TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

ANTTI MARTIKAINEN

SEGREGATION OF PROTEIN AGGREGATES IN ESCHERICHIA COLI

Master of Science Thesis

Examiners:
Associate Prof. Andre S. Ribeiro,
Prof. Risto Ritala

The examiners and topic of the thesis were approved by the Council of the Faculty of Engineering Sciences on 8th of October 2014

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Automation Engineering
MARTIKAINEN, ANTTI: Segregation of protein aggregates in *Escherichia coli*
Master of Science Thesis, 52 pages
March 2015
Major: Systems analysis
Examiners: Associate Prof. Andre S. Ribeiro, Prof. Risto Ritala
Keywords: *Escherichia coli*, image based measurements, cellular ageing, protein quality control, protein misfolding, protein aggregates, inclusion bodies

Not long ago it was believed that symmetrically dividing bacteria didn't age, but fairly recently this belief has been challenged, as the development of microscopy techniques has allowed more detailed study of these organisms. The discovery of bacterial ageing has made it possible to study fundamental ageing related mechanisms in organisms where they can be observed on a molecular level with relative ease. As a result, the accumulation of misfolded proteins, corrupted end products of gene expression, has been identified to be the main ageing factor in *Escherichia coli*. This accumulation is also associated with many diseases in higher organisms, including humans. Studying the basic concepts related to ageing could thus potentially provide clues to develop methods of managing these diseases in the future, as well as improve our general understanding of ageing and it's evolutionary origin.

The research this thesis is a part of takes a look at the effect of different types of stress on protein production and the mechanisms that cells use to cope with corrupted proteins. The main focus here is on the robustness of one of the mechanism the cells use to mitigate protein damage, which is the segregation, retention and eventual asymmetric inheritance of unwanted protein aggregates. The nucleoid, a denser region containing the genetic material of the cell, has been recently identified to have an instrumental role in this mechanism in *Escherichia coli* cells. Here we study the effects of stress, which alters the size of this region on the robustness of this mechanism. Some related results are also presented regarding the effects of stress on gene expression dynamics.

The segregation and retention mechanisms are studied here using time lapse microscopy measurements to observe the relative movement of the nucleoids and unwanted protein aggregates within individual cells. In addition to presenting the results, this thesis focuses on the statistical and image analysis methods used during the project, as the majority of the work done for this thesis was done on this part. These methods are introduced in chapter 3 of the thesis.

After that, we present the results which show that changes in the relative size of the nucleoid within the cell do cause significant changes in the spatial distribution and dynamics of the aggregates. Based on our observations, we conclude that even though the segregation and retention mechanisms are fairly robust to these changes, they are not completely immune. Additionally, we show that the functioning of these mechanisms seems to be optimized with moderate nucleoid sizes, whereas significant increases or decreases in nucleoid size lead to diminished functionality.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Automaatiotekniikan koulutusohjelma
MARTIKAINEN, ANTTI: Väärintaittuneiden proteiinien solun päihin eriytymisen robustisuus *Escherichia coli* bakteerissa
Diplomityö, 52 sivua
Maaliskuu 2015
Pääaine: Systeemien analysointi
Tarkastajat: Associate prof. Andre S. Ribeiro ja Prof. Risto Ritala
Avainsanat: *Escherichia coli*, kuvaan perustuvat mittaukset, solujen vanheneminen, proteiinien väärintaittuminen

Vielä jokin aika sitten uskottiin, ettei vanheneminen koskenut bakteereita. Tämä uskomus on kuitenkin tullut haastetuksi uusien mikroskopiamenetelmien mahdollistamien havaintojen myötä. Vanhenemisen merkkien havaitseminen myös bakteereissa on puolestaan mahdollistanut vanhenemisen tutkimisen organismeissa, joissa molekyylitason mekanismien havainnoiminen on verrattain yksinkertaista. Tämän tutkimuksen tuloksena on tunnistettu väärintaittuneiden proteiinien kertymisen olevan merkittävin solujen ikääntymistä aiheuttavista tekijöistä *Escherichia coli* bakteerin tapauksessa. Vahingoittuneiden proteiinien kertyminen on yhdistetty myös moniin sairauksiin monimutkaisemmissa eliöissä, kuten ihmisissä. Siten ikääntymiseen ja vahingoittuneisiin proteiineihin liittyvien perusmekanismien tutkiminen voi vanhenemisen ja sen alkuperän paremman ymmärryksen lisäksi tulevaisuudessa kenties edistää uusien hoitomuotojen kehitystä.

Tässä työssä esitelty tutkimus perehtyy stressitekijöiden vaikutuksiin proteiinien tuotannossa ja niissä mekanismeissa, joita solut käyttävät proteiinivaurioiden vaikutusten rajoittamiseen. Työn päähuomio keskittyy erään tällaisen mekanismin robustisuuteen. Kyseinen mekanismi on proteiinikertymien eristäminen solun päätyihin, mikä puolestaan useamman jakautumisen jälkeen johtaa solujen välisiin eroihin perittyjen kertymien määrissä ja sitä myötä vanhenevan ja nuorenevan fenotyypin syntyyn. Solun keskellä olevan nukleoidi-rakenteen on havaittu vaikuttavan huomattavasti näiden erojen syntyyn. Tässä työssä tutkimme nukleoidin kokoon vaikuttavien olosuhteiden vaikutusta mekanismin toimintaan. Projektin taustana esittelemme lyhyesti myös stressin vaikutuksia geenien ilmentymiseen.

Projektissa käytetään eristysmekanismin tutkimiseen multimodaalimikroskopiaa proteiinikertymien ja nukleoidien vuorovaikutuksen havainnoimiseksi yksittäisten solujen sisällä. Tulosten lisäksi tässä työssä esitellään mittausten analysointiin käytettyjä tekniikoita, koska suurin osa opinnäytteen työpanoksesta tehtiin projektin tähän osaan. Näitä ja muita projektiin liittyviä menetelmiä esitellään tarkemmin luvussa 3.

Tämän jälkeen esittelemme tulokset, joiden perusteella voidaan todeta nukleoidin koon muuttaneen proteiinikertymien jakautumista solun sisällä merkittävästi. Tutkimme myös kertymien liikettä ja sijainteja solujen sisällä nukleoidin koon funktiona. Havaintojemme perusteella voimme todeta tutkitun mekanismin olevan hyvin robusti tutkituille olosuhteille, mutta ei kuitenkaan ole täysin immuuni. Mekanismin toiminta vaikuttaa olevan jotakuinkin optimaalista nukleoidin ollessa keskikokoinen, kun taas merkittävä kasvu tai kutistuminen vaikuttaisi vähentävän mekanismin toimintatehoa.

PREFACE

The work for this thesis was carried out in the Laboratory of Biosystem Dynamics, part of the Department of Signal Processing at Tampere University of Technology. The work presented here was mainly carried out between April 2014 and January 2015, but also includes material from a previous related project.

The author's contribution to the research presented here in this thesis was the image and data analysis of the microscopy measurements, as well as participation to the preparation of the two resulting scientific publications. At the time of writing the research paper presenting the main results within this thesis is under review at Journal of Bacteriology, whereas the other paper containing results used in this thesis has been previously published in the peer reviewed journal “PLoS ONE” (see [1]). Additionally, the research done during this thesis has resulted in presentations at the following scientific conferences: Lyon SysBio (19th to 21th of November 2014, INSA de Lyon, Villeurbanne, France), Molecular Genetics of Bacteria and Phages meeting (Aug. 5-9, 2014, Madison, WI, USA), Systems Biology (WCSB 2014) (May 15-16, 2014, Lisbon, Portugal) and conference on Molecular genetics of bacteria and phages (6 - 10 August 2013, University of Wisconsin, Madison, U.S.A). The author of this thesis was credited as a co-author in each of these publications.

Thanks to all the colleagues alongside whom I had the pleasure to work for more than a year. Special thanks to all those who were co-authors in the resulting research papers, Andre Ribeiro for the patience and guidance when supervising the projects and this thesis, as well as Abhisekh Gupta, Antti Häkkinen and Jason Lloyd-Price for their valuable day-to-day advice regarding the methods and intricacies of the used software tools.

Finally, thanks to my family for all the support over the years.

Tampere, 18th of March, 2015

Antti Martikainen

CONTENTS

TERMS, ABBREVIATIONS AND SYMBOLS

Chloramphenicol	Antibiotic, which causes the condesation of the nucleoid
HupA-mCherry	Fluorescent DNA binding protein used for tagging the nucleoid
Nucleoid	A structure within a procaryotic cell consisting of the compacted genome of the cell
Rifampicin	Antibiotic, which causes the expansion of the nucleoid
Streptomycin	Antibiotic, which prevents the transition from translation initiation to chain elongation and also causes miscoding of proteins. Used here for inducing the production of misfolded proteins as well as the IbpA chaperone
DAPI	4',6-diamidino-2-phenylindol, a fluorescent DNA dye for staining nucleoids, used here for validation purposes
E. coli	*Escherichia coli*, a rod shaped bacteria, the model organism used in this thesis
IbpA	Inclusion body protein A, chaperone which associates with damaged proteins
KDE	Kernel density estimate
LB	Lysogeny Broth, type of nutrient rich cell culturing media, used here as the control condition
M63	Minimal cell culturing media, used here to change the size of the nucleoid
MAMLE	Multi-resolution Analysis and Maximum Likelihood Estimation, an automated technique used for segmentation of microscopy images in this thesis
mRNA	Messenger ribonucleic acid
RNA	Ribonucleic acid
TB	Terrific broth, a type of nutrient rich cell culturing media, used here to alter the size of the nucleoid
YFP	Yellow fluorescent protein
Δt	Time interval between consecutive RNA production events
t_0	Activation time, the time that it takes for the first RNA

	molecule to be produced after induction
$\boldsymbol{\sigma}$**, *std***	Standard deviation
$\boldsymbol{\mu}$	Mean
$\mathbf{CV^2}$	Measure of variability, $\boldsymbol{\sigma^2/\mu^2}$
K(•)	Kernel function for the kernel density estimate
h	parameter defining the width of the kernel function
v_i	Intensity value of the ith pixel
x_i	Normalized location of the *i*th pixel along the major axis of the cell

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Proteins are the functional expression of the genetic code in each organism. They are the results of multi-stepped, stochastic processes, which first transcribe the DNA to intermediary RNA form and from there on translate it to form the complex, three dimensional arrangements of polypeptide chains that are proteins. All cellular functions are regulated by modulating these gene expression processes and the correct functioning of the resulting proteins is critical for the well being of the organism. Different types of stress can affect these processes at any of the stages by changing the dynamics of the production [1-8] or by increasing the number of errors in production, resulting in non-functional or even malfunctioning proteins [9-12].

Due to the stochastic nature of the steps involved in the production of proteins and the high number of possible disturbances, some proteins will eventually get corrupted or damaged and this process is further exacerbated by environmental stress. The accumulation of these damaged proteins has been linked to cellular ageing in simple single cell organisms [9, 11, 13-19] and several diseases, such as Alzheimer's, Huntington's and other neurodegenerative diseases in higher organisms [9, 12, 20-22]. Due to the severe consequences of this aggregate accumulation, cells have various protein quality control mechanisms in place to prevent, repair or mitigate the damage [9-11, 13-18, 23] and many of them aren't yet well understood. By studying the functioning of these mechanisms, it is possible to gain better understanding of the underlying principles related to ageing and various diseases, as well as gain knowledge, which might be relevant for the development of applications in the field of synthetic biology.

The main focus of this thesis is on one of the mechanisms cells use for damage mitigation, which is the compartmentalization of the unwanted proteins to minimize the detrimental effects. Once damaged proteins start accumulating within the cell, due to their undesired chemical properties they start forming larger and larger aggregates as they collide with each other as well as other macromolecules within the cell [9, 11]. If the cell isn't able to repair or destroy the damaged proteins, it has to limit negative the effects of the accumulating protein damage somehow to avoid a significant reduction in vitality. The damage mitigation is achieved by segregating the aggregates to specific locations within the cell and retaining them there. It will result in the unwanted proteins causing less harm and eventually, asymmetric inheritance of the aggregates into progeny after cell division [11-19, 22]. This in turn will lead to ageing of the cell inheriting the majority of the aggregates and rejuvenation of the other [11, 13-19, 22], which proves to be beneficial for the population as a whole in the case of procaryotes [12-18, 22]. This process is described in more detail in the second chapter.

Segregation and retention are studied here *in vivo*, from multi-modal single cell level microscopy measurements, using *Escherichia coli* as the model organism. More specifically, the focus of the study is on the robustness of this mechanism when subjected to stress or growth conditions, which alter the size of the nucleoid. The nucleoid is a well defined, relatively stiff structure within the cell, consisting of the compacted genetic material of the cell [24]. The nucleoid has recently been identified to have an instrumental role in the process by limiting the movement of the protein aggregates within the cell though volume exclusion [11, 13-19, 25, 26]. We study the effect of changes in nucleoid size on the movement and spatial distributions of the aggregates to gain information about the effectiveness of the compartmentalization mechanism. To do this, multiple approaches are used for changing the relative size of the nucleoids within the cells. The aggregates, as well as the nucleoids, are tagged with fluorescent proteins to capture their interaction within individual cells using time lapse microscopy. The resulting measurements are then analysed utilizing methods previously developed within the research group, as well as some custom made during the project. The measurements are used to examine the spatio-temporal characteristics of the aggregates in different conditions and correlate them with changes in nucleoid size, as well as infer metrics to describe the effectiveness of the segregation and retention mechanisms.

The results presented here are from research projects done in collaboration with other members of the Laboratory of Biosystem Dynamics at Tampere University of Technology. The research related to the robustness of the segregation and retention mechanisms, which is the main topic of this thesis, has resulted in an article, which at the time of writing is under review at the Journal of Bacteriology. Some of the research results that are presented here to illustrate concepts related to the background of the main topic have also been previously published in the peer reviewed journal “PLoS ONE” (see [1]). The scope of the work done for this thesis is mainly limited to performing the image and data analysis parts of the projects, with additional contributions to writing and preparation of the resulting scientific publications. As a result, there is a larger focus on the used analysis methods and the analysis process than is required solely for the purpose of presenting the results.

This thesis is divided into five chapters. Chapter 2 gives a brief introduction to gene expression, accumulation of protein damage, the segregation and retention of accumulated proteins and how it relates to cellular ageing in procaryotes, as well as briefly presents some related results from a previous project describing the effects of stress on gene expression. This provides the necessary background for the rest of the thesis. The third chapter moves on to describe the methods used to study the segregation and retention processes. The chapter mainly concentrates on the analysis process and utilized methods, such as preprocessing of the images, extraction of the relevant

measures and classification of the cells based on nucleoid count, but also includes a brief description of the biological methods used to perturb the cells and tag the parts of interest within the cell, as well as a description of the microscopy measurement system to provide a better understanding of the project as a whole and the type of data that was analysed to produce the results. The fourth chapter presents the results and discusses the possible implications, whereas the fifth chapter sums up the final conclusions.

2 BACKGROUND

Escherichia coli, the model organism used in this thesis, is a rod shaped bacterium, which reproduces by dividing symmetrically in the middle, resulting in morphologically identical daughter cells [22]. In the wild it is commonly found in the digestive tract of mammals. It is one of the most studied model organisms in molecular biology due to the ease of culturing, the ability to augment the strains relatively easily using synthetic plasmids and the simple structure of the cell making the analysis of microscopy measurements easier. It is a procaryote and as such doesn't have actual cell organelles or a nucleus separated by a membrane, as in eukaryotic organisms. Instead, the cell consist of a single compartment, which is filled with a fluid suspension, the cytoplasm, containing all the functional components of the cell. Even though the genetic material isn't contained within a nucleus separated by a membrane, it still forms a well defined denser structure, called the nucleoid, located at the mid-cell region.

The following subsections of this chapter will briefly cover topics in the molecular biology of *E. coli*, which are relevant to provide the sufficient background for the later parts of the thesis. First, we'll start with gene expression, which is the process of transcribing and translating genetic code within DNA molecules into functional end products, proteins.

2.1 Transcription in procaryotes

The first step in the gene expression process is transcription, which transcribes a section of the DNA into a strand of mRNA, which will in turn codes for the protein synthesis later on. Transcription starts with transcription initiation where the RNA polymerase holoenzyme, which is a combination of the RNA polymerase core enzyme and a DNA binding protein called a σ-factor, collides with the DNA molecule, slides along it and locates the promoter, which is a sequence of DNA indicating the start site for the transcription of a specific gene. The polymerase then binds to the promoter site and unwinds a part of the DNA, separating the two strands to form the open complex, where the nucleotides of the gene are exposed and ready to be transcribed. This transcription initiation is a major point of regulation in gene expression, as it is a complex process which requires a series of conformational changes from both the DNA and the RNA polymerase holoenzyme [10]. After this, the transcription starts, first slowly, but after transcribing about ten first nucleotides, the polymerase breaks its interactions with the promoter site and shifts to elongation mode to continue the transcription. Transcription then continues until the polymerase encounters the transcription termination site in the DNA, which releases the completed RNA polymerase. The transcribed RNA then folds to the three dimensional configuration

which is required for its correct functioning. For some genes the RNA molecule itself is the final end product performing structural, catalytic or regulatory roles within the cell, but here we concentrate on mRNA which codes for protein synthesis [10].

The control of gene expression in *E. coli* is believed to mainly happen at the stage of transcription, particularly its initiation [1, 10]. This is because the subsequent stages, such as elongation and termination, are much faster and the translation can begin while the mRNA molecule is still being transcribed [10, 27-30]. The relative slowness and significance of transcription initiation in the regulation is a result of it being a multi-stepped process, which requires several complex conformational changes from the involved molecules [10, 31].

Regulation of the gene expression process is essential for the correct functioning and survival of the cell. This is especially the case under stressful environmental conditions. Before moving on to translation, in the following section we will take a brief look at phenomena related to transcription regulation by going through some results from a previous related project.

2.2 Effects of stress on transcription regulation in *Escherichia coli*

Stress and various environmental conditions can affect the gene expression at various stages of the process. One of the more significant ones in terms of the regulation is transcription initiation. Stress induced damage and errors in the transcription process can also result in the eventual production of misfolded proteins. Here we will take a brief look at the effects stress can have on transcription by briefly going through the main results of a previous related project. The results presented in this section have been previously published in [1].

Escherichia coli has genetic pathways, which respond to specific types of stress, but little is known how stress affects the functioning of other, non stress related genes. In a previous related project we took a look at how mild acidic shift and oxidative stress affect the *in vivo* transcriptional kinetics of a gene under the control of a synthetic promoter (lac/ara-1) uninvolved in stress-response pathways. The gene codes for a RNA target for fluorescent proteins (MS2d-GFP), which allow its detection with single-molecule sensitivity as soon as it is produced [32].

The acidic shift and oxidative stress used to the perturb the cells were realized by adding sub-lethal amounts of 4-morpholine-methanesulfonic acid (MES) or hydrogen peroxide to the growth media. To quantify the degree of stress resulting from these additions, division times were extracted from the time-lapse measurements. For the control grown in optimal conditions we measured 66.7 minute mean division time and

38.4 minute standard deviation, whereas in acidic shift we recorded division times of 87.2 ± 52.3 minutes and in oxidative stress 91.0 ± 70.3 minutes respectively.

To study dynamics of this gene, we used fluorescence time-lapse microscopy techniques largely similar to those described in more detail later on in chapter 3. However there are few differences related to the analysis methods and these are briefly introduced below in conjunction with the results.

From the florescence microscopy measurements, we measured two main variables to characterize the gene expression dynamics in each condition. Those were the time interval between each consecutive RNA production event within a cell (later denoted also as Δt), as well as the transcription activation time, which is the time it takes for each cell to produce the first target RNA molecule after induction. Note that the activation time (later referred to also as t_0) includes both the time for a cell to uptake at least one inducer as well as the time for the first transcription event to be completed.

These measures were extracted from the microscopy images after segmentation and aggregate detection (described in more detail later in chapter 3) by using the total fluorescence of all the target RNAs within the cell over time, with method introduced in [33]. This was done by first subtracting the background fluorescence of the cell from the summed up aggregate fluorescence for each time point. Then, since the MS2d-GFP tagged RNA molecules do not degrade during measurements of a few hours [34], the moments of appearance for the target RNAs within the cell could be obtained as in [33], by least squares fitting a monotonically increasing, piecewise-constant function to the corrected total spot intensity over time. The number of terms for the fit was selected by F-test with P value of 0.01. Each jump corresponds to the production of one target RNA. This fit to the fluorescence intensity time series is illustrated below in figure 1.

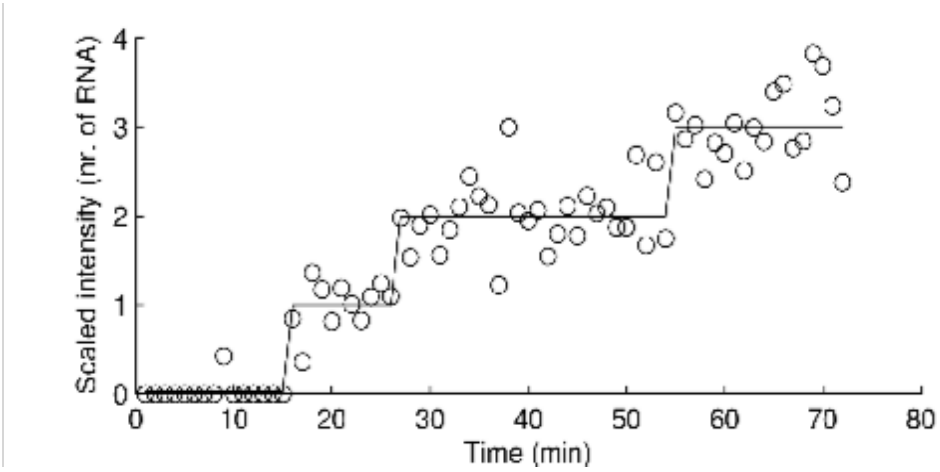

Figure 1: Illustration of the mRNA production event detection. Circles indicate the detected background corrected and normalized intensity within the cell at each time point, whereas the solid line shows the least squares fitted monotonically increasing piecewise-constant function. Each jump in the fitted function corresponds to a time when a target RNA molecule is produced. Adapted from [1]

By using this data extraction process we gained the activation time and production interval distributions for each condition, which are shown below in figures 2 and 3.

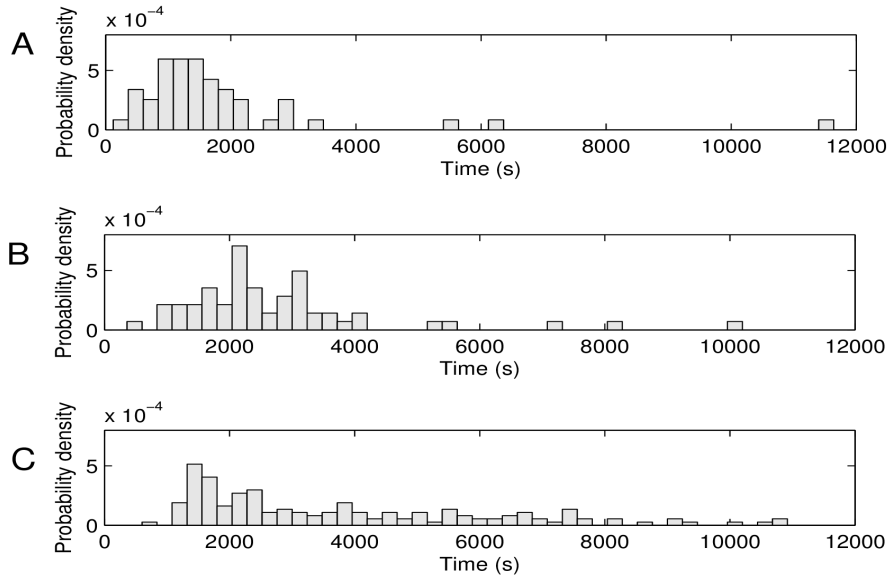

Figure 2: Transcription activation time distributions. Probability density distributions of measured activation times, t_0, in individual cells subject to (A) optimal growth conditions (55 cells), (B) acidic shift (61 cells), and, (C) oxidative stress (158 cells). Mean and standard deviation of the distributions equalled to, (A) mean(t_0) = 1871 s and std(t_0) = 1819 s, (B) mean(t_0) = 2960 s and std(t_0) = 2217 s, and (C) mean(t_0) = 3931 s and std(t_0) = 2678 [1]

The observed substantial elongations of the activation time provide strong evidence that the intake kinetics of the inducer, IPTG, are significantly altered in both stress conditions. This is likely due to a stress induced decrease in the permeability of the cell membranes, which in general leads to delays in transcriptional response to external signals [35-38].

When we moved on to study the production dynamics via the mean and standard deviations of the RNA production intervals in individual cells between conditions, we observed that in the case of acidic shift only the second hour differs significantly from the control, whereas in the case of oxidative stress the change occurs faster, as can be seen from the table 1 below.

Table 1: Comparison of measured intervals between consecutive RNA production events initiated during the first and second hours of observation in each stress condition. The p-values are from two sample KS-test of the measured interval distribution vs. that of the control. Adapted from [1]

	Control	**Acidic shift, 1st hour**	**Acidic shift, 2nd hour**	**Oxidative stress, 1st hour**	**Oxidative stress, 2nd hour**
Mean, Δt (s)	898	866	1452	1351	1904
Std, σ^2 (s)	696	599	1054	1061	1063
CV^2, σ^2/μ^2	0.60	0.48	0.53	0.62	0.31
p-value	-	0.82	< 0.01	< 0.01	< 0.01
No. of intervals	135	87	94	116	178

Based on the measured differences in the values of the CV^2 of the duration of the intervals initiated during the first hour and of the intervals initiated during the second hour, we expected the cell to cell diversity in numbers of RNA molecules produced during those two time periods to differ significantly. To assess this, we obtained the number of RNA molecules in each cell at the end of the first hour of the measurements and at the end of the second hour of the measurements. From this, we found that under acidic shift this quantity equalled 1.7 at both time moments. However, under oxidative stress, it equalled 1.8 at the end of the first hour, and 1.2 at the end of the second hour. Based on this, we concluded that the decrease over time in the variability of the production dynamics observed in the case of oxidative stress also leads to significant reduction the cell to cell variability in terms of RNA numbers.

Previous research has shown that the transcription initiation process can be modelled using a multiple rate limiting step model [31], with each step following an exponential distribution in duration [39].

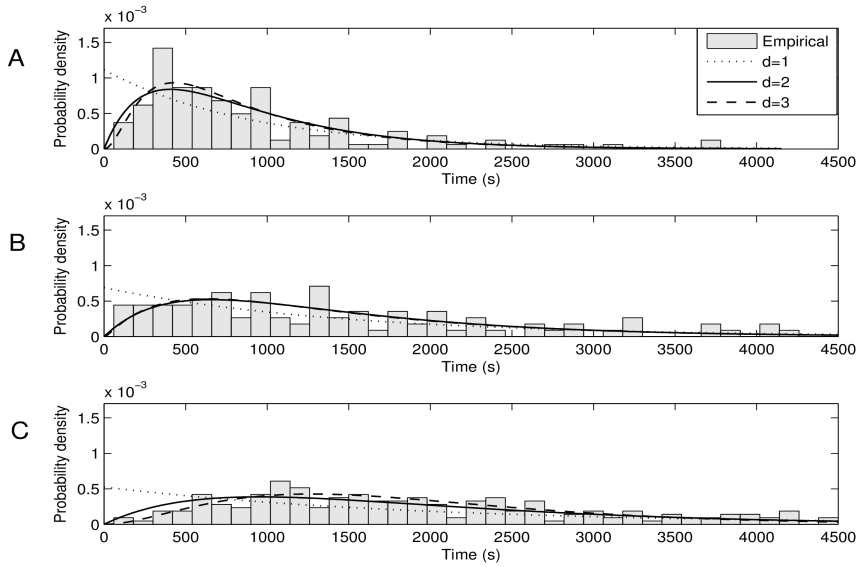

Figure 3: Experimental distributions of intervals between RNA production events and the inferred models with 1, 2 or 3 steps. A) control, optimal conditions, 139 cells, B) acidic shift, 167 cells, C) oxidative stress, 507 cells. Dotted line indicates the 1-step model, solid line indicates the 2-step model and dashed line the 3-step model inferred for each condition. Note that in some conditions the 2-step and 3-step models cannot be visually distinguished. [1]

Using this information, the approximation for the duration of these steps can be inferred from the observed distribution of intervals as in [39]. This assumes that the duration of each step follows an independent exponential distribution. The most likely durations can be determined by maximum likelihood from the distributions for all the candidate degrees of the model. The degree of the model (i.e. the number of rate limiting steps) can then be determined by a likelihood-ratio test between pairs of models. The degree of the model is grown to maximize the likelihood for the given data, until the likelihood ratio test shows no significant improvement, in which case the higher degree model is rejected in favour of the lower degree one. Finally, the goodness of the fit for the chosen model is evaluated by a KS-test between the inferred model and the measured data [10].

Using this method we inferred the following durations for rate limiting steps in each condition: control (284 s, 614 s), acidic shift (456 s, 996 s), oxidative stress (635 s, 635 s, 635 s). The p-values resulting from a KS-test between each of the inferred models and the corresponding measured distribution were all larger than 0.01 (0.2 for control, 0.7 for acidic shift, and 0.6 for oxidative stress), from which we conclude that all models match the corresponding empirical data in a statistical sense. In the case of acidic shift a

significant elongation in the durations of the steps can be seen, but the number of the rate limiting steps remains unchanged, whereas in the case of oxidative stress we observe the appearance of a third rate limiting step. The detected rate-limiting steps likely occur in transcription initiation rather than elongation, since elongation only lasts for tens of seconds [40, 41], while the changes in the intervals between consecutive RNA production events were of the order of hundreds of seconds.

Note that the described inference procedure doesn't provide information about the temporal order of the steps. Also note that due to a previously reported unknown artefact [39, 42], duration of the steps inferred by this method is biased towards a solution which gives identical values to the steps if the actual underlying durations for them are similar enough. We performed a rough estimation based on the number of samples, for the minimum ratio between these durations that would allow the inference algorithm to be more likely to return a non-gamma solution. This estimation resulted in ratio of 1.30. Thus, using this method of inference, we can only conclude that in this case the three rate-limiting steps differ in duration by less than 30%.

As a conclusion based on the results, application of either of the tested stress conditions reduces the mean rates of transcription activation and subsequent RNA production of the probe gene. This is particularly true under oxidative stress. Meanwhile, the noise in RNA production decreases under oxidative stress, but not under acidic shift. Based on the inferred values for the rate limiting steps, both conditions cause distinct changes in the durations of the steps and overall the changes differ in the two stress conditions. These changes are likely to be caused by global stress response mechanisms. These results also support the hypothesis that the genome-wide reduction in RNA and protein numbers of genes unrelated to stress [35-37] is achieved by *E. coli* by decreasing transcription rates rather than, expending more energy by increasing the rates of RNA and/or protein degradation.

2.3 Translation, misfolding and formation of protein aggregates

Next in the gene expression process comes translation, which synthesizes the specific polypeptide chains of proteins by linking together amino acids according to the information coded in the transcribed RNA. Protein synthesis is performed in the ribosome, which is a complex catalytic machine made from tens of different proteins and multiple ribosomal RNA molecules. In the case of procaryotes, the transcribed mRNA needs no additional preprocessing to code for the protein synthesis. As the ribosome has access to the mRNA during the transcription process, translation can also begin while the elongation step of transcription is still in progress [10].

Even though protein synthesis is a fairly rapid process in bacteria, having rates of around 20 amino acids per second per ribosome, accuracy of translation is on the order of one mistake per 10^4 amino acids. As higher degree of order always has it cost and proofreading and correction mechanisms add to this, reaching this accuracy requires the expenditure of a great deal of energy [10]. Failures in this process can result from mutations or post-transcriptional modifications among other things [9].

After creating the polypeptide chain, folding of the molecule still needed for the molecule to obtain the chemical properties required for fulfilling its intended biological role within the cell. Even if the translation phase completes without errors, protein folding is still a fairly inefficient and error prone stochastic process. Errors can be caused for example by environmental stress or molecular crowding due to high concentrations of other macro-molecules within the cell [9, 10]. The resulting misfolded proteins can have undesired reactive properties due to abnormally exposed hydrophobic regions, which among other things, allow them to aggregate together with other proteins and macromolecules [9, 10, 23]. Molecules called chaperones guide the correct folding of the proteins as well as function as repair, disposal and damage control mechanisms in case of misfolded proteins [10, 23]. We will be using a fluorescent variant of one of these chaperones (IbpA) later on in this thesis to tag the protein aggregates for observation using fluorescence microscopy.

2.4 Bacterial ageing

Bacterial ageing is a phenomenon closely related to protein misfolding. A major part of the research related to ageing has previously been done on organisms where the signs of ageing and the fact that reproduction creates distinct, rejuvenated offspring are easily identifiable. For a long time the functional asymmetries between the offspring in organisms with morphologically symmetric division, such as *E. coli*, remained unknown and so it was believed that they didn't age and were as such functionally immortal [12, 14, 22].

With recent advances in microscopy techniques data analysis capabilities, it has become possible to study these organisms in more detail and this has revealed that even in species where division is morphologically seemingly symmetric, there are distinct functional asymmetries after division. This includes the general loss of fitness typically associated with ageing in other organisms: diminished growth, decreased reproduction rate and higher chance of death [11, 14, 22]. This discovery has enabled the study of underlying concepts related to ageing in organisms where it is easier to accurately identify the relevant molecular mechanisms.

The following subsections provide a brief outline covering the the basics of cellular ageing in procaryotes to provide the necessary background for the later parts of this thesis.

2.4.1 Accumulation of damaged proteins

In bacteria the accumulation of aggregates containing damaged proteins, also often referred to as inclusion bodies, leads to loss of function, decreased reproductive capability and increased incidence of cell death. [9, 19, 22, 23]. This is the bacterial equivalent for ageing. In more complex organisms, such as humans, the accumulation of misfolded proteins has been associated with a variety of hereditary and neurodegenerative diseases, such as Alzheimer's disease [9, 10].

Due to the frequency of errors in the folding process and the severity of potential consequences, cells have various protein quality control mechanisms in place to fight this problem. One of the major components responsible for the quality control within cells are molecules called chaperones which function in a variety of roles. First line of defence is prevention, where chaperones guide the correct folding. After that come repair mechanisms, and if they fail, chaperones mark damaged proteins to be targeted by proteolytic degradation processes used to break down the harmful molecules [10, 23]. The final option for damage mitigation is the compartmentalization of the aggregates to minimize the toxic effects [9, 12, 23]. This is the mechanism to which the rest of this thesis is dedicated.

2.4.2 Segregation of protein aggregates and partitioning upon division

To mitigate the build up of harmful proteins, *E. coli* isolates the above mentioned aggregates by containing them at the cell poles [9, 12, 23, 25]. Based on current knowledge [11, 25], this is achieved in *E. coli* through the combined effect of purely diffusive Brownian motion of the proteins within the cytoplasm and coagulation of proteins into larger aggregates, coupled with the occlusion effect of the nucleoid, which is a denser structure at the mid-cell region, containing the compacted genome of the organism [24]. Current research indicates that the denser nucleoid obstructs the diffusive movement of the aggregates through the mid-cell region to a sufficient degree to generally retain the aggregates at the pole region once they enter it. Similar nucleoid occlusion effects have also been observed when studying the movement of synthetic RNA molecules [26, 44]. Additionally, several *in silico* experiments have been conducted [11, 25, 26, 43] to explore the feasibility of the nucleoid occlusion hypothesis and its consequences, leading to the result that the observed behaviour could be mainly, if not purely, the result of nucleoid occlusion coupled with the diffusion and coagulation of the aggregates.

The role of the volume exclusion caused by the nucleoid is further reinforced by recent experiments where *E. coli* cells were put through special treatments, which led to formation of either multinucleoid or anucleoid cells [11]. In the case of multinucleoid cells it was observed that in addition to localizing at the pole regions, the aggregates also accumulated in the nucleoid free regions between the nucleoids, whereas no significant preference for the pole region was observed in the case of anucleoid cells.

Even though the movement of the aggregates seems to be purely diffusive [23, 25], and as a result the partitioning of the aggregates follows an unbiased binomial distribution with aggregates having no preference for either of the poles, when combined with cell division, this process actually ends up increasing heterogeneity within the population [11, 14, 23, 26]. This is due to the fact that when a cell divides, even if the unwanted proteins would divide up equally between the daughter cells, each of the resulting daughter cells ends up with one “old pole”, which already was a cell pole previously and one “new pole”, which was just created through division. The old pole contains the unwanted aggregates that have been inherited through division, but the new pole is clean until new aggregates form or drift there. If we assume that all (or at least the majority) of the aggregates already at a pole will stay at the same pole through the lifetime of the cell, this means that we now have a heterogeneity in the amount of unwanted proteins between the poles, even though the segregation process itself is unbiased. Now, when a second cell division occurs, this translates into clear heterogeneity between the resulting daughter cells in the number of inherited aggregates [11, 14, 23, 26]. However, it has to be noted that there are also other possible sources of non-genetic phenotypic diversity between the resulting daughter cells, such as imperfections in the random partitioning of other low copy number macro molecules within the cell.

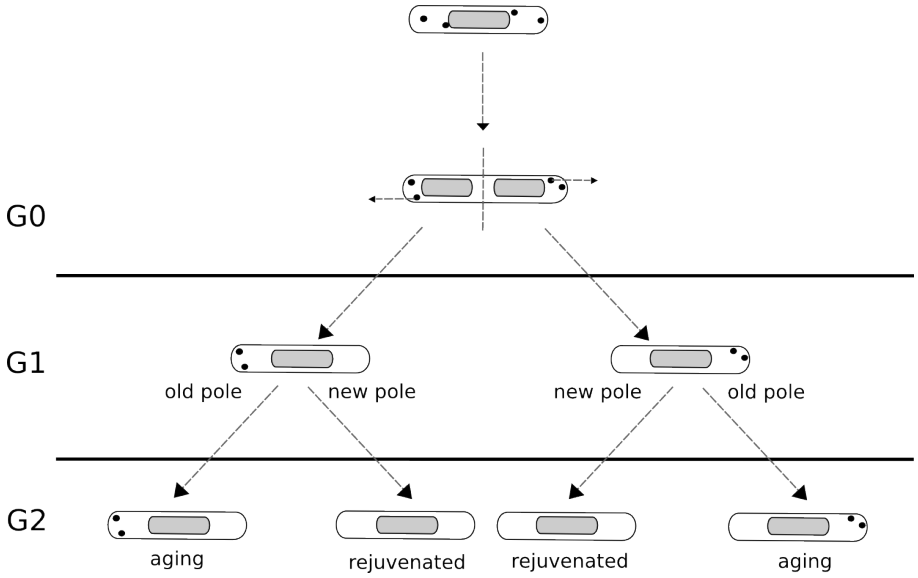

Figure 4: Illustrations of partitioning of protein aggregates upon division leading to increased variability in progeny, causing the emergence of aging and rejuvenated branches of the lineage. Black dots represent protein aggregates and the grey ellipsoids the nucleoids. This is a simplified illustration of the principle and assumes no aggregate production between generations.

This resulting asymmetry in the inheritance of protein damage actually proves to be very beneficial in terms of maintaining the health of the lineage as a whole [14-18, 25]. This would be true especially if the other methods for disposing of the unwanted aggregates prove to be insufficient or too costly due to rapid accumulation of damage, for example in a highly hostile environment [14, 17]. As the accumulation of these aggregates is linked to decreased vitality [9, 12, 14-19, 22, 23], failing to address the problem would lead to decrease in the overall fitness of the population. The existence of this mechanism, which results in asymmetric partitioning, allows the bacteria to create rejuvenated offspring even in such conditions, at the cost of the fitness of the aging “mother cell” and thus eliminating the harmful aggregates from the future population. Similar segregation and asymmetric partitioning of potential ageing factors in cell division has also been observed in other organisms beside *E. coli*, however in many other organisms additional transport mechanisms or morphological asymmetries exist to aid this process [12, 18, 45]. The fact that *E. coli* seemingly doesn't possess such active transport mechanisms to aid the disposal of unwanted aggregates has been hypothesized to be due to the smaller size and simpler structure of *E. coli* cells when compared to eukaryotes [11].

In evolutionary theory related to the origin of ageing, it has been proposed that asymmetric partitioning of accumulated damage in the reproduction process is actually a selected for trade-off to optimize the fitness of the genetic line as a whole, rather than paying the cost to optimally maintain each individual organism acting as the vessel for it [12, 17]. This would mean that ageing isn't necessarily just a mandatory evil for the bacteria, but it might actually be instrumental for ensuring the survival of the genetic line in hostile or highly competitive environments.

3 METHODS

This chapter introduces the methods used to study the functioning of the segregation and retention mechanisms as a function of the nucleoid size. First, we will briefly go through the methods used for imaging and visualization of the cells as well as the nucleoids and aggregates within them. Following this, the methods used to change the nucleoid size are briefly introduced. Finally, we will go through the image analysis methods and explain how the the main variables used for the statistical analysis were derived from the microscopy measurements.

3.1 Bacterial strain and tagging system

The strain used here for all the measurements is the *Escherichia coli* MG1655 (MGAY) strain [19] carrying the IbpA-YFP sequence in the chromosome under the control of endogenous chromosomal IbpA promoter. This IbpA-YFP sequence codes for the production of a fluorescent variant the native IbpA chaperone, which associates with damaged proteins. The production of this chaperone, as well as the accumulation of damaged proteins can be induced using Streptomycin, which prevents the transition from translation initiation to chain elongation and also causes miscoding of proteins [10]. The use of Streptomycin unfortunately also sets an upper limit of 1 hour for the usable length of the experiments, as after this the antibiotic starts to significantly affect the growth of the cells [19].

The original strain was modified by adding a plasmid expressing HupA-mCherry (a nucleoid-tagging fluorescent protein) under the control of a constitutive promoter with Ampicillin resistance. The plasmid from which HupA-mCherry originates from [24] is constitutively expressed (MGAY-HupA-mCherry strain), thus no activation procedure is required for it to be produced. We refer to the resulting strain as MGAY-HupA-mCherry. Composite image with both fluorescent channels overlaid on the phase contrast illustrating the tagging system can be seen below in the figure 5.

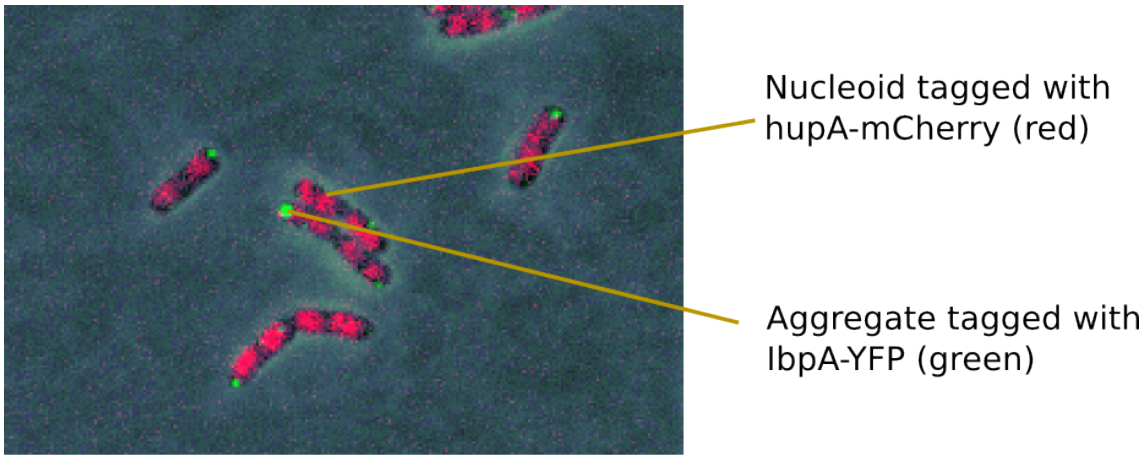

Figure 5: Composite image combining phasecontrast image with both of the confocal channels to illustrate the tagging system. Nucleoids are shown as red and aggregates as green.

Some experiments were also performed using DAPI (4',6-diamidino-2-phenylindol) to stain the nucleoids. DAPI labels the nucleoid by binding strongly to A-T rich regions of the DNA, staining it blue under UV excitation [46]. As using the DAPI dye requires the cells to go through a process where they are fixed using formaldehyde, *in vivo* observation of the nucleoid dynamics is not possible using this technique. Thus DAPI was used mainly for validation purposes, whereas the main results are gained here using HupA-mCherry tagging. DAPI measurements were performed on the original MGAY strain.

3.2 Microscopy measurement system and imaging

Multi-modal time-lapse microscopy was used for the visualization of the cells, as well as the IbpA-YFP labelled aggregates and HupA-mCherry-tagged nucleoids within. Cells were imaged using a Nikon Eclipse (Ti-E) inverted microscope with a 100x objective. Phase-Contrast images were acquired using a CCD camera (DS-Fi2, Nikon). Confocal microscopy was used to detect IbpA-YFP aggregates (exposure using 488 nm laser) and HupA-mCherry tagged nucleoids (exposure using 543 nm HeNe laser). Epifluorescence microscopy using a mercury lamp for UV exposure was used to detect the DAPI-stained nucleoids. Temperature controlled chamber was used to keep the microscopy slides at 37 °C during image acquisition.

When capturing the time series, images from the confocal channels were acquired every minute for one hour to track the dynamics of the aggregates and the nucleoids.

Phase contrast images were captured every 5 minutes for the purpose of cell segmentation.

Note that even though it would be desirable to sample the aggregate locations with a higher sampling rate for more accurate tracking of the movement and characterization of the dynamics, in practice the sampling frequency is strongly limited by a phenomenon called photo-bleaching. This phenomenon occurs due to the fact that each fluorescent protein molecule has only a limited photon budget, after which it loses its fluorescent properties [47]. As higher sampling rates would lead to higher exposure to the laser, the photon budget would deplete more rapidly, bleaching the protein and rendering the tagging non-functional after a shorter time period. In practice, this leads to trade-offs between sampling rate and observation time in fluorescent microscopy experiments. Here we aim to study the long term dynamics, so the once per minute sampling rate was chosen to gain the 1h usable observation period.

Initially, prior to performing the time series measurements, population measurements from a single time point were also conducted, using both the DAPI staining as well as HupA-mCherry-tagged nucleoids. This was done to evaluate the suitability of the used test conditions. In the case of DAPI-staining, images were taken 1h after aggregate induction, whereas in the case HupA-mCherry-tagged nucleoids, images were taken 30 min after the induction.

3.3 Methods to change the nucleoid size

The morphology of the nucleoid is sensitive to availability of nutrients as well as various stresses, such as those caused by antibiotics that inhibit transcription or translation [48-52]. Due to this, the conditions chosen for the experiments conducted here include different types of growth media, as well as addition of Rifampicin or Chloramphenicol. Chloramphenicol halts translation elongation by blocking access of charged tRNA to the ribosomal A site [53], leading to nucleoid compaction [48, 51, 54]. Meanwhile, Rifampicin blocks transcription initiation by binding to RNA Polymerases [52, 55] resulting in nucleoid expansion [50, 52]. These changes can be seen in the figure 6 below. Chloramphenicol and Rifampicin were added after aggregate induction. We observed that, with the used concentrations, both antibiotics reduced the division rate significantly, but did not completely halt it.

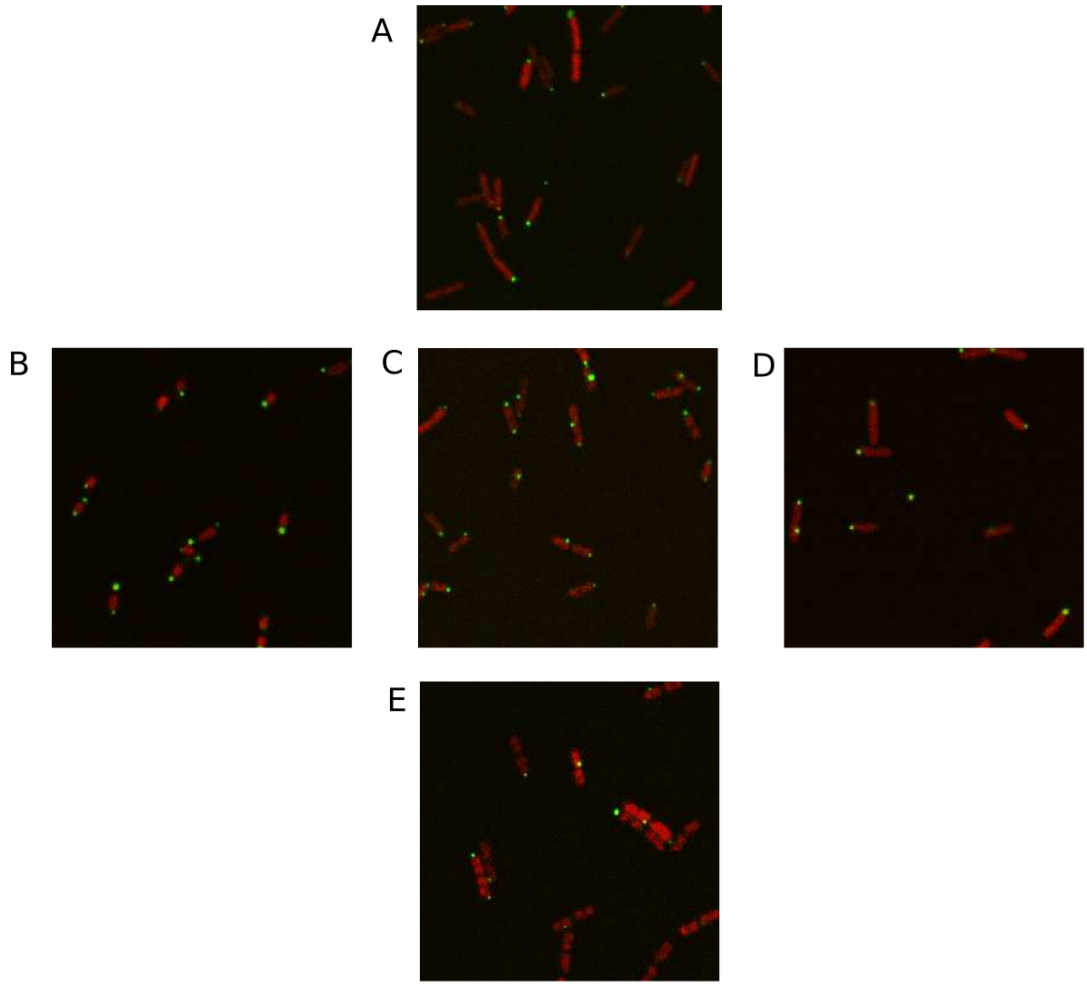

Figure 6: Composite images gained by combining the red and green channels from the confocal microscope to illustrate the changes in nucleoid size for cells grown in each of the main conditions. A) Cells grown in M63 media (increased relative nucleoid size), B) cells grown in LB media with Chloramphenicol added (decreased relative nucleoid size), C) cells grown in LB media with no added antibiotics (control), D) cells grown in LB media with Rifampicin added (increased relative nucleoid size) E) cells grown in TB media (increased relative nucleoid size). HupA-mCherry tagged nucleoids are shown in red and IbpA-YFP tagged aggregates in green.

The types of growth media used here include two types of rich media (Lysogeny Broth and Terrific broth) as well as minimal media (M63 with glycerol as carbon source) [56, 57]. Use of each of the media leads to different growth rates for the cells, which is known to affect the relative size of the nucleoid in *E. coli* [57-60]. For example, in the minimal M63 media, nucleoids are relatively larger than in LB media

[59]. The same occurs in rich TB media, perhaps due to lacking undefined nutrients that are present in LB [60].

3.4 Image analysis

Image analysis tools and statistical methods were an essential part of the project, as without tools capable of providing a relatively high throughput and the capability to meaningfully represent and interpret the large amounts of data, it wouldn't have been possible to uncover relevant information about the underlying biological mechanisms.

The image analysis was mostly done utilizing an existing comprehensive semi-automated microscopy image analysis tool chain developed for Matlab [61]. The tool chain has evolved within the research group through few years of continuous development. However, the existing set of tools had to be modified for use in this project, as previously it had been mainly used, and thus developed, for studying transcription dynamics utilizing only two microscopy channels. Among other things, this meant that a method for segmenting and measuring the nucleoids within the cells had to be found, the limitation of two channels had to be worked around and some modifications had to be done to the way the dimensions of the cell were measured. The analysis steps and used methods are outlined in more detail in the following subsections.

3.4.1 Pre-processing and segmentation

Prior to the analysis of the data, the intensity values of the confocal green channel (used to detect aggregates) were subtracted from the red channel (used to detect nucleoids) in order to reduce crosstalk from the green channel to the red channel. The crosstalk resulted from the fact that the emission spectrum of IbpA-YFP overlaps slightly with the pass band of the microscope filter used for the red channel. We observed by inspection that due to the much smaller size of the aggregates relative to the nucleoids, this subtraction does not significantly affect the measurements of nucleoids from the resulting images. Following this, the images from each channel were run through alignment, which compensates for possible minor misalignments between consecutive frames due to drift of the sample.

Analysis of the microscopy images was then performed using a semi-automated method similar to one described in [61], augmented with an initial segmentation acquired from phase contrast images using the MAMLE based method introduced in [62].

MAMLE stands here for Multi-resolution Analysis and Maximum Likelihood Estimation. The used method performs the segmentation of individual cells from the

image in two main steps. The first one over-segments the cell using multi-resolution edge detection, which incorporates prior information about the morphological properties of the cell. Following this, the over segmentation is corrected by merging together regions to maximize the “cell likeness” objective function for all cells using a maximum likelihood based method.

In total, the whole segmentation process includes seven steps: image de-noising, foreground and background segmentation, multi-scale morphological edge detection, threshold decomposition, learning stage for the shape parameter, binary splitting of the regions and finally, the over-segmentation correction, which results in the final segmentation result from this automated process.

The performance of this method is excellent for images which are densely populated with cells and have a very sharp focus. This is good for experiments where the observation time is long and the cells grow and divide rapidly, forming dense cell clusters. However, for this project the cells in the used images rarely had time to form any significant clusters. This, in combination with the occasional suboptimal focus of the images, led to the fact that in practice significant amounts of manual corrections were required to obtain accurate segmentation results. The result from the segmentation process is a collection of masks, indicating the area occupied by each individual cell at each time point.

Following the segmentation, the resulting masks were manually aligned to the corresponding confocal images based on the information from the red channel.

3.4.2 Detection of nucleoids and aggregates

Next, nucleoids and IbpA aggregates were detected inside the cells by utilizing a method which is essentially a type of outlier detection based on prior knowledge about the shape of the intensity distribution for the background pixels. This method defines the nucleoids and aggregates as connected components with each pixel having fluorescence intensity above a certain threshold. For this, it was assumed that the background pixel intensities follow a Gaussian distribution with the same median and upper quartile as the pixels inside the cell. A threshold was selected so that the probability of mislabelling a pixel from this distribution was less than 0.005 for the aggregates and less than 0.01 for the nucleoids. For nucleoids, a low pass filter was also applied to obtain a smoothly connected nucleoid area. Additionally, limits were separately set for the maximum and minimum area occupied by a nucleoid or aggregate to obtain accurate results. This detection phase results in masks indicating the region occupied by each nucleoid or aggregate within the cell. The results, as well as the segmentation process, are illustrated in the figure 7 below.

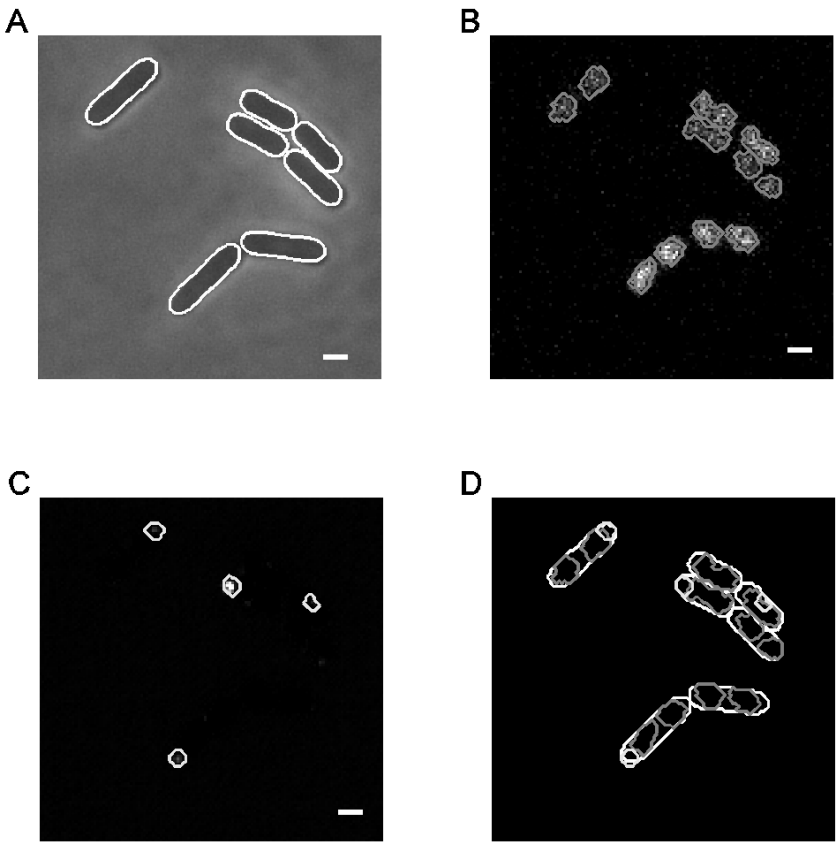

Figure 7: Illustration of the image analysis process. Images of E. coli cells, protein aggregates and nucleoids. (A) phase-contrast microscopy image of E. coli cells with the cell segmentation result overlaid as white curves. (B) confocal image of HupA-mCherry tagged nucleoids of the cells in (A). Results from the segmentation of the nucleoids (grey lines) are also shown. (C) confocal images of IbpA- YFP tagged aggregates of the cells in (A). Detection results for the aggregates (white lines) are also shown. (D) The three images (A-C) are aligned and then their segmentation results are merged. Scale bars are 1 µm. In the confocal images, the contrast was enhanced for easier visualization and, in (B) the effects of crosstalk from the the green channel to the red channel was removed by subtraction.

The detection of nucleoids from the validation images using DAPI-staining was performed with a slightly different method, which doesn't result in individual masked nucleoids, but rather an estimate of the mean intensity along the major axis of the cell in the whole population. First, backgrounds from the epifluorescence images were removed by subtracting a cubic polynomial surface, fitted to the image by L1-norm

minimization as in [26]. Following this, the raw pixel intensity values were extracted from each cell along the major axis of the cell. Position of each pixel was then normalized in relation to the cell size. The derived values were used to calculate an estimate for the the average intensity along the cell major axis within the population, using a non-parametric estimation method known as kernel density estimation, also sometimes referred to as Parzen window method. The kernel function *K* was chosen to be Gaussian and to estimate the distribution of fluorescence intensity along the major axis of the cell, the KDE was weighed with the pixel intensity values and then normalized. This estimate is given by:

$$\hat{f}_h(x) = \frac{\sum_{i=1}^{n} v_i K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^{n} v_i} \quad , \qquad (1)$$

where $\hat{f}_h(x)$ denotes the value of the estimate at location x along the major axis of the cell, *n* denotes the number of samples used to calculate the estimate, v_i is the pixel intensity value for the *i*th sample, x_i is the normalized location of the the *i*th sample on the major axis of the cell, *K* is the chosen kernel function and *h* is a parameter defining the width of the kernel function (here essentially the σ-parameter of a Gaussian distribution), sometimes also referred to as bandwidth or window size.

An estimate of the mean location for the boundary between the nucleoid and the poles was then obtained by maximum likelihood fitting a piecewise constant function to the KDE similarly as in [26].

3.4.3 Measurement of cell dimensions as well as aggregate and nucleoid locations

Once the locations and areas occupied by the cells, nucleoids and aggregates had been detected, we then performed the extraction size data by measuring the length and width for the cells and nucleoids. The locations of the nucleoids and aggregates were then normalized in relation to the rest of the cell to give values that are comparable between the cells.

The cells size was measured by first determining the orientation of the cell by applying principal component analysis to the cell mask, and then measuring the dimensions for the smallest bounding box for the cell mask with the same orientation as the cell. This results in two values, size on major axis and size on minor axis, measured in pixels. Similar measurement was then performed for the mask indicating the area

occupied by the nucleoid, but the resulting major and minor axis sizes were then scaled to give the ratio in relation to those of the cell. Finally, the location within the cell for the center of each aggregate and nucleoid was then normalized similarly to a scale, where -1 and 1 indicate the opposite cell extremities on each axis.

3.4.4 Measurements of aggregate dynamics

Once the normalized size and location measures had been obtained, we analysed the intracellular dynamics of aggregates. The analysis was performed only during the period of time when only one nucleoid and only one aggregate were detected in the cell. For these time moments, displacement vectors along the major cell axis were calculated for the aggregates between each consecutive frame as in [26].

Due to the practical limitations regarding observation time and sampling rate mentioned in section 3.2 leading to the use of once per minute sampling rate, we couldn't use the measured displacement vectors to calculate accurate estimates for the mean squared displacements of the aggregates over time, which could in turn be used to identify diffusion coefficients to more accurately characterize the Brownian motion of the aggregates via the Stokes-Einstein relation as in [25, 44, 63]. However, such measurements have already been performed on IbpA-YFP tagged aggregates [25]. The results from the measurements in [25] indicate that the movement is purely diffusive, with coefficient constants of the order of 500 nm^2/s. The speed of diffusion was also shown to decrease with increasing aggregate size in line with the hypothesis of pure Brownian motion. As the focus here is on the long term dynamics and consequences of interaction with the nucleoids, we instead used the displacements to characterize the anisotropy of the aggregate movement along the major axis as in [26], for which the sampling rate is sufficient.

Additional study of aggregate dynamics as a function of nucleoid size using *in silico* methods was also initially considered within the scope of this thesis. However, as multiple papers exploring nucleoid occlusion and its effect on protein aggregation, segregation and retention using various modelling techniques already exist [11, 25, 26, 43], the scope of this thesis was limited to analysis of the *in vivo* measurements, which has more potential for discovering novel new information.

3.4.5 Cell classification based on nucleoid count

After segmenting the cells, as well as detecting the nucleoids and protein aggregates, it was found that the nucleoid detection process in many cases detected two close by nucleoids as a single connected one. Because of this, the following heuristic classification method was developed to mitigate this problem. It seeks for significant

drops in the fluorescence of the nucleoid within the mid-cell region, which would indicate the presence of two distinct nucleoids. Other methods, such as fitting a two component Gaussian mixture model and testing it against a single component one, were briefly considered, but as the initial tests using the method described here gave sufficiently good results, further testing of other methods was abandoned.

Before performing the classification, all pixel intensity values from the nucleoid channel within the segmented cell area of a single cell were extracted from the image along with their coordinates. The coordinates of each pixel were then transformed to a relative position on the major axis of the cell. Pixel intensity weighed Gaussian kernel density estimate was then calculated for the pixel intensity along the major axis of the cell using the formula 1 (see figure 8 below).

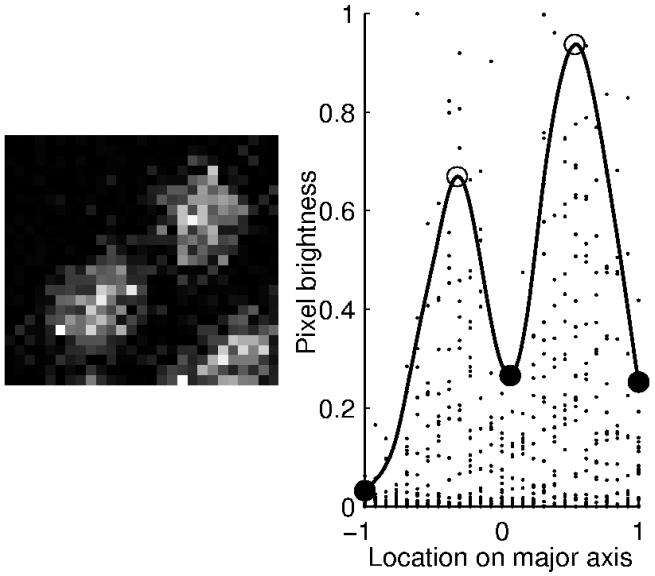

Figure 8: Illustration of the feature extraction for cell classification. The image on the left is the image from the nucleoid channel at the cell location for a double nucleoid cell. The intensity of the pixels in the image has been scaled to maximize the contrast for illustration purposes. On the right we have a scatter plot of the pixel intensity values along with their relative location on the major axis of the cell. The plotted curve indicates the Gaussian KDE for the intensity using σ equal to 0.1. Open circles indicate the local maxima and filled circles the local minima of the KDE used for the classification calculations. Note that the two nucleoids within the cell being classified are the ones completely contained within the image and only pixels within the cell are used for the classification.

The classification was performed using the local maxima and minima values from the obtained KDE. A mean was calculated for all local minima which fall within an experimentally determined mid-cell region (±0.31 cell lengths from mid-cell used here) and all local maxima. Cell was then classified to have two nucleoids if the ratio of mean of minima per mean of maxima was lower than an experimentally determined threshold value of 0.7.

The performance of the classifier was evaluated against a well representative set of 6898 manually classified samples, 65.21% of the samples being single nucleoid. The total classification accuracy improved from 91.66% using just the result from nucleoid detection to 94.30% when using this classifier. More importantly the percentage of double nucleoid cells falsely labelled as single nucleoid decreased from 23.17% to 9.57% of all double nucleoid samples. This came with the trade off of increasing the percentage of single nucleoid samples being falsely labelled as double nucleoid from 0.02% to 3.74%.

4 RESULTS AND DISCUSSION

Here we present the results obtained from the microscopy measurements. In the first subsection we compare nucleoid sizes and spatial distributions between conditions to see their effect, to show that the behaviour of the cells differs significantly between the conditions and to get initial results on the relation between nucleoid size and behaviour of the aggregates. Then in the second subsection, we take a look at the relation between asymmetries in nucleoid positioning and the preferred locations of the aggregates. In the third subsection we move on to studying dynamics of aggregate movement as a function of the nucleoid size, for which we divide the samples based on nucleoid size and transition to studying these fractiles. Finally, in the two last subsections we present some results related to the efficiency of the retention process and discuss how this possibly impacts the lineage as the cells undergo division.

Unless otherwise stated, all of the results presented in this chapter are measured from cells which were determined to have only one nucleoid using the classification technique described in the section 3.4.5.

4.1 Nucleoid sizes and spatial distribution of aggregates in different conditions

First, we wanted to study was nucleoid size and how it related to the spatial distributions of aggregates in each condition. This was done to assess the extent to which the conditions affect the relative size of the nucleoid and to see if the relation between nucleoid size and the positioning of the aggregates was consistent even when different types of stress and environmental conditions were used to induce the changes in the nucleoid.

For this, we took *in vivo* measurements in 9 conditions, which consist of all pairings of each chosen media and antibiotic, as well as each media without antibiotic, using the multi-modal microscopy and fluorescent tagging techniques described in the previous chapter to visualize the cells. From the 9 conditions, the condition “LB with no antibiotics” is considered to be the control condition.

Once the aggregates and nucleoids had been detected from the images, their locations and sizes were normalized in relation to the cell size to give values that are comparable between cells and conditions. For some of the location calculations presented later, such as the values in table 2, we folded the two halves of a cell on top of each other, resulting in a scale, where 0 is the cell center and +1 is either of cell poles. Mean normalized dimensions of the nucleoid along the major and minor cell axes and

mean normalized positioning of the aggregates along the major cell axis are shown in table 2 for each condition.

Table 2: Positioning of aggregates along the major cell axis and relative nucleoid size. For each condition the mean and standard deviation of the relative location of the aggregates along the major cell axis are shown. Also shown are the mean and standard deviation of the relative size of the nucleoid along the major and minor cell axes. The '0' position corresponds to mid-cell and the '1' position to the cell extremities.

		Chloramphenicol	**No antibiotic**	**Rifampicin**
M63	**Aggregate location**	0.83±0.22	0.86±0.23	0.87±0.24
	Nucleoid major	0.71±0.14	0.81±0.13	0.88±0.09
	Nucleoid minor	0.78±0.13	0.79±0.15	0.87±0.12
LB	**Aggregate location**	0.73±0.22	0.79±0.24	0.83±0.22
	Nucleoid major	0.46±0.07	0.60±0.13	0.88±0.07
	Nucleoid minor	0.60±0.13	0.59±0.13	0.89±0.09
TB	**Aggregate location**	0.78±0.22	0.85±0.18	0.84±0.26
	Nucleoid major	0.76±0.17	0.85±0.11	0.92±0.06
	Nucleoid minor	0.80±0.16	0.78±0.11	0.91±0.07

In tables 3 and 4, we show the results of two sample Kolmogorov-Smirnov (KS) tests comparing between pairs of conditions of the distribution of relative nucleoid sizes along the major cell axis and of the distribution of aggregates location along the major cell axis, respectively. In these KS tests, as well as subsequent ones, we consider that for p-values smaller than 0.05, the null hypothesis that the two sets of data are from the same distribution can be rejected. In this case, we only compared conditions that differ from the control in one variable only, that being the media or the addition of an antibiotic.

Table 3: P-values resulting from two sample Kolmogorov-Smirnov tests between the distributions of relative nucleoid sizes along the major cell axis between pairs of conditions. Comparison of distributions of relative nucleoid sizes along the major cell axis between pairs of conditions that differ in one variable from the control (presence or absence of antibiotic, or media richness). For p-values smaller than 0.05, the null hypothesis that the two sets of data are from the same distribution can be rejected. "Chlor." stands for Chloramphenicol, "Rif." stands for Rifampicin, and "no AB" stands for "no antibiotic".

	LB, no AB	**LB, Rif.**	**TB, no AB**	**M63, no AB**
LB, Chlor.	$< 10^{-15}$	0	$< 10^{-15}$	0
LB, no AB	-	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$
LB, Rif.	-	-	$< 10^{-15}$	$< 10^{-15}$
TB, no AB	-	-	-	$< 10^{-6}$

Table 4: P-values resulting from two sample Kolmogorov-Smirnov tests between distributions of aggregate locations along the major cell axis between pairs of conditions. Comparison of distributions of aggregate locations along the major cell axis between pairs of conditions that differ in one variable from the control (presence or absence of antibiotic, or media richness). For p-values smaller than 0.05, the null hypothesis that the two sets of data are from the same distribution can be rejected. "Chlor." stands for Chloramphenicol, "Rif." stands for Rifampicin, and "no AB" stands for "no antibiotic".

	LB, no AB	**LB, Rif.**	**TB, no AB**	**M63, no AB**
LB, Chlor.	$< 10^{-14}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$
LB, no AB	-	$< 10^{-6}$	$< 10^{-15}$	$< 10^{-15}$
LB, Rif.	-	-	$< 10^{-15}$	$< 10^{-4}$
TB, no AB	-	-	-	$< 10^{-15}$

From table 2, along with the p-values in table 3, we conclude that the relative nucleoid size differs significantly between all pairs of conditions differing either in used growth media or application of antibiotic stress. Similarly, from tables 2 and 4, we conclude that the spatial distribution of aggregates along the major cell axis also differs significantly between any pair of conditions.

From table 2, it can be seen that in the conditions where the mean relative nucleoid size is larger, the aggregates preferentially locate closer to the pole, in agreement with the existence of a functional volume-exclusion mechanism in all conditions. One slightly surprising result in table 2 was the increase in the relative size of the nucleoid in the nutrient rich TB media, which was initially expected to cause a reduction in the relative nucleoid size. This is perhaps due to a lack of some nutrients, which are present in LB media. Qualitatively identical results were gained when assessing the nucleoid size by DAPI staining (see table 5 below).

Also note that the changes in size along the minor and major axes with changing conditions are heavily positively correlated (Pearson correlation of 0.93 with a p-value smaller than 10^{-3} in a t-test with the null hypothesis that the data are uncorrelated), indicating that the increase in size in one direction is not made at the cost of size reduction in the other.

Table 5: Relative nucleoid size along the major cell axis as measured by DAPI staining. The size of the major cell axis was normalized to equal 1.

	Nucleoid size along major axis of the cell
M63	0.72
LB	0.69
TB	0.75
LB + Chloramphenicol	0.56
LB + Rifampicin	0.77

Finally, from table 2 it is visible that simultaneous changes in media richness and antibiotic stress have combined effects on nucleoids that are at least as strong as the strongest individual change. Further, as before, the mean positioning of the aggregates changes accordingly, in agreement with the existence of a fully or partially functional nucleoid segregation and retention mechanism.

For further, more in-depth measurements using time-lapse microscopy, we chose to proceed with five conditions, which consist of the control and all the conditions which differ from it in only one variable, i.e. presence of antibiotic or different composition of growth media.

4.2 Asymmetries in nucleoid positioning and correlation with aggregate distribution

As we measured the center locations of the nucleoids within the cells in each condition, we observed that there was some variance associated with the location of the nucleoid center along the major axis of the cell and the degree of this variance was different between conditions. To assess how these changes in the positioning of the nucleoid affect the spatial distribution of aggregates and if this affects the effectiveness of the volume exclusion mechanism, we obtained for each condition the correlation between the positions of aggregates and nucleoids in individual cells, during their lifetime. Results are shown in table 6. Also shown is the p-value of statistical significance of the correlation from a t-test with the null hypothesis that the data is uncorrelated.

Table 6: Pearson correlation between the positions of the nucleoid centers and of individual aggregates along the major cell axis. Also shown are the p-values from a t-test with the null hypothesis that the data is uncorrelated.

		Chloramphenicol	**No antibiotic**	**Rifampicin**
M63	**Correlation**	-0.17	-0.15	-0.23
	p-value	$< 10^{-15}$	$< 10^{-6}$	$< 10^{-15}$
LB	**Correlation**	+0.11	-0.09	-0.19
	p-value	$< 10^{-5}$	0.02	$< 10^{-15}$
TB	**Correlation**	-0.25	-0.37	-0.29
	p-value	$< 10^{-6}$	$< 10^{-15}$	$< 10^{-12}$

From table 6, it can be seen that the positioning of aggregates and nucleoid centers along the major cell axis is negatively correlated in most conditions. Exception from this are the cells in LB media with Chloramphenicol, which have the smallest relative nucleoid of all the conditions. Also, in general, the correlation strength increases with nucleoid size. This negative correlation is a result of the fact that in most cells where the nucleoid is off-centered, i.e. positioned in a biased fashion towards one side of the cell, the aggregates preferentially locate on the opposite side, as expected from volume exclusion, as less of the cell volume is occupied by the nucleoid on this side of the cell.

In addition to this cell wide negative correlation, another interesting phenomenon was found when studying the relation between the nucleoid center and the aggregate positions. This is the fact that for the aggregates that are already located at the pole regions of the cell, the correlation between the nucleoid center location and the position

of the aggregates are positively correlated in all cases. This positive correlation is illustrated in the figure 9.

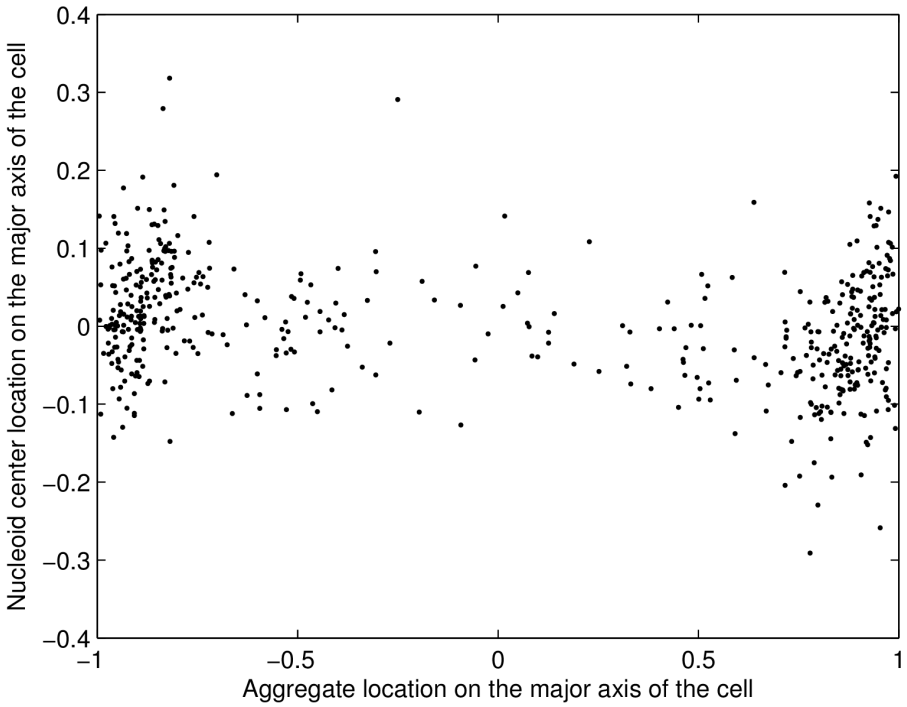

Figure 9: Relation between the aggregates and the center of the nucleoid for the whole population at single time point in LB media with Rifampicin. Clearly visible is the local positive correlation at the pole regions and slight cell wide negative correlation can also be observed. Varying degrees of similar correlations exist for other conditions as well, as can be seen from tables 6 and 7.

The positive correlation for the aggregates at the poles can be easily seen as the slightly tilted nature of the aggregate clusters seen at the poles. This local positive correlation at the poles indicates that the as the nucleoid shifts along the major axis, the aggregates are forced to shift accordingly and is consistent with the nucleoid occlusion effect. The Pearson correlations as well as the p-value of statistical significance of the correlation from a t-test with the null hypothesis that the data is uncorrelated for only the aggregates within the functional pole region are shown in table 7. For the calculations in table 7, the side of the cell containing the aggregate was always regarded as the positive side of the cell and position of the nucleoid was measured relative to this orientation. Also, only the aggregates located within the functional pole region, i.e. the aggregates within the clusters like seen at the poles in figure 9, were included in the calculation.

Table 7: Pearson correlation between the positions of the nucleoid centers and of individual aggregates along the major cell axis for the aggregates within the functional pole region. Also shown are the p-values from a t-test with the null hypothesis that the data is uncorrelated.

		Chloramphenicol	**No antibiotic**	**Rifampicin**
M63	**Correlation**	0.34	0.32	0.20
	p-value	3.4×10^{-62}	2.1×10^{-22}	2.5×10^{-12}
LB	**Correlation**	0.58	0.44	0.31
	p-value	3.7×10^{-138}	1.5×10^{-29}	1.2×10^{-12}
TB	**Correlation**	0.27	0.36	0.17
	p-value	3.0×10^{-7}	3.5×10^{-5}	5.6×10^{-5}

Surprisingly, in LB with Chloramphenicol, the condition where the relative nucleoid size is the smallest, the overall cell-wide correlation becomes positive, while remaining statistically significant, but behaves similarly to other conditions when it comes to aggregates already at the poles. This implies that the nucleoid is no longer large enough to have a significant role in determining on which side the aggregates preferentially locate at. Sill, when shifting along the major axis, it is capable of forcing the aggregates to move accordingly in a positively correlated fashion. The shift to significantly positive overall correlation as the nucleoid size gets sufficiently small is likely the result of weakening in the former phenomenon, while the latter one becomes more pronounced due to the smaller nucleoid size making a larger movement range of the nucleoid possible. This would be consistent with the nucleoid occlusion phenomenon.

4.3 Dynamics of the aggregates as a function of nucleoid size

As we moved on to study the dynamics of the aggregate movement as a function of nucleoid size, it became apparent that we couldn't accomplish this just by studying each of the conditions as a whole, as there is a wide cell to cell variability in some of the cell properties we study, even between cells in the same conditions. This can be clearly seen from the values in table 2 as well as the distribution of nucleoid sizes within the used samples, illustrated in figure 10. Thus, in order to study how the changes in nucleoid size affect the dynamics of the aggregates, we gathered the data from all conditions and then partitioned the cells into quartiles based on their relative nucleoid size (figure 11). Not surprisingly, cells in different conditions contribute differently, in numbers, to the

various quartiles. Table 8 shows the fraction of cells from each condition that belongs to each quartile.

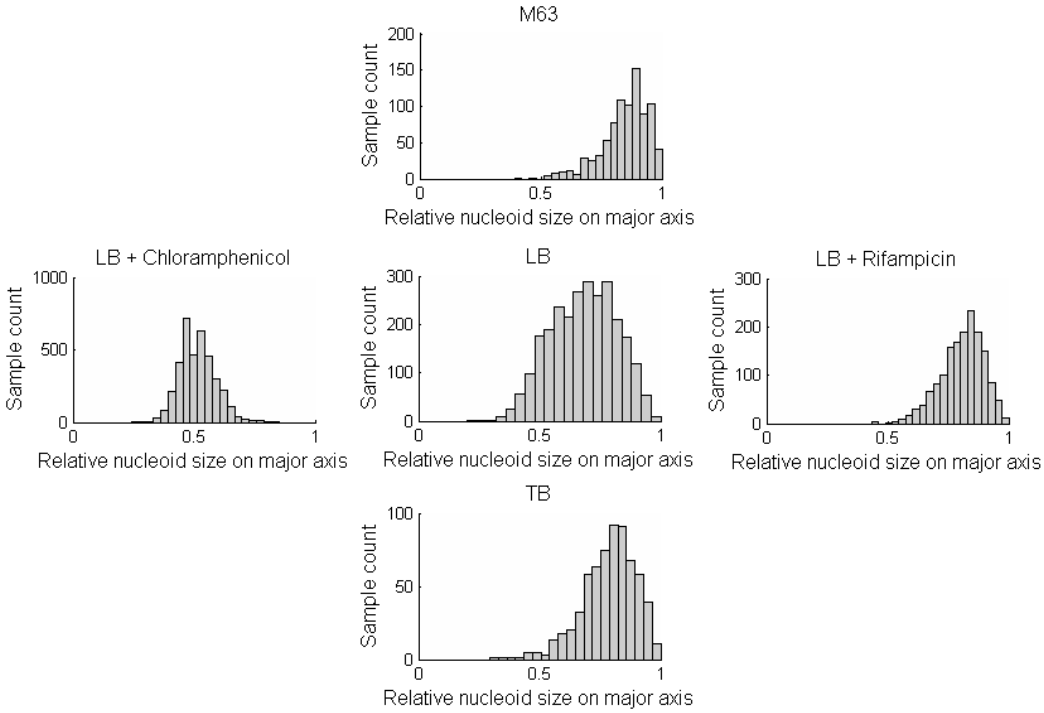

Figure 10: Nucleoid size distributions of the data sets used for the study of spatio-temporal dynamics of the aggregates. X-axis indicates the size of the nucleoid relative to the size of the cell, whereas Y-axis indicates the number of samples. Note the relatively high variance within each condition, especially in the case of LB with no antibiotics (middle).

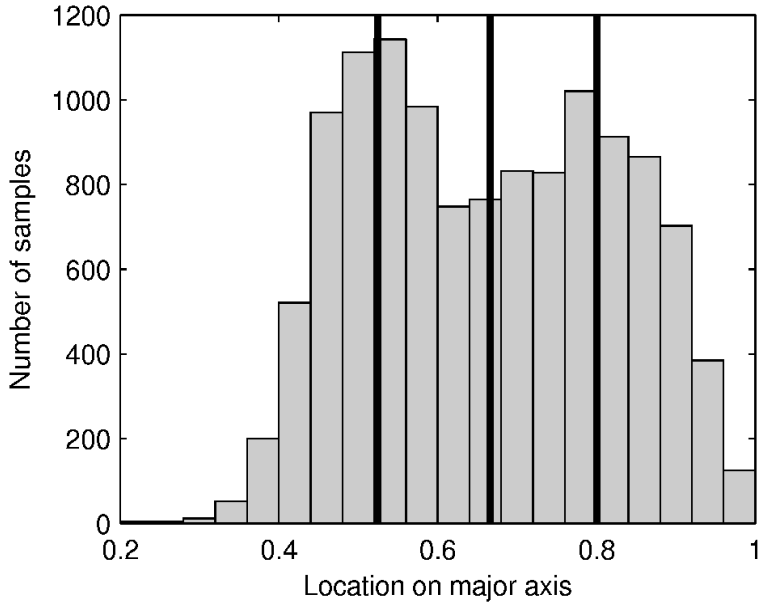

Figure 11: Distribution of relative nucleoid sizes in the dataset gained by combining the data from all five conditions shown in the previous figure. The black vertical lines separate the quartiles of relative nucleoid size.

Table 8: Fraction of cells in each condition belonging to each of the quartiles of relative nucleoid size shown in the figure 11.

	1st Q	**2nd Q**	**3rd Q**	**4th Q**
LB, Chlor.	0.591	0.385	0.023	2.7×10^{-4}
LB, no AB	0.140	0.310	0.326	0.210
LB, Rif.	0.003	0.081	0.321	0.570
TB, no AB	0.023	0.114	0.343	0.488
M63, no AB	0.005	0.054	0.182	0.725

From Table 8, as expected, the fraction of cells belonging to each of the quartiles differs between conditions. Nevertheless, in all conditions there are at least a few cells in each of the quartiles.

Next, we compared the spatial distributions and dynamics of aggregates of cells in different quartiles of nucleoid size. First, we investigated whether the location of the

nucleoid borders along the major cell axis is related to the position along the same axis where the the local maxima of positive anisotropy in the motion of the aggregates occurs. This local maxima is responsible for the accumulation of aggregates at the poles [26]. Anisotropy curves extracted from the motion of aggregates in cells of each quartile are shown in below in figure 11. From these, we extracted the location on the major cell axis where the maximum peak of positive anisotropy occurs. These locations are shown in table 9. Also in table 9 are shown the numbers of cells analysed per quartile, the number of aggregates, the fraction of aggregates at the pole regions, as well as the the mean and standard deviation of the relative nucleoid size along the major and minor axes.

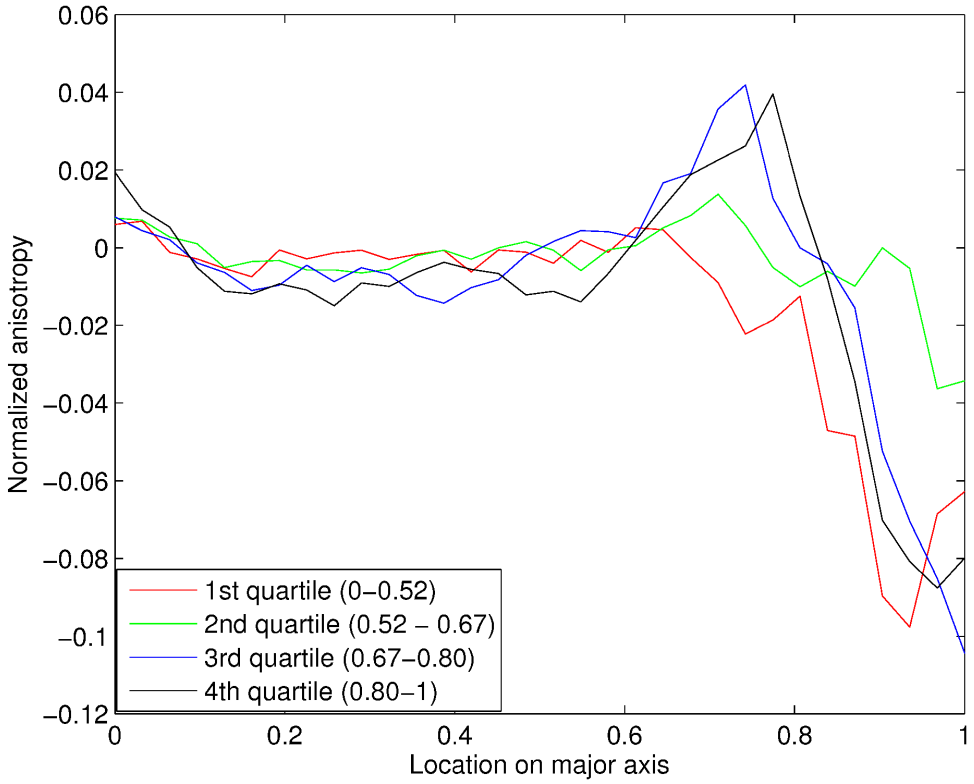

Figure 12: Anisotropy curves. Difference between the numbers of displacement vectors that are directed toward the poles and toward the mid-cell along the major cell axis. The differences were calculated from the displacement vectors originating within a window extending 0.05 normalized cell lengths around that point. The resulting anisotropy curves are scaled with the sample count to give curves that can be easily compared between the quartiles. The analysis was performed solely during the period of time when only one nucleoid and only one aggregate were detected in the cell.

The general shape of the anisotropy curves shown above is similar to that seen previously in [26] and is consistent with what is known about the movement of the aggregates within the cytoplasm in *E. coli* based on previous research [11, 25]. The apparent lack of anisotropy at the mid-cell region supports the hypothesis that movement of the aggregates is at least mainly diffusive, whereas the positive peak coinciding with the nucleoid border provides support for the nucleoid occlusion mechanism. The drop to the negative side close to the extremity of the cell is the result of aggregate interaction with the cell wall.

Table 9: Aggregates positioning and dynamics along the major cell axis in cells of different quartiles of nucleoid size. The table shows the number of aggregates analysed in cells of each quartile, the fraction of aggregates at the pole regions, the mean and standard deviation of the relative nucleoid size along the major and minor axes, and the location of the positive maximum peak of anisotropy along the major cell axis. The number of cells observed for these measurements was 2138.

Quartile	No. aggregates	Fraction aggregates in pole regions	Relative nucleoid size along major axis	Relative nucleoid size along minor axis	Peak location along major axis of anisotropy curve
1st	2404	0.92	0.47±0.04	0.64±0.14	(~0.03)
2nd	2681	0.86	0.59±0.04	0.71±0.13	0.71
3rd	2104	0.70	0.73±0.04	0.82±0.11	0.74
4th	2147	0.40	0.87±0.04	0.88±0.08	0.77

Additionally, from table 9 we can see that the fraction of aggregates in the pole region differs, as expected given the definition of pole and the changes in relative pole size. More importantly, the location of the peak of positive anisotropy follows the same trend as the relative position of the nucleoid borders. This can be seen in the fact that as the latter become relatively closer to the cell extremities, so do the positive peaks of anisotropy.

Nevertheless, it is worth noting that, while for cells of the 2nd quartile the peak of anisotropy is closer to the cell extremity than the nucleoid border, for cells of the 3rd quartile the peak of anisotropy and the nucleoid border match in position, and for cells of the 4th quartile, the peak of anisotropy is closer to cell center than the nucleoid border. We hypothesize that this is due to the increasing space constraints at the poles, which in cells of the 4th quartile may force aggregates to move more often into mid-cell region, while not necessarily entering the nucleoid itself.

Also, interestingly, we were unable to find a clear anisotropy peak in cells of the 1st quartile. We hypothesize that this is because, in these cells, the nucleoids are relatively smaller along both the major and minor axes, which likely reduces their ability of retention of aggregates at the poles by nucleoid occlusion. Further, the fraction of events that correspond to encounters between an aggregate and the nucleoid become rarer. This is due to reduced relative nucleoid size leading to larger pole regions, which in turn leads to dilution of aggregate density close to the nucleoid border.

4.4 Polar retention ability as a function of nucleoid size

Next, we studied the extent to which the ability of retention of aggregates at the cell poles differs between cells of different quartiles. In Figure 13, we show the distribution of locations of aggregates along the major cell axis (from all time moments) in cells of each quartile (top row). Also shown (bottom row) is their location along the major and minor cell axes along with the mean position of the nucleoid border (vertical black line).

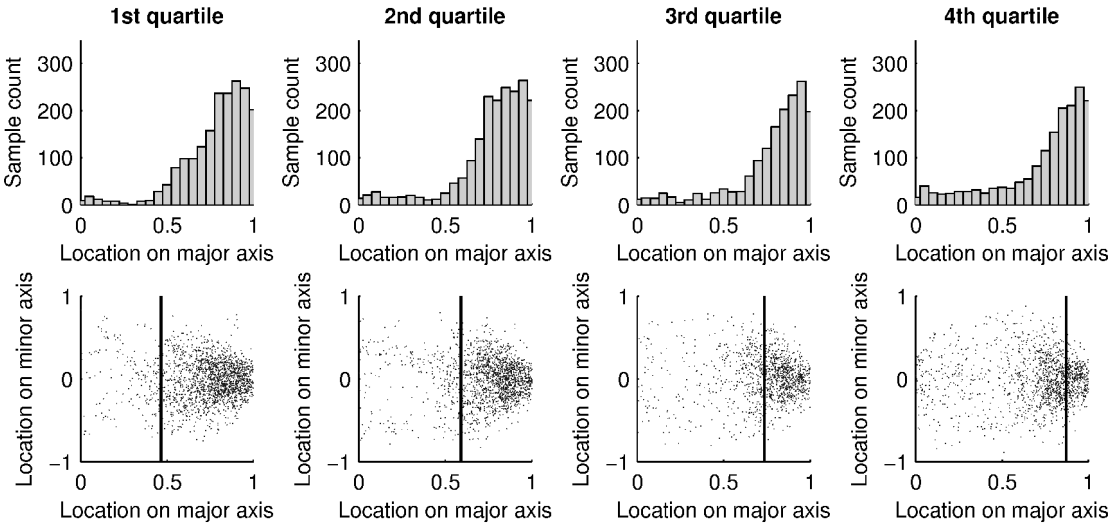

Figure 13: (Top) Distributions of aggregates location along the major cell axis as a function of nucleoid size. (Bottom) Scatter plots of aggregates locations along the major and the minor cell axes, for cells of each quartile. (Top) The figures show the number of aggregates in each position along the major cell axis during cells lifetime, for cells with nucleoids belonging to each of the quartiles. (Bottom) The figures show all positions occupied by aggregates during cells lifetime along with the mean relative position of the nucleoid border in cells of each of the quartiles along the major and minor axes.

Finally, in Table 10, the results of KS tests of comparison between the distributions of aggregates positioning along the major cell axis from cells of different quartiles are shown. The results show that all pairs of distributions differ, in a statistical sense.

Table 10: P-values resulting from two sample Kolmogorov-Smirnov tests between the distributions of aggregates positioning along the major cell axis from cells of different quartiles. For p-values smaller than 0.05 the null hypothesis that the two sets of data are from the same distribution can be rejected.

	2nd Q	**3rd Q**	**4th Q**
1st Q	$< 10^{-4}$	$< 10^{-5}$	$< 10^{-7}$
2nd Q	-	0.002	$< 10^{-6}$
3rd Q	-	-	$< 10^{-4}$

Figure 13 (bottom) suggests that, in cells with the smallest nucleoids (1st quartile), there is little interaction between aggregates and nucleoids, as aggregates rarely locate close to the nucleoid borders or at mid-cell. Meanwhile, in the cells with the largest nucleoids (4th quartile), the aggregates are more often found in the mid-cell region. This is in agreement with the observation in the previous section of a loss of efficiency by cells with larger nucleoids in retaining aggregates outside the mid-cell region. This was observed as the position of the anisotropy peak moving relative to the position of the nucleoid border as the nucleoid size grows. Note also that in histograms shown on the top row in figure 13, the number of samples in the tail of the distribution located at the mid-cell region grows as the size of the nucleoid grows, indicating a larger fraction of aggregates being located a the mid-cell region. Given this and all the results shown above, it would seen that mid-sized nucleoids are optimal for segregating and retaining protein aggregates at the cell poles.

4.5 Effects on the partitioning of aggregates upon cell division

Based on previous research [22, 26, 29, 64, 65] it is expected that aggregates located in between the nucleoids the moment prior to cell division will not, in general, contribute to the emergence of asymmetries in the amount of inherited protein damage between sister cells of future generations (see illustration in figure 14). This in turn would be expected to have implications in the functioning of ageing and rejuvenation processes and thus would affect the fitness of the population. Because of this, studying

the fraction of the aggregates located at the gap between the two nucleoids of a cell moments prior to division would be of great interest.

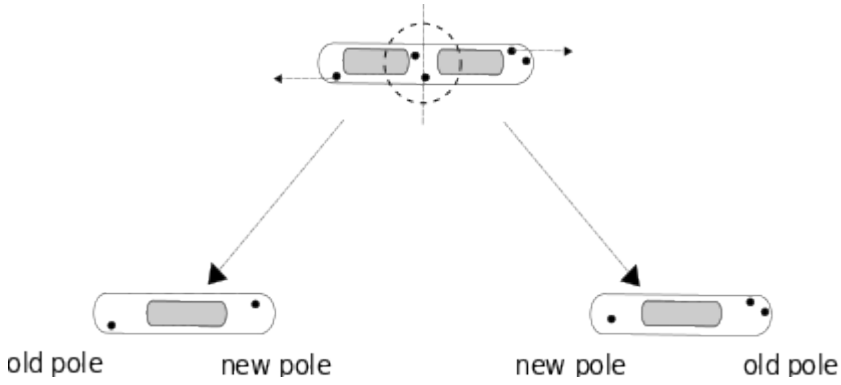

Figure 14: Illustration of partitioning of aggregates located at the mid-cell gap moment prior to division. Unlike the aggregates retained at the pole regions, the aggregates at the mid-cell gap (marked with dasher circle) prior to division do not contribute to an asymmetry in the aggregate count between the new and old poles of the daughter cells, as they are inherited in a random and unbiased fashion. Thus the cell to cell diversity in the number of protein aggregates increases to a lesser extent after the next cell division.

However, the fraction of aggregates located at the mid-cell region prior to division depends on a large number of variables. It depends on the degree of retention functionality, which most likely in turn depends on the size of the nucleoid and the size of the aggregates. Additionally, it also very likely depends on the time that the state with two nucleoids exists, how large the free space between the two nucleoids is, the production rate of the damaged proteins, where within the cell the damaged proteins are produced and their their diffusion rate. The division and growth rates also likely play a role by affecting the time spent with two nucleoids as well as diluting the concentration of damaged proteins in the cytoplasm.

Due to the contribution and complex interaction of all these variables, some of which we were unable to accurately measure using our experimental system, deducing the effect of individual underlying mechanisms responsible for the observed differences between the conditions is extremely difficult, if not impossible given the data available to us. Because of this, we weren't able to gather any conclusive direct evidence about the relation of nucleoid size to the number of aggregates located in the mid-cell gap moments prior to division.

However, given the results presented in the previous sections, we conclude that both increasing, as well as decreasing the relative size of the nucleoid of cells likely leads to an increase in the fraction of aggregates between the nucleoids the moment prior to cell division. This is based on the observed increase in fraction of aggregates seen at mid cell region in single nucleoid cells as the nucleoid size grows, as well as the fact that nucleoid size on both axes are strongly correlated, leading to a significantly diminishing obstruction of aggregate movement as the relative size of the nucleoid becomes small. Nevertheless, it is worth noting that, in no condition is the ability to retain aggregates at the poles completely lost.

While the observation that smaller nucleoids lead to a decrease in the ability to segregate to and then retain aggregates at the cell poles might have been expected, the loss of efficiency in retention of the largest nucleoids was less expected. Within the scope of this project, we couldn't identify the root cause for this phenomenon, but it is most likely closely related to the extreme space constraints at the pole region. One possible explanation would be decreased density of the nucleoid, leading to aggregates being able to enter the space occupied by the nucleoid, while another possibility would be the extreme crowding at the pole forcing some aggregates outside the pole by pushing them towards the mid-cell along the channel between the nucleoid and the cell wall.

5 CONCLUSIONS

In previous research it had been shown that the segregation and retention of protein aggregates in *Escherichia coli* is achieved via the combined effect of nucleoid occlusion, diffusive motion of proteins within the cytoplasm and the coagulation of damaged proteins into larger aggregates upon collision. Here we studied the robustness of this mechanism to different stress conditions and nutrient concentrations, which affect the relative size of the nucleoid. This was done by observing cells containing fluorescent tagged nucleoids and aggregates in nine different conditions using multi-modal microscopy.

To analyse the microscopy data, we used a previously developed tool chain as well as a collection of Matlab scripts written specifically for this project. External components were developed to expand the functionality of the original analysis tool kit, for example to utilize the three microscopy channels required for this study. Also, some parts of the tool chain were re-purposed to achieve the detection of the nucleoids, whereas others, such as the measurement of cell dimensions, were overwritten using project specific external components.

The use of automated image analysis tools and statistical methods was invaluable to us in order to gain the results shown here. Combining these methods with expertise in biology and experimental techniques enabled us to gain valuable information about the underlying biological and biophysical processes. This was due to the fact that automated tools made it possible to analyse the thousands of microscopy images required to gain sufficient statistics with relative ease. Once sufficient amounts of data were gained for statistical analysis and relevant partitioning and filtering techniques were applied, the refined data could then be used to visualize different aspects related to the functioning of the segregation and retention processes to enable easier interpretation. This in turn led to discovering the relevant correlations and derived measures, which inform about the functioning and robustness of the mechanisms, as well as their relation to nucleoid size.

Based on the analysis results, we found that the segregation and retention processes are highly robust to the used perturbations, however not completely immune. Even though the size and positioning of the nucleoid varied significantly in the used conditions, no complete breakdown of the mechanisms could be observed under any of the conditions. However, at the both extremes of nucleoid size, a partial loss of the functionality could be observed.

As the size of the nucleoid on both axes was observed to be highly correlated, an extremely small nucleoid size reduces the retention ability by leaving a larger gap between the nucleoid and the cell wall, increasing the chance of any aggregates

escaping from the pole towards the mid-cell region, passing the nucleoid. Also more free space exists between the nucleoids prior to division as well as at the pole regions throughout the whole cell life time. This leads to overall higher degree of freedom in the movement of the aggregates as well as the nucleoids.

At the other extreme, a nucleoid occupying a very large portion of the cell seems to also cause an increase in the fraction of aggregates seen at the mid-cell region. This might be due to the extreme space limitations at the pole regions, leading to more aggregates being forced into the mid-cell region. Further studies are needed to show whether this is due to perhaps decreased density of the nucleoid, allowing the aggregates to enter the space occupied by the nucleoid, or possibly by the aggregates simply being pushed towards the mid-cell region along the gap between the nucleoid and the cell wall.

The main role of the segregation and retention processes most likely is to maintain the effectiveness of the ageing and rejuvenation through asymmetric inheritance of accumulating protein damage. As this goal is achieved by keeping the mid-cell region free of aggregates, while the cell undergoes division, both of the extremes in nucleoid size would seem to lead to reduction in the efficiency of this rejuvenation process for very distinct reasons.

Overall, our findings suggest that the process of the segregation and retention of aggregates at the cell poles is important for the fitness of the bacterial population, as natural selection appears to have, at least to some extent, optimized the functioning of this mechanism under favourable conditions. However, our results also suggest that the outcomes of this process are likely not essential for short-term survival, since under stress or less-than-optimal conditions, cells neglect it and allow the nucleoid to vary in size.

Our research left some questions to be answered in future research, such as what is the root cause behind the weakening of the mechanism as the nucleoid grows extremely large. This could possibly be investigated if a novel technique for measuring the density of the nucleoid could be found or by using synthetic, fluorescent molecules of various known sizes to study if the molecules become able to enter the nucleoid as the size passes a certain limit. Additional topics of interest left for later studies are the relation of nucleoid size to aggregate accumulation at mid-cell region in cells with two nucleoids as well as the quantification of the degree to which the changes in nucleoid size actually translate to asymmetries in the partitioning of aggregates after multiple divisions and how this affects the vitality of the cells. All these topics require further careful experiment design and possibly the development of some new analysis techniques to better account for the effects of all the involved variables.

Still, our study here has provided us with useful insight into the functioning and robustness of the segregation and retention mechanisms and the gained knowledge and experience with the analysis methods will likely prove useful in future research.

REFERENCES

[1] Muthukrishnan, A-B., Martikainen, A., Neeli-Venkata, R., Ribeiro, A.S., “In Vivo Transcription Kinetics of a Synthetic Gene Uninvolved in Stress-Response Pathways in Stressed Escherichia coli Cells,” *PLoS ONE*, 2014, 9(9): e109005. doi:10.1371/journal.pone.0109005

[2] Gunasekera, T.S., Csonka, L.N., Paliy, O., “Genome-wide transcriptional responses of Escherichia coli K-12 to continuous osmotic and heat stresses,” *J Bacteriol,* 2008, 190: 3712–3720. doi: 10.1128/JB.01990-07

[3] Battesti, A., Majdalani, N., Gottesman, S., “The RpoS-mediated general stress response in Escherichia coli,” *Annu Rev Microbiol*, 2011, 65: 189–213. doi: 10.1146/annurev-micro-090110-102946

[4] Hengge-Aronis, R., “Interplay of global regulators and cell physiology in the general stress response of Escherichia coli,” *Curr Opin Microbiol*, 1999, 2: 148–152.

[5] Hengge-Aronis, R., “Recent insights into the general stress response regulatory network in Escherichia coli,” *J Mol Microbiol Biotechnol*, 2002, 4: 341–346.

[6] Weber, H., Polen, T., Heuveling, J., Wendisch, V.F., Hengge, R., “Genome-Wide Analysis of the General Stress Response Network in Escherichia coli: sS-Dependent Genes, Promoters, and Sigma Factor Selectivity,” *J Bacteriol*, 2005, 187:1591–1603.

[7] Baez, A., Shiloach, J., “Escherichia coli avoids high dissolved oxygen stress by activation of SoxRS and manganese-superoxide dismutase,” *Microb. Cell Fact*, 2013, 12: 23, doi:10.1186/1475-2859-12-23

[8] Kannan, G., Wilks, J.C., Fitzgerald, D.M., Jones, B.D., Bondurant, S.S., et al. “Rapid acid treatment of Escherichia coli: transcriptomic response and recovery,” *BMC Microbiol*, 2008, 8: 37, doi:10.1186/1471-2180-8-37

[9] Bednarska, N.G., Schymkowitz, J., Rousseau, F., Van Eldere, J., “Protein aggregation in bacteria: the thin boundary between functionality and toxicity,” *Microbiology,* 2013, 159: 1795–1806 , doi: 10.1099/mic.0.069575-0

[10] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., *Molecular biology of the cell*, 5th ed., Garland Science, USA, 2002.
ISBN 978-0-8153-4105-5

[11] Winkler, J., Seybert, A., König, L., Pruggnaller, S., Haselmann, U., Sourjik, V., Weiss, M., Frangakis, A.S., Mogk, A., Bukau, B., “Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality and cellular ageing,” *EMBO J.*, 2010, 29: 910-923.

[12] Rujano, M.A., Bosveld, F., Salomons, F.A., Dijk, F., van Waarde, M.A.W.H., et al., “Polarised asymmetric inheritance of accumulated protein damage in higher eukaryotes,” *PLoS Biol*, 2006, 4(12): e417.
doi: 10.1371/journal.pbio.0040417

[13] Nyström, T., “Aging in bacteria,” *Curr Opin Microbiol.*, 2002, 5(6):596-601.

[14] Nyström, T., “A bacterial kind of aging,” *PloS Genetics*, 2007, 3(12): e224, doi: 10.1371/journal.pgen.0030224

[15] Liu, B. et al. “Segregation of protein aggregates involves actin and the polarity machinery,” *Cell*, 2011, 147(5): 959-961,
doi: 10.1016/j.cell.2011.11.018

[16] Nyström, T., “Spatial protein quality control and the evolution of lineage-specific ageing,” *Philos Trans R Soc Lond B Biol Sci.*, 2011, 366(1561):71-5 , doi: 10.1098/rstb.2010.0282

[17] Nystrom, T., Liu, B., "Protein quality control in time and space - links to cellular aging," *FEMS Yeast Res*, 2014, 14(1):40-8 , doi: 10.1111/1567-1364.12095

[18] Nyström, T., Liu, B., “The mystery of aging and rejuvenation – a budding topic,” *Curr Opin Microbiol.*, 2014, 18:61-7, doi: 10.1016/j.mib.2014.02.003.

[19] Lindner, A.B., Madden, R., Demarez, A., Stewart, E.J., Taddei, F., "Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation," *Proc. Natl. Acad. Sci. USA* 105:3076-3081, doi: 10.1073/pnas.0708931105

[20] Dobson, C.M., "The structural basis of protein folding and its links with human disease," *Philos Trans R Soc London B,* 2001, 356:133-145

[21] Harding, J.J., "Viewing molecular mechanisms of ageing through a lens," *Ageing Res Rev*, 2002, 1:465-479.

[22] Stewart, E.J., Madden, R., Paul, G., Taddei, F., "Aging and death in an organism that reproduces by morphologically symmetric division," *PLoS Biol.*, 2005, 3:0296-0300

[23] Sabate, R., de Groot, N.S., Ventura, S., "Protein folding and aggregation in bacteria," *Cell Mol Life Sci.*, 2010, 67(16):2695-715, doi: 10.1007/s00018-010-0344-4

[24] Fisher, J.K., Bourniquel, A., Witz, G., Weiner, B., Prentiss, M., Kleckner, N., "Four-dimensional imaging of E. coli nucleoid organization and dynamics in living cells," *Cell.* 153:882-895, doi: 10.1016/j.cell.2013.04.006

[25] Coquel, A.S., Jacob, J.P., Primet, M., Demarez, A., Dimiccoli, M., Julou, T., Moisan, L., Lindner, A.B., Berry, H., "Localization of protein aggregation in Escherichia coli is governed by diffusion and nucleoid macromolecular crowding effect," *PLoS Comput Biol.*, 2013, 9(4):e1003038, doi: 10.1371/journal.pcbi.1003038

[26] Gupta, A., Lloyd-Price, J., Neeli-Venkata, R., Oliveiro, S.M.D, Ribeiro, A.S, "In vivo kinetics of segregation and polar Retention of MS2-GFP-RNA Complexes in Escherichia coli," *Biophysical Journal*, 2014, 106:1928-1937, doi: 10.1016/j.bpj.2014.03.035

[27] Golding, I., Cox, E.C., "RNA dynamics in live Escherichia coli cells," *Proc Natl Acad Sci U.S.A*, 2014, 101: 11310–11315, doi: 10.1073/pnas.0404443101

[28] Arndt, K.M., Chamberlin, M.J., "Transcription termination in Escherichia coli. Measurement of the rate of enzyme release from Rho-independent terminators," *J Mol Biol*, 1988, 202: 271–285.

[29] Young, R., Bremer, H., “Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate,” *Biochem J*, 1976, 160: 185–194.

[30] Dennis, P.P., Bremer, H., “Differential rate of ribosomal protein synthesis in Escherichia coli *B/r,” J Mol Biol*, 1974, 84: 407–422.

[31] McClure, W.R., ”Mechanism and control of transcription initiation in procaryote,” *Ann Rev Biochem*, 1985, 54: 171-204.

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

[33] Häkkinen, A., Kandhavelu, M., Garasto, S., Ribeiro, A.S., “Estimation of fluorescence-tagged RNA numbers from spot intensities,” *Bioinformatics*, 2014, doi:10.1093/bioinformatics/btt766

[34] Mäkelä, J., Kandhavelu, M., Oliveira, S.M.D., Chandraseelan, J.G., Lloyd-Price, J., et al. “In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter,” *Nucleic Acids Res*, 2013, 41:6544-6552

[35] Farr, S.B., Kogoma, T., “Oxidative stress responses in Escherichia coli and Salmonella typhimurium,” *Microbiol Rev*, 1991, 55: 561–585.

[36] Farr, S.B., Touati, D., Kogoma, T., “Effects of oxygen stress on membrane functions in Escherichia coli: role of HPI catalase,” *J Bacteriol*, 1988, 170: 1837–1842.

[37] Yuk, H., Marshall, D.L., “Adaptation of Escherichia coli O157: H7 to pH Alters Membrane Lipid Composition, Verotoxin Secretion, and Resistance to Simulated Gastric Fluid,” *Appl Environ Microbiol*, 2004, 70: 3500–3505.

[38] Chung, H.J., Bang, W., Drake, M.A. “Stress response of Escherichia coli,” *Compr Rev food Sci food Saf*, 2006, 5: 52–64.

[39] Kandhavelu, M., Mannerström, H., Gupta, A., Häkkinen, A., Lloyd-Price, J., et. al. “In vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps,” *BMC Syst Biol*, 2011, 5:149, doi:10.1186/1752-0509-5-149

[40] Greive, S.J., von Hippel, P.H., “Thinking quantitatively about transcriptional regulation,” *Nat Rev Mol Cell Biol*, 2005, 6: 221–232.

[41] Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., et al. “Sequence-resolved detection of pausing by single RNA polymerase molecules,” *Cell*, 2006, 125: 1083–1094.

[42] Muthukrishnan, A-B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., et al. “Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells,” *Nucleic Acids Res*, 2012, 40: 8472-8483.

[43] Saberi, S., Emberly, E., “Non-equilibrium polar localization of proteins in bacterial cells,” *PLoS ONE*, 2013, 8(5): e64075, doi: 10.1371/journal.pone.0064075

[44] Golding, I., Cox., E. C., “Physical nature of bacterial cytoplasm,” *Phys. Rev. Lett.*, 2006, 96:098102.

[45] Aguilaniu, H., Gustafsson, L., Rigoulet, M., Nyström, T., “Asymmetric inheritance of oxidatively damaged proteins during cytokinesis,” *Science*, 2003, 299(5613):1751-3

[46] Kapunscinski, J., “DAPI: a DNA-specific fluorescent probe,” *Biotech. Histochem.*, 1995, 70:220-233.

[47] Bernas, T., Zarebski, M., Cook, R.R. and Dobrucki. J.W. “Minimizing photobleaching during confocal microscopy of fluorescent probes bound to chromatin: Role of anoxia and photon flux,” *Journal of Microscopy*, 2014, 215: 281-296.

[48] Van Helvoort, J.M., Kool, J., Woldringh, C.L., “Chloramphenicol causes fusion of separated nucleoids in Escherichia coli K-12 cells and filaments,” *J. Bacteriol.*, 1996, 178:4289–93.

[49] Zimmerman, S., Murphy, L., “Release of compact nucleoids with characteristic shapes from Escherichia coli,” *J. Bacteriol.*, 2001, 183.

[50] Cabrera, J.E., Cagliero, C., Quan, S., Squires, C.L., Jin, D.J., “Active transcription of rRNA operons condenses the nucleoid in Escherichia coli: examining the effect of transcription on nucleoid structure in the absence of transertion,” *J. Bacteriol.*, 2009, 191:4180–5.

[51] Chai, Q., Singh, B., Peisker, K., Metzendorf, N., Ge, X., Dasgupta, S., Sanyal, S., “Organization of ribosomes and nucleoids in Escherichia coli cells during growth and in quiescence,” *J. Biol. Chem.*, 2014, 289:11342–52.

[52] Sun, Q., Margolin, W., “Effects of perturbing nucleoid structure on nucleoid occlusion-mediated toporegulation of FtsZ ring assembly,” *J. Bacteriol.*, 2004, 186.

[53] Weber, M.J., DeMoss, J.A., “The inhibition by chloramphenicol of nascent protein formation in E. coli,” *Proc Natl Acad Sci* USA, 1966, 407:1224–1230.

[54] Zimmerman, S.B., “Shape and compaction of Escherichia coli nucleoids,” *J. Struct. Biol.*, 2006, 156:255–61.

[55] McClure, W., Cech, C., “On the mechanism of rifampicin inhibition of RNA synthesis,” *J. Biol. Chem.*, 1978, 8949–8956.

[56] Jin, D.J., Cabrera, J.E.. “Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in Escherichia coli,” *J. Struct. Biol.*, 2006, 156:284–91

[57] Hadizadeh Yazdi, N., Guet, C.C., Johnson, R.C., Marko, J.F., “Variation of the folding and dynamics of the Escherichia coli chromosome with growth conditions,” *Mol. Microbiol.*, 2012, 86:1318–33.

[58] Tao, H., Bausch, C., Richmond, C., “Functional genomics: expression analysis of Escherichia coli growing on minimal and rich media,” *J. Bacteriol.*, 1999, 181:6425–6440.

[59] Jin, D., Cagliero, C., Zhou, Y., “Growth rate regulation in Escherichia coli,” *FEMS Microbiol*., 2012, 36:269–287.

[60] Jin, D., Cagliero, C., Zhou, Y., “Role of RNA polymerase and transcription in the organization of the bacterial nucleoid,” *Chem. Rev.*, 2013, 113:8662–8682.

[61] Häkkinen, A., Muthukrishnan, A.B., Mora, A., Fonseca, J.M., Ribeiro, A.S., “CellAging: A tool to study segregation and partitioning in division in cell lineages of E. coli,” *Bioinformatics*, 2013, 29(13):1708-9. doi: 10.1093/bioinformatics/btt194

[62] Chowdhury, S., Kandhavelu, M., Yli-Harja, O., Ribeiro., A.S., “Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE),” *BMC Bioinformatics*, 2013, 14 (Suppl 10):S8

[63] English, B.P, Sanamrad, A., Tankov, S., Hauryliuk, V., Elf, J., “Tracking of individual freely diffusing fluorescent protein molecules in the bacterial cytoplasm,” 2010, arXiv:1003.2110

[64] Lloyd-Price, J., Häkkinen, A., Kandhavelu, M., Marquesa, I.J., Chowdhurya, S., Lihavainen, E., Yli-Harja, O., Ribeiro, A.S., “Asymmetric disposal of individual protein aggregates in Escherichia coli, one aggregate at a time,” *J. Bacteriol.*, 2012, 194:1747–1752, doi: 10.1128/JB.06500-11

[65] Huh, D., Paulsson, J., “Random partitioning of molecules at cell division,” *Proc. Natl. Acad. Sci. U.S.A.*, 2011, 108:15004–9.