TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Abhishekh Gupta

Effects of Intracellular and Partitioning Asymmetries in *Escherichia coli*

Julkaisu 1335 • Publication 1335

Tampere 2015

Tampereen teknillinen yliopisto. Julkaisu 1335
Tampere University of Technology. Publication 1335

Abhishekh Gupta

Effects of Intracellular and Partitioning Asymmetries in *Escherichia coli*

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB104, at Tampere University of Technology, on the 27th of November 2015, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2015

Supervisor: Assoc. Prof. Andre S. Ribeiro
Tampere University of Technology,
Finland

Pre-examiners: Dr. Cecilia Garmendia Torres
Institut de Génétique et de Biologie Moléculaire et Cellulaire,
France

Dr. Hidde de Jong
INRIA Grenoble – Rhône-Alpes,
France

Opponent: Prof. Nir Gov
The Weizmann Institute of Science,
Israel

ISBN 978-952-15-3601-4 (printed)
ISBN 978-952-15-3657-1 (PDF)
ISSN 1459-2045

Abstract

Cell divisions in *Escherichia coli* are, in general, morphologically symmetric. However, in a few cases, significant asymmetries between sister cells exist. These asymmetries between sister cells result in functional differences between them. For example, cells inheriting the older pole, over generations, accumulate more unwanted protein aggregates than their sister and, consequently, have a reduced growth rate. The reduced ability of these cells to reproduce shows that even these unicellular organisms are susceptible to the effects of aging. To understand senescence in these organisms, it is critical to investigate the sources as well as the functional consequences of asymmetries in division.

In this thesis, we characterize mechanisms responsible for functional and morphological asymmetries in division in *E. coli* cells, using live, single-cell, single-molecule imaging techniques and detailed stochastic models. First, to understand the functional asymmetries due to the heterogeneous spatial distribution of large, inert protein complexes, we study the kinetics of segregation and retention of such complexes by observing these events, one event at a time. For that, we track individual MS2-GFP tagged RNA complexes, as they move in the cell cytoplasm, and characterize the mechanisms responsible for their long-term spatial distribution and resulting partitioning. Next, to understand the morphological asymmetries, we study the difference in cell sizes between sister cells at division under different environmental conditions. Finally, we present the models and simulators developed to characterize and mimic these processes, as well as to explore their functional consequences.

Our results suggest that functional and morphological asymmetries in division, in the growth conditions studied, appear to be mostly driven by the nucleoid. In particular, we find that the fluorescent complexes are retained at the poles due to nucleoid occlusion. Further, the positioning of the point of division is also regulated by the degree of proximity between the two replicated nucleoids in the cell at the moment preceding division. Finally, based on simulation results of the models in extreme conditions, we suggest that asymmetries in these processes in division can enhance the mean vitality of *E. coli* cell populations. Overall, the results suggest that nucleoid occlusion contributes, in different ways, to heterogeneities in *E. coli* cells that ultimately generate phenotypic differences between sister cells.

Preface

This study was carried out at the Laboratory of Biosystem Dynamics (LBD) in the Department of Signal Processing, Faculty of Computing and Electrical Engineering, Tampere University of Technology (TUT) under the supervision of Associate Professor Andre S. Ribeiro.

I would like to express my sincere gratitude to Andre S. Ribeiro for his constant support, guidance and motivation throughout this thesis. He has always been an inspirational mentor. I am also obliged to DTech. Jason Lloyd-Price, who persuaded me to start this doctoral study. Without his consistent instructions and advice from the beginning, the thesis would not have been in the present form. I express my sincere gratitude to Dr. Cecilia Garmendia-Torres and Dr. Hidde de Jong for the pre-examination of this thesis. Their insightful comments and critiques helped to further improve this thesis.

I would like to thank the Tampere Doctoral Programme in Information Science and Engineering (TISE) for funding the first year of my doctoral studies and the Finnish Cultural Foundation (Teresia ja Rafael Lönnströmin rahasto and Aino ja Einari Haakin rahasto) for funding the remaining two years of this research work. I also thank Tampereen kaupungin tiederahasto for funding my dissertation process. Furthermore, I thank the Doctoral Education Network on Intelligent Systems (DENIS) for funding one of my conference trips.

I am grateful to all the members of LBD, which being a multidisciplinary group is an eternal source of ideas from different fields of science. The beneficial discussions with each of the group members were of great help in fulfillment of this work. I thank Antti Häkkinen, Jarno Mäkelä, Eero Lihavainen, Huy Tran, and Teppo Annilla for their valuable comments and ideas on the theoretical aspects of this study. I also thank Dr. Anantha-Barathi Muthukrishnan, Samuel Matos Dias Oliviera, Ramakanth Neelivenkata, Jerome Geoffrey Chandraseelan, and Vinodh Kandavalli for their support on the experimental side of this thesis. I thank Professor Olli-Yli Harja, Assistant Professor Meenakshisundaram Kandhavelu and Dr. Ilya Potapov for their support during this study.

I sincerely thank the secretaries and personnel of the Department of Signal Processing for creating an environment which allows the researchers to focus on their work. In this regard, I would like to specially thank the departmental secretaries Virve Larmila, Johanna Pernu, Noora Rotola-Pukkila, Susanna Anttila and Lab Manager Juha Peltonen for their constant support. I am also thankful to coordinator of international affairs, Ulla Siltolappi, for helping manage my stay at TUT. Finally, I would like to thank the academic coordinator Elina Orava and the faculty secretary Sari Salo for their help in completing all the formalities during this dissertation process.

On a lighter note, I would like to acknowledge the Nepalese community in Tampere which has been very supportive during my stay here. I would like to specially mention the Tampere Nepali Sports Club, which has always been a positive influence. I would like to thank all the cricket and football players as well as the supporters of the club who have shared many joyous moments with me. These moments helped me to remain enthusiastic and focused during my doctoral study. I would also like to thank Dr. Subas Neupane and his family, Tek Prasad Gautam and his family, and Sachin Raj Mishra for their valuable suggestions throughout my time in Tampere.

I am deeply and forever indebted to my parents, Mr. Kritya Nand Gupta and Mrs. Ranjana Gupta, for their love, support and encouragement throughout my entire life. My elder brother, Aalok Kumar Gupta, inspired me and ensured that I would never compromise in what I wanted to achieve. This would be perfect opportunity to thank him for his belief and generosity. Also, I thank my sister-in-law, Suman Gupta, and nephew, Siddhant Gupta, for all their love and support. Additionally, I thank my in-laws, Mrs. Sita Khadka, Mr. Mukunda Khadka and Mrs. Narayani Khadka, for being affectionate and considerate. Finally, I thank all my friends, relatives and well-wishers in Nepal and around the globe for their appreciation and support.

Last but not the least, I am deeply grateful to my wife, Uma Khadka, for being understanding and patient all these years. With all those ups and downs we faced during this study, it would not have been possible without her unconditional love, care and support.

Abhishekh Gupta
Tampere, September 2015

Contents

Abstract **i**

Preface **iii**

Acronyms **vii**

List of Publications **ix**

1 Introduction **1**
- 1.1 Background and Motivation 1
- 1.2 Thesis Objectives 2
- 1.3 Thesis Outline 3

2 Biological Background **5**
- 2.1 Biology of *E. coli* 5
- 2.2 Gene Expression 6
- 2.3 Cytoplasmic Movement 8
- 2.4 Cell Division 10
- 2.5 Partitioning of Cellular Components 12
- 2.6 Cell Aging 13

3 Theoretical Background **17**
- 3.1 Biological Models 17
- 3.2 Chemical Master Equation 17
- 3.3 Stochastic Simulation Algorithm 19
  - 3.3.1 Direct Method 20
  - 3.3.2 First Reaction Method 21
  - 3.3.3 Next Reaction Method 22
- 3.4 Delayed SSA 24
- 3.5 Spatial Inhomogeneity 25
  - 3.5.1 Dynamic Compartments 25
  - 3.5.2 Reaction Diffusion Master Equation 26
- 3.6 Modeling Partitioning in Division 26

4 Methods and Approach **31**
4.1 The Experimental System 31
4.1.1 MS2−GFP Fusion Protein 32
4.1.2 Nucleoid Staining 32
4.2 Image Processing and Computation 34
4.2.1 Cell Segmentation 34
4.2.2 Spot Segmentation 35
4.2.3 Nucleoid Detection 37
4.3 Simulator and Models 38
4.3.1 Compartments 38
4.3.2 Delayed Reactions 40
4.3.3 Division and Partitioning 40
4.3.4 Retention of Complexes 43

5 Summary of Results **45**

6 Discussion **49**

Bibliography **51**

Publications **63**

Acronyms

CME	Chemical Master Equation
DAPI	4',6-diamidino-2-phenylindole
DM	Direct Method
DNA	Deoxyribonucleic acid
FRM	First Reaction Method
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KDE	Kernel Density Estimation
LB	Lysogeny Broth
LDM	Logarithmic Direct Method
NAP	Nucleoid Associated Proteins
NRM	Next Reaction Method
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PDM	Partial-propensity Direct Method
RDME	Reaction-Diffusion Master Equation
RNA	Ribonucleic acid
SNR	Signal-to-Noise Ratio
SSA	Stochastic Simulation Algorithm

List of Publications

This thesis is compound of following publications. In the text, these are referred to as **Publication I**, **Publication II** and so on. The enumeration of the publications is made chronologically, based on their published date.

I J. Lloyd-Price, **A. Gupta**, and A. S. Ribeiro (2012). SGNS2: A Compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations. *Bioinformatics*, **28**(22):3004-5, doi: 10.1093/bioinformatics/bts556.

II **A. Gupta***, J. Lloyd-Price*, R. Neeli-Venkata, S. Oliveira, and A. S. Ribeiro (2014). *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*. *Biophysical Journal*, **106**(9):1928-37, doi: 10.1016/j.bpj.2014.03.035.

III **A. Gupta**, J. Lloyd-Price, S. Oliveira, O. Yli-Harja, A.-B. Muthukrishnan, and A. S. Ribeiro (2014). Robustness of the division symmetry in *Escherichia coli* and functional consequences of symmetry breaking. *Physical Biology*, **11**(6):066005, doi: 10.1088/1478-3975/11/6/066005.

IV **A. Gupta**, J. Lloyd-Price, and A. S. Ribeiro (2014). Modelling Polar Retention of Complexes in *Escherichia coli*. *Lecture Notes in Computer Science*, **8859**:239-43, doi: 10.1007/978-3-319-12982-2_17.

V **A. Gupta**, J. Lloyd-Price, and A. S. Ribeiro (2015). *In silico* analysis of division times of *Escherichia coli* populations as a function of the partitioning scheme of non-functional proteins. *In Silico Biology*, **12 (1,2)**: 9−21, doi: 10.3233/ISB-140462.

In **Publication I**, which is about a stochastic simulator that includes multi-delayed events, dynamic compartments and partitioning in division, the author contributed to codes of molecule partitioning schemes in division. The author, along with J. Lloyd-Price, also conducted the example simulations and analysis of results. The author participated actively in writing of the manuscript. For **Publication II**, the study about the kinetics of segregation and polar retention of large fluorescent complexes in live *Escherichia coli* cells was jointly designed and conceived by the author, by J. Lloyd-Price and by A. S. Ribeiro. Furthermore, the author and J. Lloyd-Price performed the data extraction and analysis of the results. The author had an active participation in the writing of the manuscript. In this study, microscopy measurements were performed by R. Neeli-Venkata and S. Oliveira. **Publication III**, where the morphological symmetry breaking and its functional consequences in *E. coli* is studied, was designed and conceived by the author, by J. Lloyd-Price and by A. S. Ribeiro. The author performed all the data extraction and analysis and participated thoroughly in the writing of the manuscript. In this study, microscopy measurements were performed by S. Oliveira and A.-B. Muthukrishnan.

Publication IV is an study of verifying and validating the most plausible source of polar retention of complexes in *E. coli* using models parameterized by the measurements. The study was jointly designed and conceived by the author and by J. Lloyd-Price. The author performed all the analysis and solely wrote the manuscript. **Publication V** is the study of the functional effects on a cell population due to different partitioning schemes in division. This study was performed using a computational model of a growing *E. coli* cells. The author and A. S. Ribeiro designed and conceived the study. The author built the model, conducted the simulations and performed analysis of the results. The author also participated thoroughly in the writing of the manuscript.

Publication I has been used by J. Lloyd-Price in his PhD dissertation.

1 Introduction

1.1 Background and Motivation

In all living beings, cell division is one of the most important biological events, through which they reproduce. In cell division, a cell performs a sequence of events in which it duplicates its contents and then ultimately divides into two genetically identical daughter cells (Alberts et al. 2002). It is frequently assumed that the cellular components, such as organelles, plasmids, ribosomes and proteins, among others, are equally distributed into the daughter cells (Birky 1983; Birky 1984; Marshall 2007). However, there are several examples of cell divisions where the daughter cells, after a division, differ in protein and DNA content, cell size, inherited organelles, or developmental potential (Chia et al. 2008; Doe 2008; Neumuller and Knoblich 2009; Knoblich 2008). For example, the division in *Saccharomyces cerevisiae* (also known as budding yeast) is highly asymmetric as the daughter cells differ in size and in amount of inherited cellular components (Chant 1999; Erjavec et al. 2008). This asymmetric division, which results in one visibly smaller daughter cell, has been linked to aging in these organisms (Jazwinski 2002; Nyström 2007; Spokoini et al. 2012).

In unicellular organisms, such as *Escherichia coli*, whose division appears to be morphologically symmetric, asymmetries exists between daughter cells (Stewart et al. 2005; Lindner et al. 2008). Recent studies have focused on the sources of such differences as well as on their functional consequences. One such source of asymmetries is the stochasticity in partitioning of the cellular components in cell division (Huh and Paulsson 2011b; Huh and Paulsson 2011a). This randomness is affected by the cellular machinery responsible for segregation of cellular components (Huh and Paulsson 2011b). For example, unwanted protein aggregates segregate to and remain at the cell poles (Winkler et al. 2010). After several generations, large amounts of these aggregates accumulate in the daughter cells with older poles, when compared to the ones with newer poles (Lindner et al. 2008). This accumulation of unwanted constituents at the older pole has been linked with a reduction in vitality of these organisms (Stewart et al. 2005).

Another source of asymmetries between daughter cells is the variability in the positioning of the constriction plane that defines the point of division in *E. coli*

(Lele et al. 2011; Sullivan and Maddock 2000). The constriction plane prior to division is initiated by a septum almost precisely at the midpoint of the longer cell axis in these rod-shaped organisms (Cullum and Vicente 1978; Koch and Schaechter 1962; Sullivan and Maddock 2000). Consequently, in optimal growth conditions, cell divisions are usually morphologically symmetric, with few exceptions (Männik et al. 2012). If the cellular components are evenly distributed on either side of the constriction plane, after a morphologically symmetric division the daughter cells are expected to be phenotypically similar (Lin et al. 1971). As such, any error in locating the constriction plane prior to cell division is bound to create differences between daughter cells.

As the rejuvenation in *E. coli* cells is linked with the asymmetries in division (Stewart et al. 2005; Lindner et al. 2008), a detailed molecular understanding of these sources of asymmetries as well as the mechanisms regulating them is important. Furthermore, mechanisms responsible for aging of *E. coli* cells may inform us on the aging of mitochondria, since mitochondria in eukaryotic cells have a bacterial origin (Alberts et al. 2002). Understanding of these mechanisms is therefore of utmost importance to understand aging in all living beings, as no life strategy is immune to them (Stewart et al. 2005).

With recent advances in fluorescent live cell imaging, improved image analysis techniques, and advances in cell tracking and lineage analysis tools, it is now possible to further investigate the sources of asymmetries in cell division at a single-cell level. Aside from these, realistic models are required to properly assess the effects and long-term consequences of these asymmetries. The models need to mimic the biological processes that involve, in general, a small number of molecules and many events that take place in each cell over generations. Because of the nature of these processes, the models must rely on the stochastic formulation of chemical kinetics. Furthermore, to simulate these models, stochastic simulators are required to sample the trajectories of the components involved in the model.

1.2 Thesis Objectives

This thesis aims to quantify the intracellular asymmetries in division and their functional consequences in *E. coli* cells. For that, we first measure the effects of segregation and partitioning from *in vivo* measurements of live cells with fluorescent MS2-GFP-RNA complexes, single-cell, single-molecule at a time, as in (Golding and Cox 2004). From the analysis of time series images of cells, we study the spatio-temporal distribution of these complexes within the cells. We aim to study if cell divisions introduce asymmetries in their numbers in older and newer cell poles, as in the case of unwanted protein aggregates (Lindner et al. 2008). We investigate and identify the mechanisms responsible for the observed behavior of these complexes using measurements and models. We further aim to assess the robustness of the cellular mechanisms responsible for these processes to sub-optimal temperature conditions.

Next, we study morphological differences between sister cells following a division event in different environmental conditions. In optimal growth conditions, most divisions are known to be symmetric and very few are asymmetric (Männik et al. 2012). We aim to assess the robustness of the symmetry in division to mild chemical stresses and sub-optimal temperatures. We further investigate the functional consequences of any possible morphological asymmetry. Specifically, we investigate if morphological asymmetries in division introduce differences in the inherited components and consequently functional asymmetries between sister cells. Lastly, we aim to find the changes in mechanisms responsible for any observed differences in morphological asymmetries in division between these environmental conditions.

Finally, based on the identified mechanisms, we develop computational models of growing cell populations taking into account the cumulative effects of cell growth (Koch and Schaechter 1962; Koch 1966), asymmetric cell division (Huh and Paulsson 2011a; Männik et al. 2012), and partitioning of unwanted aggregates (Golding and Cox 2004; Lindner and Demarez 2009; Lindner et al. 2008; Llopis et al. 2010; Lloyd-Price et al. 2012b) over cell generations. These models, as much as possible, are parameterized by aforementioned experimental measurements. We use a stochastic model of *E. coli* that includes gene expression, non-functional protein generation, aggregation and polar retention, and stochastic molecule partitioning in division to explore the effects of various partitioning schemes *in silico*. We use the division times of cell populations under different partitioning schemes as a measure of the vitality of the populations. From this, we aim to attain a better understanding of how cells employ asymmetric partitioning schemes in division to cope with aging. The models are implemented in a simulator that we specifically design to simulate these models.

1.3 Thesis Outline

This thesis is organized as follows: First, Chapter 2 briefly introduces the biological background, emphasizing on the required topics to understand this thesis. In particular, the general information about *E. coli* along with the description of some sources of asymmetries in this organism namely, gene expression, cell division, partitioning of cellular components and cellular aging in this organism are presented in this chapter. Next, Chapter 3 introduces the fundamental concepts of stochastic modeling of biochemical processes. For that, the basis of building a stochastic simulator, such as the Chemical Master Equation (CME), the Stochastic Simulation Algorithm (SSA), the Reaction-Diffusion Master Equation (RDME), to construct various models used in this thesis work are described in this chapter. Chapter 4 presents the methods and approaches required to perform data acquisition, and analysis, namely the experimental systems for the measurements, image processing techniques, statistical tools, and methods for modeling of cellular processes, among others. Chapter 5 presents a summary of the results from each publication included in this thesis. Finally, a discussion and future directions of this work are presented in Chapter 6.

2 Biological Background

This chapter is an overview of the biological concepts used in this thesis. It includes biological description of *E. coli* along with its cellular processes such as gene expression, cell division, partitioning of cellular components in division, and aging.

2.1 Biology of *E. coli*

E. coli, a gram-negative rod-shaped bacterium which is usually 2.0 μm long and 0.25-1.0 μm in diameter (Kubitschek 1990), is the most widely studied prokaryote in the field of molecular biology (Alberts et al. 2002). It can live on a wide variety of substrates and is facultative anaerobic. The optimal growth of *E. coli* occurs at 37°C, but there are some laboratory strains that grow in sub-optimal temperature conditions. Although commonly found in the lower intestine of warm blooded organisms, it has the ability to survive outside the body. This bacterium is known to grow easily and inexpensively in a laboratory setting, and has been intensively studied for more than half a decade (Alberts et al. 2002; Goeddel et al. 1979).

Since these model organisms can grow easily and their genetics is relatively easy to manipulate, there has been plenty of studies where genes are introduced using plasmids to allow high level of protein expression (Taniguchi et al. 2010; Goeddel et al. 1979; Gottesman and Maurizi 1992; Kuhlman and Cox 2012; Muthukrishnan et al. 2012; Tao et al. 1999). One of the significant application of *E. coli* cells was to produce human insulin (Goeddel et al. 1979), in which these cells were manipulated using recombinant DNA technology. Furthermore, these cells were used to understand the bacteriophage genetics (Beckett et al. 1991; Arkin et al. 1998; Svenningsen et al. 2005; St-pierre and Endy 2008; Zeng et al. 2010).

More recently, genetically modified *E. coli* cells, such as DH5α-PRO strain (Golding and Cox 2004; Lutz and Bujard 1997), have been used to study the physical properties of the cytoplasm (Golding and Cox 2004; Golding et al. 2005; Golding and Cox 2006), the localization of cellular components such as plasmids and ribosomes (Peabody 1993; Bakshi et al. 2012; Gordon et al. 2004; Kandhavelu et al. 2012a; Parry et al. 2013; Reyes-Lamothe et al. 2013). These organisms have also been used to investigate the aggregation of protein molecules (Lindner et al.

2008; Tyedmers et al. 2010; Winkler et al. 2010; Coelho et al. 2014; Coquel et al. 2013; Maisonneuve et al. 2008), the segregation of unwanted aggregates (Häkkinen et al. 2013; Lloyd-Price et al. 2012a; Lloyd-Price et al. 2012b), and the plane of cell divisions (Woldringh et al. 1994; Woldringh et al. 1991; Donachie and Begg 1996; Meinhardt and Boer 2001; Wang et al. 2005), among others. Because of these, *E. coli* serves as an ideal organism to further investigate the intracellular and partitioning asymmetries, and consequently the cellular senescence.

2.2 Gene Expression

Gene expression is the process by which information encoded in a chromosomal Deoxyribonucleic acid (DNA) sequence is read to synthesize, first, a Ribonucleic acid (RNA) molecule and eventually a protein molecule using the same RNA (Crick 1970; Alberts et al. 2002). This sequence of events is known as the central dogma of molecular biology and is illustrated in Figure 2.1. Stochastic events in gene expression are known to produce fluctuating time patterns of RNA and protein numbers in individual cells (Arkin et al. 1998). These fluctuations are a source of diversity of protein concentrations between cells at any given time, and thus of phenotypic diversity in a population.

Bacteria have been used as model organism to study stochasticity in gene expression. In these organisms, most RNA molecules exist in very small numbers as they vary from only one to a few molecules at any given moment in a cell (Taniguchi et al. 2010). The phenotype of these cells is thus affected by the number of RNA molecules produced by each gene (Choi et al. 2008) as well as the timing at which they are produced. This is because protein numbers follow the RNA numbers (Bernstein et al. 2002; Kæ rn et al. 2005).

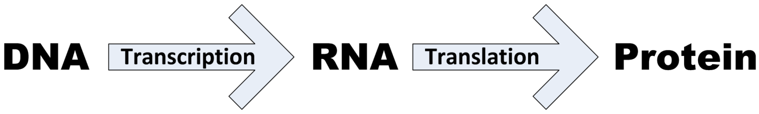


Figure 2.1: An overview of the central dogma of molecular biology. RNA is transcribed from a DNA which then translates to a protein. The arrows show the direction of flow of genetic information as stated in (Crick 1970)

Transcription is the process by which genetic information stored in a DNA strand is copied into a complementary strand of RNA, with the aid of RNA polymerases. One transcription event results in an RNA copy of a gene, which generally codes for a functional protein. Protein coding RNAs are called messenger RNA (mRNA). Transcription begins as the RNA polymerase binds to the promoter region of the gene and then unwinds a small section of the DNA. This is followed by elongation where one strand, known as the coding strand, is used as a template to synthesize an exact RNA copy (Kennell and Riezman 1977).

In prokaryotes, transcription is coupled with translation, which is a process of protein synthesis from a specific sequence of amino acids. This is accomplished by a protein/RNA hybrid known as ribosome. Ribosomes bind to the translation initiation sequence on the mRNA, and elongate the protein in a similar manner to transcription, creating a new protein. In *E. coli*, the events in transcription and in translation are probabilistic in nature (Roussel and Zhu 2006; Kæ rn et al. 2005; Pedraza and Paulsson 2008; Rajala et al. 2010), and their kinetics is sequence dependent (Sorensen et al. 1989; Yarchuk et al. 1992; Herbert et al. 2006; Mcclure 1985). The stochastic events that cause fluctuations in the mRNA count (Golding et al. 2005; Fusco et al. 2003; Paulsson 2005; Yu et al. 2006) mainly occur during:

(a) **Transcription Initiation**: It is a process in which the RNA polymerase binds at a sequence of double stranded DNA called the promoter (as illustrated in Figure 2.2). The level of initiation largely controls the dynamics of gene expression and thereby is one of the key determinants of the phenotypic diversity in monoclonal cell population (Peccoud 1995; Bernstein et al. 2002; Kæ rn et al. 2005). Several studies on the kinetics of transcription initiation have measured the mean durations of steps involved in the initiation process using *in vitro* measurements (Lutz and Bujard 1997; Lutz et al. 2001; Mcclure 1985; Buc and McClure 1985). The results suggest that the two most rate-limiting steps in initiation are: the closed complex formation and the open complex formation (Mcclure 1985; Kandhavelu et al. 2012b). The durations of these steps vary widely between promoters (Herbert et al. 2006), as well as with temperature (Browning et al. 2004; Kandhavelu et al. 2012b) and the concentration of possible activator and repressor molecules (Mcclure 1985).

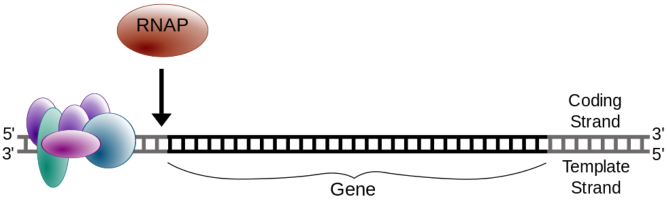


Figure 2.2: Schematic representation of transcription initiation. Reprinted from the work released into the public domain via Wikimedia Commons.

(b) **Elongation**: It is a process in which the RNA polymerase traverses the template strand to create an RNA copy (as illustrated in Figure 2.3). This phase of transcription refers to the process through which nucleotides are added to the growing RNA chain. In bacteria, since transcription and translation are coupled, translation can begin before the completion of transcription, and as a result several translation events can occur in parallel from one transcript. However, since translation cannot complete to produce functional proteins before the completion of transcription, events during transcription elongation fluctuates the protein levels. Studies have shown how the events during transcription elongation can affect translation elongation and thereby produce fluctuations in protein levels (Mäkelä et al. 2011).

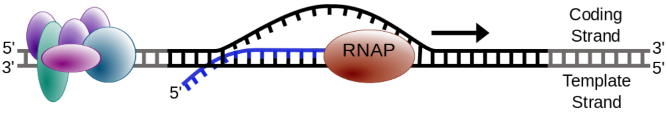


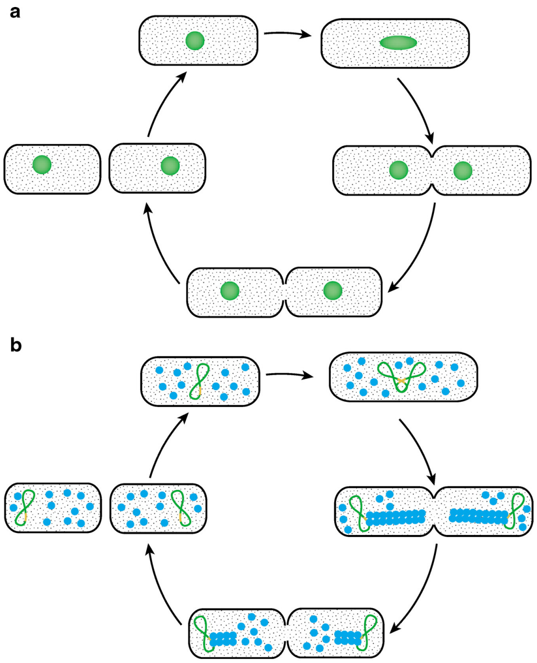
Figure 2.3: Schematic representation of transcription elongation. Reprinted from the work released into the public domain via Wikimedia Commons.

(c) **mRNA degradation**: It is an essential process required to control steady-state concentration of mRNA in all living beings. In bacteria, the half-life of an mRNA typically ranges from 3 to 8 minutes (Bernstein et al. 2002). This process is usually modelled as a first order reaction as the cell-to-cell diversity introduced by such decay process has been studied (Pedraza and Paulsson 2008).

2.3 Cytoplasmic Movement

Cytoplasm is a gel-like substance that fills the cell and is composed mostly of water (Alberts et al. 2002). Cytoplasm, usually clear and colorless, is referred as cytosol, which means substance of the cell. The cytoplasm is also referred to as a molecular soup. In a bacterial cell, all the contents of the cell are contained within the cytoplasm. Due to the absence of cytoskeleton-dependent process of cytoplasmic streaming, in bacteria, diffusion is the most likely means of intracellular movement (Elowitz et al. 1999). The intracellular movement is also affected by the presence of heterogeneous macromolecules in the crowded cytoplasm. Several other studies have investigated the nature of bacterial cytoplasm (Errington 2003; Hippel and Berg 1989; Shimamoto 1999) as well as the movement of macromolecules within it (Golding and Cox 2006; Gordon et al. 2004; Coquel et al. 2013).

Recent studies have focused on the segregation of molecules such as plasmids, ribosomes and protein aggregates in *E. coli* cells (Bakshi et al. 2012; Coquel et al. 2013; Lindner et al. 2008; Gordon et al. 2004; Reyes-Lamothe et al. 2013). One of these studies reported that, high-copy number plasmids such as *ColE1*-type molecules move freely in the cytoplasm (Reyes-Lamothe et al. 2013). Furthermore, these molecules tend to localize at the cell poles due to exclusion from the nucleoid and diffuse occasionally between poles along the major axis of the cell. Similar diffusion-based movement of the low-copy number plasmids to the quarter-cell regions is reported by (Gordon et al. 2004). One interesting study suggests that the segregation of unit-copy or low-copy bacterial plasmids is not only governed by the free diffusion in cytoplasm but is also regulated, spatially and temporally, by their partitioning systems (Ghosh et al. 2006). For example, the plasmids P1 and R1 in *E. coli* have different segregation behavior (as shown in Figure 2.4). In particular, plasmid P1 is known to diffuse freely in the cell and localizes at mid-cell for replication while plasmid R1 is found in the cell poles. R1 also localizes at mid-cell for replication but, afterwards, is pushed back to the poles after replication.



Ghosh SK, et al. 2006.
Annu. Rev. Biochem. 75:211–41

Figure 2.4: Segregation of P1 and R1 plasmids in *E. coli*. (a) The focus of P1 plasmid, which is free to diffuse in a newborn cell, localizes to the mid-cell position just before replication. The foci of duplicated plasmid molecules are extruded in opposite directions. (b) The R1 plasmid (*green*) has a pole-proximal location in the newborn cell. The plasmid translocates to the cell center for replication. ParM (*blue*)-ATP polymerization into filaments at the ParR-bound parC loci (*gold*) transports the replicated plasmid molecules toward opposite cell poles. The filament depolymerizes, starting at the tail end, to generate ParM-ADP. Reprinted from (Ghosh et al. 2006) with permission from the Annual Review of Biochemistry.

Similar studies have been carried out to study the spatial distribution of ribosomes and RNA polymerases (Bakshi et al. 2012; Mondal et al. 2011). It has been suggested that, while there is strong nucleoid-ribosome segregation due to excluded-volume effects, RNA polymerases mostly localize to the nucleoid lobes. It is hypothesized that ribosomes have been found to occupy all the nucleoid free space because of its high-copy numbers, and this localization has been found to change with cell growth (Bakshi et al. 2012). On the other hand, larger non-functional molecules like fluorescently tagged complexes are found to be sub-diffusive (Golding and Cox 2004) and are driven by macromolecular crowding, aggregation and volume exclusion (Coquel et al. 2013). The sub-diffusive motion of macromolecules in the bacterial cytoplasm is suggested to be independent of the main cytoskeletal elements, and is more likely explained by the heterogeneous environment, volume exclusion, as well as the nature and abundance of the molecules.

2.4 Cell Division

E. coli cells, being rod-shaped, grow by elongation under stable growth conditions and has little variation in its width from one generation to the next (Marr et al. 1966; Trueba and Woldringh 1980). At a certain stage of elongation, the constriction plane that defines the point of division (Marr et al. 1966; Errington et al. 1965) is initiated by a septum, almost precisely at the midpoint of the longer cell axis (Cullum and Vicente 1978; Koch and Schaechter 1962; Sullivan and Maddock 2000). This moment of division, that results in two morphologically identical cells each with a copy of chromosome, appears to be strongly correlated with the attainment of a specific cell length during elongation (Donachie et al. 1976; Donachie and Begg 1996; Osella et al. 2014). Cell division in these organisms is considered to be largely deterministic as there is a very small variance in the point of division as well as the moment of division (Marr et al. 1966; Koch 1966).

Cell division in *E. coli* is regulated by the action of at least ten proteins (Joseleau-petit et al. 1999). Recent studies have led to a substantial progress in understanding the assembly of these proteins at the cell septum (Wang et al. 1997; Weiss et al. 1999; Ma et al. 1996). In particular, it has been suggested that the expression of proteins encoded by the filamentous temperature-sensitive genes (*fts*), such as *ftsA, ftsI, ftsK, ftsL, ftsN, ftsQ, ftsW, ftsZ,* and *zipA*, among others, localize to the division plane during septation (Wang et al. 1997; Weiss et al. 1999; Ma et al. 1996; Chen et al. 1999; Ma and Margolin 1999). Among the encoded proteins, *FtsZ* is the one that acts from the start of septation by forming *FtsZ*-ring (Ma et al. 1996). As this *FtsZ* protein is required until the final step of division, it is one of the best characterized and most thoroughly-studied cell division proteins (Joseleau-petit et al. 1999; Yu and Margolin 1999; Ma et al. 1996; Ma and Margolin 1999).

Another important set of proteins used by *E. coli* during division are *MinC, MinD* and *MinE*, and are also known as Min system (Yu and Margolin 1999; Kerr et al. 2006; Huang et al. 2003; Hu et al. 2003). These proteins generate a dynamic oscillation of *ftsZ* protein inhibition for the proper localization of the septum prior to division. In particular, the Min system produces a dynamic distribution of Min proteins whose minimum, at mid-cell, is used as a signal to select the division site (Kerr et al. 2006). On the other hand, it is also suggested that the variability of

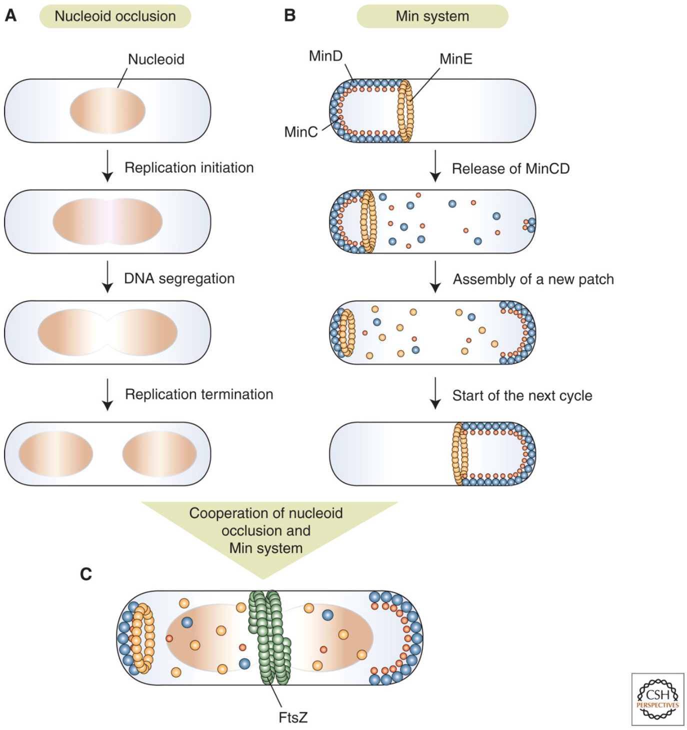

A
Nucleoid occlusion
Nucleoid
Replication initiation
DNA segregation
Replication termination
B
Min system
MinD
MinE
MinC
Release of MinCD
Assembly of a new patch
Start of the next cycle
Cooperation of nucleoid occlusion and Min system
C
FtsZ
CSH PERSPECTIVES

Figure 2.5: Schematic representation model for the positioning of the FtsZ ring by the nucleoid occlusion and Min systems in *E. coli*. (A) Temporal and spatial regulation of cell division by nucleoid occlusion. (B) Inhibition of polar cell-division events by the Min system. (C) Cooperation of the nucleoid occlusion and Min systems. Reprinted from (Thanbichler 2009) with permission from the Cold Spring Harbor Perspectives in Biology.

the location of the minimum in *MinD* concentration is too high to account for the accuracy of the symmetry in division (Kerr et al. 2006). It is therefore reported that, while the Min system is responsible for placing the division point far from the cell poles (Yu and Margolin 1999; Kerr et al. 2006), it is volume exclusion due to the nucleoid that confers the observable degree of precision, i.e. symmetry, to the division process in *E. coli* (Mulder and Woldringh 1989; Woldringh et al. 1990; Woldringh et al. 1991). This co-operation of the Min system and the volume exclusion due to nucleoid to ensure the proper positioning of *FtsZ*-ring, precisely in the mid-cell region, is illustrated in Figure 2.5 as in (Thanbichler 2009).

Similar conclusions are drawn from other observations. It has been shown that irregular nucleoid movements affect the angle and position of the constriction plane and, consequently, the division site (Valkenburg and Woldringh 1984; Woldringh et al. 1994). Furthermore, there is a striking co-localization between the nucleoid-free region at mid-cell and the division point, observed in both normal and aberrant-shaped cells (Männik et al. 2012).

2.5 Partitioning of Cellular Components

An important aspect in cell division is the partitioning of cellular components, such as nucleoids, organelles, and proteins, among others, between the sister cells. In bacteria, the replication of the nucleoid, which envelopes the chromosomal content, ends long before the start of septum formation for cell division (Wang et al. 2005). The replicated nucleoids are then segregated to different centers at the first and the third quarter of the cell (Wang et al. 2005; Mulder and Woldringh 1989). After segregation, the nucleoids in conjunction with the Min system (as shown in Figure 2.5), ensure the proper spatial and temporal regulation of cell division (Yu and Margolin 1999; Mulder and Woldringh 1989; Woldringh et al. 1991; Thanbichler 2009).

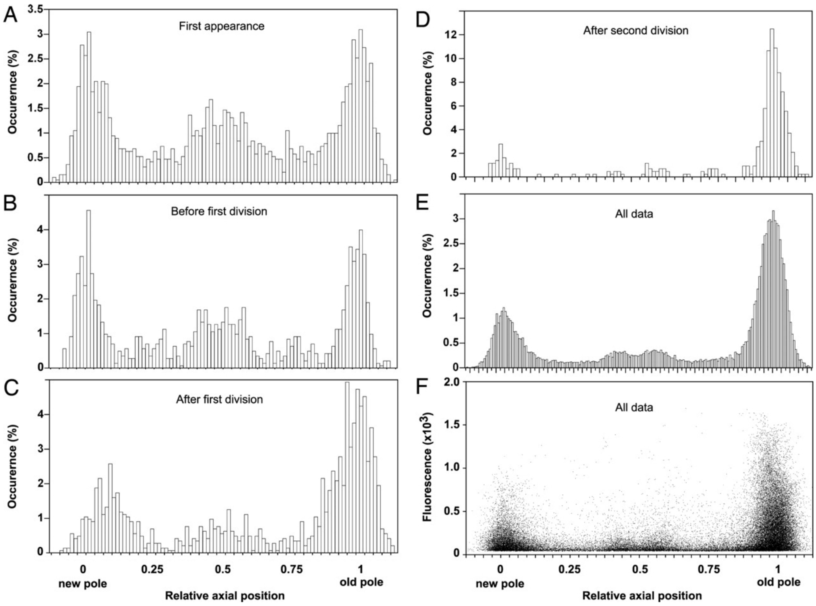


Figure 2.6: Aggregate distribution and associated fluorescence levels along the cell axis. Shown is IbpA-YFP foci localization along the cells' normalized longitude internal coordinate, oriented from the new pole (0) to the old pole (1) of each cell. Binned histograms show foci's localization at first appearance (A), at the last movie frame before first division (B), at the first movie frame after first division (C), at the first movie frame after two consecutive divisions (D), and cumulative over all movies' frames (E). (F) Foci's maximal fluorescence intensity (arbitrary gray-level units) as a function of their localization. Figure from (Lindner et al. 2008), Copyright (2008) National Academy of Sciences, U.S.A.

Equally relevant is the partitioning of other cellular components in division. Since there are low copy numbers of organelles as well as macromolecules in bacteria, stochastic partitioning in division contributes to the non-genetic heterogeneity in a monoclonal population (Huh and Paulsson 2011a; Huh and Paulsson 2011b). For instance, if molecule species are segregated independently and randomly into either daughter cell, upon division this will lead to a distribution similar to that obtained from a Binomial partitioning. Due to this and based on the normalized variance of the number of molecules in the population, the cell-to-cell diversity will double after division. On the other hand, molecules that form dimers before being partitioned to the daughter cells will lead to a distribution with less variance than the Binomial. As a result, this partitioning, also referred to as "ordered" in (Huh and Paulsson 2011a), will introduce less diversity into the population after division. In contrast, the molecules that form clusters, which are then segregated independently, will lead to a distribution with greater variance than independent Binomial. This "disordered" partitioning then leads to considerably greater diversity in the population. It has been suggested that such errors in partitioning are hard to correct (Huh and Paulsson 2011a).

The partitioning of a specific molecule in division is also most likely related to its behavior in the cytoplasm and is reflected on its long-term spatial distribution. For example, if all non-functional proteins are to be enclosed into a single inclusion body, only one of the daughter cells would inherit them in division. Also in case of aggregation or localized accumulation of molecules, these will be asymmetrically partitioned in division among the daughter cells, for example, IbpA aggregates tend to accumulate at the cell poles and, after few cell divisions, are partitioned asymmetrically with daughter cells inheriting the older pole getting more (as shown in Figure 2.6)(Lindner et al. 2008; Winkler et al. 2010). On the other hand, if the molecules do not aggregate or are in high-copy numbers (like ribosomes (Bakshi et al. 2012)), one could expect them to be less retained at the poles and thus, to be randomly partitioned in division. Further, it is also possible that the partitioning scheme will depend on the location of the functional form of the proteins, provided that the loss of function does not affect the long-term spatio-temporal distribution. The process of partitioning in division of all such intracellular molecules is thus likely to be under regulation, either by direct or indirect mechanisms, and to differ between each other.

2.6 Cell Aging

E. coli cells divide symmetrically and, in suitable environments, appear to be functionally immortal as each cell perpetuates itself by dividing into cells with the same genotype as the mother cell (Marr et al. 1966; Männik et al. 2012). These cells, however, are not immune to the accumulation of unwanted substances and degradation of internal components. These organisms need mechanisms to cope with such deleterious accretion. Evidence suggests that the rejuvenation process is achieved by a deliberate asymmetry in the partitioning of intracellular material in division. This was first observed in unicellular organisms that exhibit highly asymmetric divisions (for example, *Saccharomyces cerevisiae* and *Caulobacter* *crescentus*) which are tightly related to the juvenile stage and aging in these organisms (Ackermann et al. 2003; Avraham et al. 2013; Jazwinski 2002). However organisms such as *E. coli* and *Schizosaccharomyces pombe* have seemingly morphologically symmetric division, and thus the knowledge of an aging process is less clear (Barker and Walmsley 1999; Johnson and Mangel 2006; Woldringh 2005).

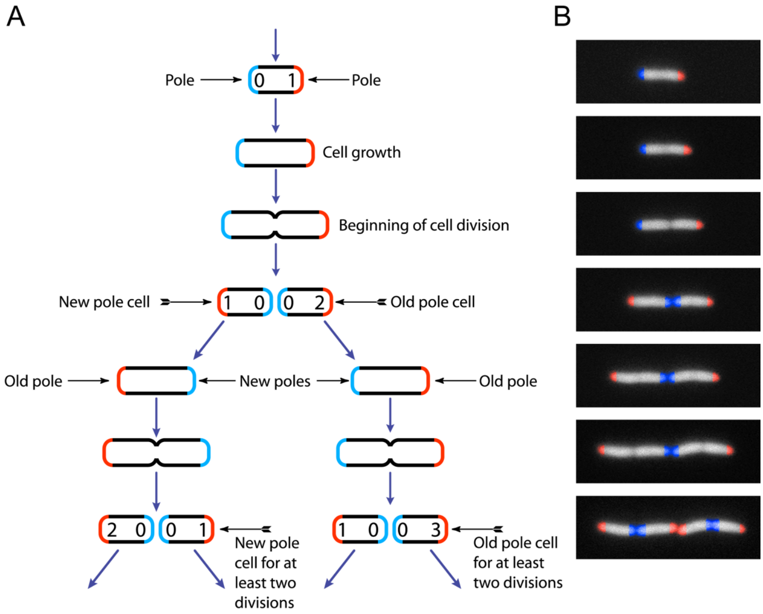

A
B
Pole
0 1
Pole
Cell growth
Beginning of cell division
New pole cell
1 0
0 2
Old pole cell
Old pole
New poles
Old pole
2 0
0 1
New pole cell for at least two divisions
1 0
0 3
Old pole cell for at least two divisions

Figure 2.7: Schematic diagram of the life cycle of *E. coli* as shown in (Stewart et al. 2005). During cell division, two new poles are formed, one in each of the daughter cells (shown in *blue*). The old poles of those cells were formed during a previous division (shown in *red*). (A) The number of divisions since each pole was formed is indicated by the number inside the pole. Using this number, it is possible to assign an age in divisions to that cell, as indicated. (B) Time-lapse images of growing cells corresponding to the stages in (A). False color has been added to identify the poles. Figure used from (Stewart et al. 2005) which permits the usage under Creative Commons Attribution License.

Stewart and colleagues showed that two apparently identical sister cells of *E. coli* exhibit functional asymmetries (Stewart et al. 2005) (shown in Figure 2.7 from (Stewart et al. 2005)). This asymmetry appears to be achieved in division, with one of daughter cell inheriting most unwanted substances (Lindner et al. 2008). It was suggested that this accumulation causes those daughter cells to have a slower rate of division (Stewart et al. 2005). Namely, during their lifetime cells accumulate the substances at the poles. In division, each daughter inherits one pole from the mother cell, and creates a new one. As this continues, older poles accumulate large amounts of 'trash'. Aging in *E. coli* can thus be observed as an accumulation of cell constituents with limited diffusion and long half-lives at the old pole of the mother cell, resulting in larger old poles and cumulatively slower growth of the daughter cells receiving these substances (Stewart et al. 2005; Lindner et al. 2008; Lindner and Demarez 2009).

Recent studies have further investigated different aging-related protein aggregates as well as the mechanisms that trigger aging in these organisms (Ackermann et al. 2007; Coquel et al. 2013; Erjavec et al. 2008; Häkkinen et al. 2013; Maisonneuve et al. 2008; Coelho et al. 2014). The mechanisms have been found to be associated to the asymmetries in segregation of aging-related proteins in *E. coli* which are stochastic (Winkler et al. 2010; Bakshi et al. 2012; Lloyd-Price et al. 2012a; Lloyd-Price et al. 2012b) and operate on a generational time-frame (Stewart et al. 2005; Lindner et al. 2008; Lindner and Demarez 2009).

To better understand the asymmetries that most likely result in aging, in this thesis, the mechanisms of segregation and partitioning of large, inert fluorescent complexes are studied in live *E. coli* cells, from single cell measurements with single-molecule sensitivity. Furthermore, software tools, statistical and analytical tools are developed to study these events. The effect of different environmental factors, such as temperature shifts and chemical stresses, to these intracellular asymmetries are also studied as these are known to enhance aging (Ballesteros et al. 2001; Hsu et al. 2003; Soti and Csermely 2003; Winkler et al. 2010).

3 Theoretical Background

This chapter is an overview of the theoretical concepts of simulation and modeling approaches used in this thesis. It includes the basics about modeling biological systems, description of stochastic simulation methods and concepts to incorporate biological processes such as cell division and partitioning to the models.

3.1 Biological Models

Modeling the kinetics of a process in biophysical system is aimed to improve the understanding of that system. Biological systems differ from physical and chemical systems as biological processes are mostly stochastic in nature and involve low copy numbers of interacting species (Fusco et al. 2003; Gordon et al. 2004; Glick 1995). One such example is the transcriptional regulation of gene expression. Noise due to the fluctuations in number of biomolecules play an important role in the dynamics of such systems (Gillespie 1976; Gillespie 1977; Gillespie 2007; Choi et al. 2008). Because of this, a deterministic approach alone is not sufficient in modeling cellular systems.

Models of the biophysical processes in bacterial cells must therefore account for the stochastic nature of the events occurring in the system. The stochastic formulation of chemical kinetics has been successfully applied to these models (McQuarrie 1967). This formulation describes the time-evolution of a well-stirred set of chemically interacting molecules in thermal equilibrium in a fixed reaction volume. Furthermore, numerous studies have demonstrated that these systems must capture the effect of noise in gene networks as well as the inherent stochasticity in the chemical reactions (Mcadams and Arkin 1997; Arkin et al. 1998; Ozbudak et al. 2002; Lahav et al. 2004; Kæ rn et al. 2005).

3.2 Chemical Master Equation

In the stochastic formulation, the state of a system with N homogeneously spread chemical species at time t is represented by an N-dimensional vector $\mathbf{x}$ containing the molecules of each species in the volume (McQuarrie 1967). The occurrence of one of M chemical reactions that can take place between the molecules results in the change in the population of a species. Consequently, the time-evolution of $\mathbf{x}$ takes the form of a random walk through the N-dimensional space of the populations of the reacting species. This is based on the probabilities of occurrence per unit time of each of these reactions, R_μ, which in turn is based on the current state vector that is defined by the propensity function, a_μ. The a_μ is defined as:

$$\begin{aligned} a_\mu(\mathbf{x})dt = {} & \text{the probability that a particular combination of the molecules} \\ & \text{that are presently in the system will react via reaction } R_\mu \text{ in} \\ & \text{the next infinitesimal time interval } [t, t+dt). \end{aligned} \tag{3.1}$$

Using the laws of probability, this definition of propensity function can be used to derive the master equation for a chemical system. It is therefore considered to be the fundamental premise for the stochastic formulation of chemical kinetics. The form that the propensity function, a_μ, takes depends on the type of the reaction that it represents. It is thus necessary to inspect the physical rationale for unimolecular and bimolecular reactions.

In case of the unimolecular reactions which are known to occur internally within each molecule of a given species S_i, the underlying physics dictates that there is some constant c_μ such that, $c_\mu dt$ gives the probability that some molecule of S_i will spontaneously react via reaction R_μ in the next infinitesimal time dt (Gillespie 2007). It follows from the laws of probability that if, at any given instance, there are X_i molecules of S_i in the system, the probability that one of them will react via R_μ in the next time interval is $X_i c_\mu dt$. Thus, the propensity function for unimolecular reactions is given as:

$$a_\mu(\mathbf{x}) = X_i c_\mu.$$

On the other hand, bimolecular reactions occur when two molecules of different species, say S_i and S_j, meet and react. Using the homogeneity assumption, it can be shown that there exists a constant c_μ such that, $c_\mu dt$ gives the probability that a given pair of molecules of species S_i and S_j will meet and react via a reaction R_μ (Gillespie 1977). This constant, c_μ, can be derived from micro-physical properties (Gillespie 1992). If there are X_i molecules of S_i and X_j molecules of S_j currently in the system, then there are $X_i X_j$ pairs of these molecules. The probability that one of these pairs will meet and react via R_μ in the next infinitesimal time dt is therefore $X_i X_j c_\mu dt$. The propensity function for bimolecular reactions with different chemical species is then given as:

$$a_\mu(\mathbf{x}) = X_i X_j c_\mu.$$

However, the derivation of this function is not completely correct. For bimolecular reactions between two molecules of the same species S_i, the number of pairs does not grow as X_i^2, since a molecule cannot react with itself and the pairs $X_i - X_j$ and $X_j - X_i$ must be counted only once. In such scenarios, the number of pairs grow as $X_i(X_i - 1)/2$, making the propensity function for such a reaction to be:

$$a_\mu(\mathbf{x}) = X_i(X_i - 1)c_\mu/2.$$

In the stochastic formulation, the joint probability distribution $P(\mathbf{x}, t|\mathbf{x}_0, t_0)$ of having a given state vector $\mathbf{x}$ at time t after the initial conditions $\mathbf{x} = \mathbf{x}_0$ at $t = t_0$ is described. Also if it is assumed that v_i represents the absolute number of each reactant that change when reaction R_i occurs, the rate of change of the probability of being in a given state $\mathbf{x}$ can be expressed as the sum of the probabilities of all reactions that can change the system's state into $\mathbf{x}$ in the next infinitesimal time interval, subtracting the sum of the probabilities of all reactions that can cause the system to leave that state. The result is a partial differential equation for P, the Chemical Master Equation (CME) (Gillespie 2007):

$$\frac{\partial P(\mathbf{x}, t|\mathbf{x}_0, t_0)}{\partial t} = \sum_{\mu=1}^{M} \left[a_\mu(\mathbf{x} - v_\mu)P(\mathbf{x} - v_\mu, t|\mathbf{x}_0, t_0) - a_\mu(\mathbf{x})P(\mathbf{x}, t|\mathbf{x}_0, t_0)\right] \tag{3.2}$$

This equation determines the probability that each species will have a specified molecular population at a given time in future. The function that satisfies the CME simultaneously describes the probability of all possible trajectories through the N-dimensional state space of reactant populations. Because of the explicit handling of every possible state that the system can be in, the CME can take an accurate account of the effects of both fluctuations. This has been a major justification for using the stochastic approach over the mathematically simpler deterministic approach. However, solving the CME analytically is usually cumbersome and, at times, impossible. This is because the number of molecular species involved scales the dimension of the system exponentially. A commonly used alternative is simulating the trajectories of $\mathbf{x}$ versus t to sample the distribution of $\mathbf{x}$.

3.3 Stochastic Simulation Algorithm

The approach of simulating the trajectories results in the probability density function of $\mathbf{x}$ instead of random sample of $\mathbf{x}$, as given by the CME. Because of this, $P(\mathbf{x}, t|\mathbf{x}_0, t_0)$ is substituted by a new probability function $p(\tau, \mu|\mathbf{x}, t)$. This function gives us the probability that the next reaction, R_μ, in the system will occur in the infinitesimal time interval $[t + \tau, t + d\tau)$.

Specifically, if the system is currently in state $\mathbf{x}$, this function $(p(\tau, \mu|\mathbf{x}, t))$ is the joint probability density function of the two random variables: the time until the next reaction occurs (τ) and the index of this reaction (μ). By applying the laws of probability, an exact formula for $p(\tau, \mu|\mathbf{x}, t)$ is derived from the fundamental premise equation (shown in 3.1)(Gillespie 1977). It is given as:

$$p(\tau, \mu|\mathbf{x}, t) = a_\mu(\mathbf{x}) exp(-a_0(\mathbf{x}))\tau \tag{3.3}$$

where,

$$a_0(\mathbf{x}) = \sum_{\mu=1}^{M} a_\mu(\mathbf{x}) \tag{3.4}$$

These equations(3.3 and 3.4) are the mathematical basis for the stochastic simulation approach. One implementation of this approach is Stochastic Simulation Algorithm (SSA), a Monte Carlo procedure for numerically generating time trajectories of the molecular populations in exact accordance with the CME. The procedure that SSA follows, as proposed by Gillespie (Gillespie 1976), are listed in Algorithm 1.

Algorithm 1 : Procedure of Stochastic Simulation Algorithm

```
1: Set time $t = 0$. Set up the initial state vector $\mathbf{x} = \mathbf{x}_0$.
2: With the system in state $\mathbf{x}$ at time $t$, evaluate all the $a_\mu(\mathbf{x})$ and their sum $a_0(\mathbf{x})$.
3: Using a suitable sampling procedure, generate a random pair $(\tau, \mu)$ according to the joint probability distribution defined above by $p(\tau, \mu|\mathbf{x}, t)$.
4: Output the system state for each of the sampling points in the time interval $[t, t + \tau)$.
5: If $t + \tau \geq t_{stop}$, terminate.
6: Set $t = t + \tau$, and $\mathbf{x} = \mathbf{x} + v_\mu$.
7: Recalculate $a_i$ for all $i$ such that any $X_\mu$ that was changed in step 4 appears as an updated reactant in $\mathrm{R}_i$.
8: Go to step 3.
```

For Step 3, there are several exact procedures to generate the samples of τ and μ based on the joint probability distribution $p(\tau, \mu|\mathbf{x}, t)$. The two original, and statistically equivalent, sampling procedures of the SSA are the Direct Method (DM) and the First Reaction Method (FRM) (Gillespie 1976). Based on these, there are other sampling approaches such as the Next Reaction Method (NRM), the Logarithmic Direct Method (LDM),and the Partial-propensity Direct Method (PDM), among others.

3.3.1 Direct Method

The Direct Method (DM) applies the standard inversion generating method of Monte Carlo theory (Gillespie 1976). In this sampling approach two random numbers, namely r_1 and r_2, are drawn from the uniform distribution in the unit interval, and the random pair (τ, μ) is computed as:

$$\tau = (1/a_0(\mathbf{x}))ln(1/r_1) \tag{3.5}$$

$$\mu = \text{the smallest integer satisfying} \sum_{\mu'=1}^{\mu} a_{\mu'}(\mathbf{x}) > r_2 a_0(\mathbf{x}) \tag{3.6}$$

The formulae for this method is derived using the fact that, a two-variable probability density function can be written as the product of two one-variable probability density functions. Specifically, if $P_1(\tau)d\tau$ is assumed to be the probability that the next reaction will occur between $t + \tau$ and $t + \tau + d\tau$ irrespective of the reaction; and $P_2(\mu|\tau)$ is assumed to be the probability that the next reaction will be an R_μ, given the next reaction occurs in $[t + \tau, t + \tau + d\tau)$, the joint probability density function can be written as:

$$P(\tau|\mu) = P_1(\tau) \cdot P_2(\mu|\tau) \tag{3.7}$$

This can be written as:

$$P_2(\mu|\tau) = \frac{P(\tau|\mu)}{\sum_{v=1}^{M} P(\tau|v)} \tag{3.8}$$

also because,

$$P_1(\tau) = \textstyle\sum_{v=1}^{M} P(\tau|v)$$

From these equations and using the laws of probability, as in (Gillespie 1976), it has been shown to yield:

$$P_1(\tau) = a_\mu exp(-a_0\tau) \tag{3.9}$$

$$P_2(\mu|\tau) = a_\mu/a_0 \tag{3.10}$$

The main idea of the DM is to first generate τ according to $P_1(\tau)$ and then generate μ using the $P_2(\mu|\tau)$. The resulting random pair (τ, μ) will be distributed according to $P(\tau|\mu)$. The DM, given a fast, reliable uniform random number generator, and is easily programmable. This method is therefore referred to as a simple, rigorous procedure for implementing Step 3 of the SSA (Gillespie 1976).

3.3.2 First Reaction Method

The First Reaction Method (FRM) is another sampling approach described in the original formulation of the SSA (Gillespie 1976). As this method is not as efficient as the DM, it is mostly used to provide the insights to the stochastic simulation approach. The core idea of this method is based on the generation of "tentative reaction times", τ_v using a random number (r_1) from a uniform distribution in the unit interval, according to the probability density function in equation 3.3. The generation of reaction times is repeated for all the M reactions and is given as:

$$\tau_v = (1/a_\mu(\mathbf{x}))ln(1/r_1) \tag{3.11}$$

From the resulting M tentative reaction times, the pair (τ, μ) is then chosen such that τ is the smallest τ_v and μ is the corresponding value of v for that particular τ_v . It is proven that the probability density function for this random pair (τ, μ) is statistically equivalent to the $P(\tau|\mu)$ prescribed by the CME (Gillespie 1976). Although the FRM is as rigorous and exact as the DM, in its original form it is generally slower to compute since it requires M separate random numbers from the uniform random number generator for each of the M reactions. However, since its original publication (Gillespie 1976), this method has been further optimized for computational efficiency without affecting the statistics.

3.3.3 Next Reaction Method

The Next Reaction Method (NRM) (Gibson and Bruck 2000) is one of the approaches that reduce the computational costs of the FRM significantly. The FRM needs to perform $O(M)$ operations per iteration of the SSA, since it takes a time proportional to M to both update a_is, as well as to generate and find the smallest tentative reaction time (τ_v). The NRM instead uses a special data structure, an Indexed Priority Queue, to store the tentative reaction times (τ_v) generated in previous iterations and to extract them whenever required. This results in a significant improvement in the runtime performance when compared to the FRM.

The NRM uses a directed graph G(V,E) to represent the set of all reactions $\{\mathrm{R}_v\}$ and their relationship on how they affect the molecules which their propensity functions depend on (Gibson and Bruck 2000). In this “dependency graph”, the Vertex set V $= \{\mathrm{R}_v\}$ and there is a directed edge from v_i to v_j when reaction R_i changes the number of a molecule, which affects a_j. In other words, the dependency graph is a data structure that informs on the a_is that need to be changed after a reaction occurs. The use of the dependency graph limits the number calculations of a_is to a minimum. Also for each time step, as the number of edges from a given vertex is typically small, only a few propensities are to be updated. The nodes to be updated in the dependency graph are changed in the place where they are stored, which results in the bubbling up or down the tree structure. This bubbling continues until the property of priority queue is satisfied again. This approach takes $O(logr)$ time, where r is number of reactions in priority queue. But, if there are a small number of reactions, r', that have rate constants faster than the others, then most updates will occur in those reactions and take approximately $O(logr')$ time. This algorithm does not need to continue further once it reaches a node that is already in the desired location.

This procedure requires a data structure that can quickly find the smallest τ_i and can efficiently update the sub-set of the changed τ_i's. To serve this purpose, the

NRM uses an indexed priority queue. It can be implemented efficiently with a tree of ordered pairs of the form (i, τ_i), where i is the number of a reaction and τ_i is the putative time when reaction i occurs; and an index structure whose ith element is a pointer to the position in the tree that contains (i, τ_i). Also, the hierarchy of this tree maintains the heap property, in that the parents always have a lower value of τ_i than either of their children. Consequently, identifying the reaction with smallest τ_μ can then be done by examining only the root node, which makes the selection of a reaction achievable in constant time. The steps of NRM, as described in (Gibson and Bruck 2000), are shown in Algorithm 2.

Algorithm 2 : Next Reaction Method

```
1: Initialize:
     (a) Set initial numbers of molecules, set $t = 0$, generate a dependency graph G;
     (b) Calculate the propensity function, $a_i$, for all $i$;
     (c) For each $i$, generate a putative time, $\tau_i$, according to an exponential distribution with parameter $a_i$;
     (d) Store the $\tau_i$ values in an indexed priority queue P.
2: Let $\mu$ be the reaction whose putative time, $\tau_\mu$, stored in P, is minimum.
3: Let $\tau$ be $\tau_\mu$.
4: Change the number of molecules to reflect the execution of reaction $\mu$. Also, set $t \leftarrow \tau$.
5: For each edge $(\mu, \alpha)$ in the dependency graph G,
     (a) Update $a_\alpha$;
     (b) If $\alpha \neq \mu$, $\tau_a \leftarrow (a_{\alpha,old}/a_{\alpha,new})(\tau_a - t) + t$;
     (c) If $\alpha = \mu$, generate a random number, $\rho$, according to an exponential distribution with parameter $a_\mu$, and set $\tau_a \leftarrow \rho + t$;
     (d) Replace the old $\tau_a$ value in P with the new value.
6: Go to Step 2.
```

This algorithm is exact as well as efficient. The run time of this algorithm is $O(log(M))$ which is, in case of large reaction networks and loosely coupled reaction channels, significantly faster than both the FRM and the DM. Furthermore, as it generates one random number per iteration, the total number of random numbers generated is the sum of the number of reactions with the number of simulation events. This is half the number required by the DM. Also, the NRM requires $O(log(M))$ time to find the index μ of the next reaction when compared to $O(M)$

required by the DM. However, for small systems, this advantage may not be significant as the computational cost of maintaining the data structures in the NRM dominates the simulation time. There are several other efforts to improve and optimize the DM for large systems, as well as to decrease the computational as well as data structure complexity of NRM (Slepoy et al. 2008; Cao et al. 2004; McCollum et al. 2006).

3.4 Delayed SSA

The SSA also needs to take into account the complex processes which take non-negligible amount of time to complete once started. One example of these processes can be multi-stepped processes composed of many simple reactions, such as stepwise elongation. For example, the elongation of an RNA molecule by an RNA polymerase can take as long as a few minutes in *E. coli*, which is in order of magnitude comparable to that of the cell's lifetime. A variant of the SSA was proposed in (Gibson and Bruck 2000) to simulate such processes by representing them as single-step delayed reactions. In these reactions, some of the products are released some time later than the depletion of the substrates, and not necessarily all at the same time. However, as the system's evolution in time is no longer a purely Markov process, it cannot be simulated with any of the methods described above.

Nevertheless, these events need to be considered as they affect many cellular processes, such as the dynamics of the gene regulatory network (Ribeiro 2010). The construction of explicit model for these events may not be feasible as the exact nature of these processes may not be known. On the other hand, allowing reactions that produce their products at an arbitrary time later in the simulation has two main advantages. First, delays of arbitrary distributions can be inserted, even if the underlying process is not completely known. Next, it potentially removes many reactions from the system without affecting its dynamics, and hence speeding up the simulations considerably (Gibson and Bruck 2000).

Algorithm 3 : Delayed SSA

```
if $\tau < t_{min}$, where $t_{min}$ is the earliest entry in the wait list then
    Perform the normal SSA Execution step
else
    Set $t = t_{min}$, and $X_i = X_i + n$.
    Remove the earliest entry from the wait list.
end if
```

The reactions that can have delayed products with different delays are implemented using "wait list". The products of these reactions are placed on the wait list as a tuple (t_r, i, n) where t_r is the time at which the n molecules of the product S_i needs to be released. In the NRM, the wait list can be implemented by adding the products on the list as nodes of the indexed priority queue. This results in a logarithmic time per addition and removal of the products. Alternatively, a heap-based priority queue with the same runtime bounds can be implemented to run alongside the DM implementation. The steps in this variant of the SSA execution, as in (Roussel and Zhu 2006), are shown in Algorithm 3.

3.5 Spatial Inhomogeneity

The simulation approaches, discussed in previous sections, derives the time-evolution of a spatially homogeneous mixture of chemicals in a single reaction volume. Specifically, these approaches consider the reactions in the biological cells as primarily non-spatial and treat each cell as a well-mixed volume. However, there are complex spatial environments within a cell, which contains different organelles, irregular membrane structures, macromolecules, and many other geometrically complex structures. These are systems that have dynamically-relevant inhomogeneity. A stochastic formulation that takes account of these spatial effects is required to build a realistic model of a cell.

3.5.1 Dynamic Compartments

A simple approach, commonly used, is to divide the space into discrete regions called compartments. Each of these compartments is assumed to be spatially homogeneous. To model biological processes, the cell can be visualized as a hierarchy of compartments, each of which enclosed by a membrane. Each of such compartments may contain elementary molecules as well as other compartments. Mitochondria, the Golgi complex, and other organelles are examples of such structures within a cell. In case of tissues or organs, the cells that they are made up of can be considered as example of such structures. In this approach, all the processes within a compartment such as a chemical reaction within a compartment, transport of molecules outside of or into a compartment, and creation and dissolution of compartments are viewed as sequences of discrete events.

The P system formalism is another approach to simulate biochemical processes by making use of dynamically changing; nested compartments (Paun 2001). The presence of the nested compartments in P systems, however, violates the assumption of homogeneous distribution of molecules of the SSA (Spicher et al. 2008). It can be easily modified by maintaining the assumption of spatial homogeneity within the compartments. After this, any variant of the SSA can be used to simulate the dynamics in each compartment (Spicher et al. 2008). Events such as transport of molecules (from one compartment to other), creation and deletion of compartments can be then treated as reactions of the specific compartment. Furthermore, the multiple concurrent SSA simulations can then be integrated by using an NRM-like algorithm to determine the compartment where the next reaction takes place, effectively turning it into a “Next Compartment Method”.

3.5.2 Reaction Diffusion Master Equation

While studies have modeled the effects of these well-mixed structures on the dynamics of chemical reactions within biological cells (Lloyd-Price et al. 2014; Spicher et al. 2008; Mäkelä et al. 2011), several others have investigated the effect of the spatially distributed nature of the cell on biochemical signaling using reaction-diffusion models (Kerr et al. 2006; Isaacson 2009; Zon and Wolde 2005). Although there have been deterministic reaction-diffusion models for the biochemical systems with high concentrations of reactant species, there are not many standard models for systems with noise in the chemical reactions.

Recently, some studies have suggested mathematical models to represent the stochastic reaction-diffusion systems (Zon and Wolde 2005; Andrews and Bray 2004). In these studies, molecules are modeled as points undergoing spatially continuous Brownian motion. The bimolecular reactions occur between the molecules as they cross a specified reaction radius. Although both (Zon and Wolde 2005) and (Andrews and Bray 2004) differ in their simulation algorithm, the idea of discretization of time while the approximations being spatially continuous is the same.

There are other studies that discretized space and approximated the diffusion of the molecules as a continuous-time random walk on the lattice (Elf and Ehrenberg 2004; Isaacson and Peskin 2006). In these studies, the reactions occur with a fixed probability per unit time for molecules within the same lattice site. This is based on the idea of Reaction-Diffusion Master Equation (RDME) in (Gardiner et al. 1976), where the space is divided into a collection of equally-sized mesh cells. An RDME is then derived to give the probability of the system being in a given state. Exact realizations of the RDME can be created using the Gillespie method (Gillespie 1977).

3.6 Modeling Partitioning in Division

Cellular processes such as gene expression, gene regulation, protein aggregation, and motion of macromolecules, among others, can be modeled by creating chemical reactions using the ideas discussed above. These stochastic chemical reactions are known to create fluctuations during the cell cycle and are sources of cell-to-cell diversity (Kæ rn et al. 2005; Choi et al. 2008; Mäkelä et al. 2013). Along with these, stochastic partitioning of components in cell division creates further fluctuations, consequently, enhancing the cell-to-cell variability (Lloyd-Price et al. 2012b; Lloyd-Price et al. 2014; Huh and Paulsson 2011a). The errors in partitioning contribute equally to the overall heterogeneity of cells as these are suggested to be the origin of much of the noise attributed to, for example gene expression (Huh and Paulsson

2011a). This is illustrated in Figure 3.1 from (Huh and Paulsson 2011a), wherein a distinct drop in copy number of segregating units is visible at each division event. To model these effects of partitioning errors, it is important to theoretically understand the mechanisms of partitioning of cellular components in division.

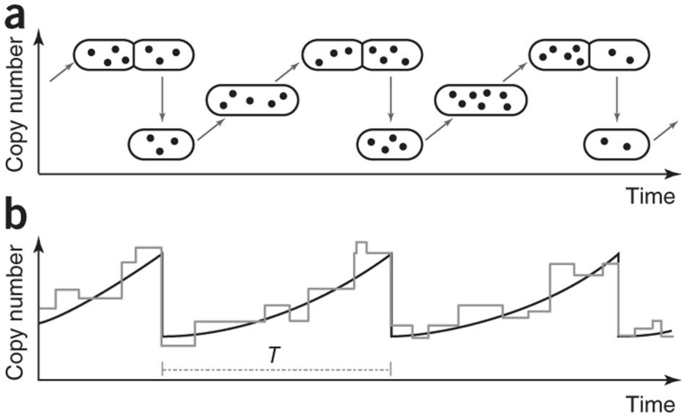


Figure 3.1: (a) Cartoon of an individual cell line and segregating units (*dots*) followed through rounds of growth and division. (b) Sample-time trace of copy number per cell (*gray*) and their average (*black*). Random changes are due to births and deaths during the cell cycle and segregation at cell division (time T). Reprinted from (Huh and Paulsson 2011a) with permission from Nature Publishing Group.

The mathematical understanding of the basic guiding principles of partitioning and their effect on cell heterogeneity was recently studied in (Huh and Paulsson 2011b). This study investigates heterogeneities introduced by three different segregation mechanisms which results in differing variance in the molecule numbers introduced in cell division. More specifically, a simple independent segregation mechanism where each segregating component has a constant and independent probability of being inherited by either daughter cell is compared, to disordered segregation where variation in the partitioning machinery further randomizes levels of segregating components between daughters, or to ordered segregation where components directly or indirectly interact with each other to create a more even distribution of inherited components between daughters.

The partitioning error in division of parent cells with x components and each daughter inheriting L and R copies, respectively, after division has been defined, in (Huh and Paulsson 2011b), to be:

$$Q_x^2 = \frac{\langle (L-R)^2 \rangle}{\langle x^2 \rangle} = CV_L^2 - CV_x^2 \tag{3.12}$$

where the brackets $\langle .. \rangle$ denote averages over all dividing cells in the population while CV_x^2 and CV_L^2 are the squared coefficients of variation (CV^2, defined as the variance over the squared mean) of the number of molecules in parent cells, and in daughter cells immediately after division, respectively. This equation captures the statistical differences between the daughter cells as well as the contributions to population heterogeneity introduced by cell division.

In the simplest case, if molecules are partitioned independently and randomly, i.e. each molecule is inherited by either daughter with equal probability, the resulting distribution of number of molecules inherited by a daughter will be binomial. The Equation 3.12, regardless of the fluctuations in x across cells, decreases with the number of molecules (Huh and Paulsson 2011b), and thus is written as:

$$Q_x^2 = 1/\langle x \rangle \tag{3.13}$$

The independent segregation, however, does not seem to be the likely mechanism for most of the cellular components as it has been hypothesized that various mechanisms, such as active transport mechanisms (Rokney et al. 2009), macromolecular crowding in bacterial cells (Golding and Cox 2006; Coquel et al. 2013), regulate the segregation of cellular components.

A more realistic disordered segregation has been suggested to increase partitioning errors (Huh and Paulsson 2011b). For example, if there is a difference in cytoplasmic volume available for the daughters because of the error in locating the division plane or due to occupancy by large macromolecules, the individual molecules to be partitioned become statistically dependent. Therefore, even-though these molecules do not interact, the partitioning error is increased and, as in (Huh and Paulsson 2011b), is given as:

$$Q_x^2 = \frac{1 - Q_{vol}^2}{\langle x \rangle} + Q_{vol}^2 (CV_x^2 + 1) \tag{3.14}$$

where Q_{vol} is the partitioning error of relative available cytoplasmic volume. This shows that the variation in available volume has an impact on the partitioning error. This should be common for cells that divide into daughter cells of different sizes (Trueba and Woldringh 1980; Shehata and Marr 1975; Trueba et al. 1982; Schaechter and Kjeldgaard 1958).

On the other hand, in case of the ordered segregation such as “pair formation”, the partitioning error has been shown to decrease (Huh and Paulsson 2011b). For example, the partitioning error is reduced if it is assumed that a fraction k of x segregating molecules forms pair and are partitioned to different daughters with a probability p, to same daughter with probability *(1-p)*, and the unpaired molecules *(1-k)* are partitioned independently. The partitioning error for all such mechanisms, in (Huh and Paulsson 2011b), is shown to follow:

$$Q_x^2 = \frac{1 - (2p - 1)k}{\langle x \rangle} \tag{3.15}$$

Similar partitioning errors have been computed for other examples of disordered as well as ordered segregation mechanisms (Huh and Paulsson 2011b). Although such errors for some examples can be computed directly from assumptions, it might not be the case for more complex partitioning mechanisms. For those, Markov processes (which are merely mock processes) to capture the segregation statistics of each partitioning mechanisms has also been designed (Huh and Paulsson 2011b). Thus, the probabilistic rules of other partitioning examples explained are applicable to model partitioning of different cellular components in cell division (Huh and Paulsson 2011b).

In this thesis, aside from the measurements, we need realistic models to properly assess the effects of segregation and partitioning at the molecular level. We, therefore, develop computational models of cell populations that account for cell growth, asymmetric divisions, and partitioning of unwanted aggregates over cell generations. To simulate these models, we design a stochastic simulator based on the the stochastic formulation of chemical kinetics. Using the results from this specifically tailored simulator, we numerically assess long-term consequences of these cellular processes that involve a number of events which take place in each cell over generations.

4 Methods and Approach

This chapter is an overview of the methods employed in this thesis. These methods include the description of the experimental system, image processing techniques, and data extraction approaches, among others used for the studies in **Publication II** and **Publication III**. In addition to that, the simulation and modeling approaches used for **Publication I**, **Publication IV** and **Publication V** are presented in this chapter.

4.1 The Experimental System

The experiments, required for **Publication II** and **Publication III**, are conducted in DH5α-PRO strain of *E. coli*, generously provided by I. Golding (Baylor College of Medicine, U.S.A.). This strain contains two constructs: (i) PROTET-K133 carrying P$_{LtetO-1}$$-MS2d-GFP$, and (ii) a pIG-BAC (P$_{lac/ara-1}$$-mRFP1-MS2-96bs$) vector, carrying a 96 MS2 binding site array under the control of P$_{lac/ara-1}$ (Golding and Cox 2004).

This DH5α-PRO strain has been chosen because of its relatively slow division rate when compared to a wild-type strain of *E. coli* (Jung et al. 2010). This slower growth rate facilitates the automated tracking of individual fluorescent complex and of the precise moment of cell divisions from time-lapse microscopy images. Moreover, as multiple MS2-GFPs are known to bind a specific RNA target to form MS2-GFP-RNA complexes (Golding and Cox 2006; Golding and Cox 2004), these complexes can be easily detected, quantified and tracked as they move in the cytoplasm (Häkkinen et al. 2014). Also, once the RNA target is bound to MS2-GFP, it is immortalized (Golding and Cox 2004), possibly due to reduced or complete prevention of degradation. Because of these, it is possible to track the long-term spatial distributions and partitioning in division of these complexes, across several generations. Finally, it is also relevant that, since the cells are flooded with MS2-GFPs, the extraction of exact shape and size of the cells from confocal microscopy images becomes easier.

Bacterial cell cultures are grown in Lysogeny Broth (LB) media for all experiments. For **Publication II**, cells are grown at optimal (37°C) and sub-optimal (24°C) temperatures. On the other hand, for **Publication III**, cells are grown under two mild stresses, namely acidic shift and oxidative stress, as well as at five different temperatures (42°C, 37°C, 33°C, 30°C, and 24°C).

4.1.1 MS2–GFP Fusion Protein

The dimeric MS2 fused to GFP (MS2-GFP fusion protein) used as a detection tag (Golding and Cox 2004) is expressed from a medium-copy vector under the control of the $P_{LtetO-1}$ promoter (Lutz and Bujard 1997), regulated by the tetracycline repressor. The RNA target for MS2-GFP is located on a single-copy F-based plasmid, and is controlled by the $P_{lac/ara-1}$ promoter (Lutz and Bujard 1997), regulated by Isopropyl β-D-1-thiogalactopyranoside (IPTG) and Arabinose.

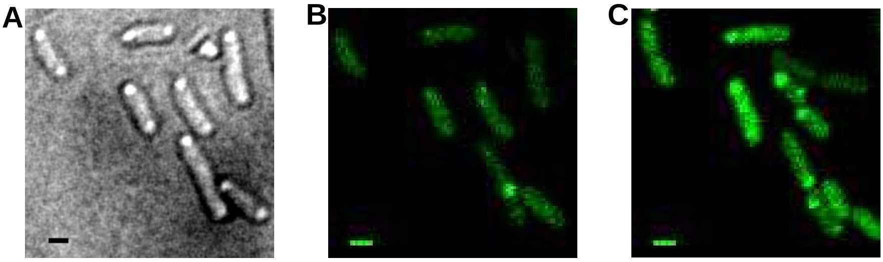


Figure 4.1: (A) Example bright-field image of *E. coli* cells expressing MS2-GFP and target RNA, from a time-series at t = 1 *min*. (B) Fluorescence channel of the image shown in (A). (C) Fluorescence image from the same time-series at t = 85 *min*, showing the outcome of a few cell divisions and partitioning of complexes. Scale bars are 1 μm. Contrast is enhanced for easier visualization. Image from **Publication III**.

The induction of the target RNA production is completed by adding 1 *mM* IPTG and cells are incubated for 5 minutes prior to preparation of the microscope slide. This induction procedure is found to be necessary for the accumulation of sufficient numbers of MS2-GFP to detect the target RNA and for the full induction of the target gene (Golding et al. 2005; Mäkelä et al. 2013). Usually, 2 to 4 tagged RNAs are produced each hour by a fully induced cell (Kandhavelu et al. 2011). Furthermore, the MS2-GFP molecules distribute themselves homogeneously throughout the cytoplasm, and fluorescent spots appear only in the presence of target RNAs (Golding and Cox 2004; Lloyd-Price et al. 2012a). For all experimental conditions, cells are visualized using time-lapsed images captured every minute for 2 hours using fluorescence microscopy (as illustrated in Figure 4.1 B and C). The green fluorescence was measured using a 488 *nm* laser and a 515/30 *nm* detection filter.

4.1.2 Nucleoid Staining

In **Publication II**, to understand the role of nucleoid in spatial distribution of the MS2-GFP-RNA complexes, the nucleoids are stained using 4’,6-diamidino-

2-phenylindole (DAPI) to obtain their sizes. DAPI is known to stain nucleoids specifically, with very little or no cytoplasmic labeling (Kapuscinski 1995). For live cell nucleoid staining, DAPI (2 $\mu g/ml$) is added to the cells suspended in Phosphate Buffered Saline (PBS) and incubated for 20 minutes in the dark. Cells are then washed twice with PBS, and placed on a 1% agarose gel pad prepared with LB media. The cells are simultaneously observed by epifluorescence microscopy, using a mercury lamp with a DAPI filter, and by phase contrast microscopy. DAPI is excited at 359 nm and emits at 461 nm. The representative images are shown in Figure 4.2.

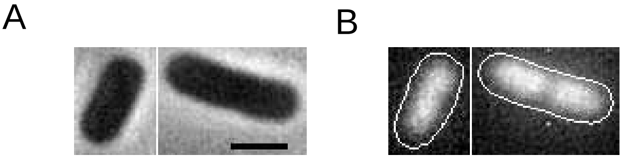


Figure 4.2: (A) Example phase contrast images of *E. coli* cells and (B) corresponding background-subtracted epifluorescence with the nucleoids stained by DAPI, and the detected cell contours superimposed. Scale bar is 1 μm. In both (A) and (B), contrast is enhanced for easier visualization. Image from **Publication II**.

However, the concentration of DAPI, required for live cell staining, is generally very high and therefore can be toxic for growing cells (Zink et al. 2003). Furthermore, as DAPI is an intercalating dye (Kapuscinski 1995), the observed nucleoids sizes are usually slightly larger than expected. Due to these reasons, a different approach is used for **Publication III**.

The new approach is based on several Nucleoid Associated Proteins (NAP) which participate in structural organization of *E. coli*. The major NAP found in this organism are *H-NS, HU, Fis, IHF* and *StpA* (Dillon and Dorman 2010). The *HU* protein, a dimeric histone-like protein, is the most abundant one and has been thoroughly studied (Azam et al. 1999; Claret and Rouviere-Yaniv 1997). For these reasons, this protein has been modified by tagging mCherry to study the spatial distribution of nucleoids (Fisher et al. 2013). This modified protein serves the purpose for this study.

The plasmid pAB332 carrying hupA-mCherry was generously provided by N. Kleckner (Harvard University, U.S.A.). This plasmid is inserted into DH5α-PRO strain of *E. coli* cells for the measurements concerning this study. The nucleoids in individual cells are observed using Highly Inclined and Laminated Optical sheet (HILO) microscopy (Tokunaga et al. 2008). Nucleoids tagged by hupA-mCherry are observed using a 543 nm HeNe laser and a band-pass 608-683 nm emission filter. The cells borders for the analysis are detected from images taken by phase contrast microscopy.

4.2 Image Processing and Computation

For the studies of **Publication II** and **Publication III**, multiple sets of time-lapse, microscopy images were collected for different temperature as well as stress conditions. Based on the requirement of the study, different image processing and data analysis methods as well as statistical tools were developed and then used.

4.2.1 Cell Segmentation

From the confocal microscopy images used in both studies, cells are detected using a semi-automatic method as in (Kandhavelu et al. 2012b). For that, the images are first temporally aligned using cross-correlation. This alignment removes any possible drift occurred during the image acquisition process. Such drifts can occur due to several reasons such as temperature change, unexpected slight movement of objective or stand. Because of such drifts same cells are at slightly different position in the adjacent frames which complicates the tracking of cells over time.

This effect is eliminated by cross correlating the consecutive frames to find the number of shifted pixels along the both axes. The number of shifted pixels is an integer value, and thus round-off errors can accumulate. The round-off error is corrected by comparing each frame with several preceding frames and computing the average shift. Once this process is complete for all frames in the time-series, the maximum area that is common to all the frames is computed. This area gives the amount of drift that occurred from the start of the time-series. All the frames are thus cropped to this maximum common area. This ensures the least possible drift of cells between frames, and thus makes cell tracking over time easy and effective.

4.2.1.1 Cell Features Extraction

The cells are segmented by manually drawing a mask over the region that each cell occupies during the time-series. The masking of cells is done using a simple computer graphics program, which in this study was GIMP (the GNU Image Manipulation Program). To track a division event, two new masks are drawn over the new cells in the frame where the division is first observed. Once masking of all the time-series images is complete, the automated analysis is run at each time point and for each mask. During this, Principal Component Analysis (PCA) of the fluorescence distribution under the mask is used to obtain the features of the cell, such as position, axis lengths, and orientation, among others. While using the PCA, there is a strong possibility of biasing the centroid of the cell towards the location of the bright spots and getting erroneous measures of the cell. To remove this problem, the fluorescence distribution within each mask is thresholded to enforce a uniform fluorescence within the cell. This is achieved by reducing the intensity of the 30 brightest pixels to the intensity of the 30th brightest pixel prior to subjecting the cells to PCA. By inspection, for the measurements in this study, it is found to have removed all such biases.

4.2.1.2 Cell Lineage Construction

To construct cell lineages, each cell at each time moment is assigned a 'parent'. The assignment is done such that for a cell in a given frame, 'parent' is the cell in the previous frame with the nearest centroid to its centroid. This assignment is done after transforming the previous frame's cell centroids into the cell's space (i.e. poles at (-1,0) and (1,0), and sides at (0,-1) and (0,1)) to avoid incorrectly assigning adjacent cells as the cell's parent. A division is assumed to have occurred when two cells are assigned the same parent (verified also by inspection). The results of cell segmentation are illustrated in Figure 4.3 B.

For **Publication III** which has temporal images obtained by bright-field microscopy (as illustrated in Figure 4.1 A), cells are automatically segmented using CellAging (Häkkinen et al. 2013). Manual correction is used to remove any errors resulting from the automated segmentation. After manual correction, CellAging tracks the cells across time and finds cell divisions. It further extracts cell features such as area over time, division time, and parent, among others. The asymmetry in cell size between sister cells, for this study, for each division event is then calculated as:

$$\Delta S = (S_c - S_s)/(S_c + S_s) \tag{4.1}$$

where, S_c and S_s are the areas of the daughter cell of interest and of its sister cell, respectively, immediately after division. This measure, where -1 and +1 denote maximum asymmetries and 0 denotes symmetry, is a factor of two different from the 'percent difference' used by, for example (Männik et al. 2012).

From the set of temporal images obtained by phase contrast microscopy (to extract the nucleoid measurements for **Publication II** and **Publication III**), cells were automatically segmented using multi-resolution analysis and maximum likelihood estimation (MAMLE) tool (Chowdhury et al. 2013). MAMLE is a multi-step automated segmentation tool that has been effective in segmenting the images from phase contrast microscopy. After automated segmentation, manual corrections by inspection are done to remove any errors. Subsequent analysis for lineage construction and feature extraction is performed, as above, using CellAging (Häkkinen et al. 2013).

4.2.2 Spot Segmentation

Both **Publication II** and **Publication III** requires segmentation of fluorescent MS2-GFP-RNA complexes. These complexes are detected in each cell, at each frame of the time-series, using the same method as in (Kandhavelu et al. 2012b). The fluorescent complexes inside each mask are automatically segmented using a Kernel Density Estimation (KDE) method for spot detection as in (Ruusuvuori

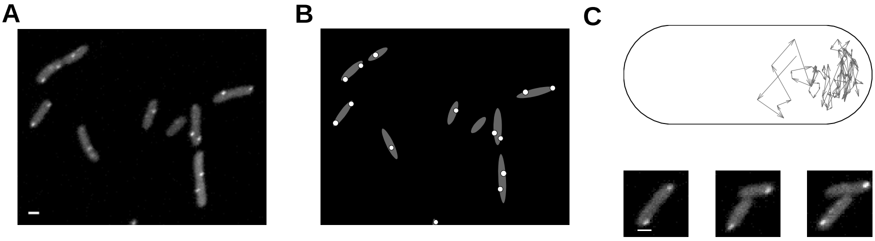


Figure 4.3: (A) Example image of *E. coli* cells with fluorescent MS2-GFP-RNA complexes within. (B) Segmentation and principal component analysis results of the image in (A) with cells (*gray*) and complexes (*white*). (C) One example of the extracted displacement vectors of a complex from its consecutive positions in the cell. Three images of the cell are shown below, taken at 40, 80, and 120 *min* (displacement vectors are from the upper cell). Scale bars are 1 μm. The contrast of these images was enhanced for easier visualization. Image from **Publication II**.

et al. 2010). Gaussian kernel, as in (Kandhavelu et al. 2012b), is used for KDE. This method measures the local smoothness of the image and then determines spot locations by designating areas with low smoothness as spots. Once a spot area is picked, more features of the spot, such as position, total fluorescence intensity, and area, among others, are extracted. Cell-background-corrected complex intensities are then calculated by subtracting the mean cell background intensity multiplied by the area of the complex from the total fluorescence intensity of each complex. The results of spot segmentation are illustrated in Figure 4.3 B.

In both publications, the issues with quantizing the number of complexes in each cell (Häkkinen et al. 2014) is avoided by measuring directly the fluorescence intensities of the complexes. In particular, the long-term spatial distribution of complexes in **Publication II** is obtained using the fluorescence intensities while the difference in complex intensities between daughter cells in **Publication III** at a division event is calculated as:

$$\Delta I = (I_c - I_s)/(I_c + I_s) \tag{4.2}$$

where, I_c and I_s are the total background-corrected complex intensities of the daughter cell of interest and of its sister cell, respectively, immediately after division.

In **Publication II**, to study the kinetics of segregation of individual complex, cells that contained at most one complex throughout their lifetime (either inherited or produced) are selected. This is done to ensure that the complexes could be reliably tracked. From these cells, once the complexes are detected at each time point, displacement vectors from their positions in consecutive frames is obtained. An example of the extracted displacement vectors of a complex from its positions at consecutive frames is illustrated in Figure 4.3 C.

4.2.3 Nucleoid Detection

For **Publication II**, to compare the spatial distribution of MS2-GFP-RNA complexes with the position of nucleoid, the sizes of nucleoids are estimated based on the fluorescence intensities resulting form DAPI staining. An example image of DAPI stained nucleoid obtained from epifluorescence microscopy is shown in Figure 4.2. The background of such images is removed by subtracting a cubic polynomial surface, fitted to the image by L1-norm minimization. More specifically, this is achieved by minimizing the absolute difference between the surface and the image; see for example (Portnoy and Koenker 1997). The fluorescence intensities in each cell were then extracted and the borders of the nucleoid are determined by fitting the intensity distribution to a piecewise-constant probability density function with two pieces by maximum likelihood.

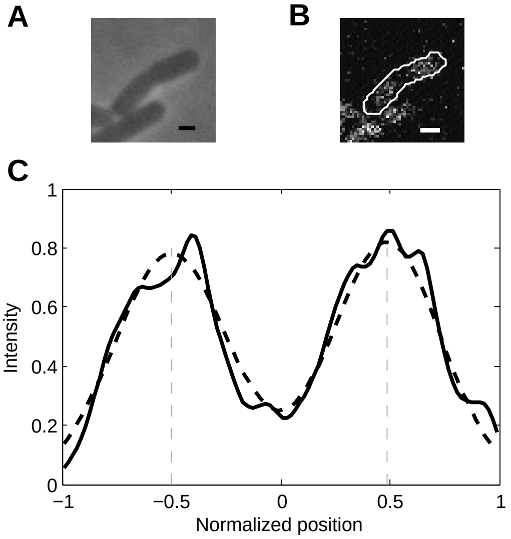


Figure 4.4: (A) Example image of a cell expressing hupA-mCherry, in the frame before division is detected, taken with phase contrast microscopy to visualize the cell borders. (B) Corresponding image from confocal microscopy to visualize the fluorescence intensity of hupA-mCherry (cell outline obtained from (A) is shown in *white*). (C) KDE of the fluorescence intensity distribution (*solid*) along the major axis of the cell shown in (B), with the Gaussian mixture model fit (*dashed*). Scale bars are 1 μm. Image from **Publication III**.

In **Publication III**, to compute the distance between the replicated nucleoids and relate it to cell size asymmetry in division, the nucleoid positions are estimated in cells the moment prior to their division. The centers and widths of the two replicated nucleoids are estimated by fitting a Gaussian mixture density distribution with two Gaussian with equal weight to the fluorescence intensity distribution along the major axis of the cells, normalized by half the cell length. For that, a least-squares fit is done to the KDE of fluorescence intensity distribution, using a Gaussian kernel and a bandwidth of 0.05. The centers of the resulting two Gaussian are interpreted to be the centers of the replicated nucleoids. This approach is illustrated in Figure 4.4 and has been verified by inspection as well as by estimating Signal-to-Noise Ratio (SNR) in the measurements.

4.3 Simulator and Models

The design of simulator, built in **Publication I** and used in **Publication V**, is based on the NRM and the delayed SSA. The simulator also uses the concepts of dynamic compartments as described in (Spicher et al. 2008). The efficiency of the NRM is exploited to simulate discrete events which also includes delayed events. Also, the flexibility of the NRM to incorporate other simulation algorithms, termed sub-simulations, is also utilized. For example, the reaction system within each compartment is represented by its own NRM priority queue whereas the overall NRM has its own priority queue. The priority queue of each compartment publishes a “next firing time” to the overall NRM and the indexed priority queue in the overall NRM provides the flexibility needed to add and remove entire sub-simulations (in this case, compartments) at runtime.

The main simulation loop consists of the following steps:

1. **Selection**: The event and the time of its occurrence is determined. This step runs in constant time since the next compartment in which an event occurs is stored at the front of the indexed priority queue, and the next event to occur in that compartment is stored at the front of its indexed priority queue.

2. **Execution**: The event is performed, moving time forward and modifying the state of the simulation according to the type of the event that occurred. Reaction propensities that depend on the changed state are flagged as *dirty*.

3. **Update**: The *dirty* propensities and putative reaction times are recalculated and the changes are propagated back up the data structures. This step takes $O(logS \cdot U(S))$, where S is the number of sub-simulations in the system and U(S) is the mean time required to update the *dirty* sub-simulations.

Thus, the runtime complexity of the simulation depends solely on the Execution and Update steps which in turn depend on the nature of the sub-simulations that occur.

4.3.1 Compartments

Compartments contain a subset of the reactants in a simulation, which interact differently with the rest of the system. In each compartment, the molecules only react with other molecules in the same compartment. For example, a reactant α in compartment P will react with other molecules in P, but not with reactants (including the α molecules) in a different compartment, say Q. Each compartment has its own state vector as it contains a set of molecules. Furthermore, as a compartment is created, there are reactions created for it. This is accomplished by introducing the notion of compartment type whereby every compartment is of a type and contains set of molecules and reactions associated to it. A reactant α in compartment P is thus denoted as α@P. Reactions in the model are defined as occurring within or between compartment types. A separate instance of each reaction occurring in type P will be created in each compartment of type P.

Compartments, in this simulator design, are hierarchy organized. It means that higher-level compartments always contain lower-level compartments. Compartment types are similarly organized, creating a hierarchy of compartment type. Based on this hierarchy, the reactions that occur in the system are classified into two types, namely:

(a) **Intra-compartment reactions**: The reactions that occur within a compartment are intra-compartment reactions. Each of the compartments has their own NRM priority queue that contains all the information about reactions that occur in that specific compartment. These priority queues publish "next firing time" to the overall NRM. The simulation proceeds through the main steps of simulation loop, namely, Selection, Execution and Update.

    In **Publication V** which models a population of growing *E. coli* cells, cellular processes such as gene expression, non-functional protein production, aggregation, and cell growth, among others are modeled using intra-compartment reactions. As all of such reactions occur due to the interactions between components within a cell, intra-compartment reactions perfectly capture them.

(b) **Inter-compartment reactions**: The reactions that span between the compartments are inter-compartment reactions. Because of its complexity in implementation, inter-compartment reactions are only allowed to occur between compartments and their containing compartments. It means that reactions may only span vertically across the compartment hierarchy (i.e. between parents and children), but not horizontally (i.e. between siblings). Due to this restriction, a child compartment can only affect its sibling indirectly by first changing the parent compartment.

    To simulate a system with such reactions, the propensities of all these reactions needs to be computed. For simplicity, the vertical reaction is always assumed to occur in the lowest level compartment. For example, if a compartment Q is within compartment P, any vertical reaction between these compartments is assumed to occur in compartment Q. The propensities of vertical reactions are then calculated from the possible combinations of reactants in the compartment the reaction occurs in, and its containing compartments. The propensity a_μ that a molecule 'a' in P will react with a molecule 'b' in Q, based on concepts of bimolecular reaction, is given as:

$$a_\mu = X_{a@P} X_{b@Q} c_\mu$$

The vertical reaction occurring between the parent and the child compartments are considered as a separate sub-simulation of the parent compartment's NRM. This sub-simulation has its own time that passes as $X_{super}dt$, where X_{super} is the population of the molecule in the parent compartment. For a given reaction time in the simulation, the reaction with the earliest tentative firing time remains at the front of the priority queue of the sub-simulation and can be retrieved in a constant time. Any change in the population of the molecule in the parent compartment (X_{super}) then only requires the next firing time of the sub-simulation to be updated.

In **Publication V**, these reactions are used to model the interaction between the cell environment and the cell. One basic example is cell division which occurs when an *E. coli* cell attains a specific length and results in a new cell in the environment. This reaction including the ones required in modeling bacteriophage's infection of a cell population for the supplementary of **Publication I** is modeled using inter-compartment reactions.

4.3.2 Delayed Reactions

Following the delayed SSA, the delayed products of a reaction are inserted into a wait-list. These are retrieved and reinserted into the simulation when the actual simulation time crosses the delayed time. A separate wait list is created for each compartment, and is inserted as a sub-simulation of the global NRM. The wait list itself is priority queue of delayed species. The wait list for each compartment is also beneficial during the compartment destruction. When modeling a dynamic population of cells, the construction and destruction of the compartments can be a regular event. Since the wait list exists are distributed to the compartments, the compartments can be destructed without any global dependencies of delayed molecules within it.

The most common example of a delayed product is the delay in the release of the protein during translation. The translation reaction with delayed release of proteins has been used in **Publication V** as well as in the examples of **Publication I**.

4.3.3 Division and Partitioning

The division of cells, used in **Publication V**, is modeled by the construction of compartments in the runtime of simulation. This is handled by inter-compartment between the original compartment and the system. Specifically, when a division occurs, the original compartment (the parent) remains in the simulation as one of the daughter compartments, while a new compartment is created in the system (environment) for the other daughter. This is handled efficiently using the NRM to add and remove the compartment's reactions from the system.

Another important aspect of cell division is the partitioning of molecules from parent cells to the newly created cells. The physical mechanisms behind the partitioning are known to differ between molecule species. Thus, upon division, a "mock partitioning process" for each divided species describing the partitioning statistics is executed. Several partitioning schemes and their mock processes have been described and their statistical properties have also been examined in (Huh and Paulsson 2011b). The three groups that these processes are divided into, as classified in (Huh and Paulsson 2011b), are implemented in the simulator.

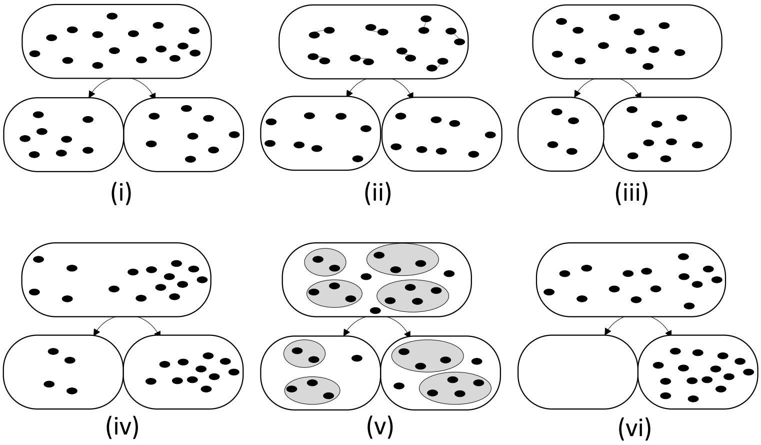


Figure 4.5: Schematic representation of different partitioning schemes. (i) Perfect Partitioning in which the molecules are equally partitioned into daughter cells (ii) Pair formation in which the molecules form pairs and each molecule in a pair is partitioned to different daughter cell. The ones that do not form pair are independently partitioned.(iii) Random size in which the daughter cells are of different size and inherit molecules based on their size.(iv) Preferential in which the molecules are preferentially inherited by one of the daughters. (v) Clustered in which the molecules aggregate to form clusters and then partitioned. (vi) All or Nothing in which all the molecules are inherited by one of the daughters.

(a) **Independent Partitioning**. This is a simple partitioning distribution where each partitioned molecule has an independent probability of ending up in either of the daughter cells. This type of independence can be realized by having a well-mixed cytoplasm, by having immobile molecules independently appear in either cell half, or by randomly picking the molecules and moving those to either cell half. This results in a partitioning scheme that is unbiased and the molecules after partitioning are distributed binomially.

(b) **Ordered Partitioning**. In ordered partitioning, mechanisms in the cell interact directly or indirectly with each other to create a more even distribution of the molecules between daughter cells during division. These mechanisms may be used by a cell as internal control mechanisms to make the partitioning more evenly distributed, or to compensate for disordered mechanisms. An ideal example is perfect partitioning in which each daughter cell gets exactly half of the molecules in the parent cell. It can be implemented by dividing the molecules into exactly half and then putting them to the daughter compartments, as illustrated in Figure 4.5(i).

Another example is molecules that form pairs prior to division and upon division are segregated to different daughter cells. This strategy lowers the variance in partitioning when compared to independent partitioning. This is implemented by probabilistically forming pairs of molecules when the division occurs. The pairs are then split with a certain probability, with one molecule of each successfully split pair going to one of the daughter compartment. The molecules that failed to make pairs and the pairs that fail to split are then partitioned independently to the daughter cells. It is illustrated in Figure 4.5(ii).

(c) **Disordered Partitioning**. In disordered partitioning, the variation in the partitioning system or the intracellular environment randomizes population levels between daughter cells more than independent partitioning. There are several examples of disordered partitioning schemes. If daughter cells do not have the same volume following a division, the larger daughter is likely to inherit more molecules than the smaller one. The size asymmetry here is modeled using a Beta distribution. The molecules to be partitioned are then segregated binomially depending upon the ratio of the daughter cell sizes. This partitioning results in significant differences in the daughter cells, and is referred to as random size partitioning. It is illustrated in Figure 4.5(iii).

For molecules, such as protein aggregates, that preferentially segregate to one of the cell poles (Winkler et al. 2010), the partitioning is expected to result in a scheme where most of the molecules will end up in one of the daughter cells. This can be implemented by a biased binomial partitioning of the molecules such that the parent compartment likely retains more molecules than the newly formed compartment, as illustrated in Figure 4.5(iv). After few generations, this preferential partitioning scheme will result in older cells with very large number of molecules.

For the cellular components or molecules that form or exist in clusters, the partitioning, given that these are partitioned independently, is expected have a even greater variance from division. To model their partitioning during division, we first group the molecules into N clusters and then independently segregate the clusters into the daughter cells. This partitioning scheme is referred to as clustered partitioning and is illustrated in Figure 4.5(v).

> The extreme of the disordered partitioning is the partitioning scheme in which all the molecules are always partitioned into one of the daughter cells. This would introduce the most variance possible during division. This is implemented by randomly selecting one of the daughter cell and putting all the molecules to it. This scheme is referred to as all or nothing and is illustrated Figure 4.5(vi).

Using non-functional proteins as probes that is assumed to affect cell health, the effect of all these partitioning schemes on vitality of *E. coli* population is studied in **Publication V**.

4.3.4 Retention of Complexes

For **Publication IV**, the cell is modeled as a compartmentalized one-dimensional space divided into N homogeneous sub-volumes, which is indexed from [1,N]. For all the models, N is set to 100, and the diffusion coefficient (D) is set to $1.43 * 10^{-2} \mu m^2/min$ based on previous measurements (Golding and Cox 2006) and on measurements from **Publication II**. The movement of the complexes along the major axis of the cell is modeled with unimolecular reactions based on the RDME (Gardiner et al. 1976).

To model the motion of complex from position x to position $x+1$, propensity of the forward reaction represented as $\overrightarrow{\alpha}(x)$ is used. Similarly, the propensity of backward reaction, $\overleftarrow{\alpha}(x)$, models the motion of a complex from position x to $x-1$. These propensity functions account for the combined effects of the rod shape of the cell and the nucleoid on the motions of the complexes. Based on these, stochastic models are developed and simulated. The results are used to distinguish between the possible mechanisms responsible to retain the complexes at the cell poles.

5 Summary of Results

The effects of intracellular and partitioning asymmetries in *E. coli* using measurements and models is studied in this thesis. This chapter summarizes the results of individual study that led to each of the publications.

In **Publication I**, a simulator (SGNS2) that can simulate models of systems with multi-delayed events, dynamic compartments and partitioning of molecules is developed. The design of SGNS2 is based on a combination of the efficient implementations of the SSA and incorporates the concept of dynamic compartments. It can simulate an array of biological processes. The usage is exemplified in a simulation of gene expression at the nucleotide and codon levels (Mäkelä et al. 2011). Furthermore, SGNS2 is also suited to study partitioning in cell division, which is known to have effect on aging, among other processes. In such studies where data from populations over multiple generations is required, SGNS2 is more relevant as performing experimental measurements might be complex as well as tedious.

Some example models are built to demonstrate the utility of SGNS2. Of particular interest is the model of biased partitioning of protein aggregates in *E. coli.* These aggregates are known to accumulate at the cell poles and reduce cell vitality (Lindner et al. 2008; Stewart et al. 2005). The result of simulation using SGNS2 shows the growth rate of cells for few generations is strikingly similar to the one measured in (Stewart et al. 2005). Another model is used to study the effects of cell-cycle synchrony on the population-level statistics of RNA numbers as in (Lloyd-Price et al. 2012b). A final model is built to study bacteriophage infection of a dynamic bacterial population.

In **Publication II**, the kinetics and long-term spatial distribution of biologically inert complexes, composed of RNA molecules tagged with multiple MS2-GFP proteins, in the cytoplasm of *E. coli* cells is studied in optimal and sub-optimal temperature conditions. The previously reported tendency of these complexes to travel toward the poles is found to be a symmetric process in all the tested conditions. It is found that, cell divisions introduce asymmetries in their numbers between the old and the new poles of cells of subsequent generations, and after few generations, between sister cells. Similar observations for protein aggregates were made in (Lindner et al. 2008). The kinetics of movement of these complexes is further studied to investigate the mechanisms responsible for their long-term spatial distribution.

A strong anisotropy in the displacement distribution, due to obstruction of the motion of complexes, is found at approximately half way between the cell extremes and the center. The hypothesis that the excluded volume effect at mid-cell determines the positioning of these complexes is found to be supported by the overall two-dimensional distributions of positions occupied by complexes throughout their lifetimes. The complexes are found to avoid the nucleoid, both axially and radially. Further support is found because of the agreement between the location of the nucleoid and the location where movement of complexes is obstructed. In addition, in the longest cells, where the nucleoid(s) occupy a relatively wider region due to nucleoid replication, the complexes are found to occupy a relatively smaller region at the poles. From all these evidences, it is concluded that the macromolecular crowding due to the nucleoid at mid-cell region is responsible for the retention of complexes at the cell poles. Additional support for this conclusion is provided by modeling. From the models, in absence of the effects due to nucleoid, namely, in absence of the anisotropy in the velocity distribution of complexes, the retention at the poles is severely hampered as the spatial distribution of complexes is uniform throughout the cell. On the other hand, introduction of the anisotropy in the region between poles and mid-cell regenerates an accurate long-term spatial distribution of the complexes.

In **Publication III**, the robustness of division symmetry and functional consequences due to breaking of this symmetry in *E. coli* is studied. For that, previous assessment of the degree of morphological symmetry in division in these organisms, that most divisions are symmetric with some exceptions, is verified in optimal conditions. The morphological asymmetries that exist in division is found to be a source of functional asymmetries between sister cells as the differences in cell sizes correlates with the differences in the numbers of fluorescent MS2-GFP-RNA complexes inherited by these cells from the mother cell.

The asymmetry in sizes in division is then studied for non-optimal environmental conditions. In mild chemical stresses (acidic shift and oxidative stress), the distributions of asymmetry in sizes between sister cells is found to be statistically indistinguishable from those at optimal growth conditions. However, this is not the case under sub-optimal temperature conditions. The frequency of asymmetric divisions is found to increase as temperature changed from optimal to sub-optimal. From these, it is concluded that the mechanism responsible for the symmetry observed in division is robust to the mild chemical stresses but not to temperature changes. The mechanisms that have been identified to affect the selection of point of division are the ‘Min system’ (Yu and Margolin 1999; Kerr et al. 2006) and the positioning of the nucleoids prior to division (Wang et al. 2005; Mulder and Woldringh 1989; Woldringh et al. 1991). To investigate the observed variance in the point of division, the replicated nucleoids at time moment prior to division is studied at sub-optimal temperatures. Even though the relative sizes of the nucleoids are unchanged, there is an increase in the mean relative distance between nucleoids. Given that the positioning of the division point is near random in between nucleoids (Mulder and Woldringh 1989), the increase in the mean relative distance necessarily increases the variance in the division point. Furthermore, since no mini-cells is observed in any of the tested conditions, to indicate the disruption of Min system, it is concluded that the role of nucleoid is, to some extent, affected at sub-optimal temperatures.

In **Publication IV**, stochastic modeling of polar retention of complexes is used to distinguish between the functional consequences of possible retention mechanisms. Namely, in **Publication II**, it is hypothesized the nucleoid occlusion is the responsible mechanism, as a strong anisotropy in displacement distribution is observed at the border of nucleoid. However, the observed spatial distribution of complexes could also, theoretically, arise from other sources such as heterogeneities in the speed of the complexes along the major axis of the cell. In order to distinguish between these mechanisms, it is necessary to study the functional consequences of the models based on these. These models, with parameters tuned to match the measurements reported in **Publication II**, is found to generate good fits to the long-term spatial distribution. However, the model with varying speed along the major cell axis, at the time scale of the measurements, is unable to reproduce the observed anisotropic displacement distribution at the border between the mid-cell and poles. Based on this, it is concluded that the polar retention most likely relies on these anisotropies in the displacement distribution rather than differences in speeds. This is consistent with the hypothesis that the observed long-term behavior is the result of macromolecular crowding, likely due to the nucleoid.

In **Publication V**, possible selective advantages of different partitioning schemes of non-functional proteins in *E. coli* are investigated using a stochastic model. For that, a model of growing *E. coli* cells is developed and is parameterized by previous measurements. The model is verified to mimic the cellular processes such as gene expression, non-functional protein production, cell growth, cell division and partitioning of components, among others, within a cell of DH5α-PRO strain. Given the assumption that the accumulation of non-functional proteins reduces cell vitality, it is found that increase in the degree of asymmetry in partitioning of non-functional proteins increases the division rates of cell population. The degree of asymmetry is found to be further enhanced by properties such as preferential movement and clustering of the non-functional proteins. Furthermore, it is found that retention of the inherited proteins at the old pole can, on its own without any asymmetry, increase the mean division rate of the cell population. This retention is a non-energy consuming mechanism as it is driven by volume exclusion of nucleoid. It is therefore advantageous for the cells. The results suggest that the mechanisms of intracellular organization of non-functional proteins, including clustering and polar retention, affect the vitality of *E. coli* populations.

6 Discussion

The results reveal nucleoid as one of the regulating mechanism of the sources of asymmetries in *E. coli*, which ultimately generate phenotypic differences between sister cells. In particular, it is found that the nucleoid plays a vital role in segregation of large inert macromolecules and, consequently, on their the spatial distribution within the cytoplasm of *E. coli* cells. Furthermore, the positioning of nucleoid is found to regulate the location of septum that defines the division point in these organisms. Also it is shown that, both of the processes, namely segregation mechanism and cell division, are sources of asymmetries in these organisms and have detrimental consequences on cellular functionality.

It is interesting that both of these processes are regulated either by the presence or by the positioning of the nucleoid. It is indicative that aside from the central role of carrying the genetic information, nucleoid plays a vital role in variety of other cellular functions. In optimal growth condition, the nucleoid ensures the retention of inert macromolecules at the least functional region of cell i.e. at the cell poles. It further ensures that the sister cells in division have least possible morphological differences between them. Although the mechanism of retention is unaffected in sub-optimal temperatures, the symmetry in division starts to fail due to the increased distance between replicated nucleoids. Since the mechanism is not so robust, it is of interest to study the features of nucleoid in extreme environmental conditions. Further studies are required to investigate the changes in such conditions and, more importantly, their impact on both retention and symmetry in division.

It is important to mention that all the experimental and model results in this thesis are based on the measurements from *E. coli* strain DH5α-PRO. The slower growth rate of this strain (Jung et al. 2010) facilitates the automated tracking of cell divisions and partitioning of complexes from time-lapse microscopy images. Even with this difference in strain, the observations made here can be considered representative of a wild-type strain due to the behavioral similarity of these strains. For example, the segregation and the accumulation of molecules at the cell poles observed in DH5α-PRO has also been reported in wild-type strains(Lindner et al. 2008; Bakshi et al. 2012; Zhang et al. 2007; Yu et al. 2006). In addition, the morphological symmetry in division for both the strains are found to be strikingly similar (Männik et al. 2012). Furthermore, the results of models for wild-type strain should be qualitatively similar to those reported from models in this study. The models are flexible enough that the parameters of models can be easily tuned to match the measurements of a wild-type strain. Nonetheless in future, it would definitely be interesting to study the strain-to-strain variability in this organism.

Finally, a simulator (SGNS2) has been developed to simulate diverse biological processes such as gene expression, molecule interactions, delayed reactions, cell division, and stochastic partitioning, among others. It provides novel functionalities to model such stochastic cellular processes, and thus it is useful in studying *in silico* the phenotypic diversity of cell populations. This is exemplified in a model that accounts for the effect of various partitioning schemes on the division times of growing *E. coli* population. The model reveals the importance of asymmetries in partitioning of harmful proteins in overall vitality of the population. This general model is tunable for a different parameter-set, for example, parameters for a different *E. coli* strain or for a completely different species. Furthermore, although the model is already complete, it can be easily extended to accommodate the detailed features and effects of the nucleoid. For that, measurements of the nucleoid that informs on the parameters such as size, location, and replication time, among others are required. It would be interesting to add these features to the model and explore the processes such as retention at cell poles in detail.

Overall, it can be concluded that the functionality of *E. coli* cells is not immune to asymmetries at birth, which may explain why these organisms have evolved a robust process of symmetric division under optimal conditions. We expect our results to be valid for organisms or cells with similar division process. It can be hypothesized rejuvenation of such organisms has led to evolutionary processes that result in asymmetries in division, as one of the daughter is favored. Since complete understanding of the underlying complex and multifaceted mechanisms that characterize aging is one of the great challenges in biology, in future, it would be of interest to investigate sources of such asymmetries in eukaryotes. These organisms have a more complex internal structure and mechanisms that may allow more ingenious means to cope with the effects of aging, which is probably a fundamental property of all living beings.

Bibliography

Ackermann, M., L. Chao, C. T. Bergstrom, and M. Doebeli. “On the evolutionary origin of aging”. *Aging Cell* 6.2 (2007), pp. 235–244. DOI: `10.1111/j.1474-9726.2007.00281.x`.

Ackermann, M., S. C. Stearns, and U. Jenal. “Senescence in a bacterium with asymmetric division.” *Science* 300.5627 (2003), p. 1920. DOI: `10.1126/science.1083532`.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science, 2002.

Andrews, S. S. and D. Bray. “Stochastic simulation of chemical reactions with spatial resolution and single molecule detail.” *Physical Biology* 1.3-4 (2004), pp. 137–51. DOI: `10.1088/1478-3967/1/3/001`.

Arkin, A., J. Ross, and H. H. Mcadams. “Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage Lambda Infected *Escherichia coli* Cells”. *Genetics* 149 (1998), pp. 1633–48.

Avraham, N., I. Soifer, M. Carmi, and N. Barkai. “Increasing population growth by asymmetric segregation of a limiting resource during cell division.” *Molecular Systems Biology* 9.656 (2013), p. 656. DOI: `10.1038/msb.2013.13`.

Azam, T. A., A. Iwata, A. Nishimura, and S. Ueda. “Growth Phase-Dependent Variation in Protein Composition of the *Escherichia coli* Nucleoid”. *Journal of Bacteriology* 181.20 (1999), pp. 6361–6370.

Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. “Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells.” *Molecular Microbiology* 85.1 (2012), pp. 21–38. DOI: `10.1111/j.1365-2958.2012.08081.x`.

Ballesteros, M., A. s. Fredriksson, J. Henriksson, and T. Nyström. “Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes.” *The EMBO Journal* 20.18 (2001), pp. 5280–9. DOI: `10.1093/emboj/20.18.5280`.

Barker, M. G. and R. M. Walmsley. “Replicative Ageing in the Fission Yeast *Schizosaccharomyces pombe*”. *Yeast* 15 (1999), pp. 1511–1518.

Beckett, D., S Koblan, and G. K. Ackers. “Quantitative Study of Protein Association at Picomolar Concentrations : The lambda Phage cl Repressor”. *Analytical Biochemistry* 196 (1991), pp. 69–75.

Bernstein, J. A., A. B. Khodursky, P.-h. Lin, S. Lin-chao, and S. N. Cohen. "Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays". *Proceedings of the National Academy of Sciences of the United States of America* 99.15 (2002), pp. 9697–9702.

Birky, C. "Random partitioning of cytoplasmic organelles at cell division: The effect of organelle and cell volume". *Journal of Theoretical Biology* 106.4 (1984), pp. 441–447.

– "The partitioning of cytoplasmic organelles at cell division". *International Review of Cytology* 15 (1983), pp. 49–89.

Browning, D. F., S. J. W. Busby, and C. S. J. W. B. "The regulation of Bacterial Transcription Initiation". *Nature Reviews* 2.January (2004), pp. 1–9. DOI: `10.1038/nrmicro787`.

Buc, H. and W. R. McClure. "Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps". *Biochemistry* 24.11 (1985), pp. 2712–2723. DOI: `10.1021/bi00332a018`.

Cao, Y., H. Li, L. Petzold, and J Bruck. "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems". *Journal of Chemical Physics* 121.9 (2004), pp. 4059–4067. DOI: `10.1063/1.1778376`.

Chant, J. "Cell Polarity in yeast". *Annual Review of Cell and Developmental Biology* 15.1 (1999). DOI: `10.1146/annurev.cellbio.15.1.365`.

Chen, J. C., D. S. Weiss, J.-m. Ghigo, and J. Beckwith. "Septal Localization of FtsQ , an Essential Cell Division Protein in *Escherichia coli*". *Journal of Bacteriology* 181.2 (1999), pp. 521–530.

Chia, W., W. G. Somers, and H. Wang. "Drosophila neuroblast asymmetric divisions: cell cycle regulators, asymmetric protein localization, and tumorigenesis." *The Journal of Cell Biology* 180.2 (2008), pp. 267–72. DOI: `10.1083/jcb.200708159`.

Choi, P. J., L. Cai, K. Frieda, and S. Xie. "A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell". *Science* 4142.October (2008), pp. 442–446.

Chowdhury, S., M. Kandhavelu, O. Yli-harja, and A. S. Ribeiro. "Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE)". *BMC Bioinformatics* 14 (2013), pp. 1–13.

Claret, L. and J. Rouviere-Yaniv. "Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival." *Journal of Molecular Biology* 273.1 (1997), pp. 93–104. DOI: `10.1006/jmbi.1997.1310`.

Coelho, M., S. J. Lade, S. Alberti, T. Gross, and I. M. Tolić. "Fusion of protein aggregates facilitates asymmetric damage segregation." *PLoS Biology* 12.6 (2014), e1001886. DOI: `10.1371/journal.pbio.1001886`.

Coquel, A.-S., J.-P. Jacob, M. Primet, A. Demarez, M. Dimiccoli, T. Julou, L. Moisan, A. B. Lindner, and H. Berry. "Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect.” *PLoS Computational Biology* 9.4 (2013), e1003038. DOI: `10.1371/journal.pcbi.1003038`.

Crick, F. “Central Dogma of Molecular Biology”. *Nature* 227 (1970), pp. 561–563.

Cullum, J. and M. Vicente. “Cell Growth and Length Distribution in *Escherichia coli*”. *Journal of Bacteriology* 134.1 (1978), pp. 330–337.

Dillon, S. C. and C. J. Dorman. “Bacterial nucleoid-associated proteins, nucleoid structure and gene expression.” *Nature Reviews Microbiology* 8.3 (2010), pp. 185–95. DOI: `10.1038/nrmicro2261`.

Doe, C. Q. “Neural stem cells: balancing self-renewal with differentiation.” *Development* 135.9 (2008), pp. 1575–87. DOI: `10.1242/dev.014977`.

Donachie, W. D. and K. J. Begg. “"Division potential" in *Escherichia coli.*” *Journal of Bacteriology* 178.20 (1996), pp. 5971–6.

Donachie, W. D., K. J. Begg, and M. Vicente. “Cell length, cell growth and cell division”. *Nature* 264 (1976), pp. 328–333.

Elf, J. and M. Ehrenberg. “Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases”. *Systems Biology* 1.2 (2004), pp. 230–236. DOI: `10.1049/sb`.

Elowitz, M. B., M. G. Surette, P.-e. Wolf, J. B. Stock, and S. Leibler. “Protein Mobility in the Cytoplasm of *Escherichia Coli*”. *Journal of Bacteriology* 181.1 (1999), pp. 197–203.

Erjavec, N., M. Cvijovic, E. Klipp, and T. Nyström. “Selective benefits of damage partitioning in unicellular systems and its effects on aging.” *Proceedings of the National Academy of Sciences of the United States of America* 105.48 (2008), pp. 18764–9. DOI: `10.1073/pnas.0804550105`.

Errington, F. P., E. O. Powell, and N. Thompson. “Growth Characteristics of Some Gram-negative Bacteria”. *Journal of General Microbiology* 39.1 (1965), pp. 109–123. DOI: `10.1099/00221287-39-1-109`.

Errington, J. “Dynamic proteins and a cytoskeleton in bacteria.” *Nature Cell Biology* 5.3 (2003), pp. 175–8. DOI: `10.1038/ncb0303-175`.

Fisher, J. K., A. Bourniquel, G. Witz, B. Weiner, M. Prentiss, and N. Kleckner. “Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells.” *Cell* 153.4 (2013), pp. 882–95. DOI: `10.1016/j.cell.2013.04.006`.

Fusco, D., N. Accornero, B. Lavoie, S. M. Shenoy, J. M. Blanchard, R. H. Singer, and E. Bertrand. “Single mRNA molecules demonstrate probabilistic movement in living mammalian cells.” *Current Biology* 13.2 (2003), pp. 161–167.

Gardiner, C. W., K. J. McNeil, D. F. Walls, and I. S. Matheson. “Correlations in Stochastic Theories of Chemical Reactions”. *Journal of Statiscal Physics* 14.4 (1976), p. 307.

Ghosh, S. K., S. Hajra, A. Paek, and M. Jayaram. “Mechanisms for chromosome and plasmid segregation.” *Annual Review of Biochemistry* 75 (2006), pp. 211–41. DOI: `10.1146/annurev.biochem.75.101304.124037`.

Gibson, M. A. and J. Bruck. “Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels”. *Journal of Physical Chemistry* 104 (2000), pp. 1876–1889.

Gillespie, D. T. “A general method for numerically simulating coupled chemical reactions”. *Journal of Computational Physics* 22 (1976), pp. 403–434.

– “A rigorous derivation of the chemical master equation”. *Physica A* 188 (1992), pp. 404–425.

– “Exact Stochastic Simulation of Coupled Chemical Reactions”. *Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361.

– “Stochastic Simulation of Chemical Kinetics”. *Annual Review of Physical Chemistry* 58 (2007), pp. 35–55. DOI: `10.1146/annurev.physchem.58.032806.104637`.

Glick, B. R. “Metabolic Load and Heterologous Gene Expression.” *Biotechnology Advances* 13.2 (1995), pp. 247–261.

Goeddel, D. V., D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirosef, A. Kraszewskit, K. Itakuraf, and A. D. Riggs. “Expression in *Escherichia coli* of chemically synthesized genes for human insulin”. *Proceedings of the National Academy of Sciences of the United States of America* 76.1 (1979), pp. 106–110.

Golding, I. and E. C. Cox. “RNA dynamics in live *Escherichia coli* cells”. *Proceedings of the National Academy of Sciences of the United States of America* 101.31 (2004), pp. 11310–11315.

Golding, I. and E. Cox. “Physical Nature of Bacterial Cytoplasm”. *Physical Review Letters* 96.9 (2006), p. 098102. DOI: `10.1103/PhysRevLett.96.098102`.

Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. “Real-Time Kinetics of Gene Activity in Individual Bacteria”. *Cell* 123 (2005), pp. 1025–1036. DOI: `10.1016/j.cell.2005.09.031`.

Gordon, S., J. Rech, D. Lane, and A. Wright. “Kinetics of plasmid segregation in *Escherichia coli.*” *Molecular Microbiology* 51.2 (2004), pp. 461–469. DOI: `10.1046/j.1365-2958.2003.03837.x`.

Gottesman, S. and M. R. Maurizi. “Regulation by proteolysis: energy-dependent proteases and their targets.” *Microbiological Reviews* 56.4 (1992), pp. 592–621.

Häkkinen, A., A.-B. Muthukrishnan, A. Mora, J. M. Fonseca, and A. S. Ribeiro. “CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli.*” *Bioinformatics* 29.13 (2013), pp. 1708–9. DOI: `10.1093/bioinformatics/btt194`.

Häkkinen, A., M. Kandhavelu, S. Garasto, and A. S. Ribeiro. “Estimation of fluorescence-tagged RNA numbers from spot intensities.” *Bioinformatics* 30.8 (2014), pp. 1146–1153. DOI: `10.1093/bioinformatics/btt766`.

Herbert, K. M., A. L. Porta, B. J. Wong, R. A. Mooney, K. C, R. Landick, and S. M. Block. “Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules”. *Cell* 125.6 (2006), pp. 1083–1094.

Hippel, P. H. V. and O. G. Berg. “Facilitated Target Location in Biological Systems”. *Journal of Biological Chemistry* 264.2 (1989), pp. 675–678.

Hsu, A.-L., C. T. Murphy, and C. Kenyon. “Regulation of aging and age-related disease by DAF-16 and heat-shock factor.” *Science* 300.5622 (2003), pp. 1142–5. DOI: `10.1126/science.1083701`.

Hu, Z., C. Saez, and J. Lutkenhaus. “Recruitment of MinC , an Inhibitor of Z-Ring Formation , to the Membrane in *Escherichia coli* : Role of MinD and MinE”. *Journal of Bacteriology* 185.1 (2003), pp. 196–203. DOI: `10.1128/JB.185.1.196`.

Huang, K. C., Y. Meir, and N. S. Wingreen. “Dynamic structures in *Escherichia coli*: Spontaneous formation of MinE rings and MinD polar zones”. *Proceedings of the National Academy of Sciences of the United States of America* 100.22 (2003), pp. 12724–12728.

Huh, D. and J. Paulsson. “Non-genetic heterogeneity from stochastic partitioning at cell division”. *Nature Genetics* 43.2 (2011), pp. 95–99. DOI: `10.1038/ng.729`.

– “Random partitioning of molecules at cell division”. *Proceedings of the National Academy of Sciences of the United States of America* 108.36 (2011), pp. 15004–15009. DOI: `10.1073/pnas.1013171108`.

Isaacson, S. A. “The reaction-diffusion master equation as an asymptotic approximation of diffusion to small target”. *Journal of Applied Mathematics* 70.1 (2009), pp. 77–111.

Isaacson, S. A. and C. S. Peskin. “Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations”. *SIAM Journal of Scientific Computation* 28.1 (2006), pp. 47–74.

Jazwinski, S. M. “Growing old: metabolic control and yeast aging.” *Annual Review of Microbiology* 56 (2002), pp. 769–92. DOI: `10.1146/annurev.micro.56.012302.160830`.

Johnson, L. R. and M. Mangel. “Life histories and the evolution of aging in bacteria and other single-celled organisms.” *Mechanisms of Ageing and Development* 127.10 (2006), pp. 786–793. DOI: `10.1016/j.mad.2006.07.004`.

Joseleau-petit, L. E., D. Vinella, and R. D’Ari. “Metabolic Alarms and Cell Division in *Escherichia coli*”. *Journal of Bacteriology* 181.1 (1999), pp. 9–14.

Jung, S.-C., C. L. Smith, K.-S. Lee, M.-E. Hong, D.-H. Kweon, G. Stephanopoulos, and Y.-S. Jin. “Restoration of growth phenotypes of *Escherichia coli* DH5alpha in minimal media through reversal of a point mutation in *purB*.” *Applied and Environmental Microbiology* 76.18 (2010), pp. 6307–9. DOI: `10.1128/AEM.01210-10`.

Kæ rn, M., T. C. Elston, W. J. Blake, and J. J. Collins. “Stochasticity in Gene Expression: from Theories to Phenotypes”. *Nature Reviews* 6.June (2005), pp. 451–464. DOI: `10.1038/nrg1615`.

Kandhavelu, M., H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-price, O. Yli-Harja, and A. S. Ribeiro. “In vivo kinetics of transcription initiation of the lar promoter in *Escherichia coli.* Evidence for a sequential mechanism with two rate-limiting steps”. *BMC Systems Biology* 5.1 (2011), p. 149. DOI: `10.1186/1752-0509-5-149`.

Kandhavelu, M., J. Lloyd-Price, A. Gupta, A.-B. Muthukrishnan, O. Yli-Harja, and A. S. Ribeiro. “Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter.” *FEBS Letters* 586.21 (2012), pp. 3870–3875. DOI: `10.1016/j.febslet.2012.09.014`.

Kandhavelu, M., A. Häkkinen, O. Yli-Harja, and A. S. Ribeiro. “Single-molecule dynamics of transcription of the lar promoter.” *Physical Biology* 9.2 (2012), p. 026004. DOI: `10.1088/1478-3975/9/2/026004`.

Kapuscinski, J. “DAPI: a DNA-specific fluorescent probe.” *Biotechnic & Histochemistry* 70.5 (1995), pp. 220–33.

Kennell, D. and H. Riezman. “Transcription and translation initiation frequencies of the *E. coli* lac operon”. *Journal of Molecular Biology* 114 (1977), pp. 1–21.

Kerr, R. A., H. Levine, T. J. Sejnowski, and W.-J. Rappel. “Division accuracy in a stochastic model of Min oscillations in *Escherichia coli.*” *Proceedings of the National Academy of Sciences of the United States of America* 103.2 (2006), pp. 347–352. DOI: `10.1073/pnas.0505825102`.

Knoblich, J. A. “Mechanisms of asymmetric stem cell division.” *Cell* 132.4 (2008), pp. 583–97. DOI: `10.1016/j.cell.2008.02.007`.

Koch, A. L. “On evidence supporting a deterministic process of bacterial growth.” *Journal of General Microbiology* 43 (1966), pp. 1–5.

Koch, A. L. and M Schaechter. “A model for statistics of the cell division process.” *Journal of General Microbiology* 29 (1962), pp. 435–54.

Kubitschek, H. “Cell Volume Increase in *Escherichia coli* after Shifts to Richer Media”. *Journal of Bacteriology* 172.1 (1990), pp. 94–101. DOI: `10.1016/S0006-3495(68)86521-X`.

Kuhlman, T. E. and E. C. Cox. “Gene location and DNA density determine transcription factor distributions in *Escherichia coli*”. *Molecular Systems Biology* 8 (2012), p. 610. DOI: `10.1038/msb.2012.42`.

Lahav, G., N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz, and U. Alon. “Dynamics of the p53-Mdm2 feedback loop in individual cells.” *Nature Genetics* 36.2 (2004), pp. 147–50. DOI: `10.1038/ng1293`.

Lele, U. N., U. I. Baig, and M. G. Watve. “Phenotypic plasticity and effects of selection on cell division symmetry in *Escherichia coli.*” *PloS One* 6.1 (2011), e14516. DOI: `10.1371/journal.pone.0014516`.

Lin, E. C. C., Y Hirota, and F Jacob. “On the Process of Cellular Division in Escherichia coli VI . Use of a Methocel-Autoradiographic Method for the Study of Cellular Division in *Escherichia coli*”. *Journal of Bacteriology* 108.I (1971), pp. 375–385.

Lindner, A. B. and A. Demarez. “Protein aggregation as a paradigm of aging.” *Biochimica et biophysica acta* 1790.10 (2009), pp. 980–96. DOI: `10.1016/j.bbagen.2009.06.005`.

Lindner, A. B., R. Madden, A. Demarez, and E. J. Stewart. “Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation”. *Proceedings of the National Academy of Sciences of the United States of America* 105.8 (2008), pp. 3076–3081.

Llopis, P. M., A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-wagner. “Spatial organization of the flow of genetic information in bacteria”. *Nature* 466.7302 (2010), pp. 77–81. DOI: `10.1038/nature09152`.

Lloyd-Price, J., A. Häkkinen, M. Kandhavelu, I. J. Marques, S. Chowdhury, E. Lihavainen, O. Yli-Harja, and A. S. Ribeiro. “Asymmetric disposal of individual protein aggregates in *Escherichia coli*, one aggregate at a time.” *Journal of Bacteriology* 194.7 (2012), pp. 1747–1752. DOI: `10.1128/JB.06500-11`.

Lloyd-Price, J., H. Tran, and A. S. Ribeiro. “Dynamics of small genetic circuits subject to stochastic partitioning in cell division.” *Journal of Theoretical Biology* 356 (2014), pp. 11–9. DOI: `10.1016/j.jtbi.2014.04.018`.

Lloyd-Price, J., M. Lehtivaara, M. Kandhavelu, S. Chowdhury, A.-B. Muthukrishnan, O. Yli-Harja, and A. S. Ribeiro. “Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*”. *Molecular BioSystems* 8 (2012), pp. 565–571. DOI: `10.1039/c1mb05100h`.

Lutz, R. and H. Bujard. “Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR / O , the TetR / O and AraC / I 1 -I 2 regulatory elements”. *Nucleic Acids Research* 25.6 (1997), pp. 1203–1210.

Lutz, R., T. Lozinski, T. Ellinger, and H. Bujard. “Dissecting the functional program of *Escherichia coli* promoters : the combined mode of action of Lac repressor and AraC activator”. *Nucleic Acids Research* 29.18 (2001), pp. 3873–3881.

Ma, X. and W. Margolin. “Genetic and Functional Analyses of the Conserved C-Terminal Core Domain of *Escherichia coli* FtsZ”. *Journal of Bacteriology* 181.24 (1999), pp. 7531–7544.

Ma, X., D. W. Ehrhardt, and W. Margolin. “Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living *Escherichia coli* cells by using green fluorescent protein.” *Proceedings of the National Academy of Sciences of the United States of America* 93.23 (1996), pp. 12998–13003.

Maisonneuve, E., B. Ezraty, and S. Dukan. “Protein aggregates: an aging factor involved in cell death.” *Journal of Bacteriology* 190.18 (2008), pp. 6070–6075. DOI: `10.1128/JB.00736-08`.

Mäkelä, J., M. Kandhavelu, S. M. D. Oliveira, J. G. Chandraseelan, J. Lloyd-Price, J. Peltonen, O. Yli-Harja, and A. S. Ribeiro. “In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter.” *Nucleic Acids Research* 41.13 (2013), pp. 6544–52. DOI: `10.1093/nar/gkt350`.

Mäkelä, J., J. Lloyd-price, O. Yli-Harja, and A. S. Ribeiro. “Stochastic sequence-level model of coupled transcription and translation in prokaryotes”. *BMC Bioinformatics* 12.121 (2011), pp. 1–13. DOI: `10.1186/1471-2105-12-121`.

Männik, J., F. Wu, F. J. H. Hol, P. Bisicchia, D. J. Sherratt, J. E. Keymer, and C. Dekker. “Robustness and accuracy of cell division in *Escherichia coli* in diverse cell shapes.” *Proceedings of the National Academy of Sciences of the United States of America* 109.18 (2012), pp. 6957–69562. DOI: `10.1073/pnas.1120854109`.

Marr, A. G., R. J. Harvey, and W. C. Trentini. “Growth and division of *Escherichia coli*.” *Journal of Bacteriology* 91.6 (1966), pp. 2388–9.

Marshall, W. F. “Stability and robustness of an organelle number control system: modeling and measuring homeostatic regulation of centriole abundance.” *Biophysical Journal* 93.5 (2007), pp. 1818–33. DOI: `10.1529/biophysj.107.107052`.

Mcadams, H. H. and A. Arkin. “Stochastic mechanisms in gene expression”. *Proceedings of the National Academy of Sciences of the United States of America* 94 (1997), pp. 814–819.

Mcclure, R. “Mechanism and Control of Transcription Initiation in Prokaryotes”. *Annual Reviews Biochemistry* 54 (1985), pp. 171–204.

McCollum, J. M., G. D. Peterson, C. D. Cox, M. L. Simpson, and N. F. Samatova. “The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior.” *Computational Biology and Chemistry* 30.1 (2006), pp. 39–49. DOI: `10.1016/j.compbiolchem.2005.10.007`.

McQuarrie, D. “Stochastic approach to chemical kinetics”. *Journal of Applied Probability* 4.3 (1967), pp. 413–478.

Meinhardt, H. and P. A. de Boer. “Pattern formation in *Escherichia coli*: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site.” *Proceedings of the National Academy of Sciences of the United States of America* 98.25 (2001), pp. 14202–14207. DOI: `10.1073/pnas.251216598`.

Mondal, J., B. P. Bratton, Y. Li, A. Yethiraj, and J. C. Weisshaar. “Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells.” *Biophysical Journal* 100.11 (2011), pp. 2605–13. DOI: `10.1016/j.bpj.2011.04.030`.

Mulder, E. and C. L. Woldringh. “Actively Replicating Nucleoids Influence Positioning of Division Sites in *Escherichia coli* Filaments Forming Cells Lacking DNA”. *Journal of Bacteriology* 171.8 (1989), pp. 4303–4314.

Muthukrishnan, A.-B., M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro. “Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells.” *Nucleic Acids Research* 40.17 (2012), pp. 8472–83. DOI: `10.1093/nar/gks583`.

Neumuller, R. A. and J. A. Knoblich. “Dividing cellular asymmetry : asymmetric cell division and its implications for stem cells and cancer”. *Genes & development* 23 (2009), pp. 2675–2699. DOI: `10.1101/gad.1850809.within`.

Nyström, T. “A bacterial kind of aging.” *PLoS Genetics* 3.12 (2007), e224. DOI: `10.1371/journal.pgen.0030224`.

Osella, M., E. Nugent, and M. Cosentino Lagomarsino. “Concerted control of *Escherichia coli* cell division.” *Proceedings of the National Academy of Sciences of the United States of America* 111.9 (2014), pp. 3431–5. DOI: `10.1073/pnas.1313715111`.

Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. “Regulation of noise in the expression of a single gene.” *Nature Genetics* 31.1 (2002), pp. 69–73. DOI: `10.1038/ng869`.

Parry, B. R., I. V. Surovtsev, M. T. Cabeen, C. S. O’Hern, E. R. Dufresne, and C. Jacobs-Wagner. “The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity.” *Cell* 156 (2013), pp. 1–12. DOI: `10.1016/j.cell.2013.11.028`.

Paulsson, J. “Models of stochastic gene expression”. *Physics of Life Reviews* 2.2 (2005), pp. 157–175. DOI: `10.1016/j.plrev.2005.03.003`.

Paun, G. “From cells to computers : computing with membranes (P systems)”. *Biosystems* 59 (2001), pp. 139–158.

Peabody, D. S. “The RNA binding site of bacteriophage MS2 coat protein”. *The EMBO Journal* 12.2 (1993), pp. 595–600.

Peccoud, J. “Markovian Modelling of Gene Product Synthesis”. *Theoretical Population Biology* 48 (1995), pp. 222–234.

Pedraza, J. M. and J. Paulsson. “Effects of molecular memory and bursting on fluctuations in gene expression.” *Science* 319.5861 (2008), pp. 339–43. DOI: `10.1126/science.1144331`.

Portnoy, S. and R. Koenker. “The Gaussian Hare and the Laplacian Tortoise : Computability of Squared-Error versus Absolute-Error Estimators”. *Statistical Science* 12.4 (1997), pp. 279–300.

Rajala, T., A. Häkkinen, S. Healy, O. Yli-Harja, and A. S. Ribeiro. “Effects of transcriptional pausing on gene expression dynamics.” *PLoS Computational Biology* 6.3 (2010), e1000704. DOI: `10.1371/journal.pcbi.1000704`.

Reyes-Lamothe, R., T. Tran, D. Meas, L. Lee, A. M. Li, D. J. Sherratt, and M. E. Tolmasky. “High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division.” *Nucleic Acids Research* 42.2 (2013), pp. 1042–1051. DOI: `10.1093/nar/gkt918`.

Ribeiro, A. S. “Stochastic and delayed stochastic models of gene expression and regulation.” *Mathematical Biosciences* 223.1 (2010), pp. 1–11. DOI: `10.1016/j.mbs.2009.10.007`.

Rokney, A., M. Shagan, M. Kessel, Y. Smith, I. Rosenshine, and A. B. Oppenheim. “*E. coli* transports aggregated proteins to the poles by a specific and energy-dependent process.” *Journal of Molecular Biology* 392.3 (2009), pp. 589–601. DOI: `10.1016/j.jmb.2009.07.009`.

Roussel, M. R. and R. Zhu. “Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression”. *Physical Biology* 3 (2006), pp. 274–284. DOI: `10.1088/1478-3975/3/4/005`.

Ruusuvuori, P., T. Aijö, S. Chowdhury, C. Garmendia-Torres, J. Selinummi, M. Birbaumer, A. M. Dudley, L. Pelkmans, and O. Yli-Harja. “Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images.” *BMC Bioinformatics* 11 (2010), p. 248. DOI: `10.1186/1471-2105-11-248`.

Schaechter, B. Y. M. and N Kjeldgaard. “Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium”. *Journal of General Microbiology* 19 (1958), pp. 592–606.

Shehata, T. E. and A. G. Marr. “Effect of temperature on the size of Escherichia coli cells .” *Journal of Bacteriology* 124.2 (1975), pp. 857–862.

Shimamoto, N. “One-dimensional Diffusion of Proteins along DNA: Its Biological and Chemical Significance Revealed by Single-Molecule Measurements”. *Journal of Biological Chemistry* 274.22 (1999), pp. 15293–15296. DOI: `10.1074/jbc.274.22.15293`.

Slepoy, A., A. P. Thompson, and S. J. Plimpton. “A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks.” *The Journal of Chemical Physics* 128.20 (2008), p. 205101. DOI: `10.1063/1.2919546`.

Sorensen, M. A., C. G. Kurland, and S. Pedersen. “Codon Usage Determines Translation Rate in *Escherichia coli*”. *Journal of Molecular Biology* 207 (1989), pp. 365–377.

Soti, C. and P. Csermely. “Aging and molecular chaperones”. *Experimental Gerontology* 38.10 (2003), pp. 1037–1040. DOI: `10.1016/S0531-5565(03)00185-2`.

Spicher, A., O. Michel, M. Cieslak, J.-l. Giavitto, and P. Prusinkiewicz. “Stochastic P systems and the simulation of biochemical processes with dynamic compartments”. *Biosystems* 91 (2008), pp. 458–472. DOI: `10.1016/j.biosystems.2006.12.009`.

Spokoini, R., O. Moldavski, Y. Nahmias, J. L. England, M. Schuldiner, and D. Kaganovich. “Confinement to organelle-associated inclusion structures mediates asymmetric inheritance of aggregated protein in budding yeast.” *Cell Reports* 2.4 (2012), pp. 738–47. DOI: `10.1016/j.celrep.2012.08.024`.

St-pierre, F. and D. Endy. “Determination of cell fate selection during phage”. *Proceedings of the National Academy of Sciences of the United States of America* 105.52 (2008), pp. 20705–10.

Stewart, E. J., R. Madden, G. Paul, and F. Taddei. “Aging and death in an organism that reproduces by morphologically symmetric division.” *PLOS Biology* 3.2 (2005), pp. 0295–0300. DOI: `10.1371/journal.pbio.0030045`.

Sullivan, S. M. and J. R. Maddock. “Bacterial division: Finding the dividing line.” *Current Biology* 10.6 (2000), R249–52.

Svenningsen, S. L., N. Costantino, D. L. Court, and S. Adhya. “On the role of Cro in lambda prophage induction”. *Proceedings of the National Academy of Sciences of the United States of America* 102.12 (2005), pp. 4465–4469.

Taniguchi, Y., P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. “Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells.” *Science* 329 (2010), pp. 533–8. DOI: `10.1126/science.1188308`.

Tao, H., C. Bausch, C. Richmond, R Frederick, T. Conway, H. A. N. Tao, and F. R. Blattner. “Functional Genomics : Expression Analysis of *Escherichia coli* Growing on Minimal and Rich Media”. *Journal of Bacteriology* 181.20 (1999), pp. 6425–6440.

Thanbichler, M. “Synchronization of Chromosome Dynamics and Cell Division in Bacteria”. *Cold Spring Harbor Perspectives in Biology* 2.a000331 (2009), pp. 1–15.

Tokunaga, M., N. Imamoto, and K. Sakata-sogawa. “Highly inclined thin illumination enables clear single-molecule imaging in cells”. *Nature Methods* 5.2 (2008), pp. 159–161. DOI: `10.1038/NMETH.1171`.

Trueba, F. J. and C. L. Woldringh. “Changes in cell diameter during the division cycle of *Escherichia coli.*” *Journal of Bacteriology* 142.3 (1980), pp. 869–78.

Trueba, F. J., E. A. van Spronsen, J. Traas, and C. L. Woldringh. “Effects of temperature on the size and shape of *Escherichia coli* cells.” *Archives of Microbiology* 131.2 (1982), pp. 235–40.

Tyedmers, J., A. Mogk, and B. Bukau. “Cellular strategies for controlling protein aggregation.” *Nature Reviews. Molecular Cell Biology* 11.11 (2010), pp. 777–788. DOI: `10.1038/nrm2993`.

Valkenburg, J. A. and C. L. Woldringh. “Phase separation between nucleoid and cytoplasm in *Escherichia coli* as defined by immersive refractometry.” *Journal of Bacteriology* 160.3 (1984), pp. 1151–7.

Wang, X., C. Possoz, and D. J. Sherratt. “Dancing around the divisome: asymmetric chromosome segregation in *Escherichia coli.*” *Genes & Development* 19.19 (2005), pp. 2367–2377. DOI: `10.1101/gad.345305`.

Wang, X., J. Huang, A. Mukherjee, C. Cao, and J. Lutkenhaus. “Analysis of the Interaction of FtsZ with Itself , GTP , and FtsA”. *Journal of Bacteriology* 179.17 (1997), pp. 5551–5559.

Weiss, D. S., J. C. Chen, J.-m. Ghigo, J. Beckwith, and D. Boyd. “Localization of FtsI (PBP3) to the Septal Ring Requires Its Membrane Anchor , the Z Ring , FtsA , FtsQ , and FtsL”. *Journal of Bacteriology* 181.2 (1999), pp. 508–520.

Winkler, J., A. Seybert, L. König, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk, and B. Bukau. “Quantitative and spatiotemporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing.” *The EMBO Journal* 29.5 (2010), pp. 910–923. DOI: `10.1038/emboj.2009.412`.

Woldringh, C. L. “Is *Escherichia coli* getting old?” *BioEssays* 27.8 (2005), pp. 770–4. DOI: `10.1002/bies.20271`.

Woldringh, C. L., A Zaritsky, and N. B. Grover. “Nucleoid partitioning and the division plane in *Escherichia coli.*” *Journal of Bacteriology* 176.19 (1994), pp. 6030–8.

Woldringh, C. L., E. Mulder, J. Valkenburg, F. Wientjes, Zaritsky, A., and N. Nanninga. “Role of the nucleoid in the toporegulation of division.” *Research in Microbiology* 141 (1990), pp. 39–49.

Woldringh, C. L., E. Mulder, P. G. Huls, and N. Vischer. “Toporegulation of bacterial division according to the nucleoid occlusion model.” *Research in Microbiology* 142 (1991), pp. 309–320.

Yarchuk, O., J. Guillerez, and M. Dreyfus. “Interdependence of Translation , Transcription and mRNA Degradation in the lacZ Gene”. *Journal of Molecular Biology* 226 (1992), pp. 581–596.

Yu, J., J. Xiao, X. Ren, K. Lao, and X. S. Xie. "Probing gene expression in live cells, one protein molecule at a time." *Science* 311.5767 (2006), pp. 1600–3. DOI: `10.1126/science.1119623`.

Yu, X. C. and W Margolin. "FtsZ ring clusters in min and partition mutants: role of both the Min system and the nucleoid in regulating FtsZ ring localization." *Molecular Microbiology* 32.2 (1999), pp. 315–326.

Zeng, L., S. O. Skinner, C. Zong, J. Sippy, M. Feiss, and I. Golding. "Decision making at a subcellular level determines the outcome of bacteriophage infection." *Cell* 141.4 (2010), pp. 682–91. DOI: `10.1016/j.cell.2010.03.034`.

Zhang, P., C. M. Khursigara, L. M. Hartnell, and S. Subramaniam. "Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy". *Proceedings of the National Academy of Sciences of the United States of America* 104.10 (2007), pp. 3777–81. DOI: `10.1073/pnas.0610106104`.

Zink, D., N. Sadoni, and E. Stelzer. "Visualizing chromatin and chromosomes in living cells". *Methods* 29 (2003), pp. 42–50.

Zon, J. van and P. ten Wolde. "Simulating Biochemical Networks at the Particle Level and in Time and Space: Green's Function Reaction Dynamics". *Physical Review Letters* 94.12 (2005), p. 128103. DOI: `10.1103/PhysRevLett.94.128103`.

Publications

Publication I

J. Lloyd-Price, **A. Gupta**, and A. S. Ribeiro (2012). SGNS2: A Compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations. *Bioinformatics*, **28**(22):3004-5.

BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 22 2012, pages 3004–3005
doi:10.1093/bioinformatics/bts556

Systems biology Advance Access publication September 26, 2012

SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations

Jason Lloyd-Price*, Abhishekh Gupta and Andre S. Ribeiro
Department of Signal Processing, Tampere University of Technology, 33101 Tampere, Finland

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Cell growth and division affect the kinetics of internal cellular processes and the phenotype diversity of cell populations. Since the effects are complex, e.g. different cellular components are partitioned differently in cell division, to account for them in silico, one needs to simulate these processes in great detail.

Results: We present SGNS2, a simulator of chemical reaction systems according to the Stochastic Simulation Algorithm with multi-delayed reactions within hierarchical, interlinked compartments which can be created, destroyed and divided at runtime. In division, molecules are randomly segregated into the daughter cells following a specified distribution corresponding to one of several partitioning schemes, applicable on a per-molecule-type basis. We exemplify its use with six models including a stochastic model of the disposal mechanism of unwanted protein aggregates in *Escherichia coli*, a model of phenotypic diversity in populations with different levels of synchrony, a model of a bacteriophage's infection of a cell population and a model of prokaryotic gene expression at the nucleotide and codon levels.

Availability: SGNS2, instructions and examples available at www.cs.tut.fi/~lloydpri/sgns2/ (open source under New BSD license).

Contact: jason.lloyd-price@tut.fi

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 31, 2012; revised on August 30, 2012; accepted on September 6, 2012

1 INTRODUCTION

Recent evidence suggests that even in cellular organisms whose division is morphologically symmetric, there are a number of asymmetries between daughter cells. These arise, among other things, from the stochasticity in the partitioning of components in division (Huh and Paulsson, 2011) and from biased partitioning schemes for some components. For example, in *Escherichia coli*, unwanted protein aggregates follow biased partitioning schemes dependent on the age of the daughter cells' poles (Lindner *et al*., 2008).

These and other recent findings suggest that the phenotypic diversity of cell populations, among other factors, depends on errors and biases in the partitioning of RNA, proteins and other molecules. This is of relevance since most RNAs exist in small numbers (Bernstein *et al*., 2002) and small fluctuations in these numbers can alter the behavior of genetic circuits (Ribeiro and Kauffman, 2007) and trigger visible phenotype changes (Choi *et al*., 2008).

These sources of phenotypic heterogeneity are difficult to distinguish from, e.g. noise in gene expression (Huh and Paulsson, 2011). Although some effects can be assessed analytically (Huh and Paulsson, 2011), others are too complex and must be assessed numerically. A simulator is thus needed that accounts for noise and delays (Kandhavelu *et al*., 2012) in gene expression and for compartmentalization of processes and components.

Presently, simulators of the dynamics of noisy biochemical systems rely on the Stochastic Simulation Algorithm (SSA) (Gillespie, 1977), e.g. (Blakes *et al*., 2011; Hattne *et al*., 2005; Hoops *et al*., 2006; Lok and Brent, 2005). Some support compartmentalization, simulating reaction-diffusion systems in either static (Hattne *et al*., 2005) or dynamically sized compartments (Blakes *et al*., 2011; Versari and Busi, 2008). Others support rule-based creation of reactions at runtime (Lok and Brent, 2005; Spicher *et al*., 2008), and thus can simulate a dynamic cell population. Very few support delays on the release into the system of one or more products of a reaction (Roussel and Zhu, 2006). These delays are essential to accurately model the kinetics of some processes, e.g. transcription, as RNA production is mostly regulated by the duration of events in transcription initiation (Muthukrishnan *et al*., 2012).

Here, we present SGNS2, an extension of SGN Sim (Ribeiro and Lloyd-Price, 2007) that incorporates dynamic compartments and multiple partitioning distributions at cell division, applicable on a per-molecule-type basis.

2 METHODS

SGNS2 is an extension of SGNS, the stochastic simulator of SGNSim (Ribeiro and Lloyd-Price, 2007). It contains all the features of SGNS, such as reactions with multi-delayed events. The two key additions in SGNS2 are (i) it supports dynamic, interlinked, hierarchical compartments and (ii) it supports multiple molecule and compartment partitioning schemes, applicable on a per-molecule-type basis. The novel features considerably extend the class of models that can be simulated.

SGNS2 uses a modified version of the Next Reaction Method (NRM) (Gibson and Bruck, 2000). Namely, the NRM was adapted to stochastic P-systems (Spicher *et al*., 2008) by using a hierarchy of indexed priority queues (IPQ, an ordered list of elements that keep track of their position in the list) and further modified to allow multiple delays in reactions. The IPQ data structure, implemented with a binary heap, is described in Gibson and Bruck (2000). We use a separate IPQ for each compartment, which publish a 'tentative next event time' to an overall IPQ which determines the next event time in the entire simulation. We optimize the update step when molecule populations in a parent compartment change by using a hierarchical refinement of the IPQs with appropriate scaling of tentative firing times (see Supplementary Material). Delayed events were implemented by creating wait lists, implemented by binary heap-based priority queues, whose earliest event is published to each compartment's indexed priority queue. The simulation's elementary SSA steps scale logarithmically with

---

*To whom correspondence should be addressed.

Downloaded from http://bioinformatics.oxfordjournals.org/ at Tampere University of Technology Library on November 25, 2014

© The Author 2012. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

the number of reactions, compartments and delayed events, allowing complex models to be simulated in reasonable time.

To simulate cell division, we introduced a special reaction event, whose timing follows the SSA rules. When executed, instead of subtracting substrates from the system, a random number is generated based on one of the several partitioning distributions available, including some of those listed in Huh and Paulsson (2011). Each of these mimics a specific molecule partitioning process during cell division. SGNS2 allows both biased and unbiased partitioning of molecules and sub-compartments. The results of these events can be instantaneous or be placed on the wait list. Compartment division and molecule partitioning are represented in the following form:

$$\mathrm{split}(p) : \mathrm{Protein@Cell} \overset{c_\mu}{\rightarrow} \mathrm{@Cell+} : \mathrm{Protein@Cell}$$

When this reaction occurs, a new cell compartment is created (@Cell in the product list). Proteins in the original cell are partitioned according to a biased binomial partitioning scheme. In this, each protein is independently partitioned into the new cell with probability p. Other common partitioning distributions include the independent partitioning of molecules into daughter cells with random (beta-distributed) sizes and the binding of molecules to spindle binding sites which are segregated evenly between daughter cells such as during mitosis. Available distributions are listed in the manual.

SGNS2 is a command line utility, designed to fit into a toolchain, supporting various input and output formats. Input can be specified in two formats: SBML (Hucka *et al*., 2003) and SGNSim's native format (Ribeiro and Lloyd-Price, 2007). A subset of SBML Core level 3 version 1 is supported, allowing simulation of most SBML models. Output can be in csv, tsv or in binary format. A text editor may be used to write models in SGNSim format. SBML-based graphical interfaces such as CellDesigner (Funahashi *et al*., 2008) or Cytoscape (Smoot *et al*., 2011) may be used to manage SBML models. The results of simulations are interpretable by programs like MATLAB, R or Excel. An example of running a model in SGNSim format of a growing cell population is shown in Supplementary Figure S1.

3 DISCUSSION

SGNS2 is the first stochastic simulator that includes multi-delayed events, dynamic compartments and molecule partitioning schemes in division. To test its correctness, we simulated models from the Discrete Stochastic Model Test Suite (Evans *et al*., 2008). All showed the expected behavior (Supplementary Figs S2 and S3).

SGNS2, though making use of existing and slightly modified versions of existing algorithms, can simulate an array of biological processes not previously possible. For example, it is ideal for simulating gene expression at the nucleotide and codon levels (see 'Availability' section) and study features such as how events in transcription elongation affect protein production kinetics (Mäkelä *et al*., 2011).

SGNS2 is also suited to study partitioning in cell division, which affects aging, among other processes, and is of particular relevance when modeling populations over multiple generations. To exemplify this, we modeled the biased partitioning of protein aggregates in *E. coli*, known to accumulate in cells with older poles, reducing vitality (Lindner *et al*., 2008). The results in Supplementary Figure S4 agree with measurements (Stewart *et al*., 2005). We further studied how cell-cycle synchrony affects the population-level statistics of RNA numbers [Supplementary Fig. S5, in agreement with measurements in Lloyd-Price *et al.* (2012)]. As a side note, we expect the partitioning of RNA and proteins to affect the dynamics of genetic circuits, particularly the stability of their noisy attractors (Ribeiro and Kauffman, 2007). To further demonstrate the simulator's utility, we modeled the viral infection of a dynamic bacterial population.

In conclusion, SGNS2 provides novel functionalities to model and simulate cellular processes not previously possible, as seen from the examples. In general, SGNS2 enables the modeling of stochastic processes in live cells that require compartmentalization, multi-delayed complex processes and complex stochastic partitioning schemes at a per-molecule type in cell division. These features are necessary to study *in silico*, among other phenomena, phenotypic diversity in cell populations.

Funding: Work supported by Academy of Finland (126803).

Conflict of Interest: none declared.

REFERENCES

Bernstein,J.A. *et al.* (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 9697–9702.

Blakes,J. *et al.* (2011) The infobiotics workbench: an integrated in silico modelling platform for systems and synthetic biology. *Bioinformatics*, **27**, 3323–3324.

Choi,P.J. *et al.* (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, **322**, 442–446.

Evans,T. *et al.* (2008) The SBML discrete stochastic models test suite. *Bioinformatics*, **25**, 285–286.

Funahashi,A. *et al.* (2008) Celldesigner 3.5: a versatile modeling tool for biochemical networks. *Proc. IEEE*, **96**, 1254–1265.

Gibson,M.A. and Bruck,J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, **104**, 1876–1889.

Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.

Hattne,J. *et al.* (2005) Stochastic reaction-diffusion simulation with mesord. *Bioinfomatics*, **21**, 2923–2924.

Hoops,S. *et al.* (2006) Copasi a complex pathway simulator. *Bioinformatics*, **22**, 3067–3074.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Huh,D. and Paulsson,J. (2011) Random partitioning of molecules at cell division. *Proc. Natl Acad. Sci. USA*, **108**, 15004–15009.

Kandhavelu,M. *et al.* (2012) Random partitioning of molecules at cell division. *Phys. Biol.*, **9**, 026004.

Lindner,A.B. *et al.* (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl Acad. Sci. USA*, **105**, 3076–3081.

Lloyd-Price,J. *et al.* (2012) Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of Escherichia coli. *Mol. Biosys.*, **8**, 565–571.

Lok,L. and Brent,R. (2005) Automatic generation of cellular reaction networks with moleculizer 1.0. *Nat. Biotech.*, **23**, 131–36.

Mäkelä,J. *et al.* (2011) Stochastic sequence-level model of coupled transcription and translation in prokaryotes. *BMC Bioinf.*, **12**, 121.

Muthukrishnan,A.-B. *et al.* (2012) Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Res.*, **40**, 8472–8483.

Ribeiro,A.S. and Kauffman,S. (2007) Noisy attractors and ergodic sets in models of gene regulatory networks. *J. Theor. Biol.*, **247**, 743–755.

Ribeiro,A.S. and Lloyd-Price,J. (2007) SGN sim, a stochastic genetic networks simulator. *Bioinformatics*, **23**, 777–779.

Roussel,M.R. and Zhu,R. (2006) Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys. Biol.*, **3**, 274–284.

Smoot,M. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Spicher,A. *et al.* (2008) Stochastic P systems and the simulation of biochemical processes with dynamic compartments. *Biosystems*, **91**, 458–472.

Stewart,E.J. *et al.* (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, **3**, e45.

Versari,C. and Busi,N. (2008) Efficient stochastic simulation of biological systems with multiple variable volumes. *Elec. Notes Theor. Comp. Sci.*, **194**, 165–180.

Downloaded from http://bioinformatics.oxfordjournals.org/ at Tampere University of Technology Library on November 25, 2014

Supplement to “SGNS2: A Compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations”

Jason Lloyd-Price, Abhishekh Gupta, and Andre S. Ribeiro

Implementation Details

SGNS2 uses the Next Reaction Method[1] (NRM) to simulate the dynamics according to the Stochastic Simulation Algorithm[2] (SSA). This method is an efficient implementation of the SSA, which begins by randomizing a ‘next firing time’ for each possible reaction in the system and storing these tentative reaction times in an indexed priority queue (IPQ). The reaction with the soonest tentative firing time is then taken from the queue, performed, and its next firing time is re-randomized. Any reaction whose propensity depends on the set of molecule species affected by this reaction then have their tentative firing times transformed to follow the new distribution of firing times prescribed by the Chemical Master Equation. Then, their positions in the priority queue are updated. These reactions are determined by pre-generating the graph depicting which reactions potentially affect the propensities of other reactions (the reaction dependency graph). We implement the NRM’s IPQs using array-based binary heaps, which provide logarithmic scaling of the runtime with the number of reactions for the SSA steps in a sparsely-coupled model (i.e. a model whose reaction dependency graph is sparse).

To allow compartments to be quickly created and destroyed, a separate IPQ is created for each compartment. These IPQs are inserted into a higher-level IPQ which acts as a “Next Compartment Method”, allowing us to determine which compartment the next reaction will occur in, in logarithmic time with the number of compartments. Creating/destroying compartments is then done by constructing/destructing these IPQs and inserting/removing them from the overall IPQ. In this arrangement, compartment creation takes $O(\log C + R\log R)$ time, while compartment destruction takes $O(\log C)$ time, where R is the number of reactions in the new compartment and C is the current number of compartments in the simulation.

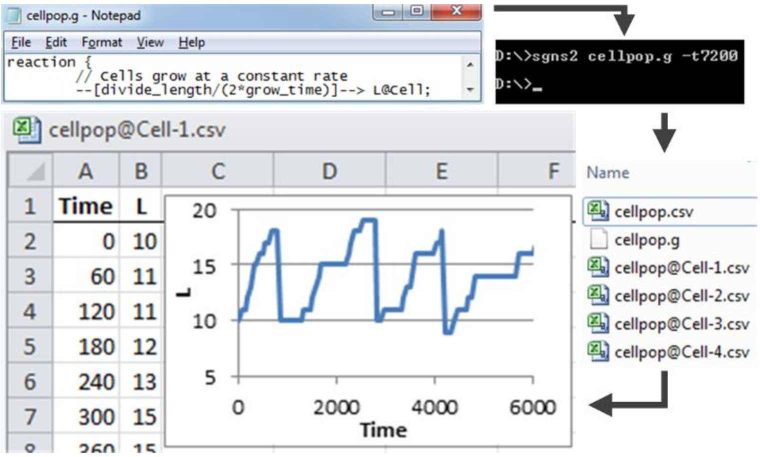
Communication between compartments is accomplished by reactions that affect molecules in both a ‘parent’ and a ‘child’ compartment. Since the propensity of each instance of such a reaction depends on the population of the reactant in the parent compartment, O(C) propensities must be recalculated when this quantity changes, an $O(C\log R)$ operation. Since each reactant of a reaction factors independently into the propensity of the reaction, the reactant in the parent compartment can be factored out from all of the instances of the reactions in the sub-compartments. This calculation is similar to the partial propensity methods [3]. To accomplish this without requiring an O(C) operation, we create a separate IPQ for the sub-compartment’s reaction instances in which the local simulation time, t_{sub}, is advanced such that $dt_{sub} = Xdt$, where X is the current population size of the reactant in the parent compartment and t is the global simulation's time variable. This sub-simulation then publishes a next firing time to the parent compartment's IPQ, adjusted according to the NRM’s propensity update formula. When the parent compartment reactant's population changes, only the adjusted next firing time must be recalculated and only one element of an IPQ may change position, reducing the cost of this operation to $O(\log R)$. SGNS2 assumes that there are no direct interactions between compartments at the same level of the hierarchy.

To include multi-delayed reactions as well, which are simulated according to the Delayed SSA[4], we implement a wait list using a binary heap-based priority queue. The transient nature of compartments makes it necessary for each to contain its own wait list. The earliest event in a compartment's wait list is then inserted into the compartment's IPQ. All operations on the wait lists are therefore $O(\log W + \log R + \log C)$, where W is the number of delayed events on that wait list. When a compartment is destroyed, all delayed events in that compartment are forgotten, assuring that no delayed molecules of that compartment are released following this event.

References

1. M. A. Gibson and J. Bruck, J. Phys. Chem. A 104, 1876, 2000.
2. D. T. Gillespie, J. Phys. Chem. 81, 2340, 1977.
3. R. Ramaswamy, N. González-Segredo, and I. F. Sbalzarini, J. Chem. Phys. 130, 244104, 2009.
4. M. R. Roussel and R. Zhu, Phys. Biol. 3, 274, 2006.

Supplementary Figures


cellpop.g - Notepad
File
Edit
Format
View
Help
cellpop@Cell-1.csv
A
B
C
D
E
F
20
15
10
5
0
2000
4000
6000
Time
L
Name

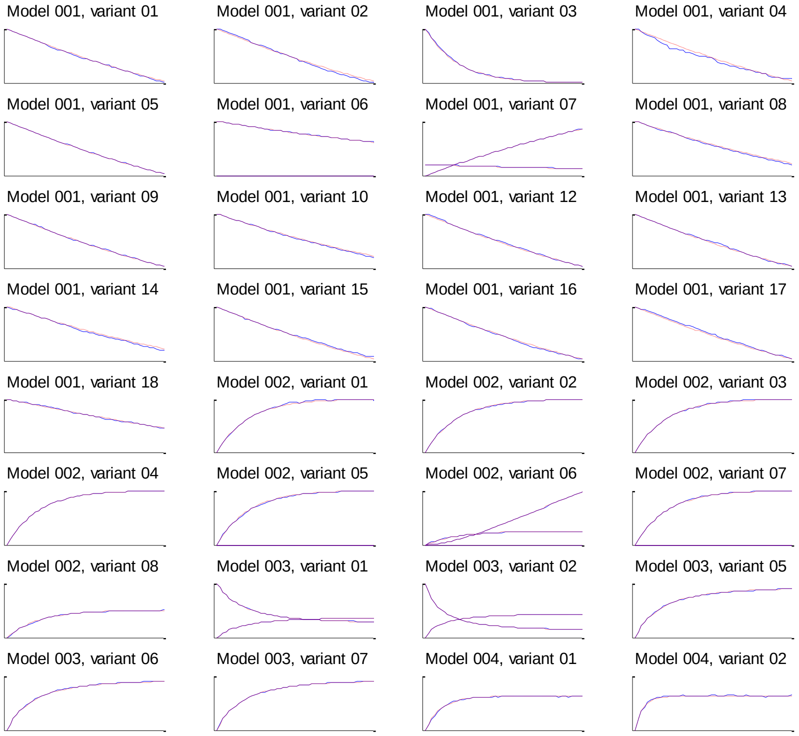
```
reaction {
        // Cells grow at a constant rate
        --[divide_length/(2*grow_time)]--> L@Cell;
```

```
D:\>sgns2 cellpop.g -t7200
D:\>_
```

	A	B
1	Time	L
2	0	10
3	60	11
4	120	11
5	180	12
6	240	13
7	300	15
8	360	15

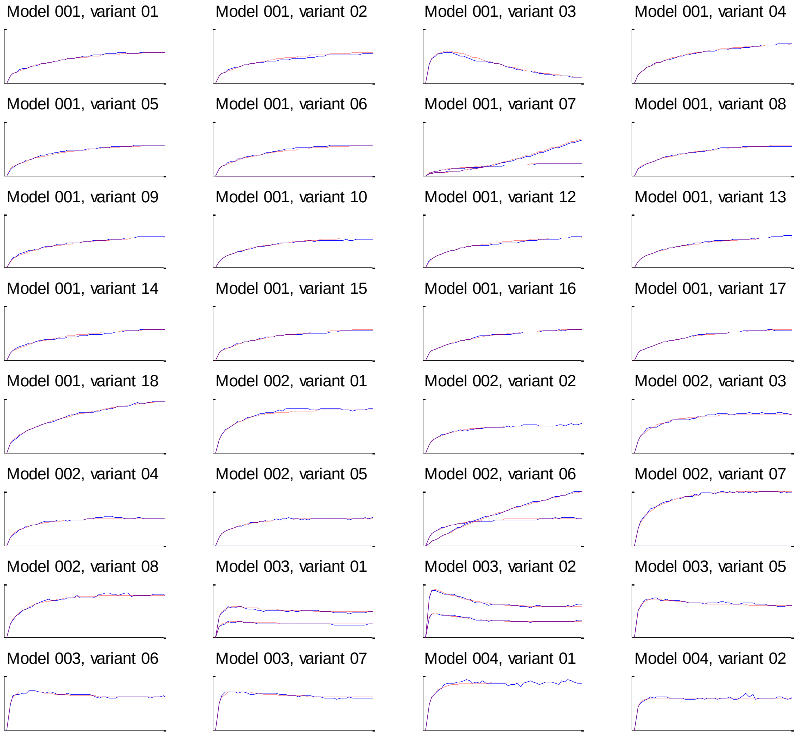
cellpop.csv
cellpop.g
cellpop@Cell-1.csv
cellpop@Cell-2.csv
cellpop@Cell-3.csv
cellpop@Cell-4.csv

Fig S1: Example of SGNS2 in use. A model is created in a text editor, here Notepad (upper left), and is simulated with SGNS2 (upper right). The csv files output (lower right) are loaded and analyzed in Excel (lower left).



Model 001, variant 01
Model 001, variant 02
Model 001, variant 03
Model 001, variant 04
Model 001, variant 05
Model 001, variant 06
Model 001, variant 07
Model 001, variant 08
Model 001, variant 09
Model 001, variant 10
Model 001, variant 12
Model 001, variant 13
Model 001, variant 14
Model 001, variant 15
Model 001, variant 16
Model 001, variant 17
Model 001, variant 18
Model 002, variant 01
Model 002, variant 02
Model 002, variant 03
Model 002, variant 04
Model 002, variant 05
Model 002, variant 06
Model 002, variant 07
Model 002, variant 08
Model 003, variant 01
Model 003, variant 02
Model 003, variant 05
Model 003, variant 06
Model 003, variant 07
Model 004, variant 01
Model 004, variant 02

Fig S2: Means of molecule populations over time for the models in the Discrete Stochastic Model Test Suite which do not use Rules or Events. Solid blue lines are the means of the results from 500 runs of SGNS2, while dashed red lines show the analytical solutions. The overlap of these lines results in a purple-like line.



Model 001, variant 01
Model 001, variant 02
Model 001, variant 03
Model 001, variant 04
Model 001, variant 05
Model 001, variant 06
Model 001, variant 07
Model 001, variant 08
Model 001, variant 09
Model 001, variant 10
Model 001, variant 12
Model 001, variant 13
Model 001, variant 14
Model 001, variant 15
Model 001, variant 16
Model 001, variant 17
Model 001, variant 18
Model 002, variant 01
Model 002, variant 02
Model 002, variant 03
Model 002, variant 04
Model 002, variant 05
Model 002, variant 06
Model 002, variant 07
Model 002, variant 08
Model 003, variant 01
Model 003, variant 02
Model 003, variant 05
Model 003, variant 06
Model 003, variant 07
Model 004, variant 01
Model 004, variant 02

Fig S3: Standard deviations of molecule populations over time for the models in the Discrete Stochastic Model Test Suite which do not use Rules or Events. Solid blue lines are the standard deviations of the results from 500 runs of SGNS2, while dashed red lines show the analytical solutions. The overlap of these lines results in a purple-like line.

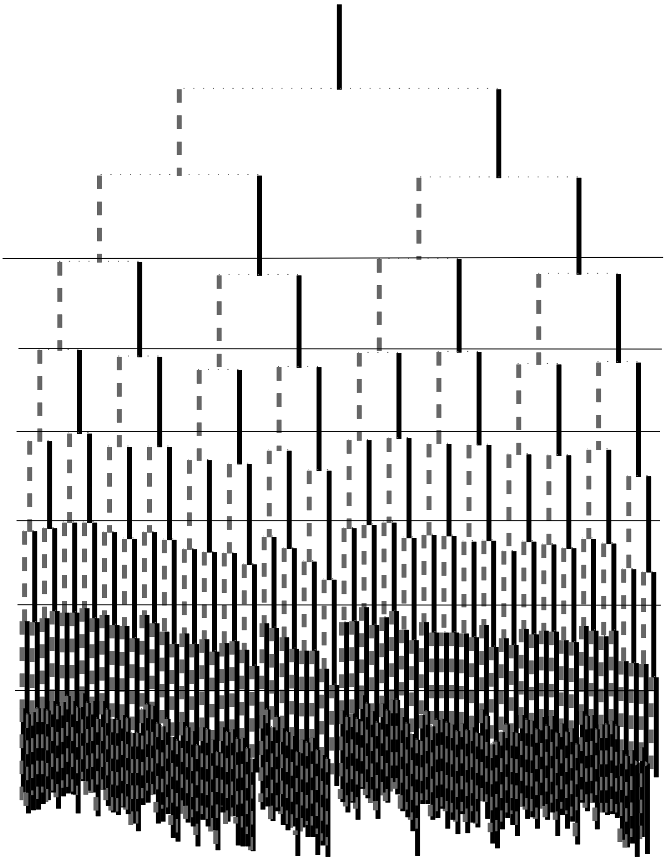


Fig S4: Cell lineage with biased partitioning of vitality-reducing protein aggregates. Cells with older poles are placed on the right. The length of each cell's line is proportional to its lifetime.

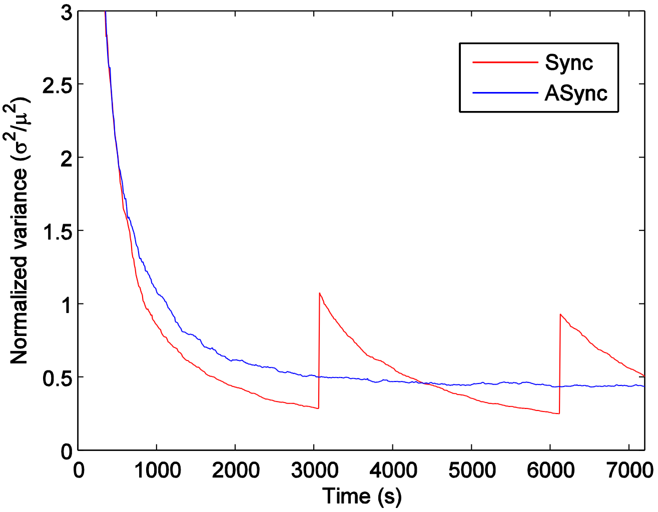


Fig S5: Normalized variance of RNA numbers over time in perfectly synchronous (red) and asynchronous (blue) cell populations. 500 cells were simulated in each population.

Publication II

A. Gupta*, J. Lloyd-Price*, R. Neeli-Venkata, S. Oliveira, and A. S. Ribeiro (2014). *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*. *Biophysical Journal*, **106**(9):1928-37.

1928 Biophysical Journal Volume 106 May 2014 1928–1937

In Vivo Kinetics of Segregation and Polar Retention of MS2-GFP-RNA Complexes in *Escherichia coli*

Abhishekh Gupta, Jason Lloyd-Price, Ramakanth Neeli-Venkata, Samuel M. D. Oliveira, and Andre S. Ribeiro*
Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT The cytoplasm of *Escherichia coli* is a crowded, heterogeneous environment. From single cell live imaging, we investigated the spatial kinetics and heterogeneities of synthetic RNA-protein complexes. First, although their known tendency to accumulate at the cell poles does not appear to introduce asymmetries between older and newer cell poles within a cell lifetime, these emerge with cell divisions. This suggests strong polar retention of the complexes, which we verified in their history of positions and mean escape time from the poles. Next, we show that the polar retention relies on anisotropies in the displacement distribution in the region between midcell and poles, whereas the speed is homogeneous along the major cell axis. Afterward, we establish that these regions are at the border of the nucleoid and shift outward with cell growth, due to the nucleoid's replication. Overall, the spatiotemporal kinetics of the complexes, which is robust to suboptimal temperatures, suggests that nucleoid occlusion is a source of dynamic heterogeneities of macromolecules in *E. coli* that ultimately generate phenotypic differences between sister cells.

INTRODUCTION

Even single-celled organisms, such as *Escherichia coli*, possess a far from random internal organization. Proteins involved in chemotaxis are preferentially located at the cellular poles (1–4), whereas proteins (e.g., RNA polymerases) and transcription factors involved in gene expression mostly locate within a structure known as the nucleoid that, before its replication, is generally located in the central region of the cell (5–9).

At least some of the heterogeneities in the cytoplasm of *E. coli* cells influence their functioning. One example is that cells inheriting the older pole of the mother cell exhibit diminished growth rate (10), which suggests that some contents in the older pole are harmful, and exist in smaller amounts in the newer pole. Subsequent studies hypothesized that one possibly harmful component inherited with the older pole is protein aggregates (10–14).

It is well known that *E. coli* cells, apart from the nucleoid, lack internal organelles (15). They also lack transport mechanisms for proteins (12,15). Thus, the generation and maintenance of most heterogeneity are likely based on the physical properties of the cells, namely, the presence of the nucleoid at midcell (16) and the shape of the cell (17), and on the physical properties of the components (18).

Recently, to study the nature of the cytoplasm of *E. coli*, Golding and Cox (15,19) used live cell microscopy and a synthetic RNA coding for multiple binding sites for a synthetic protein MS2-GFP, based on the MS2 capsid protein (20). By tracking the MS2-GFP tagged RNA molecules, they observed that, at short timescales, their motion was subdiffusive with an exponent that is robust to physiological changes, such as the disruption of cytoskeletal elements (15). In addition, they showed that, at long timescales, these complexes tend to localize at the cell poles. It was hypothesized that this was due to hydrodynamic coupling between the complexes and the cell walls of the poles (19).

Here, using the same approach as in (15), we further investigate the behavior of these large, inert complexes within the cytoplasm of *E. coli*. We choose to use this complex because of its long lifetime (19) (longer than 2 h) and its robustness to photobleaching (21). Furthermore, the individual MS2-GFP proteins are known to distribute uniformly within the cell (22), whereas the motion of the complexes appears to be dominated by physical interactions (15). This provides strong evidence that there are no significant biological interactions between the MS2-GFP proteins or between the MS2-GFP-tagged RNA and other components of the cytoplasm.

From the analysis of time series images of cells expressing MS2-GFP and the target RNA, we address the following questions. Is the accumulation of these complexes at the cell poles a symmetric process? Do cell divisions introduce asymmetries in their numbers in older and newer cell poles, as in the case of unwanted protein aggregates (11)? Are they retained at the poles, and if so, for how long? What heterogeneities and anisotropies in their motion exist along the major cell axis? Does their spatial distribution change in the course of a cell's lifetime? Finally, we investigate to what extent the nucleoid is involved in the observed behavior of these complexes.

MATERIALS AND METHODS

Chemicals

Bacterial cell cultures were grown in lysogeny broth (LB) media. The chemical components of LB (Tryptone, Yeast extract, and NaCl) were purchased

---

Submitted February 10, 2014, and accepted for publication March 28, 2014.

*Correspondence: andre.ribeiro@tut.fi

Abhishekh Gupta and Jason Lloyd-Price contributed equally to this work.

© 2014 by the Biophysical Society

0006-3495/14/05/1928/10 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.03.035

from LabM (Topley House, Bury, Lancashire, UK) and the antibiotics from Sigma-Aldrich (St. Louis, MO). Isopropyl β-D-1-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc) used for induction of the target genes are from Sigma-Aldrich. Agarose (Sigma-Aldrich) was used for microscope slide gel preparation for cell imaging. 4′,6-diamidino-2-phenylindole (DAPI) from Sigma-Aldrich was used to stain cell nucleoids.

Cells and plasmids

Experiments were conducted in *E. coli* strain DH5α-PRO, generously provided by I. Golding (Baylor College of Medicine, Houston, TX), which contains two constructs: i), PROTET-K133 carrying $P_{LtetO-1}$-*MS2d-GFP* (19), and ii), a pIG-BAC ($P_{lac/ara-1}$-*mRFP1-MS2-96bs*) vector, carrying a 96 MS2 binding site array under the control of $P_{lac/ara-1}$ (19).

Induction of production of fluorescent complexes

The dimeric MS2 fused to green fluorescent protein (GFP) (MS2-GFP fusion protein) used as a detection tag (19) is expressed from a medium-copy vector under the control of the $P_{LtetO-1}$ promoter (23), which is regulated by the tetracycline repressor. The RNA target for MS2-GFP is located on a single-copy F-based vector, and is controlled by the $P_{lac/ara-1}$ promoter (23). For our measurements, precultures were diluted from the overnight culture to an OD_{600} of 0.1, in fresh LB media supplemented with the appropriate antibiotics and kept at 24°C or 37°C at 250 RPM in a shaker for ~2 h at 24°C or ~1.5 h at 37°C until they reached an $OD_{600} \approx 0.5$. At this point, cells were induced with 50 ng/ml of aTc and 0.1% L-arabinose for 45 min, at which point the OD_{600} was ~0.8. Induction of the target RNA production was then completed by adding 1 mM IPTG and cells were incubated for 5 min before preparation of the microscope slide. We note that this induction procedure is necessary for cells to have sufficient numbers of MS2-GFP to detect the target RNA and to achieve full induction of the target gene during the microscopy measurements (24,25).

Imaging

After induction of the target gene, a few microliters of culture were placed on a microscope slide between a coverslip and a 0.8% agarose gel pad set with the LB media, followed by the assembly of a thermal imaging chamber (Bioptechs, FCS2) set at the appropriate temperature. Cells were visualized by fluorescence microscopy using a Nikon Eclipse (TE2000-U, Nikon, Tokyo, Japan) inverted microscope with a C1 confocal laser scanning system using a 100× Apo TIRF (1.49 NA, oil) objective. GFP fluorescence was measured using a 488 nm laser (Melles-Griot) and a 515/30 nm detection filter. Images of cells were taken from each slide by the Nikon software EZ-C1, starting ~10 min after induction of the target gene, 1/min, for 2 h. The pixel dwell was 1.33 μs, resulting in a line scanning time of ~1.4 ms; this is significantly faster than the diffusion speed of the MS2-GFP-RNA particles (see (15)), and should therefore not introduce any time-averaging artifacts.

Imaging of nucleoids

DAPI stains nucleoids specifically, with little or no cytoplasmic labeling (26). Precultures were grown for ~3 h with the same previous protocol (but without inducing the target or reporter genes). After reaching an $OD_{600} \approx 0.8$, cells were centrifuged and suspended in phosphate buffered saline (PBS). For live cell nucleoid staining, DAPI (2 μg/ml) was added to the cells suspended in PBS and incubated for 20 min in the dark. Cells were then washed twice with PBS, and placed on a 1% agarose gel pad prepared with LB. The cells were simultaneously observed by epifluorescence microscopy, using a mercury lamp with a DAPI filter, and by phase contrast microscopy. DAPI is excited at 359 nm and emits at 461 nm. Images were acquired with Nikon software NIS-Elements. Cells in phase contrast images were segmented using the software MAMLE (27). The background of the images obtained by epifluorescence microscopy was removed by subtracting a cubic polynomial surface, fitted to the image by L1-norm minimization (i.e., minimizing the absolute difference between the surface and the image; see, e.g. (28)). The fluorescence intensities in each cell were then extracted. See Fig. 6, which shows the fluorescence intensities were projected along the major axis of the cell.

Detection of cells and individual complexes within from the images

Cells were detected from the images by a semiautomatic method as in (22). First, the time series images from confocal microscopy were aligned, so that the cells stayed in the same position throughout the time series. Next, a mask was manually drawn over the region that each cell occupied during the time series. When a cell divided, separate masks were drawn in the frame where the division was first observed, to represent daughter cells after division. After thresholding the fluorescence distribution within each mask to enforce a uniform fluorescence within the cell, principal component analysis was used to obtain, at each frame, the position, dimension, and orientation of the cell inside each mask.

To construct cell lineages, we automatically assigned a parent to each cell in each frame, as the cell in the previous frame with the nearest centroid. This was done after transforming the previous frame's cell centroids by the inverse of the transform that maps a unit circle to the cell's ellipse, to avoid incorrectly assigning adjacent cells as parents. When two cells are assigned the same parent, a division is assumed to have occurred. We verified the efficiency of this method by inspection, and found the rate of error to be negligible.

Next, we detected fluorescent MS2-GFP-RNA complexes in each cell, at each frame, as in (22). We segmented the fluorescent complexes automatically inside each mask with the kernel density estimation (KDE) method for spot detection (29). This method measures the local smoothness of the image, and determines spot locations by designating areas with low smoothness as spots. We used a Gaussian kernel as in (22). Cell background corrected complex intensities were then calculated by subtracting the mean cell background intensity multiplied by the area of the complex from the total fluorescence intensity of the complex.

For cells containing only one complex, once the complexes were detected at each time point, we obtained displacement vectors from their positions in consecutive frames. In Fig. 1, we show an example image of cells (Fig. 1 *A*), along with the segmented cells and detected complexes within (Fig. 1 *B*), and an example of the extracted displacement vectors of a complex from its positions at consecutive frames (Fig. 1 *C*).

By inspection, we observed that the spot detection is reliable. Although we are unable to determine the precision with exactness, as it depends on many variables including noise in the image, we can estimate a conservative upper bound for the error. Assuming that the method of detection is perfect, the discrete nature of the pixels implies that the error in the estimate of the spot's position is up to $2^{-1/2}$ pixels. If the spot detection, e.g., misidentifies pixels at the borders of spots, in the worst case scenario, it would misidentify all pixels only on one side of the spot. This could introduce a further $2^{-1/2}$ error into the estimation of the spot's position. Given this, the error in the estimate of the spot position should have an upper bound of $\sqrt{2}$ pixels, or ~0.17 μm. As several rare events are required to produce this error, the real expected error is considerably lower. Nevertheless, even this upper bound is much smaller than the cell length (2–4 μm), rendering this error negligible.

Models of long-term spatial distributions of large molecules in the cytoplasm of *E. coli*

A cell is modeled as a one-dimensional space, which is divided into N homogeneous subvolumes. The motion of the complexes is modeled by

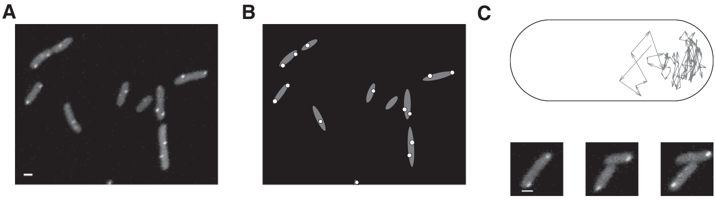


FIGURE 1 (*A*) Example image of cells with fluorescent MS2-GFP-RNA complexes within. (*B*) Segmentation and principal component analysis results of the image in (*A*) with cells (*gray*) and complexes (*white*). (*C*) One example of the extracted displacement vectors of a complex from its consecutive positions in the cell. Three images of the cell are shown below, taken at 40, 80, and 120 min (displacement vectors are from the upper cell). Scale bars are 1 μm. The contrast of these images was enhanced for easier visualization.

unimolecular reactions following the reaction-diffusion master equation (30). The propensities in the reaction-diffusion master equation are functions of the position of the subvolume, and are presented in the Supporting Material. Three models were implemented: one without internal heterogeneities or anisotropies, one with heterogeneities, and one with anisotropies. Also in the Supporting Material, we present the methods used to analyze the results of the models.

RESULTS

Spatial distribution of the complexes

To study the spatial distribution of the complexes, we imaged cells for 2 h following the induction of the target RNA and tagging MS2-GFP proteins (see Methods). Images were taken once per minute, in optimal and suboptimal growth conditions (LB media, 37°C and 24°C, respectively). An example image is shown in Fig. 1 *A*. During this period, cells grew, divided, and produced MS2-GFP-RNA complexes, which moved within the cytoplasm of the cells and were partitioned in cell divisions.

In general, these complexes are first observed at midcell (where the F-plasmid carrying the target gene is located (31)) and then travel toward the cell extremes, where they tend to remain (19). To study whether the side to which they travel is a symmetric (i.e., unbiased) process with respect to the age of the cell halves, we observed this process in cells that initially contained no complexes and that were born during the measurement period (107 cells at 37°C and 156 cells at 24°C), so that the older half of each cell could be identified. In these cells, at each time moment, we identified the locations of fluorescent complexes along the major axis of the cell (positions are normalized by half the length of the major axis), and determined the background-corrected fluorescence intensity of each. Colocalized complexes will, approximately, exhibit a fluorescence intensity that is the sum of the intensity of its component complexes (25). For each condition, we summed the intensities of the complexes at each location along the major axis of the cell, over all time points and cells, thus obtaining the spatiotemporal distribution of the complexes. We used KDE (32) with a Gaussian kernel to perform this sum, resulting in a smooth distribution (Fig. 2 *A*). We note that we did not separate cells of different sizes when obtaining this distribution. Given the time length of the measurements and the fact that most cells divided during this period, the resulting distribution and all conclusions drawn from it should be considered to be the average behavior over the cell cycle.

From the distribution for each condition, we computed the fraction of complexes in the older half. We found this fraction to be 0.46 at 37°C and 0.47 at 24°C, which are both statistically indistinguishable from the expected fraction assuming an unbiased partitioning of complexes (p-values of the binomial test with N equal to the number of observed cells were larger than 0.2, and it is usually accepted that, for p-values smaller than 0.01, the null hypothesis is rejected). We therefore find no evidence that

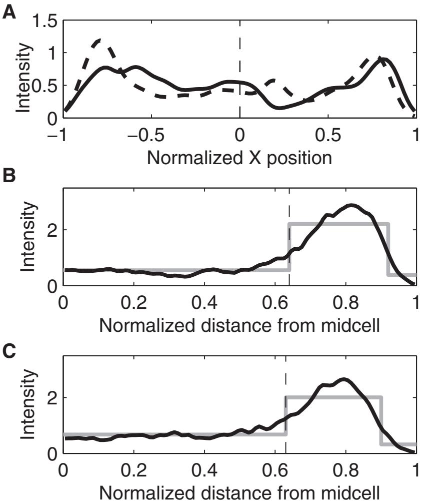


FIGURE 2 KDE of spatial distributions of the fluorescence intensity (in arbitrary units) of complexes along the major axis of the cells (bandwidths 0.05). All cells were born during the measurement period. (*A*) Data from all cells that inherited no complexes but produced one or more. The old pole is at +1 and the new pole is at −1. Data are from 107 cells at 37°C (*solid line*) and 156 cells at 24°C (*dashed line*). The dashed vertical line represents the cell center. (*B* and *C*) Data from all cells (*black line*). Complex positions were normalized by half the cell length and mirrored about the cell center. Also shown is the fit of a piecewise constant probability density function by maximum likelihood (*gray line*). The vertical dashed line represents the detected separation points between the midcell and poles. Measurements are from (*B*) 531 cells at 37°C, with separation point at 0.64 and (*C*) 372 cells at 24°C, with separation point at 0.61.

the complexes are asymmetrically distributed between the older and newer halves of the cell before division events.

Also from Fig. 2 *A*, in both conditions, the complexes were preferentially located close to the cell extremes (i.e., at the polar regions) for most of the measurement period, as reported in (25). To quantify the degree of polarization of the complexes, one needs to formally define such polar regions, along the major axis of the cells. We did this from a functional perspective, i.e., based on the heterogeneities of the spatial distributions of the complexes visible in Fig. 2 *A*. Given the symmetry in these distributions between the old and new halves of the cells, from here onward, we folded the spatial distribution around 0, and summed the intensities from both halves. As such, it is possible to include all cells born during the measurements in this analysis. The resulting distributions of each condition are shown in Fig. 2, *B* and *C*.

To distinguish the functional regions in each condition, we fitted a piecewise constant probability density function with three pieces to each intensity distribution by maximum likelihood (*gray line* in Fig. 2, *B* and *C*). The separation points between the regions from the fit were found to be at 0.64 for 37°C and at 0.61 for 24°C.

Based on this separation between poles and midcell regions, we calculated the concentration of complexes at the poles and at midcell, in each condition, to assess the degree of polarization of the complexes. In this case, concentrations >1 indicate that the complexes are located in this region beyond what would be expected from a uniform distribution. This concentration was found to be 1.72 at 37°C and 1.45 at 24°C.

From this separation, we also determined whether cell divisions introduced biases in the numbers of complexes at the old and new poles in subsequent generations. In the measurements conducted at 37°C, sufficient divisions occurred in the 2 h measurement period of to assess this. From these, we selected cells that inherited one complex but produced none during their lifetime (111 cells), to ensure that the complexes analyzed are only inherited ones. During the lifetime of these cells, the old pole contained 65% of the complexes located in a pole. The *p*-value that this fraction arises from an unbiased binomial distribution with the number of trials equal to the number of cells is 0.004, from which we conclude there is a significant bias that favors the old pole.

As a control, similar experiments were performed in minimal media (M63) at 37°C (for details, see the Supporting Material). The results (Fig. S1 and Fig. S2) are qualitatively the same as in LB media. Specifically, the complexes accumulated at the poles, in well-defined regions. Furthermore, cell divisions introduced asymmetries between the numbers of complexes at the old and new poles of the daughter cells. These asymmetries following divisions are possible if, to some extent, the complexes are retained at the pole where they are inherited (otherwise, the bias would be rapidly lost).

Retention of complexes at the poles

To study the retention of the complexes at the poles, for each condition, we selected cells that contained at most one complex throughout their lifetime (either inherited or produced), so that they could be reliably tracked. We recorded the position along the major axis where the complex was first observed within a polar region, as previously defined, and the complex's final position, i.e., either before division or at the end of the measurement period. This information is presented in Fig. 3. In general, a complex's final position is within the polar region to which it first traveled to, as expected from a strong polar retention. However, in a few cases, the opposite occurs, which shows that complexes can move across the entire major cell axis (as reported in (19)). This indicates that there is a degree of leakiness, or noise, in the retention at the poles.

To quantify the strength of polar retention, we measured the escape times of the complexes from the poles, in each condition (mean values in Table 1). Note that, on some (rare) occasions, the complexes appeared to leave the pole for only one time moment, and then promptly returned. We did not count these cases as escapes. Furthermore, when complexes did not leave the pole until the end of the measurement period or until a division event, we consider that they remained at the pole only until the next time

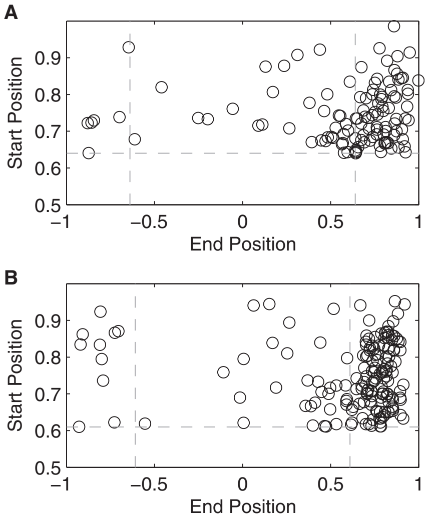


FIGURE 3 Relationship between the position along the major axis where each complex was last observed and the absolute position where it was first observed at a pole. Here, an end position of +1 indicates that the complex remained at the same pole as it was first observed, whereas −1 indicates that it traveled to the other pole. The horizontal and vertical dashed lines represent the detected separation between the midcell and poles from Fig. 2, *B* and *C*. All cells were born during the measurement period and contained only one complex during their lifetime. Measurements are from (*A*) 160 cells at 37°C and (*B*) 198 cells at 24°C.

TABLE 1 Mean and standard deviations of escape times of complexes from the poles, in each condition

T (°C)	No. cells	Mean division time (min)	Escape times (min)	p-value of *t*-test
37	160	63	14.8 ± 19.3	0.062
24	198	91	18.0 ± 18.4	

Also shown is the number of cells observed, their mean division time, and the *p*-value of the *t*-test with the hypothesis that the mean escape times are identical in both conditions.

moment. Thus, the measured mean escape times are underestimates of the real escape times.

From Table 1, within the range tested, we find no evidence that temperature affects the mean escape time from the poles. Furthermore, in both temperature conditions, the standard deviations of escape times were similar to the means, a characteristic of exponential distributions, which is the expected solution of, e.g., a barrier crossing problem or of a particle trying to escape from a region through small passages (33).

Spatial dynamics of complexes

To better understand how the complexes are retained at the poles, from cells containing at most one complex during their lifetime, we obtained their displacement vectors along the major cell axis between frames. These inform on the directionality of a complex between consecutive images (assessed by the sign of the displacement vector). In addition, they inform on the speed at which the complexes are able to move along the major cell axis during the intervals between consecutive images (assessed by the magnitude of the displacement vector). Cell growth between consecutive frames was accounted for by projecting the origin of each displacement vector into the cell space in the following frame, before calculating the magnitude and direction.

First, for each condition, we extracted the speeds from the displacement vectors going toward a pole and going toward the midcell, as a function of their point of origin. For this, we defined a sliding window with a width of 0.1 cell lengths and determined which displacement vectors originated within that window and their direction.

At midcell, the mean speed of complexes going toward a pole (0.13 μm/min at 37°C and 0.14 μm/min at 24°C) was statistically indistinguishable from the mean speed of those going toward midcell (0.12 μm/min at 37°C and 0.14 μm/min at 24°C) (p-values of the t-tests >0.4). At the poles, in both temperature conditions, the mean speed of complexes going toward a pole (~0.07 μm/min) and toward midcell (~0.11 μm/min) was distinguishable, in a statistical sense (p-values of the t-tests <0.01). This difference is visible in Fig. 4, for both conditions. The decrease in mean speed as the complexes approach the cell extremes is expected, given the proximity to the cell wall.

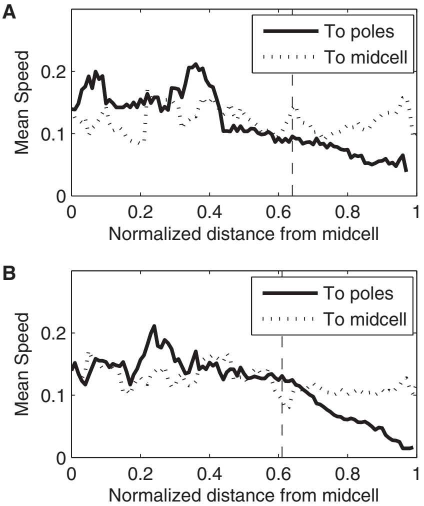


FIGURE 4 Mean speed (μm/min) of the displacement vectors directed toward the poles and toward midcell along the major cell axis. The mean speeds were calculated from the displacement vectors originating within a window extending 0.05 normalized cell lengths around that point. The dashed vertical lines represent the functional separation between midcell and poles (obtained from Fig. 2, *B* and *C*). All cells were born during the measurement period and contained one complex in their lifetime. Measurements are from (*A*) 49 cells at 37°C and (*B*) 101 cells at 24°C.

Relevantly, the mean speed of the vectors going toward the cell center of complexes at the poles and at midcell is indistinguishable in a statistical sense, in both temperature conditions (p-values of the t-tests >0.01). Given that the mean speeds at midcell of complexes traveling toward poles and toward midcell are also indistinguishable (see above), we conclude that the speed of complexes is fairly homogenous throughout the major cell axis, except for the complexes traveling toward a pole that are already near the cell extreme (which, therefore exhibit lower mean speed).

We next analyzed the directionality of the displacement vectors. Using the same sliding window as before, we counted the number of displacement vectors originated in the window, which were directed toward the midcell and toward the poles. In Fig. 5, we show the difference between these two numbers along the major cell axis. In both conditions, a characteristic spatial heterogeneity is observable. At midcell, the complexes have equal probability of moving in either direction, whereas at the poles, there are local biases in the directionality of the displacement vectors. In particular, if close to the cell extremes, the complexes tend to move inward, toward the midcell. Meanwhile, if they are close to the border between midcell and the pole (as defined from the spatial distributions of complexes shown in Fig. 2, *B* and *C*), the opposite occurs. As a result, once reaching a pole, the complexes tend to remain there.

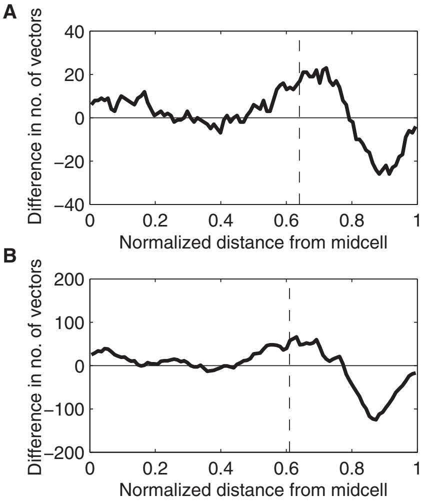


FIGURE 5 Difference between the numbers of displacement vectors that are directed toward the poles and toward the midcell along the major cell axis. The differences were calculated from the displacement vectors originating within a window extending 0.05 normalized cell lengths around that point. The dashed vertical lines represent the functional separation between midcell and poles (obtained from Fig. 2, *B* and *C*). All cells were born during the measurement period and contained one complex in their lifetime. Measurements are from (*A*) 49 cells at 37°C and (*B*) 101 cells at 24°C.

It is noted that although the anisotropic displacement distribution at the extremes of the cell is expected from the geometry of the cell wall at the poles, the source of the opposite anisotropy in the transition between midcell and poles is less clear. Its existence suggests that the motion of the complexes going from poles to midcell is, to a degree, obstructed. This effect is possible if the complexes are encountering a more dense structure that hampers their entrance into that region.

Relevantly, the location of the anisotropy, namely, 0.64 at 37°C and 0.61 at 24°C, is in agreement with previous measurements of the nucleoid size in *E. coli* cells grown in LB media at 37°C (e.g., its length relative to the major axis length is 0.53 ± 0.05 in wild-type DJ2599 cells (34)). We thus hypothesized that the nucleoid is involved in this phenomenon, and measured its length in the cells of the strain used here, in the same conditions as above.

Spatial distribution of nucleoids

To test whether the regions of anisotropies of the displacement distributions along the major cell axis are consistent with the borders of the nucleoid, we measured the nucleoid size (see Methods) from 220 cells at 37°C and 143 cells at 24°C (see example, Fig. 6, *A* and *B*). The intensity of the DAPI signal, summed along the minor axis for each position along the major axis of the cells, and summed over all cells, is shown in Fig. 6, *C* and *D*.

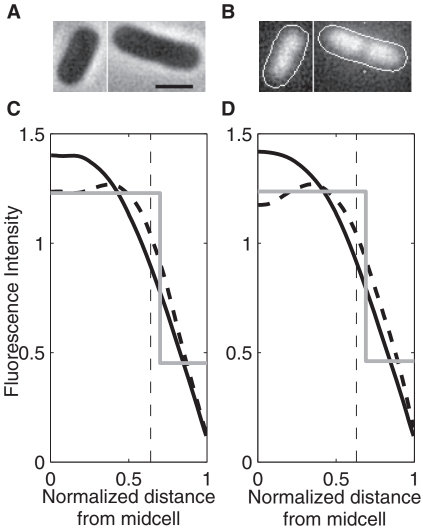


FIGURE 6 (*A*) Example phase contrast images of cells and (*B*) corresponding background-subtracted epifluorescence with the nucleoids stained by DAPI, and the detected cell contours superimposed (contrast enhanced for easier visualization in both *A* and *B*). Scale bar is 1 μm. The two bottom figures (*C*) and (*D*) show the spatial distributions of the fluorescence intensity (in arbitrary units) of the DAPI signal along the major cell axis (*black line*) and the fit of a piecewise constant probability density function by maximum likelihood (*gray line*). The vertical dashed lines represent the detected separation between midcell and poles (from Fig. 2, *B* and *C*). Also shown is the fluorescence distribution from the 10% longest cells in each condition (*dashed line*). Measurements are from (*C*) 220 cells at 37°C and (*D*) 143 cells at 24°C.

To determine the edge of the nucleoid, we fitted a piecewise constant probability density function with two pieces to the intensity distribution by maximum likelihood. We found the separation point to be at 0.69 in both conditions. This is close to the measured separation points between the midcell region, which the complexes avoid, and the poles, where they accumulate (0.61 and 0.64 at 24°C and 37°C, respectively, see Fig. 2, *B* and *C*). The slightly larger size of the nucleoid may be due to DAPI being an intercalating dye (26). Note that the nucleoid size was not altered by the differences in temperature. This is in agreement with the lack of differences in the width of the polar regions where the complexes tend to accumulate in the two temperature conditions.

If the heterogeneities in the spatial distribution of the complexes depend on the positioning of the nucleoid(s), as these results suggest, the replication of the nucleoid before cell division should then affect this distribution. To determine this, we first selected the 10% longest cells detected in the DAPI measurements and searched for differences in their spatial distribution of fluorescence intensities, when compared to the total population. Results in Fig. 6, *C* and *D*, confirm the existence of differences, as the region occupied by the nucleoids along the major cell axis is relatively wider in these cells. Note also a decrease in fluorescence intensity, precisely at midcell, as expected if several of these cells contain two nucleoids.

Next, we studied the spatial distribution of the complexes in the 10% longest cells as extracted from the time series measurements of the complexes' positions and kinetics (from the same data used in Fig. 2, *B* and *C*). If the nucleoid(s) affect the spatial distribution of complexes, in these cells one can expect the complexes to accumulate relatively closer to the poles. Fig. 7 shows that this is the case both at 37°C and 24°C, thus providing supporting evidence that the nucleoid(s) influence(s) the complexes' preferential positioning.

As a side note, we did not find evidence for cells with more than two nucleoids at any stage of their lifetime (see, e.g., that Fig. 6, *C* and *D* do not show evidence for more than two lobes). This could be due to the strain used (DH5α-PRO divides slower than wild-type *E. coli* (35)) along with the measurement conditions. In addition, we found no significant accumulation of complexes at the center in the longer cells (i.e., in between the two nucleoids). This is likely due to the very small number of complexes (~1 to 3 per cell before division). Furthermore, many are created before nucleoid partitioning and thus simply remain at the poles as the nucleoid divides.

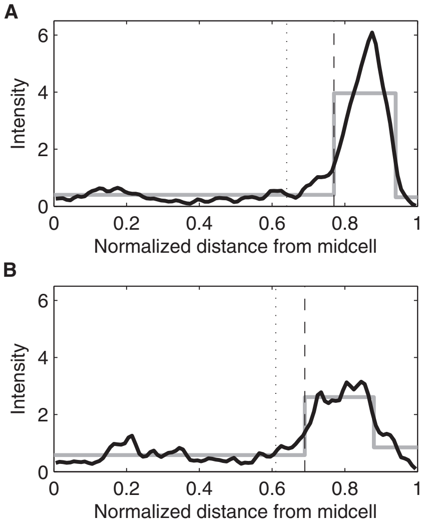


FIGURE 7 KDE of the spatial distribution of the fluorescence intensity (in arbitrary units) of complexes along the major axis of the cells, extracted from all time points when the cells were among the 10% longest cells (*black line*, bandwidth 0.05). All cells were born during the measurement period. The dashed vertical line represents the detected separation point for (*A*) 531 cells at 37°C and (*B*) 372 cells at 24°C, both grown in LB media.

Finally, because we observed several complexes traveling from one pole to the other, it is of interest to ask whether they do so by traveling through the nucleoid or around it. To address this, we obtained the KDE of the two-dimensional distributions of complexes from all time points in both temperature conditions (Fig. S3, *A* and *B*). From the figures, it is visible that the complexes tend to avoid the nucleoid region both axially and radially, concentrating at midcell close to the cell walls (despite the greater width in the center of the bacteria, at $Y = 0$). From this, we conclude that the complexes tend to go around the nucleoid, when traveling through the midcell region. Relevantly, this result is in agreement with both the homogeneity in the distributions of speeds along the major axis of the cells as well as with the localized anisotropies between midcell and poles.

Models of the spatial kinetics of complexes

To test whether the localized anisotropies in directionality, given the homogeneity of the speeds, can generate the observed heterogeneity in the long-term spatial distributions of the complexes, we constructed two one-dimensional models to simulate the diffusion of the complexes within the cell. Both models contain spherical cell caps and their effects. Meanwhile, in one model, we also introduced a localized anisotropy (see Fig. 8, *A* and *B*; for a complete description of the models see the Supporting Material).

Given that the mean speeds of the complexes (see Fig. 4) are sufficiently large to allow them to travel from one pole to the other within the cell's lifetime (more than once), we assume that the initial positions of the complexes do not have a significant effect on their long-term spatial distribution, and thus this information is not included in the models. Additionally, for both models, we set N, the number of subvolumes in the cell, to 100, and D, the diffusion coefficient, to 1.43×10^{-2} μm^2/min (measured from the displacement distribution depicted in Fig. 4 *B*), which we normalized by half the mean cell length (~1 μm), in agreement with previous measurements (15).

Next, for each model, we varied all parameters and, for each set of values, obtained the distribution of complex positions that would be observed at infinite time. We then selected the set of parameters whose resulting distribution best fit the measured distribution of complex positions at each time point (Fig. 2, *B* and *C*). This fitting was done to the distribution obtained from the measurements at 24°C (from which we extracted more data points), using the earth-mover's measure of distance between distributions (36,37), which measures the amount of work required to make two distributions identical (see methods in the Supporting Material). Similar parameter values were obtained when fitting to the measurements at 37°C.

The results from the two models, each using the best-fit parameter values, are shown in Fig. 8. The propensities of the two models are shown in Fig. 8, *A* and *B*. From Fig. 8,

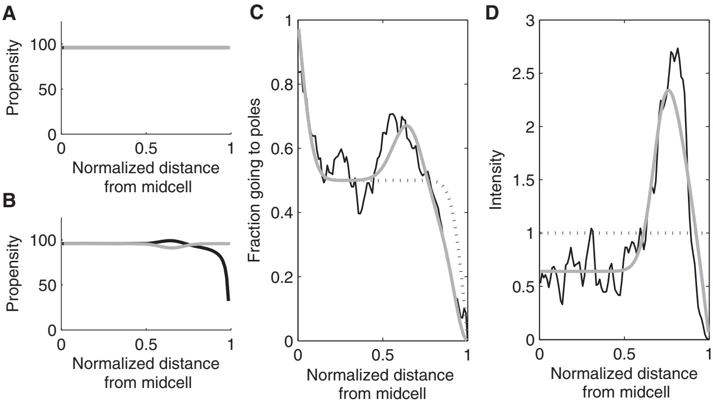


FIGURE 8 (*A*) Propensities $\overrightarrow{a}(x) = \overleftarrow{a}(x)$ (*gray line*) of the best-fitting model without a nucleoid ($B = 1$). (*B*) Propensities $\overrightarrow{a}(x)$ (*black line*) and $\overleftarrow{a}(x)$ (*gray line*) of the best-fitting model with a nucleoid ($B = 0.46$, $\mu = 0.65$, $\sigma = 0.07$, and $h = 0.05$). (*C*) Measured fraction of displacement vectors originating within a window extending 0.05 normalized cell lengths around that point which are directed toward the pole (*black line*), model prediction without (*dashed line*), and with (*gray line*) a nucleoid. Note that the dashed line is superimposed by the gray line in the left side of the graph. (*D*) Measured spatial distribution of fluorescence intensities of complexes (*black line*) model prediction without (*dashed line*) and with (*gray line*) a nucleoid.

C and *D*, the model without the anisotropy fails to reproduce the displacement distribution (Fig. 8 *C*), and the consequent heterogeneity in the spatial distribution of complexes that favors their presence at the poles (Fig. 8 *D*). Meanwhile, the other model captures both of these properties of the dynamics of the complexes with significant accuracy.

DISCUSSION

We studied the kinetics and spatial distribution of biologically inert complexes, composed of RNA molecules tagged with multiple MS2-GFP proteins, in the cytoplasm of *E. coli* cells. The advantages of using these complexes are that their numbers can be controlled by regulating the activity of the target gene, both the target gene and the tagging MS2-GFP molecules are functional in a wide range of environmental conditions including stresses, and the complexes formed have a long lifetime (38,39). This last advantage ensures that we can observe how they are partitioned in division and thus, how they become distributed across cell lineages.

We found that their previously reported tendency to travel toward the poles (19) is a symmetric process in the conditions tested, in that equal numbers of newly produced complexes travel toward the old and the new pole of the cells. Once reaching the poles, they are robustly retained there. Because of this, cell divisions introduce asymmetries in their numbers between the old and new poles of cells of subsequent generations, and after two generations, between sister cells.

There are two possible mechanisms by which the complexes could accumulate at the cell poles, without a transport mechanism. Either their speed distribution is heterogeneous along the major cell axis (i.e., slower at the poles, see Fig. S4), or there is an excluded volume effect at midcell. We found that the retention is solely based on the latter. First, we showed that the speed distribution is homogenous, which rejects the first mechanism. Next, we showed that there is a strong anisotropy in the displacement distribution at approximately half way between the cell extremes and the center, where the motion of the complexes toward midcell is, to a great extent, obstructed. This is consistent with volume exclusion effects due to the presence of the nucleoid.

The existence of the anisotropy and the absence of heterogeneity in the speeds suggest that to go from one pole to the other, the complexes go around the nucleoid. The overall two-dimensional distributions of positions occupied by complexes throughout their lifetimes supported this hypothesis, as they showed that the complexes avoid the nucleoid, both axially and radially. Meanwhile, the escape times of these complexes from the poles were found to be approximately equal to a third of the cell's lifetime and to follow exponential distributions. This implies that the escape from the pole is a Poisson process, which is consistent with the behavior of a particle trying to escape from a region through a small passage (33).

To further support the hypothesis that the polar retention is primarily driven by the nucleoid positioning, we tested whether there is an agreement between the location of the nucleoid and where the movement of complexes is obstructed. We found that the region where the complexes are retained agrees with the region where the nucleoid ends. In addition, we observed that in the longest cells, where the nucleoid(s) occupy a relatively wider region (due to nucleoid replication), the complexes occupy a relatively smaller region at the poles. Additional support was provided by modeling. From the models, in the absence the nucleoid's effects, namely, in the absence of the anisotropy in the region between poles and midcell, the retention at the poles was severely hampered. Introducing anisotropy in the velocity distribution resulted in an accurate long-term spatial distribution.

The size of the region occupied by the nucleoid(s) during the cell lifetime was found to be robust for optimal and suboptimal temperatures. This explains the robustness of the width of the region where the complexes accumulated, in the range of temperatures tested here. In the future, measurements in wider temperature ranges may help to establish if the retention at the poles is affected, as the number of genes transcribed changes, altering the nucleoid size (40).

It is known that, when in exponential growth phase, some cells can contain more than two nucleoids. Although we did not observe this here (perhaps due to the strain used), if these nucleoids become widely spread across the major cell axis long enough, the regions in between the nucleoids may become regions of accumulation of complexes, until multiple cell division events separate the nucleoids. Future research may determine whether this occurs and to what extent.

Finally, by using the methodology employed here on other *E. coli* strains, it should be possible to determine whether our observations are representative of the behavior of wild-type *E. coli*. Given the physical nature of the underlying processes suggested by our results, we expect this to be the case. In this regard, it is worth noting that the long-term spatial distribution of the complexes observed here is strikingly similar to other protein complexes in *E. coli*. In particular, their accumulation at the poles of the cells is similar to aggregate processing chaperones (11) and ribosomes (9). Furthermore, clusters of Tsr proteins, involved in chemotaxis, are known to accumulate at the poles (4,41), although the mechanisms by which these heterogeneities are achieved remain unknown. Future research may establish if all these processes in *E. coli* are regulated by the same means.

SUPPORTING MATERIAL

Four figures and supporting data are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00336-1.

This work was supported by the Finnish Funding Agency for Technology and Innovation (S.O., R.N.V., A.S.R.), the Academy of Finland (A.S.R.), the Portuguese Foundation for Science and Technology (A.S.R.), the Tampere Doctoral Programme in Information Science and Engineering (A.G.), and the TUT President's graduate programme (J.L.P.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Sourjik, V., and H. C. Berg. 2004. Functional interactions between receptors in bacterial chemotaxis. *Nature.* 428:437–441.
2. Wadhams, G. H., and J. P. Armitage. 2004. Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* 5:1024–1037.
3. Parkinson, J. S., P. Ames, and C. A. Studdert. 2005. Collaborative signaling by bacterial chemoreceptors. *Curr. Opin. Microbiol.* 8:116–121.
4. Zhang, P., C. M. Khursigara, ..., S. Subramaniam. 2007. Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA.* 104:3777–3781.
5. Kuhlman, T. E., and E. C. Cox. 2012. Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol. Syst. Biol.* 8:610.
6. Montero Llopis, P., A. F. Jackson, ..., C. Jacobs-Wagner. 2010. Spatial organization of the flow of genetic information in bacteria. *Nature.* 466:77–81.
7. Jin, D. J., C. Cagliero, and Y. N. Zhou. 2012. Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev.* 36:269–287.
8. Bratton, B. P., R. A. Mooney, and J. C. Weisshaar. 2011. Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. *J. Bacteriol.* 193:5138–5146.
9. Bakshi, S., A. Siryaporn, ..., J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* 85:21–38.
10. Stewart, E. J., R. Madden, G. Paul, and F. Taddei. 2005. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* 3:0295–0300.
11. Lindner, A. B., R. Madden, ..., F. Taddei. 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. USA.* 105:3076–3081.
12. Winkler, J., A. Seybert, ..., B. Bukau. 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J.* 29:910–923.
13. Ebersbach, G., A. Briegel, ..., C. Jacobs-Wagner. 2009. A multimeric pole-organizing protein critical for chromosome attachment, division and protein localization in *Caulobacter*. *Cell.* 134:956–968.
14. Maisonneuve, E., B. Ezraty, and S. Dukan. 2008. Protein aggregates: an aging factor involved in cell death. *J. Bacteriol.* 190:6070–6075.
15. Golding, I., and E. C. Cox. 2006. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.* 96:098102.
16. Fisher, J. K., A. Bourniquel, ..., N. Kleckner. 2013. Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells. *Cell.* 153:882–895.
17. Coquel, A.-S., J.-P. Jacob, ..., H. Berry. 2013. Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. *PLOS Comput. Biol.* 9:e1003038.
18. Mondal, J., B. P. Bratton, ..., J. C. Weisshaar. 2011. Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells. *Biophys. J.* 100:2605–2613.
19. Golding, I., and E. C. Cox. 2004. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA.* 101:11310–11315.
20. Peabody, D. S. 1993. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* 12:595–600.
21. Kandhavelu, M., A. Häkkinen, ..., A. S. Ribeiro. 2012. Single-molecule dynamics of transcription of the lar promoter. *Phys. Biol.* 9:026004.
22. Kandhavelu, M., J. Lloyd-Price, ..., A. S. Ribeiro. 2012. Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter. *FEBS Lett.* 586:3870–3875.
23. Lutz, R., and H. Bujard. 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 25:1203–1210.
24. Mäkelä, J., M. Kandhavelu, ..., A. S. Ribeiro. 2013. In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter. *Nucleic Acids Res.* 41:6544–6552.
25. Golding, I., J. Paulsson, ..., E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123:1025–1036.
26. Kapuscinski, J. 1995. DAPI: a DNA-specific fluorescent probe. *Biotech. Histochem.* 70:220–233.
27. Chowdhury, S., M. Kandhavelu, ..., A. S. Ribeiro. 2013. Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE). *BMC Bioinformatics.* 14 (*Suppl 10*):S8.
28. Portnoy, S., and R. Koenker. 1997. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Stat. Sci.* 12:279–300.
29. Ruusuvuori, P., T. Aijö, ..., O. Yli-Harja. 2010. Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics.* 11:248.

30. Gardiner, C. W., K. J. McNeil, …, I. S. Matheson. 1976. Correlations in Stochastic theories of chemical reactions. *J. Stat. Phys.* 14:307.

31. Gordon, S., J. Rech, …, A. Wright. 2004. Kinetics of plasmid segregation in *Escherichia coli*. *Mol. Microbiol.* 51:461–469.

32. Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London, UK.

33. Chandrasekhar, S. 1943. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* 15:1–89.

34. Cabrera, J. E., C. Cagliero, …, D. J. Jin. 2009. 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.* 191:4180–4185.

35. Jung, S.-C., C. L. Smith, …, Y. S. Jin. 2010. Restoration of growth phenotypes of *Escherichia coli* DH5alpha in minimal media through reversal of a point mutation in *purB*. *Appl. Environ. Microbiol.* 76:6307–6309.

36. Rubner, Y., C. Tomasi, and L. J. Guibas. 2000. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* 40:99–121.

37. Dobrushin, R. L. 1970. Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.* 15:458–486.

38. Lloyd-Price, J., M. Lehtivaara, …, A. S. Ribeiro. 2012. Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*. *Mol. Biosyst.* 8:565–571.

39. Muthukrishnan, A.-B., M. Kandhavelu, …, A. S. Ribeiro. 2012. Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucleic Acids Res.* 40:8472–8483.

40. Tao, H., C. Bausch, …, T. Conway. 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J. Bacteriol.* 181:6425–6440.

41. Yu, J., J. Xiao, …, X. S. Xie. 2006. Probing gene expression in live cells, one protein molecule at a time. *Science.* 311:1600–1603.

Supporting Material for "*In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*"

Abhishekh Gupta[†,‡], Jason Lloyd-Price[†,‡], Ramakanth NeeliVenkata[†], Samuel M.D. Oliveira[†], and Andre S. Ribeiro[†]

† Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, FI-33101 Tampere, Finland.

‡ Equal contributions

Measurements in M63 Media

We imaged cells grown in M63 media at 37°C for 2 hours, following induction of the target RNA and of the tagging MS2-GFP proteins (one hour before), using the same experimental and analysis protocols as in the experiments in LB described in the main text. The spatial distribution of newly-produced complexes, corresponding to Fig. 2 *A* in the main text for LB, is presented in Fig. S1.

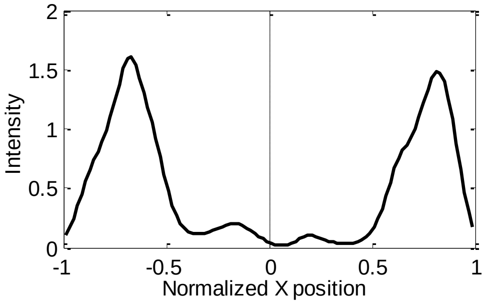


FIGURE S1 KDE of the spatial distribution of the fluorescence intensity (in arbitrary units) of complexes along the major axis of the cells, extracted from all cells and time points (*black line*, bandwidth 0.05). Data is extracted from cells that inherited no complexes but produced one or more. The old pole is at +1 and the new pole is at -1. All cells were born during the measurement period. The dashed vertical line represents the cell center. Measurements are from 63 cells grown in M63 media at 37°C. The fraction of complexes observed in the older half of the cells was 0.45 which is statistically indistinguishable from an unbiased partitioning of complexes (p-value of the binomial test with *N* equal to the number of observed cells is 0.45).

The folded, spatial distribution of all complexes, corresponding to Figs. 2, *B* and *C* in the main text for LB, is presented in Fig. S2. Also shown are the results from the 'region detection' method.

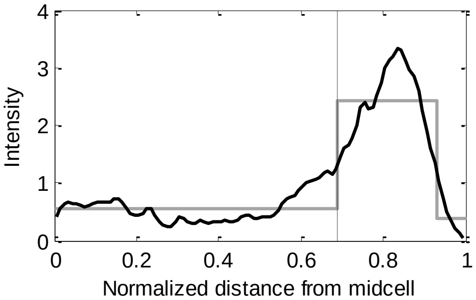


FIGURE S2 KDE of the spatial distribution of the fluorescence intensity (in arbitrary units) of complexes along the major axis of the cells, extracted from all cells and time points (*black line*, bandwidth 0.05). Complex positions were normalized by half the cell length. Also shown is the fit to a piecewise-constant probability density function by maximum likelihood (*gray line*). All cells were born during the measurement period. The vertical dashed lines represent the detected separation points between the midcell and poles. Measurements are from 221 cells grown in M63 media at 37°C, with separation point detected at 0.69.

2D Spatial distribution of complexes

We obtained the KDE of the 2D distributions of complexes from all time points in both temperature conditions. Results are shown in Figs. S3, *A* and *B*, for cells at 37°C and 24°C in LB, respectively.

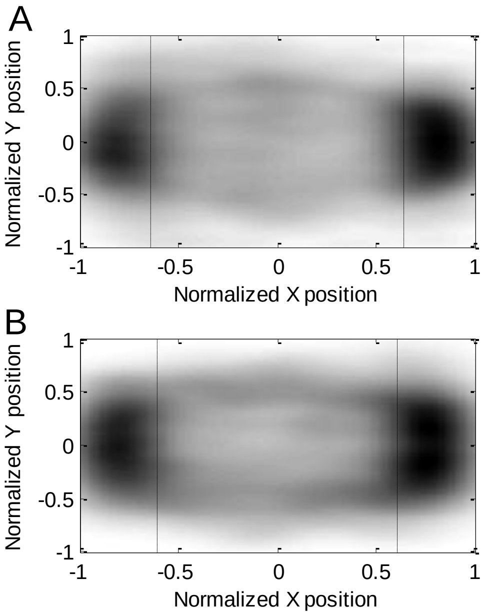


FIGURE S3 KDE of the 2D spatial distributions of complexes from all time points. All cells were born during the measurement period. Measurements are from (*A*) 531 cells at 37°C and (*B*) 372 cells at 24°C, both grown in LB media. Separation points (obtained from Figs. 2, *B* and *C*) are also shown.

Models of long-term spatial distributions of large molecules in the cytoplasm of *E. coli*

A cell is modeled as a 1-dimensional space, which is divided into *N* homogeneous sub-volumes, indexed from $[1, N]$. The motion of the complexes is modeled along the major cell axis by unimolecular reactions following the Reaction-Diffusion Master Equation (1). Collisions between complexes are not modelled. We define $\vec{a}(x)$ as the propensity of the forward reaction (modeling the motion of a complex from position *x* to position *x*+1) and $\overleftarrow{a}(x)$ as the propensity of the reverse reaction (from *x* to *x*-1).

These two propensities account for the combined effects of the cell geometry (rod shape and pole caps) and of the nucleoid on the displacements of the complexes. Let $\mathbf{P}(t)$ be the *N*-by-1 vector describing the probability of observing a complex in each sub-volume at time *t*. $\mathbf{P}(t)$ therefore evolves according to the following master equation, in matrix-vector form:

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{AP}(t)$$

where

$$\mathbf{A} = \begin{bmatrix} -\vec{a}(1) & \overleftarrow{a}(2) & 0 & & 0 & 0 \\ \vec{a}(1) & -\overleftarrow{a}(2)-\vec{a}(2) & \overleftarrow{a}(3) & \cdots & 0 & 0 \\ 0 & \vec{a}(2) & -\overleftarrow{a}(3)-\vec{a}(3) & & 0 & 0 \\ & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & & -\overleftarrow{a}(N-1)-\vec{a}(N-1) & \overleftarrow{a}(N) \\ 0 & 0 & 0 & \cdots & \vec{a}(N-1) & -\overleftarrow{a}(N) \end{bmatrix}$$

Since a complex can travel from any sub-volume in the cell to any other sub-volume, given enough time, the system is ergodic. Therefore, as $t \to \infty$, $\mathbf{P}(t)$ will converge to a unique solution, $\mathbf{P}_\infty$. This solution can be found by solving the linear system of equations $\mathbf{0} = \mathbf{AP}_\infty$, with the constraint that the total probability must sum to 1. As this is the long-term spatial distribution of the complexes predicted by the model, this was the distribution we fit to the measurements.

In a model not accounting for the caps of the cells, the propensities of the forward and reverse diffusion reactions would be proportional to the diffusion constant of the complexes, *D*:

$$\vec{a}(x) = \overleftarrow{a}(x) = \frac{N^2 D}{2}$$

To account for the rod shape, i.e. a cylinder capped with two half-spheres, the length of the cell was parameterized by $B \in [0,1]$, the normalized distance from midcell at which the cap begins. The forward propensities were attenuated by $\phi(x)$, the ratio between the areas of the cross sections of the cell (denoted $S(x)$) at adjacent positions. As such, $\overleftarrow{a}(x)$ remains the same and $\vec{a}(x)$ becomes:

$$\vec{a}(x) = \frac{N^2 D}{2}\phi(x)$$

where

$$\phi(x) = \frac{S(x+1)}{S(x)}$$

$$S(x) = \begin{cases} \pi, & c(x) < B \\ \pi\left[1 - \left(\dfrac{c(x)-B}{1-B}\right)^2\right], & c(x) \geq B \end{cases}$$

$$c(x) = \frac{x-0.5}{N}$$

Here, $c(x)$ translates the index of a sub-volume into the normalized distance from the midcell to the center of the sub-volume. In this case, $B = 1$ recovers the cylindrical cell from above, and $B = 0$ produces a spherical cell.

The effects of a nucleoid are introduced in the above model by adding a Gaussian function to $\vec{a}(x)$ while subtracting it from $\overleftarrow{a}(x)$. This anisotropy was parameterized with center $\mu \in [0,1]$, standard deviation *σ,* and height *h*. Specifically:

$$\vec{a}(x) = \frac{N^2 D}{2}\left[\phi(x) + h \cdot \exp\left\{-\frac{(c(x)-\mu)^2}{2\sigma^2}\right\}\right]$$

$$\overleftarrow{a}(x) = \frac{N^2 D}{2}\left[1 - h \cdot \exp\left\{-\frac{(c(x)-\mu)^2}{2\sigma^2}\right\}\right]$$

To fit the models to the measurements, we use the Earth-Mover's metric (2), otherwise known as the first Wasserstein metric (3), defined as:

$$W(F,G) = \int_{-\infty}^{\infty} \left|F(x) - G(x)\right| dx$$

where *F* and *G* are the cumulative distribution functions of the model and the measurements.

To obtain the fraction of complexes moving towards the pole from sub-volume *x*, we first initialized the model with all probability in sub-volume *x*, denoted $\mathbf{P}^x(0)$, and numerically integrated the system over one minute using the Matlab function `ode23s` to obtain the probability distribution at $t = 1$, $\mathbf{P}^x(1)$. This fraction was then calculated as:

$$\frac{1}{2}\mathbf{P}_x^x(1) + \sum_{i=x+1}^{N} \mathbf{P}_i^x(1)$$

Spatial distribution without anisotropy

We constructed a 1-dimensional model with the forward and backwards propensities of diffusion events set to be equal, and inversely proportional to the observed spatial distribution (results in Fig. S4). When both propensities are equal, the probability that a complex will travel in one direction is 0.5, and thus there is no velocity anisotropy. When the propensities are inversely proportional to the observed distribution, in the long term, the complexes tend to linger in the areas where they were observed with high probability. Thus, the long-term spatial distribution is exactly as observed (the lines are indistinguishable in Fig. S4 *B*), while producing a negligible anisotropy in the predicted displacement distribution (Fig. S4 *A*).

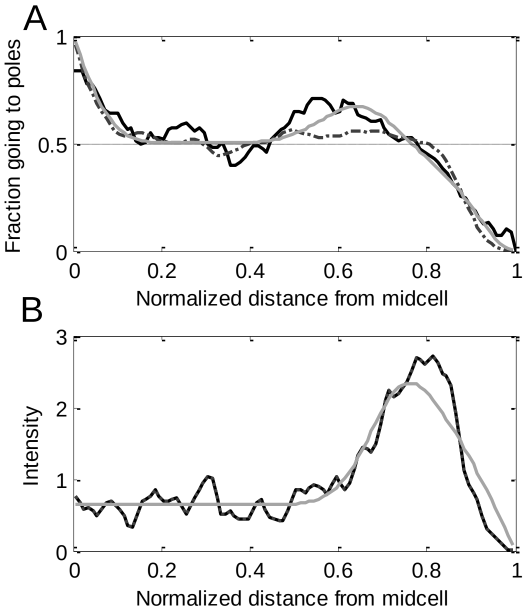


FIGURE S4 (*A*) Predicted fraction of complexes travelling towards the pole from each position along the major axis. (*B*) Long-term spatial distribution of complexes. Results are from a model with a localized anisotropy as in Fig. 8 (*gray lines*), a model with heterogeneous speeds (*dashed lines*), and the measurements at 24°C (*black lines*). Note that the dashed line in (*B*) is superimposed by the black line.

SUPPORTING REFERENCES

1. Gardiner, C.W., K.J. McNeil, D.F. Walls, and I.S. Matheson. 1976. Correlations in Stochastic Theories of Chemical Reactions. *J. Statiscal Phys.* 14: 307.

2. Rubner, Y., C. Tomasi, and L.J. Guibas. 2000. The Earth Mover's Distance as a Metric for Image Retrieval. *Int. J. Comput. Vis.* 40: 99–121.

3. Dobrushin, R.L. 1970. Prescribing a system of random variables by conditional distributions. *Theory Probab. Its Appl.* 15: 458.

Publication III

A. Gupta, J. Lloyd-Price, S. Oliveira, O. Yli-Harja, A.-B. Muthukrishnan, and A. S. Ribeiro (2014). Robustness of the division symmetry in *Escherichia coli* and functional consequences of symmetry breaking. *Physical Biology*, **11**(6):066005.

IOP Publishing Physical Biology
Phys. Biol. **11** (2014) 066005 (12pp) doi:10.1088/1478-3975/11/6/066005

Robustness of the division symmetry in *Escherichia coli* and functional consequences of symmetry breaking

Abhishekh Gupta, Jason Lloyd-Price, Samuel M D Oliveira, Olli Yli-Harja, Anantha-Barathi Muthukrishnan and Andre S Ribeiro

Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, FI-33101 Tampere, Finland

E-mail: andre.ribeiro@tut.fi

Received 12 August 2014, revised 15 October 2014
Accepted for publication 21 October 2014
Published 10 November 2014

Abstract

The morphological symmetry of the division process of *Escherichia coli* is well-known. Recent studies verified that, in optimal growth conditions, most divisions are symmetric, although there are exceptions. We investigate whether such morphological asymmetries in division introduce functional asymmetries between sister cells, and assess the robustness of the symmetry in division to mild chemical stresses and sub-optimal temperatures. First, we show that the difference in size between daughter cells at birth is positively correlated to the difference between the numbers of fluorescent protein complexes inherited from the parent cell. Next, we show that the degree of symmetry in division observed in optimal conditions is robust to mild acidic shift and to mild oxidative stress, but not to sub-optimal temperatures, in that the variance of the difference between the sizes of sister cells at birth is minimized at 37 °C. This increased variance affects the functionality of the cells in that, at sub-optimal temperatures, larger/smaller cells arising from asymmetric divisions exhibit faster/slower division times than the mean population division time, respectively. On the other hand, cells dividing faster do not do so at the cost of morphological symmetry in division. Finally we show that at suboptimal temperatures the mean distance between the nucleoids increases, explaining the increased variance in division. We conclude that the functionality of *E. coli* cells is not immune to morphological asymmetries at birth, and that the effectiveness of the mechanism responsible for ensuring the symmetry in division weakens at sub-optimal temperatures.

S Online supplementary data available from stacks.iop.org/PB/11/066005/mmedia

Keywords: cell division, morphological asymmetry, molecule partitioning, cell division times, nucleoid occlusion

Introduction

Cell division in *Escherichia coli* is known for its morphological symmetry, and sister cells appear to be phenotypically similar under optimal conditions [1–3]. This phenotypic similarity relies on an even distribution of many of the cellular components among all progeny cells [4].

E. coli cells are rod-shaped and grow by elongation under stable growth conditions, with little variation in the width of the minor axis from one generation to the next [1, 5]. Under normal conditions, the constriction plane that defines the point of division [1, 6] is initiated by a septum almost precisely at the midpoint of the longer cell axis [7–9]. The moment of division appears to be strongly correlated with reaching a specific cell length [10–12]. The small variance in both the point and the moment of division led to the division process being considered as largely deterministic [13].

The 'Min system' (Min proteins and FtsZ ring localization) and the positioning of the nucleoids prior to division have been identified as factors affecting the selection of the

© 2014 IOP Publishing Ltd Printed in the UK

point of division [3, 14–18]. In particular, the Min system produces a dynamic distribution of Min proteins whose minimum, at midcell, is believed to be used as a signal to select the site of division [18]. However, measurements suggest that the variability of the location of the minimum in MinD concentration is too high to account for the accuracy of the symmetry in division [18]. As such, there ought to exist other contributing factors. Currently, it is believed that while the Min system places the division point far from the cell poles, i.e. at the midcell region [19], it is nucleoid occlusion that confers the observable degree of precision, i.e. symmetry, to the division process of *E. coli* [14–17]. This is supported by several observations. First, the selection of the point of division in between the two nucleoids appears to be close to random [14]. Also, irregular nucleoid movements affect the angle and position of the constriction plane and thus of the division site [20, 21]. Finally, in both normal and aberrant-shaped cells, there is striking co-localization between the nucleoid-free region at midcell and the division point [22].

However, this system is not perfect. Even in monoclonal cell populations, there is some variance in the division point that results in a small set of morphologically asymmetric divisions (for example see [12], or figure 2(C) of [22]). It is unknown whether these asymmetries in division generate functional asymmetries between sister cells, but there is a strong possibility that this is the case, since *E. coli* is not known to possess mechanisms that could counteract the asymmetries in the partitioning of cellular components that would arise from morphological asymmetries [23, 24].

Even morphologically symmetric division events are expected to not be immune to functional asymmetries arising, e.g. from random errors in the partitioning of RNA and protein molecules [23, 25]. The lowest expected variance, in the absence of a transport mechanism, is obtained by independent partitioning of the molecules resulting in a binomial distribution of the number of inherited molecules, which, for a small number of molecules, can result in large partitioning errors [25]. Provided spatial constraints, the errors and, thus, the functional asymmetries, should accumulate over the generations [26]. For example, unwanted protein aggregates, which accumulate at the cell poles, quickly become non-evenly distributed among the cells [27, 28]. Evidence suggests that the cells inheriting more aggregates have reduced vitality [27–29]. Morphological asymmetries in cell divisions should further enhance these functional asymmetries.

It is known that the morphology of *E. coli* cells is environment-dependent. For example, temperature affects the average cell length [30, 31]. However, so far, little is known about the extent to which the morphological symmetry in the division of *E. coli* cells is robust to non-optimal environmental conditions.

Here, from time-lapse microscopy measurements and using semi-automated image analysis methods, we first verify previous assessments of the degree of morphological symmetry in division in *E. coli* cells under optimal growth conditions. Next, we search for functional asymmetries resulting from morphological asymmetries in division. In particular, we search for asymmetries in the partitioning of protein complexes in cell division. Afterwards, we explore the effects of mild chemical stresses and sub-optimal temperatures on the degree of morphological symmetry in division. Subsequently, we investigate whether morphological asymmetries in division lead to functional asymmetries between sister cells, in particular, asymmetries in their reproduction rate. Finally, we search for mechanisms capable of explaining our observations.

For these studies, we use *E. coli* strain DH5α-PRO [32], since the division rate of these cells is relatively slow [33] which facilitates the automated tracking of cell divisions from time-lapse microscopy images. Also, since the cells are flooded with MS2-GFP proteins, it is possible to extract their shape and size from images taken by confocal microscopy [26]. Further, multiple MS2-GFP proteins bind a specific RNA target, forming MS2-GFP-RNA complexes that can be easily detected, quantified and tracked as they move in the cytoplasm [24, 32, 34]. Finally, these complexes are very long-lived [32], and thus it is possible to track their long-term spatial distributions and partitioning in division, across several generations [26].

Methods

Chemicals

Bacterial cell cultures were grown in Lysogeny Broth (LB) media. The chemical components of LB (Tryptone, Yeast extract and NaCl) were purchased from LabM (Topley House, 52 Wash Lane, Bury, Lancashire, UK) and the antibiotics from Sigma-Aldrich (3050 Spruce Street, St. Louis, MO, USA). Isopropyl β-D-1-thiogalactopyranoside (IPTG), L-arabinose, and anhydrotetracycline (aTc) used for induction of the target genes are from Sigma-Aldrich. Agarose (Sigma-Aldrich) was used for microscope slide gel preparation for cell imaging. 30% Hydrogen peroxide (H_2O_2) and 4-morpholine-methanesulfonic acid (MES) used to induce stress are also from Sigma-Aldrich.

Cells and plasmids

Experiments were conducted on *E. coli* strain DH5α-PRO, generously provided by I. Golding (Baylor College of Medicine, USA), which contains two constructs: (i) PROTET-K133 carrying $P_{LtetO\text{-}1}$-*MS2d-GFP*, and (ii) a pIG-BAC ($P_{lac/ara\text{-}1}$-*mRFP1-MS2-96bs*) vector, carrying a 96 MS2 binding site array under the control of $P_{lac/ara\text{-}1}$ [32].

Induction of production of fluorescent complexes

The dimeric MS2 fused to GFP (MS2-GFP fusion protein) used as a detection tag [32] is expressed from a medium-copy vector under the control of the $P_{LtetO\text{-}1}$ promoter [35], regulated by the tetracycline repressor. The RNA target for MS2-GFP is located on a single-copy F-based vector, and is controlled by the $P_{lac/ara\text{-}1}$ promoter [35], regulated by IPTG and Arabinose.

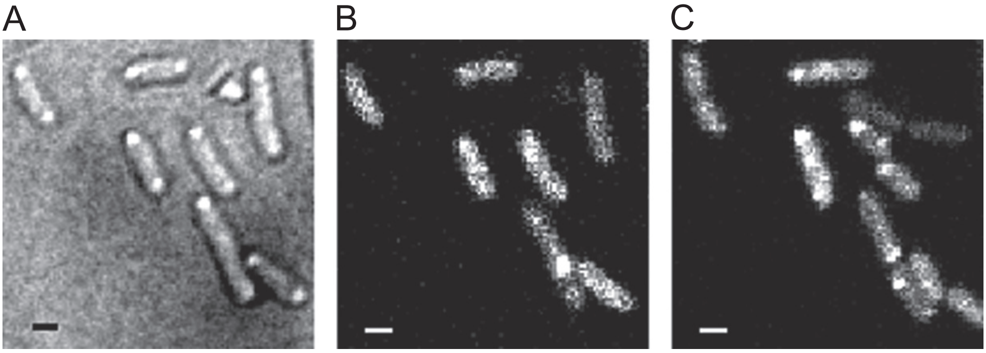


Figure 1. (A) Example bright-field image of cells expressing MS2-GFP and target RNA, from a time-series at $t = 1$ min. (B) Fluorescence channel of the image shown in (A). (C) Fluorescence image from the same time-series at $t = 85$ min, showing the outcome of a few cell divisions and partitioning of complexes. Scale bars are 1 μm. Contrast enhanced for easier visualization.

For our measurements, pre-cultures were diluted from the overnight culture to an OD_{600} of 0.1, in fresh LB media supplemented with the appropriate antibiotics and kept at an appropriate temperature at 250 RPM in a shaker until they reached an $OD_{600} \approx 0.5$. Cells were then induced with 50 ng ml^{-1} of aTc and 0.1% L-arabinose for 45 min, at which point the OD_{600} was ~0.8. Induction of the target RNA production was completed by adding 1 mM IPTG, and cells were incubated for 5 min prior to preparation of the microscope slide. This induction procedure is necessary for cells to accumulate sufficient numbers of MS2-GFP to detect the target RNA and to achieve full induction of the target gene [36, 37]. Usually, in one hour, two to four tagged RNAs are produced by a fully induced cell [38]. Finally, note that the MS2-GFP molecules alone distribute themselves homogeneously throughout the cytoplasm, and only in the presence of target RNAs do fluorescent spots appear [32, 39].

In vivo nucleoid staining

There are several nucleoid associated proteins (NAP) that participate in its structural organization. In *E. coli*, major NAPs include H-NS, HU, Fis, IHF and StpA [40]. The dimeric histone-like protein HU is one of the most abundant [41, 42]. Because of this, we use a version of this protein that has been tagged with mCherry to study the spatial distribution of nucleoids [43]. We are grateful to Nancy Kleckner (Harvard University) for providing us with the plasmid pAB332 carrying hupA-mCherry. For our measurements, we inserted this plasmid into DH5α-PRO cells. This allows the detection of nucleoids in individual cells during the course of live cell microscopy sessions (see figure 3).

Stress conditions

To determine which concentrations of stressors (H_2O_2 [44] or MES [45]) would provide significant, while not lethal, stress responses, we measured growth rates in liquid culture containing various concentrations of H_2O_2 or MES from the OD_{600}, measured every 30 min up to 4.5 h. Once appropriate (sub-lethal but affecting growth rate) concentrations were found (0.6 mM for H_2O_2 and 150 mM for MES), they were applied in the time lapse microscopy measurements, where division times were assessed from tracking individual cells, so as to confirm the effectiveness of the stressors (see Results section 'Effects of chemical stress on the morphological symmetry of the division process').

For fluorescence images of cells under stress, prior to microscopy we used the same culturing protocols as described above (section 'Induction of production of fluorescent complexes'). Stress conditions (oxidative stress and acidic shift) were applied at the start and then maintained during the microscopy measurements by a peristaltic pump that provided a continuous flow of fresh LB media containing either H_2O_2 or MES at the appropriate concentrations, at the rate of 1 ml min^{-1}, through the thermal imaging chamber (Bioptechs, FCS2). In this setting, fresh media and stressors continously diffuse through the agarose. The sub-lethal stress conditions allow long live-cell imaging sessions. The addition of the stress agents to the warmed media began ~2–5 min. prior to the start of the imaging procedure since, upon addition of these agents, the media conditions shift to the desired levels in a matter of seconds.

Imaging fluorescent complexes and morphological asymmetries in division

We placed 4 μl of culture on a 0.8% agarose gel pad of LB media set between a microscope slide and a cover slip. Next, we assembled the thermal imaging chamber (Bioptechs, FCS2) [46]. Time-lapsed florescence images were taken every minute for two hours. We also imaged cells without inducing the reporter gene (coding for MS2-GFP). For that, following the procedure described above, except for the induction of the reporter plasmid, we acquired bright-field images of cells, every minute for two hours, in optimal growth conditions (LB media, 37 °C).

Cells were visualized (figures 1(B) and (C)) using a Nikon Eclipse (TE2000-U, Nikon, Japan) inverted microscope with a 100x Apo TIRF (1.49 NA, oil) objective.

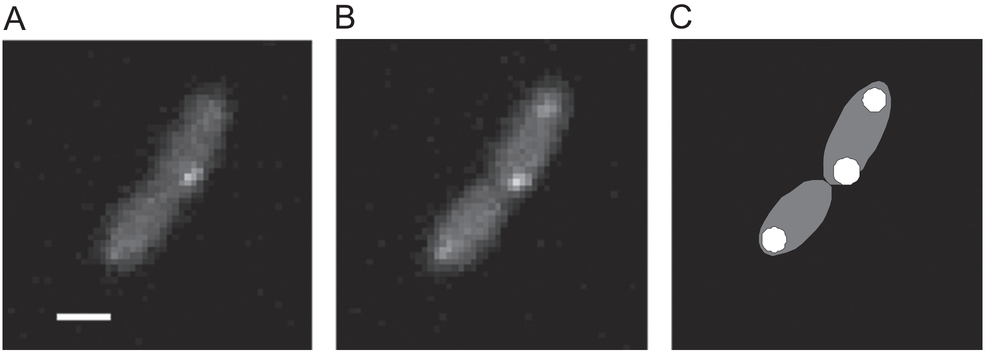


Figure 2. Example images of a cell expressing MS2-GFP and the target RNA, ten minutes before the division was detected (A), and in the frame where the division was detected (B). (C) Image segmentation results for the frame (B), with detected cell ellipses from PCA (gray), and detected MS2-GFP-RNA complexes (white). Scale bar is 1 μm.

Fluorescence images were taken by a C1 confocal laser-scanning system with pixel dwell of 1.33 μs, resulting in a line-scanning time of ~1.4 ms. GFP fluorescence was measured using a 488 nm laser (Melles–Griot) and a band-pass 500–530 nm emission filter (HQ515/30 m, Nikon). The pinhole size was set to 1.2 AU, giving an optical sectioning depth of ~0.83 μm (estimated as in [47]), which suffices to cover the depth of the cells (see, e.g. figure 1 and scale bars within). Image acquisition was performed by Nikon EZ-C1 software. Bright-field images, as seen in figure 1(A), were taken by a CCD camera (DS-Fi2, Nikon, Japan).

When performing time-lapse measurements of mCherry-tagged-nucleoid(s), we used Highly Inclined and Laminated Optical sheet (HILO) microscopy [48]. The imaging was performed using a Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with an EMCCD camera (iXon3 897, Andor Technology) and a 100x Apo TIRF objective (1.49 NA, oil). Nucleoids tagged by hupA-mCherry were observed using a 543 nm HeNe laser (Melles–Griot) and a band-pass 608–683 nm emission filter (Texas Red® filter block, Nikon). Cell borders were detected from Phase-Contrast images (example shown in figure 3(A)), which were captured using a CCD camera (DS-Fi2, Nikon, Japan).

Segmentation of cells and lineage construction from the images

Cells were detected from the time-lapse, confocal microscopy images by a semi-automatic method as in [49]. First, the images were temporally aligned using cross-correlation, to ensure that cells 'move' as little as possible over time. Next, cells were segmented by manually drawing a mask in GIMP (the GNU Image Manipulation Program [50]) over the region that each cell occupied during the time-series. If a division occurred, two new masks were drawn over the new cells in the frame where the division was first observed. Afterwards, at each time point and for each mask, Principal Component Analysis (PCA) of the fluorescence distribution under the mask was used to obtain the position and orientation of the cell. To avoid biasing the centroid of the cell towards the poles with bright spots, we reduced the intensity of the 30 brightest pixels to the intensity of the 30th brightest pixel prior to PCA (we observed by inspection that this removed all such biases). Cell lineages were then constructed as in [26], i.e. each cell at each time moment was assigned, as parent, the cell in the previous frame with the nearest centroid. To avoid incorrectly assigning adjacent cells as the cell's parent, this assignment was done after transforming the previous frame's cell centroids into the cell's space (i.e. poles at (−1, 0) and (1, 0), and sides at (0, −1) and (0, 1)). A division is assumed to have occurred when two cells are assigned the same parent (verified also by inspection).

From the set of temporal images obtained by bright-field microscopy, cells were automatically segmented using CellAging [51]. After the automated segmentation, manual correction was performed to correct any errors. After manual correction, CellAging tracks the cells across time and finds cell divisions. Cell features such as area over time, division time, and parent, among others, are then extracted. The asymmetry in size for each division event is then calculated as: $\Delta S = (S_c - S_s)/(S_c + S_s)$, where S_c and S_s are the areas of the daughter cell of interest and of its sister cell, respectively, immediately after division. This measure, where −1 and +1 denote maximum asymmetries and 0 denotes symmetry, is different by a factor of two from the 'percent difference' used by [22], for example. Example images of a cell expressing MS2-GFP and target RNA that divided asymmetrically are shown in figure 2. This particular division occurred with $|\Delta S| = 0.11$.

From the set of temporal images obtained by phase contrast, cells were automatically segmented using MAMLE [52]. Subsequent analysis (and manual correction of the segmentation and of the lineage construction) was performed using CellAging [51], as above.

Detection of individual fluorescent complexes from the images

We detected fluorescent MS2-GFP-RNA complexes in each cell, at each frame, as in [26, 49]. We segment fluorescent complexes automatically inside each mask with a kernel density estimation (KDE) method for spot detection [53], using a Gaussian kernel. This method measures the local smoothness of the image and determines spot locations by designating areas with low smoothness as spots. The results for the image in figure 2(B) are shown in figure 2(C).

Cell-background-corrected complex intensities were then calculated by subtracting the mean cell background intensity multiplied by the area of the complex from the total fluorescence intensity of each complex. To avoid issues with quantizing the number of complexes in each pole [34], we directly compared the fluorescence intensities of the complexes. The difference in complex intensities between daughter cells at a division event is calculated as: $\Delta I = (I_c - I_s)/(I_c + I_s)$, where I_c and I_s are the total background-corrected complex intensities of the daughter cell of interest and of its sister cell, respectively, immediately after division.

Detection of hupA-mCherry-tagged nucleoids from the images

Nucleoid positions were estimated in cells the moment prior to their morphological separation into sister cells. As such, each cell contains two nucleoids. The centers and widths of the nucleoids were estimated by fitting a Gaussian mixture density distribution with two Gaussians with equal weight to the fluorescence intensity distribution along the major axis of the cells, normalized by half the cell length. The fit was done in a least-squares sense to a KDE of the fluorescence intensity distribution (see example in figure 3), using a Gaussian kernel and a bandwidth of 0.05. The centers of the Gaussians are interpreted to be the centers of the nucleoids. We determined the error resulting from the fit for differing levels of noise in the input data by simulating intensity traces with zero-mean Gaussian noise (see supplementary figure S1). We estimated the signal-to-noise ratio in the measurements to be ~68, resulting in an error in the nucleoid separation measurement of ~0.022.

Results

Symmetry of the division process

We first searched for morphological asymmetries between sister cells by following each division event in cells under optimal growth conditions. We imaged cells for two hours in LB media at 37 °C, using both confocal and bright-field microscopy. In one case, we induced the expression of MS2-GFP and the RNA target for MS2-GFP, while in the other case these were not induced (see Methods). Example images of cells are shown in figures 1(A), (B) and (C) and in supplementary figures S2–S4.

From the images of cells expressing MS2-GFP and the RNA target obtained by confocal microscopy, we extracted the distribution of normalized differences in size (ΔS) between sister cells from the 170 division events observed during the measurement period. The standard deviation of this distribution (see supplementary figure S5(A)) was found to be 0.10, which shows that most divisions are morphologically symmetric, with the bulk of the distribution lying near the origin. Note that this distribution is perfectly symmetric, since each division contributes with a positive and a negative value of ΔS.

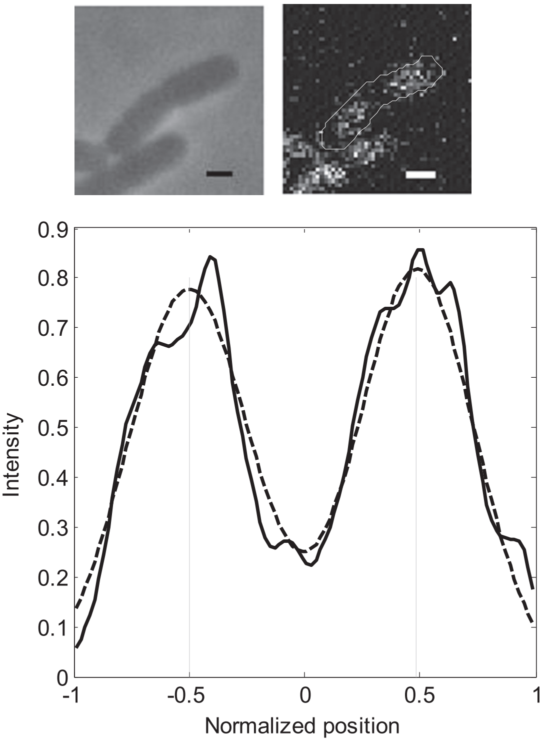


Figure 3. Example image of a cell expressing hupA-mCherry, in the frame before division is detected, taken with (top left) phase contrast microscopy to visualize the cell borders, and (top right) confocal microscopy to visualize the fluorescence intensity of hupA-mCherry (cell outline obtained from (top left) is shown in white). (Bottom) KDE of the fluorescence intensity distribution (solid line) along the major axis of the cell shown in (top right), with the Gaussian mixture model fit (dashed line). Scale bars are 1 μm.

Next, to determine if the size asymmetries can be reliably measured from the fluorescence images, we obtained the same distribution of morphological asymmetries between sister cells from bright-field images of cells in the same conditions (figure S5(B)). In these, 97 divisions were detected. The standard deviation of this distribution is 0.13, which is similar to the standard deviation obtained from the fluorescence images. Finally, to determine if the production of MS2-GFP alters the distribution of size asymmetries, we obtained this distribution from bright-field images of cells, without inducing MS2-GFP (figure S5(C)). In these measurements, 130 divisions were detected.

We compared the three distributions (figure S5). In doing so, we considered only the positive side of each distribution since the Kolmogorov–Smirnov (K–S) test assumes that all samples are independent and, in our case, the right and left sides of the distribution are necessarily identical. The K–S

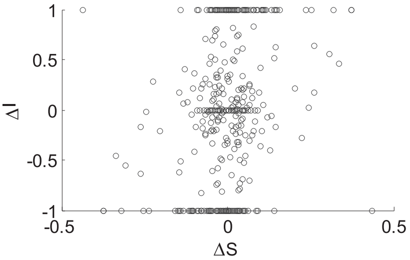


Figure 4. Asymmetry in inherited complexes (ΔI) versus asymmetry in size (ΔS) between sister cells the moment following division. Data extracted from 170 divisions in optimal growth conditions (LB media, 37 °C).

tests resulted in p-values of 0.74 and 0.84 between figures S5 (A) and (B), and between figures S5(B) and (C), respectively. We conclude that they cannot be distinguished in a statistical sense (usually, for p-values smaller than 0.05, it is accepted that two distributions differ significantly).

These measurements confirm previous reports that the division of *E. coli* cells is, in general, morphologically symmetric under optimal conditions [1, 2], though some highly asymmetric divisions can occur (2.3% of divisions exhibited $|\Delta S| > 1/3$). It is also noted that the distributions in figure S5 are strikingly similar to the one reported in [22]. Finally, we find that the expression of MS2-GFP and/or RNA target does not affect the symmetry in the division of *E. coli* cells.

Functional effects of morphological asymmetries: partitioning of complexes in cell division under optimal growth conditions

In figure S5, we observed that a few divisions were morphologically asymmetric. It is unknown whether pronounced morphological asymmetries in division introduce functional asymmetries between sister cells. It is reasonable to expect some, particularly from large morphological asymmetries, due to the expected differences in the numbers of components inherited by the two daughter cells from the mother cell.

We first searched for functional asymmetries between sister cells by analyzing the partitioning in division of MS2-GFP-RNA complexes as these are easy to detect and quantify [34]. Further, these complexes are highly stable [32] in that they are ‘immortal’ for periods much longer than the cells’ division times [54]. Consequently, when cells divide, the complexes are partitioned between the daughter cells [26]. Finally, it has been established that these complexes first appear close to midcell, but are quickly segregated to the poles, where they tend to be retained [55]. The choice of pole is known to be unbiased, under both optimal and sub-optimal temperatures [26].

From the 170 division events observed, we extracted ΔS, the normalized difference in size at birth, as well as ΔI, the difference in intensity of the complexes inherited by sister cells the moment following the division of their mother cell. The relationship between these quantities is shown in figure 4. Since the distribution of ΔI is far from normal, we computed the Kendall’s tau correlation [56] between ΔI and ΔS in each division event (see Methods). We found a positive correlation of 0.19 (the p-value of the null hypothesis that there is no relationship between ΔS and ΔI was 0.0004). From this, we conclude that larger cells tend to inherit more complexes, even though most of these are retained at the poles. As such, we also conclude that, as expected, asymmetries in division have tangible functional consequences.

Note also from figure 4 the stochasticity in the localization of the complexes: although overall there is a positive correlation between cell size and the intensity of its inherited complexes, one can also observe that, in a few cases, the smaller cell inherited all complexes. As noted, such events are possible because of the low number of complexes (see [49]) and the somewhat stochastic nature of their positioning in the cell [26].

Effects of chemical stress on the morphological symmetry of the division process

To assess the effects of acidic shift and oxidative stress, we selected concentrations of MES and H_2O_2, respectively, that produced mild stress conditions (see Methods) [57]. To quantify the degree of stress during the measurements, we measured the division time of cells born during the measurement period. Example images are shown in supplementary figures S6 and S7. For comparison, we obtained the same data from cells under optimal growth conditions (control).

The mean cell division time was found to be 66.7 min under optimal conditions, 87.2 min under acidic shift, and 91.0 min under oxidative stress, indicating that the cells are mildly stressed in the two latter conditions. To confirm this, we performed K–S tests to determine whether the distributions of division times in both stress conditions could be statistically distinguished from the distribution obtained from control cells. In both comparisons, the p-value was much smaller than 0.05, from which we conclude that the division times differed from the control.

Next, we studied the effects of the two mild chemical stresses on the degree of morphological asymmetries in division. In both, we extracted the distribution of the normalized difference in the sizes of sister cells at birth. The resulting distributions, along with the result from optimal conditions, are shown in figure 5. The standard deviations of the distributions were 0.095 and 0.099 for oxidative and acidic shift stresses, respectively. Meanwhile, in cells under optimal conditions the standard deviation was 0.10. In agreement with the apparent similarity, we could not distinguish between any of these distributions using the K–S test (p-values > 0.05). Thus, we conclude that the distributions of cells under these two mild chemical stresses do not differ, in a statistical sense, from the distribution of cells under optimal conditions. Consequently, we conclude that the degree of symmetry of division is not significantly affected by either of these mild chemical stresses.

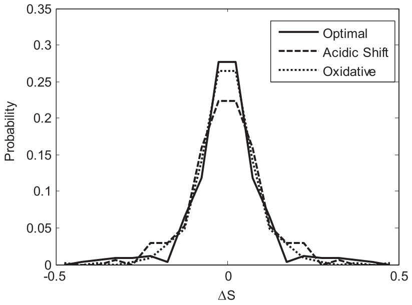


Figure 5. Distributions of asymmetry in size of sister cells at birth (ΔS), obtained from cells under acidic shift, oxidative stress, and optimal conditions. Data is from 269 divisions under oxidative stress (dotted line), 85 divisions under acidic shift (dashed line), and 170 divisions of cells under optimal conditions (solid line).

Effects of suboptimal temperatures on the morphological symmetry of the division process

We imaged cells for two hours in LB media at 42 °C, 37 °C (optimal), 33 °C, 30 °C, and 24 °C. From the images, we extracted the distributions of normalized difference in size between the daughter cells from all divisions observed in each condition (figure 6(A)).

The standard deviation of the distribution appears to increase for temperatures further from optimal (figure 6(B)). We performed K–S tests to determine if these changes are statistically significant. We found that distributions for temperatures of 30 °C and below are statistically distinguishable from the distribution for 37 °C. Also, the distribution at 42 °C differs significantly from the distributions at 37 °C and below (all $\text{p-values} < 0.05$). We conclude that changing the temperature from optimal to sub-optimal decreases the degree of morphological symmetry in cell division.

Effects of morphological asymmetries on cell division times

If there are significant functional asymmetries between sister cells due to morphological asymmetries at birth, their division times should differ, since cells that are healthier or that possess more resources tend to divide faster (see e.g. [27]). To test this, from the data used in the previous section, for each temperature condition, we extracted the asymmetry in size at birth (ΔS) and subsequent division time (τ_{div}) of each cell that was born and then divided during the measurements. Then, given the definition of ΔS (see Methods), we tested whether relatively larger cells tend to divide faster on average. That is, we searched for statistically significant negative correlations between ΔS and τ_{div}. Note that all cells considered necessarily participate in two division events, first as a daughter and then as a parent.

The correlation between asymmetries in size at birth and subsequent division times is shown in table 1 (Kendall's tau between ΔS and τ_{div}), for each temperature condition, along with the number of cells analyzed. As temperature decreases, the correlation (negative) becomes statistically significant ($\text{p-value} < 0.05$ for 30 °C and lower temperatures) and gradually increases, implying that relatively larger daughter cells divide faster than the average division time of the population. The most significant increase is between 30 °C and 24 °C. Further, at 42 °C, the correlation also becomes significant. Overall, we conclude that for temperatures beyond the optimal temperature, the two quantities become significantly anticorrelated. The degree of anticorrelation increases as the temperature gets further from optimal.

Additionally, we examined whether the division time of each daughter cell is related to the degree of morphological asymmetry in the next division, i.e. we assessed whether faster dividing cells tend to exhibit larger morphological asymmetries in division. For this, we extracted the division times of each cell (τ_{div}) and the magnitude of the asymmetry in its division (referred to as $|\Delta S'|$). The results in table 1 (last column) show no significant correlation for any temperature condition. Overall, these results suggest that the faster dividing cells do not do so at the cost of morphological symmetry in division.

Finally, we searched for a relationship between ΔS and differences between the τ_{div}'s of sister cells. That is, we assessed if asymmetries in division are correlated with subsequent asymmetries in division times. However, we did not find such a relationship in the available data.

Mechanisms underlying the morphological asymmetries under sub-optimal temperatures

When the Min system is disrupted, it is common to observe the appearance of 'minicells' [3]. Since we did not observe minicells in any temperature condition, we hypothesized that the increased variances in the division point at sub-optimal temperatures were either due to higher variability in the cell sizes or due to increased relative distance between nucleoids.

To test the former possibility, we extracted the distribution of non-normalized asymmetry in sizes between sister cells at 24 °C, 37 °C and 42 °C (supplementary figure S9). The K–S tests to distinguish between these distributions resulted in $\text{p-values} < 0.05$, which implies that they differ between conditions. Since we obtained the same conclusion using the normalized values (figure 6(A)), we next investigate the differences in relative distance between nucleoids.

To test this, we made use of hupA-mCherry to visualize the nucleoids *in vivo* [43] and measured the positions and sizes of the two nucleoids in each cell in the moment prior to division, at 24 °C, 37 °C and 42 °C (see Methods). First, we saw no significant change in relative nucleoid size with temperature (data not shown). Next, we tested for changes in the mean relative distance between nucleoids. For simplicity, as their size was unaffected, we measured the distance between nucleoid centers in individual cells in each condition.

In figure 7, for each cell in each condition, we show the position of the center of each of the nucleoids along the major cell axis, the moment prior to division (normalized by half the cell length). In addition, in table 2 we show the mean distance between the centers of the nucleoids in each cell, along with the standard deviation, and the p-value of the t-test to distinguish between the means of pairs of conditions.

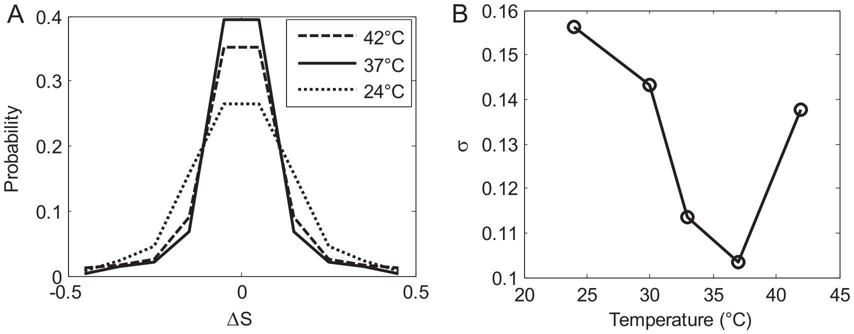


Figure 6. (A) Distributions of asymmetries in the size of sister cells at birth, obtained from cells at 42 °C, 37 °C and 24 °C. (B) Standard deviations of the distributions of asymmetries in size against temperature. Data from 231 divisions at 42 °C, 170 divisions at 37 °C (optimal) 545 divisions at 33 °C, 197 divisions at 30 °C, and 280 divisions at 24 °C.

Table 1. For each condition, the table shows the temperature, the number of cells observed that divided at least twice, the correlation between the asymmetry in size at birth (ΔS) and the subsequent division time (τ_{div}), and between the division time (τ_{div}) and the magnitude of the subsequent asymmetry in size in division ($|\Delta S'|$). The p-value of the null hypothesis that there is no relationship (i.e. that Kendall's tau is 0) is shown in parentheses. Scatterplots of the relationships between these variables are shown in supplementary figure S8.

Temperature	No. of cells	Kendall's tau between ΔS and τ_{div}	Kendall's tau between τ_{div} and $\lvert\Delta S'\rvert$
42 °C	133	−0.13(0.04)	0.06(0.28)
37 °C	75	−0.12(0.13)	0.03(0.72)
33 °C	182	−0.12(0.02)	−0.09(0.07)
30 °C	86	−0.18(0.02)	−0.06(0.39)
24 °C	17	−0.41(0.03)	−0.35(0.09)

From figure 7 and table 2, the mean distance between the centers of the nucleoids in each cell is visibly minimized at 37 °C, and is consistent with the lesser variance in the division point. Also the smaller standard deviation of the distances between the nucleoids in each cell is in agreement with such lesser variance. In addition, the t-tests in table 2 show that the distributions of positions are statistically distinct, except between 24 °C and 42 °C, as expected from the equality of the means in these two conditions. Finally, note that there is no overlap between the confidence intervals of the means (column 2 in table 2) from different conditions (these intervals include the variability in the centers' position arising from biological sources or from any uncertainty in the method of determination of these positions from the images).

It also interesting to note that, assuming that the border of a nucleoid is one standard deviation away from the center (based on the Gaussian fit), it is possible to determine that the measured increase of 5% in the mean normalized distance between the nucleoid centers (at sub-optimal temperatures) corresponds to an ~8% increase in the mean normalized distance between the borders of the two nucleoids. This increase should suffice to explain the observed increase in the errors in positioning of the division plane at sub-optimal temperatures (figure 6).

Overall, these results are in agreement with the hypothesis that the source of increased asymmetry in the division process of *E. coli* cells at sub-optimal temperatures is the increased relative distances between nucleoids (see schematic figure 8).

To further assess whether the distance between nucleoids at the moment of division influences the degree of randomness (and thus asymmetry) in the selection of the point of division, for each temperature condition we compared the mean normalized distances between nucleoid centers the moment before division between the 25% cells exhibiting higher asymmetry in division (ΔS) and the remaining 75% of the cells. We found that, in all conditions, the cells with larger mean distances between nucleoid centers at the moment of division also exhibit higher mean degrees of morphological asymmetry in division, in agreement with our hypothesis. In particular, in each condition, the mean normalized distances between nucleoid centers of cells with larger and smaller morphological asymmetries in division equaled, respectively, 1.02 and 0.99 at 42 °C, 0.97 and 0.94 at 37 °C, and, finally, 1.01 and 0.99 at 24 °C.

As a side note, the results in figure 7 and table 2 depend on the accuracy of the determination of the centers of the nucleoids. Unfortunately, such accuracy is too complex to calculate as it depends on numerous variables (e.g. noise in fluorescence of hupA-mCherry, noise in mCherry localization, noise from the fluorescence detector device, accuracy of the image processing, goodness of the model fit, among other). Instead, as the centers are obtained from the models that are fitted to the fluorescence intensity curves (figure 3), we estimated how well these models fit the data by obtaining the median of the coefficient of determination obtained from each cell, for each condition (last column in table 2). From these results, we find that the model fits the data well, in that it explains at least 70% of the observed variance in the fluorescence intensities along the major axis of the cells as obtained from the hupA-mCherry staining.

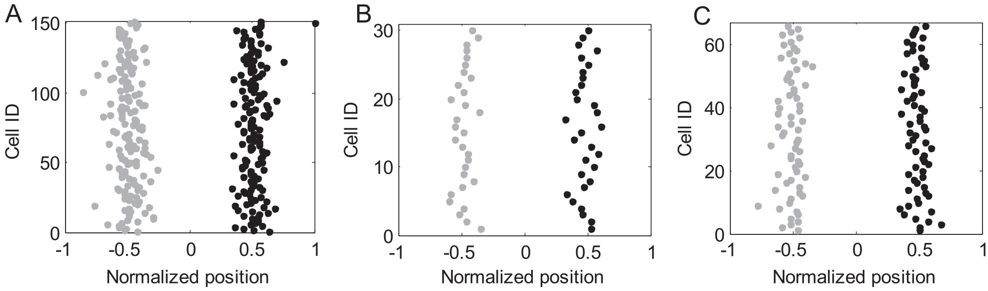


Figure 7. Position of the two nucleoids along the major cell axis in individual cells, the moment prior to division. Positions were normalized by half the length of the major axis of each cell. Results are shown for cells at (A) 42 °C, (B) 37 °C and (C) 24 °C.

Table 2. For each condition, the temperature and the mean and standard deviation of the distance (normalized by half the cell length) between the centers of the nucleoids are shown. Finally, the p-value of the t-test between pairs of conditions is shown (conditions compared shown in parentheses).

Temperature (°C)	Mean normalized distance	Standard deviation of normalized distances	p-value of t-test between the mean distances	Median R-squared
42	1.00 ± 0.008	0.098	0.0076 (42 °C vs 37 °C)	0.75
37	0.95 ± 0.012	0.067	0.0099 (37 °C vs 24 °C)	0.88
24	1.00 ± 0.009	0.077	0.5551 (24 °C vs 42 °C)	0.72

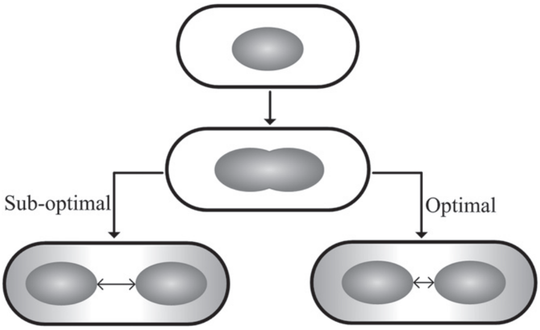


Figure 8. Schematic representation of the role of the nucleoids in the definition of the point of cell division. Two situations are represented: optimal and sub-optimal. Under the optimal temperature, the relative distance between nucleoids is minimal, which minimizes the uncertainty of the point of division. Meanwhile, under sub-optimal temperatures, the larger relative distance will result in distributions of normalized asymmetry in the size of sister cells at birth with greater variance, while the mean division point remains unaffected. Also represented (grey gradient in background) is an illustration of the regions (darker polar regions) where the formation of the point of division is inhibited by the Min system.

Discussion

Previous studies showed that, in optimal conditions, the division process of *E. coli* cells is highly symmetric, with only a few divisions exhibiting morphological asymmetry. This symmetry is made possible by at least two mechanisms, the Min system and nucleoid occlusion. The existence of regulatory mechanisms responsible for achieving a near-symmetric division suggests that such morphological symmetry in division is advantageous to *E. coli*, at least in optimal growth conditions. However, little is known of how the symmetry of this process is affected by physiological stress. Such effects, which we observed here to be tangible, may influence the survival of these organisms under such conditions and, thus, their pathogenicity or antibiotic-resistance.

Here, we first investigated whether morphological asymmetries in division are a source of functional asymmetries between sister cells under optimal conditions. Our measurements showed that morphological asymmetries between sister cells are strongly, positively correlated to the differences in the numbers of MS2-GFP-RNA complexes that these cells inherit from the mother cell. As such, we expect morphological asymmetries in division to also create asymmetries in the partitioning of cellular components, which should create asymmetries in the division times [12].

Next, we studied the extent to which the morphological symmetry in the division of *E. coli* cells is robust to non-optimal environmental conditions. First, we studied whether mild chemical stresses (acidic shift and oxidative stress) affected the degree of morphological symmetry. Even though the growth rate of the cells was significantly hampered due to stress, the distributions of asymmetry in sizes between sister cells were statistically not different from those at optimal growth conditions. A previous study made similar conclusions concerning cells subject to physical stress [22]. One plausible explanation is that the mild chemical stresses applied here did not affect the mechanisms associated with the degree of symmetry in division (nucleoid occlusion and the Min system). Afterwards, we studied how sub-optimal temperatures affected the morphological symmetry in division. The frequency of asymmetric divisions was found to increase as temperature changed from optimal to sub-optimal. From this, we conclude that the mechanism responsible for the symmetry observed in division is not robust to temperature changes.

Relevantly, we also found a statistically significant correlation between asymmetries in size at birth and subsequent division times. This is in agreement with [12], where it was reported that the size of an *E. coli* cell affects when it divides. We hypothesize that this correlation might explain how the precision in division at optimal temperatures evolved. Namely, if there is no need for cells to adopt a bet-hedging strategy under optimal conditions, then symmetry in division is advantageous. At sub-optimal temperatures, the decrease in precision of the point of division may also be beneficial, in that it allows a cell to hedge its bets towards one of the daughter cells, which would explain why the mechanisms determining the division point have not evolved robust temperature resistance. However, it is also possible that such temperature-resistance is not possible, due to the nature of the mechanisms responsible for the symmetry.

It is worth mentioning that we found no correlation between the division time of a cell and the degree of asymmetry in size in its next division (τ_{div} and $|\Delta S'|$ above). This suggests that the mechanisms responsible for symmetry in division are not significantly affected by a cell delaying or accelerating its moment of division. This might not be surprising, given the properties of the Min system (oscillation period of 40 s [58]) and of the time it may take the nucleoid(s) to change position with changing temperature [43].

Finally, we investigated the physical mechanism responsible for the observed increase in variance in the point of division. Our results suggest a simple explanation. Namely, at suboptimal temperatures, while the relative sizes of the nucleoids are unchanged, the mean relative distance between the two nucleoids increases. Provided that the positioning of the division point is near random in between nucleoids [14], this increase in the mean relative distance necessarily increases the variance in the division point. Future studies on why temperature changes affected the nucleoid's positioning should provide valuable information on the mechanisms responsible for this positioning. It is likely that such a study will require analyzing the potential relationship between growth dynamics, nucleoid positioning, and variability of the point of division, as a function of temperature.

Finally, it is worth noting that we do not expect the observed morphological asymmetries in division to generate significant asymmetries in numbers of inherited molecules between sister cells when, for example, these molecules exist in large numbers in the cells, or when they are evenly located at the cell poles. However, significant asymmetries in numbers of inherited molecules due to the morphological asymmetries in divisions (of a degree reported here) ought to occur when, first, the molecule exists in small numbers and, second, either it distributes itself homogenously throughout the cytoplasm, or it preferentially locates at midcell. Relevantly, several RNA molecules appear to fit both conditions. Namely, usually they exist in numbers ranging from 1 to a few [59, 60] and are located either at midcell, or randomly throughout the cytoplasm [61–63].

We conclude that the functionality of *E. coli* cells is not immune to asymmetries at birth, which may explain why *E. coli* has evolved a robust process of symmetric division under optimal conditions. We further conclude that, even though it is based on at least two mechanisms, this process lacks robustness at sub-optimal temperatures, leading to larger asymmetries, and consequently, increased population diversity in cell functionality, including vitality.

Acknowledgments

This work was supported by the Academy of Finland (grant no. 257603) (ASR, SO), the Portuguese Foundation for Science and Technology (grant no. PTDC/BBB-MET/1084/2012) (ASR), the Finnish Cultural Foundation (AG) and the TUT President's graduate programme (JLP, AB). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Marr A G, Harvey R J and Trentini W C 1966 Growth and division of *Escherichia coli J. Bacteriol.* **91** 2388–9

[2] Trueba F J 1982 On the precision and accuracy achieved by *Escherichia coli* cells at fission about their middle *Arch. Microbiol.* **131** 55–9

[3] Yu X C and Margolin W 1999 FtsZ ring clusters in min and partition mutants: role of both the Min system and the nucleoid in regulating FtsZ ring localization *Mol. Microbiol.* **32** 315–26

[4] Lin E C C, Hirota Y and Jacob F 1971 On the process of cellular division in Escherichia coli VI. Use of a methocel-autoradiographic method for the study of cellular division in *Escherichia coli J. Bacteriol.* **108** 375–85

[5] Trueba F J and Woldringh C L 1980 Changes in cell diameter during the division cycle of *Escherichia coli J. Bacteriol.* **142** 869–78

[6] Errington F P, Powell E O and Thompson N 1965 Growth characteristics of some gram-negative bacteria *J. Gen. Microbiol.* **39** 109–23

[7] Cullum J and Vicente M 1978 Cell Growth and length distribution in *Escherichia coli J. Bacteriol.* **134** 330–7

[8] Koch A L and Schaechter M 1962 A model for statistics of the cell division process *J. Gen. Microbiol.* **29** 435–54

[9] Sullivan S M and Maddock J R 2000 Bacterial division: finding the dividing line *Curr. Biol.* **10** R249–52

[10] Donachie W D, Begg K J and Vicente M 1976 Cell length, cell growth and cell division *Nature* **264** 328–33

[11] Donachie W D and Begg K J 1996 'Division potential' in *Escherichia coli J. Bacteriol.* **178** 5971–6

[12] Osella M, Nugent E and Cosentino LagoMarsino M 2014 Concerted control of *Escherichia coli* cell division *Proc. Natl. Acad. Sci. USA* **111** 3431–5

[13] Koch A L 1966 On evidence supporting a deterministic process of bacterial growth *J. Gen. Microbiol.* **43** 1–5

[14] Mulder E and Woldringh C L 1989 Actively replicating nucleoids influence positioning of division sites in *Escherichia coli* filaments forming cells lacking DNA *J. Bacteriol.* **171** 4303–14

[15] Wang X, Possoz C and Sherratt D J 2005 Dancing around the divisome: asymmetric chromosome segregation in *Escherichia coli Genes Dev.* **19** 2367–77
[16] Woldringh C, Mulder E, Valkenburg J, Wientjes F, Zaritsky A and Nanninga N 1990 Role of the nucleoid in the toporegulation of division *Res. Microbiol.* **141** 39–49
[17] Woldringh C L, Mulder E, Huls P G and Vischer N 1991 Toporegulation of bacterial division according to the nucleoid occlusion model *Res. Microbiol.* **142** 309–20
[18] Kerr R A, Levine H, Sejnowski T J and Rappel W-J 2006 Division accuracy in a stochastic model of Min oscillations in *Escherichia coli Proc. Natl. Acad. Sci. USA* **103** 347–52
[19] Sun Q and Margolin W 2004 Effects of perturbing nucleoid structure on nucleoid occlusion-mediated toporegulation of FtsZ ring assembly *J. Bacteriol.* **186** 3951–9
[20] Valkenburg J A and Woldringh C L 1984 Phase separation between nucleoid and cytoplasm in *Escherichia coli* as defined by immersive refractometry *J. Bacteriol.* **160** 1151–7
[21] Woldringh C L, Zaritsky A and Grover N B 1994 Nucleoid partitioning and the division plane in *Escherichia coli J. Bacteriol.* **176** 6030–8
[22] Männik J, Wu F, Hol F J H, Bisicchia P, Sherratt D J, Keymer J E and Dekker C 2012 Robustness and accuracy of cell division in *Escherichia coli* in diverse cell shapes *Proc. Natl. Acad. Sci. USA* **109** 6957–69562
[23] Huh D and Paulsson J 2011 Random partitioning of molecules at cell division *Proc. Natl. Acad. Sci. USA* **108** 15004–9
[24] Golding I and Cox E 2006 Physical nature of bacterial cytoplasm *Phys. Rev. Lett.* **96** 098102
[25] Huh D and Paulsson J 2011 Non-genetic heterogeneity from stochastic partitioning at cell division *Nat. Genet.* **43** 95–9
[26] Gupta A, Lloyd-Price J, Neeli-Venkata R, Oliveira S M D and Ribeiro A S 2014 *In Vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli Biophys. J.* **106** 1928–37
[27] Stewart E J, Madden R, Paul G and Taddei F 2005 Aging and death in an organism that reproduces by morphologically symmetric division *PLOS Biol.* **3** 0295–300
[28] Lindner A B, Madden R, Demarez A and Stewart E J 2008 Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation *Proc. Natl. Acad. Sci. USA* **105** 3076–81
[29] Ackermann M, Stearns S C and Jenal U 2003 Senescence in a bacterium with asymmetric division *Science* **300** 1920
[30] Maclean F I and Munson R J 1961 Some environmental factors affecting the length of *Escherichia coli* organisms in continuous cultures *J. Gen. Microbiol.* **25** 17–27
[31] Begg K J and Donachie W D 1985 Cell shape and division in *Escherichia coli*: experiments with shape and division mutants *J. Bacteriol.* **163** 615–22
[32] Golding I and Cox E C 2004 RNA dynamics in live *Escherichia coli* cells *Proc. Natl. Acad. Sci. USA* **101** 11310–5
[33] Jung S-C, Smith C L, Lee K-S, Hong M-E, Kweon D-H, Stephanopoulos G and Jin Y-S 2010 Restoration of growth phenotypes of *Escherichia coli* DH5alpha in minimal media through reversal of a point mutation in *purB. Appl. Environ. Microbiol.* **76** 6307–9
[34] Häkkinen A, Kandhavelu M, Garasto S and Ribeiro A S 2014 Estimation of fluorescence-tagged RNA numbers from spot intensities *Bioinformatics* **30** 1146–53
[35] Lutz R and Bujard H 1997 Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I 1 -I 2 regulatory elements *Nucleic Acids Res.* **25** 1203–10
[36] Mäkelä J, Kandhavelu M, Oliveira S M D, Chandraseelan J G, Lloyd-Price J, Peltonen J, Yli-Harja O and Ribeiro A S 2013 *In vivo* single-molecule kinetics of activation and subsequent activity of the arabinose promoter *Nucleic Acids Res.* **41** 6544–52
[37] Golding I, Paulsson J, Zawilski S M and Cox E C 2005 Real-time kinetics of gene activity in individual bacteria *Cell* **123** 1025–36
[38] Kandhavelu M, Mannerström H, Gupta A, Häkkinen A, Lloyd-price J, Yli-Harja O and Ribeiro A S 2011 *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.* **5** 149
[39] Lloyd-Price J, Lehtivaara M, Kandhavelu M, Chowdhury S, Muthukrishnan A-B, Yli-Harja O and Ribeiro A S 2012 Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli Mol. Biosyst.* **8** 565–71
[40] Dillon S C and Dorman C J 2010 Bacterial nucleoid-associated proteins, nucleoid structure and gene expression *Nat. Rev. Microbiol.* **8** 185–95
[41] Claret L and Rouviere-Yaniv J 1997 Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival *J. Mol. Biol.* **273** 93–104
[42] Azam T A, Iwata A, Nishimura A and Ueda S 1999 Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid *J. Bacteriol.* **181** 6361–70
[43] Fisher J K, Bourniquel A, Witz G, Weiner B, Prentiss M and Kleckner N 2013 Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells *Cell* **153** 882–95
[44] Imlay J A and Linn S 1987 Mutagenesis and stress responses induced in *Escherichia coli* by hydrogen peroxide *J. Bacteriol.* **169** 2967–76
[45] Weber H, Polen T, Heuveling J, Wendisch V F and Hengge R 2005 Genome-wide analysis of the general stress response network in *Escherichia coli*: sigma S -dependent genes, promoters, and sigma factor selectivity *J. Bacteriol.* **187** 1591–603
[46] *Bioptechs Focht Chamber System (FCS2)* 2014 www.bioptechs.com/Products/FCS2/fcs2.html
[47] Park J, Choi C and Kihm K 2004 Optically sliced micro-PIV using confocal laser scanning microscopy (CLSM) *Exp. Fluids* **37** 105–19
[48] Tokunaga M, Imamoto N and Sakata-sogawa K 2008 Highly inclined thin illumination enables clear single-molecule imaging in cells *Nat. Methods* **5** 159–61
[49] Kandhavelu M, Lloyd-Price J, Gupta A, Muthukrishnan A-B, Yli-Harja O and Ribeiro A S 2012 Regulation of mean and noise of the *in vivo* kinetics of transcription under the control of the lac/ara-1 promoter *FEBS Lett.* **586** 3870–5
[50] *GIMP—The GNU Image Manipulation Program* 2014 www.gimp.org
[51] Häkkinen A, Muthukrishnan A-B, Mora A, Fonseca J M and Ribeiro A S 2013 CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli Bioinformatics* **29** 1708–9
[52] Chowdhury S, Kandhavelu M, Yli-Harja O and Ribeiro A S 2013 Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE) *BMC Bioinformatics* **14** S8
[53] Ruusuvuori P, Aijö T, Chowdhury S, Garmendia-Torres C, Selinummi J, Birbaumer M, Dudley A M, Pelkmans L and Yli-Harja O 2010 Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images *BMC Bioinformatics* **11** 248
[54] Muthukrishnan A-B, Kandhavelu M, Lloyd-Price J, Kudasov F, Chowdhury S, Yli-Harja O and Ribeiro A S 2012 Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells *Nucleic Acids Res.* **40** 8472–83

[55] Lloyd-Price J, Häkkinen A, Kandhavelu M, Marques I J, Chowdhury S, Lihavainen E, Yli-Harja O and Ribeiro A S 2012 Asymmetric disposal of individual protein aggregates in *Escherichia coli*, one aggregate at a time *J. Bacteriol.* **194** 1747–52

[56] Kendall M G 1938 A new measure of rank correlation *Biometrika* **30** 81–93

[57] Muthukrishnan A-B, Martikainen A, Neeli-Venkata R and Ribeiro A S 2014 *In Vivo* transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells *PLoS One* **9** e109005

[58] Raskin D and Boer P D 1999 Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli Proc. Natl. Acad. Sci. USA* **96** 4971–6

[59] Taniguchi Y, Choi P J, Li G-W, Chen H, Babu M, Hearn J, Emili A and Xie X S 2010 Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells *Science* **329** 533–8

[60] Bernstein J A, Khodursky A B, Lin P, Lin-chao S and Cohen S N 2002 Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays *PNAS* **99** 9697–702

[61] Llopis P M, Jackson A F, Sliusarenko O, Surovtsev I, Heinritz J, Emonet T and Jacobs-wagner C 2010 Spatial organization of the flow of genetic information in bacteria *Nature* **466** 77–81

[62] Valencia-Burton M, Mccullough R M, Cantor C R and Broude N E 2007 RNA visualization in live bacterial cells using fluorescent protein complementation *Nat. Methods* **4** 421–7

[63] Guet C C, Bruneaux L, Min T L, Siegal-Gaskins D, Figueroa I, Emonet T and Cluzel P 2008 Minimally invasive determination of mRNA concentration in single living bacteria *Nucleic Acids Res.* **36** e73

Supplementary material for “Robustness of the division symmetry in *Escherichia coli* and functional consequences of symmetry breaking” by A. Gupta et al.

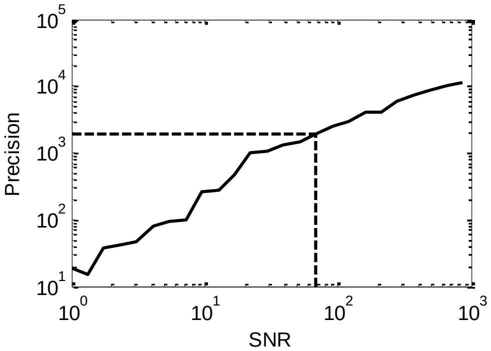


Figure S1: Precision (inverse of variance) of the estimate of the nucleoid separation as a function of the Signal-to-Noise Ratio (SNR), assuming zero-mean Gaussian noise. Data is from 100 simulated intensity distributions with a nucleoid separation of 0.95 and widths (standard deviations of the Gaussians) of 0.3. The dashed line represents the precision for the estimate of the real SNR, calculated from the measured intensity distributions at 37°C and the Gaussian mixture model fits to them.

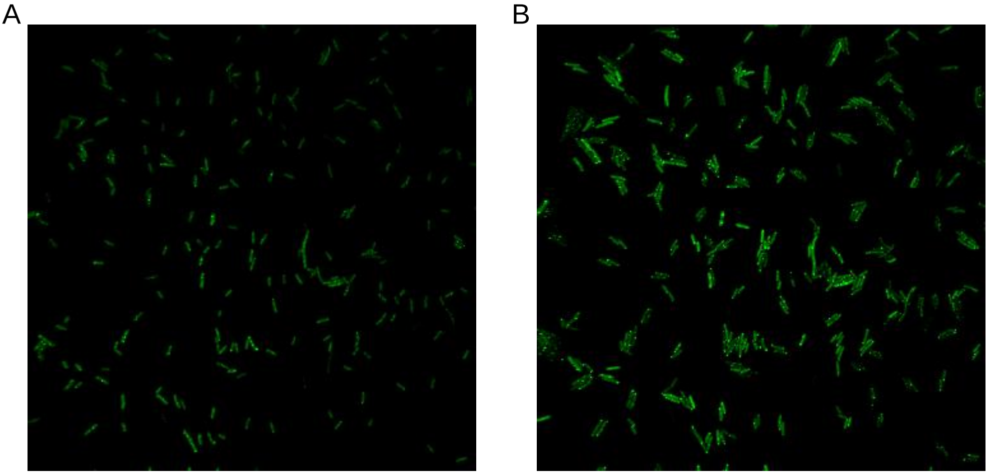


Figure S2: Example time-series images from confocal microscopy grown in LB media, at 37°C, (A) image at 5 minutes (B) image at 120 minutes.

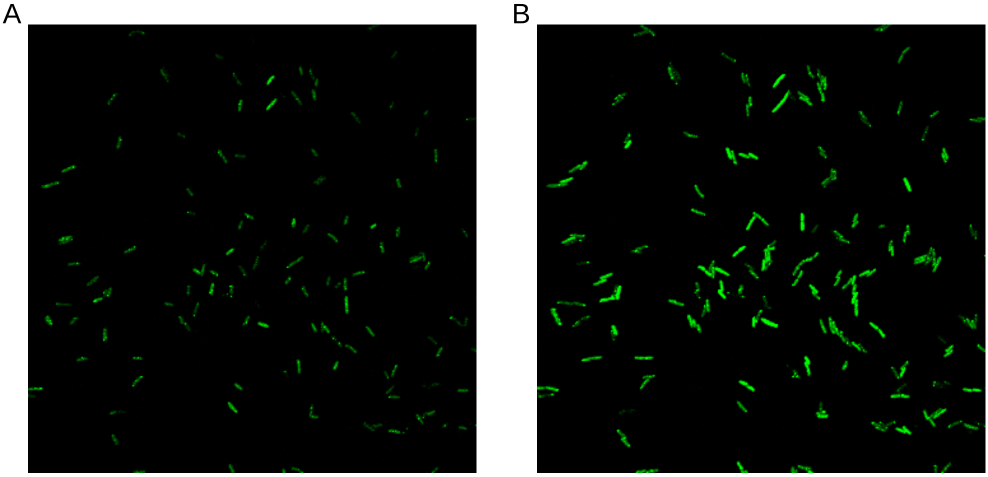


Figure S3: Example time-series images from confocal microscopy grown in LB media, at 24°C, (A) image at 5 minutes (B) image at 120 minutes.

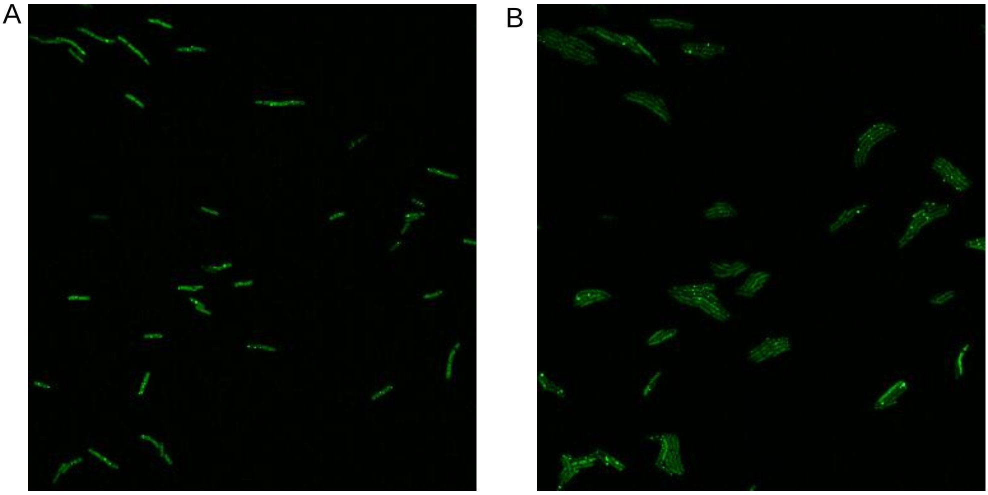


Figure S4: Example time-series images from confocal microscopy grown in LB media, at 42°C, (A) image at 5 minutes (B) image at 120 minutes.

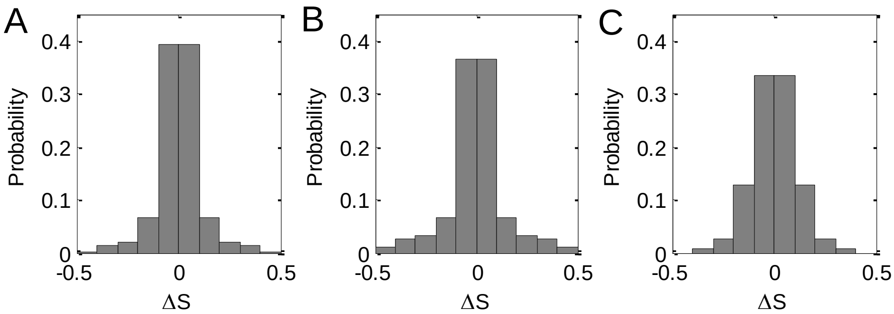


Figure S5: Distributions of normalized asymmetry in size of sister cells at birth (ΔS) in optimal growth conditions (LB media, 37°C), (A) expressing MS2-GFP and target RNA using confocal microscopy (170 divisions), (B) expressing MS2-GFP and target RNA using bright-field microscopy (97 divisions), and (C) not expressing MS2-GFP or target RNA using bright-field microscopy (130 divisions).

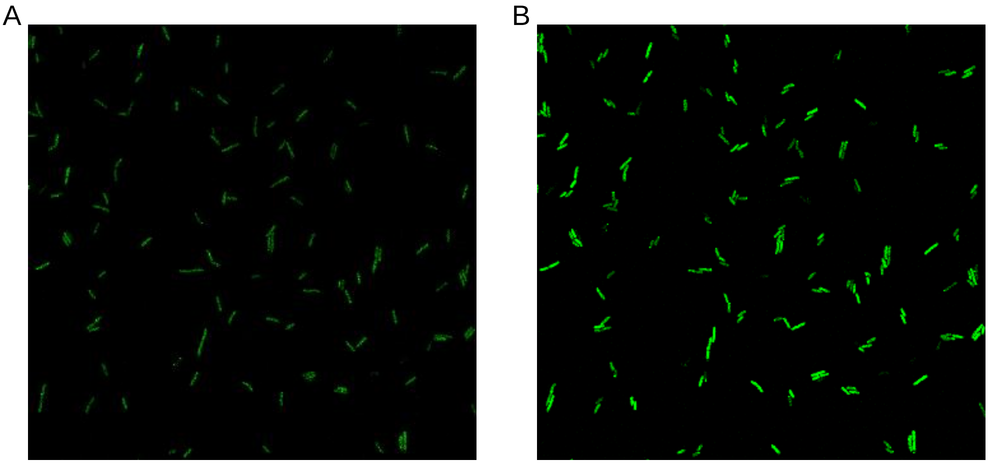


Figure S6: Example time-series images from confocal microscopy grown in LB media at 37°C and subjected to mild oxidative stress, (A) image at 5 minutes (B) image at 120 minutes.

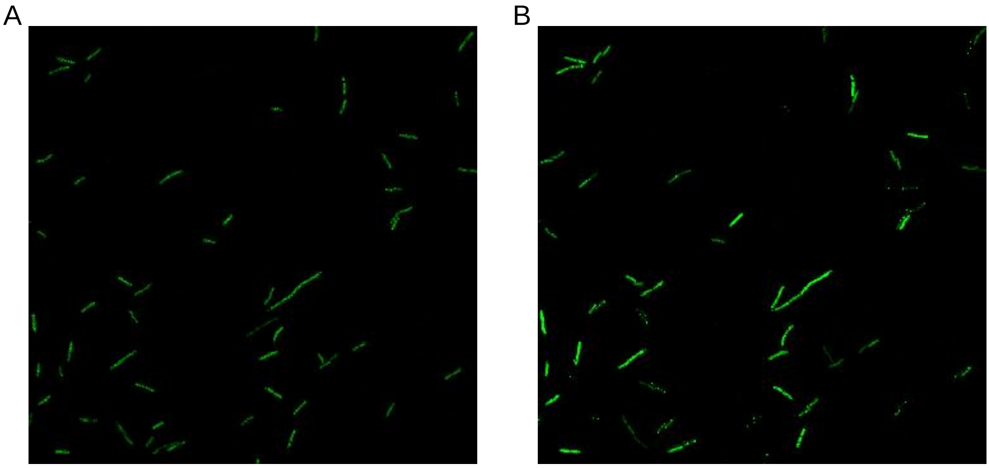


Figure S7: Example time-series images from confocal microscopy grown in LB media at 37°C and subjected to mild acidic shift stress, (A) image at 5 minutes (B) image at 120 minutes. Note the lack of cell divisions in this period of time, due to the stress conditions.

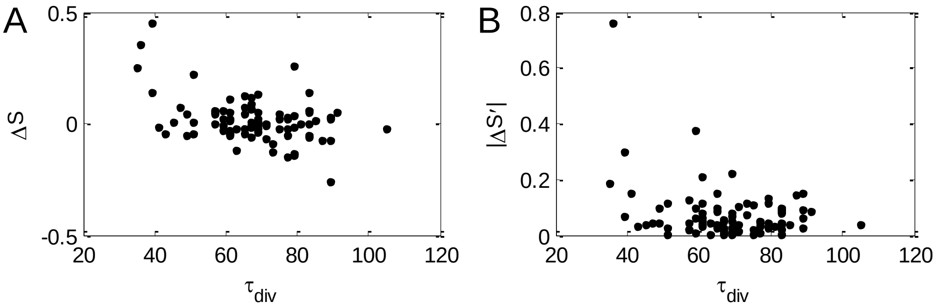


Figure S8: (A) Scatter plot between asymmetry in size at birth of a cell (ΔS) and its division time (τ_{div}) (B) Scatter plot between the asymmetry in size in division of a cell ($\Delta S'$) and τ_{div} of its mother cell. Data from 86 cells that were born and divided during measurements in LB media at 30°C.

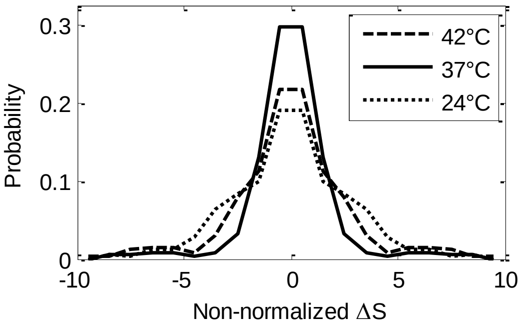


Figure S9: Distributions of non-normalized asymmetries in size of sister cells at birth, obtained from cells at 42°C, 37°C and 24°C. Data from 231 divisions at 42°C, 170 divisions at 37°C (optimal), and 280 divisions at 24°C.

Publication IV

A. Gupta, J. Lloyd-Price, and A. S. Ribeiro (2014). Modelling Polar Retention of Complexes in *Escherichia coli*. *Lecture Notes in Computer Science*, **8859**:239-43.

Modelling Polar Retention of Complexes in *Escherichia coli*

Abhishekh Gupta, Jason Lloyd-Price, and Andre S. Ribeiro

Laboratory of Biosystem Dynamics, Computational Systems Biology Research Group,
Department of Signal Processing, Tampere University of Technology,
P.O. Box 553, 33101 Tampere, Finland
`abhishekh.gupta@tut.fi`

Abstract. The cytoplasm of *Escherichia coli* is a crowded, heterogeneous environment. The spatial kinetics and heterogeneities of synthetic RNA-protein complexes have been recently studied using single-cell live imaging. A strong polar retention of these complexes due to the presence of the nucleoid has been suggested based on their history of positions and long-term spatial distribution. Here, using stochastic modelling, we examine likely sources, which can reproduce the reported long-term spatial distribution of the complexes. Based on the anisotropic displacement distribution observed at the border between the mid-cell and poles, we conclude that the original hypothesis that the observed long-term behavior is the result of macromolecular crowding holds.

Introduction. Even single-celled organisms, such as *Escherichia coli*, possess a far from random internal organization, as the cytoplasm is a crowded, heterogeneous environment. Some proteins preferentially locate at the cell poles (e.g. those involved in chemo-taxis), while others, e.g. those involved in gene expression, locate at mid-cell, within a structure known as the nucleoid.

Recent single-cell live microscopy measurements have studied the spatio-temporal distributions of a large complex, composed of a synthetic RNA tagged with multiple MS2-GFP proteins [1][2]. In one of these studies it was observed that, at short time scales, the motion of the complexes is sub-diffusive with an exponent that is robust to physiological changes and, at long time scales, the complexes tend to localize at the cell poles [1]. Further, it has been shown that these complexes are retained at the poles, as shown in Figure 1A, most likely due to the presence of the nucleoid at mid-cell [2]. This hypothesis arises from the observation of a strong anisotropy in the displacement distribution where the border of the nucleoid is expected to be (Figure 1B). However, the observed long-term spatial distribution of complexes could also, theoretically, arise from other sources, e.g. heterogeneities in the speed of the complexes along the major axis of the cell. Here, we use stochastic modelling to distinguish, from the observations, between the possible retention mechanisms taking place.

Methods. We model the cell as a compartmentalized 1-dimensional space which is divided into N homogeneous sub-volumes, indexed from [1,N]. The motion of

P. Mendes et al. (Eds.): CMSB 2014, LNBI 8859, pp. 239–243, 2014.
© Springer International Publishing Switzerland 2014

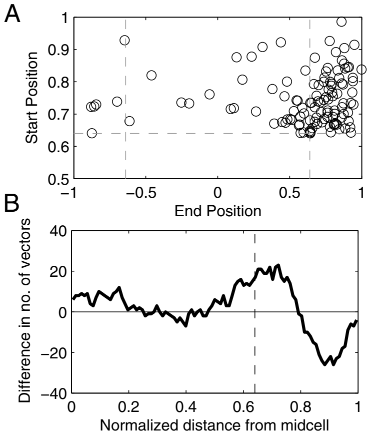


Fig. 1. (A) Relationship between the position along the major axis where each complex was last observed and the absolute position where it was first observed at a pole. Here, an end position of +1 indicates that the complex remained at the same pole as it was first observed in, while -1 indicates that it traveled to the other pole. (B) Difference between the numbers of displacement vectors that are directed towards the poles and towards the mid-cell along the major cell axis. The differences in numbers were calculated from the displacement vectors originating within windows extending 0.05 normalized cell lengths around that point. All 160 cells were born during the measurement period and contained one complex in their lifetime. In both figures, the horizontal and vertical dashed lines represent the detected separation between the mid-cell and poles.

the complexes along the major cell axis is modeled with unimolecular reactions following the Reaction-Diffusion Master Equation [3]. We define $\overrightarrow{\alpha}(x)$ as the propensity of the forward reaction (modeling the motion of a complex from position x to position $x+1$) and $\overleftarrow{\alpha}(x)$ as the propensity of the backward reaction (from x to $x-1$). These propensity functions account for the combined effects of the rod shape of the cell and the nucleoid on the motions of the complexes.

Let $P(t)$ be the $N \times 1$ vector describing the probability of observing a complex in each sub-volume at time t, and A be the $N \times N$ transition rate matrix of propensities. $P(t)$ therefore evolves according to the following master equation, in matrix-vector form:

$$\frac{dP(t)}{dt} = AP(t). \tag{1}$$

Since a complex can travel from any sub-volume in the cell to any other sub-volume, given enough time, the system is ergodic and as $t \to \infty$, $P(t)$ will converge to a unique solution, P_∞ . Solving the linear system of equations $0 = AP_\infty$, with the constraint that the total probability must sum to 1, we obtain the long-term spatial distribution of the complexes predicted by the model.

The propensities of the both diffusion reactions without accounting effects due to the rod shape and nucleoid are proportional to the diffusion constant of the complexes, D, given by:

$$\overrightarrow{\alpha}(x) = \overleftarrow{\alpha}(x) = \frac{N^2 D}{2}. \tag{2}$$

To account for the rod shape, i.e. a cylinder capped with two half-spheres, the length of the cell was parameterized by $B \in [0, 1]$, the normalized distance from midcell at which the cap begins.The forward propensities were attenuated by $\phi(x)$, the ratio between the areas of the cross sections of the cell (denoted $S(x)$) at adjacent positions. As such, $\overleftarrow{\alpha}(x)$ remains the same and $\overrightarrow{\alpha}(x)$ becomes:

$$\overrightarrow{\alpha}(x) = \frac{N^2 D}{2} \phi(x). \tag{3}$$

where,

$$\phi(x) = \frac{S(x+1)}{S(x)},$$

$$S(x) = \begin{cases} \pi & \text{if c(x) < B} \\ \pi \left[1 - \left(\frac{c(x)-B}{1-B} \right)^2 \right] & \text{if c(x)} \geq \text{B} \end{cases}$$

Here, $c(x)$ translates the index of a sub-volume into the normalized distance from the midcell to the center of the sub-volume. In this case, $B = 1$ recovers the cylindrical cell from above, and $B = 0$ produces a spherical cell.

The effects of a nucleoid are introduced in the above model by adding a Gaussian function to $\overrightarrow{\alpha}(x)$ while subtracting it from $\overleftarrow{\alpha}(x)$. This anisotropy was parameterized with center $\mu \in [0, 1]$, standard deviation σ, and height h. Specifically:

$$\overrightarrow{\alpha}(x) = \frac{N^2 D}{2} \left[\phi(x) + h * exp \left\{ \frac{-(c(x) - \mu)^2}{2\sigma^2} \right\} \right], \tag{4}$$

and

$$\overrightarrow{\alpha}(x) = \frac{N^2 D}{2} \left[1 - h * exp \left\{ \frac{-(c(x) - \mu)^2}{2\sigma^2} \right\} \right]. \tag{5}$$

To fit the models to the measurements, we use the Earth-Mover's metric [4][5]:

$$W(F, G) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx \tag{6}$$

where F and G are the cumulative distribution functions of the model and the measurements. This metric is a measure of the amount of work required to make two distributions identical.

Results. We constructed three 1-dimensional models to simulate the diffusion of the complexes within the cell. For all the models, we set N to 100, and D, the diffusion coefficient, to $1.43 * 10^{-2} \mu m^2/min$ based on previous measurements [1][2]. Two of the models contain spherical cell caps and their effects. We introduced a localized anisotropy in one of these models to test whether it, as observed in [2], can generate the observed long-term spatial distributions of the complexes (see Methods). In the last model, we set the forward and backwards propensities of diffusion events to be equal, and inversely proportional to the observed spatial distribution. Due to this, in the long term, the complexes tend to linger in the areas where they were observed with high probability.

Next, for each model, we varied all parameters and, for each set of values, obtained the distribution of complex positions that would be observed at infinite time. We then selected the set of parameters whose resulting distribution best fit the measured distribution of complex positions reported in [2].

The results from all three models, each using the best-fit parameter values, are shown in Figure 2. The model without the anisotropy fails to reproduce the displacement distribution (Figure 2A), and the consequent heterogeneity in the spatial distribution of complexes that favors their presence at the poles (Figure 2B). Meanwhile, the second model captures both of these properties of the dynamics of the complexes with significant accuracy. Interestingly, even though the third model reproduces the long-term spatial distribution exactly as observed (the lines are indistinguishable in Figure 2B), it produces a negligible anisotropy in the predicted displacement distribution (Figure 2A).

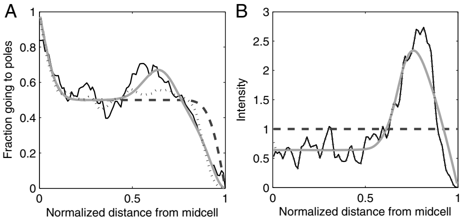


Fig. 2. (A)Measured fraction of displacement vectors originating within a window extending 0.05 normalized cell lengths around that point which are directed towards the pole (black line), model prediction with homogenous speed (without nucleoid (dashed line) and with nucleoid (gray line)), and with differing speed without nucleoid (dotted line). Note that the dashed line is superimposed by the gray line in the left side of the graph. (B) Measured spatial distribution of fluorescence intensities of complexes (black line) model prediction with homogenous speed (without nucleoid (dashed line) and with nucleoid (gray line)), and with differing speed without nucleoid (dotted line). Note that the dotted line is superimposed on the black line of the graph.

Using these models, with parameters tuned to match the measurements reported in [2], we show that both an anisotropy in the displacement vectors and a reduced speed at the poles produce good fits with the measurements. However, the model with varying speed along the major cell axis, at the time scale of the measurements, was unable to reproduce the observed anisotropic displacement distribution at the border between the mid-cell and poles. We conclude that polar retention most likely relies on these anisotropies in the displacement distribution rather than differences in speeds, consistent with the hypothesis that the observed long-term behavior is the result of macromolecular crowding, likely due to the nucleoid. Overall, the spatiotemporal kinetics of the complexes suggests that nucleoid occlusion is a source of dynamic heterogeneities of macromolecules in *E. coli* that ultimately generate phenotypic differences between sister cells.

References

1. Golding, I., Cox, E.: Physical Nature of Bacterial Cytoplasm. Phys. Rev. Lett. 96, 098102 (2006)
2. Gupta, A., Lloyd-Price, J., Neeli-Venkata, R., Oliveira, S.M.D., Ribeiro, A.S.: In Vivo Kinetics of Segregation and Polar Retention of MS2-GFP-RNA Complexes in *Escherichia coli*. Biophys. J. 106, 1928–1937 (2014)
3. Gardiner, C.W., McNeil, K.J., Walls, D.F., Matheson, I.S.: Correlations in Stochastic Theories of Chemical Reactions. J. Statiscal Phys. 14, 307 (1976)
4. Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Movers Distance as a Metric for Image Retrieval. Int. J. Comput. Vis. 40, 99–121 (2000)
5. Dobrushin, R.L.: Prescribing a system of random variables by conditional distributions. Theory Probab. Its Appl. 15, 458 (1970)

Publication V

A. Gupta, J. Lloyd-Price, and A. S. Ribeiro (2015). *In silico* analysis of division times of *Escherichia coli* populations as a function of the partitioning scheme of non-functional proteins. *In Silico Biology*, **12 (1,2)**: 9–21.

In Silico Biology 12 (2012-2014) 9–21
DOI 10.3233/ISB-140462
IOS Press

In silico analysis of division times of *Escherichia coli* populations as a function of the partitioning scheme of non-functional proteins

Abhishekh Gupta, Jason Lloyd-Price and Andre S. Ribeiro*
Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, Tampere, Finland

Received 6 August 2014
Revised 23 September 2014
Accepted 30 September 2014

Abstract. Recent evidence suggests that cells employ functionally asymmetric partitioning schemes in division to cope with aging. We explore various schemes *in silico*, with a stochastic model of *Escherichia coli* that includes gene expression, non-functional proteins generation, aggregation and polar retention, and molecule partitioning in division. The model is implemented in SGNS2, which allows stochastic, multi-delayed reactions within hierarchical, transient, interlinked compartments. After setting parameter values of non-functional proteins' generation and effects that reproduce realistic intracellular and population dynamics, we investigate how the spatial organization of non-functional proteins affects mean division times of cell populations in lineages and, thus, mean cell numbers over time. We find that division times decrease for increasingly asymmetric partitioning. Also, increasing the clustering of non-functional proteins decreases division times. Increasing the bias in polar segregation further decreases division times, particularly if the bias favors the older pole and aggregates' polar retention is robust. Finally, we show that the non-energy consuming retention of inherited non-functional proteins at the older pole via nucleoid occlusion is a source of functional asymmetries and, thus, is advantageous. Our results suggest that the mechanisms of intracellular organization of non-functional proteins, including clustering and polar retention, affect the vitality of *E. coli* populations.

Keywords: Non-functional proteins, partitioning schemes, functional asymmetries, division times, *In silico* models

1. Introduction

The mechanisms employed by *Escherichia coli* in protein production and maintenance [60] are not free from a number of errors, e.g., during folding and activation, particularly when cells are stressed, or following mutations or overexpression [40]. Consequently, there are several non-functional proteins present in a cell at any given moment. How many is unknown but these numbers should depend on the environment and differ between proteins. The excessive accumulation of non-functional proteins has been linked to decreasing

---

*Corresponding author: Andre S. Ribeiro, Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland. Tel.: +358 408490736; Fax: +358 331154989; E-mail: andre.ribeiro@tut.fi.

1386-6338/12-14/$35.00 © 2012-2014 – IOS Press and the authors. All rights reserved
This article is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License.

growth rates, not only in *E. coli* [14, 18, 36, 44] but in other organisms as well [15, 25, 50].

Not surprisingly, *E. coli* has evolved a complex machinery to enhance proteins' functionality. For example, chaperones both catalyze the proper folding of proteins, which prevents aggregation as it assists in reaching stable structures, as well as participate in the rescue of misfolded proteins [6, 60]. Under optimal growth conditions, at least 10% to 20% of newly formed polypeptides appear to be associated to chaperones, such as GroEL and DnaK [10, 16, 53].

When these rescue mechanisms fail, some misfolded proteins can still be targeted for destruction by the protease cell machinery [21, 59]. Evidence suggests that approximately 20% of newly synthesized polypeptides are degraded [20, 38]. It is reasonable to assume that this number roughly corresponds to the percentage of polypeptides that are non-functional following their production. However, this is likely to be an overestimation as a certain fraction of proteins is likely to be degraded at all times, even in the absence of non-functional proteins, to ensure the existence of raw material for the production of novel proteins [20]. Evidence suggests that as much as 10% of the proteins can be degraded per hour in non-growing bacteria [19, 37, 45, 61].

When the above strategies fail, bacteria still have additional strategies to cope with non-functional proteins [54]. For example, misfolded polypeptides of some proteins form aggregates [22, 32, 58, 62]. Relevantly, the process of formation of these aggregates exhibits some similarities to events in eukaryotic cells that have been linked to the emergence of diseases such as Huntington's, Alzheimer's, and Parkinson's [5, 11, 24, 41, 56].

Once formed, some of these aggregates can be gathered into one or two inclusion bodies, which tend to locate at the cell poles, particularly in cells are under heat or oxidative stress [22, 32]. While these inclusion bodies are stable structures, there is evidence that the process of inclusion of damaged proteins is not necessarily irreversible [6]. Interestingly, the process of aggregate formation appears to be protein-specific [51] and under regulation (e.g., in *E. coli* grpE280 mutant aggregates are small-sized [32]).

The partitioning in division of a specific non-functional polypeptide is likely related to its long-term spatial distribution. For example, if all non-functional polypeptides of a given protein are gathered into a single inclusion body, only one of the daughter cells will inherit them. On the other hand, if the non-functional polypeptides do not aggregate, one could expect them to be less retained at the poles and thus, to be more randomly distributed. As such, it would be more likely that such polypeptides are partitioned by the daughter cells in an unbiased fashion. It is also possible that the partitioning scheme will depend on the location of the functional form of the proteins, provided that the loss of function does not affect the long-term spatio-temporal distribution. The process of partitioning in division of non-functioning polypeptides is thus likely to be under regulation, either by direct or indirect mechanisms, and to differ between proteins. This is supported by recent evidence that some protein aggregates, due to polar retention and cell division, accumulate primarily at the older pole of the cells [23, 34].

Here, we explore this issue *in silico*. For that, we developed a stochastic model that includes cell division and the partitioning of intracellular molecules, along with a delayed stochastic model of gene expression that includes the generation of both functional and non-functional proteins. First, we show that, using parameter values extracted from measurements, the model is capable of emulating the reduced division times as a function of the increasing number of non-functional proteins in the cells, over several generations [52]. Next, we investigate how different partitioning schemes of non-functional proteins in division affect cell division times.

2. Results & discussion

2.1. Tuning the effects of non-functional proteins on the growth rate of model cell populations

After setting the parameter values (Table 1) of the model (see Methods), we first studied how the growth rate of the population is affected by β, the number of non-functional proteins necessary for the cell growth rate to be halved, and by n, the exponent of $f(Q)$, the function regulating the degree to which the growth rate is reduced by the presence of non-functional proteins (Equation (8)). Here, n determines how fast the cell growth rate transitions from high to low with increasing numbers of non-functional proteins in the cells. For simplicity, we assume that the environment poses no growth constraints (neither on the rate of growth nor on the total number of cells). At moment zero, each population contains one cell.

By inspection, we verified that decreasing β increases the mean division time, as expected. To exemplify this, in Fig. 1 (left) we show the number of cells over time for three values of β. Fig. 1 (right) shows the corresponding distributions of division times. In these simulations, n was arbitrarily set to 1.

Though the mother cell of each lineage is initiated without non-functional proteins, these will accumulate over time as they are produced. Meanwhile, cell division should act as a decay rate (by 'dilution') on the number of non-functional proteins in individual cells. Thus, we expect the mean division time to differ with the cell generation, until reaching quasi-equilibrium.

The dependence of the division time on the presence of non-functional proteins should differ with β. Higher β should increase the number of generations required for the effects of aging on division times to become visible (and to reach quasi-equilibrium). This can be seen in example Fig. 2, where we plot the division time of individual cells of each generation, for each value of β, up to 6 generations. Note that, for $\beta = 10$, the effects of aging are visible already in the first generation (much longer division time than for the other values of β) and are maximized approximately from the fourth generation onwards. Meanwhile, for β equal to 35 and 50, the effects of aging (i.e. the increase of the division time with the generation) appear more gradually, and are (approximately) maximized only in generations 5 and 6, respectively.

Table 1
Parameters used in the model, unless stated otherwise

Parameter	Description	Value	Source
t_{div}	Value set to the mean division time of *E. coli* DH5α-PRO	3600 s	[29]
k_{cc}	Closed complex formation	$1/574\ \mathrm{s}^{-1}$	[29]
k_{oc}	Open complex formation	$1/391\ \mathrm{s}^{-1}$	[29]
k_{tr}	Translation rate	$1/3\ \mathrm{s}^{-1}$	[43]
τ_p	Mean of time delay for protein production	420 s	[43]
σ_p	Std. dev. of time delay for protein production	300 s	[43]
k_{drbs}	Degradation rate of RNA	$0.011\ \mathrm{s}^{-1}$	[43]
k_{dp}	Degradation rate of proteins	$3.9 \times 10^{-5}\ \mathrm{s}^{-1}$	[43]
k_D	Rate at which functional proteins become non-functional	$1.9 \times 10^{-5}\ \mathrm{s}^{-1}$	[9]

From here onwards, we arbitrarily set β to 35 as this value (and similar ones) allows the accumulation of non-functional proteins to affect cell division times in a manner that is consistent with observations (see e.g. [34]), whereas much lower values disrupt too strongly cellular well-functioning and much higher values do not affect cell growth significantly.

Next, we study the effects of n on the cell growth rate. Different non-functional polypeptides are expected to have different effects on the cell growth rate, to differ in their degree of aggregation, etc. Also, previous observations [34, 52, 62] suggest that, starting from a single healthy cell, only after a few generations do the effects of the accumulation of non-functional proteins on division times become visible. This can be captured by the model by tuning n. From Fig. 3, as n increases, cell division rate increases. This is due to the fact that, for high values of β, as n increases, more non-functional proteins need to accumulate in a cell to tangibly affect its division time. Nevertheless, it is worth noting that no qualitative changes are observed as n is increased. As such, for simplicity, from here onwards we set n to 1.

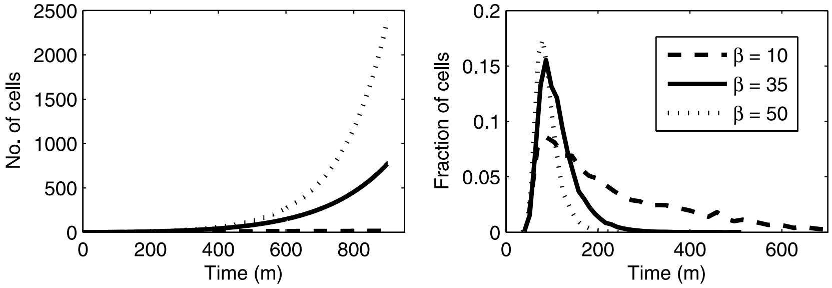


Fig. 1. (*Left*) Mean number of cells over time for different values of β. As β is lowered, the number of cells drastically decreases. Also shown (*Right*) are the corresponding distributions of division times.

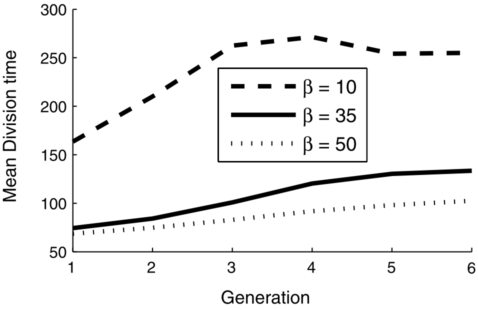


Fig. 2. Mean division time of cells as a function of their generation and of β, the number of non-functional proteins necessary for the cell growth rate to be reduced to half the maximum rate.

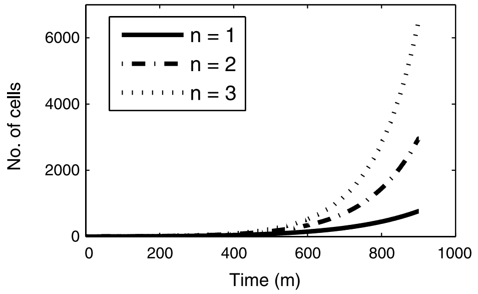


Fig. 3. Effect of changing parameter n that models the effect of non-functional proteins on the growth rate.

Using these values of β and n, the model is able to emulate (see example Fig. 4) empirical distributions of division times in near-optimal growth conditions (see e.g. [52]). Next, we use the model to study the effects of different partitioning schemes of non-functional proteins on the population dynamics.

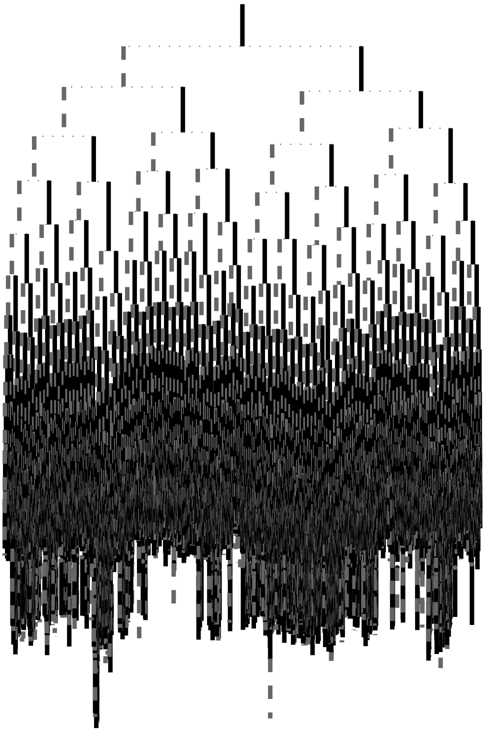


Fig. 4. Example lineage showing individual division times across lineages for $\beta = 35$ and $n = 1$. The length of the line that connects a cell to its progeny is proportional to the average growth rate of that cell (longer lines represent higher growth rates). At each division, the cell inheriting the old pole is placed on the right side of the division pair, while new poles are placed on the left side of each pair. The mean doubling time of the cells equals 135 min.

2.2. *Effects of different partitioning schemes*

We investigate the effects of different partitioning schemes of non-functional proteins in division. In particular, we consider six schemes [26] (schematic representations in Fig. 5, Top): (i) perfect partitioning, where individual non-functional proteins are equally partitioned between the daughter cells, (ii) pair formation, where non-functional proteins form pairs, which are split evenly into the two daughter cells, while the remaining individuals are independently partitioned, (iii) random size partitioning, where the non-functional proteins are assumed to be uniformly distributed in the mother cell, and the daughter cells differ significantly in size (iv) preferential partitioning, where individual non-functional proteins are partitioned according to a biased binomial distribution, (v) cluster formation, where non-functional proteins form clusters of variable size which are then independently partitioned, and (vi) all or nothing partitioning, where all non-functional proteins are inherited by only one of the daughter cells.

Using the parameter values in Table 1, for each partitioning scheme, we simulated 1 cell lineage for $6 \times t_{div\ s}$ (i.e. 6 times the expected division time). This allows, on average, 6 cell generations to be observed. Aside from these parameters, we have

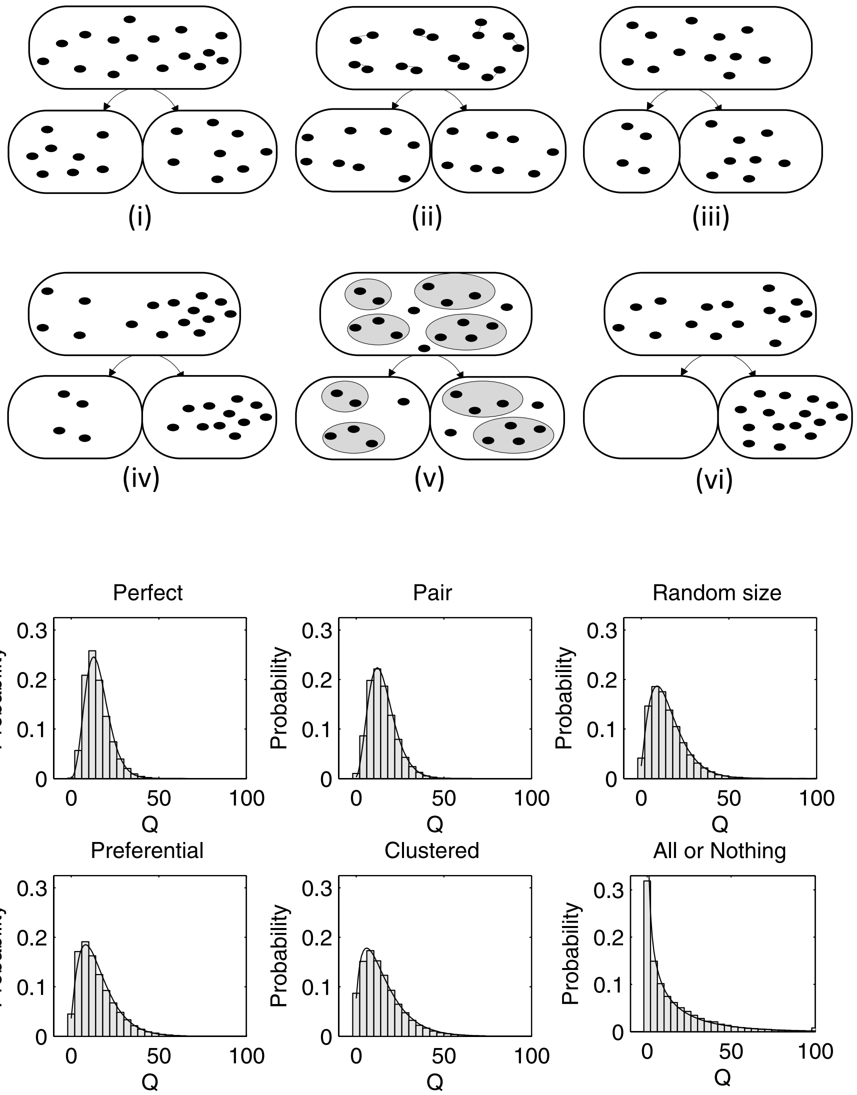


Fig. 5. (*Top*) Different partitioning schemes. (i) Perfect (ii) Pair formation (iii) Random size (iv) Preferential (v) Clustered (vi) All or Nothing. Also shown (*Bottom*) are the non-functional protein distributions resulting from each of the partitioning schemes.

used the following additional set of parameter values (necessary to implement some of the partitioning schemes). For pair formation, we set the fraction of non-functional proteins which form pairs to 0.25. In random size partitioning, the position of the division point of the daughter cells was drawn from a Beta distribution with $\alpha = \beta = 2$. In preferential partitioning, the bias was set to 0.25 (i.e. the cell inheriting the older pole retains, on average, 75% of the non-functional proteins produced by its mother cell). Finally, in the cluster formation scheme, we set the degree of clustering, k_{cl}, to 0.03.

As stated in the methods section, inherited non-functional proteins are not subject to these schemes, only *novel* ones are. This assumes perfect retention of inherited non-functional proteins at the older pole of the cell, which, while not being exact, is a close approximation to observations [23].

For each partitioning scheme, we obtained the number of non-functional proteins in each cell at the end of the simulation time. Distributions are shown in Fig. 5 (Bottom). Table 2 shows the means and standard deviations of these distributions. As expected, the standard deviation increases from a perfectly symmetric to a perfectly asymmetric partitioning scheme. Meanwhile, as protein production rates (functional and non-functional) are identical in all models, mean numbers of non-functional protein are necessarily also identical.

To investigate if the differences in the standard deviation generate different mean cell division times in cells with different partitioning schemes, we performed 1000 simulations of the model for each partitioning scheme. Each simulation was performed from one initial cell, for 15 times the expected division time (i.e., $15 \times t_{div}$ s). For each scheme, we obtained the mean increase in the number of cells in the last 30 minutes of the simulation, to eliminate the effects of the transient observed in Fig. 2. Also, we obtained the division times of all cells that divided during the simulation time, from which we extracted mean and standard deviation. Results are shown in Table 3.

From Table 3, increasing the degree of asymmetry in the partitioning of novel non-functional proteins in division decreases the mean division time, while not significantly affecting the variance (except in the all-or-nothing case, where a significant decrease is observed).

Table 2
Mean and standard deviation of the distributions of the number of non-functional proteins in individual cells (one lineage, after 360 min, or approximately 6 generations), for different partitioning schemes

Partitioning Scheme	Mean no. of non-functional proteins	Standard deviation of the no. of non-functional proteins
Perfect	15.4	7.1
Pair	15.4	7.8
Random Size	15.4	10.4
Preferential	15.3	10.9
Clustered	15.4	12.1
All or Nothing	15.4	20.4

2.3. Effects of the degree of clustering and of preferential partitioning

We next investigate two partitioning schemes in more detail, Clustered and Preferential, since these are employed by *E. coli*. For example, proteins such as Tsr form clusters at the poles [64]. Further, in some, the proteins preferentially locate at the older cell pole (e.g. IbpA [62]).

First, to study how the clustering rate of non-functional proteins affects division times, we ran simulations setting the clustering parameter, k_{cl}, to values between 0.01 and 0.09. Note that higher k_{cl} produces a weaker ‘degree of clustering’. E.g., for $k_{cl} = 0.01$, there are very few clusters, with each necessarily including many non-functional proteins, while for $k_{cl} = 0.09$ the number of clusters is higher, each with fewer non-functional proteins.

From 1000 independent lineages simulated for $15 \times t_{div}$ s (each starting from a single cell), we extracted the mean increase in the number of cells in the last 30 minutes of the simulations and the cells’ division times. Results are shown in Table 4. Visibly, both the mean and the standard deviation of the division time

Table 3
Number of cells at the end of the simulation time and mean and standard deviation of the division time of individual cells in the various partitioning schemes

Partitioning Scheme	Mean increase in no. of cells at the last 30 mins of simulation time	Mean division time (m)	Standard deviation of division time (m)
Perfect	104.7	114.2	42.6
Pair	107.0	113.7	42.6
Random Size	147.7	105.7	41.9
Preferential	163.5	103.8	41.1
Clustered	189.9	100.6	41.5
All or Nothing	402.9	85.8	36.6

Table 4
Number of cells at the end of the simulation time and mean and standard deviation of the division time of individual cells in lineages subject to the clustering partitioning scheme, for different degrees of clustering (k_{cl})

k_{cl}	Mean increase in no. of cells at the last 30 mins of simulation time	Mean division time (m)	Std of division time (m)
0.01	337.3	89.2	38.7
0.03	189.9	100.6	41.5
0.05	155.6	105.1	42.0
0.07	140.1	107.4	42.1
0.09	131.7	108.8	42.2

increase with increasing k_{cl}, (i.e. decreasing degree of clustering). These results confirm that increasing the degree of asymmetry in partitioning of non-functional proteins (in this case via clustering) decreases mean division times. Interestingly, these results agree with reports that the fusion of protein aggregates supports population survival in yeast [7].

Next, we tested the effects of preferential polar segregation and partitioning in division of non-functional proteins produced by the cell (again, inherited proteins are not affected, remaining at the old pole). For this, we introduced an additional parameter in the model, so as to regulate the degree of bias in partitioning, named '*Bias*', and ran the model for values of *Bias* from 0 to 1. For *Bias* equal to 0.5, *new* non-functional proteins will be partitioned unbiasedly and independently between daughter cells. Also, according to this scheme, e.g., for a bias of 0.1, each non-functional protein has a 90% chance to be inherited by the daughter cell inheriting the older pole of the mother cell. Finally, note that since inherited proteins will remain at the old pole throughout a cell's life time, divisions will unavoidably introduce asymmetries. Because of this, a *Bias* of ∼0.7 (and not 0.5) is expected to result in an unbiased partitioning in division of *all* non-functional proteins of the mother cell.

We obtained the mean increase in the number of cells in the last 30 minutes of the simulations, as well as the mean and standard deviation of the division time of the cells from 1000 independent lineages, for $15 \times t_{div}$ s. The results are shown in Table 5. First, the mean number of cells at the end of the simulation is smallest for a *Bias* of 0.7, since this leads to more symmetry in partitioning in division than setting *Bias* to 0.5.

Table 5
Number of cells at the end of the simulation time and mean and standard deviation of the division time of individual cells in lineages subject to the preferential partitioning scheme, for different degrees of *Bias*

Bias	Mean increase in no. of cells at the last 30 mins of simulation time	Mean Division time (m)	Standard deviation of division time (m)
0.00	888.2	75.0	31.4
0.10	572.3	80.7	33.9
0.30	238.8	95.4	39.3
0.50	109.1	113.8	42.3
0.70	95.9	118.0	41.6
0.90	217.3	99.2	38.6
1.00	384.9	88.0	35.8

Another interesting observation is the difference in mean division time between biases towards older and newer cell poles. In particular, one observes that cells that bias the partitioning of novel non-functional proteins towards the older pole, where the inherited non-functional proteins are located, have faster division times. This is because this asymmetry in pole choice maximizes the asymmetry in the numbers of non-functional proteins (inherited plus produced) at the poles.

To validate this further, we loosened the condition of polar retention of inherited non-functional proteins. For this, we replace the constant 'Retention' by a parameter, named '*R*', which can range from 0 to 1. For $R = 1$, *all* inherited non-functional proteins remain retained at the old pole. Decreasing *R* increases the number of inherited proteins that escape the old pole during a cell's lifetime. If escaping, they 'merge' with the population of newly formed non-functional proteins at midcell, and then partitioned in division accordingly. For $R = 0$, all inherited non-functional proteins are partitioned following the same scheme as produced ones, which controlled by the parameter *Bias*.

We extracted, as a measure of cell growth rate, the mean increase in number of cells during the last 30 min of each simulation for different values of *Bias* and *R*. The results, averaged from 1000 simulations per condition (each $15 \times t_{div}$ s long), are shown in Fig. 6. The mean increase in number of cells is maximized for a *Bias* of 0 (which favors the old pole), regardless of the value of *R*.

From the above, we conclude that the most advantageous solution for cells is to place all non-functional proteins, produced and inherited, into the older pole. Note that, combining a *Bias* of 1 (all new non-functional proteins move to the new pole) with an *R* of 0 (no inherited protein is retained at the poles), one attains similar growth rates as for a *Bias* of 0 (since all proteins, inherited and produced, will be placed at the new pole, creating a full bias). However, we do not expect this latter scheme to be found in nature (at least, not commonly) since it would likely require an energy-consuming mechanism to be implemented to ensure that inherited proteins moved to the newer pole, while the other option is less energy consuming.

Meanwhile, visible in the inset of Fig. 6, in the absence of any bias in pole choice by newly produced proteins, retention of the inherited proteins at the old pole can, on its own, increase the mean rate of cell

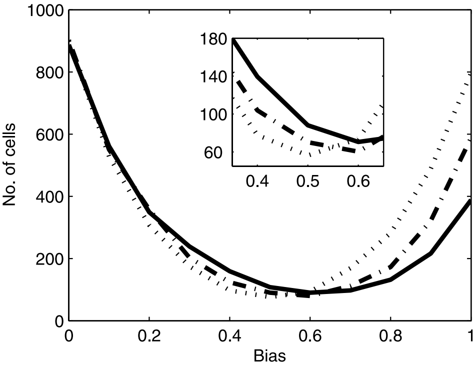


Fig. 6. Mean number of cells produced during the last 30 minutes for different biases in preferential partitioning with varying levels of retention (R). Each simulation started with a single cell and the results for each parameter set are averaged over 1000 runs. Dotted line corresponds to a model with R set to 0.0, while dashed line is for R set to 0.5, and solid line is for R set to 1.0. Inset shows the results in detail for Bias between 0.35 and 0.65.

division of the population, leading to higher cell numbers. In particular, in the present case, we observe an increase of 10.2% in the number of cells that appear in the last 30 min when increasing R from 0 to 0.5, and an increase of 34.1% when increasing R from 0 to 1.

Also visible in the inset are the points at which the various lines cross. From these, one can observe that the *Bias* at which the crossing occurs depends on the values of R. Visibly, the higher the difference in values of R of the two models, the higher is the value of Bias at which these models exhibit equal cell division rates.

3. Conclusion

We explored, *in silico*, using a stochastic model and parameter values extracted from measurements (Table 1), possible selective advantages of different partitioning schemes of non-functional proteins in *E. coli*. Assuming that the accumulation of non-functional proteins is harmful by decreasing cell division rate, we observed that increasing the degree of asymmetry in their partitioning in division increases the overall division rates of cell lineages.

Relevantly, a source for such asymmetries in *E. coli* has been identified. Namely, the nucleoid at midcell is denser than the cytoplasm [55]. Evidence suggests that its presence appears to force protein aggregates to travel to and then remain at the poles, by generating anisotropies in the borders between nucleoid and cytoplasm that favor the motion towards the poles [23]. Cell divisions subsequently introduce asymmetries between the number of such non-functional proteins at the old and new poles of the daughter cells, favoring the older pole. The latter was observed to occur to aggregates of non-functional proteins as well [8, 23, 34, 62].

These observations also inform on the means used by cells to differentiate between functional and non-functional proteins. Namely, as the segregation is based on occlusion, one natural means of differentiation is to aggregate the unwanted, non-functioning proteins (and only these), into large clusters, thus ensuring segregation from midcell. Such aggregation has been observed in live cells, and is likely enhanced by chaperones such as IbpA [34]. Here, we further observed that increasing the efficiency of the process of clustering enhances the generation of asymmetries between the numbers of non-functional proteins at the two poles of a cell. As such, we expect the cellular mechanisms of detection and aggregation of non-functional proteins to be under selective pressure for efficiency.

To further verify that the generation of asymmetries in non-functional protein numbers across the cell populations enhances the mean division rate within cell lineages, we tested combining retention of inherited proteins at the old pole with a mechanism able to force newly formed non-functional proteins to move into the old pole as well, prior to cell division. In agreement with the hypothesis, division rates further increased. However, there is no known mechanism of selective transport of unwanted protein aggregates to a *specific* cell pole (old or new) in *E. coli*. Also, there is no evidence for asymmetries in the choice of pole (see e.g. [23]). It may be that the division rate of *E. coli* renders such selective, energy-consuming, transport mechanism not selectively advantageous.

Finally, we tested the effects of removing the polar retention mechanism. Overall, the test showed that this mechanism alone is selectively advantageous, as expected, given that it is a non-energy consuming source of functional asymmetries. Interestingly, its effects could be combined with partitioning schemes capable of enhancing asymmetries in the number of non-functional proteins at the old and new pole of the cells to further enhance the mean rate of cell divisions.

It is worth noting that our model assumes that the accumulation of damage in cells (in this case, non-functional proteins) leads to a gradual increase in division times, as suggested by previous studies [33, 34, 36, 44, 52]. However, one recent study suggests that the accumulation of damage leads only to death, rather than affecting growth rates gradually [57]. In this study, death occurs when a lethal "factor" in a cell crosses a threshold. This factor was hypothesized to correspond to non-functional protein aggregates [57]. We expect that, using such a model, rather than a change in mean division times and number of cells at the end of the simulation time as a function of, e.g., the partitioning scheme, one would observe only the changes in the number of cells. These changes should be qualitatively similar to those reported here for high *n*, in the number of cells at the end of the simulation time as a function of the partitioning scheme, the bias in partitioning and, the degree of clustering.

There are several means by which non-functional proteins can interfere with the normal cellular functionality. For example, reduced ribosomal fidelity due to damage accumulation in the protein assembly machinery [3] would not produce functional proteins but would still waste resources, reducing growth rates [49]. As such, it is not surprising that cells evolved mechanisms to handle non-functional proteins, such as polar retention [62]. This retention, among other things, prevents these proteins from reaching binding sites in the DNA, which, in the case of non-functional transcription factors, could lead to harmful interference with normal gene expression dynamics [9].

While *E. coli* has a morphologically symmetric division process, and thus we expect our results to be valid for other such organisms [1, 2, 4, 15], we also expect the results to be applicable to species with morphologically asymmetric division processes, such as budding yeast [27, 28, 30]. In the case of morphologically asymmetric divisions, if the daughter cell is smaller (e.g. in *Saccharomyces cerevisiae*), non-functional proteins and other harmful substances should be easily retained in the mother cell. However, if the daughter cell was to be larger, transport mechanisms may be required. We hypothesize that this is one of many factors that has led to evolutionary processes resulting in the production of daughter cells that are smaller than the mother cells.

In the future, it would be of interest to investigate sources of asymmetries in eukaryotic cells as their more complex internal structure and mechanisms may allow more ingenious means to cope with the effects of aging, among other.

4. Methods

4.1. Simulator

Simulations were performed by SGNS2 [35], a simulator of chemical reaction systems whose dynamics are driven according to the delayed Stochastic Simulation Algorithm [48]. While based on the Stochastic Simulation Algorithm (SSA) [17], the delayed SSA differs in that it allows for multi-delayed reaction events. The main advantage of SGNS2, which was developed from SGN Sim [46], is that it allows multi-delayed reactions to be implemented within hierarchical, interlinked compartments that can be created, destroyed and divided during the simulation. In cell divisions, molecules are randomly segregated into the daughter cells following a specified distribution corresponding to one of several partitioning schemes, applicable on a per-molecule-type basis. As such, SGNS2 allows easy implementation and simulation of different models of partitioning of proteins (functional and/or non-functional) in division.

4.2. Modeling gene expression, cell growth and division, and long-term spatial distributions of non-functional proteins

Our model of a 'cell' has four main components: (i) a gene expression mechanism; (ii) a mechanism of generating non-functional proteins; (iii) a mechanism of cell growth and division, including partitioning schemes of molecules and effects of non-functional proteins on the growth rate; and, (iv) a mechanism of polar retention of non-functional proteins.

The first component, the model gene expression mechanism, was proposed in [47] and validated in [65], using measurements of protein production in *E. coli* cells with single molecule sensitivity [63]. More recently, this model, in particular, reactions (1) and (2), were shown to also capture with accuracy the *in vivo* kinetics of RNA production, when measured with single-molecule sensitivity by MS2-GFP tagging [29, 42]. This component includes reactions (1)–(5):

$$Pro_C \xrightarrow{k_{cc}} Pro{\cdot}RNAP_C \qquad (1)$$

$$Pro{\cdot}RNAP_C \xrightarrow{k_{oc}} Pro_C + RBS_C \qquad (2)$$

$$RBS_C \xrightarrow{k_{tr}} RBS_C + P_C(normal : \tau_p,\ \sigma_p) \qquad (3)$$

$$RBS_C \xrightarrow{k_{drbs}} \emptyset \tag{4}$$

$$P_C \xrightarrow{k_{dp}} \emptyset \tag{5}$$

Here, C denotes the index of the cell in which the reaction is taking place. Reaction (1) models the finding by an RNA polymerase (RNAP) of the transcription start site ('*Pro*', located in the promoter region) and the formation of the closed complex ('*Pro·RNAP*'). Reaction (2) models the formation of the open complex, from the closed complex, along with the promoter escape by the RNAP and the formation of the ribosome binding site region of the RNA ('*RBS*'). The time lengths of the closed complex formation and of the open complex formation are accounted for by the values set for k_{cc} and k_{oc}, respectively [29]. Also, for simplicity, the RNAP is not explicitly modeled in the left side of reaction (1), since it is assumed that the concentration of free RNAP in the cell does not change significantly during a cell's lifetime. As such, its release is not represented in the right side of reaction (2).

Reaction (3) models the translation of a protein P along with its folding and activation [65]. These latter events are accounted for by the delay τ_P in the release of the protein. Note that the ribosome is not explicitly modeled, for the same reason that the RNAP is not explicitly modeled. Note also that this reaction can initiate as soon as a *RBS* is formed [39]. As such, and since our model does not include either transcription or translation elongation at the nucleotide level, it is not necessary to represent complete RNA molecules explicitly. Finally, reactions (4) and (5) model the degradation of RNA (by degradation of its *RBS* region) and proteins, respectively, as first-order events [65].

The second component of the model is responsible for generating non-functional proteins, '*Q*', from functional proteins:

$$P_C \xrightarrow{k_D} Q_C \tag{6}$$

According to this model, the fraction of non-functional proteins in a cell is controlled by the rate k_D [9]. As noted in the introduction, the percentage of non-functional proteins in a cell at any given time is unknown. Since approximately 20% of formed polypeptides are non-functional and given the existence of several error-correction mechanisms, we chose to set k_D, the rate at which functional proteins become non-functional, to 1.9×10^{-5} s^{-1} (see Table 1), since we observed by inspection that, with this value, usually 1% to 5% of the proteins will become non-functional during the lifetime of a cell. Finally, note that these non-functional proteins, unlike functional ones, do not degrade (as we assume that they 'survived' the error correction and subsequent degradation mechanisms).

The third component of the model consists of cell growth, division and partitioning schemes of proteins and RNA molecules in division. As such, this component needs to account for the effects of non-functional proteins on a cell's division rate. Division and partitioning of cellular components are not represented in the form of chemical reactions (see below).

In optimal conditions, the moment of division of an *E. coli* cell appears to be strongly correlated with reaching a specific cell length [12, 13], and there is a very small variance in cells' lifetime. The division process is therefore considered to be largely deterministic [31]. As such, we model division as an instantaneous event that occurs once the cell reaches a specific length. When a division occurs, the DNA (i.e. the promoter region, Pro_C) is replicated and one copy is placed in each daughter cell. If the RNAP has formed the closed complex (i.e. $Pro{\cdot}RNAP_C$) at the point of division it is inherited by the cell with the older pole while the cell inheriting the newer pole inherits a free promoter. This arbitrary choice does not affect our conclusions as it only affects (very mildly) the protein numbers in each cell.

Meanwhile, *RBS* and *P* molecules are partitioned between the daughter cells according to an unbiased, independent partitioning scheme (unless stated otherwise), resulting in a binomial distribution of molecules inherited by a given daughter cell.

We use reaction (7) to simulate cell growth and the effect of non-functional proteins on this growth rate. This reaction controls the quantity 'cell length' (denoted l_C, where C is the cell ID).

$$\emptyset \xrightarrow{l_C \times \left(\frac{\ln(2)}{t_{div}}\right) \times f(Q)} l_C \tag{7}$$

In reaction (7), the rate constant $l_C \times \left(\frac{\ln(2)}{t_{div}}\right)$ (where t_{div} is the division time in the absence of non-functional proteins) is multiplied by a hill function, *f(Q)*. This function ranges from 0 to 1 and regulates the degree to which the optimal growth rate is reduced in the presence of non-functional proteins (Equation (8)):

$$f(Q) = \frac{\beta^n}{Q^n + \beta^n} \tag{8}$$

In this equation, β is the number of non-functional proteins required for $f(Q)$ to equal 0.5 and n is the exponent of the hill function and will determine the how quickly $f(Q)$ transitions from high to low as the number of non-functional proteins increases. Here, we assume that the length, l_C, initially equals 100 (an arbitrary value). Division, modeled as an instantaneous process, is triggered automatically when l_C reaches the value 200. Note that this reaction is such that, in the absence of non-functional proteins, the cell length grows exponentially in time [49].

We set the division time of the cells (t_{div}) to match the empirical mean division times of DH5α-PRO cells in LB media (∼60 minutes) [29] when $f(Q)$ equals 1 (i.e. no non-functional proteins present). These times are longer than for other strains (e.g. K12). They were chosen since most of the other rate constants used were obtained from measurements in DH5α-PRO. Also, this rate does not qualitatively affect the results. Namely, only the time-scale of events is affected.

Finally, to implement the 'cluster formation' partitioning scheme (see 'Effects of different partitioning schemes' in the Results section), the model requires one additional reaction. Clusters form during a cell's lifetime according to the following reaction, where the index C is the cell ID:

$$\emptyset \xrightarrow{l_C \times \left(\frac{\ln(2)}{t_{div}}\right) \times k_{cl}} Cluster_C \tag{9}$$

At division, clusters are partitioned following an unbiased binomial distribution. The non-functional proteins are partitioned by assigning each non-functional protein to one of the partitioned clusters [26]. All cells begin with *Cluster* set to 1, which then increases depending on the rate of k_{cl} . Larger values for k_{cl} will produce a larger number of clusters in the cell, and thus a weaker 'degree of clustering'.

The fourth and final component of the model is a retention mechanism of non-functional proteins at the cell poles. Based on previous observations [23, 34], we assume that non-functional proteins inherited by a cell will remain retained at the old pole of that cell during its lifetime, and will be partitioned accordingly in division. Meanwhile, new non-functional proteins produced during a cell's lifetime will be partitioned between the two daughter cells following the same mechanism used to partition the functional proteins, namely, the unbiased, independent partitioning scheme.

All parameter values, aside from β, along with the references from which they were extracted, are shown in Table 1.

Author contributions

A.S.R. and A.G. designed research; all authors performed research; A.G. and J.L.P. implemented the models. A.G. analyzed data; A.S.R. and A.G wrote the paper.

Acknowledgments

Work supported by Academy of Finland (grant no. 257603) (ASR), Portuguese Foundation for Science and Technology (grant no. PTDC/BBB-MET/1084/2012) (ASR), Finnish Cultural Foundation (AG) and TUT President's graduate programme (JLP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] M. Ackermann, A. Schauerte, S.C. Stearns and U. Jenal, Experimental evolution of aging in a bacterium, *BMC Evol Biol* **7** (2007), 126.

[2] M. Ackermann, S.C. Stearns and U. Jenal, Senescence in a bacterium with asymmetric division, *Science* **300** (2003), 1920.

[3] M. Ballesteros, A. Fredriksson, J. Henriksson and T. Nyström, Bacterial senescence: Protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes, *EMBO J* **20** (2001), 5280–5289.

[4] M.G. Barker and R.M. Walmsley, Replicative ageing in the fission yeast *schizosaccharomyces pombe, Yeast* **15** (1999), 1511–1518.

[5] N.G. Bednarska, J. Schymkowitz, F. Rousseau and J. Van Eldere, Protein aggregation in bacteria: The thin boundary between functionality and toxicity, *Microbiology* **159** (2013), 1795–1806.

[6] M.M. Carrió, J.L. Corchero and A. Villaverde, Proteolytic digestion of bacterial inclusion body proteins during dynamic transition between soluble and insoluble forms, *Biochim Biophys Acta* **1434** (1999), 170–176.

[7] M. Coelho, S.J. Lade, S. Alberti, T. Gross and I.M. Tolić, Fusion of protein aggregates facilitates asymmetric damage segregation, *PLoS Biol* **12** (2014), e1001886.

[8] A.-S. Coquel, J.-P. Jacob, M. Primet, A. Demarez, M. Dimiccoli, T. Julou, et al., Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect, *PLoS Comput Biol* **9** (2013), e1003038.

[9] X. Dai, S. Healy, O. Yli-Harja and A.S. Ribeiro, Tuning cell differentiation patterns and single cell dynamics by regulating proteins' functionalities in a toggle switch, *J Theor Biol* **261** (2009), 441–448.

[10] E. Deuerling, A. Schulze-Specking, T. Tomoyasu, A. Mogk and B. Bukau, Trigger factor and DnaK cooperate in folding of newly synthesized proteins, *Nature* **400** (1999), 693–696.

[11] C.M. Dobson, The structural basis of protein folding and its links with human disease, *Philos Trans R Soc Lond B Biol Sci* **356** (2001), 133–145.

[12] W.D. Donachie and K.J. Begg, "Division potential" in *Escherichia coli, J Bacteriol* **178** (1996), 5971–5976.

[13] W.D. Donachie, K.J. Begg and M. Vicente, Cell length, cell growth and cell division, *Nature* **264** (1976), 328–333.

[14] H. Dong, L. Nilsson, C.G. Kurland, H. Dong, L. Nilsson and C.G. Kurland, Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction, *J Bacteriol* **177** (1995), 1497–1504.

[15] N. Erjavec, M. Cvijovic, E. Klipp and T. Nyström, Selective benefits of damage partitioning in unicellular systems and its effects on aging, *Proc Natl Acad Sci U S A* **105** (2008), 18764–18769.

[16] K.L. Ewalt, J.P. Hendrick, W.A. Houry and F.U. Hartl, *In vivo* observation of polypeptide flux through the bacterial chaperonin system, *Cell* **90** (1997), 491–500.

[17] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, *J Phys Chem* **81** (1977), 2340–2361.

[18] B.R. Glick, Metabolic load and heterologous gene expression, *Biotechnol Adv* **13** (1995), 247–261.

[19] A.L. Goldberg, A role of aminoacyl-tRNA in the regulation of protein breakdown in *Escherichia coli*, *Proc Natl Acad Sci U S A* **68** (1971), 362–366.

[20] A.L. Goldberg, Degradation of Abnormal Proteins in *Escherichia coli, Proc Natl Acad Sci U S A* **69** (1972), 422–426.

[21] S. Gottesman and M.R. Maurizi, Regulation by proteolysis: Energy-dependent proteases and their targets, *Microbiol Rev* **56** (1992), 592–621.

[22] A.I. Gragerov, E.S. Martin, M.A. Krupenko, M.V. Kashlev and V.G. Niikrforov, Protein aggregation and inclusion body formation in *Escherichia coli* rpaH mutant defective in heat shock protein induction and inclusion body formation, *FEBS Lett* **291** (1991), 222–224.

[23] A. Gupta, J. Lloyd-Price, R. Neeli-Venkata, S.M.D. Oliveira and A.S. Ribeiro, *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*, *Biophys J* **106** (2014), 1928–1937.

[24] J.J. Harding, Viewing molecular mechanisms of ageing through a lens, *Ageing Res Rev* **1** (2002), 465–479.

[25] A.-L. Hsu, C.T. Murphy and C. Kenyon, Regulation of aging and age-related disease by DAF-16 and heat-shock factor, *Science (80)* **300** (2003), 1142–1145.

[26] D. Huh and J. Paulsson, Random partitioning of molecules at cell division, *Proc Natl Acad Sci U S A* **108** (2011), 15004–15009.

[27] S.M. Jazwinski, Growing old: Metabolic control and yeast aging, *Annu Rev Microbiol* **56** (2002), 769–792.

[28] M. Kaeberlein, Lessons on longevity from budding yeast, *Nature* **464** (2010), 513–519.

[29] M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-price, O. Yli-Harja, 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* **5** (2011), 149.

[30] T.B.L. Kirkwood, Understanding the odd science of aging, *Cell* **120** (2005), 437–447.

[31] A.L. Koch, On evidence supporting a deterministic process of bacterial growth, *J Gen Microbiol* **43** (1966), 1–5.

[32] E. Laskowska, J. Bohdanowicz, D. Kuczynska-Wisnik, E. Matuszewaska, S. Kedzierska and A. Taylor, Aggregation of heat-shock-denatured, endogenous proteins and distribution of the IbpA/B and Fda marker-proteins in *Escherichia coli* WT and grpE280 cells, *Microbiology* **150** (2004), 247–259.

[33] A.B. Lindner and A. Demarez, Protein aggregation as a paradigm of aging, *Biochim Biophys Acta* **1790** (2009), 980–996.

[34] A.B. Lindner, R. Madden, A. Demarez and E.J. Stewart, Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation, *Proc Natl Acad Sci U S A* **105** (2008), 3076–3081.

[35] J. Lloyd-Price, A. Gupta and A.S. Ribeiro, SGNS2: A compartmentalized stochastic chemical kinetics simulator for dynamic cell populations, *Bioinformatics* **28** (2012), 3004–3005.

[36] E. Maisonneuve, B. Ezraty and S. Dukan, Protein aggregates: An aging factor involved in cell death, *J Bacteriol* **190** (2008), 6070–6075.

[37] J. Mandelstam, The intracellular turnover of protein and nucleic acids and its role in biochemical differentiation, *Bacteriol Rev* **24** (1960), 289–308.

[38] C.G. Miller and L. Green, Degradation of abnormal proteins in peptidase-deficient mutants of Salmonella typhimurium, *J Bacteriol* **147** (1981), 925–930.

[39] O.L. Miller, B.A. Hamkalo and C.A. Thomas, Visualization of bacterial genes in action, *Science* **169** (1970), 392–395.

[40] M. Miot and J.-M. Betton, Protein quality control in the bacterial periplasm, *Microb Cell Fact* **3** (2004), 4.

[41] M. Morell, R. Bravo, A. Espargaró, X. Sisquella, F.X. Avilés, X. Fernàndez-Busquets, et al., Inclusion bodies: Specificity in their aggregation process and amyloid-like structure, *Biochim Biophys Acta* **1783** (2008), 1815–1825.

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

[43] J. Mäkelä, J. Lloyd-price, O. Yli-Harja and A.S. Ribeiro, Stochastic sequence-level model of coupled transcription and translation in prokaryotes, *BMC Bioinformatics* **12** (2011), 121.

[44] T. Nyström, Aging in bacteria, *Curr Opin Microbiol* **5** (2002), 596–601.

[45] M.J. Pine, Metabolic control of intracellular proteolysis in growing and resting cells of *Escherichia coli*, *J Bacteriol* **92** (1966), 847–850.

[46] A.S. Ribeiro and J. Lloyd-price, SGN Sim, a stochastic genetic networks simulator, *Bioinformatics* **23** (2007), 777–779.

[47] A. Ribeiro, R.U.I. Zhu, S.A. Kauffman and A. General, Modeling strategy for gene regulatory networks with stochastic dynamics, *J Comput Biol* **13** (2006), 1630–1639.

[48] M.R. Roussel and R. Zhu, Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression, *Phys Biol* **3** (2006), 274–284.

[49] M. Scott, C.W. Gunderson, E.M. Mateescu, Z. Zhang and T. Hwa, Interdependence of cell growth and gene expression: Origins and consequences, *Science (80)* **330** (2010), 1099–1102.

[50] C. Soti and P. Csermely, Aging and molecular chaperones, *Exp Gerontol* **38** (2003), 1037–1040.

[51] M.A. Speed, D.I. Wang and J. King, Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition, *Nat Biotechnol* **14** (1996), 1283–1287.

[52] E.J. Stewart, R. Madden, G. Paul and F. Taddei, Aging and death in an organism that reproduces by morphologically symmetric division, *PLOS Biol* **3** (2005), 0295–0300.

[53] S.A. Teter, W.A. Houry, D. Ang, T. Tradler, D. Rockabrand, G. Fischer, et al., Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains, *Cell* **97** (1999), 755–765.

[54] J. Tyedmers, A. Mogk and B. Bukau, Cellular strategies for controlling protein aggregation, *Nat Rev Mol Cell Biol* **11** (2010), 777–788.

[55] J.A. Valkenburg and C.L. Woldringh, Phase separation between nucleoid and cytoplasm in *Escherichia coli* as defined by immersive refractometry, *J Bacteriol* **160** (1984), 1151–1157.

[56] L. Wang, S.K. Maji, M.R. Sawaya, D. Eisenberg and R. Riek, Bacterial inclusion bodies contain amyloid-like structure, *PLoS Biol* **6** (2008), e195.

[57] P. Wang, L. Robert, J. Pelletier, W.L. Dang, F. Taddei, A. Wright, et al., Robust growth of *Escherichia coli*, *Curr Biol* **20** (2011), 1099–1103.

[58] S. Ventura and A. Villaverde, Protein quality in bacterial inclusion bodies, *Trends Biotechnol* **24** (2006), 179–185.

[59] E. Viaplana, J.X. Feliu, J.L. Corchero and A. Villaverde, Reversible activation of a cryptic cleavage site within *E. coli* beta-galactosidase in beta-galactosidase fusion proteins, *Biochim Biophys Acta* **1343** (1997), 221–226.

[60] S. Wickner, Posttranslational quality control: Folding, refolding, and degrading proteins, *Science (80)* **286** (1999), 1888–1893.

[61] N.S. Willetts, Intracellular protein breakdown in non-growing cells of *Escherichia coli*, *Biochem J* **103** (1967), 453–461.

[62] J. Winkler, A. Seybert, L. König, S. Pruggnaller, U. Haselmann, V. Sourjik, et al., Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing, *EMBO J* **29** (2010), 910–923.

[63] J. Yu, J. Xiao, X. Ren, K. Lao and X.S. Xie, Probing gene expression in live cells, one protein molecule at a time, *Science (80)* **311** (2006), 1600–1603.

[64] P. Zhang, C.M. Khursigara, L.M. Hartnell and S. Subramaniam, Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy, *Proc Natl Acad Sci U S A* **104** (2007), 3777–3781.

[65] R. Zhu, A.S. Ribeiro, D. Salahub and S.A. Kauffman, Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models, *J Theor Biol* **246** (2007), 725–745.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3601-4
ISSN 1459-2045