THE UNIVERSITY *of* EDINBURGH

THE UNIVERSITY
of EDINBURGH

Doctoral Thesis

Cell size regulation of *Escherichia coli* at high osmolarity

---

Dario Miroli

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

March 15, 2020

Abstract

The goal of the study of bacterial physiology, and in the study of biophysics more broadly, is to apply quantitative predictive models to the behaviour of complex living organisms. Just as is the case with thermodynamics, the challenge comes from the fact that the subject systems are far too intricate to be modelled from the bottom up using first principles. Instead, the focus is on the development of coarse-grained models which reduce the description of the system down to a small number of key state variables. As late as the mid 20th century the identity, and even existence, of these variables was an open question, because living cells are a non-equilibrium system. However, between the 1950s and 1970s a golden era of bacterial physiology would give rise to a series of robust quantitative models, that linked a number of important physiological parameters. Starting with the Copenhagen school in 1958, it would soon be discovered that cell size was connected to growth rate via a single exponential relation that was independent of the makeup of the media used to achieve it. Shortly after this key result, the work of Helmstetter and Cooper would reveal that the time taken to replicate the chromosome of *E. coli* was constant over a wide range of growth rates. This counter intuitive observation would lead to the realisation that in order to grow at doubling times shorter than their replication time, bacterial cells must overlap multiple rounds of replication simultaneously. Taking these findings and running with them, Donachie at the University of Edinburgh was able to develop a model in which cells accumulate a critical initiation mass in order to start new rounds of replication. Using previous data, he was able to show that this initiation mass remained constant across a wide range of growth conditions, and in doing so, was able to relate neatly the average cell size, growth rate, chromosome replication time and time to divide. Since its development Donachie's constant initiation mass model has been tested under numerous different environmental conditions and found to hold. Although examples exist where the initiation mass can be shown to vary, the underlying relationship continues to stand up to scrutiny. Despite this

extensive testing, one condition underrepresented in literature is that of growth at high osmolarity. Much like nutrient limitation, high osmolarity conditions have been shown to reduce the growth rate although the mechanism by which this happens remains poorly understood.

To this end the objective of this doctoral thesis is to examine the balanced growth of *E. coli* cells in high osmolarity conditions, and determine how the cell physiology under such conditions fits within the paradigm of constant initiation mass, as described by Donachie. To achieve this, I characterise the bulk steady state growth rates of cells and show that in agreement with previous work they decrease significantly with increasing salt concentration. I go on to measure cell volume both using a custom built microscope I constructed, and a high throughput Coulter counter. By comparing the results from these independent measurement I am able to show that, in steady state, there is very little volume change that accompanies the significant decrease in growth rate at high osmolarity. Following the analysis of cell size, I proceed to measure how the DNA content of cells changes at high osmolarity, using flow cytometry. I argue that, in contrast to some recent findings, there is little change in cellular DNA content as the salt concentration is increased. Combining these results with measurements of the chromosome replication time from real time PCR reactions, I also demonstrate an increase in the cell cycle time with increasing osmolarity, in agreement with recent findings. Combining these results, I am able to show what effect high osmolarity has on the initiation mass. Additionally, I examine the results from a small number of experiments where the osmolarity was increased using soribtol instead of sodium chloride. I discuss if the nature of the osmolyte used changes the effect a high osmolarity condition has on cellular physiology, and suggest some interesting experiments going forward to further understand this relationship.

Lay summary

Bacteria are some of the simplest organisms on earth, yet they are surprisingly hardy. They can survive extremes in temperature, pressure, and acidity that human beings often can not. As a result, they can live in almost any environment on earth including on, and inside, both animals and humans. For bacteria like *E. coli*, which live in the intestines of humans, one of the main challenges they have to face is changes to the so called "osmolarity" of their environment. Osmolarity is just a measure of the amount of all the different solutes which are dissolved in a liquid. These solutes can be anything, such as sugars, proteins or salts. If a liquid has more of these solutes, then we say it is of a higher osmolarity. If, for example, one takes a drink of a very sugary soda, when it reaches our intestines it will cause an increase in the osmolarity for any *E. coli* living there. This can cause a big problem for *E. coli* because each cell is permeable to water. As a result, osmosis will cause water to flow out of the *E. coli* cell until the concentration of solutes inside and outside the cell are equal. If too much water flows out of the cell, *E. coli* can shrivel up and eventually die. To stop this from happening, *E. coli* has evolved to have a number of defences, such as special pumps which can transport solutes from the environment into the cell, reducing the osmolarity difference between the cell and its environment, in that hope that water will again start to flow back into the cell. A lot of work has been done by other researches to understand how *E. coli* copes in the short term with these changes in osmolarity, but a lot less is known about how *E. coli* changes, when it has to live and grow at very high osmolarities for many generations. In my project, I grew *E. coli* cells in the lab on a liquid medium containing all the nutrients which *E. coli* needs. I then added different amounts of salt to change the osmolarity of the medium, making it larger than the osmolarity which *E. coli* would normally find itself in. I measured how fast cells were able to grow and divide in these conditions, and I found that they grow more slowly as the osmolarity gets larger. I also built myself a custom microscope and used it to take pictures of cells in these conditions

to try and work out if the size of the cells changed. I found out that *E. coli* get smaller as the osmolarity increases. I hope that by finding out how *E. coli* changes when it has to survive for long periods of time at high osmolarity that it will help us understand how to better kill it when it gets into places which are at a high osmolarities. A good example of this is the urinary tract where serious infections can occur if *E. coli* settles there.

Declaration

Except otherwise stated, the research undertaken in this thesis was the work of the author. Where work was done in collaboration with others, a significant contribution was made by the author. Collaborative work is indicated as such at the beginning of the relevant section.

Dario Miroli
March 15, 2020

Acknowledgements

I would like to first and foremost thank my supervisor Teuta Pilizota. Not only for making this entire endeavour possible, but for being a constant source of advice, inspiration, and guidance throughout. I could not have made a better choice of supervisor for these last four years. I would also like to thank my thesis committee: Meriem El Karoui and Filippo Menolascina for their advice at various times during the project.

To the TP lab I owe a great debt. To Alex, thank you for being the adult in the room on far too many occasions, and for your help throughout, but especially early on during the alignment of the scope. To Jerko and Katya, thank you for the trading of rants, tips and expertise regarding all things optics. The existence of my microscope owes much to your combined efforts. To Smitha I am thankful for her devotion to the tenant that "a problem shared is a problem halved", she has generated a great deal of positive karma over the last few years. To Leonardo I am thankful for the many useful conversations regarding the philosophy of life (and death), as well as more practical matters. To Guillaume, I am very grateful for his theoretical expertise, and the numerous times he allowed me to borrow them. To Keiran I am thankful for his help regarding all things IT and OD. Finally to Xavier as an honouree member of the TP lab, I am thankful for his constant advice, optimism, company, entertainment, and readings.

To Benura Azeroglu and Elliott Chapman for teaching me everything I know regarding QPCR, and for their saint like patience during that process. Thank you.

To Ross Mckenzie I am deeply grateful for his help regarding machine learning, and other similar matters of a spiritual nature.

To the members of Teviot, especially Miles, Dale and Nick, who have been the solution to, and cause of, an equal numbers of problems throughout. I am grateful to you all.

To Rupesh I am eternally thankful for his unwavering faith, support and friendship.

To my wife, I am so very grateful. Without her constant support, patience and compassion throughout, but especially during these last few trying weeks, I certainly would never have made it this far. In this regard, I count myself extraordinarily lucky.

I owe much to my parents Karen and Mark, and my grandmother Brenda. I am thankful for their constant support, but especially at that start for encouraging me to pursue this project.

Finally I would like to dedicate this thesis to my late grandfather Jack. I doubt he would of understood much of what is to follow, but he was none the less an instrumental part of its creation.

"Success is stumbling from failure to failure with no loss of enthusiasm."

Winston Churchill

Contents

1 Introduction **3**

1.1 The need for biophysical models . . . 3
1.2 Balanced growth and steady state . . . 4
1.3 The nutrient growth law . . . 6
1.4 The bacterial cell cycle . . . 7
1.5 Constant initiation mass . . . 11
1.6 Average DNA content . . . 15
1.7 The general growth law . . . 17
1.8 Osmolarity . . . 20
1.9 Motivation for thesis and overview . . . 22

2 Methodology **25**

2.1 A custom microscope for microfluidics . . . 25
2.1.1 Bright-field imaging . . . 26
2.1.2 Fluorescent alignment . . . 27
2.1.3 Perfect focus alignment . . . 29
2.1.4 Software for microscope control . . . 33
2.2 Bacterial strains and media . . . 34
2.2.1 Strains . . . 34
2.2.2 Media and osmolytes . . . 35
2.2.3 Cell culturing . . . 37
2.3 Growth rate measurements . . . 38
2.4 Cell size from fluorescent microscopy . . . 39
2.4.1 Experimental protocol . . . 39
2.4.2 Image analysis and volume extraction . . . 42
2.4.2.1 The need for a custom algorithm . . . 42

2.4.2.2 Selecting cells for analysis . . . 43
2.4.2.3 Convoloutional neural networks as image segmenters 43
2.4.2.4 Network architectures . . . 51
2.4.2.5 Data normalisation, augmentation and training sets . 55
2.4.2.6 Network training and comparison . . . 57
2.4.2.7 Cell morphology extraction from segmented images . 60
2.5 Cell size measurement using the Coulter counter . . . 65
2.5.1 Principle of operation . . . 65
2.5.2 Experimental protocol . . . 70
2.6 DNA content from flow cytometry . . . 71
2.6.1 Principle of operation . . . 71
2.6.2 Calibration . . . 74
2.6.3 Experimental protocol . . . 78
2.7 C period determination from QPCR . . . 79
2.7.1 Principal of operation . . . 79
2.7.2 Experimental protocol . . . 86

3 Results **89**
3.1 Growth rate at high osmolarities . . . 89
3.2 Cell size at high osmolarity . . . 94
3.2.1 Cell size from microscopy . . . 94
3.2.1.1 Recreating the nutrient growth law . . . 94
3.2.1.2 Effect of high osmolarity . . . 98
3.2.2 Cell size from Coulter counter . . . 105
3.2.2.1 Comparison to standard beads and microscopy data 105
3.2.2.2 The nature of the bump . . . 108
3.2.2.3 Cell size along the growth curve . . . 111
3.2.2.4 Cell size at high salt concentrations . . . 112
3.2.2.5 Cell size at high sorbitol concentrations . . . 119
3.2.2.6 Cell size at high osmolarity . . . 121
3.2.2.7 LB and high osmolarity . . . 123
3.2.3 Discussion and future work . . . 127
3.3 DNA content from flow cytometry . . . 133
3.3.1 Mannose grown cells as a callibration standard . . . 133
3.3.2 DNA content at high osmolarity . . . 138

3.3.3 Discussion and future work . . . 146
3.4 Cell cycle parameters at high osmolarity . . . 148
3.4.1 C period determination from QPCR . . . 148
3.4.1.1 A comparison of methods for computing QPCR primer efficiencies . . . 148
3.4.1.2 C period at high osmolarities . . . 152
3.4.2 D period at high osmolarity . . . 158
3.4.3 Cell cycle time at high osmolarity . . . 159
3.4.4 Number of origins at high osmolarity . . . 159
3.4.5 The unit cell size at high osmolarity . . . 161
3.4.6 Discussion and future work . . . 163

4 Conclusions and further work **167**

A Code **173**

Bibliography **174**

Chapter 1

Introduction

1.1 The need for biophysical models

Bacterial cells are highly complex systems. Take for instance, the model bacterium *Escherichia coli*. An average *E. coli* cell can be less than $1\mu m^3$ in volume, yet it can contain over 10^6 proteins per cell (Milo 2013). The intricate interaction of these proteins, and the control of their expression from the over 4,000 protein coding genes of the *E. coli* genome, is the primary driving force for life (Serres et al. 2001). Understanding these interactions, and how they give rise to a system which can not only self replicate, but also sense and adapt to its environment, has been a key goal of Biology for over a century (Lane-Claypon 1909). One strategy for this task is to attempt to understand the function and interaction of every gene and protein in the entire *E. coli* cell. To this end, much progress has been made. For example, of the thousands of genes on the *E. coli* genome only around one third have not yet had their function experimentally determined (Ghatak et al. 2018). Likewise, recent advances in computational power have led to the development of "whole cell" models, which aim to simulate all of the key components of a cell, and their interaction. These models have had some real success, simulating simpler bacteria such as *Mycoplasma genitalium*, which has only around 500 genes in its genome (Karr et al. 2012). Despite this it appears that, at least for now, *E. coli* may be sufficiently complex to still be out of reach of whole cell approaches, though progress is being made (O'Brien, Monk, and Palsson 2015).

An alternative to this "bottom up" approach is the "top down" view of coarse

grained biophysical models. Instead of aiming to understand every intricate detail of cellular function, these approaches instead attempt to express the complex biological system in terms of a small number of key "state variables". The term state variable is not accidental, it draws an intentional comparison with thermodynamics. In fact, a useful analogy is to liken the study of *E. coli* to that of the study of ideal gases. In the case of an ideal gas, a bottom up approach would involve the computation of the trajectories of each individual atom in the gas. While in principle possible, this is a highly impractical task. Instead, a better solution is to realise that the state of the gas can be expressed in terms of only three variables, namely: volume, temperature and pressure. By describing the system in terms of these variables we are now able to predict how perturbations to the system will affect it. This is precisely the goal of biophysical models, to gain predictive power of a complex system with as simple a model as possible. The difficulty with this approach, in the case of biological systems, is correctly identifying the state variables. It is not at all obvious what the biological equivalent of pressure and temperature should be, and the field has not yet settled on a single answer to this question. This difficulty is compounded by the fact that, unlike ideal gases, life as a system is almost defined by its departure from thermodynamic equilibrium. That is not to say that no progress has been made however. In the following sections we will outline how the first coarse grained biological model for bacteria was developed, and how this led to an improvement of our understanding of bacterial physiology in a number of diverse ways. As we will see, the key advantage of this approach is that it allows us to improve our understanding and ability to predict the system, even when many of the more intricate mechanisms remain unknown. To see why this approach is so vital consider the minimal genome project, which aims to develop a synthetic organism with the smallest possible genome. In 2015 the result of this work was the creation of an organism which had some 473 genes, 149 of which had completely unknown biological function (Hutchison et al. 2016)

1.2 Balanced growth and steady state

One of the first challenges for developing simple, coarse grained models of bacterial cells, is the complexity of the bacterial growth curve. It was known at the beginning of the 1900s that bacteria grown in liquid culture could be inoculated into fresh media, and that the resulting growth would show multiple characteristic phases

(Lane-Claypon 1909). These phases can be summarised as a lag phase: where cells exhibit little growth, exponential phase: where cell growth is constant, and stationary phase: where cell growth arrests and/or decreases. The existence of these phases posed a complication for the development of coarse grained models, because it was soon noticed that bacteria change their physiology significantly as they transition between phases (McKee 1929; Henrici 1928). To again draw analogy with the ideal gas law, it would be difficult to model the effect of a perturbation to a sample of ideal gas, if the sample itself was already undergoing transitions between various temperatures and pressures. In essence, what was required was the concept of steady state for bacterial growth, and a means of achieving it.

The first step towards defining the concept of steady state, came in the late 1940s when a mathematical treatment of the lag, stationary and exponential phases was succinctly compiled (Monod 1949). In particular, the exponential phase was characterised by a single parameter which described the growth rate. This parameter was derived from the simple starting assumption that cells divide once every division time τ. In other words, given an initial number of cells N_0, the number of cells N, at time t later can be computed by the expression:

$$N = N_0 2^{\frac{t}{\tau}}. \tag{1.1}$$

We can rewrite this equation in terms of the growth rate λ, by taking the natural logarithm of both sides as follows:

$$ln(N) = ln(N_0) + \frac{ln(2)}{\tau}t, \tag{1.2}$$

$$ln(N) = ln(N_0) + \lambda t, \tag{1.3}$$

where $\lambda = \frac{ln(2)}{\tau}$. Equation 1.3 is exactly of the form $y = c + mx$, and so if one plots the logarithm of cell number against time, the gradient of the resulting line will yield the growth rate. By being able to express the complexity of exponential growth in terms of a single well defined parameter, the stage was set for the development of the concept of balanced growth. In 1957 balanced growth was defined as a state: "*where every extensive property of the growing system increases by the same factor*" (Campbell 1957) [1]. Today, as in 1957, balanced growth can be achieved in two

[1] By extensive we mean properties of the entire bulk system such as biomass, quantity of DNA or of a particular protein. Specifically, we distinguish extensive properties from intensive ones which do not double as the population doubles, *e.g* average cell size, or DNA content per cell.

ways. Firstly, using continuous culture devices such as chemostats, which control the addition of fresh media and removal of spent media from a fixed volume reactor, so as to maintain the same growth rate indefinitely (Novick and Szilard 1950). Secondly, by a traditional batch growth, which through a high level of dilution of the seed culture, maintains exponential growth for many generations. With a definition of balanced growth now established, and the means to experimentally achieve it, all the ingredients for the development of a coarse grained biophysical model were in place. As we will see in the following sections, the reason that this development was so crucial is that it allowed a baseline to be established, against which the effects of perturbations could be measured, without any of the complications that arise from shifts between complex growth phases. However, it is important to note that while functionally treated as steady state in many cases, balanced growth is a necessary but not sufficient condition for steady state. In addition to requiring the constant growth of all extensive properties with time, steady state requires the further condition that all intensive variables remain constant with time. For example, a population of cells in balanced growth, but where individual cells do not divide, will still double all of their extensive properties with time, however the cell size distribution will clearly not be invariant with time and therefore the system will not be in steady state.

1.3 The nutrient growth law

Not long after the development of balanced growth, in 1958, a series of two papers were published which outlined experiments that would later become known as some of the most "*fundamental experiments in bacterial physiology*" (Cooper 1993). Both were produced by what has since become known as the "Copenhagen school", and detailed work on the growth of the bacterium *Salmonella typhimurium*.

The first paper, which will be our primary focus, examined how the composition of cells changed, as the nutritional properties of the media they were grown in was varied (Moselio Schaechter, Maaløe, and Kjeldgaard 1958). To do this, cells were grown in large batch cultures with frequent dilution to ensure balanced growth. The DNA content per cell, cell number, cell mass and RNA content were then measured for many different media. Strikingly, it was found that the cell composition was not dependent on the particular chemical makeup of the media used, but only on the

growth rate it produced. For example, if one medium had a good carbon source but a poor nitrogen source, and a second medium had a poor carbon source but a good nitrogen source, and both had a doubling time of 60 minutes, then the DNA, RNA and mass per cell would not differ between the two media. This is an astonishing result. It suggests that despite the complex biochemical pathways that differ when cells are grown in different media, the variable of growth rate alone is sufficient to determine the macroscopic composition and properties of the cell. Furthermore the Copenhagen school found that the relationship between growth rate and these macroscopic properties could be easily explained by a single, simple, exponential function (as shown in figure 1.1). Mathematically we can relate how the average cell volume (or mass) S of the population varies with the growth rate λ, in the following way:

$$S \propto e^{\gamma\lambda}. \tag{1.4}$$

Where γ is the gradient of the line plotted in figure 1.1:A. The fact that a parameter such as cell size, which is both the result of, and maintained by, a vast regulatory system can be related to the growth rate (itself the end result of a vast biochemical pathway) by such a simple equation, shows the heights which coarse grained biological models can achieve. Additionally, note the predictive power implicit in this finding. Given the introduction of a new medium, determination of the growth rate is sufficient to allow an accurate prediction of the population averaged cell size, as well as RNA and DNA content. Future models which we will discuss in the following sections, aim to precisely replicate this success.

1.4 The bacterial cell cycle

Implicit in the work of the Copenhagen school was the idea that the cell cycle of bacteria was very much like that of mammalian and other eukaryotic cells. While there are some bacteria such as *Caulobacter crescentus* which share similarities to the eukaryotic cell cycle, in general the the cell cycle of bacteria and eukaryotes are very different (Laub, Shapiro, and McAdams 2007). In 1968 there was a series of two papers by Helmstetter and Cooper which revolutionised the way the cell cycle was thought of in *E. coli*. The first showcased a series of experiments utilising the so called "baby machine" which was a method of partially synchronising the cell cycle of *E. coli* (C. E. Helmstetter 1967). An excellent retrospective review of the baby

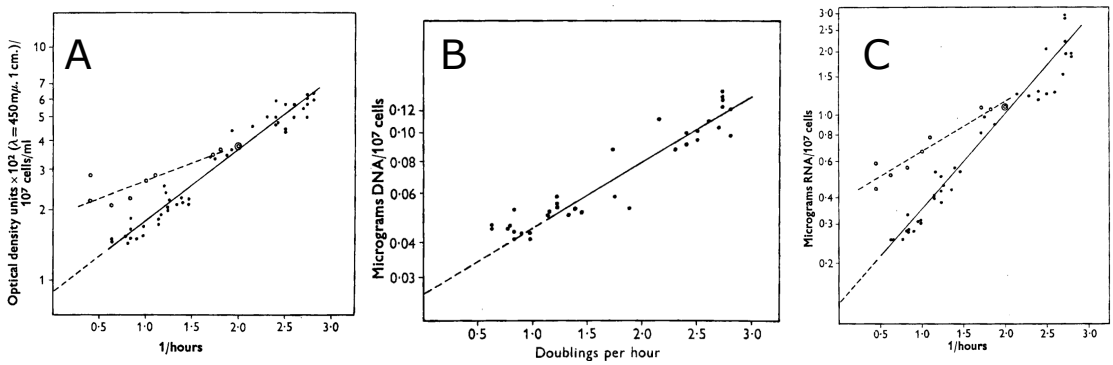

Figure 1.1: The nutrient growth law as measured by the Copenhagen school in 1958. **A:** Growth rate on the x axis is plotted against the logarithm of cell mass as measured by optical density measurement. **B:** Growth rate against logarithm of DNA content per 10^7 cells. **C:** Growth rate against logarithm of RNA content per 10^7 cells. In figures A and C the solid line represents data from batch culture growth while the dashed line represents restricted growth in continuous culture. Figures taken from Moselio Schaechter, Maaløe, and Kjeldgaard 1958.

machine and its impact can be found in C. E. Helmstetter 2015, but a brief outline will be given here for completeness. In essence a baby machine is a specialised set up of filters, where initially a culture of exponentially growing cells in balanced growth are filtered through the membrane and adhered to its surface. The filter is then turned upside down and fresh media is allowed to filter through the membrane again. The mother cells which are tightly bound to the membranes surface can not be eluted out, while younger daughter cells can be. Therefore, cells in the elution have a very narrow age distribution and can, as a result, be considered synchronised. By allowing the exponential cells to be exposed to radioactive thymine for a period of time (pulse labelled), and then measuring the amount of radiation in the daughter cells (and importantly how it varies with age) the pair were able to measure the DNA replication rates in *E. coli* (C. E. Helmstetter and Cooper 1968). From this data they developed a so called two timer model to explain the mechanics of DNA replication during the cell cycle (Cooper and C. E. Helmstetter 1968). Today it is better known as the Helmstetter Cooper model.

One of the key findings of Helmstetter and Cooper was that the time it takes for *E. coli* to replicate its chromosome is essentially constant (at around 40 minutes) for doubling times between 20 and 60 minutes. The fact that the replication time

can be twice as long as the doubling time is a finding that should cause perpetual amazement. To explain this remarkable finding the pair suggested the cells carry out overlapping rounds of replication, where new replication forks begin at the origin before the previous ones have completed replication (figure 1.2). In this way cells are able to continue replication throughout the division process, and through generations. Indeed, at high growth rates it is possible that replication begins in the grandmother generation to produce chromosomes for the current generation. The constant period of time between initiation of replication and its completion is denoted as the C period.

The pair also suggested that after the completion of each round of replication there was a constant D period (measured at around 20 minutes) after which division would be completed. In principle this D period corresponded to the time taken to segregate the chromosomes and fully fission the cell, and was not dependent on the growth rate over a wide range. Together the C and D periods constituted the cell cycle time τ_{cyc}, which was itself invariant over a wide range of growth conditions. Conceptually the way to approach the model is that each new cell born at a time t was preceded by an initiation of replication event at a time $t - \tau_{cyc}$.

The Helmstetter Cooper model allowed for the prediction of a number of useful quantities. The number of replication forks active at any particular time throughout the cell cycle could be determined just from a measurement of the C, D and τ periods. This is turn allowed the computation of the cellular DNA content throughout the cycle and, given knowledge of the age distribution, the average DNA content of exponentially growing cells[2]. Importantly this allowed a theoretical understanding of the finding by the Copenhagen school, that the DNA content of cells increases with increasing growth rate. Using their model, this could be understood as a natural consequence of the increased number of overlapping cell cycles at high growth rates, leading to higher overall average DNA content per cell.

Again, what we have seen here is the development of a coarse grained biological model which sheds light on the complex way cells coordinate the doubling of their

[2]We will not derive the expression for the average DNA content here because we will do so later using the method by Bremer & Churchwood.

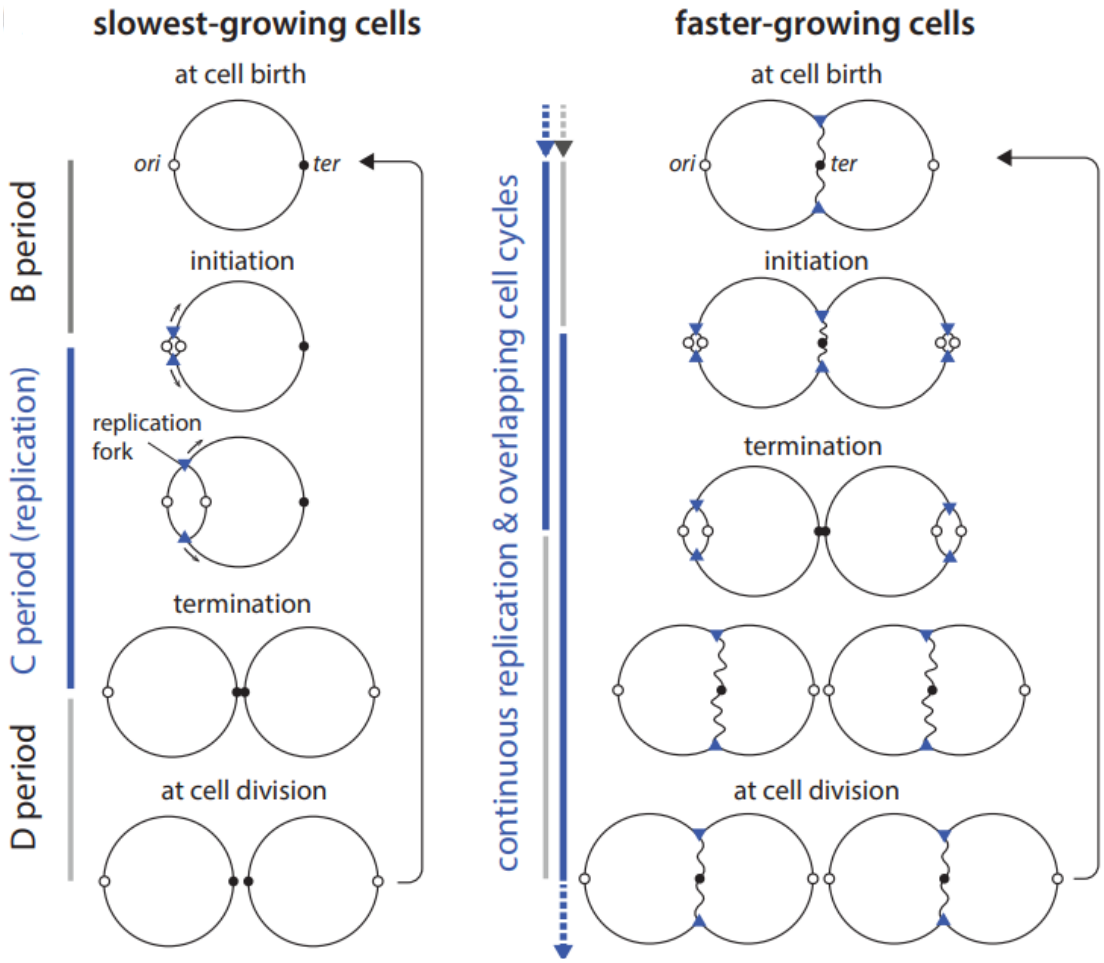

Figure 1.2: Replication of the *E. coli* chromosome. **Left:** During slow growth (doubling times greater than 60 minutes) replication rounds do not overlap. Instead after cell birth there is an initial period where no replication occurs (B period). When initiation occurs a pair of replication forks are formed at the origin and replication proceeds in both directions towards the terminus. Once the forks reach the terminus replication is completed and the two newly produced chromosomes can be separated. The time between initiation and the end of replication is the *C* period. The time between the end of replication and division is the *D* period. **Right:** During fast growth rounds of replication overlap. As shown by the blue arrows, cells inherit chromosomes which are in the process of being replicated. By the time this newly inherited replication completes, new rounds of replication have already been initiated at the newly formed origins. This means that DNA synthesis never stops throughout the cell cycle. Figure taken from S. Jun et al. 2018.

biomass with the doubling of their genome, and subsequent division. The model is able to make predictions (for example of average DNA content per cell) which are verified by experimental data. Importantly, as with the nutrient growth law, it does so using a small number of important parameters that interact in a simple manner, and most of which are constant over a wide range of growth rates. The Helmstetter Cooper model remains another foundational example of successful models applied to biology.

1.5 Constant initiation mass

The same year that Helmstetter and Cooper revealed the inner workings of the bacterial cell cycle, Donachie was able to combine their findings with the work of the Copenhagen school, to unveil a powerful observation about the interaction between cell mass and coordination of replication. In short, Donachie found that cells initiate replication at a constant cell mass per origin (Donachie 1968). While his finding can be expressed succinctly, to understand how it came about is an involved process which will bring together much of the work we have seen in the previous three sections, and more besides. To say that initiation mass is constant does not do justice to the impact of the observation, or to the sustained effort and insight required to make it. In this section we will hope to gain a full appreciation of this result.

To start, we will examine a result obtained by the consideration of the age distribution in exponentially growing cells in balanced growth. The frequency of cell ages in such a population has a curious property that newborn cells are twice as abundant as cells just about to divide. Intuitively, this make sense as each time a cell divides it produces two newborn cells. To see a full derivation of the age distribution one can consult Powell 1956, here we will simply quote the result. If a is the age of a bacteria then the probability of finding a cell in an exponentially growing culture $p(a)$ is given by the expression:

$$p(a) = \frac{2ln(2)}{\tau} 2^{\frac{-a}{\tau}} \quad ; \quad 0 \leq a \leq \tau, \tag{1.5}$$

where τ is the doubling time. One can note that as expected $p(0)$ yields $\frac{2ln(2)}{\tau}$ which is precisely twice the value of $p(\tau) = \frac{ln(2)}{\tau}$. It should also be noted that the general form of the equation is exponentially decreasing, this reflects the exponential increase in

the number of cells with each doubling time that passes. From this age distribution, if one additionally assumes that cells grow exponentially between their birth length l_b and their division length l_d (which of course must satisfy the relation that $2l_b = l_d$) then the probability of finding a cell in the population at a length l, can be computed as (Cullum and Vicente 1978; A. Koch and Schaechter 1962):

$$p_l(l) = \begin{cases} \frac{l_d}{l^2} & l_b \leq l \leq l_d \\ 0 & l < l_b \ ; \ l > l_d. \end{cases} \tag{1.6}$$

Where $p_l(l)$ is the probability of finding a cell at length l. Using this equation we can now compute the expectation value of $p_l(l)$:

$$\bar{l} = \int_{l_b}^{l_d} l p_l(l) dl = 2 l_b ln(2) = l_d ln(2). \tag{1.7}$$

Therefore, we see that given the theoretical age distribution, where cells divide exactly after τ minutes of growth, the average cell size in the population is proportional to both the size at birth and the size at division. This will be a key observation in showing that the initiation mass is constant.

The next ingredient we require in order to derive the constant initiation mass, is a measure of the time at which initiation occurs during each cell cycle. As Cooper and Helmstetter showed, a consequence of their model is that at fast growth rates cells must overlap replication cycles in order to maintain division times lower than the $C + D$ period. As a result, the time at which replication is initiated during the cell cycle is dependent on the relative lengths of τ and τ_{cyc}. To compute the time at which initiation occurs between cell birth and division one can use the formula:

$$t_i = \tau \left(1 - Frac\left[\frac{\tau_{cyc}}{\tau}\right]\right). \tag{1.8}$$

Where t_i is the initiation time and $Frac[]$ represents the fractional part of the argument. While correct, this equation can seem obscure on first reading. Indeed in their papers Helmstetter and Cooper do not write this expression explicitly but instead show that the initiation time can be extracted by diagrammatic means. The most detailed explanation on the construction of these diagrams is found in C. Helmstetter et al. 1968, although Helmstetter himself laments later that this source contains "*in retrospect, some impossibly complex figures*" (C. E. Helmstetter 2015). Here we will present some hopefully less complex diagrams based on the advice that Helmstetter

gives in his 2015 review.

First, consider that during balanced growth in order for the DNA content to remain invariant across generations, it is required that cells initiate replication once and only once per doubling time. If initiation occurs more than once cells will accumulate DNA over time, and likewise if it occurs less than once, cells will lose DNA over time. It follows therefore, that the time interval between initiation events must be precisely τ minutes. According to the Helmstetter Cooper model after each initiation event replication will start and last precisely C minutes. Immediately following the end of replication the D period will occur at the end of which cell division is complete. In figure 1.3 there are plotted two diagrams showing how the initiation time can be derived from these assumptions. The green arrows in both diagrams indicate the time between initiation events. The orange arrows represent the replication time, and the red arrows the division time. If one drops vertical lines (shown in blue) from the end of two sequential D periods (indicating cell division) then one finds, as expected, that the time between divisions is precisely the doubling time τ. If we continue to tile this pattern of initiation, replication and division we find that initiation occurs once between the drawn vertical lines. We can then read off the position of this initiation event to get the time in the cell cycle at which initiation occurs. In figure 1.3:A this process is shown for a doubling time of 35 minutes, a C period of 40 minutes and a D period of 20 minutes. Likewise in figure 1.3:B we show the process for a doubling time of 25 minutes, a C period of 40 minutes and a D period of 20 minutes. As can be seen, this yields an initiation time of 10 and 15 minutes respectively. If we take the example of a doubling time of 25 minutes, we can confirm that the obtained value for the initiation time matches the calculated value using equation 1.8 as follows:

$$t_i = \tau\left(1 - Frac\left[\frac{C+D}{\tau}\right]\right) = 25\left(1 - Frac\left[\frac{60}{25}\right]\right) = 25\left(1 - \frac{10}{25}\right) = 15. \qquad (1.9)$$

We can now see that the the Helmstetter Cooper model contains sufficient information to compute the initiation time in exponentially growing populations. With that we can now turn finally to examine Donachie's observation.

To show that the initiation mass was constant Donachie primarily used a graphical approach which combined the data of the Copenhagen school with that of Helmstetter and Cooper. First consider single cells growing exponentially between their

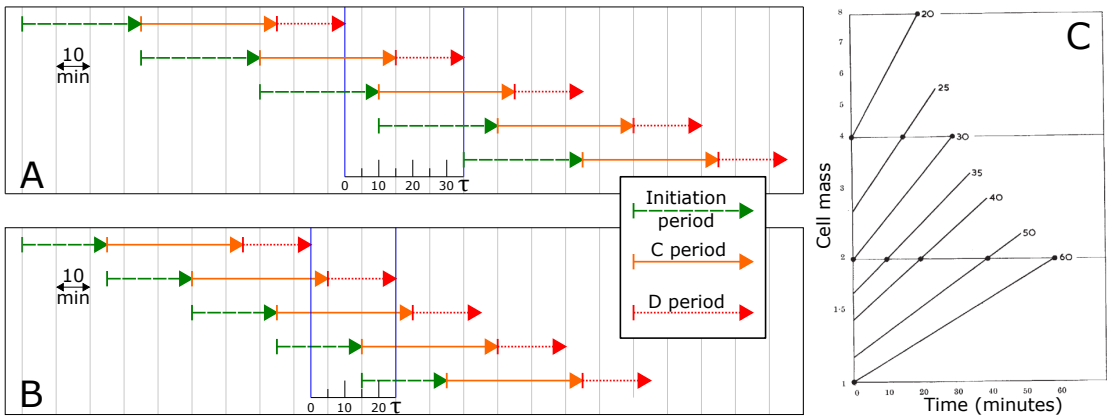

Figure 1.3: Timing of initation of replication. **A & B:** Diagrams based on those presented in C. E. Helmstetter 2015. The coloured arrows represent the three phases of the cell cycle, initiation (green), replication (orange), and division (red). In both diagrams initiation is continuous, *i.e* immediately after an initiation event has finished a new cycle begins. A doubling time τ lies between divisions, as shown by the vertical blue lines. The initiation time can be determined by reading off the position of the start of the green arrow in between the blue lines. Each vertical grey line represents 10 minutes of time. In both cases the C period is 40 minutes and the D period is 20 minutes. In subfigure A the doubling time is 35 minutes and in subfigure B it is 25 minutes. **C:** Plot of the mass of of individual cells against time for different doubling times between 20 and 60 minutes. Solid circles represent the initiation time t_i. Figure taken from Donachie 1968.

birth mass (m_b) and division mass ($m_d = 2m_b$). If one plots on a logarithmic scale cell mass against time from birth, then the result is a straight line that doubles in height over τ minutes. One can do this for a range of doubling times between 20 and 60 minutes as shown in figure 1.3:C. The contribution of Moselio Schaechter, Maaløe, and Kjeldgaard 1958, was to show how the average mass of the population varied with the doubling time. As can be seen in figure 1.3:C. Donachie uses this result to determine the birth size of each cell. We now see why it was necessary to demonstrate that the population average size is proportional to the birth size. The data from Moselio Schaechter, Maaløe, and Kjeldgaard 1958 was for the population average, and here we are plotting mass of individual cells, therefore to be able to make this connection we require that the two be proportional. The next step is to mark on each trace the initiation time as calculated using the Helmstetter Cooper model. This is indicated in figure 1.3:C by the position of the black dots. As can be

seen from the plot it appears that, in all the conditions shown, initiation occurs at two discrete mass values. We are now dangerously close to showing that initiation mass is constant. To do that we will need one quantity from Cooper and Helstetter, and that is the number of origins present at the moment of initiation. Again this quantity can be extracted from further analysis of diagrams of the kind shown in figure 1.3:A, or mathematically from the ratio of the τ_{cyc} and τ periods. Here we will just briefly state the result, which is that there are two origins present at the time of initiation for cells with doubling times between 30 and 60 minutes, and four origins for cells with doubling times between 20 and 30 minutes. If we use this result to divide the mass at initiation by the number of origins at initiation we find that mass required to initiate replication at a single origin is constant across the range of doubling times tested.

This powerful observation completes a chain of understanding that started with the Copenhagen school. Observations that cell mass, and DNA content, were exponentially dependent on the growth rate led to the development of the Helmstetter Cooper model of DNA replication. This showed that at high growth rates cells had to overlap replication rounds, thus increasing the DNA content of the population in a manner dependent on the growth rate itself. Finally, the Helmstetter Cooper model allowed Donachie to show that the mass required to initiate replication at each origin was constant. Therefore, we see that the increase in cell size at high growth rates is a result of the need for overlapping rounds of replication which increases the number of origins per cell and thus the required mass per cell.

1.6 Average DNA content

In both the work of Donachie, and that of Helmstetter and Cooper, appeal was made to the age distribution of exponentially growing cultures. In both cases it was normally the ideal age distribution that was used, rather than a more complex one involving stochasticity in the timings of division. It turns out, as Bremer & Churchward showed, that this assumption about the age distribution of the population is often not required (Bremer and Churchward 1977). Here we will derive a key result regarding an expression for the average DNA content of cells, using the approach of Bremer & Churchwood.

First, as before consider an exponentially growing culture in balanced growth such that the number of cells N at time t can be computed as:

$$N = N_0 2^{\frac{t}{\tau}}, \tag{1.10}$$

where again, τ is the doubling time. One of the key logical steps from Bremer & Churchwood, is that given the assumption that growth is balanced, it is not only the cell number which must grow exponentially with time, but also the number of origins and termini. If we let I be the number of origins and T be the number of termini, then we can write the following expressions:

$$I = I_0 2^{\frac{t}{\tau}}, \tag{1.11}$$

$$T = T_0 2^{\frac{t}{\tau}}. \tag{1.12}$$

Where I_0 and T_0 are the initial number of origins and termini in the population at $t = 0$. Next, consider that as consequence of the Cooper Helmstetter model, each cell born at time t must have been preceded by an initiation event which occurred at a time $t - (C + D)$. Where C and D have their by now familiar meaning of the replication and division periods respectively. Similarly, one can say that each cell is the result of the creation of a new termini at a time $t - D$ in the past. We can therefore express the initial number of origins and termini as:

$$I_0 = N_0 2^{\frac{C+D}{\tau}}, \tag{1.13}$$

$$T_0 = N_0 2^{\frac{D}{\tau}}. \tag{1.14}$$

Using the above equations we can compute the mean number of origins (or terminators) in the population as:

$$\bar{I} = \frac{I}{N} = \frac{I_0 2^{\frac{t}{\tau}}}{N_0 2^{\frac{t}{\tau}}} = \frac{N_0 2^{\frac{C+D}{\tau}}}{N_0} = 2^{\frac{C+D}{\tau}}, \tag{1.15}$$

$$\bar{T} = \frac{T}{N} = 2^{\frac{D}{\tau}}. \tag{1.16}$$

From these equations, we can compute the average number of replication forks $\bar{F}$ in the population. To do this, consider that each origin initially creates two replication

forks which are then destroyed when a terminus is formed. Therefore, the average number of forks is equal to twice the difference between the number of origins and termini:

$$\bar{F} = 2(\bar{I} - \bar{T}) = 2\left(2^{\frac{C+D}{\tau}} - 2^{\frac{D}{\tau}}\right). \tag{1.17}$$

Given that each fork replicates half the chromosome in C minutes, we can write that each fork produces DNA at a rate of $\frac{1}{2C}$. Therefore, the average rate of DNA generation is as follows:

$$\frac{d\bar{G}}{dt} = \bar{F}\frac{1}{2C} = \frac{1}{C}\left(2^{\frac{C+D}{\tau}} - 2^{\frac{D}{\tau}}\right). \tag{1.18}$$

Finally in exponential growth, the rate of DNA generation will also be exponential. If we let the rate of growth be λ then the DNA content, or number of genomes per cell, can be computed as follows:

$$\frac{Genomes}{cell} = \bar{G} = \frac{1}{C\lambda}\left(2^{\frac{C+D}{\tau}} - 2^{\frac{D}{\tau}}\right), \tag{1.19}$$

where λ is equivalent to $\frac{ln(2)}{\tau}$. This is the same result derived by Cooper and Helmstetter but without appealing to either the time of initiation or the age distribution.

1.7 The general growth law

As we have seen in the previous sections, the nutrient growth law as described by Moselio Schaechter, Maaløe, and Kjeldgaard 1958, was a powerful tool for the development of our understanding of bacterial physiology. Naturally, the next question to ask is how robust, or general, is this law. The most obvious way to test this is to perturb the growth rate in a way not related to nutrient limitation, and determine if the relationship between growth rate and cell size still holds. This is essentially the approach taken by Si et al. 2017, where they perturbed the physiology of exponentially growing cells in a number of ways, and then plotted the average cell size against growth rate (figure 1.4). From this plot we can see that the nutrient growth law is not in fact a general explanation of the relationship between cell size and growth rate. Take for example, the case of growth rate reduction via chloramphenicol treatment (yellow points and grey lines in figure 1.4). Increasing concentrations of chloramphenicol reduce growth rate but do not lead to a corresponding exponential decrease in cell size, as would be expected using the nutrient growth law.

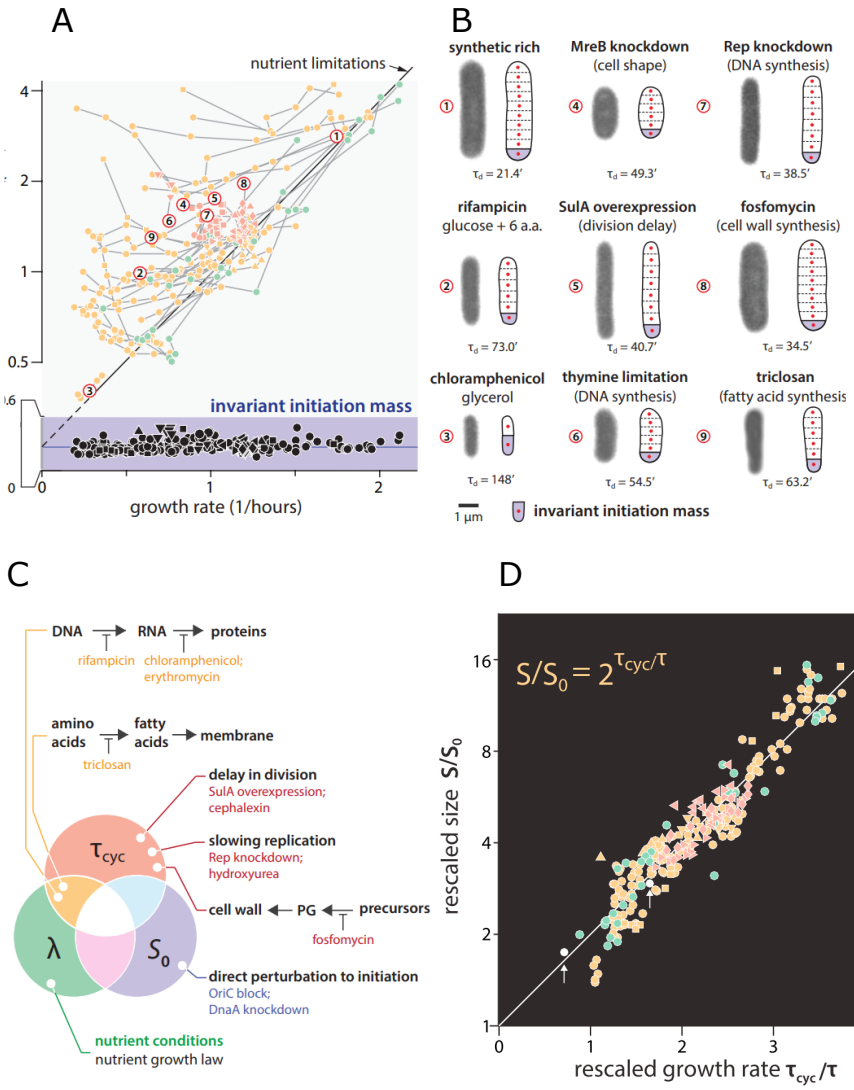

Figure 1.4: The general growth law. **A:** Growth rate against cell size for a number of different conditions including nutrient limitation (green points) and various antibiotic treatments and genetic alterations (red and yellow points). The exact conditions are indicated in subfigure C. The unit cell size (S_0) is shown for each condition along the bottom of the figure. **B:** Image showing how cells change their size in various conditions. The red numbers correspond to those in subfigure A. Each cell is also shown as a composition of unit cells of size S_0 **C:** Venn diagram showing the three parameters of the general growth law. Each section is labelled with the ways in which it was perturbed in Si et al. 2017. **D:** Same data as subfigure A rescaled by the general growth law. Figure taken from Si et al. 2017.

To explain this deviation, Si et al develop the idea of the general growth law, which they derive in a tautological manner which we will outline here. First consider that from equation 1.15, the average number of origins can be expressed in terms of τ_{cyc} and τ. We can therefore state that:

$$S = S\frac{\bar{I}}{\bar{I}}, \tag{1.20}$$

where S is the average cell size in the population and $\bar{I}$ is the average number of origins. We can then define S_0 as $\frac{S}{\bar{I}}$ and rewrite the above equation as:

$$S = S_0 2^{\frac{t_{cyc}}{\tau}}. \tag{1.21}$$

This is the general growth law as stated by Si et al. 2017. The quantity S_0 is the average cell size divided by the average number of origins. It can be shown that S_0 is also proportional to the initiation mass of cells (Donachie 1968; Bremer and Churchward 1977). S_0 can be thought of conceptually as the minimum cell volume needed to hold all of the resources required for initiation, and to continue biosynthesis until cell division. As a result, Si et al also refer to S_0 as the unit cell size. As shown in figure 1.4:A, the measured value of S_0 stays relatively constant across a wide variety of conditions. In figure 1.4:D Si et al show that if the data is rescaled, by dividing the average cell size by S_0, and plotting this against the logarithm of $\frac{\tau_{cyc}}{\tau}$, then the data from figure 1.4:A collapses onto a single straight line. Thus, unlike the nutrient growth law, the general growth is robust to a wide variety of perturbations. In fact, the nutrient growth law can be thought of as a special case of the general growth law, where τ_{cyc} and S_0 are held fixed. It is interesting to note that the three parameters of the general growth law λ, τ_{cyc} and S_0 have their origin in the three important results we have discussed previously. The Copenhagen school were responsible for showing that λ is a key state variable, which influences cell size. Cooper and Helmstteter were responsible for developing the concept of τ_{cyc} as a combination of C and D periods. Finally, Donachie noticed that the initiation mass (S_0) was constant over a range of nutrient limited conditions. It turns out from the work of Si et al that Donachie's result is in fact more general than just nutrient limitation, and that S_0 is constant over a wide range of perturbations. As a consequence, the general growth law maintains the predictive capability of the nutrient growth law, but expands it to have a much wider applicability.

1.8 Osmolarity

Like all living things *E. coli* is made distinct from its environment by the presence of a cell envelope. In the case of *E. coli*, this envelope is a complex structure consisting of inner and outer phospholipid bilayers, with a peptidoglycan cell wall running in between (Silhavy, Kahne, and Walker 2010). This envelope is permeable to water, but not to larger molecules such as DNA, ribosomes or proteins (Delamarche et al. 1999). As a result, an osmotic pressure (π) is generated across the cell membrane in response to differences in concentration of solutes between the cell interior and its environment. This pressure can be expressed as (Dick 1966):

$$\pi = \phi RT(C_{ext} - C_{int}), \tag{1.22}$$

where R is the gas constant, T is the temperature, and C_{ext} and C_{int} are the concentration of solutes in the environment and inside the cell respectively. The variable ϕ is the osmotic coefficient, which is experimentally determined for each different species of solute. For ideal solutes the value is unity, but for larger molecules such as proteins or sucrose, it can vary significantly. As the interior of *E. coli* tends to have a higher concentration of solutes than the environment, the osmotic pressure is usually positive, that is to say it acts to expand the cell envelope. The physiological role of osmotic pressure is not clear. It has been suggested that osmotic pressure is necessary for new material to be incorporated into the cell wall, and thus for growth (Arthur L Koch, Higgins, and Doyle 1982; Arthur L Koch 1990), although recent experimental work has indicated that this may not be the case (Rojas, Theriot, and K. C. Huang 2014a). Regardless of its function, it is known that *E. coli* maintains osmotic pressure homeostasis (Wood 2011). Indeed, a great deal of work has been put into elucidating the physiological response of *E. coli* to both hyperosmotic and hypoosmotic shocks. For example during hyperosmotic shocks *E. coli* initially shrinks due the movement of water out of the cell (Pilizota and Shaevitz 2012). Volume is then recovered by a complex, multi-step process involving the active import of potassium ions and synthesis of glutamate (McLaggan et al. 1994). Over time, imported potassium is replaced with either synthesised or important osmoprotectants such as glycine betaine, proline or trehalsoe, which increase the osmolarity of the cytoplasm but do not interfere with enzymatic function (Larsen et al. 1987). Similarly, the response to hypoosmotic shock is also well studied. During hypoosmotic shock water rushes into the cell which can lead to volume expansion and eventually cell

lysis (Kung 2005; Reuter et al. 2014). To combat this *E. coli* cells contain a number of mechanosensitive channels which can gate to allow solutes to exit the cell thus equilibrating osmotic pressure (Buda et al. 2016; Haswell, Phillips, and Rees 2011). The sophisticated response that *E. coli* has to both hyper and hypoosmotic shocks is perhaps not surprising, given that such changes in osmolarity are common in many of the environments in nature where *E. coli* are found, such as the gastrointestinal or urinary tracts of mammals (Reppas et al. 2015; Sands and Layton 2009).

Despite the extensive progress made in understanding the transient responses of *E. coli* to hyperosmotic shock, relatively little progress has been made in understanding the steady state response to such conditions. We know for example that at high osmolarity the growth rate substantially decreases (Pilizota and Shaevitz 2014; D. S. Cayley, Guttman, and Record Jr 2000). However the cause for this reduction remains largely unknown. It has been suggested that at high osmolarity, the reduction in cytoplasmic free water, and the resulting increase in crowding, may lead to reduced global diffusion and therefore lower growth rates (S. Cayley and Record Jr 2004). However, so far direct measurements of diffusion of cytoplasmic GFP molecules have indicated a complex story (Konopka et al. 2009). While diffusion decreased significantly in cells just after a hyperosmoric shock, it was found to be much less reduced in cells that had reached exponential growth at high osmolarity, even though the biopolymer volume fraction was about the same in both cases. This indicates that the level of crowding may not uniquely determine the level of diffusion and therefore growth rate. Recent work has also shown that the translational elongation rate also decreases at high osmolarity although again the exact cause for this remains unclear, with the authors suggesting again reduced diffusion and crowding as a likely cause (Dai, Zhu, et al. 2018).

Until recently there had been very little work looking at high osmolarity conditions from the point of view of the growth laws described in the previous section. In 2018 Dai and Zhu, examined a number of important physiological parameters of *E. coli* cells during exponential growth at high osmolarity (Dai and Zhu 2018). In particular they demonstrated that average cell volume is reduced at high osmolarity, while the C and D periods increase substantially, along with cellular DNA content. By plugging their measured values into the general growth law (equation 1.21) they

show that the unit cell size appears to decrease at high osmolarity. However when they consider initiation mass (as opposed to volume) they find that it is invariant at high osmolarity. Thus suggesting that it not initiation volume but initiation mass which is kept constant under the general growth law. One possible complication in the results of Dai and Zhu et al., is that the primary media used in their experiments was Lysogeny broth (LB) media. Previous work has shown that it is difficult to obtain balanced growth in LB media due to the presence of diauxic shifts (Sezonov, Joseleau-Petit, and d'Ari 2007). It will be a primary objective of this work to examine cell physiology at high osmolarity in a very similar manner as Dai and Zhu et al., but in defined media where balanced growth is more likely to be achieved.

1.9 Motivation for thesis and overview

In this thesis the primary goal will be to examine the effect of high osmolarity conditions on the physiology of *E. coli* cells undergoing balanced growth. Much like Dai and Zhu et al., I will approach this task from the point of view of the general growth law, as described by Si et al. 2017. I will therefore, make independent measurements of each of the parameters of the general growth law, and use them to make an estimate of the unit cell size at high osmolarity. In doing so I will be able to add to the findings of Dai and Zhue et al, by showing how results compare for slower growing cells, in defined media where balanced growth is more easily obtained. The hope is that by understanding growth at high osmolarity from the perspective of the general growth law, one might be able to obtain new information about the possible mechanisms which lead to slower growth at high osmolarity.

In detail, I use spectrophotometer measurements to determine how the exponential growth rate λ changes as the media osmolarity is increased by adding sodium chloride. Next, I make a measurement of how the population averaged cell size (S) varies with increasing sodium chloride concentrations. I do this through direct imaging of cells using a custom fluorescent microscope, which I construct myself. I compare the results from microscopy with those obtained using a high throughput electronic Coulter counter. I then use the Coulter counter to examine how cell size changes across the growth curve in both normal and high osmolarity conditions. To

allow extraction of the cell cycle parameters, I measure how cellular DNA content ($\bar{G}$) changes at high salt concentrations, using high throughput flow cytometry. To obtain estimates for the C period I use quantitative polymerase chain reaction (QPCR) to extract the relative abundance of origins versus termini in exponentially growing cells. Combining this with the DNA content measurement I am able to compute the D period at high sodium chloride concentrations. These results are combined to show how the unit cell size (S_0) changes at high sodium chloride concentrations. Finally, I perform a limited number of experiments where osmolarity is increased using sorbitol instead of sodium chloride. By measuring cell size and growth rate in these conditions, and comparing the results with similar sodium chloride experiments, I discuss whether osmolarity uniquely determines the physiological state of cells, or if instead the nature of the osmolyte used plays an important role.

This thesis will be split into two main chapters. In Chapter 2 I will outline the techniques used to measure each parameter of cell physiology (λ,S,$\bar{G}$, C and D). This includes an overview of the working principle in each case as well as a detailed explanation of the experimental protocol. In Chapter 3 I apply these techniques to the measurements of exponentially growing cells at high osmolarity, and discuss the results of these measurements. Finally, I propose further experiments to further elucidate the physiological effect of high osmolarity.

Chapter 2

Methodology

2.1 A custom microscope for microfluidics

One of the main goals of this project was to study the size of cells using microscopy based techniques. To facilitate this, part of the work I carried out was to construct a custom microscope that could be used not only for this task, but for a number of other possible experiments as well. Indeed, one of the main benefits of constructing a custom microscope is that the resulting set up tends to be far more flexible in the kinds of experiments that can be carried out using it, compared to traditional commercial systems. In this case, the microscope was constructed with attention paid to the design requirements of using microfluidic devices. Microfluidics are ubiquitous in modern bacterial studies, and can have a wide range of applications, from observation of long term growth of single cells, to rapid high throughput cell screening (P. Wang et al. 2010; Okumus et al. 2016). Even though microfluidics were not used in this work, it was important to consider their use for other experimenters wishing to utilise the microscope. To this end the scope was built around Nikon CFI Plan Apochromat Lambda 100x objective and a Nikon TI-C-LWD condenser Lens. The large working distance of the condenser allowed for adequate space above the sample to accommodate most microfluidic devices. Similarly in addition to the mechanical stage for coarse x,y positioning of the sample we utilised a PI U-751 XY stage which gave 25mm of motion in each dimension. This stage was chosen to accommodate the large traversal distances often required for microfluidic experiments. In this section I will outline the various optical components of the microscope that were utilised to achieve sample illumination, imaging and stabilisation.

2.1.1 Bright-field imaging

The most basic imaging condition available to almost all light microscopes is that of bright-field. In principle, almost any illumination light source could be used but we opted for a white light LED (Luxeon Star, USA). The LED was powered using a constant current driver that allowed brightness to be controlled through the sending of 0-10 volt analogue signal (A011-D-V-700 FlexBlockTM, Luxdrive, USA). To keep the LED within operating temperatures during extended use, it was attached to an aluminium heat sink, before being placed in the optical train. In between the LED and the sample are: two lenses, two irises, and a condenser. These components act together to provide even illumination. As shown in the schematic of the microscope in figure 2.1, the first lens closest to the LED has a focal length of $20mm$. This lens is used to image the LED onto the second iris, which is known as the condenser iris. The condenser iris would normally sit at the back focal plane of the condenser, but in our setup there was no convenient place to mount the iris at that location. Instead the f=$200mm$ lens is used to relay the image of the LED and the condenser iris onto the back focal plane of the condenser. Finally, the field iris sits at a location such that it is imaged onto the sample by the combined action of the condenser and f=$200mm$ lens. By opening and closing the condenser iris the intensity of illumination can be controlled. The field iris on the other hand serves as a reference for the height of the condenser, as an image of it can be seen in the sample plain only when the condenser is in the correct vertical position. The combined effect of this setup is Köhler illumination which ensures a uniform image illumination intensity across the sample plain.

After the sample has been illuminated it is imaged by the infinity corrected objective. The resulting collimated light then passes through a telescope formed of a f=$175mm$ lens and one of three possible lenses (f=60,100 or $150mm$) depending on the desired magnification. Flipper lens mounts allow the desired lens to be easily and quickly selected. Downstream of the telescope, a periscope and flipper mirror allow the light to be sent to either a standard CCD camera (Wat-902H ultimate, Watex France), which was used mostly for alignment of the optics, or an EMCCD camera (iXon Ultra 888, Andor UK) which was the primary camera used for image acquisition. In the case of the CCD camera a f=100 mm lens was used to image the collimated light from the telescope onto the sensor, while for the EMCCD a

f=200mm lens was used. The reason for this difference is that the iXon Ultra 888 camera has a much larger chip size of 1,024x1,024 pixels compared to the CCD and so a larger focal length lens is required to ensure that the entire image receives an undistorted part of the field of view.

2.1.2 Fluorescent alignment

For fluorescent imaging we utilised an OptoLED system from Cairn Research UK. This system provides a well controlled power source to a high intensity white light LED. Together the system outputs light at both excellent intensity and stability, which is necessary for accurate quantification of fluorescent signal. To control the wavelength of the illuminating light, a filter wheel unit (Optospin IV Cairn Research UK) was used to switch between a number of optical band pass filters (Chroma). To achieve uniform illumination as shown in figure 2.1 an f=50mm lens is used to collimate the light produced from the LED. This collimated light is then imaged on to the back focal plane of the objective using an f=50mm lens, and making sure to overfill the objective aperture. In this way an even fluorescent illumination can be achieved. To ensure that the excitation light is reflected and sent to the object while simultaneously allowing the emitted light to be collected by the down stream optics, a dichroic mirror is used which reflects many common excitation wavelengths and transmits most common emission wavelengths (59022bs-UV, Chroma)

In addition to the OptoLED LED system we also coupled in an alternative fluorescent illumination light source. This was the Polychrome IV from Texas instruments, USA. The polychrome unit uses a mercury vapour lamp and a diffraction grating to output light at any desired wavelength between 320 and 680nm. In addition to being finely tunable, the output light also has a narrow bandwidth of around 8nm. This allows for the excitation of almost any fluorophore without the need to purchase expensive excitation filters for each one. Illumination of the sample using the polychrome is very similar to the method used with the optoLED. An f=50mm lens is positioned as before to collimate the light coming out of the optical fibre which couples to the polychrome unit. A flipper mirror is then used to divert the light from the fibre through a second f=50mm lens again making sure to overfill the objective. An image of the actual set up can be seen in figure 2.2C. The addition of the polychrome into the microscopy set up allows for far more fluorophores to be

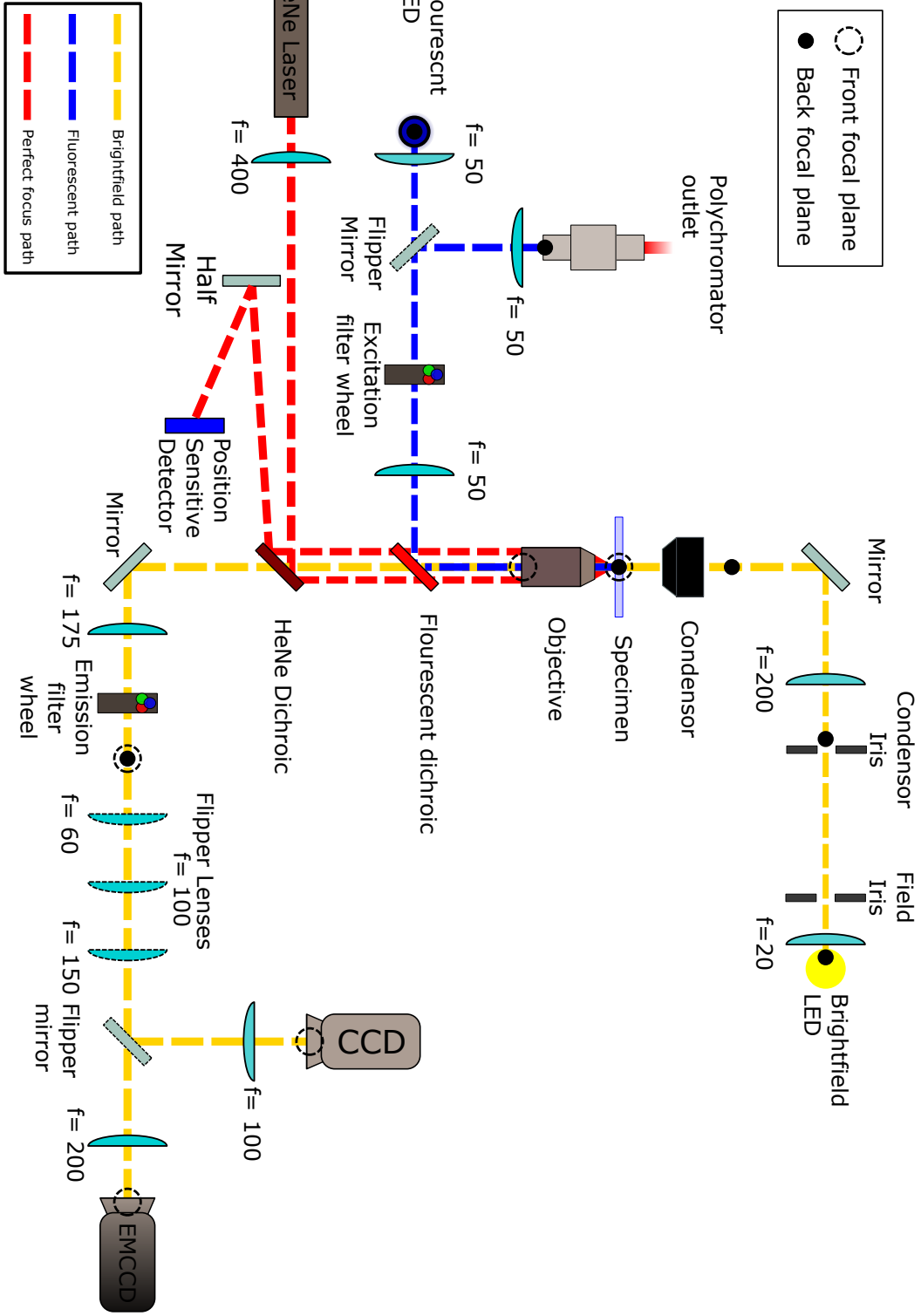

Figure 2.1: Schematic of the custom microscope built for and used in this work. The different optical trains used for bright field, fluorescence and perfect focus are indicated by the coloured dashed lines. Lenses are labelled by their focal lengths.

used. However, as in this investigation I was exclusively interested in the imaging of GFP, I opted to use the LED as the primary fluorescent light source due to its quicker start up and shut down times.

2.1.3 Perfect focus alignment

In order to facilitate long term microscopy experiments where sample imaging may occur over hours, or days, it is necessary to have some form of image stabilisation. Over time mechanical drift in the microscope stage, and movement of the immersion oil between the objective and sample, lead to the sample moving out of the focal plane of the objective. To counteract this and allow images to remain in focus over prolonged periods of time, we employed a form of image stabilisation based on the same principles as the Nikon Perfect focus system (J. Huang et al. 2015; B. Huang et al. 2008).

The basic principle of perfect focus is to send a laser up through the objective in such a way that it reflects off the boundary between the glass coverslip and the sample. This is made possible by virtue of the the sample often having a refractive index substantially less than that of the glass coverslip. For example, water has a refractive index of 1.33 while glass coverslips tend to be around 1.5. This, in combination with the high numerical aperture of oil immersion objectives, means that the incident laser can be made to hit the interface at an angle greater than the critical angle, and thus be reflected. The reflected light is then collected by the objective and can then be sent, by a series of mirrors, to a sensor far from the objective itself. If the sample is now moved vertically along the z axis, then the the angle at which the laser light is incident on, and reflected from, the cover slip will change. This in turn causes a change in the position of the reflected beam emitted from the objective. If the sensor is sensitive to the change in position of the reflected spot, then this information can be used to determine how far the sample has moved in z. This allows for the construction of a feedback loop whereby the stage position initially drifts causing a change in the observed position of the reflected spot, the microscope z stage is then given instructions to move in such a way to return the spot to the initial position. The effect of this process is accurate image stabilisation.

In practice we used a heavily attenuated $1mW$ Helium Neon laser as the illumi-

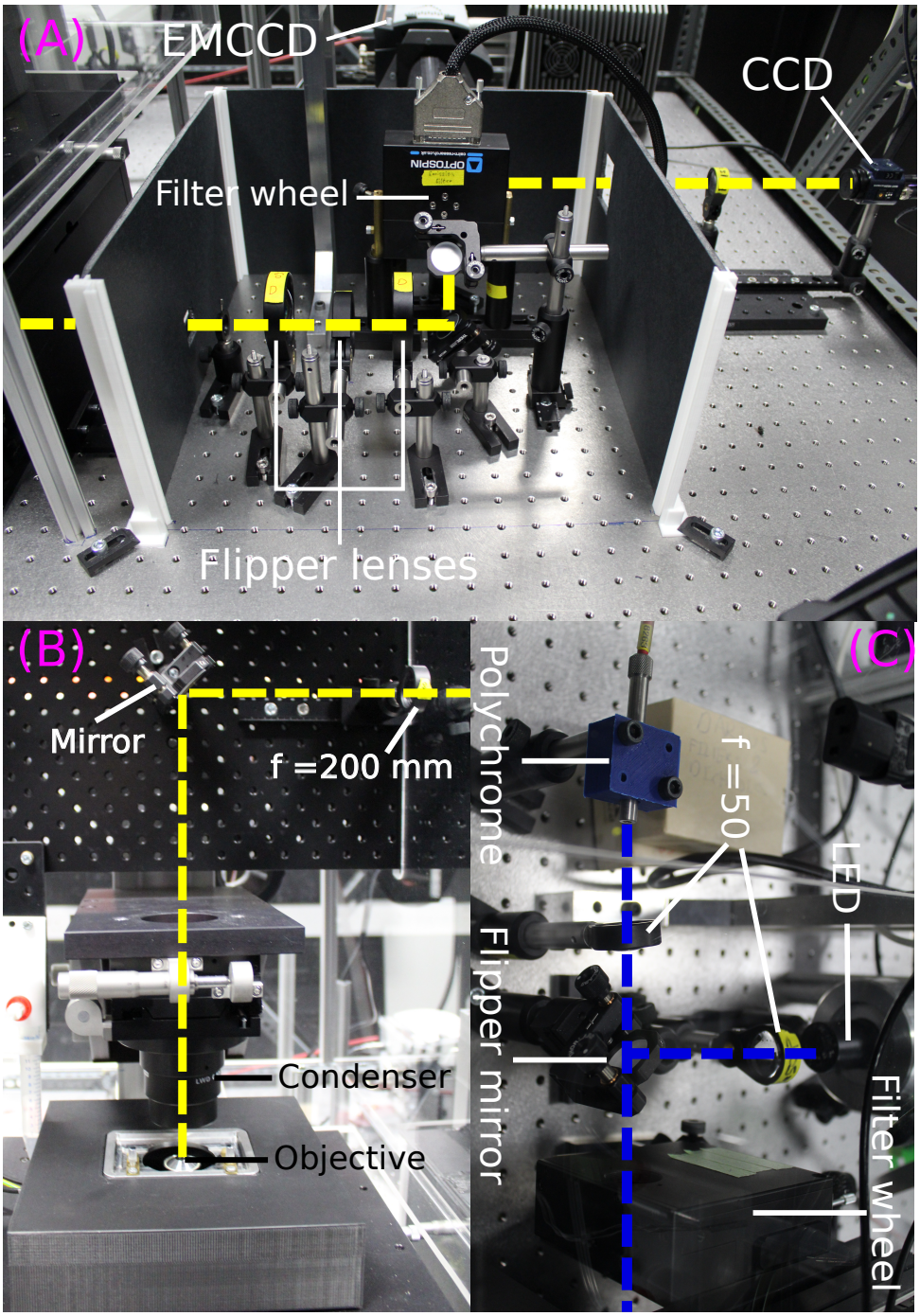

Figure 2.2: Pictures of microscope layout. **A:** Downstream optics of the objective including the flippable telescope lenses and cameras. **B:** Main optics of microscope including objective and condenser. **C:** Fluorescent illumination optics and light sources.

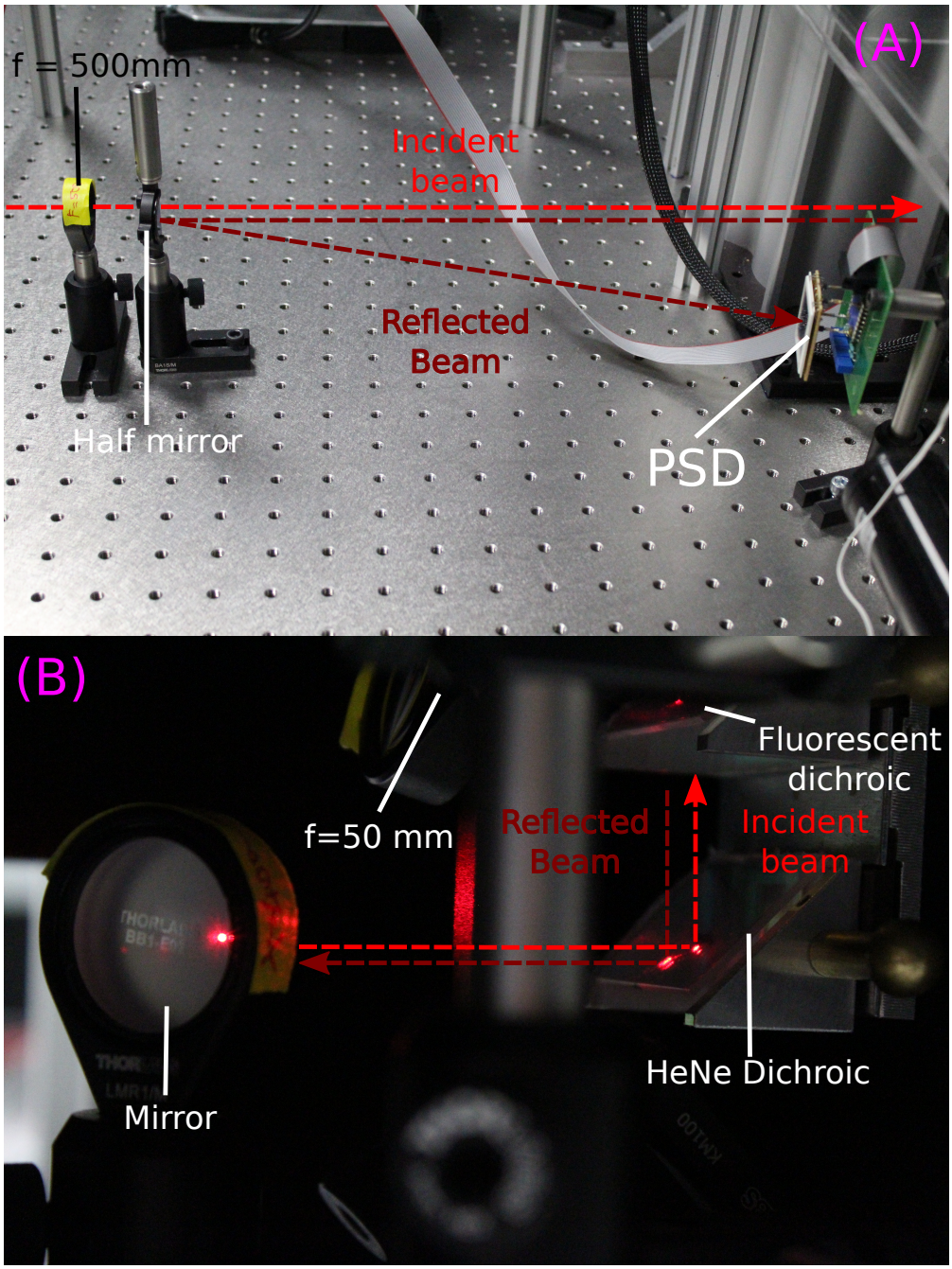

Figure 2.3: Pictures of perfect focus alignment. **A:** Picture showing optics at the back of the microscope which send in the incident beam to the objective and aim the reflected beam onto the position sensitive detector (PSD). **B:** Picture showing an example incident and reflected spot from a sample being imaged in the microscope.

nation source for the perfect focus set up. We chose this for two reasons. Firstly, it provided a bright, coherent source of illumination that was already well shaped and stable, and which could be easily focused down into a tight spot. Secondly, at a wavelength of $633nm$ the HeNe laser is less damaging to biological samples than blue or green light for example, but remains visible to the naked eye unlike infrared sources. Given the possibility for stray reflections, and transmission through the sample, having the light be in the visible spectrum was a significant safety advantage.

To align the laser a series of mirrors were used to guide the beam up through the objective perfectly vertically, and parallel to the z axis. Rather then aiming for the centre of the objective however, the goal was to send the beam as close as possible to the edge of the back aperture of the objective. This allowed the beam to be collected from the opposite edge increasing the space between the incident and return beams. To make the laser spot as tight as possible by the time it reached the back aperture of the objective, a f=$500mm$ lens was placed $500mm$ away from the aperture. To allow the laser to reach the objective while not interfering with with the imaging light path a dichroic mirror was used which reflected only a narrow range of wavelengths around $633nm$ and transmitted all other wavelengths (ZT635dcrb, Chroma). As shown in figure 2.3 the reflected and incident beams are positioned very close to one another. Therefore, to collect the reflected beam without obscuring the incident one we utilised a half D-shaped mirror to catch the reflected beam and aim it onto the surface of the sensor of interest, which in this case was a position sensitive detector (PSD).

To track the position of the reflected spot we utilised a SiTek 2L10 PSD, from Laser Components. This PSD has a large sensor area of 20x20mm which allowed the detector to be placed further down stream of the objective, this increased the lateral distance over which the beam would move for a given z displacement increasing the overall sensitivity of the system. The PSD sensor itself is composed of a semi conductor substrate which generates photo currents wherever light falls on its surface. This current then flows to one of four electrodes placed on each side of the square sensor. By comparing the currents from each electrode the position of the centre of the spot can be determined irrespective of the spots intensity or shape. The

accompanying electronics for the PSD output two voltages which represent the x and y position of the spot. In our setup these voltages are used as the input for a PID controller where only the proportional P term is used. The output of this PID controller is used to generate movements for the z stage of the microscope to keep the spot in the same position over time, and thus the sample in focus. The perfect focus system was tested by running time lapse microscopy experiments for 24 hours examining both beads and cells. We found that there was no perceptible loss of focus over this period of time, although the sample did move in x and y in many cases. This lateral movement usually occurred within the first few hour, and was significantly less than a field of view. The main use of the perfect focus system in this investigation was in the collection of images for cell size determination. As the system could maintain focus during stage movement at approximately $10\mu ms^{-1}$, I was able to automate image acquisition. The stage would move in one direction by approximately one field of view and acquire a new image. This process would be repeated multiple times to generate around 50 images per sample. The perfect focus system allowed all of these images to be taken in the same focal plane ensuring that all of the cells at the coverslip surface were in focus. This was particularly important in our setup, as coarse positioning of the sample in z was obtained by tilting of the x,y stage. Therefore the sample tended to not lie perfectly perpendicular to the objective, and as a result, lateral movements in x, y caused effective changes in z.

2.1.4 Software for microscope control

One of the main challenges of building a custom microscope is controlling all of the different elements through a single unified interface. In this work I developed a custom LabView program that utilised manufacturers VIs to control all aspects of the scopes operation. This included coordinating EMCCD and illumination settings with stage positioning, filter wheel selection and perfect focus operation. The user interface allows complex experiments to be easily set up including time lapses, multi position acquisitions and three dimensional z stacks. By using an event driven state machine approach, each complex task is broken down into a series of simpler steps ensuring that events such as camera shuttering or filter wheel selection always occur in the correct sequence. Figure 2.4 shows the main interface for the programs UI.

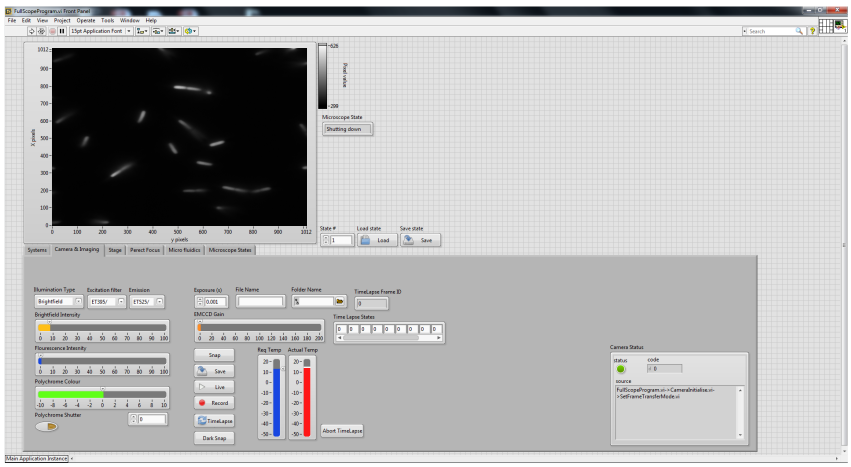

Figure 2.4: Software developed for control of the custom microscope in this investigation.

2.2 Bacterial strains and media

2.2.1 Strains

Where possible during this work it would have been preferable, to use the ubiquitous wild type *E. coli* strain MG1655 (Blattner et al. 1997). However, in many cases, such as in the measurement of cell size, it was necessary to make use of a non-motile strain that lacked the ability to swim. Additionally, it was often advantageous that the strain be visible under fluorescent microscopy. To this end I primarily used the strain YD133 with the addition of the pWR20 plasmid which encodes for both an enhanced green fluorescent protein (EGFP) and kanamycin resistance (Pilizota and Shaevitz 2012; Vats, Shih, and Rothfield 2009). The YD133 strain itself is a derivative of the Keio collection parent strain BW25113, with a number of deletions which remove the pili, flagella and hook, thus rendering the strain "*bald*" (Baba et al. 2006; Datsenko and Wanner 2000). Both BW25113 and MG1655 are descendants of the K-12 wild type strain, with the main difference being that BW25113 has been altered to block its utilisation of arabinose, rhamnose and lactose. Where possible throughout this work comparisons were made between YD133+pWR20 and BW25113 to ensure that any observed trends were not due to the presence of the pWR20 plasmid, or the deletions unique to YD133. Table 2.1 shows a summary of the main strains used in this work.

Table 2.1: Summary of strains used in this work. Genotype information taken from Baba et al. 2006 and Pilizota and Shaevitz 2012

Strain name	**Description**	**Genotype**
MG1655	Wild type strain K-12 derivative	$F^-, \lambda^-, rph-1$
BW25113	Keio collection parent strain, K12 derivative	F^-, $\Delta(araD-araB)567$ $\Delta lacZ4787(::rrnB-3)$, λ^-, $rph-1$, $\Delta(rhaD-rhaB)568$, $hsdR514$
YD133	BW25113 derivative lacking pili, flagella and hooks	$BW25113+$ $\Delta FimA$, $\Delta FliC$, $\Delta FlgE$
YD133+ pWR20	YD133 with EGFP and kanamycin resistance on plasmid	$YD133 + EGFP$ $+Kan^R$

2.2.2 Media and osmolytes

When deciding on the media to use for this investigation there were a number of factors to consider. To aid in reproducibility of results a defined media would be preferable to one with undefined components. Additionally, to allow experiments to be carried out more rapidly the media should have a high nutritional capacity and offer as fast a growth rate as possible. As in most cases throughout this work we were interested in achieving steady state growth it is also advantageous that the media not exhibit any form of diauxic shift, categorised by a change in the growth rate that results from a change in the nutrients being metabolised by cells. Finally, as we are often interested in imaging of cells using fluorescent microscopy, it would also be of use that the media have low levels of auto-fluorescence, in order to increase signal to noise ratio during these experiments. To meet as many of these competing demands as possible I decided to utilise M63 media as the base media for the majority of this work. The exact formulation of M63 media used was taken from Scott et al. 2010a, and as standard was supplemented with glucose as the primary carbon source and

cas amino acids as the main nitrogen source. To manufacture one litre of M63 media first a stock of 5x M63 salts was made as shown in table 2.2. This solution was then adjusted to pH 7.0 by the addition of potassium hydroxide pellets before being filter sterilised by passage through a $0.22\mu m$ filter.

Table 2.2: Recipe for 1L of 5X M63 salts from Scott et al. 2010a

Component	**Chemical Formula**	**Quantity ($1L$)**	**Final Concentration**
Monobasic Potassium Phosphate	KH_2PO_4	$68g$	$0.5M$
Iron sulphate hepta hydrate	$FeSO_4\ 7H_2O$	$13.214g$	$0.1M$
Thiamine hydrochloride	$C_{12}H_{17}CIN_4OS$ $.HCl$	$6.35mg$	$19\ \mu M$
Sterile water	H_2O	up to $1L$	-

To create one litre of M63 media the M63 5X salts were diluted to their working (1X) concentration by the addition of sterile water. Simultaneous to this, one of a choice of possible carbon sources was added (usually glucose). Cas amino acids were also optionally included to increase the growth rate. Finally Magnesium Sulphate is also added to provide a source of Magnesium ions, it was not included in the M63 salts as it has a tendency to cause precipitation when left for extended periods of time. Finally, the resulting solution is then filter sterilised again by passage through a $0.22um$ filter. Table 2.3 summarises this process.

In addition to the M63 media listed above it was also necessary at times to utilise other media and buffers. For example, some experiments were performed in Lysogeny broth (LB) in order to facilitate comparison with previous works, and Phosphate buffered saline (PBS) solution was used at various stages during flow cytometry experiments. Table 2.4 shows the formulation of all the other media and buffers used throughout this work.

As one of the main goals of this work was to determine the effect of osmolarity

Table 2.3: Recipe for 1L of M63 media from Scott et al. 2010a

Component	**Quantity (1L)**	**Final Concentration**
5X M63 Salts	200ml	1X
Cas amino acids	2g	0.2% (W/V)
Carbon source (Glucose/glycerol etc)	25ml of 20% (W/V)	0.5% (W/V)
Magnessium Sulphate ($MgSO_4$. $7H_2O$)	0.5g	2mM
Sterile water	Up to 1L	-

on cell physiology it was often necessary to increase the osmolarity of many of the standard media and buffers described above. To this end three different osmolytes were used including: sodium chloride, sorbitol and sucrose. In the case of M63 media, high osmolarity solutions were achieved by adding varying amounts of concentrated (1-4M) aqueous osmolyte solution during the preparation of the final M63 media. For the remaining buffers and media the desired mass of solid osmolyte was dissolved into the already prepared media to yield the required concentration. In all cases the resulting high osmolarity solution was again filter sterilised.

2.2.3 Cell culturing

For the experiments conducted in this work great effort was made to ensure that the bacterial cultures grown were not only exhibiting exponential, but also steady state growth. To this end the protocol for culturing cells was as follows. First, cells from an agar plate were inoculated into 20ml of fresh media (M63 + glucose + cas amino acids) and grown overnight well into stationary phase. The next day 0.2μl of overnight culture was inoculated into 20ml of fresh media and allowed to grow to early exponential phase, between 0.1 and 0.3 optical density at a wavelength of 600nm (OD_{600}). Next a frozen stock was created by mixing 3ml of culture with 2ml of 50% glycerol solution. The resulting mixture was dispensed into 1.5ml micro centrifuge tubes in aliquots of 100μl each, and flash frozen in liquid nitrogen. The addition of glycerol to the mixture helps reduce the number of cells which are lost during the freezing and thawing process (Sharp 1984; Sprouffske, Aguilar-Rodríguez, and Wagner 2016). These frozen aliquots were stored at -80°C for several weeks until

Table 2.4: Recipe for other buffers and media used in this investigation.

Media / Buffer	**Constituents**	**Main uses**
Lysogeny Broth (LB)	Yeat extract 5g, Tryptone 10g, NaCl 10g, Up to 1L in H_2O	Rich growth media used to compare results with previous work
Phosphate buffered Saline (PBS)	154 mM NaCl, 5mM Na_2HPO_4, 1.5 mM KH_2PO_4, adjusted to pH 7.5	Used as a wash buffer for flow cytometry experiments.
Rich defined media (RDM)	1X MOPS, 1.32mM K_2HPO_4, 1X AGCU, 1X EZ supplement	High growth rate defined media.

required. To prepare a sample of cells for each experiment, a single tube of this frozen stock was thawed at room temperature and again $0.2\mu l$ was added to $20ml$ of fresh media of the kind required by the particular experiment. Cells were again allowed to grow to an OD_{600} of between 0.1 and 0.3 at which point they were harvested for the required experimental protocol. At all stages in this process cells were grown in $50ml$ conical flasks, at $37°C$ and with constant shaking at $200rpm$.

2.3 Growth rate measurements

By far the most common means of measuring the growth rate of bulk cultures of bacteria is to make use of spectrophotometers (Maia et al. 2016; Cashel et al. 1996; Scott et al. 2010b). Typically the optical density is measured at a wavelength of $600nm$ (OD_{600}) and is related to the concentration of cells C via the Beer–Lambert law (Stevenson et al. 2016):

$$OD_{600} = C\sigma Z, \tag{2.1}$$

where σ is the cross section of the bacteria and relates to its size and Z is the path length of the sample. When concentrations are low and cell sizes are fixed the OD_{600} value responds linearly with increasing cell concentration. However as the concentration increases and the effects of multiple scattering become apparent this

linear relationship is lost (Arthur L Koch 1970). For this reason when measuring cell number using the spectrophotometer the sample was diluted to ensure that the read value fell within the linear range between 0.05 and 0.1. Each sample was measured using a $1.6ml$ cuvette (fisher scientific) and a Shimadzu UV 1280 spectrophotometer. To determine the growth rate a series of time points were taken for each sample and gradient of the semi log plot taken to compute the growth rate and doubling time.

2.4 Cell size from fluorescent microscopy

2.4.1 Experimental protocol

The cellular size of an *E. coli* is one of the most obvious and accessible physiological parameters that can be measured. Despite this, there still appears to be no consensus on a singular best experimental method for preparing cells, such that their size and shape can be measured. For example, one very widespread method involves the use of microfluidic mother machine devices (P. Wang et al. 2010; Taheri-Araghi et al. 2015). However unless these devices are very carefully designed and manufactured they can have significant affects on cellular dimensions, growth rates and the mechanical forces which cells experience (Yang et al. 2018). Similarly, another common method involves dropping cells in liquid cultures onto agar pads, and allowing them to adhere to the surface before imaging (M. W. Bailey et al. 2014; Adiciptaningrum et al. 2015; Young et al. 2012; Basan et al. 2015). This method often involves a period of drying where the liquid from the bacterial culture is allowed to evaporate before a cover slide is applied to the agar surface. It seems possible, that as the water evaporates during this drying process, many of the solutes present in the culture media remain. This could lead to an increased osmolarity at the surface of the agar pad, potentially exposing cells to a hyper-osmotic shock. As hyperosmotic shocks have been shown to reduce cell volume (Pilizota and Shaevitz 2012) there seems to be legitimate cause for concern that cell volume measurements made on agar pads may not accurately reflect the steady state volumes cells obtain in liquid culture, especially if the media used is already of a high initial osmolarity. Additionally, cells growing in agar pads experience significant mechanical forces which can act to lower their elongation rates in correspondence with the mechanical properties of the agar they are growing in (Tuson et al. 2012). For these reasons it seems likely that growth on agar pads may initially be accompanied by a departure from the steady

state conditions of liquid culture, where our interest primarily lie. As a result in this investigation we chose to use a much simpler assay for the measurement of cell size by optical microscopy.

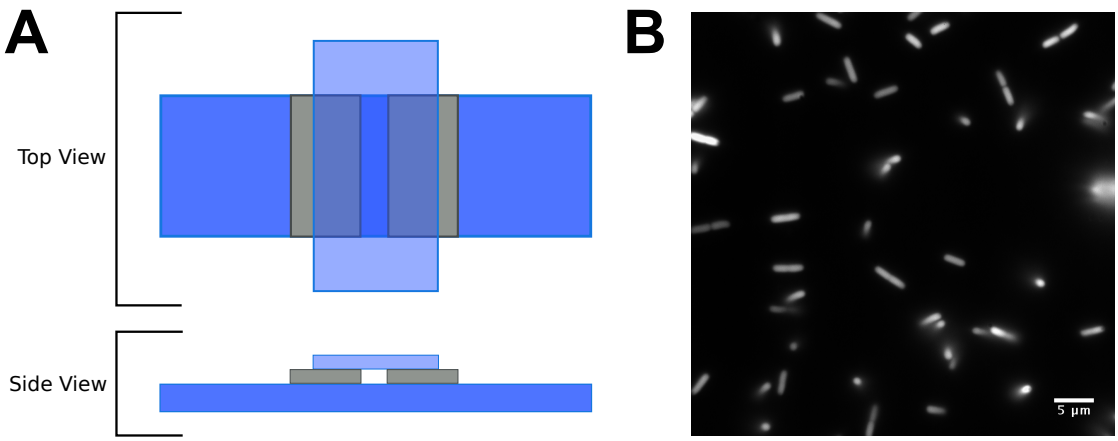

Figure 2.5: **A:** Schematic of tunnel slide assay used in this work. Two pieces of double-sided adhesive tape (shown in grey) are attached to a standard microscope slide (shown in dark blue). A standard 22x44mm coverslip (shown in light blue) is then placed on top of the tape to from a tunnel which is approximately 25mm in length, 5-10mm in width and 100μm in height. **B:** Example image taken using tunnel slide and fluorescent microscopy in the GFP channel. Cells are YD133 + pWR20.

In order to image cells while perturbing them as little as possible we opted to use a tunnel slide assay as described in Pilizota, Bilyard, et al. 2007. Figure 2.5 shows a schematic of the assay which consists of two parallel pieces of double-sided adhesive tape placed a few millimeters apart on a standard microscope slide. A thin glass coverslip is then placed on top of the channel to form a tunnel approximately 100μm in height. The sample can then be taken to the microscope and imaged through the coverslip surface.

Using this tunnel slide assay cells were prepared for microscopy measurements as follows. First cells were grown from frozen stock as discussed in section 2.2.3. When the culture reached an OD_{600} of between 0.1 and 0.3, approximately 20μl of culture was then loaded via capillary action into a tunnel slide that had been prewarmed for at least 30 minutes at 37°C. The tunnel slide was then orientated with the coverslip

downwards and allowed to incubate at $37°C$ for 15 minutes. As non-motile YD133 + pWR20 cells were used for these experiments by the end of the incubation period the vast majority of cells in the sample had sedimented on to the coverslip surface. This sedimentation is advantageous for imaging for two reasons. First, cells at the coverslip surface tend to become attached and therefore are not moved as significantly by Brownian motion compared to free floating cells. Secondly, cells at the surface have a tendency to lie with their length parallel (or very roughly parallel) to the coverslip surface. This tends to keep the cell body of more cells in a single focal plane, making the resulting imaging and analysis far simpler than in free floating culture, where cells are aligned in a random orientation often at an angle or even perpendicular to the focal plane.

In general we chose to image cells using fluorescent microscopy, as apposed to bright-field microscopy, for two main reasons. Firstly fluorescent microscopy tends to have higher signal to noise ratios and contrast when compared to bright-field imaging. Secondly fluorescent images are less affected by cells lying outside of the image plane. In bright field imaging, cells leave visible airy patterns even when several microns from the actual focal plane. While most cells in our sample were sedimented at the time of imaging, not all were. To reduce the affect of these non sedimented, and therefore out of focus, cells we opted to use fluorescent imaging.

Fluorescent images were taken using an exposure time of $0.1s$ and around 50 EMCCD gain at full LED brightness. The gain was altered slightly around this value to allow for day to day variation in cell brightness and imaging conditions. For each condition around 50 fields of view were randomly selected and imaged per slide, ensuring no overlap between images. At least two slides from different days were imaged per condition resulting in the final analysis of at least 150 cells per condition. We estimate that between removing cells from flask growth to the end of imaging took no longer than 30 minutes, which in most cases was significantly less than a single doubling time.

2.4.2 Image analysis and volume extraction

To extract cell volumes from the gathered fluorescent images we employed the use of custom algorithms and software which are described in detail in this section. Briefly we first manually selected regions which contained cells which appeared to be parallel to the image plane, as apposed to those angled out of it. Next we threshold the image using a custom machine learning algorithm and then vetted the results of that segmentation manually by comparison to the original selected region. Finally we used a custom python routine to locate the approximate location of the poles, and by fitting a polynomial through the poles and a cellular back bone we were able to compute the length and corresponding width of each cell.

2.4.2.1 The need for a custom algorithm

In section 2.4.1 I outlined in detail how fluorescent micrographs of cells were obtained. The purpose of collecting these images was to allow the extraction of the size and shape of the cell populations which the images contained. Given that this is a very common task performed in the literature one might be forgiven for assuming that a standardised, and wide spread method had emerged to tackle it. Unfortunately, this is not the case. In general, to determine the size and shape of a cell there are two independent tasks which must be performed: cell segmentation, and measurement of cell morphology. Cell segmentation refers to the ability to correctly label each pixel in an image as either belonging to a cell, or belonging to the image background and can be performed in a number of different ways. In the literature it is common to see a wide variety of methods either using custom code and algorithms (Taheri-Araghi et al. 2015; Ursell et al. 2017; Guberman et al. 2008) or algorithms constructed using readily available commercial software such as microbe tracker (and its successor oufti) or cell profiler (Carpenter et al. 2006; Volkmer and Heinemann 2011a; Paintdakhi et al. 2016). Likewise, when it comes to determining cell morphology a number of different strategies are employed. Most frequently cell volume is determined by assuming that cells are spherocylinders and then measuring their length and width directly. However the length and widths can themselves be measured in a number of different ways from the segmented data. For example, by using the ferret diameter (Basan et al. 2015) or pole to pole distance (Guberman et al. 2008). Part of the reason for this wide range of solutions is that the needs of each experimenter vary, and as as result there is no one algorithm or method which will yield the best results

in all situations. It is for this reason that in this investigation I too developed a custom pipeline for extracting the length and width of cells from fluorescent images. In this section I will outline exactly how this process operates and the results which we were able to obtain by using it.

2.4.2.2 Selecting cells for analysis

One of the unique challenges of the experimental set up I used in this investigation was that not all cells were aligned with the image plane. Therefore, it was necessary to somehow screen each cell in the image and come to a determination if the cell was sufficiently aligned to allow an accurate estimate of its size and shape. I invested significant effort in attempting to develop an automated method for making this discrimination, but ultimately I was not successful in finding a satisfactory solution. It is in fact very difficult, even manually by eye, to discriminate "flat" and "non-flat" cells *i.e* those aligned and unaligned with the focal plane. Given this, the decision was made to instead manually select which cells would be considered for analysis, non ideal though this solution was. To help make this manual selection process as easy and as quick as possible I developed custom python code, based on the matplotlib package, which allowed me to quickly select all the cells in an image which I thought were good candidates for further analysis. The GUI for this step in the pipeline is shown in figure 2.6.

2.4.2.3 Convoloutional neural networks as image segmenters

With areas containing flat cells now selected the next step is to segment the cells from their background. As we mentioned before there are many ways in which this is typically done and I tried a number of different methods to determine which gave the best results. In particular I tried a number of methods based around intensity thresholding of cells. These methods utilise either a global threshold for the whole image, or an adaptive threshold which is set locally for each pixel based on some property of the surrounding area. In general I found that neither of these approaches worked well. Global thresholds tended to fail as a result of cell to cell variability in intensity, which meant that dim cells were not identified and bright cells were segmented much larger than they should have been. Adaptive thresholds on the other hand performed well in areas close to single isolated cells but performed poorly in cell free areas, where they tended to erroneously identify background noise as cells.

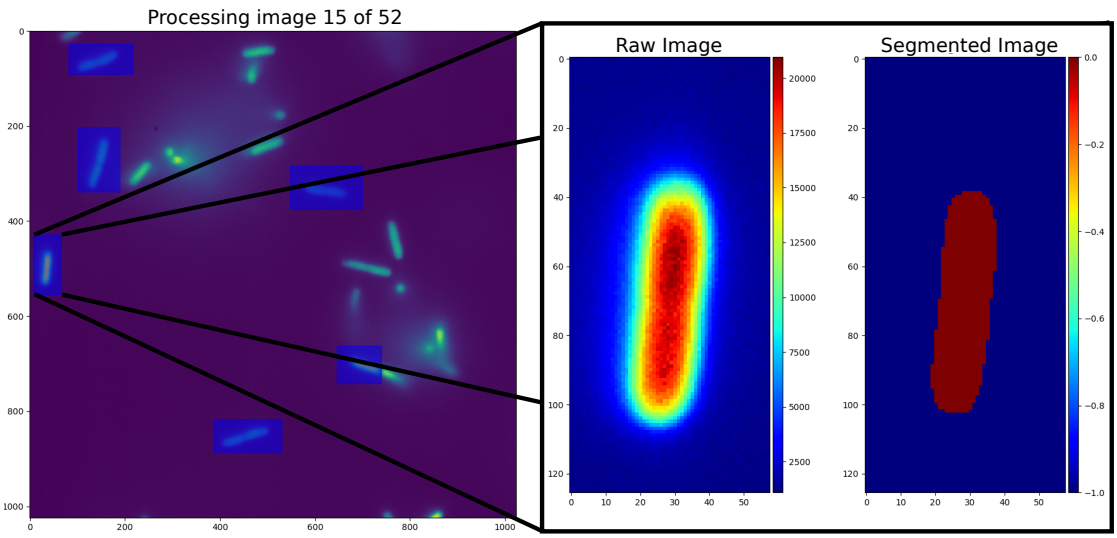

Figure 2.6: Manual selection of flat and non-flat cells. The left hand image shows the raw microscopy image as presented to the user. The user then marquee selects any number of regions which they think contain flat cells. These sections are shown highlighted in blue. After all selections have been made the user is shown each cropped region of the raw image next to the segmented image. They are then able to keep or discard the images based on how successfully they have been segmented.

Additionally both methods struggled to accurately segment cells when they were in close proximity such as shortly after division. In these cases they often segment multiple cells into a single aggregate. To segment cells in a more accurate and robust manner than threshold based methods can achieve I turned to machine learning based methods and in particular convolutional neural networks or CNNs.

To understand how CNNs can be used to segment cells I will give a brief overview of their original role as image classifiers. More detailed explanations can be found widely, including in the papers in which they make some of their first appearances such as LeCun et al. 1998 and Cireşan, Meier, and Schmidhuber 2012. Simply, the task for CNNs when acting as image classifiers is to take a large two dimensional array of pixel values which make up the input image, and reduce this down to a small number of values: with each value representing the probability that the image belongs to a particular category or class. Take for example the MNIST (Modified National Institute of Standards and Technology) data set, which is commonly used for the benchmarking of new image processing algorithms. The MNIST data set consist of

60,000 images of hand written digits, in the range zero through nine. The goal for a CNN when applied to MNIST is to take a given image and correctly identify which digit it represents. To do this a CNN starts by convolving the input image with a number of small square matrices known as filters or kernels (see figure 2.9 for an example). These filters, normally around 3x3 in size, contain learnable weights, which are altered during the training phase (which I will discuss in more detail shortly). After convolution the next step in a typical CNN is max pooling. During max pooling the image dimensions are reduced by splitting the image into small regions (typically 2 by 2) and replacing each region by the maximum pixel value it contains. Usually multiple rounds of convolution and max pooling are repeated to substantially reduce the image size before the flattening stage occurs. During flattening, all the values in the now reduced image are reorganised into a one dimensional vector. This vector then forms the input for a traditional fully connected neural network. It is common to see multiple (increasingly small) fully connected layers which lead to a final output layer containing the same number of neurons as categories (10 in the case of MNIST for the digits 0 through 9). Figure 2.7 shows an example schematic of a CNN classifier for a MNIST like data set.

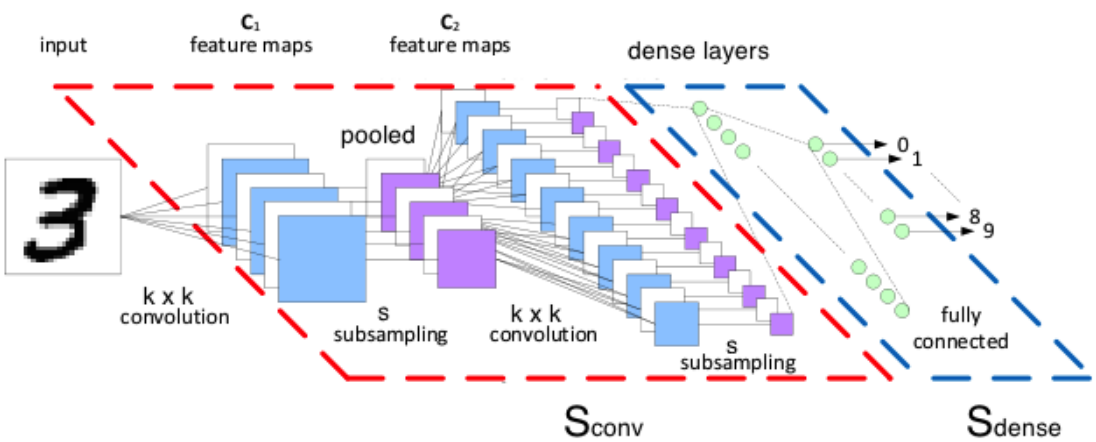

Figure 2.7: Example of a classifier convolutional neural network. Convolutions using a kxk kernel result in feature maps shown in blue. Down pooling operations lead to reduced image sizes shown in purple. The fully connected layer is formed by flattening all of the final feature maps into a one dimensional array and is shown in green. Image was taken from: `http://www.mriaz.me/`

To ensure the CNN is able to capture the non linear mapping between input image and output classification the result of each convolution is passed through a

non linear function. A number of possible functions exist for this task but the one encountered most frequently (and used in this investigation) is the rectified linear units or ReLU function. ReLU introduces non linearity in the network by leaving positive neuron values unaltered and mapping negative neuron value to zero. Neuronal values are passed through the ReLU activation function both in the dense fully connected layers, as well as in between convolutions and max pooling steps.

To train the network such that it correctly identifies each character, a learning phase is performed. During learning, an image of a known character is input to the network and the output layer computed using random weights. By weights we mean both the values inside the filter matrices as well as the strength of connection between neurons in the fully connected layers. As random weights are used, initially the output values of the network will also be random. To improve the network over time each weight has its value altered in a direction which acts to minimise the chosen error function. The error function is simply some metric which measures the difference between the actual output of the network and the expected output for that training image. For example, the error function could simply be the mean of the squared differences between each value in the output layer and the desired output value. The success of neural networks comes from choosing appropriate error functions and in the ability of sophisticated optimisers to accurately locate minima in this chosen error function. This is a difficult task as the error function is both guaranteed to be non convex[1] and has as many dimensions as weights in the network (often in excess of hundreds of thousands). For this reason much work has been put into the development of elaborate optimisers such as Adam (Adaptive moment estimation) or momentum based optimisers (Kingma and Ba 2014; Qian 1999). By presenting the network with many training images where the outcome is already known, and by updating weights to minimise the error function after each one, the network slowly learns to classify images correctly such that when a new image of unknown category is presented the network will correctly identify it[2]. However, as learning can continue to improve network performance even after all the training data has been shown, it is very common to see networks repeatedly trained on the

[1] providing the number of weights greater than two

[2] In practice the networks are not shown one image at a time but instead a small collection of images called a batch. The weights are then updated such that the error for the entire batch is reduced.

same data set. Each showing of a full training data set to the network is referred to as an epoch and it is not unusual to see networks trained for many epochs on the same data set. The number of epochs a network is trained on is a balance between two competing factors. Obviously, more epochs allow for more adjustment of weights and therefore better network performance. However, too many epochs can lead to over-fitting, where the network learns the specific nature of the training set rather than the more general principles involved. This results in poorer performance for new data sets which have not been used to train the network.

While a mechanistic understanding of exactly what happens during the training of a CNN classifier is useful, it would also be of benefit to have an intuitive understanding of why the network architecture works so well. While neural networks are famously difficult to interpret (Boz and Hillman 2000) we can get some intuition by examining intermediate layers of trained networks and by considering convolutions as "feature" extractors. Each CNN has many different filters in each convolutional layer and, in general, they are all unique in a fully trained network. The result of a convolution is often referred to as a "feature map" as each filter is thought to be picking out one particular feature in the image, such as edges, vertical or horizontal lines or other particular patterns or shapes. By having multiple such feature maps and by combining them into a single fully connected layer CNNS are able to integrate information about the presence or absence of many different features in the image and use this to decide which category the image belongs to. Additionally, by down pooling and reducing the image size before subsequent convolutions, the network is able to integrate both local and global variations in pixel values, combining them to help improve categorisation. Figure 2.8 shows an example of a CNN trained for analysing hand drawn digits much like the MNIST data set. In the figure all of the intermediate layers of the trained network are visible and their responses are shown for a previously unseen image containing a four. It can be seen that all of the convolved images in the first layer are unique and different, despite the fact that they all still resemble a four. Once can imagine from these layers how each filter can be picking a different feature to extract. Interpretation becomes much more difficult however after passing the second convolution step, and by the time we reach the fully connected layer it becomes almost impossible to see a connection between the input image and pattern of layer activity.

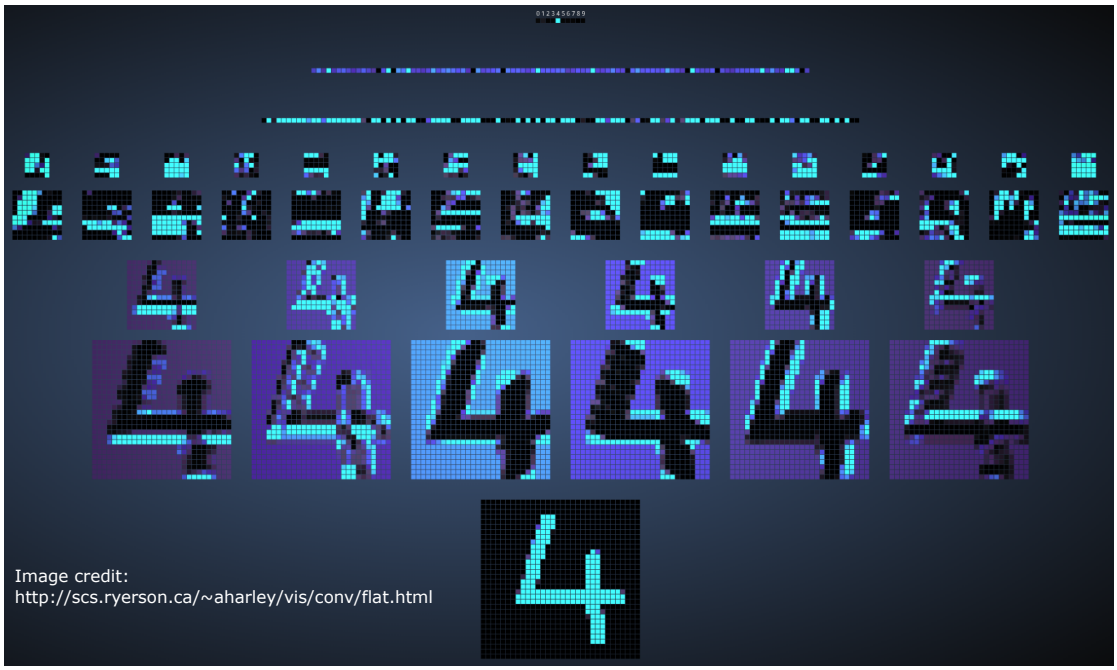

Figure 2.8: Example of a convolutional neural network to identify hand written digits similar to the MNIST data set. Here the activation values for all intermediate layers are shown. From bottom to top the layers shown are: input layer; first convoloutional layer; first down pool, second convoloutional layer; second downpool; first fully connected layer; second fully connected layer and finally output layer. Image was generated using the excellent online tool available at: `http://scs.ryerson.ca/~aharley/vis/conv/flat.html`

With an understanding of how CNNs can be used to classify images we can now turn to the problem of segmentation. Segmentation can in essence be thought of as a problem of classifying each pixel in an image into one of two categories: cell or not cell. To achieve this goal there are a number of possible methods, but I will focus on the two which I utilised in this investigation. The first method involves an individual assessment of each pixel in the image. The image is divided into a number of smaller image regions which are centred around each and every pixel in the image. Each region can then be classified individually by a standard CNN, exactly as with the example above. The only difference is that the final classification relates not to the entire image but to only the central pixel in each image. By taking the classification of each individual pixel, one can assemble a new segmented version of the original image. For this reason these networks are called sliding-window or pixel classifier

networks. Examples of such networks used in the analysis of biological data sets can be found in Van Valen et al. 2016 and Ciresan et al. 2012. The advantages of this approach is that the network architecture is quite simple, and has relatively few parameters compared to other methods, resulting in comparatively shorter training times. However , there is also a number of disadvantages to this approach. Firstly, to categorise each pixel one requires to process a small image region. These regions have high overlap with their neighbours and so a lot of computation is wasted processing almost, but not quite, identical images. Additionally, because the network needs to be forward propagated for each pixel in the image, the time to segment a whole image is quite large and grows with the square of the image size. Finally, because each pixel is classified based on a small region around it, the network is not sensitive to patterns which have length scales larger than this region. This makes choosing the size of this region an important parameter in determining the effectiveness of the model.

To try to address some of these shortcomings a category of networks has been developed, sometimes referred to as fully convolutional neural networks (FCNNs). These networks are exemplified by the likes of U-net (Ronneberger, Fischer, and Brox 2015). FCNNs are very similar in structure to standard CNNs, the exception being the absence of a fully connected layer. Rather, FCNNs are composed of two sections an: encoder and decoder. The encoder is functionally identical to the beginning of a CNN, consisting of alternative convolutions and down poolings. The decoder section, which takes the place of the fully connected layers in a a traditional CNN, performs exactly the opposite job of the encoder network. It does this by alternating convolutions with up sampling steps. These up-sampling steps can take different forms but, so called "transpose convolutions" are common. Transpose convolutions are difficult operations to intuitively understand and can be read about in more detail in Dumoulin and Visin 2016, but I will give a brief explanation here. In essence transpose convolutions perform the opposite task from a normal convolution. While, a standard convolution maps many pixels in the input image to a single pixel in the smaller output image, a transpose convolutions maps a single pixel in the input image to many pixels in the output image. The number of pixels in the output layer that each pixel in the input layer effects is controlled by the size of the filter used. The filter weights control how much each pixel in the

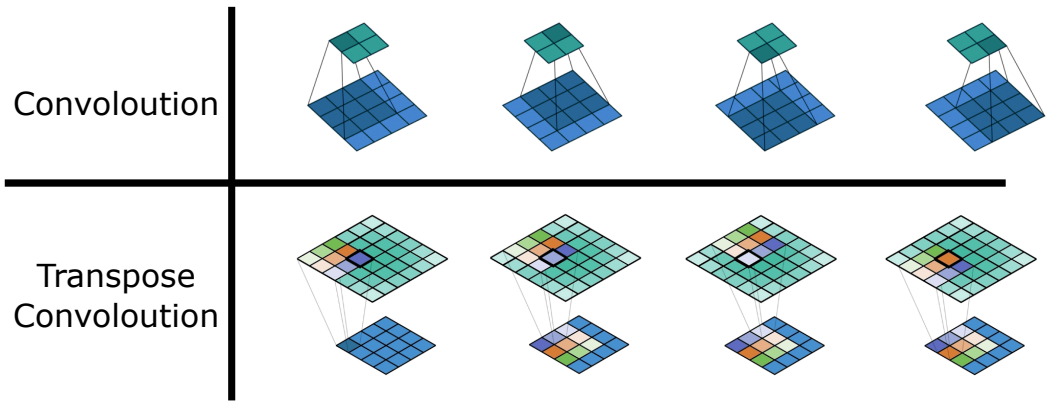

Figure 2.9: Example of a convolution and transpose convolution for a 4x4 input array shown in blue, and a 3x3 filter. In the case of convolution, each pixel in a 3x3 area is multiplied by a corresponding weight and the results summed to give rise to a single output pixel. In the case of a transpose convolution each input pixel contributes some portion of its value (dictated by the filter weights) to each pixel in a 3x3 area in the output image. These contributions are then summed to obtain the final pixel value in the output. The different colours in the transverse convolution images indicate the different weight strengths. Image adapted from Dumoulin and Visin 2016.

input contributes to each connected pixel in the output. Figure 2.9 shows a comparison between a convolution and transpose convolution to help clarify this distinction.

The effect of the decoder network in FCNNs is to regenerate an image of the same dimensions as the input, but where each pixel is now correctly assigned a class label. To ensure that information from large, and small, length scales are incorporated together when computing the output, a common practice is to combine information from the encoder and decoder layers. For example, in U-net, after each up sampling step the resulting feature maps are concatenated together with the corresponding maps from the encoder network. This helps to ensure that the small scale variation present in the encoder images can be analysed together with the larger scale information present in the fully encoded images. Figure 2.10 shows a schematic of the U-net architecture as presented in Ronneberger, Fischer, and Brox 2015.

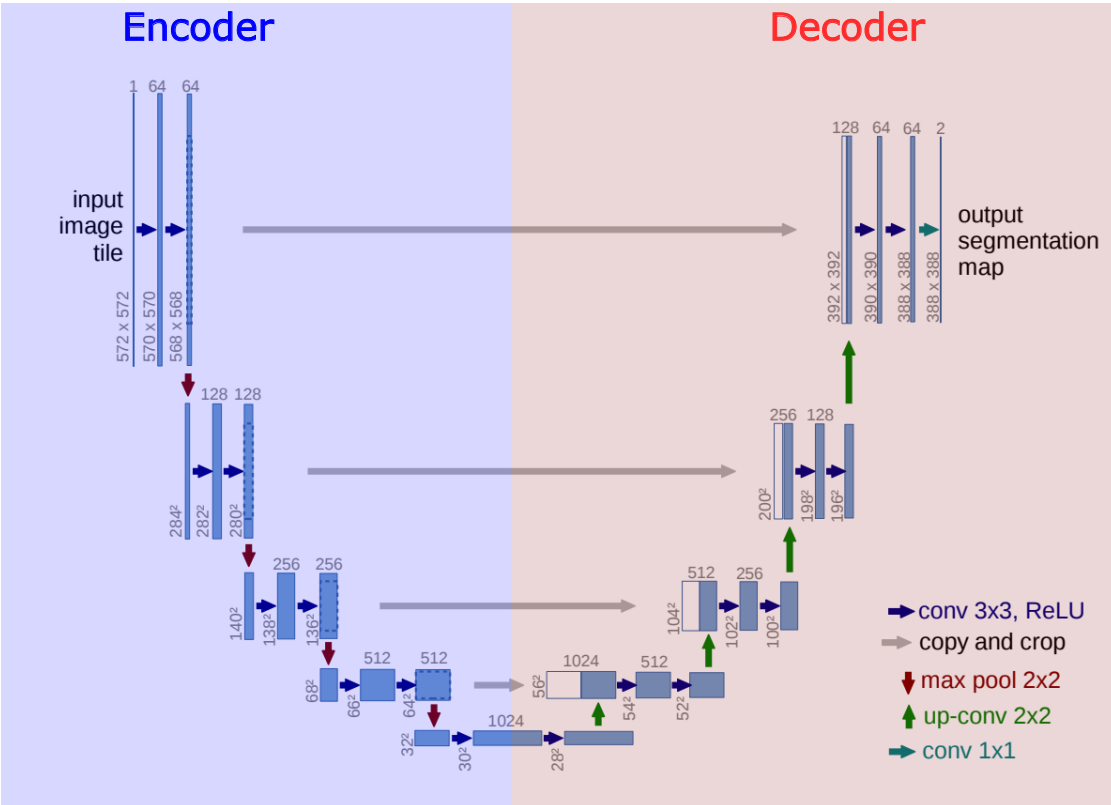

Figure 2.10: U-net architecture. U-net combines an encoder network which looks like a typical CNN with a decoder network which performs up sampling using transpose convolution. The horizontal grey arrows represent the combining of feature maps from the encoder and decoder network by concatenation, and are one of the defining features of U-net architectures. Image adapted from Ronneberger, Fischer, and Brox 2015.

2.4.2.4 Network architectures

In this investigation I created two networks for image segmentation. The first a sliding window network and the other based on U-net architecture. The reason for constructing two networks was to determine which was more suitable, and effective, for my particular application. In this section I will outline the exact make up of these networks.

As mentioned in section 2.4.2.3 the size of the area analysed around each pixel in a sliding window network (called the receptive field) is an important factor which determines the success of the network. If the receptive field is too large, too much

surrounding context will alter the networks ability to correctly identify the class of the central pixel. Likewise if it is too small, there will be insufficient context to make a correct determination of the pixels class. To exemplify this, consider a sliding window network where the receptive field was significantly smaller than the width of a single cell. In this case when trying to determine the class of a pixel inside a cell, the receptive field could contain no background pixels. As it is normally the contrast between background and cells which allow us to differentiate them, in the absence of any background pixels it would prove very difficult for the network to correctly distinguish the cell interior from the background noise. Indeed, while we tried many different receptive field sizes, we found best results when the receptive field was around the same size as the cells in the image. Therefore, we used a receptive field size of 25 pixels which was slightly larger than the width of the average cell. Following this, the model includes two convolution and down pool stages using ReLU activation functions after each convolution. The number of filters used in these stages were 100 and 200 respectively. We found this made a suitable compromise between the time taken to learn and predict, and segmentation accuracy. Finally, we utilised a single fully connected dense layer which was 100 neurons in size. Finally, we output into a two neuron layer using a soft max activation function. The error function used for learning was categorical cross entropy and an Adam optimiser was used. As with all networks in this investigation the code was implemented using the Keras python package and a tensor flow back end (Chollet et al. 2015; Martin Abadi et al. 2015). The model summary, showing the layer types and dimensions as output from Keras is shown in figure 2.11.

The second network I constructed was based on U-net architecture. The main difference is that while U-net was designed to analyse small portions of an image and tile the results, my network was designed to operate on entire microscopy images in a single pass. As a result my starting layer was around twice the size of U-nets at 1,024x1,024 pixels. From there, I followed a very similar process to the original U-net with four stages (or blocks) in the encoder and likewise four stages in the decoder. The encoder blocks, which we call down blocks, each consisted of two sequential convolutions and a down-pool, which halved the image size. After each down pool a dropout layer was inserted. Dropout layers are only relevant during the training phase and are used to randomly switch of a fraction of weights in the network in

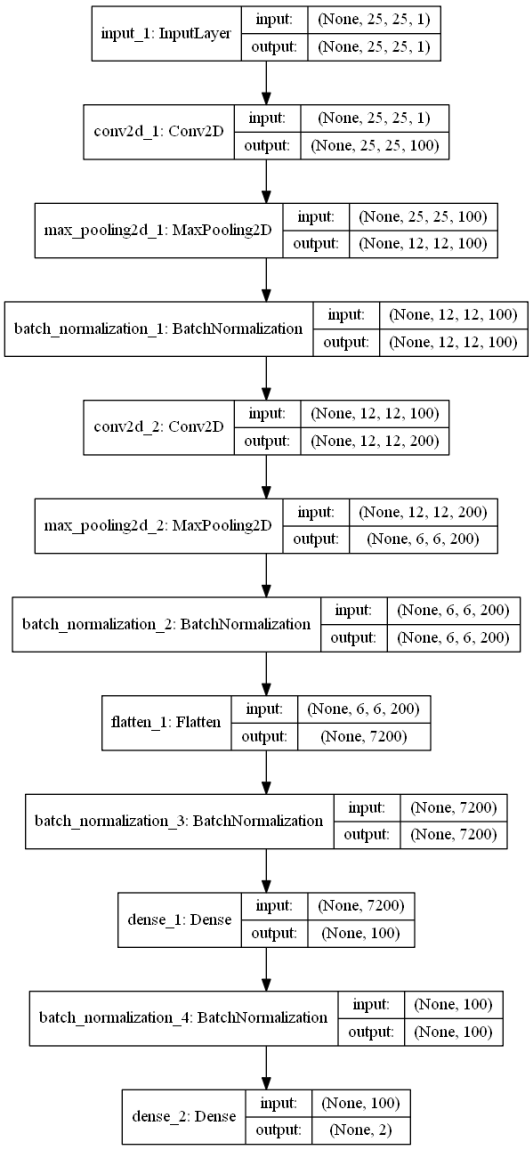

Figure 2.11: Sliding window network architecture. This figure shows the Keras layers which are assembled to form the sliding window network used in this investigation. Each box represents a layer of the model. Each layer is labelled on the left with its name and type (max pool, 2D convolution etc). On the right hand side of each box the dimensions of the input and output layer are shown. These dimensions are listed as four dimensional by default in Keras. Therefore a layer of dimensions (None,21,21,100) represents a 21x21 pixel image, and 100 filter layers. The batch normalisation layers are used to increase learning speed by normalising the output activation of each layer.

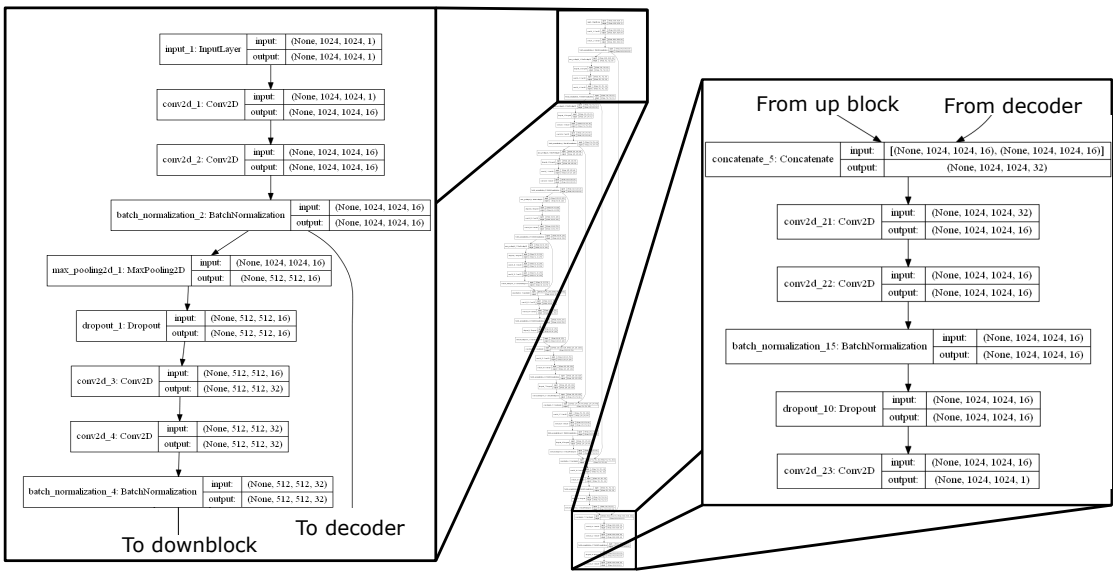

Figure 2.12: U-net like architecture. The central image shows the entire network summary as output by Keras. The enlarged image to the left shows the input layer and and first down block section of the network. Likewise, the rightmost enlarged image shows a detailed view of the final up block section and network output. In both cases the layer types and dimensions are shown, just as with the sliding window network in figure 2.11.

attempt to reduce over-fitting, and to improve network robustness. Similarly, each up block consisted of a single transverse convolution (doubling the image size) followed by two standard convolutions. In both the up and down blocks, the filter size was 3x3, and the number of filters doubled from 16 for the largest layers, to 256 for the smallest. Again, after each convolution, the ReLU activation function was applied and an Adam optimiser was chosen. Unlike with the sliding-window network I chose not to use two classes for the image output. Instead I mapped the output between 0 and 1 using a sigmoidal activation function, where 1 represented a cell and 0 the background. This allowed the entire network to map from a single monochrome image to a single monochrome image of the exact same size. As a result of this choice, I opted for Keras's binary cross entropy error function. Figure 2.12 shows a summary of the U-net model used as generated directly by Keras.

2.4.2.5 Data normalisation, augmentation and training sets

To aid in training and to attempt to increase robustness, images must be normalised before they are fed into the networks. There are a number of ways this is typically done for example by rescaling by the median, or the mean. I found that for the sliding window network it was best to rescale each receptive field individually by dividing each pixel value by the median. By contrast, with the U-net model I found best results when normalising each image between -1 and 1. Both training sets and unknown images were rescaled before being processed by the networks

To train the models in a way which best allowed comparison we used the same training set for both networks. This training set was selected to give as wide a range of examples of fluorescent cells as possible. To this end, images contained cells of different sizes, shapes and intensities. For each training image an answer key had to be created showing the desired output of the network. To do this I first used the adaptive threshold tool in fiji (Schindelin et al. 2012) to create binary images from the raw microscopy data. I then manually corrected the binary image using Gimp image editing software, until I was satisfied that the new image correctly represented a segmented version of the input. The training images and created answer keys are shown in figure 2.13.

To increase the quantity of training data without having to laboriously create more answer keys, each training image (and its corresponding answer key) was transformed a number of times with each transformation appearing in the training data set. Images were rotated 90°,180° and 270° and flipped vertically and horizontally. This meant that for each unique training image the network actually received 13 images. Additionally, to further reduce the number of answer keys that had to be manually generated, a few training images were collected using multiple different EMCCD gains and exposure times. This allowed multiple images with different signal to noise ratios to be added to the data set while only having to generate a single answer key.

Paying close attention to the training data set in figure 2.13 one might notice that almost all the cells are aligned with the focal plane, a scenario that is not always the case during the experimental measurement of cell size. Indeed, a deliberate

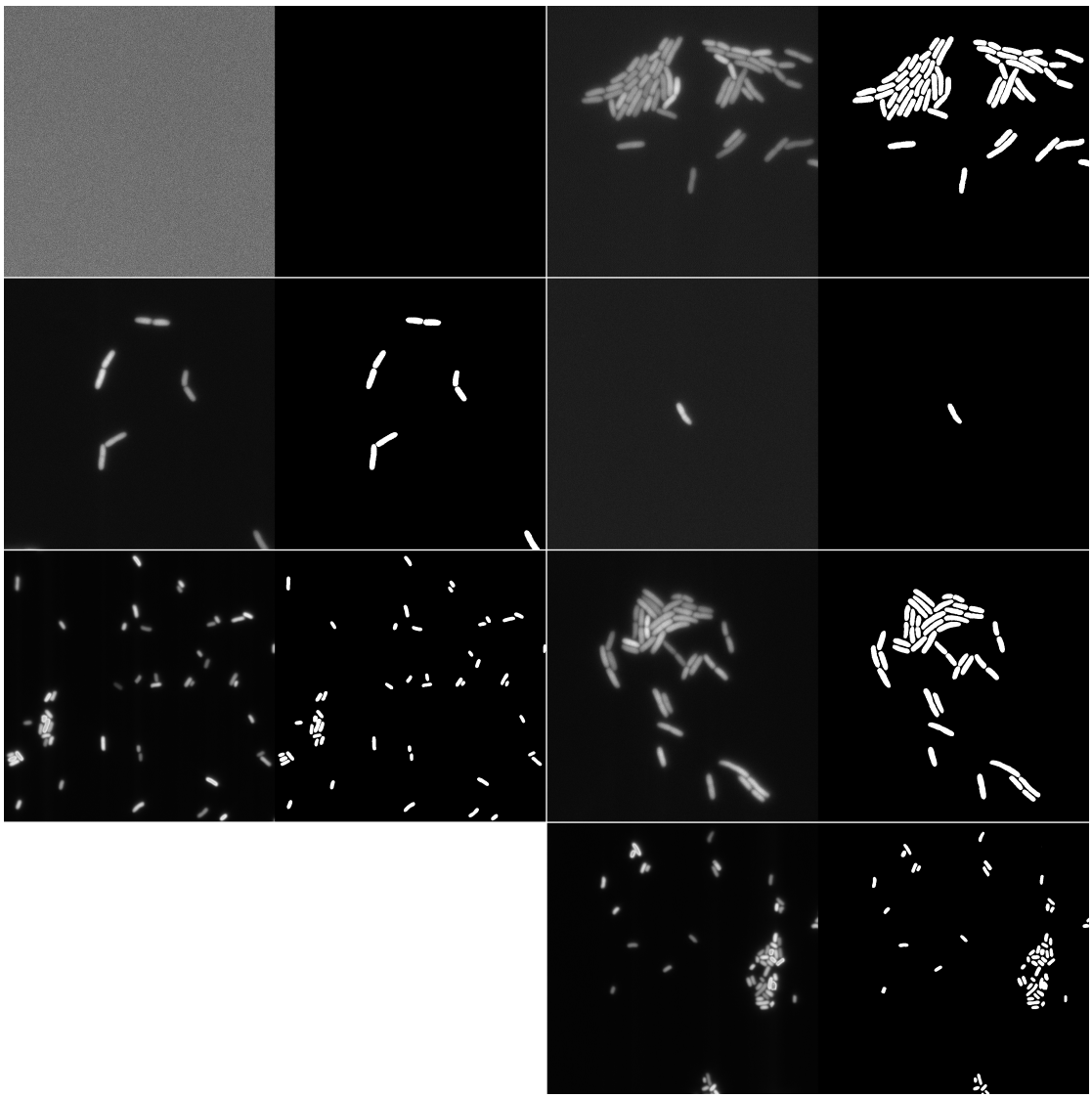

Figure 2.13: Training set for neural networks used in this investigation. The left image in each pair is the raw microscopy data and the right image is the manually augmented binary answer key. The top left image pair are of an empty field of view with no cells present. All other images contain cells.

choice was made to have all training images be of cells on agar pads, to ensure they remained in the focal plane. The reason for this has to do with the generation of answer keys. If some cells were at an angle to the focal plane in the training images then it raises the question of what should be the desired segmentation outcome for these cells. One could simply ignore them and attempt to teach the network to ignore cells at angles to the image plane, although this may result in conflicting

goals which ultimately hinder network learning. Alternatively one might try to approximate a two dimensional projection of the cell and segment based on that projection. Another option would be to try to segment all parts of the cell which are above a certain brightness but then the question becomes how to choose the correct threshold. Rather than trying to answer these questions definitively, and come up with a means of generating answer keys for these out of plane cells, it was decided to instead only train the network on flat cells. We found that in doing so, both networks were still able to learn to segment the in focus part of cells that were at an angle to the image plane. The networks also correctly ignored (in most cases) the faint blurs of cells, or parts of cells which were out of the focal plane. Given this, and given that cells were being manually selected, and the segmentation manually vetted, I decided this was an acceptable measure to reduce the time required to generate answer keys.

2.4.2.6 Network training and comparison

With the model architectures and training images set we can now turn to examine the performance of the networks during training. It is at this stage that we can begin to see the different strengths and weaknesses of the two approaches. Training was performed on a windows PC using an Intel i5-6600k quad core processor running at 3.7GHz, and a factory clocked Nividia GeForce GTX 1070 graphics card. After data augmentation the data set consisted of 168 images. In the case of the U-net like architecture that corresponded to 151 training samples, and 17 validation samples. However this was not the case for the sliding window network. As this network treats each pixel individually this resulted in approximately thirteen million data samples for that network, and approximatley one million validation samples.

One of the first parameters which distinguishes the two networks, is that of training time. Training time can be expressed in terms of a number of epochs, or in the actual time taken to complete the training. I found that due to the much larger number of training samples, the sliding window network took far less epochs to train. However, in terms of minutes of training time, the sliding network was much slower. It took around 15 minutes to train 1 epoch in the sliding window network compared to approximately 45 seconds for the U-net like architecture. This is not a surprising outcome. As the U-net training set contains less than 200 images, the entire network and training set can be loaded into GPU memory at once. This allowed practically

all of the computations to be handled on the GPU, vastly increasing performance. On the other hand, the sliding window network requires that a small image be processed for each pixel in each image. This vastly increases the memory requirements, and as a result, requires that images be constantly transferred between the RAM and GPU memory. This process is itself quite slow, which leads to reduced performance.

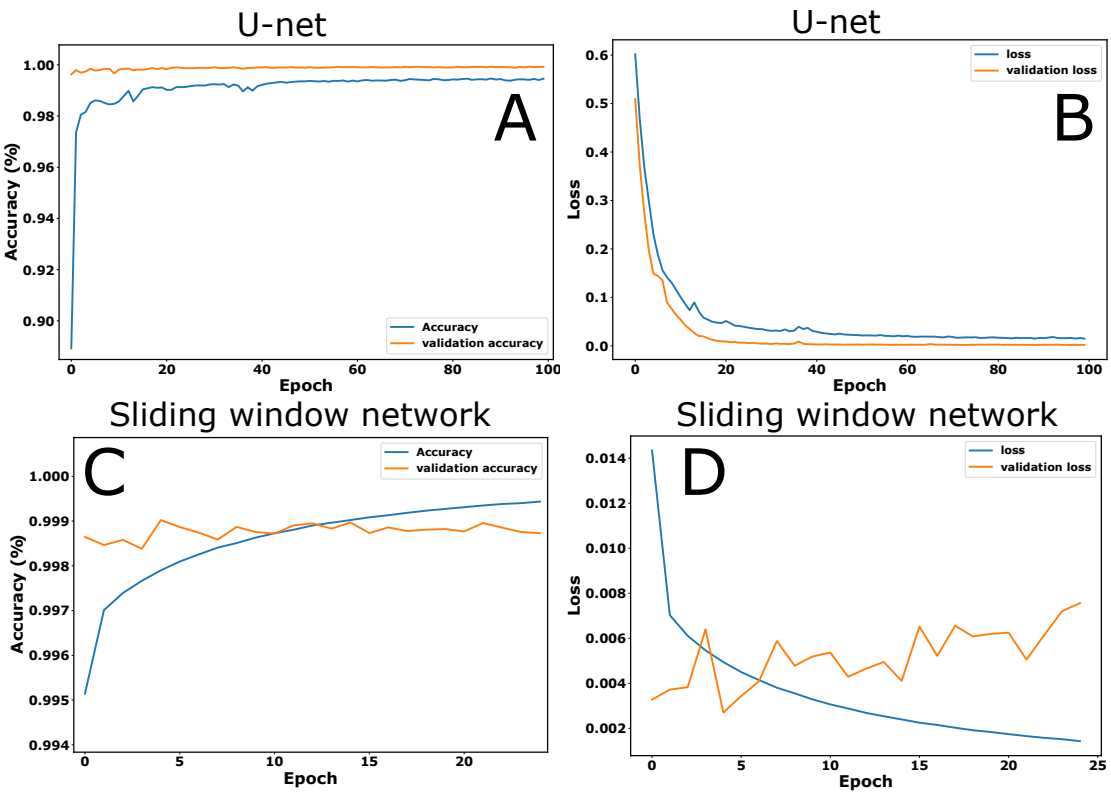

Figure 2.14: Training loss and accuracy for the two networks developed in this investigation. **A & C:** Training accuracy (blue) and validation accuracy (orange) for the U-net and sliding window networks. **B & D:** Training loss (blue) and validation loss (orange) for the U-net and sliding window networks.

To compare differences in how effectively the two networks were trained, one can examine how the loss and accuracy of the network changes with each successive epoch. The accuracy is defined as the percentage of pixels which have been assigned the correct class. To compute the accuracy, each pixel in the network output is rounded to either 0 or 1, and then compared to the answer key to determine what percentage of pixels were assigned the correct class. The loss on the other hand, is simply the mean value of the error function computed for each image.

In figure 2.14 I plot the accuracy and loss for both networks, against the number of epochs of training. The blue line in each plot indicates the loss or accuracy as computed using 90% of the training data. The remaining 10% of the total data was kept aside as a validation set, and was not used to train the network. Instead, after each epoch, the validation set was put through the network and the accuracy and loss computed. The orange line in each figure shows the validation loss or accuracy, as computed at the end of each epoch.

From figures 2.14:A and 2.14:B, we can see that for the U-net model, the accuracy follows a characteristic logarithmic increase with epoch, while the loss shows an exponential decrease. The high value of accuracy obtained by the model ($>$ 99%) suggests that the network was able to effectively learn to segment the images. The graphs also indicate that better results are obtained on the validation set, as compared to the training set. This is unusual as one would normally expect worse performance on validation images, which have not been used to train the network. There a number of possible reasons for this. Firstly, only 10% of the data is used in the validation set. As this 10% is chosen at random it may be that there are more difficult examples in the training set compared to the validation set. Secondly, the drop out layers present in the model are active only during training. During prediction of the validation data, drop out is not active, and as a result, one might expect a better result than during training. Given that both the validation and training loss decrease in a similar manner with epoch, I am confident that the model has still been trained well, despite the odd relationship between training and validation loss.

In figures 2.14:C and 2.14:D I plot the accuracy and loss for the sliding window network. As can be seen, the training loss and accuracy both follow the characteristic logarithmic increase, and exponential decrease respectively. However, the validation accuracy in this case appears flat, while the validation loss appears to actually increase in time. This suggests that the network might be over fitting to the training set. This would lead to better results for the training images, but worse results for images which the network has not been trained on. It is for this reason that the network was not trained for as many epochs as for the U-net model. However, given that the final validation loss for the sliding network was still lower than the validation

loss for U-net, the results obtained from this model were still quite good overall.

To get a more intuitive idea of how well the two networks performed, I segmented three images which were not used at all during the training process. These images are shown in figure 2.15. Unlike the training data, these images were not taken of cells on agar pads. Instead the first two were taken of cells in a tunnel slide, and as a result not all the cells were aligned with the image plane. The third image was taken using a microfluidic mother machine, and as a result, all the cells in this image were aligned vertically. As figure 2.15 shows, both networks preform quite well. Both networks appear to have difficulty separating cells which have just divided, but this is something which is very difficult to do by eye, and ultimately a subjective choice must be made about what constitutes "divided". Otherwise, the networks perform quite similarly, and given that segmentation is manually vetted in this work, I would have been satisfied using either network for the task, at least as far as segmentation accuracy was concerned.

While the segmentation accuracy of the two networks was very similar, where they were not similar, was in terms of prediction speed. To segment a single image took over two minutes for the sliding window network, while the U-net network took only around 0.1 seconds. It was for this reason that most of the segmentation performed in this work was done using the U-net architecture. This does not mean that there is no place for the sliding window network however. Due to the pixel wise way it segments image this network is suited for data sets where image size is not constant between images. The U-net model on the other hand, once trained, can only segment images of a fixed size. However, given that all of my images in this work were taken using the same camera, this was not an issue in my case.

2.4.2.7 Cell morphology extraction from segmented images

With a means of selecting in focus cells from microscopy images, and the ability to satisfactorily segment the result, we can now turn to the problem of extracting the length,width and volume from segmented cells. To do this I used a custom algorithm to first locate the cell poles and then fit a polynomial function to the cell backbone. My approach was similar in style to that found in Guberman et al. 2008. Figure 2.16

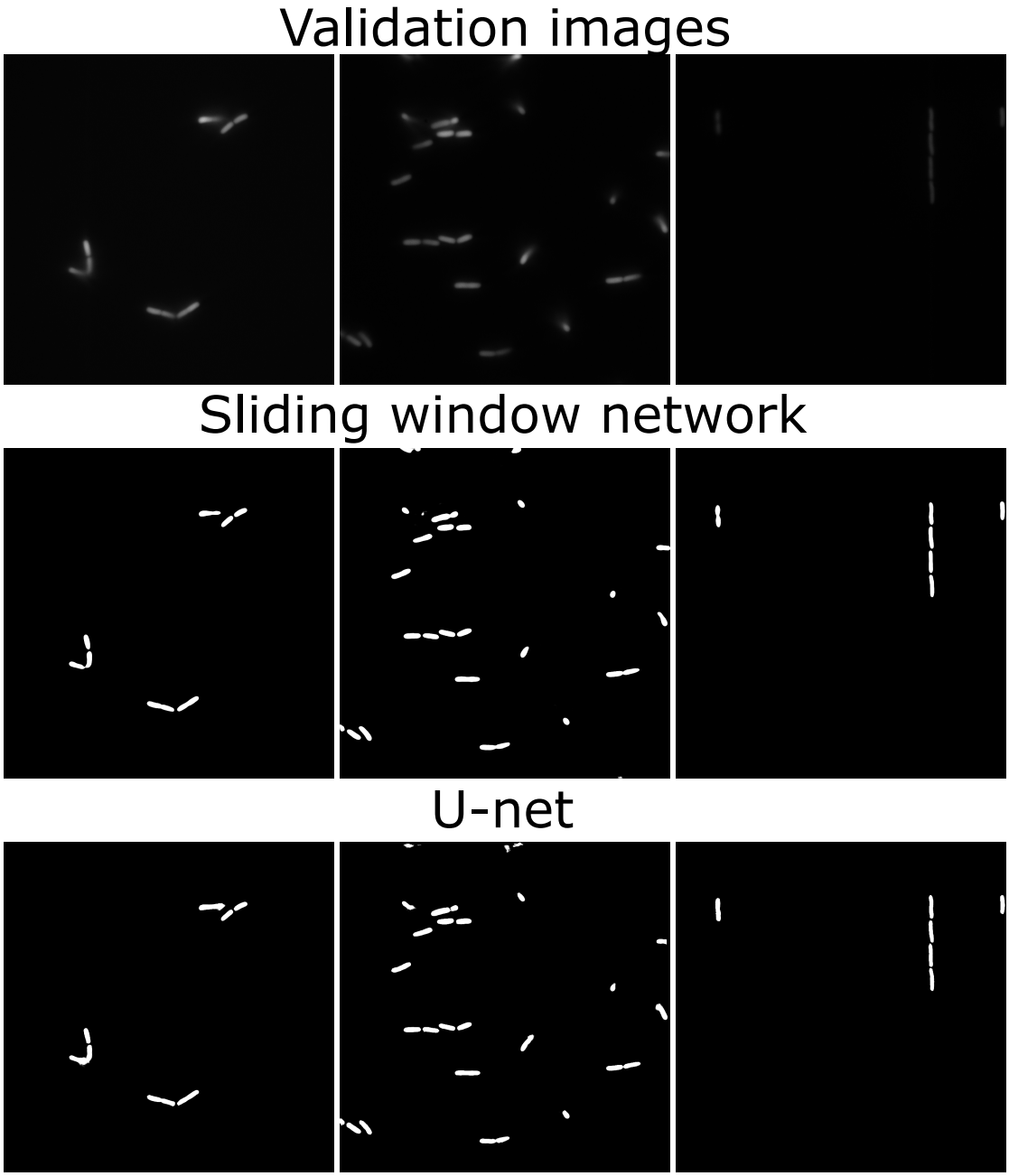

Figure 2.15: Validation images used to examine segmentation performance of the two networks developed in this work. The top row shows the original images. The middle row shows segmentation as performed by the sliding window network. The bottom row shows segmentation as performed by the U-net model.

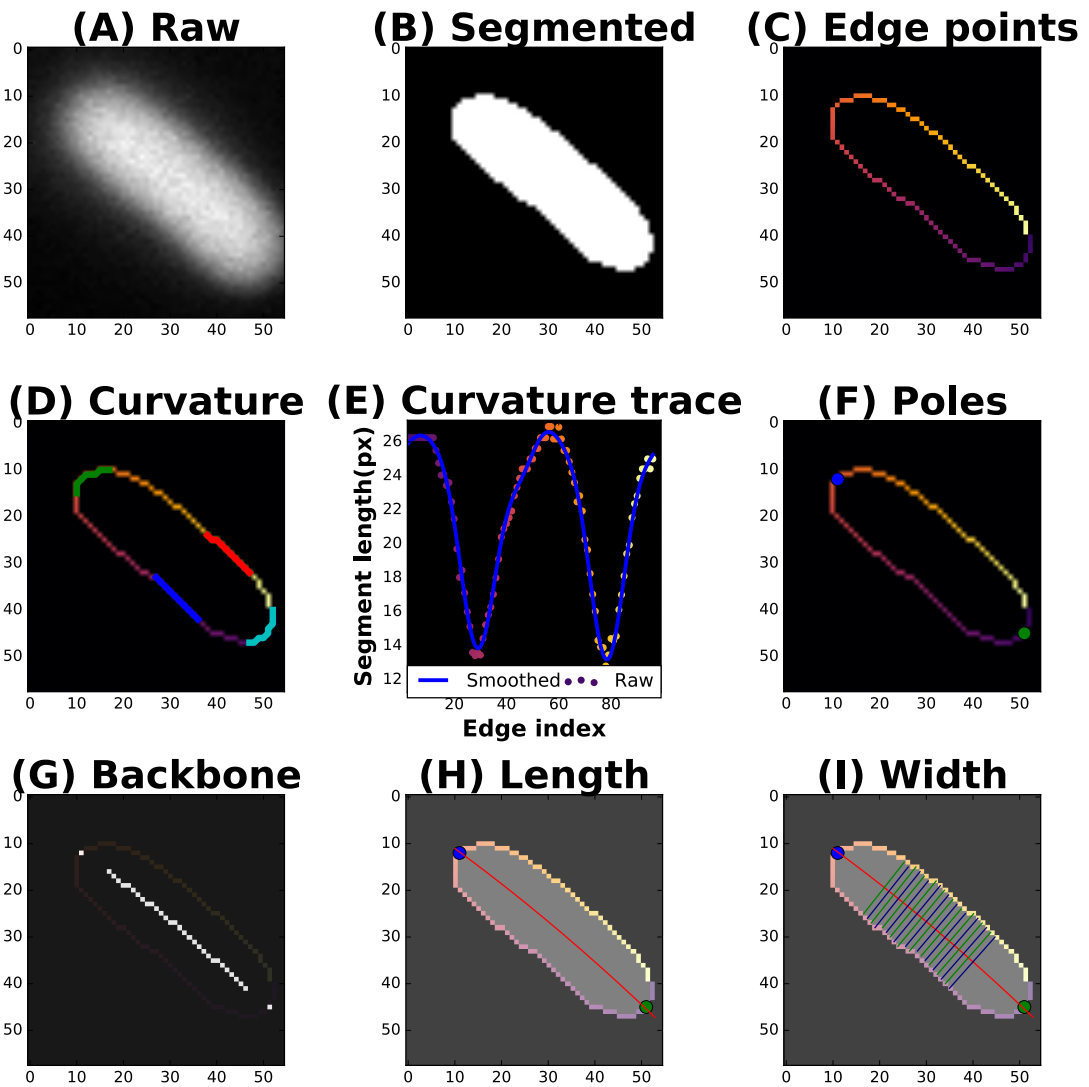

Figure 2.16: Various stages of the algorithm used to extract cell lengths and widths. See text for details on each lettered step in the process.

shows the results for each stage in the algorithm, as performed on an example cell.

First, each manually chosen cell is segmented by the neural network (2.16:B) and the edge pixels selected. These edge pixels are then labelled in order going either clockwise or anti clockwise around the cell (2.16:C). Next, the poles are identified by locating the region along the edge of the cell with the maximum curvature. To do this, each pixel along the edge of the cell is examined in turn. For each one I analyse

a sequence of n[3] consecutive pixels with the currently examined pixel falling in the middle of this sequence. I then measure the straight line distance between pixels at either extreme of this sequence and use this as a proxy for the local curvature. Figure 2.16:D illustrates this process by showing four example coloured line segments, of equal length, placed along the edge of the cell. As can be see, when the line segment is in between the poles the line is straight and the distance between its ends is maximal. Conversely, when the line segment reaches the pole it curves causing the distance between its ends to drop. If I plot the end to end length for each consecutive pixel along the edge as shown in figure 2.16:E we see that the plot has two clear dips which correspond to the poles. By smoothing the point to point trace by removing high frequency components I can recover the blue line in figure 2.16:E. From this curve I can find the minimum and plot points on the corresponding pixels in figure 2.16:F. As shown, the position of the poles is correctly identified. Next a back bone is found by skeletonising the full segmented image. This was done using the skeletonise function from the Skimage python package, which makes successive passes on the image removing edge pixels on the condition that their removal does not break continuity. In figure 2.16:G I show the backbone and located poles in white, on top of the ordered edge pixels. To find the length of the cell I fit a second order polynomial through the backbone and poles, and truncate it using the boundaries of the segmented cell. I then compute the length of the remaining line and take this as the final length of the cell (see the red line in figure 2.16:D). To compute the width I examine the midpoint of the length line, and take a line perpendicular to the tangent of the length line at that point. I then compute the width by determining the length of that line between the points where it intersect the cell boundary. This is done for 10 additional points either side of the midpoint, and the results averaged to yield the width (see blue and green lines in figure 2.16:I). Figure 2.17 shows the length and widths determined from this algorithm for a number of representative cells. As we can see in most cases the algorithm places the poles very close to their actual position. However in some cases, due the limited pixel resolution, the identified poles are slightly to one side of the actual position. This tends to cause the length line to curve more than it otherwise would. However, when I examined these cases more closely I found that the algorithm generated length that was within a few percent of a manually placed line. For example in a sample of 30 cells taken

[3] where n is an odd integer

from two different conditions, I found that on average there was a 1.3% difference between the length as computed by the algorithm, and as determined by a manually drawn line. Additionally of these 30 cells I found 5 where the poles had been placed at a position where I would not have manually placed them. I therefore concluded that the algorithm was "good enough" to draw meaningful conclusions about cell size. Finally, to compute the volume of each cell I assumed a spherocylinder shape and used the following formula, where l is the length of the cell and w is the width ad V is the volume (Buda et al. 2016):

$$V = \frac{1}{6}\pi w^3 + \pi\frac{w^2}{4}(l - w). \tag{2.2}$$

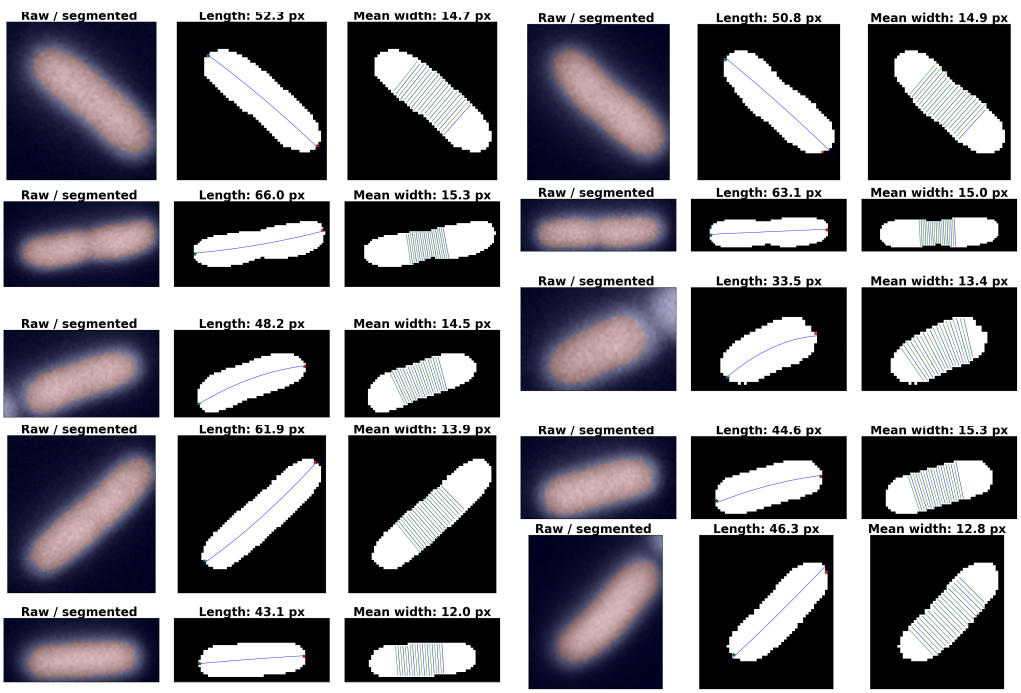

Figure 2.17: Example of cell length and width extraction. Each panel shows the original image overlaid with the segmentation mask in red. The computed length and width lines are also plotted, and the figure tittles contain the measured length and width values in pixels. Finally the positions of the poles are also plotted as red and green stars.

2.5 Cell size measurement using the Coulter counter

2.5.1 Principle of operation

In section 2.4 I outlined how cell size can be determined using fluorescent microscopy. The advantage of this technique is that it allows for a detailed visual examination of cells, and therefore, extraction of not only their volume, but also their length and width. The method also allows for as little perturbation of the cells from their initial steady state growth as possible. However, because of the need to manually select those cells which lie directly in the focal plane, this technique also suffers from relatively limited throughput, being able to analyse less than a thousand cells per experiment, and even then only after considerable man-hours of analysis. In order to provide a confirmation of the results we achieved through fluorescent microscopy, and to provide a higher resolution measurement of cell size distribution, we employed the use of a Multisizer 4e Coulter Cell Analyzer (Beckman Coulter).

The Coulter counter was first developed in 1958 as a high throughput method for measuring cell size (Kubitschek 1958). Since then it has seen widespread use, both for its ability to accurately measure cell volumes, and as a means to count the number of cells in a liquid sample (Trueba et al. 1982; Basan et al. 2015). In principle, the operation of a Coulter counter is quite simple (see figure 2.18:A). The main component is a glass aperture tube which is a continuous vessel, except for a single small aperture near the base of the tube. The tube is filled with a particular electrolytic solution, and then submerged in a suspension of electrolyte and cells. Electrodes are then immersed in both the internal and external reservoirs, and a current run between them. As the glass tube is highly insulating, the only way for current to flow is through the aperture itself. This means that the resistance between the electrodes is inversely correlated with the volume of the aperture, with larger aperture dimensions causing lower levels of resistance. In order to measure the volume of cells in the suspension, liquid is drawn from the external reservoir through the aperture. When a cell enters the aperture, it displaces its own volume in electrolyte. This displaced electrolyte is no longer available to allow current to flow through the aperture, so the resistance between the electrodes in increased. As most modern multisizer systems work by maintaining constant current between the electrodes, this increase in aperture resistance, due to the presence of a cell, is actually measured

by an increase in voltage across the electrodes. As shown in figure 2.18:B there is a small volume (typically two and a half times the aperture volumes) around the aperture where the presence of a cell causes a detectable increase in voltage. The presence of this "sensing zone", as it is often called, means that as a particle transits the aperture, it is not detected at a single point in space and time, but rather continuously along its trajectory inside the sensing zone. This means that each particle gives rise not to an instantaneous spike in voltage put to a broad pulse as shown in figure 2.18:C, where the height of the pulse corresponds to the particles total volume.

The range of cell volumes which the multisizer can measure is dependent upon two factors. One is the aperture size itself, and the other is the strength of electrolyte used. Typically the multisizer voltage response is linear with cell volumes only for particles which have an equivalent spherical diameter between 2% and 60% of the aperture diameter (Berge 1990; Anderson and Quinn 1971; Coulter 2010). This means that for example a $20\mu m$ diameter aperture will only be able to accurately size particles with equivalent spherical diameters between $0.4\mu m$ and $12\mu m$. On top of this size restriction, the strength of electrolyte used can also alter the minimal detectable cell size. This effect is caused by an increase in electrical noise which occurs when higher voltages have to be utilised to pass detectable currents through very poor conducting electrolytes. Exactly what constitutes a sufficiently conductive electrolyte is dependent on a number of factors including aperture diameter, size of cells to be measured, and noise requirements, but typically for a $20\mu m$ aperture to have a noise level at or below the 2% limit a conductivity of around $15mScm^{-1}$ is required (Coulter 2010). Given a particular aperture and electrolyte combination, a noise threshold can be set by measuring the voltage response when a fresh, well filtered electrolyte is passed through the aperture. When an actual sample is being run, any fluctuation in voltage which falls below this threshold can be ignored.

In order to convert the raw voltage signals into cell size distributions the multisizer first employs an analogue pre-amplifier followed by an analogue pulse processor to distinguish pulses from noise and register each particle as an individual event. These pulses are then digitised, and reduced to meta data such as height, width and area by the digital pulse analyser, before being sent to the controlling computer for storage in a data file. Using modern electronics, the multisizer can process and anal-

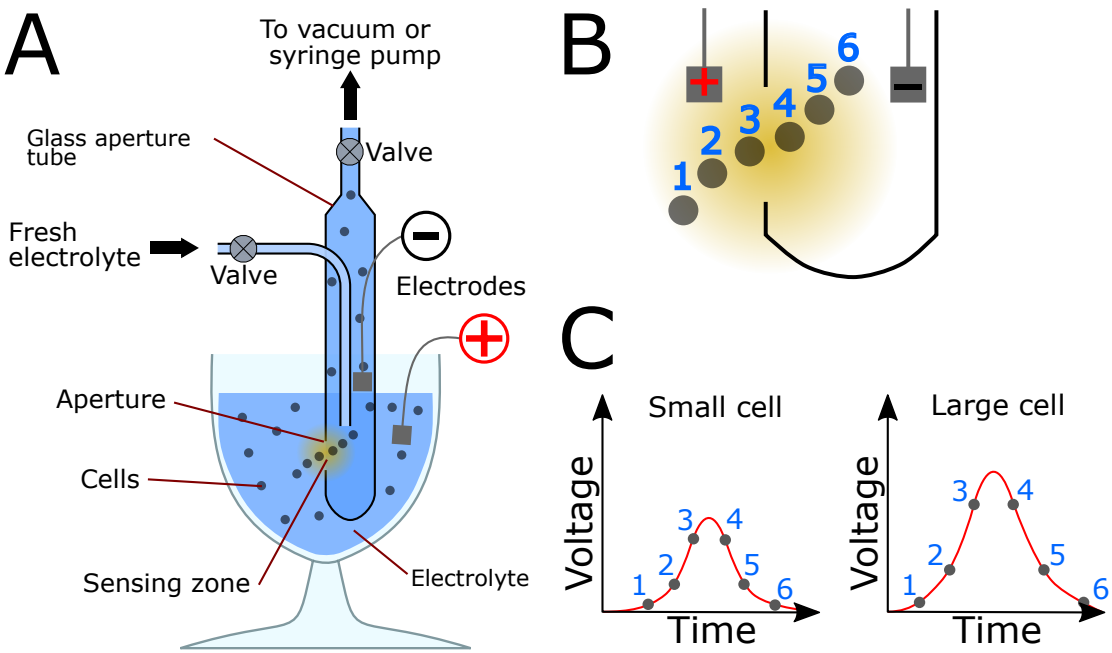

Figure 2.18: Coulter counter principle of operation see text for full details. **A**: Schematic of Coulter counter apparatus showing aperture tube and connections to external electrolyte supply and vacuum/syringe pumps. **B**: Zoomed in schematic of a particle transiting through the aperture. Six points in the particles trajectory are shown and the sensing zone is highlighted in yellow. **C**: Traces representing the change in voltage across the electrode as a large and small particle transits the sensing zone. The height of the traces correspond to the particles total volume and the numbered points corresponds to the trajectory shown in sub-figure B.

yse up to ten thousand particles each second. Despite this, in practice actual event rates tend to be considerably lower. The reason for this is the coincident arrival of particles in the sensing zone[4]. In very dilute samples where there is considerable time between cells entering the aperture each individual event can be easily distinguished and separated. However as cell density increases and the time between cells entering the aperture shrinks, voltage pulses begin to overlap and it becomes increasingly difficult to distinguish two separate events (figure 2.19:A). When these coincidence events occur it results in erroneously large volume values leading to altered cell size

[4]Technically, what is described here is primary coincidence. Secondary coincidence can also occur when two particles which are individually too small to be detected enter the aperture at the same time and are thus counted as a single larger particle. However in this work, where cells are typically much larger than the minimum detectable size this is not as relevant.

distributions. To reduce the effect of coincidence the multisizer software is able to utilise pule meta data to correct volume distributions (Wales and Wilson 1961). However this correction is limited and insufficient at coincidence rates beyond 10%. To guarantee therefore that distributions remain largely unaffected by coincidence, it is necessary to ensure that samples remain sufficiently dilute, and that as a result, actual throughput is considerably less than ten thousand events per second. This leads to a trade-off that the experimenter must make, between high throughput measurements which can be done quickly at the cost of higher levels of coincidence versus, slower, more time consuming measurements which are more accurate.

A second source of error, that leads to deviations from the ground truth cell size distributions, has to do with the trajectories that cells take through the aperture. While the vacuum and metering pumps are designed to ensure a steady rate of flow through the aperture, it does not guarantee that cells follow the exact same path each time. For example, cells which travel through the aperture closer to the walls, where the fluid velocity is lower, will spend more time in the aperture giving rise to pulses which are wider and higher than those which travel closer to the centre of the aperture (see figure 2.19:B). In samples with a narrow cell size distribution (narrow here meaning less than a factor of two difference in diameter, or factor of eight difference in volume) where this effect is particularly noticeable, the multisizers "pulse edit" mode can be utilised to reduce skewing of the results (Göransson 1990). In pulse edit mode, each pulse is analysed for the characteristic skew typical of particles which did not travel directly through the middle of the aperture. These identified pulses are then removed from the cell size distribution data, but not from measurements of total numbers of cell counts. Approximately 50% of events are removed in this way when using pulse edit mode.

Finally, the use of the multisizer in acting as a cell counter should be mentioned. Leaving aside the problem of coincidence discussed above, the multisizer is able to detect and count each individual cell which enters the aperture. Therefore, provided the amount of volume which passes through the aperture can be measured, an accurate number of cells per millilitre of culture can be determined. In order to measure the volume of liquid drawn through the aperture, the multisizer utilises a motor driven syringe pump, which can precisely measure the position of the plunger,

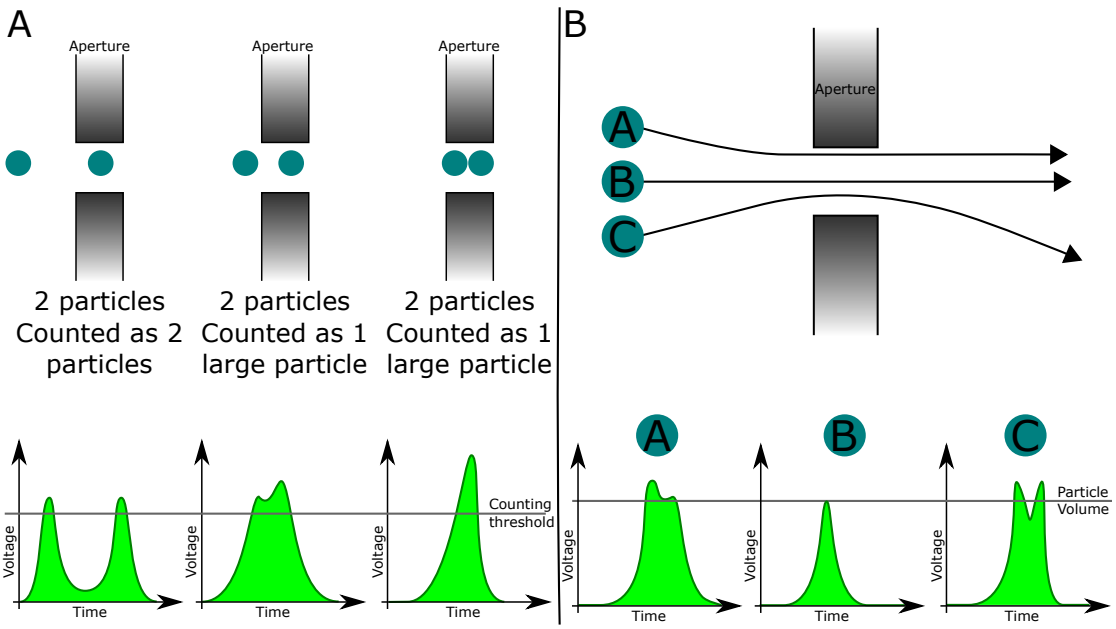

Figure 2.19: Coincidence events and pulse edit mode in the Coulter counter. **A**: Three scenarios showing how the distance between particles as they enter the aperture can cause two particles to be counted as one. The horizontal line in the voltage traces shows the level above which a signal is considered to be a discrete event. **B**: Schematic showing how three particles with different trajectories can lead to different measurements of volume. The horizontal line represents the correct voltage peak height when particles travel directly through the center of the aperture.

and therefore the volume of liquid drawn. However, as the syringe itself has a relatively small volume, of only a few hundred micro litres, it is not possible to both accurately measure the volume of liquid drawn, and collect large number of cells in order to gather more robust cell size distributions. For this reason the multisizer is equipped with a vacuum pump, which can draw limitless quantities of liquid through the aperture, but at the cost of being unable to measure the volume drawn. In this investigation we were primarily concerned with determining accurate cell size distributions rather than precise estimates of cell number, therefore we chose to operate the cell sizer in the vacuum pump mode, in almost all cases.

2.5.2 Experimental protocol

As mentioned in the previous section, the range of cells which the multisizer can measure is dependent on the aperture size used. In this investigation we opted to use the $30\mu m$ diameter aperture, as we felt this represented a good balance between smaller apertures, with smaller minimum detectable sizes, and larger apertures which are far less prone to blockages and clogging. In almost all cases, measurements of volume were performed using 400 linearly spaced bins between volume values of $0.11\mu m^3$ and $10\mu m^3$. Pulse edit mode was also employed to ensure the highest accuracy possible.

At the start of each round of experiments, the multisizer was flushed and then filled with the desired media, which had first been passed through a $0.22um$ filter to remove any small particulates. The aperture tube was then submerged in approximately $10ml$ of this media, and the noise threshold value automatically determined by the software. At this point note of the automatically determined value was taken, to ensure it was below 2%, thus indicating that our media had sufficient conductivity that it was not increasing noise levels. For all media trailed in this work this was found to be the case.

To get an estimate of how particle free the media and apparatus was, before any experimental samples were run, a blank was performed using the fresh filtered media. To do this $50\mu l$ was drawn through the aperture, using the multisizers syringe pump driven volumetric mode. This was repeated four times after each change of media. Typically when the system was well cleaned, we counted less than 100 particles per $50\mu l$ of media.

In order to harvest cells for measurement of volume distributions in the multisizer, cultures were grown as stated in section 2.2.3. At various stages of growth, $1ml$ of culture was removed and an OD_{600} measurement taken. Immediately following this measurement a small volume of culture was diluted into $10ml$ of fresh, filtered, room temperature media and a measurement of the cell size distribution taken. The severity of dilution depended on the measured OD_{600} of the sample, but typically ranged from between 2-100μl of sample per $10ml$ of media. The goal of this dilution step was to strike a balance between the low coincidence rates of a very dilute sample, and the fast measurement times of a dense sample. In practice the percentage

concentration as measured by the multizier lay in the 0.1-3% range, with coincidence values falling well below 10%. Each measurement of cell size involved four runs, where in each run liquid was drawn through the aperture using the vacuum pump, until 25,000 events had been registered. By splitting the measurement into four separate runs it allowed us to check if there were any changes in the size distribution with time. Where a change was found with time (for example late in the growth curve where late exponential phase cells were being heavily diluted into fresh media) only the first run was taken. We estimate that it took approximately 2 minutes between removing the sample from flask growth to diluting the sample, and another 2-5 minutes between diluting the sample and completion of the first run. After each run the aperture tube was flushed with fresh media and the outside of the aperture tube was rinsed thoroughly with deionised water. When tested, we found no change to background counts of blank media made directly after measurements containing cells. This indicates that our rinsing protocol was sufficient in ensuring that very few cells remained on the aperture tube or electrodes after rinsing, to contaminate subsequent samples.

2.6 DNA content from flow cytometry

2.6.1 Principle of operation

Though significantly more difficult to access than cell size, the DNA content of cells remains a key physiological parameter, closely tied to both growth rate, and the cell cycle. It was not long after the development of the flow cytometer in the late 60s and early 70s that it was applied as a high throughput method for determining the DNA content of individual bacterial cells (J. E. Bailey et al. 1977). Although other methods have been used for the task, such as fluorescent microscopy (Si et al. 2017), it remains in wide use predominately for its maintained superior throughput (Hill et al. 2012; Volkmer and Heinemann 2011b).

The main concept in a flow cytometer is to bring particles or cells one by one to a single point in space, known as the interrogation point, where a number of measurements can be made. To achieve this, a technique known as hydrodynamic focusing is employed. As shown in figure 2.20:A, hydrodynamic focusing involves flowing the sample fluid through a nozzle, such that upon exiting it is surrounded

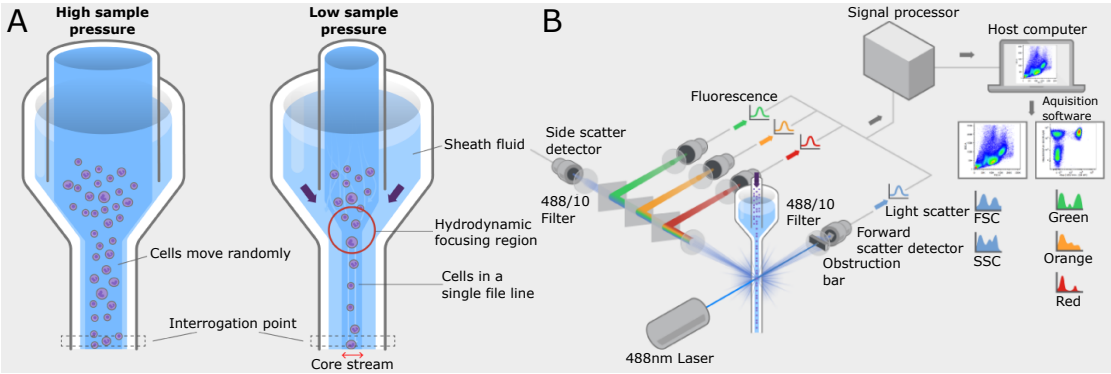

Figure 2.20: Hydrodynamic focusing and flow cytometry. **A**: Schematic showing how the fluid stream is hydrodynamically focused by the surrounding sheath fluid at different sample fluid pressures. **B**: Schematic of a typical flow cytometer showing: flow cell, forward and side scatter detectors and example output traces. Image taken and modified from: Scientific 2019

by another liquid, called the sheath fluid. The sheath and sample fluids are kept at different pressures, such that the fluid velocity of the two is different. This helps to ensure laminar flow, and the existence of an interface between the two liquids which inhibits mixing. The relative area of the channel which is taken up by flow or sheath fluid can be altered by adjusting the pressure of the sample fluid. At high pressures, the sample fluid takes up more of the channel, and the flow of cells is not well ordered. However, as the pressure of the sample fluid is reduced, it contracts and takes up significantly less area in the channel. At sufficiently low pressures the sample stream diameter is comparable to the diameter of cells, and as a result the cells flow in essentially single file down the channel towards the interrogation point.

Once at the interrogation point cells are illuminated with one or more lasers of different wavelengths. Typically a wavelength of $488nm$ or $405nm$ is used to measure both the amount of light scattered at small angles from the incident beam (known as forward scatter) and at angles around 90° from the incident beam (side scatter). Alongside scattering, fluorescent information is also collected by exciting the sample with additional lasers, and then collecting the emitted light using dichroic mirrors, and narrow band bass filters as shown in figure 2.20:B. To actually convert the emitted or scattered photons into an electrical signal, two kinds of detectors are commonly used. For the forward scatter channel typically a photodiode (PD) is used as the signal tends to be quite strong, with most of the light being scattered at small

angles. However, for the much weaker side scatter and fluorescent signals, photon multiplier tubes (PMTs) are used for their far greater sensitivity (Adan et al. 2017). The analogue signal generated by the PMTs or PD is converted to a digital signal by the flowcytometer's on board electronics, in a very similar manner to that of the multisizer as described in section 2.5.1. First, as the cell traverses the interrogation point it gives rise to a voltage pulse with the maximum value obtained when the cell is located at the centre of the interrogation point. This voltage pulse is then amplified by either a linear, or logarithmic amplifier depending on the desired application. For situations where very large dynamic range is required, logarithmic amplification is used, while for situations with smaller dynamic range and an emphasis on quantitative results (such as DNA content measurement) linear amplification is preferred. After amplification the pulse is digitised and its width, height and area calculated[5] and stored to a file on the host computer. The height and area channels are proxies for overall strength of the signal, while the width is useful for detecting doublet events where two cells enter the interogation point at the same time.

The number and kinds of fluorescent signals measured in a flow cytometry experiment depends on the exact nature of the investigation. However, in almost all cases, the forward and side scatter channels will be collected. The reason for this is that the forward scattered channel (FSC) is often interpreted as being indicative of overall cell size, while the side scatter channel (SSC) is thought to be related to the internal complexity or "granuallity" of cells (Ambriz-Aviña, Contreras-Garduño, and Pedraza-Reyes 2014). While these statements tend to somewhat oversimplify the underlying scattering theory, it has allowed flow cytometery operators to discriminate different kinds of cells using only their different FSC and SSC responses. For example, it is possible to differentiate yeast cells from *E. coli* cells, and between different types of mammalian cells, but not between many different bacterial strains (Veal et al. 2000; Boyum 1968). As a result, for the analysis of bacterial samples, the fluorescent channels are predominately used to extract information, particularly when trying to determine DNA content.

[5]The order of digitisation and extraction of the pulse parameters actually depends on the exact flow cytometer used. Older machines tend to extract the parameters using analogue pulse processing circuits and then digitise the results, while newer machines tend to first digitise the raw pulse by sampling the signal at high frequencies, before extracting the relevant information.

To measure DNA content using flow cytometry the DNA must first be labelled using an appropriate fluorophore, such as propidium iodide. Propidium iodide makes an ideal nuclear stain as it experiences a change in excitation wavelength and intensity when intercalated with DNA, meaning that molecules bound to DNA contribute the vast majority of the the fluorescent signal. (Krämer, Wiechert, and Kohlheyer 2016; Lopez-Amoros, Comas, and Vives-Rego 1995). However, as propidium iodide is only permeable to dead cells, the sample must be fixed before staining to allow the DNA to be properly labelled. After cells are correctly labelled, the fluorescence response of each cell can then be measured at the appropriate wavelength, and the corresponding fluorescent intensity used to compute the quantity of DNA per cell. In order to be able to quantify the relationship between fluorescent signal and DNA content, a standard is required. This standard is typically achieved by treating slow, exponentially growing cells with rifampicin and cephalexin during what are known as "run out" experiments (Skarstad, Boye, and Steen 1986). During run out experiments cephalexin acts to stop cells from dividing, while rifampicin inhibits the initiation of new rounds of replication (while allowing existing rounds to complete). The combined effect, is that if cells are allowed time to grow after antibiotic treatment, they will all end up with integer numbers of fully replicated chromosomes. In particular, if the culture was growing sufficiently slowly before antibiotic treatment, most cells will end up with either one or two chromosomes, allowing the fluorescent signal which correlates to a genome worth of DNA to be quantified.

2.6.2 Calibration

In order to utilise the fluorescent signal of cells stained with propidium iodide to measure DNA content, we must do two things. First, we must be able to measure the fluorescent signal of a sample with a known DNA content, and second we must know the relationship between sample brightness and measured fluorescent signal. As mentioned above, to achieve the first requirement, we use run out cells to act as a DNA content standard. These cells have integer number of chromosomes, so we can relate the measured signal to genome equivalents. To understand how the measured signal changes with actual intensity however, we make use of specialised 8-peak, fluorescently labelled, rainbow calibration particles (BD biosceinces). These rainbow particles get their name from the fact that they are coated with a mix of fluorophores that are excited at any wavelength across most of the visible spectrum,

between 365-650nm. The rainbow particles themselves are microspheres with well controlled diameters between 3-3.4μm. While the beads are monodisperse in size, they contain eight populations, each of which have a different, tightly controlled quantity of fluorophore. The quantity of fluorophore present in each population is measured by the manufacturers in a unit called molecules of equivalent soluble fluorescence or MESF (L. Wang et al. 2002). The MESF unit aims to be able to quantify the brightness of a fluorescent sample, by asking the question how many molecules of free fluorophore in solution would be required to generate the same intensity of emitted light. For example, if a microsphere has a stated MESF value of 500, that means that the particle emits same intensity of light as 500 free fluorophore molecules. Rather counter intuitively a MESF value of 500 does not mean, necessarily, that the particle contains exactly 500 fluorophore molecules. To see why this is we can examine the formula for total measured intensity as described in A. Schwartz et al. 2004:

$$i_f = [ge\Omega\epsilon\phi \int Q(\lambda)S(\lambda)T(\lambda)d\lambda]c. \tag{2.3}$$

Where: i_f is the measured fluorescence intensity; g is the photomultiplier tube gain; e is the elementary charge; Ω is the aperture and collection optics of the cytometer; ϵ is the molar extinction coefficient at the excitation wavelength; ϕ is the quantum yield of the fluorophore; $Q(\lambda)$ is the quantum efficiency of the PMT at wavelength λ; $S(\lambda)$ is the normalised emission spectra of the fluorophore at wavelength λ; $T(\lambda)$ is the filter characteristics at wavelength λ and finally c is the concentration or number of fluorophores. From this equation we can note a few things. Firstly, we can notice that the intensity (and therefore the MESF value) is dependent on the emission spectra of the fluorophore used. In general each fluorophore has a unique emission spectra and therefore it follows that the MESF value can only be specified for a single, specific fluorophore and not in general. Secondly as the intensity is dependent on both the quantum efficiency and the emission spectra, the MESF value has the potential to vary with the environmental conditions surrounding the fluorophore. For example pH is known to have an affect on the spectra of many fluorescent molecules, and so care must be taken that the pH is matched between the sample, and control run using the rainbow particles (A. Schwartz et al. 2004). As a consequence of this environmental dependence, it can often be the case that the same number of molecules of fluorophore will have different intensities when located inside

cells, compared to on the surface of a bead. For this reason many manufacturers of calibration particles go to great lengths to ensure that the florescent molecules they use are bound to the beads using linkers which emulate the spectra shift fluorophores experience when in typical physiological conditions inside cells (L. Wang et al. 2002).

As discussed above in order to quantify an unknown fluorescent signal in terms of MESF, we require that our sample and calibration particles be stained with the same fluorophore. Unfortunately, as calibration beads for flow cytometers tend to be more heavily aimed towards standard immunological stains more commonly used in the study of mammalian cells, I was not able to obtain beads stained with propidium iodide. Instead I opted for a fluorophore which had as similar an emission spectrum as I could find, namely Texas red-X. The spectra for both propidium iodide and Texas red-X are shown in figure 2.21:A. While relatively small, this discrepancy in emission spectra prevented us from determining the absolute MESF value of our samples. It did not however, stop us from quantifying the relative changes in fluorescence, which was sufficient to determine the relationship between intensity, and measured fluorescent signal. The reason for this is that, as shown in equation 2.3, the intensity remains proportional to the number of fluorophores, even if the filters or emission spectra are different. This means that while the quoted absolute MESF value for each of the eight bead populations may not be correct for propidium iodide, the relative difference in MESF values between them will still hold.

To calibrate the relative intensity of the rainbow beads using the flow cytometer, a sample was run and the fluorescent signal collected in the propidium iodide channel. This channel is denoted as YG 610/20 because the excitation is provided by a yellow/green (YG) laser, and the light is collected after passing through a band pass filter which only transmits light of a wavelength between $600nm$ and $620nm$ (610/20). The resulting 8 peak plot is shown in figure 2.21:B. Linear amplification was used for this run along with a gain of 1 and a PMT voltage of $451V$.

To get the fluorescent signal corresponding to each peak a Gaussian function was fit individually to each, and the mean and standard deviation extracted. The mean value was then plotted against the manufacturers quoted MESF values on a linear-linear and log-log plot as shown in figure 2.21:C. To these points both a linear

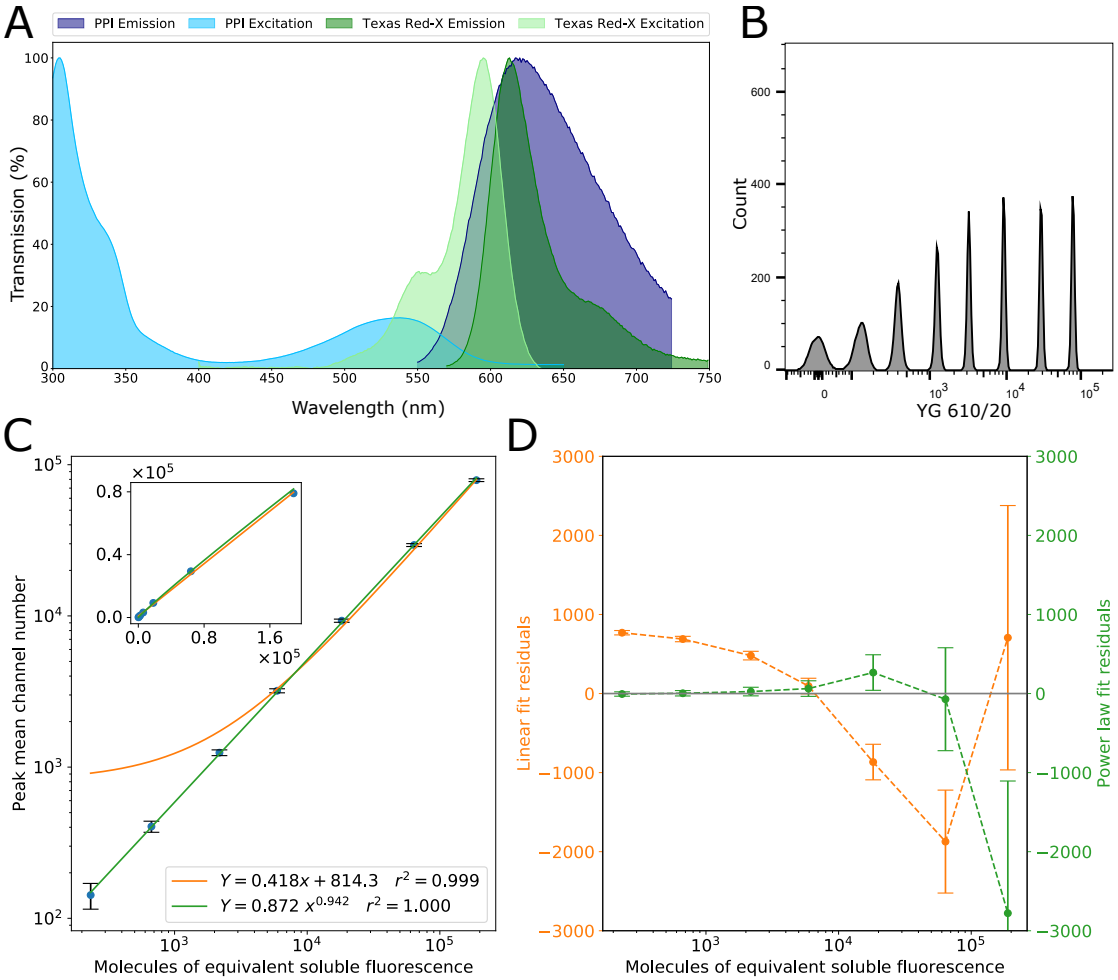

Figure 2.21: Eight peak calibration beads for DNA quantification. **A**: Excitation and emission spectra for propidium iodide, and its closest matching flurophore present on the rainbow calibration particles, Texas red-X. Data taken from the excellent spectra viewer resource provided by Chroma® at `https://www.chroma.com/spectra-viewer`. **B**: Plot of fluorescent intensity of eight peak rainbow calibration particles against frequency. Data and plot were both kindly provided by Anna Raper using the same Fortessa flow cytometer used throughout this work. **C**: Plot of mean peak position against manufacturers MESF value for each bead. The main figure is a log-log plot while the inset is the same data represented on linear axes. The data points in blue were fitted to both a straight line (yellow) and a power law function (green). Error bars are standard deviations of the peaks in subplot B. **D**: Residuals for both the linear and power law fits shown in subplot C.

(shown in yellow) and power law function of the form $y = Ax^B$ (shown in green) were fitted. The error bars are simply the standard deviation from the Gaussian fits of the fluorescent peaks. When plotted on a linear plot as is the case in the inset figure in 2.21:C the data, and both fits, seem to be approximately linear. However when plotted on a log-log plot and when the residuals are examined (figure 2.21:D) we can see that a power law function outperforms a liner one at low MESF values. This is also reflected in the higher r^2 value for the power law fit. At fluorescence channels above 2000 both models seem to fit the data equally well and the differences between them reduce as the fluorescent value increases. Therefore, throughout this work I used power law fits to calibrate each sample before analysis.

2.6.3 Experimental protocol

Each flow cytometry experiment began with the culturing of cells as described in section 2.2.3. At an OD_{600} between 0.1 and 0.3 part of the culture was removed for fixation while part was treated with antibiotic for the run out experiment. The details of these two processes are described below.

To fix cells from culture media the following protocol was followed. First $2ml$ of culture at the desired OD_{600} was removed and spun down in $2ml$ microcentrifuge tubes for two minutes at maximum RPM (18,000g). The supernatent was then removed and the pellet resuspended in 1ml of PBS solution that had been filtered using a $0.22um$ filter. This process was repeated two more times for a total of three washes into $1ml$ of PBS. Finally, for a fourth time the sample was spun down at maximum RPM, before the pellet was suspended, this time in only $100\mu l$ of PBS. To this, $900\mu l$ of ice cold 100% ethanol was added to fix the cells. After thoroughly, but gently mixing the contents of the tube by inverting several times, the sample was stored at $4°C$ until staining was performed. To prevent cell lysis this process was adjusted for each different condition, by adding the relevant amount of osmolyte to the PBS solution used to wash the cells. For example, if the experiment used cells grown in M63 with 600mM NaCl, then the PBS would have NaCl added to it, to obtain a concentration of 600mM.

For run out experiments cells were treated with cephalexin and rifampicin in the following way. First 5ml of culture was transferred to a prewarmed flask at $37°C$. To

this transferred media $15\mu l$ of $50mg/ml$ rifampicin and $50\mu l$ of $1mg/ml$ cephalexin solution was added. The sample was then allowed to incubate at $37°C$ for 3 hours to finish ongoing rounds of replication before being fixed as described above.

After fixation cells were stained with propidium iodide as follows. First a 5X concentrated staining solution was made which contained $50\mu g/ml$ propidium iodide and $500\mu g/ml$ RNase dissolved in well filtered PBS. The presence of RNase was important to allow the digestion of any RNA which propidium iodide also binds to. The 5X staining solution was also filter sterilised to remove any particulates that may have had similar FSC or SSC values to *E. coli*. To create a 1X staining solution of the correct concentration the 5X solution was diluted using well filtered PBS. The actual staining protocol itself began with the centrifugation of fixed cells for 3 minutes at maximum RPM . The supernatent was then removed, and the pellet left for 20 minutes in a sterile environment to allow evaporation of any residual ethanol. Following this, the pellet was washed twice by first resupending it in $1ml$ of well filtered PBS before centrifugation at maximum RPM. Finally, the pellet was spun down and resuspended in $1ml$ of 1X staining solution and left to incubate in the dark, at room temperature for one hour. After incubation, the sample was stored at $4°C$ until run in the flow cytometer (usually the next day).

The flow cytometer used was primarily the LSR Fortessa (BD biosciences). The Fortessa uses a $488nm$ laser for forward and side scatter collection and a $561nm$ laser for the excitation of the propidium iodide. The gain settings used were chosen to allow one to eight chromosomes to be displayed at resolvable separation, using as much of the dynamic range as possible of the linear amplifier. Samples were ran at slow speeds to minimise noise.

2.7 C period determination from QPCR

2.7.1 Principal of operation

Few techniques have been as revolutionary in the field of molecular biology as that of the polymerase chain reaction (PCR). For his work in developing this method, Kary B. Mullis was awarded part of the joint 1992 Nobel prize in chemistry (Mullis et al. 1986). A PCR reaction combines: deoxyribonucleotide triphosphate (dNTPS);

oligonucleotide primers; and a temperature stable polymerase derived from the thermophile bacterium *Thermus aquaticus* (Chien, Edgar, and Trela 1976), in order to copy and amplify a specific region of a DNA template. Each PCR reaction is broken into a number of repeated cycles with each cycle consisting of three steps: first a high temperature denaturing step which separates the double stranded target DNA; second a lower temperature annealing step which allows the primers to bind to the now single stranded DNA; and third an extension phase which allows the polymerase to amplify the targeted region of DNA. The severe temperature ranges traversed during a single PCR cycle, and the need to repeat that cycle approximately thirty times during a typical reaction, demanded that the polymerase used be stable at a wide range of temperatures, hence the ubiquitous use of *Thermus aquaticus* or TAQ derived polymerases. Typically during the early days of PCR, and still in many current applications, quantification of the amplified product is performed once at the end of the reaction, using techniques such as gel electrophoresis or spectrophotometry. However in the early nineties the first "real time" PCR reactions were performed which allowed for continuous measurement of the PCR product between each cycle of amplification (Higuchi et al. 1993). By measuring the amplification at each stage the amount of initial DNA in each sample could be determined, opening up the possibility of quantitative PCR or QPCR for short.

In detail, a QPCR reaction is carried out in a very similar manner to that of a standard PCR reaction, the main difference being simply the measurement of PCR product after each cycle of amplification. To determine the amount of product, two broad categories of QPCR reaction are used namely: non-specific fluorescent DNA dyes; and fluorescently labelled oligonucleotide probes (Kralik and Ricchi 2017). When using a probe based methodology a short oligonucleotide sequence is added to the standard PCR reaction. This probe binds specifically to the target DNA sequence at some location between the standard primer pairs and is labelled with both a fluorescent reporter molecule and corresponding quenching molecule. While both molecules are bound to the probe, any excitation of the reporter molecule is immediately quenched via Förster resonance energy transfer (FRET), and as a result no fluorescent signal is emitted. However, during the extension phase of the PCR reaction the TAQ polymerase encounters the probe and, as a result of its exonuclease activity, digests it allowing separation of the quencher and reporter molecules.

With the reporter now no longer within FRET range of the quencher, the fluorescent intensity of the sample increases, and consequently the amount of PCR product generated can be measured. A simpler alternative to these probe based methods are non-specific dye based reactions. In a dye based reaction a fluorescent dye is added, which is fluorescent only when intercalated with double stranded DNA. As the PCR reaction progresses, the number of double stranded product molecules increases, and as a result so to does the fluorescent intensity. The first dye used in this manner was propidium iodide but now much better, and safer, specially designed dyes are used, predominately SYBR green (Zipper et al. 2004). While much cheaper and simpler than probe based methods dye based methods are referred to as non specific because they bind to all double stranded DNA, not just that of the target sequence. This means that they are more sensitive than probe based methods to contaminate DNA and also to primer dimers, which may form in some reactions. However, despite these considerations I opted to use SYBR green based reactions for my investigation. This is predominaty a result of the much reduced cost of these reactions, and the fact that I was using previously published primer sets, where no primer interactions were anticipated.

To collect the fluorescent signal from a QPCR reaction, each reaction volume is excited using either lasers or LEDs and appropriate filters, the resulting fluorescence is then collected using either PMTs or photodiodes. Measurement of the fluorescent signal after each cycle leads to sigmoidal traces as shown in figure 2.22:A. During the first few cycles the quantity of DNA is below the detection threshold of the system, and as a result the traces appear flat. This flat region is often used as baseline to background subtract the signal for the whole curve, and as a result we refer to it as the baseline or ground phase. As the reaction progresses and the amount of PCR product continues to double after each cycle, we see a section of the trace which appears to be exponential in nature. This exponential phase is quickly followed by a levelling of the fluorescent signal known as the plateau phase. The cause for the plateau in fluorescence, seen towards the end of the reaction, is actually not very well understood, but recent work suggests it is caused by the degradation of the primer sequences due to repeated temperature cycling (Jansson and Hedman 2019).

Most QPCR applications are concerned primarily with the analysis of this ex-

ponential region that occurs early to mid reaction. In principle one might expect a perfect doubling of template, and consequently of fluorescent signal, after each cycle. This would imply that the fluorescence at cycle t could be modelled as:

$$f(t) = f_0 2^t, \tag{2.4}$$

where $f(t)$ is the fluorescence signal at cycle t, and f_0 is the original fluorescence at reaction start. However, in reality the amount of product is rarely exactly doubled after each reaction. The reason for this is normally to do with primer designs. Primers which form dimers, hairpins or which have significantly different melting points, can all contribute to lower the reaction efficiency bellow 100%. While steps can be taken to improve primer design, generally efficiencies even for well designed pairs, are not exactly 100% and values between 90-110% are generally accepted (Taylor et al. 2010). The gold standard for primer efficiency determination is the measurement of so called standard curves. Standard curves are created by taking a sample of a known quantity of DNA, and creating a dilution series to generate several samples with known concentrations. A QPCR reaction is then run for each sample, and a c_t value computed for each one. A c_t value is simply the number of cycles required for a sample to reach a given threshold level of fluorescent intensity. The exact threshold value is chosen somewhat arbitrarily, but usually lies within the exponential region of amplification. When the threshold value lies between two consecutive fluorescent measurements, linear interpolation is usually performed to extract the c_t value as a fractional number of cycles. In order to see why dilution series and c_t values allow us to extract primer efficiencies, consider the application of equation 2.4 to two reactions with some unknown efficiency, and different initial concentrations of DNA. Let $f_{0,1}$ and $f_{0,2}$ be the initial concentrations and we will consider $f(t)$ at the threshold value of fluorescence f_t such that $f(c_t) = f_t$. This yields the following expression:

$$f_{0,1}\alpha^{c_{t,1}} = f_t = f_{0,2}\alpha^{c_{t,2}}, \tag{2.5}$$

where α varies between 1 and 2 depending on the primer efficiency, with 1 indicating no amplification and 2 indicating a perfect doubling of DNA each cycle. We can take the left and right hand sides of equation 2.5 and rearrange them as follows:

$$\frac{f_{0,1}}{f_{0,2}} = \frac{\alpha^{c_{t,2}}}{\alpha^{c_{t,1}}}, \tag{2.6}$$

$$\frac{f_{0,1}}{f_{0,2}} = \alpha^{c_{t,2}-c_{t,1}}. \tag{2.7}$$

By substituting Δc_t for $c_{t,2} - c_{t,1}$ and taking the logarithm of both sides we yield the following expression:

$$log_{10}\left(\frac{f_{0,1}}{f_{0_2}}\right) = log_{10}\left(\alpha^{\Delta c_t}\right). \tag{2.8}$$

By applying rules of logarithm manipulation we can rearrange to yield:

$$\Delta c_t log_{10}(\alpha) = log_{10}(f_{0,1}) - log_{10}(f_{0,2}). \tag{2.9}$$

Rearranging for α we find:

$$log_{10}(\alpha) = \frac{log_{10}(f_{0,1}) - log_{10}(f_{0,2})}{\Delta c_t}, \tag{2.10}$$

$$\alpha = 10^{-\frac{log_{10}\left(f_{0,2}\right) - log_{10}\left(f_{0,1}\right)}{\Delta c_t}}. \tag{2.11}$$

Now at this point we can realise, that if we plot on the horizontal axis of a graph the logarithm of the initial DNA concentration, and on the vertical axis the corresponding c_t value, the gradient of the resulting line will be precisely the reciprocal of the exponent in equation 2.11. If we let the gradient of this line be m and remember that $1 \leq \alpha \leq 2$ we can finally write the expression for the primer efficiency expressed as a percentage:

$$E = 10^{-\frac{1}{m}} - 1. \tag{2.12}$$

With an understanding of how primer efficiency is calculated we can now turn our attention to the determination of the replication or C period. To do this we can follow a derivation outlined in Bremer and Churchward 1977. To start with we can remind ourselves, that in a steady state population experiencing exponential growth, the number of cells N as a function of time t follows the relationship:

$$N = N_0 2^{\frac{t}{\tau}}, \tag{2.13}$$

where N_0 is the initial number of cells at $t = 0$ and τ is the doubling time. Similarly, to maintain steady state conditions the number of origins of replication (I) and terminators (T) must also follow an exponential dependence:

$$I = I_0 2^{\frac{t}{\tau}}, \tag{2.14}$$

$$T = T_0 2^{\frac{t}{\tau}}. \tag{2.15}$$

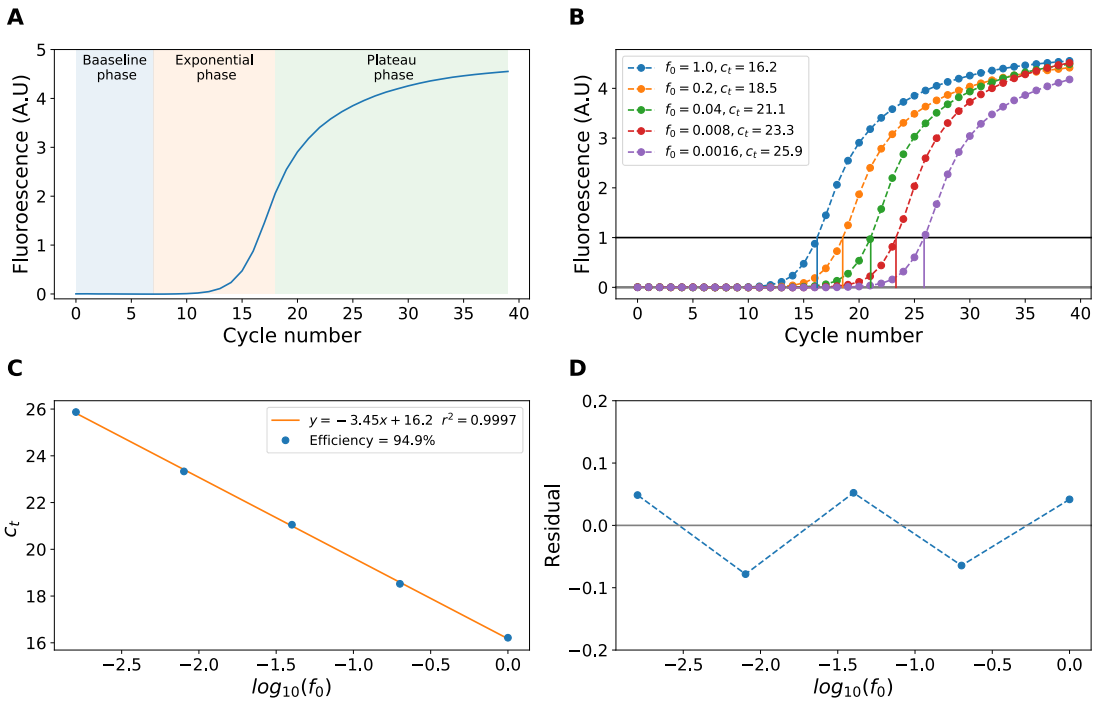

Figure 2.22: QPCR and primer efficiency. **A**: Plot showing an example trace from a QPCR reaction. The baseline, exponential and plateau phases are highlighted and the flouresccnece is expressed in arbitrary units (A.U). **B**: Plot showing a set of QPCR traces where each trace is part of a serial dilution. In this case the DNA is diluted by a factor of five between each reaction mix. The threshold line is shown in black and the c_t values are determined from where each trace crosses the threshold line. **C**: Plot of c_t values against the logarithm of the initial DNA concentration (f_0) for the traces shown in subfigure B. The primer efficiency is computed from the gradient of the line of best fit (shown in orange) using equation 2.12. **D**: Residuals for the line of best fit shown in subfigure C. The residuals indicate that a straight line fit is a suitable model for the data.

To determine the starting number of initiators (I_0), and terminators (T_0), we remember that each initiation of replication leads to a division, and creation of a new cell, at a time $C + D$ minutes later. Therefore, if at a time $t = 0$ we have N_0 cells, that means that at $t - (C + D)$ we had N_0 origins of replication. As we already saw, the number of origins of replication grows exponentially, and therefore at a time $t = 0$ the number of origins must be $N_0 2^{\frac{t}{\tau}} = N_0 2^{\frac{C+D}{\tau}}$. This is precisely our value of I_0. In a similar manner we can obtain a value for T_0 by noticing that each time a terminator is replicated, a new cell is formed via division, exactly D minutes later. This allows us to rewrite equations 2.14 and 2.15 as follows:

$$I = N_0 2^{\frac{C+D}{\tau}} 2^{\frac{t}{\tau}}, \tag{2.16}$$

$$T = N_0 2^{\frac{C+D}{\tau}} 2^{\frac{t}{\tau}}. \tag{2.17}$$

Now to determine the C period we can simply take a ratio of these two equations to yield the following:

$$\frac{I}{T} = \frac{N_0 2^{\frac{C+D}{\tau}} 2^{\frac{t}{\tau}}}{N_0 2^{\frac{D}{\tau}} 2^{\frac{t}{\tau}}} = \frac{2^{\frac{C+D}{\tau}}}{2^{\frac{D}{\tau}}}, \tag{2.18}$$

and finally:

$$\frac{I}{T} = 2^{\frac{C}{\tau}}. \tag{2.19}$$

We can rearrange this equation for C giving:

$$C = \tau log_2\left(\frac{I}{T}\right). \tag{2.20}$$

Therefore, if we know the number or origins and termini in a sample, along with the doubling time of the population, the C period can be determined. To extract the number of origin and termini using QPCR, we simply use two sets of primers to amplify a small regions of DNA very close to each. We then run the QPCR reaction and compute the c_t values generated. Following this, we use equation 2.5 to determine the number of origins (and termini) in the following way. First we rewrite equation 2.5 for each primer set at the c_t value:

$$f_t = f_{ori} \alpha_{ori}^{c_{t,ori}}, \tag{2.21}$$

$$f_t = f_{ter} \alpha_{ter}^{c_{t,ter}}. \tag{2.22}$$

Where: f_{ori} and f_{ter} are the initial amount of origin and terminus DNA in the sample; $c_{t,ori}$ and $c_{t,ter}$ are the c_t values for the set of primers which amplify part of

the origin and terminus respectively; and α_{ori} and α_{ter} are the corresponding primer efficiencies. From here we can rearrange each equation for the initial amount of DNA and take the ratio as follows:

$$\frac{f_{ori}}{f_{ter}} = \frac{f_t \alpha_{ori}^{-c_{t,ori}}}{f_t \alpha_{ter}^{-c_{t,ter}}}, \tag{2.23}$$

$$\frac{f_{ori}}{f_{ter}} = \frac{\alpha_{ter}^{c_{t,ter}}}{\alpha_{ori}^{c_{t,ori}}}. \tag{2.24}$$

Now, if we assume that the the fluorescent intensity is proportional to the total quantity of DNA, at least in the exponential region of the QPCR reaction, then we can rewrite f_{ori} and f_{ter} as kI and kT (where k is some constant of proportionality) and substitute this into equation 2.24 yielding:

$$\frac{kI}{kT} = \frac{I}{T} = \frac{\alpha_{ter}^{c_{t,ter}}}{\alpha_{ori}^{c_{t,ori}}}. \tag{2.25}$$

Finally substituting this expression into 2.20 we can express the C period in terms of variables extracted from QPCR:

$$C = \tau log_2 \left(\frac{\alpha_{ter}^{c_{t,ter}}}{\alpha_{ori}^{c_{t,ori}}} \right). \tag{2.26}$$

2.7.2 Experimental protocol

To acquire samples for QPCR, cells were grown as described in section 2.2.3. At an OD between 0.1 and 0.3 cells were harvested, and processed in a very similar manner as described in Azeroglu et al. 2016. First, approximately 15ml of cells were transferred into a $50ml$ Falcon tube and immediately put on ice. The sample was then centrifuged at $2000g$ for 7 min at $4°C$. After discarding the supernatent the cell pellet was resuspended in $1ml$ of ice cold PBS solution which had already been filter sterilised. Cells were then washed three times by centrifugation, followed by suspension in $1ml$ of ice cold PBS. At this point cells were pelleted a fourth time by centrifugation, the supernatent discarded, and the sample stored at -80$°C$ for several days, until it was time to perform the DNA extraction protocol.

To extract cellular DNA, frozen cell pellets were defrosted by placing on ice for 30 min. Following this, each pellet was resuspended in $250\mu l$ of ice cold CHIP buffer (200mM Tris-HCl (pH 8.0), 600mM NaCl, 4% Triton X-100). To disable any enzyme activity that may act on the DNA, the CHIP buffer used to resuspend cells

contained cOmplete^{TM} EDTA free protease inhibitor cocktail (Roche). Before resuspending cells $10ml$ of CHIP buffer was used to dissolve one protease tablet at $4°C$. To mix the sample after resuspension, a $10\mu l$ pipette was taken, and with care not to introduce bubbles, the sample was very gently mixed. To lyse cells and remove DNA samples were sonicated using a Diagenode Bioruptor at 30 sec intervals for 10 min, at high amplitude. During the entire process, from defrost to sonication, care was taken to make sure the sample did not exceed $5°C$. To extract the now fragmented cellular DNA from the soncicated sample a standard PCR purification kit was used, following the manufacturers instructions (QIAcquick PCR purifcation Kit, Qiagen). DNA was eluted in the final step using $50\mu l$ of sterile deionised water.

With the sample DNA prepared, the QPCR reaction could be performed. I used a LightCycler 96 System (Roche) which allowed the recording of 96 separate QPCR reactions simultaneously. For the reaction chemistry I used Brilliant II SYBR Green qPCR Master Mix (Agilent Technologies). For the primer sequences I used those utilised in Si et al. 2017, namely: TTCACCAGATGGGCAATCACTTCGG and ACATTGTGGTGGTGCATAACGGCATC for the origin and ATGGATCTTCCTGCTATACCGCCAC and GGCCAACATTTGCGTCCACAGATAAG for the terminus. In both cases the region amplified is around 100 base pairs long. Each reaction volume was $10\mu l$ which contained: 5 μl of SYBR green master mix; $0.5\mu l$ of forward and reverse primer (of $10\mu M$ concentration); $2\mu l$ of target DNA at around $0.5ng/\mu l$ concentration and $2\mu l$ of deionised water. Typically each 96 well plate was divided into two lots of 48 reactions, one half per sample being tested. Normally one plate would contain two biological repeats of the same condition being tested. Each half plate was then divided into two sections, one section with primers to amplify the origin, while the other amplified the terminus. For each primer set a dilution series was created, by diluting the sample DNA five times with each subsequent aliquot being diluted by a factor of five compared to the previous. The sixth concentration was a blank, which contained no template DNA and was used as a negative control. For each sample, primer set, and concentration four repeats were run per plate to allow averaging of c_t values during the analysis. Figure 2.23 shows a schematic of the layout of a typically 96 well plate. After filling, each plate was sealed and spun down gently for approximately 15 seconds at low RPM, to ensure all the reaction mixture lay at the bottom of the well. The plate was then run through a simple two

step reaction protocol starting with an initial denaturation at $95°C$ each cycle consisted of 30 *sec* of denaturation followed by 30 *sec* of annealing at $60°C$. A separate extension phase was not required due to the small region being amplified in each reaction, as the polymerase was able to act during the transitions between annealing and denaturing phases. At the end of each reaction, a melting curve was performed by slowly increasing the temperature and measuring the change in fluorescence of the sample to ensure that the reaction had generated a single PCR product.

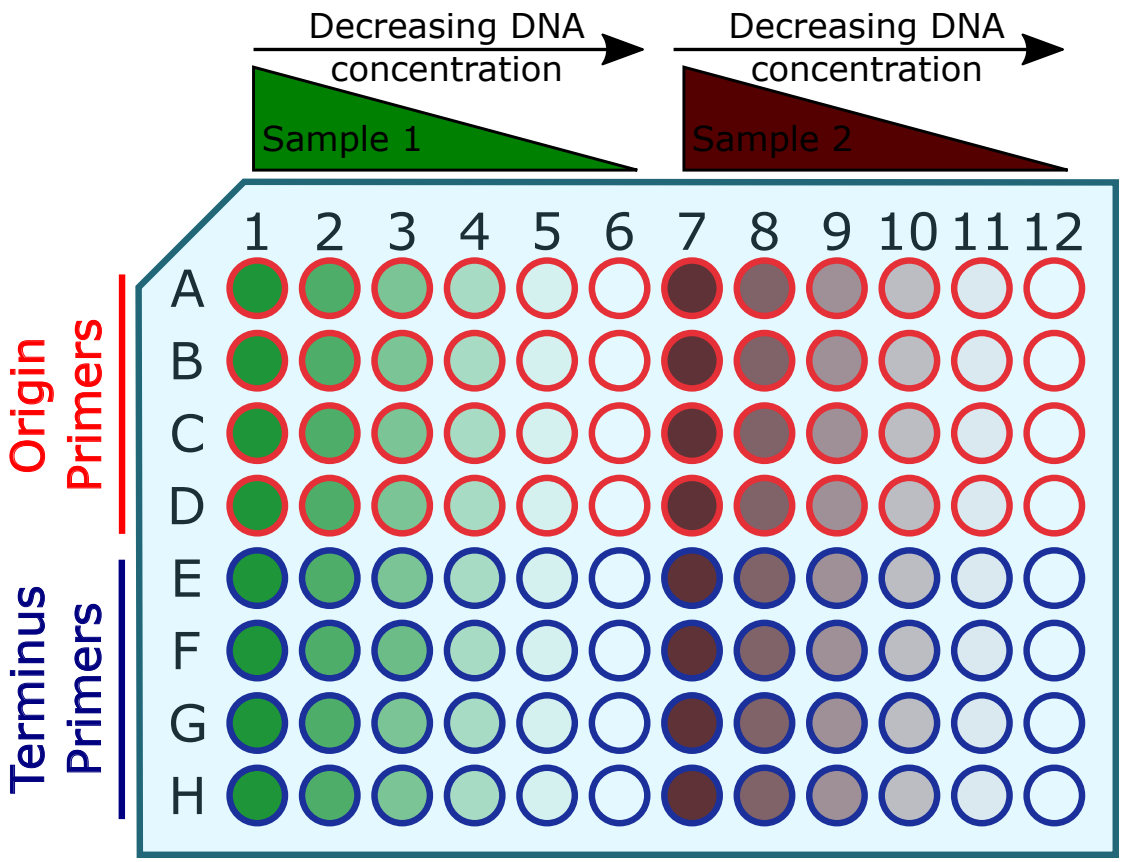

Figure 2.23: Typical QPCR plate layout. Each plate normally held two samples which tended to be biological repeats of the same condition. To allow primer efficiencies to be calculated on a per reaction basis each sample was run with six different concentrations of DNA, including no DNA, as indicated by the red and green colour gradients. The two sets of primers which amplify the origin and terminus respectively are shown in red and blue, and outline each well. For each sample, primer and DNA concentration four repeats were carried out to allow averaging of c_t values during the analysis.

Chapter 3

Results

3.1 Growth rate at high osmolarities

To begin my investigation into the effects of high osmolarity conditions I set out to determine how the bulk growth rate varies with increasing concentration of solute. To this end, I grew cells as discussed in 2.2.3 and measured cell density via spectrophotometer measurement. This was done at a number of times points in the early to mid exponential region. The medium used for these experiments was M63 with a glucose carbon source and supplemented with cas amino acids. I altered the osmolarity by adding varying concentrations of NaCl and focused on four main conditions which included: 0, 200, 400 and 600mM of additional NaCl. Figure 3.1:A shows the optical density measurements against time for all four NaCl concentrations, with the horizontal axis indicating the time in hours from the inoculation of the culture. Already from figure A we can see qualitatively that both the doubling time and lag time increase, with increasing NaCl concentration. To compute the value of the growth rate in each case we restrict our analysis to measurements in the range between 0.05 and 0.4 OD_{600}. The reason for this minimum restriction is that according the spectrophotometer manual readings are not accurate below this minimum threshold. An upper limit was necessary because at some point we expect the culture to depart from exponential growth and enter into stationary phase, with a corresponding reduction in the growth rate. I chose 0.4 for this threshold as it represented a good balance between being sufficiently low to exclude parts of the curve entering stationary phase, while also being sufficiently high to allow a substantial number of time points to be gathered in the permissible range. As shown in

figure 3.1:B the growth rate can be extracted by plotting the natural logarithm of the OD_{600} value for each trace against time, and computing the gradient of the resulting line of best fit. As indicated by the computed r^2 values, and the residuals shown in figure 3.2, a linear fit is a good approximation of the data in this restricted region between 0.05 and 0.4 OD_{600}, this would indicate that we have correctly identified a region of exponential growth.

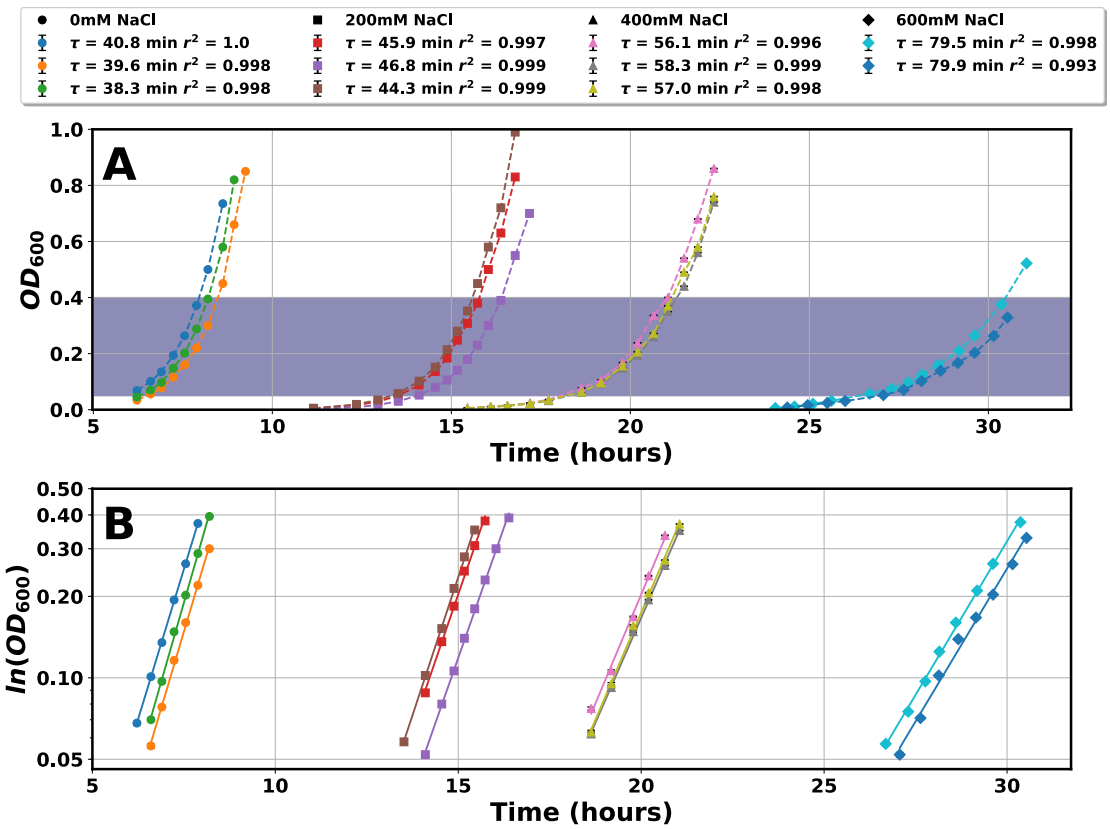

Figure 3.1: Growth curves at high osmolarity. **A**: Plot showing OD_{600} against time for four different osmolarities. For each condition two to three repeats are included that were inoculated using different frozen stock aliquots. The purple highlighted region indicates the range of OD_{600} values which were considered for analysis. **B**: Semi log plot of the data in the highlighted region in sub plot A. The solid line represents a line of best fit obtained by least squares regression in the python package numpy. The r^2 value for each fit as well as the computed doubling time is shown in the figure legend. In both cases a measurement error of $\pm$ 0.001 was plotted but is smaller than the size of the data points.

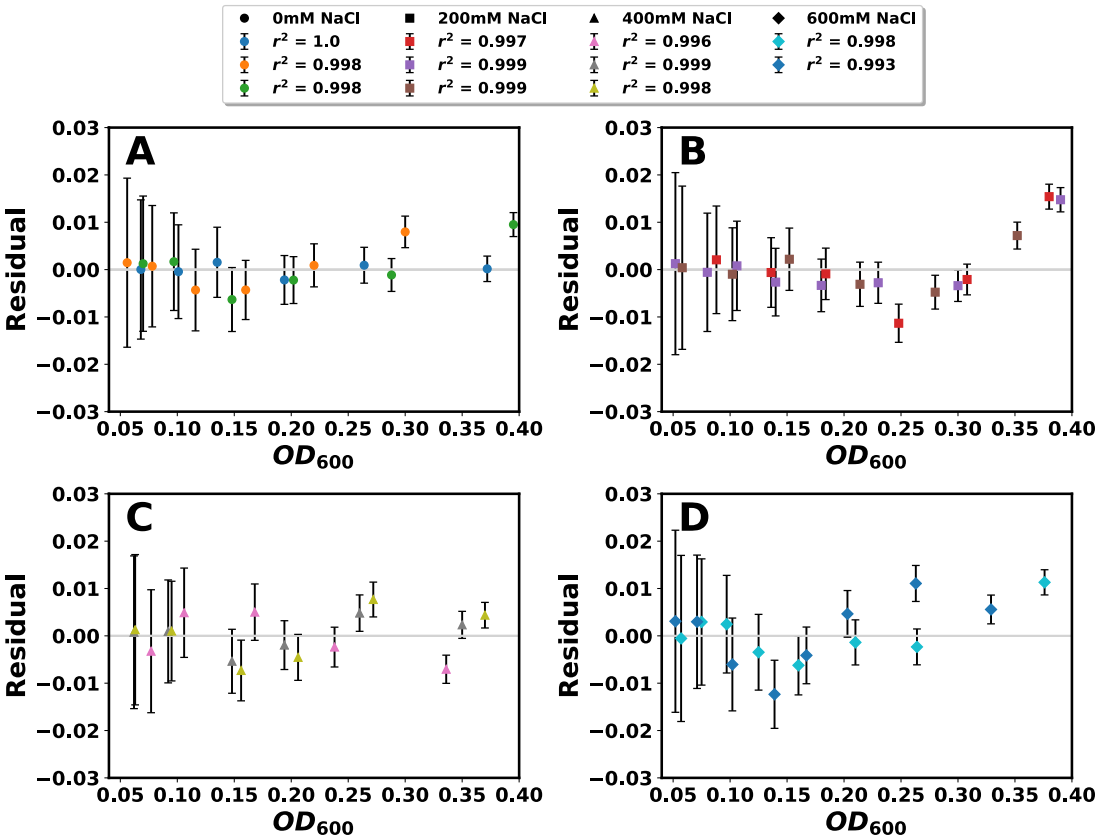

Figure 3.2: Growth curve residuals. Residual plots for the lines of best fit shown in figure 3.1:B. As indicated by the legend each panel contains one of the four conditions tested. Error bars represent a constant ± 0.001 measurement uncertainty in linear space however as the y value is logged to obtain lines of best fit the error bars appear to decrease in size with with increasing OD_{600} value.

To get a better appreciation of how the growth rate of cells change with increasing osmolarity, figure 3.3:A plots the average doubling time and growth rate as computed in figure 3.1 against the osmolarity of each media. The osmolarity was measured using a Micro-Digital Osmometer Type 15M (Camlab). As we can see from the plot there is a clear, almost linear, decrease in growth rate with increasing osmolarity, a result which matches a number of previous findings (Dai, Zhu, et al. 2018; Pilizota and Shaevitz 2014). In particular in figure 3.3:B I have shown the results from Dai and Zhu 2018 which examined the effect of high NaCl concentrations on a number of physiological parameters. In this study the media used was LB and so the growth rates are consequently higher at all osmolarities. However, despite this, their data shows a good qualitative agreement with the decreasing growth rates we observe

with increasing media osmolarity.

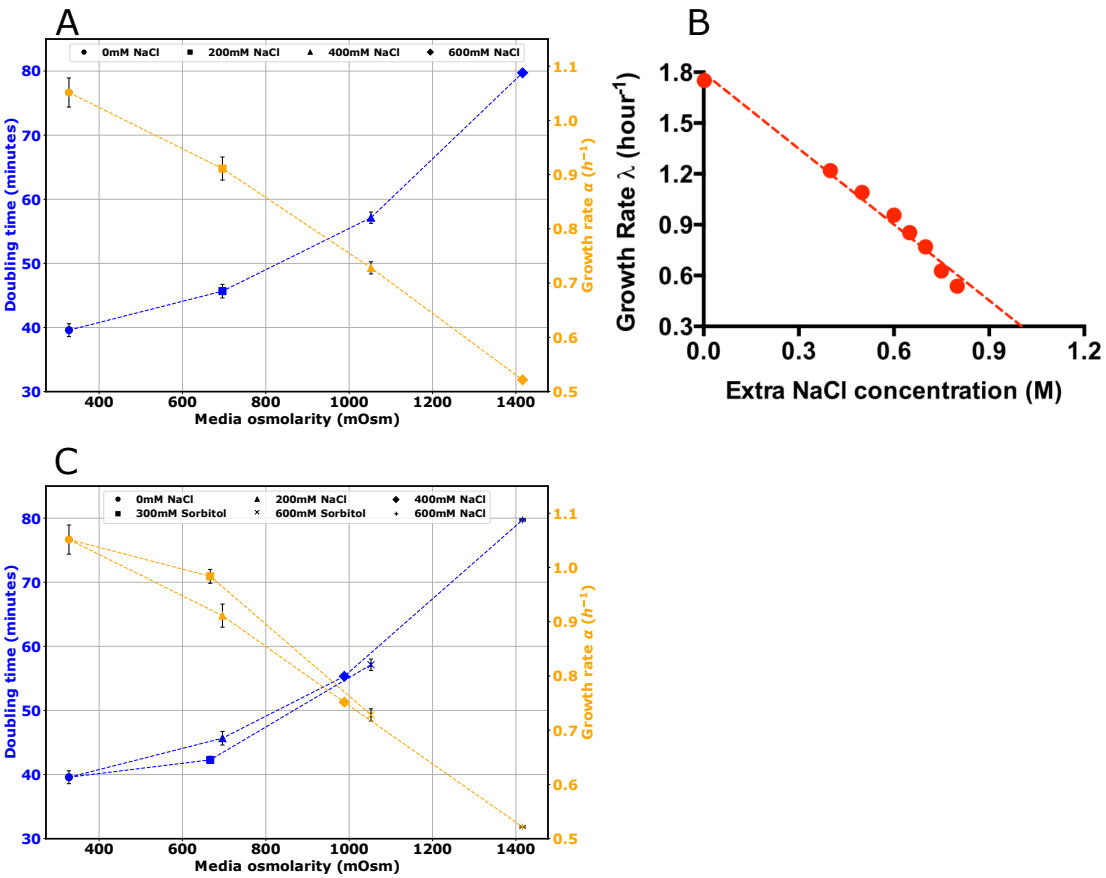

Figure 3.3: Osmolarity vs growth rate. **A**: For each condition in figure 3.1 the computed growth rates and doubling times were averaged and plotted against the osmolarity of the media. To measure the osmolarity three readings were taken and averaged to yield the final value. The horizontal error bars which represent the standard deviation of the osmolarity measurements are too small to be shown while the vertical error bars represent the standard deviation of the growth rates or doubling times. **B**: Plot showing growth rate against added NaCl concentration as measured by Dai and Zhu 2018. **C**: Same plot as subfigure A but with added data from the 300mM and 600mM sorbitol conditions.

To determine if the observed decrease in growth rates was dependent on the osmolyte used, I performed a small number of experiments utilising sorbitol instead of sodium chloride to increase the osmolarity. Sorbitol was chosen for two reasons. Firstly, because it is often used in studies examining high osmoliarty conditions, and

secondly because in contrast to Sodium Chloride it is a large and uncharged molecule (Rojas, Theriot, and K. C. Huang 2014b). As can be seen in figure 3.3C, sorbitol appears to follow a broadly similar downward trend in growth rate as sodium chloride. While it is difficult to analyse the quantitative difference between NaCl and Sorbitol with only two data points, it does appear that the trend in growth rate diverges, and is not identical, with very close agreement at high osmolarity and more divergence at lower osmolarity. This is a trend that we will see repeated when we turn to cell size analysis in the following section.

3.2 Cell size at high osmolarity

3.2.1 Cell size from microscopy

3.2.1.1 Recreating the nutrient growth law

Before measuring how high osmolarity conditions affect cell size, I first wanted to ensure that the algorithm I developed for segmenting cells and measuring their volumes was able to reproduce known results. To this end I decided to reproduce the nutrient growth law, that showed an exponential dependence of cell volume on the nutrient limited growth rate. To achieve this goal, I grew cells in three media of different nutrient quality. The richest media was RDM, followed by M63 with glucose and cas amino acids and finally M63 with glycerol. In each condition I measured between 150 to 300 cells across two biological replicates. Figure 3.4 A and B show the volume distribution gathered from cells in these conditions. As we can see, their is an approximately three fold increase in volume with increasing nutrient quality (and decreasing doubling time). Figure 3.4 shows that if I plot the logarithm of the mean volume against the growth rate, I recover a linear relationship indicative of the nutrient growth law. To compare my results with previous work I can express the dependence between the growth rate and cell volume as an exponential relationship of the form:

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

Where: S is the average cell volume; λ is the growth rate and S_0 and γ are constants. As can be seen from figure 3.4:C, I obtain a value for γ of 1.29 ± 0.03 . Likewise, for S_0 I compute a value of 0.23 ± 0.03. These values compare well with those obtained in Si et al. 2017, which found a value of 1.33 for γ and a value of 0.26 for S_0[1]. The good agreement shown here between my results, and literature values implies that my experimental set up, and subsequent image analysis algorithms, are working effectively to accurately measure relative changes in cell volume. To determine if changes in cell volume came from changes in width or length, I plot their distributions independently in figure 3.5. We see that both the length and the width increase with increasing growth rate, in a way the mirrors the overall volume changes. In fact

[1]The value for S_0 was not quoted directly in the paper and instead was taken from Harris and Theriot 2018 where they re-plot the data from Si et al. 2017 and explicitly state S_0.

both the width and length increase by about one and a half times when the media is changed from M63 + glycerol to RDM. This shows at least qualitative agreement with Taheri-Araghi et al. 2015 which showed an exponential dependence of both length and width (of approximately equal magnitude) on increasing growth rate.

An additional check I wanted to make on the robustness of my cell size measurement, was to examine if cell intensity had any effect on the measured size of

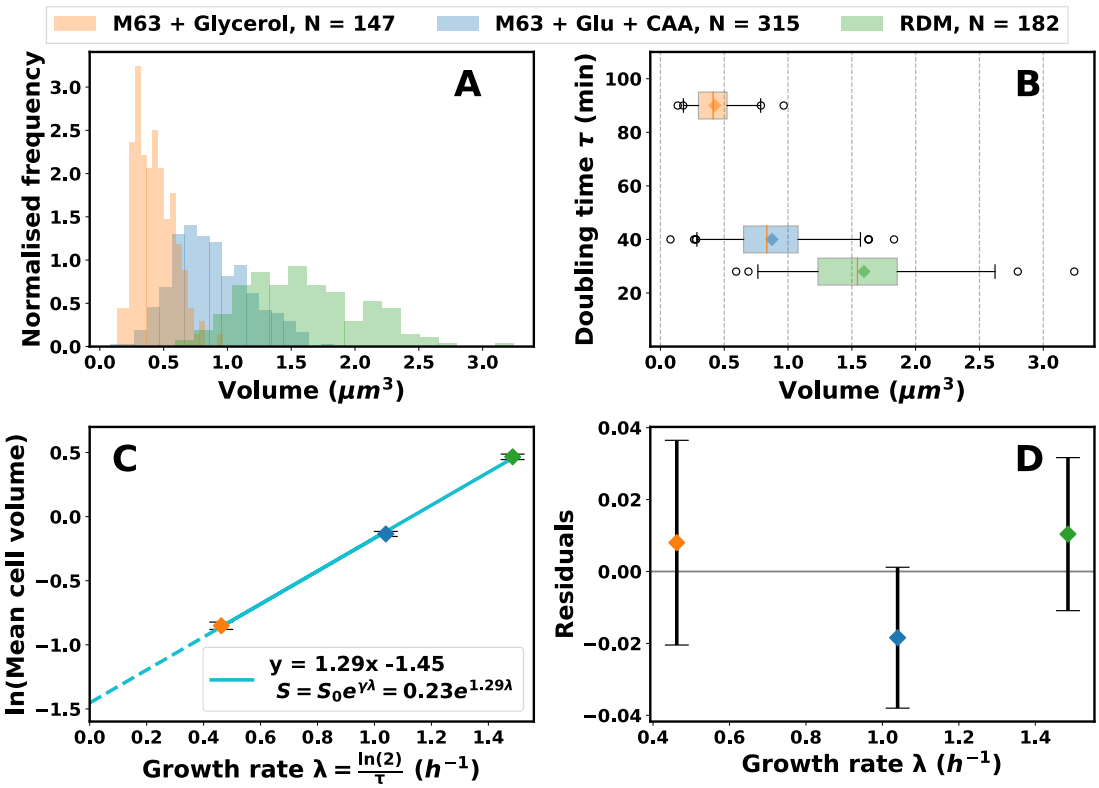

Figure 3.4: Cell size and the nutrient growth law. **A:** Histogram of cell volume for each of the three conditions. The histograms were normalised such that the area under each curve is unity. **B:** Box plots showing the spread of cell volumes for each condition against the doubling time in that condition. Diamonds represent means while orange lines indicate medians. The whiskers in each plot indicate the area including 98% of the data. **C:** Plot recovering the nutrient growth law. The growth rate λ is plotted against the natural logarithm of the mean cell volume for each condition. The blue line indicates line of best fit, and error bars are standard error on the mean **D:** Residuals for the lines of best fit shown in subfigue C. Error bars are standard error on the mean.

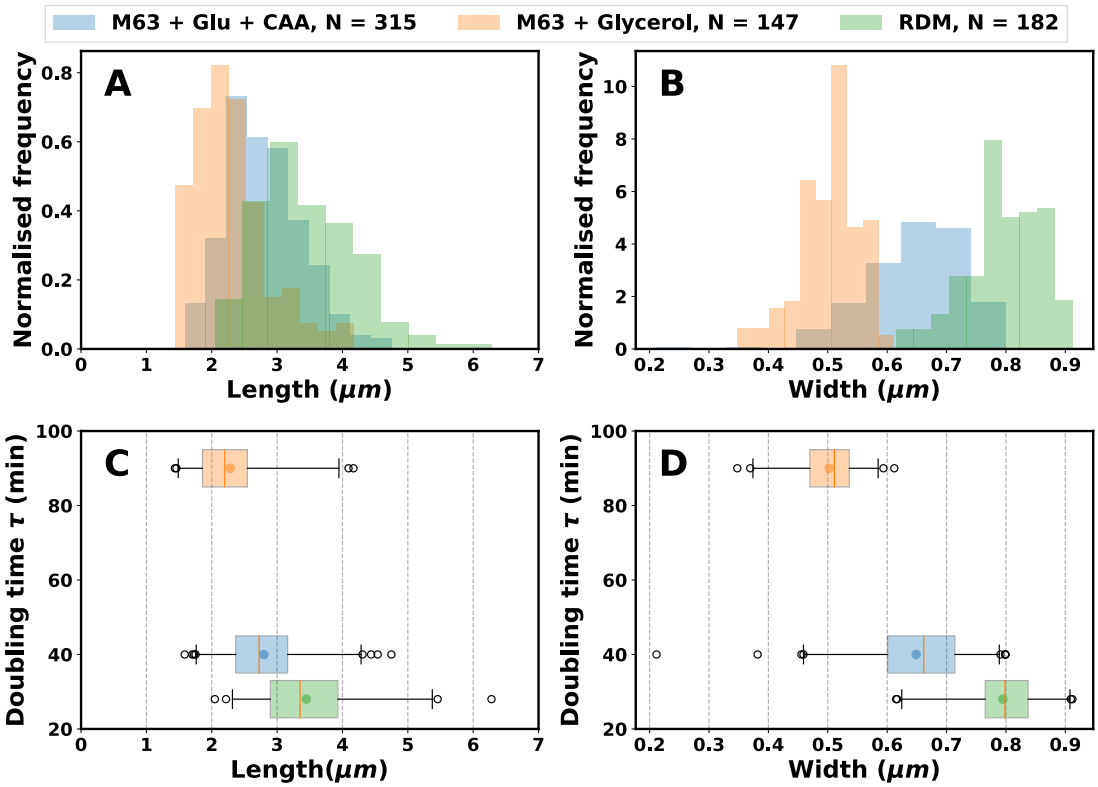

Figure 3.5: Length and width distributions of cells under nutrient limitation. **A & B**: Histograms showing the width and length distribution of cells grown in M63 media with glucose and cas amino acids, RDM and M63 with glycerol. Histograms have been normalised to have unit integral. **C & D:** Box plots showing cell length and width distributions under nutrient limitation. Dots represent means while vertical orange lines indicate medians. The whiskers span an area which includes 98% of the data.

the cells. To do this, I measured the average pixel intensity of all pixels in the segmentation mask for each cell which was measured. I then was able to plot the mean pixel intensity of each cell, against its measured volume as shown in figure 3.6. From the scatter plot alone it appeared that there was no obvious correlation between cell intensity and size. However, to attempt to quantify the magnitude of any correlation, I computeD the Spearman correlation coefficient for each data set, which can be seen in the legend of figure 3.6:D. In all cases the Spearman coefficient suggested at most a weak correlation between intensity and volume. This indicated that my results were not being strongly biased by cell intensity differences.

As a final control I wanted to ensure that my experiments were not being biased as a result of imaging the fluorescent channel. In essence, I wanted to ensure that there was not a population of non-fluorescent cells which I was failing to consider in my analysis. To ensure this, I grew cells in M63 media with glucose and cas amino acids. I then placed cells on an agar pad (rather than a tunnel slide so that cells were immobilised) allowing me to image each cell twice: once in the fluorescent GFP channel and once in bright field. I was then able to overlay the bright field and fluorescent images to determine how many cells lacked a fluorescent signal. As can be seen in figure 3.7 A through E all cells imaged were fluorescent under normal

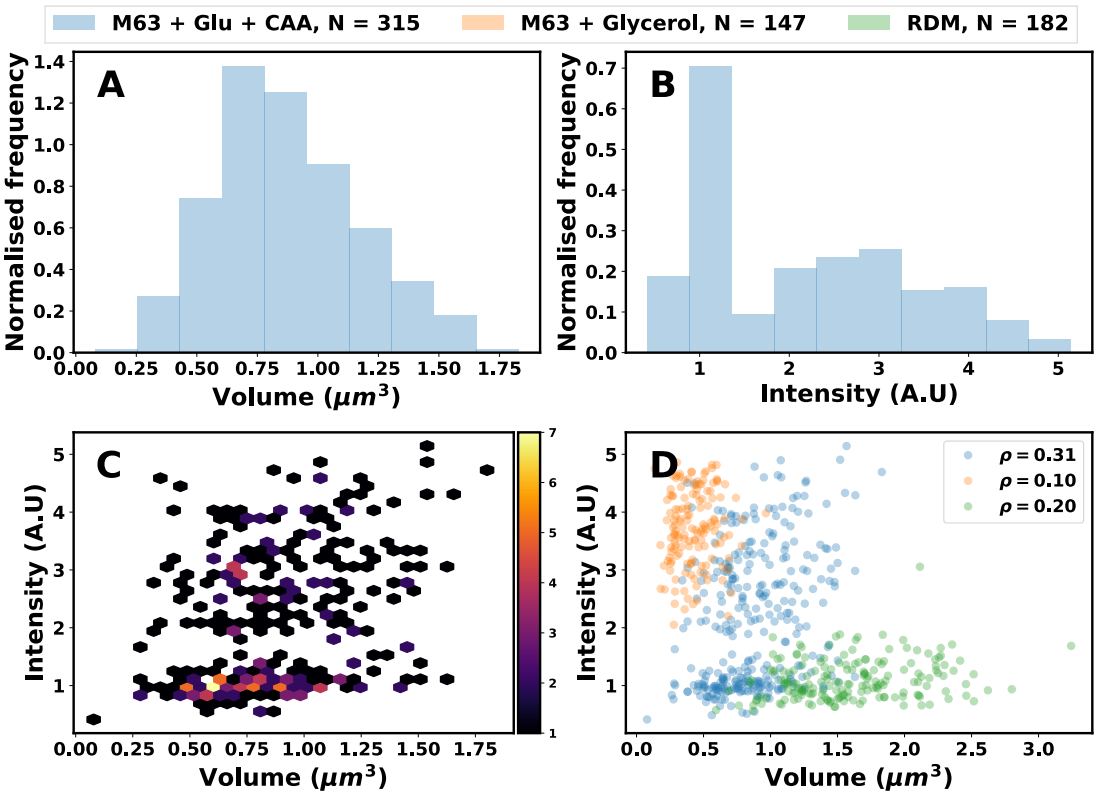

Figure 3.6: Effect of cell intensity on measured volume. **A:** Volume histogram of cells grown in M63 media with glucose and cas amino acids. **B:** Histogram of mean cell intensity for cells grown in M63 media with glucose and cas amino acids. **C:** Two dimensional histogram showing the relationship between volume and mean intensity. The colour bar represents the number of cells in each hex shaped bin. **D:** Scatter plot showing relationship between volume and mean intensity for all three nutrient conditions tested. The symbol ρ in the figure legend indicates the value of the computed Spearman coefficients.

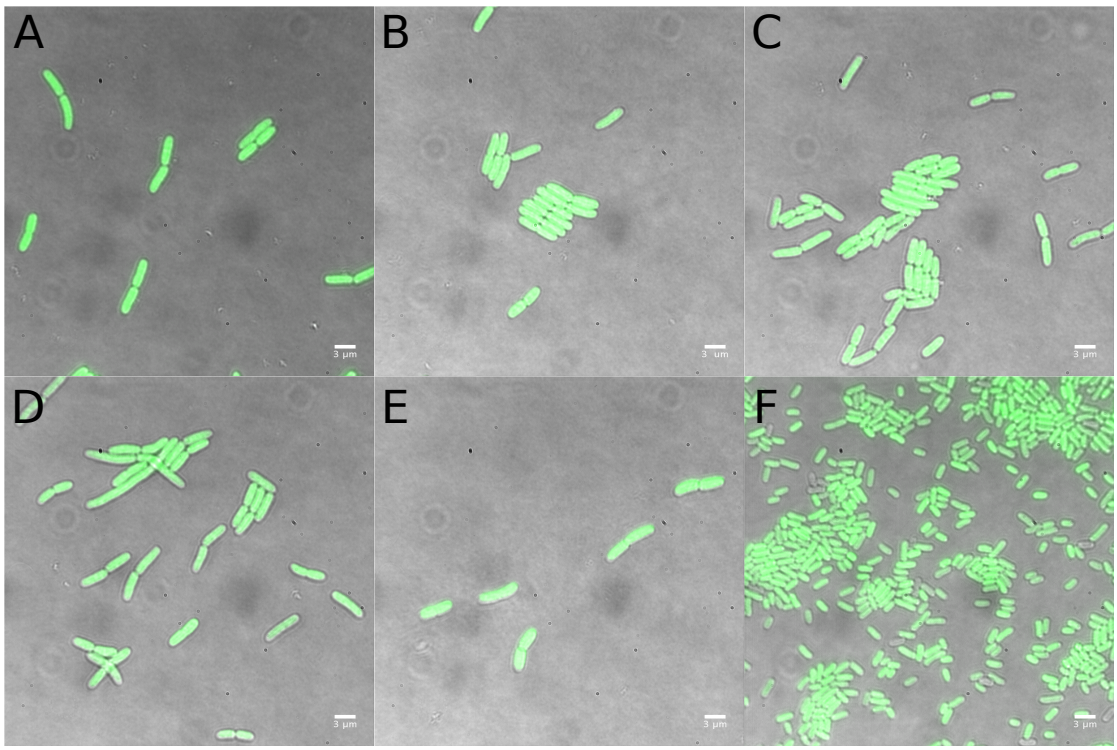

Figure 3.7: Bright field and fluorescence imaging of YD133 + pWR20. **A-E:** Cells grown as normal for a volume measurement experiment. **F:** Cells grown into late stationary phase in the absence of antibiotic selection. In all cases the bright field image (grey scale) has been overlaid with the fluorescent image (green). Scale bars are three microns.

conditions. To test a "worst case" scenario I also grew cells into stationary phase in the absence of antibiotic to test if a lack of positive plasmid selection caused its loss, and therefore a reduced number of fluorescent cells. As shown in figure 3.7:F, even under these conditions the vast majority of cells maintained their fluorescent signal. I therefore concluded that by imaging in the fluorescent channel only, I was still measuring a representative sample of the population.

3.2.1.2 Effect of high osmolarity

With some confidence that my experimental protocol and subsequent image analysis is accurately reflecting changes to cell size in different conditions, we now turn to examine what effect high osmolarity conditions have on cell volume. To do this I measured the size of cells grown in M63 with glucose and cas amino acids under different osmotic conditions. The osmolarity was altered by adding 0, 200, 400 and

600mM NaCl and 300mM sorbitol. As before, at least two biological replicates were measured and the results pooled to yield between 300 and 750 cells per condition. Figure 3.8 A and B show the volume distributions for each condition. From these distributions themselves it is quite difficult to pick out a discernible pattern. However, we can gain some clarity by seeing how our results compare to the nutrient growth law measured in the previous section. In figure 3.8:C I have plotted the mean volume against growth rate for each osmotic condition. I have also plotted the results obtained from the previous section for both RDM and M63 and glycerol. When I plot lines of best fit through the high osmolarity conditions we see that there is clear departure from the nutrient growth law. Although the growth rate decreases substantially, there is very little decrease in average cell size. In fact, although the overall slope of the line is downward with increasing osmolarity, if we consider just the NaCl conditions, we see that there is actually an initial increase in cell size at 200mM NaCl. From this point there then proceeds to be a monotonic decrease with increasing osmolarity. Another interesting observation is the apparent difference between 200mM NaCl and 300mM sorbitol. While these two conditions are very close in growth rate and osmolarity (figure 3.3), they show a substantial difference in cell size. In particular, sorbirtol has a significantly lower cell size than even 600mM NaCl even though it has much lower osmoalrity. This is the first strong indication we have that there may be a difference in effect between conditions of the same osmolarity, depending on the nature of the osmolyte used.

As before, to gain an understanding of how length and width changes contribute to the overall volume change, I plotted their individual distributions in figure 3.9. From the box plots in particular, we can see that there is a complex relationship where the length decreases while the width increases, with increasing osmolarity. Again, examination of the 300mM sorbitol condition proves interesting. Although it shows a very similar width distribution to the 200mM NaCl condition, it has a significantly lower length by comparison. To better exemplify these differences in length and width I have plotted the surface area to volume ratio for each condition in figure 3.10. I have also included the nutrient limitation conditions for comparison. By taking the ratio of the surface area and volume for each condition we summarise the length and width changes into a single value, which is easily interpreted. Cells with a low surface area to volume ratio tend be thicker, shorter and therefore more

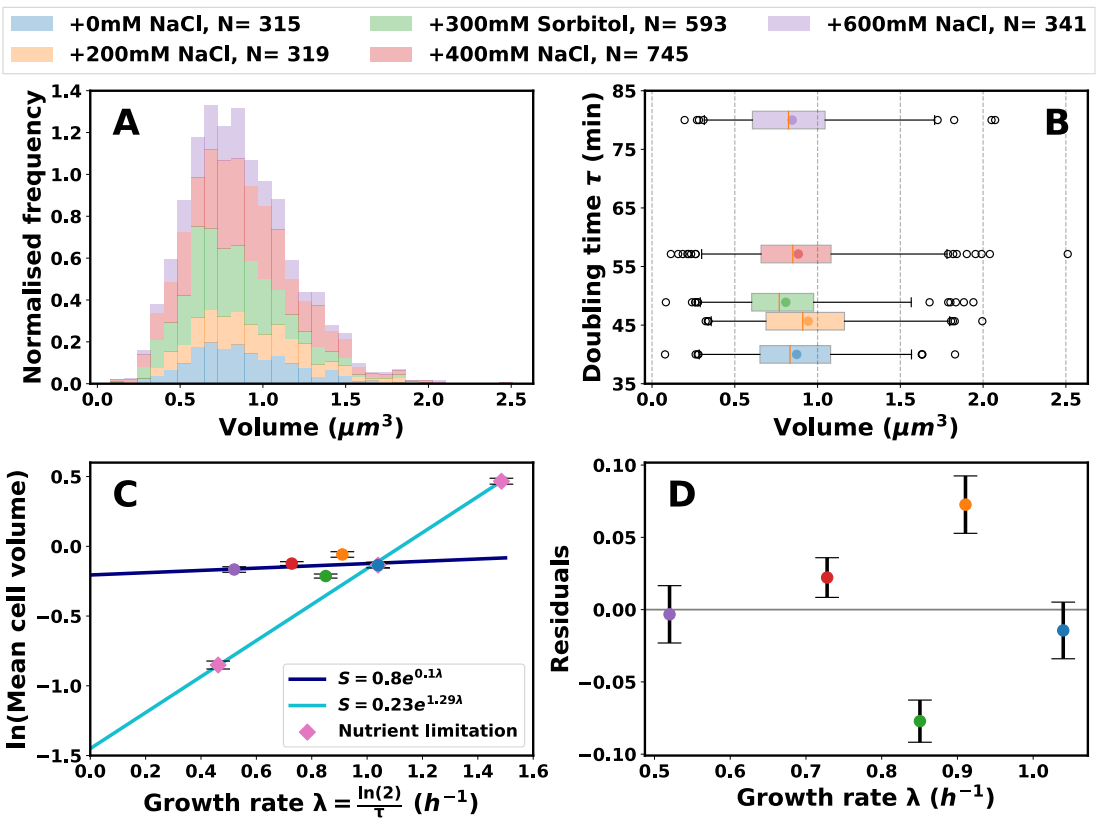

Figure 3.8: Effect of high osmolarity on cell volume. **A:** Stacked histogram showing distribution of cell volumes for each condition. Histograms were normalised to have an integral of 1. **B:** Box plots showing distribution of cell volume against doubling time. The mean is represented by a dot for each condition, while the median is indicated by a vertical orange line. Whiskers cover a range containing 98% of the data. **C:** Plot of the logarithm of mean cell volume against growth rate for both the high osmolarity conditions, and nutrient limitation conditions. Solid lines represent lines of best fit. The equation for each line in shown in the figure legend. Error bars are standard error on the mean **D:** Residuals for the line of best fit for the high osmolarity conditions in subfigure C.

spherical in shape. Conversely cells with a high surface area to volume ratio tend to be narrower, longer and less sphere like. Figure 3.10 shows that under nutrient limitation, cells become more spherical with increasing growth rate. This is a finding that allows for ready comparison with previous work. In particular, Harris and Theriot 2018 showed recently that the surface area to volume ratio did follow a linear relationship with growth rate. Specifically they found that the surface area to volume ratio ($\frac{SA}{V}$) depended on the growth rate λ in the following manner:

$$\frac{SA}{V} = -2.8\lambda + 9.3 \tag{3.2}$$

This relationship shows good agreement with the relationship I found for the surface

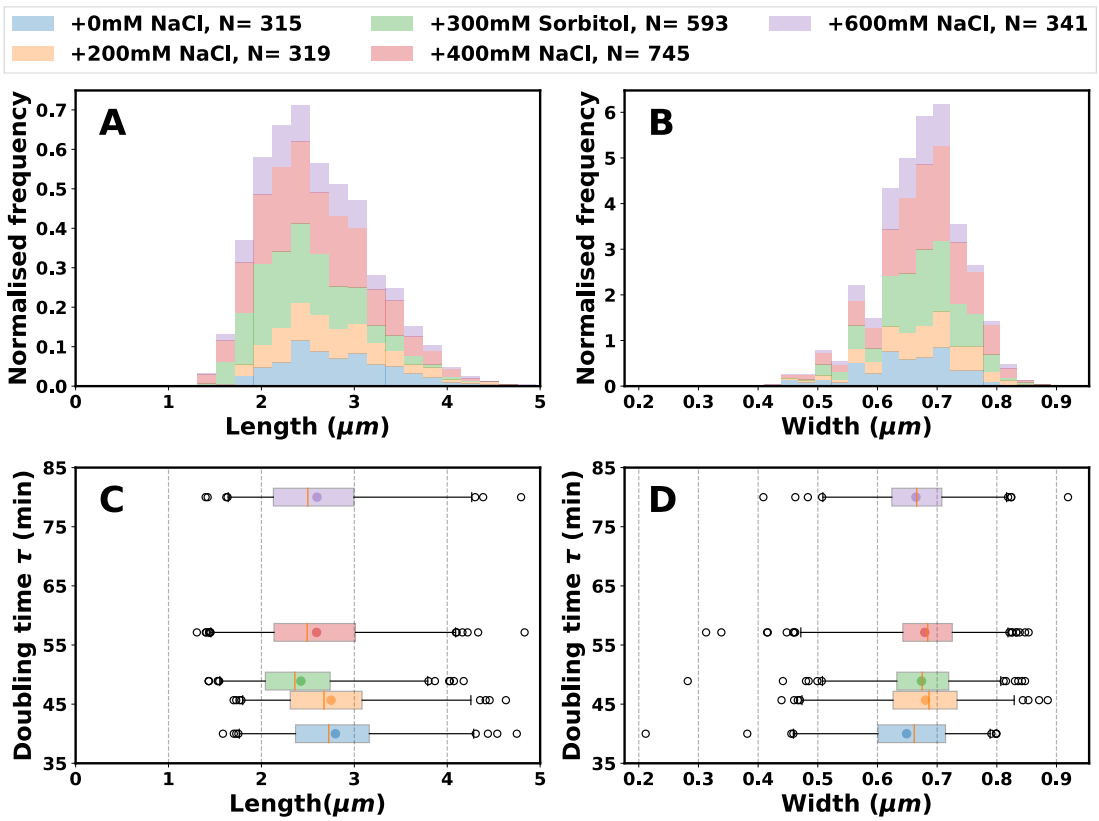

Figure 3.9: Length and width distributions of cells at increased osmolarity. **A & B**: Stacked histograms showing the width and length distribution of cells grown in various high osmolarity conditions. Histograms have been normalised to have unit integral. **C& D:** Box plots showing cell length and width distributions at high osmolarity. Dots represent means while vertical orange lines indicate medians. The whiskers span an area which includes 98% of the data.

area to volume ratio for the nutrient growth law, as shown in the legend of figure 3.10 (where errors are taken from uncertainty in the fitting):

$$\frac{SA}{V} = (-3.0 \pm 0.2)\lambda + (9.8 \pm 0.2). \tag{3.3}$$

This close agreement in the slope of the line between the data shown here, and that

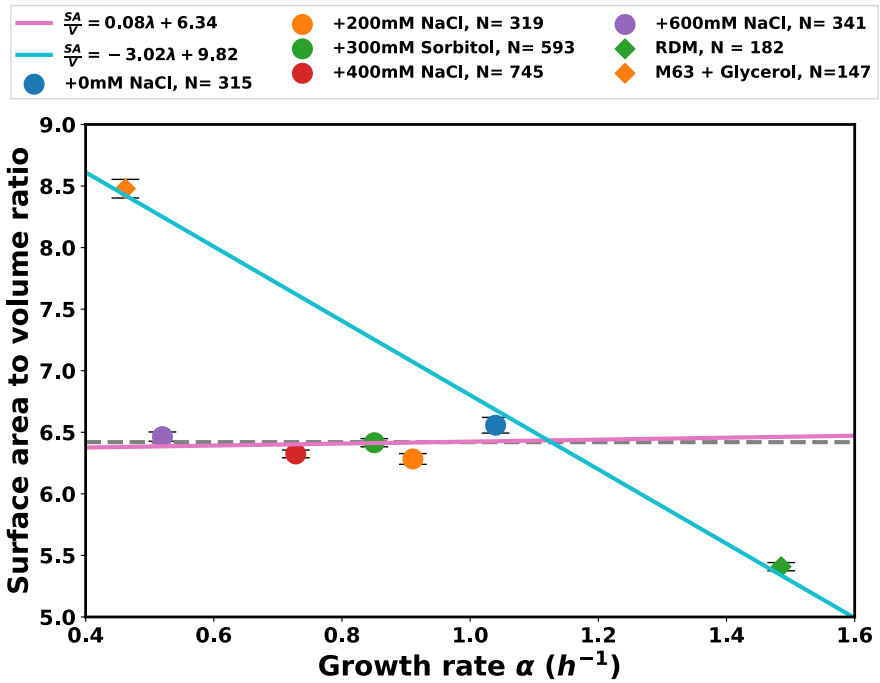

Figure 3.10: Surface area to volume ratio at high osmolarity. Plot showing how the mean surface are to volume ratio varies with growth rate for both nutrient limited conditions (diamonds) and high osmolarity conditions (dots). Error bars represent the standard error on the mean. Solid lines are lines of best fit. The grey dashed line is a perfectly horizontal line for comparison with the trend with increasing osmolarity.

of Harris and Theriot 2018, indicates that the measurement of length and width, are independently verified. Combined, with the agreement in observation of the nutrient growth law in the previous section we can state with high confidence that our methods are accurately portraying changes to cell length, width and volume.

Figure 3.10 also indicates that, as with volume, high osmolarity has a profound effect on cell shape. There is a clear departure from equation 3.3 as the osmolarity increases. In fact, if we fit the high osmolarity data, the result is an almost horizontal

line that shows very little change in surface area to volume ratio with growth rate (the dashed grey line in figure 3.10 shows a perfectly horizontal line for comparison).

To compare my results regarding the effect of high osmolarity on cell size, I plot in figure 3.11:A, the mean cell volume against media osmolarity. Adjacent to this plot in figure 3.11:B, I have included the results from Dai and Zhu 2018 which shows how cell volume depends on sodium chloride concentration, for cells grown in LB media. My results indicate that there is small but complex dependence between cell size, and concentration of NaCl. At 200mM NaCl there is an approximately ten percent increase in cell size, followed by a monotonic decrease thereafter with increasing NaCl concentration. At 600mM NaCl cell volume is reduced by less than five percent of the standard osmolarity condition. This result contrasts the findings of Dai and Zhu 2018 which show an approximately twenty percent decrease in cell size at 600mM NaCl. Even accounting for difference in media and strains, it seems that I observe a less significant change in cell size as osmolarity increases.

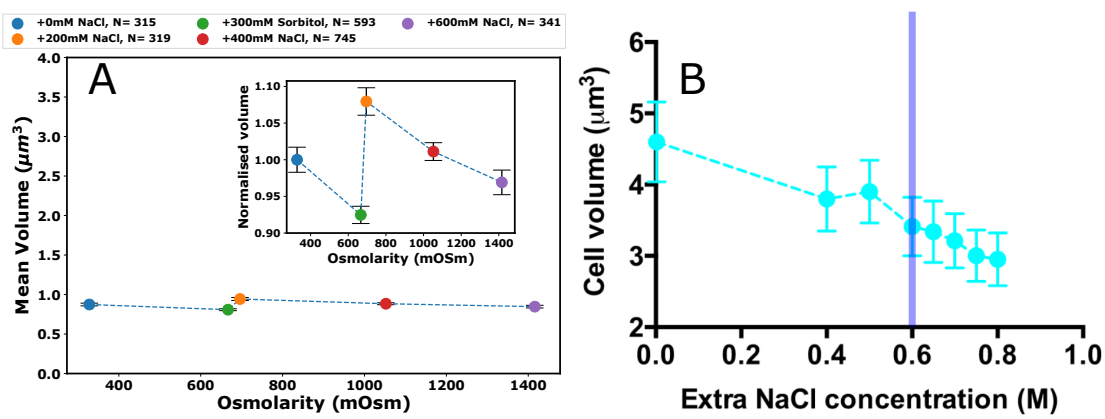

Figure 3.11: Cell volume at high osmolarity. **A:** Mean cell volume against osmolarity. The figure inset shows the same data on a normalised plot where the standard osmolarity condition has a volume of 1. Error bars indicate the standard error of the mean. **B:** Figure from Dai and Zhu 2018 showing effect of increased NaCl concentration on mean cell volume. Vertical blue line indicates the maximum concentration I tested in my experiments (600mM NaCl).

Finally, as in the previous section looking at nutrient limitation, I wanted to ensure that the segmentation of cells was not being affected by variation in cell intensity. In figure 3.12, just as in figure 3.6, I have plotted how the intensity of each

cell varies with its mean pixel intensity. As before, both the qualitative shape of the plots and the computed Spearman coefficients suggest that the intensity is only very weakly correlated with the cell volume. I conclude therefore, that intensity is unlikely to be a significant source of bias.

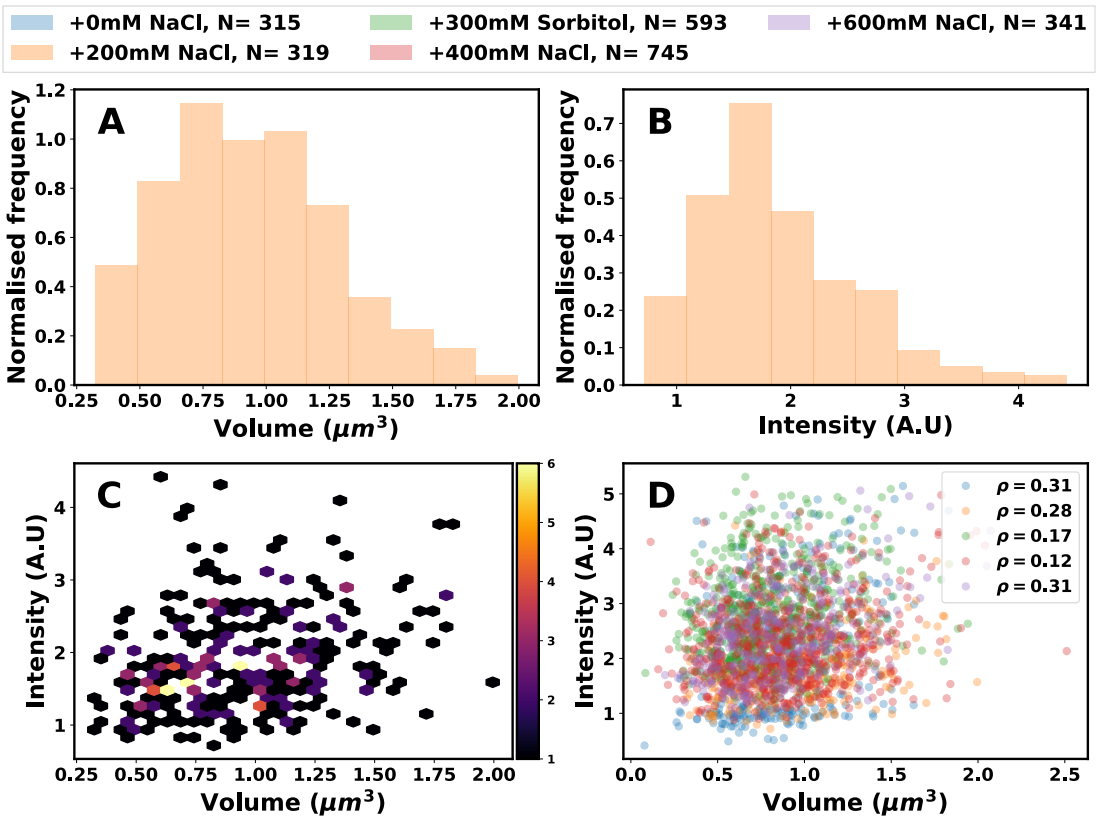

Figure 3.12: Intensity and cell volume correlations at high osmolarity. **A:** Volume histogram of cells grown at 200mM NaCl. **B:** Histogram of mean cell intensity for cells grown at 200mM NaCl. **C:** Two dimensional histogram showing the relationship between volume and mean intensity. The colour bar represents the number of cells in each hex shaped bin. **D:** Scatter plot showing relationship between volume and mean intensity for all high osmolarity conditions tested. The symbol ρ in the figure legend indicates the value of the computed Spearman coefficients.

3.2.2 Cell size from Coulter counter

3.2.2.1 Comparison to standard beads and microscopy data

As discussed in section 2.5, I aimed to verify the results obtained from microscopy measurements using the alternative high throughput technique of the Coulter counter. As we will see in this section, there are two main benefits to measuring cell size using the Coulter counter, as compared to microscopy based techniques. Firstly, it allows for much more detailed size distributions to be measured as a result of the orders of magnitude increase in cell number per measurement. Secondly, the speed with which distributions can be measured, allows for measurements to be taken at a number of time points along the growth curve. These advantages come at the cost however, of losing all information regarding cell shape, length and width. Instead, I extract only a single variable of cell volume for each measurement.

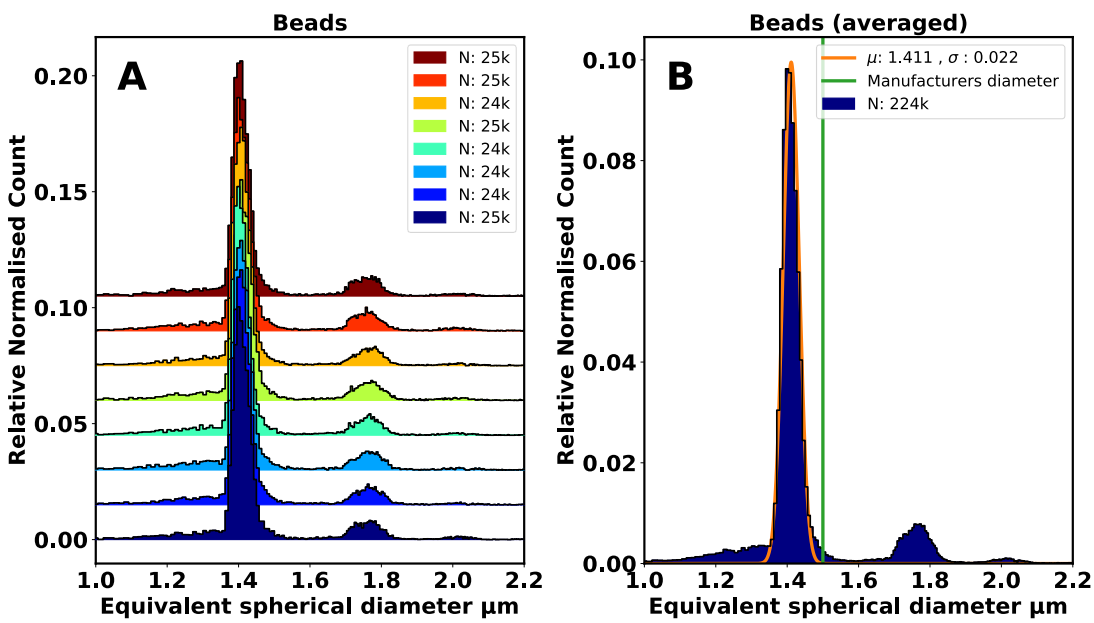

Figure 3.13: Size distributions of polystyrene beads as measured by the Coulter counter. **A:** Particle diameter against count for eight repeated measurements, each of around 25,000 particles. **B:** Combined measurement of all the runs in subfigure A into a single trace. The orange curve shows a Gaussian function fitted to the data. The green line represents the manufactures quoted diameter for the beads.

The first sample I measured with the Coulter counter was polystyrene beads,

with a diameter of $1.5\mu m$ (Polysciences Inc). The reason for this, was to determine how the absolute measurement of particle size compared for a sample of known dimensions. In figure 3.13 I show the measured size distribution for a sample of beads. Subfigure A shows eight repeated runs of the same beads sample. Each histogrma in the figure has been normalised, however, they have been artificially spaced along the y axis to improve readability in a style reminiscent of that used by Ostriker 1971 to visualise pulsar radio data[2]. Each repeat measured around 25,000 particles, which were then combined in figure 3.13:B to form a single histogram containing 224,000 particles. They x axis on both plots shows the equivalent spherical diameter, which is the diameter a sphere would have to be to have the same volume as the measured particle. The vertical green line in the subfigure B indicates the manufacturers stated diameter of $1.5\mu m$. As we can see there is about a 10% difference between the stated and measured diameters of the beads. There are in principle a number of possible reasons for this. Firstly the stated particle size quoted by the manufacturer was almost certainly not measured using a Coulter counter. Differences in technique may produce small difference in quoted results. Secondly, although the plot shows diameter, it is actually volume which is being measured. If the particles are not perfectly spherical then there will be a difference between the average geometric diameter and the equivalent spherical diameter. More likely though, the result can be explained by imperfect callibration of the instrument. Calibration is normally performed by the manufacturer for each individual aperture tube before they are sold. However over time erosion of the aperture can cause drift in the relationship between detected voltage and interpreted volume. Although it is possible to re-calibrate tubes based on standard samples, as we are predominately interested in relative changes rather than absolute ones, I will simply note that my measured value seems to underestimate the actual size of particles by approximately 10%.

The next sample I examined with the Coulter counter was YD133 +pWR20 cells grown in M63 media with glucose and cas amino acids. As I had already measured the size distribution of this condition using microscopy, it allowed a comparison to be made between the two methods. In figure 3.14 I show two cell volume distributions for the same condition. One was generated by analysing 315 cell using microscopy,

[2]This style of plotting histograms will be used frequently in this section. Technically they are called ridge line plots, however the term joy plots seems to have gained favour recently as a result of their appearance on the cover art of Joy Division's 1971 album: Unknown Pleasures.

and the other from Coulter counter measurements of 127,000 cells. As we can see from the figure, there are a number of differences between the two distributions. First, the mean for the Coulter counter data is about 50% larger than that of the microscopy data. There are a number of reasons why this might be the case. Firstly, due to the difficulty in determining the exact position of the cell edge from standard light microscopy, segmentation tends to err on the side of being too aggressive, and as a result, tends to underestimate the length and width of cells. As the cell volume scales as the cube of the width, a small underestimate of the width would lead to a larger underestimate of the volume. Secondly, one might expect their to be a degree of uncertainty accompanied from extracting a three dimensional parameter like volume from a two dimensional image of a cell. The success of this depends on the extent to which *E. coli* cells can be approximated as spherocylinders, which is itself difficult to measure, although one might expect this to have a smaller effect on the overall volume.

The second key way in which the two distributions differ is in their spread and shape. The red and orange lines in figure 3.14:A show Gaussian fits to the data. As we can see the Coulter counter data has an odd dual peak shape, which is absent from the microscopy data. If we consider the standard deviation of just the first peak then the microscopy data appears wider, however, if the standard deviation is taken for the entire distribution then the Coulter counter data appears wider as shown in figure 3.14:B. Again there are a number of possible explanations for this discrepancy in distribution shape and spread. Most obviously the Coulter counter data has a distribution composed of far more cells than than that of the microscopy data. Therefore it is possible that had I analysed thousands more cells in the microscopy data, I would recovered a distribution that looked more similar in shape. Another explanation is that the Coulter counter distribution is composed from a single sample, where as the microscopy data is composed of multiple biological repeats each taken at a slightly different OD_{600}. If the distribution changes with time across the growth curve, then the shape of the distribution might be lost by pooling data from multiple time points. While one might be initially concerned to see such differences in a measurement of a single parameter by two independent methods, I will for now move on under the assumption that, once again, our interests lie in relative changes and not in the absolute value (or shape) of the cell volume distributions.

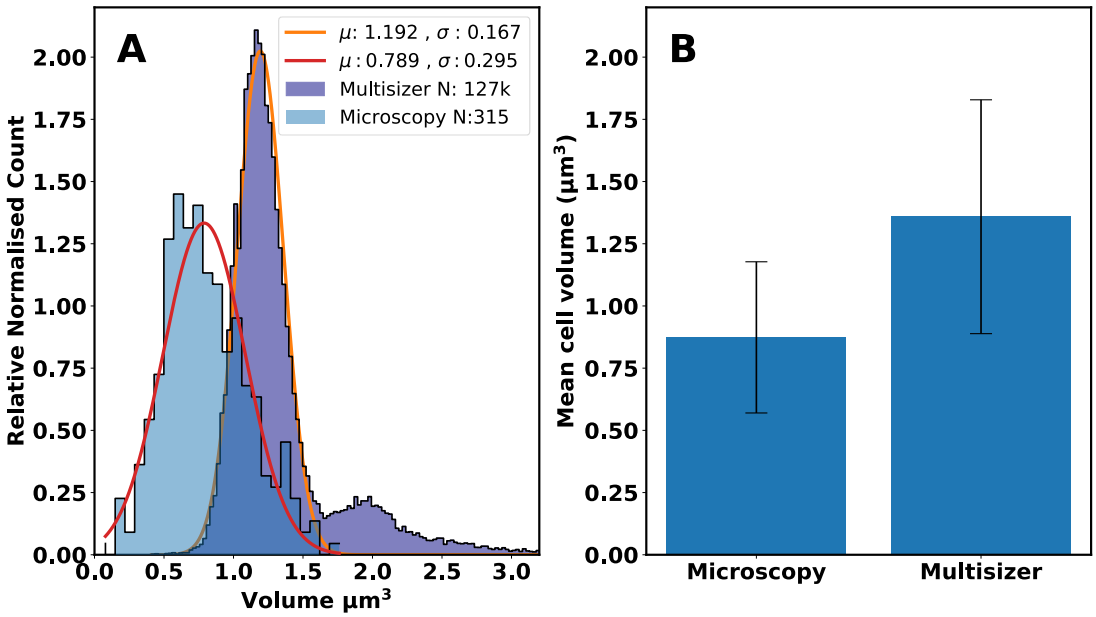

Figure 3.14: Size distributions as measured by microscopy and Coulter counter data. **A:** Normalsied volume histograms of cells grown in M63 media with glucose and cas amino acids as measured by microscopy and Coulter counter methods. The red and orange lines represent best fit Gaussian functions, the mean and standard deviations of which are shown in the figure legend. **B:** Comparison of mean volume between the two methods. The mean and standard deviation of the whole distributions are plotted, not just the values from the fits in subfigure A.

3.2.2.2 The nature of the bump

In the last section we noticed that the size distribution for YD133 + pWR20 cells grown in M63 with glucose and cas amino acids had this characteristic, and surprising, dual peak shape. One might at first glance think that the second peak (or bump) was simply an observation of doublets. Doublets, where multiple cells are stuck together, for example following division, would lead to a second population of cells at a volume approximately twice the population of the first. However in this case, the second peak is at a volume significantly lower than twice the value of the first. This would suggest that doublets are not responsible. Additionally as the Coulter counter software attempts to identify and remove doublets from analysis, I did not suspect that doublets were the cause.

An alternate explanation would be that the second smaller bump is caused by a

population of cells which have either left steady state early (with an accompanying change in volume) or have yet to reach steady state (and have thus not reached the equilibrium volume). To test this hypothesis I grew cells at two different dilutions from frozen stock: 10^5 and 10^7. The idea being, that cells grown for more generations would be more likely to be in steady state while the culture was at a low OD_{600}. In figure 3.15 I plot volume distributions for 10^5 and 10^7 dilutions, at both a low OD_{600} of around 0.05 and at a higher OD_{600} of around 0.2. As the plot shows, in all cases the bump remains, although it appears to change slightly in shape between low and high OD_{600}. I concluded based on the absence of any change with dilution, that the cause was unlikely to be a departure from, or late entry into, steady state.

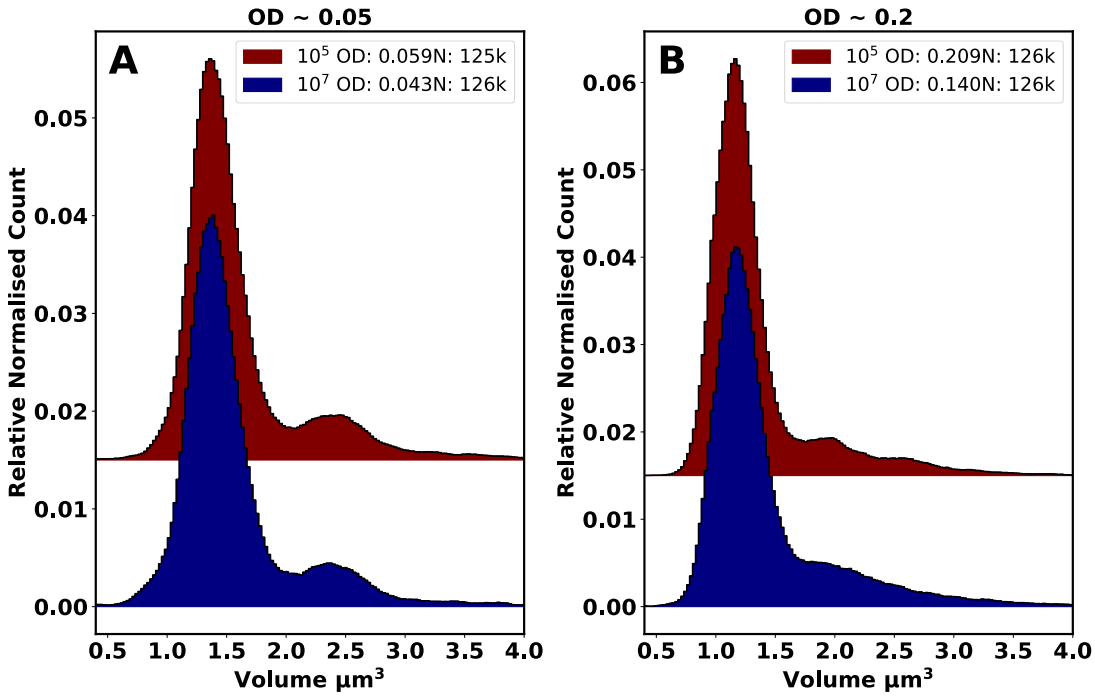

Figure 3.15: Cell volume distributions at different dilutions. **A:** Cell volume distributions measured via Coulter counter for YD133 + pWR20 cells grown in M63 media with glucose and cas amino acids. The blue curve is from cells grown at a 10^5 dilution from frozen stock while the red is from cells grown at a 10^7 dilution. The OD_{600} each sample was harvested at is shown in the figure legend. **B:** Same plot as subfigure A but cells were harvested at a higher OD_{600}.

After failing to find an obvious cause for the bump in the size distribution plots, I decided to investigate if the shape of the distribution was strain dependent. To test

this I measured the volume distribution of cells from four different strains including: YD133 + pwR20, YD133, BW25113 and MG1655. In each case, cells were harvested between an OD_{600} of 0.08 and 0.2. As we can see from figure 3.16, all BW25113 derived strains had some bump present, while MG1655 had essentially no bump. To better exemplify the difference I fitted a Gaussian to each trace. As can be seen from the figure, for MG1655 there were very few cells which sit outside the Gaussian fit while for other strains there is significant portion of the population that lies outside the curve. I therefore concluded that the bump was a strain dependent feature, most likely linked to the age distribution of cells. However, as there are very few mutations separating MG1655 and BW25113 (Baba et al. 2006), it is quite surprising that I observe such a dramatic difference. Finally, as an added bonus of this comparison we can see that the mutations that separate YD133 and BW25113, and the addition of the pWR20 plasmid have very little effect on the cell size distribution.

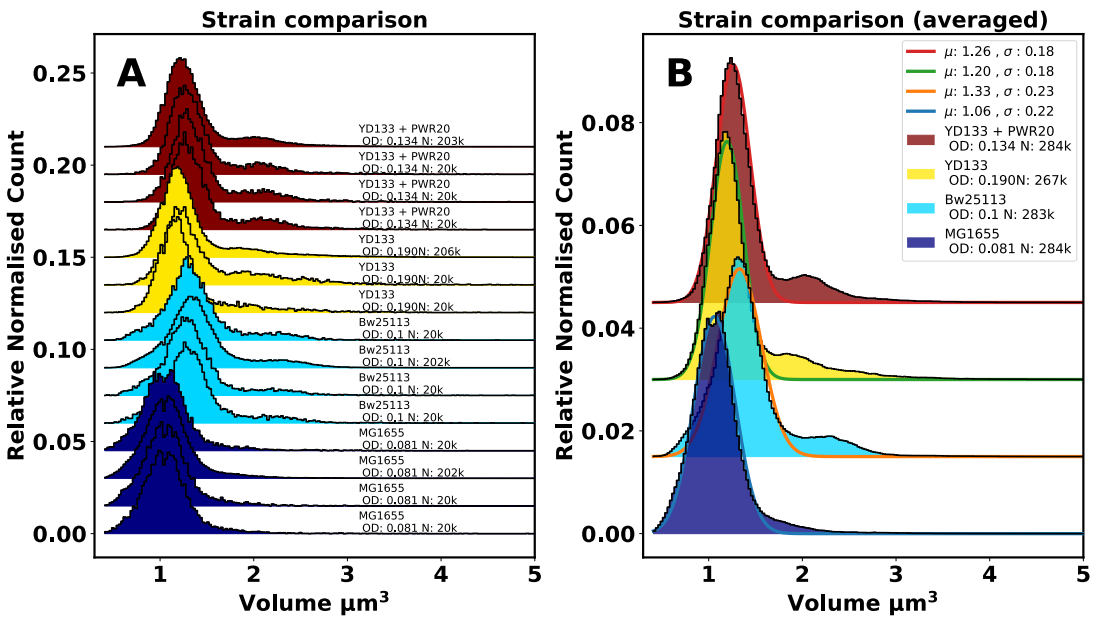

Figure 3.16: Cell size distribution for different *E. coli* strains. **A:** Cell volume distributions for a number of *E. coli* strains. Each colour represents a different strain with the plot showing a number of repeats for each strain. **B:** Distributions from subfigure A combining all repeats into a single trace. Gaussian fits are shown as coloured lines with their mean and standard deviations shown in the figure legend.

3.2.2.3 Cell size along the growth curve

As previously stated, one of the key advantages of the Coulter counter as a method for measuring cell size, is that measurements can be taken very quickly. This allows the cell size distribution to be followed across the growth curve, just as OD_{600} measurements allow the following of cell number across the growth curve. We have seen already in figure 3.15, that the distribution does appear to change with OD_{600} and therefore, it seems pertinent to explore this relationship further. To do this I grew cells in M63 with glucose and cas amino acids as normal. Then, at approximately twenty minute intervals, I removed 1ml of culture, from which the OD_{600} value was measured. Immediately following this, the 1ml smaple was then measured in the Coulter counter to obtain the size distribution. As samples often needed to be diluted before they could be measured in the Coulter counter, there was a concern that as cells adjust to the fresh media their size distribution could change. To ensure this was not the case, each measurement was broken up into four separate runs spaced approximately one minute apart. This allowed me to confirm that there was no change in the size distribution over the time it took to measure the sample. Figure 3.17 shows how the size distribution changes over a wide range of OD_{600} values in a number of different ways. Figures 3.17:A and 3.17:B show the raw and averaged cell volume distributions while figure 3.17:C shows a box plot for each OD_{600} at which a measurement was taken. Figure 3.17:D shows how the mean cell volume changes with OD_{600}.

As the plot shows, the size distribution does not change below 0.1 OD_{600}. However, at around 0.1 there is a significant reduction in average cell size which is approximately one standard deviation in magnitude. Following this dip, the size again holds constant till around OD_{600} 1.0 after which it starts to decrease monotonically with increasing OD_{600}. This is both surprising and concerning. The large drop in size at around OD_{600} 0.1 surely indicates a departure from steady state growth, as it is not possible for intensive parameters such as cell size to change and the population to remain in steady state. It is also interesting because of the apparent sharpness of the fall off. One might have expected a more gradual dip in cell volume to accompany a departure from steady state. Beyond this initial drop, the continued decrease of size as OD_{600} increases in consistent with repeated observation that cell size decreases as cells enter stationary phase.

The observed distributions in cell size, and the relationship between OD_{600} and cell size, are very reproducible across repeats. In fact the data in figure 3.17 is actually composed of two days worth of experiments, each following a different culture's growth curve. If the data is interleaved by OD_{600} as it is in figure 3.17, it becomes impossible to distinguish which data points come from which day. This confirms that both the data, and OD_{600} dependence, are very reproducible across repeats.

3.2.2.4 Cell size at high salt concentrations

With an understanding how cell size varies across the growth curve in the case of normal osmolarity, I next examined the effect of high sodium chloride concentrations. As with the microscopy measurements I looked at three concentrations of NaCl: 200mM, 400mM and 600mM. In each case I followed how the volume distribution changed across the growth curve at least between 0.1 and 1 OD_{600}. In figure 3.18, I show how the size distribution changes with OD_{600} for the case of 200mM NaCl.

There are a few things to note immediately from these distributions. Firstly the shape has drastically changed from the standard osmolarity condition. No longer do we observe two, approximately Gaussian, peaks. Instead we see a single peak which is skewed to the right. To exemplify this in figure 3.18:A, I have fitted a log normal distribution to each trace. As can be seen, a log normal appears to be a very good approximation of the distributions shape. To highlight changes to the distribution with OD_{600}, one would like to represent each measurement by a single value. Given that the distributions are log-normal in shape, and therefore not normally distributed, I opted to use the median (rather than the mean) to measure the central position of each distribution. Accordingly, to measure the spread of each distribution I chose to use the median absolute deviation (MAD) instead of the standard deviation. In figure 3.18:B I plot the median cell volume against OD_{600} with error bars indicating the MAD. From this we can see that there is an overall trend of lower cell volume at high OD_{600}, and in particular, a drop at around OD_{600} 0.1, although the drop is lower in magnitude in this case compared to the normal osmolarity condition.

Figure 3.19 shows the same kind of plot as figure 3.18 but for cells grown with

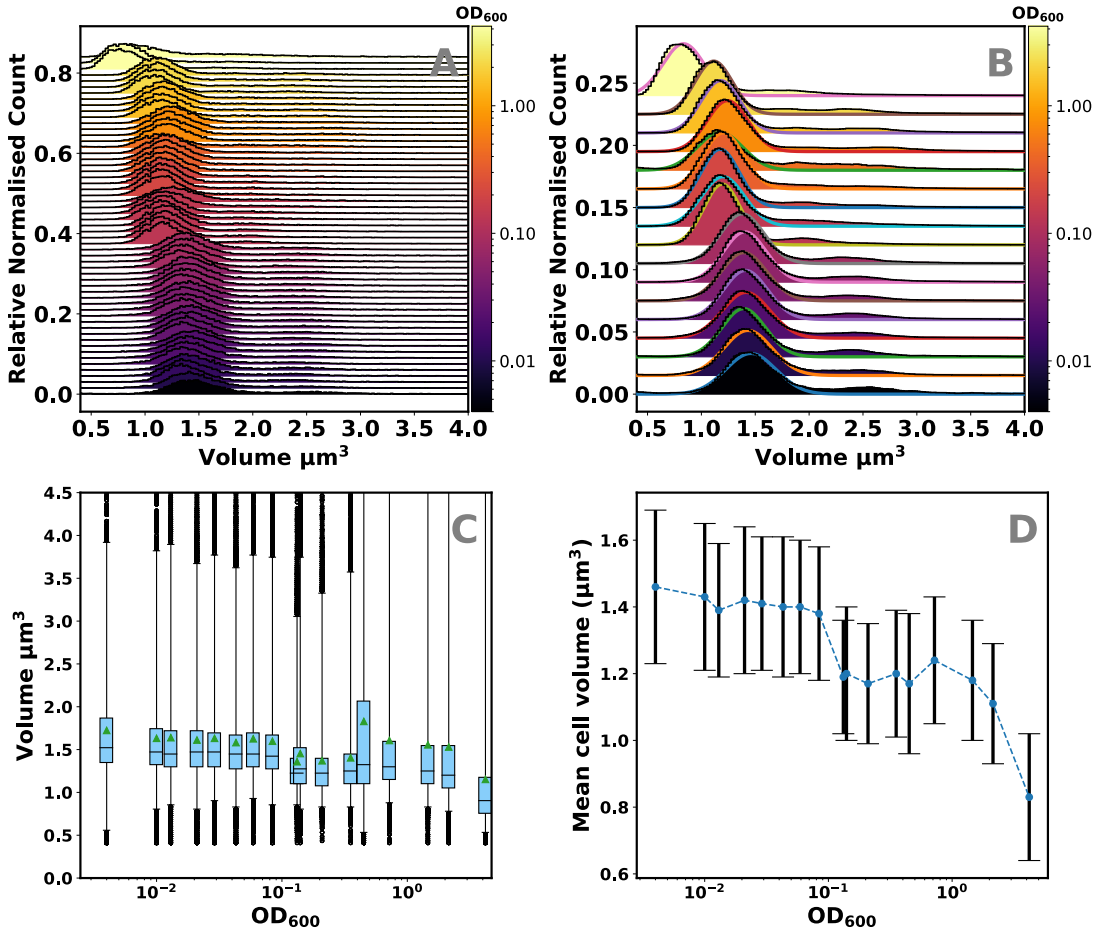

Figure 3.17: Cell size across the growth curve. **A:** Size distribution for cells grown in M63 media with glucose and cas amino acids. The colour of each plot indicates the OD_{600} value at which the measurement was made. Here, all repeats measured for each OD_{600} value are shown. **B:** Same data as in subfigure A but repeats are pooled together to form a single distribution. Additionally a Gaussian function is fit to each trace to show the position of the first peak, and to highlight that the shape stays approximately constant across the curve. **C:** Box plot summarising the data in subfigure B. Triangle represent means, and whiskers span 98% of the data. **D:** Mean and standard deviation of each size distribution in subfigure B plotted against OD_{600} on a semi-log plot.

added 400mM NaCl. Again we see a similar pattern with cells showing a log normal volume distribution. The same downward trend in median cell volume is present although this time there does not appear to be a downward step in volume at OD_{600} 0.1. Similarly in figure 3.20, I show the data for the 600mM NaCl condition, which

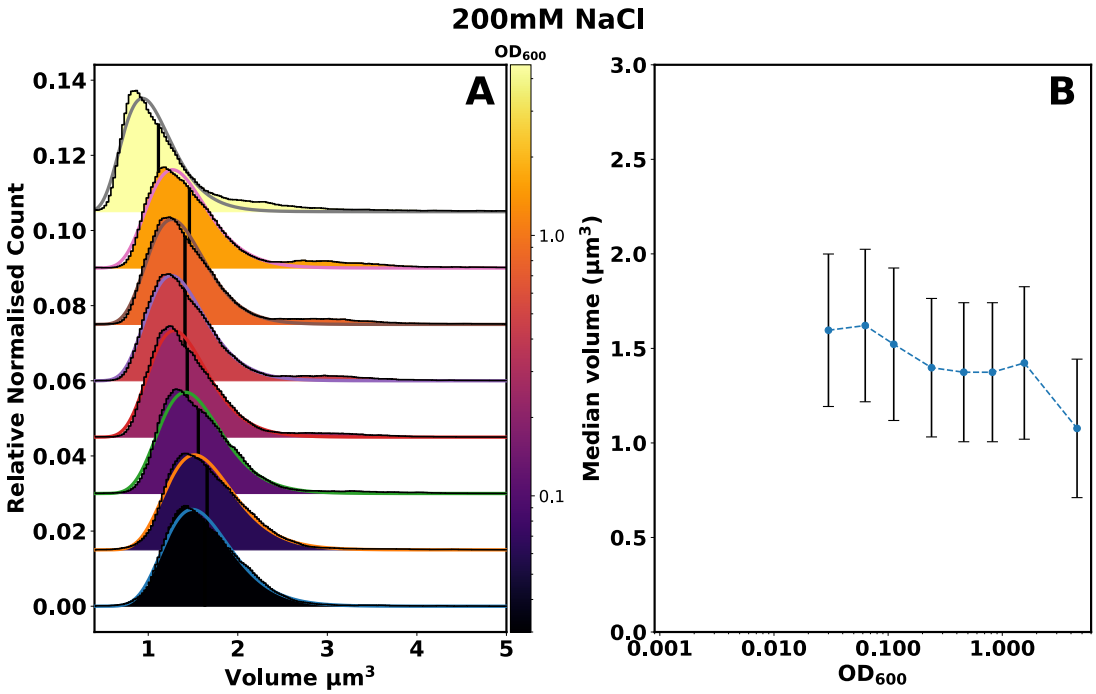

Figure 3.18: Cell size against OD_{600} at 200mM NaCl. **A:** Plot shows cell volume distributions averaged by repeats. Each distribution contains $\approx$ 125, 000 cells. Black lines indicated median volume, and solid coloureds lines are log-normal fits. Colour-bar indicates OD_{600} at which cells were harvested. **B:** Median cell volume against OD_{600}. Error bars represent median absolute deviations.

again shows a log-normal distribution and a general decrease with cell size at high OD_{600}.

To be able to gather a full picture of how cell volume changes at high osmolarity and along the growth curve, we need to combine the data we have seen so far in this section. Given the change in shape of the distribution at high salt concentrations, it is difficult to chose a single parameter to represent the centre or spread of each distribution. However, one must be chosen, and so I decided to take the median and MAD. The median is less sensitive to outliers than the mean, and given that none of the data seems particularly normally distributed, I felt it was a better choice. In figure 3.21A I plot the median cell volume against OD_{600} for all sodium chloride concentrations. We can see that as a result of the unfortunate step decrease in volume at around 0.1 OD_{600} for the normal osmolarity condition, we must restrict our analysis to comparing volumes at similar OD_{600} values. The triangles and square

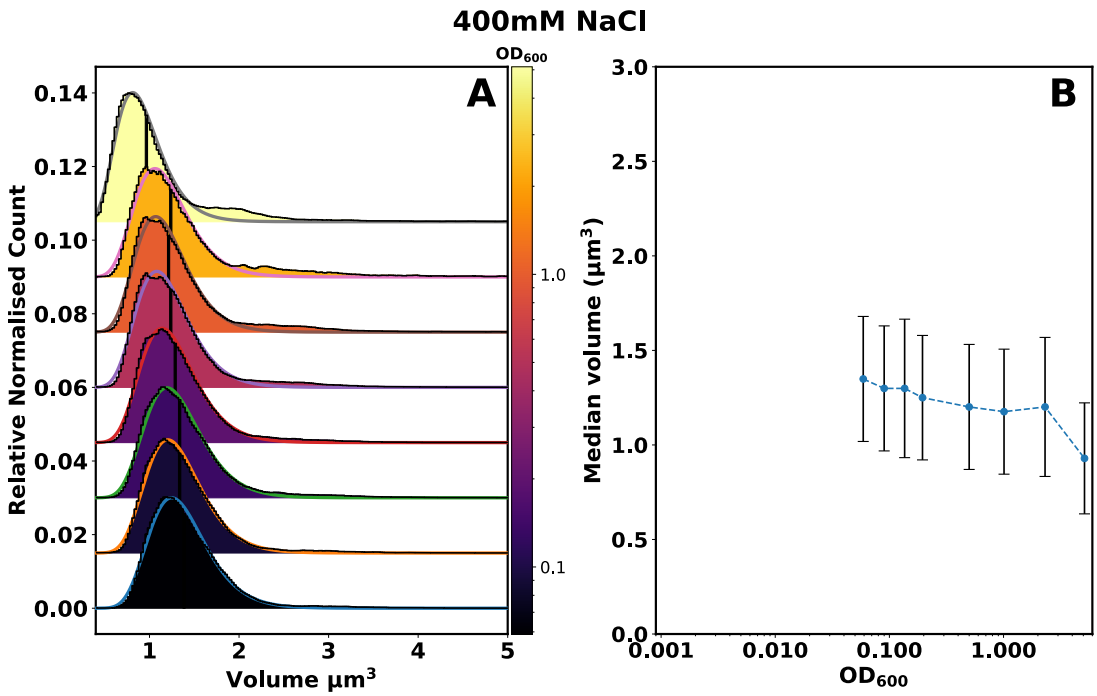

Figure 3.19: Cell size against OD_{600} at 400mM NaCl. Plot show data for cells grown in M63 media with: glucose, cas amino acids and 400mM added NaCl. For full figure details see caption of figure 3.18.

points in figure 3.21:A, show the measurements in each trace which are closest to a OD_{600} value of either 0.06 for the triangles, or 0.15 for the squares. The full volume distribution for each of these points are then plotted in figure 3.21 B and C.

There a few important observations we can make from figure 3.21. Firstly, we can say that the location along the growth curve can alter the perceived comparisons made between different conditions. For example, if OD_{600} values lower than 0.1 are considered one will conclude that cell size is significantly reduced at 400mM NaCl. However, if the same comparison is made at an OD_{600} of 0.2 then the conclusion will be that growth at 400mM NaCl has little effect on cell size. Examination of volume distributions at various locations along the growth curve yields, I argue, a more nuanced conclusions. Exposure to even low concentrations of NaCl alter the shape of the cell size distribution, causing it to become tailed towards high volumes, and therefore increasing the apparent cell size. Further increases to NaCl concentration then cause a monotonic (but small) decrease in cell volume from there on, at least over the range examined here.

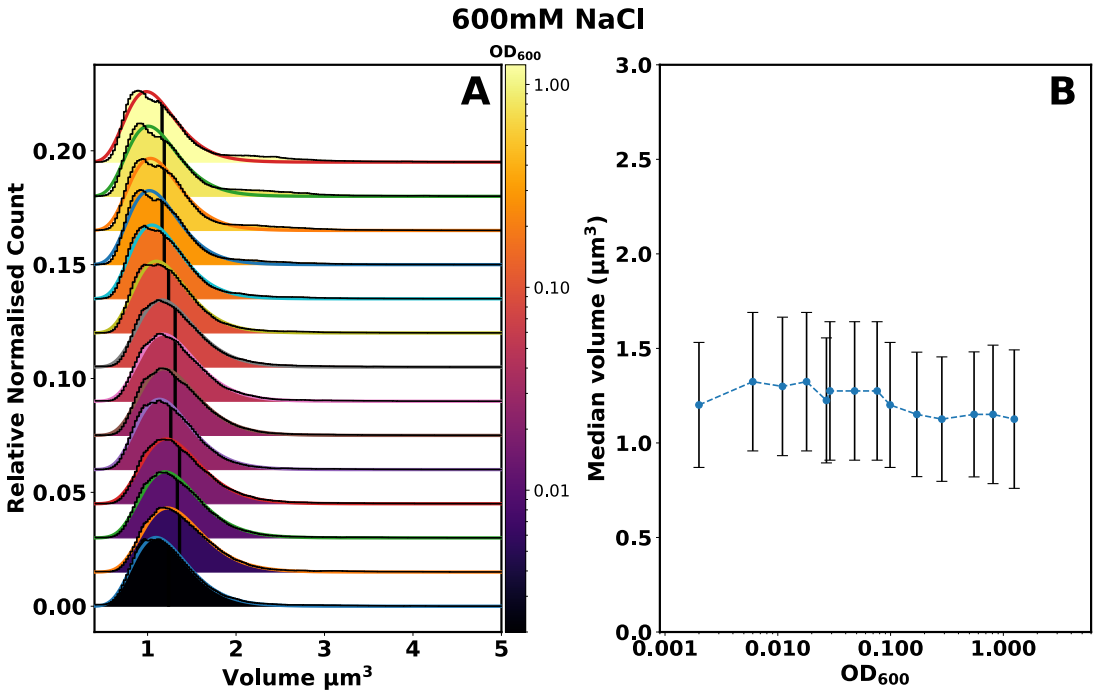

Figure 3.20: Cell size against OD_{600} at 600mM NaCl. Plot show data for cells grown in M63 media with: glucose, cas amino acids and 600mM added NaCl. For full figure details see caption of figure 3.18.

To determine how the results gathered here from the Coulter counter measurements compare to the microscopy measurements made in section 3.2.1.2, we can examine figure 3.22. In subfigure A I have plotted the median volume for each condition, as measured by microscopy (squares) and Coulter counter experiments (diamonds). As we saw in section 3.2.2.1, the microscopy data seems to consistently underestimate the cell volume compared to the Coulter counter data. Also in that section, I suggested that a reason for this underestimation could simply be overly aggressive segmentation algorithms, which give slightly narrower widths than the ground truth. Interestingly, if I normalise the data for both methods, such that the normal osmolarity condition has a volume of unity, I observe very similar patterns in both data sets, which can be seen in figure 3.22:B. Both methods show an initial increase in cell volume at 200mM NaCl, and a decrease in cell volume with increasing NaCl concentration from there. As previously discussed, I chose to use the median and MAD to summarise the distributions, due to the log normal shape of the distributions at high salt concentration. However in figure 3.22:C, I have also plotted the

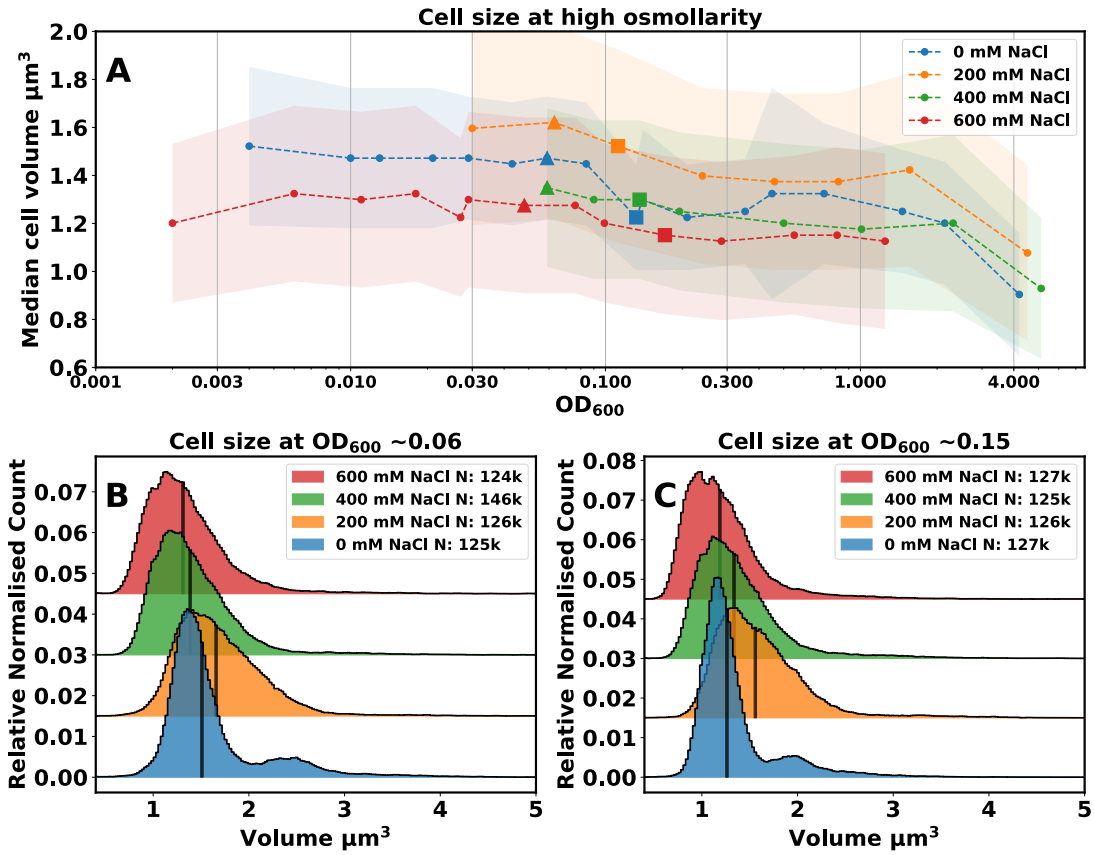

Figure 3.21: Cell size at high salt concentrations. **A:** Plot showing median cell volume changes with OD_{600} at different sodium chloride concentrations. Shaded areas represent median absolute deviations (MADs). **B:** Volume distributions for the triangular points in subfigure A. Black lines indicate medians. **C:** Volume distributions for square points in subfigure A.

mean, and standard error on the mean, for both the microscopy and Coulter counter data. To compute the mean cell volume for the Coulter counter data, I averaged the results for all available data points in the OD_{600} range 0.1 to 0.3. This is because that was the range of OD_{600} values at which cells were harvested for measurement in the microscopy experiments. I took the standard error on the mean of these data points to yield the error bars in figure 3.22. Again, if we normalise this data as has been done in figure 3.22:D, then we see both that there is good agreement between the data sets, and that the data looks very similar irrespective if it is the mean or median used to summarise it.

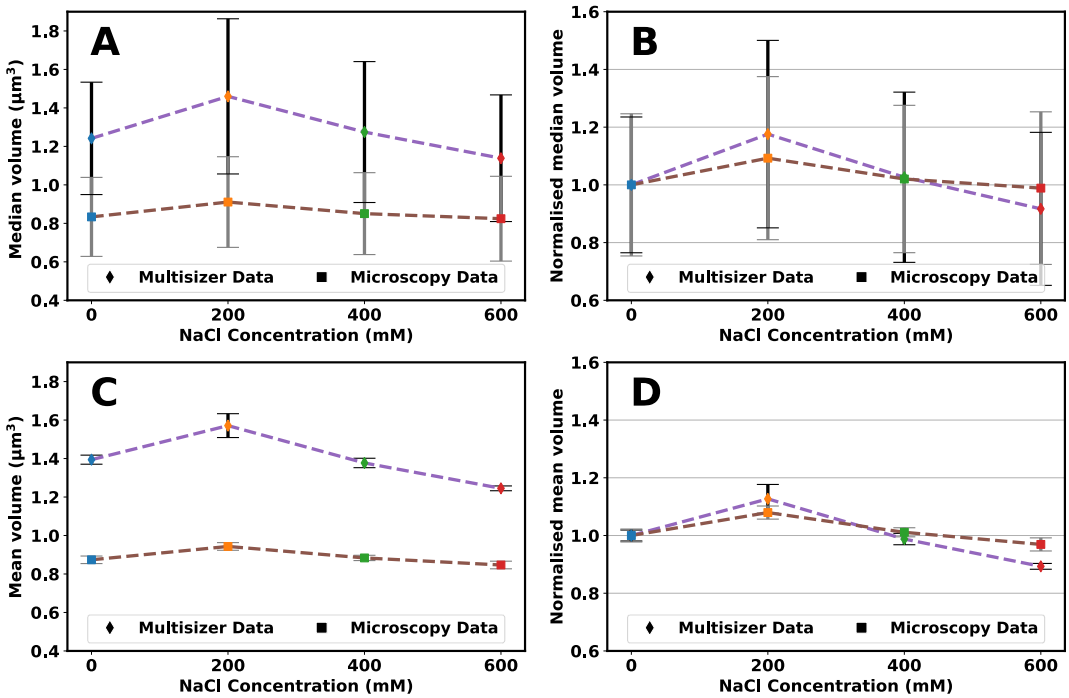

Figure 3.22: Cell size at high NaCl concentration and comparison with microscopy data. **A:** Median cell volume against NaCl concentration for both microscopy and Coulter counter data. Error bars indicate MADs. **B:** Median cell volumes normalised such that the 0mM NaCl condition has a volume of 1. Error bars as before are MADs. **C:** Mean cell volumes against NaCl concentration for both microscopy and Coulter counter data. Error bars are standard errors on the mean. **D:** Normalised mean cell volume against NaCl concentration for microscopy and Coulter counter data.

While the relative change in volume as measured by the two methods agree broadly, the Coulter counter data seems to show a larger increase in cell volume at 200 mM NaCl and a larger decrease in cell volume at 600mM NaCl. Given the tailed nature of the volume distributions at high salt concentrations, it is imaginable, that by virtue of sampling far fewer cells, the microscopy data might underestimate the mean (or median) volume. It is harder to imagine why the Coulter counter would yield a smaller average cell size at 600mM NaCl as a result of collecting far more cells. There may be two alternative explanations however. One might be the method by which I have averaged the data from the two experiments. The microscopy data is pooled from a number of repeats where cells were harvested at different OD_{600} values in the range 0.1 to 0.3. Likewise, the data for the Coulter counter was averaged from

the available data points in this range. However, if the cell size distribution varies across the range of OD_{600}s then the combined averaged value would be sensitive to the exact number of cells measured at the low end of the range, versus the high end of the range. Given that the distribution of OD_{600} values from which cells were taken is not constant between the two methods, this very well could explain the difference in average size. An alternate explanation could be that there was some bias in the selection of cells for the microscopy data. For example, it is very difficult to distinguish very small cells which are more spherical in shape, from cells standing on end. This could lead to omission of cells on the extreme left of the cell size distribution thus causing an artificially large mean volume. However, as overall both methods seem to show the same pattern in relative change in volume I am confident that I have correctly characterised the effect of high salt concentrations on cell size.

3.2.2.5 Cell size at high sorbitol concentrations

In section 3.2.1.2, I showed that there was some evidence that the volume response of *E. coli* cells to high NaCl concentrations was not identical to the response at high sorbitol concentrations. In this section, I will look at how that observation matches with Coulter counter measurements of cell volume at both 300mM and 600mM sorbitol. Cell were again grown in M63 media with glucose and cas amino acids, and the relevant concentration of sorbitol. The volume distribution was measured at a range of OD_{600} values across the growth curve.

In figure 3.23 I show the volume distribution for cells grown with 300mM of sorbitol. Interestingly the size distribution, unlike with NaCl, remains unchanged from the normal osmolarity case. The duel peaks that characterised the standard osmolarity condition are still present, and fitting a Gaussian to the main peak still yields a good result. The same holds true for the 600mM sorbitol condition, as shown in figure 3.24. This is a surprising finding as it seems to confirm the results from microscopy measurement, that osmolarity can have a differing effect on cell physiology depending on the nature of osmolyte used to obtain that high osmolarity. We observe here a significant change to the cell size distribution which presumably indicates a substantial effect on either the age distribution of cells or on their size regulation.

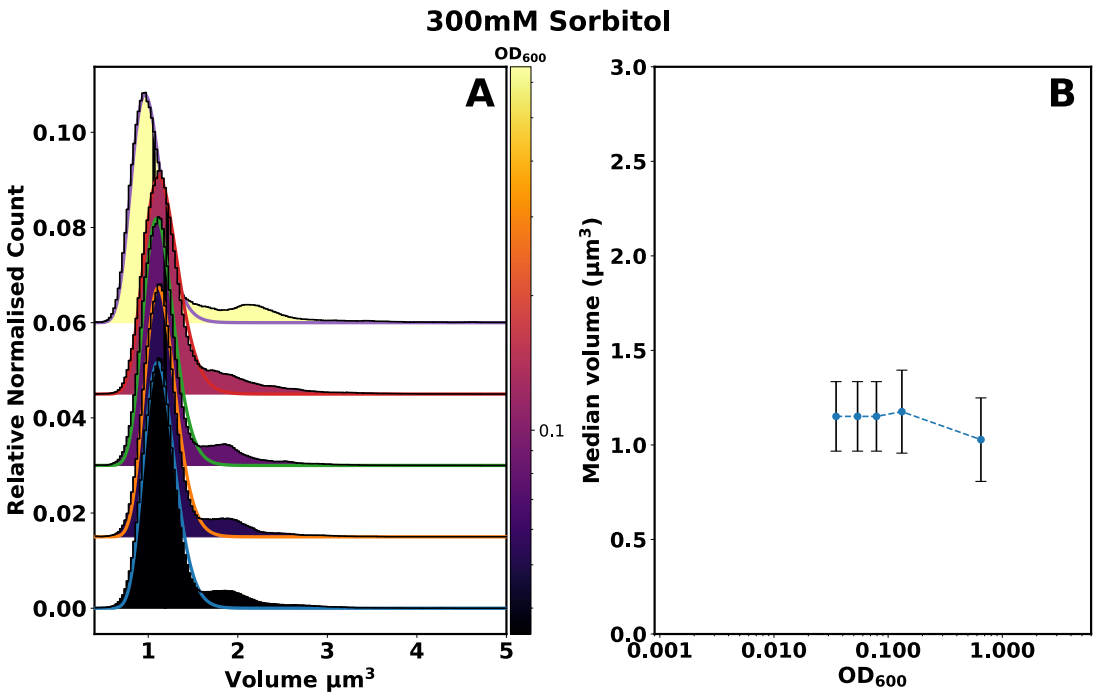

Figure 3.23: Cell size against OD_{600} at 300mM sorbitol. **A:** Plot shows cell volume distributions averaged by repeats. Each distribution contains ≈125, 000 cells. Black lines indicate median volume, and solid coloureds lines are Gaussian fits. Colour-bar indicates OD_{600} at which cells were harvested. **B:** Median cell volume against OD_{600}. Error bars represent median absolute deviations.

As before, to really inspect the effect of high sorbitol concentrations on cell volume, it is necessary to plot all of the data together on single plot. Even though the data in this case could be described by the position and standard deviation of the fitted Gaussian's, I chose instead for the sake of consistency to continue using medians and MADs as the chosen measure of distribution centre and spread. In figure 3.25, I plot a very similar figure to 3.21, but with the sorbitol data instead of sodium chloride. Again we break the analysis into two sections either side of the drop in cell volume that occurs at around OD_{600} 0.1 in the normal osmolarity condition. As before, there is a slight downward trend with cell size across the growth curve. There also appears to be no step change in size at OD_{600} 0.1, although it is harder to draw firm conclusions regarding that for this data set, due to the reduced number of time points. Across the curve it appears that size decreases with increasing sorbitol concentration. In fact there is a larger decrease than the corresponding concentration of sodium chloride, with a shift larger than two MADs in the case of 600mM sorbitol.

Although we do not have microscopy data for the 600mM sorbitol condition we can still compare the 300mM data measured via the Coulter counter to the micrsocopy measurements. If we do this in exactly the same way as was done in 3.22, we see again very good agreement. Additionally, there still appears to be no significant difference in pattern shown by analysing the median or mean of the distributions. I should note also, for the 300mM sorbitol condition, that as there was only a single data point taken by Coulter counter measurments in the OD_{600} range between 0.1 and 0.3, the error bars for that point in figure 3.26:C were therefore taken to be the same size as the 600mM Sorbitol case.

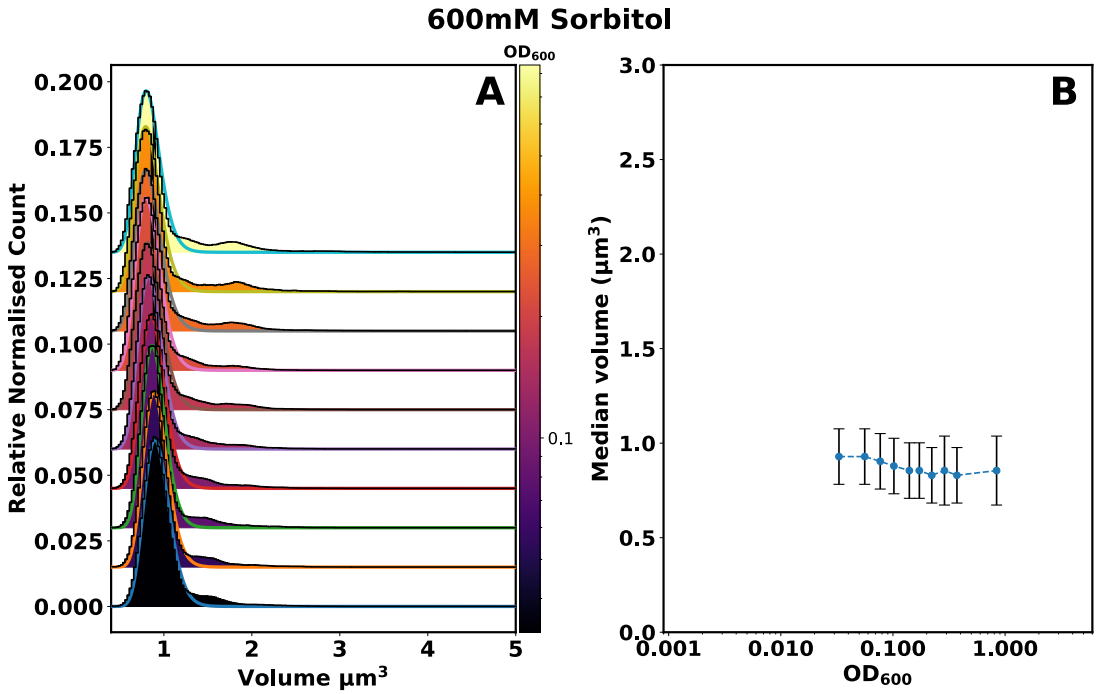

Figure 3.24: Cell size against OD_{600} at 600mM sorbitol. For more setails see caption of figrue 3.23.

3.2.2.6 Cell size at high osmolarity

As previously discussed, there is little recent data on steady state volume changes for *E. coli* cells grown at high osmolarity. None the less, I wanted to compare my results to some literature values and therefore we looked again at Dai and Zhu 2018.

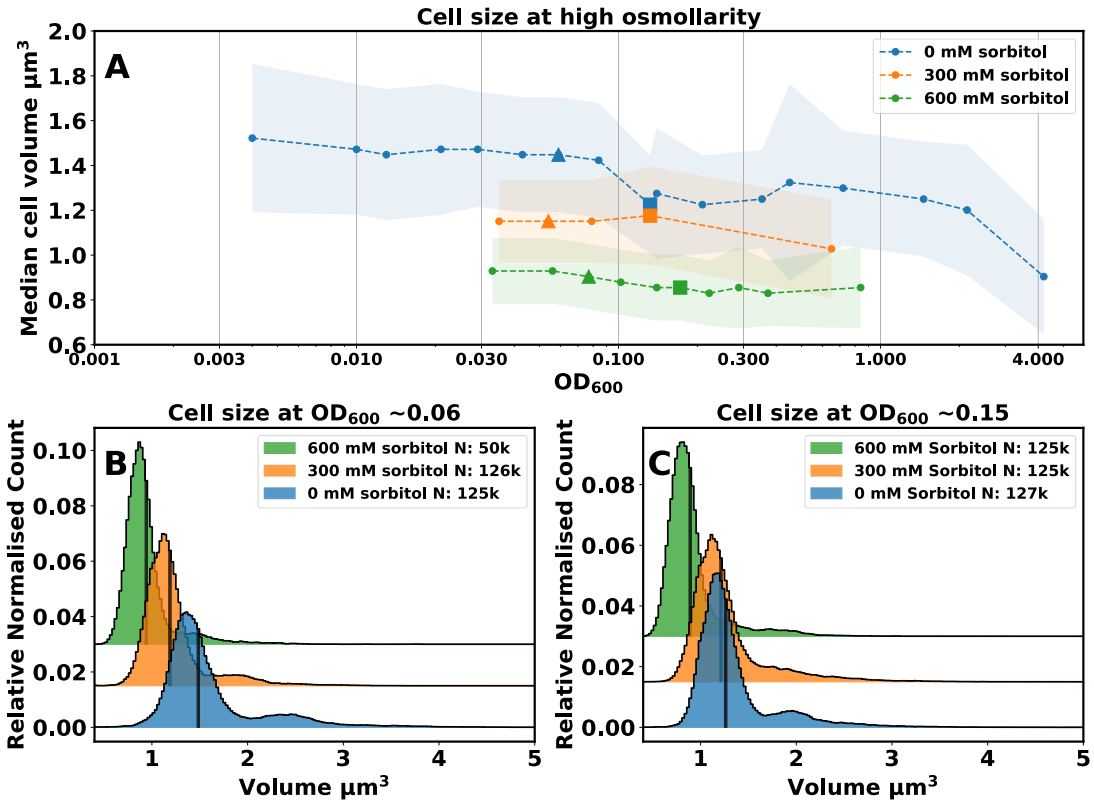

Figure 3.25: Cell size at high sorbitol concentrations. **A:** Plot showing median cell volume changes with OD_{600} at different sorbitol concentrations. Shaded areas represent median absolute deviations (MADs). **B:** Volume distributions for the triangular points in subfigure A. Black lines indicate medians. **C:** Volume distribution for square points in subfigure A.

Although the data in this case was measured in LB media I still felt it a useful comparison to examine if the overall trend in cell size matched between my result and theirs. In figure 3.27, I plot the absolute mean cell volume in both sorbitol and sodium chloride as measured by the Coulter counter. I then compare this to the results from Dai and Zhu 2018. To aid the comparison I normalise my data as before by taking the standard osmolarity condition to be 1. I then plot the data from Dai and Zhu 2018 on top. Unfortunately, as the data for points in the plot was not made available I had to estimate the values by reading off the graph in figure 3.27C. Still, given these limitation both data sets seem to show a similar downward trend in cell size with increasing sodium chloride concentration. However, the data from Dai and Zhu 2018 suggests a larger decrease than I observe with increasing

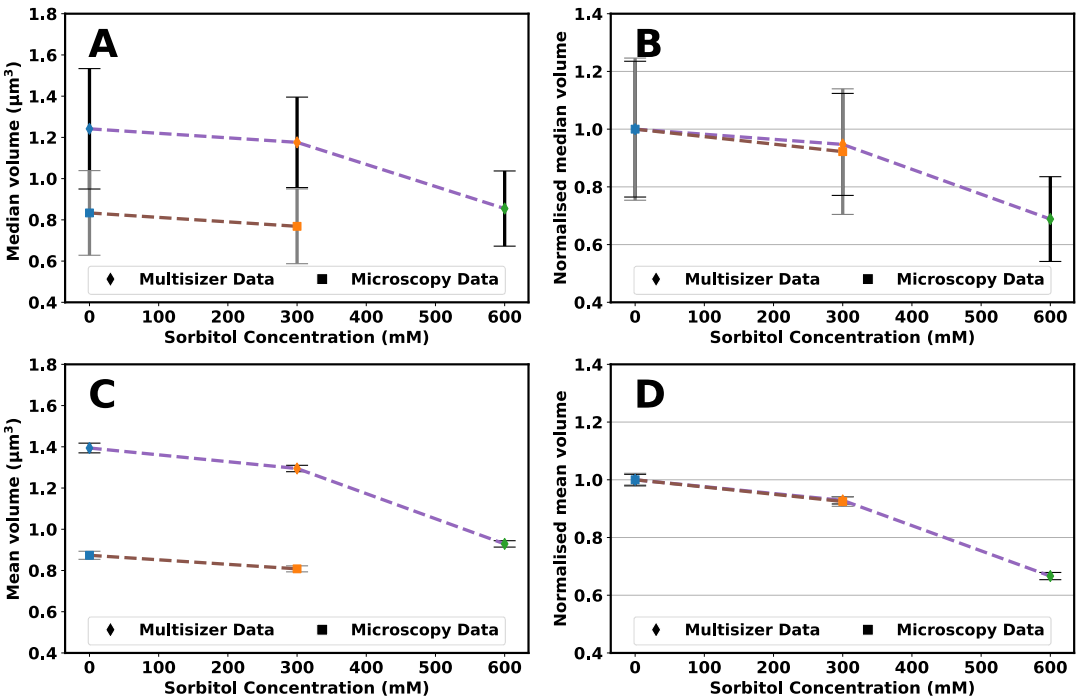

Figure 3.26: Cell size at high sorbitol concentration and comparison with microscopy data. **A:** Median cell volume against sorbitol concentration for both microscopy and Coulter counter data. Error bars indicate MADs. **B:** Median cell volumes normalised such that the 0mM sorbitol condition has a volume of 1. Error bars as before are MADs. **C:** Mean cell volumes against sorbitol concentration for both microscopy and Coulter Counter data. Error bars represent the standard error on the mean. **D:** Normalised mean cell volume against sorbitol concentration.

sodium chloride. Additionally, there is no suggestion that the size initially increases in their data, although they only examine sodium chloride concentrations starting at 400mM. Furthermore, given that LB has a significantly higher growth rate than M63 media with glucose and cas amino acids it may be reasonable to expect a different response to increased concentrations of osmolyte.

3.2.2.7 LB and high osmolarity

To investigate if the difference in the observed decrease in cell volume at high NaCl concentrations, between the results presented here, and those of Dai and Zhu 2018, were the result of the difference in media, I decided to perform a small number of experiments in LB media. I grew cells in two conditions: LB and LB + 600mM

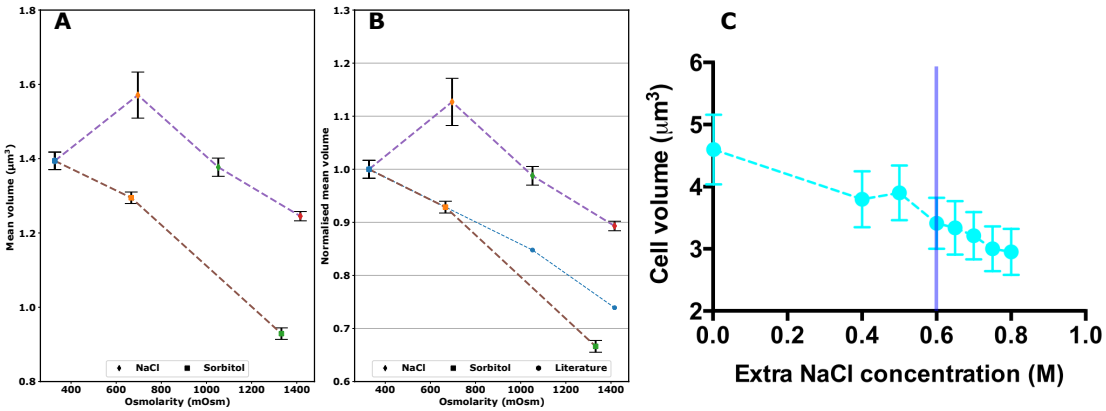

Figure 3.27: Cell size at high osmolarity. **A:** Mean cell volume against osmolarity for sodium chloride and sorbitol. Error bars are standard error on the mean. **B:** Normalised mean volume against osmolarity. Blue points indicate best estimate for data in subfigure C. **C:** Data from Dai and Zhu 2018 showing mean cell volume against NaCl concentration for *E. coli* cells grown in LB.

NaCl. As shown in 3.28 the LB condition looked very much as expected. The characteristic two peak distribution was present, and cells were larger in size compared to the M63, glucose and casmino acids condition. However, the LB + 600mM NaCl condition was a very different story. As can be seen in figure 3.28:B, the distribution of cell volumes in this condition was completely different. Rather than a peaked distribution, there appeared to be a significant proportion of cells at almost any size in the measured range. This was very odd, and initially I suspected a blockage of the Coulter counter to be the cause. However, after seeing the same result on different days, and after unblocking protocols had been carried out on the Coulter counter, I began to suspect that the measurement was in fact working as intended. To confirm this suspicion I imaged the sample under the microscope in both bright field and fluorescence. Some example images of cells grown in LB + 600mM NaCl are shown in figure 3.29.

From these microscopy images we can conclude that Coulter counter measurements were accurate. Cells in this condition show unusual morphology. Many cells show "chaining" where newborn cells, although apparently divided, stay attached to the mother, as well as subsequent daughter cells. Additionally filamentous cells are also common, where cells continue to grow but do not divide. Additionally, many cells show (what I assume to be) blebs: where cell membrane bulges outward form

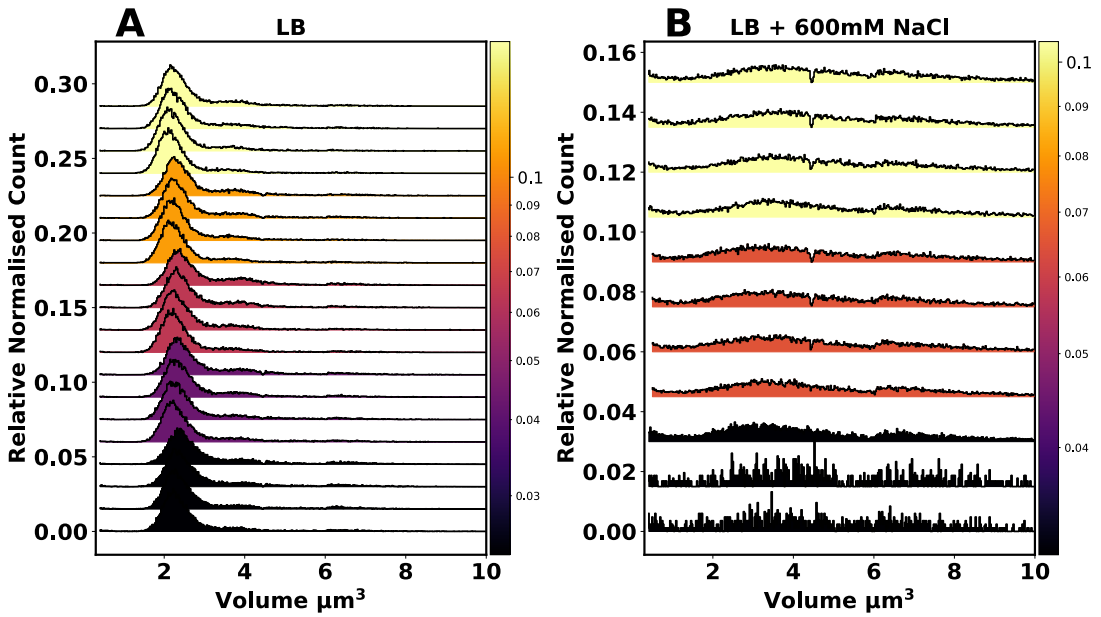

Figure 3.28: Cell volume distributions in LB media at high salt concentration. **A:** Volume distributions along the growth cuve for *E. coli* cells grown in LB media. **B:** Volume distributions for *E. coli* cells grown in LB media with 600mM NaCl.

the cell wall (Yao, Kahne, and Kishony 2012). Finally, as in figure 3.29:A, I also observe many cells with an apparent "branching" structure, where the cell forks and appears to grow from both sides of the fork. All of these observations are unusual and I have never observed them in other conditions. Despite the odd morphology, the OD_{600} curve for cells in this condition did not appear different from any other condition.

Upon initial observation of this strange morphology, I first suspected a contaminant of some kind to be responsible. I grew cells in a number of different ways to try to rule out this is as cause, checking in each case the cell morphology under the microscope. For example: I grew cells in plastic falcon tubes rather than glassware to eliminate the possibility that a contaminate present in the washing process was to blame. Next, I tried growing cells in pre-made LB media bought directly from the manufacturer rather than the standard media that was made by our lab on site from individual components. Neither of these experiments changed the strange morphology I observed. Next, I attempted to see if the morphology varied with strain. I grew BW25113 cells and MG1655 cells and found that both had exactly the same

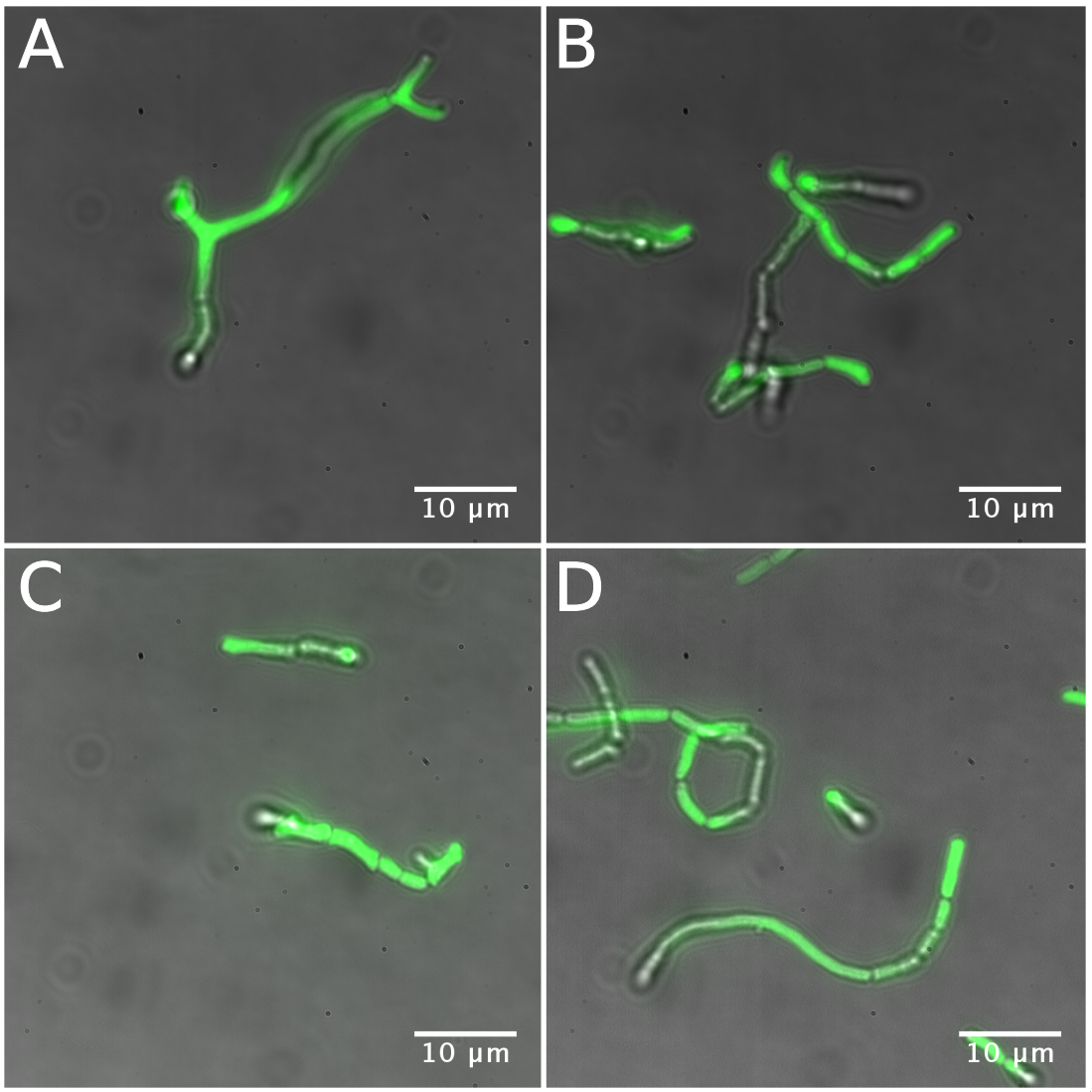

Figure 3.29: Cell morphology at 600mM NaCl. Examples showing some representative images of the unusual cell morphology cells exhibit when grown at 600mM NaCl in LB.

result as YD133 + pWR20. Following this, I examined if the observation was media, or growth rate dependent by growing cells in RDM with 600mM NaCl. I found that cell chaining was less prevalent but that blebbing and branching were still common, although it is of course difficult to quantify the degree of these phenomenon given the wide variety in morphology. To see if the effect was present at lower salt concentrations I grew cells at 400mM NaCl. In this condition I found basically no change to the cell physiology, compared to the 600mM NaCl condition. Interestingly when I grew cells in LB with 600mM sucrose as way of comparison, I found that cell mor-

phology was not notibly perturbed from the normal state.

Taken in their entirety these results suggest that cell size regulation is strongly perturbed at high salt concentrations and high growth rates in a strain independent manner. While exploratory, these results suggest that the effect osmolarity has on cells is both dependent on the osmolyte used, and on the nutrient conditions of the base media to which the osmolytes are added. Why such physiology has not been previously reported especially, by Dai and Zhu 2018, is an open question that does not have an immediate obvious answer.

3.2.3 Discussion and future work

In the previous section, I have sought to answer the question of how the average cell size changes with osmolarity, in a population of exponentially growing *E. coli*. To answer this I used a combination of fluorescent microscopy and image analysis, as well as direct volume measurement, using the Coulter counter. In this section, I will discuss the main key findings of this work, as well as what could be done in the future to build on them. In particular I will focus on: the comparison between the two methods; the effect of high sodium chloride concentration; what the size distributions along the growth curve tell us; and finally the differences between different osmolytes.

The first method I used to measure cell size was fluorescent microscopy. As described in section 2.4, I created a custom pipeline for the manual selection of cells, followed by segmentation using a neural network, and finally extraction of cell length and width using a custom algorithm. To validate this complex pipeline, in section 3.2.1.1, I recreated the nutrient growth law by examining cell size in three different nutrient conditions. I found that I was able to reproduce the exponential dependence of cell size on growth rate, as previously measured by other researchers, and as described by equation 3.1 (Si et al. 2017). Given the difference in experimental methodology, strain, and analysis, it is reassuring that I was able to so closely match previous work, and it proves a strong source of validation for my analysis methods. Specifically, it validates my measurement of relative changes in volume, with growth rate. It is more difficult, to compare direct absolute volume measurements, given these difference in media and strain. We can however, get some idea of how the

absolute volume compares with literature values, by comparing the value of S_0 in equation 3.1, between my results and literature. In this case S_0 represents the cell size at zero growth rate, and thus gives an indication of the absolute measurement of cell volume. Although my measured value of S_0 is within error of the value obtained in Si et al. 2017, it is still around 12% lower than the literature value. This may indicate, that I am comparatively underestimating cell volume. In a similar vein, in section 3.2.1.2, I plotted the surface area to volume ratio of the three different nutrient conditions, and showed that again one can retrieve an exponential dependence of this quantity on growth rate. The mathematical expression for this relationship, as stated in equation 3.3, showed good agreement with literature values (Harris and Theriot 2018). This provided a verification, that my independent measurements of relative changes in length and width, were in good agreement with previous work. All of this is to say: that I am confident that I am accurately measuring changes in cell size, and am probably slightly underestimating absolute cell volume.

To provide an alternate validation of my cell volume determination, in section 3.2.2.1, I compare results from micrsocopy measurements, with those obtained from the Coulter counter. In agreement with the discussion above, I find that the microscopy data measures a lower cell volume than the Coulter counter data. I also suggest in that section, that it is difficult to draw firm conclusions about the different shapes of the cell size distributions, as computed by the two methods, due to the comparatively small number of cells measured using microscopy. Furthermore, I suggest that the necessary step of selecting flat cells for the micrsocopy data, adds an extra source of possible bias, due to the human error associated with this task. This step in the process also limits throughput, leading to the reduced number of cells available for the cell size distribution. Therefore, going forward I think it would be of great benefit to develop an automated method to discriminate flat cells, from not flat cells. This may be achievable using a machine learning approach similar to the one presented here. By first segmenting images using a network like U-net, and then passing the segmented regions to a second network for classification, as “flat” or “non flat” cells, it may be possible to determine which cells should be considered for analysis, in an automated way. This would greatly increase the throughput for cell size measurement using the tunnel slide assay, and as a result, a better comparison could be drawn between the two methods of measuring cell size.

One of the main goals of section 3.2 was to determine how cell size changes with high NaCl concentration. In section 3.2.2.4, I show that using both microscopy measurments and the Coulter counter, a complex relationship between NaCl concentration and cell size is revealed. At low concentrations of NaCl (200mM) the cell size initially increases. This is followed by a monotonic decrease in cell size at higher NaCl concentrations. The overall reduction in cell size is also significantly lower than that seen in Dai and Zhu 2018, for the same concentration of sodium chloride (figure 3.27). This discrepancy could be explained by a difference in the measurement technique, but far more likely it indicates that the effect on cell size of high salt concentrations is dependent on growth rate of the base media. Ideally I would have been able to measure cell size in LB media with added NaCl, to confirm this hypothesis. However, as detailed in section 3.2.2.7, I was not able to grow cells in LB at high salt concentrations without a severe effect on cell morphology, as shown in figure 3.29. The fact that this staggering change to cell behaviour was not reported by Dai and Zhu 2018 is quite surprising. To my knowledge, there are no other reports where *E. coli* cell size is measured in steady state, using LB media, and at high amounts of sodium chloride. However, given that I observed this effect using different strains, in both RDM and LB, and only when using NaCl as the osmolyte, it is hard to imagine how this could be an artefact, rather than a genuine result. Indeed, the kinds of morphology that I observe in figure 3.29, are not often found in *E. coli*. For example, chaining of *E. coli* cells has been observed previously, but only during biofilm formation (Vejborg and Klemm 2009). Filamentation on the other hand, is not uncommon, but can have a number of different causes. These include the DNA damage induced SOS response, or inhibition of the septum formation by antibiotics such as cephalexin (Schoemaker, Gayda, and Markovitz 1984; Chung et al. 2009). Additionally, in figure 3.29, cells also showed signs of blebbing similar to that previously described under exposure to beta-lactam antibiotics (Yao, Kahne, and Kishony 2012). To see all of these apparent affects simultaneously is very unusual, and clearly more work would have to be done to gain a concrete idea of what is causing this strange morphological state. That being said, we can speculate on possible explanations. Given the combination of blebbing, filamentation, and the effect only being present at high growth rates, it is possible that at high osmolarity, inclusion of peptidoglycan into the cell wall becomes more difficult. This could lead

to cell wall thinning, allowing the formation of blebs, and preventing the rapid incorporation of peptidoglycan into the hemispherical end caps in the cell wall, which are required for division (Harris and Theriot 2016). These affects would be more pronounced at higher growth rates where cells already have a lower surface area to volume ratio, and have less time to double the amount of cell wall material between birth and division. Investigating this hypothesis would be an interesting direction for future work.

As mentioned previously, one of the great advantages of the Coulter counter is the ability to measure cell size at various points across the growth curve. In section 3.2.2.3, I showed that for cells growing in M63 media with glucose and cas amino acids, there was a significant drop in average cell size at around OD_{600} 0.1. This drop is important for two reasons. Firstly it indicates that size comparisons must be done carefully, ensuring samples only of the same OD_{600} are in general compared. Additionally, as previously stated, it indicates departure from steady state growth. It is a little surprising that cells leave steady state so early on in the growth curve, as typically OD_{600} values of around 0.1-0.5 are considered a safe range for "steady state" (Dai, Zhu, et al. 2018; Dai and Zhu 2018; Moselio Schaechter, Maaløe, and Kjeldgaard 1958). It is possible that this early departure is a rather unique property of the M63 media with glucose and cas amino acids. However, it is interesting to notice that from the graphs of OD_{600} against time in figure 3.1, there is no obvious change in the growth rate over this range of OD_{600} values. As the presence of steady state is normally determined by examination of the growth rate, it may be possible that early departure is more common the we are aware. Previous work has found that departure from steady state is more obvious when the cell number per OD_{600} (effectively cell mass or size) is consulted, compared to the growth rate (Sezonov, Joseleau-Petit, and d'Ari 2007). This same work found that in LB media, steady state ends at around OD_{600} 0.3 for the *E. coli* strain MG1655, and even earlier for some other strains. Interestingly, Dai and Zhu 2018, use both MG1655 and LB media. They measure growth rate over a range of OD_{600} values between 0.05-0.5. They also measure cell size at an OD_{600} value of 0.3, and measure dry mass at an OD_{600} value of 0.4. This indicates that at least part of their results were measured most likely after departure from steady state. Given this, it is clear that going forward, it would be of great benefit to develop a standard protocol for determining the limits

of steady state, that is not based directly on growth rate. Given that in both my work and in the work of Sezonov, Joseleau-Petit, and d'Ari 2007, it was effectively cell size that suggested departure from steady state, this would be a good place to start. It would also be good going forward, to determine the reason why cells in M63 media with glucose and cas amino acids depart steady state at an OD_{600} value of 0.1, for example by determining if it is caused by a lack of nutrients, and if so which ones.

In section 3.2.2.5, I compared the cell size distributions for cells grown with high concentrations of sorbitol as apposed to NaCl. I found that strikingly, the shape of the cell size distribution was not altered by the addition of sorbitol. This is in stark contrast to the NaCl case, which had a large effect on the cell size distribution's shape. Furthermore, even the lowest concentration of sorbitol tested (300mM) reduced the average cell size almost as much as the highest NaCl condition examined (600mM). Taken together, these results suggest that osmolarity is not a unique identifier of cell physiology. This is an important result of this thesis. It is exceedingly common in the literature, for "high osmolarity" to be taken as any state with an increased concentration of solute, irrespective of what that solute is. For example NaCl, sorbitol and sucrose are all commonly referred to as high osmolarity conditions, by work where only that single osmolyte is examined (Dai and Zhu 2018; Dai, Zhu, et al. 2018; Pilizota and Shaevitz 2014; Rojas, Theriot, and K. C. Huang 2014b). The reason for this is in someways understandable, after all, the osmotic pressure as defined in equation 1.22, has no dependency on the type of solute used, only on the difference in osmolarity between the cell and its environment Therefore, if osmotic pressure is the key variable affected at high osmolarity, one would expect no difference between different osmolytes. On the other hand, previous work has shown that different osmolytes can cause varying degrees of plasmolysis during hyperosmotic shock, depending on their level of outer membrane permeability (Pilizota and Shaevitz 2013). It may not be so strange then, that osmolytes have different effects on cell physiology in the steady state. Going forward, it would be of great interest to further compare the relationship between osmolarity, growth rate, and cell size, for the three most common osmolytes found in the literature (NaCl, sucrose and sorbiotl). This would allows us to better understand the applicability of the the term "high osmolarity", and determine if instead, each unique osmolyte should be considered independently. It may be, for example, that the results of these experi-

ments indicate that osmolyte charge is a relevant factor. Perhaps charged osmolytes have adverse effects on proton motive force, thus explaining the different behaviour between sorbitol and NaCl. However, much more work would need to be done to draw firm conclusions on such conjecture.

3.3 DNA content from flow cytometry

3.3.1 Mannose grown cells as a callibration standard

As discussed in section 2.6.2 one of the main challenges of using flow cytometry techniques to measure the DNA content of cells is correctly determining the relationship between fluorescent signal and quantity of DNA. In that section I showed that, using rainbow callibration particles, a power law relationship between the quantity of fluorophore and the measured fluorescent signal, can be observed. In this section I will show that using this observation, and cells grown in M63 media with mannose, the fluorescent signal can be linked to DNA content.

As the growth rate of cells decrease under nutrient limitation, the number of overlapping cell cycles decrease. Eventually, cells will initiate and complete replication in the same generation, if the doubling time is sufficiently low. Therefore, if we take cells under these conditions, and carry out a run out experiment with them, then the result should be a population of cells with either one or two chromosomes. Specifically I will make use of cells grown in Mannose, which have been shown previously to exhibit exactly this behaviour (Si et al. 2017) .This forms the basis of our calibration process, allowing us to determine the fluorescent signal associated with a single chromosome. However, before we can get to this stage we must first discuss the strategy used for gating the flow cytometry events.

Unlike with the Coulter counter, there are no automatic means of removing doublets with the flow cytometer used in this investigation. Instead we have to manually determine a method of removing likely doublets from the analysis. To do this I decided to gate based on the width and height of the recorded pulses in the propidium iodide (PPI) channel. In the case of single cells, we expect that an increase in fluorescent signal will lead to larger area pulses which have a taller height, but are not significantly wider. On the other hand, if two cells pass through very close to each other to form a doublet, then we expect a much wider pulse for a given pulse area. We can use this to eliminate most doublets from the analysis. Take for example figure 3.30:A which shows data from a run out experiment using mannose treated cells. The plot shows a two dimensional histogram with pulse width on the x axis and pulse area on the y axis. As the data is from a run out experiment, there are a

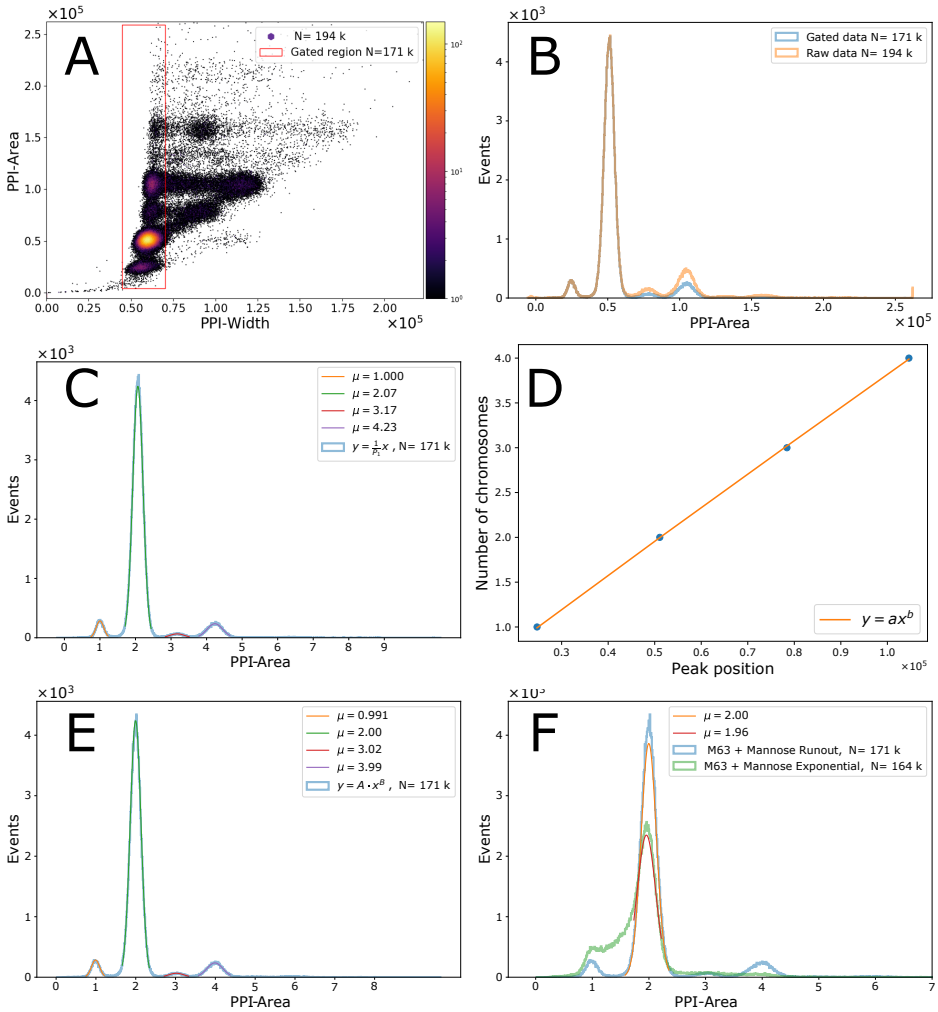

Figure 3.30: DNA content from cells grown in M63 media with mannose. **A:** Density plot showing frequency of cells plotted against the width of pulses in the PPI channel and the area of pulses in the PPI channel. Red box indicates gated area. **B:** PPI-Area against frequency for both the gated (blue) and raw (orange) data in subfigure A. **C:** Gated data from subfigure B rescaled such that the first peak has a mean of 1. Solid coloured lines indicate Gaussian fits to each of the peaks. The mean of these fits are shown in the figure legend. Note that they deviate from integer values considerably. **D:** Mean of peak position against number of chromosomes the peak represents. Orange line is power law fit to the data. **E:** Gated data rescaled by power law in subfigure D. Notice that peaks now lie very close to integer values as indicated by the Gaussian fits. **F:** Rescaled data for both run outs and exponential culture. Notice that the peaks in both data sets align very well.

number of discrete peaks which represent different populations of cells with integer numbers of chromosomes. By selecting only the peaks which are arranged vertically, such that they have different peak areas but the same peak widths, we can eliminate a large amount of doublets from the analysis. The gated area is indicated by the red box in figure 3.30:A, and contains approximately 171,000 events of the 194,000 recorded during that particular run. The effect of gating can be determined by examining figure 3.30:B. Here I plot the area of each pulse against frequency for both the gated and raw data in figure 3.30:A. As can be seen, gating has little effect on the lowest pulses but has substantially more effect for higher peaks. This make sense as the smallest doublet must be at least twice the lowest valued cells, so we expect no effect on the first peak. To clarify, there is only one combination of cells which can generate a doublet at twice the value of the minimum cell's intensity. Conversely, there can be multiple combinations of cells, which lead to the higher valued peaks, and therefore gating will reduce the number of events in these peaks more significantly.

With the data now gated, we can turn to the problem of relating fluorescent signal to DNA content. As mentioned, we expect the run outs from cells grown in mannose to have one or two chromosomes per cell. What we actually observe in figure 3.30:B are four peaks. I will address the possible reason for this shortly, but for now I will continue under the assumption that these peaks represent cells with one, two, three and four chromosomes respectively. In figure 3.30:C I have plotted the gated data again, with the addition of a fitted Gaussian to each of the four peaks. I have also rescaled the data by dividing the data by the mean of the first peak. The new mean of each fitted Gaussian is shown in the figure legend. As expected, performing a simple linear rescaling does not lead to each peak sitting at an integer value. Instead we have to resale using a power law, as was done in the case of the rainbow calibration particles. To do this, I plot the actual measured position of each peak (as obtained from the fitted Gaussians) against the number of chromosomes I believe it to represent. I then fit to this plot a power law function, and use this to rescale the data such that each peak now falls onto an integer number. In figure 3.30:D I show the plot of peak position against number of chromosomes, and the obtained power law fit. I then show the data rescaled by this power law in figure 3.30:E. When Gaussian functions are again fitted to each peak in the rescaled data

I find that each peak has a mean that is well within 1% of the ground truth value.

Now that I have obtained a clear relationship between fluorescent signal and DNA content in the case of the run out data, the next step is to determine if the same relationship holds for the exponential culture that was not treated with antibiotic. As each run out and exponential sample was taken from the same culture, and stained in the exact same way, we would expect that both would follow the same relationship between intensity and DNA content. To verify this, I take the data from cells grown to exponential phase in mannose and rescale this using the same power law obtained for the run out cells. I then plot this on top of the run out data as shown in figure 3.30:F. The exponential trace shown in the figure, also appears to have two peaks. This is because cells grown on mannose in exponential phase spend significant time before initiation with a single chromosome, and likewise after completion of replication, cells spend time with two fully replicated chromosomes. The fact that the peaks in the exponential and run out traces align well, when both are rescaled based on the fitting to the run out data, indicates that the relationship between intensity and DNA content holds at least for samples prepared from the same culture. I will use this trick of rescaling the exponential phase data based on the run out results, for all of the measurements in this section.

Now that we have successfully converted the fluorescent signal for the mannose condition to number of chromosomes, the question is how to utilise this to perform the same task for other conditions. To showcase the procedure used for every condition in this work, we will look at in detail, the condition of M63 media with glucose and cas amino acids. The first step, as before, is to gate both sets of data based on the PPI width and PPI height channels. This is shown in figure 3.31:A. Next the PPI area channel is plotted against number of events, for both the M63 glucose and cas amino acids, and M63 mannose data. This is shown in figure 3.31:B. As can be seen, the peaks for both data sets line up very well. This is again a good indication that sample to sample variation in staining and measurement of fluorescent intensity is minimal. All that is required from this step is that there is sufficient similarity between the two traces, to allow an assignment of a number of chromosomes to each peak in the M63, glucose and cas amino acid data. Once each peak has been assigned a corresponding number of chromosomes, then the data can be rescaled

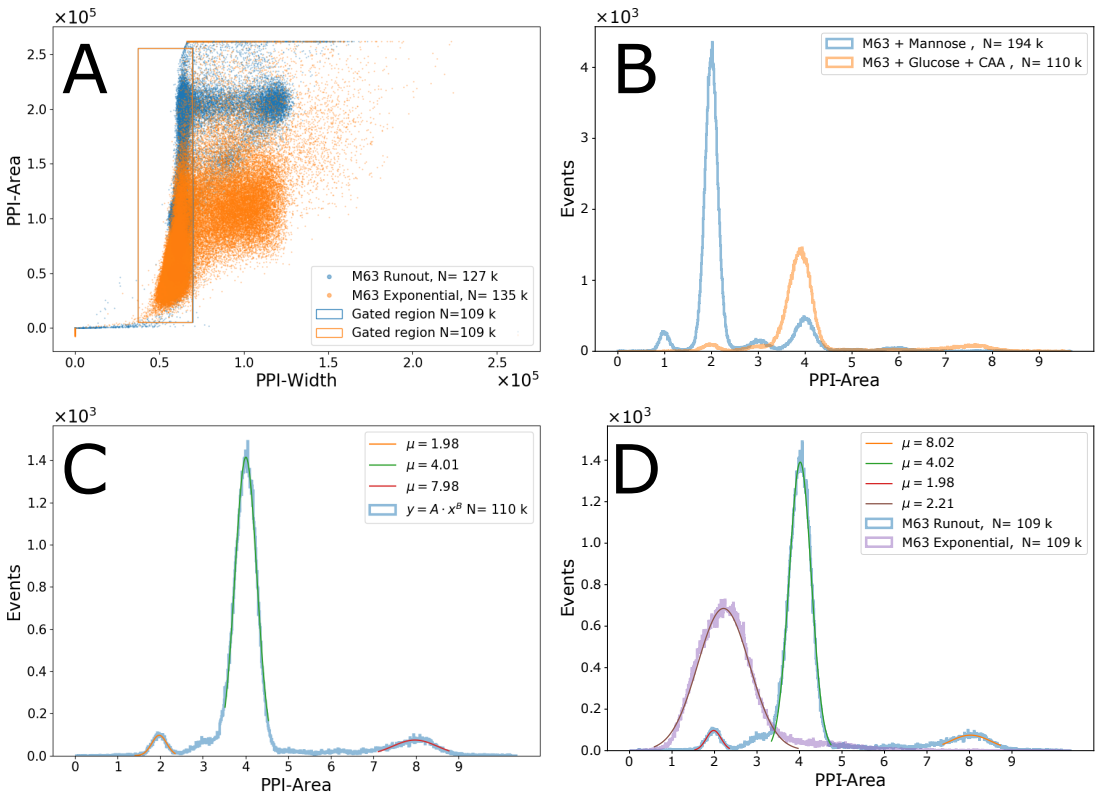

Figure 3.31: DNA content for cells grown in M63 media with glucose and cas amino acids. **A:** Scatter plot showing PPI pulse width against PPI pulse area for both the M63 mannose and M63 glucose data. Orange and blue squares indicated the gated areas. **B:** PPI area against frequency for the gated data in subfigure A. Notice that with no rescaling the peaks for the two run outs still align very well. **C:** PPI area against frequency for the data rescaled by a fitted power law function. The peaks now sit at integer number of chromosomes as indicated by the fitted Gausssian functions. The means of the Gaussians are shown in the figure legend. **D:** PPI area against frequency for the run out and exponential data. Both sets of data have been rescaled by a power law function fit to the peaks of the run out data.

by a power law function, obtained by fitting as described above for the mannose case. The results of this rescaling are shown in figure 3.31:C. Again there is less than a 1% difference between the apparent and ground truth position of the peaks, suggesting that a power law relationship is at least close to the actual relationship between fluorescent intensity and DNA content. Finally, we can rescale the data for exponential phase cells by the same power law relationship obtained for the run outs, and plot the two together as done in figure 3.31:D. With the exponential data now expressed in terms of units of DNA, we can trivially obtain the average DNA content per cell by averaging the purple curve in figure 3.31:D.

3.3.2 DNA content at high osmolarity

In the previous section I outlined how by appealing to run out data, and slow growing cells one can link fluorescent intensity to DNA quantity. In this section I will repeat this process for cells grown at high osmolarity. In particular, we will again look at cells grown in M63 media with glucose and cas amino acids and supplemented with either 200, 400 or 600mM NaCl.

The process starts in exactly the same way as in the previous section, with gating of the run out experiments based on the pulse width and pulse height in the propidium iodide channel. This is shown in figure 3.32:A, where the data for the 0 and 200mM NaCl conditions are compared. Following this, the run outs for the two conditions are overlaid in figure 3.32:B. Here, there is a slight departure from the previous examples, as the two run out traces do not overlap very closely. The reason for this is due to the differing salt concentrations in the staining solutions. I found that the standard protocol for preparing cells for the flow cytometry experiments led to cell lysis, as it called for a wash into a much lower osmolarity PBS solution. The lysis was caused by hypoosmotic shock as reported in Buda et al. 2016. To avoid this, I altered the protocol slightly by washing cells into PBS media with added sodium chloride, at a concentration that matched that of the experimental condition being examined. To remove any possibility of cell lysis during staining, I also added the appropriate amount of NaCl to the staining solution. This however, led to a difference in staining efficiency and, as a result, a lack of overlap in peak position between samples stained in the presence of different amounts of sodium chloride. I was still able to identify the number of chromosomes each peak corresponded to

in two ways. Firstly by comparison to a condition with known peaks as shown in figure 3.32:B. Even with differential staining it was often still clear which peaks were which. For example in both traces in figure 3.32:B, the largest peak corresponded to two chromosomes. For this to have not been the case, would have required a huge change to the cellular DNA content, which was not likely given the modest increase in NaCl concentration, and correspondingly modest reduction in growth rate (at 200mM NaCl). A second way I was able to verify the number of chromosomes each peak corresponded to, was by comparison with a small number of run out experiments I performed where I did not add NaCl to the staining or PBS solutions. In this case, though there was significant cell lysis, the peaks in the run outs did overlap well between conditions. I was therefore able to use this data (not shown) to help identify which peaks were which, in the data which had been stained with added sodium chloride.

With the number of chromosomes that each peak corresponded to establish, as before a power law function was fit, and the data rescaled using that function. This is shown in figure 3.32:C. With the data rescaled, the exponential data can be transformed in the same way and overlaid with the run out data as shown in figure 3.32:D. Again the mean DNA content per cell can then by trivially extracted.

This process was repeated for 400 mM NaCl (figure 3.33) and 600mM NaCl (figure 3.34). In each case alignment of the run out peaks was achieved by comparison to the next lowest osmolarity condition. Having rescaled the DNA content for exponentially growing cells in each NaCl condition, we can now examine how the average DNA content changes with osmolarity. To do this I can compare the rescaled DNA content histograms for each condition as shown in figure 3.35, where the x axis now shows genomes per cell and the y axis shown the normalised frequency. As can be seen from the figure, after rescaling into units of genomes, there appears to be very little change in the DNA content distribution of exponentially grown cells, as the sodium chloride concentration increases. If I take the mean DNA content per cell for each condition, and plot this against the osmolarity (as done in figure 3.35:C) we can see that there is basically no change in DNA content as osmolarity increases. This is in stark contrast to the result of Dai and Zhu 2018, which show an exponential increase in DNA content per cell, with increasing sodium chloride concentration. The

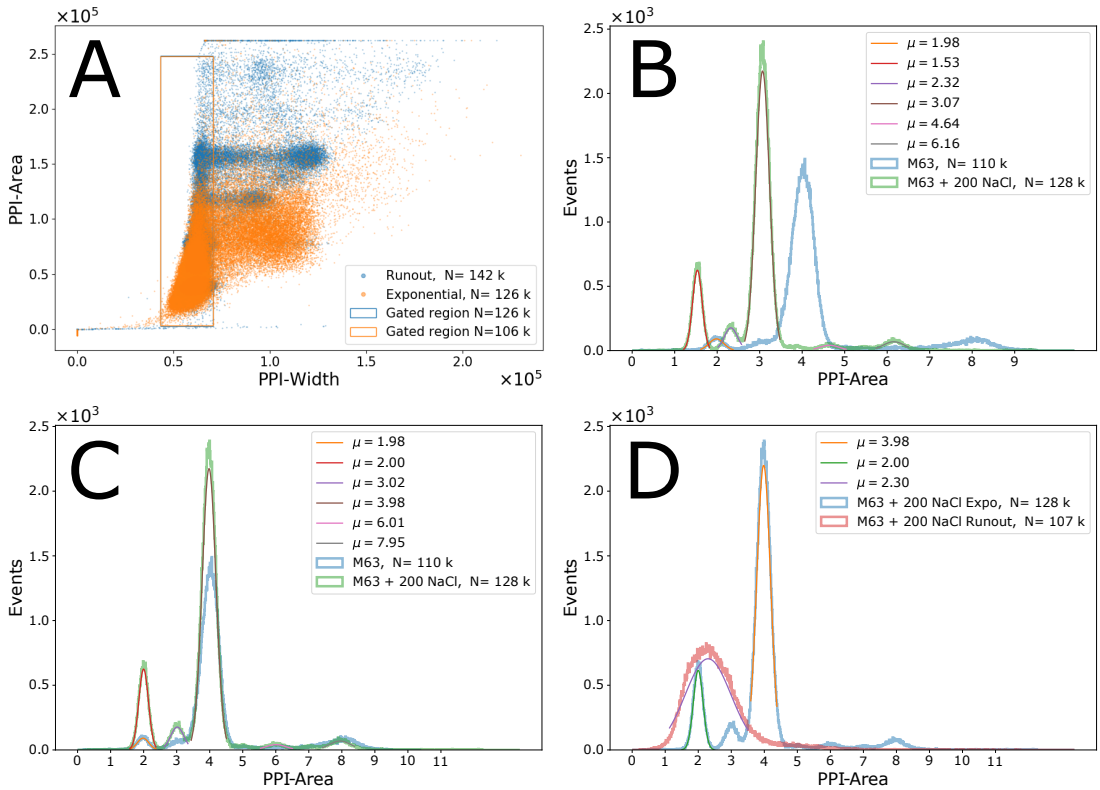

Figure 3.32: DNA content of cells grown in M63 media with added: glucose, cas amino acids and 200mM NaCl. **A:** Gating of data for 0mM and 200mM NaCl conditions based on pulse width and area in the PPI channel. Orange and blue boxes represent gated area. **B:** Raw data showing PPI area against frequency for both conditions. **C:** PPI area against frequency for each condition individually rescaled by a power law function fit to the mean of each peak. **D:** Exponential and run out traces for the 200mM NaCl condition. Both have been rescaled using the power law function fit to the run outs.

reason for this discrepancy remains unknown. It is possible that the effect high NaCl concentrations have on cellular DNA content is dependent on the nutrient limitations of the base media. It may be therefore, that the difference can be explained by the difference in media as Dai and Zhu 2018, used LB in their study. In any case, given the excellent alignment between the distributions shown in figure 3.35:C, I have high confidence that the DNA content is invariant under osmolarity changes. It seems highly improbable that the distributions would align so well had an error been made in correlating fluorescent signal with DNA content in each of the different conditions.

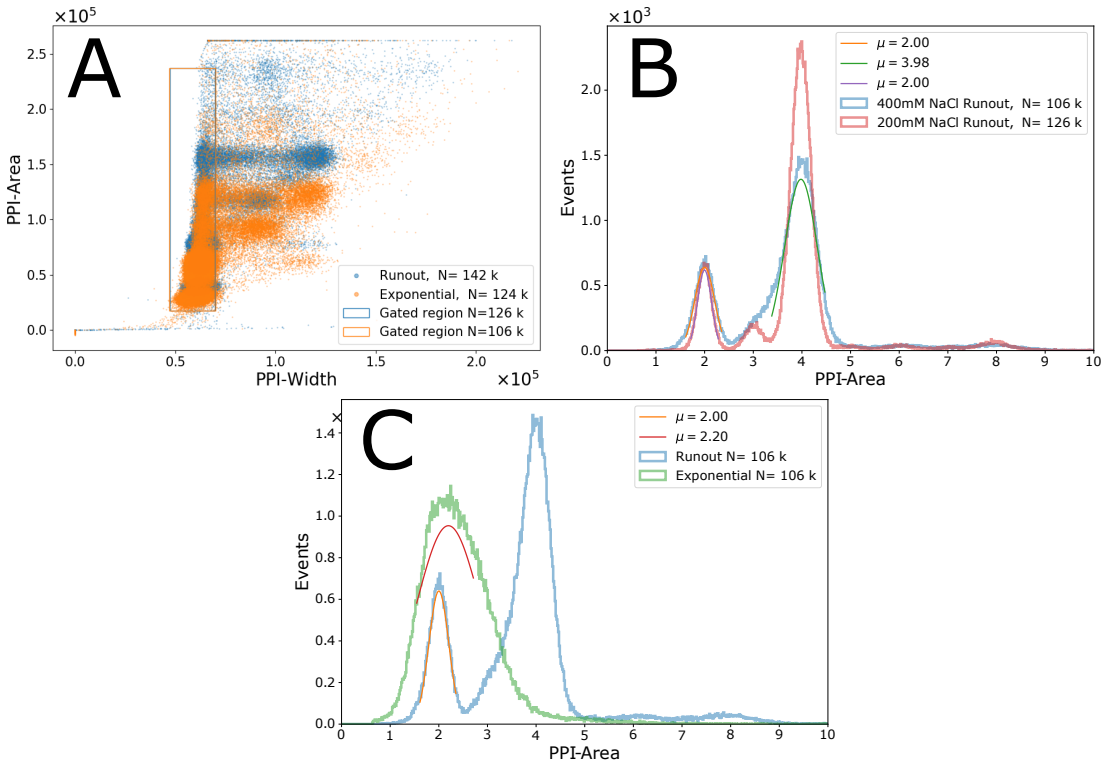

Figure 3.33: DNA content of cells grown in M63 media with added glucose and cas amino acids and 400mM NaCl. **A:** Gating of data from the 400mM and 200mM NaCl conditions. Boxes represent gated area. **B:** Plot of PPI area against frequency for the gated data in subfigure A. **C:** Plot of PPI area against frequency for the run out and exponential data from the 400mM NaCl condition. Both data sets have been rescaled by a power law function obtained from fitting of the peak positions in the run out data.

As well as average DNA content, the number of origins per cell can also be examined using the flow cytometry data. In particular from the run out experiments I can plot the percentage of cells in the population with a given number of genomes. This can be used to quantify the difference that occur in the run out traces as the osmolarity is increased. In figure 3.36:B I plot the rescaled DNA content distributions for the run out in each of the NaCl conditions. It appears from this trace, that the number origins in lower peaks increase as the osmolarity increases. To summarise this trend in figure 3.36:C, I plot the percentage of the population present with two chromosomes, four chromosomes and less than five chromosomes. From this we see that as the osmolarity increases, the percentage of cells with two chromosomes increases, while the number of cells with four chromosomes decreases. This suggest

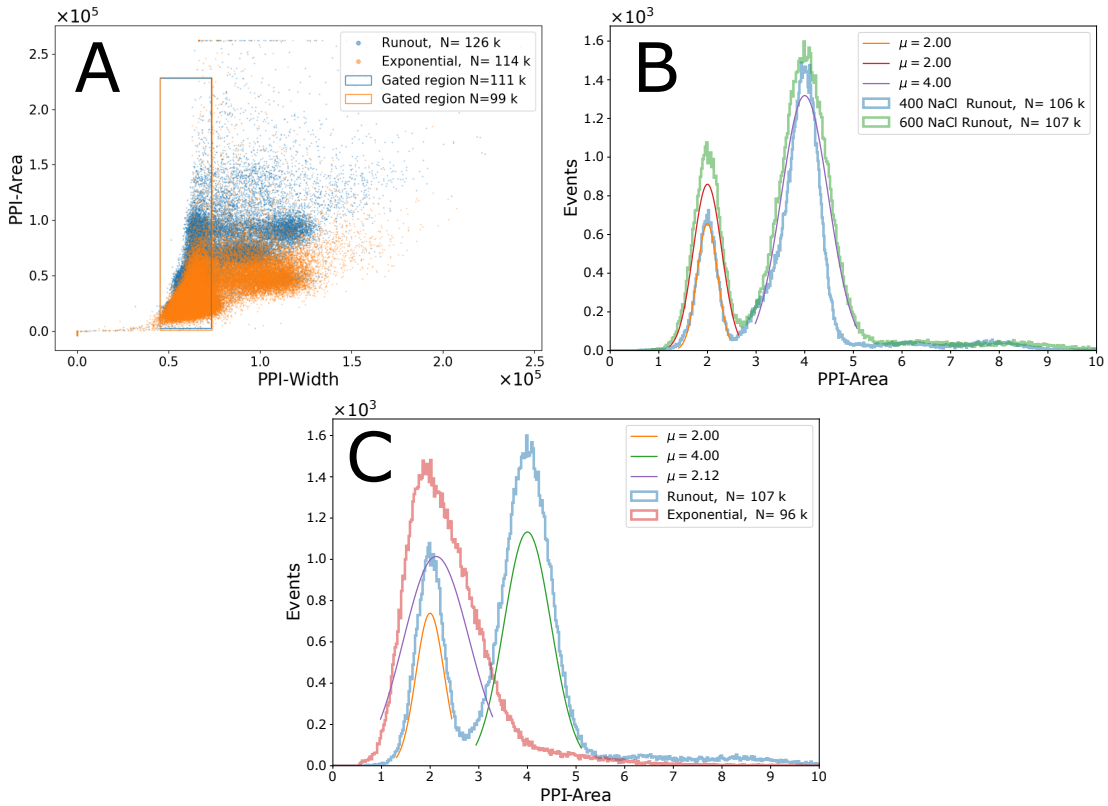

Figure 3.34: DNA content of cells grown in M63 media with added glucose and cas amino acids and 600mM NaCl. Plots shown here are same as for figure 3.33, see figure caption for details.

that the the number of origins tends to decrease overall slightly at high osmolarity. As the average DNA content of exponential cells stays fixed at high osmolarity, the decrease in number of origins suggests that time at which initiation occurs in the cell cycle is pushed later as the osmolarity increases.

Finally, I wanted to examine the extent of the day to day variation in the flow cytometry results, and how they varied with strain. To this end, in figure 3.37 I plot the exponential and run out data for three biological repeats that were grown, collected, and ran in the flow cytometer on different days. As can be seen there is very little variation in the exponential and run out DNA distributions of the three repeats, indicating that our results are robust to day to day variation.

To examine how the DNA content profile changes with strain I plot the run out

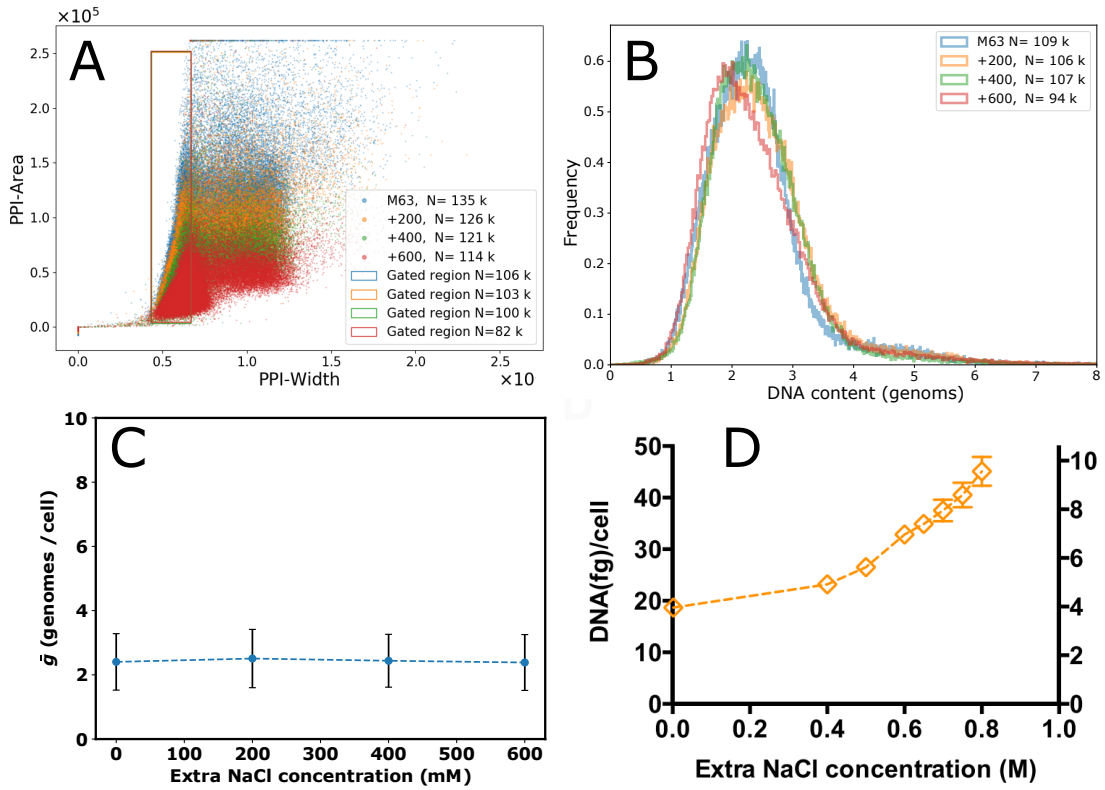

Figure 3.35: Average DNA content at high sodium chloride concentrations. **A:** PPI width against area for each sodium chloride concentration. The box represents the gated area. **B:** DNA content per cell in units of genomes against normalised frequency for each NaCl concentration. Concentration is indicated in the figure legend in units of mM. **C:** Mean DNA content per cell in units of genomes against NaCl concentration. Error bars represent the standard deviation of the distribution in each condition. **D:** DNA content per cell in fg per cell and genomes per cell at different sodium chloride concentrations as measured by Dai and Zhu 2018.

DNA distribution for both YD133 + pWR20 and BW25113 cells in figure 3.38. As can be seen, there is very little variation between the two strains which indicates that the DNA content measurement is not significantly affected by the presence of the pWR20 plasmid, or the mutations that separate BW25113 and YD133.

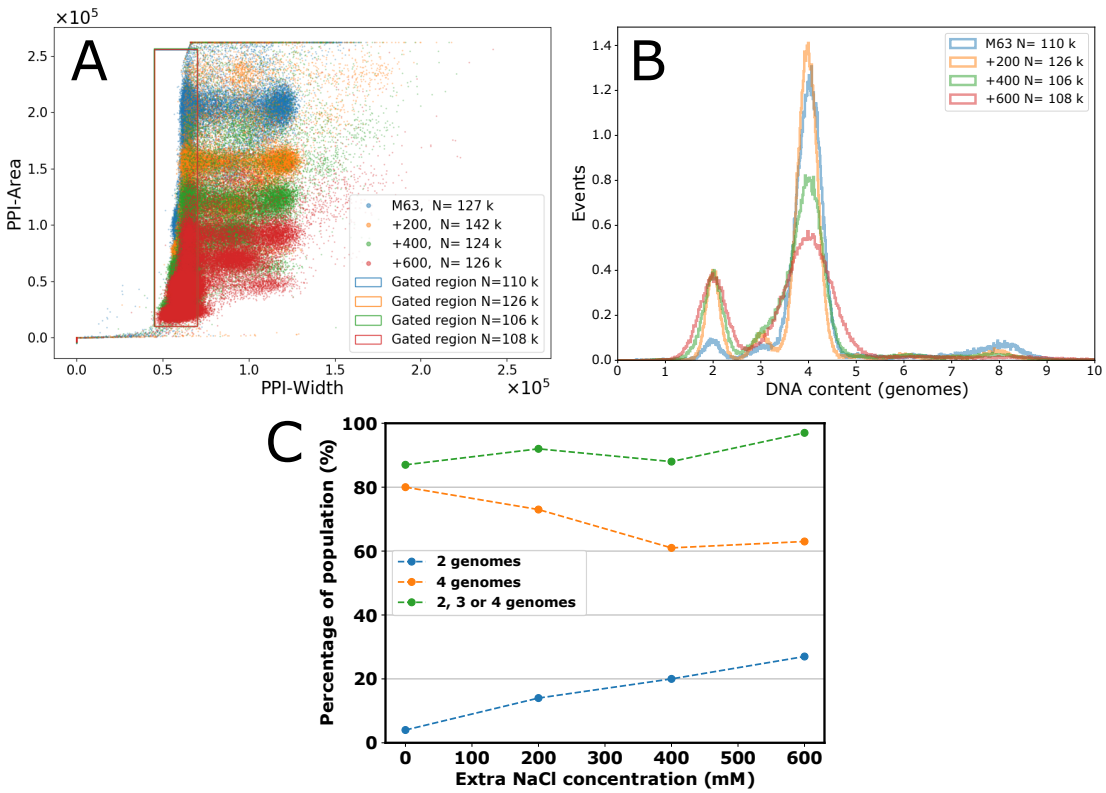

Figure 3.36: DNA content of runouts at high osmolarity. **A:** Plot showing gating of each run out condition based on width and area of PPI channel. Boxes indicate gated area. **B:** DNA content expressed in genomes per cell against frequency for run outs in each of the NaCl conditions. **C:** Percentage of cells with 2 genomes (blue), 4 genomes (orange) and 2,3, or 4 genomes (green) against NaCl concentraiton.

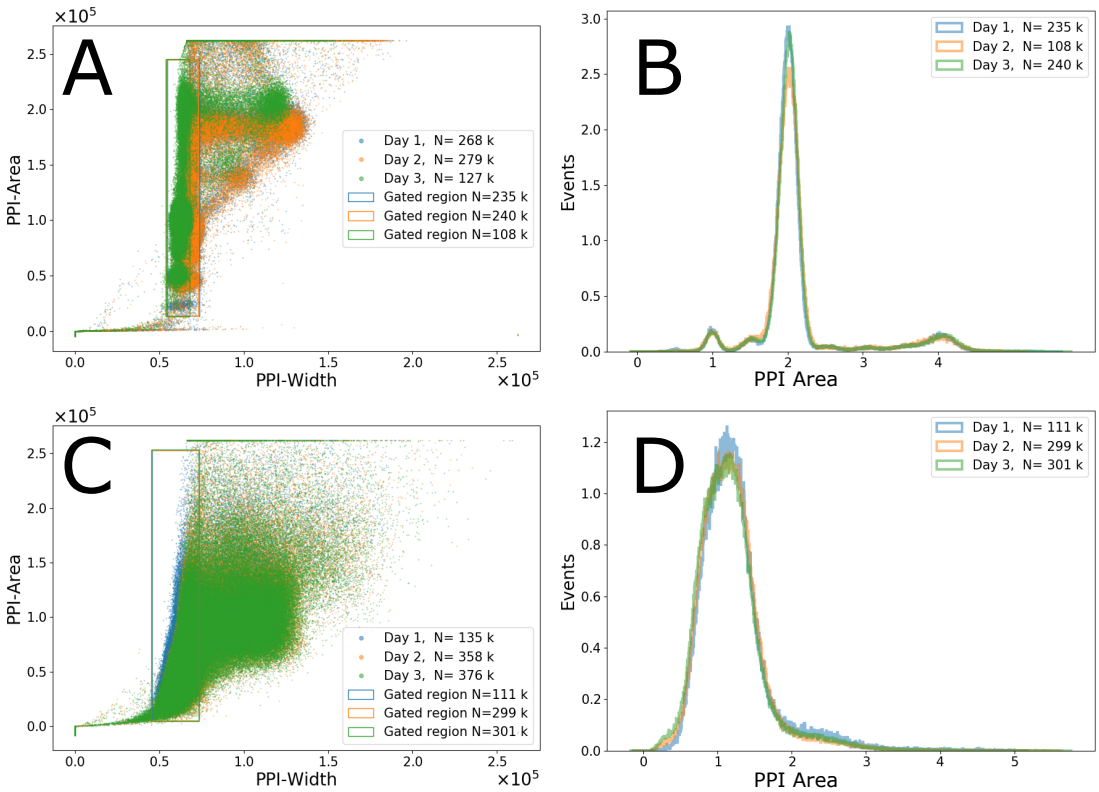

Figure 3.37: Day to day variability in flow cytometry measurements. **A & C:** Plots showing the gating of run out data (figure A) and exponential data (figure C) for three biological replicates measured on different days. **B & D:** Day to day variability in DNA content from run outs (figure B) and exponential cells (figure D).

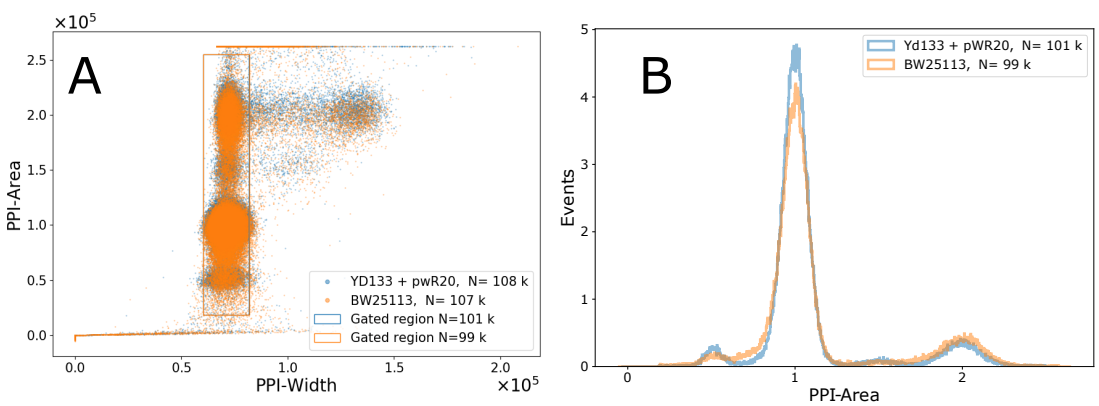

Figure 3.38: DNA content for BW25113 and YD133 + pWR20 cells. **A:** Gating of both strains based on pulse width and height of the PPI channel. Boxes indicated gated events. **B:** PPI area against frequency for both cell types.

3.3.3 Discussion and future work

In the previous section my primary goal was to determine how the DNA content of cells changes with increasing concentration of NaCl. In this section I will discuss how successfully I was able to achieve this aim, and what work could be done to in the future to build on my findings.

To measure the DNA content of cells I fixed exponentially growing, and run out, cells with ethanol and labelled their DNA using propidium iodide. I then measured the fluorescent intensity of cells using flow cytometry. As discussed in section 3.3.1, I used M63 and mannose media to obtain slow growing cultures, which after run out, contained predominately cells with one or two chromosomes. Having previously established in section 2.6.2, that there is a powerlaw relationship between quantity of fluorophore, and measured intensity, I was able to transform the run out data, from units of arbitrary fluorescent intensity, to genome equivalents. By applying this same rescaling to the data for exponentially growing cells, I was able to measure the average DNA content of cells in exponential growth. Through alignment of the run out traces, I was able to convert each subsequent sample from arbitrary fluorescent intensity to genome equivalents. Again, by using the same rescaling for exponential samples as run outs, I was able to determine how the average DNA content changed, for exponentially growing cells at different NaCl concentrations. Given the very good levels of alignment between the exponential DNA content traces at different osmolarities, as shown in figure 3.35:B, I concluded that this process of alignment and rescaling had been carried out correctly, as it would be very improbable that all the traces would align otherwise.

In figure 3.35:C, I showed that the average DNA content per cell in exponentially growing cultures, was essentially unaltered by the increase in NaCl concentration, between 0mM and 600mM. This was in stark contrast to the results of Dai and Zhu 2018, which showed an exponential increase in average DNA content per cell, with increasing NaCl concentration. In fact Dai and Zhu 2018, observed an almost 100% increase in DNA content at 600mM of additional NaCl. It is quite surprising to see such a drastic difference between the two results. It is possible that the difference can be explained by high osmolarity conditions having an effect on cell physiology, which is dependent on the growth rate of the base media. However, as discussed in

section 3.2.2.7, I was unable to successfully culture cells at high growth rates and at high NaCl concentrations, without severely affecting cell morphology. Therefore, while it would be a good idea going forward to measure the DNA content of cells grown in LB and additional NaCl, it is difficult to see how this could be done, without first addressing and understanding the changes I observed to cell morphology. An alternative set of experiments, would be to instead look at how the DNA content varies with the osmolyte used. Specifically, it would be interesting to determine what affect increasing concentrations of sorbitol and sucrose have on cell DNA content. This could shed more light on how cell physiology depends on the osmolyte used, and inform the discussion on the validity of "high osmolarity" as a concept which uniquely determines cell physiology.

As previously mentioned, cells grown at a low growth rate on mannose initiate and terminate chromosome replication in the same generation. Therefore, when these cells are used in a run out experiment, the result should be cells which have either one or two chromosomes. However, we saw previously in figure 3.30, that in practice the run out of cells in this condition actually leads to a population of cells with three and four chromosomes. Given that cells normally require a doubling time of 60 minutes or less to contain four origins of replication (and therefore to end up with four chromosome during run out) the existence of these four chromosome cells is quite surprising, especially considering the doubling time in mannose is around 90 minutes. I suggest that this population of cells may be a feature of BW25113 derived strains, and therefore that going forward, work could be done to determine if the same population of four chromosome cells, is seen in MG1655 or other wild type strains. My reason for suggesting this is that in section 3.2.2.2, we saw that there was a sub-population of cells which had significantly larger volume. I labelled this sub-population "the bump", and showed that it was unique to BW25113 derived strains. It seems possible, that cells in this sub-population may differ, not only in cell size, but also in other ways such as DNA content. Therefore by comparing MG1655 and BW25113 run outs, this hypothesis could be tested. The results of such an experiment, may elucidate the cause for such a bump in the cell size distribution, and aid the explanation of the observed differences between the two strains.

3.4 Cell cycle parameters at high osmolarity

3.4.1 C period determination from QPCR

3.4.1.1 A comparison of methods for computing QPCR primer efficiencies

As discussed in section 2.7 the time to replicate the chromosome in *E. coli* (C period) can be determined by quantifying the number of origins vs termini in a sample of genomic DNA, taken from an exponentially growing culture. Section 2.7 also explained how, by analysing real time PCR traces, the relative copy number of the origin and termini can be computed. As part of that explanation, the concept of standard curves was introduced as a gold standard method for determining the efficiency of QPCR primers, which is a critical step in the extraction of the C period. However, standard curves are not the only method for this task, and examples can be found where alternatives are used. A key example of this is Si et al. 2017, where primer efficiency is computed by fitting a straight line to the logarithm of the fluorescent trace generated from each reaction. In this section, a comparison will be made between this logarithm fitting method and standard curves, to determine the effect each has on the derived primer efficiencies. How these results vary with the chosen threshold value will also be discussed.

One of the disadvantages of computing primer efficiencies via standard curves is that it requires the running of at least five, and often tens, of reactions to compute the efficiency of a single primer pair. For this reason there is demand for a method of computing primer efficiency simply from a single QPCR trace. This can be achieved, in principle, by examining the fluorescent trace for a single reaction on a semi log plot, where the y axis is the logarithm of the fluorescent value, as shown in figure 3.39 :B. Typically, a threshold value is set somewhere in the exponential phase of the reaction and between four and six points are examined around this threshold value. By determining a line of best fit for these points, a gradient can be computed, which can be related to the primer efficiency as follows. First, recall from equation 2.5, that we expect the fluorescence f at cycle t to follow an exponential dependence like so:

$$f = f_0 \alpha^t, \tag{3.4}$$

where f_0 is the initial fluorescence and α is the primer efficiency. Next if we take the logarithm of both sides we can yield the following expression:

$$log_2(f) = log_2(\alpha) \cdot t + log_2(f_0), \tag{3.5}$$

which is precisely of the form $y = mx + c$. Therefore, we can relate the gradient of the line of best fit (m) to the primer efficiency as follows:

$$2^m = 2^{log_2(\alpha)} = \alpha. \tag{3.6}$$

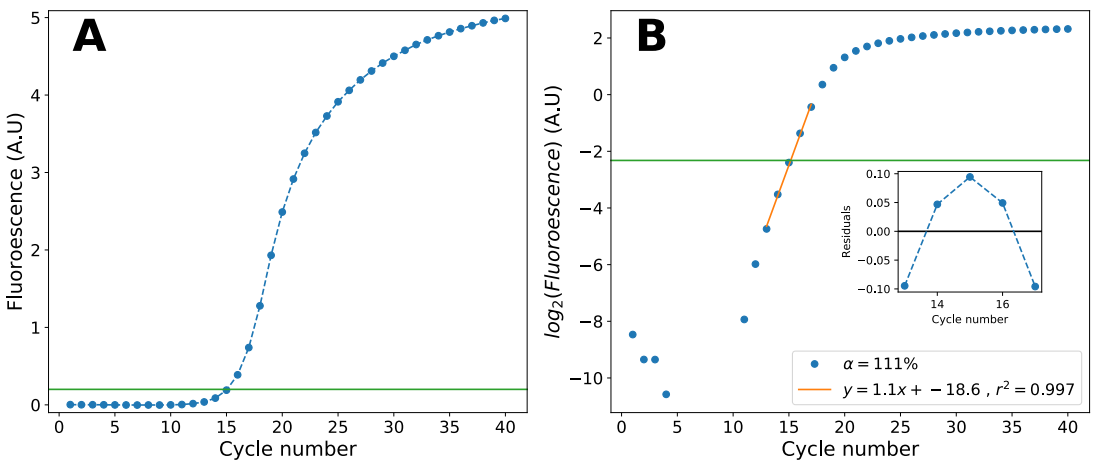

Figure 3.39: Primer efficiency from single QPCR traces. **A:** Fluorescent signal against cycle number for an example QPCR trace. The green line indicates the value around which points are examined to determine the primer efficiency. **B:** Logarithm of fluorescence against cycle number for the QPCR trace in subgifure A. The green line indicates the point around which the data is fitted. The fit is shown in orange and its equation (and extrapolated efficiency) are shown in the figure legend. Inset shows the residuals for the fitted line.

In figure 3.39:B, the legend shows the equation for the line of best fit for five points surrounding the threshold value, as indicated by the green line. The derived primer efficiency α (expressed as a percentage) is also indicated in the figure legend. Two natural questions to ask at this point are firstly: how does the derived efficiency change with the chosen threshold value, and secondly: how does the efficiency compare with that derived from standard curves. To answer these question, I took a set of 20 QPCR reactions composed of four repeat reactions for each of five different

concentrations of target DNA. This allowed a standard curve to be computed and compared to the efficiencies extracted from the semi log fits. The results of this are shown in figure 3.40.

In figure 3.40:A, the fluorescent traces for the 20 chosen PCR reactions are shown. These reactions were performed using template DNA taken from the strain YD133+pWR20. The template DNA was diluted five times, with each subsequent aliquot being a five fold dilution from the previous one. For each dilution, four repeats were performed and, as can be seen form the plot, there exists very little variation between repeats. The multiple horizontal green lines represent the different threshold values that were used to determine the primer efficiency. Figure 3.40:B, shows the same data simply plotted on a semi-log scale. For each threshold value the efficiency was computed individually for each trace as shown in figure 3.39:B, by fitting of a straight line to five points straddling the threshold value. In addition, a standard curve efficiency was computed by first calculating and then averaging the c_t values for each repeat, before then plotting the average c_t against the logarithm of the DNA concentration, to extract the efficiency (as shown in figure 2.22). Figure 3.40:C shows how the calculated efficiency varies with the threshold value for each individual trace as compute by semi-log fits. This data is then averaged and plotted on figure 3.40:D. This plot also shows how the efficiency computed from standard curves changes with the threshold value.

From this data we can see that there is a drastic change in primer efficiency with threshold value, when using the semi-log fit method. In fact, it would appear that there is a monotonic decrease in primer efficiency with increasing threshold value. It would therefore, seem difficult to determine where to place the threshold, such that results remain robust to small variations around the chosen value. In comparison, when using the standard curve method, results seem far more robust to the choice of threshold, with the efficiency curve appearing almost flat compared to that of the semi-log method. However as the inset in figure 3.40:D shows, there still remains a small change in computed efficiency with threshold value. Across the examined range, the computed standard curve efficiency varies by around 5% although the variation is not monotonic, as is the case with the semi-log fit method. From this we can conclude, that while both results are sensitive to the chosen value of the

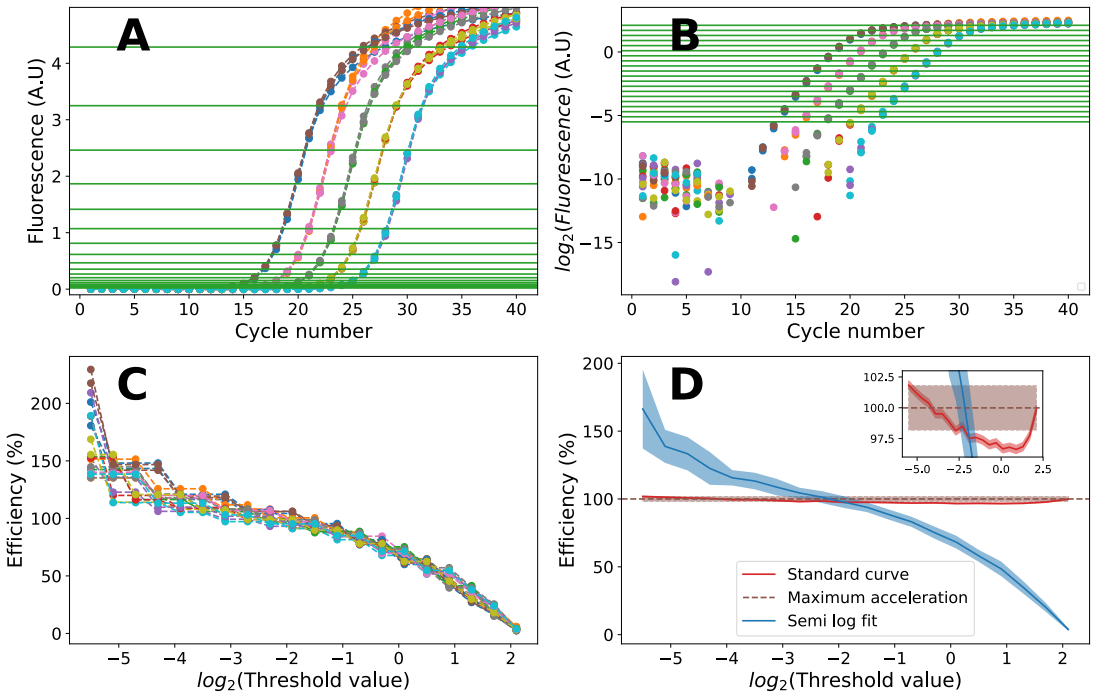

Figure 3.40: Primer efficiency as a function of threshold value. **A**: Fluorescent traces for 20 QPCR reactions. Green lines represent threshold values at which primer efficiency was computed. **B**: Same data as in sub-figure A but plotted on a semi-log scale. **C**: Primer efficiency against the logarithm of threshold value. The efficiency for each QPCR trace was computed by fitting of a straight line to the semi-log fits in sub-figure B at each threshold value. **D**: Figure showing how primer efficiency changes with threshold value for both methods. The blue trace is the average of the data in sub-figure C with the highlighted interval indicating the standard deviation. The red trace is the computed standard curve efficiency with the highlighted region indicating the uncertainty in efficiency as a result of uncertainty in the computed gradient. The brown trace indicates the mean efficiency computed using the maximum acceleration method with the highlighted area being the standard deviation of the four repeats.

threshold, the standard curve is significantly less sensitive, and as such, is a far better choice. However the problem still remains of where to place the threshold. Luckily there exists a third method which allows the extraction of c_t values, and therefore computation of primer efficiency, without the need to chose a particular threshold. This method known as the double derivative, or maximum acceleration, method involves choosing the c_t value at the point where the second order derivative of the fluorescent trace is maximised. This method is based on the idea that the point where the curves acceleration is maximal, will also be the point which lies most clearly in the exponential region. The advantage of this method is that it does not rely on the absolute value of the fluorescent trace curves. Therefore, the method does not require the selection of a threshold value nor the background subtraction of the data. It should correspondingly be more robust to changes in the variation of the fluorescent signal during early cycles. For each of the four repeats in 3.40, the primer efficiency was computed using the standard curve method, with c_t values derived from the maximal acceleration method. These four repeats were then averaged and plotted in brown in figure 3.40:D, with the standard deviation shown as the brown highlighted region. As we can see, there is very good agreement between the value generated from the maximal acceleration and threshold methods. Due to this good agreement, and because of the simplification that comes from not having to choose a threshold value, I decided to use the maximal acceleration method as provided by the light-cycler 96 software, to determine the c_t values for all QPCR experiments in this investigation.

3.4.1.2 C period at high osmolarities

With an understanding of how I computed the QPCR primer efficiency, we can now turn to the measurement of the C period, and in particular, how it changes at high osmolarity. In this case we will focus on the familiar conditions of M63 media with glucose and cas amino acids, with varying concentrations of additional sodium chloride including: 0, 200, 400 and 600mM. For each condition I grew two biological repeats, and extracted their DNA by sonication, as described in section 2.7.2. For each primer set, five concentrations of DNA were ran, with each one having four technical duplicates. Each repeat was averaged and used to plot a standard curve as shown in figure 3.41:A. In this figure the four lines represent the standard curves for the origin and terminus primers, for each of the two biological replicates. The

residuals for each line are shown in figure 3.41;B.

Both the r^2 values for the standard curves and the residuals indicate that the fits are very good. However, we found that outliers within the technical duplicates were a significant issue that hampered plate to plate consistency in the final measurement of the C period. Take for instance, the first point of the red curve in figure 3.41. From the residuals we can see that one repeat lies significantly lower than the other three. It turns out, that if this outlier point is removed then the the overall efficiency of the primer pair decreases. As the equation for the C period, raises the efficiency to the power of the C_t value, small changes in the efficiency can generate large changes in the final C value. We could of course, manually remove outliers, however I wanted to instead use an automated approach to eliminate the possibility of bias, in the choice of outliers. To do this, I implemented a simple algorithm which looked at the technical repeats of each DNA concentration. If the standard deviation of these points was larger than a threshold value the point furthest from the mean was discarded. This process was repeated until either, two points were left or the new standard deviation fell within the threshold value. The choice of threshold value obviously had some effect, but I found that in the range 0.6-0 there was little change to the C value as the threshold was moved. In the end I decided to use a value of 0.3 as the threshold which lay in the middle of this range. This method was certainly not perfect however. If all four data points were evenly spaced such that there was a large standard deviation but no obvious outlier, then the algorithm would choose a random point furthest from the mean and throw that point away, even if in this case the point was not really an outlier. However, as there are multiple points per standard curve, I think the effect of this averages out over the coarse of determining a single C period. In figure 3.41;C and D I show the results of applying the algorithm to remove outliers. As can be seen it has a very small effect on the final efficiency determination. For completeness the data for the 200,400 and 600mM conditions are shown in figures 3.42, 3.43 and 3.44 respectively.

With the primer efficiencies now computed we can turn to the computation of the C period for each condition. The C period is calculated using equation 2.26. This is done using the average C_t of the technical replicates (after removal of outliers). A C value is calculated for each DNA concentration in both biological replicates. This

gives a total of 10 individual estimates for the C period which are then averaged to get an overall value for each condition. In figure 3.45 I show each individual estimate and the mean C period for each condition. I then plot the mean C period against osmoalrity in figure 3.46, and compare this to the results from Dai and Zhu 2018. As can be seen there appears to be a very good agreement between my results shown here and the literature measurements. In both case there is a monotonic increase in C period as the salt concentration increases. At 600mM NaCl both data sets measure a C period of approximately 100 minutes. This may be an indication that the C period, unlike cell size, is affected by high osmolarity conditions in a manner that is not dependent on the growth rate of the cell. However to confirm this we would need to measure the C period again for a different nutritional media such as LB or RDM.

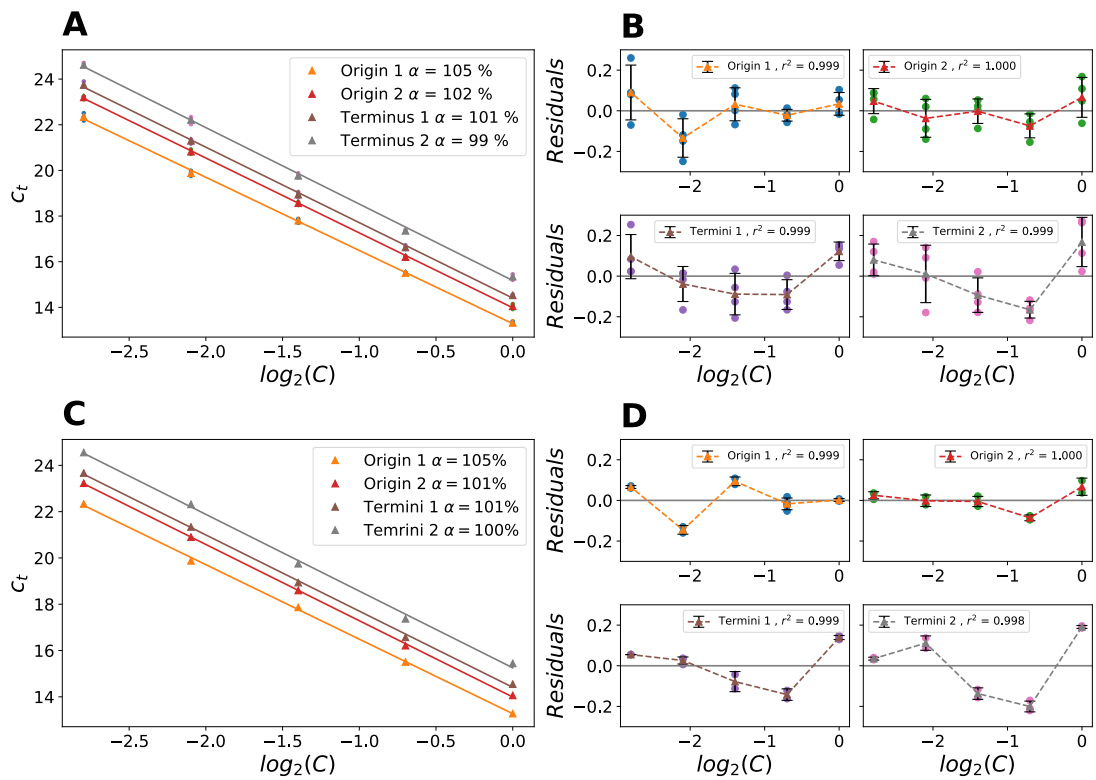

Figure 3.41: Primer efficiency determination for M63 media with glucose and cas amino acids. **A:** Standard curves for two biological repeats showing data for both the origin and terminus primer sets. Triangles represent means of technical duplicates for each DNA concentration while dots represent individual replicates. **B:** Residuals for lines of best fit to the data in subfigure A. Error bars are standard deviations of technical replicates. Again dots are individual replicates and triangles are means. **C:** Same plot as subfigure A but data has been passed through a process to remove outliers (see text for details). **D:** Same plot as subfigure B but again data has been processed to remove outliers.

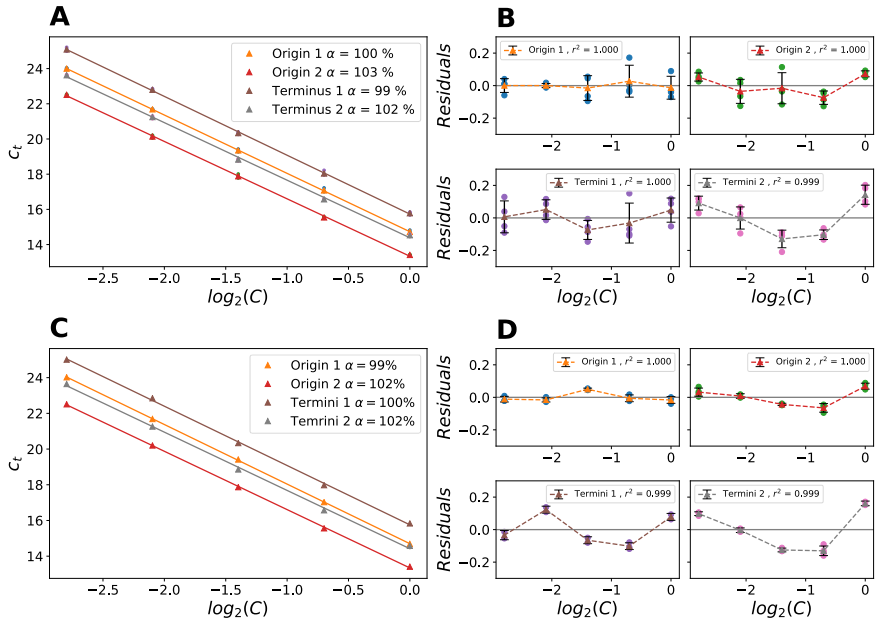

Figure 3.42: Primer efficiency determination for cells grown in 200mM NaCl. See figure 3.41 for details.

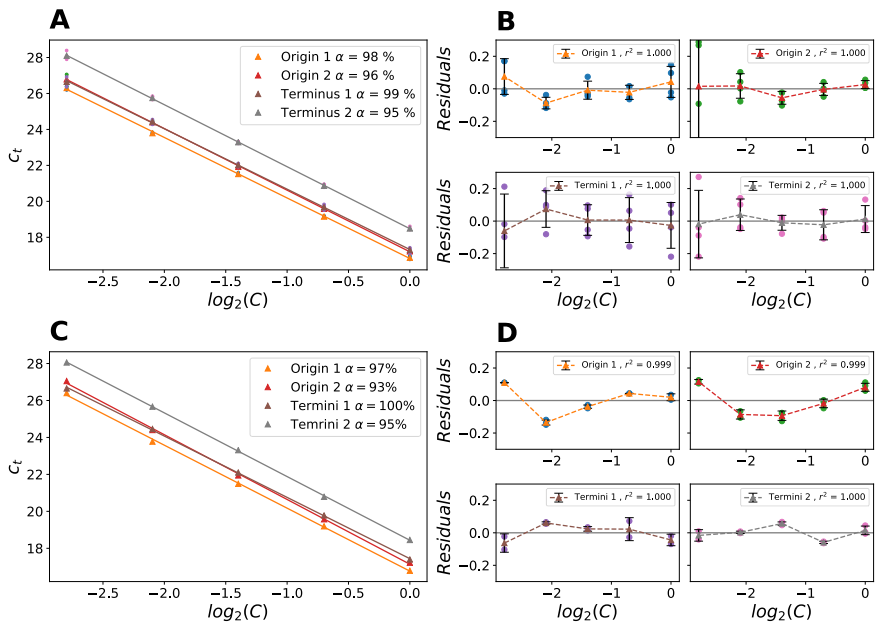

Figure 3.43: Primer efficiency determination for cells grown in 400mM NaCl. See figure 3.41 for details.

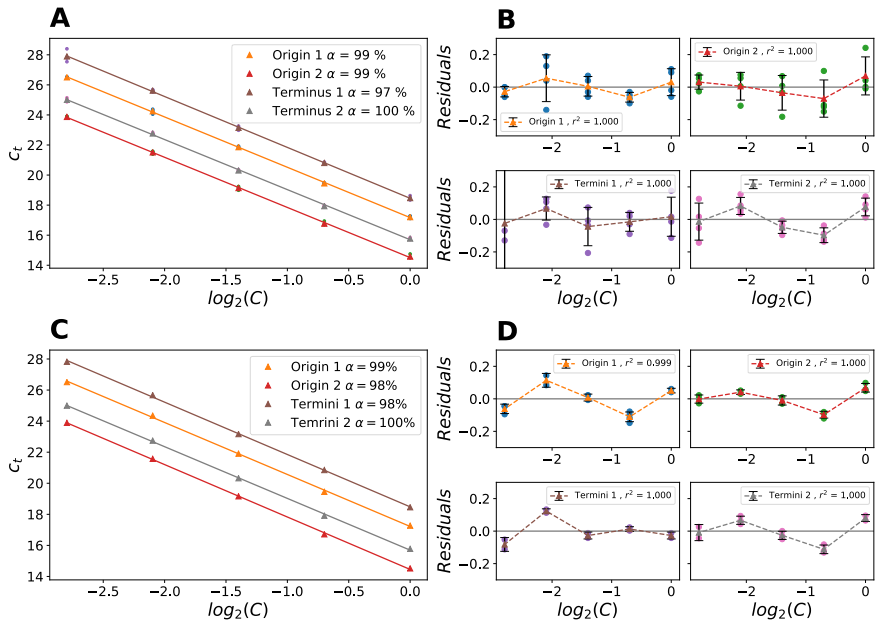

Figure 3.44: Primer efficiency determination for cells grown in 600mM NaCl. See figure 3.41 for details.

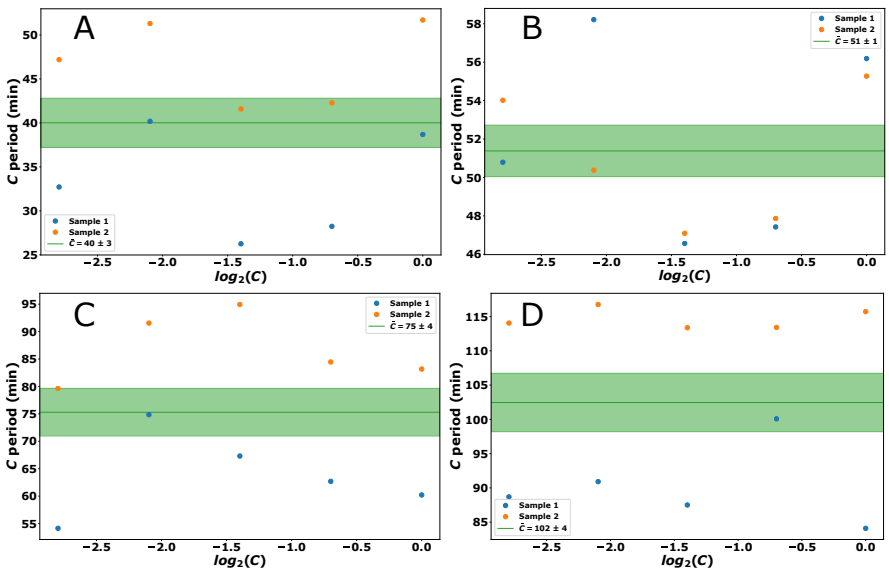

Figure 3.45: C period determination. Plot of C period estimate against logarithm of DNA concentration for each high salt condition. Green line represents mean of all points in each plot. Shaded green area represents standard devaition. **A:** 0mM NaCl. **B:** 200mM NaCl. **C:** 400mM NaCl. **D:** 600mM NaCl.

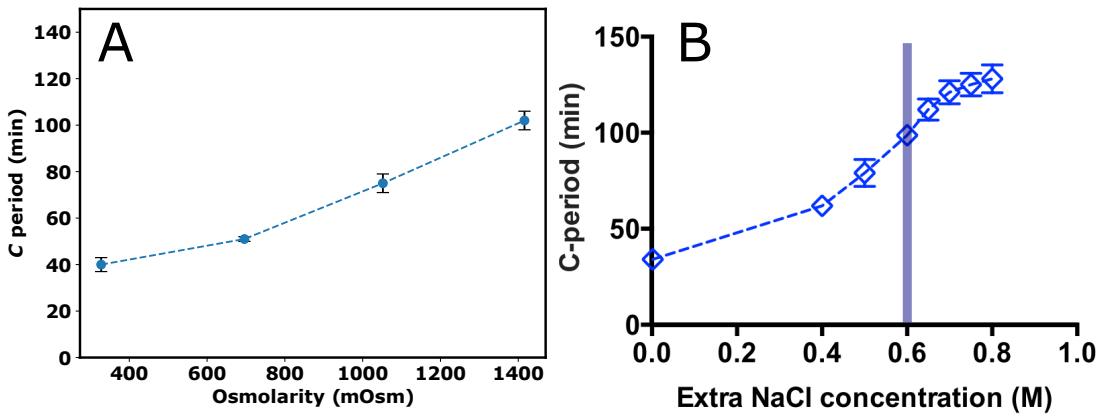

Figure 3.46: *C* period at high osmolarity. **A:** *C* period vs osmolarity obtained using different concentrations of NaCl. **B:** *C* period against NaCl concentration as measured by Dai and Zhu 2018.

3.4.2 D period at high osmolarity

With the *C* period and DNA content at high osmolarity measured, we are now in a position to compute the *D* period. We do this using an equation which has been derived by Bremer and Churchward 1977, assuming only exponential growth, and that each initiation of replication is followed $C + D$ time later by division. This equation (shown below) is simply an inversion of equation 1.19:

$$D = \tau log_2 \left(\frac{C \cdot \bar{G} \cdot ln2}{\tau \left(2^{\frac{C}{\tau}} - 1 \right)} \right), \tag{3.7}$$

where $\bar{G}$ is the average DNA content of cells, τ is the doubling time and C is the chromosome replication time. As the error in the C period was the largest I considered only the error in this parameter to determine the error in the D period. To compute error in D I computed a D value for each estimate of C in figure 3.45. I then took the standard deviation of these D periods to be the error. As can be seen in figure 3.47 the D period scales with osmolarity in what appears to be a very good approximation to an exponential way. This matches closely the results from Dai and Zhu 2018 which are also shown in 3.47.

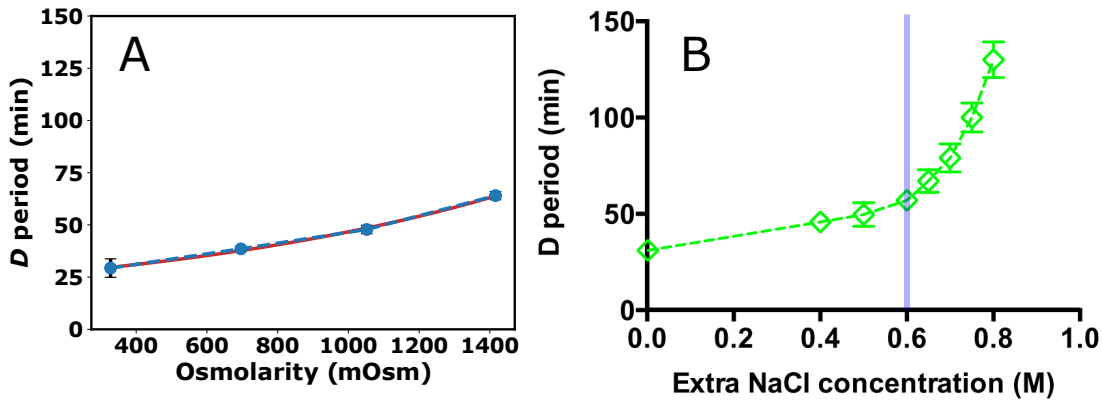

Figure 3.47: D period at high osmolarity. **A:** D period as computed in this work. The red line indicates an exponential fit of the form: $y = \alpha e^{\beta x} + \gamma$ where α, β, and γ are fitted constants. **B:** D period against NaCl concentration as measured by Dai and Zhu 2018. The vertical blue line indicates the highest concentration of NaCl examined in this work.

3.4.3 Cell cycle time at high osmolarity

With the C and D period now measured we can combine them to examine how the cell cycle time (τ_{cyc}) varies with osmolarity. To compute τ_{cyc} we simple sum the C and D periods. As before, in order to get the associated error in τ_{cyc} we compute an individual value for each C (and associated D) period in figure 3.45 to get ten estimates of τ_{cyc} for each condition. We then take the standard deviation of these ten estimates to be the associated error in the D period. As with the D period the cycle time shows an exponential dependence on the osmolarity. Again this is in good agreement with Dai and Zhu 2018 as shown in figure 3.48.

3.4.4 Number of origins at high osmolarity

With the cycle time computed we are now able to extract the number of origins per cell at high osmolarity. We do this using the following expressions which again can be derived from very few assumptions as shown by Bremer and Churchward 1977:

$$\bar{I} = 2^{\frac{C+D}{\tau}}. \tag{3.8}$$

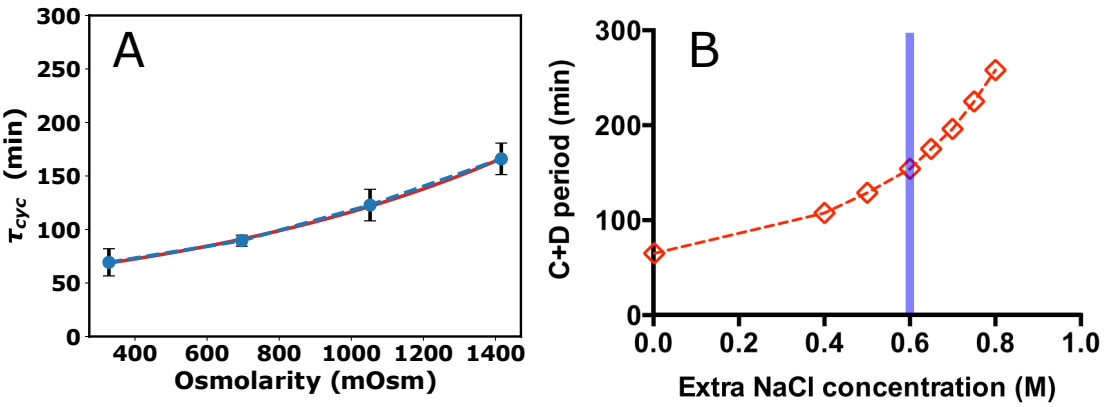

Figure 3.48: τ_{cyc} at high osmolarity. **A:** τ_{cyc} as computed in this work. The red line indicates an exponential fit of the form: $y = \alpha e^{\beta x} + \gamma$ where α, β, and γ are fitted constants. **B:** τ_{cyc} against NaCl concentration as measured by Dai and Zhu 2018. The vertical blue line indicates the highest concentration of NaCl examined in this work.

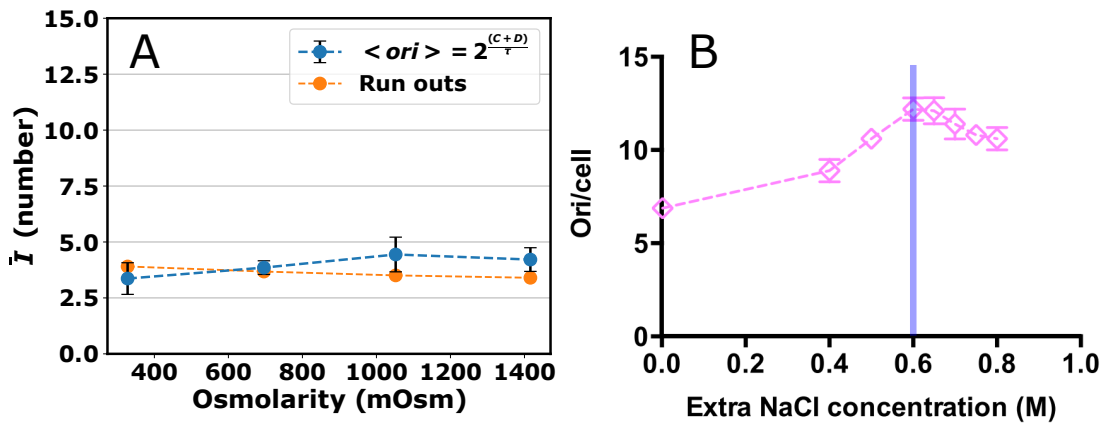

Figure 3.49: Number of origins at high osmolarity. **A:** Number of origins as computed using flow cytometry data (orange) and estimates of C and D periods (blue). Errors are standard devaition of multiple estimates of the number of origins. **B:** Number of origins against NaCl concentration as measured by Dai and Zhu 2018.

In figure 3.49 I plot the number of origins against the osmolarity and compare the results to Dai and Zhu 2018. As before the error was taken as the standard deviation for the individual measurements of $\bar{I}$ using the ten points for each condition shown in figure 3.45. We can see from this plot there is a discrepancy between the two data sets. Dai and Zhu 2018 show an almost doubling of the number of origins between 0 and 600mM NaCl, whereas my data shows a flat, to slightly upward trend with about a 20% difference over the same range.

Also in figure 3.49 I have plotted a measure of the number of origins per cell using the flow cytometry data. In particular I used the data from the run out experiments for each condition. In principle, after a run out each cell will contain a number of chromosomes equal to the number of origins and therefore by counting the average number of chromosomes per cell in the run out, one can determine the average number of origins the moment before the run out was started. As can be seen in figure 3.49 the two methods show fairly similar results with most points from the run outs coming within error of the data points computed using the C and D period estimates. This is relatively reassuring as it suggests that the data obtained by two different methods (flow cytometry and QPCR) seem to agree quite well.

3.4.5 The unit cell size at high osmolarity

At this stage we have now measured sufficient number of cell cycle parameters that we can turn our attention to the question of the unit cell size. Without a direct measurement of the unit cell size S_0 we cannot determine if the general growth law holds at high osmolarity. Instead, we can assume it holds, and measure the effect that would imply on the unit cell size. To do this we arrange the familiar general growth law equation into the following form:

$$S_0 = \frac{S}{2^{\frac{C+D}{\tau}}}. \tag{3.9}$$

We can then plot the unit cell size against osmolarity as in figure 3.50. As before an estimate of the error was obtained by calculating a S_0 value using each individual estimate for C shown in figure 3.45, as the error in C was the dominant source of error. As the unit cell size is dependent on the average size of cells, we can compute an estimate of S_0 using both the microscopy estimates of cell volume, and those obtained via the Coulter counter. Both results are shown in figure 3.50, and as can

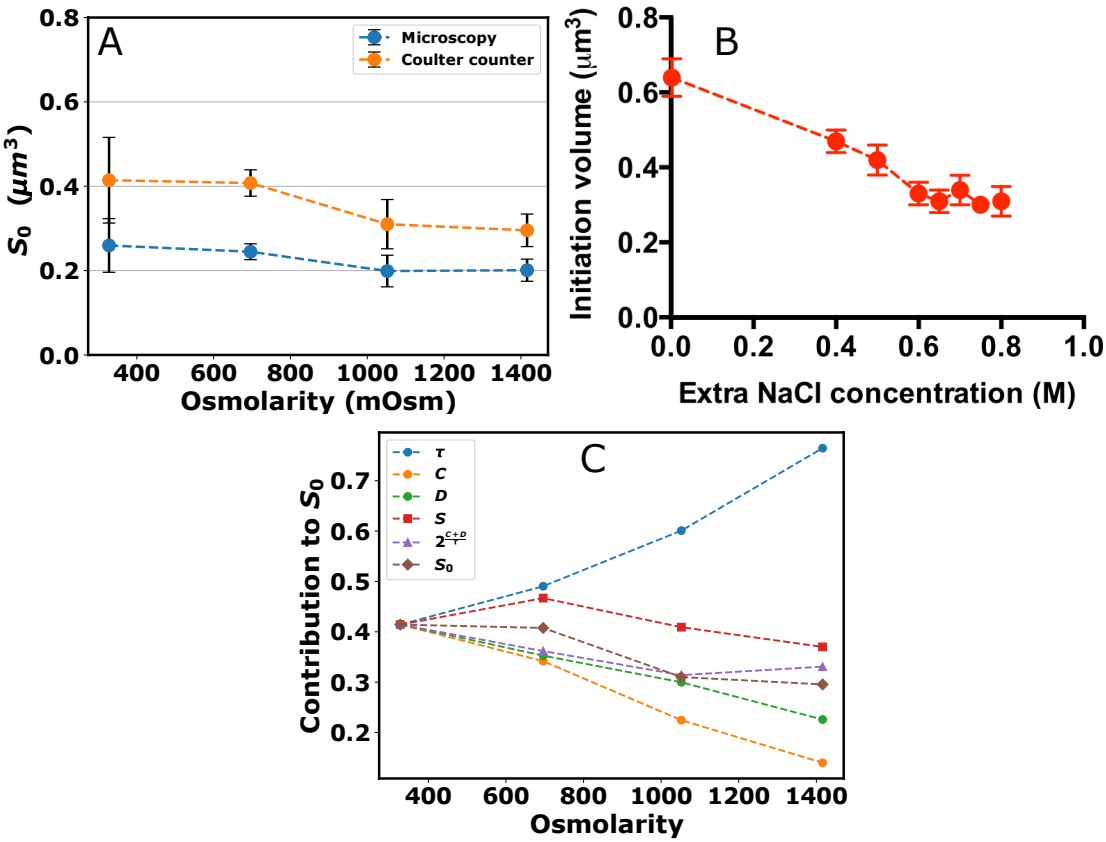

Figure 3.50: The unit cell size S_0 at high osmolarity. **A:** S_0 against osmolarity as computed using the measurement of cell volume as obtained by microscopy (blue points) and via the Coulter counter (orange points). Error bars represent standard deviations. **B:** S_0 against NaCl concentration as measured by Dai and Zhu 2018. **C:** Plot showing contribution to S_0 each variable makes. For each trace the value for the stated variable was varied while keeping all others fixed at their value for the normal osmolarity condition. The plot shows therefore, how S_0 would have varied with osmolarity, had only that particular variable changed. The brown diamonds represent the measured change in S_0.

be seen, while there is a difference in absolute value, the relative change with osmolarity is very similar between the two methods. We can see from the plot that there is about a 25% decrease in the value of S_0 over the measured range in osmolarity, this is significantly less than the trend observed by Dai and Zhu 2018, which shows an approximately 50% decrease in S_0 over the same range. It should also be noted that our absolute estimate for S_0 at normal osmolarity lies between 0.26 for the microscopy data and 0.41 for the Coulter counter data. Both of these values, agree

with the estimate of the unit cell size as determined by Si et al. 2017 (0.28) much more closely than the value obtained by Dai and Zhu 2018 (≈ 0.65).

In order to see which cell cycle parameter is most responsible for the changes to S_0 at high osmolarity we can examine figure 3.50:C. Here I have taken each variable (S, C,D and τ) and plotted how S_0 would change with osmolarity if all other variables were held constant at their value for normal osmolarity. I compare this to the actual observed change in S_0 which is shown in brown. From this I can conclude that the dominant variables that lead to to a decrease in S_0 at high osmolarity are the cell cycle parameters C and D.

3.4.6 Discussion and future work

In the previous section I examined the cell cycle parameters of *E. coli* cultures growing exponentially at different sodium chloride concentrations. In this section I will discuss these results, and suggest some interesting experiments going forward.

In section 3.4.1.2, I demonstrated how the C period of exponentially growing cells could be calculated using QPCR reactions. Also in this section, I described how the presence of outliers in the C_t values obtained from technical replicates, had a strong effect on the final value obtained for the C period. I consider this to be one of the main criticisms of applying the technique of QPCR to the determination of the C period. In equation 2.26, we can see that the C period is calculated by taking the primer efficiency (itself a product of C_t values) and raising it to the power of the C_t value. As a result of this, the C period is extremely sensitive to fluctuations in the estimate of C_t. Given the intrinsic variability present in QPCR reactions, from sources such as pipetting errors, contaminants in the DNA sample, and even well evaporation, it is not surprising that the C_t values showed such significant variation (Reiter and Pfaffl 2008; Svec et al. 2015). That being said, in this work I performed a few controls to check that data obtained was sensible. Firstly, I confirmed that the measured C period in the case of M63 media with glucose and cas amino acids was around 40 minutes. It is well recorded that media with doubling times less than 60 minutes have C periods around the 40 minute mark (Si et al. 2017; C. E. Helmstetter and Cooper 1968). As a second check I analysed the C_t values for run out cells.

These cells have in principle, the same number of origins as termini because all of their chromosomes are fully replicated. Consequently, if one takes the ratio of the C_t values obtained for the origin and termini primers, a value of 1 should be achieved. In practice I achieve a value very close to one of 0.996, which is a good check on the experimental protocol and analysis. Going forward, if more C period determinations were to be required, I would suggest increasing the number of technical replicates in an effort to reduce the effect of outliers, without having to remove them. Previous work has suggested that there is a benefit in increasing the number of technical replicates to as high as 16 (Svec et al. 2015). Alternatively, results could be compared with other methods of measuring the C period, such as radioactive pulse labelling as used by C. E. Helmstetter and Cooper 1968.

In section 3.4.1.2, I showed that as the NaCl concentraiton increases the C period also increases. I compared my findings to the results of Dai and Zhu 2018, and showed that there was relatively good agreement in the absolute values of the C periods between the two data sets. For example, both data sets showed an increase, from around 40 minutes at 0mM extra NaCl, to around 100 minutes at 600mM extra NaCl. Given that different media were used in the two experiments, this initial agreement may seem surprising. Indeed, in section 3.2.3, I showed that I observed a different trend between cell size and NaCl concentration compared to Dai and Zhu, and I suggested that the difference in media was a likely cause of this. However, while cell size is tied to media via the nutrient growth law, cell cycle times are constant for doubling times between 20 and 60 minutes. This may suggest, that the cell cycle parameters are set only by osmolarity, while other parameters carry a growth rate and osmolarity dependence. This would imply, that the relationship between cell size and osmolarity is dependent on the initial size of cells in the normal osmolarity condition. The replication time on the other hand, would not be dependent on the media, because it is not dependent on the growth rate. Therefore, the same replication time is achieved at a given osmolarity independent of the media used. Another datum in favour of this idea, is the similarity in osmolarity between M63 media with glucose and cas amino acids and LB ($330 mOsm$ and $240 mOsm$ respectively). In fact, the difference between the media is almost four times smaller than the $369 mOsm$, separating normal M63 media and M63 media with 200mM NaCl. Therefore, if the C period is dependent only on the osmolarity, then one would ex-

pect to see very similar values between my results and those of Dai and Zhu 2018. Going forward, it would be good to test this hypothesis by measuring the C period at different osmolarities in different media. This would allow one to determine if the C period is in fact dependent only on osmolarity.

In section 3.4.2, I used the C periods from section 3.4.1.2 and the DNA content measurments from section 3.3.2, to compute the D periods of cells grown at high osmolarity. As with the C periods, the results showed very good agreement with those obtained by Dai and Zhu 2018. Both data showed an increase from around 30 minutes at 0mM extra NaCl, to around 55 minutes at 600mM NaCl. This close agreement is actually quite surprising for two reasons. Firstly as with the C period, there is a difference in media between the two experiments which one might initially expect to lead to a difference in the absolute values of the D period. Secondly, the D period is dependent on the average DNA content per cell ($\bar{G}$) which I found to be essentially constant with NaCl concentration. Dai and Zhu 2018 on the other hand, found a large increase in cellular DNA content with increasing NaCl concentration. However, it appears that the differences in our measured doubling times τ, are sufficient such that the computed D periods end up being very similar over the 0-600mM range. This is an interesting observation, which may lend further credence to the idea that the cell cycle parameters are sensitive to the total osmolarity, independent of the media used. The arguments for this follow the exact same train of thought as above with regard to the C period. Consequently, going forward the natural thing to do, would be to measure how the D period varies with osmolarity using different media with different growth rates. This would allow a confirmation to be made on the validity of this idea.

In section 3.4.4, I computed the average number of origins per cell using both equation 3.8 and direct measurement from flow cytometry. I showed that both measurments suggested a basically flat trend with increasing NaCl concentration. This was in stark contrast to Dai and Zhu 2018, which showed that the number of origins doubled between 0 and 600mM NaCl. Given that the average number of origins is calculated from only the cell cycle time (τ_{cyc}) and τ, this suggests that differences in τ are responsible for the differences in outcome between my experiments and Dai and Zhu 2018. On one hand, this is surprising given the agreement in τ_{cyc} (figure 3.48).

On the other hand, given the differences in $\bar{G}$ between the two results, it follows that there would be a similarly large difference in the average number of origins. It is interesting to note that the average number of origins, and cell size are both altered by different media according to the nutrient growth law, while the C and D periods are not. Therefore, the fact that a different trend is observed for the number of origins at high osmolarity between my resuts and those of Dai and Zhu 2018, is at least consistent with the idea discussed above that the cell cycle parameters are dependent only on the osmolarity, while other parameters have both an osmolarity and growth rate dependence. Again, in the future it would be good to measure how the number of origins changes with osmolarity, at different media controlled growth rates to confirm this idea.

Finally, in section 3.4.5, I examined how the unit cell size changes as the NaCl concentration is increased. I showed that while my results, and those of Dai and Zhu 2018, both suggest a downward trend with increasing NaCl concentraiton, I observe a far less significant decrease. From figure 3.50 we can conclude that the reason for this, is that I observe a less sever decrease in cell size with osmolarity, and I do not observe the significant increase in origins that Dai and Zhu 2018 do. These factors compound to produce a far less significant decrease in the observed value of S_0. Going forward some interesting experiments would be to see how the value of S_0 changes at high osmolarity in different media, to determine if the observed difference between this study and others can be explained entirely by the choice of media. Additionally, in light of the discussion in section 3.2.3, it would be interesting to see how the unit cell size changes at high osmolarity in M63 media using different osmolytes such as sucrose or sorbitol, again this would shed light on how well osmolarity by itself uniquely determines cell physiology. Finally, another interesting experiment going forward would be to determine how the dry mass changes for cells grown in M63 media at high NaCl concentrations. Dai and Zhu 2018, showed that at high osmolarity the density of cells increase, such that that mass per origin remains constant even as the size per origin (S_0) decreases. By measuring the dry mass I could confirm if that holds true at different growth rates.

Chapter 4

Conclusions and further work

Bacteria, such as *E. coli*, are highly complex systems which must adapt to their environment in order to survive, and continue growing. One of the ways in which *E. coli*'s environment can change is by an increase in external osmolarity. Such changes are common in the environments where *E. coli* finds itself, such as in the digestive tracts off mammals. Consequently, *E. coli* has developed a complex response to hyperosmotic shocks, which allow it increase internal osmolyte concentration and restore turgor pressure, without severely damaging enzymatic function. Though the short term response of *E. coli* to such shocks is well studied, comparatively little work has been done in examining the steady state growth of *E. coli* at high osmolarity. In this thesis the aim has been to study *E. coli* cells at high osmolarity during balanced growth, and consequently determine what affect this condition has on their long term physiology. Furthermore, this task has been approached from the point of view of the general growth law, which is itself the product of a 60 year long endeavour to develop coarse grained biological models of the steady state behaviour of bacteria.

In section 2.1 of this thesis, I outlined how I developed the hardware and software for a custom microscope, which I then used to take images of exponentially growing *E. coli* cells, at various NaCl concentrations. These images were manually vetted, before being segmented using custom machine learning algorithms which I also developed, and outlined in section 3.2.1. The length, width, and volume of cells in these images were extracted using yet another custom algorithm. Using this pipeline of analysis, in section 3.2.1.1, I was a able to recreate the nutrient growth law, and show that cell size has an exponential dependence on growth rate under nutrient

limitation. I then showed that in agreement with previous work, I observe a decrease in cell size at high osmolarity, although to a much lesser extent than what has previously been reported (section 3.2.1.2). I confirmed these results in section 3.2.2 by appeal to measurments obtained using the high throughput Coulter counter. I showed that while the absolute volumes were not identical with those obtained from microscopy, the trend with increasing osmolarity was very similar.

In section 3.3 I discussed the measurment of cellular DNA content using flow cytometry. In this section I demonstrated how through careful callibration using slow growing cells metabolising mannose, one could determine the link between fluorescent signal and DNA content per cell. Utilising this I showed that I observe essentially no change in the DNA content per cell as osmolarity is increased, through the addition of NaCl. This result stands in stark contrast to previous findings which showed a doubling of the cellular DNA content over the range of osmolarities examined in this work.

In section 3.4 I showed that using QPCR one can measure the chromosome replication time, or C period, of exponentially growing cells. I also measured how the C period changed with increasing NaCl concentration. I found that increasing the NaCl concentration led to a significant increase in the C period. Not only did this increase in C match the trend found in previous work, but I also found good agreement in the absolute values of the C periods for each given concentration of NaCl. Using this result in combination with the measurment of average DNA content described above, I was able to compute how the division time (D period) changed with increasing NaCl concentration. I found that as with the measurment of the C period, there was good agreement both in terms of the upwards trend with increasing osmolarity, and in the absolute values measured. Using the combined measurments for the C and D periods, along with the measurments of growth rate obtained in section 3.1, I was able to compute the average number of origins of replication, and how this changed with osmolarity. I found that that there was very little change in the number of origins per cell as the NaCl concentration was increased. This finding was confirmed by comparison with flow cytometry data from run out experiments. Just as was the case with measurments of average DNA content, my findings here disagreed with those obtained previously, which showed a large increase in the number of origins as

the NaCl concentration increased. Finally, I combined measurments of the number of origins, with those of average cell size described above, to compute how the unit cell size (S_0) changed at high NaCl concentrations. Once again, my results agreed with those obtained previously, showing a decrease in S_0 as the NaCl concentration increased, although I observed only about 50% of the decrease previously reported, over the same change in osmolarity.

Taken together, the results presented here suggest that cell physiology is severely perturbed by steady state growth at high NaCl concentrations. Growth rate, and cell cycle time (τ_{cyc}) both show a strong dependence with increasing salt concentration. Conversely, the cell size shows only a weak trend, while the DNA content and number origins are barely altered as the NaCl concentration increases. Throughout this thesis I have made constant comparison to Dai and Zhu 2018, which to my knowledge is the only other work to address the steady state behaviour of cells at high osmolarity. In doing so I have generated a mixture of: good agreements as in the case of the cell cycle parameters; partial agreements as in the cas of cell size and unit cell size; and no agreement in the case of DNA content and number of origins. I have explained that Dai and Zhu 2018 used LB media for their study, and that as a result, comparison is made more complicated. In section 3.2.3 I suggested that a possible way to interpret the differences between my work and theirs, is that the cell cycle parameters are increased as a function of the osmolarity only, while the cell size, growth rate and DNA content are variables which depend on both the osmolarity and the media induced growth rate. Going forward then, I have suggested that the next step should be to measure the parameters of the general growth law at high salt concentration, in media with different growth rates in order to test this hypothesis. In this work I attempted to grow cells at high NaCl concentration using LB, and as shown in section 3.2.2.7, I was unable to do so without causing severe morphological changes to cells. I have shown that the phenomenon I observed here is reproducible across strains, and other high growth rate media, but I have been unable to find a compelling explanation for why this was not observed in previous work. It is possible that differences in how cells were cultured between the two studies are relevant, and future work should explore this to help narrow down the possible causes for this strange morphology.

It has been known for some time that the average cell size changes as cells move between the different phases of the growth curve. It is understood that as cell number increases in a batch culture, eventually the growth rate will slow, cell size will diminish, and cells will escape from balanced growth. In fact, previous work showing that this can happen as early as OD_{600} 0.3 in LB media was one of the primary reasons I opted to use defined M63 media for this work. Ironic then, that in section 3.2.2.3 I showed that as early as OD_{600} 0.1 there is a significant reduction in cell size in M63 media with glucose and cas amino acids. This drop in cell size must by definition accompany a departure from steady state. Interestingly, this apparent departure from steady state is not at all noticeable from the growth curves shown in section 3.1. Given that examination of the growth curve is the primary way in which experimenters determine the presence of steady state, this suggests that a standard protocol for verifying the OD_{600} at which steady state ends is required. This is actually quite a difficult task because steady state is not defined by the presence of any one phenomenon, instead it is defined as the absence of change in any intensive or extensive property. Therefore, without measuring a large number of physiological parameters across the growth curve, it is difficult to determine with certainty when steady state is, or is not, being achieved. That being said, from both previous work, and this thesis, it appears that measurement of cell size provides a more sensitive signal, earlier in the growth curve that the end of steady state has been reached, compared to analysis of the growth rate. Therefore, going forward I suggest that whenever new media is being used, the cell size should be examined to determine the range of OD_{600} values where steady state might apply.

One of the most interesting results of this thesis is described in section 3.2.2.5, where I showed how cell size changes as osmolarity is increased using sorbitol instead of sodium chloride. I showed that at high sorbitol concentrations, cell size is reduced to a much greater extent than caused by the equivalent osmolarity of NaCl. Furthermore, I showed that while high NaCl concentrations lead to a significant change in the shape of the cell size distributions, high osmolarity as achieved by increasing concentrations of sorbitol leaves the shape of these distributions virtually unchanged. This is compelling evidence that the effect of osmolarity is not independent of the osmolyte used to achieve it. This is an important result. As I mention in section 3.2.3, it is very common in the literature to see investigations on the effect of "high

osmolarity" conditions, where in fact only the effect of a single osmolyte is examined. These investigations would take on new interpretation if in fact osmolarity is not a single variable which defines cell physiology, and instead there is also a dependency on the type of osmolyte used. Furthermore, combining this result with the discussion above on the comparison between my results and those obtained from Dai and Zhu 2018, suggests that the effect on cell physiology of a high osmolarity condition may be dependent on the nutritional capacity of the media also. This suggests that rather than the effect of high solute concentrations being dependent on the single variable of osmolarity, the exact physiological response, might actually be the result of a combination of three parameters, namely: the growth rate, osmolyte used, and osmolarity.

Appendix A

Code

All the code used in this work can be found on my GitHub page at: https://github.com/DarioMiroli. All of the individual repositories are listed below. In all cases python was the primary programming language.

The software used to manually select cells and extract lengths, widths and volumes can be found at:
https://github.com/DarioMiroli/PillizotaGroupImageAnalysisCode

The neural networks used for image segmentation can be found at: https://github.com/DarioMiroli/ImageSegmentationWithKeras

The GUI I developed for analysing flow cytometry data can be found at: https://github.com/DarioMiroli/FlowCytometry_Analyser

The GUI I developed for analysing plate data from QPCR experiments can be found at: https://github.com/DarioMiroli/QPCR_Analyser

The software I developed to analyse growth curves can be found at: https://github.com/DarioMiroli/GrowthCurve_Analyser

Bibliography

Adan, Aysun et al. (2017). "Flow cytometry: basic principles and applications". In: *Critical reviews in biotechnology* 37.2, pp. 163–176.

Adiciptaningrum, Aileen et al. (2015). "Stochasticity and homeostasis in the E. coli replication and division cycle". In: *Scientific reports* 5, p. 18261.

Ambriz-Aviña, Verónica, Jorge A Contreras-Garduño, and Mario Pedraza-Reyes (2014). "Applications of flow cytometry to characterize bacterial physiological responses". In: *BioMed research international* 2014.

Anderson, John L and John A Quinn (1971). "The relationship between particle size and signal in Coulter-type counters". In: *Review of Scientific Instruments* 42.8, pp. 1257–1258.

Azeroglu, Benura et al. (2016). "RecG directs DNA synthesis during double-strand break repair". In: *PLoS genetics* 12.2, e1005799.

Baba, Tomoya et al. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection". In: *Molecular systems biology* 2.1.

Bailey, James E et al. (1977). "Characterization of bacterial growth by means of flow microfluorometry". In: *Science* 198.4322, pp. 1175–1176.

Bailey, Matthew W et al. (2014). "Evidence for divisome localization mechanisms independent of the Min system and SlmA in Escherichia coli". In: *PLoS genetics* 10.8, e1004504.

Basan, Markus et al. (2015). "Inflating bacterial cells by increased protein synthesis". In: *Molecular systems biology* 11.10, p. 836.

Berge, Lars Inge (1990). "Particle transit time distributions in single pores by the resistive pulse technique". In: *Journal of colloid and interface science* 135.1, pp. 283–293.

Blattner, Frederick R et al. (1997). "The complete genome sequence of Escherichia coli K-12". In: *science* 277.5331, pp. 1453–1462.

Boyum, AI (1968). "Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g". In: *Scand. J. Clin. Lab. Invest. Suppl.* 97, p. 77.

Boz, Olcay and Donald Hillman (2000). *Converting a trained neural network to a decision tree dectext-decision tree extractor*. Citeseer.

Bremer, H and G Churchward (1977). "An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions". In: *Journal of Theoretical Biology* 69.4, pp. 645–654.

Buda, Renata et al. (2016). "Dynamics of Escherichia coli's passive response to a sudden decrease in external osmolarity". In: *Proceedings of the National Academy of Sciences* 113.40, E5838–E5846.

Campbell, Allan (1957). "Synchronization of cell division". In: *Bacteriological reviews* 21.4, p. 263.

Carpenter, Anne E et al. (2006). "CellProfiler: image analysis software for identifying and quantifying cell phenotypes". In: *Genome biology* 7.10, R100.

Cashel, M et al. (1996). "Escherichia coli and Salmonella: cellular and molecular biology". In:

Cayley, D Scott, Harry J Guttman, and M Thomas Record Jr (2000). "Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress". In: *Biophysical Journal* 78.4, pp. 1748–1764.

Cayley, Scott and M Thomas Record Jr (2004). "Large changes in cytoplasmic biopolymer concentration with osmolality indicate that macromolecular crowding may regulate protein–DNA interactions and growth rate in osmotically stressed Escherichia coli K-12". In: *Journal of Molecular Recognition* 17.5, pp. 488–496.

Chien, Alice, David B Edgar, and John M Trela (1976). "Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus." In: *Journal of bacteriology* 127.3, pp. 1550–1557.

Chollet, François et al. (2015). *Keras*. `https://keras.io`.

Chung, Hak Suk et al. (2009). "Rapid β-lactam-induced lysis requires successful assembly of the cell division machinery". In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21872–21877.

Cireşan, Dan, Ueli Meier, and Jürgen Schmidhuber (2012). "Multi-column deep neural networks for image classification". In: *arXiv preprint arXiv:1202.2745*.

Ciresan, Dan et al. (2012). "Deep neural networks segment neuronal membranes in electron microscopy images". In: *Advances in neural information processing systems*, pp. 2843–2851.

Cooper, Stephen (1993). "The origins and meaning of the Schaechter-Maaloe-Kjeldgaard experiments". In: *Microbiology* 139.6, pp. 1117–1124.

Cooper, Stephen and Charles E Helmstetter (1968). "Chromosome replication and the division cycle of Escherichia coliBr". In: *Journal of molecular biology* 31.3, pp. 519–540.

Coulter, Beckman (2010). *Multisizer 4 Particle Analyzer, User's manual*.

Cullum, J and M Vicente (1978). "Cell growth and length distribution in Escherichia coli." In: *Journal of bacteriology* 134.1, pp. 330–337.

Dai, Xiongfeng and Manlu Zhu (2018). "High Osmolarity Modulates Bacterial Cell Size through Reducing Initiation Volume in Escherichia coli". In: *mSphere* 3.5, e00430–18.

Dai, Xiongfeng, Manlu Zhu, et al. (2018). "Slowdown of translational elongation in Escherichia coli under hyperosmotic stress". In: *MBio* 9.1, e02375–17.

Datsenko, Kirill A and Barry L Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products". In: *Proceedings of the National Academy of Sciences* 97.12, pp. 6640–6645.

Delamarche, Christian et al. (1999). "Visualization of AqpZ-mediated water permeability in Escherichia coli by cryoelectron microscopy". In: *Journal of bacteriology* 181.14, pp. 4193–4197.

Dick, David Andrew Thomas (1966). *Cell water*. Butterworths.

Donachie, William D (1968). "Relationship between cell size and time of initiation of DNA replication". In: *Nature* 219.5158, p. 1077.

Dumoulin, Vincent and Francesco Visin (2016). "A guide to convolution arithmetic for deep learning". In: *arXiv preprint arXiv:1603.07285*.

Ghatak, Sankha et al. (2018). "The y-ome defines the thirty-four percent of Escherichia coli genes that lack experimental evidence of function". In: *BioRxiv*, p. 328591.

Göransson, Billy (1990). "Improved Accuracy in the Measurement of Particle Size Distribution with a coulter counter equipped with a hydrodynamically focused aperture". In: *Particle & Particle Systems Characterization* 7.1-4, pp. 6–10.

Guberman, Jonathan M et al. (2008). "PSICIC: noise and asymmetry in bacterial division revealed by computational image analysis at sub-pixel resolution". In: *PLoS computational biology* 4.11, e1000233.

Harris, Leigh K and Julie A Theriot (2016). "Relative rates of surface and volume synthesis set bacterial cell size". In: *Cell* 165.6, pp. 1479–1492.

— (2018). "Surface area to volume ratio: A natural variable for bacterial morphogenesis". In: *Trends in microbiology* 26.10, pp. 815–832.

Haswell, Elizabeth S, Rob Phillips, and Douglas C Rees (2011). "Mechanosensitive channels: what can they do and how do they do it?" In: *Structure* 19.10, pp. 1356–1369.

Helmstetter, Charles E (1967). "Rate of DNA synthesis during the division cycle of Escherichia coli B/r". In: *Journal of Molecular Biology* 24.3, pp. 417–427.

— (2015). "A ten-year search for synchronous cells: obstacles, solutions, and practical applications". In: *Frontiers in microbiology* 6, p. 238.

Helmstetter, Charles E and Stephen Cooper (1968). "DNA synthesis during the division cycle of rapidly growing Escherichia coliBr". In: *Journal of molecular biology* 31.3, pp. 507–518.

Helmstetter, C et al. (1968). "On the bacterial life sequence". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 33. Cold Spring Harbor Laboratory Press, pp. 809–822.

Henrici, AT (1928). "Morphologic Variation and the Rate of growth of Bacteria". In:

Higuchi, Russell et al. (1993). "Kinetic PCR analysis: real-time monitoring of DNA amplification reactions". In: *Bio/technology* 11.9, p. 1026.

Hill, Norbert S et al. (2012). "Cell size and the initiation of DNA replication in bacteria". In: *PLoS genetics* 8.3, e1002549.

Huang, Bo et al. (2008). "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy". In: *Science* 319.5864, pp. 810–813.

Huang, Jiaqing et al. (2015). "3D multifocus astigmatism and compressed sensing (3D MACS) based superresolution reconstruction". In: *Biomedical optics express* 6.3, pp. 902–917.

Hutchison, Clyde A et al. (2016). “Design and synthesis of a minimal bacterial genome”. In: *Science* 351.6280, aad6253.

Jansson, Linda and Johannes Hedman (2019). “Challenging the proposed causes of the PCR plateau phase”. In: *Biomolecular detection and quantification* 17, p. 100082.

Jun, Suckjoon et al. (2018). “Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review”. In: *Reports on Progress in Physics* 81.5, p. 056601.

Karr, Jonathan R et al. (2012). “A whole-cell computational model predicts phenotype from genotype”. In: *Cell* 150.2, pp. 389–401.

Kingma, Diederik P and Jimmy Ba (2014). “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980*.

Koch, AL and M Schaechter (1962). “A model for statistics of the cell division process”. In: *Microbiology* 29.3, pp. 435–454.

Koch, Arthur L (1970). “Turbidity measurements of bacterial cultures in some available commercial instruments”. In: *Analytical biochemistry* 38.1, pp. 252–259.

— (1990). “Growth and form of the bacterial cell wall”. In: *American Scientist* 78.4, pp. 327–341.

Koch, Arthur L, Michael L Higgins, and Ronald J Doyle (1982). “The role of surface stress in the morphology of microbes”. In: *Microbiology* 128.5, pp. 927–945.

Konopka, Michael C et al. (2009). “Cytoplasmic protein mobility in osmotically stressed Escherichia coli”. In: *Journal of bacteriology* 191.1, pp. 231–237.

Kralik, Petr and Matteo Ricchi (2017). “A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything”. In: *Frontiers in microbiology* 8, p. 108.

Krämer, Christina EM, Wolfgang Wiechert, and Dietrich Kohlheyer (2016). “Time-resolved, single-cell analysis of induced and programmed cell death via non-invasive propidium iodide and counterstain perfusion”. In: *Scientific reports* 6, p. 32104.

Kubitschek, Herbert E (1958). “Electronic counting and sizing of bacteria”. In: *Nature* 182.4630, pp. 234–235.

Kung, Ching (2005). “A possible unifying principle for mechanosensation”. In: *Nature* 436.7051, p. 647.

Lane-Claypon, Janet E (1909). "Multiplication of bacteria and the influence of temperature and some other conditions thereon". In: *Epidemiology & Infection* 9.2, pp. 239–248.

Larsen, PI et al. (1987). "Osmoregulation in Escherichia coli by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose". In: *Archives of microbiology* 147.1, pp. 1–7.

Laub, Michael T, Lucy Shapiro, and Harley H McAdams (2007). "Systems biology of Caulobacter". In: *Annu. Rev. Genet.* 41, pp. 429–441.

LeCun, Yann et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Lopez-Amoros, R, J Comas, and J Vives-Rego (1995). "Flow cytometric assessment of Escherichia coli and Salmonella typhimurium starvation-survival in seawater using rhodamine 123, propidium iodide, and oxonol." In: *Appl. Environ. Microbiol.* 61.7, pp. 2521–2526.

Maia, Margarida RG et al. (2016). "Simple and versatile Turbidimetric monitoring of bacterial growth in liquid cultures using a customized 3D printed culture tube holder and a miniaturized spectrophotometer: application to facultative and strictly anaerobic bacteria". In: *Frontiers in microbiology* 7, p. 1381.

Martin Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: http://tensorflow.org/.

McKee, CS (1929). "Morphologic Variation and the Rate of Growth of Bacteria". In: *Canadian Medical Association Journal* 21.1, p. 128.

McLaggan, Debbie et al. (1994). "Interdependence of K+ and glutamate accumulation during osmotic adaptation of Escherichia coli." In: *Journal of Biological Chemistry* 269.3, pp. 1911–1917.

Milo, Ron (2013). "What is the total number of protein molecules per cell volume? A call to rethink some published values". In: *Bioessays* 35.12, pp. 1050–1055.

Monod, Jacques (1949). "The growth of bacterial cultures". In: *Annual review of microbiology* 3.1, pp. 371–394.

Mullis, Kary et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 51. Cold Spring Harbor Laboratory Press, pp. 263–273.

Novick, Aaron and Leo Szilard (1950). "Experiments with the chemostat on spontaneous mutations of bacteria". In: *Proceedings of the National Academy of Sciences of the United States of America* 36.12, p. 708.

O'Brien, Edward J, Jonathan M Monk, and Bernhard O Palsson (2015). "Using genome-scale models to predict biological capabilities". In: *Cell* 161.5, pp. 971–987.

Okumus, Burak et al. (2016). "Mechanical slowing-down of cytoplasmic diffusion allows in vivo counting of proteins in individual cells". In: *Nature communications* 7, p. 11641.

Ostriker, Jeremiah P (1971). "The nature of pulsars". In: *Scientific American* 224.1, pp. 48–63.

Paintdakhi, Ahmad et al. (2016). "Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis". In: *Molecular microbiology* 99.4, pp. 767–777.

Pilizota, Teuta, Thomas Bilyard, et al. (2007). "A programmable optical angle clamp for rotary molecular motors". In: *Biophysical journal* 93.1, pp. 264–275.

Pilizota, Teuta and Joshua W Shaevitz (2012). "Fast, multiphase volume adaptation to hyperosmotic shock by Escherichia coli". In: *PLoS One* 7.4, e35205.

— (2013). "Plasmolysis and cell shape depend on solute outer-membrane permeability during hyperosmotic shock in E. coli". In: *Biophysical journal* 104.12, pp. 2733–2742.

— (2014). "Origins of Escherichia coli growth rate and cell shape changes at high external osmolality". In: *Biophysical journal* 107.8, pp. 1962–1969.

Powell, EO (1956). "Growth rate and generation time of bacteria, with special reference to continuous culture". In: *Microbiology* 15.3, pp. 492–511.

Qian, Ning (1999). "On the momentum term in gradient descent learning algorithms". In: *Neural networks* 12.1, pp. 145–151.

Reiter, M and MW Pfaffl (2008). "Effects of plate position, plate type and sealing systems on real-time PCR results". In: *Biotechnology & Biotechnological Equipment* 22.3, pp. 824–828.

Reppas, Christos et al. (2015). "Characterization of contents of distal ileum and cecum to which drugs/drug products are exposed during bioavailability/bioequivalence studies in healthy adults". In: *Pharmaceutical research* 32.10, pp. 3338–3349.

Reuter, Marcel et al. (2014). "Mechanosensitive channels and bacterial cell wall integrity: does life end with a bang or a whimper?" In: *Journal of the Royal Society Interface* 11.91, p. 20130850.

Rojas, Enrique, Julie A Theriot, and Kerwyn Casey Huang (2014a). "Response of Escherichia coli growth rate to osmotic shock". In: *Proceedings of the National Academy of Sciences* 111.21, pp. 7807–7812.

— (2014b). "Response of Escherichia coli growth rate to osmotic shock". In: *Proceedings of the National Academy of Sciences* 111.21, pp. 7807–7812.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Sands, Jeff M and Harold E Layton (2009). "The physiology of urinary concentration: an update". In: *Seminars in nephrology*. Vol. 29. 3. Elsevier, pp. 178–195.

Schaechter, Moselio, Ole Maaløe, and Niels O Kjeldgaard (1958). "Dependency on medium and temperature of cell size and chemical composition during balanced growth of Salmonella typhimurium". In: *Microbiology* 19.3, pp. 592–606.

Schindelin, Johannes et al. (2012). "Fiji: an open-source platform for biological-image analysis". In: *Nature methods* 9.7, p. 676.

Schoemaker, JM, RC Gayda, and A Markovitz (1984). "Regulation of cell division in Escherichia coli: SOS induction and cellular location of the sulA protein, a key to lon-associated filamentation and death." In: *Journal of bacteriology* 158.2, pp. 551–561.

Schwartz, Abe et al. (2004). "Formalization of the MESF unit of fluorescence intensity". In: *Cytometry Part B: Clinical Cytometry: The Journal of the International Society for Analytical Cytology* 57.1, pp. 1–6.

Scientific, ThermoFisher (2019). *How a Flow Cytometer Works*. URL: https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/cell-analysis-learning-center/molecular-probes-school-of-fluorescence/flow-cytometry-basics/flow-cytometry-fundamentals/how-flow-cytometer-works.html (visited on 05/22/2019).

Scott, Matthew et al. (2010a). "Interdependence of cell growth and gene expression: origins and consequences". In: *Science* 330.6007, pp. 1099–1102.

— (2010b). “Interdependence of cell growth and gene expression: origins and consequences”. In: *Science* 330.6007, pp. 1099–1102.

Serres, Margrethe H et al. (2001). “A functional update of the Escherichia coli K-12 genome”. In: *Genome biology* 2.9, research0035–1.

Sezonov, Guennadi, Danièle Joseleau-Petit, and Richard d’Ari (2007). “Escherichia coli physiology in Luria-Bertani broth”. In: *Journal of bacteriology* 189.23, pp. 8746–8749.

Sharp, RJ (1984). “The preservation of genetically unstable microorganisms and the cryopreservation of fermentation seed cultures.” In: *Advances in biotechnological processes* 3, p. 81.

Si, Fangwei et al. (2017). “Invariance of initiation mass and predictability of cell size in Escherichia coli”. In: *Current Biology* 27.9, pp. 1278–1287.

Silhavy, Thomas J, Daniel Kahne, and Suzanne Walker (2010). “The bacterial cell envelope”. In: *Cold Spring Harbor perspectives in biology* 2.5, a000414.

Skarstad, Kirsten, Erik Boye, and Harald B Steen (1986). “Timing of initiation of chromosome replication in individual Escherichia coli cells.” In: *The EMBO journal* 5.7, pp. 1711–1717.

Sprouffske, Kathleen, José Aguilar-Rodríguez, and Andreas Wagner (2016). “How archiving by freezing affects the genome-scale diversity of Escherichia coli populations”. In: *Genome biology and evolution* 8.5, pp. 1290–1298.

Stevenson, Keiran et al. (2016). “General calibration of microbial growth in microplate readers”. In: *Scientific reports* 6, p. 38828.

Svec, David et al. (2015). “How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments”. In: *Biomolecular detection and quantification* 3, pp. 9–16.

Taheri-Araghi, Sattar et al. (2015). “Cell-size control and homeostasis in bacteria”. In: *Current Biology* 25.3, pp. 385–391.

Taylor, Sean et al. (2010). “A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines”. In: *Methods* 50.4, S1–S5.

Trueba, Frank J et al. (1982). “Effects of temperature on the size and shape of Escherichia coli cells”. In: *Archives of microbiology* 131.3, pp. 235–240.

Tuson, Hannah H et al. (2012). “Measuring the stiffness of bacterial cells from growth rates in hydrogels of tunable elasticity”. In: *Molecular microbiology* 84.5, pp. 874–891.

Ursell, Tristan et al. (2017). "Rapid, precise quantification of bacterial cellular dimensions across a genomic-scale knockout library". In: *BMC biology* 15.1, p. 17.

Van Valen, David A et al. (2016). "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments". In: *PLoS computational biology* 12.11, e1005177.

Vats, Purva, Yu-Ling Shih, and Lawrence Rothfield (2009). "Assembly of the MreB-associated cytoskeletal ring of Escherichia coli". In: *Molecular microbiology* 72.1, pp. 170–182.

Veal, DA et al. (2000). "Fluorescence staining and flow cytometry for monitoring microbial cells". In: *Journal of immunological methods* 243.1-2, pp. 191–210.

Vejborg, Rebecca Munk and Per Klemm (2009). "Cellular chain formation in Escherichia coli biofilms". In: *Microbiology* 155.5, pp. 1407–1417.

Volkmer, Benjamin and Matthias Heinemann (2011a). "Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling". In: *PloS one* 6.7, e23126.

— (2011b). "Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling". In: *PloS one* 6.7, e23126.

Wales, M and JN Wilson (1961). "Theory of coincidence in Coulter particle counters". In: *Review of scientific instruments* 32.10, pp. 1132–1136.

Wang, Lili et al. (2002). "Quantitating fluorescence intensity from fluorophores: practical use of MESF values". In: *Journal of research of the National Institute of Standards and Technology* 107.4, p. 339.

Wang, Ping et al. (2010). "Robust growth of Escherichia coli". In: *Current biology* 20.12, pp. 1099–1103.

Wood, Janet M (2011). "Bacterial osmoregulation: a paradigm for the study of cellular homeostasis". In: *Annual review of microbiology* 65, pp. 215–238.

Yang, Da et al. (2018). "Analysis of factors limiting bacterial growth in PDMS mother machine devices". In: *Frontiers in microbiology* 9, p. 871.

Yao, Zhizhong, Daniel Kahne, and Roy Kishony (2012). "Distinct single-cell morphological dynamics under beta-lactam antibiotics". In: *Molecular cell* 48.5, pp. 705–712.

Young, Jonathan W et al. (2012). "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy". In: *Nature protocols* 7.1, p. 80.

Zipper, Hubert et al. (2004). "Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications". In: *Nucleic acids research* 32.12, e103–e103.