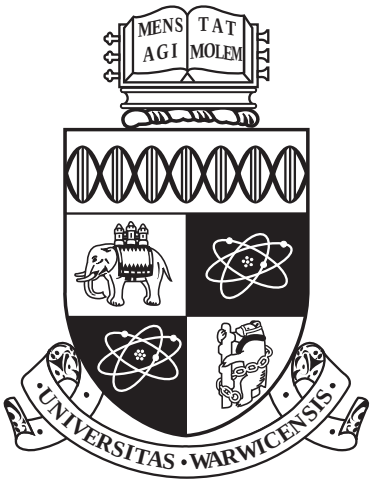


Modelling and Simulation of Heterogeneous Growth Dynamics in Bacterial Populations using a Novel Multiphasic Growth Method

by

Melchior du Lac

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

School of Engineering

September 2017

THE UNIVERSITY OF
WARWICK

Contents

List of Figures

Acknowledgments

I would like to thank my advisor Professor Declan Bates for the support, giving me the opportunity to fulfil my PhD and trusting my humble expertise on this complex subject. I would like to also thank Joshua Leonard for giving me the opportunity to visit his lab and cooperate with him in the fulfilment of the work. His guidance and problem solving of all the hurdles of our project is second to none. I would like to thank Andrew Scarpelli for generating the data and showing me around Chicago during my stay there. I would like to thank Professor Orkun Soyer for also giving me the opportunity to fulfil my PhD and giving me access to his lab to perform some experiments. I have found in his lab and all his members a great source of support and help for laboratory work and great number of conversations that they all seemed to never tire from. I would like to thank my colleagues from the Engineering department for their friendship throughout all these years. I would like to thank my grandfather for taking the time to explain complex ideas of science to a wide eyed kid, that eventually inspired me to pursue the sciences. My mother for guiding me, and giving me the tools to pursue the field of biology. My father who developed my critical thinking and through his support permitted me to be where I am today. I would like to thank uncle Ben and aunt Mary as well as my grandmother for their support throughout all these trying years. Finally I would like to thank my glip glops, William Rostain and Corentin Raffaillac for their friendship throughout this thesis.

Declarations

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted to any other for another degree.

Experimental work form Section 4.2 was performed by Dr. Andrew Scarpelli under the supervision of Dr. Joshua Leonard from the University of Northwestern University. The analysis and parsing of the data was done by myself and published in the journal "ACS Synthetic Biology" under the title "Predicting the Dynamics and Heterogeneity of Genomic DNA Content within Bacterial Populations across Variable Growth Regimes" (du Lac et al. [2016]).

The code was written by myself, with packages clearly specified in the Materials and Method, and in the code itself. It may be found on the CD-ROM along with the compilation instructions.

Abstract

The cell cycle is an inevitable source of population heterogeneity, that creates predictable discontinuities. By summarising the canonical understanding of the major steps within the bacterial cell cycle into a mechanistic model, the Cooper-Helmstetter model is able to formally describe a number of population properties such as age, DNA and volume distributions. Although this model successfully describes many different attributes of a bacterial population, it is limited to exponential growth conditions. Outside of rigorous growth environments, bacterial populations contain innate temporal features that make them difficult to formalise theoretically using traditional mechanistic or equation based mathematical models. To model bacterial population cell cycle outside of exponential growth, the single cell cycle mechanistic model was inspected and expanded. A new individual based model was developed and a novel method to track the growth of a population using measured optical density data alone was developed. Together these new features made for the Heterogeneous Multiphasic Growth simulator, and were used to explore the chromosomal DNA dynamics of bacterial populations in disparate growth regimes. The effects of the recA1 mutation on the dynamics of the cell cycle was examined through optimisation to measured data. Furthermore, predictive modelling of theoretical effects of gene copy number and partition noise on synthetic genetic constructs expressed as ordinary differential equations were explored theoretically. By explicitly simulating each member of a population using such a method, a wide range of different aspects of bacterial population may be approached theoretically with more ease, and throughout more diverse growth dynamics.

Chapter 1

Introduction

1.1 Overview

Single celled populations are, by their very nature, heterogeneous systems. In an effort to control for such a feature, it is standard practice to be experimentally rigorous instead of embracing it as an inevitable feature of the system. For example, the broad experimental steps of quantifying bacterial genetic or protein expressions, involves sampling after long-term incubation in a constant rich media, such that each cell in the population grows at their metabolic maxima in the given growth conditions and assures that each member of the population experiences the same growth environment (Cooper [1969]). Furthermore, it is seen as valid to compare population quantifications if the two different populations are grown in the same environment and at the same growth rate, since it is assumed that the same heterogeneity is experienced by the two populations. Nevertheless, such a methodology renders the output of a measurement the reflection of a statistical property of the predominant molecular state of the population that cannot be used as a reliable estimation of discrete intracellular concentrations (Lidstrom and Konopka [2010], Dubnau and Losick [2006]). Although the origins of heterogeneity are diverse and predominantly still to be appreciated, under these stringent experimental conditions and thanks to decades of experimental work, some aspects of their heterogeneity may be predicted when they depend on known biological mechanisms (Cooper [1969], Lidstrom and Konopka [2010]). For instance, bacteria in a culture having similar growth rates do not have coordinated cell cycles (Skarstad et al. [1985], Keasling et al. [1995]) and many genes have differential expression based on the growth cycle state of the cell (Sobetzko et al. [2012]). Understanding sources and the dynamics of this heterogeneity helps to mathematically formalise such systems that are in turn used to

predict them (Portle [2009], Stokke et al. [2012]).

The field of Engineering has adopted, with very high success, abstraction of complex dynamical systems into simplified models, usually described as a series of ordinary differential equations (ODE) (Mellodge and Kachroo [2008], Gershenfeld [1999]). This has proven to be a very useful technique at obtaining insights into the behaviour of a system without the need to physically build it, and has led to improved design. The field of biology, since the 1990's and the growth of "omics" technologies, has experienced an aggressive increase in the amount and precision of data available from a range of different biological systems. Thanks to this, the same type of mathematical formalisation has been attempted on biological systems, mainly through the field of systems biology, in the hopes of achieving the same degree of predictive power (Alberghina and Westerhoff [2007]). Although the field has accomplished a great deal, it has overall fallen short of its promises, due to biological systems being greatly more complex and sophisticated than human engineered devices (Lazebnik et al. [2003]). Where physical systems lend themselves relatively well to this type of abstractions, biological systems have proven to be more difficult (Lazebnik et al. [2003]). In this writer's perspective, this emanates from the very nature of biological systems. Where physical engineered devices are created with subunits that perform a particular and well-defined function whose input and output may be robustly tested, and can be connected to one another in a controlled manner; it is nearly impossible to achieve the same level of confinement of biological subunits. The efficiency of a particular protein can be nonlinearly and asynchronously influenced by relatively small changes in the system, that causes a lot of uncertainty in the validity of measurements and in turn causes imprecisions in different ranges of conditions of simulations (Moser et al. [2012]). Whereby modelling promised to help explain many aspects of biological systems that are difficult to study, biologists find themselves to be increasing needed to measure with higher precision such systems to satisfy the modellers needs.

Synthetic biology on the other hand tries to leverage knowledge of biological systems by their forward design. *In silico* designs are implemented *in vitro* by inserting synthetic genetic materials or genetically modifying a cell to perform a novel function using well defined wet lab biological practices (Andrianantoandro et al. [2006]). Just as electrical devices have well defined datasheets from manufacturers that contain schematic diagrams, connection diagrams and minima and maximum requirements and characteristics; catalogues such as the bioBrick project attempt to quantify genes and genetic modules in terms of their function and performance with the same approach (Shetty et al. [2008]). Using an engineering approach to describe

biological systems, there have been many successes such as the repressilator (Elowitz and Leibler [2000]), the toggle-switch (Gardner et al. [2000]) to name a few. However, optimising a desired function via an engineered metabolic pathway requires careful and labour-intensive optimization of the degree to which various genes are expressed (Moser et al. [2012]). Even well-defined processes; such as transgene copy number, genomic integration site, promoter strength, translational efficiency, and culture conditions, that generate proteins of interest, all impact and are impacted by the state of the cell in manners that is difficult to predict and typically requires high throughput screening or evolutionary selection for such properties. Indeed, because the plasmid vectors contain the blueprint of the system requires a scaffold to be expressed (i.e. the cell), and its expression is influenced by the state of the cell, it is equally important to formally describe the properties of the cell as well as the integrated designed system. However, the cell has received little attention due to the belief that similar stringent growth conditions across different experiments is enough to make the population experience "steady state", and thus can be easily mathematically described (Abner et al. [2014]).

In this research, single cell and population "steady state" is considered to be an oversimplification that clouds the predictive capabilities of the field. A new growth strategy is proposed that does not assume any steady state properties of the population and thus allows for its simulation in a wider range of growth conditions. This is combined with a mechanistic model of the bacterial cell cycle, simulated in parallel and heterogeneously to reflect the dynamics of measured bacterial population. The result is a robust framework for the simulation of a bacterial populations in a wide range of growth conditions, where any native or foreign genetic system may be explored *in silico* in situations that reflect with more accuracy the heterogeneous conditions in which they are expressed in.

1.2 Biological Background

1.2.1 The Central Dogma of the Cell Division Cycle

Escherichia coli (*E. coli*) has long been the focus of a large swathe of scientific study due to its very large prevalence in nature and its ease of culture in laboratory conditions (Blount [2015]). This motile, fast-growing, unicellular, gram-negative, facultative anaerobe is particularly tenacious and able to grow on a multitude of carbon sources and growth conditions (ex: aerobic and anaerobic). Today, it is the preferred organism for genetic manipulation due to its ability to host conjugative plasmids as well as non-conjugative plasmids. Methods for genetic manipulation,

chromosomal and extra-chromosomal, are well established and relatively easy to use (Shetty et al. [2008]). Furthermore, its simple circular chromosome of 4.6 million base pairs has been sequenced and extensively studied including many knock-out mutants that facilitate its use as the organism for genetic manipulation (Blount [2015]).

The central dogma of the cell division cycle (in allusion to Francis Crick's molecular biology dogma) fundamentally defined how one thinks of the links between chromosome replication and cellular division (Haeusser and Levin [2008]). It breaks down the events of the cell cycle into three successive, temporally defined, physiological phases referred to as the BCD (See Figure 1.1). The C phase represents the minimal time required to replicate a single chromosome, experimentally measured to be $\approx$ 40 min (Michelsen et al. [2003], Keasling et al. [1995]). Upon termination of replication, another phase starts called the D phase, experimentally measured to last a minimal time of $\approx$ 20 min. After that time has passed, the cell divides into two daughter cells. If the doubling time of the cell is larger than that of the C and D phases combined, then the B phase emerges as the difference in time between the doubling time of the cell and the sum of the C and D phases ($B = \tau - C + D$, where τ is the doubling time of the cell). Separating the cell cycle events in this manner laid the groundwork in defining the control mechanisms of the cell cycle, and formalised the timing of initiation and replication that permits the cell to have overlapping rounds and thus have a doubling time smaller than $C + D$ (Zaritsky [2015]). Unlike eukaryotic cells, that have very stringent phases controlled by checkpoints and that need a complete replication event to begin mitosis, bacteria are able to divide faster than the minimal recorded time to divide a chromosome. To achieve this paradoxical condition, the cell undergoes multiple rounds of replications, where it either inherits an already replicating chromosome or performs multiple replication initiations within a single cell cycl. Subsequently, this enables the cell to initiate the division process faster than if a full replication event was required (Browning et al. [2004], Keasling et al. [1995], Abner et al. [2014]).

The development of the BCD model can be traced back to the 60's "Copenhagen School" movement of study of bacterial physiology, with the motto "Look, do not touch", that investigated the physiological properties of bacteria under diverse growth conditions and growth rates (Bremer et al. [1996], Fishov et al. [1995], Neidhardt [1999]). During that time, examination of populations state by way of stringent exponential growth conditions, uncovered many fundamentals of the physiology and the connections between chromosome replication, cell growth and viability of the bacterial cell (Zaritsky [2015]). Indeed, development of the "baby machine" (also

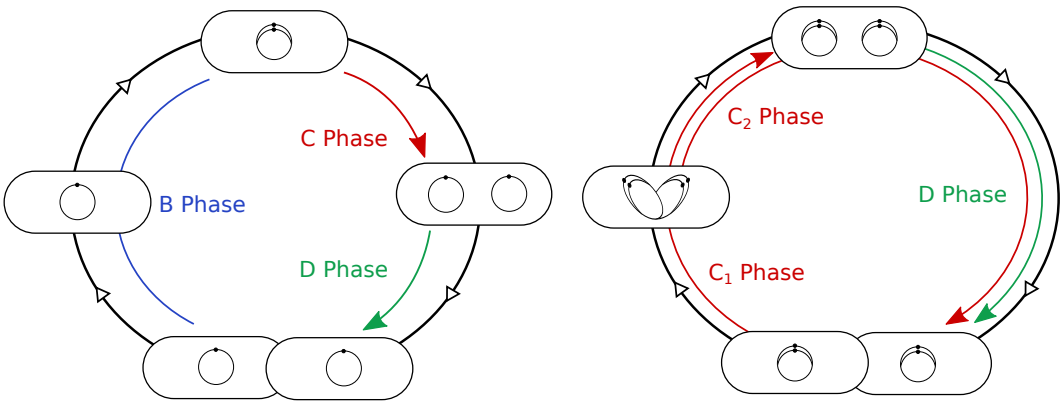


Figure 1.1: Illustration of the BCD model of the *E. coli* cell cycle. The left-hand side figure shows the cell cycle for slow growing bacteria. The different phases of growth are spatially well defined. The-right hand side shows the cell cycle with overlapping rounds of replication, where the cell inherits a partially replicated chromosome that terminates soon after division and starts the division process.

called the membrane-elution technique), enabled the generation of cell cycle coordinated populations. Through such a device, the rate of replication was observed not to be a direct linear consequence of the rate of growth (Helmstetter and Cummings [1964], Helmstetter and Cooper [1968]). Furthermore, through the use of thymine mutants grown in thymine limited media, the division rate was observed not to be a direct consequence of growth rate and instead proved to have strong links with the chromosome replication pattern (Zaritsky and Pritchard [1973]). These observations articulated the BCD model bacterial cell cycle with chromosome replication patterns as the main mode of control of the bacterial cell cycle.

Replication Initiation

By studying the number of replication events compared to mass changes of bacteria, Donachie [1968] derived that the mass at initiation was a function of the number of origins of replications, where each could be assigned a constant mass regardless of the growth rate (also called critical mass). This observation led to the theory of a growth-independent positive-acting factor initiating chromosome replication, whereby accumulation to a critical concentration of a factor enabled the activation of replication of each chromosome simultaneously (Donachie [1968]). Although, as will be seen, the true mechanism of action is more complicated, the elegance of this formulation can be attested by the vast amount of literature and the still highly predictive models of the bacterial cell cycle that have been derived from this constant (Stokke et al. [2012], Abner et al. [2014], Keasling et al. [1995], Michelsen

et al. [2003]).

With the advent of molecular biology and knock-out experiments, a clearer picture of the molecular actors for the replication events has been uncovered (see Figure 1.2). All bacteria contain a highly conserved initiator protein, DnaA, that interacts with a 9-bp recognition sequence commonly called the DnaA box (Kaguni [2006], Messer [2002]). This protein becomes activated by ATP hydrolysis and thus has two forms: activated, referred to as DnaA-ATP, and inactivated form called DnaA-ADP (Messer [2002]). The *E. coli* chromosome contains a single origin of replication (*oriC*) that includes five successive DnaA boxes whereby, upon binding of roughly 10-20 DnaA-ATP, unwinds the downstream AT rich region of the chromosome (Boye et al. [2000], Messer [2002], Schumann [2006]). Although the concentration of DnaA is constant in a cell, its activated form peaks at initiation and drops drastically thereafter by a number of mechanisms that will later be described (see Section 1.2.1). This initial complex imports the rest of the replication actors to form the replication bubbles, including a replicative DNA helicase (DnaB), a helicase-loader (DnaC) for further unwinding, a DNA primase (DnaG) and DNA polymerase III complex (Messer [2002], Schumann [2006]). The two replication bubbles then continue bi-directionally along the circular chromosome until they meet at the termination sequence on the geographical opposite side of *oriC* (see Section 1.2.1).

Thus, the critical mass theory posed by Donachie [1968], that predicted the existence of a positive actor on initiation, seems to have a molecular candidate in DnaA. However, the control of chromosome replication is not solemnly controlled by a positive actor, where many aspects of the initiation mechanism are negatively controlled (Likhoshvai and Khlebodarova [2014]).

Eclipse Period

In the 70's, an interesting observation was made using a thymine-limited strain of *E. coli*, a nucleobase required for the DNA strand. By limiting the concentration of thymine in the media, Zaritsky [1970] is able to extend the replication rate without affecting the growth rate of the cell. The author noticed that after a significant time following replication initiation, although the conditions for critical mass is achieved, the chromosome does not initiate a new replication event. A predictable amount of time was required before a new replication event could occur. This minimal time between successive replication events was called the eclipse period and revealed the existence of at least a single negative regulation for initiation. From these findinfs, further knock-out experiments and molecular techniques have identified three nega-

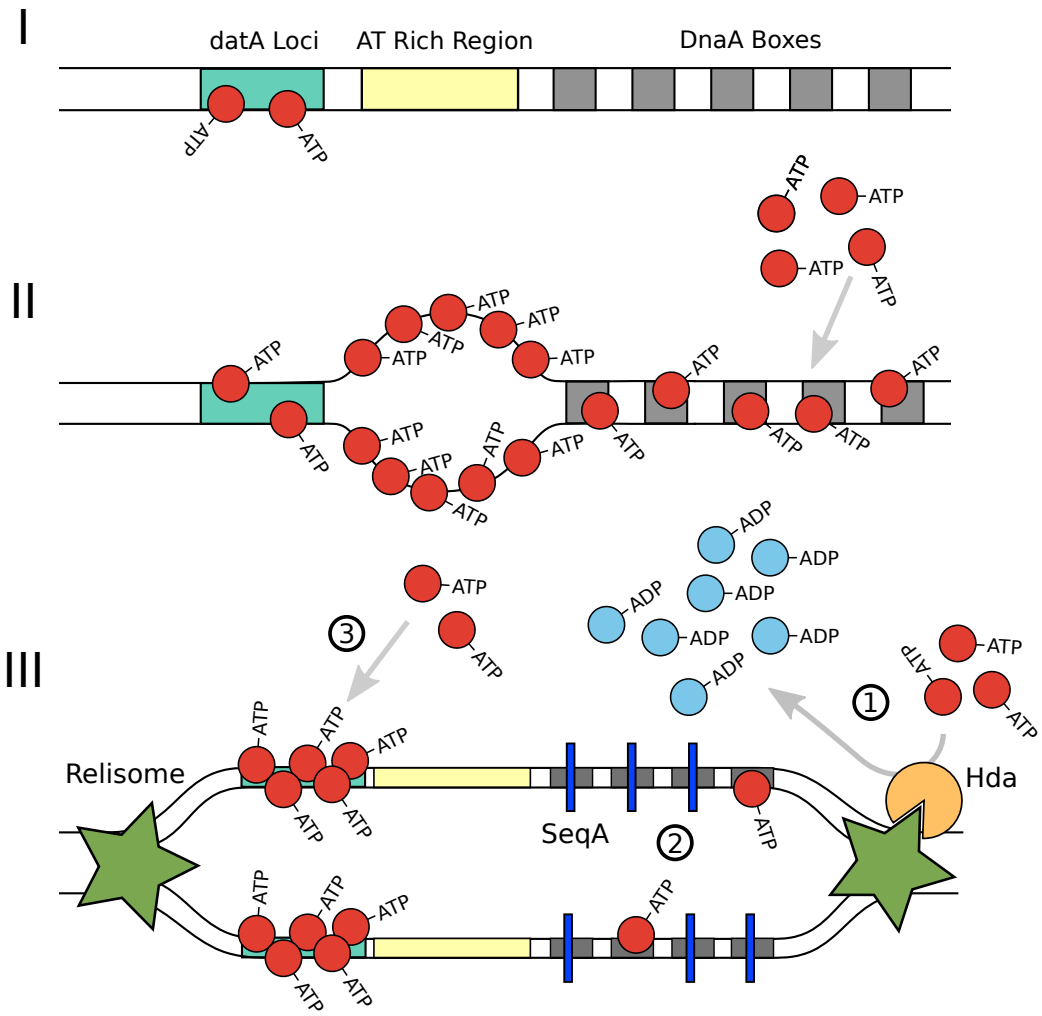


Figure 1.2: Simplified illustration of the main steps in the initiation of replication and the actors in the eclipse period. Sub-figure I shows the *oriC*, composed of five DnaA boxes and an AT rich region downstream and further down of a datA loci. Upon the growing pool of activated DnaA (DnaA-ATP) (sub-figure II), an increasing number of them bind to the DnaA boxes causing the unwinding of the At rich regions. Upon unwinding the single strand DNA is coated with DnaA and imports the replisome machinery that include helicases for further unwinding of the DNA and polymerases for synthesising a new strand. Sub-figure III shows the identified mechanisms that cause the eclipse period: 1) RIDA, where the protein Hda upon binding of DNA polymerase causes the inactivation of DnaA into DnaA-ADP (Kaguni [2006]). 2) Competitive binding of the protein SeqA to the DnaA boxes with DnaA (Schumann [2006]). 3) Sequestration of DnaA, where upon synthesising of a new strand past the DatA loci, a higher number of DnaA are bound to these loci and sequestered away from the *oriC* DnaA boxes (Bogan and Helmstetter [1997]).

tive regulation mechanisms: sequestration, regulatory inactivation of DnaA (RIDA) and titration. All prevent premature re-initiation, where mutations in many of the actors of these regulators cause the over-initiation of replication (Schumann [2006], Cooper [2012]).

The first, sequestration of newly replicated *oriC*, came from the surprising observation that the origin of replication contains a high number of GATC sequences, targets of Deoxyadenosine (*dam*) methyltransferase (Bogan and Helmstetter [1997]). These loci before and upon initiation of replication are methylated, but as the replication progresses the addition of new unmethylated nucleotides to the chromosome generates a hemimethylated site at the *oriC*. Although *in vitro* these sites are methylated by *dam* methyltransferase within a minute, *in vivo* these remain hemimethylated for up to one third of the replication time (Wilkinson [2009], Boye et al. [2000]). It was later uncovered that the protein SeqA, that has a high affinity to hemimethylated sites, binds specifically to the *oriC* in that state, and sequesters the *oriC* from further initiation (Condition 2 in Figure 1.2). Although for *E. coli* K-12 *dam* methyltransferase is not required for viability, it has still been measured to be an important is central to the initiation of replication. Indeed, *oriC* isolated *in vitro* from *dam*- cells have a greatly reduced initiation effeciency, which may be rescued with the addition of purified WT *dam* methyltransferase proteins (Lobner-Olesen et al. [2005], Smith et al. [1985]).

Next, RIDA is the main mechanism by which the active form of DnaA is inactivated (Condition 1 in Figure 1.2) (Donachie and Blakely [2003]). The formation of a complex composed of the protein Hda (named as homologous to DnaA) and DNA Polymerase III hydrolyses DnaA-ATP into DnaA-ADP and consequently, this negative feedback system reduces the concentration of active DnaA in the cell after the successful opening of *oriC* (Schumann [2006], Cooper [2012]). The last identified mechanism is the titration of DnaA (Condition 3 in Figure 1.2). Bioinformatics analysis of the bacterial chromosome revealed the existence of the datA locus containing a high number of DnaA boxes (Robert [2015]). Although not essential for viability, it is theorised that upon initiation the replication of the chromosome once it progresses past the loci, causes the formation of four datA loci that titrate excess DnaA proteins and restrains the possibility of another initiation event due to transient localised high concentration of DnaA (Hansen et al. [1991]). Indeed, adding extra copies of this loci causes the delay of initiation (Løbner-Olesen and Skarstad [2003]).

Even if these control elements can be summarised by the existence of an eclipse period, the consequence on initiation timing is not settled. von Freiesleben

et al. [2000] report 25 to 30 min while Browning et al. [2004] report that the replication forks must progress past $\approx 33\%$ of the chromosome before a new replication event is permitted. Furthermore, many report that these three negative control elements ensure that initiation occurs once and only once per cell cycle (Robert et al. [2014], Messer [2002], Olsson [2003], Boye et al. [2000]). Although this rule is true for cells with a doubling rate of ≥ 60 min, defining it in this manner is misleading since under fast growth, multiple initiation events have been observed with the formation of multifork replication events (Trojanowski et al. [2017], Youngren et al. [2014], Nielsen et al. [2007]), an event that is particularly visible upon nutritional shift-up (Kepes et al. [1987], Wallden et al. [2016], Ho and Amir [2015]). More research must be done to uncover the role of the three aformentioned negative controls have on the pattern of chromosomal initiation.

Chromosome Elongation and DNA damage

After a successful initiation event the replication of the chromosome proceeds bidirectionally withg a lagging and a leading strand (Cooper [2012], Streips et al. [2002]). As previously mentioned, for fast growing cells ($\tau \leq 60$ min), the replication time has been measured to last ≈ 40 min and progresses at a constant rate. Considering the size of the genome (4.6 mbp), this translates to ≈ 1000 nucleotides per seconds (Alberts [2017]). During that replicative process, it is commonplace for the cell to experience replication fork pause, stall or even collapses, during which the prospect of double stranded breaks increases substantially (Schumann [2006]). These arise not only through replicative errors, mispairing of bases, abnormal DNA structures (such as cross-linked DNA) or from ongoing RNA transcription, but also from outside sources such chemical or physical agents (Ultraviolet (UV) radiation) that cause DNA damage before or during the replicative process (Streips et al. [2002], Courcelle and Hanawalt [2003], Schumann [2006]).

Because it is paramount for all organisms to faithfully copy their genetic material, bacterial cells have a number of mechanisms such as nucleotide-excision repair, base-excision or mismatch excision that ensure that small errors are swiftly repaired (Courcelle and Hanawalt [2003]). However, if the damage becomes too extensive, the SOS repair mechanism is activated and blocks the cell cycle to repair such deleterious incidents. The basic mechanism is the following: the LexA protein negatively regulates transcription of at least 32 genes identified to have a range of actions including cell cycle arrest and DNA repair (Goodman [2000]). Upon severe DNA damage that leads to the exposure of large portions of single stranded (ssDNA) either by exonucelolytic digestion, double stranded breaks converted to

ssDNA by exonuclease V or replication of lesion containing templates, the protein RecA is upregulated (Cox et al. [2000], Schumann [2006], Michel [2000]). The RecA protein then binds to ssDNA with high affinity, whereby upon binding it activates the autocatalytic cleavage of LexA while simultaneously maintaining the structural integrity of the chromosome. Depressing the concentration of LexA causes the up-regulation of the genes that it repressed (among them RecA) who repair the DNA lesions (Cooper [2012], Schumann [2006]). As the damaged strand gets repaired, the number of bound RecA decreases and subsequently the concentration of LexA increases, that in turn switches off its regulon genes. For a schematic representation of the SOS repair mechanism see Figure 1.3.

The C time, as described by the BCD model, encompasses all the replication events that have been described here, including the extension time of the replicative strand and the hurdles due to DNA damage and replication errors that must be dealt with for the successful production of a sister chromosome.

Termination

To conclude a successful replication event, termination of replication and separation of the two chromosomes must occur. Due to the bi-directionality of replication, one may speculate it simply terminates when the two bubbles randomly meet. However, there are stringent control mechanisms that prevent the replication forks from not only "crashing" into each other, but obligate the two to meet precisely at a location 180° from the *oriC* on the chromosome, guided by termination regions (Ter) (Duggin and Wilce [2005], Schumann [2006]). Ter sites, contrary to their name, are not sequences that dissemble the replication complex, but instead act as unidirectional barriers of replication. Indeed, these highly conserved 22 base pair sequences interact specifically with a single Tus protein, a polar contra-helicase, that stops separation of DNA strands in a polar manner by inhibiting the unwinding activity of DnaB helicase (Duggin and Wilce [2005], Schumann [2006]).

Although most representations of the bacterial chromosome describe a single termination point (*terC*), there are in fact ten different Ter regions along the chromosome, organised and orientated in such a fashion that five Ter sites block the clockwise fork movement and the other five block the anti-clockwise fork movement (Duggin and Wilce [2005], Schumann [2006]). *terC* is commonly referred to as the termination site because it is the first that is met by either replication bubble. Moreover, the reason for such a high number of Ter sites is still unknown. *Bacillus subtilis* (*B. subtilis*) contains only two while having the same replication mechanism (Schumann [2006]). Some argue that the others act as fail-safe in case the replication

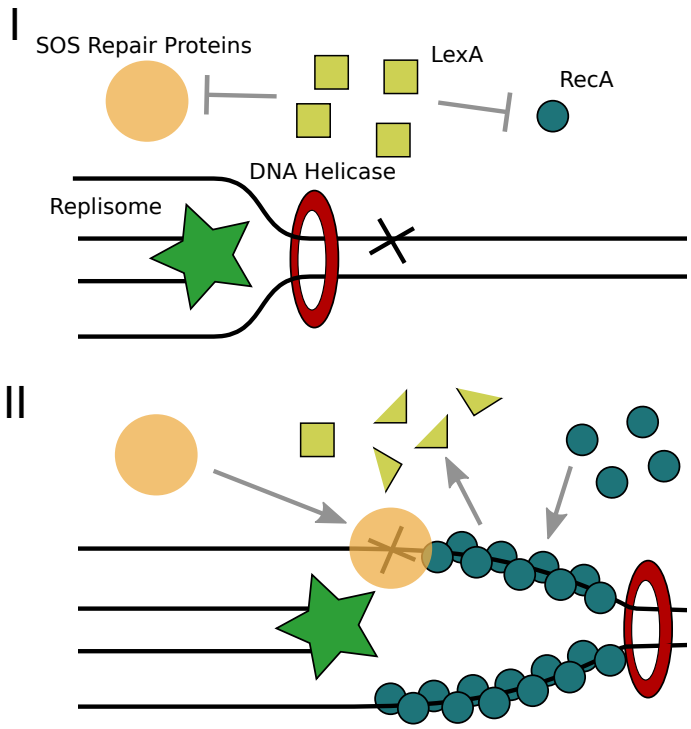


Figure 1.3: Simplified illustration of the main steps in the SOS repair mechanism. I) Before activation, LexA represses the expression of the RecA protein, and the family of SOS repair proteins. The replisome, upon reaching an error in the DNA strand, cannot progress. II) The exposed single stranded DNA strands are coated with RecA, whereby cause the degradation of the LexA protein. This causes a positive feedback on the expression of RecA while expressing the SOS repair protein family that fixes the error. Upon repair the replisome may progress and the system returns to WT.

fork does go past any *Ter* sites, while others argue that they act more like "speed bumps", that slow the replication fork progression (Duggin and Wilce [2005], Schumann [2006]). Although, knock-out experiments demonstrate that even without *Ter* sites or Tus proteins viability of the cell is not affected, some experimental evidence links the *Ter* sites to the correct partitioning of chromosome at division (Duggin and Wilce [2005]).

Chromosome Partitioning

The mechanisms of chromosome partition and division are much less known than that of chromosome replication. To date, there are three different major theories including the train-on-track model (where the chromosome move freely inside the cell), the replication factory model and the replicon model that propose different ways bacterial cells have such a stringent repartition of chromosomes between daughter cells (Schumann [2006], Haeusser and Levin [2008]). What is known from fluorescent in situ hybridization (FISH) and fluorescent tagging microscopy generally, is that *terC* and the *oriC* have distinct patterns of localisation along the cell (Nielsen et al. [2006], Haeusser and Levin [2008]). Chromosome initiation has been observed to be geographically located at the mid cell (or 1/4 and 3/4 for overlapping rounds of replication with two *oriC* to be initiated). This led to the development of the replication factory model, that states that all replication activities remain at the mid cell while the sister strands migrate to the poles of the cell (Nielsen et al. [2006]). Fluorescent tagging of *oriC* shows that the two subsequently separate to each cell pole upon termination of replication at around the same speed as the cell elongates, and remain there for the rest of the cell cycle (a process known as nucleoid exclusion) (Wu and Errington [2012]). Due to the similarity in segregational speed and cell elongation some have proposed, through the the replicon model, that the chromosome attaches to the membrane and progresses to each pole in a passive manner due to cell growth (Lemon and Grossman [2001]). The *Ter* regions have been observed to generally be located at the midcell during the whole process, right until the onset of septation, where the two chromosomes migrate to the centre of each daughter cell. The importance of the *Ter* sites for the correct partition of the chromosome, may explain the existence of such high copies (Haeusser and Levin [2008], Wang et al. [2005]).

Cell Septation

The final step in the cell cycle is the separation of the mother cell into the daughter cells by the division of the cell membrane (cytokinesis). The process begins with the formation of a septal ring at the midcell by the polymerization of the tubulin-like protein FtsZ into the Z ring (Weiss [2004], Haeusser and Levin [2008], Adams and Errington [2009]). The correct positioning of the Z ring is guided by the MinC, MinD and MinE proteins, who repress the formation of the septal ring on either of the cell poles, leaving only the centre of the cell available. The mechanism of action is one of oscillation, where an activated form of the protein (MinD-ATP) forms a complex with MinC and assembles at a single cell pole that extends toward the midcell. Once a MinC-minD complex is formed the MinE protein breaks the complex by stimulating MinD ATPases activity (forming MinD-ADP). The low concentration of MinE on the other cell pole causes the formation of the MinD-ATP and MinC complex as the other cell pole is dissembling causing the oscillatory behaviour. This is repeated many times during a cell cycle (with a single oscillation taking roughly 40 seconds) and keeps the inhibitory effect on FtsZ away from the mid cell (Weiss [2004], Lutkenhaus [2007], Lutkenhaus and Addinall [1997]).

The timing and mechanism of cell constriction contains many unanswered questions, where the Ftsz protein alone has been shown not to have sufficient contractile force to separate the daughter cells (Ghosh and Sain [2008]). Some argue that the Z ring provides a scaffold for other proteins that actually do the contractile work (Weiss [2004], Adams and Errington [2009]), while others argue that the FtsZ is the main actor but requires the assistance of other proteins (Errington et al. [2003], Adams and Errington [2009]). Nevertheless, a constriction force causes the "pinching" of the cell membrane to form two daughter cells.

All these events in the chromosome partitioning and cell septation are summarised by the D time in the BCD model. The formulation of a single rate for all these cell cycle events has its origin in the lack of experimental evidence of the molecular mechanisms of cell spetation and chromosome segregation at the time of the model formulation. Indeed, upon discovery of a predictable time of replication and the existence of overlapping rounds of replications, a similar predictable time was inferred until division when compared with the doubling time (Cooper [1969]). Numerous studies of the timing of FtsZ ring formation in the cell cycle all agree that the termination of replication is not the trigger for the formation of the Z ring, which happens before that, but do not exclude another part of the replication process as the trigger (Den Blaauwen et al. [1999], Tsukanov et al. [2011]). More experimental work is required to understand in more depth the control elements of

the bacterial septation and its link to the chromosome dynamics is required to be more conclusive.

1.2.2 Population Heterogeneity and Biological Noise

Intrinsic and Extrinsic Noise

Noise in biology may be categorised as either intrinsic or extrinsic in relation to the system of interest (Hilfinger and Paulsson [2011], Elowitz et al. [2002], Lidstrom and Konopka [2010]). The former relates to the inherent stochasticity of biochemical reactions that may have large downstream noise propagation effects on the dynamics of the system as a whole. The latter encompasses all other influences, that may include physical properties of the environment such as acidity and temperature, as well as the biological properties such as concentrations of proteins or of required compounds from upstream biochemical networks that the system depends on. Defining what constitutes intrinsic and extrinsic noise is very much dependent on the system of interest (Lidstrom and Konopka [2010]). For example, consider the reaction between a promoter sequence and its transcription machinery alone. One may define the intrinsic noise as the stochasticity and efficiency in binding of the promoter to the different members of the transcription machinery, while the concentration of the transcription machinery and the promoter sequence may be sources of extrinsic noises.

Perhaps the best example of the difference between the two types of noise originates from the elegant experimental work by Elowitz et al. [2002]. The authors incorporated two different fluorescent proteins, (cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP)), both with the same promoters on the *E. coli* chromosome, on opposite sides and at equidistance from the *oriC*. This setup yields three potential outcomes. At any given time either the cell expresses the two fluorescent proteins at equal proportion or the cell express one of the two preferably. The former would translate to the main source of noise originating from an extrinsic source since it is equally experienced by both genes, while the latter would translate to an intrinsic noise source since each has different patterns of stochasticity. Using inducible promoters with Isopropyl β -D-1-thiogalactopyranoside (IPTG) or strong constitutive promoters, the cells expressed an equal portion of the two fluorescent proteins. At the scope of the single cell, the fluctuations in the IPTG concentrations from an external source or of the replication machinery is equally experienced by both promoters. On the other hand, using a weak constitutive promoter made a cell express a majority of either CFP or YFP, evidence of the higher levels of intrinsic

noise, probably due to transient and stochastic effects of expression of the genes.

Balanced Growth and Population to Cell Quantification

The importance of one source of noise over another is very much context and system specific. Indeed, as the Elowitz et al. [2002] example shows, if one measures the fluorescent intensities at the population level, then regardless of the construct or experimental conditions, the measurements would return the same proportion of CFP and YFP. Thus, at the population level, all the conditions described by the authors would be sources of intrinsic noises. Because population averages are still commonly used as a means of studying synthetic genetic constructs (SGC) and cell expressions in general, it may prove to be problematic when trying to draw conclusions of their dynamics at the single cell level. As stated by Taheri-Araghi et al. [2015], "population-averaged data and correlations are inconclusive as the averaging process masks causal effects at the cellular level". Today a large amount of single cell quantification techniques are available including flow cytometry, single cell microscopy and microfluidic devices, yet a large portion of the field of synthetic biology still relies on population averages. For example, consider the use of fluorescence over OD, a method that is commonly used as a characterisation technique of synthetic genetic circuits (SGC) (Heinze [2012]). This experimental procedure ensures that one can compare the fluorescent signal of two independent time points or experiments, by taking into account the increase in fluorescence differences due to cell concentration changes (Carbonell and François [2015]). From this type of experiment no conclusions of the behaviour of the SGC at the single cell level can be made, since the expression of one cell over another cannot be distinguished. With such data, one cannot ascertain if for example the circuits behave in a bi-stable manner or if indeed each cell expresses average values (Portle [2009]).

Furthermore, cell heterogeneity within a population is an inevitable source of noise. Again, in the Elowitz et al. [2002] work, the authors noticed that RecA knock-out with their two IPTG inducible constructs caused the transient expression of one fluorescence protein preferably. As seen in Section 1.2.1, RecA is paramount to the SOS repair mechanism, where such a mutation probably causes chromosome heterogeneity by replication errors and DNA damage, such that it has an effect on the expression of the two constructs on the chromosome. Another example is the work by Tan et al. [2009], that observed non-intended bistable behaviour of their SGC due to its metabolic load combined with innate cell metabolic heterogeneity.

To mitigate the problem, it is common practice to grow cells in balanced growth, achievable in batch culture when the population has been growing expo-

nentially for a substantial amount of time, or under chemostat conditions where exponential growth may be maintained indefinitely as a result of the constant influx of fresh media and efflux of the device (Moser et al. [2012]). Under these conditions, the population heterogeneity has been found to be predictable, and the physiological parameters of bacterial populations have been found to be uniquely correlated with growth rate (Bremer et al. [1996], Bipatnath et al. [1998], Abner et al. [2014]). However, such features of the population are rarely taken into consideration when studying SGC dynamics.

1.3 Theoretical Background

1.3.1 Cooper-Helmstetter Model

As covered in Section 1.2.1, the BCD model of the cell cycle separates into three distinct phases. The C phase represents the time required to complete one round of genomic DNA replication. Once the cell completes at least one round of replication, the D phase represents the period during which the cell then undergoes segregation of the chromosomes into two daughter cells to complete cell division. If the doubling time is greater than the sum of the C and D periods, then another phase arises called the B phase, which is simply the time required for the cell to accumulate enough mass to initiate a new round of replication. Combined with the concept of critical mass by Donachie [1968], and the age distribution of an exponentially growing bacterial population (Figure 1.4), the following mathematical equations formalise the main steps of the cell cycle, thereafter referred to as the Cooper-Helmstetter (CH) model (Abner et al. [2014], Keasling et al. [1995]).

The DNA content of a single cell at an age a assuming the timer model (Abner et al. [2014]); where $0 \leq a \leq \tau$ given that 0 is the age of a newly divided cell and τ is the age of a cell about to divide:

$$G_a = G_{mother} + \sum_{n=0}^{k} 2^n G_n \tag{1.1}$$

where G_a is the genetic content of the cell at time a, G_{mother} is the DNA content inherited from the mother cell of a newly divided bacterial cell and n is the replication fork number. This value depends on the constant integer k where 2^k represents the number of origin of replication and l where 2^l represents the number of origin of termination, both for a cell at average age:

$$k = \left\lfloor \frac{C+D}{\tau} \right\rfloor \tag{1.2}$$

$$l = \left\lfloor \frac{D}{\tau} \right\rfloor \tag{1.3}$$

The DNA content of a cell may be calculated given the replication rate (C), segregation rate (D) and doubling time (τ) (note that the sign "|" represents modulo):

$$G_{mother} = (2^k - 1)\frac{\tau}{C} + 2^k \frac{(C+D) \mid \tau}{C} - (2^l - 1)\frac{\tau}{C} - 2^l \frac{D \mid \tau}{C} \tag{1.4}$$

In equation 1.1, G_n represents the amount of DNA that is synthesized for a given replication fork (n):

$$G_n = \frac{a - d_n}{b_n} \tag{1.5}$$

Where d_n is the inherited replication fork DNA content and b_n is the actively replicated DNA content, both calculated using the following conditions:

1. If $a_{i(n)} < 0$, then $d_n = 0$, else $d_n = a_{i(n)}$
2. If $a \leq a_{i(n)}$, then $b_n = 0$
   Else if $a_{i(n)} + C > a > a_{i(n)}$, then $b_n = C$
   Else if $a \geq a_{i(n)}$, then $b_n = (a - d_n) \cdot \frac{C}{a_{i(n)}+C}$

Condition 1 above removes the length of time from the start of the initiation to enable the calculation of time past the initiation point only. The second condition quantifies the progress of the replication forks, where the first line considers if time a is before the initiation point, the second if it actively is being replicated and the last line if it is past the initiation point in question.
The age distribution of a bacterial population growing exponentially has been theoretically determined to follow the following probability density function (Powell [1958]):

$$PD(a) = 2\ln(2)e^{-a\ln(2)}, \quad 0 \leq a \leq 1 \tag{1.6}$$

where $PD(a)$ is the probability for a cell to be at age a, where $0 \leq a \leq 1$, such that $a = 0$ would correspond to a newly divided cell, and $a = 1$ is the age at which a cell divides. An example of the output of the Cooper-Helmstetter (CH) model can be found written in ANSI-C in the ori.c and ori.h attached files.

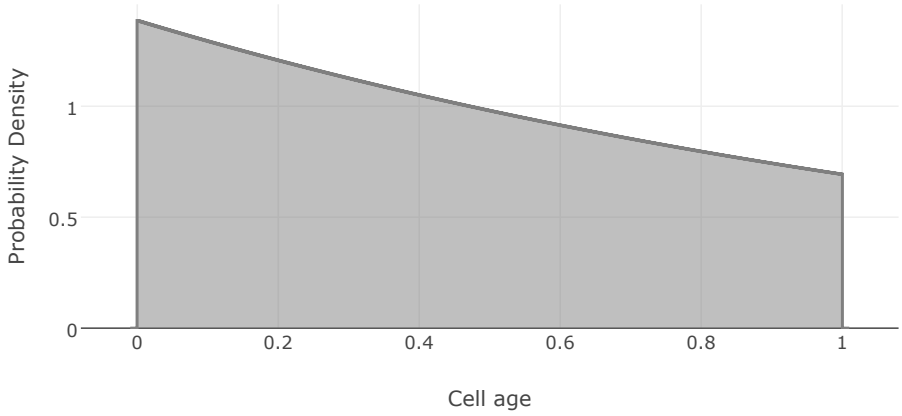


Figure 1.4: Illustration of the age (PDF of Equation 1.6) distribution of an exponentially growing population assuming balanced growth.

Examples of the output of the CH model on different individual cells at different growth rates is illustrated in Figure 1.5. The CCSIM simulation software provides a good interactive display of the timer's model describing the DNA content of a cell in an interactive way (Zaritsky et al. [2011]).

The right end plots include the idealised DNA distributions for populations of bacterial cells in balanced growth, growing at their respective growth rate. Indeed by combining the CH model with the probability density function (PDF) describing the theoretical age distribution of a population growing exponentially (Equation 1.6, and illustrated in Figure 1.4), the DNA distributions of such a population can be calculated using the doubling time (τ), C and D times alone. This strategy has been widely used to determine the C and D parameters, for example by fitting simulated DNA distributions to experimentally measured DNA distributions sampled from exponentially growing cultures (Stokke et al. [2012]).

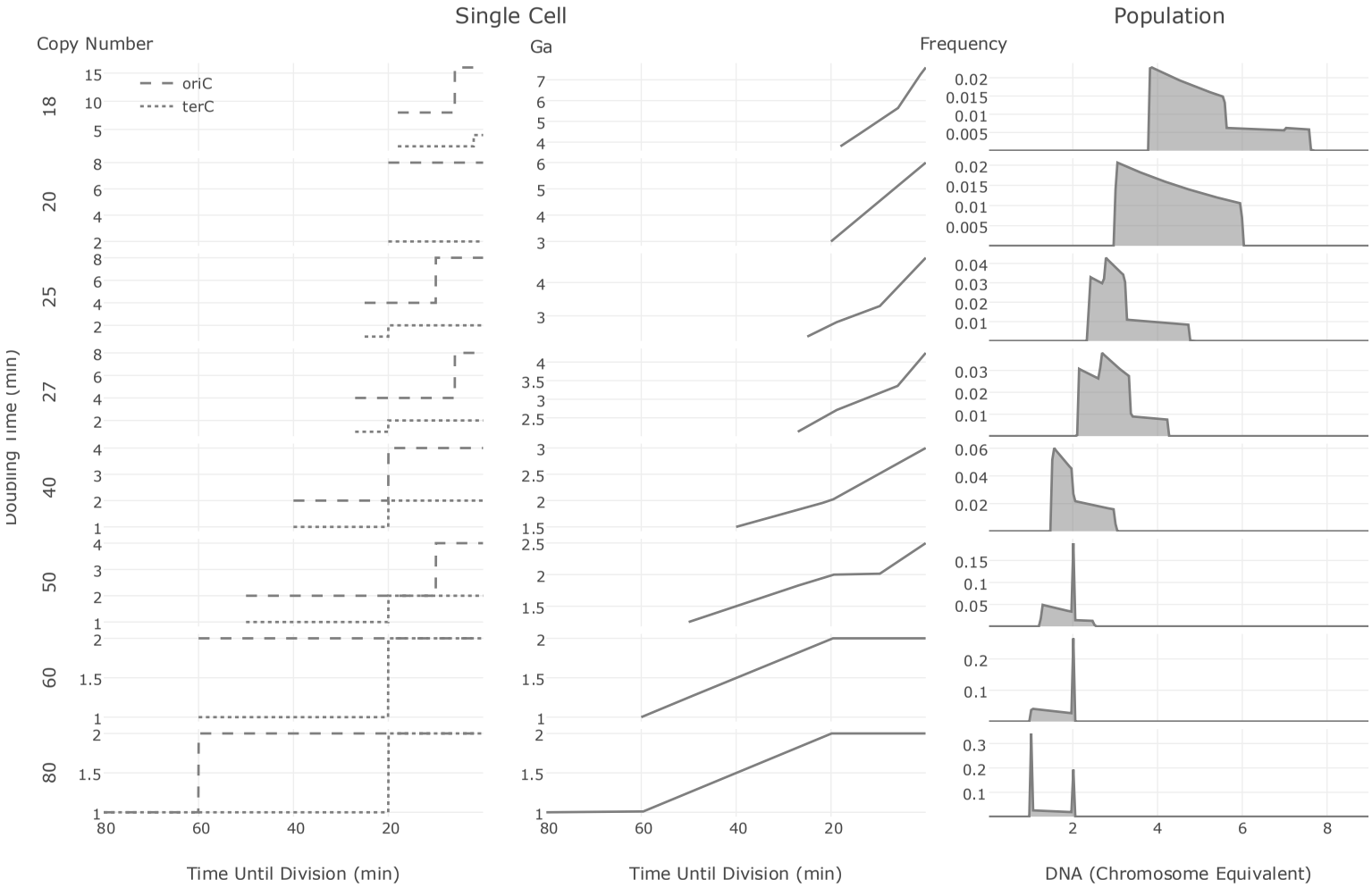


Figure 1.5: Example of single cell and population DNA distribution predictions from the original CH model. The first column represents the number of origin of replications ($|oriC|$) and origin of termination ($terC$) for the doubling of a single cell. The second column represents the DNA content (Ga) during the doubling of a single cell. The third column represents the DNA distribution of a population of cells, assuming that an exponentially growing population has an age distribution described by Equation 1.6. The far-right label shows the doubling rate for each row. ($C = 40.0$ min $D = 20.0$ min)

1.3.2 Individual Based Simulation

Individual based models (IBM) (also called agent-based models (ABM)) have been described as "simulation models that treat individuals as unique and discrete entities which have at least one property in addition to age that changes during the life cycle" (Grimm and Railsback [2005]). The forefather of this type of approach may be attributed to cellular automata (CA), one of the first examples of a discrete dynamical system (Hadeler and Müller [2017]). Where continuous equations describe time and space in a continuous fashion, such as ordinary differential equations (ODE) and partial differential equation (PDE); with IBM style modelling one or all aspects of the model are defined in a discrete manner, be it time, space or state of individuals. The classic example of CA called "the game of life", depicts a two-dimensional grid structure where each grid contains binary states (on/off), and given an initial state at time 0, the model is advanced following deterministic rules based on the state of the grid and its surrounding grid state. Given a starting position and the rules provided, a wide range of states may be achieved including homogeneity and oscillatory behaviours (Hadeler and Müller [2017]).

There are a certain number of biological problems for which continuous models do not capture the essentials, that become highly complex making them non-approachable problems. For example, consider Turing pattern formation of biological systems through the inhibitor-activator model, where diffusive instabilities cause local fluctuations that form complex patterns observable in nature (Hadeler and Müller [2017]). These have been described precisely at the macroscopic level, using Partial differential equations (PDE). However, such a representation does not inform of the microscopic behaviour (MacDonald et al. [2011], Kondo and Miura [2010]). If one considers Turing patterns chemically then a continuous representation is beneficial by its elegance and ability to derive more complex states from such a definition. If however one considers the cellular organisation of such patterns (in perhaps a grid-like structure) and would like to know the "state" of each cell, then such mathematical representation does not capture the necessary complexities (Hadeler and Müller [2017]). Thus the strategy used to explore theoretically different biological structures and functions depends on the system of interest and the specific properties that one wishes to formalise through them (see Figure 1.6).

One important feature of the IBM technique, is the ability to infer a novel property from a system that is not a property of the individuals that it is made of, a behaviour sometimes referred to as self-organisation or emergence (Lints [2005], Tack et al. [2015]). A good example of the complex emergent behaviours that this strategy may model is the development of "boids". In 1986 in an effort to simulate

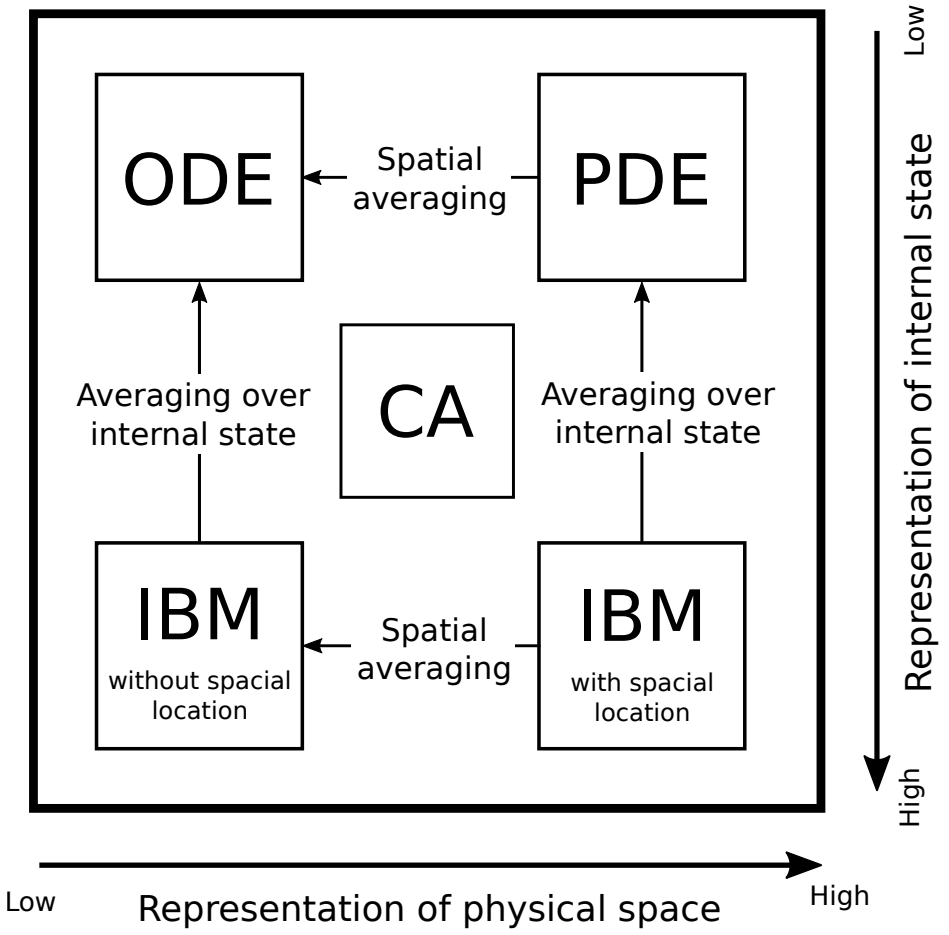


Figure 1.6: Illustration of the link between equation based versus individual based simulation techniques. CA: Cellular Automata, IBM: Individual Based Model, ODE: Ordinary Differential Equation, PDE: Partial Differential Equation. (Adapted from Hadeler and Müller [2017])

the flocking patterns of birds, individual entities where given the following rules in a 3-D plane; to not move too close to its neighbours (separation), to not move too far from its neighbours (cohesion) and to follow the same direction as its neighbours (alignment) (Reynolds [1987]). Without explicitly defining rules on how the cohort as a whole should behave, posing the problem in this fashion returned the same patterns as observed in nature. Thus, for such non-linear and stochastic problems, IBM simulation strategies have proven to be quite useful (Hellweger and Bucci [2009]).

1.3.3 Genetic Optimisation

Genetic optimisation or genetic algorithms are a special class of evolutionary algorithm inspired from natural selection, that use inheritance, mutation, selection, and crossover; behaviours that one typically finds in biological organisms (Tamaki et al. [1996], Kumar et al. [2010]). The basic steps in the algorithm are the following: given a number of individuals that constitute the population (the number of individuals is preserved throughout the optimisation procedure) each has their fitness evaluated occrding to a user defined fitness function. The best scoring individuals within that population are selected to be preserved for the next generation. This same population has its parameters mutated and crossover occurs between individual members to generate a new population that in turn, has its fitness evaluated for the next generation. The process is repeated until an end condition is satisfied (Kumar et al. [2010]). By performing the optimisation in this fashion, statistically the best individuals are preserved throughout the optimisation process, but where lower scoring ones are able to co-exist and generate new offspring that might have a higher fitness. This method is good at solving problems that contain a high number of variables, such that an exact solution is difficult to find. This type of algorithm is classed as a global search heuristic method (Kumar et al. [2010]).

1.4 Aims

The CH model describes the major events of the cell cycle mechanistically and combined with the age distribution PDF in equation 1.6, is an undeniable robust method of simulating for population cell cycle heterogeneity. However, by its assumption of balanced growth, it is fundamentally limited to bacterial populations grown in very stringent and particular experimental conditions. This limits the scope and practical usability of the CH model, since researchers are commonly faced with populations that are not grown in such ideal environments or do not produce perfectly sigmoidal

growth curves where balanced growth is valid.

By removing equation 1.6 one can separate the aspects of the model where assumptions of balanced growth are required (in the CH model presented above, that is the age distribution and thus the assumption that the single cell and population grow exponentially) and be left with a mechanistic model that has been shown to be a valid summary of the major steps in the cell cycle. This research aims at using this proven single cell model, expanding it with newer findings of the cell cycle, to develop a new individual based modelling (IBM) framework for the simulation of population cell cycle heterogeneity.

Thereafter, this research aims are using it in combination with a newly developed growth method to simulate for populations outside of assumptions of balanced growth. Indeed, one difficulty of modelling populations that are not in balanced growth are its temporal features. That is, its state is a consequence of a previous state. It is theorised that by the emergent properties of the IBM simulation method and by using a valid cell cycle model, one needs only substitute for a description of growth that is valid under conditions and phases of growth outside of balanced growth to simulate for temporal features of a population. Lastly, this research aims at using the framework as an explorative tool for the optimisation of cell cycle related features and mutations, and for the exploration of the potential effects of population cell cycle heterogeneity on synthetic biology.

Chapter 2

Materials and Methods

The experimental procedures were developped by Dr. Joshua Leonard and Dr. Andrew H. Scarpelli. All experiments for TOP10 cells were performed at Northwestern University in Chicago by Dr. Andrew H. Scarpelli and were used in the following publication du Lac et al. [2016]. All experiments involving WT cells were generated for this work at the University of Warwick by me.

2.1 Media and Bacterial Strains

Two different strains were used in this study. The MG1655 K-12 (F λ ilvG rfb-50 rph-1) and the TOP10 strain (F- mcrA δ(mrr-hsdRMS-mcrBC) ϕ r-hsdRMS-δ r-hsd nupG recA1 araD139 δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-) (Invitrogen).
The bacterium were grown in LB and glucose supplemented M9: Lysogeny Broth Lennox (LB) media composition (Diluted in 1L H_2O and 0.2 μm filtered):

10g	Tryptone
5g	Yeast extract
5g	N_aCl

M9 salts solution (with pH adjusted with $NaOH$ to 7.2):

$47.748722349 \cdot mM$	$MgSO_4$
$22.044883 \cdot mM$	KH_2PO_4
$8.555784 \cdot mM$	N_aCl
$18.695083193 \cdot mM$	NH_4Cl

M9 media for MG1655:

$47.748722349 \cdot mM$	$MgSO_4$
$22.044883 \cdot mM$	KH_2PO_4
20% (w/v)	Glucose

For TOP10 cells the M9 media composition was the following M9 salts with 0.4% glycerol, 0.2% casamino acids, and 1 mM thiamine hydrochloride. For the TOP10 cells, streptomycin was used at a final concentration of 50 $\mu g \cdot mL^{-1}$. All cultures were incubated at 37°C with various shaking rates in batch culture (23 and 230 rotations per minute (rpm)), incubated from overnight cultures grown overnight in LB. For TOP10 cells in $3mL$ with appropriate levels of streptomycin and for MG1655 in $250mL$ shake flasks with $100mL$ of media for both the overnight and experimental cultures.

2.1.1 Chromosomal DNA Quantification

For K-12 bacterial strain examination by flow cytometry, $500\mu L$ of sampled cultures were fixed with $500\mu L$ pre-chilled high grade ethanol at -20 °C in $1.5\mu L$ eppendorf tubes. The samples where then centrifuged at 5000rcf for 5min and the supernatant was disregarded. The pellet was re-suspended in $500\mu L$ of a solution of ice-chilled $0.2\mu m$ filtered PBS with $12\mu g \cdot mL^{-1}$ of Hoechst 33342 stain for an OD_{600} of 0.01. The solution was then left on ice in the dark for a minimum of 30 minutes and a maximum of 8 hours until flow cytometry analysis. The samples were analysed using the flow cytometer BD LSRFortessa (FSC: 646, SSC: 243, UV53: 498) with a maximum of 100000 cells and a minimum of 10000 cells. The experiments were performed three times, and the results presented are the sum of these three runs.

For TOP10 bacterial strains examination by flow cytometry, sampled cultures were diluted 5 μL culture into 200 μL ice-chilled PBS. For OD_{600} under 0.5, 1 μL culture into 200 μL ice-chilled PBS. For OD_{600} between 0.5 and 2.0, and 0.5 μL culture into 200 μL ice-chilled PBS. For OD_{600} over 2.0, 800 μL ice chilled ethanol was then added to this solution. The solution was gently shaken, and then immediately spun with a microcentrifuge for 5 minutes at 1000xg. Ethanol solution was discarded and pelleted cells were then re-suspended in 500 μL chilled PBS, and spun a second time in the same conditions. Cells were then re-suspended in 500 μL chilled PBS with 1 $\mu L \cdot mL^{-1}$ DAPI, and immediately placed in a 4 °C refrigerator until samples could be run on the flow cytometer. Cells were run on the flow within 24 hours of collection. Samples were analysed on an LSR II (BD)). A minimum of 2000 individual cells (typically out of 25000 events) was analysed per sample.

2.2 Computational Methods and Packages

2.2.1 Normalising Fluorescence to Genomic Content

Flow cytometry returns the relative fluorescence intensities (in this work) of stained bacterial chromosomes. These are a collection of discrete values that are typically binned into histograms, for easier human interpretation, by most analysis software. To convert these fluorescence intensities to their DNA content equivalents, one must use a standard population with known DNA content as a means of comparison (Michelsen et al. [2003]). To generate such standard the bacterial population was grown for a prolonged amount of time in batch culture, such that the population copy number falls to integer amounts of chromosome (Michelsen et al. [2003], Skarstad and Boye [1993b]).

Furthermore, to enable the quantitative comparison between simulated and measured DNA distribution, the simulated population must be convoluted to reflect experimentally measured DNA distributions. Indeed, the latter reflect the accumulated experimental variation associated with both labelling (e.g., variable efficiencies of fixation, permeabilisation, and binding of DAPI or Hoechst to DNA) and detection (e.g., variability in the signal measured by flow cytometry given a fixed quantity of DAPI or Hoechst in a single sample). Therefore, even if a population of cells included a discrete and uniform amount of DNA per cell, the measured DNA distribution would be "spread" by these other sources of variability, and this spread must be quantified. It has been observed that the spread increases linearly with increasing fluorescence intensity (Michelsen et al. [2003]).

From these fits, two important parameters were calculated. From the means of the Gaussians, a calibration curve for relating discrete numbers of chromosomes to flow cytometry channels was generated (levels of DAPI or Hoechst associated fluorescence per cell). Notably, the standard deviation of the peaks followed the same pattern observed in previous reports (Michelsen et al. [2003]) (Figure 2.1); with an increasing amount of DNA, the standard deviation of the peaks increases linearly. This makes sense physically, since when more DAPI or Hoechst labelling takes place, the variability associated with both labelling and quantifying cell-associated DAPI or Hoechst would also increase. These parameters enabled the conversion of simulated DNA distributions to simulated measured DNA distributions by Gaussian blurring. Due to the potential for variation in measured fluorescence values between flow cytometry runs performed on different days, the calibration was repeated on each day. For K-12 cells, each peak may be interpreted as having 2^n, where n is an integer for each subsequent peak, while for TOP10 each peak corresponds to an

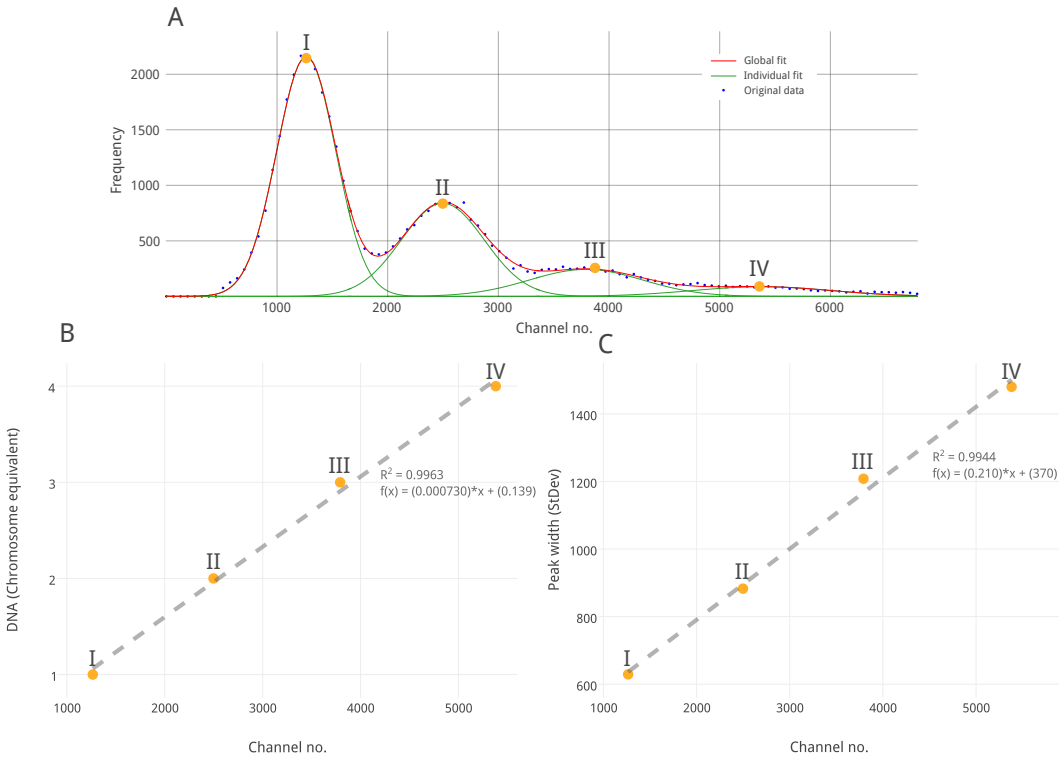


Figure 2.1: Example of computational deconvolution of DNA distributions using the ipf11 MATLAB package to identify a sum of Gaussian distributions that is maximally consistent with the measured data. The scatter plot in blue in the top subplot represents TOP10 DNA distribution measured by flow cytometry with DAPI stained chromosomes after 24h incubation in LB at a shaking rate of 230 rpm. The global, sum of Gaussian fit, is in red and the individual Gaussian fits to the peaks are in green. Because recA1 expresses the aberrant phenotype (see Setion 3.2.2), each peak corresponds to an increase each peak, from left to right, is interpreted to have 1, 2, 3, 4 chromosome copy numbers. The bottom left hand subplot plots the chromosome copy number to the fluorescent channels, and fit to a first order polynomial function. The bottom right hand subplot is the width of the fit Gaussian (standard deviation) plotted against the fluorescent channels, also fit with a first order polynomial function. These former functional form is used to convert measured DNA distributions to their chromosome equivalents, while the latter is used to convolute the simulated data to enable the quantitative comparison between the two. Taken from du Lac et al. [2016].

increase in one chromosome equivalent.

2.2.2 Computational Packages

All up to date files may be found here: https://github.com/Melclic/HMG.

Files returned by the flow cytometer (fcs), were analysed using the python package fcsparser (version 0.1.4) with python3. For the spline fitting procedure, a script written in python3 using the following packages: the spline method was imported from the Wave Analysis for Fatigue and Oceanography (WAFO) package (version 0.3.1), the scientific packages numpy (version 1.13.1) and scipy (version 0.19.1). For noise convolution and nonlinear regression scipy packages signal and stats were used. For sensitivity analysis the Sensitivity Analysis Library in Python (SAlib) package was used (version 1.1.0). For computing the sum of Gaussians the interactive peak finding package (ipf11) in MATLAB was used (version 12.1).

For the cell cycle model, the program was written in ANSI-C including C Standard Libraries: stdbool.h for boolean data types, time.h for measuring execution time, string.h for string handling, math.h for basic mathematical operations, stdio.h for input/output of files and stdlib.h for memory management. For solving ODE in ANSI-C, the GNU Scientific Library (GSL) was used including: gsl_matrix.h for allocating and creating matrices, gsl_odeiv2.h for defining the ODE system, and gsl_errno.h for error handling. To compile the model the gcc compiler was used (version 6.3.0).

For the genetic optimisation, the Distributed Evolutionary Algorithms (DEAP) package (version 1.0.2) was used with the AeMuLambda algorithm, with a Gaussian mutation function ($\mu = 0.0$, $\sigma = 0.005$, probability $= 0.75$) in combination with the native python3 package multiprocessing (version 2.6). For the generation of Sobol sequences for the start of the optimisation procedure the python sobol_seq (version 0.1.2) packages was used. To connect to the ANSI-C model, the native ctypes (version 2.5) was used.

All were compiled and ran on a Linux Ubuntu (64bit) 24-core Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10 GHz, 250GB RAM.

Chapter 3

Model Development

This chapter reviews the currently available formal mathematical descriptions of the cell cycle at the single cell and population levels, and examines the implications of this theory and the experimental evidence to support it. Three different aspects of the cell cycle, and their interplay, are studied:

1. Growth dynamics: the observed single cell growth dynamics and the mathematical formalisation that ensued. The growth dynamics of bacterial populations are also considered, how population growth is modelled and how single cell growth links to the population growth heterogeneity. A new strategy of tracing bacterial growth based on the conversion of OD readings to their volumetric equivalent is introduced.

2. Chromosome dynamics: Modelling of chromosome dynamics for a bacterial cell population is discussed. How and when the cells trigger the initiation of their chromosomes, the speed at which they replicate, and the consequence on the DNA content for the cell and population.

3. Cell division: This process and its consequences are considered fundamental and yet arguably the least well understood aspect of the cell cycle. Particular focus is given to how single cell division patterns influences the volume and DNA homoeostasis of the population.

The three distinct physiological features of the cell cycle; cell growth, chromosome replication and cell division are intimately linked, and thus in the literature they are often described as one unbroken chain of events (Abner et al. [2014], Keasling et al. [1995], Zaritsky et al. [2012]). However as will be explored, there is no consensus as to the degree of their interdependence. More research is required to understand the degree and pattern of control each distinct physiological

feature has over another, and the molecular links between them. The current way one describes either qualitatively or quantitatively the links between cell growth, chromosome replication and cell division is largely due to the fundamental work of Cooper, Helmsteer and Bremer (CH model; see Section 1.3.1) in developing the first rational mathematical description of how a population maintains volume homeostasis and its link to the DNA dynamics at the single cell and the population levels (Bremer et al. [1996], Abner et al. [2014], Keasling et al. [1995]). However, their model was built using population data that do not truly represent individual cells, and can thus obscure the understanding of the behaviour of individual cells within that population (Osella and Lagomarsino [2013]). Recently, with the availability of more precise single cell data, various features of this model have been able to be revisited (Basan et al. [2015], Sauls et al. [2016], Harris and Theriot [2016], Wallden et al. [2016]). Here, many of the assumptions of the CH model will be re-investigated.

3.1 Growth Dynamics

3.1.1 Single Cell Growth Dynamics

From experimental data, sometimes dating back as far as the 1970's, of time series confocal microscopy measuring the bacterial length or volume over the cell cycle, two different theories have been proposed as to the dynamics of single cell growth: linear growth and exponential growth (Pruitt and Kamau [2007], Abner et al. [2014], Reshes et al. [2008]). Abner et al. [2014]'s theoretical investigation of single cell growth reports a maximum difference of 6% between the two. Because of this small difference, high single cell resolution of cell volume or length is required to be able to distinguish with certainty if the exponential or linear growth law applies to single cell growth. More accurate techniques, such as electron microscopy, usually requires the fixation of the cell and the use of treatments that might inadvertently alter the size of the cell (Kemp et al. [1993a]). More non-invasive methods include confocal microscopy, light microscopy and flow cytometry (Reshes et al. [2008], Kemp et al. [1993b]). Below are the two mathematical expressions of the exponential and linear elongation laws:

$$V_{exp} = V_0 \cdot e^{\frac{ln2}{\tau} \cdot a} = V_0 \cdot e^{\mu \cdot a} \tag{3.1}$$

$$V_{lin} = V_0 + \frac{V_0}{\tau} \cdot a \tag{3.2}$$

where V_{exp} is the exponential functional form and V_{lin} is the linear function form.

τ is the doubling time, and μ is the instantaneous growth rate that are related in the following manner: $\mu = \frac{ln2}{\tau}$. V_0 is the volume at birth and a is the age of the cell where $0 \leq a \leq \tau$.

More recent experimental quantifications of single cell growth using high powered confocal microscopes combined with microfluidic baby-machine devices, suggests that cells grow exponentially (Wang et al. [2010], Taheri-Araghi et al. [2015]). However, most of these studies rely on the length increase of a cell over time and overlook the width, assuming that the latter does not change significantly over a single doubling event when under the same growth conditions (Wang et al. [2010], Taheri-Araghi et al. [2015], Campos et al. [2014]). Because of its rod shape, a change in its width would have a greater impact on the volume compared to an equal change in its length, and is thus a dangerous assumption (Reshes et al. [2008]). There is a significant body of work that demonstrates that the width of a cell does indeed change over time (Cooper [1988], Zheng et al. [2016], Trueba and Woldringh [1980], Wang et al. [2010]). While some suggest that cell width decreases during the cell cycle (Trueba and Woldringh [1980]), investigation of the single cell data from Wang et al. [2010] (Figure 3.1) suggests that the width of the cell increases $\approx 0.13\mu m$ over a single cell cycle. The former uses invasive experimental techniques that may influence the measurement outcomes, while the latter uses a non-invasive experimental method but generates such noisy width data that in both cases it is difficult to be conclusive. Because more experimental work is required to be able to distinguish between linear and exponential growth, this model will assume that individual bacterial cells grow exponentially.

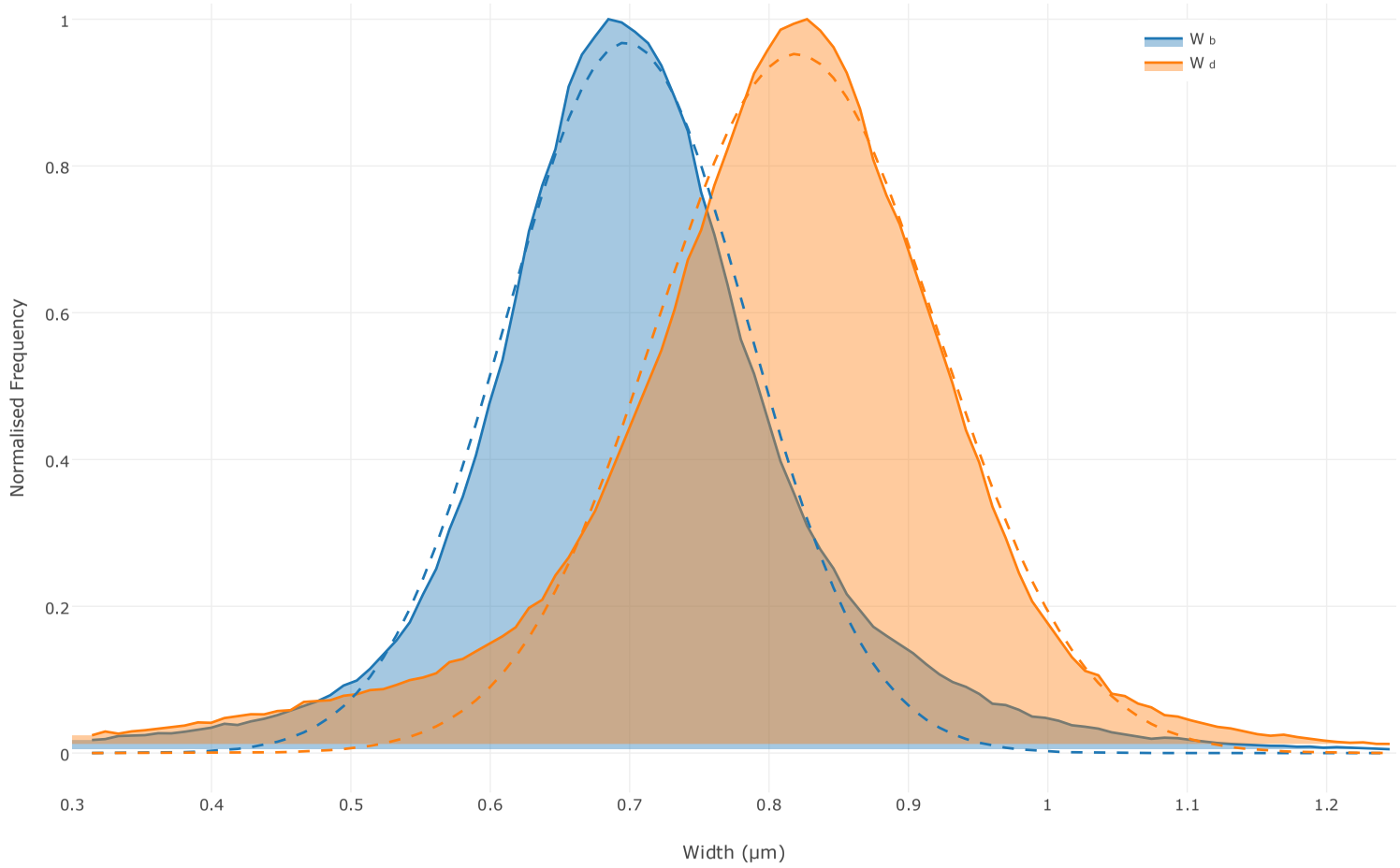


Figure 3.1: Single cell data (Wang et al. [2010]) histogram of the width at birth (W_b) and against the width at division (W_d). The dotted lines are the Gaussian fits to the data with means of 0.7, 0.82 and standard deviations of 0.123, 0.143 for width at birth and width at division respectively.

3.1.2 Balanced Growth

With the advent of the chemostat in the 1920's, microbiologists were able to grow bacteria in the same conditions and in the same physiological state for a prolonged amount of time, something only achievable in batch culture with continuous reinoculation (Dalgaard et al. [1994]). This enabled the precise reproducible examinations of the effects of defined media on bacterial population physiology and uncovered that without external modifications of the environment, bacterial cells as a population did not change over time (Dalgaard et al. [1994]). Under such growth conditions the population is said to be in "steady state", a thermodynamic term used to describe a particular condition of a system that is in equilibrium, where its variables are constant in time (Fishov et al. [1995]). This idealised formalisation for a system is commonly applied in many different fields including engineering, economics, biology, chemistry and others; typically as an approximation method since it offers solid mathematical simplifications (Gonze [2013]).

To apply the same approach to bacterial populations it is necessary to introduce the concepts of extensive and intensive variables. The former are variables that describes a physical quantity of a system that depends on its size (for single cells that includes volume, mass, DNA and protein for example) or describes the quantity of a system (Akerlund et al. [1995]). The latter are all variables that do not depend on the size of the system (for example temperature, molarity and distributions of extensive parameters), where regardless of how many times the system is subdivided the variable does not change (Fishov et al. [1995], Akerlund et al. [1995]). Thus, when considering a growing population of single celled organisms, balanced growth is defined as a situation where "every extensive property of the... system increases by the same factor over a time interval" (Campbell [1957]), while steady state growth applies only if both intensive and extensive parameters are independent with time (Fishov et al. [1995]). True steady state for a bacterial population is only achievable under growth in a chemostat, with a constant efflux of bacterial cells in combination with a constant influx of fresh media (Wallden et al. [2015]). Hence, a population that is in steady state is automatically in balanced growth, but the opposite is not necessarily true. Often when describing populations, steady state and balanced growth are used alongside each other, when they describe two distinct conditions. Furthermore, populations can be in steady state or balanced growth while single cells, because of discontinuous processes such as chromosome replication that make the extensive parameters of the cell change with time, are not in balanced growth themselves, let alone steady state (Akerlund et al. [1995]). In conclusion, it is perfectly valid to consider a bacterial population to be in balanced growth while

the cells that constitute it are not. However, this top down approach studying the state of individuals within a population makes it difficult to deduce the dynamics and state of the individuals within a population if experimental procedures are not stringent enough to assure that all the above conditions are met.

From the early works on chemostats, perhaps one of the most fundamental findings ensued. Given two populations in two different environmental conditions but with the same growth rate during balanced growth, the measured macromolecular composition of these populations was found to be identical (Herbert [1961], Bremer et al. [1996]). This includes all extensive parameters of the model, that are observed to be unchanging for a population under conditions of balanced growth, and depend only on the doubling time (τ) of the population (unless of course the rate limiting factor for the growth of the population was an antibiotic that impeded the ability of the bacterial cell to replicate its chromosome) (Michelsen et al. [2003], Bremer et al. [1996]). If every cell in a population experiences the same growth limiting factor, and is considered to be in balanced growth, it may be stated that this population contains these following features. First the population is atemporal, that is, regardless of when a sample is taken in a population that is in balanced growth, every feature of the population should be the same. This implies the second feature of the population, that it contains a predictable heterogeneity that may be quantified deterministically (Keasling et al. [1995], Bremer et al. [1996], Allman et al. [1991]). This is indeed how the CH model is framed and why it has such powerful predictive tool for population cell cycle dynamics. However, as previously mentioned, the experimental requirements to achieve a condition where the CH model is valid can be difficult to acheive equally in batch culture as in chemostat (Moser et al. [2012]).

3.1.3 Modelling Population Growth

Continuous Growth Models

A typical bacterial growth curve includes the lag phase, log phase, stationary phase and death phase; all representing distinct conditions (Schumann [2006]). There are many mathematical formulations that describe the growth of bacterial populations, including but not limited are the models of Gompertz, Roberts, Monod, Logistic, Malthusian and different derivations of the latter. It is important to explain their characteristics and implied assumptions when one fits these growth models to infer growth parameters from OD growth curves such as growth rate, lag time and other features.

The first and most distinguished model of population growth is the Malthusian growth model, that proposes that a population grows at a rate which is proportional to its size at any given time (Pruitt and Kamau [2007]). This simple exponential growth theory suggests that growth is constant since the population experiences no growth limiting factors. It is thus only valid under early batch culture with an abundance of media and space or in continuous cultures (Pruitt and Kamau [2007]). Given that all populations experience limiting factors that inevitably restrict their growth, the logistic model introduced a growth limiting factor dictated by the availability of a single essential factor (Pruitt and Kamau [2007], Tsoularis and Wallace [2002]). This is usually defined as the carrying capacity of the environment of the culture, which in the case of bacterial cells, would be dictated by conditions such as available resources for growth, physical space, toxicity, etc... However, the logistic functional form of growth does not explain the phenomenon of lag phase, and implies that the growth of a population contains an inflexion point of maximal growth, a concept which is disputed (Buchanan et al. [1997]). To overcome the limitations of the Logistic equation, other models of population growth, such those proposed by Gompertz, Richard and Bertalanffy have emerged, largely in order to describe the lag phase quantitatively that are all children of the logistic function and thus if the Malthusian growth model. The Monod equation is another type of population growth model that uses the same functional form as the Michaelis-Menten equation, describing enzyme kinetics, to summarise population growth by the consumption of a single substrate (Zwietering et al. [1990], Buchanan et al. [1997]). This growth model is widely used since it fits growth curves relatively well, while at the same time providing an empirical explanation of the relationship between growth and substrate availability (Monod [1949]).

It is sometimes difficult to identify one continuous growth model over another given even a typical sigmoidal shaped growth curve. The R package grofit is a good illustration of this problem for biologists, where the software provides the possibility of iterating through a series of models and lets the user select the model that best fits the growth curve input (Kahm et al. [2010]). Furthermore, it is often the case that growth curves, especially from OD data, do not follow the traditional lag, log, and stationary phases. For example, cells grown with complex media experience diauxic growth (meaning double growth) reflecting the consumption of the easiest accessible carbon source before the consumption of a more difficult one (Loomis and Magasanik [1967]). Overall, continuous growth models may contain parameters that, by not being tied to any specific biological function, lack mechanistic meaning (Horowitz et al. [2010], Zhang et al. [2006]).

Injection Growth: A Condition Independent Growth Strategy

Because of the lack of consensus, the non-mechanistic parameters of some continuous growth models, and the many growth conditions that lead to non-classical sigmoidal growth curves; a new method of tracing population growth was developped that better suits the need for a reliable fit to OD data (Buchanan et al. [1997]).

Given that the growth curve reflects changes in population volume in the following manner (Zhang et al. [2006]):

$$V_{mean} = \frac{\Delta N}{\Delta t} \tag{3.3}$$

where N is the number of bacterial cells, V_{mean} depends on the scale of the considered time, t, the instantaneous growth is the first order derivative of that change where (Zhang et al. [2006]):

$$V_{inst} = \frac{dN}{dt} = \lim_{\Delta t \to 0}(\frac{\Delta N}{\Delta t}) = \mu \tag{3.4}$$

Since V_{mean} and $\frac{dN}{dt}$ both represent volumetric changes, it is theorised that the instantaneous growth rate (V_{inst}) may be used outside of balanced growth (Zhang et al. [2006]). Indeed, previously V_{inst} was used to deduce the state of a population only under assumption of steady state (Abner et al. [2014], Keasling et al. [1995]). Thus, instead of using a model and curve fit measured OD data to inform us of the dynamics of the population, growth curves were fitted using a spline interpolation method (Zhang et al. [2006]). This procedure ensures that the growth curve reflects precisely the measured data, regardless of its shape while making no assumption as to the dynamics of the population.

Often, OD data is used as a means of calculating the concentration of bacteria in a sample where the agreed estimate is that an OD of 1 gives $8.0 \cdot 10^8 \cdot mL^{-1}$ cells (Sezonov et al. [2007]). OD is simply a quantitative measure of the refracted light through a sample, and as a consequence such an estimate is known to be an oversimplification (Volkmer and Heinemann [2011]). Certainly an increase in the cell numbers in a sample would lead to an increase in diffracted light from the OD reader. However, consider two bacterial samples, one with cells that are grown in a rich media such as LB and another that is grown in a poor media such as M9 at equal cell concentration. Cells growing in a richer media are larger than the ones grown in a poorer media (Volkmer and Heinemann [2011]). Thus, at equal concentrations, the LB sample would return a larger OD than that of the M9 grown population. From this knowledge, Volkmer and Heinemann [2011] measured the OD of bacterial cells grown in a wide range of conditions with different growth rates,

and normalised the measurements to the concentration of cells per mL through flow cytometry analysis. The authors calculated the following constant:

$$3.6 \quad \mu L \cdot OD_{600}^{-1} \cdot mL^{-1} \tag{3.5}$$

where the total volumetric concentration of a sample is unchanging regardless of the growth conditions of the sample and the size of the cells. The authors conclude that OD quantitatively reflects the total volume of a sample and not the cell concentratio ot the volume of a single cells that populate such populations.

From the constant in Equation 3.5, one can calculate the change in volume of a population (per mL^{-1}) for a given time step:

$$V_{a+dt} = 3.6 * OD_{600}(a + dt) - 3.6 * OD_{600}(a) \tag{3.6}$$

where OD_{600} is the measured OD at wavelength of 600 nm and a is the current simulation time and dt the time step. Using Equation 3.6, one can convert an OD growth curve to its volumetric equivalents and calculate the volumetric increase given any arbitrary defined time step. By distributing these volumetric changes among members of a population the growth of a population is simulated, where regardless of the conditions of growth the simulated population would always follow the input or measured volumetric changes. Indeed, simulating individual cells as growing exponentially, as Keasling et al. [1995], is contingent on assuming that the population is in balanced growth, while this method of growth may be used whatever the dynamics of growth for a given population.

The question remains as to the relevance of this growth strategy to reflect individual cell growth within a population. Indeed, knowledge of total volumetric increase of a population does not reveal the true growth dynamics of individuals within that population, since these volumetric changes must be distributed among the members of the population. Experimental evidence suggests that single cells grown in a range of different environments express growth rate distributions that collapse onto their respective mean in a Gaussian manner (Wallden et al. [2015], Taheri-Araghi et al. [2015]). However, not all members of a bacterial population grow within the mean. The most striking example is the phenomenon of persistence, where a subpopulation within an isogeneic culture grows at a reduced growth rate as a means to resist the occurrence of environmental changes (Balaban et al. [2004], Patra and Klumpp [2013]). Other related states include the non-culturable state, the stationary phase contact-dependent inhibition and filamentation (Llorens et al. [2010]). And indeed, it seems that this heterogeneity in conditions of growth

is fundamental in the way bacterial cells adapt so quickly to changing environments. Furthermore, cell death is an inevitable factor that, considering that the cell's membrane is not compromised, would also influence the distribution of recorded volumetric changes from Equation 3.6 to the individual members of a population (Stewart et al. [2005]). Because it is unclear what are the exact conditions and the rate of occurance of these bacterial states, it will be assumed that they play a small role in affecting total volumetric changes of a population as they grow in disparate growth regimes.

Since growth rate has been measured to be normally distributed, both under assuption of exponential growth and using the injection method, growth rate distributions are assumed to be normally distributed. For exponential growth, each cell sees its doubling time (τ) assigned at birth and division, randomly selected from a normally distruted pool with a mean that is user defined at the start of the simulation and standard deviation σ_{V_a} (see Table 3.1) (Keasling et al. [1995]). However, during injection growth, the HMG simulation framework forgoes the growth rate parameter. To assure that the same observed Gaussian distribution of growth is maintained, it was required to implement a new parameter called "injection deviation":

$$V_{a+dt} = V_a + V_{inc} * InjectionDeviation \tag{3.7}$$

where V_{inc} is the individual cell's increase in volume as per the injection growth. And $InjectionDeviation$ is a parameter that is assigned at birth and division of a cell, that is randomly selected from a normally distributed pool with mean of 1 and the same standard deviation σ_{V_a} as for exponential growth in Table 3.1. By keeping the same "injection deviation" parameter throughout the cell cycle just as the doubling rate is assigned, the same volume distributions effects can be seen compared to an exponentially growing population.

Furthermore growth rate is the only dependent parameter that links the C and D times to the state of the cell. So in injection growth, there is a need to link the C and D times to the physiological state of a cell to respect their respective functional forms. Given that with the injection growth, each cell increases its volume in a passive manner, this increase is employed to back-calculate the growth rate of individual cells in the following manner:

$$\mu = \frac{V_{a+dt} - V_a}{V_a \cdot dt} \tag{3.8}$$

where μ is the instantaneous growth rate, V_a is the volume of the cell at time a

and dt the time step of the simulation. This calculation is performed at creation or division only to resolve appropriate C and D times, in exactly the same manner when a population is growing under exponential growth. This results in individual cells containing a C and D rate that follows the growth pattern of the individual cell. The individual cells are thus either grown assuming exponential growth or injection growth as ilustrated by Figure 3.2.

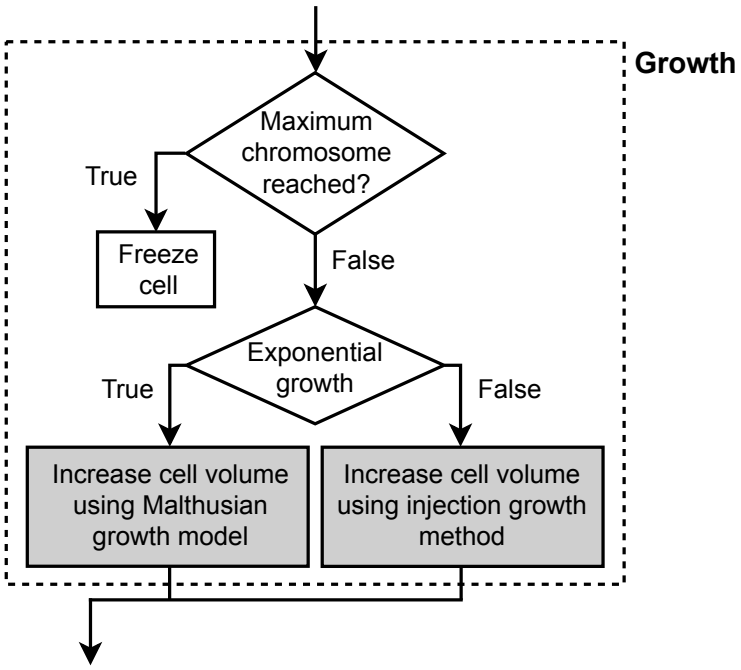


Figure 3.2: Flow chart describing the growth algorithm for a single cell in HMG. The grey boxes represents the steps with some degree of stochastic noise (see Table 3.1). The cell was either grown using an injection growth method or assumed exponential growth. Growth was "frozen" if the maximal number of chromosomes was reached (32 chromosomes maximum in this work), as to avoid crashes in the simulation.

Identifying Exponential Growth

Because population heterogeneity is extensively studied and well formalised using the CH model, this knowledge is leveraged by generating the initial "seed" population assuming exponential growth and thus balanced growth. A method for identifying the exponential (and post-exponential) phases of growth was required. The results of the spline fit were analysed where the user is guided to the theoretical area where exponential growth applies using a minimising the following equation (Zhang et al. [2006]):

$$\frac{d^2(\ln(f(t)))}{dt^2} \tag{3.9}$$

Given the typical noise associated with OD measurements, particularly at low OD values, this methodology leads to the definition of exponential growth as representing a temporal window over which the minimization condition is satisfied (thus, any early inflection point, which also satisfies the minimization condition, is not identified as the window of exponential growth).

The growth curves were analysed and fit using a spline fitting method. Because the HMG modelling framework does not implement cell death, if the decrease is small, then the fitting protocol interprets it as stagnation. However when that reduction is too large, when for example a bacterial culture is in death phase, then that section of the growth curve is ignored. For example in Figure 4.7, the M9 23 rpm population, although cultivated for 24h as all other populations, returns a substantial portion of the OD growth curve that is ignored due to the OD that is excessively getting smaller. Furthermore, typical spectrophotometers have an accurate range between 0.1 and 1.0. Consequently, even if from visual inspection one could consider that the exponential section of the growth curve appears earlier than the ones that where returned, because earlier OD's are smaller than 0.1, they are considered to be unreliable. Furthermore, earlier time points return low cell concentration that may be too small for accurate analysis using flow cytometry.

As an illustrative example, consider Figure 4.1b, which demonstrates the steps in identifying the section of the growth curve where exponential growth applies without using a growth model. The top figure shows the spline fit given the original measured OD data points, on a log scale. In this case, one would fit a first order polynomial to the linear section of the growth curve and its slope would represent the instantaneous growth rate of the exponentially growing section of the OD data. The middle subplot shows the first order derivative of the natural log fit while the bottom shows the second order derivative and is intended to show the section of the growth curve that minimisation of Equation 3.9 identifies. Inspecting the growth curve, one may consider that the exponential growth curve happens at around 100 min, because of the low OD, these values are ignored. More generally exponential growth applies when the first order differential of the fit plateaus and thus when the second order differential of the fit = 0. However, due to the noise of the OD data, values close to 0 are accepted.

3.1.4 Non-Balanced Growth

Although laboratory strains of bacteria are typically grown in ideal conditions, saturated with a carbon source; naturally bacteria live in conditions where they experience long times in starvation and/or conditions of stress, where availability of carbon source is intermittent at best (Kolter et al. [1993]). Genetic expression analysis of cells in stationary phase compared to exponentially grown ones show that a predictable set of genes are activated, designated the Pex (postexponential) proteins (Kolter et al. [1993]). It would thus make sense that naturally, bacteria prepare themselves for starvation and stressful conditions, and it is of particular interest to investigate this state of growth as much as exponential growth. For example, it was noticed that industrial scale bacterial bioreactors with constant feed, after inoculation's exponential growth stabilised to a distinct population state that did not resemble exponential one (Moser et al. [2012]). Despite this, research has largely focused on exponential growing populations.

The bacterial population dynamics outside of stringent exponential growth conditions was explored, where balanced growth does not apply. A population that is outside of the exponential phase has innate temporal features that are indeed difficult to predict with deterministic models, since its heterogeneity is a direct consequence of the previous state of the population. A typical batch culture progressing from log phase to stationary phase, is a good example of populations transitioning from a condition of balanced to non-balanced growth conditions. Although the stationary phase is usually thought of as a phase of growth where the culture is depleted of a carbon source, by its very definition, it applies when the number of cells does not increase substantially with time (Ferenci [1999]). This may occur due to a wide range of conditions, such as acidification of the environment, to a lack of oxygen, or carbon limitations (Kolter et al. [1993], Ferenci [1999]).

It is proposed that the distribution and heterogeneity of extensive parameters of bacterial population in non-balanced growth can be formally predicted given an appropriate simulation method. Using the injection growth strategy, this method of distributing the volumetric changes experienced at the population to single cell level will be examined to see if it contains enough information to describe the same elements of the population heterogeneity as when assuming balanced growth with the CH model.

Rate-Maintenance

In the 1970's bacteriologists first performed nutritional shift-up experiments to investigate the growth patterns of bacteria. Upon shifting a population of cells to a richer media, it is observed that "the rates of cell division were unchanged for about an hour", where following that time the cells increase their rates abruptly (Kubitschek [1969], Cooper [1969], Zaritsky et al. [2012]). It was assumed that this delay is caused by the minimal time required for chromosome replication, segregation and division, while other features of the cell increase immediately (Ho and Amir [2015]). This dichotomy puzzled the biologists at the time, and coined the term "rate-maintenance" to describe features of the cell that are preserved for a significant amount of time after the shift in growth. In the cell cycle model these include replication rate (C) and the division rate (D). The rate-maintenance feature of bacterial populations was implemented in the single cell model, as described in Figure 3.12, by updating all cell cycle parameters upon division and not continuously as a function of growth rate (dotted square on the bottom left describing cell division). Implementing the cell cycle model in this fashion would in theory return the rate-maintenance features as described in the literature. Indeed, with both injection and exponential growth, the volumetric changes of the populations are immediately reflected in the individual cells, while the cell parameters required the event of a cell division before being updated. Although seemingly trivial, the correct implementation of such a feature is important in the faithful reproduction of the population dynamics. Because of this difference between volume growth and DNA growth, it is observed that cells undergo multiple rounds of replication during this one hour adaptation (Ho and Amir [2015]).

Case Study: The Shift-Up Experiment

To illustrate the requirement for a new modelling strategy to quantify population heterogeneity throughout more diverse conditions of growth, a nutritional shift-up experiment was performed (Figure 3.3 and Figure 3.4). This experiment consists of growing a population of bacterial cells in a carbon poor growth environment, and then transferring the cells to a richer carbon source. As a consequence, the bacterial culture experiences two exponential growth conditions and a transition period between the two that is akin to a controlled reversed diauxic growth; a known behaviour of bacterial cells that are grown in two different sugars, where the most easily accessible carbon source is used before the other (Zhang et al. [2006], Narang A [2007]). In order of appearance: the first transition period is the lag phase from the

transfer to the poor carbon source. Then the cells experience the carbon poor exponential growth. Once the cells are transferred to a carbon rich environment, the cells experience another transition period before entering the exponential phase of the carbon rich environment.

Figure 3.3 shows the OD measurement of the experiment, where the two exponential sections of growth are easily identified, and return doubling rates of 31.17 and 20.83 min for M9 and LB respectively. Once the culture was transferred to a new growth media ($\approx$ 375 min post-inoculation), there is a clear period of time where the OD is unchanging. The DNA distributions that correspond to each section are presented in Figure 3.4 along with the CH simulated DNA distributions. For the transition period, an infinitely large doubling rate was used to illustrate that this classic simulation method cannot reflect the measured DNA distributions, while the other two were reasonably simulated using the CH model with the doubling rates extracted from these sections of the growth curve. Later, this shift-up experiment will be explored using the injection growth strategy.

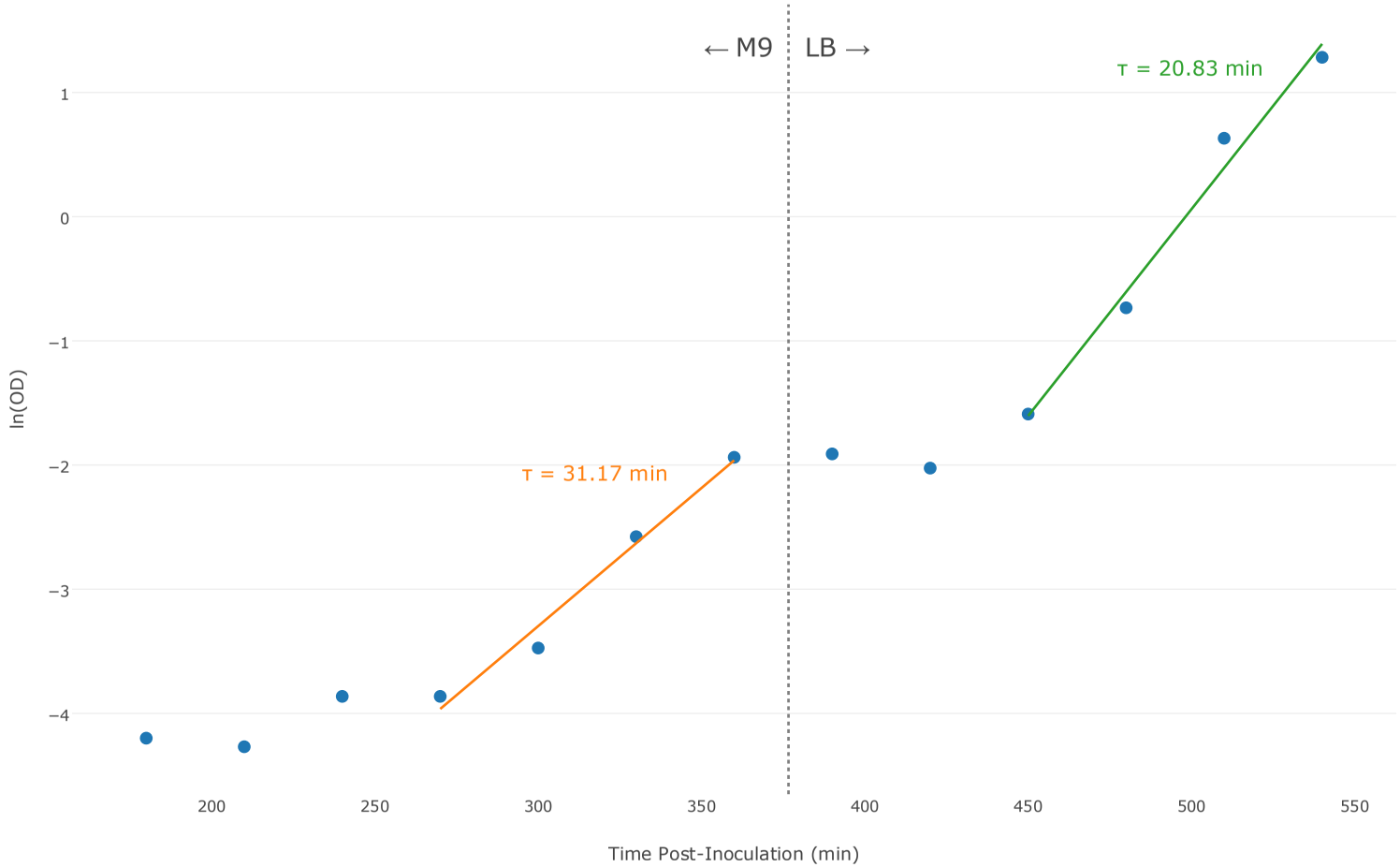


Figure 3.3: Nutritional shift-up experimental OD results. WT K-12 bacteria where initially grown in M9 minimal media supplemented with 20% glucose, and then transferred to LB ($\approx$ 375 min, dotted line). The population experiences two distinct exponential growth conditions where they were fit with first order polynomials (orange and green lines). Their slope is used to calculate the growth rate of 31.17 min doubling time for M9 and 20.83 min doubling time for LB.

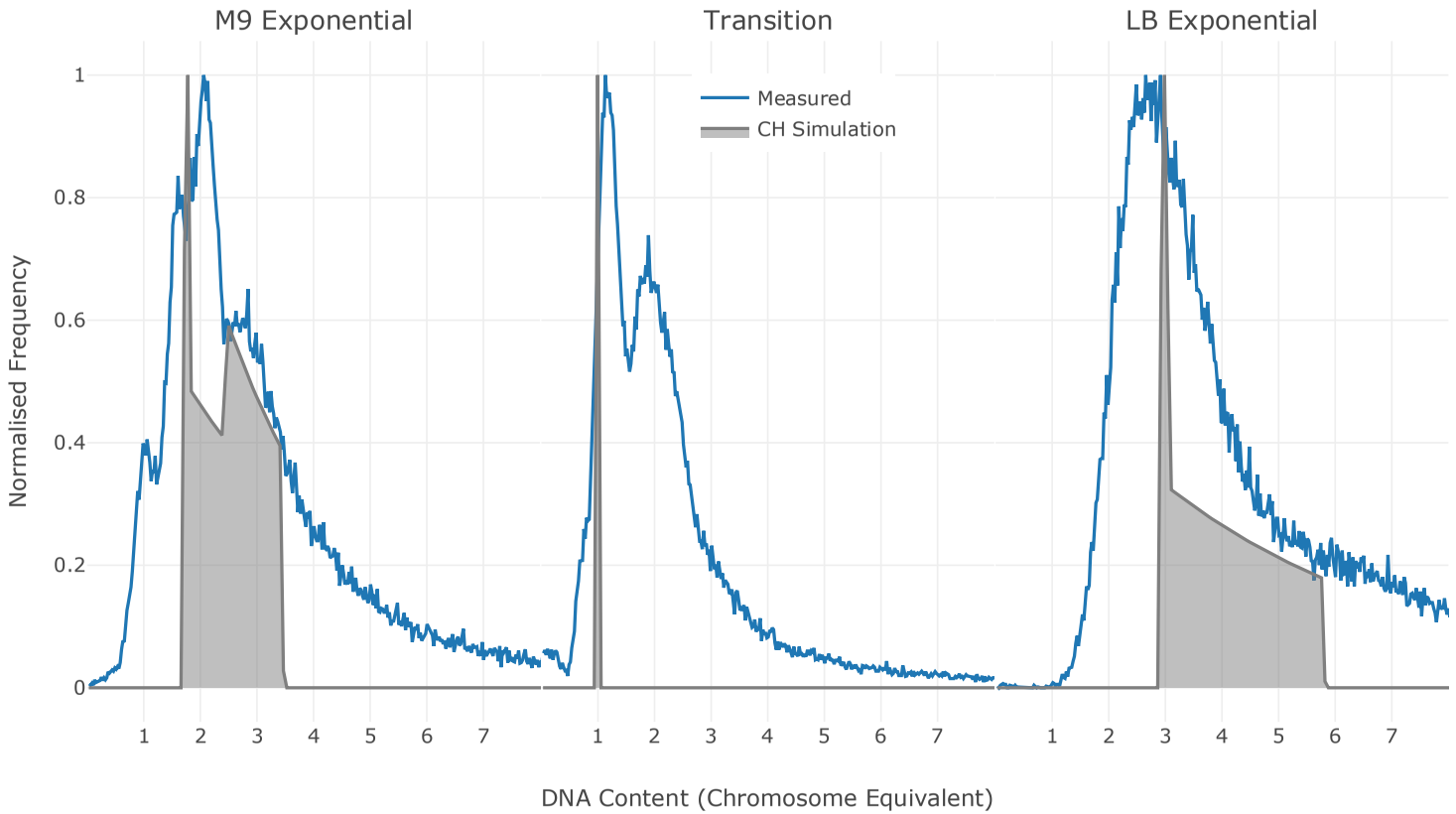


Figure 3.4: Flow cytometry measured compared with theoretical CH DNA distributions from the nutritional shift-up experiments (Figure 3.3). In blue lines are the measured DNA distributions where grey filled lines are the simulated DNA distribution from the CH model with C and D of 40.0 min and 20.0 min respectively with 31.17 min, infinite and 20.83 min doubling time.

3.2 Chromosome Dynamics

3.2.1 Critical Mass and the Eclipse Period

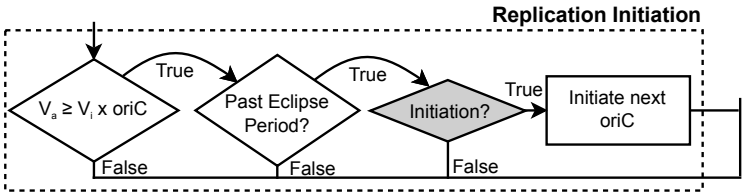


Figure 3.5: Flow chart describing the replication initiation algorithm for a single cell in HMG. When the volume of cell is larger or equal to the factor of critical mass and the sum of *oriC*, then the HMG attempts to open a new replication fork. If all previous replication forks are open, the algorithm checks if they are all past the eclipse period (see Section 1.2.1) or if there are no opened replication forks, and then the next replication fork is opened stochastically (grey box) (see Table 3.1).

The concept of critical mass is based on an observation made by Donachie [1968], where the ratio of cell mass to replication origin is a constant. The general consensus remains that critical mass is independent of the cell's growth rate, (Keasling et al. [1995], Osella and Lagomarsino [2013], Osella M [2017]), even if some studies report that the initiation mass does change in different growth conditions (Bates et al. [2005], Boye E [2003], Wold et al. [1994]). The literature consensus was accepted, that the critical mass is indeed an accurate estimate of the ratio between DNA content and volume of the cell and is independent of the cells growth condition. Implementation of critical mass was performed when the following condition is met:

$$V_a \geq V_i \cdot \sum |oriC| \tag{3.10}$$

where a is the time age of the cell, V_i is the critical volume constant and $|oriC|$ is the sum of replication origins for a single cell. If the above condition is met a new replication event would occur, where each available *oriC* opens for a new replication event. The observations of constant mass suggested a mechanistic function of control of initiation related to its size, and this equation proposes that all *oriC* open synchronously and non-randomly (Koppes and Von Meyenburg [1987]). For synchronous replication, this constant elucidates the interplay between cellular mass and DNA content, and formally explaines the observation that smaller cells had less DNA content than larger cells. The central role of DnaA and other molecular mechanism such as RIDA as the molecular machinery behind this theory, was only later

uncovered (see Section 1.2.1) (Amir [2014]).

Some reports suggest that initiation of an *oriC* occurs once and only once per cell cycle. Where, upon initiation, an *oriC* may not initiate once again even if the condition in Equation 3.10 is met (Messer [2002], Robert [2015], Boye et al. [2000], Wang et al. [2010], Skarstad et al. [1986]). The combination of sequestration of the *oriC*, titration of DnaA, and RIDA suggests that *E. coli* has a molecular mechanisms by which it limits the initiation of its chromosome (Boye et al. [2000]). Single cell analysis shows that multiple initiations within a single cell cycle has been observed and thus must arise more than once per cell cycle on some occasions (Zaritsky et al. [2011]). Thus in the HMG, no such restriction was made, where an *oriC* is able to replicate twice (or more) within the cell cycle's lifetime if the conditions of Equation 3.10 is met. However, because of the existance of many negative regulators in the initiation of replication (see section 1.2.1) the "eclipse" period was implemented (see Section 1.2.1). This is based on the observation that a minimal time is required before that same origin of replication is able to be re-initiated after initiation, and avoid the unwanted situation where two replication bubbles would run into one another. The manner in which we implement the eclipse period is based on empirical observations that the earliest observed reinitiation happens once the chromosome has replicated at least 33% of its chromosome (Browning et al. [2004]). Thus, using a rule based method, if the conditions are met where critical mass is reached and all chromosomes have reached this 33% threshold, then another round of initiation is made possible. A summary of the steps in the algorithm can be found in Figure 3.5.

3.2.2 Replication

Although some studies suggest that the rate of replication is constant for a given strain and species (Abner et al. [2014], Zaritsky et al. [2012]), measured C rates in different growth conditions and thus growth rates indicate that this is not the case (Keasling et al. [1991], Allman et al. [1991]). Indeed, concatenation of measured and calculated MG1655 strain C rates from four different sources plotted together in Figure 3.6 suggests that constant replication time is an oversimplification. Keasling et al. [1995] reports that the rate of C, is a function of population growth rate in the form of a one-phase exponential function. Where when $\tau \leq 60$ min, the replication time of bacteria is 43.2 min and at slower growth rates the time for a chromosome to replicate increases exponentially:

$$C = 43.2 \cdot (1 + 4.86 \cdot e^{\frac{-4.50}{\tau}}) \tag{3.11}$$

The logic behind this equation makes instinctive biological sense: under fast growth (doubling rate lower than $\approx$ 60 min) the cell contains all the macromolecular content to saturate the replication machinery. The rate limiting step in this case becomes the replication itself, and the speed at which the chromosome may incorporate nucleotides on the replicating DNA strand. To accommodate this rate limiting step in growth, bacterial cells undergo multiple rounds of replication on chromosomes that have not yet finished replication, also called overlapping rounds of replication (Cooper and Keasling [1998]). As a consequence, a chromosome that is actively replicating may be inherited to the daughter cells upon division. This permits the daughter cells to terminate the replication of a chromosome earlier than the time it takes to replicate a complete chromosome and thus have a doubling time that is less than that of the time it takes to replicate a complete chromosome (Cooper [2012]). Because the successful replication of at least one chromosome is required for division, if such a mechanism did not exist, then their doubling rate could never be less than that of the replication time. During slower growth, however, the replication machinery is not as saturated and thus the rate limiting step in replication becomes the availability of the macromolecular content of the cell that can be dedicated to the replication of chromosomes. From the measured C rates and measured growth rates, the relationship between the two is reported to be in the form of exponential decay in relation to the doubling rate of the cell (Keasling et al. [1995]).

Furthermore, there are many cases where different bacterial strains have different rates of replication. This may include different growth environments, or a mutant strain, or a combination of the two. For example, the K-12 AB1157 has a C rate of 55 min with a doubling time of 28 min, whilst the K-12 CM735 has a C rate of 44 min with a doubling time of 29 min (Allman et al. [1991]). Because replication involves a complex machinery of DNA polymerases, helicases and β-clamps and others (see Section 1.2.1), it can be expected that any slight difference in the efficiency of any of these proteins, or their interaction, may have a large downstream effect on the rate of replication.

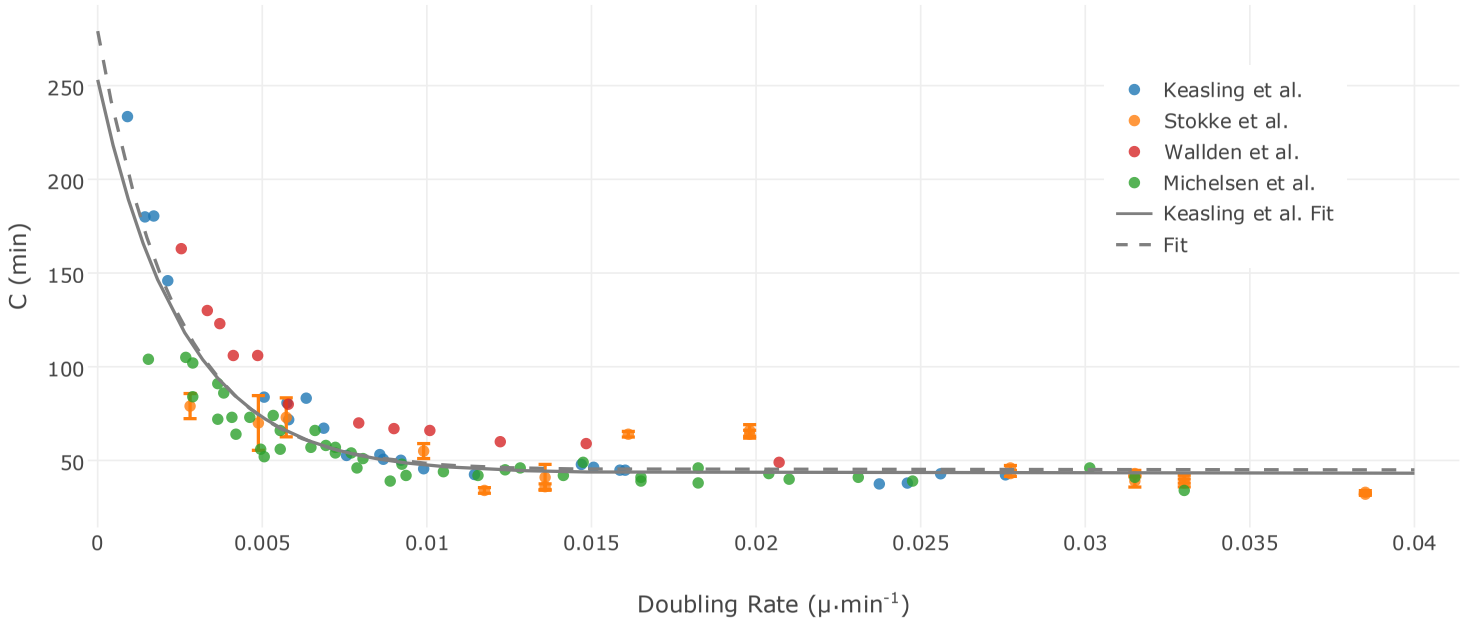


Figure 3.6: Concatenation of all the literature C rates of K-12 MG1655 cells grown under various growth conditions, and measured using different methods (circles) (see corresponding papers). The solid line was taken from Keasling et al. [1995] , the dotted line corresponds to the best fit to the data (circles), where $C = 44.98 \cdot (1.0 + (5.206 \cdot e^{\frac{-4.917}{\tau_h}}))$.

Synchronous and Asynchronous Replication

For WT cells, the initiation of replication is synchronous on all *oriC* available at the time of critical mass, such that the chromosome copy number for an individual cell always follows a 2^n copy number (where n is a non-negative integer) (Katayama et al. [2010], Skarstad and Boye [1993a], Brewer and Fangman [1991]). However, reports of mutant bacterial strains reveal that this rule does not always hold, where either long-term incubation (Akerlund et al. [1995]), or drug treatments demonstrates that some populations may contains cells with 3, 5, 6 etc... copy numbers (Allman et al. [1991], Boye et al. [1988], Skarstad and Boye [1993a]). A review of the literature shows that many different strains express this phenotype is known as the asynchronicity of replication (or aberrant chromosome copy number) (Boye et al. [1988]). Knockout experiments have also shown that many different proteins are important actors in the process that yields the WT synchronous phenotype (Boye et al. [1988]). Among the proteins where the mutant forms have been attributed to losing the synchronous timing of replication include DnaA, Dam, RecA, SeqA, DatA and others (Boye et al. [1988], Skarstad and Boye [1993a], Katayama et al. [2010]).

Globally the types of mutations that affect this phenotype can be classified into two distinct categories: the ones that are involved in the initiation machinery and the ones that are involved in faithful replication of the chromosome. The former category includes either mutations of the mechanisms involved in the eclipse (see Section 1.2.1), such as the SeqA protein, as well as mutations in the initiation of replication itself such as DnaA, that causes either random opening of *oriC* or over-initiation (Boye et al. [1988], Skarstad and Boye [1993a]). Because chromosome replication is by nature error prone, and because cells are constantly exposed to stresses that cause DNA damage such as oxidative stress and UV radiation, bacterial cells possess different mechanisms to manage damage to their DNA, including excision repair, mismatch repair, and the SOS response system (Courcelle and Hanawalt [2003]). For repair involving templating by homologous DNA sequences, the RecA protein plays a central role in matching the damaged DNA to its complementary sequence (Courcelle and Hanawalt [2003], Kuzminov and Stahl [1997], Michel [2000]). RecA not only alleviates repression of DNA repair protein expression through a co-proteolytic function that cleaves the repressor LexA, but upon coating single stranded DNA, RecA catalyses the pairing to a complementary strand (Courcelle and Hanawalt [2003], Michel [2000], Little et al. [1980]). A particularly popular mutated version of this protein called recA1 has been observed to be active active, albeit at a reduced rate than its WT counterpart (Goodman [2000], Allman et al. [1991], Cox et al. [2000], Courcelle and Hanawalt [2003]).

Aberrant Chromosome Copy Number Case Study: recA1

The RecA is often mutated in laboratory strains used for the purpose of genetic engineering (such as the TOP10 cell type). In the case of recA1, although still functional, this version of the protein has an impaired ability to perform recombination of inserted plasmid genetic material and as a consequence this increases the stability of the inserted plasmid (Kuzminov [1995]). Unsurprisingly, such bacterial strains are particularly sensitive to any type of DNA damage, and during normal laboratory growth, a large population of anucleate cells may be observed under normal cultivation methods (Skarstad and Boye [1993a], Kuzminov [1995]). A few key observations are particularly informative of the consequences of this RecA mutation. Early studies demonstrate that although recA1 cells contain aberrant chromosome copy numbers, timing and coordination of initiation is shown to not be affected when exposed to high doses of UV (Skarstad and Boye [1993a], Zyskind et al. [1992]). Moreover recA1 bacteria contain a higher number of free floating nucleotides than their WT counterparts, and this phenotype is exacerbated by faster growth rates (Horii and Suzuki [1968]). In vitro experiments show that WT RecA inhibits the nuclease activity of RecBCD, and knock out experiments for this protein are more lethal to the cell than are RecA knockouts (Kuzminov and Stahl [1997]). If, for any reason, the replication fork is arrested long enough, then the arrest leads to double-stranded breaks, and RecBCD is recruited and degrades the replicating strand (Courcelle and Hanawalt [2003], Kuzminov [1995], Kuzminov [1999]). Furthermore, there seems to be a RecA-independent damage avoidance mechanism which involves suppression and removal of damaged strand by RecBCD during replication, which leads to the collapse of the replicating strand, and as a consequence the chromosome returns to its original form (Miranda and Kuzminov [2003], Kuzminov [1995]). Other evidence suggests that recA1 mutant bacteria also experience whole chromosome degradation (Skarstad and Boye [1993a]). Finally, the lack of functional RecA may affect the synchronous segregation of chromosomes at division, such that a given complement of chromosomes is divided less evenly between daughter cells in recA mutant strains (Zyskind et al. [1992]).

Altogether, these observations led to summarise these recA-associated mechanisms for inclusion in the model in the following way, illustrated in Figure 3.8 dashed boxes. If the replicating strand encounters DNA damage (which it is unable to repair due to the lack of a functional RecA), then a double stranded break occurs. If the break occurs upstream of the replication fork, then RecBCD degrades the replicating strand until it reaches the end of the other replication fork (Kuzminov [1999]), which leads to restoration of the replicating chromosome to its original

state. If the double-stranded break occurs downstream of the replication fork, then RecBCD degrades the whole chromosome. A concise mechanism to capture the impact of RecA mutation was formulated and implemented in the cell cycle model, which is summarised in Figure 3.7.

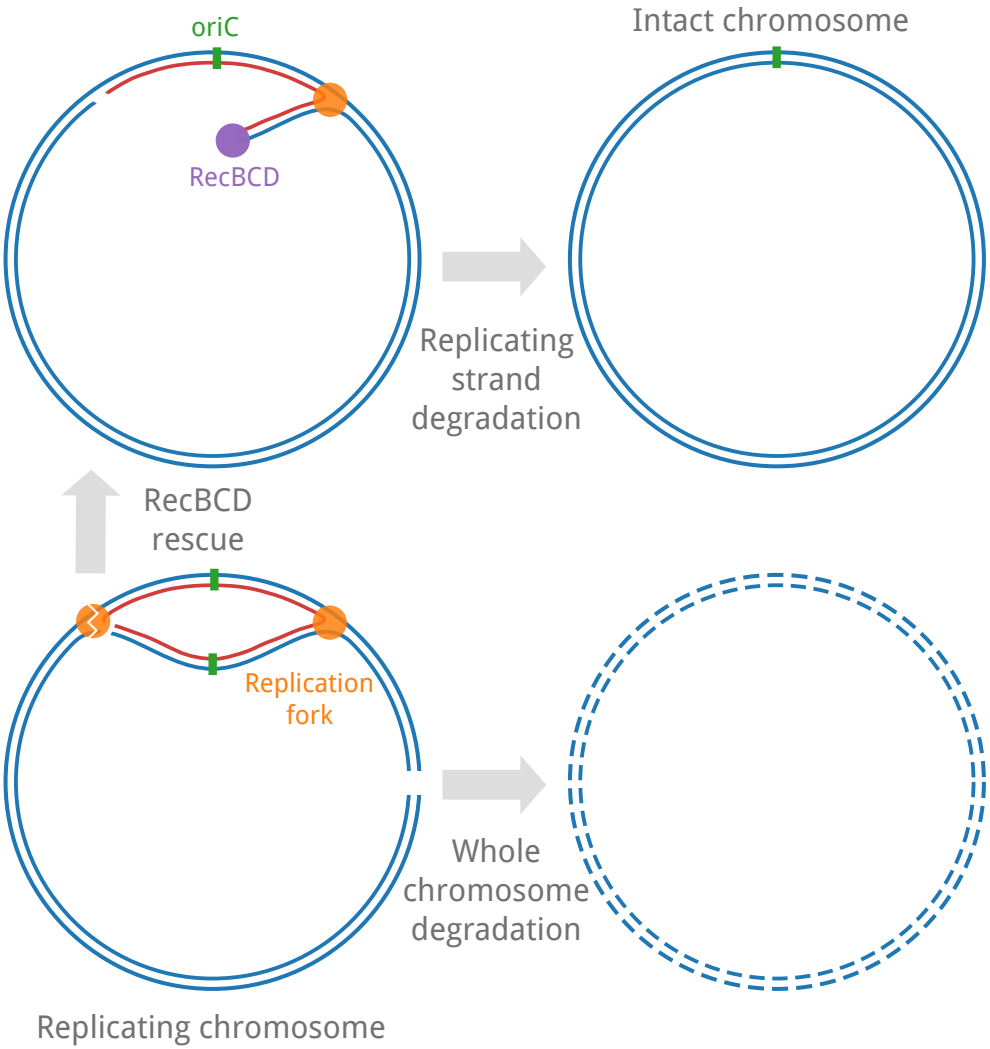


Figure 3.7: Summary of the simplified description of DNA double stranded damage responses used simulations to represent the consequence of RecA mutation. Green bars are the *oriC*, orange circles are the replication fork bubbles, the purple circle is the RecBCD enzyme, each blue line is a single chromosome strand, and each red line represents a newly replicated DNA strand. As described in the text, the simulation captures two potential outcomes following DNA damage. In the first scenario, a double-stranded break occurs at the replication fork, which leads to the collapse of the replication fork (broken orange circle), which is subsequently rescued by RecBCD, returning the chromosome to its original form. In the second scenario, the double-stranded break occurs downstream of the replication fork, which instead leads to degradation of the whole chromosome.

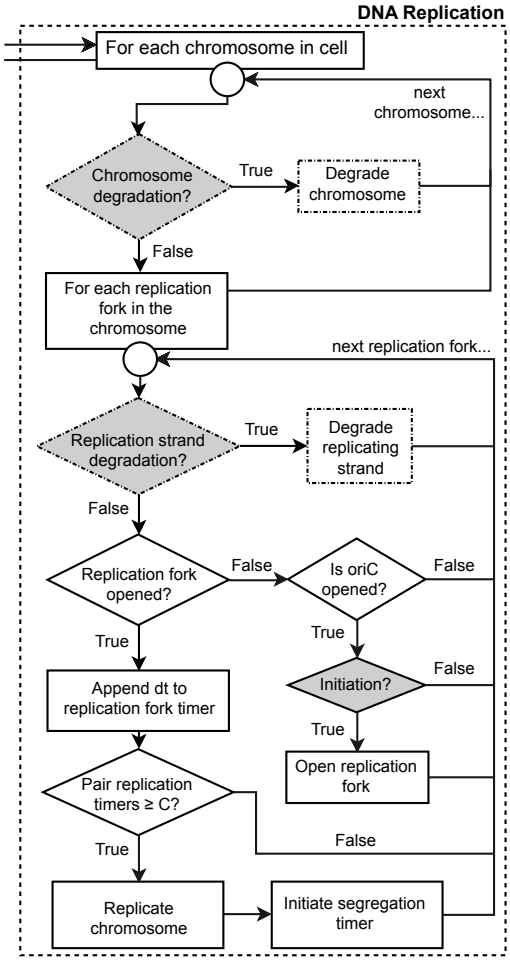


Figure 3.8: Flow chart describing the replication algorithm for WT and the aberrant chromosome copy number phenotype (see Figure 1.3) in HMG. The grey boxes represent steps in the algorithm with some degree of stochastic noise (see Table 3.1). The dashed lines represent the steps only applicable for cells that express the aberrant chromosome phenotype (TOP10 mutant), which includes whole chromosome degradation and replication fork collapse. As the algorithm loops through all chromosomes and replication forks it advances the timers by a pre-determined time step (dt). If any of these timers exceeds the replication time (C), then the chromosome is replicated and the pair of replication forks were reset and the segregation timer is initiated.

3.3 Cell Division

Early quantification of size within populations of bacteria in balanced growth showed that a given growth rate produced populations with narrow size distributions (Schaechter et al. [1958]). These findings suggest that there must be a "conserved mechanism responsible for coordinating cell composition and size with growth rate" (Vadia and Levin [2015]). The division mechanism of a bacterial cell into daughter cells is central in ensuring size homeostasis within a population, and indeed, the "very existence of a stable cell size distribution indicates the presence of intrinsic mechanisms that reduce cell size fluctuations" (Campos et al. [2014]). Many reports that investigated the size distribution of populations show that, given a growth environment, bacterial cells have a narrow distribution of sizes (Koch [1993], Sauls et al. [2016]). This is perhaps the most important feature of a population of cells, since virtually all extensive parameters are dependent on the control of division. An accurate account of how a single cell makes the decision of dividing is central to HMG. If one assumes that the critical mass theory of chromosome initiation is valid, the distribution of bacterial size would inevitably dictate the control of the DNA content of a cell. Three different theories of the control of division are the sizer, timer and adder models are presented here.

3.3.1 Sizer model

The first model of division that guaranteed this conservation of size proposed that a cell divides once it reaches a threshold size (independent of the critical mass for chromosome initiation), and was inspired by the empirical observation that a cell divides once it doubles its size (Cooper [1988]). One assumption of this "sizer" model is that a smaller cell at birth would grow for a longer time than one that is born larger, so as to accumulate enough mass to double. Another implied assumption is that the cell contains a molecular mechanism that allows it to "know" its size and consequently can divide once the size threshold is met. Even though a few molecular mechanisms and potential individual proteins have been speculated to play a role in this size detection mechanism, no consensus has so far been reached on the subject (Robert et al. [2014]).

Despite this, this theory of division is used in modelling yeast and mammalian cells, where it is believed that the commitment to division occurs upon reaching a threshold size (Campos et al. [2014]), recent single cell quantitative measurements of bacteria show that there is a strong statistical correlation between the size at birth and the size at division, in contrast to the sizer model that implies that the size at

birth and division have no correlation, since the size at division is constant (Campos et al. [2014], Wang et al. [2010], Taheri-Araghi et al. [2015]). Furthermore, the size of the mother and daughter cells seem to be correlated (Campos et al. [2014]).

3.3.2 Timer model

This model assumes that cell division is controlled by its age alone, where the cell divides after a predictable time following birth. The development of the model was based on the empirical observation that a population of cells growing in a constant environment had a predictable time from birth to division (Cooper [2012]). This theory was very much inspired by inferring that the doubling time of individuals in the population, and the start of division within single cells, is tightly controlled by the age of the cell and not its size (Robert et al. [2014]).

Unlike the "sizer" model of division, it does assume that there is a strong correlation between the size of the cell at birth and division. However, the question remains as to how population size homeostasis is maintained if all members of the population have a strict doubling time. Under such a rule, larger cells would remain larger than their smaller counterparts as they grow for the same amount of time and rate. Indeed, the lack of control of their size with such a mechanism would theoretically require the adjustment either of the elongation rate or of the divisional time to compensate for this discrepancy if the population size homeostasis is to be maintained. If not, the population would inevitably drift, and the simulated distributions using this division rule show that they are too wide compared to measured distributions (Robert et al. [2014]). Single cell analysis reveals that a timer model with a strict global doubling rate is not mirrored in measured bacterial populations in balanced growth, where larger cells grow for a shorter time than their smaller counterparts on average (Campos et al. [2014]).

3.3.3 Adder model

With the advent of microfluidic devices, better confocal microscopes and better image analysis software; accurate time-lapse tracking of single cells is greatly improved, making tracking of bacterial lineages possible (Taheri-Araghi et al. [2015], Campos et al. [2014]). This generated a new set of data that is much more accurate at the single cell level, and in a constant environment thanks to microfluidic devices (Campos et al. [2014], Taheri-Araghi et al. [2015]). A similar technique used in the past included the use of agar pads (Bennett and Hasty [2009]). Although this also enabled single cell tracking, these would essentially constitute batch culture situa-

tions, while microfluidic mother machine devices permit the continuous culture of cells and the generation of more rigorous datasets that allow the accurate measurement of cells without environmentally induced fluctuations (Campos et al. [2014]). With this new data some of the assumptions that were made about the division patterns of bacteria could be confirmed or revisited.

This new scrutiny of the bacterial cell cycle revived an old theory of cell growth called the "adder" model of growth. In short, it proposes that population homeostasis may be maintained if one adds a constant amount of volume from birth to division at every cell cycle. This was first argued theoretically by Voorn et al. [1993], and again by Amir [2014], and later experimentally investigated upon single cell population analysis (Campos et al. [2014], Taheri-Araghi et al. [2015]). The major observation that led to the validation of the adder model was the statistical link between the size at birth and the size at division. Indeed, it seems that the size at division is dependent on the size at birth which, as we have seen, refutes the "sizer" model.

One of the most interesting implications of the "adder" model is the disappearance of the D period, where the initiation of the division steps is no longer determined by the termination or initiation of replication; as the CH model seems to suggest. Indeed, where the C period for example lends itself much more easily to time control, defining segregation of chromosomes and division of the cells with time control seems to be limiting since the recorded times are very variable (Adiciptaningrum et al. [2015]). And indeed, concatenation of D times from the literature into a single plot (Figure 3.9) makes for much more variable results than that of the C times (Figure 3.6), making the generalisation of the D period much more difficult (Zaritsky [2015]). More experimental evidence would be required to uncover the molecular mechanism and actors for the control of bacterial division.

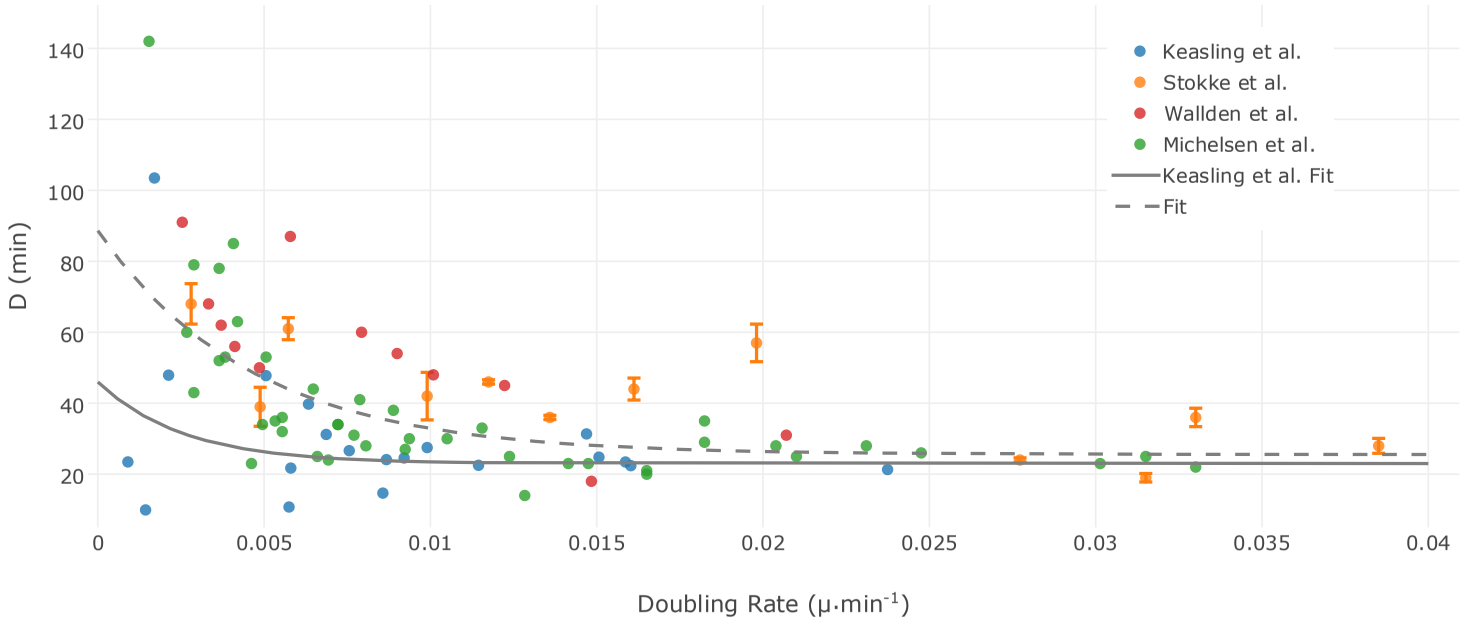


Figure 3.9: Concatenation of the measured D periods for the bacterial strain K-12 MG1655. The solid line is the one-phase exponential function as reported by Keasling et al. [1995], while the dotted line is fit to the same functional form with the reported data. We note that the data is much more spread than 3.6, where the accuracy of the fit or the literature functional forms are both not very accurate.

3.4 CH Model: Linking Chromosome and Division Dynamics

3.4.1 Mixed Timer and Sizer Model

The most popular and most robustly studied model of the cell cycle is the Cooper and Helmstetter model (CH). The authors pose that the cell cycle can be formalised using a combination of the timer and sizer models, whose steps can be summarised in the following way: a cell grows until it reaches a particular size, called the critical mass, that is determined by the Donachie [1968] constant ratio between the chromosome copy number and mass of the cell. From that point on, it becomes a timer model, where the cell continues to grow for $C+D$ amount of time, where C is the replicative time and D is the segregation and division time (see introduction for complete mathematical description of the CH model).

Termination of Replication Starts Division

Although the link between volume and DNA is robustly described through the critical mass ratio, supported by the central role of the DnaA protein, there is much more uncertainty as to the mechanism of control and timing for the start of division (Taheri-Araghi et al. [2015]). Early experimental evidence suggests that inhibiting DNA replication in turn inhibits replication (Maruyama and Lark [1961]), and supports the idea that the two processes are closely linked. Some studies have proposed that initiation of replication is the trigger for the control of division (Ho and Amir [2015]). Measured timing of FtZ ring formation in different strains and growth conditions seems to suggests that the formation of the roughly corresponds with the start of the D period as defined by the CH model in wild type cells (Den Blaauwen et al. [1999]). However overwhelming numbers of mutant perturbations demonstrate that the timing of replication does not affect the division timing of a cell, where perturbations to the replication initiation pattern have no effect on the Ftz ring formation outcome and the resulting population size distributions (Bernander and Nordström [1990], Tropini et al. [2014]). When, for example, cell dimensions were perturbed, either by Ftz or mreB titration, the replication time was not affected while the D time increased monotonically with increase in size (Zheng et al. [2016]). Although there is certainly a mechanism of control of division by replication, or else one would observe many more anucleate cells, the degree of the link doesn't have to be as strong or as direct as the one posed by the CH model. Because of the lack evidence as to what controls the timing, this question remains open and requires

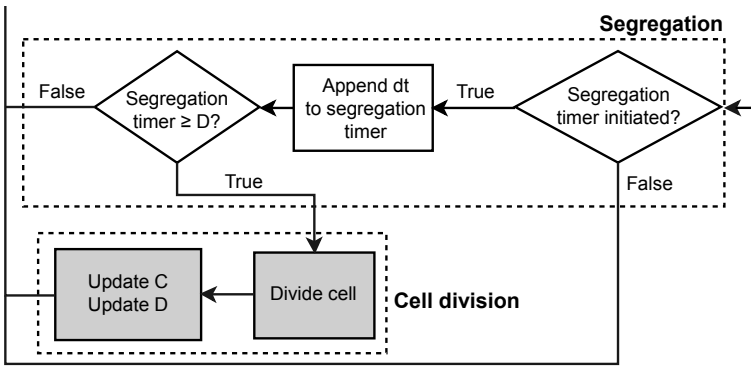


Figure 3.10: Flow chart describing the segregation and division algorithms in HMG. The grey boxes represent the steps where stochastic noise is applied (see Figure 1.3). Once the segregation timer has been initiated by the completion of at least one chromosome replication (see Figure 3.8), then a pre-determined time step (dt) is appended to the segregation timer (D). Once that timer exceeds the D time, then the cell is divided where the DNA content and D timer reset. Lastly the cells replication time (C) and D parameters are re-generated.

more research.

Although not explicitly stated, the CH model implies that the termination event of a replication triggers the segregation of the chromosomes and the constriction of the mother into the daughter cells. For example, consider a cell that has a doubling rate of 30 min, C of 40 min and D of 20 min. This situation would lead to overlapping rounds of replication where a cell inherits an already replicating chromosome and initiates a new round of replication before the current chromosome finishes. The resulting replicating time is 10 min; combined with the constant D time of 20 min leads to a doubling time of 30 min. It is clear that the termination of replication starts the D timer. If initiation was the trigger and D alone was the timer for segregation and division then the cell would have a doubling time of 20 min, which is not the case. Leaving aside the timings of the CH model, if indeed initiation of replication was triggering the situation of overlapping rounds of replication, initiation of replication would imply that the mother cell passes to its daughter cell the information to start the division, a scenario that does not seem to be likely.

One of the advantage of critical mass dictating the cell division in this fashion is that it solves the problem raised in the previous section of the timer model population drift when the cell age is the only factor dictating the growth dynamics. For example, consider a population in balanced growth with a doubling time of 70 min, C of 40 min and D of 20 min. In this scenario, a cell at any given time has a chromosome content of $1 \leq \text{chromosomes} \leq 2$ and thus has no overlapping rounds

of replication. Assuming that the critical mass is 1.0 (in this case this would be a theoretical critical mass, where it may be multiplied into any true volumetric equivalent), in a completely deterministic approach, the cell would be born with a volume ratio of 0.906, and thus the cell requires the accumulation of 0.094 before initiation and subsequent division may occur (that would take 10 min if the doubling time is to be respected). If division is stochastic, a cell that is born with more than 1.0 of volume would initiate replication upon birth, while a cell that is born with a mass of less than 1 requires time to grow before it is able to achieve the required mass to initiate replication (also called the B period of growth). The result is two different doubling times, the larger cell where $\tau = C + D$ and the smaller cell where $\tau = B + C + D$, and the population's volume would converge to a single mean. In short, the CH model assures a size homoeostasis by compensating for the drift that would occur with age-based division by correcting the doubling time of the cell. A summary of the steps in the algorithm of cell segregation and cell division may be found in Figure 3.10.

3.5 Conclusion: The Model

3.5.1 Flow Chart

Most simulations of bacterial population dynamics start with a simple model of growth, typically framed at the single cell level, which is then expanded to predict the growth dynamics of populations (Stokke et al. [2012], Michelsen et al. [2003], Keasling et al. [1995]). The objectives of this dissertation are not well served by this approach, as it would be useful to have a model that operates in the reverse direction-starting from simple, experimentally measured growth curves (OD vs time), such a model would enable one to infer the growth dynamics of the individual cells within such a population. To achieve this goal of describing chromosomal dynamics across a heterogeneous population, an individual-based simulation framework was designed, termed the heterogeneous multiphasic growth (HMG) simulator. This framework comprises of two distinct innovations: an "injection growth" mechanism and a novel individual based description of the bacterial cell cycle.

At the single cell level, the CH model approach is employed for formalising the bacterial cell cycle, where the chromosome dynamics are dictated by the critical mass and advanced using a timer (see Figure 3.12 for a complete summary of the single cell cycle algorithm). The eclipse period was implemented where a new pair of replication forks could not initiate until the previous ones are 33% completed (Browning et al. [2004]). The division dynamics are governed by the termination of

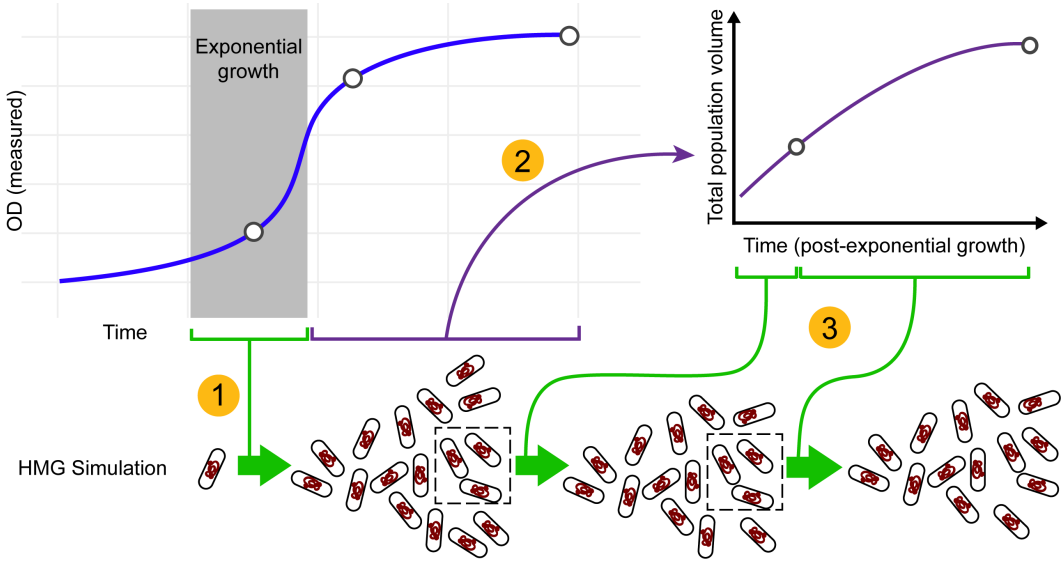


Figure 3.11: Injection-based strategy for connecting the HMG simulator to empirical growth data. This cartoon summarizes the process by which empirical growth data (example, a measured OD vs time curve) is used to "drive" the HMG simulator via the volume injection method, where the open circles represent the sections of the growth curve where DNA distributions were measured. Thus, in this illustration, the simulation would contain three independent steps: (1) The region of exponential growth is identified. This exponential growth rate is used to drive the HMG simulation from a single cell inoculate to a diversified population of exponentially growing cells; (2) During post-exponential growth, the OD curve is used to calculate the rate at which the overall cell volume (of the population) is increasing; (3) At each time point, the calculated rate of volumetric change (per cell) is "injected" into each cell in the population, each of which advances its cell state via the HMG algorithm outlined in Figure 3.12. The dashed rectangles indicate that during each time step of the simulation, a random subset of the population is taken forward into the subsequent time step of the simulation in order to keep simulations computationally tractable. Taken from du Lac et al. [2016].

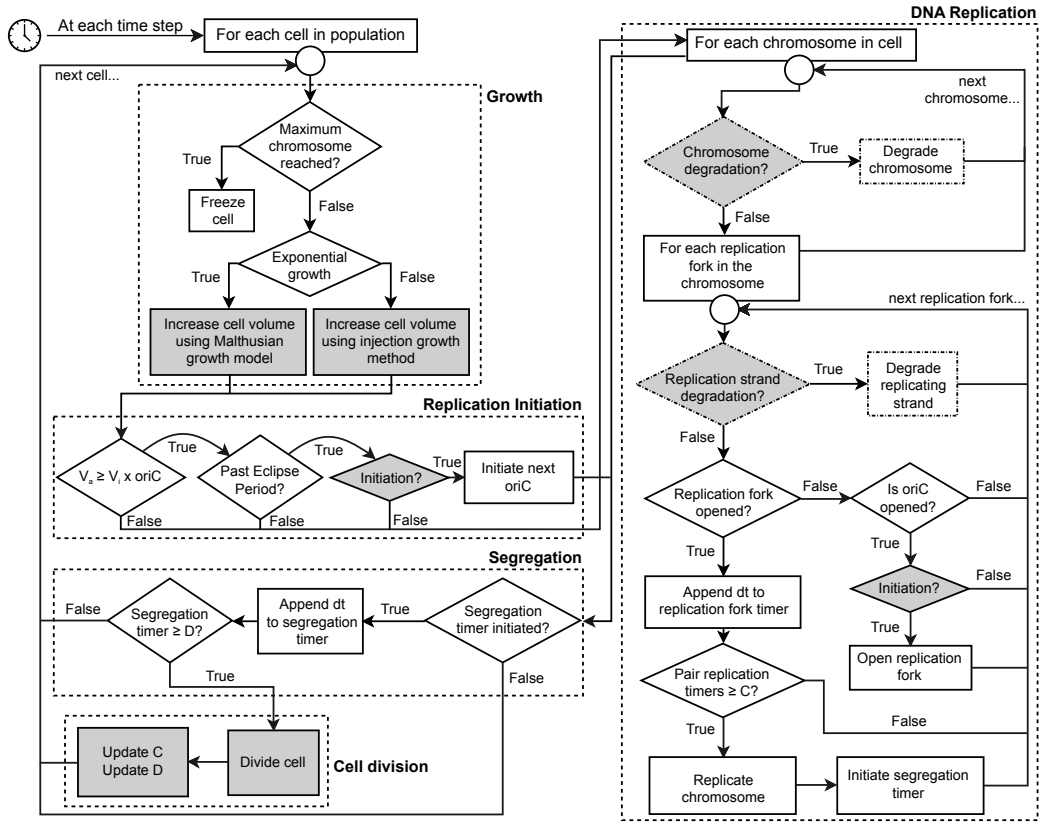


Figure 3.12: Flow chart of the single cell model. More descriptive explanations of the different steps may be found in Figure 3.2, 3.5, 3.8 and 3.10. The model follows the CH model with the addition of the injection growth method, the DNA degradation from a mutated recA1 and the eclipse period. The greyed out parameters represents the paramters that include a stochastic element, as described in Table 3.1. Taken from du Lac et al. [2016].

at least one pair of replication forks, and progresses using a timer (D phase). To model for the asynchronicity phenotype for cells that contain a RecA mutant, the model implements a stochastic possibility of a chromosome experiencing replication fork collapse as well as whole chromosome degradation.

For population growth, the raw OD data was used as input, fitted with a B-spline interpolation method with a smoothing parameter that minimises the measurement error, while providing a smooth outcome (Figure 3.11) (Zhang et al. [2006]). Using the minimizing function in Equation 3.9, the section of the growth curve where balanced growth applies is identified. From the doubling rate calculated from that window, the model assuming Malthusian (exponential) growth is populated. The rest of the growth curve was then selected, either until the end or until the OD starts to decrease. Indeed, since implementing cell death is beyond the scope of this research, the death phase of growth cannot be accounted for. Nonetheless, the selected portion of the growth curve that contains the transition phase and the stationary phase is converted to its volumetric equivalent and normalised to the balanced growth seed population. The simulation was then advanced, distributing the volumetric changes calculated from Equation 3.6 equally among all members.

Using this method, it is postulated that it is possible to simulate a population of cells outside of balanced growth. Indeed, the generation of the seed population using Malthusian growth theory generates a population of cells whose distribution has been extensively covered, and is atemporal. Using this population the model is able to progress through the other phases of growth and reflect the population in a temporal manner, where the distribution of the population is dependent on the distribution of the previous population in time.

3.5.2 Inclusion of Stochastic Effects

One of the important characteristics of an IBM simulation strategy is the ability to reflect the population heterogeneity in an emergent manner. This requires the inclusion of a degree of stochasticity in the parameters of the single cell model in a method that reflects the measured distributions. Population rates of replication (C), division rates (D), the binary fission ratio between daughter cells, and the growth deviations, all show distributions that are normally distributed (Keasling et al. [1995]). This makes sense under the central limit theorem, where each of these parameters imply a multitude of actors and thus would result in Gaussian distributions (Taylor [1997]). The initiation of replication also contains stochastic elements that need to be implemented in a different manner due to the way the model is constructed. Indeed, there are recorded differences in timing of roughly

Term	Definition	Value
V_i	Volume at initiation (also called critical volume)	$0.9\mu m^3 \equiv 0.9fL$
σ_{V_i}	Gaussian noise sd	10% for dt = 0.01min
V_a	Cell volume at cell cycle time a	N/A
σ_{V_a}	Gaussian noise sd	5% for dt = 0.01min
μ	Growth rate (min^{-1})	N/A
τ	Doubling time (min)	N/A
C	Genome replication time (min)	N/A
σ_C	Gaussian noise sd	5% for dt = 0.01min
D	Genome segregation time (min)	N/A
σ_D	Gaussian noise sd	5% for dt = 0.01min
Division asymmetry	Gaussian noise sd describing the asymmetry of binary fission	10% for dt = 0.01min
Chance of initiation	Given that critical mass is reached, probability that the *oriC* opens	4.01% for dt = 0.01min
Chance of DNA damage	Probability that the replication fork experiences any damage	tbd
Ratio of DNA damage	If DNA damage occurs, ratio if that damage causes replication fork collapse or whole chromosome degradation	tbd

Table 3.1: Nomenclature of the different parameters in the model. N/A corresponds to the parameters that depend on the conditions of growth. tbd = to be determined. All simulation where performed with a time step of 0.01 min.

4.01% between replication fork formation at the initiation moment of two *oriC* (in synchronous populations, see Section 3.2.2) (Keasling et al. [1995]). The way this feature was implemented is as follows. Given that critical mass is reached, every *oriC* is flagged as being open for replication. Every *oriC* is then iterated through a random number which is sampled from a normal distribution with a mean of 0 and a standard deviation of 1. If that number is larger than 1.75 then that *oriC* is opened, and the iteration of the *oriC* continues until all *oriC* are opened. All other Gaussian parameters are sampled randomly at birth and division from a normally distribution with mean as determined from the population assignment and the standard deviation with the parameters listed in Table 3.1. For growth, under a Malthusian growth model, a random sampling of the growth rate at birth is used. For the injection growth model, however, an injection deviation parameter was implemented with a mean of 1 and with standard deviation as described in Table 3.1, sampled at the birth of the cell. Given the amount of volume added at a given time step, this parameter ensures that at the population level a cell grows by a portion of the total, so as to reflect the same population distribution as under Malthusian growth.

Chapter 4

Model Examination and Optimisation to recA1 Mutants

In this chapter the previously discussed developed model will be tested and explored. The first part will include simulations compared to measured WT K-12 DNA distributions of bacterial population along different phases of the growth curve using the injection growth method. Different aspects of the newly developed simulation method will be discussed and tested. The second part involves optimisation of the cell cycle model developed to measured population DNA distributions from a widely used mutant bacterial strain, TOP10, that contains a known mutation in the RecA protein that causes peculiar chromosomal dynamics. This aims at showing that using HMG to optimise for mechanistic features of the cell provides robust solutions since they are not only tested in a wide range of disparate growth conditions, but the problem is organised in the same way as the experimental measurements.

4.1 HMG Examination Experiments

4.1.1 Wild Type Cells

Aim

To test the ability of the HMG simulation protocol, to simulate for cell cycle properties of a population outside of exponential growth, DNA distribution of WT K-12 MG1655 bacterial cells was measured throughout different phases of batch culture as the population progressed from exponential to stationary phases, at two different shaking regimes. The different shaking rates provided large enough oxygenation differences that it had a consequence on their growth rate (Riedel et al. [2013]). Two

particularities of the model are investigated in detail.

Although these experiments and their corresponding simulations had for goal the global validation of the HMG simulation framework, by matching the DNA distributions of measured population to the output simulated cell cycle DNA distributions, two peculiarities of the model are investigated.

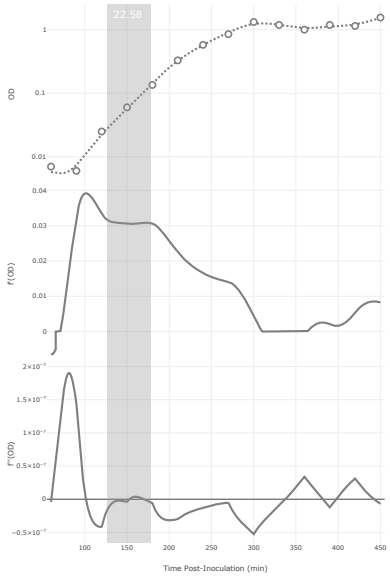
First, the injection growth method as an accurate method for the simulation of population growth throughout disparate conditions and phases of growth. At the time of writing, no experimental quantification of C and D parameters outside of balanced growth have been made. C and D functional forms as reported by Keasling et al. [1995] and supported by other experimental measurements presented in Figure 3.6 and 3.9 are estimated from exponentially growing populations.

Second, the C and D functional forms applying outside of balanced growth. Outside of these stringent growth regimes it is not unreasonable to assume that these rates do not hold. The second aspect tested is the injection growth method as a means of simulating growth of a population outside of exponential growth. Although the injection method by definition strictly follows the OD growth curve, there are some uncertainties as to how the volumetric changes are distributed. When under assumption of balanced growth (where the Malthusian theory of growth applies), the exponential function (Equation 3.1) implies that the volume increase is exponentially proportional to the volume of the cell (Wallden et al. [2015]).

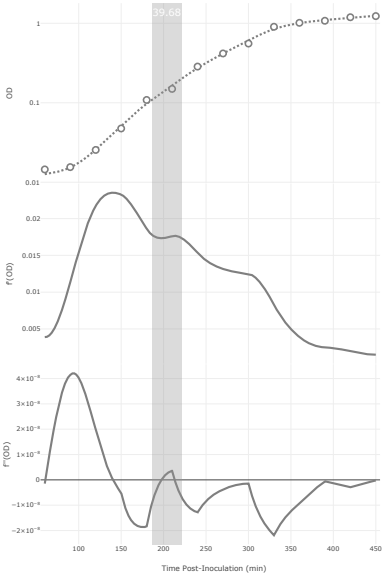
A further goal of this exercise is to inspect if formalising the heterogeneity of a population in non-balanced growth using the HMG simulation is accurate enough to reproduce measured distributions. DNA distributions are used as a proxy for the output of the cell cycle, and it is assumed that if one is able to reproduce the same DNA dynamics theoretically, one can describe the nature of the heterogeneity of the cell cycle of a population accurately.

Experimental Results

The measured DNA content of the population is presented in Figure 4.2. At 230 rpm, the first histogram that corresponds to the exponential DNA distribution ($\tau = 22.58$ min) returns a population containing between 3 to 6 chromosomes. As the population enters the transition phase, the DNA content progressively diminishes until 300 min post-inoculation, where the population contains a majority of cells with 2 chromosomes. After that time, the population can be said to be in stationary phase since the OD stagnated completely (Figure 4.1). Consequently, the frequency of initiation becomes rare and a diminishing number of cells were actively replicating their chromosomes causing the measured DNA content to fall to two distinct peaks with



(a) shaking rate of 230 rpm. Exponential growth was identified as the area between 125 and 180 minutes post-inoculation, and returns $\tau = 22.58$ min.

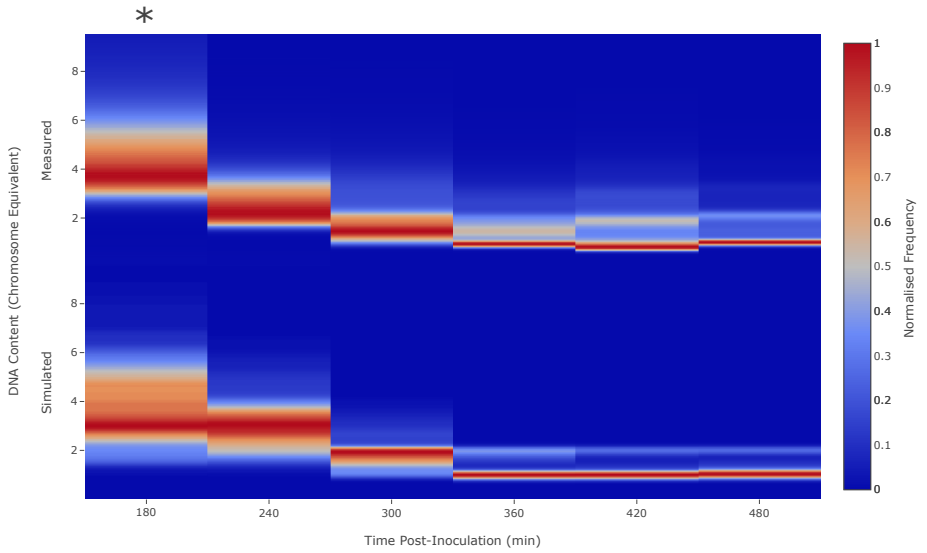


(b) shaking rate of 23 rpm. Exponential growth was identified as the area between 180 and 220 minutes post-inoculation, and returns $\tau = 39.68$ min

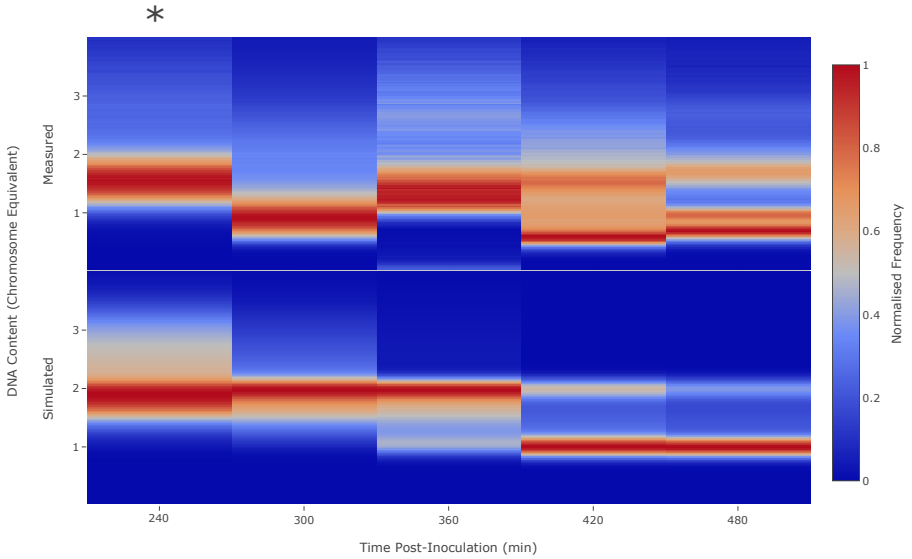
Figure 4.1: Spline fit to the measured OD from K-12 cells grown in LB with a shaking rate of 230 rpm and 23 rpm. Exponential growth applies when the first order differential stagnates, or when the second order differential is equal to 0, as shown in grey (see Section 3.1.3). The asterisk shows the measured and simulated DNA distributions that correspond to exponential growth.

1 and 2 chromosome equivalents. At the end of the experiment, the vast majority of the population contains integer amount of DNA, with 1 and 2 chromosomes.

For the culture grown at 23 rpm ($\tau = 39.68$ min), the exponential DNA distribution contains 1.5 to 3 chromosomes equivalent. This is maintained throughout the transition phase until 360 min post-inoculation, where the population enters stationary phase (Note that the distribution at 300 min post-inoculation does not conform with the other DNA distributions. The most likely explanation that an error in staining was made that caused the reduces fluorescence of this population compared to the others). Similar to the 230 rpm culture, the population's DNA content falls to two discrete peaks, where at the end of the simulation a majority of the cells have either 1 or 2 chromosomes in the population.



(a) Shaking rate of 230 rpm against a simulated population using the HMG framework with input the OD growth curve presented in Figure 4.1a. The first measured DNA distribution (180 min) corresponds to the identified exponential distribution.



(b) Shaking rate of 23 rpm against a simulated population using the HMG framework with input the OD growth curve presented in Figure 4.1b. The first measured DNA distribution (240 min) corresponds to the identified exponential distribution.

Figure 4.2: Measured WT K-12 MG1655 bacteria grown in LB at two different shaking rates, measured against simulated population using the HMG framework. The simulated DNA distribution was spread using a in-house Gaussian blurring protocol (see Section 3.1.3).

Simulation Results

The results of the simulation are shown in Figure 4.2 and quite accurately reproduce the measured exponential DNA distributions. Although not directly intended, it also strengthens the case for the CH model as an accurate formalisation of a WT bacterial population growing exponentially. The simulation was then brought forward using the injection growth method throughout the rest of the growth curve, where the DNA content of the simulated population was returned at the same time points as the ones measured.

At 230 rpm, the simulated exponential distribution (240 min) mirrors the measured DNA content well, despite containing a sub-population of cells with less than 3 chromosomes that the measured DNA content does not have. The transition phase is well emulated, with the same rate of decrease in the DNA content of the population, until 300 min post-inoculation where the population contains a majority of cells with 2 chromosome equivalents. As the population progresses through the stationary phase, the simulated population returns populations containing 1 and 2 chromosomes equivalents. Although the globally frequencies of the simulated peaks matched well the measured ones during that time, the simulated distribution perhaps underestimated the frequency of cells with 2 chromosomes equivalent, compared to the measured DNA distribution.

At 23 rpm, the exponential simulated DNA distribution returns a DNA content of 1.5 to 3 chromosome equivalents, while the measured one returns between 1 and 2. Although the measured DNA distribution returns a tail end of cells with more than 2 chromosomes, it occurs at a much smaller frequency than the simulated one. Overall it seems like the simulation returned a slight overestimation of the DNA content of the population by 0.5 DNA chromosome equivalent. The two subsequent (300 min and 360 min) simulated DNA distributions are complementary, with a majority of the cells with 2 chromosomes, and a large tail end portion of the cells with decreasing amount of DNA. At 360 min post-inoculation the simulated population starts to have a sub-population with 1 chromosome, something that does not happen in the measured population. The last two (420 min and 480 min) DNA distributions return distinct peaks of 1 and 2 chromosomes equivalent, where a majority of the population contains 1 chromosome, and where the peak at 2 chromosomes is slowly decreasing. The measured DNA distributions exhibits a similar behaviour. However, in the latter, there was a higher frequency of cells with more than 2 chromosomes, and many more cells in between 1 and 2 DNA peaks; evidence that some are actively replicating their chromosomes.

Conclusions

Although individual DNA distributions sometimes do not overlap precisely with the measured DNA content, and especially the frequency of DNA content commonly does not match recorded ones, overall the DNA content of the population was well represented using the HMG modelling framework, especially considering the large sources of potential errors. Indeed, the pattern of DNA distributions as the growth rate of the population diminishes from exponential growth to stagnation of growth is very well reflected using the HMG simulation method. This validates the HMG simulation framework, and confirms that this modelling strategy is accurate enough to simulate the dynamics and heterogeneity of a population growing in non-balanced growth conditions. It confirms that injection growth is an accurate enough method to simulate for the growth heterogeneity of a population. Furthermore, the C and D times in the functional form as described by Keasling et al. [1995] is an accurate enough description of the rate of replication and segregation for WT MG1655 *E. coli*.

4.1.2 Nutritional Shift-up Experiment

Aim

Unlike assumptions of balanced growth that may be simulated making atemporal assumptions on the state of the population, all non-balanced growth is temporal by nature, where its state is the function of a previous state. The nutritional shift-up experiment is a good case study to examine the dynamics of non-balanced and balanced growth by transitioning between two different growth regimes. This is also a good case study to examine the features of rate-maintenance (see Section 3.1.4).

Experimental Results

WT MG1655 bacterial cells were extracted from an overnight culture and inoculated into M9 minimal media supplemented with glucose. While still under exponential growth, a sample of the M9 culture was diluted into LB media culture. Measured OD from the shift up experiment, presented in Figure 4.3, reflects literature reports of a lag phase transition period between the two growth regimes as the cells adapt to the new growth environment (Kepes et al. [1985]). It must be reminded that although the OD seems to decrease and thus instantaneous growth rate is < 0, because the HMG modelling strategy does not take into account cell death, the spline fitting method interprets any reduction in OD as stagnation instead (i.e.

$\mu = 0$). Nevertheless, as expected, cells in M9 media grow slower ($\tau = 31.17$) than in LB ($\tau = 20.83$), with a transition lag phase between the two.

Simulation Results

By extracting the doubling rate from the linear sections of the OD growth curve (Figure 4.3) the two exponential M9 and LB growth conditions may be approximated using the CH model. The transition period between the two however, contained a stagnating growth curve that translates to a doubling rate that approaches 0, and thus cannot be simulated using the above-mentioned method. For the transition period the HMG model was used with the results presented in Figure 4.4. To simulate the shift-up experiment the following steps where undertaken: using the doubling rate from the M9 exponential section of the growth curve a "seed" population was generated under the assumption of exponential growth ($\tau = 31.17$ min). Then the rest of the OD growth curve was used as input to the HMG framework. This includes the LB exponential distribution ($\tau = 20.83$ min), and as a consequence the reported simulated DNA distribution was not generated under the assumption of exponential growth. It was merely inferred from the input OD that was itself growing exponentially. Theoretically, there is no difference between a simulated population under the assumption of exponential growth and one simulated using the injection growth if the measured OD is increasing exponentially at the same rate (starting from a random cell). Any difference here would thus come from temporal features of the population.

Measured M9 exponential DNA content returned a majority of 2 chromosomes with a small sub-population with 1 chromosome and the last peak at 3 chromosomes equivalent evidence that these cells chromosomes are being actively replicated. On the same sub-plot, the CH DNA distribution major peak corresponds to ≈ 1.75 chromosomes with another major portion of the population that contains 3 chromosome equivalents. The simulated HMG DNA distribution returns a slightly different distribution than the CH with the major peak returning 2 chromosomes equivalent (in theory they should be very close since the same assumptions of exponential growth are made). Next the population DNA distribution in stationary phase returns a large portion of its cells actively replicating their chromosomes with a peak at 3 chromosomes, while still having a small sub-population of cells with 1 chromosome. Overall the HMG population matches well with the measured one, albeit with the frequencies of the peaks not matching with the measured distribution. Upon entering the transition period, the population growth stagnates and as a consequence, the cell cycle of the individuals slows. This is reflected in the mea-

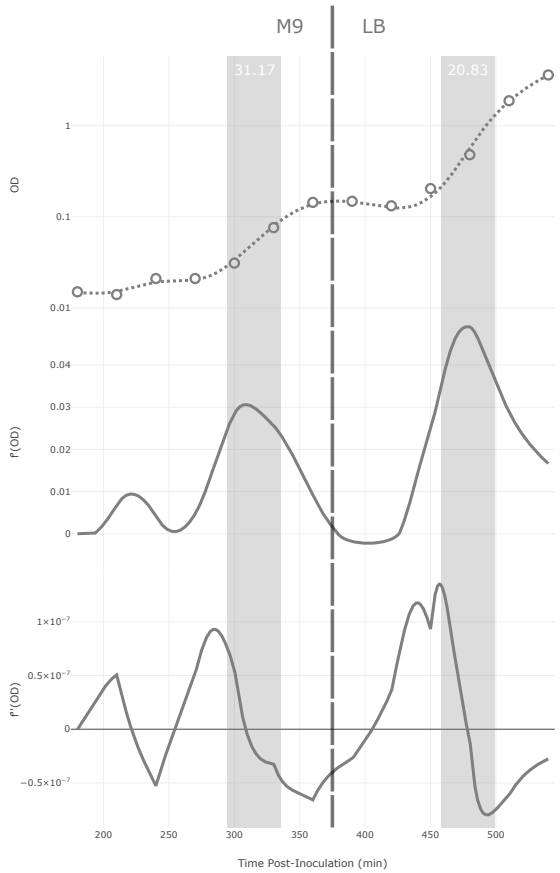


Figure 4.3: Spline fit and analysis of the nutritional shift-up OD data incubated at a shaking rate of 230 rpm and $37°C$. Top sub-plot shows the original OD data (open circles) fit with a spline function (dashed line). The middle sub-plot is the first order derivative of the natural log of the fit. The bottom sub-plot is the second order derivative of the natural log of the OD fit. In theory, exponential growth applies when the second order derivative is $= 0$. The dashed vertical line shows the time when bacteria where transitioned to a richer media (LB) from a poorer media (M9+Glucose). The grey boxes represent the identified section of the OD growth curve where exponential growth applies, where the doubling rate inferred from them are shown in those boxes in white.

sured DNA distributions that return two major peaks at 1 and 2 chromosomes with a small number of cells with more, that are still replicating their genetic material. The HMG simulation mirrors these dynamics well, with also the major peaks corresponding to 1 and 2 chromosomes equivalent, where again, the frequencies of the two do not perfectly match the measured ones. Furthermore, the number of cells that are replicating themselves seems to be underestimated with the HMG model, where it returns no cells with more than 4 chromosomes while the measured data does. As the population starts growing exponentially once again, the measured DNA distribution major peak falls at around 2.75 chromosomes equivalent but is quite spread. The HMG simulated DNA content returns the major peak at 3 chromosomes equivalent and another at 2 with a tail containing more. The CH simulated distributions contain a majority of cells with 3 chromosomes and a tail with more DNA content, until around 6 chromosomes.

Overall, although the HMG simulation population frequency does not perfectly match the measured DNA content, it still predicts the dynamics of the DNA content of the population well. The fact that the HMG and CH LB exponential DNA distributions do not match, and that the HMG distributions better fits the measured one shows a subtlety in the dynamics of the population that the HMG distributions is able to capture that the CH one cannot. Given this difference, one can argue that in fact the measured DNA distribution in LB is not in balanced growth, or at least completely. With the high frequency of cells containing a single chromosome, it is possible that the culture was not in exponential growth for a long enough time to rid the population of cells that were physiologically still in the state of lag phase.

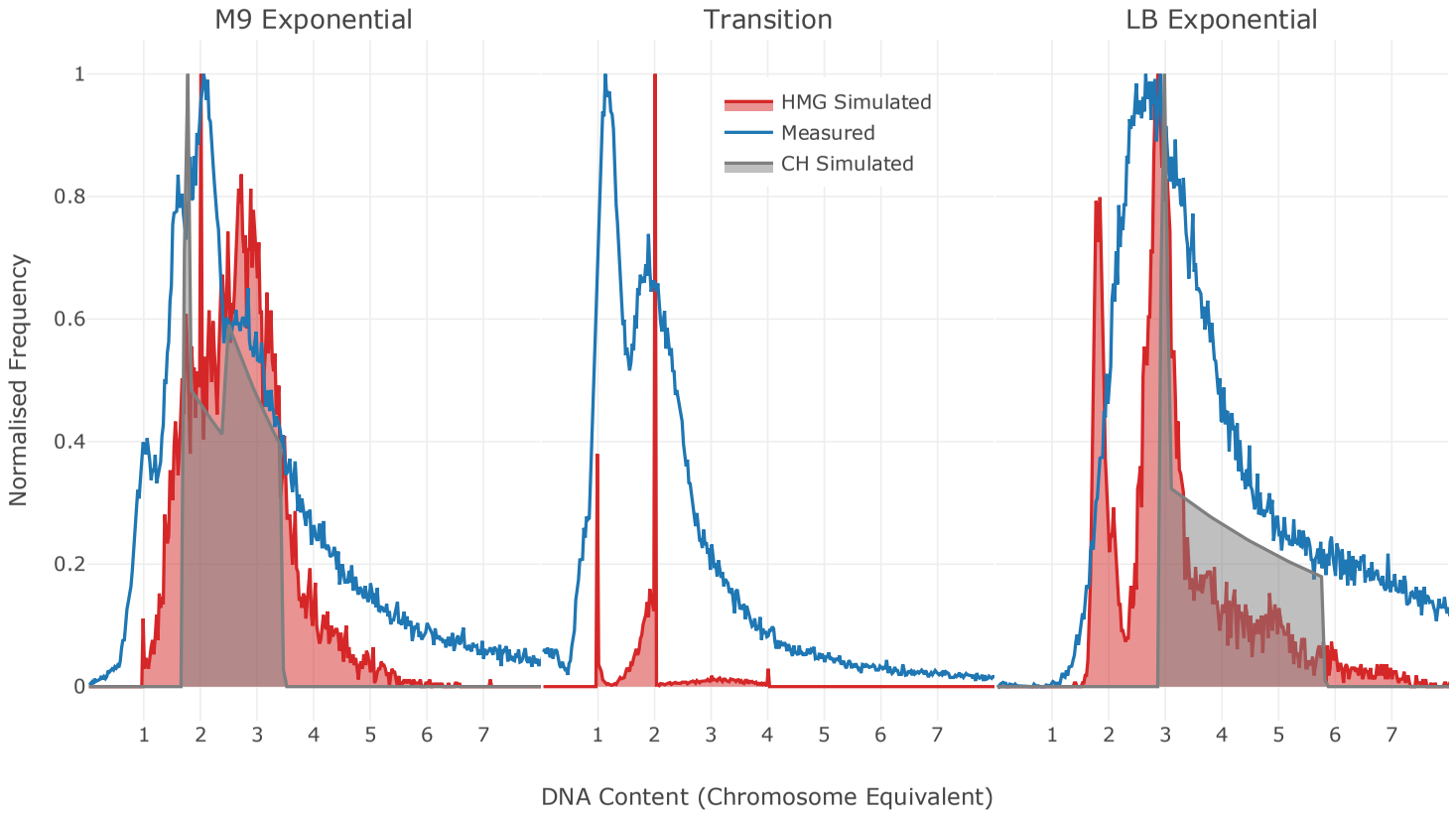


Figure 4.4: Measured against simulated DNA distributions of the three different states from the nutritional shift-up experiment. The OD growth curve was analysed as presented in Figure 4.3, where the spline fit was used as an input to the HMG model and the identified regions of exponential growth's doubling rate was used as input to the CH model. The doubling times identified are $\tau = 31.17$ min for M9 and $\tau = 20.183$ min for LB. The LB exponential DNA distribution is simulated using the HMG model where balanced growth is not assumed.

Population Properties

To gain a deeper understanding of the state of the simulated population as it transitioned through the shift-up growth dynamics and why the HMG simulated LB DNA distribution does not match the CH one, in Figure 4.5, the distributions of the models parameters are plotted. The M9 distribution corresponds to the "seed" population. It was generated assuming exponential growth using the HMG framework. All the parameters of the model are predictable since they are defined under the assumption of exponential growth. The age distribution matches well with the theoretical DNA distribution of a population growing exponentially. Both the C and D distributions return normally distributed populations and both volume and DNA have distributions that match well with the CH ones.

As the cells entered the transition phase, the population volume stagnated, where individual cells within the population stop growing. Although the portion of newborn cells is reduced, illustrated by the diminished ratio of cells with an age close to 0, the volume of the population decreased substantially. This is driven by cells where previously flagged for division expending their D time and dividing when the growth of the population stagnated. As a further consequence, newborn cells in the transition phase are assigned new C and D values that, due to the stagnation of growth, leads to their respective maximas of 253.15 and 134.78 min. The DNA distribution followed the same pattern, where population chromosome copy number contained cells that are not actively replicating; with integer number of chromosomes of either one or two copy numbers.

Upon entering the next exponential phase under LB media, the volume histogram returned a very sharp increase as the population starts accumulating more volume. The DNA content of the population also increased with the majority of the population containing 3 chromosome equivalents. However, considering that the volume of the population is so large, the DNA content was quite small. The M9 exponential distribution for example, contained between 1 and 5 μm^3 and contained between 1 and 6 chromosome equivalents. The LB distribution on the other hand contained a major peak of cells with $9\mu m^3$ and yet the maximal DNA content was 6 chromosomes equivalent. The reason for such a discrepancy becomes clear if one considers the C and D times distributions. The population was composed of many individuals with transition period parameters, while growing in the new regime. In the simulation, a cell in the transition period that is not growing would suddenly see its volume increase sharply as it enters the LB culture. These cells would reach critical mass quite quickly, but would still have a replication, segregation and division rates from the transition period. As a consequence, it would take a significant

amount of time before the cell divides and updates their C and D parameters to the LB exponential rates. This is supported by the age distribution that returned a lot of cells that are old enough that if one traces back to the time they where born would correspond to the transition period. Nevertheless, the majority of the population simulated in the LB regime was composed of cells with fast growing parameters with a significant portion containing 2 chromosomes that belong to cells that are still expressing the transition phase parameters that the CH model does not account for. It must be noted that the eclipse period here blocks a significant number of initiation events that could be occurring due to the large volume accumulation compared to the number of chromosomes, and this plays a large role in the dictating dynamics of the population under this growth regime.

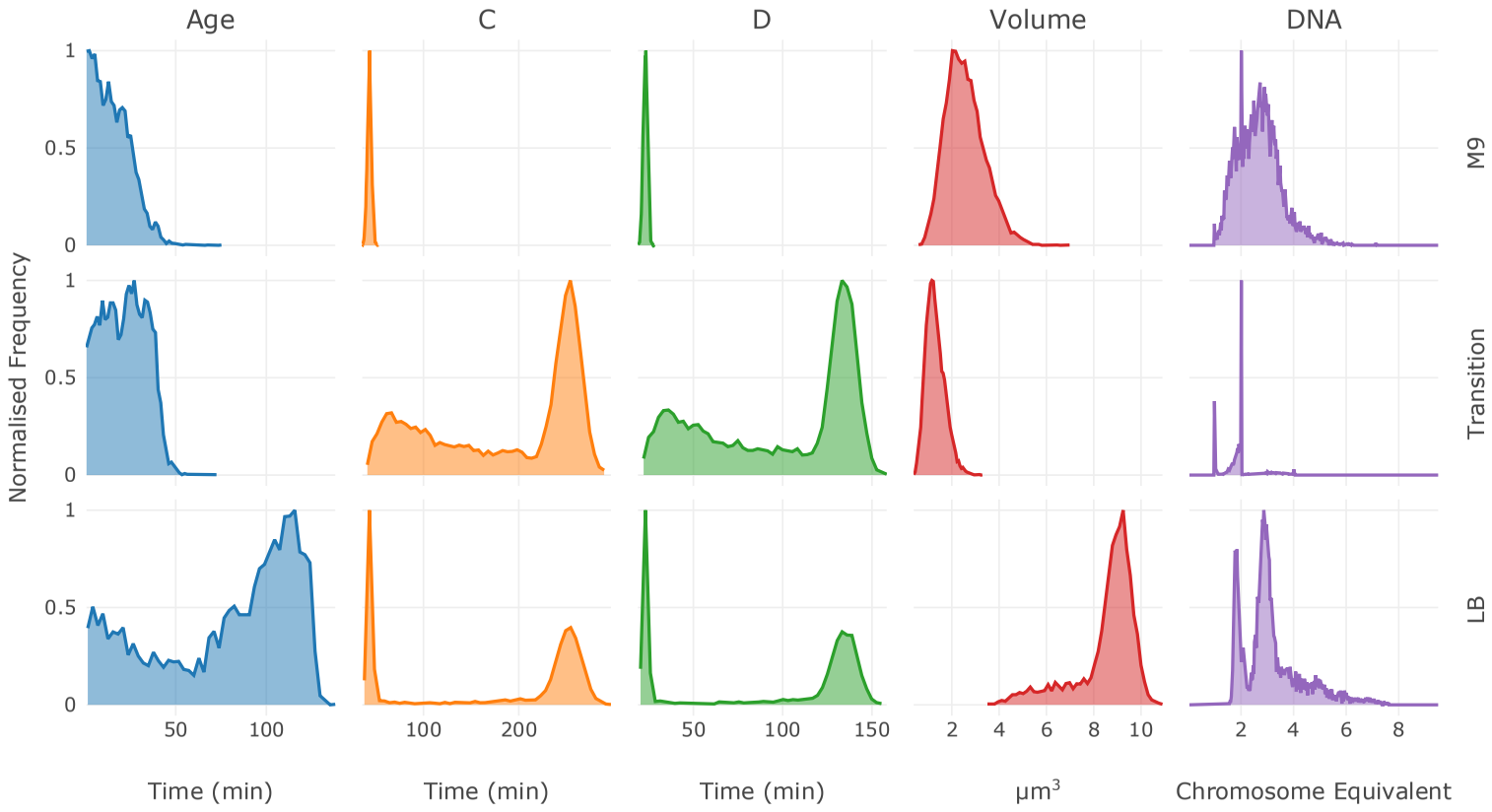


Figure 4.5: Population distribution for different parameters of the simulated model using the HMG framework with input the nutritional shift-up growth curve as presented in Figure 4.3.

Conclusions

Simulating the cell cycle heterogeneity of a bacterial culture growing in such a sharply changing environment truly tests the accuracy of the HMG framework to reflect the growth and cell cycle dynamics of a bacterial population. One characteristic of this model is the error propagation from one population state to another. Although the M9 exponential distribution HMG distribution (and CH) reasonably fits with the measured DNA content, it underestimated the number of cells in the culture with 2 and 1 chromosomes. Once growth stagnated in the transition phase the individual cells completed currently replicating chromosomes and divided while not initiating new replication forks. As a consequence, the population fell to discrete peaks of integer amounts of DNA that is very much dependent on the exponential state of growth. Because the DNA was overestimated in the previous state subsequently the stagnating phase DNA content would also have an overestimated DNA content. In the shift-up experiment case, this phenomenon may be observed where the transition phase HMG distribution missed a whole sub-population of cells with one chromosome equivalent and where the frequencies of each peak did not match each other.

Inaccuracies in simulating the DNA content of exponential growing populations illustrates a known difficulty in deducing the state of the population from the OD growth curve (Stokke et al. [2012]). Indeed, external factors such as measurement error, fitting procedure error and human error may skew the calculated doubling rate of cultures, from which the cell cycle state and heterogeneity is calculated from. Furthermore, the measured M9 DNA distribution may itself not be extracted from a population that is in balanced-growth. Although the culture was sampled when growing exponentially, it may be that the population was not in that state for a long enough time to rid its population from members that where still transitioning from the lag phase of growth. The same phenomenon explains why the results of the LB HMG DNA distribution fits with higher accuracy the measured distribution than the CH generated one. As Figure 4.5 implies, the measured DNA distribution contains a significant portion of its population with cells that are in a previous non-balanced state of growth, instead of the current balanced growth state. It can be expected that with longer simulation time at the same growth rate, the population would slowly tend towards the same DNA distribution as the CH simulated one, as these individuals that have not updated their parameters would divide and thus adopt the new parameters.

In conclusion, updating the cell cycle parameters upon division seems to be a valid approach at approximating DNA distributions from such a dynamic growth

condition. Although it is unlikely that the cells contain such a stringent control mechanisms for the replication and segregation and division, formalising the cell cycle in this manner explains the observation of rate-maintenance, and supports the view that metabolically these physiological features of the cell are not directly linked. More experimental work would be necessary to investigate the molecular actors that control such mechanisms.

4.2 Optimisation to recA1 Mutants

4.2.1 The Problem

It is common practice to optimise for the C and D rates of the CH model to fit exponentially growing populations to their measured DNA distributions (Stokke et al. [2012], Michelsen et al. [2003]). Different bacterial strains may have different rates of replication, segregation and division due to a number of different mutations or versions of proteins involved in the cell cycle (see Section 3.2.2). Consequently, it can be difficult to use the CH model to match quantitatively the cell cycle properties of mutant forms of bacteria, with the same ease as WT MG1655 or Br strains from which the models parameters have been calculated from (Keasling et al. [1995], Dennis and Bremer [1974]).

This research will focus on the recA1 mutation, a version of the DNA recombination protein commonly used to reduce its participation in homologous recombination for the purposes of genetic manipulation. Although viable, recA1 causes a range of downstream affects on the dynamics of chromosome replication, where the rescue of stalled replication forks and the rescue of damaged DNA damage have been identified as being the most severe (sketched in Figure 3.7). Because the dynamics of the rescue of DNA damage through this mutation has largely only been qualitatively investigated, and because the TOP10 bacterial strain is widely used in genetic manipulation experiments, it is of particular value to investigate if the HMG modelling framework may be able to reflect said dynamics quantitatively. At the time of writing, there has not been experimental quantification of the rate of DNA damage, and type, that this mutation causes.

4.2.2 Methods

Parameter Sensitivity

To zero in on the most likely parameters that may influence population DNA distributions, sensitivity analysis was performed on all the parameters of the model

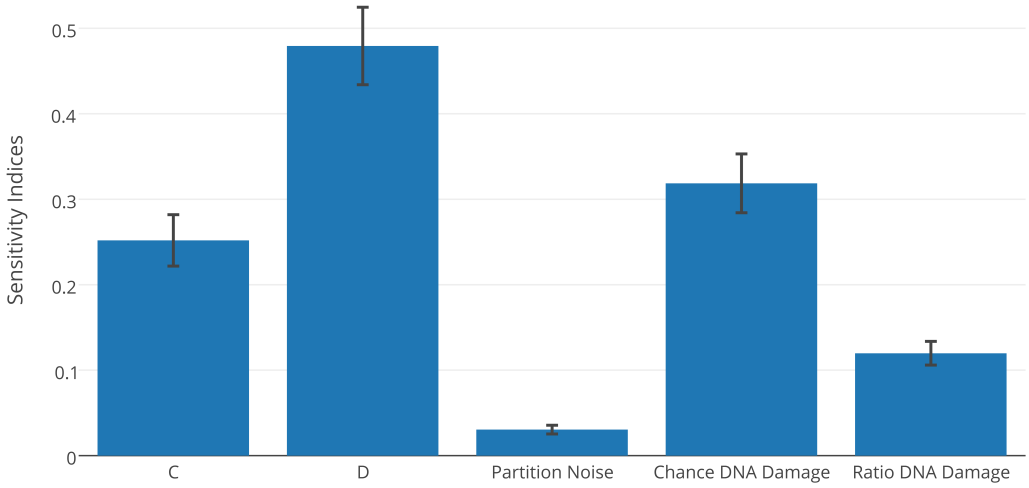


Figure 4.6: Total-order index sensitivity analysis for the optimized parameters. These sensitivities were calculated based upon an initial ensemble (SOBOL sequence) of 12000 parameter sets. The error bars represent the 90% confidence intervals. For this analysis, a base case was generated by simulating exponential growth, using the following parameters: C: 40.0 min, D: 20.0 min, Partition Noise: 0.24, Chance of DNA Damage: 3.034, Ratio of DNA damage: -1.652. Therefore, the sensitivity indices plotted here represent the degree to which changing any model parameter shifts the simulated DNA distribution (under exponential growth) compared to the base case.

(excluding the noise parameters). As presented in Figure 4.6, the parameter that influenced the greatest the DNA distribution of a population is the segregation and division rate (D). Given that the volume of the cell dictates the DNA content, through the critical mass constant, any change in the D times would cause a cell accumulating more mass once a replication event has been accomplished, and consequently would cause daughter cells to be larger with a longer D time. The ratio of "chance of DNA damage" has the second largest influence on the DNA distribution of a population, with the replication time (C) having almost the same impact. This parameter sensitivity run does not show the fact that both C and D parameters are dependent on growth rate, where with shorter doubling times a larger replication time would have a larger impact on the subsequent DNA distribution. Nonetheless, this method gives a good overview of the weight of each parameter. Surprisingly, the partition noise (the ratio of chromosomes inherited by daughter cells upon division) has very little influence on the DNA dynamics of the population, and thus this parameter was ignored in the optimisation process. This is caused by the homoeostatic nature of critical mass. For example, consider a cell that contains 4 complete chromosomes, and a volume of $4\mu m^3$ (in this scenario critical mass is 1 for simplicity) dividing into daughter cells. If partition noise is such that one daughter inherits one chromosome and the other 3 chromosomes, but the volume is divided equally, then the cell with a single chromosome would initiate replication immediately upon division, and after the eclipse period another quickly thereafter (with a birth size of $2\mu m^3$, it supports the initiation of two $oriC$). Meanwhile the other cell with 3 chromosomes and a volume of $2\mu m^3$ would require the accumulation of $1\mu m^3$ before initiating replication. The result is the cell that inherited a single chromosome would create daughter cells with a single but already replicating chromosome while the other cell would create daughter cells with 3 chromosomes that would take a longer time to divide. With time these differences would attenuate themselves depending on the growth rate.

Objective Function

For reliable mathematical optimisation one must define a valid objective function and given the complexity of the problem at hand, it is critical that it is well posed. As specified in the previous section, the parameters that were included in the optimisation protocol are C, D and the two DNA damage parameters "chance DNA damage" and "ratio DNA damage". Because as reported by Keasling et al. [1995] and measured by others (Michelsen et al. [2003], Stokke et al. [2012], Wallden et al. [2015]), the C and D rates are a function of the growth rate of the population and

thus of the cell, and because the goal of this exercise is to explore the dynamics of bacterial populations throughout a range of different growth regimes with changing population growth rates, it was required to express these two parameters in their functional forms:

$$C(\mu) = C_1 \cdot e^{-C_2\mu} + C_3 \tag{4.1}$$

$$D(\mu) = D_1 \cdot e^{-D_2\mu} + D_3 \tag{4.2}$$

the parameters for both these physiological features are C_1, C_2, C_3, D_1, D_2, D_3. Both of these functional forms are the same as expressed by Keasling et al. [1995], i.e. one phase exponential decay functions, where in this case the dependent variable, growth rate, is expressed in instantaneous (per minute) instead of a fraction of doubling time as the authors report (i.e. $\frac{1}{\tau}$ per hour). In this functional form, C_1 and D_1 correspond to part their maxima values, where when $\mu = 0.0$, $C(0) = C_1 + C_3$ and $D(0) = D_1 + D_3$. The parameters C_2 and D_2 are the rate constants, where the half-life of the equation is $= ln(2)/C_2$ and $= ln(2)/D_2$. C_3 and D_3 are the plateau values. This type of equation is commonly used in many chemical and biological processes to describe the behaviour of molecular processes where the rate is proportional to the amount left of a central chemical, such as the dissociation of a ligand to a receptor (Motulsky and Christopoulos [2004]). Regarding the other two DNA damage parameters, because there is no evidence that their rate is a direct consequence of the growth rate of the cell, there is no need to optimise them in relation to the growth rate of the cell. In this implementation, the rate of DNA damage is only indirectly a consequence of growth rate, in as much as the increase in occurrence of DNA replication events increases the possibility that DNA damage may occur.

Due to the biological realism of these parameters, the following boundary conditions were set to avoid the optimisation algorithm finding biological infeasible minima:

$$C(\mu) \geq 30min \tag{4.3}$$

$$D(\mu) \geq 15min \tag{4.4}$$

$$C(\mu) > D(\mu) \tag{4.5}$$

$$Chance \quad of \quad DNA \quad damage > 2.75 \tag{4.6}$$

The C and D times minima are based on the lowest measured times recorded in the

literature (Michelsen et al. [2003], Wallden et al. [2015], Keasling et al. [1995], Stokke et al. [2012]). The observation that the replication time must strictly be larger than segregation time is also based on literature accounts of measured replication and segregation times. From all reports (Figure 3.6 and 3.9), Michelsen et al. [2003] only documented one instance of a replication time (C) smaller than the segregation and division time (D) of the cell (a difference of only one minutes). This difference can be accounted for by the standard error of the quantification method used.

The boundary condition for the "Chance of DNA damage" parameter has a set minima. This is done for purely practical reasons, since a value of DNA damage that is too large would lead the cell to rarely experience a complete replication event, would lead to the stagnation of the simulation and thus lose optimisation time. Indeed, a value of 2.75 and simulations with $dt = 0.01$ would cause a $\approx 30\%$ chance that a replicating chromosome experiences DNA damage. Tests with the parameter "chance of DNA damage" set at such a high value causes the simulation to stagnate. The "Ratio DNA damage" represents the balance between DNA damage being either a replication bubble collapse that leads to the collapse of the replicating strand or double stranded break that leads to the degradation of the whole chromosome (See Figure 3.7). A value of 0.0 would generate an equal portion of the two, where a negative value would lead to more replication fork collapse than whole chromosome degradation and a positive value the opposite.

In the optimisation protocol, every measured DNA distribution along the growth curve is compared to its simulated counterpart, in the same chronological order, as they progress the growth curve. In addition, similar to the raw data given by the flow cytometer, the simulation returns a series of discrete DNA values from each member of the population. To enable the analysis of the raw data, it is common practice in analysis of the flow cytometry data to bin the output into histograms (Michelsen et al. [2003]). To enable for a quantitative comparison between the measured and simulated results, the same bins were used for the simulation results and the flow cytometer DNA. To compute the difference between the two, a previously reported functional form is used to define an objective function for our optimization (Skarstad et al. [1985]) that quantitatively compares the difference between two histograms, called the "similarity score":

$$\sqrt{\sum_{i=1}^{m} \frac{(\sqrt{y_i} - \sqrt{N_i})^2}{m-1}} \tag{4.7}$$

where i is the bin number, y_i and N_i are normalised value of bin i for the measured and simulated values respectively, and m is the total number of bins. Under such a

formulation, the lower the score the more similar the two histograms being compared are, where a value of 0 would represent perfectly similar histograms.

Genetic Algorithm

Because of the nature of the problem, the existence of a global optima is not guaranteed. In such a situation, a heuristic optimisation algorithm seems the most appropriate approach to finding the best possible solution (Whitley [1994]). To this end a genetic algorithm was used to optimise the model to fit the measured DNA content along the growth curve. This optimisation method contains clear advantages (reviewed in Section 1.3.3). It is multi-modal, that is, it may find multiple local minima. Indeed, given that the problem has high dimensions (in this case 8 parameters), it is unlikely that any optimisation method would find the global minima, if any exists (Whitley [1994]). Furthermore, to increase the chance of finding a minima, a SOBOL sequence was generated and evaluated at the start of the simulation, where the top sequences were used as the start to the optimisation algorithm.

A Gaussian mutation function was used as the operator to the genetic algorithm to produce variation within the sequences, with $\mu = 0.0$, $\sigma = 0.08$. The probability that an individual being mutated was set to $= 0.75$, whose function used the native Python random function, generating a pseudo-random number with the range $[0.0, 1.0)$. The result is a probability of 75% that any given member of the sequence is mutated. To generate crossover, a two-point crossover function was used, where two members of a sequence are mated with another while keeping their position. For example, the C_1 of a sequence would be switched with the C_1 of another sequence.

4.2.3 Training

Growth Curves

The identified exponential sections of the growth curves return doubling rates that are sound considering their respective growth curves. The fastest growth rate belonged to the culture grown in LB at 230 rpm with a doubling time of 47.92 min. The culture grown in M9 at 230 rpm returns the second fastest growing population with a doubling rate of 54.78 min, and the culture grown in M9 at 23 rpm but slower doubling time of 59.14 min. The last and slowest growing culture was grown in LB at 23 rpm with a doubling time of 167.88 min, and is surprising not only by its large difference in doubling rate compared to the others but also to the fact that the culture was grown in a richer media than the M9 culture at a similar shaking

rate. However, considering the input OD, it was difficult to find a section of the growth curve where the culture was clearly growing exponentially. There was most likely an experimental error that caused this growth condition that should be ideal for growth, to not be. Thanks to the nature of the fitting and simulation method, since no assumptions are made as to what should be the dynamics of a population in non-balanced growth, the simulation should accurately reflect the nature of the different parameters.

Experimental Results

The two cultures grown in LB at different shaking rates (230 rpm and 23 rpm) were used to train the model, with the histograms presented in the top of Figure 4.8. TOP10 bacterial cells grown in LB at a shaking rate of 230 rpm returned an exponential DNA distribution with relative high DNA content of 3 to 5 chromosomes. As the population entered the transition phase the DNA content of the population slowly decreased, where at around 6.5h, the DNA content settled to a population with a majority of cells with 1 and 2 chromosomes. Still the population had a substantial number of cells that were actively replicating their chromosomes, as shown from the large tail following the 2 chromosomes peak and the substantial numbers of cells between 1 and 2 chromosomes peaks. From Figure 4.7, this corresponds quite rigorously to the time when the culture entered an early stationary phase. As the culture progressed through the stationary phase, the bacterial population contained a growing number of cells with 1 and 2 chromosomes. This makes sense considering that the culture's growth has stagnated and thus the bacterial cells within the culture that were actively replicating their chromosomes with a chromosome copy number of between 2 and 4 chromosomes would have created an offspring with 1 or 2 chromosomes equivalents while the ones that were not replicating their chromosomes would not initiate a new replication event.

The culture grown at a shaking rate of 23 rpm, on the other hand, contained almost the same DNA content throughout the growth curve, that only slightly increases as the culture progressed through the stationary phase of growth. The most likely explanation is that oxygen availability being the rate limiting factor in the respiration of the cells in this particular growth condition, would cause the carbon content of the LB solution to remain high enough to support the culture for a longer amount of time than at higher oxygenation. This behaviour is also observed when the same experiment was performed with WT K-12 cells also grown in LB at the same shaking rate (Figure 4.1b).

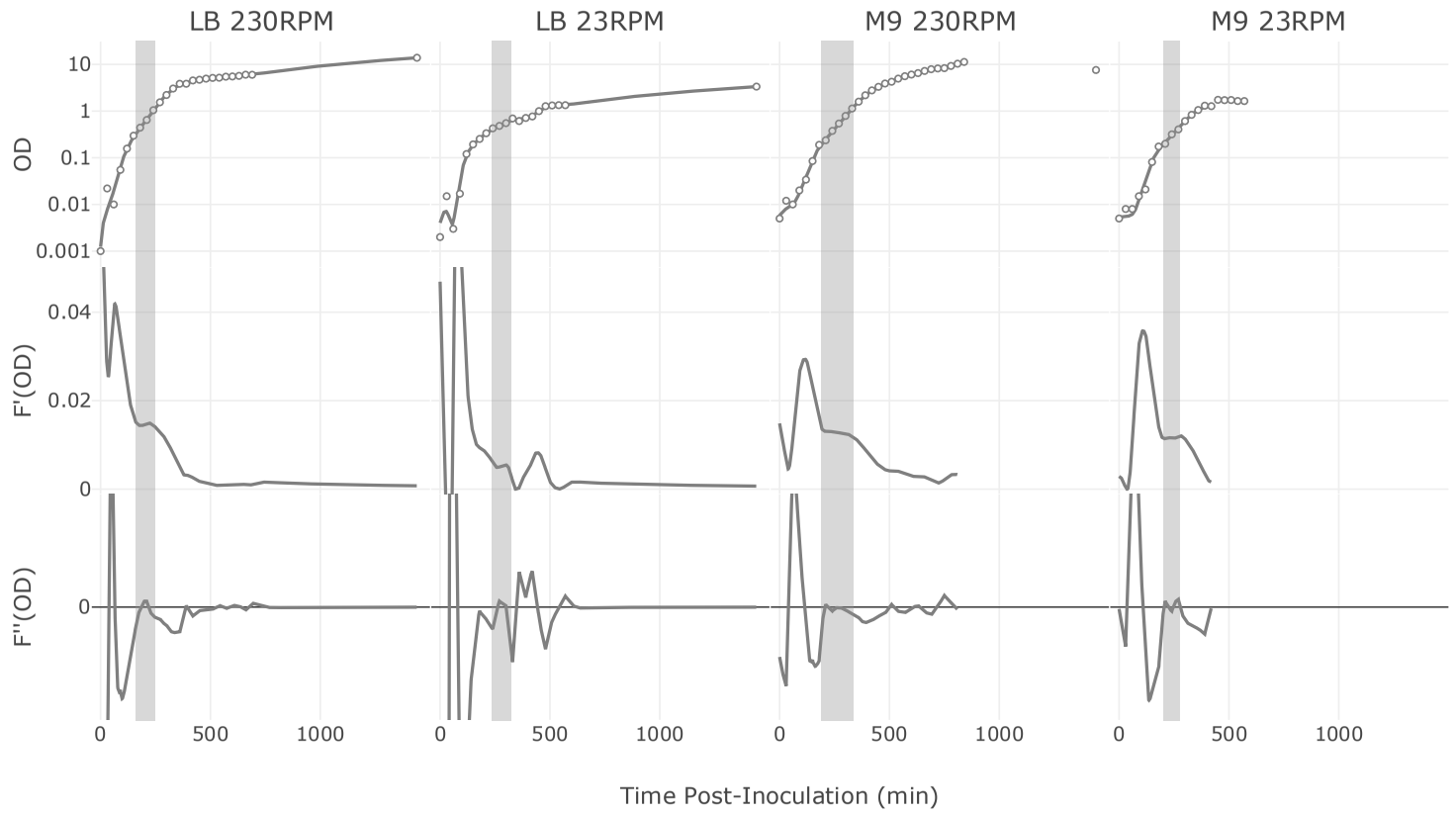


Figure 4.7: Spline fit and analysis of the OD growth curves for TOP10 bacterial cells grown in LB and M9 minimal medium supplemented with glucose at two different shaking rates (230 rpm and 23 rpm). The grey boxes show the identified section of the growth curve, that applies to the section to the first instance where first order differential function (F') is in plateau, or when the second order polynomial = 0 (F"). The returned doubling rate are 47.92 min, 167.88 min, 54.78 min and 59.14 min for the LB 230 rpm, LB 23 rpm, M9 230 rpm and M9 23 rpm respectively. The growth curve in condition M9 at shaking rate of 230 rpm has a section of the growth curve (past 840 min) that is ignored. This is due to the decreasing OD that is ignored by the fitting function, since the model does not yet implement cell death and thus decreasing OD.

Optimisation Results

The results of the optimisation for the functional forms of C and D are presented in Figure 4.9. The maxima value for the original and optimised functional forms are 231.4 and 250.02 respectively. This translates to a difference of 18.62 for the maxima C time for a cell that is not growing. Furthermore, the half-life values of the equation are 304.41 and 319.7 for the original and optimised respectively. The plateau values are 40.0 and 60.97 respectively and show a difference of 20.97. Considering the nature of C on the simulated population, the effect of a change in the rate of replication at faster growth rates controlled by the plateau value would have a larger impact on the populations DNA content than one at slower growth rates. These results are coherent with the observations that recA1 mutation causes a reduced frequency in the rescue of damaged DNA, where one can expect that stalled replication forks caused by damage that is not swiftly repaired, would lead to an increase in the overall replication time of the population (Zyskind et al. [1992]).

On the other hand, the optimised D functional form returns virtually the same plateau value, with a difference between the optimised and original of only 0.07. The major differences are between the maxima value of 80.89 and 100.59 for the original and optimised respectively, that return an increase of 20.3, very similar to the maxima increase in C. However, this increase is more significant due to the smaller original value and since the sensitivity analysis shows that any change in D has the largest impact on the downstream DNA content. As for the half-life, the original functional forms returns a value of 121.3 while the optimised half-life is 137.07. Again, even if this seems like a small value, because of the high impact of the D value on the DNA content of a population, an increase in the D time at faster growth rates would lead to a larger accumulation in the DNA content of the cell. These results of the optimisation are particularly interesting since they are closer to the literature fit to the concatenation of reported D times presented in Figure 3.9, that may suggest that the functional form from Keasling et al. [1995] may be inaccurate.

To illustrate the difference the parametrisation of the model makes on the output DNA content, and the improvement of the optimisation, a simulation of the HMG framework using the parameters as reported by the literature was plotted along with the optimised simulation DNA content, and the similarity score between the simulated and measured DNA content was reported on the bottom subplots on Figure 4.8. By literature parameters, it is implied that the DNA degradation parameters created for recA1 mutants are not used in this simulation procedure. For the culture grown at a shaking rate of 230 rpm, the HMG exponential DNA distri-

bution using the literature parameters did not mirror the measured population DNA content. The former returned between 3 and 5 chromosomes, while the simulation using parametrisation from the literature returns a DNA content of 1 and 3 chromosomes. Assuming that there were no errors in both measurements and estimation of the exponential doubling rate from the OD growth curve, the culprit must be the parametrisation of the model. Although the difference is slight, the optimised parameters returned a DNA content that is larger than the original parameters, with a DNA content in between 2.5 and 3.5. This increase may be accredited to the longer C time as described above. As the simulation progresses through the transition phase and stationary phase, at a shaking rate of 230 rpm, the DNA content is maintained until 6.5h post-inoculation. At that time, the OD growth curve starts to stagnate as the population enters the stationary phase. Before that, the DNA distribution is the same as the exponential distribution, unsurprising since the growth curve in Figure 4.7 returns a doubling rate that decreases only slightly.

The culture grown at a shaking rate of 23 rpm (Figure 4.8) returns a DNA content in between 1 and 2 chromosomes for the exponential growing population, and considering a doubling rate of 167.88 minutes, this fits well with CH model DNA distribution. Nonetheless, the optimised exponential DNA distribution returns a better fit to the measured one, where again the improvement is small. Both simulated distributions during progress through the stationary phase are quite unchanging, until the end where the DNA content increases slightly. Throughout the simulation the optimised parameters return a better fit than the original parameters, especially for the last measured time points in the simulation. The measured DNA content never returns DNA distributions less than 1 chromosome equivalent while both simulated populations do. The reason why the optimised parametrisation is better being mainly due to minimal DNA content of each histogram being larger than the original parameters. Lastly, the DNA distribution at 10h post-inoculation is very much inaccurate with the original parameters, while optimised parametrisation is much more accurate.

The bottom sub-plot in Figure 4.8 shows the similarity score of the fits to the measured data as specified in Equation 4.7. The line plot represents the mean value, with its standard deviation. Although not consistently better, the optimised parameters for the model report a better fit than the original parameters.

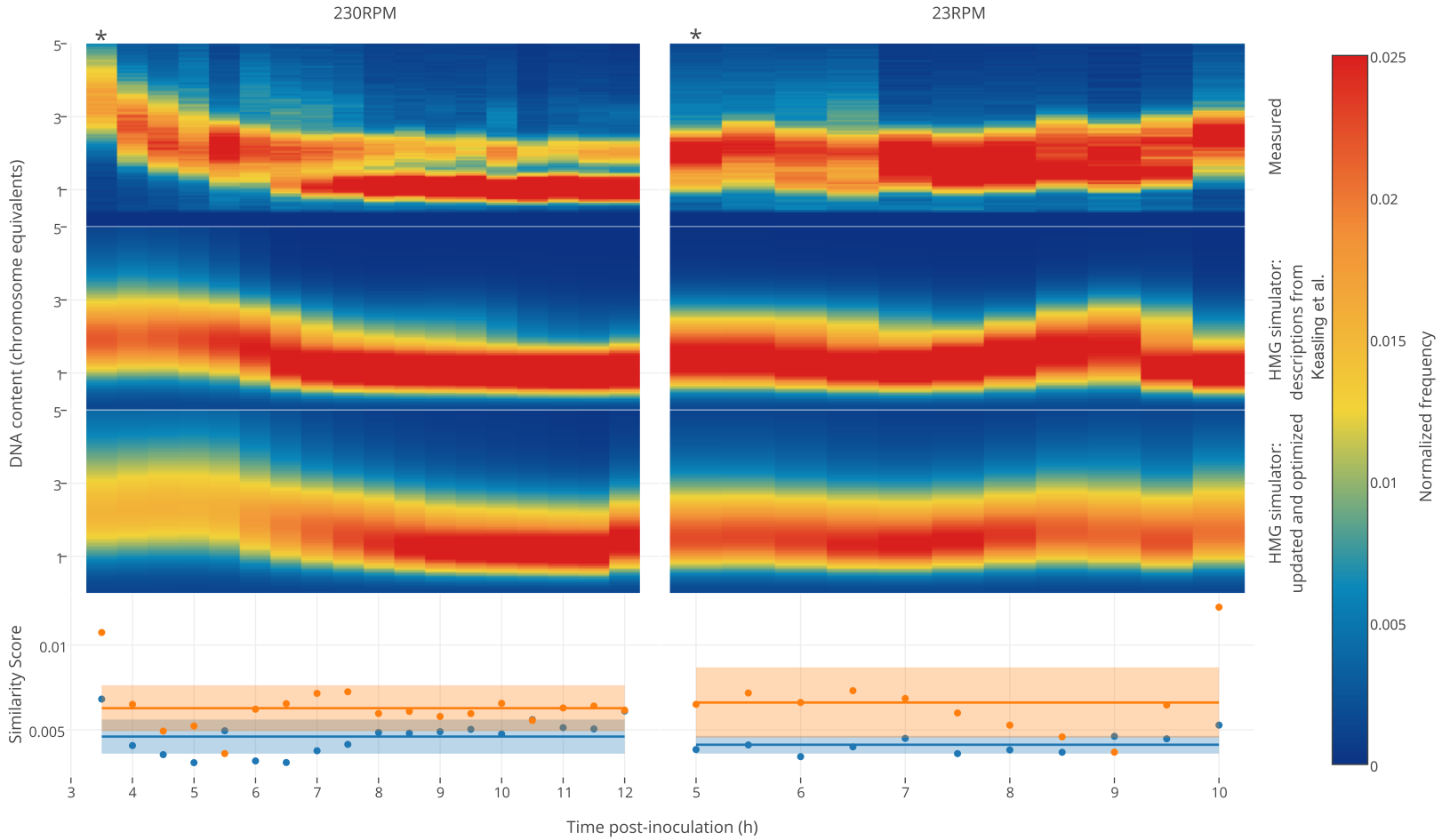


Figure 4.8: Training of the HMG simulator framework. The HMG simulator was "fed" growth curves for TOP10 cells grown in LB, shaken at 230 rpm or 23 rpm (Figure 4.7), and simulated DNA distributions were compared with those which were measured empirically. The measured DNA distributions, shown here, each represent a single experiment, each of which is representative of two or more independent experiments. The first column within each heat map represents the exponential growth phase (indicated by the asterisk (*)), and all subsequent time points represent post-exponential growth. The simulator was run using two different models and parametrisations: the first model was based upon a prior description of exponential growth by Keasling et al. [1995], which omits any consequences of RecA mutation, and the second (updated and optimized) model incorporated our description of the consequences of RecA mutation with parametric optimization. Similarity scores indicate the degree to which each prediction matches the observed DNA distribution, using a scoring function described in Equation 4.7 (lower scores represent better fits). The solid lines on the two bottom panels represent the mean similarity score across the time course, and the shaded boxes represent the standard deviation of these scores across the time course. Taken from du Lac et al. [2016].

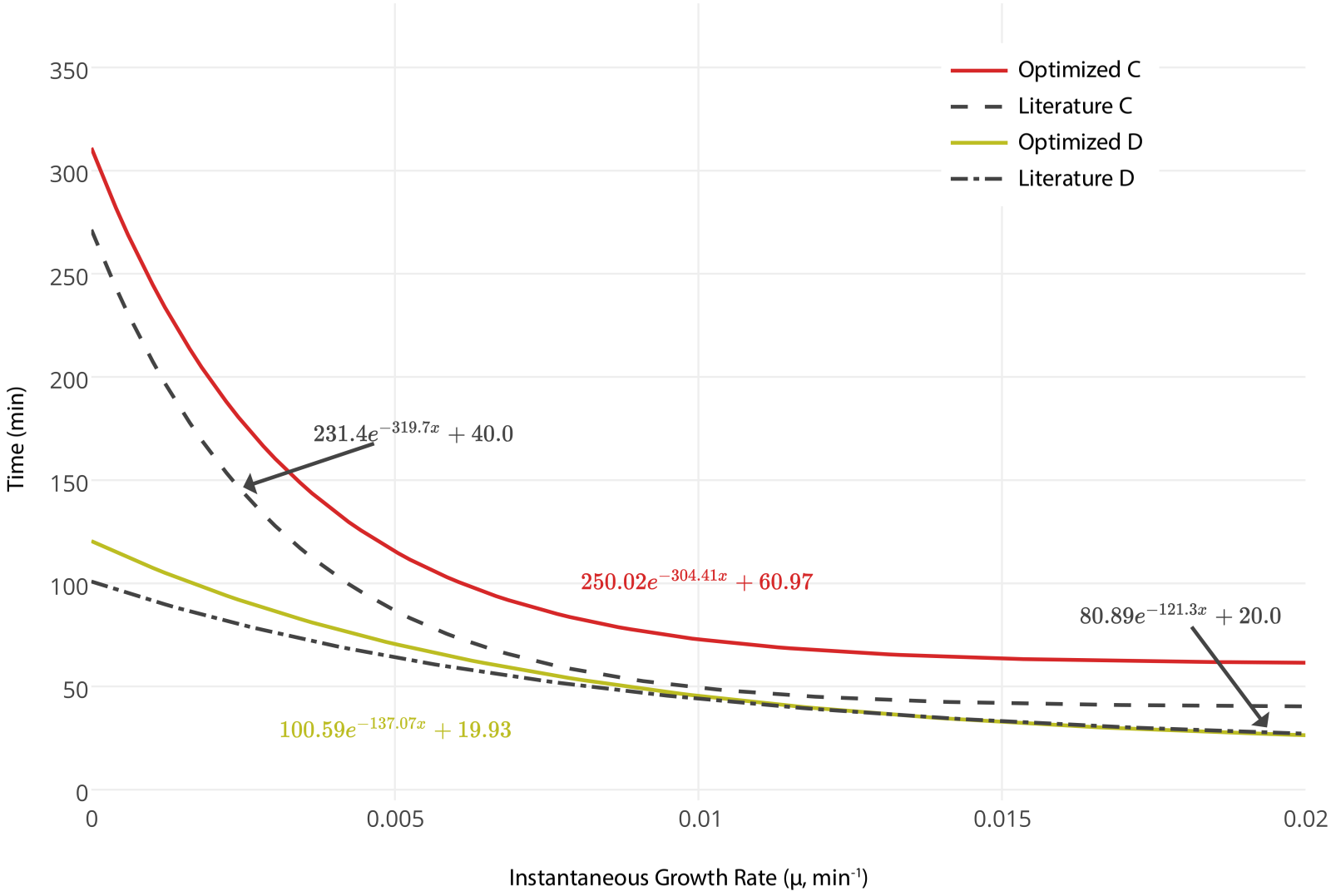


Figure 4.9: Results of the optimisation of the functional forms of C and D compared with the Keasling et al. [1995] functional forms.

4.2.4 Results

Validation

To validate the optimisation, TOP10 cells in M9 minimal medium supplemented with glucose were grown at the shaking rates of 230 rpm and 23 rpm (Figure 4.10). The overall dynamics of the populations as they progressed through the growth curve are similar as grown in LB (Figure 4.8). At a shaking rate of 230 rpm, the exponential DNA distribution returns a chromosome copy number between 2 and 4. As the population transitions to the stationary phase, the DNA content slowly descends to a majority of bacteria with 2 chromosomes. When in stationary phase, a growing number of the population contains cells with a single chromosome, until the end of the experiment, where the population is composed of an equal part of cells with 1 and 2 chromosomes equivalent and a minority of cells with more. The exponential DNA distribution returns a population between 1 and 3.5 chromosomes, a distribution maintained for the period of the transition phase. Upon entering the stationary phase, the DNA content slowly decreases as the majority of the population contains cells with 2 chromosomes. At the end of the stationary phase, a sub-population of bacteria with 1 chromosome emerges.

The optimised parameters provide a significant improvement over the original parameters to emulate the measured DNA content, as shown in the bottom similarity score sub-plots. At a shaking rate of 230 rpm, simulation of the exponential distribution using the original parameters returns an underestimated DNA content of 2 chromosomes. The optimised parameters, on the other hand, returns a DNA content that is also centred around 2 chromosomes, but has significant larger portion of its population with more and less as per the measured distribution. With both the optimised and the literature parameters, the DNA content remained the same for the first three hours after the exponential phase, as the cells are in the transition phase of growth. Upon entering the stationary phase, both simulated DNA distributions slowly reduce their DNA content to 1 chromosome equivalent, with a small overlap at 2 chromosomes. The optimised parameters showed a greater number of bacteria with 2 chromosomes equivalent compared to the original parameters and lead to a better similarity score to the measured distribution compared with the original parameters.

The exponential measured DNA distribution at 23 rpm contained a majority of cells with 2.5 chromosomes equivalents, but with a large tail above that peak. The original parameter simulation return the major peak at 2 chromosomes, that is not as spread as the measured DNA content. The optimised parameters on the other

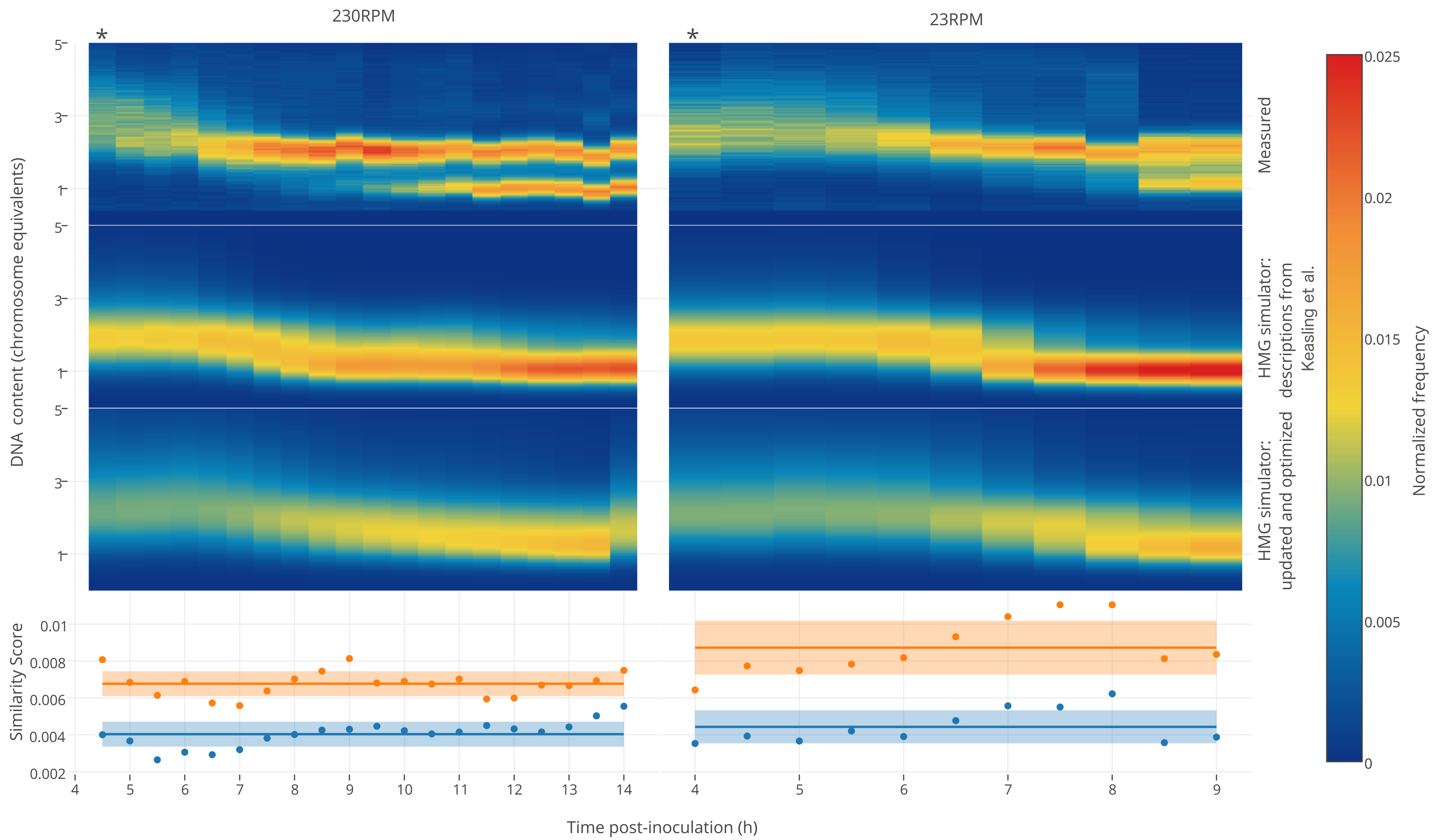


Figure 4.10: Validation of the HMG simulator framework. The HMG simulator was "fed" growth curves for TOP10 cells grown in M9, shaken at 230 rpm or 23 rpm (Figure 4.7), and simulated DNA distributions were compared with those which were measured empirically. The measured DNA distributions shown here each represent a single experiment, each of which is representative of two or more independent experiments. The first column within each heat map represents the exponential growth phase (indicated by an asterisk (*)), and all subsequent time points represent post-exponential growth. The simulator was run using two different models: the first model was based upon a prior description of exponential growth by Keasling et al. [1995], which omits any consequences of RecA mutation, and the second (updated and optimized) model incorporated our description of the consequences of RecA mutation with parameters optimized based upon growth in LB. Similarity scores indicate the degree to which each prediction matches the observed DNA distribution, using a scoring function described in Equation 4.7 (lower scores represent better fits). The solid lines on the two bottom panels represent the mean similarity score across the time course, and the shaded boxes represent the standard deviation of these scores across the time course. Taken from du Lac et al. [2016].

hand returned a very close fit, with not only a majority of cells with 2.5 chromosome equivalents, but with a DNA content that is quite spread around that peak. This is reflected through a lower similarity score. Both the original and optimised parameters returned a DNA content that remained the same as the exponential distribution to 6.5h. At that point the DNA content starts to diminish until reaching a majority of cells with one chromosome. The optimised parameters returned a better fit for two reasons. First the transition to the population with a single chromosome happened later as per the measured DNA distributions, where for the original parameters the population fell to a majority of cells with a single chromosome within the span of an hour. In both the measured and optimised DNA content this happened much more gradually. The optimised parameter showed a more progressive shift of a population of 2.5 to 2 chromosomes from exponential to 7.5h post-inoculation and when in long term stationary phase, more of the population contained 2 chromosomes as the measured one.

DNA Degradation

The results of the optimisation for DNA damage returned the following minima:

$$Chance\ of\ DNA\ damage = 3.907 \tag{4.8}$$

$$Ratio\ of\ DNA\ damage = -0.777 \tag{4.9}$$

These results may be translated as an overall 0.467% chance that a replicating chromosome experiences any type of DNA damage. Taking into account the "Ratio of DNA damage", this may further be translated as a 0.365% chance that a replicating chromosome experiences a collapse in the replicating strand and a 0.102% chance that the chromosome degrades. Although it may seem that these numbers are small; simulation of a population using these parameters, depending on the growth rate, causes the emergence of roughly 2% of anucleate cells. Zyskind et al. [1992] reports from 2% to 10% of RecA mutant population developing anucleate cells, depending on the version of the RecA mutant.

To experimentally determine if a cell strain returns the aberrant chromosome copy number (as described in Section 3.2.2), it is common practice to perform a drug treatment on an exponentially growing bacterial population with two antibiotics, rifampicin and cephalexin. The former stops the initiation of new replication forks, while allowing the ongoing replication forks to complete and the latter stops the cell dividing (Michelsen et al. [2003], Stokke et al. [2012], Hill et al. [2012]). The result is a population with integer numbers of chromosomes with no ongoing replication forks.

Using Skarstad and Boye [1993a] measured data of drug treated WT and recA1 *E. coli*, the HMG model was used to mimic the experimental steps of the authors and see if the simulated outcome matched the measured one in the following way. First the population was grown under assumption of exponential growth, to a 10000 cells. Upon reaching that number, the model simulates for drug treatment by assigning an unachievable critical mass to stop the occurrence of new replication forks, and an unachievable D time to stop the cell dividing. The simulation progresses until all replication forks complete to assure that the simulated population returns only integer numbers of chromosomes. The results are presented in Figure 4.11.

The scatter plots represent the reproduction of Skarstad and Boye [1993a] Figure 2.A. for WT MG1655 (orange) and Figure 2.E for ALS972 (blue) strains (a RecA mutant). The WT strain contains almost exclusively individual bacteria with 2 and 4 chromosomes. There is a small number of cells with 3 chromosome equivalents, caused by the stochastic nature of chromosome partitioning, chromosome initiation and replication progression that may well lead the populations to contain individuals with non 2^n number of bacteria. The mutant RecA strain on the other hand clearly expresses the aberrant chromosome copy number where the population contains a significant number of cells with 3 chromosome equivalents and especially anucleate cells.

The bar chart represents the simulated population using the original parameters (orange) as per Equation 3.11, or the optimised parameters (blue), after exponential growth with $\tau = 40.0$ min and a virtual drug treatment. The optimised parameters returned a significant higher number of cells with 3, 1 chromosomes and anucleate cells, while the majority of the population still contains two chromosomes equivalents. The simulation using the original parameters on the other hand returned a majority of cells with 2 and 4 chromosomes, with very few cells with 1 and 3 chromosomes and no anucleate cells. The latter correlates well with measured DNA content. The mutant simulation and measured did not have the same frequency of each individual chromosome numbers. This is perhaps evidence that the rate of DNA damage from the optimised parameters is too small compared to the experimental reality.

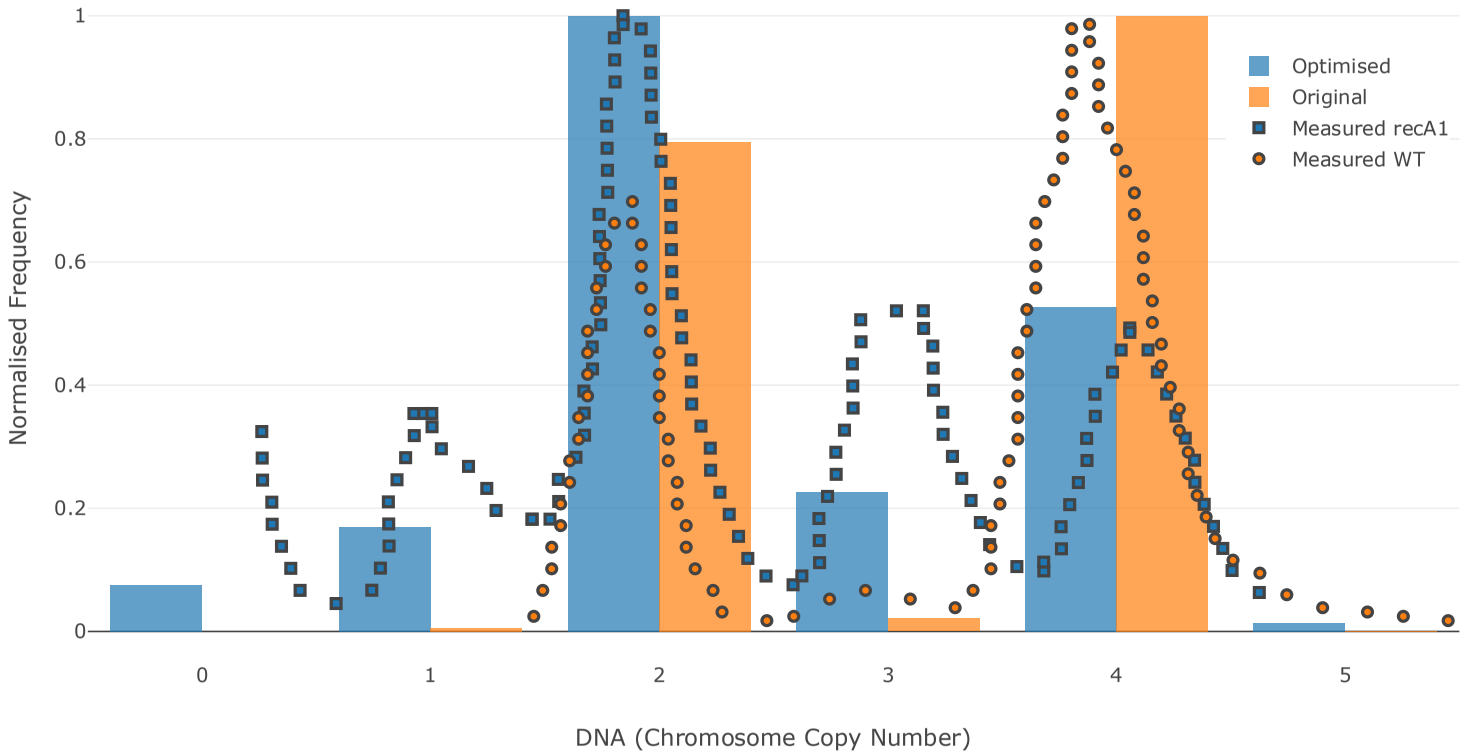


Figure 4.11: HMG Simulation with WT parameters and optimised parameters against measured DNA content for WT and recA1 mutant bacterial cells. Both simulations were virtually drug treated, where the model stops the division of a cell and initiation of new replication forks. The simulation was terminated when all cells terminated their replication forks. The measured DNA content are reproduced from Skarstad and Boye [1993a], where the WT is from MG1655 and the ALS972 (RecA mutant) strain.

4.3 Discussion and Conclusions

The HMG simulation protocol successfully reproduces the heterogeneity of populations as they progress through disparate growth regimes. Section 4.1.1 using literature parametrisation successfully simulated measured DNA distribution from exponential to late stationary growth. This demonstrates and confirms a few interesting observations. The first is the robustness of the mechanistic description of the cell cycle in the CH model. Although there may be some doubts on the molecular candidates and their dynamics in the control of the cell cycle (Section 1.2.1), the CH model is a robust formal summary of the main steps in the cell cycle. Features of rate-maintenance that were correctly simulated using HMG show that the way the parameters are updated and the eclipse period are descriptive enough to reflect such unusual growth dynamics. Next, the injection growth strategy being an accurate method to simulate for the volumetric growth of bacterial populations outside of balanced growth shows the central importance of the link between the cell cycle state of bacterial cells from their volume accumulation. This link is very much connected to the fundamental observation that the macromolecular contents of bacterial populations can be determined from the growth rate of the population alone (Bremer et al. [1996]). And growth rate is simply defined as the rate of volumetric increase of a population, or of a single cell, where this rate is enough to explain such a fundamental metabolic behaviour of the cell. The question then becomes what is the molecular mechanism that drives this behaviour?

The most widespread view is that bacterial cells actively monitor their environment and report changes to the metabolism of the cell that subsequently react to reported changes (Weart et al. [2007]). These include all types of stress responses such as the heat shock response that causes the expression of various proteins, such as chaperones, to protect the cell from the protein denaturation caused by the increased temperature (Alberts [2017]). This also involves the detection of carbon source, for example the infamous lac operon, that detects the presence of lactose as a carbon source and in the absence of glucose, causes the expression of proteins that enable the bacteria to transport and metabolise this specific carbon source (Schumann [2006]). A more recent proposition is the existence of "flux sensors", where the metabolic fluxes would first passively be altered due to changes in abundance in a particular nutrient and then these changes would be sensed by "flux sensors" who report the flux rate downstream to the cell's metabolism (Weart et al. [2007]). Such a control mechanism would logically fit well with the manner in which the cell cycle is described. Indeed, the injection growth method also increases the volume

of individual cells from the total population OD in a passive manner, and the cell cycle model then calculates the cell cycle state based on the volume accumulation patterns. The dissociation between the volume increase and the metabolism of the cell, and the causality from volume increase to the metabolic patterns seem to reflect correctly how the cell is organised.

A difficulty involved the convolution of simulated DNA distributions to enable the quantitative comparison to measured ones. Indeed, because the DNA distributions were measured using flow cytometry, a fluorescence microscopy technique, measurements of a discrete amount of DNA stained with DAPI or Hoetch returns a spread that is normally distributed (Section 2.2.1). As a consequence, one needs to convolute the DNA content from the simulation to enable the quantitative comparison. To perform such an action, the spread of the distinct peaks were measured after 24h of incubation, where the majority of the population returns integer amounts of chromosomes. It was observed that these standard deviations of the Gaussians fit to the peaks of integer amounts of chromosomes were linearly spread with the channel increase returned by the flow cytometer. However, the results of the spread may cause distinct and discrete DNA contents to overlap upon convolution, and might not return precisely the same noise convolution.

Despite these hurdles, the optimisation method returned sound minimas. Functional forms of the replication time (C) returned an increase in 21 min from the original functional forms. Considering the nature of the recA1 mutation, that produces a functional but reduced WT function in its SOS repair mechanism, replication errors would either fail to be repaired as we have implemented with the DNA damage parameters, or would reasonably take a longer time to be corrected as per the optimised C times. Furthermore, the results of the DNA damage return sound values (Section 4.2.4). Although one would need to experimentally determine the rate of replication fork collapse and whole chromosome degradation for this particular mutant strain, the optimisation results poses, soundly, that there is a higher likelihood that a cell experiences replication fork collapse compared to whole chromosome degradation (Michel et al. [1997], Kuzminov [1995], Goodman [2000]). Lastly the optimised segregation rate (D) returns a functional form that resembles the one as reported from the concatenation of literature D rates (Figure 3.9), where at slower growth rates the D times reported by Keasling et al. [1995] are much lower than both the optimised and the literature ones.

Chapter 5

Predicting Cell Cycle and SGC Properties Throughout Disparate Growth Regimes

This chapter explores some of the ways the HMG model may be used to either predict or calculate the dynamics of different aspects of the cell cycle. To illustrate the predictive capabilities, the copy number of a chromosomal gene based on its location and the input growth curve was calculated. Then to explore some the potential uses of this approach to synthetic biology, an ODE solver was implemented. The repressilator was used as an example and simulated in each cell in parallel with the mechanistic model of the cell. The effects of two inherently cell cycle properties, gene copy numbers and partition noise, on the repressilator are explored.

5.1 Determining Chromosomal Gene Copy Number

Biological noise may be categorised as being either intrinsic or extrinsic relative to the system of interest. In short, intrinsic noise may be qualified to be "local", that is, stochasticity inherent to chemical reactions. Simulating such behaviours is well characterised theoretically through the use of stochastic simulations, such as the Gillespie algorithm (Gillespie [2007], Swain et al. [2002]). Extrinsic noise on the other hand includes all other sources that may influence its dynamics. Because of the large potential sources of extrinsic noise, it can be difficult to quantitatively determine the influence of an individual source of extrinsic noise over another (Swain et al. [2002]). In this chapter the influence of different aspects of the cell cycle as a source of extrinsic noise on a given ODE is investigated through the use of the

HMG modelling strategy.

Gene copy number may have detrimental affect on the organism. For example, a human with an extra copy of chromosome 21 would develop Down's Syndrome, a condition caused by a metabolic shift due to the extra copy of this chromosome. Although bacteria commonly have many versions of chromosomes that do not seem to have such adverse effects, subtantial experimental evidence suggests that the spatiotemproal location of a gene along the chromosome can have effects on its expression patterns (Slager and Veening [2016], Sobetzko et al. [2012], Couturier and Rocha [2006]). Bryant et al. [2014] using a simple reporter cassette inserted at different positions along the chromosome, measured their corresponding outputs. The authors report up to a 300-fold differences in expression based on the location of the cassette. Although the authors conclude that copy number was not the sole culprit for these variations in expression, this combined with the natural organisation of genes along the bacterial chromosomes suggests that the location of the gene along the chromosome is an aspect that needs to be taken into consideration for robust predictions on its expression patterns (Sobetzko et al. [2012]). Furthermore, evidence suggests that due to chromosome compaction and folding into nucleoids, parts of the chromosome that are more exposed than others have been shown to have an effect on gene expression (Couturier and Rocha [2006]).

Because of chromosome replication, the amplitude and periodicity of a gene copy number changes over the course of the cell cycle; the dynamics of which depend on its location and on the dynamics of the cell cycle and leads to what is referred to as trasnsient gene dosage (Slager and Veening [2016], Couturier and Rocha [2006]). Bioinformatics investigation of chromosomal gene location based on their expression patterns throughout the cell cycle reveals that genes are naturally organised so that ones highly expressed in exponential growth are located closer to the *oriC* than ones more important in the stationary phase of growth located closer to the *terC* (Sobetzko et al. [2012]). For example, the NAP's that constitute the family of proteins that are responsible for the *E. coli* nuclei, are generally located close to the *oriC* (Sobetzko et al. [2012]). Likewise, genes with similar functions have been reported to be clustered together on the chromosome of bacteria (Slager and Veening [2016]). It is theorised that due to the single origin of replication with bidirectionality of the chromosome replication, the transient gene dosage effect causes the chromosome to naturally evolve in this fashion keeping similar gene dosage for a given cellular function (Slager and Veening [2016]).

Due to the natural organisation of the bacterial chromosome, and the potentially heterogeneous expression levels due to the gene copy number differences under

various growth conditions and within members of a same population, it can be of particular value to be able to calculate this effect. With the CH model, one is able to determine the gene copy number of a gene based on its relative location to the *oriC*, but only for exponentially growing populations (Keasling et al. [1995]). With HMG, the gene copy number from a wider range of conditions may be returned. In this dissertation, two different scenarios are investigated, one where a user would like to predict the gene copy number of a gene based on its location and a given growth curve. Another where given a growth curve and a chromosomal gene location one would like to deduce what the gene dosage affects are.

5.1.1 Predicting Gene Copy Number

To illustrate the copy number predictive feature of the model, the mean gene copy number of a series of growth curves were generated with various gene locations (Figure 5.1). Three different growth patterns are presented: fast growth (τ = 30 min), slow growth (τ = 80 min) and shift-up (τ = 50 min and τ = 30 min) for the exponential growth rates. In fast and slow growth, the simulation was initiated assuming exponential growth and then brought forward using the injection growth. Both these growth curves were generated where OD increase corresponds to their respective doubling rate for 300 min. Then the OD gradually decreases until stagnation for another 300 min. Lastly the OD remains the same for the last 100 min of the simulation to simulate for stationary phase. The OD was constructed in a similar fashion for the shift-up experiment, where the initial seed population was generated assuming exponential growth with $\tau = 50$ min and the same OD change was maintained for the first 100 min of the simulation. Then the OD gradually decreases for 100 min until stagnation that is maintained for another 100 min. Next the OD increases until reaching exponential growth ($\tau = 30$ min) and maintained over the following 200 min window. Lastly the OD decreases until stagnation for the last 200 min of the simulation.

The population with a doubling time of 80 min had a low gene copy number throughout the simulation. When growth stagnates, the population converges to single gene copy numbers. The standard deviation is low throughout, since a small number of cells were replicating themselves. At faster growth ($\tau = 30$ min) the gene copy number and the standard deviation was larger, evidence of the larger number of actively replicating chromosomes. As population growth stagnates, just like the slower growing counterpart, the gene copy number slowly decreases and converges to a single copy number.

As was covered in Section 3.1.4, the shift-up experiment is particularly inter-

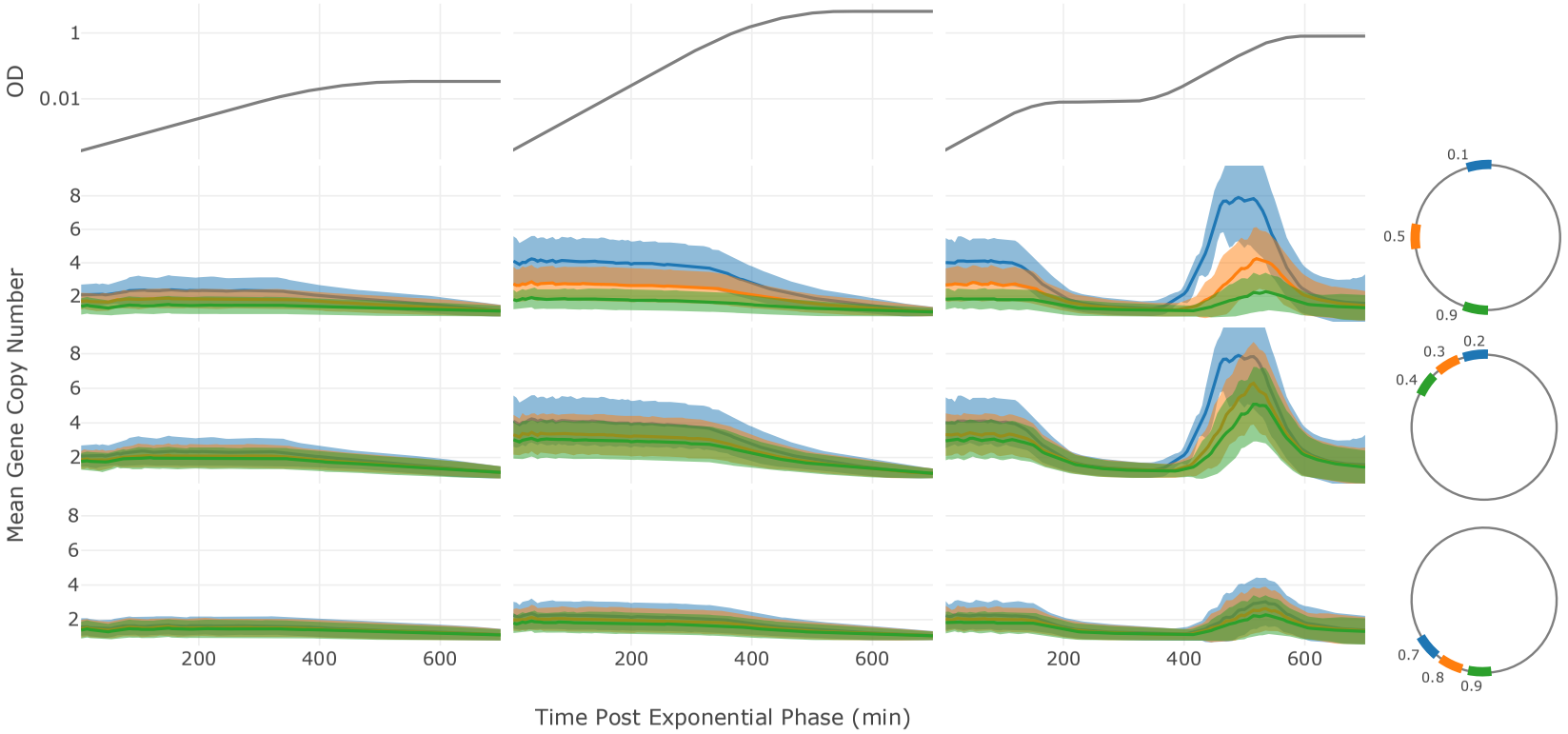


Figure 5.1: Population average gene dosage from three theoretical growth regimes (fast ($\tau = 30$ min), slow ($\tau = 80$ min), shift-up ($\tau = 50$ min and $\tau = 30$ min)). Three genes are accounted for at different distances from the *oriC*, where 0.0 is the *oriC* and 1.0 is the *terC*, illustrated on the right hand side.

esting because of the rate-maintenance feature of a population. As the simulation starts in exponential growth, the gene copy number is the same as the fast growing one. As the population begins to stagnate, the copy number of the population converges to a single copy number. However, once the population growth increases once again as it enters the second exponential phase, the gene copy number increases drastically. This burst returns a gene copy number that is particularly sharp for genes located close to the *oriC*. Although there is an increase in gene copy number for genes close to the *terC* this increase is less pronounced. As the population OD stagnates once again the gene copy number quickly falls, but with a higher gene copy number and more spread than that of the initial exponential growth. As the growth has stagnated, the population would not initiate a new rounds of replication and although with longer simulation time the gene copy number would decrease slightly, these gene copy number would be maintained. Thus, two populations that are in stagnation do not necessarily have the same DNA content, as it depends on the previous pattern of growth.

5.1.2 Calculating Gene Copy Number

Given a measured growth curve, one may use the HMG to determine the gene copy number as a population progresses through different growth regimes in batch culture. In Figure 5.2 the OD data from the SU (Figure 3.3) and WT K-12 cells grown in LB at 23 rpm and 230 rpm experiments (Figure 4.1) was used as input to the HMG reporting for the copy number as per Figure 5.2.

The results are very similar to Figure 5.1 for both the fast and slow growth conditions, where the only difference is that the measured OD's have a smoother degree of change in the gene copy number than that of the constructed ones. Globally the shift-up experiment returns the same dynamics, with only slight differences. The bottom right plot for shift-up with the three genes close to the *terC* shows a small burst in gene copy number right as the population enters the stationary phase. Although one would expect the copy number of these genes to be higher during exponential growth compared to the stationary, this burst shows that at this small moment in time, the number of completed chromosomes is high.

5.1.3 Conclusion

Overall this simulation returned three observations. The first is that genes located closer to the *oriC* would have a larger transient copy number than genes located closer to the *terC*. The second that the spread of genes copy number within a

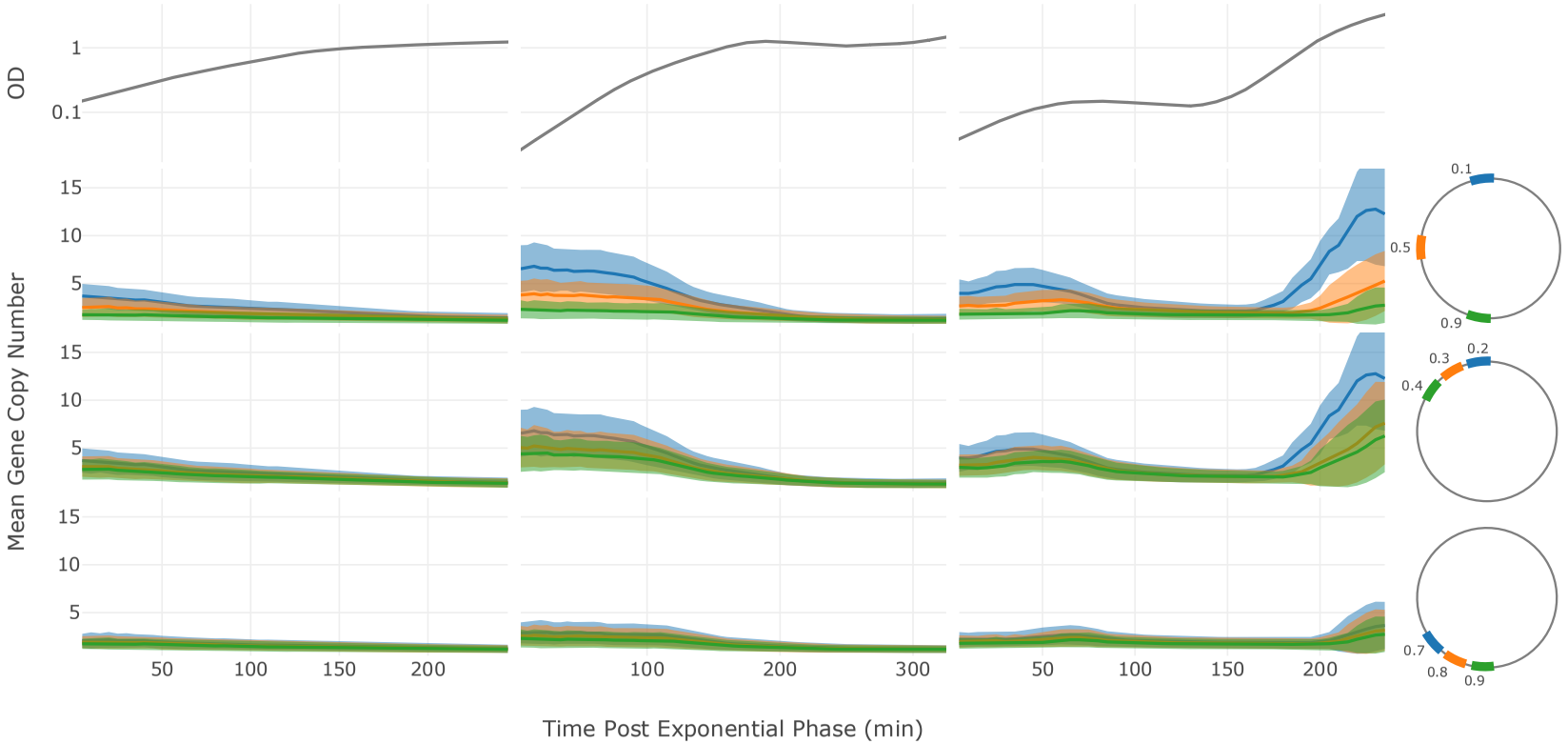


Figure 5.2: Population average gene dosage from three theoretical growth regimes (fast ($\tau = 22.58$ min), slow ($\tau = 39.58$ min), shift-up ($\tau = 31.17$ min). Three genes are accounted for at different distances from the *oriC*, where 0.0 is the *oriC* and 1.0 is the *terC*, illustrated on the right hand side.

population is greater for gene located closer to the *oriC* than ones located closer to the *terC*. And lastly even if a population is in exponential growth, the features of the population may not reflect as deterministically as the CH model predicts. Thus, depending on the growth conditions, the state of a population is determined by its growth history, where long term incubation in a constant environment is required to guarantee generating a population with cell cycle heterogeneity as predictable as the CH model does.

5.2 HMG ODE Simulation

One facet of synthetic biology is its attempt at using mathematical guidance for the rational design of genetic systems for novel purposes (MacDonald et al. [2011]). In this forward design approach, one typically creates ODE's that describe the dynamics and interactions of each part of a genetic system (Purnick and Weiss [2009]). Among the classic examples are the repressilator, the toggle switch, and other transcriptional AND/NOR gates (Osella and Lagomarsino [2013]). In the design of synthetic genetic circuit (SGC), it is common practice to ignore the state of the cell and study the dynamics of the construct in isolation when its expression is a consequence of the metabolism of the cell (Osella and Lagomarsino [2013]). If indeed the metabolism of the cell is taken into consideration, it is usually done with a simple growth rate parameter that describes the influence of the cell's metabolism on the SGC and only in rare occasions are the influence of a given SGC on the growth of the cell (Marguet et al. [2010], Osella and Lagomarsino [2013], Freudenau et al. [2015]). Making generalisation of the conditions of the expression of SGC may lead to unforeseen consequences on its expression. For example Tan et al. [2009], by expressing an inducible positive-feedback circuit system, noticed the emergence of bistability in expression of the reporter that was not intended in the design. The authors find that the interactions with the cell's metabolism is the culprit, where expression of the SGC causes growth retardation that in turn causes the observed phenotype.

To control for the inherent heterogeneity of bacterial populations that may influence the expression patterns of a SGC, a typical experimental procedure in characterisation of a part includes growing the bacteria exponentially from an overnight culture and sampling when reaching an OD of 0.1 (Heinze [2012]). The exponential growth ensures that balanced growth applies while the consistent OD ensures that the growth environment is the same. Under these growth conditions the heterogeneity of the population is considered to be predictable (see Section 3.1.2). Thus

two different constructs would theoretically endure the same influence from the population heterogeneity and single cell heterogeneity and as a result, population level measurements of the SGC or the proteins they generate are considered to be comparable. As a consequence, top down assumptions about the mechanism of the expression of a gene are made from population data since the heterogeneity is considered to be predictable (Stokke et al. [2012], Kell et al. [1991], Bremer et al. [1996], Dennis and Bremer [1974]). Although this makes sense theoretically, it can sometimes be difficult to perform such experiments. SGC frequently influence and interplay with the metabolism of the cell in unpredictable ways that may render many assumptions that are made about the population and cell invalid (Tan et al. [2009], Portle et al. [2007], Portle et al. [2009]).

With the HMG, the innate heterogeneity from the cell cycle of populations may be approached, theoretically, with more ease and outside of exponential growth. For the purposes of genetic engineering, an ODE solver was included in the HMG with the GNU Scientific Library (GSL), and combined with the mechanistic representation of the cell cycle, the influence of many aspects of the cell cycle may be studied. In this section the repressilator is used as a case study, and the influence of partition noise and transient gene dosage is investigated. This classic example of the forward design of synthetic biology is particularly good because of its reported difficult experimental application, where the amplitude of the construct are lost after a few generations within the population, and stable oscillations are difficult to perform experimentally (Potvin-Trottier et al. [2016]).

5.2.1 Partition Extrinsic Noise

Because proteins are usually quite stable (longer lifetime than the doubling of a cell), modelling their degradation rate in bacteria is usually simplified by the dilution rate experienced from the growth and division of the cell they are expressed in ($\beta = \mu \cdot ln2$) (Osella and Lagomarsino [2013]). In this model, the effect of dilution is not explicitly modelled and the original degradation terms from the ODE models are used instead. However the partition noise of a bacterial cell may have a large influence on the downstream heterogeneity of the population, such that it may affect the dynamics of a given SGC (Lloyd-Price et al. [2014]). Huh and Paulsson [2011] have shown that the noise associated with random partitioning can be indistinguishable to the stochastic effects of intrinsic noises on a system, even when one traces a reported protein at the single cell level.

Here the classic repressilator was implemented individually in each cell, varying the partition noise of its species upon division, to see the influence of the cell

cycle on a given ODE system. The original repressilator functional form was used:

$$\dot{m_i} = -m_i + \frac{\alpha}{(1 + p_i^n)} + \alpha_0 \tag{5.1}$$

$$\dot{p_i} = -\beta \cdot (p_i - m_i) \tag{5.2}$$

where m is the mRNA concentration and p is the protein concentration for the three genes (i). The parameters used in the simulation were $\alpha_0 = 0.03$, $\alpha = 300.0$, $n = 2.0$ and $\beta = 0.2$.

The instinctive way to implement the partition of the species of the ODE model seemed to be passive to the volume of daughter cells at division. A larger daughter cell would proportionally receive more members of each species of the ODE model than its smaller counterpart. However, the results return very predictable results (not shown), where the individual ODE's in the population behave as predicted since they experience very little extrinsic noise from the cell cycle. Implementing partition noise in this manner causes each specie of the ODE model to have their concentration reduced at the same ratio relative to each other. This results in the phase of the daughter cells to be exactly the same as the mother cell. Thus the population follows original oscillations.

Evidence suggests that upon division, the partition of long life mRNA is not equally distributed among daughter cells (Golding et al. [2005]). To test the effect of partition noise on the phases of the population ODE, a new random parameter was introduced that randomly distributed the partition of the ODE species into the daughter cells. This is based on a Gaussian distribution with mean 0.5 (where 0.5 is perfect partitioning between daughter cells) and varying standard deviation ($\sigma = 0.05$, $\sigma = 0.15$ and $\sigma = 0.25$). Figure 5.3 shows what happens to the oscillatory behaviour of the repressillator at the population level when the partition noise is increased. The top subplot is simulated with $\sigma = 0.05$ deviation. The amplitude of the oscillations decreases very slowly, but globally remains the same after relatively long simulation times. This means that the daughters cells at division essentially retains the same relative proportion of ODE species as their mother cell. Because there are no other sources of noise on the ODE, the simulation returns a population that shares the same phase. The middle subplot and bottom subplot show the results of the simulation with a partition noise of $\sigma = 0.15$ and $\sigma = 0.25$. In these cases, there is clear dampening of the oscillations, evidence that the population becomes less synchronised as the partition noise is large enough that some cells upon division receive a different proportion of the three proteins such that a daughter cell may end up in another phase compared to the mother cell. In other words, with a

large enough partition noise, the repressilator at the population level converges to a stable fixed point, the spread of which is dependent on the partition noise (Osella and Lagomarsino [2013]).

Since the HMG framework may handle non balanced growth parameters, and both these ODE's do not have growth rate parameters, the growth curves from Section 4.1.1 (WT LB 23 rpm and WT LB 230 rpm) and 3.1.4 (shift-up) were used as input. The results show that the partition noise is not growth rate dependent, at least under these growth rates.

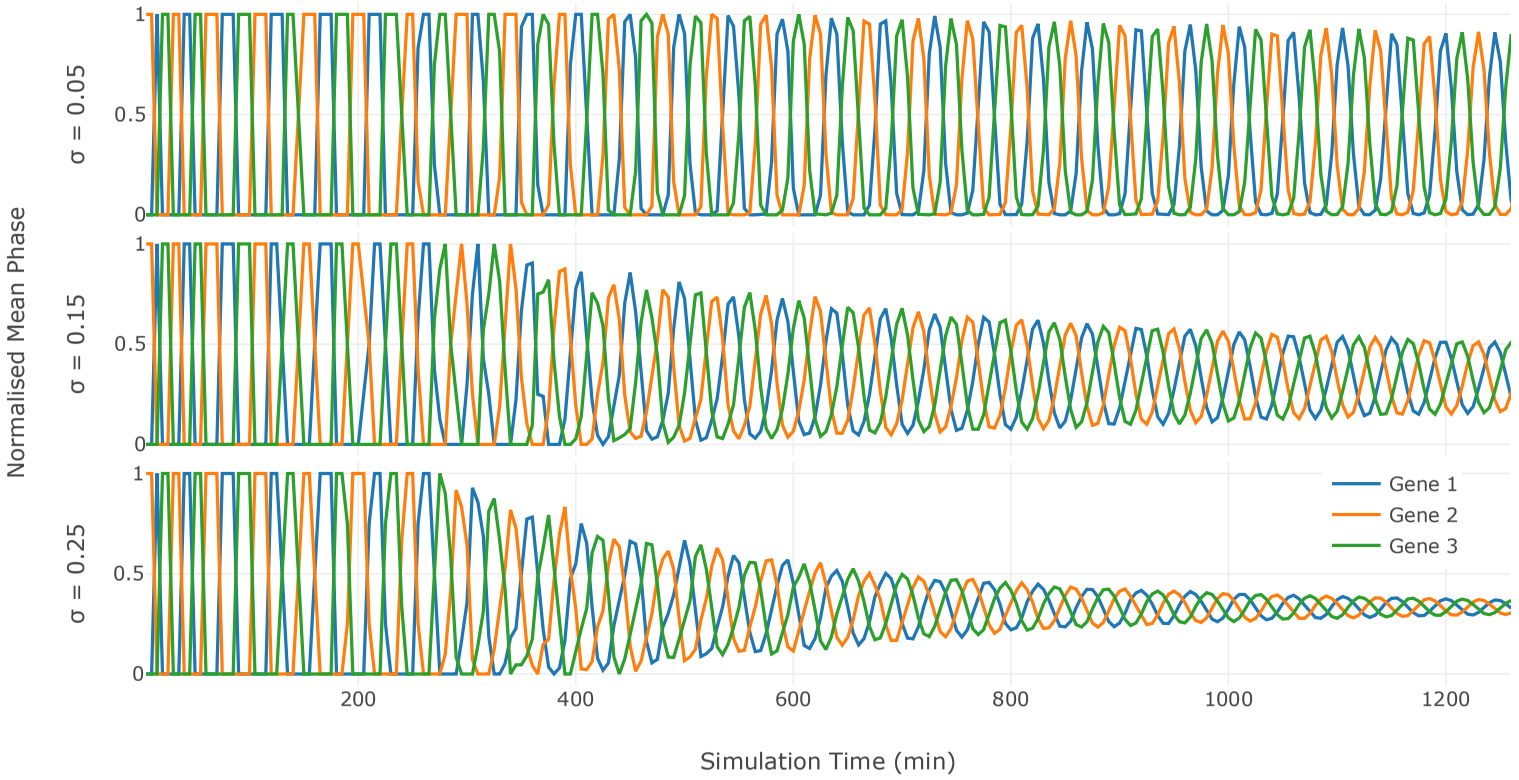


Figure 5.3: Mean population phases from original Repressilator model simulated with HMG assuming exponential growth ($\tau = 60$ min), with varying partition noises for the species of the ODE model based on Gaussian distribution ($\tau = 0.5$ and with top to bottom $\sigma = 0.05$, 0.15 and 0.25).

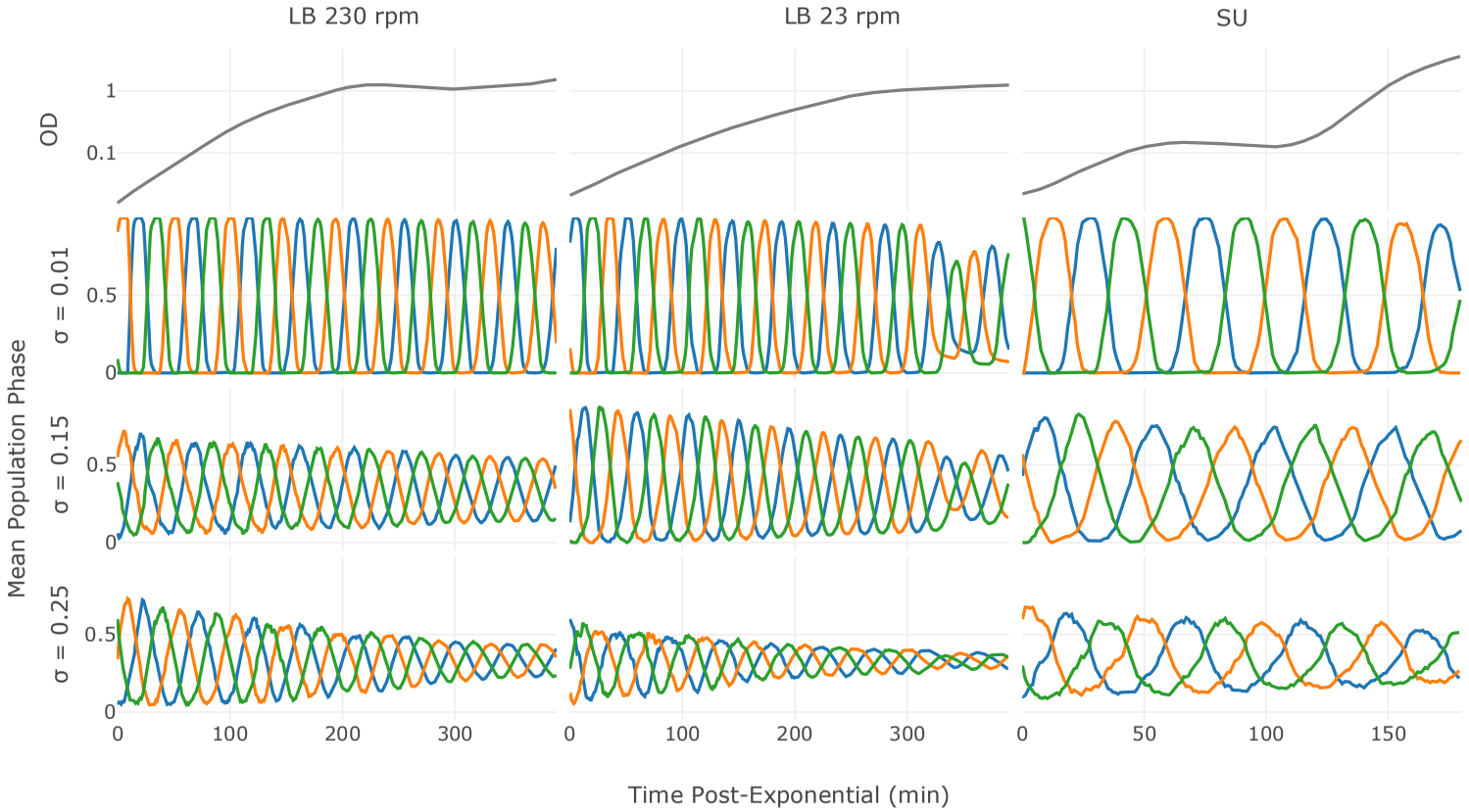


Figure 5.4: Mean population phases from ODE simulation of the original repressilator model (Equations 5.1) simulated with varying partition noise. The growth curves used correspond to the WT LB 23 rpm and WT LB 230 rpm (Section 4.1.1) and nutritional shift-up experiment (Section 3.1.4). The seed population was generated assuming exponential growth from the identified sections of exponential growth (Figure 4.1 and 3.3) with $\tau = 39.58$, $\tau = 22.58$ and $\tau = 31.17$ respectively. The rest of the simulation was brought forward using the injection growth method with the OD growth curves as presented on the top plot.

5.2.2 Transient Chromosomal Gene Copy Number

Equations 5.1 does not take into account the effects of gene copy number on the expression of mRNA's and the repressilator as a whole. Although commonly the repressilator is encoded *in vitro* in high copy pBR322 plasmids, others have investigated the theoretical effects of the oscillations at the single cell and population level based on the locations of the three cassettes that make up the repressilator based on their location on the chromosome (Elowitz and Leibler [2000], Osella and Lagomarsino [2013], Bennett et al. [2007]). To this end, Equations 5.3, 5.4, 5.5, 5.6 and 5.7 by Bennett et al. [2007] express the gene copy number as a function of the expression of the mRNA:

$$\dot{x}_i = -2\kappa_+ x_i^2 + 2\kappa_- y_i + \sigma m_i - \gamma_p x_i \tag{5.3}$$

$$\dot{y}_i = \kappa_+ x_i^2 - \kappa_- y_i - k_+ y_i d_{0,i} + k_- d_{r,i} \tag{5.4}$$

$$\dot{d_{0,i}} = -k_+ y_k d_{0,i} + k_- d_{r,i} \tag{5.5}$$

$$\dot{d_{r,i}} = k_+ y_k d_{0,i} - k_- d_{r,i} \tag{5.6}$$

$$\dot{m}_i = \alpha d_{0,i} - \gamma_m m_i \tag{5.7}$$

where $i \in \{1,2,3\}$, $j \in \{2,3,1\}, k \in \{3,1,2\}$. $d_{0,i}$ is the concentration of promoter i that is open and $d_{r,i}$ is the concentration of promoter i that is repressed and m_i the concentration of mRNA. x_i and y_i are the monomer and dimer concentration of the protein from gene i. κ are the dimerisation rates, k is binding rate of the dimers to their corresponding promoters as they are repressed, α the transcription rates, σ the translation rate and γ the degradation rates. To focus on the effect of gene copy number extrinsic noise on the repressilator, the model assumes that the partition of the species of ODE at division is done proportionally to the volume distribution between daughter cells.

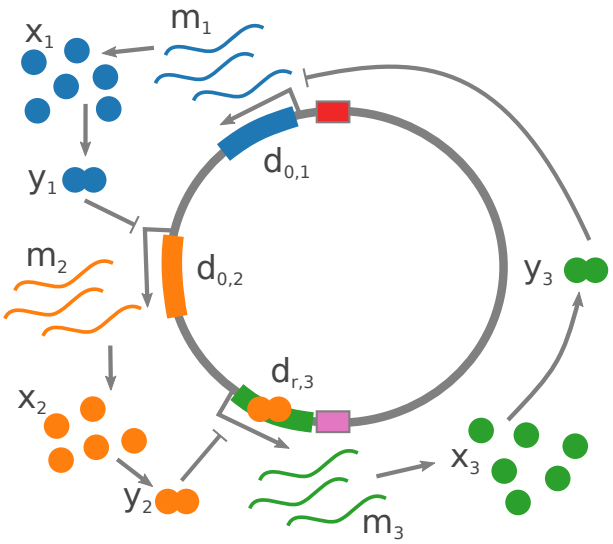


Figure 5.5: Graphic representation of Equations 5.3, 5.4, 5.5, 5.6 and 5.7 of the three promoters of the repressilator on the chromosome.

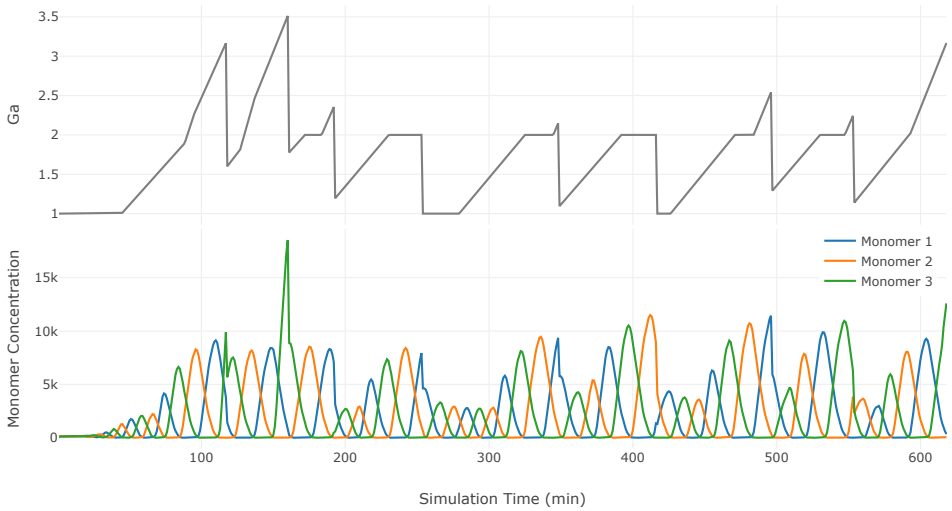


Figure 5.6: Monomer concentrations from Equations 5.3, 5.4, 5.5, 5.6 and 5.7 from a single cell grown in exponential growth ($\tau = 60$ min). The top subplot is the genetic content of the cell (Ga). The abrupt decreases in Ga represent the division moments, where a similar decrease may be observed in the monomer concentrations.

To investigate the effect of gene dosage based on the location of the gene along the chromosome on the repressilator model, the same growth rate under exponential growth was used ($\tau = 60$ min). The gene copy numbers and their standard deviation were returned for three genes located at a series of relative distances from the *oriC* in Figure 5.7. The gene dosage is illustrated in the middle subplot and shows, as expected, that the gene concentrations for the genes located closer to the *oriC* is larger than for genes located closer to the *terC*. Due to the low cell concentration at the start of the simulation the gene copy number and phases show sharp changes between 0 and 300 min simulation time. From 300 to 600 min simulation time, the oscillations become regular, albeit stochastically. During that time the amplitude of the oscillations becomes smaller while the phase become larger. From 600 min to the end of the simulation, the population contains a majority of cells in phase with the highest mean gene copy number that corresponds to the gene closest to the *oriC*.

Despite a larger mean copy number of gene 2, the frequency of cells in the population that are in that gene's phase is smaller than that of gene 3 when having a smaller copy number. Although this may seem counter-intuitive because the product of gene 1 inhibits the production of gene 2, and because the copy number of gene 1 is the highest, there is a higher expression rate of monomer for gene 1 whose dimer restricts gene 2. In turn, having a smaller protein concentration from gene 2 means that gene 3 can express itself more freely.

To have a greater picture of the gene dosage effect, the distance between each gene is brought closer together and the output population phase is returned in Figure 5.7. As the different genes are brought closer together, the difference in copy number between the three genes of the repressilator diminishes and the frequency of cells in their respective phases become increasingly similar, while at the same time the amplitude of the oscillations from each of the phases diminish. It is only once the gene locations overlap on the bottom subplot that the population actually oscillates between each phase. However, the oscillations seem to be decreasing, where the oscillatory behaviour of the population seems to be progressively lost. This is almost comparable to Figure 5.3 bottom subplot, of the original repressilator with a very high partition noise ($\sigma = 0.25$). Thus, it seems that the effect of chromosomal genetic heterogeneity alone, at this growth rate, is enough to disturb the oscillatory behaviour of this ODE model of the repressilator, in the same fashion as previously observed.

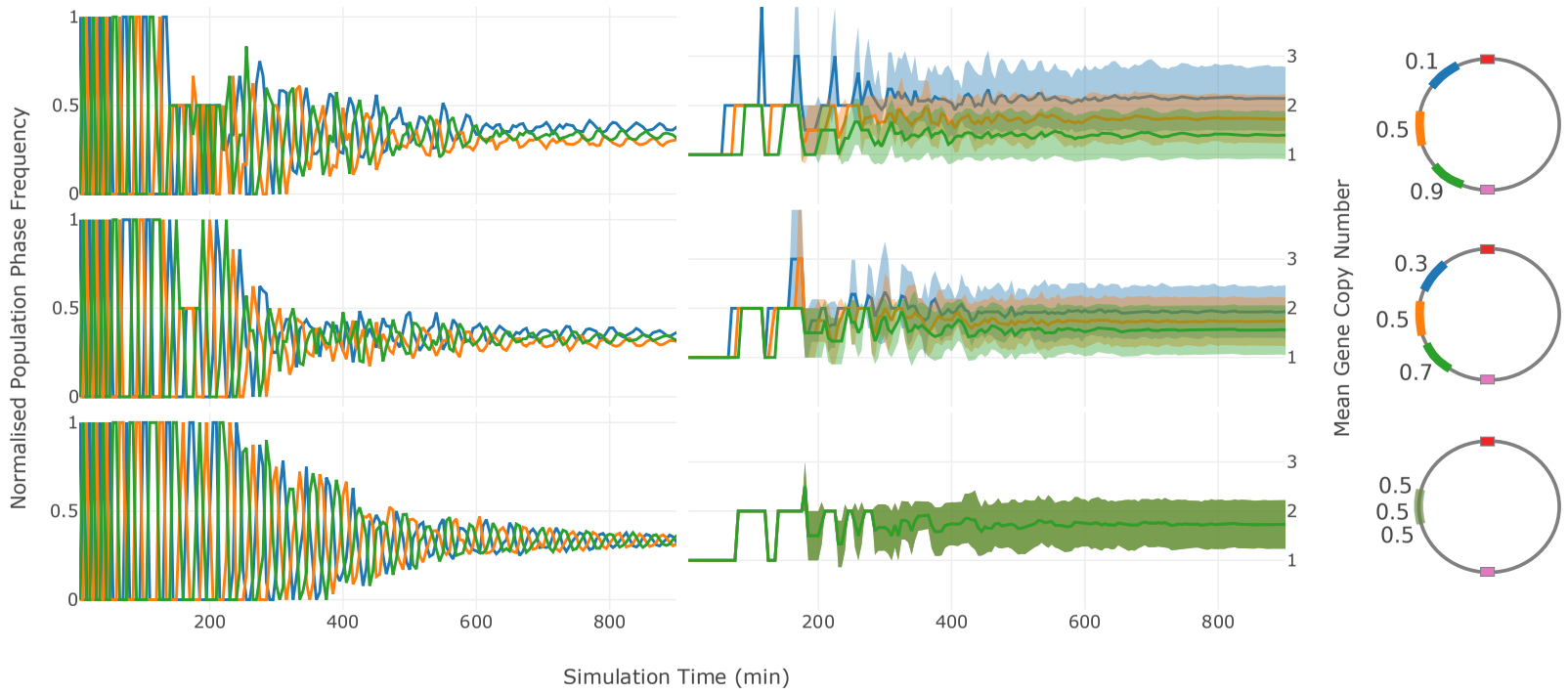


Figure 5.7: Mean population phases and gene copy numbers from the Repressilator with consideration for the gene dosage effect (Equations 5.3) simulated using the HMG assuming exponential growth ($tau = 60.0$) min with varying gene locations along the chromosome. The three genes were placed at: 10% , 50% and 90% from the *oriC* for the top subplot 30% , 50% and 70% from the *oriC* for the middle subplot, and all at 50% from the *oriC* for the bottom subplot. The left hand side shows the normalised mean population phases while the right hand side shows the gene copy number for the three genes. The phase is determined by the highest concentration of the three dimer proteins.

Similar to Section 5.2.1, as the HMG framework may handle non-balanced growth conditions, simulation of the ODE model was performed with different gene dosages across the shift-up, LB 230 rpm and LB 23 rpm conditions with WT K-12 cells to investigate the effects of changing OD of the population on a given SGC.

The results of the simulations are shown in Figure 5.8. The mean and standard deviation of the chromosome content of a population reduces significantly between the periods of exponential and stationary phase. Hence, the growth rate seems to have an influence on the oscillatory behaviour of the population since it affects the DNA content and its spread. With a faster growth rate, there is a larger amount of differences in copy number between each gene based on their location that in turns affects the oscillatory behaviour of the repressilator (see Figure 5.7). This can be seen between the LB 230 rpm and LB 23 rpm population, where the slower growing population oscillates when the genes are far apart, while the faster growing ones do not since faster growing population experience a more severe gene dosage difference than slower growing populations. Furthermore, the SU, by shifting from a slower to faster growth environment illustrates this phenomenon as well. During slower growth rates, regardless of how far apart the genes are the population oscillates and is maintained as the population enters the stationary phase of growth as the DNA content in these instances is small and consequently the gene dosage differences are small. As the population enters the faster state of growth, depending on the distance between each gene, then the population phase is either equally distributed or contains a majority of cells in phase corresponding to gene 1, then gene 3 and lastly gene 2.

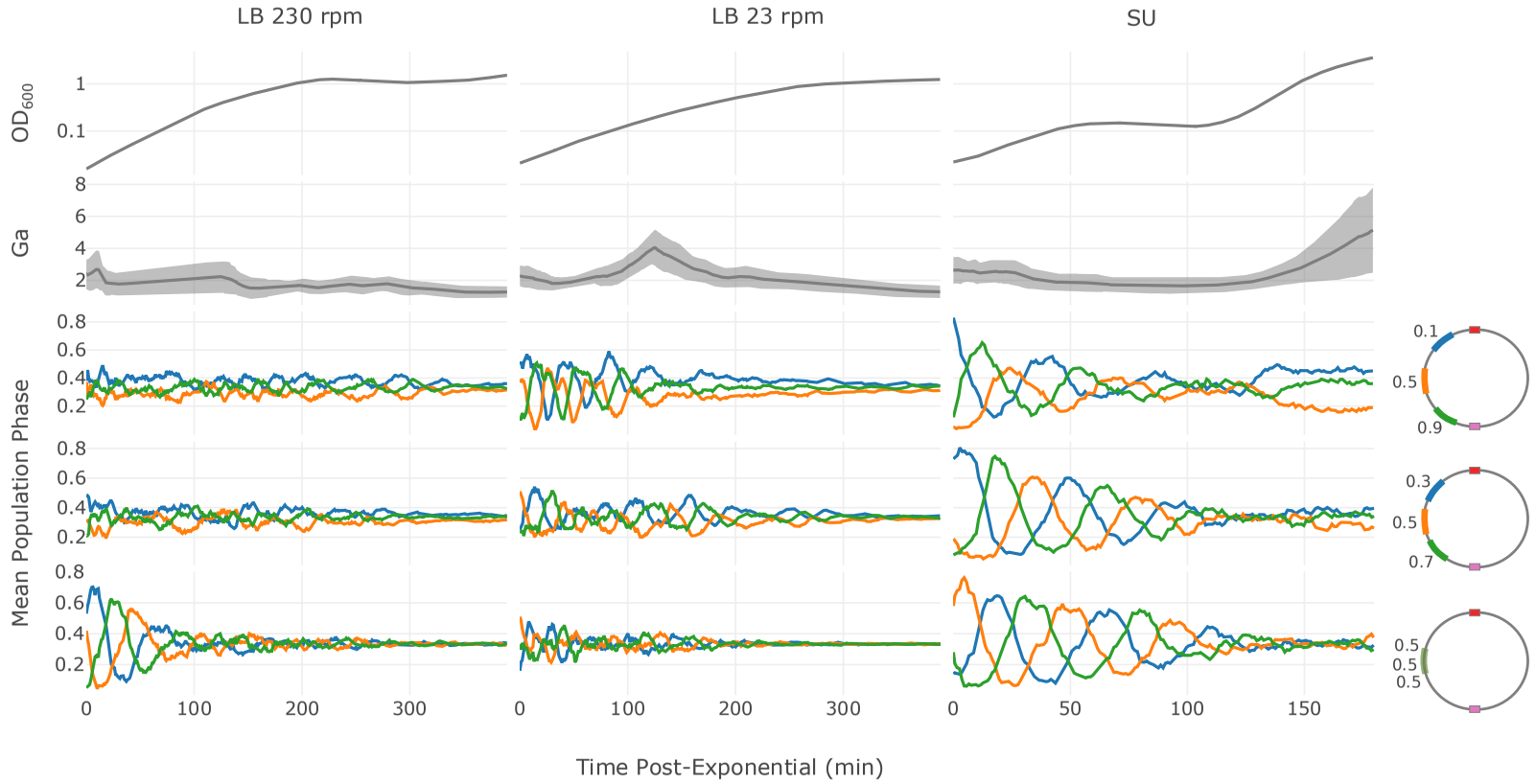


Figure 5.8: Mean population phase from the Repressilator with consideration for the gene dosage effect (Equations 5.3, 5.4, 5.5, 5.6 and 5.7) simulated using the HMG model with the injection growth method, with for input the growth curves from (Figure 4.1 and Figure 3.3). The simulation seed was initiated using the growth rate from the identified region of exponential growth for each corresponding OD growth curves, and brought forward using the injection growth method. Ga is the mean chromosomal genetic content of the population, where the grey filled line curves is the standard deviation. Mean population phase corresponds to the normalised proportion of individuals in a particular phase. The phase is determined by the highest concentration of the three dimer proteins. The right-hand side circles illustrate the chromosome where the red box is the *oriC* and the pink box is the *terC* and the location of the three genes of the repressilator, where 0.0 corresponds to the *oriC* location and 1.0 is the *terC* location along the chromosome.

5.3 Discussion and Conclusions

HMG is able to return the gene dosage of any gene on a chromosome as a population progresses throughout different growth dynamics. The results generally show that fast growing populations return a higher copy number than a population growing slowly, due to higher transient gene dosage effects. However, the degree of the copy number and the heterogeneity within a population is highly dependent on the location of the gene along the chromosome as well as the growth dynamics of the population as a whole. Genes located closer to the *oriC* have a higher gene copy number for a longer cell cycle time than ones located closer to the *terC*. This is particularly apparent with the shift-up experiment after LB growth that shows a high copy number for genes located close to the *oriC*.

Due to the nature of the HMG framework, the addition of an ODE solver facilitates the study of the effects of heterogeneity on a given SGC. In this section the effects of partition and gene dosage on the repressilator are explored. The results of the simulation assuming exponential growth are consistent with other works that have explored these sources of extrinsic noise on the system analytically (Bennett et al. [2007], Osella and Lagomarsino [2013], Gonze [2013], Bierbaum and Klumpp [2015]). The partition noise of molecular species of the repressilator only weakly affects the oscillations, while the gene dosage heterogeneity if expressed on the chromosome has a disastrous effect. One would however expect that partition noise would have a larger effect by reducing the copy number of the species of the model, where stochastic effects due to unequal partitioning would have a greater impact. Elowitz et al. [2002] showed that the intrinsic noise on a given SGC may be increased with a weaker promoter, and although there are a number of factors that come into play in the author's construct, partition noise is definitely a key player in these cases. The novelty lies in the ability to determine the extent of these types of extrinsic noises on a given SGC *a posteriori*, upon providing the growth dynamics of a given population.

Chapter 6

Summary, Conclusions and Further Work

6.1 Summary

The development of HMG first and foremost involved the revision and thorough inspection of the CH single cell cycle model. Much of the original way the cell cycle is described in the CH model remained the same with HMG. That included the trigger of the cell cycle expressed with the critical mass theory from Bremer et al. [1996] and chromosome replication and cell segregation described with two distinct timers, C and D respectively. Because the goal was not only the faithful prediction of population cell cycle dynamics but the exploration of a large parameter space through optimisation, it was important to have a deep understanding of the consequence of every assumption made in the CH model, and the evidence of such mechanisms with newer research and findings. Particular attention was given to the consequences of describing the start of segregation with the end of chromosome replication and the different theories that dictate single cell division patterns such as the sizer model, the timer model and the adder model of division. Although the adder model of cell division could be a very interesting aspect of the cell cycle model to explore, the classic mixed timer and sizer model was used for the HMG.

The eclipse period, a well established feature of the cell cycle that was not originally included within the CH model, was added in the HMG. The consequences of such a mechanism, called rate-maintenance, was experimentally measured and simulated using the HMG. Rate-maintenance was shown to be important for the faithful simulation of bacterial cell cycle throughout changing growth dynamics. This was a particularly important feature since one of the objectives of this disserta-

tion was the exploration of the cell cycle throughout non-balanced growth dynamics performed by the study of bacterial populations in the transition and stagnation phases of growth.

The cell cycle model was further expanded with the addition of the canonical mechanism of recA1 mutation on the cell cycle of a single cell. Because of the central role of the RecA protein in the SOS repair mechanism, there is a large body of work that suggests that the recA1 mutation causes the bacteria to have an impaired ability to repair errors in replication and some DNA damage. With the HMG, this is expressed with the degradation of a replicating strand or the degradation of whole chromosomes during replication. Because the rate of occurrence of such events had not been experimentally quantified, these parameters were optimised and the results were shown to accurately reflect measured aberrant chromosome copy numbers. This endeavor presented the HMG as a tool for the optimisation of new features the cell cycle *in silico*, given a growth curve and appropriate population measured experimental data.

During the development of the model, many growth models were reviewed and their assumptions and implications were inspected in an effort to identify the best type of growth model for the simulation of populations outside of assumptions of exponential growth. The conclusion was that all continuous models are inappropriate for purposes of simulation of populations during disparate growth regimes. Consequently, a novel growth method was developed that uses the constant between OD and total population volume from Keasling et al. [1995]. OD data was converted to their volumetric equivalent and distributed to individual cells in the population. This method was named the injection growth method, since it passively adds volume to each cell following a given OD input growth curve.

Lastly the HMG frameworks ability to be used as a predictive tool for gene copy numbers was illustrated in Section 5.1. Given either a measured or input growth curve, the HMG is able to determine the population mean gene copy number, and inspect the heterogeneity within the population as it progresses throughout different growth regimes. In Section 5.2, the integration of an ODE solver in each cell and simulated in parallel with the cell cycle model, created a marriage between the discrete and continuous modelling strategies. This permits the *in silico* analysis of cell cycle heterogeneity influencing a given ODE system.

6.2 Conclusion

The development of an IBM simulation method for the specific use of modelling cell cycle heterogeneity of bacterial populations, and especially the addition of an ODE solver that is influenced by the cell cycle, has not been done to date (including Keasling et al. [1995] Monte Carlo simulation). There are a large number of IBM software packages, with only a few that have been applied to specific biological problems, such as ecological and intra-species interactions (ex: INDISIM-SOM) (Gras and Ginovart [2006]), or 3-dimensional organisation of colony formation (ex: BacSim) (Kreft et al. [1998]). With HMG, a number of inherent mechanistic features of the bacterial cell cycle and its emergent behaviour as a population may be explored theoretically. Simulation using HMG under assumptions of exponential growth faithfully reproduces measured DNA distributions (that are also in par with CH model simulation results). Thereafter, in combination with the injection growth method, HMG is descriptive enough to simulate for the dynamics of populations outside of exponential growth and throughout a wider range of growth regimes than could be previously achieved. Indeed, the injection growth method developed in this research has been shown to be a valid approach at describing the volumetric changes of a population throughout a wide range of growth conditions. The assumptions made on how these population volumetric changes are redistributed to the individuals within the population are certainly an oversimplification (see Section 3.1.3), but nonetheless seem to be descriptive enough that the downstream cell cycle effects are correctly reflected using this method. Through the CH model, one was able to determine the chromosomal content of a population theoretically only if the population was growing exponentially. With the HMG, another boundary was explored of post-exponential population cell cycle dynamics. However, the accuracy of such a modelling method is bound by the accuracy of its parameterisation, and more meticulous experimental data would serve the HMG framework well.

There is a large body of work in the field of ecology that use optimisation of agent based models (ABM) or IBM to explore for the required underlying behaviours of the agents to reflect global observations. This approach is equally applicable to the study of bacteria, as shown in Section 4.2. In this example, the HMG framework was successfully used to optimise for the canonical effects of recA1 mutants on the SOS repair mechanism and ultimately the cell cycle. This version of the protein is widely used for the purposes of genetic engineering. For example, Elowitz et al. [2002] who designed two SGCs inserted in the bacterial chromosome to distinguish between extrinsic and intrinsic noise, used a RecA *null* mutant to increase the

intrinsic noise from the cell cycle. The results of the optimisation show that for recA1, the minimal replication time for a single chromosome increases by 20 minutes overall. Due to the central role of this protein in the SOS repair mechanism, and because this particular version of the mutant has a reduced function compared to its WT counterpart, it is expected that any DNA damage in the replicative process that would require the participation of RecA to be repaired, would take a longer time to be overcome thus increasing the overall replicative time. The segregation and duplication rate (D) optimised to recA1 data did not yield results that are specific to this mutant strain. Indeed, the optimisation results returned the same minimal time as literature reports at fast growth rates, while at smaller growth rates contained a longer time. This result differs from the functional form proposed by Keasling et al. [1995] and instead agrees with the functional fit to concatenation of literature parameters (Figure 3.9) for WT K-12 bacterial cells. Lastly the rate of DNA degradation returned by the optimisation, caused by the loss of a fully functional RecA seems to be sound. Loss of a fully functional RecA causes the collapse of replicating forks with a higher proportion than whole chromosome degradation, and where the production of anucleate cells and the ratio of aberrant chromosomes seem to be in par with literature reports (Section 4.2). Overall the optimisation procedure was surprisingly successful, and probably stems by the fact that the optimisation was framed in very much the same way the original data was measured in. Posing the problem in this way reflected the emergent properties of the system, where searching the parameter space of the individuals instead of the population directly makes for a more difficult problem to solve, but more precise results that are more easily interpretable.

The innate influence of heterogeneity from a growing bacterial population on the expression of genetic networks was explored in Chapter 5, through the specific examples of gene dosage and partition noise. Copy number control has been the focus of a large body of work on human disorders (Stranger et al. [2007], Schwanhäusser et al. [2011]) and eukaryotes in general, but is overall lacking in the field of bacteriology. There are a few theoretical investigations (Bennett et al. [2007], Gonze [2013], Paijmans et al. [2016]), and a few experimental procedures to control for the copy number heterogeneity in bacteria (Wang and Kushner [1991]). However, the influence of such a factor is largely ignored in the prediction of gene expression. Investigation of chromosomal gene dosage revealed that gene copy number is highly dependent on both its location on the chromosome and the growth dynamics of the population (Sobetzko et al. [2012]). One must thus take into consideration both these aspects, and with the HMG one is able to determine such features of a popula-

tion equally *a priori* and *a posteriori* throughout exponential and post-exponential growth. This may prove useful in deconvoluting experimental quantification of gene expression at the population level, or to predict the required growth dynamics to meet a particular gene copy number. The effects of gene dosage were further investigated by implementing an ODE solver in the framework, modelling the repressilator, and studying its effects. First, partitioning noise from expression of the repressilator was investigated. The results showed that the oscillatory behaviour in a population was maintained unless a very high partition noise was enforced. Then, given a range of different gene distances between the three genes of the repressilator, the oscillatory behaviour of the construct was evaluated with different growth patterns. The results showed to have a considerable impact if one considers that there is a one to one ratio between copy number of transcription rate (where for example twice the gene numbers lead to twice the transcription rate). Although the original repressilator experiment was expressed in plasmids and not on the chromosome, transient gene copy number differences combined with unequal partitioning for the three products of the genes of the repressilator could explain why this construct was observed to loose its oscillatory behaviour after a few generations.

The combination of a mechanistic description of the bacterial cell cycle and an ODE solver opens the door to a whole set of exploration of a given SGC behaviour in a population. Its design can take into account the heterogeneity of a bacterial population. For example, consider the following works: Portle et al. [2007] and Portle et al. [2009]. The authors explore the dynamics of two different well characterised SGC, the repressilator and toggle-switch. Using inducible versions with aTc and IPTG and using different GFP reporter half-lives, it was found that the repressilator expressed three different bi-threshold states and multiplicities, while the toggle switch showed two unimodal and one bimodal behaviour based on the concentration of inducers. Although the authors propose ODE models that qualitatively reflect their findings, it would be of particular interest to reproduce their findings using the HMG, that takes into account extrinsic noises from the cell cycle as well as the intrinsic noise from the construct and concentration of the inducible promoters. A diffusion reaction of IPTG or aTc from the media to the individual cells would need to be implemented, but thanks to the fact that throughout this and their work bacteria are grown in well mixed environment, it should be relatively easy to implement. Generally with the HMG, one can investigate the cell cycle's role in different SGC modelled as ODEs.

6.3 Further Work

With more recent single cell data, many assumptions of the bacterial cell cycle are put to question, that were covered in the model development section (Chapter 3). These may further be explored with the HMG simulation framework, and would require the simultaneous quantification of single cell volumetric and DNA content (technically feasible with microfluidic devices combined with microscopy). The main source of uncertainty in the formalisation of the bacterial cell cycle is the D period and its timing. Experimental quantification of the D time returns much more variable values at a single growth rate than its C counterpart (Figure 3.9), and the triggers for the period have been shown to not be controlled by the end of the replication process as proposed by the CH model. This suggests that although there must be a link between the two, or blocking the replicative process would not block the division process, the link needs not be as strong as suggested by the CH model.

Furthermore, it would be particularly interesting to implement the adder model (See Section 3.3.3), and explore to see if population size homeostasis is maintained with the same dynamics as measured experimentally and particularly if DNA distributions derived from such population also behave as measured. Indeed, the adder model of growth does not negate the critical mass theory, and the two may be combined (Campos et al. [2014]). The results would be a model with critical mass that describes the chromosome dynamics, and the adder model that describes the volumetric dynamics and division of bacterial population, only in this case the chromosome dynamics would have no influence on the division time of the cell.

Another aspect that must be taken into consideration is cell death, and may be considered a limitation of the HMG model. Indeed, if at a given timestep there is a significant amount of cell death, then the distribution of total population volumetric growth would be influenced since the increase would be subdivided to a smaller population. It is assumed that such an influence on the global volumetric changes is minor, but there are conditions and strains where cell death has a larger influence. For example, the recA1 formed a number of anucleate cells due to whole chromosome degradation. These cells were taken out of the pool of cells in the simulation and thus where treated as dead cells. Thus implementation of the canonical cell death rates would be particularly important for the HMG method.

The HMG framework may prove to be useful for the field of synthetic biology for the design and prediction of SGC, but would require the implementation of plasmid dynamics in the framework. Although there are some chromosomal genetic manipulations used in the field, the use of plasmids as vectors for the purposes of

genetic engineering is overwhelming. Furthermore, there is quite a large diversity of different origin of replication that dictate the copy number, replication dynamics and partition dynamics of plasmids that need to be taken into consideration. One family, called minichromosomes, has an *oriC* shared with chromosomes while needing an external supply of DnaA to replicate themselves. Their behaviour is very close to chromosomes ones, and thus, have already been implemented in HMG. However, more experimental work would be required to see the influence of minichromosomes on the replication of chromosomes and measure different aspects of their heterogeneity at the single cell level before it may be predictive. Other types of plasmid, have independent negative-feedback replicative initiation mechanisms with different degrees of relaxed or stringent controls that dictate their copy number, that would also require more experimental work to model their mechanistic replicative behaviour. Moreover, for some plasmids a mechanistic description would prove not to be useful, such as ColE1. Indeed, Kuo and Keasling [1996] show through their Monte Carlo investigation of the high copy number plasmid ColE1, that due to the replicative and partitioning nature of this specific plasmid, that it is more appropriate to describe it using continuous equations than through mechanistic ones. Nevertheless, an accurate description of some low copy plasmid dynamics throughout disparate growth regimes would prove to be useful for the study of plasmid behaviour *in vitro* as well for the design of the SGC; especially considering that some evidence suggests that low copy plasmids can perform as well or better than their more popular high copy counterparts (Jones et al. [2000]).

In this research, cell cycle heterogeneity was only studied as it may influence a given SGC and not the other way around. The expression of a given SGC has an influence on cell metabolism by using part of its transcription and translation machinery (also called metabolic load or burden) (Ebersbach and Gerdes [2005]). Depending on the strength of the promoter and the copy number of a SGC, then its expression has a range of different influences on the cells metabolism. With HMG, the cell's metabolism is summarised in the doubling time term (τ) and could be modulated by a SGC to simulate for metabolic load, or one could either extend the model and implement for example pools of ribosome, ATP, DNA polymerase, etc... and have a more descriptive single cell model.

On the computational side, there are a few improvements that can be done. First is the implementation of the Systems Biology Markup Language (SBML) standard into the ODE solver. Indeed, in its current form, a user must hardcode a given a set of ODE equations into the model and compile it before running it. This is impracticable and even unapproachable for many that are not familiar with coding.

Furthermore, using a quite powerful machine (see Section 2), simulation of ODE models combined with the cell cycle model returned very slow execution time as the number of individuals in the simulation > 5000, evidence of the high computing power required to solve the problem. One of the possibilities is to use Message Passing Interface (MPI) as a means of leveraging parallel computing to make simulating a larger number of individuals more efficient (a solution particularly attractive since MPI is compatible with the GSL ODE solver). Furthermore, in HMG, because ODE are solved at the single cell level, implementation of a stochastic ODE solver would be useful to account for the influence of intrinsic noise on a given a set of ODE equations.

The inherent complexity of biological systems is partly attributed to their heterogeneity, be it molecular, temporal or genetic. Ironically, this heterogeneity is the aspect of these systems that makes them so powerful and at the same time so difficult to understand, let alone control. This work takes the view that if one desires to successfully and robustly manipulate such systems, one needs to be able to understand and predict such an aspect of biological systems. With the HMG, one may theoretically study the heterogeneity of the cell cycle and its potential influence on SGC with more ease. However much more research is required to understand the dynamics of a bacterial population as it progresses in disparate growth regimes, and our limited ability to predict it only attests for our limited understanding of how bacterial cells grow and divide.

Bibliography

Abner, Kristo; Aaviksaar, Tõnis; Adamberg, Kaarel, and Vilu, Raivo. Single-cell model of prokaryotic cell cycle. *Journal of Theoretical Biology*, 341:78–87, 2014. ISSN 00225193. doi: 10.1016/j.jtbi.2013.09.035.

Adams, David W and Errington, Jeff. Bacterial cell division: assembly, maintenance and disassembly of the z ring. *Nature Reviews Microbiology*, 7(9):642–653, 2009.

Adiciptaningrum, Aileen; Osella, Matteo; Moolman, M Charl; Lagomarsino, Marco Cosentino, and Tans, Sander J. Stochasticity and homeostasis in the e. coli replication and division cycle. *Scientific reports*, 5:18261, 2015.

Akerlund, Thomas; Nordström, Kurt, and Bernander, Rolf. Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of Escherichia coli . These include : Analysis of Cell Size and DNA Content in Exponentially Growing and Stationary-Phase Batch Cultures of Escherichia coli. *Journal of Bacteriology*, 177(23):6791–6797, 1995. ISSN 00219193.

Alberghina, Lilia and Westerhoff, Hans V. *Systems biology: definitions and perspectives*, volume 13. Springer Science & Business Media, 2007.

Alberts, Bruce. *Molecular biology of the cell*. Garland science, 2017.

Allman, R; Schjerven, T, and Boye, E. Cell cycle parameters of Escherichia coli K-12. *J. Bacteriol.*, 173(2):7970–7974, 1991. ISSN 0021-9193.

Amir, Ariel. Cell size regulation in bacteria. *Physical Review Letters*, 112(20): 208102, 2014.

Andrianantoandro, E; Basu, S; Karig, DK, and Weiss, R. Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology*, 2:2006.0028, 2006. ISSN 1744-4292. doi: 10.1038/msb4100073.

Balaban, Nathalie Q; Merrin, Jack; Chait, Remy; Kowalik, Lukasz, and Leibler, Stanislas. Bacterial persistence as a phenotypic switch. *Science (New York, N.Y.)*, 305(5690):1622–1625, 2004. ISSN 0036-8075. doi: 10.1126/science.1099390.

Basan, Markus; Zhu, Manlu; Dai, Xiongfeng; Warren, Mya; Sévin, Daniel; Wang, Yi-Ping, and Hwa, Terence. Inflating bacterial cells by increased protein synthesis. *Molecular systems biology*, 11(10):836, 2015.

Bates, David; Epstein, Jessica; Boye, Erik; Fahrner, Karen; Berg, Howard, and Kleckner, Nancy. The Escherichia coli baby cell column: A novel cell synchronization method provides new insight into the bacterial cell cycle. *Molecular Microbiology*, 57(2):380–391, 2005. ISSN 0950382X. doi: 10.1111/j.1365-2958.2005.04693.x.

Bennett, Matthew R and Hasty, Jeff. Microfluidic devices for measuring gene network dynamics in single cells. *Nature reviews Genetics*, 10(9):628–638, 2009. ISSN 1471-0064. doi: 10.1038/nrg2625.

Bennett, MR; Volfson, D; Tsimring, L, and Hasty, J. Transient dynamics of genetic regulatory networks. *Biophys J*, 92(10):3501–3512, 2007. doi: 10.1529/biophysj.106.095638.

Bernander, Rolf and Nordström, Kurt. Chromosome replication does not trigger cell division in e. coli. *Cell*, 60(3):365–374, 1990.

Bierbaum, Veronika and Klumpp, Stefan. Impact of the cell division cycle on gene circuits. *Physical biology*, 12(6):066003, 2015. ISSN 1478-3975. doi: 10.1088/1478-3975/12/6/066003.

Bipatnath, Meeshel; Dennis, Patrick P, and Bremer, Hans. Initiation and Velocity of Chromosome Replication in Escherichia coli B / r and K-12. *Journal of Bacteriology*, 180(2):265–273, 1998.

Blount, Zachary D. The natural history of model organisms: The unexhausted potential of e. coli. *Elife*, 4:e05826, 2015.

Bogan, J and Helmstetter, C. DNA sequestration and transcription in the oriC region of Escherichia coli. *Molecular microbiology*, 26:889–896, 1997. ISSN 0950-382X. doi: 10.1046/j.1365-2958.1997.6221989.x.

Boye, E; Løbner-Olesen, Anders, and Skarstad, Kirsten. Limiting DNA replication to once and only once. *EMBO reports*, 1(6):479–483, 2000. ISSN 1469-221X. doi: 10.1093/embo-reports/kvd116.

Boye, Erik; Lobner-Olesen, Anders, and Skarstad, Kirsten. Timing of chromosomal replication in Escherichia coli. *BBA - Gene Structure and Expression*, 951(2-3): 359–364, 1988. ISSN 01674781. doi: 10.1016/0167-4781(88)90107-8.

Boye E, Nordstrm K. Coupling the cell cycle to cell growth. *EMBO Rep*, 4(8): 757–760, 2003. doi: 10.1038/sj.embor.embor895.

Bremer, H; Dennis, Patrick P, and others, . Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*, 2(2):1553–69, 1996.

Brewer, Bonita J. and Fangman, Walton L. Mapping replication origins in yeast chromosomes. *BioEssays*, 13(7):317–322, 1991. ISSN 1521-1878. doi: 10.1002/bies.950130702.

Browning, Samuel T.; Castellanos, Mariajosé, and Shuler, Michael L. Robust control of initiation of prokaryotic chromosome replication: Essential considerations for a minimal cell. *Biotechnology and Bioengineering*, 88(5):575–584, 2004. ISSN 00063592. doi: 10.1002/bit.20223.

Bryant, Jack A; Sellars, Laura E; Busby, Stephen JW, and Lee, David J. Chromosome position effects on gene expression in escherichia coli k-12. *Nucleic acids research*, 42(18):11383–11392, 2014.

Buchanan, RL; Whiting, RC, and Damert, WC. When is simple good enough: a comparison of the gompertz, baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiology*, 14(4):313–326, 1997.

Campbell, Allan. Synchronization of cell division. *Bacteriological Reviews*, 21(4): 263, 1957.

Campos, M; Surovtsev, IV; Kato, S; Paintdakhi, A; Beltran, B; Ebmeier, SE, and Jacobs-Wagner, C. A constant size extension drives bacterial cell size homeostasis. *Cell*, 159(6):1433–1446, 2014. doi: 10.1016/j.cell.2014.11.022.

Carbonell, Pablo and François, Jean Marie. *Synthetic Biology engineering complexity and refactoring cell capabilities*. Frontiers Media SA, 2015.

Cooper, S and Keasling, Jd. Cycle specific replication of chromosomal and F plasmid origins. *FEMS microbiology letters*, 163:217–222, 1998.

Cooper, Stephen. Cell division and dna replication following a shift to a richer medium. *Journal of molecular biology*, 43(1):1–11, 1969.

Cooper, Stephen. What is the bacterial growth law during the division cycle? *Journal of bacteriology*, 170(11):5001, 1988.

Cooper, Stephen. *Bacterial growth and division: biochemistry and regulation of prokaryotic and eukaryotic division cycles*. Elsevier, 2012.

Courcelle, Justin and Hanawalt, Philip C. RecA-dependent recovery of arrested DNA replication forks. *Annual review of genetics*, 37:611–646, 2003. ISSN 0066-4197. doi: 10.1146/annurev.genet.37.110801.142616.

Couturier, Etienne and Rocha, Eduardo PC. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular microbiology*, 59(5):1506–1518, 2006.

Cox, M M; Goodman, M F; Kreuzer, K N; Sherratt, D J; Sandler, S J, and Marians, K J. The importance of repairing stalled replication forks. *Nature*, 404(6773):37–41, 2000. ISSN 0028-0836. doi: 10.1038/35003501.

Dalgaard, Paw; Ross, Thomas; Kamperman, Laura; Neumeyer, Karina, and McMeekin, Thomas A. Estimation of bacterial growth rates from turbidimetric and viable count data. *International Journal of Food Microbiology*, 23(3-4): 391–404, 1994. ISSN 01681605. doi: 10.1016/0168-1605(94)90165-1.

Den Blaauwen, Tanneke; Buddelmeijer, Nienke; Aarsman, Mirjam EG; Hameete, Cor M, and Nanninga, Nanne. Timing of ftsz assembly in escherichia coli. *Journal of bacteriology*, 181(17):5167–5175, 1999.

Dennis, P P and Bremer, H. Macromolecular composition during steady-state growth of Escherichia coli B-r. *Journal of bacteriology*, 119(1):270–281, 1974. ISSN 0021-9193.

Donachie, William D. Relationship between cell size and time of initiation of dna replication. *Nature*, 219(5158):1077–1079, 1968.

Donachie, William D and Blakely, Garry W. Coupling the initiation of chromosome replication to cell size in escherichia coli. *Current opinion in microbiology*, 6(2): 146–150, 2003.

du Lac, Melchior; Scarpelli, Andrew H; Younger, Andrew KD; Bates, Declan G, and Leonard, Joshua N. Predicting the dynamics and heterogeneity of genomic dna content within bacterial populations across variable growth regimes. *ACS Synthetic Biology*, 2016.

Dubnau, David and Losick, Richard. Bistability in bacteria. *Molecular microbiology*, 61(3):564–572, 2006.

Duggin, Iain G and Wilce, Jackie A. Termination of replication in bacteria. *eLS*, 2005.

Ebersbach, Gitte and Gerdes, Kenn. Plasmid Segregation Mechanisms. *Annual Review of Genetics*, 39(1):453–479, 2005. ISSN 0066-4197. doi: 10.1146/annurev.genet.38.072902.091252.

Elowitz, Michael B and Leibler, Stanislas. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000. ISSN 0028-0836. doi: 10.1038/35002125.

Elowitz, Michael B; Levine, Arnold J; Siggia, Eric D, and Swain, Peter S. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

Errington, Jeffery; Daniel, Richard A, and Scheffers, Dirk-Jan. Cytokinesis in bacteria. *Microbiology and Molecular Biology Reviews*, 67(1):52–65, 2003.

Ferenci, Thomas. Regulation by nutrient limitation Ferenci. *Current Opinion in Microbiology*, 2(2):208–213, 1999. ISSN 1369-5274. doi: 10.1016/s1369-5274(99)80036-8.

Fishov, Itzhak; Zaritsky, Arieh, and Grover, NB. On microbial states of growth. *Molecular microbiology*, 15(5):789–794, 1995.

Freudenau, Inga; Lutter, Petra; Baier, Ruth; Schleef, Martin; Bednarz, Hanna; Lara, Alvaro R, and Niehaus, Karsten. ColE1-Plasmid Production in Escherichia coli: Mathematical Simulation and Experimental Validation. *Frontiers in bioengineering and biotechnology*, 3(September):127, 2015. ISSN 2296-4185 (Electronic). doi: 10.3389/fbioe.2015.00127.

Gardner, Timothy S; Cantor, Charles R, and Collins, James J. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339, 2000.

Gershenfeld, Neil A. *The nature of mathematical modeling*. Cambridge university press, 1999.

Ghosh, Biplab and Sain, Anirban. Origin of contractile force during cell division of bacteria. *Physical review letters*, 101(17):178101, 2008.

Gillespie, DT. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*, 58:35–55, 2007. doi: 10.1146/annurev.physchem.58.032806.104637.

Golding, Ido; Paulsson, Johan; Zawilski, Scott M, and Cox, Edward C. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, 2005.

Gonze, Didier. Modeling the effect of cell division on genetic oscillators. *Journal of Theoretical Biology*, 325:22–33, 2013. ISSN 00225193. doi: 10.1016/j.jtbi.2013.02.001.

Goodman, Myron F. Coping with replication train wrecks' in. *Sciences-New York*, 0004(April):189–195, 2000.

Gras, Anna and Ginovart, Marta. Indisim-som, an individual-based model to study short-term evolutions of carbon and nitrogen pools related to microbial activity in soil organic matter. *Modelling Methodologies and Simulation. Key Technologies in Academia and Industry. W. Borutzky, A. Orsoni, and R. Zobel (eds.). ECMS Digitaldruck Pirrot GmbH, Dudweiler, Berlin, Germany*, pages 554–559, 2006.

Grimm, Volker and Railsback, Steven F. *Individual-based modeling and ecology*, volume 2005. BioOne, 2005.

Hadeler, Karl-Peter and Müller, Johannes. Cellular automata: Analysis and applications. *Springer Monographs in Mathematics*, 4:467, 2017.

Haeusser, Daniel P. and Levin, Petra Anne. The great divide: coordinating cell cycle events during bacterial growth and division. *Current Opinion in Microbiology*, 11 (2):94–99, 2008. ISSN 13695274. doi: 10.1016/j.mib.2008.02.008.

Hansen, F. G.; Christensen, B. B., and Atlung, T. The initiator titration model: computer simulation of chromosome and minichromosome control. *Research in Microbiology*, 142(2-3):161–167, 1991. ISSN 17697123. doi: 10.1016/0923-2508(91)90025-6.

Harris, LK and Theriot, JA. Relative rates of surface and volume synthesis set bacterial cell size. *Cell*, 165(6):1479–1492, 2016. doi: 10.1016/j.cell.2016.05.045.

Heinze, Simon. Registry of standard biological parts: Partbbak801020, 2012. URL `http://parts.igem.org/wiki/index.php?title=Part%3ABBa_K801020`.

Hellweger, Ferdi and Bucci, Vanni. A bunch of tiny individualsindividual-based modeling for microbes. *Ecological Modelling*, 220(1):8–22, 2009.

Helmstetter, Charles E and Cooper, Stephen. Dna synthesis during the division cycle of rapidly growing escherichia coli br. *Journal of molecular biology*, 31(3): 507–518, 1968.

Helmstetter, Charles E and Cummings, Donald J. An improved method for the selection of bacterial cells at division. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 82(3):608–610, 1964.

Herbert, D. A theoretical analysis of continuous culture systems. *Continuous Culture of Microorganisms*, pages 21–53, 1961.

Hilfinger, A and Paulsson, J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A*, 108(29):12167–12172, 2011. doi: 10.1073/pnas.1018832108.

Hill, Norbert S.; Kadoya, Ryosuke; Chattoraj, Dhruba K., and Levin, Petra Anne. Cell size and the initiation of DNA replication in bacteria. *PLoS Genetics*, 8(3): 14–16, 2012. ISSN 15537390. doi: 10.1371/journal.pgen.1002549.

Ho, Po Yi and Amir, Ariel. Simultaneous regulation of cell size and chromosome replication in bacteria. *Frontiers in Microbiology*, 6(JUL):1–10, 2015. ISSN 1664302X. doi: 10.3389/fmicb.2015.00662.

Horii, Z I and Suzuki, K. Degradation of the DNA of Escherichia Coli K12 REC-(JC1569b) After Irradiation with Ultraviolet Light. *Photochemistry and Photobiology*, 8(2):93–105, 1968. ISSN 1751-1097. doi: 10.1111/j.1751-1097.1968.tb05850.x.

Horowitz, Joseph; Normand, Mark D; Corradini, Maria G, and Peleg, Micha. Probabilistic model of microbial cell growth, division, and mortality. *Applied and environmental microbiology*, 76(1):230–242, 2010.

Huh, Dann and Paulsson, Johan. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature genetics*, 43(2):95–100, 2011.

Jones, Kristala L; Kim, Seon-Won, and Keasling, JD. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metabolic engineering*, 2(4):328–338, 2000.

Kaguni, Jon M. Dnaa: controlling the initiation of bacterial dna replication and more. *Annu. Rev. Microbiol.*, 60:351–371, 2006.

Kahm, Matthias; Hasenbrink, Guido; Lichtenberg-Frat'e, Hella; Ludwig, Jost, and Kschischo, Maik. grofit: Fitting biological growth curves with R. *Journal of Statistical Software*, 33(7):1–21, 2010.

Katayama, Tsutomu; Ozaki, Shogo; Keyamura, Kenji, and Fujimitsu, Kazuyuki. Regulation of the replication cycle: conserved and diverse regulatory systems for dnaa and oric. *Nature Reviews Microbiology*, 8(3):163–170, 2010.

Keasling, J. D.; Palsson, B. O., and Cooper, S. Cell-cycle-specific F plasmid replication: Regulation by cell size control of initiation. *Journal of Bacteriology*, 173 (8):2673–2680, 1991. ISSN 00219193.

Keasling, J.D.; Kuo, H., and Vahanian, G. A Monte Carlo simulation of the Escherichia coli cell cycle. *Journal of Theoretical Biology*, 176(3):411–430, 1995. ISSN 00225193. doi: 10.1006/jtbi.1995.0209.

Kell, Douglas B.; Ryder, Hazel M.; Kaprelyants, Arseny S., and Westerhoff, Hans V. Quantifying heterogeneity: flow cytometry of bacterial cultures. *Antonie van Leeuwenhoek*, 60(3-4):145–158, 1991. ISSN 00036072. doi: 10.1007/BF00430362.

Kemp, Paul F; Cole, Jonathan J; Sherr, Barry F, and Sherr, Evelyn B. *Handbook of methods in aquatic microbial ecology*. CRC press, 1993a.

Kemp, PF; Lee, S, and LaRoche, Julie. Estimating the growth rate of slowly growing marine bacteria from rna content. *Applied and Environmental Microbiology*, 59 (8):2594–2601, 1993b.

Kepes, Francois; Joseleau-Petit, Danièle; Legros, Marcel, and Kepes, Adam. Early increases in the frequency of dna initiations and of phospholipid synthesis discontinuities after nutritional shift-up in escherichia coli. *The FEBS Journal*, 164(1): 205–211, 1987.

Kepes, Francois and others, . Postponement of cell division by nutritional shift-up in escherichia coli. *Microbiology*, 131(3):677–685, 1985.

Koch, Arthur L. Biomass growth rate during the prokaryote cell cycle. *Critical reviews in microbiology*, 19(1):17–42, 1993.

Kolter, R; Siegele, D. a., and Tormo, a. The stationary phase of the bacterial life cycle. *Annual review of microbiology*, 47:855–874, 1993. ISSN 00664227. doi: 10.1146/annurev.micro.47.1.855.

Kondo, Shigeru and Miura, Takashi. Reaction-diffusion model as a framework for understanding biological pattern formation. *science*, 329(5999):1616–1620, 2010.

Koppes, LJ and Von Meyenburg, K. Nonrandom minichromosome replication in escherichia coli k-12. *Journal of bacteriology*, 169(1):430–433, 1987.

Kreft, Jan-Ulrich; Booth, Ginger, and Wimpenny, Julian WT. Bacsim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*, 144(12): 3275–3287, 1998.

Kubitschek, H E. Growth during the bacterial cell cycle: analysis of cell size distribution. *Biophysical journal*, 9(6):792–809, 1969. ISSN 0006-3495. doi: 10.1016/S0006-3495(69)86418-0.

Kumar, Manoj; Husian, Mohammad; Upreti, Naveen, and Gupta, Deepti. Genetic algorithm: Review and application. *International Journal of Information Technology and Knowledge Management*, 2(2):451–454, 2010.

Kuo, Henry and Keasling, J. D. A Monte Carlo simulation of plasmid replication during the bacterial division cycle. *Biotechnology and Bioengineering*, 52(6):633–647, 1996. ISSN 00063592. doi: 10.1002/(SICI)1097-0290(19961220)52:6⟨633::AID-BIT1⟩3.0.CO;2-P.

Kuzminov, A. Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiology and molecular biology reviews : MMBR*, 63 (4):751–813, table of contents, 1999. ISSN 1092-2172.

Kuzminov, Andrei. Collapse and repair of replication forks in Escherichia coli. *Molecular Microbiology*, 16(3):373–384, 1995. ISSN 0950382X. doi: 10.1111/j.1365-2958.1995.tb02403.x.

Kuzminov, Andrei and Stahl, Franklin W. Stability of linear DNA in recA mutant Escherichia coli cells reflects ongoing chromosomal DNA degradation. *Journal of Bacteriology*, 179(3):880–888, 1997. ISSN 00219193.

Lazebnik, Iu; Papermaster, David, and Gall, Joe. [Can a biologist fix a radio, or what I learned while studying apoptosis]. *Advances in gerontology = Uspekhi gerontologii / Rossiskaia akademiia nauk, Gerontologicheskoe obshchestvo*, 12(September):166–171, 2003. ISSN 1561-9125.

Lemon, Katherine P and Grossman, Alan D. The extrusion-capture model for chromosome partitioning in bacteria. *Genes & Development*, 15(16):2031–2041, 2001.

Lidstrom, Mary E and Konopka, Michael C. The role of physiological heterogeneity in microbial population behavior. *Nature chemical biology*, 6(10):705–12, 2010. ISSN 1552-4469. doi: 10.1038/nchembio.436.

Likhoshvai, Vitaly A and Khlebodarova, Tamara M. Mathematical modeling of bacterial cell cycle: The problem of coordinating genome replication with cell growth. *Journal of bioinformatics and computational biology*, 12(03):1450009, 2014.

Lints, Taivo. Multiagent Modelling of a Bacterial Cell, a DnaA Titration Model Based Agent Model as an Example. *Proceedings of the Ninth Symposium on Programming Languages and Software Tools*, pages 82–96, 2005.

Little, John W; Edmiston, Susan H; Pacelli, Laura Z, and Mount, David W. Cleavage of the escherichia coli lexa protein by the reca protease. *Proceedings of the National Academy of Sciences*, 77(6):3225–3229, 1980.

Llorens, Juana María Navarro; Tormo, Antonio, and Martínez-García, Esteban. Stationary phase in gram-negative bacteria. *FEMS microbiology reviews*, 34(4): 476–495, 2010.

Lloyd-Price, J; Tran, H, and Ribeiro, AS. Dynamics of small genetic circuits subject to stochastic partitioning in cell division. *J Theor Biol*, 356:11–19, 2014. doi: 10.1016/j.jtbi.2014.04.018.

Løbner-Olesen, Anders and Skarstad, Kirsten. Titration of the escherichia coli dnaa protein to excess data sites causes destabilization of replication forks, delayed replication initiation and delayed cell division. *Molecular microbiology*, 50(1): 349–362, 2003.

Lobner-Olesen, Anders; Skovgaard, Ole, and Marinus, Martin G. Dam methylation: coordinating cellular processes. *Current opinion in microbiology*, 8(2):154–160, 2005.

Loomis, William F and Magasanik, Boris. Glucose-lactose diauxie in escherichia coli. *Journal of bacteriology*, 93(4):1397–1401, 1967.

Lutkenhaus, J. Assembly dynamics of the bacterial mincde system and spatial regulation of the z ring. *Annu Rev Biochem*, 76:539–562, 2007. doi: 10.1146/annurev.biochem.75.103004.142652.

Lutkenhaus, Joe and Addinall, SG. Bacterial cell division and the z ring. *Annual review of biochemistry*, 66(1):93–116, 1997.

MacDonald, James T; Barnes, Chris; Kitney, Richard I; Freemont, Paul S, and Stan, Guy-Bart V. Computational design approaches and tools for synthetic biology. *Integrative biology : quantitative biosciences from nano to macro*, 3(2):97–108, 2011. ISSN 1757-9708. doi: 10.1039/c0ib00077a.

Marguet, P; Tanouchi, Y; Spitz, E; Smith, C, and You, L. Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology. *PLoS One*, 5(7):e11909, 2010. doi: 10.1371/journal.pone.0011909.

Maruyama, Y and Lark, KG. Periodic synthesis of bacterial nucleic acids in the absence of protein synthesis. *Experimental cell research*, 25(1):161–169, 1961.

Mellodge, Patricia and Kachroo, Pushkin. *Model abstraction in dynamical systems: Application to mobile robot control*. Springer, 2008.

Messer, Walter. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews*, 26(4): 355–374, 2002. ISSN 01686445. doi: 10.1016/S0168-6445(02)00127-4.

Michel, Bénédicte. Replication fork arrest and DNA recombination. *Trends in Biochemical Sciences*, 25(4):173–178, 2000. ISSN 09680004. doi: 10.1016/S0968-0004(00)01560-7.

Michel, Bénédicte; Ehrlich, S. Dusko, and Uzest, Marilyne. DNA double-strand breaks caused by replication arrest. *EMBO Journal*, 16(2):430–438, 1997. ISSN 02614189. doi: 10.1093/emboj/16.2.430.

Michelsen, Ole; Teixeira de, M. Joost; Jensen, Peter Ruhdal, and Hansen, Flemming G. Precise determinations of C and D periods by flow cytometry in Escherichia coli K-12 and B/r. *Microbiology*, 149(4):1001–1010, 2003. ISSN 13500872. doi: 10.1099/mic.0.26058-0.

Miranda, Anabel and Kuzminov, Andrei. Chromosomal lesion suppression and removal in Escherichia coli via linear DNA degradation. *Genetics*, 163(4):1255–1271, 2003. ISSN 00166731.

Monod, Jacques. The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394, 1949.

Moser, F; Broers, NJ; Hartmans, S; Tamsir, A; Kerkman, R; Roubos, JA; Bovenberg, R, and Voigt, CA. Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synth Biol*, 1(11):555–564, 2012. doi: 10.1021/sb3000832.

Motulsky, Harvey and Christopoulos, Arthur. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting.* Oxford University Press, 2004.

Narang A, Pilyugin SS. Bacterial gene regulation in diauxic and non-diauxic growth. *J Theor Biol*, 244(2):326–348, 2007. doi: 10.1016/j.jtbi.2006.08.007.

Neidhardt, Frederick C. Bacterial Growth: Constant Obsession with dN/dt. *Society*, 181(24):7405–7408, 1999. ISSN 0021-9193.

Nielsen, Henrik J; Ottesen, Jesper R; Youngren, Brenda; Austin, Stuart J, and Hansen, Flemming G. The escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. *Molecular microbiology*, 62 (2):331–338, 2006.

Nielsen, Henrik J.; Youngren, Brenda; Hansen, Flemming G., and Austin, Stuart. Dynamics of Escherichia coli chromosome segregation during multifork replication. *Journal of Bacteriology*, 189(23):8660–8666, 2007. ISSN 00219193. doi: 10.1128/JB.01212-07.

Olsson, Jan. *Control of chromosome and plasmid replication in Escherichia coli.* PhD thesis, Acta Universitatis Upsaliensis, 2003.

Osella, Matteo and Lagomarsino, Marco Cosentino. Growth-rate-dependent dynamics of a bacterial genetic oscillator. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 87(1):1–12, 2013. ISSN 15393755. doi: 10.1103/PhysRevE.87.012726.

Tans SJOsella M, Cosentino Lagomarsino M. Step by step, cell by cell: Quantification of the bacterial cell cycle. *Trends Microbiol*, 2017. doi: 10.1016/j.tim.2016.12.005.

Paijmans, Joris; Lubensky, David K, and Wolde, Pieter Rein ten. Robustness of synthetic oscillators in growing and dividing cells. *arXiv preprint arXiv:1612.07992*, 2016.

Patra, Pintu and Klumpp, Stefan. Population Dynamics of Bacterial Persistence. *PLoS ONE*, 8(5), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0062814.

Portle, Stephanie. *Elucidating the connection between cell population heterogeneity and genetic regulatory architecture in specific artificial networks.* PhD thesis, Rice University, 2009.

Portle, Stephanie; Causey, Thomas B; Wolf, Kim; Bennett, George N; San, Ka-Yiu, and Mantzaris, Nikos. Cell population heterogeneity in expression of a gene-switching network with fluorescent markers of different half-lives. *Journal of biotechnology*, 128(2):362–375, 2007.

Portle, Stephanie; Iadevaia, Sergio; San, Ka-Yiu; Bennett, George N, and Mantzaris, Nikos. Environmentally-modulated changes in fluorescence distribution in cells with oscillatory genetic network dynamics. *Journal of biotechnology*, 140(3):203–217, 2009.

Potvin-Trottier, Laurent; Lord, Nathan D; Vinnicombe, Glenn, and Paulsson, Johan. Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, 538 (7626):514–517, 2016.

Powell, E O. An outline of the pattern of bacterial generation times. *Journal of general microbiology*, 18(2):382–417, 1958. ISSN 0022-1287. doi: 10.1099/00221287-18-2-382.

Pruitt, K. M. and Kamau, D. N. Mathematical models of bacteria growth, inhibition and death under combined stress conditions. *Journal of Industrial Microbiology*, 12:221–231, 2007.

Purnick, Priscilla EM and Weiss, Ron. The second wave of synthetic biology: from modules to systems. *Nature reviews Molecular cell biology*, 10(6):410–422, 2009.

Reshes, G; Vanounou, S; Fishov, I, and Feingold, M. Timing the start of division in E. coli: a single-cell study. *Physical biology*, 5(4):046001, 2008. ISSN 1478-3975. doi: 10.1088/1478-3975/5/4/046001.

Reynolds, Craig W. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH computer graphics*, 21(4):25–34, 1987.

Riedel, TE; Berelson, WM; Nealson, KH, and Finkel, SE. Oxygen consumption rates of bacteria under nutrient-limited conditions. *Appl Environ Microbiol*, 79 (16):4921–4931, 2013. doi: 10.1128/aem.00756-13.

Robert, L. Size sensors in bacteria, cell cycle control, and size control. *Front Microbiol*, 6:515, 2015. doi: 10.3389/fmicb.2015.00515.

Robert, Lydia; Hoffmann, Marc; Krell, Nathalie; Aymerich, Stéphane; Robert, Jérôme, and Doumic, Marie. Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism. *BMC biology*, 12(1):1, 2014.

Sauls, JT; Li, D, and Jun, S. Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Curr Opin Cell Biol*, 38:38–44, 2016. doi: 10.1016/j.ceb.2016.02.004.

Schaechter, M.; MaalOe, O., and Kjeldgaard, N. O. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. *Journal of General Microbiology*, 19(3):592–606, 1958. ISSN 0022-1287. doi: 10.1099/00221287-19-3-592.

Schumann, Wolfgang. *Dynamics of the Bacterial Chromosome: Structure and Function*. John Wiley & Sons, 2006.

Schwanhäusser, Björn; Busse, Dorothea; Li, Na; Dittmar, Gunnar; Schuchhardt, Johannes; Wolf, Jana; Chen, Wei, and Selbach, Matthias. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337, 2011.

Sezonov, G; Joseleau-Petit, D, and R., D'Ari. Escherichia coli physiology in luria-bertani broth. *J Bacteriol*, 189(23):8746–8749, 2007. doi: 10.1128/jb.01368-07.

Shetty, RP; Endy, D, and Knight, TF. Engineering biobrick vectors from biobrick parts. *J Biol Eng*, 2:5, 2008. doi: 10.1186/1754-1611-2-5.

Skarstad, K. and Boye, E. Degradation of individual chromosomes in recA mutants of Escherichia coli. *Journal of Bacteriology*, 175(17):5505–5509, 1993a. ISSN 00219193.

Skarstad, K. and Boye, E. Degradation of individual chromosomes in recA mutants of Escherichia coli. *Journal of Bacteriology*, 175(17):5505–5509, 1993b. ISSN 00219193.

Skarstad, K; Steen, H, and Boye, E. Escherichia coli DNA distribution measured by flow cytometry and compared with theoretical computer simulations. *J. Bacteriol.*, 163(2):661–668, 1985. ISSN 00219193.

Skarstad, Kirsten; Boye, Erik, and Steen, Harald B. Timing of initiation of chromosome replication in individual escherichia coli cells. *The EMBO journal*, 5(7): 1711, 1986.

Slager, J and Veening, JW. Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol*, 24(10):788–800, 2016. doi: 10.1016/j.tim.2016.06.003.

Smith, Douglas W; Garland, Adella M; Herman, Gary; Enns, Robert E; Baker, Tania A, and Zyskind, Judith W. Importance of state of methylation of oric gatc sites in initiation of dna replication in escherichia coli. *The EMBO journal*, 4(5): 1319, 1985.

Sobetzko, P; Travers, A, and Muskhelishvili, G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci U S A*, 109(2):E42–50, 2012. doi: 10.1073/pnas.1108229109.

Stewart, Eric J.; Madden, Richard; Paul, Gregory, and Taddei, Franc. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biology*, 3(2):0295–0300, 2005. ISSN 15449173. doi: 10.1371/journal.pbio.0030045.

Stokke, Caroline; Flatten, Ingvild, and Skarstad, Kirsten. An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS ONE*, 7(2), 2012. ISSN 19326203. doi: 10.1371/journal.pone.0030981.

Stranger, Barbara E; Forrest, Matthew S; Dunning, Mark; Ingle, Catherine E; Beazley, Claude; Thorne, Natalie; Redon, Richard; Bird, Christine P; De Grassi, Anna; Lee, Charles, and others, . Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 315(5813):848–853, 2007.

Streips, Uldis N; Yasbin, Ronald E, and others, . *Modern microbial genetics*, volume 344. Wiley Online Library, 2002.

Swain, Peter S; Elowitz, Michael B, and Siggia, Eric D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12795–800, 2002. ISSN 0027-8424. doi: 10.1073/pnas.162041399.

Tack, Ignace L M M; Logist, Filip; Noriega Fernandez, Estefan, and Van Impe, Jan F M. An individual-based modeling approach to simulate the effects of cellular nutrient competition on Escherichia coli K-12 MG1655 colony behavior and interactions in aerobic structured food systems. *Food Microbiology*, 45(PB):179–188, 2015. ISSN 10959998. doi: 10.1016/j.fm.2014.05.003.

Taheri-Araghi, Sattar; Bradde, Serena; Sauls, John T.; Hill, Norbert S.; Levin, Petra Anne; Paulsson, Johan; Vergassola, Massimo, and Jun, Suckjoon. Cell-size control and homeostasis in bacteria. *Current Biology*, 25(3):385–391, 2015. ISSN 09609822. doi: 10.1016/j.cub.2014.12.009.

Tamaki, Hisashi; Kita, Hajime, and Kobayashi, Shigenobu. Multi-objective optimization by genetic algorithms: A review. In *Evolutionary Computation, 1996., Proceedings of IEEE International Conference on*, pages 517–522. IEEE, 1996.

Tan, Cheemeng; Marguet, Philippe, and You, Lingchong. Emergent bistability by a growth-modulating positive feedback circuit. *Nature chemical biology*, 5(11): 842–848, 2009.

Taylor, John. *Introduction to error analysis, the study of uncertainties in physical measurements*. University Science Books, 1997.

Trojanowski, Damian; Hołówka, Joanna; Ginda, Katarzyna; Jakimowicz, Dagmara, and Zakrzewska-Czerwińska, Jolanta. Multifork chromosome replication in slow-growing bacteria. *Scientific Reports*, 7, 2017.

Tropini, C; Lee, TK; Hsin, J; Desmarais, SM; Ursell, T; Monds, RD, and Huang, KC. Principles of bacterial cell-size determination revealed by cell-wall synthesis perturbations. *Cell Rep*, 9(4):1520–1527, 2014. doi: 10.1016/j.celrep.2014.10.027.

Trueba, Frank J and Woldringh, Conrad L. Changes in cell diameter during the division cycle of escherichia coli. *Journal of bacteriology*, 142(3):869–878, 1980.

Tsoularis, A. and Wallace, J. Analysis of logistic growth models. *Mathematical Biosciences*, 179(1):21–55, 2002. ISSN 00255564. doi: 10.1016/S0025-5564(02)00096-2.

Tsukanov, R; Reshes, G; Carmon, G; Fischer-Friedrich, E; Gov, NS; Fishov, I, and Feingold, M. Timing of z-ring localization in escherichia coli. *Physical biology*, 8 (6):066003, 2011.

Vadia, S and Levin, PA. Growth rate and cell size: a re-examination of the growth law. *Curr Opin Microbiol*, 24:96–103, 2015. doi: 10.1016/j.mib.2015.01.011.

Volkmer, Benjamin and Heinemann, Matthias. Condition-Dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling. *PLoS ONE*, 6(7), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0023126.

von Freiesleben, Ulrik; Krekling, Martin A; Hansen, Flemming G, and Løbner-Olesen, Anders. The eclipse period of escherichia coli. *The EMBO journal*, 19 (22):6240–6248, 2000.

Voorn, WJ; Koppes, LJH, and Grover, NB. Mathematics of cell division in escherichia coli: comparison between sloppy-size and incremental-size kinetics. *Curr. Top. Mol. Gen*, 1:187–194, 1993.

Wallden, M.; Fange, D.; Gregorsson Lundius, E.; Baltekin, Ö., and Elf, J. The synchronization of replication and division cycles in individual E. coli cells (in press). *Cell*, pages 729–739, 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.06.052.

Wallden, Mats; Fange, David; Ullman, Gustaf; Marklund, Erik G., and Elf, Johan. Fluctuations in growth rates determine the generation time and size distributions of E. coli cells. *arXiv:1504.03145 [q-bio]*, 2015.

Wang, Ping; Robert, Lydia; Pelletier, James; Dang, Wei Lien; Taddei, Francois; Wright, Andrew, and Jun, Suckjoon. Robust growth of escherichia coli. *Current Biology*, 20(12):1099–1103, 2010. ISSN 09609822. doi: 10.1016/j.cub.2010.04.045.

Wang, Rong Fu and Kushner, Sidney R. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in escherichia coli. *Gene*, 100: 195–199, 1991.

Wang, Xindan; Possoz, Christophe, and Sherratt, David J. Dancing around the divisome: asymmetric chromosome segregation in escherichia coli. *Genes & development*, 19(19):2367–2377, 2005.

Weart, Richard B.; Lee, Amy H.; Chien, An Chun; Haeusser, Daniel P.; Hill, Norbert S., and Levin, Petra Anne. A Metabolic Sensor Governing Cell Size in Bacteria. *Cell*, 130(2):335–347, 2007. ISSN 00928674. doi: 10.1016/j.cell.2007.05.043.

Weiss, David S. Bacterial cell division and the septal ring. *Molecular microbiology*, 54(3):588–597, 2004.

Whitley, Darrell. A genetic algorithm tutorial. *Statistics and computing*, 4(2):65–85, 1994.

Wilkinson, Darren J. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature reviews. Genetics*, 10(2):122–33, 2009. ISSN 1471-0064. doi: 10.1038/nrg2509.

Wold, S; Skarstad, K; Steen, H B; Stokke, T, and Boye, E. The initiation mass for DNA replication in Escherichia coli K-12 is dependent on growth rate. *The EMBO journal*, 13(9):2097–2102, 1994. ISSN 0261-4189.

Wu, Ling Juan and Errington, Jeff. Nucleoid occlusion and bacterial cell division. *Nature reviews. Microbiology*, 10(1):8, 2012.

Youngren, Brenda; Nielsen, Henrik Jörk; Jun, Suckjoon, and Austin, Stuart. The multifork escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes & development*, 28(1):71–84, 2014.

Zaritsky, A. Rate of dna replication in thymine-requiring strains of escherichia coli. *Journal of general microbiology*, 63(3):vi–vii, 1970.

Zaritsky, A and Pritchard, RH. Changes in cell size and shape associated with changes in the replication time of the chromosome of escherichia coli. *Journal of bacteriology*, 114(2):824–837, 1973.

Zaritsky, Arieh. Cell-Shape Homeostasis in Escherichia coli Is Driven by Growth, Division, and Nucleoid Complexity. *Biophysical Journal*, 109(2):178–181, 2015. ISSN 1542-0086. doi: 10.1016/j.bpj.2015.06.026.

Zaritsky, Arieh; Wang, Ping, and Vischer, Norbert OE. Instructive simulation of the bacterial cell division cycle. *Microbiology*, 157(7):1876–1885, 2011.

Zaritsky, Arieh; Woldringh, Conrad L; Vischer, Norbert OE; WANG, Ping, and Helmstetter, Charles E. Simulation of the prokaryotic cell cycle at http://simon. bio. uva. nl/cellcycle. *J. Sytem. Cybernet. Informat.*, 10:18–23, 2012.

Zhang, Huaiqiang; Liu, Yuqing; Liu, Bo, and Gao, Peiji. A novel approach for estimating growth phases and parameters of bacterial population in batch culture. *Science in China, Series C: Life Sciences*, 49(2):130–140, 2006. ISSN 10069305. doi: 10.1007/s11427-006-0130-6.

Zheng, Hai; Ho, Po-Yi; Jiang, Meiling; Tang, Bin; Liu, Weirong; Li, Dengjin; Yu, Xuefeng; Kleckner, Nancy E; Amir, Ariel, and Liu, Chenli. Interrogating the escherichia coli cell cycle by cell dimension perturbations. *Proceedings of the National Academy of Sciences*, page 201617932, 2016.

Zwietering, M; Jongenburger, I; Rombouts, F, and van 't Riet, K. Modeling of the bacterial growth curve. *Appl Environ Microb*, 56(6):1875–1881, 1990.

Zyskind, J W; Svitil, A L; Stine, W B; Biery, M C, and Smith, D W. RecA Protein of Escherichia-Coli and Chromosome Partitioning. *Mol Microbiol*, 6(17):2525–2537, 1992.