertical bore of the holder, the LED and phototransistor act as a simple fixed-wavelength spectrophotometer. The LEDs are powered by a microcontroller board (Arduino Mega 2560) for each TSTAT and measurements of the current through the LEDs and the phototransistors of each TSTAT is done by taking voltage measurements across shunts in series. A Java GUI on a computer connected to the microcontroller board provides the main control and logging functionality including the ability to measure and set a blank value for each spectrophotometer.

Separate pneumatically driven media bottles supply each TSTAT with

growth media for dilution. The flow in the dilution/media lines is mechanically controlled by pinch valves (Bio-Chem Fluidics, Inc., NJ). A regulated diaphragm pump supplies driving pressure (15 kPa) to each media bottle as well as the compressed air flow which bubbles the cultures in each TSTAT vial to assure the aeration of the cell culture. A small stir bar provides additional agitation to the culture. All TSTAT experiments were performed at 37 °C in an environmental warm room to avoid temperature fluctuation.

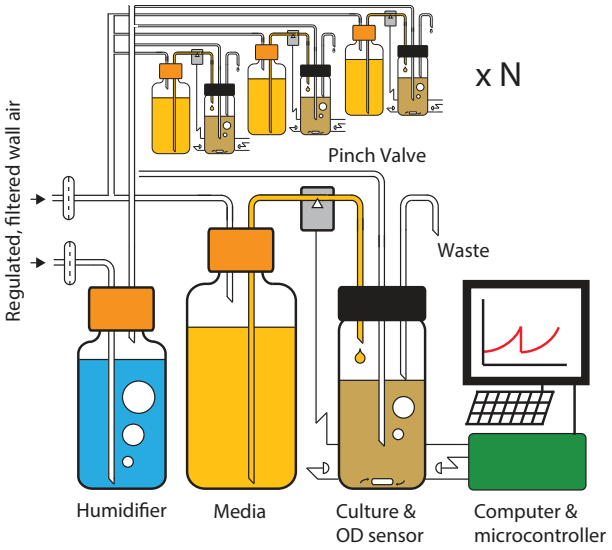

Figure A.1: **TSTAT system setup**. Figure adpated from [182].

A.1.2 Calibration of TSTAT and grwoth rate measurement

The phototransistor current is read off an analog-digital converter (ADC) connected to the phototransistor shunt. Phototransistor current readings are measured at different light intensities and an intensity-current (each normalized relative to its respective maximum value) curve was fit to a second-order polynomial. Cell cultures measured at different OD_{600}s were fixed and the working curve of a nor-

malized intensity and OD_{600} was mapped out and used to convert the readings from phototransistor currents to actual OD_{600}. $OD_{600} = -\frac{1}{\alpha}\log(I_{\text{norm}}/I_{\text{norm}}^{\text{blank}})$, where α is the slope of the normalized intensity-OD_{600} curve.

The OD_{600} vs. time curve was fit to a single exponential $I = I_0 2^{t/\tau}$ for each segment between two consecutive dilution events and the average was taken as the generation time. Growth rate is given by $\lambda = \ln(2)/\tau$. Each turbidostat was blanked with the media used in experiment.

A.1.3 Turbidostat growth experiment, sample collection and imaging

Before every turbidostat experiment, cells were inoculated into 1 mL lysogeny broth (LB) medium as seed culture from a single colony on agar plate, streaked no more than 7 days before use. After 6–12 hr in 30 °C or 37 °C water bath shaker, cells were diluted 1,000-fold into 1–2 mL of the specific growth medium as preculture and shaken at 37 °C in water bath until OD600 = 0.2. The preculture was then back-diluted 1,000-fold again into the same medium and shaken at 37 °C in water bath until OD600 = 0.2. The back-diluted culture was then inoculated into each turbidostat vial with or without specific inhibiting conditions, and the turbidostat experiment was started with $OD_{600} = 0.05$ for each vial. The turbidostat was then run for at least eight generations in steady-state growth before sample collection.

Cell samples were collected when cell culture reached $OD_{600} = 0.2$ or 0.4. Typically 20 ml of cell culture would be collected and kept on ice immediately to prevent further growth or dramatic change of cell physiology. From the same

culture, cell samples were collected for cell size, genomic DNA for qPCR, image cytometry, RNA/protein and transcription level measurements (fluorescence imaging), respectively. Cells for imaging are fixed by adding 0.24% w/v formaldehyde into the growth culture and imaged within 24 hr.

Before imaging, cells were concentrated to an appropriate cell density by centrifuging and removal of excess supernatant. 0.1 mM 2-mercaptoethanol (Sigma) was added to reduce cell clumping. Cells were pipetted onto an agarose pad made by 0.2% w/v agarose (Genesee Scientific, CA). The agarose pad was then flipped and put onto a Willco dish (WillCo Wells B.V.) with a glass coverslip on top to reduce the evaporation during imaging. All transmission light and fluorescence microscopy were performed on an in- verted microscope (Nikon Ti-E) with Perfect Focus 2 (PFS 2), 1003 oil immer- sion objective (PH3, numerical aperture [NA] = 1.45), light-emitting diode (LED) fluorescent light (Lumencor), and Andor NEO sCMOS camera (Andor Technol- ogy). Exposure time was between 100 and 200 ms with 100% transmission. From each set of samples, 140–300 images were captured and 5,000–30,000 cells were analyzed to ensure statistically significant distributions of cell measurements, such as cell size or DNA content. All cell image analysis was carried out by custom software written in Python employing the OpenCV library.

A.2 Mother machine

Mother machine is suitable for monitoring single cell growth and carrying out time lapse fluorescence imaging. The growth data (doubling time, growth rate, cell size) can be related other aspects of the cellular activity as revealed by fluores-

cence imgaing (e.g. cell cycle progression, abundance of protein of interest, Figure A.2). Detailed protocol avaiable at the Jun lab website (https://jun.ucsd.edu). The readers can be referred to other published protocols [197].

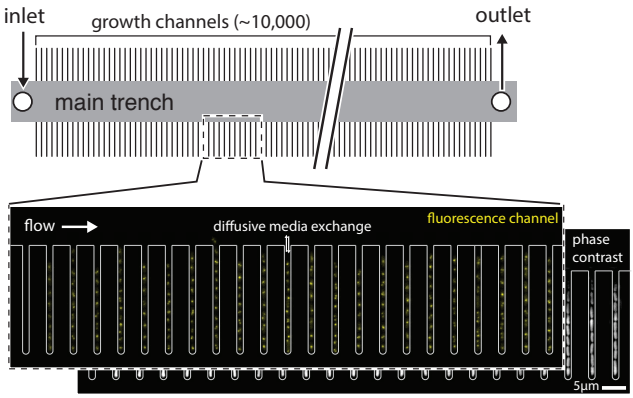

Figure A.2: **Mother machine layout**.

A mother machine experiment consists of the following steps: device making, cell preculture, microscope setup, imaging experiment and data analysis (Figure A.3).

A.2.1 Preparation of microfluidic device and device assembly

PDMS was prepared from a Sylgard 184 Silicone Elastomer kit: polymer base and curing agent were mixed in a 10 to 1 ratio, air bubbles were purged from the mixture in a vacuum chamber, the degassed mixture was poured over the master, and the devices were cured about 24 hours at 65C. Cured PDMS has a consistency like rubber; devices were peeled from the master mold. Devices were treated with pentane and then acetone to remove residual uncured polymer from the PDMS matrix.

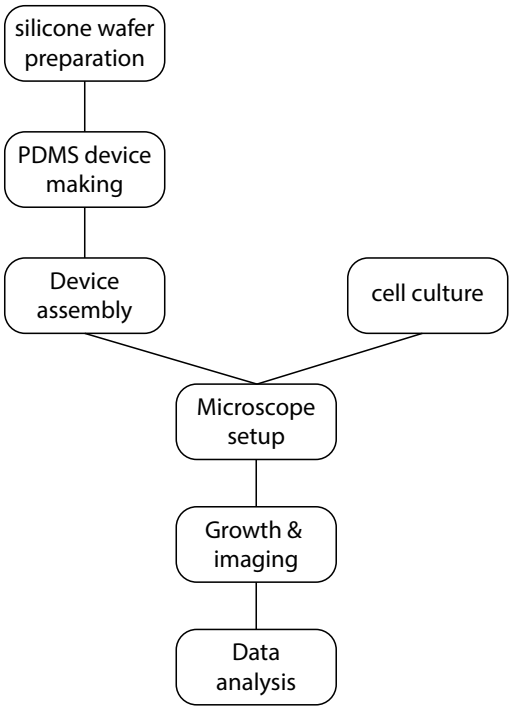

Figure A.3: **Mother machine experiment flowchart**.

Punch holes with 1.0 mm puncher to make the inlet and outlet (can also be done before acetone and pentane wash). To bond the PDMS to the glass surface of the Wilco dish (Wilco Wells), the surfaces were exposed to oxygen plasma for 30 seconds at 30 watts in a Harrick Plasma system. Oxygen plasma makes exposed PDMS and glass reactive, so that covalent bonds form between surfaces brought into contact with one another. The seal between PDMS surfaces was established for 5-10 minutes at 85 °C.

After baking of the device, infuse the device with 0.5 mg/ml BSA to make the PDMS surface hydrophilic before loading the cell. An exponential cell culture is spun down after optionally adding to final concentration of 1 mM of β-mercaptoethanol (BME, reducing reagent to prevent clumping). The culture is

then resuspended in 20-40 μl of 0.5 mg/ml BSA solution and loaded to the device with a syringe. The device is then spun on a small custom made centrifuge to force cells into channels. After visually inspecting the device and cell loading rate, the device is ready for experiment.

A.2.2 Setting up mother machine experiment

Prewarm media to be used in the experiment to the desired temperature in a syringe capped with a 0.22 μm filter to keep it sterile. Connect the media syringe to a polyethylene tubing through a 27-gauge needle, connec the other end to the inlet of the device. The outlet is then connected with another tube. Pump the media through the device to ensure a good seal, then set the flow rate to 0.5 - 1 ml/hr. Proceed with setting the microscope (powering all components, focusing sample, fixing glass dish) and aligning the wilco dish to the frame of the camera sensor.

Use Nikon NIS elements software to setup multidimensional acquisition with approriate settings for stage posisitions, illumination settings and time-lapse schemes. At the end of the experiment, data is analyzed using custom Python script.

Appendix B

Physiology variables measurement

B.1 Population level measurement

B.1.1 Genomic DNA extraction

Below is a protocol for extracting genomic DNA to be used for qPCR assay.

Genomic DNA extraction

1. Pellet 1 mL culture at OD 0.2/0.4 and quick freeze in dry ice (alternatively can be spun down immediately after sampling)
2. Pellet cells by centrifugation 15,000 RPM, 2-3 min Resuspend in 250-300 μl of TES buffer (10 mM Tris pH 8, 1 mM EDTA pH 8, 0.6% w/v SDS, 0.12 mg/mL proteinase K, 0.5 mg/mL RNase A)
3. Heat at 65 °C for 15 minutes
4. Precipitate proteins with PhOH:CHCl3 (1:1 volume)
5. Vortex till emulsion forms(DNA will shear)
6. Spin max speed, 5-7 minutes

7. Recover top aqueous layer carefully without sucking out too much of the interphase (milky, opaque pellet)
8. Add CHCl3 (30% by volume)
9. Vortex till emulsion forms
10. Spin max speed, 5-7 minutes
11. Recover top aqueous layer carefully without sucking out too much of the interphase (milky, opaque pellet)
12. Add 10% by volume of 3 M NaOAc pH 5.2, this will adjust the pH
13. Vortex
14. Add 70% by volume room temperature isopropanol
15. Close tube and mix by holding the tube horizontally and shaking along the long axis vigorously 15-20 times
16. (optional, can also leave longer) Store mixture at -20 °C for 15 minutes
17. Pellet precipitate by centrifugation at max speed for 5 minutes in cold room (4 °C)
18. Decant supernatant onto a paper towel
19. Collect remaining supernatant by spinning in a centrifuge for 30 seconds
20. Remove remaining supernatant with a micropipette
21. Leave open at room temperature overnight or completely dry out
22. Resuspend in 50 μl of Tris pH 8 and allow for 30 minutes at 37 or 10 minutes at 55 °C to dissolve the DNA.
23. Measure DNA concentration using a NanoDrop spectrophotometer.

B.1.2 qPCR assay setup

Each qPCR reaction was carried out with PowerUp SYBR Green Master Mix (ABI) with 10 ng of DNA template in a 20 μl reaction using a Mx3000P qPCR system (Agilent Technologies). All primers were designed such that the size of each amplicon/genomic locus is about 100 bp with similar annealing temperatures. The specificity of PCR amplification was validated by inspecting the melting curves of each reaction.

Fluorescence time traces from the qPCR system typically have a sigmoid shape. All traces were background-subtracted and fit to a single exponential at the amplification stage, from which the amplification efficiency (the exponential prefactor, $\alpha \in (1, 2]$) was derived. We did not observe systematic differences of amplification efficiency among different primer pairs and used the average as a global parameter for each experiment. The threshold value (H) was determined as the interpolated replication cycle number (down to two digits) at which the fluorescence signal rises above a given threshold within the exponential phase. The exact value can be varied without affecting the final results. DNA amplification can be simplified as an exponential increase such that $[DNA]_c = [DNA]_0 \cdot \alpha^H$ with $\alpha = 2$ in an ideal PCR reaction. Therefore, the amount of initial DNA is exponentially dependent on H and $-(\mathrm{H}_1 - \mathrm{H}_2)$ reflects the relative abundance of two genomic loci in the sample. As discussed below, the ratio of DNA abundance from two genomic loci is also exponentially dependent on the ratio of C/τ.

B.1.3 Image cytometry measurement

The image cytometry procedure measures the total amount of DNA and estimates its genome equivalent in reference to the standard cell. The standard cell is generated by a run-out process that blocks new initiation but allows current replication to finish. In this way, all cells contain an integer number of fully replicated chromosome (Figure B.1).

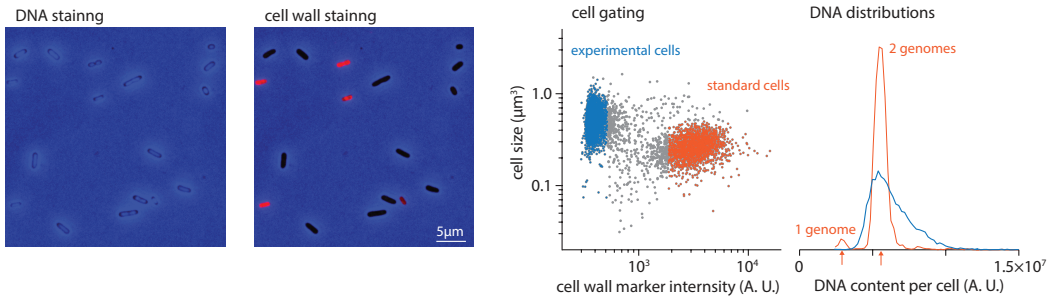

Figure B.1: **Image cytometry** The sample cells are mixed with the standard cells (highlighted in red) to ensure even illumination and fair comparison of the DNA amount. The standard cells are stained with memebrane dyes to differentiate them from the rest. The standard cells have two distinct peaks in the DNA content distribution, corresponding to one and two genome equivalents. The genome equivalent of the sample can thus be determined by normalizing based on the standards.

Below is the image cytometry protocol devloped by Fangwei Si in the lab.

Sample collection and fixation

1. Take 1 ml cell culture from flask
2. Wash with pre-cold 1X TE
3. Add 0.1 ml TE and 0.9ml pre-cold 77% EtOH (finally 70%)
4. Store at 4 °C

Staining

1. Wash cells of MOPS mannose Run-out cells (250ul) and sample cell (500ul) with TBS (10X stock: 4.8486 g/200 ml Tris (MW121.14, 200mM), 15.1944 g/200 ml NaCl (MW58.44, 1.3M), PH 7.42 (use HCl))
2. Concentrate sample cells to 250 μl
3. Stain MOPS mannose cells with WGA 594 (20ul 1 mg/ml into 250ul, finally 80 μg/ml; STOCK: 1 mg/ml in 1X PBS diluted from 10X PBS)
4. Wrap with foil and shake at 30C for 1.5 hrs
5. Wash WGA by TBS twice
6. Add 250ul MOPS mannose cells into sample cells
7. Add finally volume to 1.25 ml in 1.5 ml centrifuge tube
8. Stain mixed cells with Hoechst dye (37.5 μl 100ug/ml into 5ml finally 3 μg/ml) and BME (12.5ul 0.1M into 5ml, finally 1mM)
9. Leave on ice for 45 min
10. Concentrate to 40-70 μl before imaging

Preparation of agar pad on Wilco dish

1. 2% agarose
2. use coverslip to press onto the agar pad

Image cytometry

1. Filters
   (a) Hoechst dye: Quad

(b) WGA 594: Quad

(c) Phase contrast: None

2. Exposure time

(a) Hoechst dye: 100 ms

(b) WGA 594: 200 ms

(c) Phase contrast: 100 ms

3. Custom ND acquisition

B.1.4 Ribosome fraction measurement

The measurement of total RNA and protein follows the protocol previously reported by Hui et al. [83].

B.2 Single-cell measurement: replisome imaging

Imaging replisome in time lapse fluorescence microscopy allows tracking of replication progress. Two replisome-related proteins used are DnaN and DnaQ, which are fused with fluorescent protein Ypet or mCherry (Figure B.2, [159]).

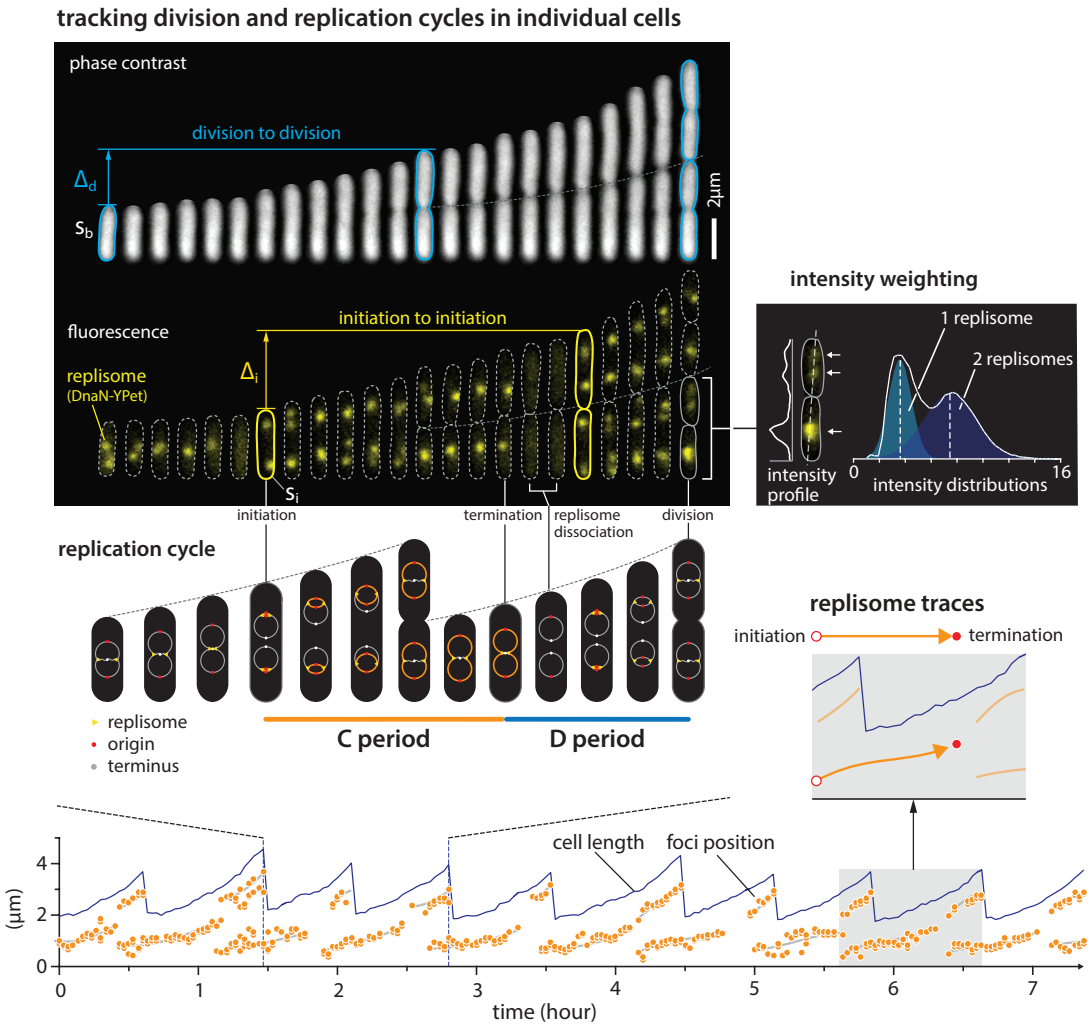

Figure B.2: **Replisome imaging**. Fluorescence time lapse imaging of Ypet-dnaN gives a replisome trace over time, where initiation and termination events can be identified. The replication period can thus be determined. Fluorescent foci close to each other cannot be optically resolved. Instead, the intensity of each focus is measurd and compared to the bimodal distribution of foci fluorescence intensity. The initiation mass S_i can be readily measurd from the frame where initiation is called.

Appendix C

Strains and Cloning

C.1 Genetic recombineering

The following protocol describes the procedure for carrying out λ Red recombineering through uses of pSIM18 plasmid. This temperature sensitive plasmid can only be replicated at 32 °C or lower temperature. The plasmid contains temperature-sensitive CI repressor (CI857), which represses expression from the λ pL promoter that drives the expression of the recominase gene [107, 203]. The plasmid is cured after extensive culture at 37 °C.

1. Grow 3 mL of cell culture in LB till OD 0.4 (assuming the strain has been transformed with pSIM18)
2. Transfer cells to 42 °C water bath for 15 minutes.
3. Transfer cells in the baffle flask immediately to an ice bath for at least 10 minutes. Cells should remain cold from this step on until electroporation.
4. Prechill Milli Q water on the ice bath.
5. Wash cells with ice cold water three times, pellet cells by centrifugation (30

seconds at most to prevent heating up).

6. Resuspend the pellet of cells with cold water, use 50 μl per reaction.
7. Mix 100 - 150 ng of PCR fragment with the cell suspension. Prepare a separte control without DNA.
8. Transfer the mixture or control into a 0.1-cm electroporation cuvettes.
9. Electroporates at 1.8 kV.
10. Transfer the electroporated cell suspension into 1 mL of LB media without any antibiotics. Recover the cells by culturing them at 32 or 37 °C 2 - 4 hours.
11. Spread a fraction of the recovery culture onto the selection plate. If low recombinantion frequency is expected, concentrate the culture before plating.
12. Incubate selection plate overnight.

C.2 P1 phage transduction

Preparation of P1 lysis

1. Add 10mL LB to 2 flasks (P1 and control)
2. Add 50 μl of 1M CaCl2.
3. Add 100 μl 20% glucose.
4. Add 200 μl cell culture.
5. Shake in 37 °C (for MG1655; may vary for other strain background) water bath for 30 minutes.
6. Add 100 μl MG1655 P1 lysis to P1 flask only.
7. Check after an hour to check that both have similar OD.

8. Check every 20 minutes afterwards. Look for noticeable differences in OD.
9. P1 lysis is ready when the solution is relatively clear.
10. Add 100-200 μl of chloroform to kill remaining cells in the P1 lysate, then spin down the lysate in a glass tube with a swing centrifuges, 5 min, take supernatant (Chloroform remains at the bottom organic phase with cell debris)
11. Store in 4 °C fridge

P1 transduction

1. Pipet 1500 μl of cell culture into a micro centrifuge tube.
2. Centrifuge at max speed for 30 seconds.
3. Discard supernatant.
4. Add 100 μl P1 Salt (10 mM $CaCl_2$/5 mM $MgSO_4$) x of transductions + 1 control.
5. Resuspend pellet in P1 salt
6. Distribute 100 μl of cell solution to a separate micro centrifuge tube for each transduction.
7. Add 100 μl of P1 Lysis to the cell solution and mix.
8. Let stand at room temperature for 30 minutes.
9. Add 1 mL LB to 1 culture tube for each transduction.
10. Add 200 μl 1M sodium citrate to each culture tube.
11. Once 30 minutes are up, transfer the micro centrifuge contents to the culture tubes.
12. Transfer to 32/37 °C water bath for 2-4 hours.

13. Once cells have had enough time to recover, transfer the cell culture to micro centrifuge tubes.
14. Centrifuge at max speed for 30 seconds, discard all but 100 μl of supernatant.
15. Resuspend pellet in remaining supernatant, and spread on the selection plate.
16. Incubate overnight (there are two possible controls, one using the P1 control on non-selecting plate to show that the P1 lysate doesn't contain any cell, the other non-transducer recipient cells on selection plate)

C.3 Strain construction

The following section lists the strategy to construct several strains used in this thesis.

C.3.1 Construction of fast-maturing fluorescence reporter cassette

To use gene expression output as reporter for cell cycle timing, I considered constructing an expression cassette with little delay in gene expression and high expression level to amplify the difference caused by gene copy number doubling. To achieve this, I used strong synthetic promoters (pLTetO, pRNA for CFP and pLacI, pRNABR1 for YFP) with optimal Shine-Dalgarno sequence to maximize translation efficiency[87] (Figure C.1). Additionally, to reduce the side effect on neighboring genes[27], I decided to add transcription terminator towards the end of the cassette to "insulate" it from the rest of the genome.

To construct this cassette, I ordered DNA fragment (gBlocks from IDT

DNA, Illinois) with drug selections marker blocks (SmR or AmpR). The cassette is then put together using Gibson Assembly (NEB, Massachusates)[55] and inserted in a low-copy plasmid (pZS*24, gift of Zhongge Zhang at UCSD) with pSC101* replication origin using NdeI sites. Low copy number plasmid was chosen to avoid toxicity of protein over-expression[10]. The resulting plasmid was then transformed into MG1655 wildtype *E. coli* cells and selected on LB agar plate with 100 μg/ml spectinomycin or 100 (μg/ml ampicillin. Transformants are inspected for their fluorescence with microscopes. The cassette can then be integrated into the chromosome by recombineering.

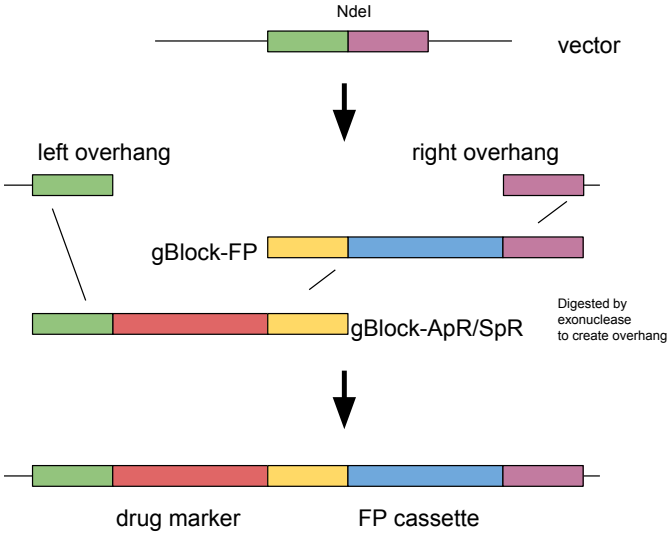

Figure C.1: **Schematics for gibson assembly of fast-maturing fluorescent reporter cassette**.

C.3.2 tCRISPRi system

The tCRISPRi system expresses dCas9 under the P^*_{BAD} promoter with sgRNA constitutively expressed. The dCas9 combined with sgRNA blocks the gene of interest from being transcrbied (Figure C.2). tCRISPRi strains are generated

by recombineering 70-nt single strand DNA oligos with complementary sequence to the gene of interest into SJ_XTL219 strain and selected with sucrose. The oligo sequence corresponds to the sgRNA sequence and can looked up from [115]. The knockdown level is quantified by a transcriptional GFP reporter inserted after the gene of interest.

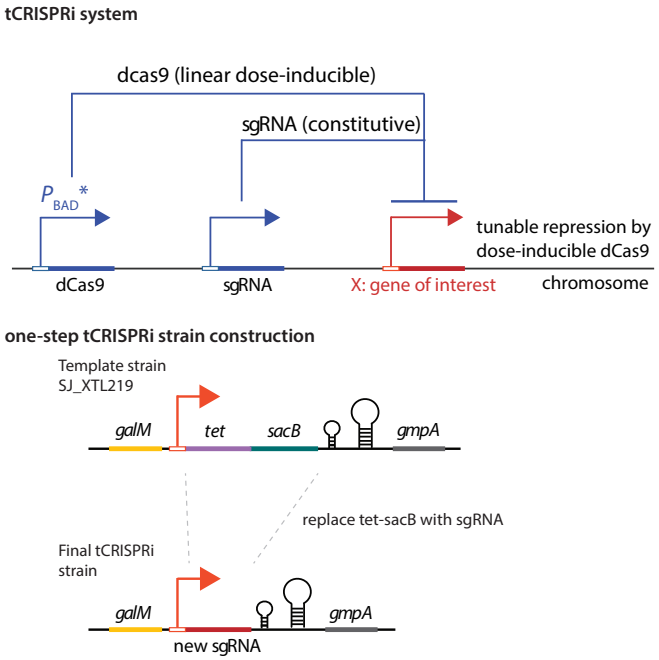

Figure C.2: **tCRISPRi strain construction**.

C.4 List of strains used in the thesis

Below is a list of strains used in this thesis work.

Table C.1: **Strain list**.

Strain number	**Alias**	**Genotype**
SJ753	SJ_XTL175	lacI, araFGH<>spec, araE, lacY A177C araD<>tet-sacB pSIM18
SJ797	SJ_XTL219	pBAD::dCas9 LacI araE araFGH<>spec lacYA177C galM//promoter-tet-sacB -handle-terminator//gmpA pSIM18
SJ1171	SJ_XTL571	pBAD::dCas9 LacI araE araFGH<>spec lacYA177C galM//dnaA sgRNA//gmpA dnaA-msfGFP-cm pSIM18
SJ1307	DL07	P1parS 89' thyA
SJ1342	DL42	pZS*24-SmR-pLTetO-CFP-T1
SJ1343	DL43	pZS*24-SmR-pRNA1-CFP-T1
SJ1344	DL44	pZS*24-ApR-pLacIq-YFP-rnpBT1
SJ1345	DL45	pZS*24-ApR-pRNA_pBR-YFP-rnpBT1
SJ1351	DL51	yifB-[SmR-pLTetO-CFP-T1]-ilvL
SJ1352	DL52	yifB-[SmR-pRNA1-CFP-T1]-ilvL
SJ1353	DL53	rsmG-[ApR-pLacIq-YFP-rnpBT1]-atpI
SJ1354	DL54	rsmG-[ApR-pRNA_pBR-YFP-rnpBT1]-atpI
SJ1415	DL115	dnaA-msfGFP, pSN300, pHSL99
SJ1416	DL116	dnaA-msfGFP, pSN305, pHSL99
SJ1417	DL117	dnaA-msfGFP, pSN306, pHSL99
SJ1424	DL124	pBAD::mCherrydnaA-msfGFP-cat
SJ1444	DL144	XTL571 dnaA-msfGFP-amp-mCherrydnaN

Bibliography

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

[2] Akiyama, M. T., Oshima, T., Chumsakul, O., Ishikawa, S., and Maki, H. (2016). Replication fork progression is paused in two large chromosomal zones flanking the DNA replication origin in Escherichia coli. *Genes to Cells*, 21(8):907–914.

[3] Armengod, M. E., Garcia-Sogo, M., and Lambies, E. (1988). Transcriptional organization of the dnaN and recF genes of Escherichia coli K-12. *J. Biol. Chem.*, 263(24):12109–12114.

[4] Atlung, T., Clausen, E., and Hansen, F. G. (1984). Autorepression of the dnaA gene of Escherichia coli. In *Proteins Involv. DNA Replication*, pages 199–207. Springer.

[5] Atlung, T., Clausen, E. S., and Hansen, F. G. (1985). Autoregulation of the dnaA gene of Escherichia coli K12. *MGG Mol. Gen. Genet.*, 200(3):442–450.

[6] Atlung, T. and Hansen, F. G. (1993). Three distinct chromosome replication states are induced by increasing concentrations of DnaA protein in Escherichia coli. *J. Bacteriol.*, 175(20):6537–6545.

[7] Atlung, T., Løbner-Olesen, A., and Hansen, F. G. (1987). Overproduction of DnaA protein stimulates initiation of chromosome and minichromosome replication in Escherichia coli. *MGG Mol. Gen. Genet.*, 206(1):51–59.

[8] Ball, C. A., Osuna, R., Ferguson, K. C., and Johnson, R. C. (1992). Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli. *J. Bacteriol.*, 174(24):8043–56.

[9] Balleza, E., Kim, J. M., and Cluzel, P. (2018). Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods*, 15(1):47–51.

[10] Basan, M., Zhu, M., Dai, X., Warren, M., Sévin, D., Wang, Y. P., and Hwa, T. (2015). Inflating bacterial cells by increased protein synthesis. *Mol. Syst. Biol.*, 11(10):836.

[11] Beattie, T. R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A. J., Leake, M. C., and Reyes-Lamothe, R. (2017). Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife*, 6.

[12] Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F., and Bensimon, D. (1994). Alignment and sensitive detection of DNA by a moving interface. *Science (80-.).*, 265(5181):2096–2098.

[13] Billen, D. (1959). Alterations in the radiosensitivity of Escherichia coli through modification of cellular macromolecular components. *Biochim. Biophys. Acta*, 34:110–116.

[14] Bird, R. E., Louarn, J., Martuscelli, J., and Caro, L. (1972). Origin and sequence of chromosome replication in Escherichia coli. *J. Mol. Biol.*, 70(3):549–566.

[15] Blau, S. and Mordoh, J. (1972). A New Element in the Control of DNA Initiation in Escherichia coli. *Proc. Natl. Acad. Sci.*, 69(10):2895–2898.

[16] Boeneman, K., Fossum, S., Yang, Y., Fingland, N., Skarstad, K., and Crooke, E. (2009). Escherichia coli DnaA forms helical structures along the longitudinal cell axis distinct from MreB filaments. *Mol. Microbiol.*, 72(3):645–657.

[17] Bonhoeffer, F. and Gierer, A. (1963). On the growth mechanism of the bacterial chromosome. *J. Mol. Biol.*, 7(5):534–540.

[18] Bordeaux, J., Welsh, A. W., Agarwal, S., Killiam, E., Baquero, M. T., Hanna, J. A., Anagnostou, V. K., and Rimm, D. L. (2010). Antibody validation. *Biotechniques*, 48(3):197–209.

[19] Bouet, J. Y., Surtees, J. A., and Funnell, B. E. (2000). Stoichiometry of P1 plasmid partition complexes. *J. Biol. Chem.*, 275(11):8213–9.

[20] Boye, E., Stokke, T., Kleckner, N., and Skarstad, K. (1996). Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 93(22):12206–12211.

[21] Braun, R. E., O'Day, K., and Wright, A. (1985). Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell*, 40(1):159–169.

[22] Breier, A. M. and Grossman, A. D. (2007). Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the Bacillus subtilis chromosome. *Mol. Microbiol.*, 64(3):703–718.

[23] Breier, A. M., Weier, H.-U. G., and Cozzarelli, N. R. (2005). Independence of replisomes in Escherichia coli chromosomal replication. *Proc. Natl. Acad. Sci.*, 102(11):3942–3947.

[24] Bremer, H., Dennis, P. P., and Bremer, H. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1):765–77.

[25] Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., and Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. *Cell*, 156(6):1312–1323.

[26] Brown, S. D. and Jun, S. (2015). Complete Genome Sequence of Escherichia coli NCM3722. *Genome Announc.*, 3(4):e00879–15.

[27] Bryant, J. A., Sellars, L. E., Busby, S. J. W., and Lee, D. J. (2014). Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Res.*, 42:11383–11392.

[28] Cairns, J. (1963a). The bacterial chromosome and its manner of replication as seen by autoradiography. *J. Mol. Biol.*, 6(3):IN3–IN5.

[29] Cairns, J. (1963b). The Chromosome of Escherichia coli. *Cold Spring Harb. Symp. Quant. Biol.*, 28:43–46.

[30] Camara, J. E., Breier, A. M., Brendler, T., Austin, S., Cozzarelli, N. R., and Crooke, E. (2005). Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of Escherichia coli DNA replication. *EMBO Rep.*, 6(8):736–741.

[31] Campbell, A. (1957). Synchronization of cell division. *Bacteriol. Rev.*, 21(4):263–272.

[32] Campbell, J. L. and Kleckner, N. (1988). The rate of Dam-mediated DNA adenine methylation in Escherichia coli. *Gene*, 74(1):189–190.

[33] Casaregola, S., Khidhir, M., and Holland, I. B. (1987). Effects of modulation of RNase H production on the recovery of DNA synthesis following UV-irradiation in Escherichia coli. *Mol. Gen. Genet.*, 209(3):494–8.

[34] Chandler, M., Bird, R. E., and Caro, L. (1975). The replication time of the Escherichia coli K12 chromosome as a function of cell doubling time. *J. Mol. Biol.*, 94(1):127–132.

[35] Chen, Y., Milam, S. L., and Erickson, H. P. (2012). SulA Inhibits Assembly of FtsZ by a Simple Sequestration Mechanism. *Biochemistry*, 51(14):3100–3109.

[36] Chiaramello, A. E. and Zyskind, J. W. (1989). Expression of Escherichia coli dnaA and mioC genes as a function of growth rate. *J. Bacteriol.*, 171(8):4272–80.

[37] Christensen, B. B., Atlung, T., and Hansen, F. G. (1999). DnaA boxes are important elements in setting the initiation mass of Escherichia coli. *J. Bacteriol.*, 181(9):2683–8.

[38] Churchward, G. and Bremer, H. (1977). Determination of deoxyribonucleic acid replication time in exponentially growing Escherichia coli B/r. *J. Bacteriol.*, 130(3):1206–1213.

[39] Cohen, S. S. and Barner, H. D. (1954). STUDIES ON UNBALANCED GROWTH IN ESCHERICHIA COLI. *Proc. Natl. Acad. Sci. U. S. A.*, 40(10):885–93.

[40] de Boer, P. A., Crossley, R. E., and Rothfield, L. I. (1990). Central role for the Escherichia coli minC gene product in two different cell division-inhibition systems. *Proc. Natl. Acad. Sci.*, 87(3):1129–1133.

[41] Delamater, E. D. (1954). Cytology of Bacteria Part II: The Bacterial Nucleus. *Annu. Rev. Microbiol.*, 8(1):23–46.

[42] Dillon, S. C. and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, 8(3):185–195.

[43] Donachie, W. (1968). Relationship between cell size and time of initiation of DNA replication. *Nature*, 219(5158):1077 – 1079.

[44] Donachie, W. D. and Blakely, G. W. (2003). Coupling the initiation of chromosome replication to cell size in Escherichia coli. *Curr. Opin. Microbiol.*, 6(2):146–150.

[45] Doudna, J. A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).*, 346(6213):1258096–1258096.

[46] Fan, H., Conn, A. B., Williams, P. B., Diggs, S., Hahm, J., Gamper, H. B., Hou, Y. M., O'Leary, S. E., Wang, Y., and Blaha, G. M. (2017). Transcription-Translation coupling: Direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res.*, 45(19):11043–11055.

[47] Flåtten, I., Morigen, and Skarstad, K. (2009). DnaA protein interacts with RNA polymerase and partially protects it from the effect of rifampicin. *Mol. Microbiol.*, 71(4):1018–1030.

[48] Frimodt-Møller, J., Charbon, G., Krogfelt, K. A., and Løbner-Olesen, A. (2016). DNA Replication Control Is Linked to Genomic Positioning of Control Regions in Escherichia coli. *PLOS Genet.*, 12(9):e1006286.

[49] Froelich, J. M., Phuong, T. K., and Zyskind, J. W. (1996). Fis Binding in the dnaA Operon Promoter Region. *J. Bacteriol.*, 178(20):6006–6012.

[50] Fujimitsu, K. and Katayama, T. (2004). Reactivation of DnaA by DNA sequence-specific nucleotide exchange in vitro. *Biochem. Biophys. Res. Commun.*, 322(2):411–419.

[51] Fujimitsu, K., Senriuchi, T., and Katayama, T. (2009). Specific genomic sequences of E. coli promote replicational initiation by directly reactivating ADP-DnaA. *Genes Dev.*, 23(10):1221–1233.

[52] Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984). The dnaA Protein Complex with the E. coli Chromosomal Replication Origin (OK) and Other DNA Sites. *Cell*, 38(3):869–900.

[53] Fuller, R. S., Kaguni, J. M., and Kornberg, A. (1981). Enzymatic replication of the origin of the Escherichia coli chromosome. *Proc. Natl. Acad. Sci. U. S. A.*, 78(12):7370–4.

[54] Ghosh, R., Gilda, J. E., and Gomes, A. V. (2014). The necessity of and strategies for improving confidence in the accuracy of western blots. *Expert Rev. Proteomics*, 11(5):549–560.

[55] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, 6(5):343–345.

[56] Golding, I. and Cox, E. C. (2004). RNA dynamics in live Escherichia coli cells. *Proc. Natl. Acad. Sci.*, 101(31):11310–11315.

[57] Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036.

[58] Goldstein, A., Goldstein, D. B., Brown, B. J., and Chou, S.-C. (1959). Amino acid starvation in an Escherichia coli auxotroph: I. Effects on protein and nucleic acid synthesis and on cell division. *Biochim. Biophys. Acta*, 36(1):163–172.

[59] Graham, J. E., Marians, K. J., and Kowalczykowski, S. C. (2017). Independent and Stochastic Action of DNA Polymerases in the Replisome. *Cell*, 169(7):1201–1213.e17.

[60] Gregory, T. R. (2001). Coincidence, coevolution, or causation? DNAcontent, cell size, and the C-value enigma. *Biol Rev*, 76:65–101.

[61] Gudas, L. J. and Pardee, A. B. (1974). Deoxyribonucleic acid synthesis during the division cycle of Escherichia coli: a comparison of strains B/r, K-12, 15, and 15T-under conditions of slow growth. *J. Bacteriol.*, 117(3):1216–1223.

[62] Hansen, E. B., Atlung, T., Hansen, F. G., Skovgaard, O., and von Meyenburg, K. (1984). Fine structure genetic map and complementation analysis of mutations in the dnaA gene of Escherichia coli. *Mol. Gen. Genet. MGG*, 196(3):387–396.

[63] Hansen, F., Koefoed, S., Sørensen, L., and Atlung, T. (1987). Titration of DnaA protein by oriC DnaA-boxes increases dnaA gene expression in Escherichia coli. *EMBO J.*, 6(1):255–258.

[64] Hansen, F. G. and Atlung, T. (2018). The DnaA Tale. *Front. Microbiol.*, 9:319.

[65] Hansen, F. G., Atlung, T., Braun, R. E., Wright, A., Hughes, P., and Kohiyama, M. (1991). Initiator (DnaA) protein concentration as a function of growth rate in Escherichia coli and Salmonella typhimurium. *J. Bacteriol.*, 173(16):5194–5199.

[66] Hansen, F. G., Christensen, B. B., and Atlung, T. (2007). Sequence Characteristics Required for Cooperative Binding and Efficient in Vivo Titration of the Replication Initiator Protein DnaA in E. coli. *J. Mol. Biol.*, 367(4):942–952.

[67] Hansen, F. G. and Rasmussen, K. V. (1977). Regulation of the dnaA product in Escherichia coli. *MGG Mol. Gen. Genet.*, 155(2):219–225.

[68] Hecht, A., Glasgow, J., Jaschke, P. R., Bawazer, L. A., Munson, M. S., Cochran, J. R., Endy, D., and Salit, M. (2017). Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.*, 45(7):3615–3626.

[69] Helgesen, E., Fossum-Raunehaug, S., Saetre, F., Schink, K. O., and Skarstad, K. (2015). Dynamic Escherichia coli SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. *Nucleic Acids Res.*, 43(5):2730–2743.

[70] Helmstetter, C. (1996). Timing of Synthetic Activities in the Cell Cycle. In Neidhardt, F. C. and Curtiss, R., editors, *Escherichia coli Salmonella Cell. Mol. Biol.*, volume 2, pages 1627–1639. ASM Press.

[71] Helmstetter, C. E. (2015). A ten-year search for synchronous cells: obstacles, solutions, and practical applications. *Front. Microbiol.*, 6:238.

[72] Helmstetter, C. E. and Cooper, S. (1968). DNA synthesis during the division cycle of rapidly growing Escherichia coli B/r. *J. Mol. Biol.*, 31(3):507–518.

[73] Helmstetter, C. E. and Cummings, D. J. (1963). BACTERIAL SYNCHRONIZATION BY SELECTION OF CELLS AT DIVISION. *Proc. Natl. Acad. Sci.*, 50(4):767–774.

[74] Hensel, Z., Feng, H., Han, B., Hatem, C., Wang, J., and Xiao, J. (2012). Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis. *Nat. Struct. Mol. Biol.*, 19(8):797–802.

[75] Herrick, J. and Bensimon, A. (1999). Single molecule analysis of DNA replication. *Biochimie*, 81(8-9):859–871.

[76] Herrick, J., Kohiyama, M., Atlung, T., and Hansen, F. G. (1996). The initiation mess? *Mol. Microbiol.*, 19(4):659–66.

[77] HERSHEY, A. D. and CHASE, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.*, 36(1):39–56.

[78] Hill, N. S., Kadoya, R., Chattoraj, D. K., and Levin, P. A. (2012). Cell size and the initiation of DNA replication in bacteria. *PLoS Genet.*, 8(3):e1002549.

[79] Hiraga, S. (1976). Novel F prime factors able to replicate in Escherichia coli Hfr strains. *Proc. Natl. Acad. Sci.*, 73(1):198–202.

[80] Hirota, Y., Mordoh, J., and Jacob, F. (1970). On the process of cellular division in Escherichia coli: III. Thermosensitive Mutants of Escherichia coli Altered in the Process of DNA Initiation. *J. Mol. Biol.*, 53(3):369–387.

[81] Hohlfeld, R. and Vielmetter, W. (1973). Bidirectional growth of the E. coli chromosome. *Nat. New Biol.*, 242(118):130–132.

[82] Hsieh, L. S., Rouviere-Yaniv, J., and Drlica, K. (1991). Bacterial DNA supercoiling and [ATP]/[ADP] ratio: Changes associated with salt shock.

[83] Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., Hwa, T., and Williamson, J. R. (2015). Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.*, 11(1):784–e784.

[84] Ishida, T., Akimitsu, N., Kashioka, T., Hatano, M., Kubota, T., Ogata, Y., Sekimizu, K., and Katayama, T. (2004). DiaA, a Novel DnaA-binding Protein, Ensures the Timely Initiation of Escherichia coli Chromosome Replication. *J. Biol. Chem.*, 279(44):45546–45555.

[85] Jacob, F., Brenner, S., and Cuzin, F. (1963). On the Regulation of DNA Replication in Bacteria. *Cold Spring Harb. Symp. Quant. Biol.*, 28(0):329–348.

[86] Jacob, F., Fuerst, C. R., and Wollman, E. L. (1957). RECHERCHES SUR LES BACTERIES LYSOGENES DEFECTIVES. 2. LES TYPES PHYSIOLOGIQUES LIES AUX MUTATIONS DU PROPHAGE. In *Ann. Inst. Pasteur (Paris).*, volume 93, pages 724–753. MASSON EDITEUR 120 BLVD SAINT-GERMAIN, 75280 PARIS 06, FRANCE.

[87] Jay, G., Khoury, G., Seth, A. K., and Jay, E. (1981). Construction of a general vector for efficient expression of mammalian proteins in bacteria: use of a synthetic ribosome binding site. *Proc. Natl. Acad. Sci. U. S. A.*, 78(9):5543–8.

[88] Jun, S., Si, F., Pugatch, R., and Scott, M. (2017). Fundamental Principles in Bacterial Physiology - History, Recent progress, and the Future with Focus on Cell Size Control: A Review. *Reports Prog. Phys.*, 81(5):056601.

[89] Kaguni, J. M. (2011). Replication initiation at the Escherichia coli chromosomal origin. *Curr. Opin. Chem. Biol.*, 15(5):606–613.

[90] Kaguni, J. M., Fuller, R. S., and Kornberg, A. (1982). Enzymatic replication of E. coli chromosomal origin is bidirectional. *Nature*, 296(5858):623–627.

[91] Kaguni, J. M. and Kornberg, A. (1984). Replication initiated at the origin (oriC) of the E. coli chromosome reconstituted with purified enzymes. *Cell*, 38(1):183–190.

[92] Kasho, K. and Katayama, T. (2013). DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc. Natl. Acad. Sci.*, 110(3):936–941.

[93] Katayama, T. and Crooke, E. (1995). DnaA protein is sensitive to a soluble factor and is specifically inactivated for initiation of in vitro replication of the Escherichia coli minichromosome. *J. Biol. Chem.*, 270(16):9265–9271.

[94] Katayama, T., Kasho, K., and Kawakami, H. (2017). The DnaA Cycle in Escherichia coli: Activation, Function and Inactivation of the Initiator Protein. *Front. Microbiol.*, 8(December):1–15.

[95] Katayama, T., Kubota, T., Kurokawa, K., Crooke, E., and Sekimizu, K. (1998). The Initiator Function of DnaA Protein Is Negatively Regulated by the Sliding Clamp of the E. coli Chromosomal Replicase. *Cell*, 94(1):61–71.

[96] Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat. Rev. Microbiol.*, 8(3):163–170.

[97] Kato, J. I. J.-i. and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli. *EMBO J.*, 20(15):4253–4262.

[98] Kellenberger-Gujer, G., Podhajska, A. J., and Caro, L. (1978). A cold sensitive dnaA mutant of E. coli which overinitiates chromosome replication at low temperature. *MGG Mol. Gen. Genet.*, 162(1):9–16.

[99] Keyamura, K., Abe, Y., Higashi, M., Ueda, T., and Katayama, T. (2009). DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J. Biol. Chem.*, 284(37):25038–50.

[100] Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'etsugu, M., Fujimitsu, K., Kagawa, W., Yokoyama, S., Kurumizaka, H., and Katayama, T. (2007). The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.*, 21(16):2083–99.

[101] Khodursky, A. B., Peter, B. J., Schmid, M. B., DeRisi, J., Botstein, D., Brown, P. O., and Cozzarelli, N. R. (2000). Analysis of topoisomerase function in bacterial replication fork movement: Use of DNA microarrays. *Proc. Natl. Acad. Sci.*, 97(17):9419–9424.

[102] Kitagawa, R., Mitsuki, H., Okazaki, T., and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the Escherichia coll chromosome. *Mol. Microbiol.*, 19(5):1137–1147.

[103] Kitagawa, R., Ozaki, T., Moriya, S., and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein. *Genes Dev.*, 12(19):3032–3043.

[104] Kjeldgaard, N. O., Maaløe, O., and Schaechter, M. (1958). The transition between different physiological states during balanced growth of Salmonella typhimurium. *Microbiology*, 19(3):607–616.

[105] Kohiyama, M., Lanfrom, H., Brenner, S., and Jacob, F. (1963). Modifications de fonctions indispensables chez des mutants thermosensibles d'E. coli sur une mutation empechant la replication du chromosome bacterien. *C. R. Hebd. Seances Acad. Sci.*, 257:1979–81.

[106] Kücherer, C., Lother, H., Kölling, R., Schauzu, M. A., and Messer, W. (1986). Regulation of transcription of the chromosomal dnaA gene of Escherichia coli. *MGG Mol. Gen. Genet.*, 205(1):115–121.

[107] Kumar, S., Boulton, A. J., Beck-Nielsen, H., Berthezene, F., Muggeo, M., Persson, B., Spinas, G. A., Donoghue, S., Lettis, S., and Stewart-Long, P. (1996). Troglitazone, an insulin action enhancer, improves metabolic control in NIDDM patients. Troglitazone Study Group. *Diabetologia*, 39(6):701–9.

[108] Lane, H. E. and Denhardt, D. T. (1975). The rep mutation. IV. Slower movement of replication forks in Escherichia coli rep strains. *J. Mol. Biol.*, 97(1):99–112.

[109] Lark, K. G. (1969). Initiation and Control of Dna Synthesis. *Annu. Rev. Biochem.*, 38(1):569–604.

[110] Lebofsky, R. and Bensimon, A. (2003). Single DNA molecule analysis: Applications of molecular combing. *Briefings Funct. Genomics Proteomics*, 1(4):385–396.

[111] Leonard, A. C. and Grimwade, J. E. (2010). Initiation of DNA Replication. *EcoSal Plus*, 4(1):29–44.

[112] Leonard, A. C. and Grimwade, J. E. (2011). Regulation of DnaA Assembly and Activity: Taking Directions from the Genome. *Annu. Rev. Microbiol.*, 65(1):19–35.

[113] Leonard, A. C. and Grimwade, J. E. (2015). The orisome: structure and function. *Front. Microbiol.*, 6(JUN).

[114] Li, G.-W. W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635.

[115] Li, X.-t., Jun, Y., Erickstad, M. J., Brown, S. D., Parks, A., Court, D. L., and Jun, S. (2016). tCRISPRi: tunable and reversible, one-step control of gene expression. *Sci. Rep.*, 6:39076.

[116] Li, Y. and Altman, S. (2004). Polarity effects in the lactose operon of Escherichia coli. *J. Mol. Biol.*, 339(1):31–39.

[117] Lim, H. N., Lee, Y., and Hussein, R. (2011). Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci.*, 108(26):10626–10631.

[118] Løbner-Olesen, A., Skarstad, K., and Hansen, F. G. (1989). The DnaA protein determines the initiation mass of Escherichia coli K-12. *Cell*, 57(5):881–889.

[119] Louarn, J., Funderburgh, M., and Bird, R. E. (1974). More precise mapping of the replication origin in Escherichia coli K 12. *J. Bacteriol.*, 120(1):1–5.

[120] Lu, M., Campbell, J. L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in E. coli. *Cell*, 77(3):413–426.

[121] Maaløe, O. and Hanawalt, P. C. (1961). Thymine deficiency and the normal DNA replication cycle. I. *J. Mol. Biol.*, 3(2):144–155.

[122] Mallik, P., Pratt, T. S., Beach, M. B., Bradley, M. D., Undamatla, J., and Osuna, R. (2004). Growth phase-dependent regulation and stringent control of fis are conserved processes in enteric bacteria and involve a single promoter (fis P) in Escherichia coli. *J. Bacteriol.*, 186(1):122–35.

[123] Margalit, H. and Grover, N. B. (1987). Initiation of chromosome replication in bacteria: analysis of an inhibitor control model. *J. Bacteriol.*, 169(11):5231–5240.

[124] Marsh, R. C. and Worcel, A. (1977). A DNA fragment containing the origin of replication of the Escherichia coli chromosome. *Proc Natl Acad Sci U S A*, 74(7):2720–2724.

[125] Marszalek, J. and Kaguni, J. M. (1994). DnaA protein directs the binding of DnaB protein in initiation of DNA replication in Escherichia coli. *J. Biol. Chem.*, 269(7):4883–4890.

[126] Martel, M., Balleydier, A., Sauriol, A., and Drolet, M. (2015). Constitutive stable DNA replication in Escherichia coli cells lacking type 1A topoisomerase activity. *DNA Repair (Amst).*, 35:37–47.

[127] Masters, M. and Broda, P. (1971). Evidence for the bidirectional replication of the Escherichia coli chromosome. *Nat. New Biol.*, 232(31):137–140.

[128] Mazo-Vargas, A., Park, H., Aydin, M., and Buchler, N. E. (2014). Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy. *Mol. Biol. Cell*, 25(22):3699–3708.

[129] Megerle, J. A., Fritz, G., Gerland, U., Jung, K., and Rädler, J. O. (2008). Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.*, 95(4):2103–2115.

[130] Meijer, M., Beck, E., Hansen, F. G., Bergmans, H. E., Messer, W., von Meyenburg, K., and Schaller, H. (1979). Nucleotide sequence of the origin of replication of the Escherichia coli K-12 chromosome. *Proc. Natl. Acad. Sci. U. S. A.*, 76(2):580–4.

[131] Meselson, M. and Stahl, F. W. (1958). The Replication of DNA in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, 44(7):671–82.

[132] Messer, W., Blaesing, F., Majka, J., Nardmann, J., Schaper, S., Schmidt, A., Seitz, H., Speck, C., Tüngler, D., Wegrzyn, G., Weigel, C., Welzeck, M., and Zakrzewska-Czerwinska, J. (1999). Functional domains of DnaA proteins. *Biochimie*, 81(8-9):819–825.

[133] Messer, W., Meijer, M., Bergmans, H. E. N., Hansen, F. G., von Meyenburg, K., Beck, E., and Schaller, H. (1979). Origin of Replication, oriC, of the Escherichia coli K12 Chromosome: Nucleotide Sequence. *Cold Spring Harb. Symp. Quant. Biol.*, 43:139–145.

[134] Michelsen, O. (2003). Precise determinations of C and D periods by flow cytometry in Escherichia coli K-12 and B/r. *Microbiology*, 149(4):1001–1010.

[135] Milo, R. and Phillips, R. (2016). Cell Biology by the Numbers (Garland Science). In *Curr. Opin. Plant Biol.*, pages 286–289. Garland Science.

[136] Morigen, Molina, F., and Skarstad, K. (2005). Deletion of the datA site does not affect once-per-cell-cycle timing but induces rifampin-resistant replication. *J. Bacteriol.*, 187(12):3913–3920.

[137] Nagata, T. (1962). POLARITY AND SYNCBRONY INTHE REPLICATION OF DNA MOLECULES OF BACTERIA*. Technical Report 5.

[138] Nagata, T. (1963). The molecular synchrony and sequential replication of DNA in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, 49(4):551.

[139] Nishida, S., Fujimitsu, K., Sekimizu, K., Ohmura, T., Ueda, T., and Katayama, T. (2002). A nucleotide switch in the Escherichia coli DnaA protein initiates chromosomal replication. Evidence from a mutant DnaA protein defective in regulatory ATP hydrolysis in vitro and in vivo. *J. Biol. Chem.*, 277(17):14986–14995.

[140] Nozaki, S., Niki, H., and Ogawa, T. (2009). Replication initiator DnaA of Escherichia coli changes its assembly form on the replication origin during the cell cycle. *J. Bacteriol.*, 191(15):4807–4814.

[141] Odsbu, I., Morigen, and Skarstad, K. (2009). A reduction in ribonucleotide reductase activity slows down the chromosome replication fork but does not change its localization. *PLoS One*, 4(10):e7617.

[142] Ogura, Y., Imai, Y., Ogasawara, N., and Moriya, S. (2001). Autoregulation of the dnaA-dnaN Operon and Effects of DnaA Protein Levels on Replication Initiation in Bacillus subtilis. *J. Bacteriol.*, 183(13):3833–3841.

[143] Oishi, M., Yoshikawa, H., and Sueoka, N. (1964). Synchronous and dichotomous replication of the Bacillus subtilis chromosome during spore germanation. *Nature*, 204(4963):1069–1073.

[144] Oka, A., Sugimoto, K., Takanami, M., and Hirota, Y. (1980). Replication origin of the Escherichia coli K-12 chromosome: The size and structure of the minimum DNA segment carrying the information for autonomous replication. *MGG Mol. Gen. Genet.*, 178(1):9–20.

[145] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.*, 31(1):69–73.

[146] Ozbudak, E. M., Thattal, M., Lim, H. H., Shraiman, B. I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of Escherichia coli. *Nature*, 427(6976):737–740.

[147] Pedersen, I. B., Helgesen, E., Flåtten, I., Fossum-Raunehaug, S., and Skarstad, K. (2017). SeqA structures behind Escherichia coli replication forks affect replication elongation and restart mechanisms. *Nucleic Acids Res.*, 45(11):6471–6485.

[148] Pfaffl, M. W. (2012). Quantification strategies in real-time polymerase chain reaction. *Quant. real-time PCR. Appl Microbiol*, pages 53–62.

[149] Pierucci, O., Rickert, M., and Helmstetter, C. E. (1989). DnaA protein overproduction abolishes cell cycle specificity of DNA replication from oriC in Escherichia coli. *J. Bacteriol.*, 171(7):3760–6.

[150] Prentki, P., Chandler, M., and Caro, L. (1977). Replication of the prophage P1 during the cell cycle of Escherichia coli. *MGG Mol. Gen. Genet.*, 152(1):71–76.

[151] Prescott, D. M. and Kuempel, P. L. (1972). Bidirectional replication of the chromosome in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, 69(10):2842–5.

[152] Pritchard, R. H., Barth, P. T., and Collins, J. (1969). Control of DNA synthesis in bacteria. *Symp. Soc. Gen. Microbiol*, 19:263–297.

[153] Pritchard, R. H., Meacock, P. A., and Orr, E. (1978). Diameter of cells of a thermosensitive dnaA mutant of Escherichia coli cultivated at intermediate temperatures. *J. Bacteriol.*, 135(2):575–580.

[154] Pritchard, R. H. and Zaritsky, A. (1970). Effect of thymine concentration on the replication velocity of DNA in a thymineless mutant of Escherichia coli. *Nature*, 226(5241):126–131.

[155] Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183.

[156] Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*, 5(10):877–879.

[157] Rao, P., Rozgaja, T. A., Alqahtani, A., Grimwade, J. E., and Leonard, A. C. (2018). Low affinity DnaA-ATP recognition sites in E. coli oriC make non-equivalent and growth rate-dependent contributions to the regulated timing of chromosome replication. *Front. Microbiol.*, 9(JUL):1673.

[158] Regev, T., Myers, N., Zarivach, R., and Fishov, I. (2012). Association of the Chromosome Replication Initiator DnaA with the Escherichia coli Inner Membrane In Vivo: Quantity and Mode of Binding. *PLoS One*, 7(5):e36441.

[159] Reyes-Lamothe, R., Sherratt, D. J., and Leake, M. C. (2010). Stoichiometry and architecture of active DNA replication machinery in escherichia coli. *Science (80-.).*, 328(5977):498–501.

[160] Riber, L., Frimodt-Møller, J., Charbon, G., and Løbner-Olesen, A. (2016). Multiple DNA Binding Proteins Contribute to Timing of Chromosome Replication in E. coli. *Front. Mol. Biosci.*, 3(June):1–9.

[161] Riber, L., Olsson, J. A., Jensen, R. B., Skovgaard, O., Dasgupta, S., Marinus, M. G., and Løbner-Olesen, A. (2006). Hda-mediated inactivation of the DnaA protein and dnaA gene autoregulation act in concert to ensure homeostatic maintenance of the Escherichia coli chromosome. *Genes Dev.*, 20(15):2121–2134.

[162] Ringquist, S. and Smith, C. L. (1992). The Escherichia coli chromosome contains specific, unmethylated dam and dcm sites. *Proc. Natl. Acad. Sci. U. S. A.*, 89(10):4539–43.

[163] Robinson, A. and Van Oijen, A. M. (2013). Bacterial replication, transcription and translation: Mechanistic insights from single-molecule biochemical studies.

[164] Roepke, R. R. (1967). Relation Between Different Thymineless Mutants Derived from Escherichia coli. *J. BACYERIOLOGY*, 93(3):1188–1189.

[165] Rosenberg, B. H., Cavalieri, L. F., and Ungers, G. (1969). THE NEGATIVE CONTROL MECHANISM FOR E. coli DNA REPLICATION. *Proc. Natl. Acad. Sci.*, 63(4):1410–1417.

[166] Rosenfeld, N. (2005). Gene Regulation at the Single-Cell Level. *Science (80-.).*, 307(5717):1962–1965.

[167] Rosenfeld, N., Perkins, T. J., Alon, U., Elowitz, M. B., and Swain, P. S. (2006). A Fluctuation Method to Quantify In Vivo Fluorescence Data. *Biophys. J.*, 91(2):759–766.

[168] Roth, A. and Messer, W. (1998). High-affinity binding sites for the initiator protein DnaA on the chromosome of Escherichia coli. *Mol. Microbiol.*, 28(2):395–401.

[169] Saito, K. and Nagai, T. (2015). Recent progress in luminescent proteins development. *Curr. Opin. Chem. Biol.*, 27(Figure 1):46–51.

[170] Sanchez, A., Cattoni, D. I., Walter, J.-C., Rech, J., Parmeggiani, A., Nollmann, M., and Bouet, J.-Y. (2015). Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. *Cell Syst.*, 1(2):163–173.

[171] Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.*, 74(12):5463–5467.

[172] Santi, I., Dhar, N., Bousbaine, D., Wakamoto, Y., and McKinney, J. D. (2013). Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria. *Nat. Commun.*, 4(May):2470.

[173] Sawitzke, J. A., Thomason, L. C., Bubunenko, M., Li, X., Costantino, N., and Court, D. L. (2013). Recombineering: Using drug cassettes to knock out genes in vivo. *Methods Enzymol.*, 533:79–102.

[174] Schaechter, M. (2015). A brief history of bacterial growth physiology. *Front. Microbiol.*, 6:289.

[175] Schaechter, M., Maaløe, O., and Kjeldgaard, N. O. (1958). Dependency on medium and temperature of cell size and chemical composition during balanced growth of Salmonella typhimurium. *Microbiology*, 19(3):592–606.

[176] Schaefer, C. and Messer, W. (1991). DnaA protein/DNA interaction. Modulation of the recognition sequence. *MGG Mol. Gen. Genet.*, 226(1-2):34–40.

[177] Schaper, S. and Messer, W. (1995). Interaction of the initiator protein DnaA of Escherichia coli with its DNA target. *J. Biol. Chem.*, 270(29):17622–17626.

[178] Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.*, 34(1):104–110.

[179] Sekimizu, K., Bramhill, D., and Kornberg, A. (1987). ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. *Cell*, 50(2):259–265.

[180] Sekimizu, K., Yung, B. Y.-m., and Kornberg, A. (1988). The dnaA protein of Escherichia coli. Abundance, improved purification, and membrane binding. *J. Biol. Chem.*, 263(15):7136–40.

[181] Sepúlveda, L. A., Xu, H., Zhang, J., Wang, M., and Golding, I. (2016). Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science (80-.).*, 351(6278):1218–1222.

[182] Si, F., Li, D., Cox, S. E., Sauls, J. T., Azizi, O., Sou, C., Schwartz, A. B., Erickstad, M. J., Jun, Y., Li, X., and Jun, S. (2017). Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Curr. Biol.*, 27(9):1278–1287.

[183] Siegele, D. A. and Hu, J. C. (1997). Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci.*, 94(15):8168–8172.

[184] Simmons, L. A., Breier, A. M., Cozzarelli, N. R., and Kaguni, J. M. (2004). Hyperinitiation of DNA replication in Escherichia coli leads to replication fork collapse and inviability. *Mol. Microbiol.*, 51(2):349–358.

[185] Skarstad, K., Boye, E., and Steen, H. (1986). Timing of initiation of chromosome replication in individual Escherichia coli cells. *EMBO J.*, 5(7):1711–1717.

[186] Skarstad, K. and Katayama, T. (2013). Regulating DNA Replication in Bacteria. *Cold Spring Harb. Perspect. Biol.*, 5(4):a012922.

[187] Skarstad, K., Steen, H. B., and Boye, E. (1983). Cell cycle parameters of slowly growing Escherichia coli B/r studied by flow cytometry. *J. Bacteriol.*, 154(2):656–662.

[188] Skinner, S. O., Sepúlveda, L. A., Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.*, 8(6):1100–1113.

[189] Snapp, E. (2005). Design and Use of Fluorescent Fusion Proteins in Cell Biology. *Curr. Protoc. Cell Biol.*, 27(1):21.4.1–21.4.13.

[190] Snapp, E. L. (2009). Fluorescent proteins: a cell biologist's user guide. *Trends Cell Biol.*, 19(11):649–655.

[191] Sompayrac, L. and Maaløe, O. (1973). Autorepressor Model for Control of DNA Replication. *Nature*, 241(109):133–135.

[192] Sueoka, N. and Yoshikawa, H. (1965). The chromosome of Bacillus subtilis. I. Theory of marker frequency analysis. *Genetics*, 52(4):747–757.

[193] Su'etsugu, M., Nakamura, K., Keyamura, K., Kudo, Y., and Katayama, T. (2008). Hda monomerization by ADP binding promotes replicase clamp-mediated DnaA-ATP hydrolysis. *J. Biol. Chem.*, 283(52):36118–31.

[194] Sugimoto, K., Oka, A., Sugisaki, H., Takanami, M., Nishimura, A., Yasuda, Y., and Hirota, Y. (1979). Nucleotide sequence of Escherichia coli K-12 replication origin. *Proc. Natl. Acad. Sci. U. S. A.*, 76(2):575–9.

[195] Taheri-Araghi, S., Bradde, S., Sauls, J. T., Hill, N. S., Levin, P. A., Paulsson, J., Vergassola, M., and Jun, S. (2015). Cell-size control and homeostasis in bacteria. *Curr. Biol.*, 25(3):385–391.

[196] Taheri-araghi, S., Brown, S. D., Sauls, J. T., Mcintosh, D. B., and Jun, S. (2015). Single-Cell Physiology. *Annu. Rev. Biophys.*, 44(February):1–20.

[197] Taheri-Araghi, S. and Jun, S. (2016). *Single-Cell Cultivation in Microfluidic Devices*, pages 5–16. Springer Berlin Heidelberg, Berlin, Heidelberg.

[198] Takai, A., Nakano, M., Saito, K., Haruno, R., Watanabe, T. M., Ohyanagi, T., Jin, T., Okada, Y., and Nagai, T. (2015). Expanded palette of Nano-lanterns for real-time multicolor luminescence imaging. *Proc. Natl. Acad. Sci. U. S. A.*, 112(14):4352–4356.

[199] Taniguchi, Y., Choi, P. J., Li, G.-W. G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science (80-.).*, 329(5991):533–538.

[200] Taylor, A. L. (1970). Current linkage map of Escherichia coli. *Microbiol. Mol. Biol. Rev.*, 34(2):155–175.

[201] Teng, S.-W., Wang, Y., Tu, K. C., Long, T., Mehta, P., Wingreen, N. S., Bassler, B. L., and Ong, N. P. (2010). Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophys. J.*, 98(9):2024–31.

[202] Thattai, M. and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci.*, 98(15):8614–8619.

[203] Thomason, L. C., Sawitzke, J. A., Li, X., Costantino, N., and Court, D. L. (2014). Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination. In *Curr. Protoc. Mol. Biol.*, volume 106, pages 1.16.1–1.16.39. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[204] Torheim, N. K., Boye, E., Løbner-Olesen, A., Stokke, T., and Skarstad, K. (2002). The Escherichia coli SeqA protein destabilizes mutant DnaA204 protein. *Mol. Microbiol.*, 37(3):629–638.

[205] Tran, Q. H. and Unden, G. (1998). Changes in the proton potential and the cellular energetics of Escherichia coli during growth by aerobic and anaerobic respiration or by fermentation. *Eur. J. Biochem.*, 251(1-2):538–543.

[206] von Freiesleben, U. (2000). The eclipse period of Escherichia coli. *EMBO J.*, 19(22):6240–6248.

[207] von Freiesleben, U., Rasmussen, K. V., and Schaechter, M. (1994). SeqA limits DnaA activity in replication from oriC in Escherichia coli. *Mol. Microbiol.*, 14(4):763–72.

[208] Walker, N., Nghe, P., and Tans, S. J. (2016). Generation and filtering of gene expression noise by the bacterial cell cycle. *BMC Biol.*, 14(1):1.

[209] Wallden, M., Fange, D., Lundius, E. G., Baltekin, Ö., and Elf, J. (2016). The Synchronization of Replication and Division Cycles in Individual E. coli Cells. *Cell*, 166(3):729–739.

[210] Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–8.

[211] Wek, R. C., Sameshima, J. H., and Hatfield, G. W. (1987). Rho-dependent transcriptional polarity in the ilvGMEDA operon of wild-type Escherichia coli K12. *J. Biol. Chem.*, 262(31):15256–15261.

[212] Wiktor, J., Lesterlin, C., Sherratt, D. J., and Dekker, C. (2016). CRISPR-mediated control of the bacterial initiation of replication. *Nucleic Acids Res.*, 44(8):3801–3810.

[213] Xia, W. and Dowhan, W. (1995). In vivo evidence for the involvement of anionic phospholipids in initiation of DNA replication in Escherichia coli. *Proc. Natl. Acad. Sci.*, 92(3):783–787.

[214] Yasuda, S. and Hirota, Y. (1977). Cloning and mapping of the replication origin of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, 74(12):5458–62.

[215] Yoshikawa, H., O'Sullivan, A., and Sueoka, N. (1964). SEQUENTIAL REPLICATION OF THE BACILLUS SUBTILIS CHROMOSOME, III. REGULATION OF INITIATION. *Proc. Natl. Acad. Sci. U. S. A.*, 52(4):973–980.

[216] Yoshikawa, H. and Sueoka, N. (1963). Sequential Replication of Bacillus Subtilis Chromosome, I. Comparison of Marker Frequencies in Exponential and Stationary Growth Phases. *Proc. Natl. Acad. Sci.*, 49(4):559–566.

[217] Youngren, B., Nielsen, H. J., Jun, S., and Austin, S. (2014). The multifork Escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes Dev.*, 28(1):71–84.

[218] Zaritsky, A. and Pritchard, E. H. (1971). Replication time of the chromosome in thymineless mutants of Escherichia coli. *J. Mol. Biol.*, 60(1):65–74.

[219] Zaritsky, A., Woldringh, C. L., Einav, M., and Alexeeva, S. (2006). Use of thymine limitation and thymine starvation to study bacterial physiology and cytology. *J. Bacteriol.*, 188(5):1667–1679.

[220] Zhang, B., Rapolu, M., Liang, Z., Han, Z., Williams, P. G., and Su, W. W. (2015). A dual-intein autoprocessing domain that directs synchronized protein co-expression in both prokaryotes and eukaryotes. *Sci. Rep.*, 5:8541.

[221] Zyskind, J. W., Cleary, J. M., Brusilow, W. S., Harding, N. E., and Smith, D. W. (1983). Chromosomal replication origin from the marine bacterium Vibrio harveyi functions in Escherichia coli: oriC consensus sequence. *Proc. Natl. Acad. Sci.*, 80(5):1164–1168.

[222] Zyskind, J. W., Harding, N. E., Takeda, Y., Cleary, J. M., and Smith, D. W. (1981). THE DNA REPLICATION ORIGIN REGION OF THE ENTEROBACTERIACEAE. In *Initiat. Dna Replication*, pages 13–25. Elsevier.

[223] Zyskind, J. W. and Smith, D. W. (1980). Nucleotide sequence of the Salmonella typhimurium origin of DNA replication. *Proc. Natl. Acad. Sci. U. S. A.*, 77(5):2460–4.