Quantifying the cell division rate of the crenarchaeon *Sulfolobus acidocaldarius* using live-cell images and molecular dynamics simulations

A thesis presented for an MSci degree in Physics

Tina Yao

Supervisors:
Dr. Anđela Šarić and Prof. Buzz Baum

University College London
Department of Physics and Astronomy
March 2020

Abstract

Cell division is a fundamental process to life, without which organisms cannot grow, repair or reproduce. The final stage of cell division in animal cells, separating the two divided cells by cutting the cell membrane between them. The group of proteins required for this process is called ESCRT-III. As yet, the underlying physical mechanism for this cutting process remains unknown. Faults in this process can lead to cell abnormalities such as tumour growth. Therefore, understanding these proteins is a vital area of research. Recent evidence shows that the proteins that drive the cell division in the single-celled organism *Sulfobolus Acidocaldarius* is genetic homologues of the ESCRT-III proteins.[1] As such, understanding cell division in these relatively simple cells can help us understand processes in more complex cells.

Previous difficulties associated with live-imaging *S. Acidocaldarius* have been overcome by Pulschen et al.[2] with the custom-built "Sulfoscope". This project uses time-lapse movies taken by Pulschen et al. and measures the midcell diameter of the dividing cells over time. These measurements are fitted with power curves of the form $d(t) = ct^\alpha + b$ to quantify "shape" and the average rate of the division, α and c. The shape of division describes how the rate of division changes over time where $\alpha = 1$ represents a constant rate of division. Analysis of two different phenotypes of *S. Acidocaldarius* provided average values of $\alpha = 0.96 \pm 0.13$ and 1.01 ± 0.16 which strongly suggests that *S. Acidocaldarius* divides at a linear rate, in agreement with previous research. The rates of division was measured to be $c = -0.04 \pm 0.02\ \mu m\ min^{-1}$ and $c = -0.05 \pm 0.02\ \mu m\ min^{-1}$ which is approximately a decrease of 5% per minute.

The Šarić group has previously developed a successful cell division simulation model consisting of a membrane and a constricting protein filament, based on experimental observations of *S. Acidocaldarius*.[3] This project furthers this work by adding cytoplasm particles into the simulated cell. Comparisons were made between the original and the modified cell by measuring the midcell diameter of the simulated cells over time for multiple configurations. The addition of cytoplasm particles have been shown to affect the successful outcome of the simulation. Furthermore, simulations show that the greater the number of cytoplasm particles within the cell, the slower the rate of division.

UCL DEPARTMENT OF PHYSICS & ASTROMOMY
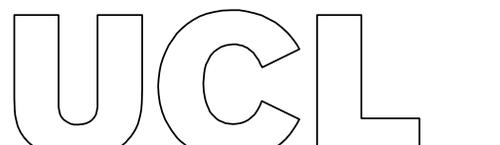
Submission of coursework for Physics and Astronomy course PHAS0097/PHAS0048

Please sign, date and return this form with your coursework by the specified deadline.

DECLARATION OF OWNERSHIP

I confirm that I have read and understood the guidelines on plagiarism, that I understand the meaning of plagiarism and that I may be penalised for submitting work that has been plagiarised.

I confirm that all work will also be submitted electronically and that this can be checked using the JISC detection service, Turnitin®.

I understand that the work cannot be assessed unless both hard copy and electronic versions of the work are handed in.

I declare that all material presented in the accompanying work is entirely my own work except where explicitly and individually indicated and that all sources used in its preparation and all quotations are clearly cited.

Should this statement prove to be untrue, I recognise the right of the Board of Examiners to recommend what action should be taken in line with UCL's regulations.

Signed

PrintName   TINA YAO

Dated   04/04/2020

Title	Date Received	Examiner	Examiner's Signature	Mark

Contents

1   Introduction

Archaea represents one of the three domains of life alongside Eukaryota and Bacteria. Archaea represent a diverse group of single-celled organisms that live in extreme conditions and have a unique mix of bacterial and eukaryotic traits. The majority of Archaeons are further classified into two main phyla Euryarchaeota and Crenarchaeota.[4] Bacteria and Euryarchaeota utilise the same proteins for cell division.[5] Recent evidence shows that the cell division proteins in crenarchaeons are genetic homologues to membrane cutting proteins in eukaryotes.[1] In eukaryotes, these proteins perform a wide range of membrane cutting roles that include cytokinesis, the last stage of cell division. The failure of cytokinesis can lead to the development of tumour cells. It is therefore necessary to understand how these ESCRT-III proteins drive cytokinesis.[6]

As a consequence of their similarities, the understanding gained from the underlying physical mechanism of cell division in simple crenarchaeons can be used to understand cytokinesis in, the more complex, eukaryotes.

The overall aim of this project is to try to gain further understanding of cell division in the crenarchaeon *Sulfolobus acidocaldarius* in hopes of applying the knowledge to membrane cutting in eukaryotes. Experimental observations and computer simulations can be used collaboratively to provide insight into this process.

In this project, the midcell diameter for live and simulated cells were measured over time. The midcell diameter is defined as the diameter across the cell's dividing region, Fig. 1.
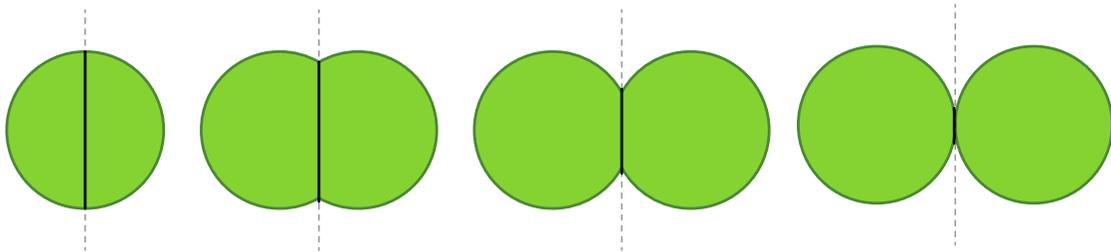
Figure 1: A diagram defining the midcell diameter and shows how it changes over division time.

This diameter decreases during cell division, and by measuring this diameter, rate of cell division can be determined and, in turn, provide insight into the dynamics within the cell.

Prior to this project, the midcell diameter was measured for live and simulated dividing *S. acidocaldarius* cells and the results are shown in Fig. 2.

Fig. 2 shows a disagreement between the two measurements where the experimental evidence shows that the cells divide linearly, i.e. at a constant rate. However, the rate of division of the simulated cell slows down over time. Although the diameter of the simulated cell was measured over simulation time, it was, at the time, not possible to measure the diameter of the live cells over real-time. Instead, the midcell diameter of a large sample of cells was measured in one snapshot of time, and the results
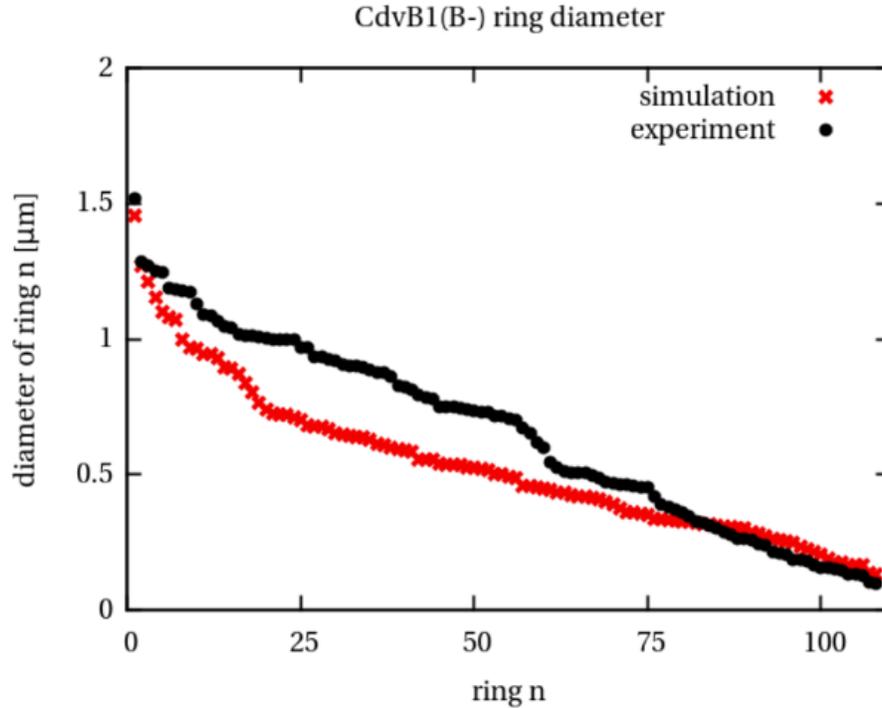
Figure 2: A graph showing midcell diameter over time as a cell divide. Results are gathered from previous experimental measurements and simulations. (Figure by Risa et al.)[3]

were ordered in size.

New advances in live-cell imaging and fluorescence microscopy means that *S. acidocaldarius* cells can be imaged over time. One of the objectives of this project is to verify whether *S. acidocaldarius* cells divide linearly over time by measuring the midcell diameter of cells over time of these live-cell images using image processing software.

In previous research, a simulated cell model was developed based on the experimental observations of *S. acidocaldarius* and proved useful in examining the internal mechanics of division.[3] Simulation experiments provide useful cross-reference for experimental observations. The simulated cell was modelled only to contain solvent, whereas, in reality, cells are filled with cytoplasm that contains lots of biological molecules. The second objective of this project is the addition of "cytoplasm" particles to the existing cell model to improve the model and to try to account for the discrepancy in Fig. 2.

2 Measurement of Live-Cell Images

2.1 Background

2.1.1 Cell Division

Cell division protein complexes are well established in Bacteria, Eukaryota and Euryarchaeota. Cell division machinery was recently discovered in the crenarchaeon *S. acidocaldarius* in 2008.[1] Protein complexes responsible for division include: FtsZ for bacteria and euryarchaeons,[5] ESCRT-III for eukaryotes[7] and CdvB -1,-2[1] for crenarchaeons. They all similarly drive cell scission; proteins polymerise as a spiral or ring at the midbody of the dividing cells before producing a constriction force that causes abscission. How these proteins produce a constricting force that drives cell division remains unknown.[8,9 ,10]

2.1.2 Eukaryotic Cell Division

Eukaryotic cells contain membrane-bound compartments and are typically $10 - 100\mu m$ in diameter. eukaryotes can reproduce asexually, producing two genetically identical daughter cells and sexually, producing four genetically similar daughter cells.[11] The Eukaryotic cell cycle is divided into phases called G1, S, G2, M and cytokinesis. The order of these phases is regulated by cyclin-CDK complexes which are regulating complexes not found in other organisms.[12] Cytokinesis is the final stage in cell division where the membrane between the two divided cells are cut, to separate them.[13]

The cutting of the cell membrane during cytokinesis is performed by a protein complex called ESCRT-III. ESCRT proteins (-I, -II, -III, Vsp4) play other vital roles in eukaryotes such as viral budding[14] and multivesicular body (MVB) biogenesis.[15] The importance of the roles performed by ESCRT-III delineate the importance of researching how it works. ESCRT-III is thought to be the only protein complex in eukaryotes that can cut cell membranes protruding away from the cytoplasm.[16] ESCRT-III has a short lifespan and is too small to see on a cellular membrane using standard measurements hence deep-etch electron microscopy and gene knockdown were used to probe the structure of ESCRT-III.[17,18] Proteins within ESCRT-III were shown to polymerise into filaments on cell membranes in a variety of shapes including rings, helices, flat and conical spirals.[19]

2.1.3 Cell Division in *S. acidocaldarius*

S. acidocaldarius is considered to be the "model" crenarchaeon as it is aerobic and its genes are well-characterised.[20] It is also a thermoacidophile that grows optimally between 75 and 80 °C and between 2 and 3 pH. *S. acidocaldarius* cells are typically spherical with a diameter between 0.8-1 μms.

In 2008, Lindås et al.[1] identified three protein complexes CdvA, - B, -C in *S. acidocaldarius*, responsible for cell division. They further showed that this division machinery is found in all crenarchaeons.

Moreover, they discovered that CdvB proteins, including CdvB1/2, are genetic homologues of the Eukaryotic protein machineries, ESCRT-III. Therefore, it is likely that ESCRT-III proteins are highly evolutionarily conserved and that Eukaryota and Crenarchaeota share a common ancestor.[3] This evidence also gives credence to the Eocyte hypothesis which suggests that crenarchaeons are related to the origin of eukaryotes.[21]

CdvB proteins are sometimes referred to as "archaeal ESCRT-III" as they cut membranes in the same way.[22] Unlike in Eukaryotic cells, ESCRT-III drives the whole cell division process in crenarchaeons as opposed to only the final stage. Understanding how archaeal ESCRT-III drives cell division in simple crenarchaeons, can elucidate the various, vital roles ESCRT-III perform in Eukaryotic cells.

2.1.4 Live-Cell Imaging of *S. acidocaldarius*

Live-imaging of living cells using time-lapse microscopy is a technique used by cell biologists to understand the dynamic processes within cells.[23] Live-imaging archaeons presents more challenges than live-imaging of either bacteria or eukaryotes due to their small size and extreme living conditions. As such, *S. acidocaldarius* has not been live-imaged until recently. Pulschen et al.[2] have created the novel "Sulfoscope", apparatus that can live-cell image *S. acidocaldarius*. The "Sulfoscope" consists of a heated chamber and an inverted fluorescent microscope.

Time-lapse movies of live *S. acidocaldarius* cells imaged using the "Sulfoscope" were of an area $\sim 150\mu m \times 150\mu m$. These "movies" contained $\sim 10^2$ *S. acidocaldarius* cells. Cells taken from eight movies recorded during August 2019 and December 2019 were measured. The cell movies were realised as a stack of sequential images which can be view using ImageJ, an image processing software. The cells within the movies had different characteristics, including:

- Phenotype of *S. acidocaldarius* - Two genetic variations of *S. acidocaldarius* were imaged, wildtype (WT) and a mutant (MW001).

- Fluorescently tagged cell components - In order for the cells to appear in the images, they must be fluorescently tagged. This is where a particular component of a cell is attached with a fluorescent substance that can be seen by the microscope. To see how the shape of the cell changes, the membrane of the cell must be fluorescently tagged. In the WT cells, both the membrane and the DNA of the cells were tagged. It was found that the toxicity of the DNA dye reduced the rate of cell division. So when images of the MW001 cells were taken, only the membrane was tagged. The cells containing DNA dye, therefore, may not be representative of cells in nature.

- *Time between image frames* - the time interval between image frames varies depending on the movie, either 1 or 2 minutes. Cell division has been recorded to take between 4-12 minutes.[2] As such, a time interval of 2 minutes between each image results in only a few data points from

which to fit the shape of the midcell diameter against time graph. For more reliable results, the time interval should be as small as possible.

- Immobilisation of the cells - The immobilisation of the cells made cells easier to measure as they would remain in approximately the same position throughout the movies. However, the drawback of immobilisation is that it slows down the rate of cell division such that division immobilised cells may not reflect the division of cells found in nature.

- Movie duration - The duration of the movies only affected the number of cells that undergo cell division because the cells divide at random times.

- Doubling time - doubling time is the time it takes for the sample to double in quantity, equivalent to the cell cycle of the cell. Immobilisation and DNA dye increase this time.

2.2 Methodology

2.2.1 Identifying Dividing Cells

Not all of the cells in a sample of imaged cells divide during the imaging duration. The dividing cells are cropped out using ImageJ, forming smaller individual image stacks. Cropping the cells from the master image stack serves to keep track of which images have already been measured. It also serves to measure the cells on a more appropriate scale, i.e. $\sim 1\mu m$ rather than $\sim 100\mu m$ thereby improving the accuracy of the measurements, as shown in Fig. 3.
Initially, the dividing cells were identified and cropped by eye by merely moving through the image stacks and looking at which cells divide. During the project, a machine learning algorithm was developed by collaborators of the Baum group that can automate the process of finding the dividing cells and crop them out individually. The use of this algorithm drastically reduced the time taken to find the dividing cell. Moreover, the algorithm was able to detect more dividing cells that were missed by human error.

2.2.2 Taking Measurements using ImageJ

The movies of *S. acidocaldarius* are low-resolution due to the small size of the cells and the limited resolution of the microscope. It is therefore difficult to determine the "edges" of the cell. The Full-Width Half-Maximum (FWHM) of an intensity profile can be used to describe the measurement of the width of an imaged object when the edges of the image are not sharp.[24] The intensity profile of the midcell diameter can be plotted calculated using ImageJ's "line tool" as shown in Fig. 4
The x (distance) and y (gray value/intensity) data can be exported to a CSV file. A Python code was written that took this CSV file as an input and fits a Gaussian function according to Eq. 1 using the *curve_fit* function in Python's *scipy* library.
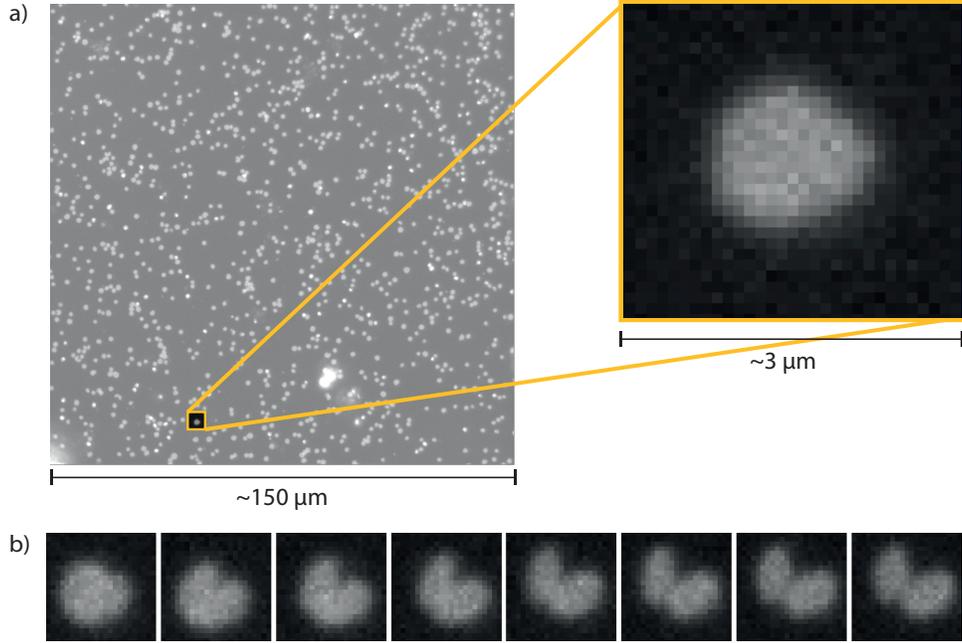
Figure 3: a) A dividing cell cropped out of a time-lapse movie of MW001 *S. acidocaldarius* cells using ImageJ software. b) A montage of images showing how the cropped cell divides over time, with one image taken every minute.
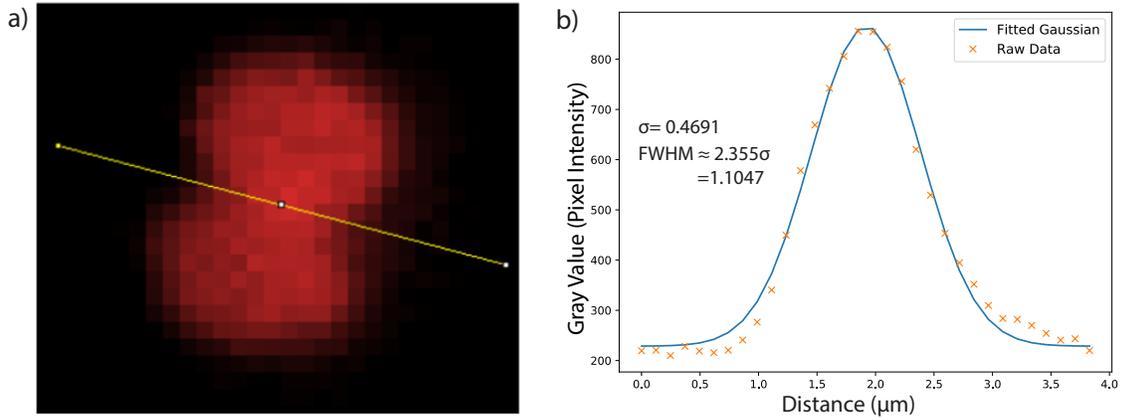
Figure 4: a) The intensity over the midcell axis of a cell is measured using the line tool in ImageJ. b) The intensity is plotted against distance and fitted with a Gaussian curve. The FWHM of the Gaussian is calculated by multiplying the standard deviation of the curve by $2\sqrt{2ln2}$.

$$Gauss(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1}$$

Where μ is the mean of the Gaussian and σ is the standard deviation. The Python code then calculates the FWHM according to by Eq. 2.

$$FWHM = 2\sqrt{2ln2}\sigma \approx 2.355\sigma \tag{2}$$

Some images of the cells have two intensity peaks corresponding to two Gaussians. The two Gaussians are present in the early stages of cell division, as shown in Fig. 5 a). This is due to the overlapping membranes of the daughter cells that create two localised areas of increased intensity. As the cell divides, the two areas move closer together until the two intensity peaks form one. Hence, initial intensity measurements are plotted to fit two Gaussian curves, and the latter measurements are plotted to fit one Gaussian.
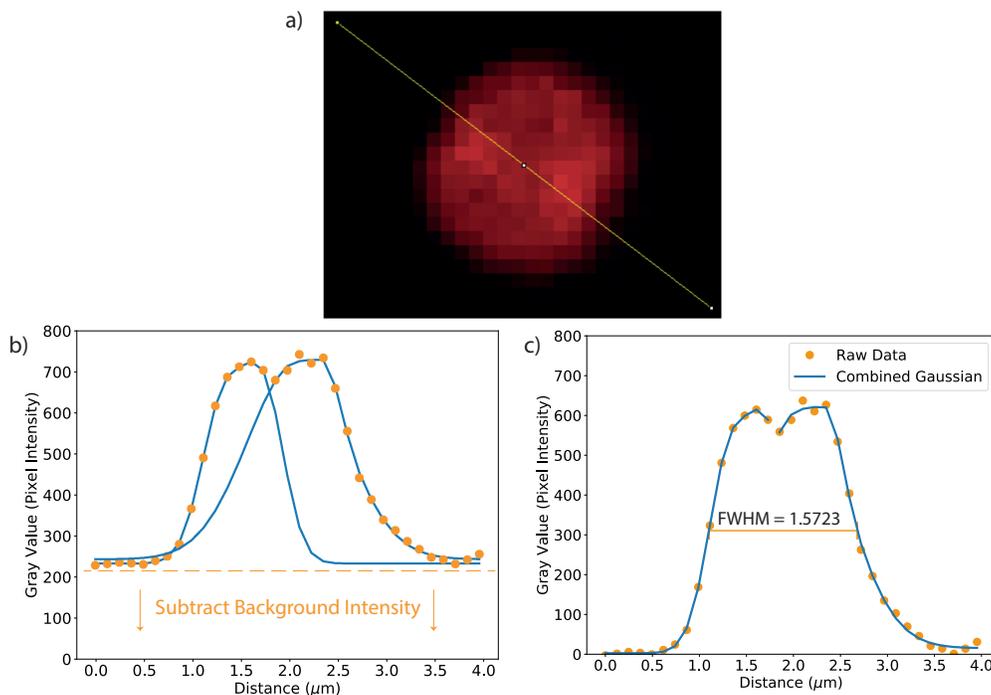
Figure 5: a) In the early stages of division, cells appear to have two intensity peaks on the midaxis of the cell due to the overlapping of daughter cells. The intensity over the midcell axis of a cell is measured using the line tool in ImageJ. b) The intensity is plotted against distance and fitted with two Gaussian curves, then zeroed by subtracting the background intensity. c) The two Gaussians are then joined, and the FWHM is calculated from the combined Gaussian.

Additional Python code was written to fit two Gaussian curves to the intensity data. The code splits the data into two halves and fits Gaussians to each half. The Gaussians are then added together to create a unified curve shown in Fig. 5 c).

Calculations of FWHM using Eq. 2 is the most accurate method as it does not depend on any changes in background intensity. However, σ cannot be calculated for two joined Gaussians. Instead, the FWHM is calculated by first "zeroing" the joined Gaussian, then taking the difference between the x values at half maximum. To "zero" the Gaussian, the background intensity needs to be subtracted

from every value such that the "tails" of the Gaussian lies on zero. Due to the small length scale of the measurements, the subtracted background intensity has a significant effect on the accuracy of FWHM, so the background intensity must be calculated accurately. The background intensity is calculated by taking the average of the "tails" of the Gaussian curves, shown in Fig. 5 b-c). Background intensity can vary from image frame to frame so it must be recalculated for every measurement.

The code for calculating the FWHM for one or two peaks can be found here:

https://github.com/TinaLanYao/Gaussian-Fitting.

2.2.3 Determining which Frames to Measure

This project is concerned with measuring the midcell diameter of cells over the time in which they divide. At present, it can not be determined when the cells start or stop dividing without observing the internal dynamics of the cells. *S. acidocaldarius* cells tend to stick together even after cell division. Measurements were therefore taken as the cell begins to "pinch", until the diameter stops decreasing. For every cell, the midcell diameter is measured over the "cell division" frames. The diameter is measured three times, and an average is calculated. A standard deviation of the three measurements provides a measure of uncertainty of the measurements.

2.2.4 Fitting Power Curves to Data

The purpose of measuring a large sample of cells is to take an average of the measurements for more accurate results. Initially, the average of the absolute measurements were going to be taken of every cell. However, cells have varying division times, and as discussed previously, there is uncertainty regarding when division truly starts and finishes. As such, there is no way to "line up" the measurements according to time for averaging. One solution was to scale, the axis to percentages rather than use absolute values, such that all the data points start at 0% and end at 100%. However, this significantly skewed the measurements, and the results were deemed invalid.

Instead, midcell diameter vs time measurements for each cell are fitted with a power curve of form Eq. 3 where the power coefficients quantify the shape and rate of division. Advantageously, averages can still be taken over all the cells measured without distorting the results.

$$d(t) = c \times t^{\alpha} + b \tag{3}$$

In Eq. 3, α quantifies the "shape" of the graph, meaning how the rate of division changes over time. $\alpha = 1$ represents a linear graph meaning that the cell divides at a constant rate, which is what is expected from previous measurements. $0 < \alpha < 1$ means that the diameter decreases faster initially and slower at later stages and $\alpha > 1$ has the opposite relation. b is the intercept of the graph, which is approximately the diameter of the undivided cell measured in μm. c is the average decrease in

midcell diameter per minute measured in $\mu\,min^{-1}$. c is, therefore, the average division rate of the cell.

Fig. 6 shows how the shape of the graph depends on α.
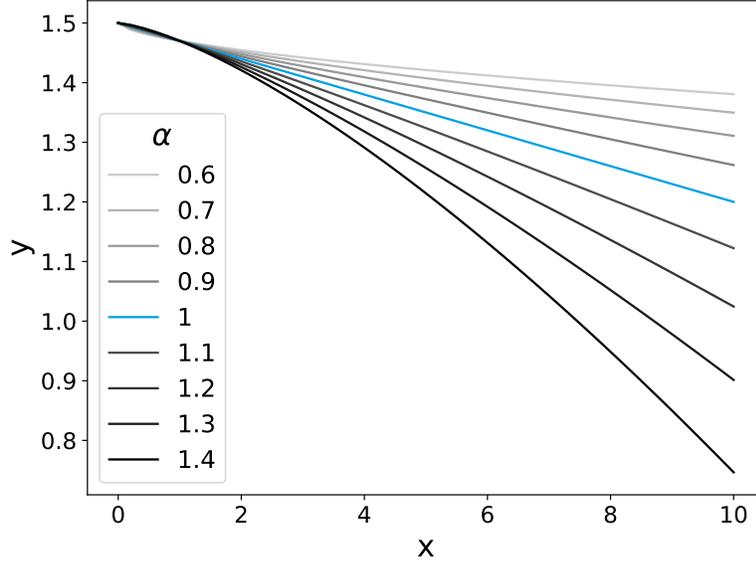
Figure 6: Power curve plots of $y = -0.03x^{\alpha} + 1.5$ for different values of α, indicating how different α values affect the shape of the curve.

2.3   Results and Analysis

54 *S. acidocaldarius* cells were measured in total. The cell measurements were divided into two groups according to their phenotypes, 32 WT and 22 MW001 cells. The details of the time-lapse movies from which these cells are taken are summarised in Table 1.

Table 1: The properties of the time-lapse movies from which the measured *S.acidocaldarius* cells were taken.

Phenotype	No. Measured	Immobilised?	Membrane Dye?	DNA dye?	Time Interval Between Image Frames /min	Duration /min	Doubling Time /h
WT	32	Yes	Yes	Yes	2	128	3
MW001	22	Yes	Yes	No	1	106	5

Power curves were plotted for the midcell diameter measurements for every cell in Fig. 7. This project is only interested in the shape and rate of division, α and c. Fig. 7 a-b) shows that the average shape of the midcell diameter against time graphs are quantified by $\alpha = 0.91 \pm 0.25$ and $\alpha = 0.90 \pm 0.23$ for WT and MW001 cells respectively. These values suggest that the rate of decrease
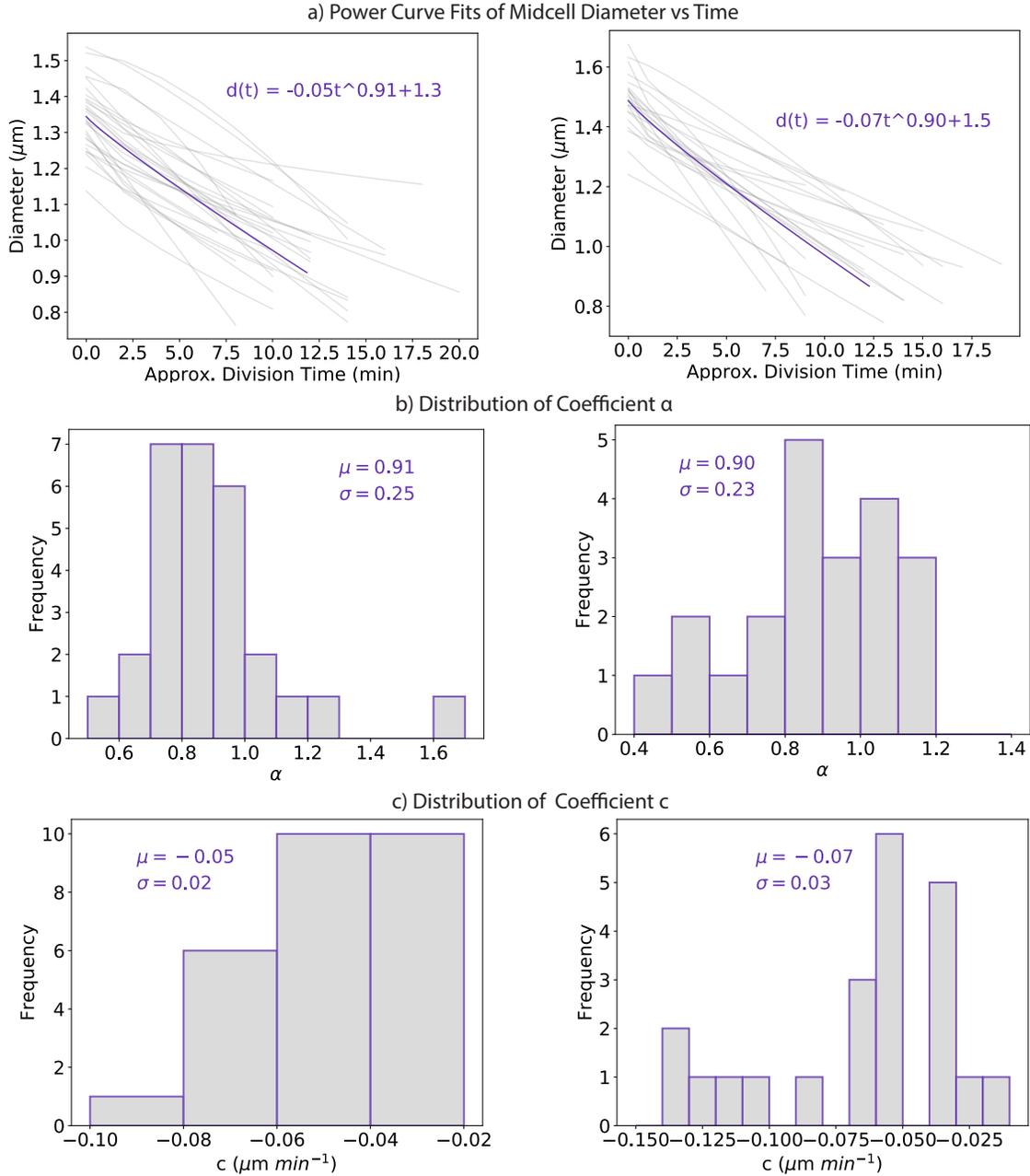
Figure 7: (Left) 32 WT cells. (Right) 22 MW001 cells. a) Purple line indicates the power curve plotted from averaged power coefficients of all the measured cells. b-c) The mean and standard deviation of fitted power coefficients are displayed alongside their distributions.

is close to linear ($\alpha = 1$) but generally the cells divide faster to begin before slowing down near the end of the division ($\alpha < 1$).

Alternatively, this $\alpha < 1$ relation could be attributed to the fact that it is not possible to know, at the time of writing, at what point the cells stop dividing. As the cells stick together even after cell

division, the midcell diameter plateaus instead of reaching zero. Human judgement is required to determine when the cell appears to have reached a plateau, and measurements are likely taken even after division stops. As a result, the rate of division would appear to slow down at the end of the measurements resulting in $\alpha < 1$.

Fig. 7 b-c) shows the distribution of the fitted coefficients α and c of all 54 cells, where the left column shows 32 WT cells, and the right column shows the 22 MW001.

The standard deviation of α values in Fig. 7b) is quite large in comparison to the mean meaning that there is a large variation between the measurements. The histograms in Fig. 7 b) show approximate bell curves centred on the mean values of α, which further validates the mean values. Similarly, Fig. 7 c) show the c values are distributed like a bell curve over the mean, $c = -0.05 \pm 0.02 \, \mu m \, min^{-1}$ and $-0.07 \pm 0.03 \, \mu m \, min^{-1}$ for WT and MW001 respectively. The presence of DNA dye and the immobilisation of the cells slow the rate of division. The WT cells measured have both the DNA dye and immobilisation. In contrast, the MW001 cells are only immobilised, which may explain why, on average, the MW001 cells have a faster rate of division. The power coefficient c, has been used to quantify the rate of division; however, it is only the average rate. If the rate during division, c would not be a good measure. c is a measure of division rate most appropriately when division is linear.

These results have multiple areas of uncertainty, including uncertainty in when the cells stop dividing, measurement error and the goodness-of-fit of the power curves. These areas of uncertainty compound when taking the final average of the power curves; therefore, it is important to try to minimise the error in each step. As yet, there is no method in determining when cells start and stop dividing.

For the FWHM method of measuring the midcell diameter of the cells, the intensity profile of the line of measurement must be a Gaussian (or two Gaussians). For accurate measurements, fitted Gaussians needs to be as well defined as possible; however, due to the tendency of $S. \ acidocaldarius$ cells to stick together, the intensity associated with other cells can interfere with the clarity of the Gaussian, as shown in Fig. 8. Some cells included in the average had a large measurement error compared to others. Although having fewer measurements in an average is generally less accurate, the measurements with large associated uncertainties decrease the overall validity of the average measurement. As such, the cells with the lowest measurement error can be averaged again for more valid results.

Similarly, it is important that the power curves represent the raw data as accurately as possible. In the 54 cells measured, there is variation in how well the power curves fit the data. For more valid results, those ill-fitting curves should also be removed. Goodness-of-fit tests are statistical measures of how well-fitted curves fit raw data. R^2 is a simple goodness-of-fit test and is calculated using Eq. 4. R^2 varies from 0 to 1, where 1 indicates a perfect fit. In this project, power curves with R-squared values less than 0.9 were rejected. 0.9 is the R^2 value at which the standard deviation of the fitted curve's errors is approximately $1/3$ the size of the standard deviation of the dependent variable.

14
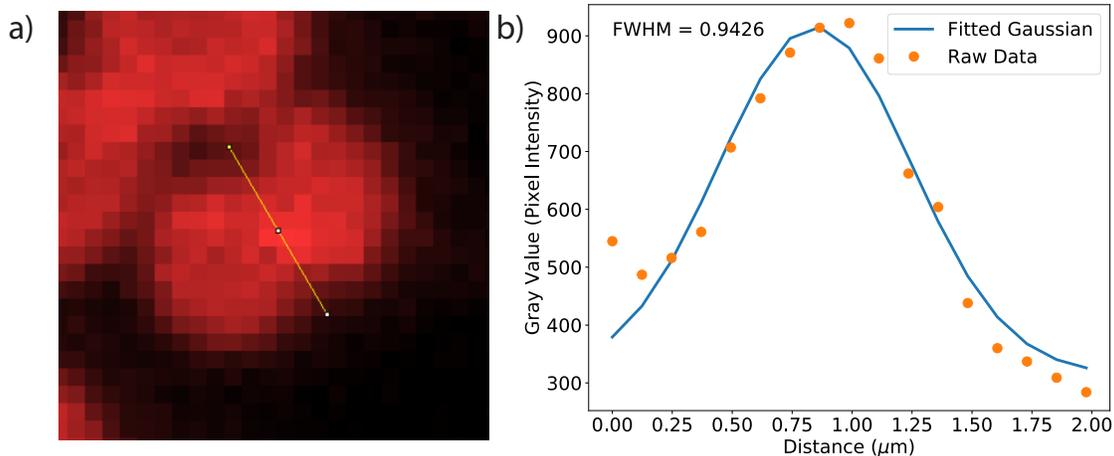
Figure 8: a) Cells that are unsuitable for measurement, using the method outlined in this project, are those in close proximity to neighbouring cells. b) Measuring the midcell diameter requires a plot of the intensity of the midcell axis of the cell and fitting a Gaussian curve. If there are other cells nearby, it affects the shape of the Gaussian and the FWHM calculated is inaccurate.

$$R^2 = \frac{\sum_i (y_i - \bar{y})^2}{\sum_i (f_i - \bar{y})^2} \tag{4}$$

Where y_i are the raw data points, $\bar{y}$ is the mean raw value, and f_i are the fitted values.

11 WT and 10 MW001 cells were left after eliminating those cells with significant measurement uncertainty and R-squared values less than 0.9. Once again, the midcell diameter of these cells was plotted over division time, fitted with power curves, and the average was plotted, the results are shown in Fig. 9.

The average coefficients plotted in Fig. 9 were calculated to be $\alpha = 0.96 \pm 0.13$ and $c = -0.04 \pm 0.02 \, \mu m \, min^{-1}$ for WT cells and $\alpha = 1.01 \pm 0.16$ and $c = -0.05 \pm 0.02 \, \mu m \, min^{-1}$.

The α values are close to 1, suggesting that the cells divide linearly as expected from previous work. The standard deviation of the fitted coefficients are smaller in these new plots suggesting that these averages are more reliable than the previous values. Although these results appear to validate previous observations, the number of cells measured is relatively small. More cells need to be measured to make a valid conclusion.

The main area of uncertainty in these measurements stems from not knowing when cell division starts and stops. Furthermore, the reduced rate of cell division due to cell immobilisation and DNA dye means that these results may not reflect *S. acidocaldarius* cells in nature.

The significance of $\alpha \sim 1$, is that ESCRT-III and CdvB are likely to exert a constant force when cutting cell membranes. Researchers can, therefore focus on physical mechanisms that exert constant force to explain the unknown mechanics behind protein filament mediated membrane cutting.

Additionally, images show that the cells do not always divide symmetrically, as shown in Fig. 10.

15
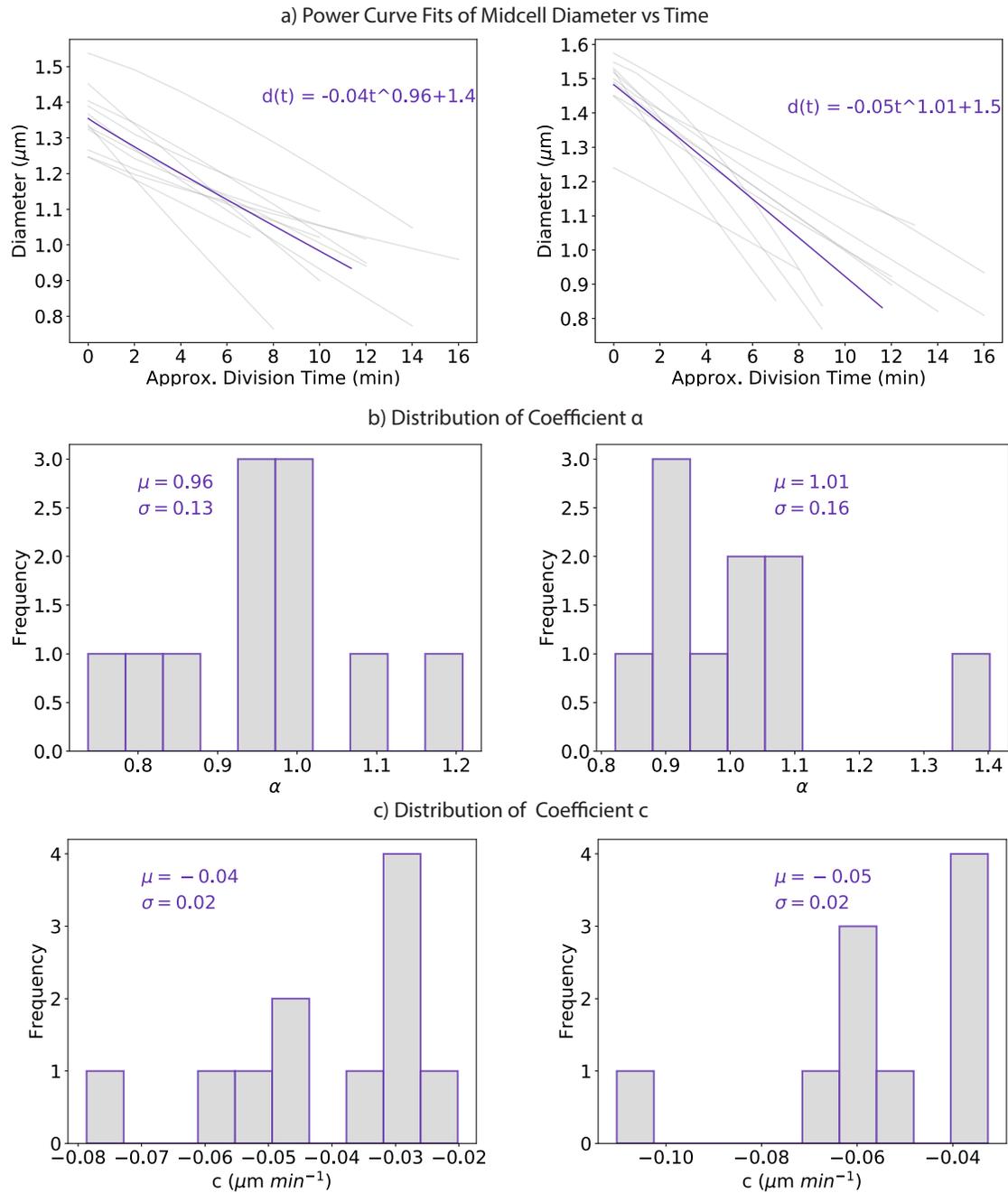
Figure 9: Selected midcell diameter measurements from Fig. 7 that have the lowest associated uncertainty. (Left) 11 WT. (Right) 10 MW001 cells

This suggests that while division protein filaments exert a constant force, it may not be constant in all orientations.
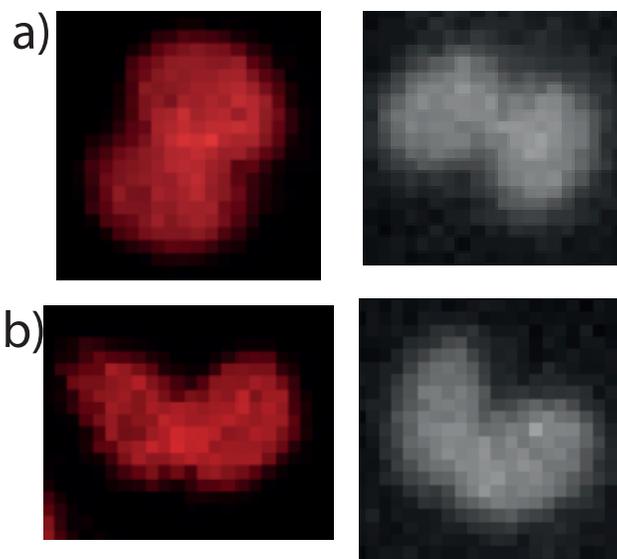
Figure 10: Selection of cells displaying a) symmetric and b) asymmetric division. Red are WT and grey are MW001 cells.

2.4   Further Work

The validity of the coefficient values α and c calculated from these cell measurements are dependent on the certainty of when cells start and stop dividing. One way to reduce this uncertainty is to measure the cell across all image frames such as in Fig. 11.

Fig. 11 shows a plateau at the beginning, because the diameter of the cell stays roughly spherical when it is not dividing. It also shows a plateau at the end as a result of cells sticking together after division. The division frames are, therefore, likely to be the highlighted area between the two plateaus. In this method, the need for human judgement is reduced.

In this project, machine learning algorithms were used to detect and crop out cells for measurement automatically. The capabilities of these algorithms can always be expanded. Currently, algorithms can be used to find the start and end of division by tracking the topology of the cell as it splits into two.

This project's main time-consuming aspects involve: hand-drawing lines in ImageJ across the midcell of the dividing cell, exporting the intensity data then fitting one or two Gaussian curves which requires minor adjustments to Python code. Each of these steps can be automated, and work has already been started to develop the algorithms to carry out these measurements.

More time-lapse movies are continually being created using the "Sulfoscope". 17 WT untampered cells, i.e. without immobilisation or DNA dye, were measured in the late stages of this project, and as such, the results are not finalised. The preliminary midcell diameter vs time results of these cells are displayed in Appendix A. These cells are more akin to those found in nature than the cells measured in this project.
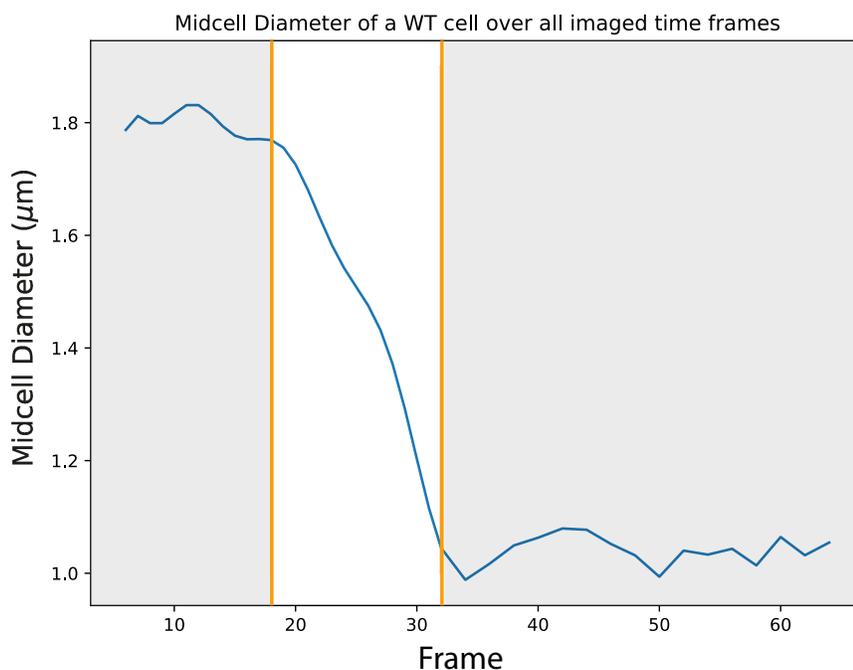
Figure 11: The midcell diameter of a WT cell plotted every two minutes for 64 image frames. It is estimated that cell division occurs in the highlighted region.

3 Development of Coarse-Grained Cell Division Model

3.1 Background

3.1.1 Computational Modelling in Biology

Computational modelling can be used to investigate the underlying dynamics of biological systems at a molecular level, including the processes within a cell. The dynamics of many-body systems, even with the applications of simple Newtonian dynamics, are unsolvable. Before computer simulations, theories were used to predict the properties of the molecules. Computer modelling experiments have the advantage over traditional experiments by providing exact numerical solutions.[25] Another distinct advantage of computational experimentation is that the parameters of an experiment, such as temperature or pressure, are unlimited by cost or practicality.[26] The accuracy and validity of the results, therefore, only depend on how a simulation is programmed. Despite the success of computational experiments, traditional experiments still play a vital role in validating results from the simulations and, by extension, improving the simulations themselves.

There are two main types of simulations used in biological systems: Monte Carlo (MC) and molecular dynamics (MD).[28]

MC methods are computational algorithms that use random sampling to obtain numerical results

and has applications ranging from computational biology to finance. MC methods rely on the law of large numbers; by randomly sampling multiple times. Then, the expected value of a measured variable can be approximated by taking the mean of a variable from the results of multiple random samples. When the probability distribution of a variable is known, Markov chain Monte Carlo (MCMC) methods can be used. MCMC is used in computational biology modelling, where the probability that the system will transition into each successive random state is calculated dependent on the previous event.[27]

Molecular dynamics (MD) has the advantage of providing a time-evolved model of the system, which is more suitable for cell division.[28]

3.1.2 Molecular Dynamics

Molecular dynamics (MD) simulations have been used since the 1970s to model the behaviour of molecules in different chemical, biological or physical systems. MD studies the behaviour of a many-particle system by computing its time evolution numerically and averaging its observable quantities over a long period.[28] This process is outlined by the flowchart in Fig. 12. Advances in computing have allowed simulation to increase in scale from a few hundred molecules to potentially millions.
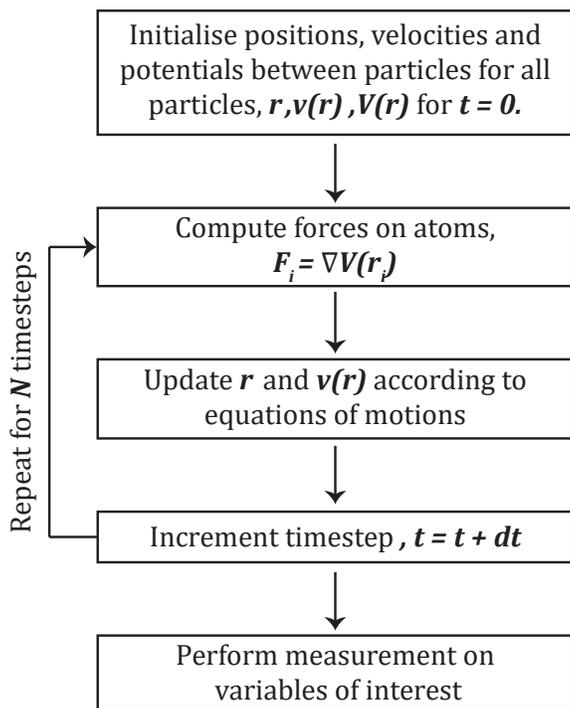
Figure 12: Molecular dynamics simulations as a simplified schematic.

MD simulations in this project are based on Langevin dynamics. Langevin dynamics recreates random particle movements that are representative of biological systems without the need to explicitly model a solvent, i.e. individual water molecules.[29] In Langevin dynamics simulations, the Langevin

equation is solved for each particle at each timestep of a given simulation, given by Eq. 5.

$$\gamma \frac{dx}{dt} = -\frac{dU}{dx} + \xi(t) \tag{5}$$

Where x represents the dynamical variables of the system, γ is the friction constant, U is the potential energy between particles and $\xi(t)$ represents the thermal noise from the environment that recreates the effect of molecules in a solvent. The thermal noise has an average force of zero so that it does not affect the system with bias in a particular direction.

3.1.3 Volume Exclusion

Eq. 6 gives the classic 12-6 LJ potential and is represented by the graph Fig. 13. It is a simple model that describes the interaction between two uncharged atoms as a function of distance between the two. The LJ potential is attractive at most distances but repulsive at small distances because the Pauli exclusion principle which prevents matter from overlapping.

$$V_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] = \epsilon\left[\left(\frac{r_{min}}{r}\right)^{12} - 2\left(\frac{r_{min}}{r}\right)^{6}\right] \tag{6}$$

Where ϵ is the measure of how strongly two particles interact occurring at $r = r_{min}$ and σ is the distance at which the potential between two particles is zero. The distances are related by $r_{min} = 2^{\frac{1}{6}}\sigma \approx 1.122$. The tail of the LJ potential tends to zero such that the potential between two particles at infinity is equal to zero, as shown in Fig. 13a). It is computationally inefficient to calculate potentials at large distances as this would require calculating the potentials between every particle in the system (within the threshold of infinity) which would scale to 2^N. Therefore, the LJ potential can be truncated at a distance r_{cut}, which is the distance at which the potential is $\frac{1}{60}$ of its minimum value. After truncation, the LJ potential needs to be shifted upwards such that the truncated potential equals 0 at r_{cut} to avoid discontinuity in the potential as shown in Fig. 13b). The truncated LJ potential is given by Eq. 7.

$$V_{LJ_{trunc}}(r) = \begin{cases} V_{LJ}(r) - V_{LJ}(r_{cut}) & r \leqslant r_{cut} \\ 0 & r > r_{cut} \end{cases} \tag{7}$$

A common interaction used in MD simulations is a special case of the truncated LJ potential where $r_{cut} = r_{min}$ called volume exclusion.[30] In MD, particles are defined by their positions and interactions without explicit radii. Volume exclusion provides a way of defining a radius for the particles. Volume exclusion does not model any attraction between particles, only repulsion if the particles are less than r_{min}. As a result, particles behave like hard spheres with radii equal to $\frac{r_{min}}{2}$.[31]

The particles in the simulations are modelled as hard-spheres with radii, $r = \frac{r_{min}}{2} \approx \frac{1.122\sigma}{2}$ because
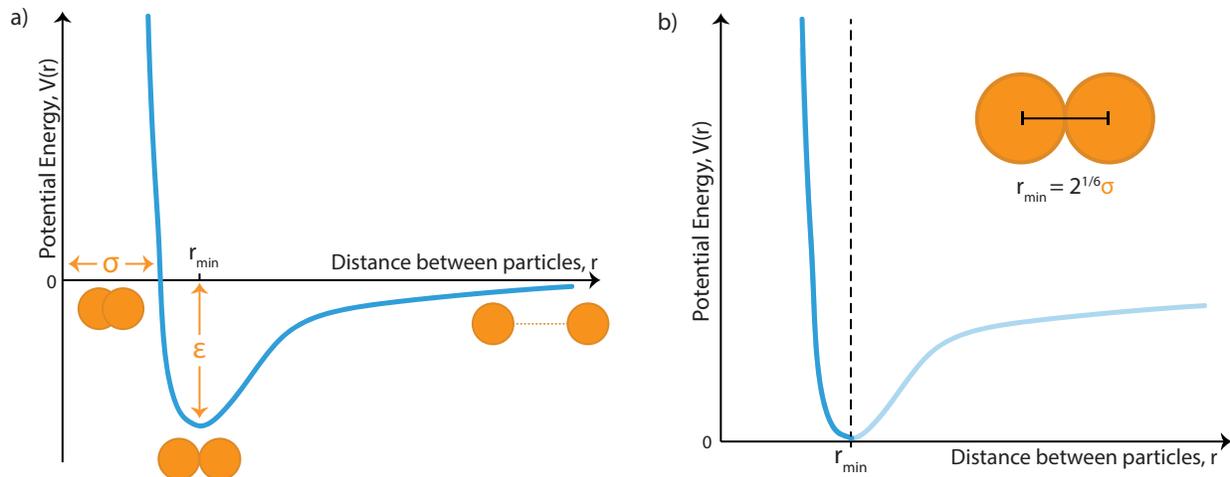
Figure 13: a) A graph of the Lennard-Jones potential that is defined by Eq. 6. b) A graph illustrating the volume exclusion potential, Eq. 8. Volume exlusions is the Lennard-Jones potential truncated at r_{min} and shifted upwards to prevent any discontinuity.

r_{min} is where interactions equal zero.

$$V_{LJ} = \begin{cases} 4\epsilon((\frac{\sigma}{r_{lj}})^{12} - (\frac{\sigma}{r_{lj}})^6) & , r_{ij} \leq r_{cut} = r_{min} \\ 0 & , r_{cut} = r_{min} > r_{ij} \end{cases} \tag{8}$$

3.1.4   Coarse-grained Modelling

Since the 1980s, simulations have been developed to study "all-atom" models for biological systems.[32] Modelling every particle is highly computationally demanding and places limitations on length and timescales. As a result, coarse-grained models were developed to reduce the number of particles required to recreate a complex system.[33] The simplicity of coarse-grained models increases efficiency at the expense of, often unnecessary, molecular detail.

3.1.5   Membrane Model

Biological membranes consist of lipid molecules that self-assemble into either a mono- or a bilayer which align according to hydrophobic interactions. Yuan et al.[34] developed a coarse-grained model of the fluid membrane that faithfully reproduces the action of fluid membranes. This model builds upon previous work by Drouffe et al.[35] by improving the fluidity of the model to meet the experimental range. Like the model by Drouffe et al., the particles in the model by Yuan et al. can self-assemble into a cell/vesicle as shown in Fig. 14. The model uses orientation-dependent pair potentials between particles to mimic the action of lipids in water. The pair potential was based on the Lennard-Jones (LJ) potential.[34]
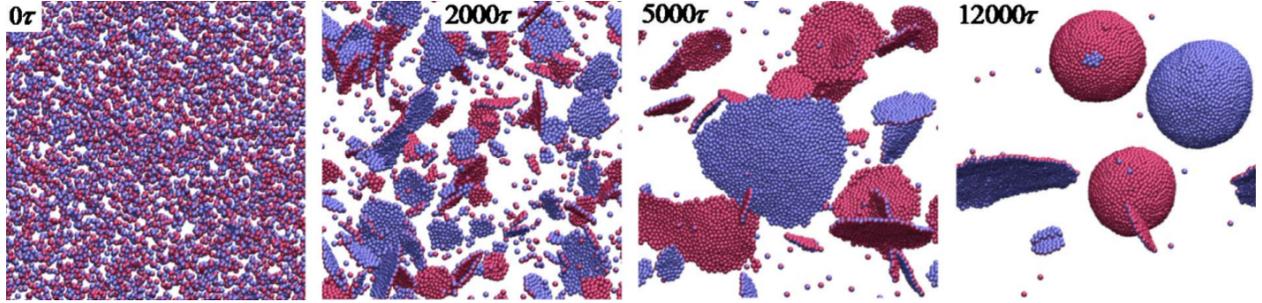
Figure 14: A diagram showing how coarse-grained membrane particles can self-assemble into cell membranes. (Figure by Yuan et al.[34])

3.1.6 ESCRT-III Filament Model

Harker-Kirschneck et al.[36] created a coarse-grained model of the ESCRT-III filament. They successfully showed how changes in its geometry can drive membrane abscission leading to cargo budding. The proteins of the constricting filament are coarse-grained in this model into units of three particles connected via nine harmonic bonds given by Eq. 9. The connected triplet units form a triple-stranded filament, as shown in Fig. 15 a).

$$E_{bond} = K_{bond} \cdot (r - r_0)^2 \tag{9}$$

Where K_{bond} is the "bond strength", or stiffness, constant, r is the distance between two subunits and r_0 is the resting bond distance. K_{bond} remained the same for each subunit.[37] These strong harmonic bonds allow for large amounts of tension on the filament without snapping. The rigidity of the filament and therefore, its geometry depends on the bond length between neighbouring triplet units which depends on bond strength, K_{bond}.

The top particles of each triplet unit do not attract or repel the membrane; instead, it interacts via volume-exclusion. The two bottom particles in each triplet subunit attracted to the membrane via the attractive region of the LJ potential Eq. 6.

When in a relaxed state, the filament forms a ring of radius R because the bonds between the subunits have the same rest length. Filaments longer than R are forced to become spirals because filaments are prevented from overlapping due to volume exclusion. Filament bonds must stretch to form spirals causing potential energy to store between the bonds. This stored energy is analogous to elastic potential energy stored in a spring shown by Eq. 9. The release of the stored energy can drive membrane deformation.

3.1.7 Archaeal Cell Division Model

The previously discussed membrane and ESCRT-III filament model were implemented together by Risa et al. to form the archaeal cell division model. The model was based on previous experimental observations of cell division in *S. acidocaldarius*.[3] The ESCRT-III filament model was used to model CdvB and CdvB1/B2, the protein complexes division proteins found in *S. acidocaldarius*. CdvB and CdvB1/B2 form the "archaeal ESCRT-III" so they behave in the same way. Images were taken of fluorescently labelled CdvB and CdvB1/B2 in *S. acidocaldarius* using SRRF super-resolution microscopy. It showed that cell division occurs in the following three-step process illustrated in Fig. 15b):
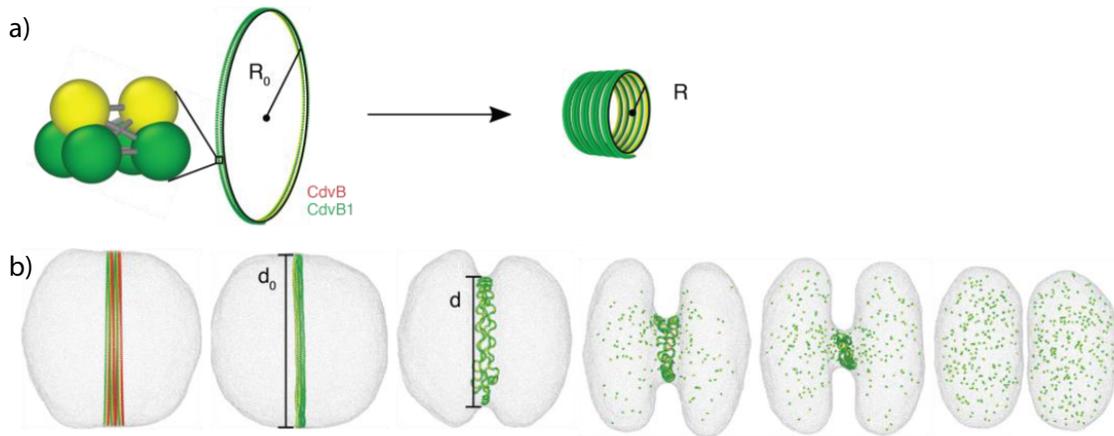
Figure 15: a) Computationally modelled archaeal cell division filament. The filament is coarse-grained and formed of triplet units joined by nine harmonic bonds. b) Initially, the CdvB1/2 filament (green) polymerises onto the CdvB scaffold (red) and is stretched to the radius of the cell R_0, when the filament relaxes it constricts to its preferred radius R. This constriction drives cell division in a simple cell. (Figure by Risa et al.)[3]

First, a CdvB ring of fixed diameter R_0 is formed in the middle of the dividing cells and acts as a scaffold. Second, CdvB1/B2 to assemble onto the CdvB. CdvB1/B2 has a smaller preferred curvature, R, so the bonds within must stretch to fit the radius of CdvB. Third, when the CdvB template is taken away, the potential energy stored in CdvB1/B2 polymer is released. This drives constriction of the ring towards their relaxed radius R, bringing along the membrane as it contracts. The paper makes the conclusion that CdvB inhibits cell division and that proteasome-mediated degradation of it triggers cell division in the first step.[3]

Only the action of CdvB1/B2 is simulated because CdvB is not present before cell division. Initially, the simulated CdvB1/B2 filaments are stretched into a helix around the circumference of the cell, R_0, imitating the real CdvB1/B2 attaching to the CdvB scaffold. Then, the constriction of the ring is modelled by shortening of the bonds between the triplet units to its preferred curvature of R, which

releases the stored harmonic energy from being stretched. The shortening of the bonds occurred according to one of two protocols: instantaneous or random. Instantaneous is when all the bonds shorten at once, and random is when random bonds shorten in successive timesteps. Experiments showed that the CdvB scaffold quickly disintegrates, which favours the instantaneous protocol. Simulations showed that the constriction of the helix alone could not cause membrane scission. After changing the model, the filaments were made to disassemble at a specific rate. Cell division successfully occurs in this model depends on the curvature of the constricting filament given by $\frac{R}{R_0}$ and the rate of disassembly of the filament. When $\frac{R}{R_0}$ is too small, the energy released from the constricting filament is not enough to drive cell division. Conversely, when $\frac{R}{R_0}$ is too large, the filament constriction energy overcomes the attraction to the membrane and detaches. If the rate of disassembly is too fast, then the filament does not have time to complete cell division. In contrast, if the rate of disassembly is too slow, then the filament forms a bottleneck which delays division. Experimental observations revealed that a diffuse CdvB1 signal was shown to accumulate as the division ring contracts and the local intensities of CdvB1/B2 increased overtime. Both observations agree with the model suggesting that CdvB1/B2 disassembles overtime. This result delineates how cross-referencing experimental evidence with simulation data can lead to new discoveries.[3]

3.2 Methodology

This project uses the cell division model by Risa et al. as a basis for further simulation tests. The original simulation consisted of only a membrane and a constricting filament that drives cell division. The model does not account for the macromolecules that constitute 20-30% of cell interiors, or cytoplasm.[38] The purpose of this project is to investigate the effect of "macromolecular overcrowding" on the dynamics of cell division, by adding cytoplasm-like particles to the centre of the simulated cell. Simulation input scripts were written using Python for LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) molecular dynamics software following documentation.[39] Simulations in this project use "LJ units" which are dimensionless units defined based on the LJ potential often used in MD for convenience, choosing length σ, mass m, time τ and energy ϵ.[40] Furthermore, Boltzmann constant, $k_B = 1$ and temperature is related to energy by $\epsilon = k_B T$.

For a simulation to run, input scripts must include the initial positions and velocities of each particle, the interactions between particles and the duration (number of timesteps) of each simulation stage. The size of each timestep needs to be large enough to reduce computation time but also not so large that the particles move too much between each timestep. Large timesteps can cause the particle positions to overlap and cause an error in the simulation. Each timestep is equal to $0.01 \ \tau$. The positions of every particle are only saved into the output file every $2 \cdot 10^4 \ \tau$ to reduce file size while retaining enough significant information. Therefore, output files contain only 92 timesteps rather than $1.84 \cdot 10^6 \ \tau$.

The cell division model consists of three components: membrane, constricting filament and cyto-

plasm.

Cell Membrane

This project uses the same simulated cell membrane used by Risa et al., who created the simulated cell following method by Yuan et al. .[34] They initialised 48,002 membrane particles that self-assembled into a roughly spherical cell of radius, $R_0 = 52.46$ σ. However, some membranes did not self-assemble and were left free-floating in the cell. Adding particles within the cell would overlap with existing particles resulting in simulation failure as repulsive energies would skyrocket. Consequently, the first step in initialising cytoplasm particles was to remove the stray membrane particles that did not form the cell. The magnitude was calculated for every membrane particle's coordinates—those less than or greater than the radius of the cell were removed, shown in Fig. 16a). The simulated cell is not a perfect sphere and not exactly centred on the origin of the simulation box. Therefore, the membranes particles that were kept were within a threshold of the radius of the cell between 51 - 62 σ. As a result, 48002 membrane particles were reduced to 47920.

Filament Protein

The filament protein was initialised following method by Risa et al. The constricting force generated by the filament is dependent on the harmonic bond between the protein particles and the curvature R/R_0 of the overall filament. In these simulations, $R/R_0 = 0.275$.

Another factor that affects the success of division is the disassembly timestep of the protein filament. If the filament disassembles too quickly, then the filament will not have enough time to drive cell division. If the disassembly is too slow, the filament will delay the time taken for cell division. In these simulations, the disassembly timestep is set equal to 4000.

Cytoplasm Particles

Cells are estimated to be filled with 20-30% biological molecules. However, it was desired that the packing of cytoplasm particles in the simulated cell could be varied to investigate its effect at different packing fractions. For maximum flexibility of the simulation, the cell must be packed maximally such that the packing fraction can be varied by reducing from the maximum. Face-centred cubic and hexagonal close-packed represent the densest possible packing of spheres. For this model, the hexagonal close packing arrangement was used where the x-, y-, z- coordinates of the particle centres are given by Eq. 10.

$$\begin{bmatrix} 2i + [(j+k)\%2)] \\ \sqrt{3}[j + \frac{1}{3}(k\%2)] \\ \frac{2\sqrt{6}}{3}k \end{bmatrix} r \tag{10}$$

Where i, j, k are indices starting from 0.

Spheres can be compactly packed into lattices up to 71%, however it is difficult to pack spheres into a larger sphere. The following method was used to try to pack a sphere with smaller spheres maximally. First, a large lattice of particles positions were calculated using hexagonal packing, Fig. 16 b). Then, those with position magnitudes greater than the radius of the cell are deleted, Fig. 16 c-d). When a desired packing fraction is entered into the Python code, the number of cytoplasm
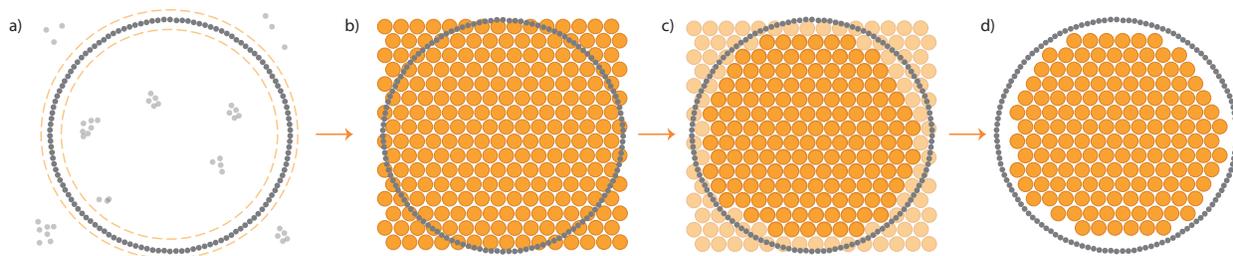
Figure 16: How cytoplasm particles were initialised in the cell model. a) Membrane particles self-assemble into a cell of radius, R_0. Some particles do not form part of the cell; these stray particles are removed to make room for the cytoplasm particles. b) Cytoplasm particles are initialised into a lattice larger R_0 in a hexagonal close packing arrangement. c) Cytoplasm particles that lie outside of the cell membrane are removed. d) The result is a maximally packed cell. The packing of the cell can be reduced by initialising fewer cytoplasm particles in the same arrangement.

particles required for the volume is calculated dependent on the radius of the cytoplasm particles. Then, only the required number of particles for the particular packing fraction are initialised. Particles in these simulations are point particles but are given an effective radius dependent on their interactions with other particles through volume exclusion. Volume exclusion potential is dependent on the strength (ϵ) and the distance at which the potential is equal to 0 ($r = r_{min} = 2^{\frac{1}{6}}\sigma$). Therefore, the effective radius of a particle is $r = \frac{r_{min}}{2}$. In these simulations, the radii of all the particles are kept the same apart from the cytoplasm. The cytoplasm radius can be parametrised as a function of σ_{cyto}.

Each simulation consists of five stages illustrated in Fig. 17:

1. Relax - In this stage, the harmonic bonds between each filament protein are relaxed while the membrane and cytoplasm particles equilibrate. In all simulations, this occurs for $1 \cdot 10^4\,\tau$.

2. Initial - The harmonic bonds between the protein particles in each triplet unit is initialised according to bond strength $K_{bond} = 600$. The filament also becomes a rigid body, meaning that the total force and torque on the body is the sum of forces and torques on the constituent particle. Therefore the filament moves as a single entity. The whole system then equilibrates for $2 \cdot 10^5\,\tau$.

3. Bonds transform - The bonds between the protein molecules are "turned off". As a result, the energy stored from the stretched bonds is released, causing a constricting force that drives cell division as the filament remains adsorbed to the membrane surface. The bonds can be released instantaneously or randomly. Instantaneous constriction is when all the bonds are released at once; this represents the fastest protocol. Random constriction is when the bonds that are released at a given time are randomly chosen. The bonds are released in one timestep for instantaneous transformation and, for random, the 18 bonds are released every one timestep for a total of 140 τ.

4. Bonds disassemble - the filament starts to disassemble. As discussed previously, cell division simulations can only reproduce cell division successfully when the filament disassembles because otherwise a bottleneck is formed. All the filament proteins (3×1440) are disassembled every 6000 τ lasting $1.44 \cdot 10^5$ τ. A bottleneck can be seen in Fig. 17 d) which diminishes over time due to disassembly.

5. Final - finally, regardless if a division is successful, the system is equilibrated for $2 \cdot 10^5$ τ.
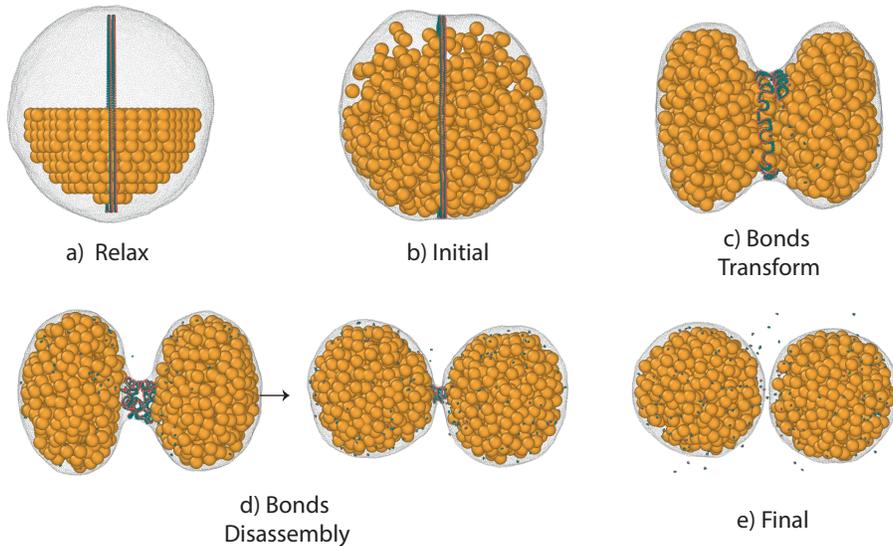
Figure 17: The five stages of the cell division simulations.

3.2.1 Simulation Procedure

An appropriate range of packing fractions and σ_{cyto} were chosen for simulation, $packing = 10, 15...40\%$ and $\sigma_{cyto} = 5, 6, 7, 8$. This range of packing fractions were chosen because it is estimated that cytoplasm is 20-30% full of proteins. At these packing fractions, σ_{cyto} values less than 5 generated too many particles (up to 10^5) which prevented any significant filament constriction. This vast number

of particles also increased the computation time required for a simulation. When the size of the cytoplasm particles are the same as the size of the membrane particles, the cell can even explode, as shown in Fig. 18 because the small cytoplasm particle size allows them to leak through the gaps in the membrane. Conversely, for σ_{cyto} values greater than 8, the number of cytoplasm particles generated were too small to affect the simulation significantly.
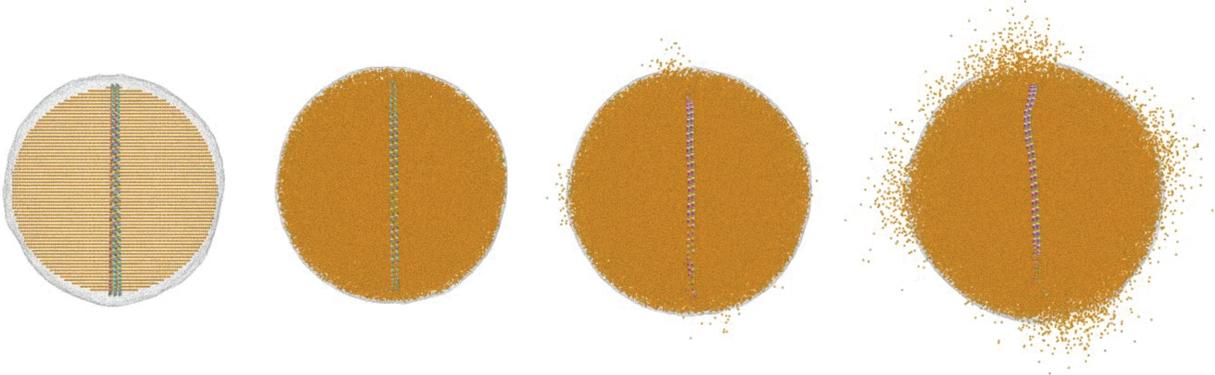
Figure 18: Simulation of a cell where $\sigma_{cyto} = 1$ and packing fraction $= 8.5\%$. The cytoplasm particles leak through the gaps in the membrane.

These simulations contain up to 10^4 particles, for more concise input scripts the particles can be separated into different groups. The interactions between particles can be described as the interactions between groups rather than individual particles, i.e. 47920 particles are in the membrane group. A summary of interactions between the particle groups are given by Table. 2.

Table 2: Interactions between different particle groups in the cell division simulations.

Particle Group 1	Particle Group 2	Interaction	Variable Values						
			σ	$\epsilon(k_BT)$	$r_{cut}(\sigma)$	$r_{min}(\sigma)$	$\theta(deg)$	μ	ζ
Membrane	Membrane	Orientation-dependent 4-2 LJ potential[34]	1	4.34	2.6	1.122	0	3	4
Membrane	Upper Protein	Volume Exclusion (Eq. 8)	1	2	r_{min}	1.122	-	-	-
Membrane	Lower 2 Proteins	Attractive LJ (Eq. 6)	1	4	r_{min}	1.122	-	-	-
Proteins	Proteins	Volume Exclusion (Eq.8)	1	2	r_{min}	1.122	-	-	-
Cytoplasm	Cytoplasm	Volume Exclusion (Eq.8)	5, 6, 7, 8	2	r_{min}	1.122	-	-	-
Cytoplasm	Membrane & Proteins	Volume Exclusion (Eq.8)	4	2	r_{min}	1.122	-	-	-

Parallel computers provides high-performance computing which is suitable for simulations of 10^6 molecules. LAMMPS simulations were run on UCL's Myriad research parallel computing cluster. The output files are XYZ files that contain the positions of every particle for every outputted time frame (every $2 \cdot 4\,\tau$. The output files can be viewed in OVITO visualisation software. The system in the simulation is subject to Langevin dynamics, as discussed previously, Eq. 5 and is represented as a microcanonical ensemble, a statistical ensemble that is used to represent the possible states of

a system that has a fixed number of particles, volume and total energy.[41]

Each simulation configuration was run 10 times and for both constriction protocols: instantaneous and random. The random seed state of the simulations were changed for each of the ten trials. Changing the seed, changes the pseudorandom processes in the simulation where cell division may be successful for some seed values but not others. Simulations were run without the presence of cytoplasm also for comparison.

3.2.2 Simulation Measurements

The midcell diameter of the simulated cell was calculated for every timestep using code provided by the Šarić group, which was adapted to accommodate the presence of cytoplasm particles. The code reads the positions of every particle at each timestep. Then, it takes a thin 3D rectangular cross-section of the centre of the simulation box represented by orange box in Fig. 19. The coordinates of the particles within the cross-section are fitted to a circle using the Taubin method for circle fitting.[42] The recorded diameter is the diameter of the fitted circle, Fig. 19 (top). When the simulated cell finishes division and there are two cells present with a space between them, the calculated diameter is not 0σ but rather $\sim 10\sigma$. The reason for this is because the gap between the two cells may not be exactly in the centre of the simulation box. Also, the 3D rectangular cross-section has a non-zero width so particles from the two daughter cells may still enter the cross-section, as shown in Fig. 19 (bottom).

A total of 500 simulations were run in this project. Code was written to automate the process of checking which simulation resulted in successful cell division. The code also plotted the average midcell diameter vs time over ten seed states for each parameter set. The measurements of midcell diameter never reaches zero but instead plateaus at a minimum value. As such, successful cell division was defined as when the midcell diameter reached below the threshold of 20 σ, where the initial diameter of the cell is always measured to be 106.52 σ.

3.3 Results and Analysis

Simulations were run for ten different seeds at packing fractions 10-40%, $\sigma_{cyto} = 5, 6, 7, 8$ and for the two different bond transform protocols. The number of successful cell division simulations out of ten simulations was plotted as a probability in the form of a phase diagram for each configuration in Fig. 20. The number of cytoplasm particles in each simulation is also included in the diagram. Fig. 20 shows that the greater the number of cytoplasm particle, the less likely that cell division is successful. This result is to be expected as the presence of particles within the simulated cell causes internal pressure that acts against the constricting force of the filament protein that causes cell division. Fig. 20 shows that there are more cytoplasm particles when packing fraction is higher or when σ_{cyto} is smaller. There is a greater number of cytoplasm particles for a given packing fraction
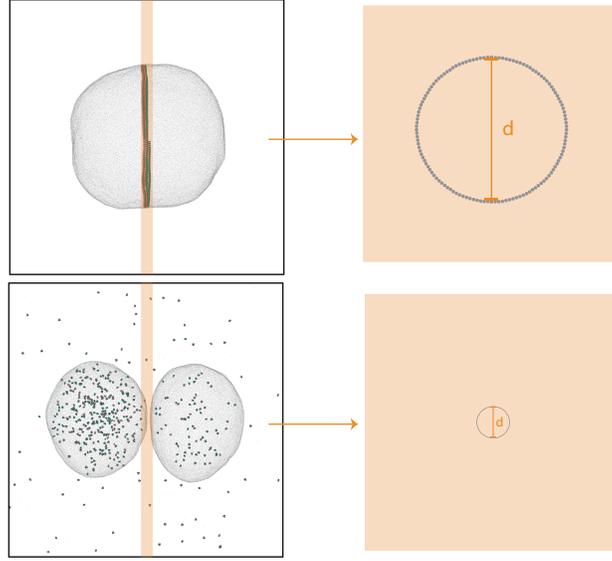
Figure 19: Diagram to show how the diameter is calculated for a simulated cell. The orange box takes a cross section of the membrane particles and then a circle is fitted to the membrane coordinates. (Bottom) When cell division is finalised, the diameter is not measured as zero because the membrane particles enter the orange box.
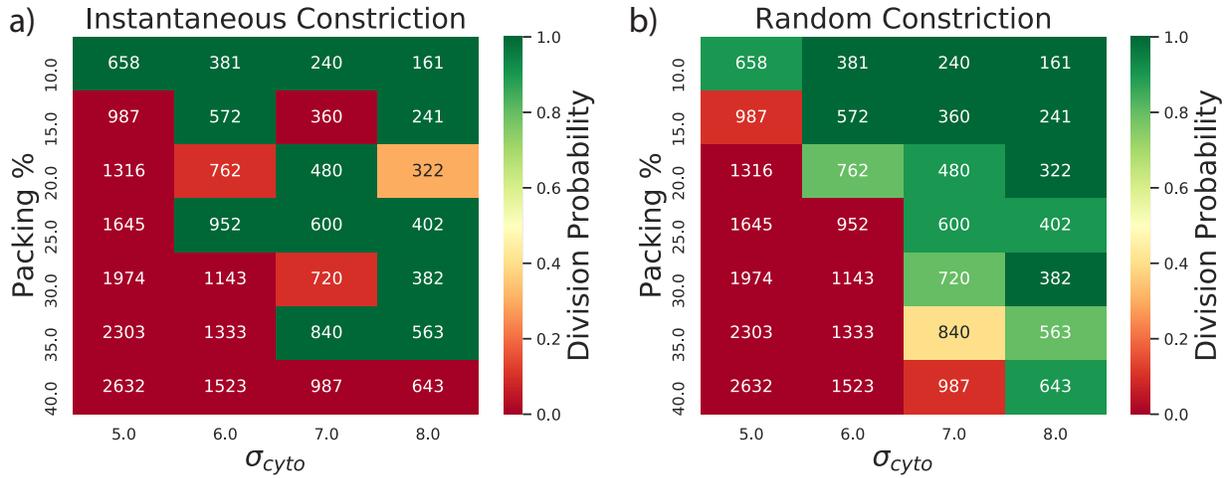
Figure 20: Phase diagrams showing the probability of successful cell division dependent on two parameters: the packing of the cell with cytoplasm particles and the radii of the cytoplasm particles (σ_{cyto}. The probability is given by the number of successful divisions out of ten trials. The phase diagrams are plotted for the two filament constriction protocols: instantaneous and random. The numbers in each box represent the number of cytoplasm particles in the simulated cell.

at smaller σ_{cyto} because more particles are required to fill the same amount of volume when the particles are smaller. The results are dependent on the number of cytoplasm particles in the system as opposed to the packing fraction or σ_{cyto} value explicitly. To further test this model, the number of cytoplasm particles could be used as a variable and choose another variable to investigate, such as the interaction strengths between different particle groups.

Unusually, the instantaneous diagram shows that division can happen at a higher packing fraction even when the division is unsuccessful at a lower packing fraction for a given σ_{cyto} value. For example, at $\sigma_{cyto} = 7$, cell division is successful at packing $= 20\%$ but not at 15% even though there is less cytoplasm particles to interfere with the division. Inspection of the individual simulations show that in these simulations, the filament detaches entirely from the membrane, rather than being "pushed out" by the presence of cytoplasm particles. This result is due to the filament having too much stored energy and overcoming the membrane attraction when that energy is released. Although it is unclear why some filaments detach and others do not given that they all initialised with the same parameters.

Fig. 21 delineates the two ways division can fail. Fig. 21 a) shows cell division failing due to the internal pressure of the cytoplasm particles pushing against the constricting filament; this usually takes a longer time. The greater the number of particles, the shorter amount of time it takes. Fig. 21 b) shows cell division failing due to the constricting filament detaching from the membrane. Detachment occurs as soon as the filament constricts.
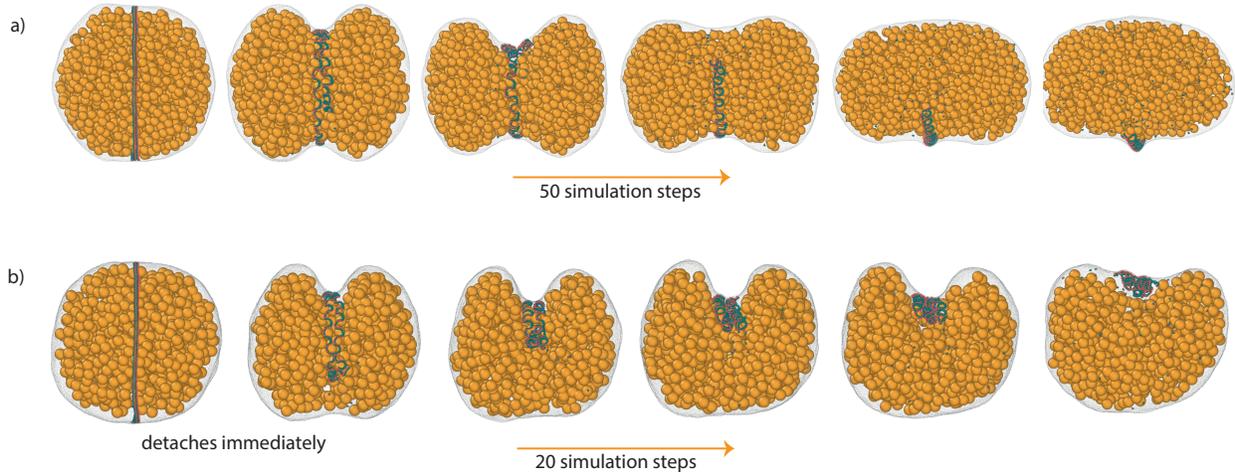
Figure 21: Two different types of cell division failure in the cell division simulations. a) Internal cytoplasm causes outward pressure against the filament and prevents division. b) Filament has too much energy and detaches from the membrane when constriction begins.

The instantaneous graph is polarised in that if cell division is successful or unsuccessful with certainty, i.e. division probability of 0 or 1. In contrast, the randomised protocol has more considerable

variation in division probability. Evolution would likely "choose" the bond transform protocol that would yield the most successful cell division probability. Therefore, it may be more likely that constricting filaments disassemble instantaneously. This argument agrees with previous observations as division is known to start quickly due to the fast degradation of the CdvB scaffold. However, this can only be verified by experimental observation which is currently not possible.

The midcell diameter was calculated over simulation time for each configuration where a mean was taken of the ten seeds. Fig. 22 is a graph of midcell diameter over time, averaged for the simulations representing randomised constriction and $\sigma_{cyto} = 7$ for all packing fractions. The results of $\sigma_{cyto} = 7$ for randomised constriction is discussed because it is the configuration with the greatest amount of variation in cell division success.
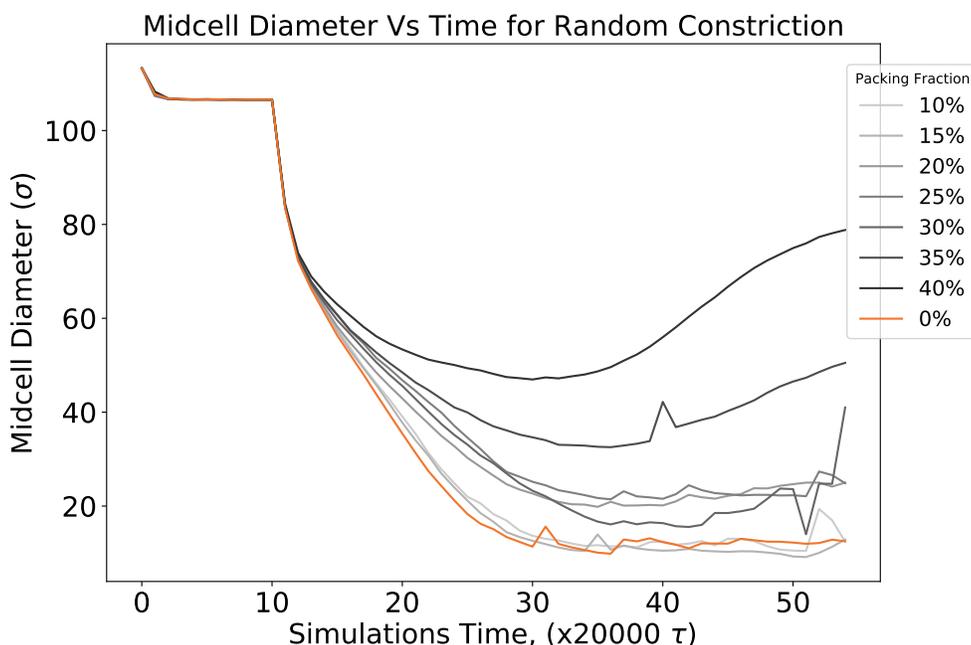
Figure 22: A diagram to show how the midcell diameter against time is dependent on the cytoplasm packing of the cell where all the particles are the same radius, $\sigma_{cyto} = 7$. Only 55 out of 92 total output timesteps are displayed because the remaining timesteps are just for equilibration. Each timestep in the graph is $2 \cdot 10^4 \, \tau$.

Although, in these simulations, positions are recalculated every $0.01 \, \tau$, the results are only outputted every $2 \cdot 10^4 \, \tau$. This is to reduce output file size. Hence, each timestep referred to in the following section is $2 \cdot 10^4 \, \tau$, and is illustrated by the x-axis of Fig 22.

As discussed previously, the simulation is initiated with the relaxed filament. The filament is then stretched and adsorbed to the membrane, which causes the cell membrane diameter to decrease

slightly, which can be seen in timestep 0 in Fig. 22. The system then equilibrates for 10 timesteps; thus, the graph plateaus at 106 σ.

Then, Fig. 22 shows that there is a sharp decrease in midcell diameter between 10-13 timesteps and that it is identical for all the simulations regardless of the number of cytoplasm particles. This is because, at the beginning, there is effectively no cytoplasm in the way of the constricting filament. Then, further constriction causes the volume of the cell to decrease so that the cytoplasm particles occupy a greater volume fraction and produces a pressure that pushes back against the filament. Fig. 22 shows that none of the simulations exhibit a linear rate of midcell diameter decrease throughout the simulation, so there is still a discrepancy between experimental and simulation measurements such as that in Fig. 2. However, if you discount the initial rapid decrease, the graph appears to be linear during the majority of the division between $\sim 13 - 30$ timesteps. At timesteps > 30, there is a plateau in diameter or an increase dependent on if cell division is successful. The gradual decrease before the plateau is due to the filament creating a bottleneck and slowly disassembling. The plateau occurs at midcell diameter ~ 10 σ rather than diameter $= 0$ σ despite the two cells being separate (see Fig. 19). When cell division fails, the constricting filament no longer exerts a force on the cell membrane. Therefore, the cell gradually returns to its spherical shape, which is why the diameter increases again over time.

Fig. 22 demonstrates how the percentage packing of the cell, or the number of cytoplasm particles, affects the midcell diameter over time. Division occurs when the diameter reaches approximately 20 σ. At higher packing fractions, the cell is prevented from decreasing to 20 σ, so division fails. Also, the greater the packing of the cell, the slower the rate of decrease, as can be seen from the shallower gradient of the curves at higher packing fractions. The orange line represents a cell without cytoplasm particles, and it has the steepest gradient and divides the fastest.

These results may explain why cells are filled 20-30% with macromolecules. It is possible that if cells were filled more, cell division would not be able to occur or would be too slow and would, therefore, require more energy.

3.4   Further Work

In the modified cell division model, cell division fails when pressure from the interacting cytoplasm particles pushes constricting filament and prevents constriction. The cytoplasm of real cells that start to divide only to fail and return to its original shape should be investigated to see if they contain an above-average density of molecules.

The dividing cells in these simulations divide symmetrically, however, as discussed in Section 2.3, some *S. acidocaldarius* cells divide asymmetrically. By changing different simulation parameters, more testing can be conducted to see under what regime asymmetric division may be successful. Arguably, for the model to replicate experimental results, it must be able to divide symmetrically and asymmetrically.

Multiple configurations of multiple factors lead to successful simulated cell division. As such, there is no way of knowing which parameters are representative of a real dividing cell. Moreover, it is difficult to find new information on the dynamics of a cell using simulations if the simulations are recreating experimental observations as the programmer of the simulations may become biased. One way to remove this bias is to use evolutionary computation. Evolutionary computation describes the use of optimisation algorithms to reproduce natural selection. These optimisation algorithm has already been previously used in conjunction with MD software to "evolve" ligand-covered nanoparticles.[43] It would, therefore, be possible to apply a similar method to evolve a cell division model. For a cell division model, the optimisation algorithm would work as follows, as shown in Fig. 23.
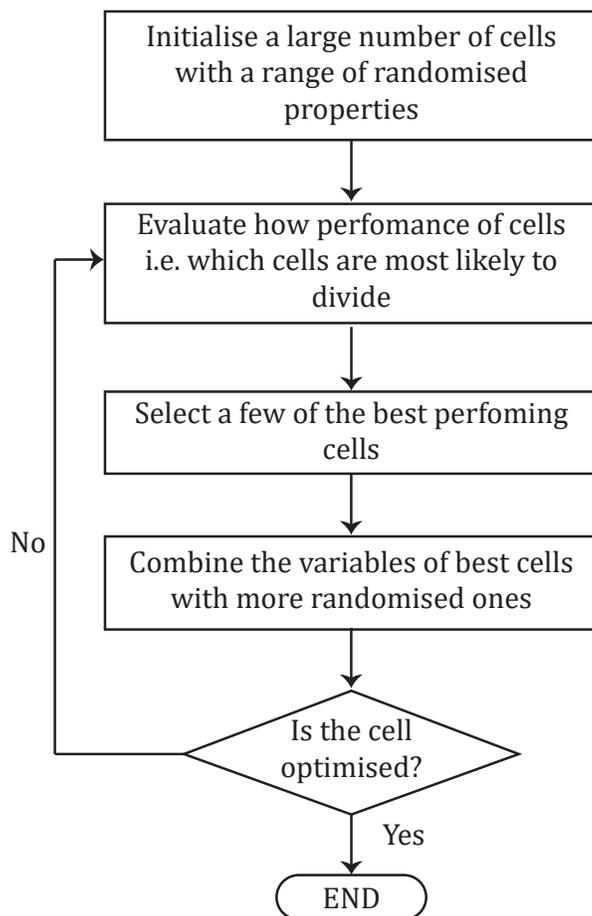
Figure 23: A simple schematic to show how evolutionary computing can be used to find a cell division model.

Usually, one would expect many optimised solutions found using an algorithm such as the one in Fig. 23. However, as discussed before, the cell division machinery found in archaeons is highly evolutionarily conserved as they are genetic homologues to cell division machinery in Eukaryotes. As such, there are likely only a few set of variables that would optimise the cell.

4 Conclusion

This project aimed to understand better how *S. acidocaldarius* cells divide and focussed on two objectives to achieve this aim. The first objective was to measure the midcell diameter of live-cell imaged dividing *S. acidocaldarius* cells over time. Then, to use these measurements to quantify the shape and rate of division. The midcell diameter of 54 *S. acidocaldarius* cells were measured using image processing software ImageJ over estimated division time. Power curves were fitted to the measurements of the form Eq. 3 where the coefficients α and c quantified the shape and rate of division, respectively. The coefficients of the fitted power curves were averaged over the 54 cells where the 54 cells were split into two groups according to their phenotypes, 32 WT and 22 MW001. On average, $\alpha = 0.91 \pm 0.25$ and $c = -0.05 \pm 0.02 \, \mu m \, min^{-1}$ for WT cells and $\alpha = 0.90 \pm 0.23$ and $c = -0.07 \pm 0.03 \, \mu m \, min^{-1}$ for MW001 cells. These measurements had a high degree of uncertainty due to the low resolution of the images. The data with the highest degrees of uncertainty were removed. This was data collected from cells that were difficult to measure due to crowding from other cells. Also, the R^2 values of each curve fit was calculated, and the fits with $R^2 < 0.9$ were removed. As a result, 11 WT and 10 MW001 cells were left and on average $\alpha = 0.96 \pm 0.13$ and $c = -0.04 \pm 0.02 \, \mu m \, min^{-1}$ for WT cells and $\alpha = 1.01 \pm 0.16$ and $c = -0.05 \pm 0.02 \, \mu m \, min^{-1}$ for MW001 cells. These values have a lower associated uncertainty and are arguably more reliable despite being taken from a smaller sample size. The results suggest that division is linear and division rate is approximately $c = -0.05 \, \mu m \, min^{-1}$, which is approximately a decrease in 5% of the total diameter per minute. The largest error associated with these measurements is the division time which had to be estimated by eye. There is currently no way of determining at what point a cell starts and stops dividing. The measurements start when the cells visibly appear to invaginate; however, it is likely that division starts occurring earlier. Measurements stop when the diameter of the cell starts to plateau because the cells stick together even after cell division. Therefore, it is likely that some measurements are taken after division stops occurring. As a result, there is a disproportionate value of $\alpha < 1$ compared to $\alpha > 1$, where the rate of division appears to slow down due to measurements entering the plateau region. Collaborators with the Baum group are developing machine learning algorithms, which would remove the human error associated with estimating when cells start and stop dividing. Furthermore, the measured cells had been immobilised, and the WT cells contained DNA dye, both factors are known to affect the rate of cell division. Non-immobilised WT cells without DNA dye were imaged by Pulschen et al. in the late stages in the project, and the preliminary results are found in Appendix A.

The second objective was to include "cytoplasm" particles to an existing cell division model, based on *S. acidocaldarius*, to see how the presence of cytoplasm particles may affect the success of the cell division model. Prior to this project code was written for an archaeal cell model by Harker-Kirschneck et al. that successfully recreated cell division. The model consisted of a cell membrane

and a constricting filament. Cells are filled with 20-30% macromolecules in the cytoplasm; therefore, the code was modified to add "cytoplasm" particles to the centre of the simulated cell. Analysis of the modified model was carried out to show that the addition of cytoplasm in the simulated cell affects the internal dynamics of the dividing cell. The greater the number of the particles inside the cell, the less likely division occurs as the internal pressure pushes away the constricting filament. Furthermore, the greater the number of cytoplasm particles, the slower the rate of division when successful.

In conclusion, evidence was given that suggests *S. acidocaldarius* cells divide at a linear rate in agreement with previous calculations. This suggests that the constricting force driving cell division in crenarchaeons and cell cutting in eukaryotes is constant. Live-cell images show that division can occur asymmetrically suggesting that the constriction force does not have to be uniform in every direction. Furthermore, cell division simulations suggest that the presence of cytoplasm particles within a cell may slow the rate of division or even prevent division at high enough concentrations. More research needs to be conducted to understand the action of the ESCRT-III/CdvB filament and under what conditions the filament fails to perform its role. If these conditions are known, research can be focussed into preventing ESCRT-III from failing in its vital roles in our own animal cells.

References

[1] Lindås, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J. and Bernander, R., 2008. A unique cell division machinery in the Archaea. Proceedings of the National Academy of Sciences, 105(48), pp.18942-18946.

[2] Pulschen, A.A., Mutavchiev, D.R., Sebastian, K.N., Roubinet, J., Roubinet, M., Risa, G.T., van Wolferen, M., Roubinet, C., Culley, S., Dey, G. and Alberts, S.V., 2020. Live cell imaging of the hyperthermophilic Archaeon Sulfolobus acidocaldarius identifies complementary roles for two ESCRTIII homologues in ensuring a robust and symmetric cell division. bioRxiv.

[3] Risa, G.T., Hurtig, F., Bray, S., Hafner, A.E., Harker-Kirschneck, L., Faull, P., Davis, C., Papatziamou, D., Mutavchiev, D.R., Fan, C. and Meneguello, L., 2019. Proteasome-mediated protein degradation resets the cell division cycle and triggers ESCRT-III-mediated cytokinesis in an Archaeon. Biorxiv, p.774273.

[4] Woese, C.R., Kandler, O. and Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences, 87(12), pp.4576-4579.

[5] Hirota, Y., Ryter, A. and Jacob, F., 1968, January. Thermosensitive mutants of E. coli affected in the processes of DNA synthesis and cellular division. In Cold Spring Harbor symposia on quantitative biology (Vol. 33, pp. 677-693). Cold Spring Harbor Laboratory Press.

[6] Normand, G. and King, R.W., 2010. Understanding cytokinesis failure. In Polyploidization and cancer (pp. 27-55). Springer, New York, NY.

[7] Hurley, J.H., 2010. The ESCRT complexes. Critical reviews in biochemistry and molecular biology, 45(6), pp.463-487

[8] Erickson, H.P., 2009. Modeling the physics of FtsZ assembly and force generation. Proceedings of the National Academy of Sciences, 106(23), pp.9238-9243.

[9] Lafaurie-Janvore, J., Maiuri, P., Wang, I., Pinot, M., Manneville, J.B., Betz, T., Balland, M. and Piel, M., 2013. ESCRT-III assembly and cytokinetic abscission are induced by tension release in the intercellular bridge. Science, 339(6127), pp.1625-1629.

[10] Ng, K.H., Srinivas, V., Srinivasan, R. and Balasubramanian, M., 2013. The Nitrosopumilus maritimus CdvB, but not FtsZ, assembles into polymers. Archaea.

[11] Vellai, T. and Vida, G., 1999. The origin of eukaryotes: the difference between prokaryotic and Eukaryotic cells. Proceedings of the Royal Society of London. Series B: Biological Sciences, 266(1428), pp.1571-1577.

[12] Swaffer, M.P., Jones, A.W., Flynn, H.R., Snijders, A.P. and Nurse, P., 2016. CDK substrate phosphorylation and ordering the cell cycle. Cell, 167(7), pp.1750-1761.

[13] Carlton, J.G. and Martin-Serrano, J., 2007. Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. Science, 316(5833), pp.1908-1912.

[14] Votteler, J. and Sundquist, W.I., 2013. Virus budding and the ESCRT pathway. Cell host & microbe, 14(3), pp.232-241.

[15] Wollert, T. and Hurley, J.H., 2010. Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. Nature, 464(7290), pp.864-869.

[16] Schmidt, O. and Teis, D., 2012. The ESCRT machinery. Current Biology, 22(4), pp.R116-R120.

[17] Cashikar, A.G., Shim, S., Roth, R., Maldazys, M.R., Heuser, J.E. and Hanson, P.I., 2014. Structure of cellular ESCRT-III spirals and their relationship to HIV budding. Elife, 3, p.e02184.

[18] Hanson, P.I., Roth, R., Lin, Y. and Heuser, J.E., 2008. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. The Journal of cell biology, 180(2), pp.389-402.

[19] Babst, M., Katzmann, D.J., Snyder, W.B., Wendland, B. and Emr, S.D., 2002. Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. Developmental cell, 3(2), pp.283-289.

[20] Leigh, J.A., Albers, S.V., Atomi, H. and Allers, T., 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. FEMS microbiology reviews, 35(4), pp.577-608.

[21] Poole, A.M. and Neumann, N., 2011. Reconciling an Archaeal origin of eukaryotes with engulfment: a biologically plausible update of the Eocyte hypothesis. Research in microbiology, 162(1), pp.71-76.

[22] Samson, R.Y., Obita, T., Freund, S.M., Williams, R.L. and Bell, S.D., 2008. A role for the ESCRT system in cell division in archaea. Science, 322(5908), pp.1710-1713.

[23] Baker, M., 2010. Taking a long, hard look. Nature, 466(7310), pp.1137-1138.

[24] Zhao, F., 2004. Confocal Microscopy Tutorial: Lateral And Axial Resolution In Confocal System. [online] http://www.hi.helsinki.fi/amu/. Available at: http://www.hi.helsinki.fi/amu/AMU%20Cf_tut/cf_tut_part1-5.htm [Accessed 6 March 2020].

[25] Hogeweg, P., 2011. The roots of bioinformatics in theoretical biology. PLoS computational biology, 7(3).

[26] Peck, S.L., 2004. Simulation as experiment: a philosophical reassessment for biological modeling. Trends in ecology & evolution, 19(10), pp.530-534.

[27] Tyson, J.J. and Hannsgen, K.B., 1986. Cell growth and division: a deterministic/probabilistic model of the cell cycle. Journal of mathematical biology, 23(2), pp.231-246.

[28] Frenkel, D. and Smit, B., 2001. Understanding molecular simulation: from algorithms to applications (Vol. 1). Elsevier.

[29] Schlick, T., 2010. Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide (Vol. 21). Springer Science & Business Media.

[30] Rapaport, D.C., 1978. Molecular dynamics simulation of polymer chains with excluded volume. Journal of Physics A: Mathematical and General, 11(8), p.L213.

[31] Eisenberg, B., Hyon, Y. and Liu, C., 2011. A mathematical model for the hard sphere repulsion in ionic solutions. Communications in Mathematical Sciences, 9(2), pp.459-475.

[32] Rohrlich, F. 1990. Computer simulation in the physical sciences. In PSA: Proceedings of the biennial meeting of the philosophy of science association (Vol. 1990, No. 2, pp. 507-518). Philosophy of Science Association.

[33] Hafner, A. E., Krausser, J., and Šarić, A. (2019). Minimal coarse-grained models for molecular self-organisation in biology. Current opinion in structural biology, 58, 43-52.

[34] Yuan, H., Huang, C., Li, J., Lykotrafitis, G. and Zhang, S., 2010. One-particle-thick, solvent-free, coarse-grained model for biological and biomimetic fluid membranes. Physical review , E 82, 011905.

[35] Drouffe, J.M., Maggs, A.C. and Leibler, S., 1991. Computer simulations of self-assembled membranes. Science, 254(5036), pp.1353-1356.

[36] Harker-Kirschneck, L., Baum, B. and Šarić, A., 2019. Transitions in filament geometry drive ESCRT-III-mediated membrane remodelling and fission. bioRxiv, p.559898.

[37] Harker-Kirschneck, L., Baum, B. and Šarić, A., Supplementary Information Transitions in filament geometry drive ESCRT-III-mediated membrane remodeling and fission.

[38] Ellis, R.J., 2001. Macromolecular crowding: obvious but underappreciated. Trends in biochemical sciences, 26(10), pp.597-604.

[39] Plimpton, S., 1993. Fast parallel algorithms for short-range molecular dynamics (No. SAND-91-1144). Sandia National Labs., Albuquerque, NM (United States).

[40] Rapaport, D. C., 2004. The Art of Molecular Dynamics Simulation. Cambridge University Press. ISBN 978-0-521-82568-9. pp.14.

[41] Gibbs, J.W., 1902. Elementary principles in statistical mechanics: developed with special reference to the rational foundation of thermodynamics. C. Scribner's sons.

[42] Taubin, G., 1991. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (11), pp.1115-1138.

[43] Forster, J.C., Krausser, J., Vuyyuru, M.R., Baum, B. and Šarić, A., 2020. Designing membrane-reshaping nanostructures through artificial evolution. bioRxiv.

APPENDIX A

Fig. A1 a) shows the power curve plots of 17 WT cells that are non-immobilised and without phototoxic DNA dye. The start and end frames measured are determined according to a machine algorithm in contrast to human judgement as outlined in the project.

As expected, Fig. A1 c) shows that the average rate of division is much faster $c = 0.09\,\mu m\,min^{-1}$ than cells discussed in Section 2.3, which had factors that reduced division rate. The preliminary results, in Fig. A1 b), show that the average α value is 1.12 ± 0.51, which adds further validation to the conclusion that the cells divide at a linear rate. However, the standard deviation of the results is substantial, as α varies between 0.12 - 2.03. These extreme values may be outliers and attributed to an algorithm deciding when division starts and stops as opposed to human judgement.
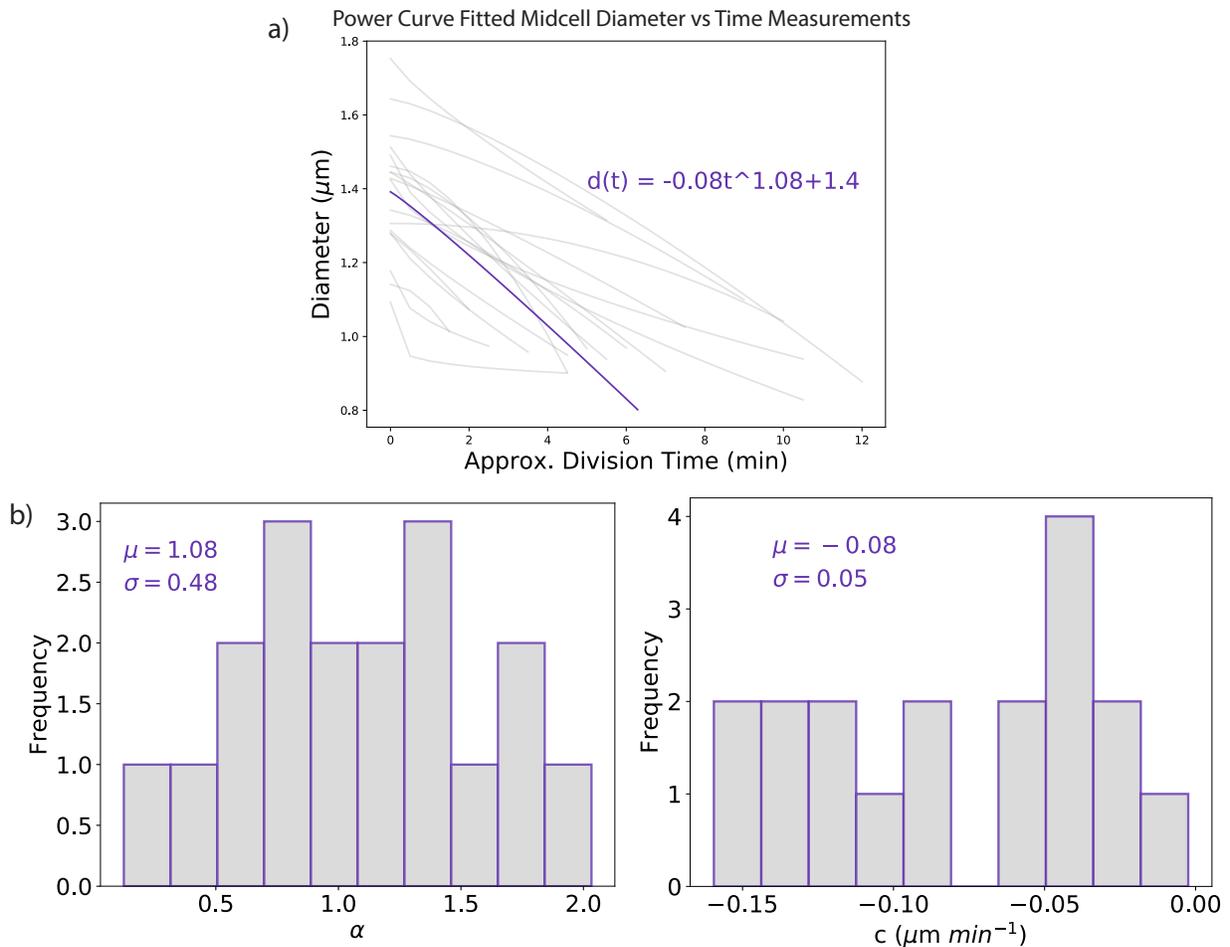
Figure A 1: a) Power curves of 17 WT cells with the average of the coefficients plotted in purple. b) Distribution of α (shape of graph) values of the 17 cells. c) Distribution of c (rate of division) values.