An image segmentation, lineage and analysis tool for bacteria: Applications in cell proliferation and cytoplasmic dynamics.

Stella Stylianidou

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Paul Wiggins, Chair

Beth Traxler

Marcel den Nijs

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

An image segmentation, lineage and analysis tool for bacteria: Applications in cell proliferation and cytoplasmic dynamics.

Stella Stylianidou

Chair of the Supervisory Committee:
Professor Paul Wiggins
Physics

Because of the stochastic nature and significant cell-to-cell variation of many biological processes, it is essential to analyze a significant number of cells to understand such processes at the single cell level. Such quantitative biological questions require fast yet reliable automated computational algorithms to analyze images containing thousands of bacterial cells. We present an automated image processing package well-suited to quantitative analysis of high-throughput live-cell fluorescence microscopy of bacterial cells. We describe several capabilities of our software, including identifying and linking cells from frame-to-frame, and characterizing the cell morphology and fluorescence. In particular, we demonstrate its strong performance in reliably segmenting micro-colonies with many cells, facilitating the analysis of cell-cycle dynamics in bacteria as well as cell-contact mediated phenomena.

We apply the package in two separate quantitative biological questions. We investigate cell proliferation with single cell resolution during lag phase, the transition of non-growing cells to a growing state. By analyzing the generation times of growing and non-growing progenitors, we suggest a qualitative model for lag phase.

As a second application, we characterize the anomalous dynamics in the cytoplasm by quantitative analysis of thousands of complete cell-cycle fluorescence trajectories of molecular complexes. Our results support two modes of nucleoid action: volume exclusion in organizing the cell and a mode of rapid motion. We also report emergent self-similar structure in the dynamics of the cytoplasm and propose a general mechanism by which scale independence emerges in the strong-disorder limit of biological systems.

TABLE OF CONTENTS

Page

LIST OF FIGURES

ABBREVIATIONS

ATC: anhydrotetracycline

CTRW: Continuous Random Time Walk

DNA: Deoxyribonucleic acid

EVT: Extreme Value Theory

FBM: Fractional Brownian Motion

FRAP: Fluorescence Recovery After Photobleaching

GFP: Green Fluorescent Protein

IPTG: Isopropyl β-D-1-thiogalactopyranoside

MRNA: Messenger Ribonucleic Acid

MSD: Mean Square Displacement

RNA: Ribonucleic Acid

RG: Renormalization group

ACKNOWLEDGMENTS

When I started my PhD degree at the University of Washington, I did not intend to study bacteria or even delve into scientific research software. My undergraduate studies did not include any biology classes so I had a vague idea what a bacterium was and any software knowledge was self-taught in my spare time. I remember walking in the Wiggins lab, and how fascinated I was by what bacteria can do, and by what we can do with microscopes and computers. My advisor, Paul Wiggins, welcomed me in the lab and spent numerous hours explaining biology and physics concepts to me and teaching me how to use different experimental apparatus. His excitement when it comes to science is contagious. None of this work would have been possible without him, his ideas, guidance and expertise.

The other person that welcomed me in the lab and helped to mold me into a scientist is our old post-doctoral scientist, Nathan Kuwada. I am thankful for his organizational skills, his patience and smile while answering numerous questions and for teaching me everything, from making slides to programming in MATLAB.

I would like to express my gratitute to the reading committee, Paul Wiggins, Beth Traxler, and Marcel den Nijs, for reading my thesis, and the rest of the committee members, Joshua Vaughan, Jens Gundlach, and Blayne Heckel, for their help. I would like to especially thank Beth Traxler for her patience while explaining the simplest biological concepts, and for her help in constructing strains for the plasmid project. I am also thankful the physics graduate advisor, Catherine Provost, for her emotional support and advice during grad school.

Some of the ideas in this thesis developed through discussions with Thomas Lampo and Andrew Spakowitz, and I am grateful for their collaboration and useful comments.

My sincere thanks to the members of the Wiggins Lab, Julie Cass, Jackie Corbitt, Sarah Mangiameli, Colin LaMont. They have been very supportive throughout the graduate school experience : helpful to bounce ideas off, offer constructive criticism, or just lend an ear. I also thank the undergraduate student I mentored, Connor Brennan, for our great long science discussions and his contributions to SuperSegger, and Silas Nissen, who worked on the SuperSegger GUIs, while he was visiting our lab for a quarter.

I made some great friends during grad school, who supported me through the years; Heather, Jackie, Tomasz, John, Isaac, Matthias, Akshay, Durmus, have been amazing at listening to my breakthroughs and frustrations or discussing science at College Inn.

Finally, I would like to thank my family and friends in Cyprus. They believed in me from the very start. My parents, since I was young, supported me financially and emotionally through my educational journeys. They were always supportive while I moved further and further away to obtain yet another degree. Being far away from family and friends has been hard, but I could always feel their love and support.

DEDICATION

to my yiayia and pappou.

PUBLICATIONS

This thesis is based on the following articles.

I Stella Stylianidou, Nathan J Kuwada, and Paul A Wiggins. Cytoplasmic dynamics reveals two modes of nucleoid-dependent mobility. Biophysical Journal, 107(11):2684-92, Dec 2014.

II Stella Stylianidou, Connor Brennan, Silas B Nissen, Nathan J Kuwada, and Paul A Wiggins. SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells. Molecular Microbiology, 102: 690-700, Sep 2016.

III Stella Stylianidou, Thomas J Lampo, Andrew J. Spakowitz, and Paul A Wiggins. Strong disorder leads to scale-invariance in complex biological systems. In preparation.

IV Thomas J Lampo, Stella Stylianidou, Mikael P. Backlund, Paul A. Wiggins, and Andrew J. Spakowitz. Cytoplasmic RNA-protein particles exhibit non-Gaussian subdiffusive behavior. Biophysical Journal, 112(3): 532 - 542, Feb 2017.

V Julie A. Cass, Stella Stylianidou, Nathan J. Kuwada, Beth Traxler, and Paul A. Wiggins. Probing bacterial cell biology using image cytometry. Molecular Microbiology. Jan 2017.

Reprints were made with the permission from the respective publishers.

Chapter 1

INTRODUCTION

We often feel that we, humans, are the most important organism on earth. But you and I are outnumbered. Wherever you look and whatever you touch, you will find those minuscule single-celled organisms, bacteria. From inhibitors of our guts that produce vitamins to pathogens that may get us sick, bacteria are part of our life.

What makes bacteria interesting is that the physics of bacteria are quite different from the physics we experience. On the microscopic scale, Newton's laws of motion, such as inertia, become less important. The viscous drag a bacterium experiences while swimming in water, would be like what you would experience swimming in molasses [7].

Furthermore the physics of molecules within a bacterial cell are quite different from the physics of our every day life. In nanoscopic scale, collisions with other molecules play a big effect in the motion of a molecule. Molecules go through a random or "drunken" walk through a very crowded environment; they stumble right and left as they bump into other molecules.

This picture of a random walk may create the impression that molecules in bacteria are randomly distributed in space and in time, and for a long time this was the common scientific understanding. But in-vivo imaging has revealed that bacteria display a complex inter-cellular structure. Different proteins and genes have been found to show temporal and spatial localization [8, 9, 10, 11]. In many cases, it is not yet clear how bacteria drive this temporal and spatial sub-cellular organization. Eukaryotic cells use membrane-bound organelles and cytoskeletal molecular motors for spatiotemporal organization, but bacteria lack those components.

To understand the mechanisms behind sub-cellular organization, it is essential to analyze paths of thousands of molecules from single cells. This is because the stochastic nature of cellular processes causes significant cell-to-cell variation [1]. Analyzing thousands of single cells would not be possible without powerful microscopes able to image time-lapses of large fields of view, but also without computational tools that can speed up the analysis of these large datasets. It is therefore necessary to develop a fast, automated and reliable tool that can identify cells in images, track an individual cell from frame-to-frame, and track fluorescent molecules of interest from frame-to-frame.

This thesis describes our solution to this problem, a software suite able to identify the regions that correspond to bacterial cells in phase images, track individual cells from frame-to-frame and track proteins, pictured as fluorescent foci, in images. Our software allows users to segment and track many cells in close contact, making it possible to collect data with enough statistics to answer quantitative biological questions. We have already applied our software to multiple quantitative biological questions, and in this thesis we present two of those applications and analysis related to those results.

1.1 Outline

As this work is in the intersection of biology and physics, the first chapter will cover some of the relevant basics of the two disciplines. We describe the basic biology of the model organism we used, *Escherichia coli* (*E. coli*), and focus on the physics of the cytoplasm. We also cover the basics of the imaging techniques used to collect our typical datasets.

The second chapter focuses on the software suite that we wrote, SuperSegger. We describe the methods and capabilities of the software, compare it with other image segmentation software for bacteria, and present the achieved error and time benchmarks.

In the third chapter, we apply the software to a biological problem, the characterization of cell proliferation in the single cell level. Specifically we are interested in the adjustment of cells to a new environment. We will compare the adjustment of five generations of cells from previously growing versus non-growing progenitors.

In the fourth chapter, SuperSegger is used to investigate cytoplasmic dynamics of large molecular complexes in *E. coli* cells. We use as a probe MS2-mRNA molecular complexes that are ectopic to the cell so that their dynamics are not affected by interaction or binding with intracellular components. We will show that our results support two modes of nucleoid action: volume exclusion in organizing the cell and a mode of rapid motion throughout the cytoplasm.

In the fifth chapter, we focus on the emergence of self similarity in the dynamics of the molecular complexes studied in chapter four. We explore strongly disordered systems and propose that extreme value theory can be applied to understand the emergence of self-similarity in the dynamics of large molecular complexes and even more biological systems limited by strong disorder.

In the sixth chapter, we conclude with a summary of the main results and some guidelines for future work.

1.2 Life of a microbe

1.2.1 Central Dogma of biology

A bacterium can be considered to be composed of three main components: the cell envelope, DNA and cytoplasm. In this section we focus on the DNA and its conversion to proteins, which are the cell machinery required for life, and we will describe the cell envelope and cytoplasm in the next section. The DNA is similar to an instruction manual containing instructions to create all components of the cell, and is written in an alphabet of four letters, A, T, C and G. These small units are combined in a series to form a long polymer, a strand of DNA. The strand of DNA pairs with another strand of complimentary sequence to form the double-helix structure.

DNA is converted to proteins in two steps. First, the DNA instructions are rewritten into another type of polymer, the RNA messages. The conversion of DNA to RNA is called *transcription*. Then, the RNA messages travel to the factories of the cell, the ribosomes, where they are *translated* into proteins.

1.2.2 Escherichia coli a model organism

One of the most studied organisms and the main focus of this thesis is a bacterium called *E. coli*. Because of it's relatively fast and easy growth in the lab, *E. coli* has been used as a model organism to understand more complex cells and life.

It is a rod-shaped, single-celled organism, about $2\,\mu\mathrm{m}$ long and $1\,\mu\mathrm{m}$ wide. It is a gram negative bacterium, meaning it does not retain the crystal violet stain when stained with Gram's method. Commonly found in the intestines of warm-blooded animals, it consists $\sim 0.1\%$ of the gut flora [12].

The outermost part of an *E. coli* is the cell envelope, composed of three parts: an outer membrane, a thin peptidoglycan layer and an inner cytoplasmic membrane. The cell envelope separates the inside of the cell from the outer world and permits the flow of molecules across it.

The *E. coli* DNA is in a single circular chromosome, the nucleoid. It is about 4,400 kilobase pairs long and has 2000 genes [13]. The nucleoid is highly compact; a stretched *E. coli* chromosome would be $1\,\mathrm{mm}$ long, about 1000 times longer than an *E. coli* cell.

The rest of the region between the inner membrane and the nucleoid is the cytoplasm, and it contains proteins, RNA, ribosomes and other particles (Fig. 1.1). Moving through the cytoplasm is not easy, it is gel-like with four times the fluid viscosity of water and its meshwork is estimated to have an average pore size between $200\,\text{Å}$ - $400\,\text{Å}$ [14]. Although there is no membrane separating the chromosome from the cytoplasm, the two can be almost regarded as separate due to the high compaction of the chromosome.

Other than the nucleoid, some cells contains smaller extrachromosomal circular DNA molecules that are called plasmids, which often carry genes for antibiotic resistance.

1.2.3 Bacterial Growth

In order for bacteria to grow, that is to increase their size and mass, they require a narrow range of environmental factors such as pH, temperature, osmotic pressure, and nutri-

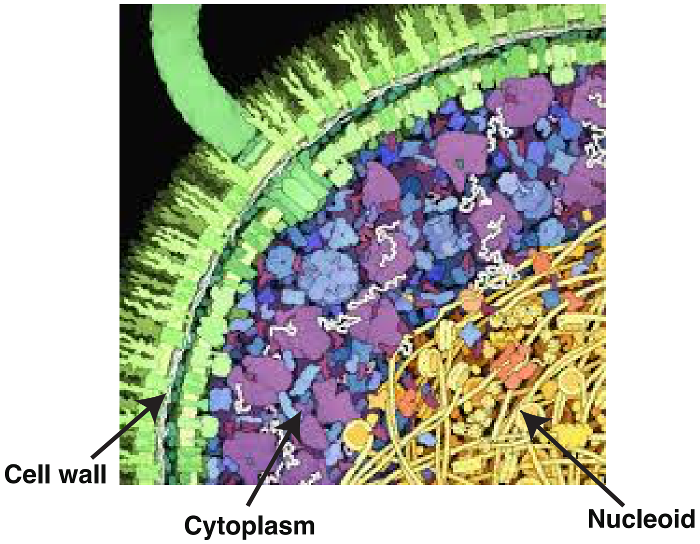


Figure 1.1: **Schematic of a cross section of an E-coli cell.** Illustration by David S. Goodsell, the Scripps Research Institute.

tional elements [13].

When *E. coli* are in the growing state, a mother *E. coli* cell elongates, doubling in length while its diameter remains about constant [13]. The mother cell then divides into two daughter cells from the inward growth of the cytoplasmic membrane and cell wall in mid-cell [15]. During this process the chromosome is replicated and each daughter receives an identical copy. The caps of the rod shaped cell are called poles. The poles inherited from the mother cell are defined as old poles and the poles formed by the last division are defined as new poles (Fig. 1.2). The generation time for *E. coli*, the time it takes for one cell to divide into two, under the best nutritional conditions is about 20 minutes [15].

The bacterial growth dynamics can be studied using the growth curve. The growth curve can be plotted by measuring the optical density of a sample of bacteria with a spectrometer. Light passing through the cell culture is scattered by cells, thus by measuring the unscattered light we can assess the cell number [15]. The growth curve has a sigmoid

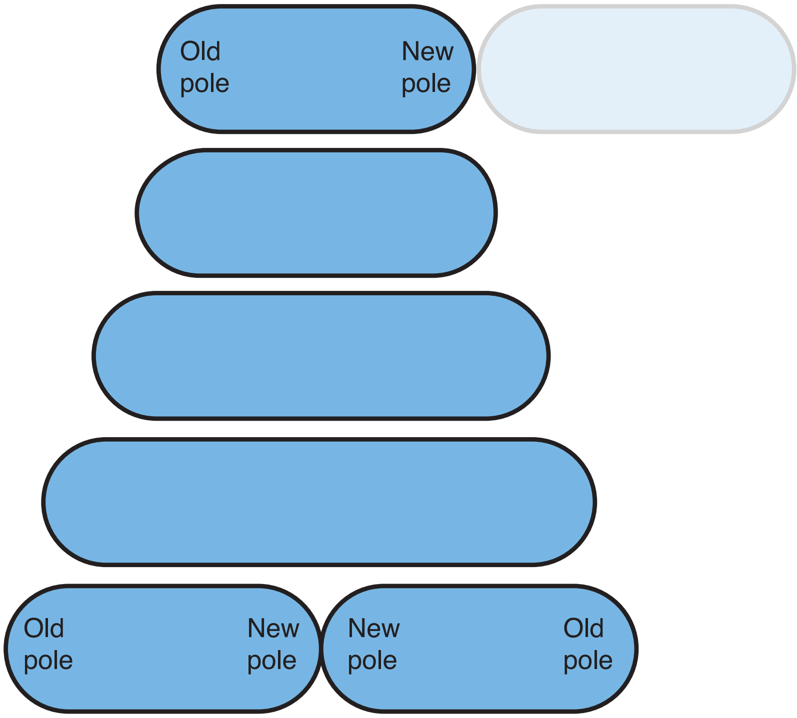


Figure 1.2: **Cell Division.** An *E. coli* cell elongates and divides in mid-cell to two daughter cells. Old poles are the the poles inherited from the mother cell and new poles are the ones formed by the last division. A new pole will become an old pole in the next division.

shape and consists of four phases:

(a) Lag phase: When bacteria are inoculated to a new environment and are adjusting to the change, growth does not immediately begin. The length of this phase depends on the environmental conditions before versus current [15].

(b) Log or exponential phase: In this state, the cells are actively growing. They double at constant rate, and the number of bacteria increases exponentially. The doubling rate can vary depending on the environmental conditions and the organism [15]. In most bacterial experiments, cells in this active growing state are studied.

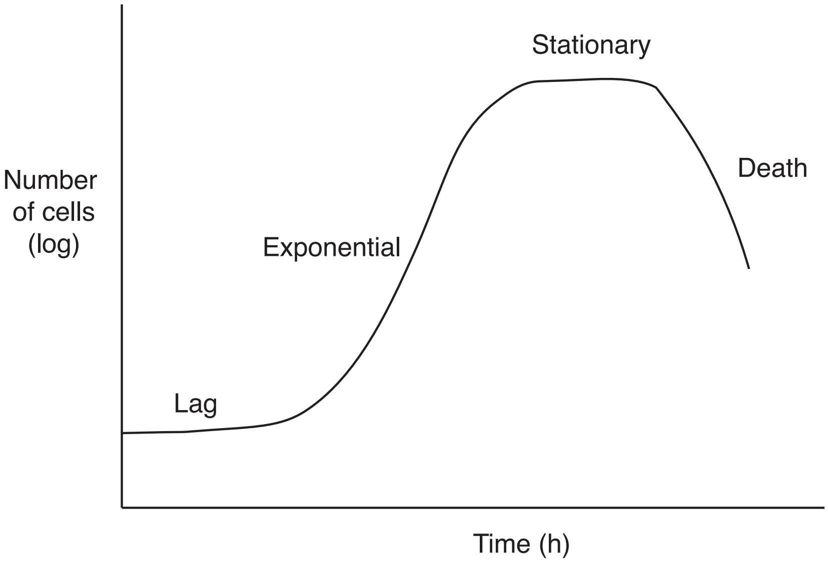


Figure 1.3: **Bacterial growth curve.**

(c) Stationary phase: As essential nutrients of the culture medium are exhausted or the waste products accumulate, growth ceases and the cells stop dividing. This phase is known as the stationary phase because the cell number does not increase or decrease and appears flat in the growth curve. Although the cell number does not grow, cells are still alive and continue to metabolize [15].

(d) Death: At this final stage, the cells eventually die and can not reproduce even when they are transferred to a new medium. During this phase, the growth rate in the growth curve decreases exponentially [15].

1.3 Inside a microbe

1.3.1 Diffusing within a cell

One of the most interesting things about bacteria is the physics of the molecules in the crowded and complex cytoplasm. Inside a bacterium, the random thermal forces, the collisions with surrounding molecules, play an equally as important role as the usual deterministic forces [16]. Due to these thermal forces, molecules observed in cells appear to take each step in a random direction and they are said to undertake brownian motion or diffusion. In order to achieve molecular transport, cells need to exploit or overcome this motion.

A common measure for diffusive motion is the mean square displacement (MSD). By taking the average over the trajectories of the particles, we can compute the MSD in one dimension

$$\mathrm{MSD} = \left\langle \Delta x^2(\delta t) \right\rangle = \left\langle (x(\delta t) - x_0)^2 \right\rangle = 2D\delta t^{\alpha} \tag{1.1}$$

where D is the diffusion coefficient and α is the scaling factor. In the case of simple diffusion the scaling factor, α, is 1. An $\alpha < 1$ and $\alpha > 1$ indicate sub-diffusive and super diffusive motion respectively. Sub-diffusive motion has been observed in cells for example in chromosomal loci [17] and the cytoskeletal protein FtsZ [18].

For a spherical particle, the diffusion coefficient is defined by the Stokes-Einstein relation

$$D = \frac{kT}{6r\pi\eta} \tag{1.2}$$

where k is the Botlzmann constant, T is the temperature, η is the viscosity of the medium and r is the radius of the particle.

Another way to describe diffusion is in terms of the change in particle concentration in time and space. In a system undergoing diffusion, if there is a gradient in particle concentration there will be a flow of particles from high to low concentration. Fick's first law describes how the flux of particles, the amount of particles through a unit area per second, changes with the concentration gradient

$$\vec{j} = -D\boldsymbol{\nabla} n \tag{1.3}$$

where j is the diffusion flux and n the particle concentration.

Assuming the number of particles is conserved, we can derive Fick's second law

$$\frac{\partial n}{\partial t} = -\boldsymbol{\nabla} \cdot \vec{j}. \tag{1.4}$$

Combining 1.3 and 1.4 we arrive at

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial^2 x}. \tag{1.5}$$

1.3.2 Bias and diffusion within a cell

The analysis above is based on the assumption that there are no external forces other than diffusion. If there is a bias towards a particular direction, equations 1.3 and 1.5 are modified as follows [16]

$$\vec{j} = -D\boldsymbol{\nabla} n + \langle v \rangle n \tag{1.6}$$

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial^2 x} - \langle v \rangle \frac{\partial n}{\partial x} \tag{1.7}$$

The drift velocity, $\langle v \rangle$, is defined as

$$\langle v \rangle = \frac{F}{\gamma} \tag{1.8}$$

where F is the bias force and γ is the friction coefficient.

1.4 Imaging Techniques

1.4.1 Phase-Contrast Microscopy

To image individual bacterial cells we use light microscopy. Light microscopy is based on the absorption of light by the specimen, which is commonly viewed as dark against a lighter background. Because bacteria are transparent to light, more advanced techniques have been developed to resolve microbes, such as phase-contrast. Phase-contrast microscopy relies on the differences in the refractive index between the cells and the surrounding region [19]. The diffracted light passing through a cell is about a quarter of a wavelength behind the phase of light passing through the surrounding medium. On the other hand, the surrounding light is advanced an additional one quarter of a wavelength by a phase plate in the microscope so that it destructively interferes with the diffracted light [19]. The result is specimens appearing in the image as dark regions in a light filled background.

Cells are also constantly moving, thus need to be immobilized to resolve cellular regions in images. To achieve this the cells are trapped in a gelatinous matrix. We place a few μl of cell culture between a glass slide and an agarose pad, which is a combination of growth media and a matrix called agarose [20]. Multiple generations of cells can be observed on the pad, until nutrients and space are exhausted and cells reach stationary phase.

1.4.2 Fluorescence

To image molecules within cells we use fluorescence microcopy. Fluorescence is the process during which light of a specific wavelength is absorbed and light of a lower wavelength (usually) is emitted. Molecules that have this capability are called fluorophores. The process is illustrated in an electronic state diagram in figure 1.4. Fluorophores have at least three energy states, E_1' where the electron arrives by absorbing a photon, E_1 where the electron relaxes to after it has been excited, and E_0 the ground state. Fluorescence emission results from the transition of an electron between E_1 and E_0. By exciting the sample with photons of the absorption energy ($E_1' - E_0$), and filtering the emitted light for photons of the excitation energy ($E_1 - E_0$), the location of the fluorophores can be recorded.

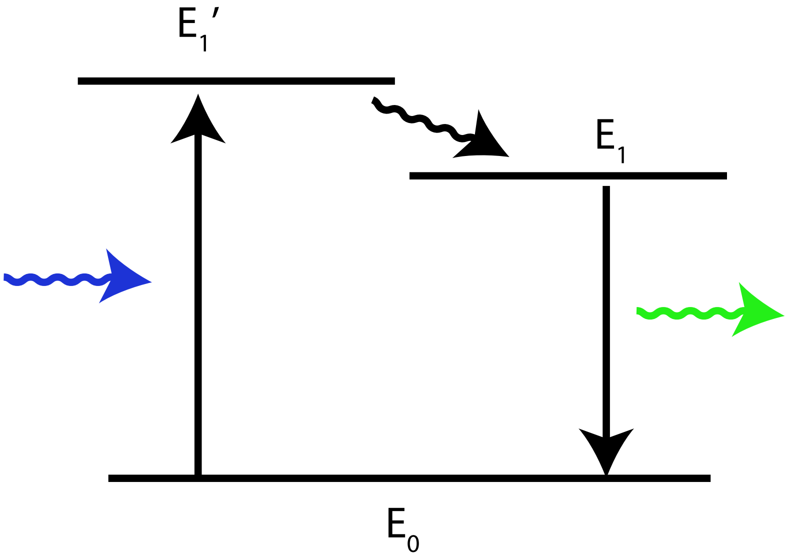


Figure 1.4: **Fluorescence Energy Diagram**. The electron absorbs a photon of energy E_1'-E_0, relaxes to E_1 state and emits a photon of energy E_1-E_0

Fluorophores after exposure to light can lose their ability to fluoresce, a process called photobleaching. Photobleaching can complicate fluorescence microscopy data, as fluorescent molecules may not appear in images for a duration of time. Scientists have taken advantage of this problem, by creating techniques that purposefully bleach the cells to study the motion or diffusion of molecules, such as FRAP (Fluorescence Recovery After Photobleaching).

1.4.3 Fluorescent Proteins

The first discovered fluorescent protein, the green fluorescent protein (GFP), was isolated in the 1962 from the jelly fish *Aequorea victoria* [21]. GFP absorbs blue light, with an excitation peak of $488\,\mathrm{nm}$, and emits green light, with an emission peak of $509\,\mathrm{nm}$. In the 1994, the genes that produce GFP were transferred to *E. coli* cells [22]. GFP has been used extensively in the last years to track gene expression, protein localization and concentration, by fusing it to other proteins or by inserting its gene after a promoter of interest (DNA region where transcription of a gene is initiated) [20]. Now, a variety of spectral variants are available (e.g. mCherry, YFP etc), and combinations of these make it possible to visualize multiple proteins simultaneously.

1.4.4 Diffraction limit

Despite the technological advancements in microscopy, the resolution by which two objects can be distinguished as separate is limited by the physical laws of light. The diffraction limit is

$$d = \frac{1.22\lambda}{\mathrm{NA}} \tag{1.9}$$

where λ is the wavelength of light and NA is the numerical aperture of the microscope. For the microscopes we use, the resolution limit is about half the wavelength of the light used to illuminate the specimen, about $250\,\mathrm{nm}$ or an eighth of a typical *E. coli* cell. Therefore fluorescent proteins with separation less than $250\,\mathrm{nm}$ can not be resolved as two separate objects with conventional fluorescent microscopy. To resolve such objects methods which modulate the activation of fluorophores either temporally or spatially, known as super resolution techniques, have been developed [23].

Chapter 2

SUPERSEGGER : AN IMAGE SEGMENTATION AND ANALYSIS TOOL

As discussed in the introduction, the cell-to-cell variation makes the quantitative statistical analysis of a significant number of cells essential to understanding many biological processes at the single cell level [1]. Although it is straightforward to image fields of view containing thousands of bacterial cells, the development of fast, reliable and automated methods for the quantitative analysis of this data, including segmentation (the identification of regions in each image corresponding to cells) and linking the cell regions from frame-to-frame, remains a significant computational challenge.

In this chapter, we present our solution to this problem: SuperSegger, an automated MATLAB-based image processing and analysis package. SuperSegger is particularly well suited for high-throughput analysis of cell-cycle dynamics of proteins and complexes by time-lapse fluorescence microscopy in single bacterial cells. Even though the software is optimized for rod-like bacterial cells, it incorporates machine-learning algorithms, to optimize cell boundaries for other cell shapes. SuperSegger can follow a cell lineage for many generations, identifying full cell cycles by linking cells from frame-to-frame in time-lapse imaging. The software performs a detailed characterization of each cell facilitating a wide range of analyses. Post-processing tools are included for analysis at both the single cell and population level, such as fluorescence kymographs, frame mosaics, consensus images and plotting tools for a variety of cellular characteristics. Powerful tools for subpopulation analysis have also been developed that allow the user to gate (generate cell subpopulations) on a wide range of cellular characteristics. The software can be downloaded from the Wiggins Lab website where the user manual and documentation on all the methods can be found. It can then either be run at the command-line or through a graphical user interface (See Appendix Fig. A.2 and A.1).

Although a number of segmentation software packages are already available [24, 25, 26, 27, 28, 20], our software was developed because we found the existing packages were not well suited to our own experiments and analysis. In our experience, these packages cannot track cells over multiple generations without significant error correction by hand. This approach is not practical when analyzing a large number of cells. In contrast we demonstrate that SuperSegger can segment multiple generations of cell division without error. Therefore, SuperSegger offers strengths complementary to these other packages, especially in experimental contexts dependent on time-lapse imaging and cultures where cells have many neighbors.

2.1 Software Methods

2.1.1 Overview

SuperSegger automates segmentation, linking and cytometry analysis and requires no user input (Fig. 2.1). In short: The user specifies the image directory and selects the appropriate segmentation parameters (typically using a set supplied with the package). SuperSegger loads images for each time point, aligns the images, identifies the cell regions from the background, tracks each cell region from frame-to-frame by linking corresponding regions, and calculates fluorescence and structural characteristics for each cell in each frame. SuperSegger generates three different types of outputs: Frame files, Clist Matrices and Cell Files, which are further described below.

2.1.2 Image Alignment

During long time-lapse imaging, stage drift often results in the accumulation of offset of the field of view between the initial and the final frames. In our own experience, this problem is often aggravated by imaging multiple xy positions during a time course. To

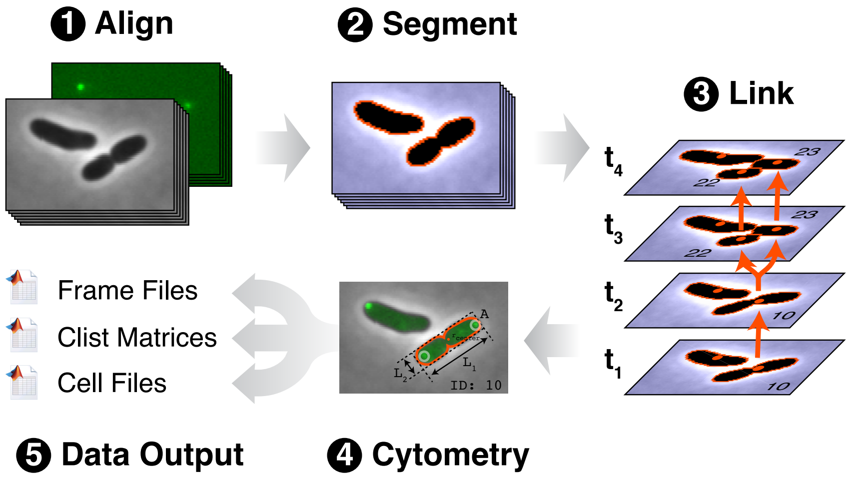


Figure 2.1: **General process of SuperSegger.** The fluorescence and phase images are processed and aligned. During segmentation the cell regions are identified from the background. Then each cell region is linked to a corresponding cell region in the next frame and the cells receive unique ID numbers. Next, the properties and fluorescence characteristics of each cell are calculated. Finally, the program outputs three different types of outputs : Frame files, Clist matrices and Cell files.

implement frame alignment, we use a fast and memory-efficient algorithm by [29] which determines the frame alignment to sub-pixel resolution using cross-correlation.

SuperSegger includes a number of important alignment features: (i) Alignment can be performed in any master channel (phase contrast, fluorescence, etc). (ii) A constant relative offset can be applied to each channel to compensate for known offsets between channels. (iii) Alignment can be performed against the previous frame, suitable for long time-lapse experiments, or against the first frame, suitable for short experiments where no significant cell growth occurs. (iv) Frames with high alignment error or out of focus are skipped automatically. (v) The image is preserved without cropping, facilitating the analysis of micro-colonies that drift in (or out) of the field of view during long time courses.

2.1.3 Cell segmentation

The process of identifying the regions of the image which correspond to cells is called image segmentation. We developed the segmentation algorithm for use on phase contrast images (Fig. 2.2A), although the algorithm also performs well on inverted fluorescence images of cells with cytoplasmic fluorescence (J.Russell, personal communication).

The first step of segmenting cells is differentiating cell micro-colonies (or clusters) from empty regions of the frame. The image is first smoothed and normalized. Next, the difference of the image filtered by gaussians of different sizes results in an image of cell clusters of high relative intensity. The initial micro-colony mask is created by a thresholding operation on this image. To further refine the micro-colony mask we filter the image with a contrast filter to compensate for the phase contrast shade-off (the loss of phase-based contrast in the middle of micro-colonies) and halo artifacts [30]. Another thresholding operation is used on the contrast filtered image to remove regions between cells and obtain the final micro-colony mask (Fig. 2.2B). (The segmentation steps are shown in more detail in Fig. A.3.) The outline of the mask of each micro-colony is set as a fixed cell

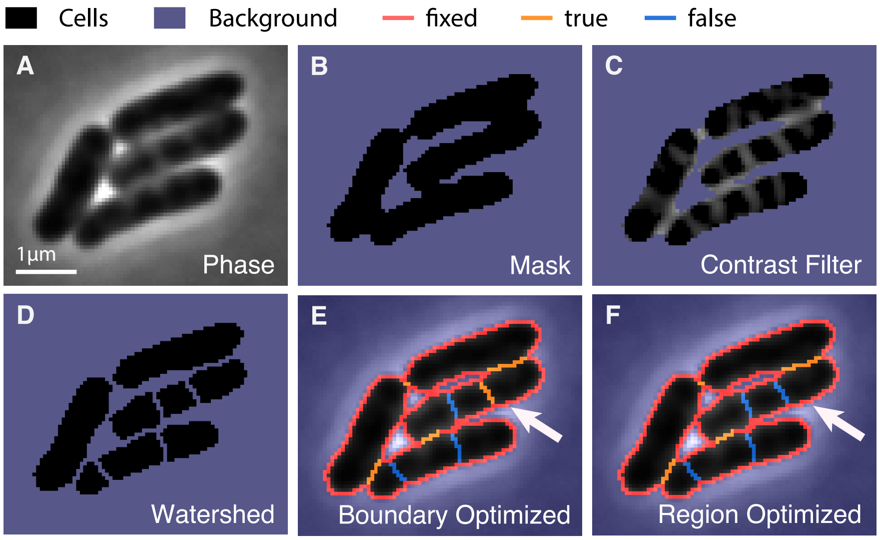


Figure 2.2: **Image segmentation procedure. Panel A:** Original phase image. **Panel B:** Mask of cells using intensity thresholding. **Panel C:** Phase image after the contrast (maximum principal curvature) filter and the mask are applied. **Panel D:** Boundaries found using the watershed function. **Panel E:** Boundaries after the boundary optimization. **Panel F:** Boundaries after region optimization. Boundaries are divided into fixed boundaries (red), boundaries that the software classified as true (orange) and false (blue). In panel E on the right side, there is a boundary incorrectly set as true after boundary optimization which is set to false after region optimization in panel F.

boundary during the rest of the segmentation procedure.

To identify cell boundaries within a micro-colony we use the watershed operation. The watershed operation identifies the catchment basins of local intensity minima, which are the dark regions of the image corresponding to cells. This results to over-segmentation: the typical cell is subdivided into multiple watershed regions (Fig. 2.2D). Watershed regions too wide to represent cells are further subdivided to ensure that each cell is represented by at least one watershed region.

The cell regions are formed by merging watershed regions. This is done by classifying the state of boundaries between them (Fig. 2.2E). The boundaries are classified as true or false using a neural networks trained on the boundaries of the specific cell type, growth conditions, pixel size etc. The input variables for each boundary include: mean and minimum intensity of the boundary, the second derivative of the intensity of the boundary, the length of the boundary, the area of neighboring regions, axes of neighboring regions etc.

The state of boundaries neighboring low scored regions are further informed by globally optimizing a score function which includes the scores of both regions and boundaries in question (Fig. 2.2F). To score a cell region, we use a second neural network trained to classify true or false regions. The input variables for each cell region include: the length of the long and short axis of the region, the area of the region, the neck of the region etc. Depending on the number of marginal boundary state classifications, either an exhaustive or stochastic (simulated anneal) search algorithm is used to identify the optimal segmentation.

2.1.4 Region linking

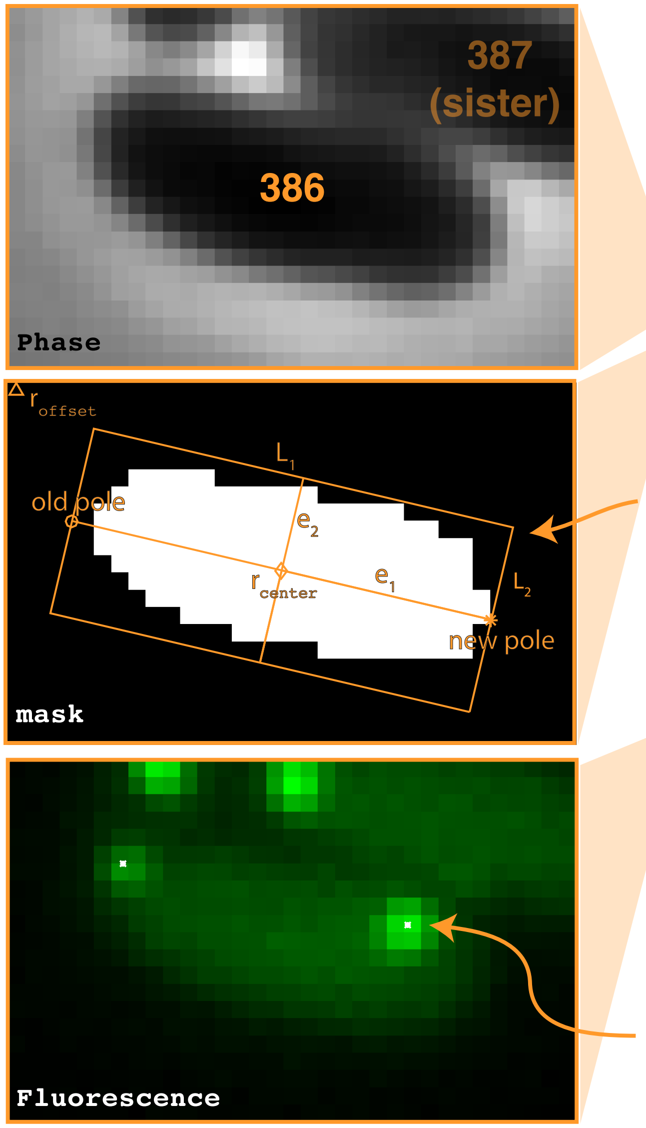
In snapshot analysis, the cell segmentation process is now complete, but for time-lapse analysis, the cell regions in successive frames must be linked to generate a cell cycle. The linking algorithm links each region to a single region (or a pair of regions) in the successive frame. A linking cost function is computed from (i) the spatial overlap between regions, (ii) the distance between the region centroids of the two regions, and (iii) the change in the areas of the regions. Because cells at the outmost part of the micro-colony move further outwards than internal cells, the distance from the center of the colony is used to bias assignment to cells further outwards in the next frame. The linking corresponding to the minimum cost is initially assigned, and the assigned regions are removed from the possible assignments until no possible assignment remains. The segmentation is further improved using the linking information from frame-to-frame to identify boundaries that may have been incorrectly classified as true or false. In the event that the area change is above a defined threshold the region is marked with an error flag. This makes it trivial to exclude incorrectly segmented cells at the end analysis.

2.1.5 Cell cytometry

After the segmentation process is complete, a detailed and time-dependent cell cytometry analysis is performed. The cell descriptors are stored in a cell data structure. Each distinct cell is given a unique ID. The cell descriptors include cropped phase and fluorescence images of the cell, the length and width of the cell, the major and minor principal axis, the area, the location of fluorescent foci in global and local coordinates, the IDs of daughters, sister, mother and neighbor cells, the frame of death and birth of the cell, etc (Fig. 2.3).

2.1.6 Analysis of fluorescence channels

SuperSegger includes a number of generic tools for the quantitation of cellular fluorescence, including the computation of statistics on fluorescence intensity and the identification of punctate foci. The mean background fluorescence intensity for each frame is subtracted from the fluorescence images. To identify and score foci, SuperSegger uses an image-curvature method which is specifically engineered to avoid the identification of false positive foci due to background intensity from cytoplasmic fluorescence (C. Bren-



```
death: 123 (frame)
birth: 2 (frame)
divide: 1 (boolean)
sisterID: 387
motherID: 140
daughterID: [3527 3528]
ID: 386
neighbors: [387]

CellA: {per frame}
    phase: [Cell Phase Image]
    mask: [Cell Mask]
    BB: [Bounding Box]
    r_offset: [Left corner of BB]
    coord: [Cell coordinates]
       A: [Cell Area]
       r_center: [Cell centroid]
       e1: [Long axis vector]
       e2: [Short axis vector]
       rcm:[Cell center of mass]
    length: [Long axis,
                   short axis]
    pole: [1x1 struct]
              op_age: [Old pole age]
              np_age: [New pole age]
    fluor1: [Fluorecence Image]
    fl1: [Florescence Statistics]
       sum: [Intensity sum]
       r: [Center of fluor]
       I: [2nd moment of fluor]
       bg: [background fluor]
    cell_dist:
            [Distance to colony edge]
    locus1: [Foci characteristics]
        r: [Global coordinates]
        score: [Foci quality]
        intensity: [Foci intensity]
        shortaxis: [Local coordinates]
        longaxis: [Local coordinates]
```

Figure 2.3: **SuperSegger outputs single cell files.** Each cell file under a unique ID contains various information regarding properties of the cell and fluorescence characteristics for each frame of the lifetime of the cell. A representative example is displayed above. The images displayed are later in the life of the cell, showing that the sister is identified even when the sister cells are not in contact pole to pole. The orange triangles point to the fields from where the phase, mask, and fluorescence images are found. Bold text shows fields displayed in the image and tabs indicate that they are subfields of the field above.

nan, unpublished). Foci detection is performed in the union of all cell regions and foci are then allocated to cells to circumvent the double counting of foci close to the membrane at the interface between cells [31, 32, 33, 34].

2.1.7 *Complete cell cycle imaging*

In order to study the dependence of fluorescence localization with the cell cycle phase or the new or old pole of the cell, it is important to identify complete cell cycles. SuperSegger identifies division and birth events. After a division is observed the program identifies the new poles, created at the last cell division, or old, created by an earlier division. During the identification of division events, the mother and daughter cells are also tracked. Starting at any frame, it is possible to track the lineage of a single cell both backwards and forwards in time and, for example, to investigate fluorescence localization through the generations.

2.1.8 *Data output*

Due to the variety of experimental demands and the large size of the typical data set, SuperSegger slices the data into three output formats: The frame files organize the data by frame and contain all the information about all cells in a single time point. The cell files organize the data by cell and contains all data for a single cell over all time points (Fig 2.3). Finally, the Clist file contains (i) a matrix of over ninety cellular descriptors (including cell age at division, cell length at birth and division, average fluorescence intensity, etc), organized by cell ID, and (ii) Clist-in-time structure with several cellular descriptors (cell length, average fluorescence intensity) for each time frame organized by cell ID.

2.1.9 *Gating functionality*

SuperSegger implements a powerful gating functionality, inspired by flow cytometry, which facilitates the selection of subpopulations of cells that match specific criteria de-

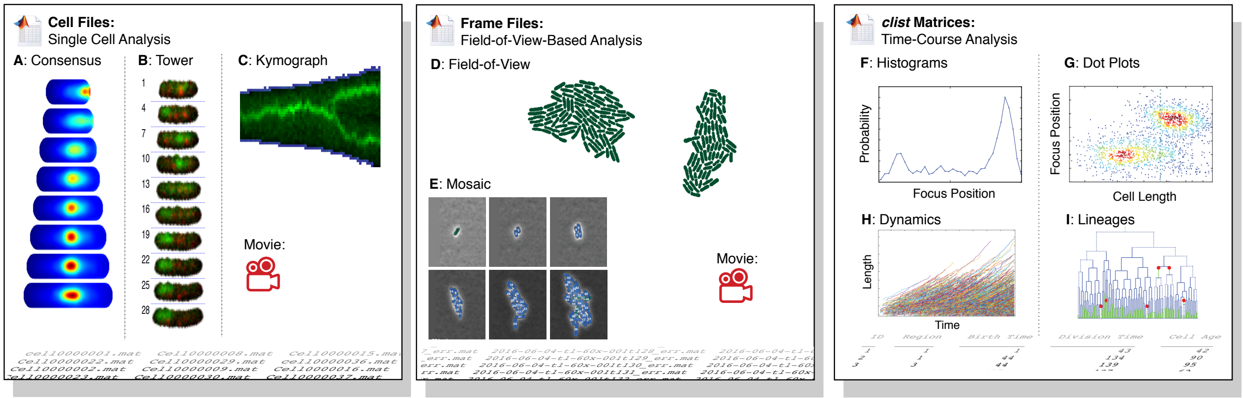


Figure 2.4: **SuperSegger output and analysis.** **Cell files** contain all time points for a single cell. Cell data can be visualized as: **Panel A:** Consensus images, the mean cell-cycle-dependent fluorescence localization pattern (example of MalI, a transcriptional repressor [1]), **Panel B:** Cell towers (example of mini-F plasmids labeled with mCherry-TetR and ParA-GFP (ATPase)[2]), **Panel C:** Kymographs (example shows of *parS* cassette inserted at *oriC* and labeled with GFP-ParB [3]) and cell movies. **Frame files** contain the data from a single field of view at a time point. Frame data can be visualized as: **Panel D:** Field-of-view images, **Panel E:** Frame mosaics, and movies. The **clist matrices** contain a summary data for all cells at all time points. Clist data can be visualized as: **Panel F:** histograms (example of probability density of position of mRNA-MS2 foci [4]), **Panel G:** dot plots (example of focus separation of replication forks versus cell length [5], **Panel H:** cell dynamics plots (example of cell length with time [4]), and **Panel I:** cell lineages.

scribed using the Clist cellular descriptors. Gating significantly streamlines many common analyses. For instance, it is straightforward to limit analysis to cells which are (i) segmented without error, (ii) observed through a complete cell cycle, and (iii) have an average fluorescence intensity above a specified cutoff. For more information about this powerful approach to quantitative cell biology the interested reader can look at [35].

2.1.10 Post processing tools

SuperSegger includes a post-processing GUI to visualize the segmentation, fluorescence localization, focus position and tools to interactively modify incorrect cell boundaries. Further, it provides a variety of analysis tools at the single cell level as well as at the population level.

At the single cell level, the user can produce through the post-processing GUI a rotated and masked fluorescence profile for a cell for each frame during its lifetime, which we refer to as cell tower (Fig. 2.4B). The fluorescence profile along the long axis of the cell can be displayed in a kymograph (Fig. 2.4C) which is the summed fluorescence along the short axis of the cell with respect to time. A movie for each cell can also be made.

At the population level, our post-processing GUI produces histograms for any of the features found in the Clist table (Fig. 2.4F), dot-plots of the Clist descriptors against each other (Fig. 2.4G) and plots for the descriptors in the Clist-in-time for all cells versus the time frame (Fig. 2.4H). In addition, by averaging over multiple single cell files, a quantity of interest for the population can be visualized, such as the mean cell-cycle-dependent fluorescence localization pattern, which we refer to as consensus image (Fig. 2.4A). The consensus images are produced by computationally interpolating each single cell onto a reference cell cycle with regular shape, growth rate, and lifetime followed by averaging the fluorescence profile [1].

2.2 *Results*

2.2.1 *High-throughput analysis*

Our software is capable of segmenting high-throughput data sets with multiple time frames and xy and z positions. Furthermore, it takes advantage of parallel computation for each separate xy frame and time frame, decreasing the processing time. Data sets taken at high frame rates can expedite the segmentation process by taking advantage of the option to segment images every k frames. In this case, the skipped frames receive the boundaries of the cells from the last frame segmented in the sequence.

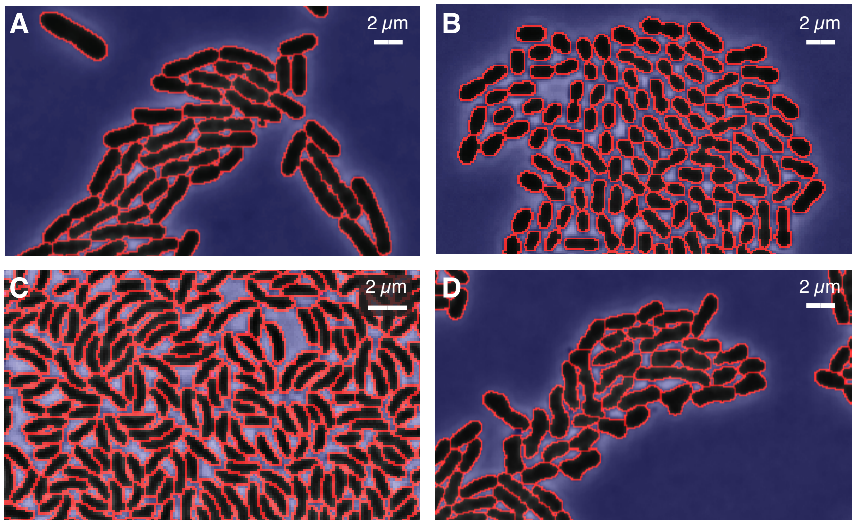


Figure 2.5: **Segmented phase images of bacteria.** In red are the computer generated boundaries. SuperSegger can successfully segment images of differently shaped cells and cells in close contact such as **Panel A:** *E. coli*, **Panel B:** *A. baylyi*, **Panel C:** *C. crescentus*, and **Panel D:** irregularly shaped *E. coli* cells.

2.2.2 *Reliable and flexible segmentation*

The segmentation parameters can be optimized on different cell shapes making SuperSegger highly flexible. We have trained parameters on *E. coli*, *P. aeruginosa*, *Caulobacter crescentus* and *Acinetobacter baylyi*. In Fig. 2.5, we show examples of segmentation of *E. coli* (Fig. 2.5A), *A. baylyi* (Fig. 2.5B), and *C. crescentus* (Fig. 2.5C) and *E. coli* with abnormal cell morphology (Fig. 2.5D). The user can interactively produce a training set of identified true and false boundaries and generate new segmentation parameters. Although SuperSegger includes an interactive tool for boundary correction after segmentation, we have not needed this feature since errors are rare enough that cells with segmentation errors can be discarded. Furthermore, the software is modular, facilitating the use of the segmentation (or other) modules in conjunction with other code.

2.2.3 *Generalization*

We note that one of the strengths of SuperSegger is that it does not require users to fine-tune the segmentation parameters from application to application and does not require retraining. In our experience the segmentation parameters generalize well between cell types and experimental conditions. For instance, we use parameters trained on *E. coli* to segment *Bacillus subtilis*. The benchmarks we report are measurements of generalization error, not training error. The segmentation parameters were trained against a different strain of *E. coli*, grown under different conditions. No data set specific optimization of the parameters was required to perform the segmentation. Other packages often require either setting specific parameters for each data set or the user to correct the boundaries by hand after segmentation.

2.2.4 *Pixel-based segmentation*

It is important to note that SuperSegger uses a pixel-based definition of the cell region, and that a number of competing package generate sub-pixel resolution definitions of the cell region. We have avoided this practice for two reasons: (i) Pixel-based definitions of the regions are convenient and make the modules of SuperSegger easier to integrate with other image analyses in our lab. (ii) A second motivation is the difficulty of generalizing these method to cells with many neighbors. A quantitative interpretation of phase contrast intensity is complicated by the two well known phase contrast image artifacts: phase halos and shade-off. When cells form micro-colonies, contrast is lost towards the center of the colony and natural cytoplasmic structures, like the nucleoid, tend to be more visibly pronounced. We therefore do not believe that the cell boundaries can be reliably segmented to higher than pixel resolution when the cell density is high.

2.2.5 Comparison with existing software

We compared the performance of SuperSegger with existing software packages (Fig. 2.6). All packages we tested could successfully segment cells that were well separated or formed small colonies of a couple cells, but since many of our applications demand full cell cycle imaging over multiple generations, it is necessary to reliably segment micro-colonies containing many cells. Segmenting cells in a micro-colony is particularly difficult because contrast is lost towards the center of the micro-colony. In our tests of performance, SuperSegger outperformed the other software packages we tried [27, 20] in segmenting these micro-colonies. Other packages either incorrectly merged or split multiple cells in the interior of the micro-colony, even in a single frame (Fig. 2.6). We note that lineage tracking requires the linking of hundreds of error-free segmented frames.

2.2.6 Data acquisition

We have advocated strongly for the strengths of the segmentation package, but it is important to note that SuperSegger does routinely generate segmentation errors. In general, SuperSegger is unable to correctly segment images where the cell outlines are not clear to the user by eye. Care should still be taken in collecting the best possible focused phase

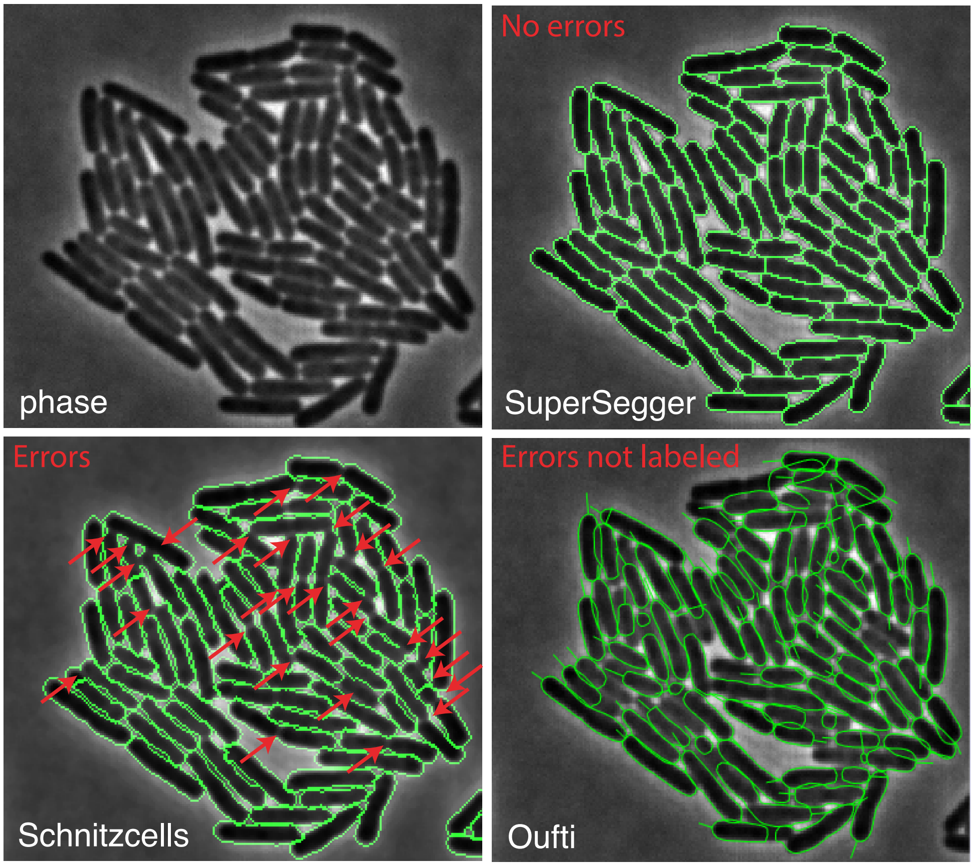


Figure 2.6: **Performance test of SuperSegger, Schnitzcells and Oufti segmenting an *E. coli* micro-colony**. For Schnitzcells and SuperSegger, we used the provided *E. coli* segmentation parameters (not trained on the data set), and for Oufti we started with the provided *E. coli* LB parameters and updated to the ones that looked best. SuperSegger outperforms the other two software packages in reliably segmenting cells in micro-colonies. Authors report Oufti performance increases for confluent cells growing in a micro-fluidic chamber.

images. We recommend that users crop out-of-focus regions of the image before the segmentation process since these parts of the image are unlikely to yield usable data.

In practice we find that it is not essential to manually correct segmentation errors in a field of view. It is much more practical to accept that some segmentation errors do occur and use the gating functionality of the software to ensure that analysis is only performed on cells that are correctly segmented. It is often sufficient to limit the analysis to cells that were tracked for a complete cell cycle without error. Although gating for cell lengths at birth and division, cell area, and cell region scores can all be used to increase the number of cells with segmentation errors excluded in the analysis.

2.3 Benchmarks

2.3.1 Snapshot benchmark

A standard metric for the performance of a segmentation algorithm is looking at the segmentation error rate. Our software performs exceptionally well in this metric. In a snapshot image of 824 *E. coli* cells in micro-colonies of about four cells, 99.4% of the 1198 boundaries were correctly classified resulting to 99.3% of the cells to be correctly segmented (Fig. 2.7). We believe that the measurement of error rate in a snapshot image is a misleading metric. The error rate in snapshot analysis depends sensitively on cell density since snapshot images of isolated bacterial cells can be segmented without error.

2.3.2 Time-lapse benchmark

What makes the bacterial cell segmentation problem challenging are the cell contacts and micro-colonies. In micro-colonies, cells at the boundary with excellent contrast and cells in the interior of the micro-colony with poorer contrast due to the shade-off artifact must be segmented simultaneously. Therefore a natural test of the algorithm is the analysis of cell proliferation on an agarose pad starting with a single isolated cell. (See Fig. 2.8.) Although the culture starts with isolated cells, a micro-colony forms and grows for roughly

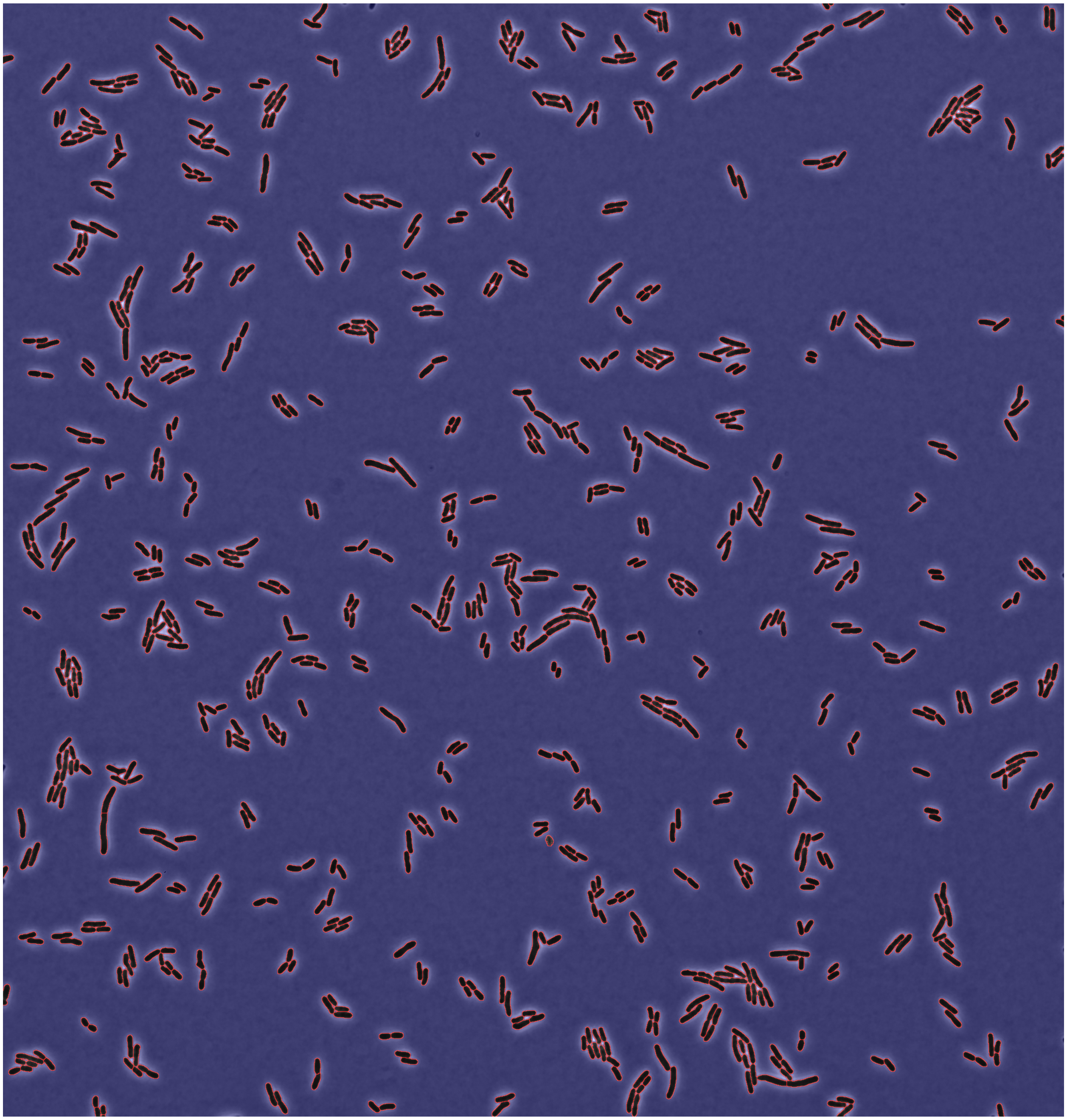


Figure 2.7: **Snapshot benchmark.** Segmentation of a snapshot image of of 824 *E. coli* cells in micro-colonies of about four cells. 99.4% of the 1198 boundaries were correctly classified resulting to 99.3% of the cells to be correctly segmented. Six cells were incorrectly divided into two cell regions. Such errors are often fixed during the linking phase from frame-to-frame.

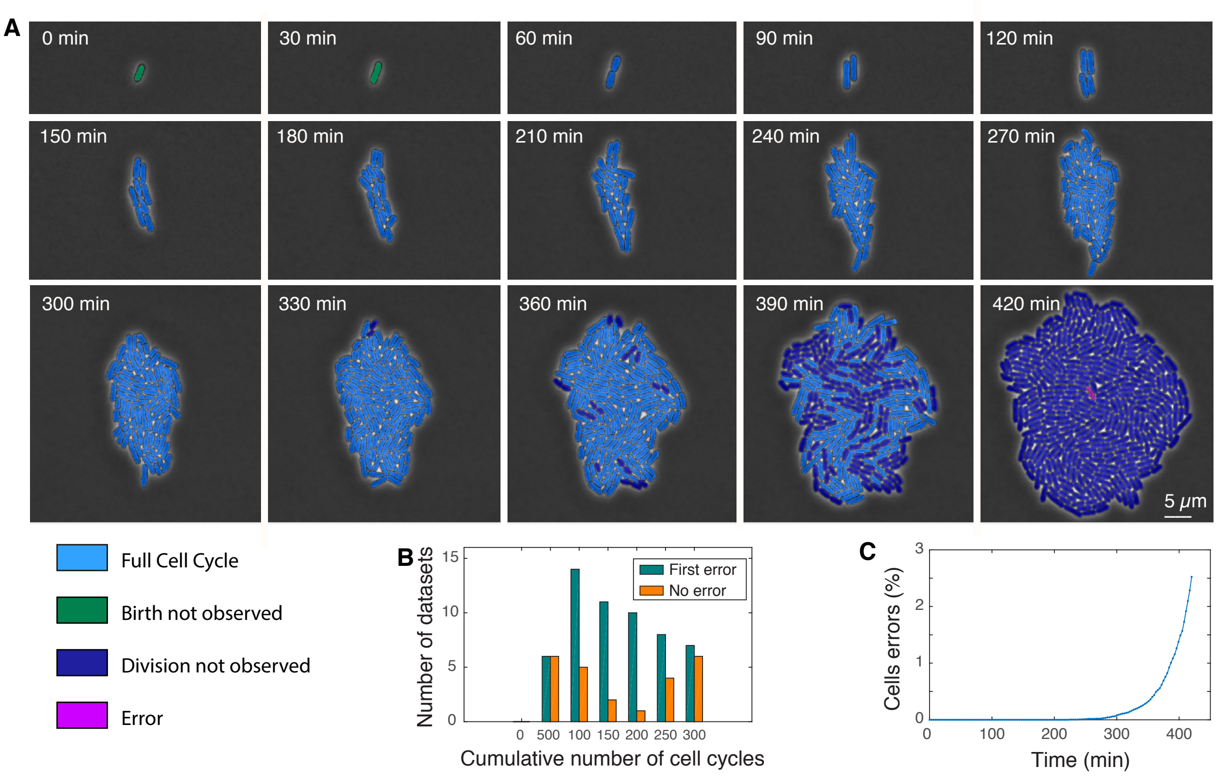


Figure 2.8: **Segmentation of proliferation from single *E. coli* cells. Panel A:** Frame mosaic of segmented images of a micro-colony. A cell is marked with an error at the end of the time-lapse. (Only 15 of 141 frames are shown.) **Panel B:** A histogram of the cumulative number of cell cycles observed before the first error or no error occurred at each data set. **Panel C:** The percentage of segmented cells with errors with time.

seven rounds of division before the culture begins to become multi-layered. SuperSegger cannot segment cells in multiple-layer micro-colonies and once cells begin to grown on top of one another, overlapping cells lead to errors. But in many applications, these cells can be excluded from the analysis, while cells at the boundary of the colony (proliferating in a single layers) can be segmented, tracked and analyzed without error.

To test the performance of SuperSegger in segmenting micro-colonies we analyzed 80 initially isolated cells grown on M9 agarose pads for 420 minutes. In each data set the software marked the first segmentation error, defined as a frame-to-frame substantial area change or loss of a cell lineage. The results are shown in Fig. 2.8.

In Fig. 2.8A, we show an example of a segmented data set with 363 cells in the final frame and 727 cells segmented cumulatively. (Cells refer to the number of actual cells observed during the experiment, not regions segmented in a single frame.) In this data set 543 cells were segmented before the first error occurred and, in the final frame, only one cell in the middle of the colony was segmented with an error. To understand why the segmentation problem is challenging, consider that in order to correctly segment this micro-colony, 30,179 boundaries needed to be classified during the time course.

In these 80 data sets, the mean number of cells segmented before an error was 174 cells, but 30% of the data sets were segmented without error (Fig. 2.8B). One of the main mechanisms resulting in segmentation errors is the overlap of cells that occurs when the colony begins to become multi-layered. In this experiment cells began to overlap at about 400 minutes and a corresponding rise in the error rate is clearly seen in Fig. 2.8C. Even with the increase, the error rate remains low: At the end of the experiment only 2.5% of the cells have segmentation errors.

2.3.3 *Runtime benchmark*

SuperSegger takes advantage of parallel processing: (i) for one *xy* position it allocates the segmentation of each time frames to the available processors, (ii) for multiple *xy* positions each *xy* position from segmentation to output runs on each available processor. SuperSegger's parallelization dramatically improves the processing time of multiple *xy* positions but does not scale with image size, and varies with cell density and number of computer processors. Nevertheless, SuperSegger is very efficient. As an example, aligning, segmenting, linking, characterizing and producing all output files for the time-lapse in Fig. 2.8 (141 frames, from a single cell to a micro-colony) took 13.6 minutes in non-parallel mode, and 9.6 minutes with 4 processors in parallel. Segmentation alone of the single cell in the first image took 0.27 seconds and the micro-colony in the last image took 10 seconds. We copied the same dataset in 4 *xy* positions and processed it in parallel and non parallel. In parallel with 4 processors all *xy* positions were fully processed in 21.2 minutes and in non-parallel in 51.5 minutes. Just segmenting the full frame image shown in Fig. 2.7 took 35 seconds.

2.4 Conclusion

SuperSegger is a powerful automated image processing and analysis package, well suited for high-throughput time-lapse fluorescence microscopy of cells [3, 36, 37, 38, 1, 39, 40, 31, 32, 41, 34, 4]. It provides reliable and flexible segmentation which can be trained to optimize performance in a wide variety of experimental contexts. The multiple data structures output by SuperSegger facilitate its use in many experimental imaging contexts. The faithful segmentation and tracking of bacterial cells with neighbors facilitates its use of quantitative imaging-based analysis in new problems, like bacterial cell-contact mediated phenomena [31, 32, 33, 34], where cell contacts are essential and applications where investigators want to study cell proliferation over multiple generations.

Chapter 3

CASE STUDY I : SINGLE-CELL ANALYSIS OF LAG PHASE

To date, we have successfully applied our package to a number of problems in the cell biology of bacterial cells, including a genome-scale characterization of protein localization dynamics in *E. coli* throughout the cell cycle [1, 38], detailed analysis of the function of the Type VI Secretion System in *Pseudomonas aeruginosa* [31, 32, 34], chromosome segregation in *E. coli* [3, 39] as well as a number of other applications [41, 4]. In this chapter, we demonstrate the power of SuperSegger for quantitative analysis by analyzing cell proliferation during lag phase in single bacterial cells, the period during which bacteria adapt to changes in growth conditions.

In addition to a good case study of analyzing many generations with SuperSegger, lag phase growth is not well studied in the single cell level [42, 43, 44, 45]. The nature of lag phase has been mainly characterized and analyzed at a population level (*e.g.* [46]). In particular, it is unclear whether the observed delay associated with lag phase is the result of slow growth of all cells in the population or a graded response with range of growth rates at the cellular level. The observation of significant cell-to-cell variation could be evidence of a *bet-hedging* strategy, the strategy of lowering the mean fitness to reduce variation in fitness [47], since the aberrant transition of cells from a dormant to a growing state would reduce cell fitness in a fluctuating environment [48, 45, 49].

3.1 Methods

3.1.1 Bacterial Strains

The *E. coli* strain used in this study for the cell proliferation experiment is the sequenced, wild-type strain K12 MG1655. The strain was obtained from CGSC, The Coli Genetic

Stock Center.

3.1.2 Growth conditions

Log phase cultures were grown overnight from a single colony in LB media. The overnight cultures were then re-inoculated (1:1000 dilution) into fresh media and grown to an OD_{600} of 0.1.

Stationary/Death phase cultures were grown overnight from a single colony in LB media. The culture was then left on the bench for 7 days at room temperature. Immediately before imaging, the culture was diluted to OD_{600} of 0.1 in fresh media so that the cell count was approximately the same as the log phase culture.

3.1.3 Sample preparation

Agarose pads were prepared by melting 1 mL of M9 media (1×M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% glycerol carbon source and 0.2% casamino acids supplement) and 4% low-melt agarose into 2 cm × 2 cm wells cut into a rubber gasket sealed onto a microscope slide (3″×1″) using VALP (1:1:1 vaseline, lanolin, paraffin). A microscope slide was then placed on top of the gasket to form a mold. The agarose gel pad was then allowed to set for 15 min. The top slide was then removed and the pad was allowed to dry for 13 min. 2 μL of the cell culture was spotted onto a coverslip, which was placed on top of the pad. The entire slide was sealed with VALP.

3.1.4 Microscopy

Imaging was performed using a Nikon Eclipse Ti-E microscope, through a 60× 1.4 NA CFI Phase objective onto an Andor Neo sCMOS camera. The microscope and sample were heated to 30°C using an environmental chamber. Images were taken every minute for a total duration of 300 minutes.

3.2 Results and Discussion

To investigate the nature of lag phase with single cell resolution, we analyzed cell proliferation upon re-inoculation on an agarose pad. We characterized single progenitor cells in both log and late-stationary/death phase growing into micro-colonies of hundreds of cells. The analysis is shown in Figure 3.1. Fig. 3.1A-C show the proliferation of a single cell over five hours to a final population of 115 cells. The micro-colony was segmented without error. Fig. 3.1A shows the lineage tree for the micro-colony and Fig. 3.1B a frame mosaic of the colony. Fig. 3.1C shows the growth curve of the micro-colony. Divisions early in growth are closely synchronized, resulting in a characteristic step-like increase in the number of cells. As subsequent divisions lead to a de-synchronization, steps later in growth are smoothed. In the limit of large time, the growth curve approaches a straight line (on a semi-log plot), consistent with the exponential growth of an un-synchronized population.

The micro-colony growth curves for log ($N = 82$) and stationary phase ($N = 119$) progenitor cells are shown in Fig. 3.1D. Over 70,000 micro-colony images were segmented in the analysis. The log (blue) and stationary phase (yellow) cells both proliferate with nearly identical growth rates, as is clear from the slopes of the growth curves. The principle difference between the curves is a delay in lag phase for stationary phase cells, as is expected from population-level analysis of lag phase after stationary phase [15]. For a number of stationary phase cells the first division event $(1 \rightarrow 2)$ is observed more than 120 minutes after the cells were inoculated.

To quantify this delay and analyze the persistence of slow growth after the first division, we plot histograms for the duration of the cell cycle for the first five generations of growth (Fig. 3.1E). The 0^* generation represents the initial incomplete cell cycle of the progenitor cell. Since the cells do not begin the experiment synchronized, the 0^* duration is not directly comparable to the duration of subsequent complete cell cycles. As expected, in generation 0^* the log-phase progenitor cells execute an incomplete cell cycle in a period shorter than later complete cell cycles (30 minutes). In contrast, the stationary-phase progenitor cells have a much longer duration 0^* generation time. About 80% of the cells divide in the first 200 minutes and show a wide distribution of cell cycle durations, centered around an hour. Nearly 20% of cells do not divide in the first 200 minutes. It is important to note that not all these cells are dead: 13% of these cells divide between 200 minutes and 300 minutes (when the time course ends).

After the 0^* generation, the first full cell cycle observed, generation 1, shows a slight shift relative to subsequent cell cycles. (Log phase cells also show a small lag phase due to changes in growth conditions on the agarose pad.) But our experiments demonstrate that this effect is small and slow growth does not appear to persist.

These observations suggest a qualitative model for stationary and death phase with three cell states: (i) stationary, (ii) stationary-dormant and (iii) dead cells. The stationary-dormant cells are those that we have identified which have a significant delay before the first cell division. It is possible that this state is an intermediate state between stationary phase and cell death. Alternatively, this state may be the result of a bet-hedging strategy where a small subset of the population of cells does not transition into the growing state. This may insure against short-lived increases in nutrients which might lead cells to transition into the growing state when the optimal behavior is to remain dormant [48].

3.3 Conclusions

In this chapter, we demonstrate image segmentation and analysis of up to five generations of cells. The order of magnitude increase in the accuracy of segmentation using SuperSegger facilitates the analysis of micro-colonies with single cell resolution and reveals new subtleties in lag phase growth phenomenology. Our experiments show a wide distribution in the delay before the first division with 20% of cells exhibiting an extremely slow transition to the growing state ($\tau > 200$ min). As noted above, not all these cells are dead and some transition after 200 minutes, but before the end of the experiment at 300 minutes. We have suggested a model for stationary and death phase of three states : (i) stationary, (ii) stationary-dormant and (iii) dead cells, where stationary-dormant may be an intermediate state or the result of bet-hedging.

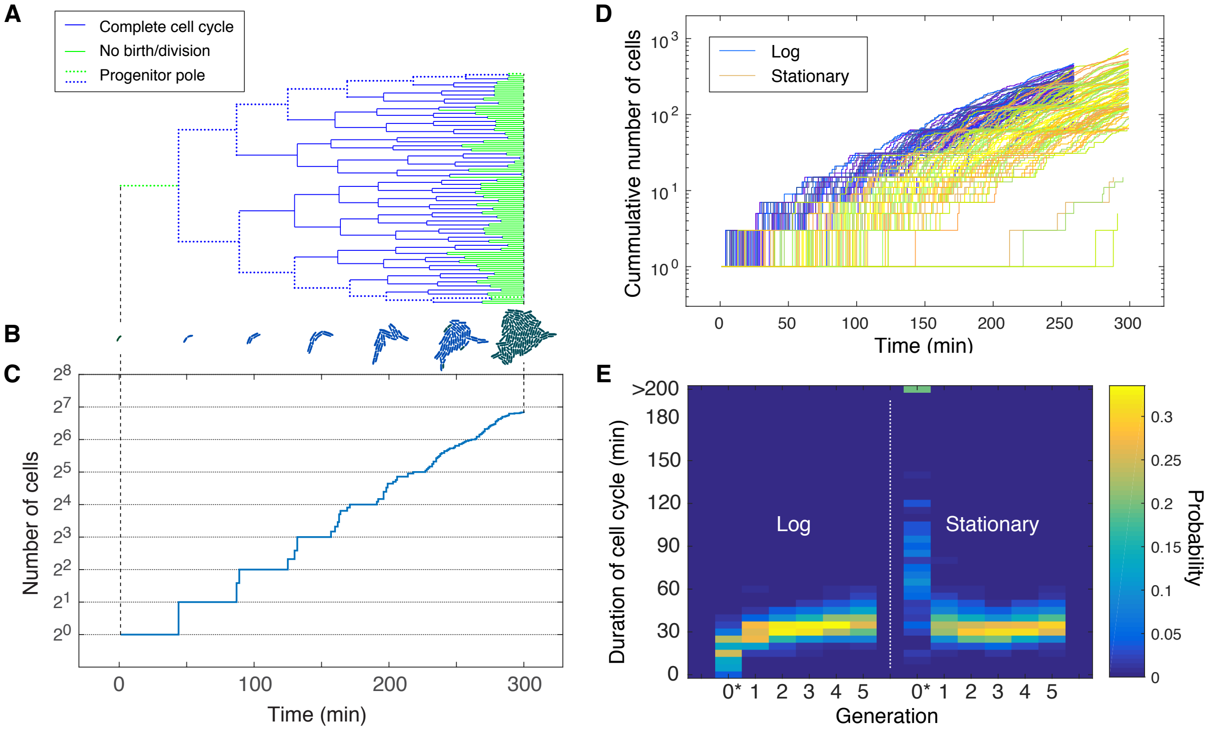


Figure 3.1: **Lag phase results. Panel A:** Cell lineage tree for a progenitor single cell. **Panel B:** Frame mosaic of micro-colony. Blue cells are observed for a complete cell cycle. **Panel C:** Growth curve for a single micro-colony. **Panel D:** Growth curves for progenitor cells in either log (blue, $N = 82$) or stationary phase (yellow, $N = 119$). **Panel E:** Duration of cell cycle by cell generation of growth on the agarose pad. The cell generation is shown on the x-axis for the two populations (log and stationary phase). The 0^* generation is defined as the incomplete cell cycle of the initial progenitor cell.

Chapter 4

CASE STUDY II : CYTOPLASMIC DYNAMICS OF MOLECULAR COMPLEXES

In the previous chapter, we tested our software in segmenting phase images of single cells growing into multiple generations of micro-colonies. Here we demonstrate its capabilities in identifying and tracking fluorescence foci, in addition to full cell cycle segmentation, in order to analyze the motion of molecular complexes. Through this analysis, we also seek to better understand the organizing principles of the bacterial cell, in particular the detailed mechanisms that drive cellular-scale spatial and temporal organization.

Specifically, we are interested in the hypothesis proposed that the nucleoid may play a central role as an organization template in the bacterial cell [50]. It has long been known that aggregates of misfolded or non-functional proteins tend to localize at the cell poles, a process that has been implicated in cellular aging effects [51, 52]. But it is not entirely clear whether the complexes move under the influence of biased forces throughout the cytoplasm or undergo unbiased Brownian motion until they become trapped at the poles. Modeling studies have suggested that membrane curvature, nucleoid exclusion, and over-expression are sufficient to cause polar localization of protein complexes [53, 54, 55, 56, 57] but a comprehensive quantitative experimental investigation has yet to confirm this hypothesis. Recent studies have also reported that the nucleoid plays a significant role in gene expression: mRNA-bound ribosomal subunits are excluded from the nucleoid while free ribosomal units are not, allowing for easy access to nascent mRNA transcripts and tight coupling between transcription and translation [58]. Furthermore, it has also been proposed that entropic effects from the nucleoid may be an essential component in cellular-scale chromosome organization, in particular in the poorly understood chromosome segregation process in *E. coli* [59, 60].

To explicitly investigate the hypothesis that forces resulting from the physical exclusion of macromolecules from the bacterial nucleoid play a central role in organizing the bacterial cell, we set out to quantitatively map the dynamics of molecular complexes in the cytoplasm of growing *E. coli* cells. Using our image and analysis tool, we track thousands of complete cell-cycle trajectories of fluorescently tagged ectopic MS2-mRNA complexes. We make use of an existing GFP-labeled MS2-mRNA system [6] as a tracer to quantitatively map the cellular-scale, spatially-dependent dynamics of the cytoplasm in thousands of complete cell cycles. The complexes are approximately twice the molecular weight of the *E. coli* ribosome and the MS2-bound messages are neither translated nor degradable, thus facilitating complete cell cycle imaging [6]. Since this system is ectopic to *E. coli*, it is believed that it is subject only to non-specific interactions that would be general to all large complexes, including blobs of condensed chromosomal DNA, plasmids, ribosomes, and large protein aggregates.

Our measurements recapitulate the results of previous reports describing a polar-directed drift of molecular complexes and provide a new spatial map of this drift throughout the cell [61, 17]. Quantitative analysis of this drift closely matches the predictions of a minimal model composed of two basic physical mechanisms: nucleoid exclusion and membrane confinement. To confirm our model, we made perturbations to nucleoid structure, cellular shape, and cell physiology, and found results that reveal an unexpectedly central role for the nucleoid in determining the mobility of protein complexes in the cytoplasm. Although extensive studies have previously been performed on characterizing the dynamics of bacterial cytoplasm [6, 62, 63], our study is the first to quantitatively characterize the spatial dependence of the dynamics throughout the entire cell cycle, which reveals that the mobility of protein complexes is significantly larger over the nucleoid region and tends to increase when nucleoid structure is perturbed. These observations may explain many of the phenomena previously attributed to more complicated physical mechanisms.

4.1 Methods

4.1.1 Bacterial Strains

The *E. coli* strain used in this study (a gift from I. Golding) is DH5α carrying the plasmids pIG-K133 and pIG-BAC2. The MS2-mRNA system consists of two components: a green fluorescent protein that is fused to the MS2 bacteriophage coat protein and a reporter RNA with a repeated sequence of MS2-binding sites (Fig. 4.1A). The low copy plasmid pIG-BAC2 is responsible for encoding the mRNA molecule with a 96-tandem repeat of MS2-binding sites under the pLac promoter. The pIG-K133 is responsible for the GFP-MS2 protein fusion under the tetracycline promoter (Fig. 4.1B). Note that in the absence of pIG-BAC2, no aggregation of GFP-MS2 is observed [6]. A detailed description of this system can be found in [6].

4.1.2 Growth conditions and microscopy

All strains were grown overnight at 30°C from a single colony in LB media with the appropriate antibiotics (Kanamycin and Chloramphenicol). Doubling time in liquid media was approximately 30 minutes. Overnights were diluted 1:1000 the following day into fresh media and grown to mid-logarithmic phase at 30°C ($OD_{600} \sim 0.2$). The cells were then induced with IPTG ($1\,\mathrm{mM}$) and aTc ($10\,\mathrm{ng/ml}$) for 15 minutes at 30°C. Cells were rinsed and grown in fresh media without IPTG and aTc for 1 hour at 30°C. Agarose pads were prepared by pouring $1\,\mathrm{mL}$ of growth media with IPTG ($0.25\,\mathrm{mM}$) and 2% agarose into $2\,\mathrm{cm} \times 2\,\mathrm{cm}$ wells cut into a rubber gasket sealed onto a standard microscope slide. $2\,\mu\mathrm{L}$ of cells were spotted onto dried pads and a coverslip was placed on top of the pad. The entire slide was sealed with VALP (1:1:1 vaseline, lanolin, paraffin). Under imaging conditions, the average doubling time for cells was 46 minutes.

For biological perturbations: Rifampicin treatment: agarose pads were treated with $100\,\mu\mathrm{g/ml}$ Rif and cells where inoculated on the pad for 30 minutes before imaging. A22: The cells were grown in LB and $10\,\mu\mathrm{g/ml}$ A22 for two hours and then spotted on agarose pad

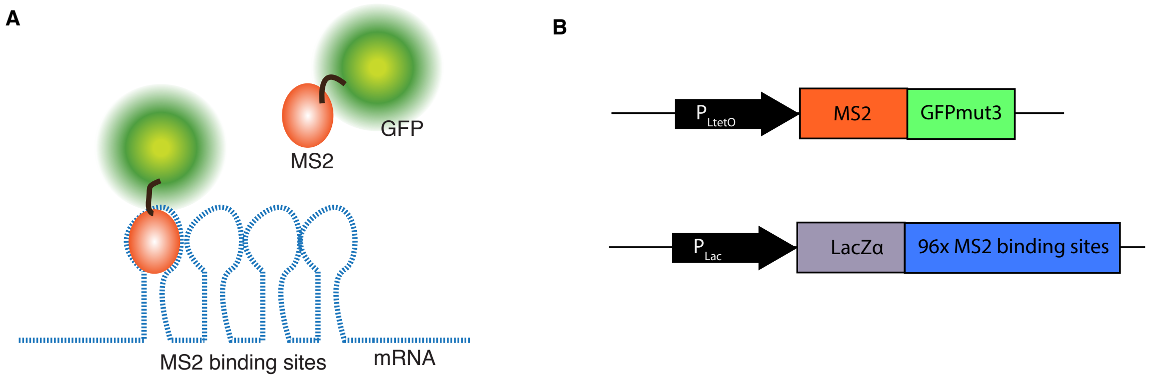


Figure 4.1: **Schematic of MS2-mRNA. Panel A:** MS2-mRNA consists of an mRNA molecule with a 96-tandem repeat of binding sites for the RNA-binding protein MS2 and a GFP-MS2 protein fusion. **Panel B:** Schematic of the two plasmids adapted from [6]. Under the control of the promoter P(LtetO-1) there is the MS2 protein fused to the GFP variant, GFPmut3, fusion. Under the P(Lac) promoter there is the tandem array of 96 MS2-binding sites (96x MS2-bs).

with the same concentration of A22. Novobiocin: Cells were spotted on agarose pads with $1000\,\mu\mathrm{g/ml}$ of novobiocin and inoculated on the pad for 30 minutes before imaging. Sodium azide: Cells were grown in Sodium Azide (0.01%) and 2-Deoxy-D-glucose ($1\,\mathrm{mM}$) for 30 minutes before being spotted on a pad with the same concentrations. UV Radiation: Cells on the pad were exposed to high intensity UV light ($\lambda = 254\,\mathrm{nm}$) for 2 seconds and then left to grow for an hour before imaging.

For the visualization of the chromosome, cells were fixed in $1\,\mathrm{ml}$ of 70% Ethanol and 30% PBS for 3 minutes, rinsed twice in PBS, and then resuspended in $1\,\mu\mathrm{g/ml}$ of DAPI.

Time-lapse phase-contrast and wide-field fluorescence microscopy images were collected at 1 minute time intervals using a large-format sCMOS camera (Andor Neo; Andor Technology/Oxford Instruments, Belfast, Northern Ireland) attached to a TiE microscope (Nikon Instruments, Melville, NY) with a 60$\times$ Plan-Apo oil immersion objective with a

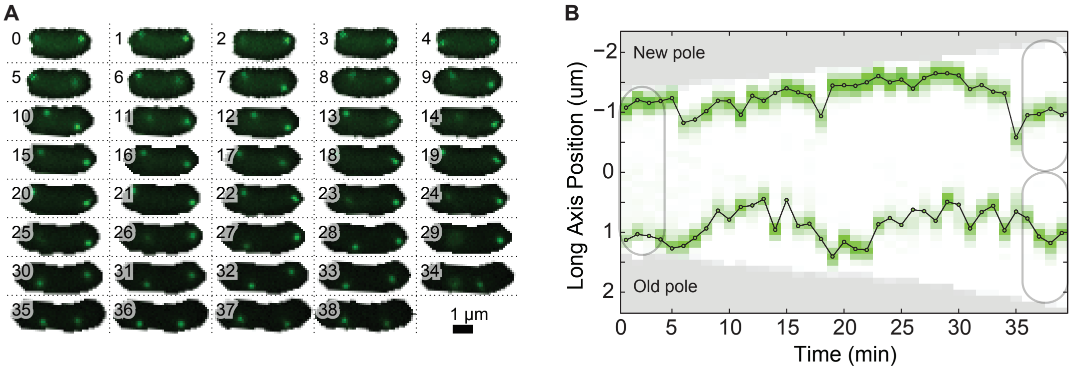


Figure 4.2: **Example of cell with MS2-mRNA complexes. Panel A:** Wide-field fluorescence images showing the dynamics of MS2-mRNA complexes in a typical cell throughout the entire cell cycle. Frame numbers are measured in minutes. **Panel B:** Kymograph of MS2-mRNA cell-cycle trajectory plotted along the long axis of the cell shows dynamic localization of protein complexes throughout the cell cycle.

1.4 numerical aperture and equipped with an environmental chamber and controlled by NIS-Elements (Nikon Instruments).

4.1.3 Image Analysis

Cells were identified and linked from frame to frame using SuperSegger. The MS2-GFP foci were tracked in each frame and linked to form trajectories. For analysis we included only cells in which both birth and division were observed, precluding filamentous and non-growing cells.

4.2 Results

4.2.1 Complete cell cycle imaging of MS2-mRNA dynamics.

To quantitatively characterize MS2-mRNA dynamics throughout the cell cycle, we captured thousands of complete cell cycles: cells in which both birth and division are explicitly observed. We analyzed 6655 complete cell cycles with at least one MS2-mRNA trajectory (Fig 4.2). For image analysis, the coordinate system in cells was oriented along the major and minor axes of the cell with the origin at cell-center. Because we combine trajectories from cells with significant cell-to-cell variation in length, it is convenient to express the long-axis positions as a fraction of total cell length, which we refer to as relative position. Finally, since we capture the entire cell cycle, we can also orient our coordinate system by the old and new cell pole (the cell pole that originated from the previous division event), where we define the center of the cell as 0, the old pole as 0.5, and the new pole as -0.5 in relative cell-length units.

It has been previously noted that MS2-mRNA complexes tend to be produced at the 1/4 or 3/4 positions of the cell and then migrate to the poles asymmetrically, with a strong bias towards the old pole of the cell [64]. We also observe that most of the MS2-mRNA molecules (76%) are indeed located less than 0.2 cell length away from the poles during the entire cell cycle, with a majority of those (70%) located near the old cell pole. We also find a smaller amount (10%) of MS2-mRNA within ± 0.1 cell length from the center of the cell, with the occupancy distribution shifting from a single peak early in the cell cycle into two centrally located peaks late in the cell cycle prior to division (Fig. 4.3).

4.2.2 MS2-mRNA motion is dominated by cell growth.

A fundamental unanswered question related to intracellular motion in growing cells is whether cytoplasmic factors tend to maintain their absolute position in the cell (in relation to some cellular landmark) or their relative position as a function of cell length, e.g. quarter-cell. We tested the two models: Complexes execute stochastic motion about their (1) absolute or (2) relative position in the cell. In the absolute motion model, we measure the positions and displacements of the MS2-mRNA complexes at all times from the center of the cell. In the relative velocity model, factors that maintain their relative position in the cell from frame-to-frame, e.g stay at quarter-cell, are defined to have zero displacement.

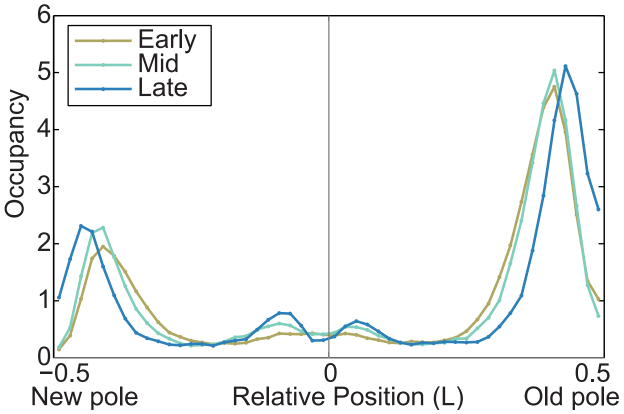


Figure 4.3: **Normalized histogram of the position of MS2-mRNA complexes in relative length units (N=282,860)**. Due to cell-to-cell variation in doubling time, cell cycles were partitioned into thirds: early, mid, and late. 76% of the molecular complexes are localized near the poles of the cell.

To calculate this relative velocity, we subtract from the absolute velocity a correction term proportional to the rate of cell growth

$$v(t) = \tilde{v}(t) - \frac{\tilde{x}(t)}{L(t)}\frac{dL}{dt} \tag{4.1}$$

where $\tilde{v}(t)$ is the absolute velocity, $\tilde{x}(t)$ is the initial position of the displacement vector, and L(t) is the length of the cell. To compare the two models, we compute the ensemble-averaged Mean Squared Displacement (MSD) for the two models of motion. The MSD is defined in Eq. 1.1. The $\Delta x(\delta t)$, the long-axis displacement over time δt is calculated in either the absolute or relative coordinates, and the averaging is performed over distinct trajectories. The model that minimizes the MSD as a function of time minimizes the bias in the motion and is therefore the best model for the motion. The two models are expected to be equal on short times but divergent on time scales over which the bias in the motion is significant in comparison to stochastic motion. (See the appendix for a more detailed explanation.)

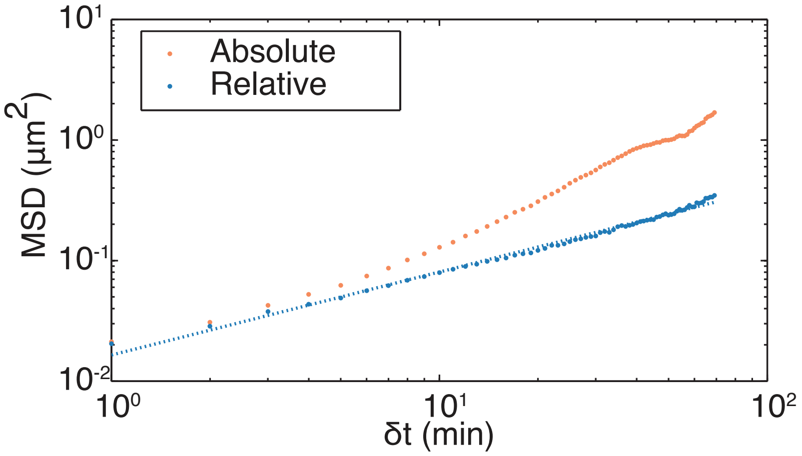


Figure 4.4: **MSD analysis of MS2-mRNA motion.** The MSD is computed using two velocity models: absolute (orange) and relative (blue). The relative model MSD has a constant slope with a scaling exponent $\alpha = 0.66$, consistent with previous measurements, while the absolute model increases in slope due to cell growth at times greater than 10 minutes (N=8053 trajectories).

The MSD plots for the two velocity models are shown in 4.4 for observation times up to 70 minutes, twenty minutes longer than previously reported for this system. As expected, the MSD measurements are consistent at short times for both models, but at times greater than 10 minutes the slope of the absolute velocity model significantly increases, and consequently the MSD for the absolute model is nearly an order of magnitude greater than the relative velocity model for times on the order of the cell cycle. If we were to use absolute velocities in the analysis of cell-cycle length trajectories, our results would be completely dominated by the apparent motion of complexes moving to maintain their relative position in the cell. Therefore, in all subsequent analysis we will use the relative velocity model.

For the relative velocity model, the MSD is well fit at all times by the power-law in Eq. 1.1 with a scaling exponent $\alpha = 0.66 \pm 0.01$. This value of the scaling exponent has been previously reported for shorter time scales using the absolute velocity model. Our observations remain consistent with these previous measurements since the discrepancy in MSD scaling between corrected and uncorrected models is only observable at timescales much longer than previous experiments explored. The observed MSD scaling exponent leads to a broad class of dynamics called subdiffusion ($\alpha < 1$).

We also observe that the motion of the molecular complexes is ergodic, with a similar power law scaling exponent for both the ensemble-averaged MSD calculated above and the time-averaged MSD (tMSD, $\left\langle \overline{[x(t+\delta t) - x(t)]^2} \right\rangle$), where the overbar denotes a time average over each trajectory before taking the ensemble average).

4.2.3 *Step-size distribution has exponential tails*

Although we observe a sub-diffusive MSD, we also observe rare but relatively large displacements of complexes, e.g. from mid-cell to the cell pole in a few minutes. These events, which have also been recently observed in the motion of the nucleoid [37], suggest variability in the mobility beyond the expected statistical variability of sub-diffusive

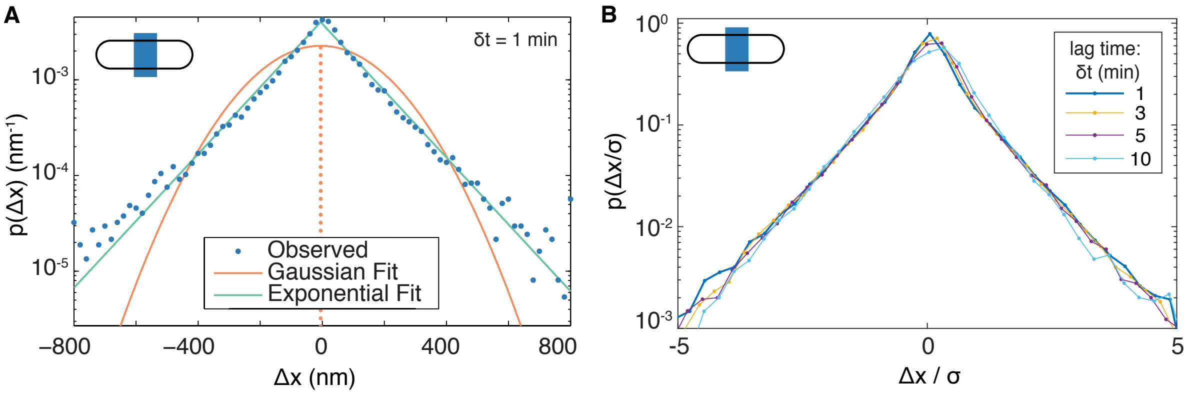


Figure 4.5: **Step-size distribution. Panel A:** The step-size distribution is shown for MS2-mRNA complexes for a $\delta t = 1\ min$ lag time for complexes with an initial position in the middle fifth of the cell, as illustrated in the inset. The data (blue circles) are compared with two models for the step-size distribution: Exponential (green line) and (ii) Gaussian (orange line). Both models have the same mean and variance as the experimental data. The observed step-size distribution is in excellent agreement with the Exponential Model (N=8053 trajectories). **Panel B:** The step-size distribution of the complexes for different lag times, normalized by the standard deviation, is said to be scale independent, since it is independent of the lag time.

motion. To quantitatively characterize these rare events, we compute the step-size distribution, defined as $\Delta x(t) = x(t + \delta t) - x(t)$, where the lag time, δt, is equal to the frame rate. Note that for clarity the data shown in Fig. 4.5A is only for MS2-mRNA complexes found in the middle of the cell. (Additional distributions from other cellular positions are included in the Appendix Fig. C.1) A standard diffusive model would predict a gaussian distribution of step sizes. Instead, we observe that the step size distribution is well characterized by a symmetric exponential decay function

$$\mathrm{p}(\Delta x; \delta t = 1\ min) \propto e^{\frac{-|\Delta x|}{\lambda}} \tag{4.2}$$

with a decay constant $\lambda = 125 \pm 2\ nm$. Both the gaussian and exponential fits in Fig. 4.5A have identical variance, but the exponential model has a higher frequency of large steps, in agreement with our qualitative observations of MS2-mRNA dynamics.

In Fig. 4.5B we show the step-size distribution for a set of different lag times, δt. We observe that the step-size distribution $\mathrm{p}(\Delta x/\sigma; \delta t)$ is scale invariant since the distribution of steps, normalized by their standard deviation, is independent of lag time δt.

4.2.4 *Viscoelastic nature of the motion*

The observation of the Laplace distribution for small displacements at long lag times has important implications. This regime of the distribution function is canonically Gaussian if subsequent steps are independent. Failure to observe a Gaussian distribution implies that the motion has a *memory* [65]. A canonical method to characterize the memory is computing the correlation between steps using the velocity autocorrelation function:

$$C_v(\Delta t; \delta t) \equiv \langle \Delta x(t)\, \Delta x(t + \Delta t) \rangle \,/\, \langle\ \Delta x^2\ \rangle\,, \tag{4.3}$$

where $\Delta x/\delta t$ is the average velocity over lag time δt and Δt is the time delay. The observed velocity autocorrelation has two important properties: it is approximately scale-invariant since since it depends only on the ratio $\Delta t/\delta t$ and it is negative for $\Delta t/\delta t = 1$,

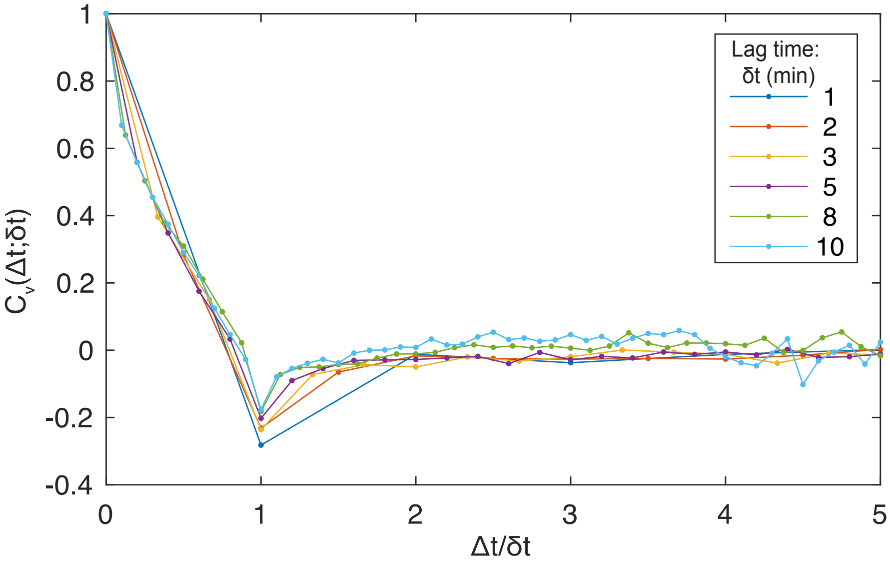


Figure 4.6: **Velocity Autocorrelation.** Velocity autocorrelation function for different velocity lag times δt. The motion is anticorrelated for $\Delta t/\delta t = 1$ for all lag times δt.

signifying anticorrelation between subsequent steps (Fig. 4.6). The negative velocity autocorrelation function is typically interpreted as viscoelasticity [17].

4.2.5 Mapping cell-cycle and position-dependent particle motion.

Fig. 4.5 displays the step-size distribution for MS2-mRNA complexes found at mid-cell, but we have trajectory information from nearly every region of the cell. To efficiently characterize the step-size distribution at different regions along the long-axis of the cell, we calculate the first two moments of the distribution: We will define the drift velocity as the mean of the frame-to-frame step-size distribution divided by the lag time (equal to the time between successive measurements),

$$\langle v(x,t)\rangle \equiv \frac{\langle \Delta x(t)\rangle}{\delta t} \tag{4.4}$$

and the diffusion coefficient as the variance of the frame-to-frame step-size distribution divided by twice the lag time,

$$D(x,t) \equiv \frac{\langle (\Delta x(t) - \langle \Delta x(t) \rangle)^2 \rangle}{2\delta t} \tag{4.5}$$

where these definitions are consistent with the standard definitions of drift velocity and diffusion coefficient for a diffusive particle in a Newtonian fluid.

In order to map the dynamics of MS2-mRNA complexes in distinct regions of the cell, the long axis of the cell was split into fifty regions and the frame-to-frame displacement of each MS2-mRNA complex was binned by the starting position of each displacement vector. We then calculate the drift velocity (Eq. 4.4) and diffusion coefficient (Eq. 4.5) for each region. Because of cell-to-cell variation in doubling time, in order to investigate cell-cycle-dependent dynamics we split the cell cycle in three relative time phases (early, mid, late), and calculated the drift velocity and diffusion coefficient map for each time phase individually.

Figure 4.7A shows an annotated drift velocity map, i.e. the drift velocity as a function of position along the long axis of the cell, averaged over the entire cell cycle for a lag time δt = 1 minute. To interpret the velocity map, we identify locations along the long-axis of the cell where the drift velocity is zero. If the slope of the velocity map is negative at the zero crossing, the position corresponds to a stable point (dashed line). If the slope of the velocity map is positive at the zero crossing, the position corresponds to a saddle point dividing two catchment regions, corresponding to regions in the cell from which a MS2-mRNA complex would move towards a stable point (in the absence of diffusive motion).

Figure 4.7B shows the diffusion coefficient of MS2-mRNA complexes as a function of position along the long axis of the cell. Rather than being constant throughout the cell, we observe a large variability in the diffusion coefficient as a function of long-axis position in the cell. The most striking feature of Fig. 4.7B is that the diffusion coefficient is highest at the cell-quarters, corresponding to the location of the nucleoid. The lowest mobility is at

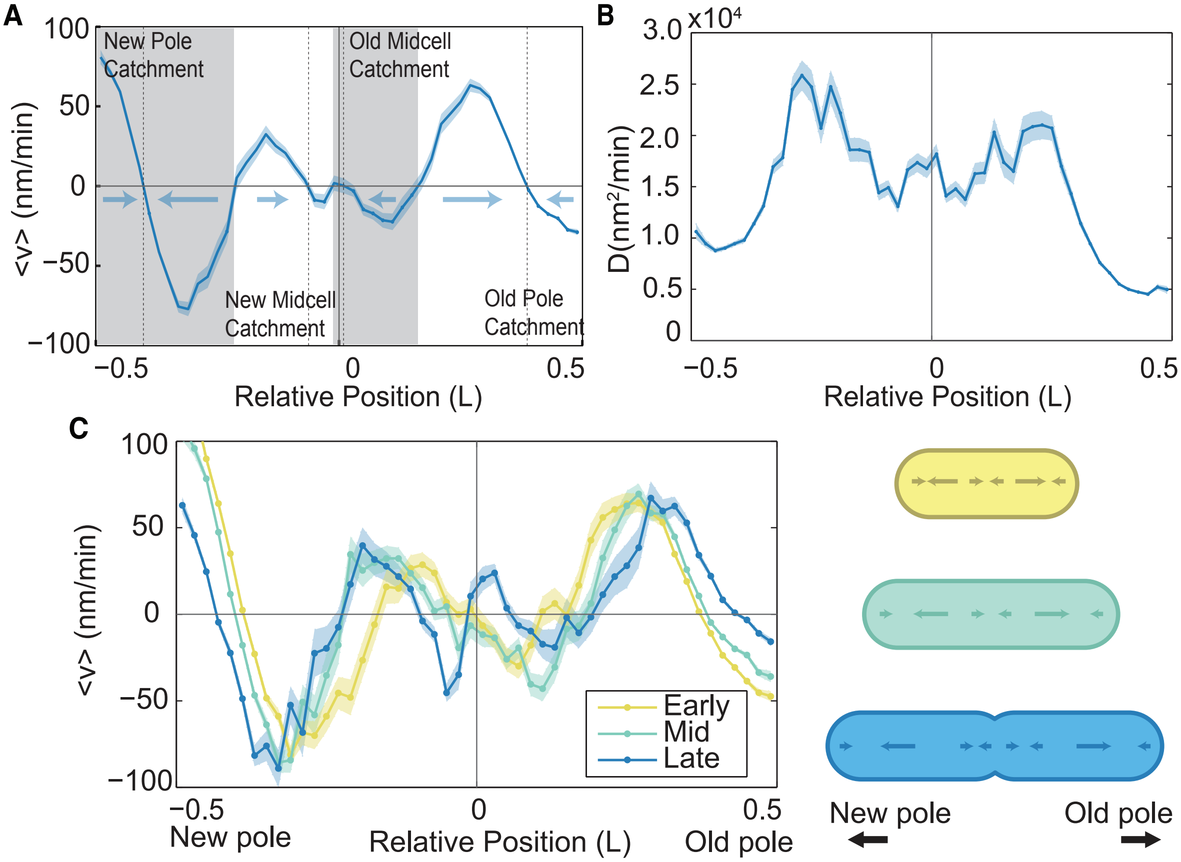


Figure 4.7: **Drift velocity and diffusion coefficient. Panel A:** Annotated map of drift velocity averaged over the cell cycle, with stable points (dashed lines) and catchment regions (grey and white regions) annotated. **Panel B:** Map of diffusion coefficient along the long axis of the cell averaged over the cell cycle. The highest mobility occurs over regions occupied by the nucleoid (± 0.25 *cell lengths*). The lowest mobility is observed at the old pole. **Panel C:** The drift velocity map shown for three relative cell-cycle times (Early, Mid, Late) displays weak cell-cycle dependence. During the last third of the cell cycle, the velocity map is qualitatively changed at midcell, where the centrally located stable point divides into two stable point prior to division. (For all N=8053 trajectories.)

the old pole, where the diffusion coefficient is roughly half the value ($4900\,\mathrm{nm}^2/\mathrm{min}$) than at the new pole ($9500\,\mathrm{nm}^2/\mathrm{min}$).

Figure 4.7C shows the drift velocity map for the three relative phases of the cell cycle. For the first two-thirds of the cell cycle (early and mid) there are three stable points, corresponding to the cell poles and mid-cell. The stable points at the old and new pole correspond to the high-occupancy regions adjacent to the cell poles while the occupancy of the stable region at midcell is much lower (see Fig. 4.3). During the last third of the cell cycle (late), the central stable point splits into two stronger stable points, which will become the new cell poles of the daughter cells after division. Because the overall cell-cycle dependence of the drift velocity map is generally weak, we will ignore the cell-cycle dependence in most of the analysis that follows.

Finally, to test whether or not it is appropriate to interpret the observed data by a steady-state drift-diffusion model, we calculated whether or not the observed occupancy, drift velocities, and diffusion coefficients satisfied Fick's Law for zero net flux

$$j(x,t) = -D(x,t)\boldsymbol{\nabla} n(x,t) + \langle v(x,t)\rangle\, n(x,t) = 0 \tag{4.6}$$

where j is the flux and n is the number density of the MS2-mRNA complexes. Using Eq. 4.6, we can infer the drift velocity required to equilibrate the observed MS2-mRNA occupancy and compare this inferred drift velocity to the observed drift velocity. As shown in the appendix C.2 and Fig. C.2, there is excellent agreement between these velocities, supporting the use of drift-diffusion model.

4.2.6 Quantitative model of nucleoid-exclusion and membrane confinement.

To quantitatively test the proposed role of nucleoid-exclusion in generating the observed drift velocities found in Fig. 4.7, we construct a minimal statistical mechanics model of nucleoid exclusion. The model is derived and discussed in detail in appendix C.3. Briefly, we treat the cytoplasm as a lattice where nucleoid DNA and MS2-mRNA cannot occupy the same site and then estimate the free energy cost from the excluded volume of the nucleoid. The model depends on a single unitless geometric parameter, Φ, that is fit by comparing the spatial distribution of MS2-mRNA complexes to the density of the nucleoid. The predicted velocity profile is

$$\langle v(x,t)\rangle = -\Phi V D(x,t)\boldsymbol{\nabla} n_{\mathrm{DNA}}(x,t) \tag{4.7}$$

where n_{DNA} is the number density of the nucleoid, V is the cell volume, and D is the measured diffusion coefficient (as a function of position and cell cycle age). To estimate the density of the nucleoid, we measure the mean fluorescence intensity of DAPI-stained nucleoids (Fig. 4.8A). The predicted velocity profile from nucleoid exclusion is plotted in Fig. 4.8B (green dotted line), which shows excellent agreement to the observed drift velocity maps throughout the long-axis of the cell, although notably fails at the cell poles.

In addition to the exclusion forces generated by the nucleoid, the inner membrane of the cell also confines the motion of cytoplasmic complexes. To estimate the effect of membrane confinement of MS2-mRNA complexes at the cell poles (where our nucleoid exclusion model fails), we model the dynamics of protein complexes near the cell poles as diffusive particles near a hard wall. The solution to this problem is worked out in detail in appendix C.4. Briefly, we calculate a theoretical step-size distribution by solving the diffusion equation with a hard-wall boundary condition using the method of images, which provides the following exact solution for the drift velocity:

$$\langle v(\delta,\delta t)\rangle = \frac{1}{\delta t}\exp\left(-\frac{\delta^2}{4D\delta t}\right) - \frac{\delta}{\delta t}(1-\mathrm{erf}(-\frac{\delta}{\sqrt{4D\delta t}})) \tag{4.8}$$

where δ is the distance from the cell pole and δt is the lag time in the velocity measurement. Note that this makes a fit-parameter-free prediction of the drift velocity at the poles of the cell solely due to membrane confinement. The predicted velocity profile from Eq. 4.8 is shown in Fig. 4.8B (red dotted line), which shows excellent agreement with the observed drift velocity map near the cell poles.

If we now combine the predicted drift velocity maps for the nucleoid exclusion and

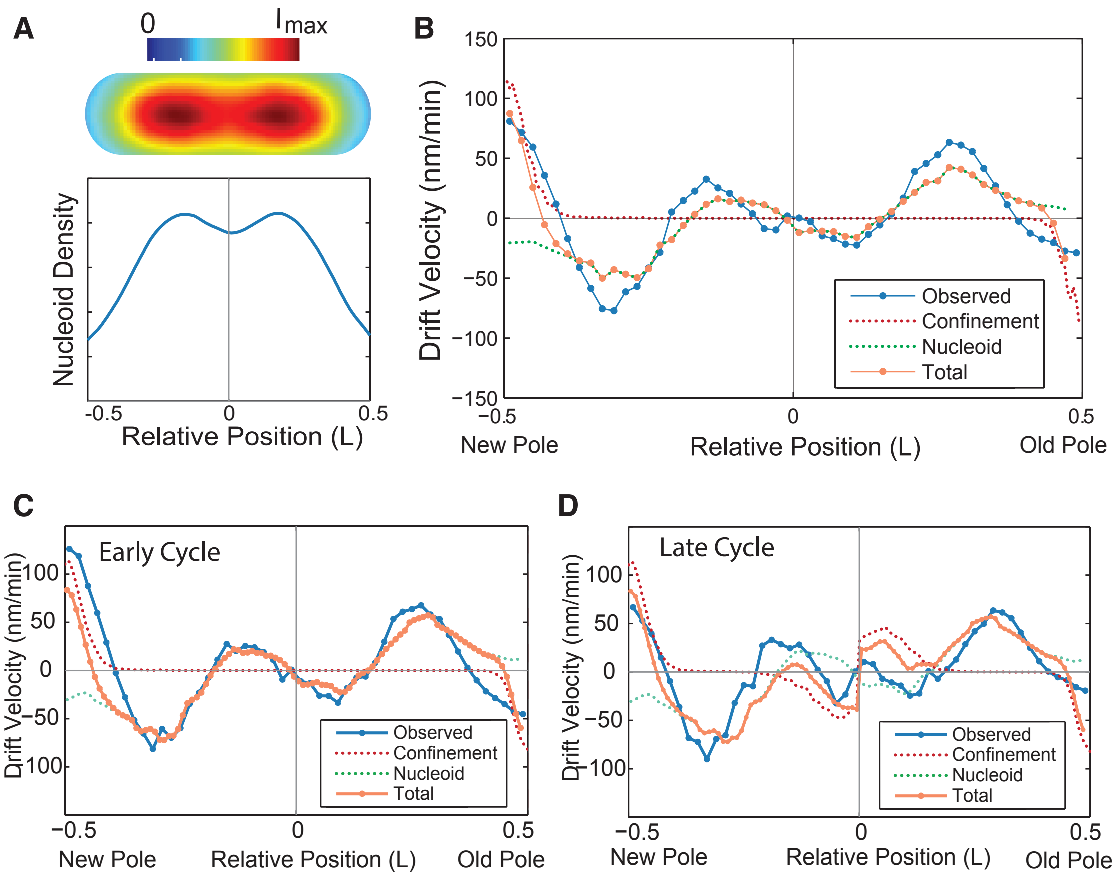


Figure 4.8: **Nucleoid exclusion statistical model. Panel A:** At the top, nucleoid density in the cell. Cell-cycle and cell-to-cell mean interpolated fluorescent image of DAPI stained nucleoids (N=369). At the bottom, nucleoid density as the sum of intensity of DAPI stained nucleoids over the minor axis of the cell for relative positions along the major axis of the cell. **Panel B:** Comparison between predicted and measured velocity profile from nucleoid-exclusion and membrane confinement at the poles during the whole cell cycle. **Panel C:** Measured velocity profile and predicted velocity profile from nucleoid-exclusion and membrane confinement at the poles during the first half of the cell cycle. **Panel D:** Measured velocity profile and predicted velocity profile from nucleoid-exclusion and membrane confinement at the poles and mid-cell during the second half of the cell cycle. The combination (orange) of the predicted velocity profile from membrane confinement (red) and nucleoid-exclusion (green) is in excellent agreement with the measured profile (blue).

membrane confinement models, we see striking agreement with the qualitative shape and quantitative scale of the observed drift velocity map throughout the entire cell, indicating that both contributions are required to accurately model cytoplasmic dynamics. More parameters can be included in the model to recapitulate subtle features of the velocity map. For instance, the velocity map can be split to an early and late cell-cycle velocity map. The late cycle shows an outward force at mid-cell due to the closing septum that can be reproduced by adding a partial membrane confinement in the middle (Fig. 4.8D). In the interest of isolating the most general aspects of cytoplasmic dynamics, for the rest of the analysis we will focus on the contributions from nucleoid exclusion and polar confinement during the whole cell cycle.

4.2.7 Perturbations of nucleoid structure, cell geometry and physiology affect drift velocities and diffusion coefficients.

To further investigate the role of the nucleoid in determining the mobility of complexes in the cell, we perturb cell physiology using a variety of small-molecule inhibitors. To identify changes to the MS2-mRNA motion we consider perturbations to the diffusion coefficient and changes to the nucleoid-induced exclusion. Based on previous work, we expect reductions in the metabolic activity through ATP-depletion would reduce the diffusion coefficient without significantly effecting nucleoid structure [63, 66], while treatments that alter cell shape or the structure of the nucleoid would affect the nucleoid-exclusion-induced forcing.

For each perturbation we again map the drift velocity and diffusion coefficient of MS2-mRNA complexes along the long axis of the cell. As can be seen from Eq. 4.9, because the predicted drift velocity distribution from nucleoid exclusion depends on both the spatially varying diffusion coefficient and the shape of the nucleoid, it is convenient to introduce the forcing ratio, defined as the ratio of drift velocity to the diffusion coefficient:

$$f(x,t) \equiv \frac{\langle v(x,t)\rangle}{D(x,t)} \approx \boldsymbol{\nabla} n_{\mathrm{DNA}}(x,t) \tag{4.9}$$

in order to isolate contributions from nucleoid structure.

It has been reported that metabolic activity [63] and ATP-hydrolysis [37] increase the mobility of complexes in vivo. To investigate the role of ATP-hydrolysis, we treated the cells with the ATP-synthesis inhibitor sodium azide. As previously reported, the mean diffusion coefficient is significantly reduced, consistent with ATP-hydrolysis playing a role in the dynamics of molecular complexes in the cytoplasm [66]. But because the forcing ratio is not significantly affected by ATP depletion (Fig. C.3A), we conclude that since ATP-driven processes do not significantly affect nucleoid structure on the time scale of the cell cycle, they do not significantly contribute to nucleoid-exclusion-driven forcing.

To further test the hypothesis that nucleoid exclusion is a key contributor to the observed drift velocity map, we first perform a non-perturbative test: The small molecule A22 inhibits the activity of MreB, an essential protein involved in cell-wall synthesis [67]. Without a proper cell wall, cells treated with A22 lose their rod-like shape and become more spherical while the nucleoid appears to be more diffusely localized throughout the cell (Fig. 4.10). The forcing ratio for cells treated with A22 is shown in Fig. 4.9A. Under these conditions, the spheroid cells were still capable of undergoing division, therefore for analysis we define the long axis of the cell to be perpendicular to the division plane. Strikingly, the qualitative shape of the forcing ratio curve for A22 treated cells is nearly opposite that of untreated cells, supporting the hypothesis that nucleoid exclusion (specifically the overall shape and density of the nucleoid) is a key component of the dynamics of molecular complexes in the cytoplasm.

The A22 treatment leads to large-scale changes in cellular structure. To more specifically test our hypothesis, we also attempt a more subtle perturbation to the shape and density profile of the nucleoid: we treat the cells with novobiocin, an antibiotic that inhibits DNA-gyrase activity and leads to DNA damage. Extended treatment with novobiocin leads to significant condensation of the nucleoid (Fig. 4.10). Because the nucleoids

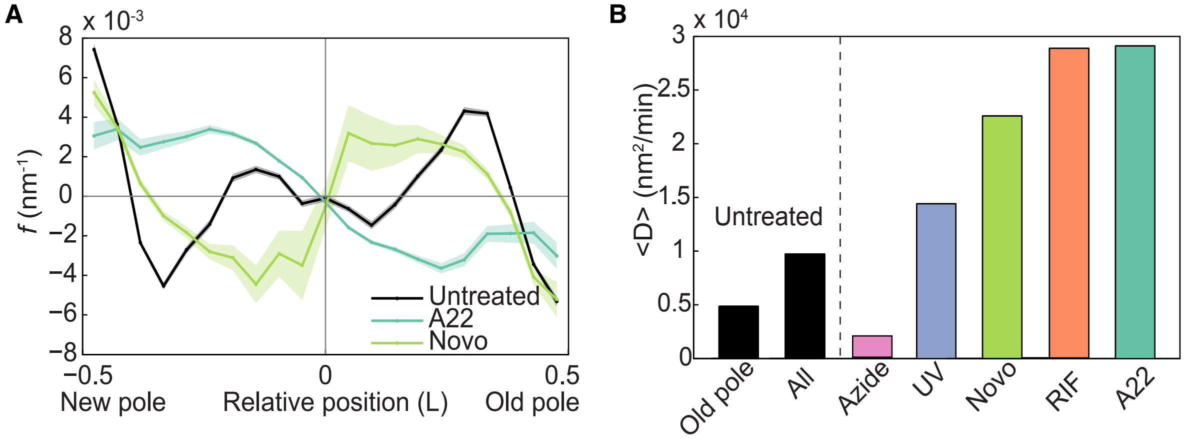


Figure 4.9: **Treated versus untreated cells. Panel A:** Forcing ratio (⟨ v ⟩ /D) versus relative position in cell for MS2-mRNA complexes in untreated, spherical (A22) (N=3207 trajectories), and gyrase inhibited cells (Novobiocin) (N=643). **Panel B:** Mean diffusion coefficient of MS2-mRNA complexes at the old pole and at all positions in untreated cells, in ATP-depleted (Azide) (N=163), DNA-damaged (UV) (N=338), gyrase inhibited (Novobiocin) (N=643), transcription inhibited (Rifampicin) (N=195), and spherical cells (A22) (N=3207).

of cells treated with novobiocin are shorter with respect to the long axis of the cell, we expect the positions of the peaks in the forcing ratio (located at ± 0.3 cell lengths in untreated cells) to shift inward towards midcell, matching the shrinking geometry of the nucleoid. As expected, the forcing ratio of novobiocin treated cells shows inward movement of the peak forces (Fig. 4.9A). In addition, the stable point at midcell is lost in novobiocin treated cells, consistent with the loss of the typical bi-lobed structure of the nucleoid due to extreme condensation.

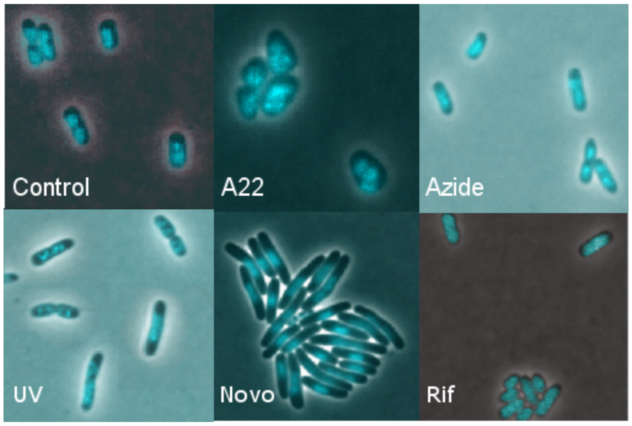


Figure 4.10: **Nucleoid visualization of untreated and treated cells.** Cells untreated and treated with either A22, Azide, Novobiocin, Rifampicin, and radiated with Ultra Violet light and stained with DAPI for the visualization of the nucleoid. Cells treated with A22 show more diffusely localized nucleoids. Azide does not appear to cause a visible difference to the shape of the nucleoid. Cells radiated with Ultra Violet radiation appear to disperse the nucleoid around the cell. Treatment with Novobiocin leads to condensation of the nucleoid and rifampicin causes expansion of the nucleoid.

In general, with the exception of ATP-depletion, we find that most physiological perturbations to the cell, e.g. altered cell shape (A22), DNA-gyrase inhibition (novobiocin), DNA damage (UV radiation), or transcription inhibition (rifampicin), tend to increase the diffusion coefficient of the MS2-mRNA complexes without significantly affecting the qualitative shape of the drift velocity map (Fig. 4.9B and Appendix Fig. C.3).

4.3 Discussion

Our results provide quantitative evidence for both the existence and relevance of forces induced by nucleoid exclusion and membrane confinement on the motion of large molecular complexes in the bacterial cell. Although nucleoid exclusion and confinement have often been invoked to explain the retention of large complexes at the poles, a detailed picture of the dynamics, including the understanding of how complexes move to the poles, has never been described from a quantitative biophysical perspective.

4.3.1 Exponential step-size distribution

The observation of non-gaussian step size for the motion of protein complexes in the bacterial cell has recently been reported by Parry et al. as a central argument for identifying the dynamics of the bacterial cytoplasm as glass-like [63]. The observation of exponential tails, and more generally non-gaussian step-size distributions, is rather generic in biology and soft-condensed matter physics (e.g. see Ref. [68]). The exponential dependence of the step-size distribution has been explained as a consequence of standard Fickian diffusion with a wide distribution of diffusion coefficients. A number of mechanisms can result in variability in diffusion coefficients, including quenched disorder in the media or transitions between states with different mobility that have lifetimes that are comparable to the lag time of observation [68]. In the context of the bacterial cytoplasm, it would be natural to assume both mechanisms may contribute to the long tails of the step-size distribution. For instance, we observe here that the diffusion coefficient of MS2-mRNA complexes varies as a function of position along the long-axis of the cell. Although these effects are typically obscured by cell-to-cell variation, the step-size distribution is explicitly not affected by cell-to-cell averaging.

4.3.2 The dynamics of large complexes is scale-invariant

We have observed that the motion of the complexes is scale invariant from the perspective of three distinct metrics: (i) The MSD corresponds to sub-diffusional motion and is scale-invariant in the sense that is it self-similar: A dilation of time and concomitant dilation of space leave the MSD invariant. (ii) Another important scale-invariant feature of the motion is the step-size distribution. The step-size distribution, normalized by their standard deviation, is independent of lag time δt and therefore scale invariant. (iii) The observed velocity autocorrelation is approximately scale-invariant since it depends only on the ratio of the delay time over the lag time, $\Delta t/\delta t$. This emergent scale invariance motivates the need for a general mechanism to describe cytoplasmic dynamics, which we will address in the next chapter.

4.3.3 Unexpectedly high mobility over regions of high nucleoid density

Classic investigations of the effects of molecular crowding on diffusive particles show that the diffusion coefficient tends to decreases with increased concentration of the crowding agent (e.g. [16, 69]). In our nucleoid-exclusion model (Eq. 4.7), we proposed that DNA crowds the MS2-mRNA complexes in the vicinity of the nucleoid, reducing both the number of local configurations overlapping the nucleoid (leading to depletion in MS2-mRNA occupancy) and frustrating diffusion by occluding neighboring sites. Therefore, the nucleoid-exclusion model would naively predict both decreased occupancy over the nucleoid as well as a decreased diffusion coefficient. Strikingly, our results show exactly the opposite phenomenology: comparison between the occupancy of MS2-mRNA (Fig. 4.3) and the position dependent diffusion coefficient (Fig. 4.7B) reveals that the highest diffusion coefficients are observed in regions with the lowest MS2-mRNA occupancy and the highest nucleoid density.

One attractive hypothesis is that the nucleoid, while acting as a crowding agent and depleting MS2-mRNA complexes from the central region of the cell, is also very dynamic, due in part to active processes [51] like transcription and replication, which can both result in large-scale rearrangements of the nucleoid. In fact, a number of lines of evidence support the central role of the nucleoid in the high MS2-mRNA mobility. For instance, the transcription inhibitor rifampicin is known to affect nucleoid structure, but its dominant effect on MS2-mRNA motion is to significantly increase the diffusion coefficient. Although one might expect transcription inhibition to diminish chromosome mobility, it in fact leads to a significant increase in mobility [62]). A tight link between MS2-mRNA and chromosome dynamics would therefore account for the large increase in diffusion coefficient and make a key prediction: The increased mobility of MS2-mRNA complexes would be most pronounced over the nucleoid in cells treated with Rifampicin. By applying our analysis to Rifampicin-treated cells, we do in fact observe nearly a four-fold increase in the diffusion coefficient of MS2-mRNA complexes in regions of highest nucleoid density (Appendix Fig. C.3B) [52].

In contrast, many large macromolecular complexes, e.g. ribosomes and protein aggregates, are known to accumulate at the cell poles. Although many processes (including translation) are active at the poles, these processes do not appear to efficiently facilitate MS2-mRNA mobility. It is possible that the presence of large complexes further frustrates the movement of the MS2-mRNA complexes [51]. Consistent with this model is the observation that the diffusion coefficient of MS2-mRNA complexes at the old pole is smaller by a factor of two than the diffusion coefficient at the new pole (Fig. 4.7B), since it is known that non-functional protein aggregates tend to accumulate at the old pole [52]. But, despite the lower mobility at the old pole, there is still strong evidence of significant non-thermal contributions to the dynamics at both poles since the mobility is significantly decreased with the depletion of ATP (Fig. 4.9B).

Parry et al. also proposed that cellular metabolism fluidizes a glass-like cytoplasm [63]. The glass model was invoked because glass-like dynamics, like cytoplasmic dynamics, are the result crowding and caging related phenomena. In particular, Parry et al. report distinct subpopulations of protein complexes with fast and slow dynamics, in analogy to distinct subpopulations observed in glass forming-liquids [63]. In a glass, rapid particle movement is the result of collective motion by neighboring particles, but there is currently no direct evidence for collective motion of this type in the bacterial cytoplasm. Our observations are consistent with the observations of Parry et al. in that we observe trajectories with a wide range of diffusion coefficients, but our spatial analysis of the diffusion coefficient as a function of position in the cell reveals that these subpopulation can be explained, at least in part, by the presence or absence of the nucleoid (Appendix Fig. C.4). Since Parry et al have reported metabolism-dependent differences in mobility in DNA free regions, we suspect the proposed metabolism-dependent mechanism must be distinct from the phenomena we report.

4.4 Conclusion

Our observations provide strong quantitative support for two modes of nucleoid action: both for the widely-accepted mechanism of nucleoid-exclusion and a newly proposed mode, facilitating rapid complex motion through the cytoplasm. These combined roles of the nucleoid in facilitating the movement of protein complexes has important biological implications and suggests that the nucleoid may play a much more active role in the organization and transport of macromolecules in the bacterial cytoplasm than previously thought.

Chapter 5

STRONG DISORDER LEADS TO SCALE INVARIANCE IN COMPLEX BIOLOGICAL SYSTEMS

Although the crowded and non-equilibrium cytoplasm of the cells is an extremely complex system, we have shown in the previous chapter that molecular complexes in the cytoplasm demonstrate scale-invariance in three distinct metrics: mean-squared displacement, velocity autocorrelation and the step-size distribution. In fact, condensed matter physics provides many example of emergent scale-invariant behavior in complex systems [70]. In a physical context, symmetry and the separation of length, time or energy scales provide powerful mathematical constraints that result in simple effective models for physical systems [71]. Although these properties may be absent in most biological contexts, emergent and scale-invariant behavior can be observed in many of these systems [72, 73, 74]. The discovery of the physical mechanisms for this emergent behavior has the potential to provide powerful new approaches to the analysis of complex systems.

The renormalization group (RG) describes the scale-dependence of physical theories [70]. An important prediction of RG is the emergence of *fixed points* corresponding to effective, long-length-scale theories with a small number of couplings that none-the-less accurately describe the physics of complex and fundamental theories at short-length-scales [70]. There is a many-to-one mapping from these complex short-length-scales theories to simple effective long-length-scale theories. In the context of random motion, diffusion is one such fixed point, but our experimental observations argue for the existence of another fixed-point for strongly disordered systems.

We propose a natural mechanism for the emergence of such a fixed point that can be understood from the perspective of the statistical properties of extreme values [75, 76].

Many biological processes are limited by an extreme value. For instance, reactions are often limited by the slowest step [77], evolutionary dynamics is limited by the growth rate of the fittest organism [73], or the first-passage time of a random walker can be limited by the largest barrier. These processes can all be modeled as depending on an extreme value of a set of random variables (representing rates, fitnesses and free energies respectively). These scenarios describe a strong-disorder limit since we are interested in processes where the difference between the extreme value and the other values is so large as to make these other values irrelevant. As the Central Limit Theorem describes the limiting distribution of the sum of N independent random variables, Extreme Value Theory (EVT) describes the limiting distribution of the extremum of N random variables [75, 76]. EVT has been used before to understand biological phenomena limited by extremes. Its implications have been investigated in fitness, beneficial mutations and adaptation of DNA sequences [78, 79, 80].

To study the emergence of scale-invariance in a complex biological system, we focus on the anomalous dynamics of large complexes in the bacterial cytoplasm that we studied in the last chapter. We will refer to these complexes as particles for the rest of the chapter. We will compare the results of dynamics hindered by extreme values and the dynamics of these particles and propose a general mechanism by which self-similarity emerges in the strong-disorder limit of biological systems.

5.1 Existing models

We provide a brief overview of the proposed models for sub-diffusive motion in the context of the current problem. Three mechanisms commonly invoked to model sub-diffusive phenomena are Continuous Time Random Walk (CTRW), fractional Brownian motion (fBm) and inhomogeneous media [65].

In CTRW, the particle waits for a random waiting time before the next jump. A power-law distributions of wait times leads to sub-diffusion. The inhomogeneous media model (patch model) consist of particles diffusing through quenched patches (*i.e.* static) with patch-specific diffusion coefficients [81, 82]. Both of these models are closely related to the Trap model that we will discuss shortly [65]. All three of these models lack the observed viscoelastic phenomenology and are non-ergodic in the strong disorder limit, in contrast with the observed dynamics [17, 83].

fBm describes random motion with Gaussian-distributed but anti-correlated steps. This model does naturally describe the observed viscoelastic phenomenology as well as the observed sub-diffusive motion. On the other hand, it does not predict the observed Laplace step-size distribution nor the scale-invariance of the velocity auto-correlation function.

Granick and coworkers have proposed a general mechanism for the observed Laplace-distributed step-size distribution [68]: independent diffusive processes with an exponential distribution of particle-specific diffusion coefficients (Exp-D). But, in order for this mechanism to give rise to a Laplace distribution, the diffusion coefficients must be quenched on a timescale comparable to those probed by experiment, otherwise transitions between diffusion coefficients result in diffusion (in the long-time limit).

We have recently proposed combining the fBm model and the exponential distribution of diffusion coefficients to describe the observed dynamics. Although this model can fit almost all of the observed phenomenology [83], there are two complications: (i) The per-particle diffusion coefficient is not quenched: the lifetime of the diffusion-coefficient memory can be measured and it is short (Appendix Fig. D.1). (ii) A satisfactory model would also have to describe a natural mechanism for the exponential distribution of per-particle diffusion coefficients without the need for fine-tuning. In this paper, we describe such a natural model where viscoelasticity, sub-diffusion and self-similarity all arise as a consequence of strong disorder.

5.2 *Description of the model*

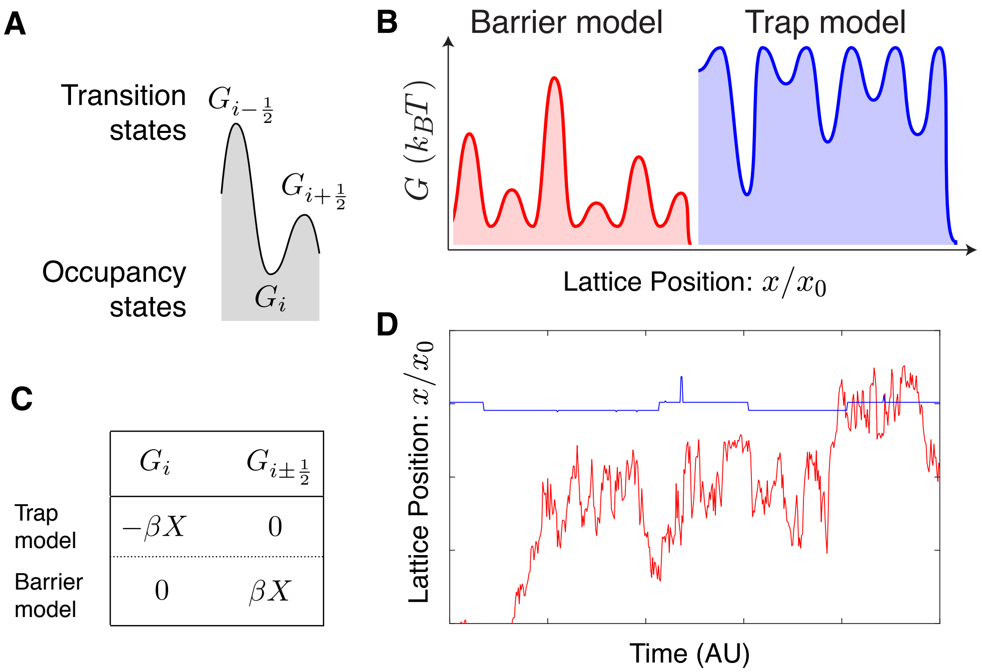
5.2.1 *Random walker on a rough landscape*

Due to the strong crowding in the cell [84, 77], it seems natural to investigate random walks in a disordered free energy landscape [65]. We model the cytoplasm as a one-dimensional lattice. Higher dimensions will be addressed later. We represent integer sites in the lattice as occupancy states and half-integer sites as transition states. (See Fig. 5.1A.) The free energy (in units of $k_B T$) of each state is G_i. The hopping rate k and average hopping time τ from site i to sites $i \pm 1$ have the Arrhenius dependence:

$$k_{i \to i\pm 1} = \tau^{-1} = t_0^{-1} \, \exp\Big(G_i - G_{i\pm\frac{1}{2}}\Big), \tag{5.1}$$

where t_0 represents a fundamental relaxation time in the system and the free energy difference in the exponent represents height of the free energy barrier to transition through the transition state. In the most general model, the free energies of both the occupancy (integer) and transition (half integer) states are represented as random variables [65]. But, it is useful to consider two limiting cases: the *barrier* and *trap* models [65], illustrated in Fig. 5.1BC. These models were originally studied in the context transport problems in the 1980s [85, 65]. In the trap model, the depth of the transition state is stochastic whereas the free energies of all transition states are 0. In the barrier model, the free energies of the barriers are stochastic whereas the free energies of the occupancy states are 0 (Fig. 5.1C).

We will treat the free energies as quenched disorder. In this context, quenched disorder implies that the free energies are static for each lattice site. In reality, this disorder in dynamic, but due to the polydispersed protein complexes in the cell, there are a wide range of relaxation times. We shall assume that many of these times are longer than the time scales of interest in our experiment, motivating the use of quenched disorder in the model. An additional motivation for the use of quenched disorder in the model is that the observed distribution functions are constructed using traces generated in many cells each which has distinct ultra-structure [1]. Finally, it is already well known from the



	G_i	$G_{i\pm\frac{1}{2}}$
Trap model	$-\beta X$	0
Barrier model	0	βX

Figure 5.1: **Random walker on a rough landscape. Panel A:** The dynamics is modeled as a random walk on a one-dimensional rough free energy landscape. We realize the dynamics as transitions between adjacent occupancy states, at integer values on the lattice, through transitions states located at half-integer positions on the lattice. The free energies of the respective states are G_i and $G_{i+\frac{1}{2}}$ and are defined in thermal units (k_BT). **Panel B:** To understand the generic model, we study two limiting cases: the trap and barrier models. (The free energy offset between models is of no significance.) **Panel C:** In the trap model, the free energies of the transition states are 0 and the free energies of occupancy states are independent-and-identically-distributed random variables X scaled by a disorder strength β. In the barrier model, the free energies of the transition states are are independent-and-identically-distributed random variables X scaled by a disorder strength β and the free energies of the transition states are 0. **Panel D:** The disorder strength $\beta = 1.5$ was chosen to match the observed MSD scaling exponent α. A representative trajectory from each model is shown. The trap and barrier models have qualitatively different dynamics. The trap model shows persistent pausing behavior in the motion corresponding to long lived trapping events. The barrier model shows a bouncing behavior caused by large energy barriers. The barrier phenomenology is difficult to distinguish from canonical Brownian motion.

investigation of models for hopping conductivity that strong disorder can lead to subdiffusion [65]. In this context, strong disorder implies that the disorder-averaged hopping time diverges:

$$\langle \tau \rangle_G \to \infty, \tag{5.2}$$

otherwise the motion is diffusive a long times [86, 87, 65].

It is necessary to choose a semi-infinite interval distribution for the free energy so that the traps and barriers preserve their nature ($G_{i\pm\frac{1}{2}} > G_i$) and to achieve the strong disorder limit (Eqn. 5.2). A canonical distribution with these properties is the one-dimensional chi-squared distribution, multiplied by a unitless disorder strength β. (The explicit functional form of the free energy distribution will not be of central importance.) It is straightforward to demonstrate that $\beta \geq \frac{1}{2}$ corresponds the definition of strong disorder.

5.2.2 Simulations

We perform explicit simulations of this model using either stochastic Gillespie simulations [88, 89] or solving the master equation for the Green's function numerically. We explore various disorder strengths, spatial dimensions and time durations.

Master Equation

By solving the master equation for the Green's function, we obtain the probability density of particles at site i :

$$\dot{p}_i = k_{i-1/2}p_{i-1} + k_{i+1/2}p_{i+1} - (k_{i-1/2} + k_{i+1/2})p_i \tag{5.3}$$

where p_i is the probability distribution of the particles at the occupancy state i and $k_i + 1/2$ is the hopping rate through the transition state i+1/2. We use periodic boundary conditions at the end points.

Gillespie Simulation

For the simulations of the trajectories we use a stochastic Gillespie simulation, which generates a statistically correct trajectory of a stochastic equation [88, 89]. To obtain the next occupancy site at regular time intervals, we use a modification of the Gillespie simulation as found in [90].

In short, for each time step $\Delta t'$ we define the total hopping rate k_{tot} equal to the sum of the hopping rates to every neighboring lattice site. The probability that the particle will not transition to a neighboring state during the time step $\Delta t'$ is dictated by $p(\Delta t') = \exp(-k_{tot}\Delta t')$. The transition occurs if a selected random number r_1, evenly distributed from 0 to 1, is larger than $p(\Delta t')$.

If the transition is to take place, the new occupancy state is found using a second random number r_2, evenly distributed from 0 to 1, and the cumulative probability distribution of the adjacent hopping rates.

The time of the transition is then determined by a third random number r_3 :

$$t_{\text{new}} = t_{\text{prev}} - 1/k_{\text{tot}} \log(r_3(p(\Delta t) - 1) + 1); \tag{5.4}$$

To allow for additional transitions between t_{new} and $t_{\text{prev}} + \Delta t'$ we repeat the same procedure with a shortened time step

$$\delta t' = t_{\text{prev}} + \Delta t' - t_{\text{new}} \tag{5.5}$$

an updated $t_{\text{prev}} = t_{\text{new}}$ and updated k_{tot} for the new adjacent states. The process is repeated until no transition takes place.

5.3 Results

5.3.1 Barriers versus traps

We begin our investigation with a numerical experiment: We simulate particle dynamics in the trap and barrier models. In both models, the free-energy difference for transition is distributed like a scaled chi-squared, but in the trap model, the free energy barriers for forward and reverse motion are identical whereas they are independent random variables in the barrier model. This distinction has significant qualitative consequences on the character of the motion, as shown in Fig. 5.1D, although their qualitatively effect on the MSD is very similar [65]. In this calculation, the disorder strength $\beta = 1.5$ in each model was tuned to generate an MSD curve with a scaling exponent $\alpha \approx 0.65$, matching the observed motion. Qualitatively, the trap model shows characteristic, long-lived trapping events which were qualitatively absent from the observed traces. (See Appendix Fig. D.2.) In contrast, the presence of barrier has a much more subtle qualitative effect on the motion. Barriers lead to the reflection of the particles but stochastic changes in direction are present in canonical Brownian motion, make reflections difficult to detect. (The reflection behavior can be measured statistically, as we discuss below.) Since strong pausing is absent from the observed trajectories, we focus on the Barrier model.

5.3.2 Disorder strength determines MSD scaling

To understand the relation between disorder strength and the dynamics, we compute the MSD for different disorder strengths. As the free energy barriers grow, we expect the motion to slow down and therefore lead to a reduction in the MSD. This reduction could occur via two distinct mechanisms: A reduction in the effective diffusion coefficient D or the scaling exponent α, or both. For weak disorder $\beta < \frac{1}{2}$, only the effective diffusion coefficient D decreases with increasing disorder strength β (Fig. 5.2), but the scaling exponent is diffusive at long times ($\alpha = 1$, [91]). But, for strong disorder $\beta > \frac{1}{2}$, the scaling exponent α decreases with increasing β (Fig. 5.3). Simulation suggests a simple empirical relation between disorder strength and scaling exponent in the strong-disorder limit: $\alpha \approx \beta^{-1}$. Simulations of the barrier model do not reveal significant ergodicity breaking (Appendix Fig. D.3), consistent with experimental observations.

The scale-invariance of the MSD is a feature of the strong-disorder limit of barrier

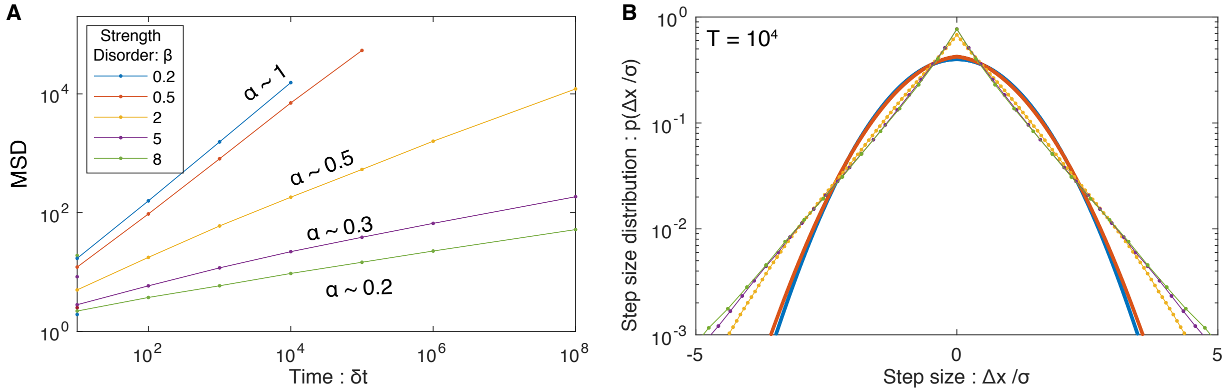


Figure 5.2: **Effect of different disorder strengths. Panel A:** The MSD has decreasing coefficient D and scaling exponent ($\alpha = 1$) with increasing disorder. For weak disorder $\beta < \frac{1}{2}$ the scaling exponent is diffusive ($\alpha = 1$), and sub-diffusive for strong disorder ($\beta > \frac{1}{2}$). **Panel B:** The step-size distribution better is a Gaussian-Like for weak disorder ($\beta < \frac{1}{2}$) and Laplace-Like for strong disorder ($\beta > \frac{1}{2}$) .

models but is not a generic feature of barrier models. For instance, if the transition-state free energies are normally distributed the definition of strong disorder (Eqn. 5.2) is not satisfied, however large the variance. In this model, the MSD is sub-diffusive at intermediate times before crossing over to diffusive at long times (Appendix Fig. D.4A).

5.3.3 Step sizes are Laplace distributed for strong disorder

Next, we compute the disorder-averaged step-size distribution. For the strong disorder limit ($\beta > \frac{1}{2}$), we discover that the step-size distribution is Laplace-like, matching the observed distribution (Fig. 5.4). Since the choice of the chi-squared distribution for the model was motivated by convenience not physics, the agreement between the model and observation suggest that the Laplace-like step-size distribution must be universal. To investigate this hypothesis, we simulate a number of other distributions: exponential, normal and Gumbel. In all cases, strong disorder results step-size distributions which

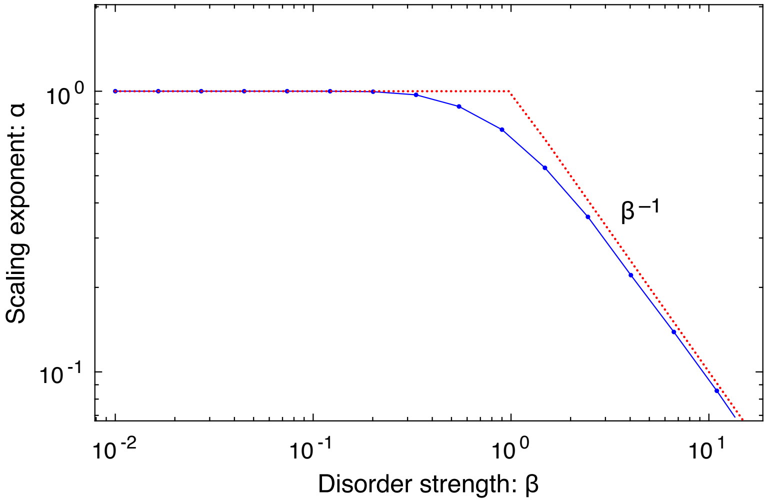


Figure 5.3: **MSD scaling coefficient α versus disorder strength.** The dynamics in the barrier model were diffusive ($\alpha = 1$) for $\beta < \frac{1}{2}$ and sub-diffusive for $\beta > \frac{1}{2}$. The disorder strength is $\alpha \approx \beta^{-1}$ in the strong disorder limit.

are Laplace-like whereas weak disorder results in Gaussian-like distributions (Appendix Fig. D.5). The normal distribution does not satisfy the strong disorder definition (Eqn. 5.2) for any variance, but a Laplace-like distribution is observed at intermediate times and a Gaussian-like distribution is observed at long times (Appendix Fig. D.4B).

These simulations suggest a RG-fixed-point-like behavior: The functional form of the macroscopic step-size distribution is insensitive to the underlying microscopic distribution of the free energy barriers. In the weak-disorder limit, the mechanism is well known (Central Limit Theorem [65]), but in the strong disorder limit, the mechanism is unknown.

5.3.4 *Anti-correlated steps*

An important feature of the observed trajectories is their viscoelastic character: The velocity-autocorrelation function $C_v(\Delta t; \delta t)$ is negative for $\Delta t > \delta t$. We compute the velocity-autocorrelation function by simulation in both the trap and barrier models. The predic-

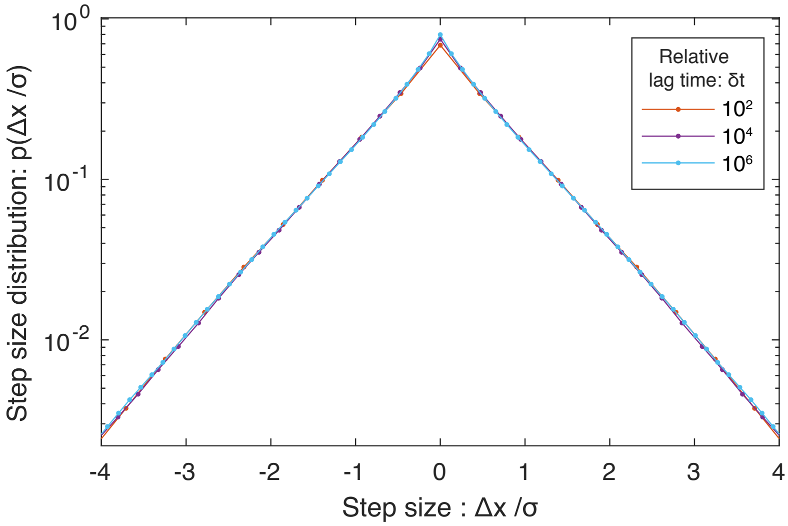


Figure 5.4: **Step-size distribution.** The step-size distribution in the strong disorder is a Laplace distribution and is therefore scale-invariant for all lag times, in agreement with the experimental data. In the strong disorder limit, the functional form of the step-size distribution is insensitive to the underlying functional form of the distribution of the random free energies G.

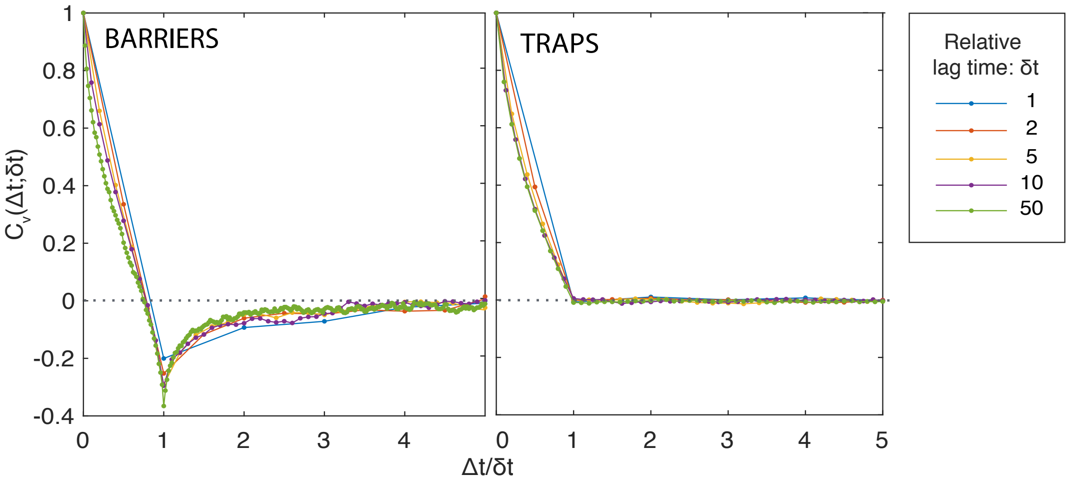


Figure 5.5: **Velocity autocorrelation for trap and barrier model.** The velocity autocorrelation function $C_v(\Delta t; \delta t)$ is shown for the trap and barrier model for strength disorder $\beta = 5$, where Δt is the delay time and δt is the lag time over which the displacements are calculated. The barrier model shows excellent qualitative and quantitative agreement with experiment, where it is zero for non-zero lag times in the trap model.

tions of the trap model do not match experiment: The velocity-autocorrelation function is approximately 0 for $\Delta t > \delta t$, consistent with our qualitative comparison between simulated and observed trajectories in Fig. 5.1. On the other hand, the barrier model prediction has excellent quantitative agreement with the experiment (Fig. 5.5A). The velocity autocorrelation function is scale invariant (depends only on $\Delta t/\delta t$) and has no additional fitting parameters since the disorder strength β is determined from fitting the MSD. The magnitude of the velocity autocorrelation function was found to depend weakly on the disorder strength (Appendix Fig. D.6A).

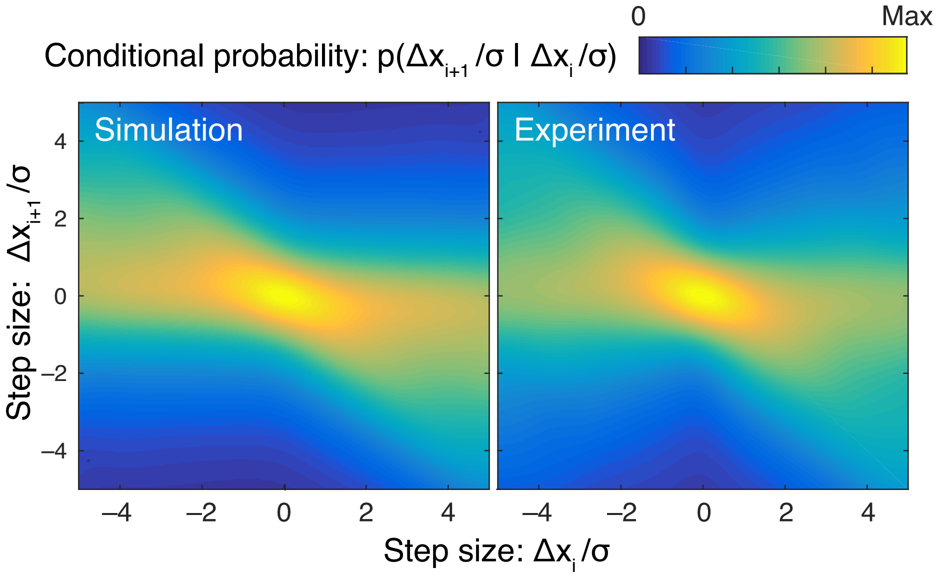


Figure 5.6: Conditional probability distribution for step-size $p(\Delta x_{i+1}|\Delta x_i)$ of experimental data (1 min) of MRNA-MS2 molecular complexes ($N = 8426$) and simulation data in the strong disorder limit. A diagonal band can be seen indicating anti-correlated steps.

The velocity-autocorrelation function can be understood as a moment of the conditional probability distribution for step-size: $p[\Delta x(t + \Delta t)|\Delta x(t)]$. This distribution can be computed from the observed traces, as well as simulated, and can therefore provide a more informative perspective on whether the model can capture the dependence of successive steps than the velocity autocorrelation function alone. The barrier model again provides excellent qualitative agreement with the observed distribution (Fig. 5.6). In both the observed data and the model, there is a strong diagonal band which represent the anti-correlation between subsequent steps. The rescaled distribution depends weakly on the disorder strength (Appendix Fig. D.6). All other models show a qualitatively different structure (Appendix Fig. D.7), again conflicting with the observed phenomenology. Since competing models make different predictions for the qualitative structure of the conditional probability, it is a powerful tool for distinguishing between models.

5.3.5 Mechanism for self-similar structure

Experimentally, it is necessary to pool the trajectories from many cells in order to generate the experimental observables. This pooling of cells and trajectories is realized in the model as a quenched-disorder average. The dynamics without averaging over the quenched disorder, although not comparable to the experimental data, can be informative to the mechanism that gives rise to self similarity.

To understand the dynamics for a fixed realization of the disorder, we compute the Green's function $p[x(t+\delta t)|x(t)]$ numerically for strong disorder (Fig. 5.7). The characteristic feature of the Green's function is the block-diagonal structure of the matrix which is the consequence of the largest stochastic barriers. These features are obscured in the disorder average (Fig. 5.4) since the position and height of the largest barriers is stochastic.

The physical mechanism for this structure can be understood intuitively: In the strong disorder limit, particles rapidly jump over barriers with average hop times τ less than the lag time δt, but the particle motion is limited by the presence of barriers where the hop time is much longer than the lag time ($\tau \gg \delta t$). The uniformity of the probability density in the blocks demonstrates that the motion is limited by these large barriers for strong disorder ($\beta \gg \frac{1}{2}$).

To understand the mechanism for scale invariance, we simulate the Green's function

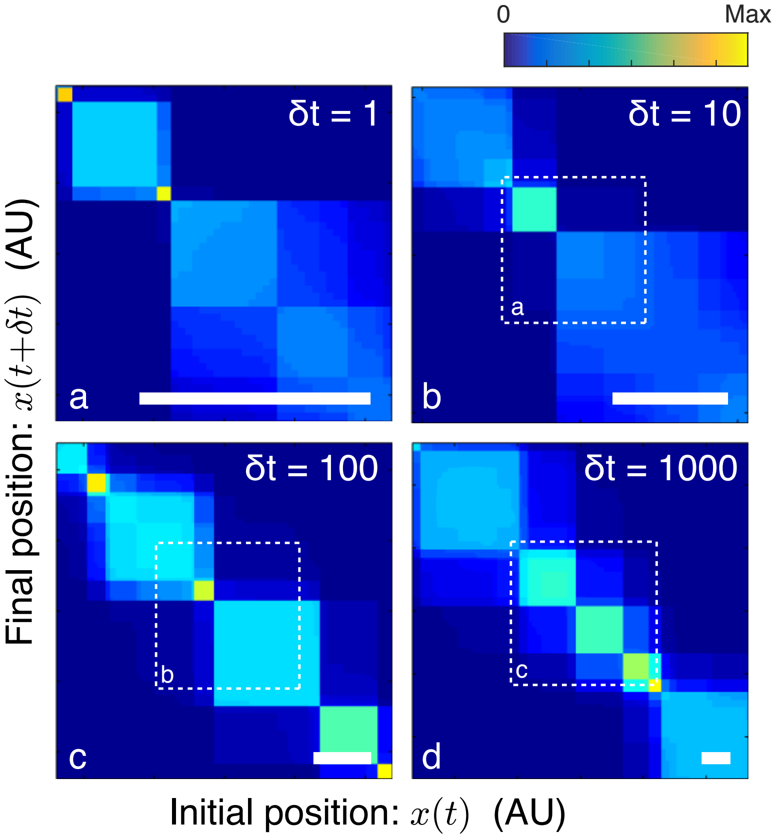


Figure 5.7: **Initial versus final position distribution.** The Green's function $p[x(t+\delta t)|x(t)]$ for $\beta = 5$ and different length scales and the corresponding rescaled time. The path of the particles gets obstructed by the largest barrier in each length scale, resulting in the visible block diagonal form.

for geometrically-increasing lag times (Fig. 5.7), while simultaneously rescaling the position x in order to keep the MSD constant. The stochastic block diagonal form of the Green's function can be observed to maintain the same characteristic size for all times (Fig. 5.7), illustrating the self-similarity. Physically, as the lag time increases, the particles can eventually jump the proximal barrier, but due to the strong disorder, they are arrested by a distal barrier with an even larger free energy. The existence of an ever larger barrier to arrest the motion is the physics that generates scale invariance.

5.3.6 The extreme value framework

The observation from simulation, that the dynamics are limited by the slowest step, motivates the use of EVT [75, 76]. The approach is schematically illustrated in Fig. 5.8: EVT theory predicts the height of the largest free-energy barrier, which grows in the displacement $\log \Delta x$ due to the *statistical multiplicity* of the stochastic barriers.

We use EVT to approximate the step-size distribution and MSD scaling. We approximate the probability for having a step larger than Δx as the probability that the largest barrier $G_{(n)} = \max_{i=1..n} G_i$ on the interval width Δx could be hopped in the lag time δt. The distribution for the maximum of n random variables G_i is easily expressed in terms of the CDF F_G for a single G:

$$\Pr\{G_{(n)} < g\} \;=\; [F_G(g)]^n, \tag{5.6}$$

where we define the number of barriers $n \equiv \Delta x / x_0$ and x_0 is the lattice spacing. In analogy to the Central Limit Theorem, there is a limiting distribution for the maximum of n independent and identically distributed random variables in the large n limit. For the free energy, the limiting distribution with the appropriate domain is the Gumbel distribution whose cumulative probability distribution is:

$$\mathrm{Gu}(g) \equiv \exp\left[-\exp\left(-\tfrac{g-\mu}{\sigma}\right)\right], \tag{5.7}$$

where the position μ and scale σ parameters depend on the distribution F_G of the random variable G. We assume this limiting form for F_G. The condition that g is the largest barrier

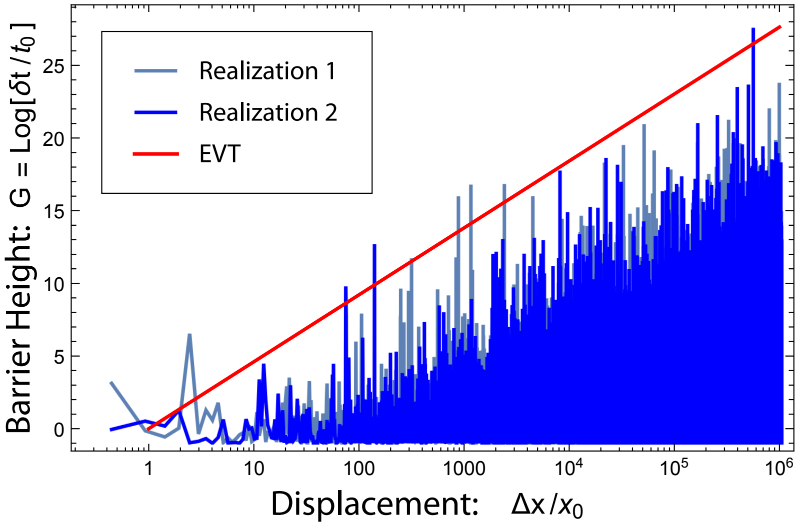


Figure 5.8: **Limiting barrier height grows with displacement.** The barrier free energies for two realization of G ($\beta = 1$) are plotted versus position on a linear-log plot. The red curve represents the extreme value prediction for the height of the maximum barrier as a function of displacement ($\Delta x/x_0$).

than can be hopped in lag time δt is $\delta t/t'_0 = \exp(g-\mu)$ where t'_0 is a time constant. Combining these approximations results in an approximate CDF for the random displacement ΔX:

$$\Pr\{|\Delta X| > |\Delta x|\} \approx \exp\left[-\tfrac{|\Delta x|}{x_0}\left(\tfrac{\delta t}{t'_0}\right)^{-1/\sigma}\right], \tag{5.8}$$

where x_0 is the lattice size. The probability distribution function is therefore a Laplace distribution, as observed. The Laplace-like step-size distribution had previously been proposed in the barrier model, but on a purely numerical basis [92].

The derivation is independent of the fine structure of the distribution of barrier heights F_G, depending only on the Gumbel parameters μ and σ. It is straightforward to derive the disorder-averaged MSD from the step-size distribution function:

$$\left\langle \Delta x^2 \right\rangle_G \propto \delta t^{2/\sigma}, \tag{5.9}$$

where there is a simple relationship between the MSD scaling exponent and the Gumbel scale parameter: $\alpha = 2/\sigma$. The detailed derivation can be found in the supplement. In the special case that the disorder is modeled by a scaled chi-squared, $\alpha = \beta^{-1}$ as was observed empirically in simulation. An analogous limiting expression has already been derived for the MSD using other methods [85].

5.4 Discussion

5.4.1 *A framework for scale invariance*

We have shown that dynamics limited by the crossing of the largest barrier produce scale-invariant phenomena, specifically MSD, step-size distribution, velocity autocorrelation and Green's function. As a result, EVT is a generic mechanism for generating scale-invariance is the strong-quenched-disorder limit, independent of any detailed assumptions about the distribution of barrier height, *etc.* With respect the analysis of the step-size distribution and MSD, there is only a single informative model parameter: the Gumbel scale parameter σ, which is inversely proportional to the MSD scaling exponent α. The parameter σ is a unitless parameter representing disorder strength and is unitless since the free energy is also unitless, when measured in thermal units. The second Gumbel position parameter absorbs the unphysical lattice spacing x_0 and time scale t'_0 and is related to the constant D in the MSD formula.

5.4.2 Applicability of the EVT model

We have made a number of formal assumptions about the model in our derivation: (i) We first assumed that the free energy landscape is quenched, or static in time. From a practical perspective, it is necessary only that the dynamics of the quenched disorder is slow compared to the observed dynamics. (ii) We also assume a strong-disorder limit in which the ensemble averaged hop time for a barrier diverges (Eqn. 5.2). This is a convenient formal statistical assumption which can be violated without significant changes in the predictions of the model. The relevant physical assumption is that dynamics is limited by the slowest step. The analysis of free energy barriers with a normal distribution is instructive for building intuition: For relatively short times, the dynamics is limited by the slowest step, but at sufficiently long times, the entropic contribution to the dynamics becomes limiting and the motion is diffusive. Therefore, if the disorder is strong enough to be limited by the slowest step, EVT-like behavior will be observed even if the average hop time is finite.

The most consequential assumption is that we have worked in one spatial dimension, along the long axis of the cell, to avoid consideration of the confinement along the short axis. The barriers are therefore interpreted as effective barrier heights between positions along the long axis. For dimension greater than one, the dynamics of random uncorrelated barrier models changes fundamentally in the long-lag-time limit: No matter how strong the disorder, large barriers can be avoided in dimension larger than one by traveling a more circuitous path. Barrier models are therefore always diffusive at *suffi-*

ciently long times (Ref. [93] and Appendix Fig. D.8). We do not believe this to be a serious shortcoming of the model for a number of reasons: (i) In practice, the crossover between EVT-type dynamics and diffusion may be extremely long. (ii) Qualitatively, the effective dynamics of these higher dimensional models tends to result in $\log p \approx -|x|^{\gamma} + ...$ with $0 < \gamma < 1$ (Appendix Fig. D.9), which is a better fit to the observed data than the Laplace distribution (Appendix Fig. D.10). (iii) Due to the structure of the nucleoid and exclusion, the motion in the cell may be effectively one-dimensional in any case. (iv) Finally this predicted crossover to diffusive dynamics at long times means that EVT-like dynamics could be ubiquitous on short times even for significantly smaller complexes, that are observed to be diffusive on the time scales probed by experiment. For example, exactly this type of crossover is in fact observed in glasses ($\dim > 1$), where the long time limit dynamics is observed to transition between sub-diffusive and diffusive, exactly as predicted qualitatively by the random barrier model [94].

5.4.3 *Realistic models*

It is important to emphasize that the proposed model is not intended to capture every aspect of the dynamics, but should be understood as an attempt to study a minimal model with the characteristics of the observed dynamics. For instance, we are convinced that in many systems there is a significant degree of quenched disorder in the diffusion coefficient. There is a natural mechanism for this quenched disorder (particle size) and there is evidence for it the context of mobility in bacterial cells [95]. Even in our own data, there is more variance in $\Delta x(t + \Delta t)$ conditioned on $\Delta x(t)$ than predicted by the model for large $\Delta x(t)$, which is consistent with some quenching in the diffusion coefficient (Fig. 5.4). In fact, the conditional probability provides a deep insight into the nature of the memory and we believe that its analysis may also be powerful in the context of other systems. It has recently been proposed that the bacterial cytoplasm behaves like a glass [95] which is broadly consistent with our approach in this paper.

5.4.4 Strong disorder is generic in biology

The EVT mechanism may predict emergent scale-invariant behavior in other systems since strong disorder appears to be ubiquitous in biology. For instance, the motion of lipids and proteins in the membrane appear to show a barrier-hopping like phenomenology. Lipids and membrane proteins appear to undergo relatively rapid motion is small confined membrane domains, then exhibit a slower hopping behavior between neighboring domains [96, 97, 98]. If the interdomain hopping rate is strongly disordered, we would expect to see EVT-like dynamics in this system. Another interesting potential application is chemical kinetics. In metabolism, reaction are often limited by the slowest step. If we treat these effective reaction rates as random variables, EVT might be used to predict the scaling of average reaction rate with reactant number. In the context of evolution, EVT has already been applied to describe fitness and beneficial mutations [78, 79, 80]. The framework may also have interesting applications beyond dynamics as well. For instance, high-resolution characterization of DNA flexibility have previously reported bending probabilities with Laplace-like distributions suggesting that similar arguments may apply in mechanics contexts as well when studying the flexibility of DNA molecules with sequence dependent flexibility [99].

5.5 Conclusion

We have demonstrated that an EVT framework can be applied understand the emergent scale-invariant properties of dynamics of large molecular complexes in the crowded cytoplasm. The framework explains how the observed behavior can be understood as the generic behavior of a system whose dynamics is limited by the slowest step, relevant for describing strongly disordered systems. The theory predicts all three observed scale-invariant phenomena: (i) MSD, (ii) step-size distribution and (iii) the velocity autocorrelation function as well as making other non-trivial predictions about the dynamics. We believe that similar arguments may apply to explain the emergent properties of many complex biological systems.

Chapter 6

CONCLUSIONS AND FUTURE OUTLOOKS

In this chapter, we discuss the main findings of the thesis and discuss limitations and possible directions for future research. Our image analysis and segmentation tool has shown benchmarking performance, however further improvements could generalize it further and make it usable in different cases. We have also revealed some exciting results on lag phase and cytoplasmic dynamics, and a model that could have applications in scale-independent phenomena in biology limited by strong disorder. We propose some experiments and analysis that could verify how general these models are.

6.1 Conclusions

6.1.1 An image analysis and segmentation tool with benchmarking performance

The error measurements and case studies of our image segmentation and analysis tool, SuperSegger, demonstrate the unique capabilities of our software. We have demonstrated that it is well suited for high-throughput time-lapse fluorescence microscopy of cells, capable of segmenting and tracking thousands of cells, and especially hundreds of cells in close contact in micro-colonies.

SuperSegger can correctly segment 99.3% of the cells in a snapshot image. But SuperSegger's performance of segmenting micro-colonies is even more impressive. In segmenting 80 datasets of initially isolated cells growing into micro-colonies, on average 174 cells were correctly segmented before the first error, and 30% of the data sets were segmented without error. At the end of the experiment, at 420 minutes, only 2.5% of the cells have segmentation errors. This is in contrast with other similar image segmentation packages that either incorrectly merge or split cells in micro-colonies even in a single image. Furthermore, the parallelization mode of SuperSegger significantly speeds up computation of multiple xy positions.

We have successfully used SuperSegger in multiple biological problems, from protein localization to cell-contact mediated phenomena. In this thesis we demonstrate its capabilities in two different case studies. For the first case study, the software was used to segment up to five generations of cells and measure the lifetimes of the cells at the single cell level. In the second case study, we used SuperSegger to identify fluorescence foci in the cells, measure their relative positions in the cell and the time in the terms of the cell cycle of the cell and calculate the displacement of foci from frame-to-frame.

We expect that our software will be of high value to many researchers in need of an image segmentation and tracking tool for fluorescence microscopy in bacteria. The multiple graphical user interfaces allow users with little programming experience to segment images, train new constants and analyze their data.

6.1.2 A model for lag phase

Our lag phase growth experiments with single cell resolution reveal subtle details about cell growth. We observe a wide distribution in the delay before the first division of stationary progenitors. We also find that 20% of cells exhibit an extremely slow transition to the growing state ($\tau > 200$ min), although some of those still transition to log phase before the end of the experiment. We propose a model of lag phase that includes a stationary-dormant cell state, in addition to the already known phases of stationary and death. Cells in the stationary-dormant state are the ones that show a significant delay before the first cell division. We believe that this state may be an intermediate state between stationary phase and cell death. Another possibility is that this state is the result of a bet-hedging strategy where a random small subset of the population of cells does not transition into log phase. This may protect the cell population on average, if the increase in nutrients is short lived cells could have transitioned into the growing state when the optimal behavior is to remain dormant.

6.1.3 Two modes of nucleoid action

Our observations, using the the MS2-mRNA system as our probe, provide strong support for the role of nucleoid exclusion in the organization and transport of macromolecules in the bacterial cytoplasm. We show a striking agreement between the qualitative shape and quantitative scale of the observed drift velocity map in the cell and a minimal statistical mechanics model of nucleoid exclusion and membrane confinement.

We also propose a new mode of nucleoid action, as a facilitator of rapid complex motion through the cytoplasm. This suggestion stems from two observations. First, in contrast with a previous hypothesis of particle sub-populations with different mobility, our data shows that the different mobility can be attributed to the location of the complexes in the cell. Specifically we observe the highest diffusion coefficients in regions with the lowest MS2-mRNA occupancy and highest nucleoid density. Second, molecular complexes in cells treated with Rifampicin, which has been found to increase nucleoid mobility, show higher diffusion coefficients. These results suggest a tight link between MS2-mRNA and chromosome dynamics, and that large rearrangements of the nucleoid, which happen during active processes, could facilitate motion in the cytoplasm.

6.1.4 Scale invariance in cytoplasmic dynamics

The observed complexes motion is scale invariant from three distinct perspectives : (i) The MSD takes the form of the power law consistent with sub-diffusion. (ii) The step-size distribution, well approximated by a Laplace distribution, is scale invariant since the distribution of steps, normalized by their standard deviation, is independent of lag time. Exponential step-size distributions are thought to be a result of a wide distribution in diffusion coefficients, which can be a result of quenched disorder in the media or transitions between states of different mobility with lifetimes comparable to the lag time of observation [68]. (iii) The velocity autocorrelation, characterizing the memory of the motion, is approximately scale-invariant since it is independent of lag time and it is negative for non-zero delay times, typically interpreted as viscoelasticity.

6.1.5 Strong disorder leads to scale invariance

The scale-independent dynamics of the complexes motivated us to seek for a general model to characterize the dynamics in the cytoplasm. We propose a minimal model of random barriers for the crowded cytoplasm, that is consistent with many aspects of the observed dynamics. We have shown that scale independent phenomena emerge from strong quenched disorder, specifically MSD, step-size distribution, velocity autocorrelation and Green's function. We demonstrated that the EVT framework from statistics can be applied to understand the underlying details of strong disorder, and that the MSD scaling factor depends only on the scale parameter of the Gumbel distribution, which is related to the disorder strength.

Our work was mainly focused in one spatial dimension. We find that in a barrier model of more than one spatial dimensions the obstacles can be avoided, and therefore the models are diffusive at sufficiently long times. In practice, this may not be a shortcoming of the model, as the crossover from EVT-type dynamics to diffusion may be extremely long and the motion in the cell may be effectively one-dimensional due to the nucleoid exclusion volume. We also note that the crossover from sub-difussion to diffusion has been observed in systems like glasses. Thus, it is possible that EVT-like dynamics could be ubiquitous on short times even for significantly smaller biological complexes, that have been observed to be diffusive on the time scales probed by experiment.

We believe that the EVT framework may apply to the emergent properties of many complex biological systems that are dominated by extreme values. For example it may be applicable to the motion of lipids and proteins in the membrane, fitness, chemical reactions limited by the slowest step and other.

6.2 Future outlooks

6.2.1 Generalized parameters for any image resolution

Our image analysis and segmentation tool has shown benchmarking performance, however further improvements could further generalize it making it viable for different use cases. The current segmentation parameters used depend on the pixel-size of the image. For every strain of different morphology or different camera resolutions new segmentation parameters need to be trained. It would be possible to avoid training new segmentation parameters for different camera resolutions by modifying the algorithm such that all quantities used during segmentation are scaled with the given pixel-size of the images.

6.2.2 Fluorescence Tracking

SuperSegger contains methods for identifying foci in fluorescence images. We also have developed methods to track foci, link foci from frame-to-frame. Linking foci from frame-to-frame is significantly more difficult than linking the cells from frame-to-frame, as foci may disappear for multiple frames due to photobleaching and then reappear, a foci may split into two foci, two foci may merge into one, or more foci are falsely detected than in the case of cells. The methods we have developed depend on the known biology of each use case, for example in the mRNA-MS2 case we do not expect the foci to split into two foci. To avoid investigator bias, a more general algorithm able to track foci independent of the biological details is needed.

6.2.3 Link between nucleoid and cytoplasmic dynamics

We believe that our observations on the dynamics of the mRNA-MS2 molecular complexes generalize to passive molecules of similar size in the cytoplasm. More experiments in other cytoplasmic molecules, such as plasmids, ribosomes etc, could verify whether our results generalize well to the dynamics of other cytoplasmic molecules. Furthermore, similar analysis in active systems, could show differences in the dynamics of passive and active systems and provide insights into the mechanisms behind active systems. For example, our analysis could be used to understand the partitioning mechanisms of plasmids.

Our hypothesis that the nucleoid action and cytoplasmic dynamics are linked could be further tested with simultaneous observation of loci in the nucleoid and the MS2-mRNA particles. This would allow one to quantify the correlation between the dynamics of the nucleoid and the cytoplasm.

BIBLIOGRAPHY

[1] N. J. Kuwada, B. Traxler, and P. A. Wiggins, "Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle," *Mol Microbiol*, vol. 95, pp. 64–79, Jan 2015.

[2] R. Ietswaart, F. Szardenings, K. Gerdes, and M. Howard, "Competing para structures space bacterial plasmids equally over the nucleoid," *PLoS Comput Biol*, vol. 10, p. e1004009, Dec 2014.

[3] J. A. Cass, N. J. Kuwada, B. Traxler, and P. A. Wiggins, "Escherichia coli chromosomal loci segregate from midcell with universal dynamics," *Biophysical Journal*, vol. 110, pp. 2597–2609, 2016/06/21 2016.

[4] S. Stylianidou, N. J. Kuwada, and P. A. Wiggins, "Cytoplasmic dynamics reveals two modes of nucleoid-dependent mobility," *Biophys J*, vol. 107, pp. 2684–92, Dec 2014.

[5] S. M. Mangiameli, B. T. Veit, H. Merrikh, and P. A. Wiggins, "The replisomes remain spatially proximal throughout the cell cycle in bacteria," *PLOS Genetics*, vol. 13, pp. 1–17, 01 2017.

[6] I. Golding and E. C. Cox, "RNA dynamics in live Escherichia coli cells," *Proc Natl Acad Sci U S A*, vol. 101, pp. 11310–5, Aug 2004.

[7] M. Krieger, "Bacteria are masters of tai-chi," 2017. `http://nautil.us/issue/37/currents/bacteria-are-masters-of-tai-chi` (visited 30-January-2017).

[8] P. H. Viollier, M. Thanbichler, P. T. McGrath, L. West, M. Meewan, H. H. McAdams, and L. Shapiro, "Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial dna replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 25, pp. 9257–9262, 2004.

[9] P. A. Wiggins, K. C. Cheveralls, J. S. Martin, R. Lintner, and J. Kondev, "Strong intranucleoid interactions organize the escherichia coli chromosome into a nucleoid filament," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 4991–4995, 2010.

[10] J. R. Maddock and L. Shapiro, "Polar location of the chemoreceptor complex in the escherichia coli cell," *Science*, vol. 259, pp. 1717–1717, 1993.

[11] D. M. Raskin and P. A. de Boer, "The mine ring: an ftsz-independent cell structure required for selection of the correct division site in e. coli," *Cell*, vol. 91, no. 5, pp. 685–694, 1997.

[12] P. B. Eckburg, E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman, "Diversity of the human intestinal microbial flora," *science*, vol. 308, no. 5728, pp. 1635–1638, 2005.

[13] F. C. Neidhardt and R. Curtiss, *Escherichia coli and Salmonella : cellular and molecular biology*. Washington, D.C.: ASM Press, 1996.

[14] K. Luby-Phelps, D. L. Taylor, and F. Lanni, "Probing the structure of cytoplasm.," *The Journal of Cell Biology*, vol. 102, no. 6, pp. 2015–2022, 1986.

[15] M. T. Madigan, J. M. Martinko, D. Stahl, and D. P. Clark, *Brock Biology of Microorganisms*. Boston : Benjamin Cummings, 13 ed., Dec. 2010.

[16] R. Phillips, J. Kondev, J. Theriot, and H. Garcia, *Physical biology of the cell*. Garland Science, 2012.

[17] S. C. Weber, A. J. Spakowitz, and J. A. Theriot, "Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm," *Physical review letters*, vol. 104, no. 23, p. 238102, 2010.

[18] L. Niu and J. Yu, "Investigating intracellular dynamics of ftsz cytoskeleton with photoactivation single-molecule tracking," *Biophysical journal*, vol. 95, no. 4, pp. 2009–2016, 2008.

[19] J. L. Slonczewski and J. W. Foster, *Microbiology: An Evolving Science: Third International Student Edition*. WW Norton & Company, 2013.

[20] J. W. Young, J. C. W. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P. S. Swain, E. Mjolsness, and M. B. Elowitz, "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy," *Nat Protoc*, vol. 7, pp. 80–8, Jan 2012.

[21] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea," *Journal of cellular and comparative physiology*, vol. 59, no. 3, pp. 223–239, 1962.

[22] M. Chalfie, "Green fluorescent protein as a marker for gene expression," *Trends in Genetics*, vol. 10, no. 5, p. 151, 1994.

[23] C. G. Galbraith and J. A. Galbraith, "Super-resolution microscopy at a glance," *J Cell Sci*, vol. 124, no. 10, pp. 1607–1611, 2011.

[24] B. Christen, M. J. Fero, N. J. Hillson, G. Bowman, S.-H. Hong, L. Shapiro, and H. H. McAdams, "High-throughput identification of protein localization dependency networks," *Proc Natl Acad Sci U S A*, vol. 107, pp. 4681–6, Mar 2010.

[25] J. M. Guberman, A. Fay, J. Dworkin, N. S. Wingreen, and Z. Gitai, "Psicic: noise and asymmetry in bacterial division revealed by computational image analysis at sub-pixel resolution," *PLoS Comput Biol*, vol. 4, p. e1000233, Nov 2008.

[26] J. C. W. Locke and M. B. Elowitz, "Using movies to analyse gene circuit dynamics in single cells," *Nat Rev Microbiol*, vol. 7, pp. 383–92, May 2009.

[27] A. Paintdakhi, B. Parry, M. Campos, I. Irnov, J. Elf, I. Surovtsev, and C. Jacobs-Wagner, "Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis," *Mol Microbiol*, vol. 99, pp. 767–77, Feb 2016.

[28] Q. Wang, J. Niemi, C.-M. Tan, L. You, and M. West, "Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy," *Cytometry A*, vol. 77, pp. 101–10, Jan 2010.

[29] M. Guizar-Sicairos, G.-S. Manuel, S. Thurman, and J. Fienup, "Efficient subpixel image registration algorithms.," *Opt. Lett.*, vol. 33, p. 156, 2008.

[30] S. Inoué and K. R. Spring, *Video Microscopy. The Fundamentals*. Plenum Press: New York and London, 2nd ed., 1997.

[31] M. LeRoux, J. A. De Leon, N. J. Kuwada, A. B. Russell, D. Pinto-Santini, R. D. Hood, D. M. Agnello, S. M. Robertson, P. A. Wiggins, and J. D. Mougous, "Quantitative single-cell characterization of bacterial interactions reveals type vi secretion is a double-edged sword," *Proc Natl Acad Sci U S A*, vol. 109, pp. 19804–9, Nov 2012.

[32] M. LeRoux, R. L. Kirkpatrick, E. I. Montauti, B. Q. Tran, S. B. Peterson, B. N. Harding, J. C. Whitney, A. B. Russell, B. Traxler, Y. A. Goo, D. R. Goodlett, P. A. Wiggins, and J. D. Mougous, "Kin cell lysis is a danger signal that activates antibacterial pathways of pseudomonas aeruginosa," *Elife*, vol. 4, 2015.

[33] M. LeRoux, S. B. Peterson, and J. D. Mougous, "Bacterial danger sensing," *J Mol Biol*, vol. 427, pp. 3744–53, Nov 2015.

[34] A. B. Russell, M. LeRoux, K. Hathazi, D. M. Agnello, T. Ishikawa, P. A. Wiggins, S. N. Wai, and J. D. Mougous, "Diverse type vi secretion phospholipases are functionally plastic antibacterial effectors," *Nature*, vol. 496, pp. 508–12, Apr 2013.

[35] J. A. Cass, S. Stylianidou, N. J. Kuwada, B. Traxler, and P. A. Wiggins, "Probing bacterial cell biology using image cytometry," *Molecular Microbiology*, vol. 103, no. 5, pp. 818–828, 2017.

[36] C. Garmendia-Torres, A. Skupin, S. A. Michael, P. Ruusuvuori, N. J. Kuwada, D. Falconnet, G. A. Cary, C. Hansen, P. A. Wiggins, and A. M. Dudley, "Unidirectional p-body transport during the yeast cell cycle," *PLoS One*, vol. 9, no. 6, p. e99428, 2014.

[37] A. Javer, N. J. Kuwada, Z. Long, V. G. Benza, K. D. Dorfman, P. A. Wiggins, P. Cicuta, and M. C. Lagomarsino, "Persistent super-diffusive motion of escherichia coli chromosomal loci," *Nat Commun*, vol. 5, p. 3854, 2014.

[38] N. J. Kuwada, B. Traxler, and P. A. Wiggins, "High-throughput cell-cycle imaging opens new doors for discovery," *Curr Genet*, vol. 61, pp. 513–6, Nov 2015.

[39] N. J. Kuwada, K. C. Cheveralls, B. Traxler, and P. A. Wiggins, "Mapping the driving forces of chromosome structure and segregation in escherichia coli," *Nucleic Acids Res*, vol. 41, pp. 7370–7, Aug 2013.

[40] T. J. Lampo, N. J. Kuwada, P. A. Wiggins, and A. J. Spakowitz, "Physical modeling of chromosome segregation in escherichia coli reveals impact of force and dna relaxation," *Biophys J*, vol. 108, pp. 146–53, Jan 2015.

[41] J. R. O'Connor, N. J. Kuwada, V. Huangyutitham, P. A. Wiggins, and C. S. Harwood, "Surface sensing and lateral subcellular localization of wspa, the receptor in a chemosensory-like system leading to c-di-gmp production," *Mol Microbiol*, vol. 86, pp. 720–9, Nov 2012.

[42] D. Madar, E. Dekel, A. Bren, A. Zimmer, Z. Porat, and U. Alon, "Promoter activity dynamics in the lag phase of escherichia coli," *BMC Syst Biol*, vol. 7, p. 136, 2013.

[43] O. Gefen, O. Fridman, I. Ronin, and N. Q. Balaban, "Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity," *Proc Natl Acad Sci USA*, vol. 111, pp. 556–61, Jan 2014.

[44] I. Levin-Reisman, O. Gefen, O. Fridman, I. Ronin, D. Shwa, H. Sheftel, and N. Q. Balaban, “Automated imaging with scanlag reveals previously undetectable bacterial growth phenotypes,” *Nature Methods*, vol. 7, no. 9, pp. 737–739, 2010.

[45] A. Jõers and T. Tenson, “Growth resumption from stationary phase reveals memory in escherichia coli cultures,” *Scientific reports*, vol. 6, 2016.

[46] M. D. Rolfe, C. J. Rice, S. Lucchini, C. Pin, A. Thompson, A. D. S. Cameron, M. Alston, M. F. Stringer, R. P. Betts, J. Baranyi, M. W. Peck, and J. C. D. Hinton, “Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation,” *J Bacteriol*, vol. 194, pp. 686–701, Feb 2012.

[47] H. Olofsson, J. Ripa, and N. Jonzén, “Bet-hedging as an evolutionary game: the trade-off between egg size and number,” *Proceedings of the Royal Society of London B: Biological Sciences*, p. rspb20090500, 2009.

[48] J.-W. Veening, W. K. Smits, and O. P. Kuipers, “Bistability, epigenetics, and bet-hedging in bacteria,” *Annu Rev Microbiol*, vol. 62, pp. 193–210, 2008.

[49] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, “Bacterial persistence as a phenotypic switch,” *Science*, vol. 305, no. 5690, pp. 1622–1625, 2004.

[50] J. L. Ptacin and L. Shapiro, “Chromosome architecture is a key element of bacterial cellular organization,” *Cellular microbiology*, vol. 15, no. 1, pp. 45–52, 2013.

[51] D. Landgraf, B. Okumus, P. Chien, T. A. Baker, and J. Paulsson, “Segregation of molecules at cell division reveals native protein localization,” *Nature methods*, vol. 9, no. 5, pp. 480–482, 2012.

[52] A. B. Lindner, R. Madden, A. Demarez, E. J. Stewart, and F. Taddei, “Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 8, pp. 3076–3081, 2008.

[53] K. C. Huang, R. Mukhopadhyay, and N. S. Wingreen, “A curvature-mediated mechanism for localization of lipids to bacterial poles,” *PLoS Comput Biol*, vol. 2, no. 11, p. e151, 2006.

[54] M. Howard, “A mechanism for polar protein localization in bacteria,” *Journal of molecular biology*, vol. 335, no. 2, pp. 655–663, 2004.

[55] J. Winkler, A. Seybert, L. König, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk, and B. Bukau, "Quantitative and spatio-temporal features of protein aggregation in escherichia coli and consequences on protein quality control and cellular ageing," *The EMBO journal*, vol. 29, no. 5, pp. 910–923, 2010.

[56] S. Saberi and E. Emberly, "Chromosome driven spatial patterning of proteins in bacteria," *PLoS Comput Biol*, vol. 6, no. 11, p. e1000986, 2010.

[57] S. Saberi and E. Emberly, "Non-equilibrium polar localization of proteins in bacterial cells," *PloS one*, vol. 8, no. 5, p. e64075, 2013.

[58] A. Sanamrad, F. Persson, E. G. Lundius, D. Fange, A. H. Gynnå, and J. Elf, "Single-particle tracking reveals that free ribosomal subunits are not excluded from the escherichia coli nucleoid," *Proceedings of the National Academy of Sciences*, vol. 111, no. 31, pp. 11413–11418, 2014.

[59] S. Jun and A. Wright, "Entropy as the driver of chromosome segregation," *Nature Reviews Microbiology*, vol. 8, no. 8, pp. 600–607, 2010.

[60] B. Youngren, H. J. Nielsen, S. Jun, and S. Austin, "The multifork escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer," *Genes & development*, vol. 28, no. 1, pp. 71–84, 2014.

[61] A. Gupta, J. Lloyd-Price, R. Neeli-Venkata, S. M. Oliveira, and A. S. Ribeiro, "In vivo kinetics of segregation and polar retention of ms2-gfp-rna complexes in escherichia coli," *Biophysical journal*, vol. 106, no. 9, pp. 1928–1937, 2014.

[62] A.-S. Coquel, 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 Comput Biol*, vol. 9, no. 4, p. e1003038, 2013.

[63] B. R. Parry, 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*, vol. 156, no. 1, pp. 183–194, 2014.

[64] J. Lloyd-Price, 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*, vol. 194, no. 7, pp. 1747–1752, 2012.

[65] J.-P. Bouchaud and A. Georges, "Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications," *Physics Reports*, vol. 195, no. 4 & 5, pp. 127–293, 1990.

[66] S. C. Weber, A. J. Spakowitz, and J. A. Theriot, "Nonthermal atp-dependent fluctuations contribute to the in vivo motion of chromosomal loci," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7338–7343, 2012.

[67] Z. Gitai, N. A. Dye, A. Reisenauer, M. Wachi, and L. Shapiro, "Mreb actin-mediated segregation of a specific region of a bacterial chromosome," *Cell*, vol. 120, no. 3, pp. 329–341, 2005.

[68] B. Wang, J. Kuo, S. C. Bae, and S. Granick, "When brownian diffusion is not gaussian," *Nature materials*, vol. 11, no. 6, pp. 481–485, 2012.

[69] N. Muramatsu and A. P. Minton, "Tracer diffusion of globular proteins in concentrated protein solutions," *Proceedings of the National Academy of Sciences*, vol. 85, no. 9, pp. 2984–2988, 1988.

[70] M. E. Fisher, "Renormalization group theory: Its basis and formulation in statistical physics," *Reviews of Modern Physics*, 1998.

[71] R. Pathria and P. Beale, *Statistical Mechanics*. Elsevier Science, 1996.

[72] G. B. West, J. H. Brown, and B. J. Enquist, "A general model for the origin of allometric scaling laws in biology," *Science*, vol. 276, no. 5309, pp. 122–126, 1997.

[73] M. Hegreness, N. Shoresh, D. Hartl, and R. Kishony, "An equivalence principle for the incorporation of favorable mutations in asexual populations," *Science*, vol. 311, pp. 1615–7, Mar 2006.

[74] R. Milo and R. Phillips, *Cell Biology by the Numbers*. Garland Science, 2015.

[75] S. I. Resnick, *Extreme values, regular variation and point processes*. Springer, 2013.

[76] M. Leadbetter, G. Lindgren, and H. Rootze, *Extremes and related properties of random sequences and processes*. Springer-Verlag, 1983.

[77] R. Phillips, J. Kondev, and J. Theriot, *Physical Biology of the Cell*. Garland Science, 2008.

[78] S. A. Frank, "Generative models versus underlying symmetries to explain biological pattern," *Journal of evolutionary biology*, vol. 27, no. 6, pp. 1172–1178, 2014.

[79] H. A. Orr, "The population genetics of beneficial mutations," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 365, no. 1544, pp. 1195–1201, 2010.

[80] P. Joyce, D. R. Rokyta, C. J. Beisel, and H. A. Orr, "A general extreme value theory model for the adaptation of dna sequences under strong selection and weak mutation," *Genetics*, vol. 180, no. 3, pp. 1627–1643, 2008.

[81] P. Massignan, C. Manzo, J. Torreno-Pina, M. García-Parajo, M. Lewenstein, and G. Lapeyre Jr, "Nonergodic subdiffusion from brownian motion in an inhomogeneous medium," *Physical review letters*, vol. 112, no. 15, p. 150603, 2014.

[82] C. Manzo, J. A. Torreno-Pina, P. Massignan, G. J. Lapeyre Jr, M. Lewenstein, and M. F. G. Parajo, "Weak ergodicity breaking of receptor motion in living cells stemming from random diffusivity," *Physical Review X*, vol. 5, no. 1, p. 011021, 2015.

[83] T. J. Lampo, S. Stylianidou, M. P. Backlund, P. A. Wiggins, and A. J. Spakowitz, "Cytoplasmic rna-protein particles exhibit non-gaussian subdiffusive behavior," *Biophysical Journal*, 2017.

[84] S. B. Zimmerman and A. P. Minton, "Macromolecular crowding: biochemical, biophysical, and physiological consequences," *Annual review of biophysics and biomolecular structure*, vol. 22, no. 1, pp. 27–65, 1993.

[85] S. Alexander, J. Bernasconi, W. Schneider, and R. Orbach, "Excitation dynamics in random one-dimensional systems," *Reviews of Modern Physics*, vol. 53, no. 2, p. 175, 1981.

[86] J. Machta, "Random walks on site disordered lattices," *Journal of Physics A: Mathematical and General*, vol. 18, no. 9, p. L531, 1985.

[87] J.-P. Bouchaud, A. Georges, and P. Le Doussal, "Diffusion anormale dans les milieux désordonnés : piégeage, corrélations et théorèmes de la limite centrale,," *Journal de physique*, vol. 48, no. 11, pp. 1855–1860, 1987.

[88] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[89] D. T. Gillespie, "Stochastic simulation of chemical kinetics," *Annu. Rev. Phys. Chem.*, vol. 58, pp. 35–55, 2007.

[90] S. Mehraeen, N. Cordella, J. S. Yoo, and A. J. Spakowitz, "Impact of defect creation and motion on the thermodynamics and large-scale reorganization of self-assembled clathrin lattices," *Soft Matter*, vol. 7, no. 19, pp. 8789–8799, 2011.

[91] R. Zwanzig, "Diffusion in a rough potential," *Proc Natl Acad Sci U S A*, vol. 85, pp. 2029–2030, 1998.

[92] J. Bernasconi, W. Schneider, and W. Wyss, "Diffusion and hopping conductivity in disordered one-dimensional lattice systems," *Zeitschrift für Physik B Condensed Matter*, vol. 37, no. 2, pp. 175–184, 1980.

[93] S. Alexander, "Anomalous transport properties for random-hopping and random-trapping models," *Physical Review B*, vol. 23, no. 6, p. 2951, 1981.

[94] E. R. Weeks, J. C. Crocker, A. C. Levitt, A. Schofield, and D. A. Weitz, "Three-dimensional direct imaging of structural relaxation near the colloidal glass transition," *Science*, vol. 287, no. 5453, pp. 627–631, 2000.

[95] B. R. Parry, 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*, vol. 156, pp. 183–94, Jan 2014.

[96] H. Feinberg, D. A. Mitchell, K. Drickamer, and W. I. Weis, "Structural basis for selective recognition of oligosaccharides by dc-sign and dc-signr," *Science*, vol. 294, no. 5549, pp. 2163–2166, 2001.

[97] J.-B. Masson, P. Dionne, C. Salvatico, M. Renner, C. G. Specht, A. Triller, and M. Dahan, "Mapping the energy and diffusion landscapes of membrane proteins at the cell surface using high-density single-molecule imaging and bayesian inference: application to the multiscale dynamics of glycine receptors in the neuronal membrane," *Biophysical journal*, vol. 106, no. 1, pp. 74–83, 2014.

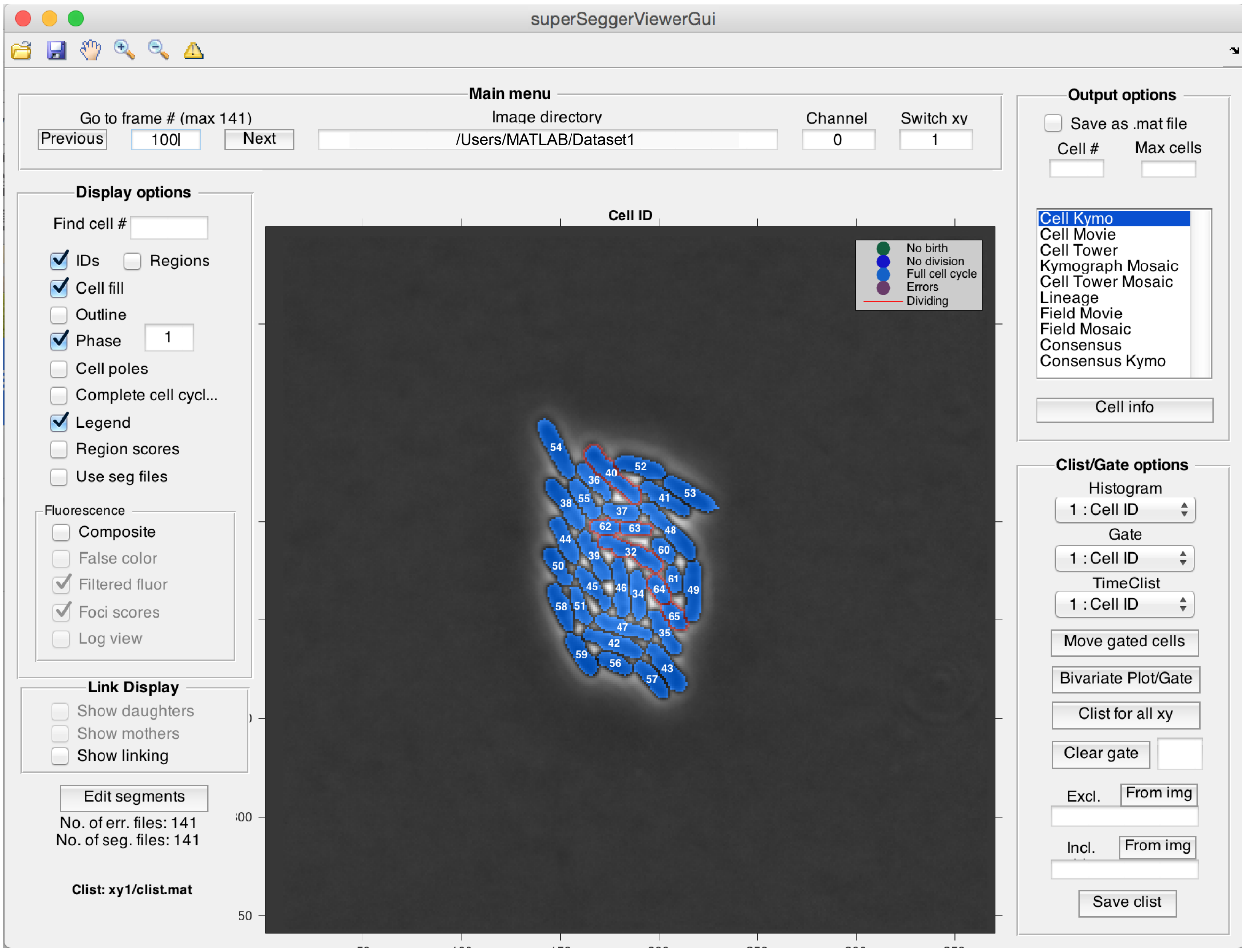
[98] A. Sergé, N. Bertaux, H. Rigneault, and D. Marguet, "Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes," *Nature methods*, vol. 5, no. 8, pp. 687–694, 2008.

[99] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, "High flexibility of dna on short length scales probed by atomic force microscopy," *Nat Nanotechnol*, vol. 1, pp. 137–41, Nov 2006.

Appendix A

SUPERSEGGER APPENDIX

A.1 Figures



Go to frame # (max 141): 100

Image directory: /Users/MATLAB/Dataset1

Channel: 0

Switch xy: 1

Find cell #:

☑ IDs ☐ Regions

☑ Cell fill

☐ Outline

☑ Phase 1

☐ Cell poles

☐ Complete cell cycl...

☑ Legend

☐ Region scores

☐ Use seg files

☐ Composite

☐ False color

☑ Filtered fluor

☑ Foci scores

☐ Log view

☐ Show daughters

☐ Show mothers

☐ Show linking

No. of err. files: 141

No. of seg. files: 141

Clist: xy1/clist.mat

☐ Save as .mat file

Cell #:

Max cells:

Histogram: 1 : Cell ID

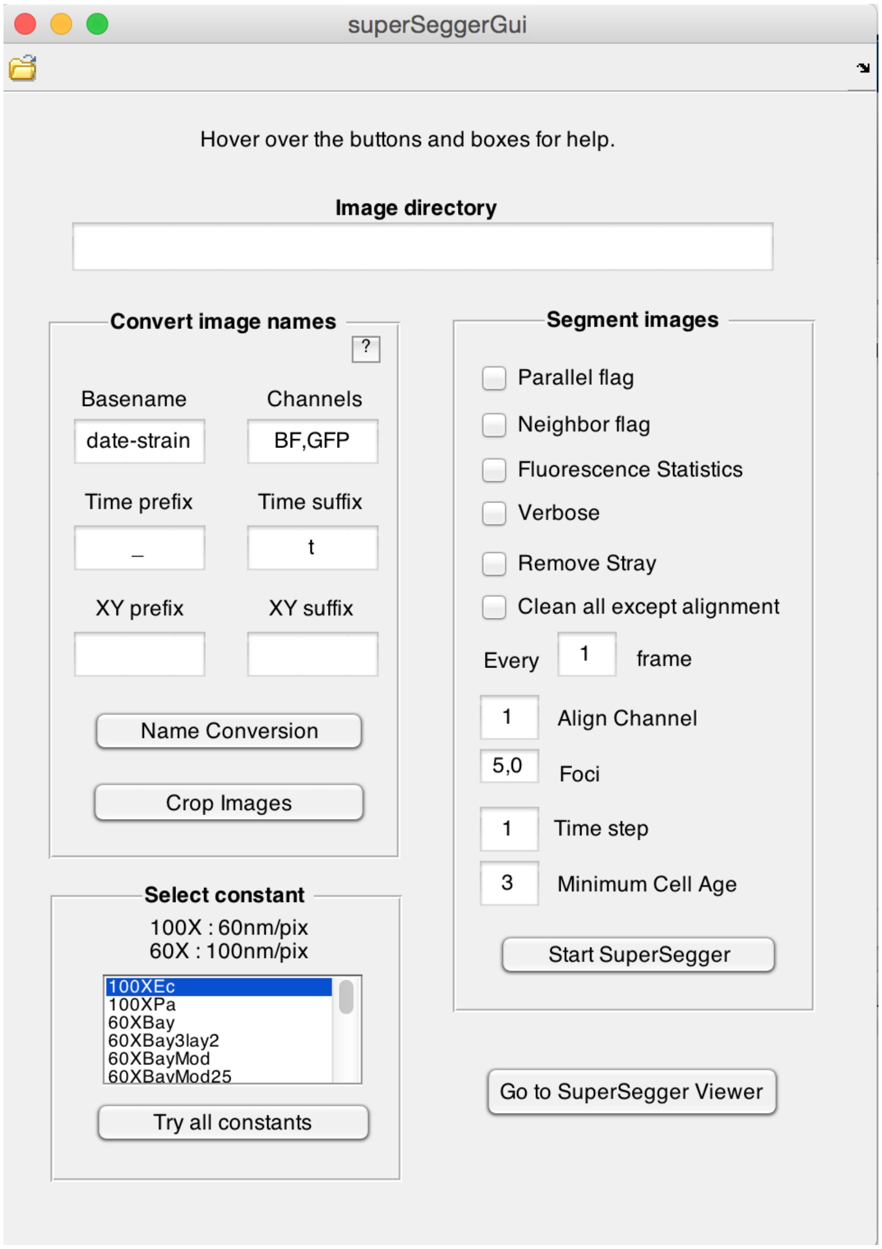
Gate: 1 : Cell ID

TimeClist: 1 : Cell ID

Excl.:

Incl.:

Figure A.1: **Graphical user interface for post-processing analysis.** superSeggerViewerGui provides different modes of visualization of segmented data and data manipulation and analysis tools such as gating, histograms, kymographs, movies, cell towers and consensus images.



Hover over the buttons and boxes for help.

Image directory

Convert image names

Basename	Channels
date-strain	BF,GFP
Time prefix	Time suffix
_	t
XY prefix	XY suffix

Select constant

100X : 60nm/pix
60X : 100nm/pix

- 100XEc
- 100XPa
- 60XBay
- 60XBay3lay2
- 60XBayMod
- 60XBayMod25

Segment images

☐ Parallel flag
☐ Neighbor flag
☐ Fluorescence Statistics
☐ Verbose
☐ Remove Stray
☐ Clean all except alignment

Every 1 frame

1 Align Channel
5,0 Foci
1 Time step
3 Minimum Cell Age

Figure A.2: **Graphical user interface for image segmentation.** superSeggerGui contains pre-processing tools for renaming and cropping your images, testing the different constants on one image and can be used to initiate image processing.

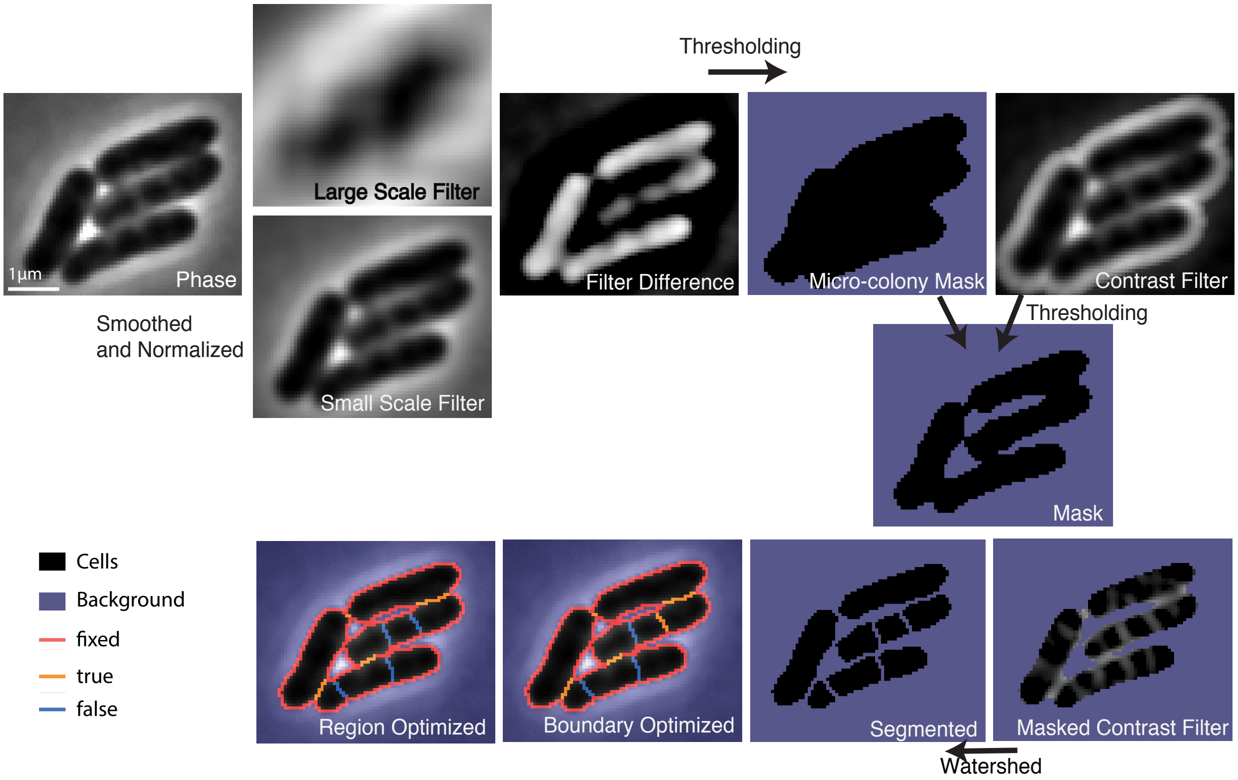


Figure A.3: **Detailed image segmentation procedure.** The phase image is smoothed and normalized. Next, two gaussian filters of different scales are applied to the image. The difference of the two filters results in an image where cell clusters have high relative intensity. The initial micro-colony mask is created by a thresholding operation on this image. To further refine the micro-colony mask we filter the image with a contrast filter and apply another thresholding operation. The result combined with the micro-colony mask is the final mask shown. The outline of the mask of each micro-colony is set as a fixed cell boundary during the rest of the segmentation procedure. To identify the boundaries between cells we enhance the contrast and apply a watershed operation to the image. The boundaries are then classified as true (orange) or false (blue) using a neural network trained on examples of boundaries. Further optimization of the boundaries takes place during region optimization using a neural network trained on examples of correctly shaped regions. In this example there is a boundary incorrectly set as true after boundary optimization which is set to false after region optimization.

A.2 Software Availability and Documentation

The website for the software can be found at `http://mtshasta.phys.washington.edu/website/SuperSegger.php` and the software can be downloaded at the GitHub repository `https://github.com/wiggins-lab/SuperSegger/`. At the time of writing, the version of SuperSegger is v.1.1. The GitHub wiki contains tutorials on how to use the software, information about the fields of the output, and a general overview of the methods (`https://github.com/wiggins-lab/SuperSegger/wiki`). In the wiki, the *Segmenting your images* section has a tutorial on how to start image segmentation. The *SuperSeggerViewer* section explains the post processing and image visualization tools of *SuperSeggerViewerGui*. The Output section contains all the field definitions of the clist, cell files and frame files. For more information on the code, all available methods and their dependencies can be found at `http://mtshasta.phys.washington.edu/website/superSegger/`.

A.3 Starting with SuperSegger

The *Segmenting your images* section in the wiki contains a full tutorial on how to segment your images. In short, to load the GUI to segment your images the user types superSeggerGui at the command line (Fig A.2). Click on the folder icon to select the folder containing the images. The filenames of the images need to follow the NIS-elements file naming convention:

`[basename]t`[num]`,xy`[num]`c`[num]`.tif`,

where *t* is followed by the time frame, *xy* by the xy position and *c* by the channel number. A script is provided to rename your filenames. Once your filenames have the NIS-Elements convention, choose the segmentation parameters from the list that best match your cells and pixelsize, select the options that best match your analysis (for example if you need fluorescence statistics or neighbor identification) and click ”Start SuperSegger”.

Once the data processing has finished click ”Go to SuperSegger Viewer” to see the results and analyze your data (Fig. A.1).

A.4 Training your own segmentation parameters

The *Creating your own constants* section in the wiki contains a tutorial on how to train your own segmentation parameters. As mentioned, we rarely train new segmentation parameters, but in some cases better results can be obtained by retraining the software. The current version of SuperSegger (v.1) depends on pixel size, and we provide parameters for 100 nm and 60 nm pixels, which are the most common pixel sizes due to the diffraction limit. Users that use a different pixel size will either need to retrain the software or resize their images to 100 nm or 60 nm pixels. Different species, or mutants with extremely abnormal morphologies may also require retraining. The GUI allows the user to set some of the segmentation parameters that are used to further optimize segmentation, such as the maximum length and width of the cells, intensity thresholds, and minimum area of regions. To train new segmentation parameters we usually use about 50 images from a time-lapse of a single cell growing to a micro-colony or 10 images of a full field of view with a variety of cell densities.

A.5 Tables

Id	Definition
1	Cell ID
2	Region num birth
3	Region num death
4	Cell birth time
5	Cell death time
6	Cell age
7	Old pole age
8	Error frame
9	stat0
10	Long axis (L) birth
11	Long axis (L) death
12	Short axis birth
13	Short axis death
14	Area birth
15	Area death
16	Region score birth
17	Region score death
18	X position birth
19	Y position birth
20	Fluor1 sum
21	Fluor1 mean
22	Fluor2 sum
23	Fluor2 mean

24	Num of neighbors
25	Region gray val
26	Focus1(1) long axis birth
27	Focus1(1) short axis birth
28	Focus1(1) score birth
29	Focus1(1) intensity birth
30	Focus1(2) long axis birth
31	Focus1(2) short axis birth
32	Focus1(2) score birth
33	Focus1(2) intensity birth
34	Focus1(3) long axis birth
35	Focus1(3) short axis birth
36	Focus1(3) score birth
37	Focus1(3) intensity birth
38	Focus1(4) long axis birth
39	Focus1(4) short axis birth
40	Focus1(4) score birth
41	Focus1(4) intensity birth
42	Focus1(5) long axis birth
43	Focus1(5) short axis birth
44	Focus1(5) score birth
45	Focus1(5) intensity birth
46	Focus1(1) long axis pole align
47	Focus1(1) long axis norm pole align
48	Focus1(1) long axis normalized

49	Focus1(2) long axis normalized
50	Focus1(3) long axis normalized
51	Focus1(4) long axis normalized
52	Focus1(5) long axis normalized
53	Focus1(1) short axis normalized
54	Focus1(2) short axis normalized
55	Focus1(3) short axis normalized
56	Focus1(4) short axis normalized
57	Focus1(5) short axis normalized
58	Focus1(1) gaussianFitWidth
59	Focus1(2) gaussianFitWidth
60	Focus1(3) gaussianFitWidth
61	Mother ID
62	Daughter1 ID
63	Daughter2 ID
64	dL max
65	dL min
66	L death / L birth
67	Fluor1 sum death
68	Fluor1 mean death
69	Fluor2 sum death
70	Fluor2 mean death
71	Focus1(1) long axis death
72	Focus1(1) short axis death
73	Focus1(1) score death

74	Focus1(1) intensity death
75	Focus1(2) long axis death
76	Focus1(2) short axis death
77	Focus1(2) score death
78	Focus1(2) intensity death
79	Focus1(2) long axis death
80	Focus1(3) short axis death
81	Focus1(3) score death
82	Focus1(3) intensity death
83	Focus1(4) long axis death
84	Focus1(4) short axis death
85	Focus1(4) score death
86	Focus1(4) intensity death
87	Focus1(5) long axis death
88	Focus1(5) short axis death
89	Focus1(5) score death
90	Focus1(5) intensity death
91	Focus1(1) gaussian fit width death
92	Focus1(2) gaussian fit width death
93	Focus1(3) gaussian fit width death
94	Long axis/Short axis birth
95	Long axis/Short axis death
96	Neck width
97	Maximum width
98	Cell dist to edge

Table A.1: Clist Descriptors

Field name	Definition
CellA	a cell array with frame-by-frame information (details below)
death	frame at which the cell divided or disappeared
birth	frame at which the cell was born or appeared
divide	1 if successful division was observed
sisterID	sister cell ID
motherID	mother cell ID
daughterID	daughter cells ID
ID	cell ID
neighbors	IDs of cells neighboring
stat0	0: birth was not observed, 1: division was not observed or error, 2: Full cell cycle
ehist	history of errors, sum of all errors
contactHist	number of neighboring cells

Table A.2: Cell file field definitions.

CellA.	Definition
xx	coordinates of padded bounding box in x-axis
yy	coordinates of padded bounding box in y-axis
mask	logical image of the cell mask, not oriented
r_offset	global coordinates of the top left edge of the cell's bounding box, useful for plotting quantities at global coordinates on the mask, phase and fluorescence with the size of the bounding box
BB	start and end coordinates of padded cell bounding box
edgeFlag	1 if cell is at the edge of the image
phase	cropped phase image of the cell
coord	structure with coordinates, area, and orientation
length	(1) length and (2) width of the box surrounding mask
cellLength	(1) max length and (2) mean width of the cell
pole	orientation of the cell pole, see below
fluor1	image for the first fluorescence channel (fluor2 if there is a 2nd channel)
fuor1mm	min and max of fluorescence channel 1 (fluor2mm if there is a 2nd channel)
fl1	statistics of fluor1, see below (fl2 if there is a 2nd channel)
cell_dist	distance to the edge of the colony
gray	average phase gray value in cell region
locus1	Focus fitting for channel 1, see below (locus2 if there is a second channel)
r	global coordinates of cell centroid (midpoint of cell)
error	segmentation error list
ehist	ehist is the sum of all errors in the region's history
contactHist	1 if cell is in contact with other cell
stat0	0 if cell birth was not observed, 1 if cell birth was observed but no good division at the end of the lifetime, 2 if good cell birth and division were observed.

Table A.3: Field definitions of CellA structure in cell files.

coord.	Definition
A	Area of cell mask
r_center	geometrical center of the cell (global coordinates)
box	coordinates of box surrounding cell (along e1 and e2 with length and width of the cell)
xaxis	coordinates of major principal axis (end and start point)
yaxis	coordinates of minor principal axis (end and start point)
I	Moment of inertia of cell mask
e1	major principal axis unit vector
e2	minor principal axis unit vector
rcm	position of center of mass for the cell mask (global coordinates)

Table A.4: Field definitions of CellA.coord structure in the cell files.

pole.	Definition
e1	major principal axis unit vector
op_ori	1 if old pole is in the direction of e1, -1 if the old pole is in the opposite direction.
op_age	age of old pole in cell cycles, NaN if no birth is observed
np_age	age of new pole in cell cycles

Table A.5: Field definitions of CellA.pole structure in the cell files.

locus1.	Definition
r	locus position (global coordinates)
score	score from focus fitting (Intensity / Intensity standard deviation)
intensity	raw intensity
shortaxis	focus position along the cell's short axis in local coordinates (0 is the cell center)
longaxis	focus position along the cell's long axis in local coordinates (0 is the cell center)

Table A.6: Definitions of fields in CellA.locus1 structure in the cell files.

fl1	Definition
sum	sum of fluorescence
r	the coordinates of the center of mass of the fluorescence
Ixx	2nd moment of fluorescence along X
Iyy	2nd moment of fluorescence along Y
Ixy	2nd moment of fluorescence along XY

Table A.7: Field definitions of CellA.fl structure for the fluorescence in the cell files.

Appendix B

LAG PHASE ANALYSIS APPENDIX

B.1 Figures

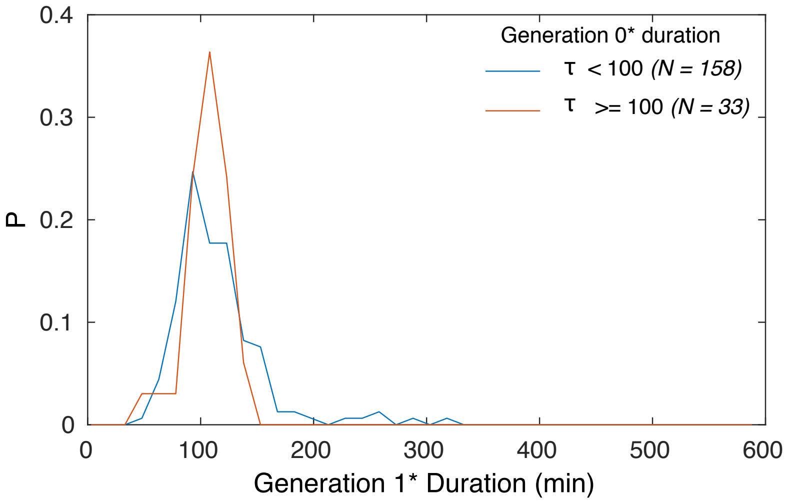


Figure B.1: **Duration of first generation of cells from stationary and non stationary progenitors.** Duration of generation 1* cells originating from progenitors with short ($\tau < 100$) and long ($\tau \geq 100$) generation 0* duration. The samples are not statistically different, indicating no memory effect on the duration of generation 1*.

B.2 Statistical Methods

In the lag analysis experiment, we run the two-sample Kolmogorov-Smirnov test to verify that the distributions of the two populations (stationary and log progenitor cells) are statistically different. The null hypothesis, that the samples are drawn from the same distribution, is rejected for all generations, with a p-value of 5×10^{-36} for the 0^* generation.

We also use the two-sample Kolmogorov-Smirnov test to test for a memory effect in the duration generation 1. Our two samples are the ages of cells with generation 1, that were observed for a full cell cycle and originate from a generation 0 with cell cycle duration < 100 minutes or ≥ 100 minutes. The null hypothesis, that the samples are drawn from the same distribution, was not rejected with a p-value of 0.0664.

Appendix C

CYTOPLASMIC DYNAMICS APPENDIX

C.1 Figures

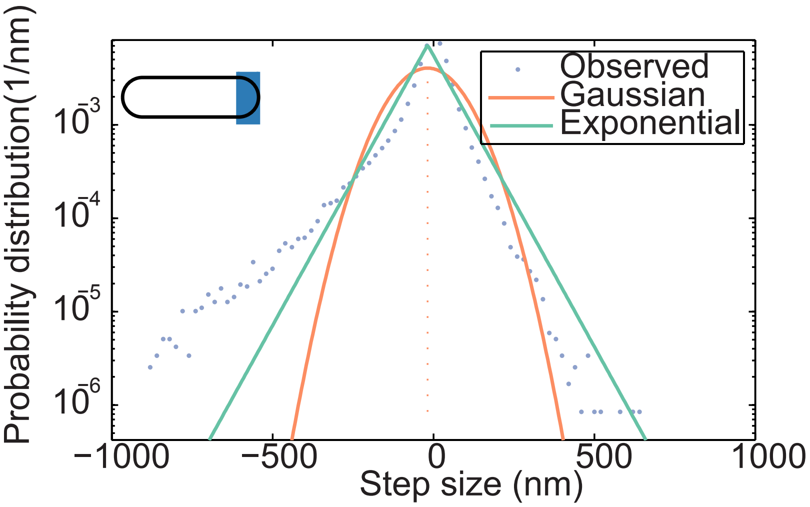


Figure C.1: **Step-size distribution has exponential tails.** The step-size distribution (using the relative velocity model) is shown for MS2mRNA complexes for a 1 minute lag time for complexes with starting position of their displacement vector near the old pole of the cell 1/5 of the cell), as illustrated in the inset. The data (blue circles) are compared with two models for the step-size distribution: Exponential (green line) and (ii) Gaussian (orange line). Both models have the same mean and variance as the experimental data. The observed step-size distribution shows the best agreement with an Exponential Model with tails of different decay constants because large steps towards the pole are mitigated by membrane exclusion forces from the pole.

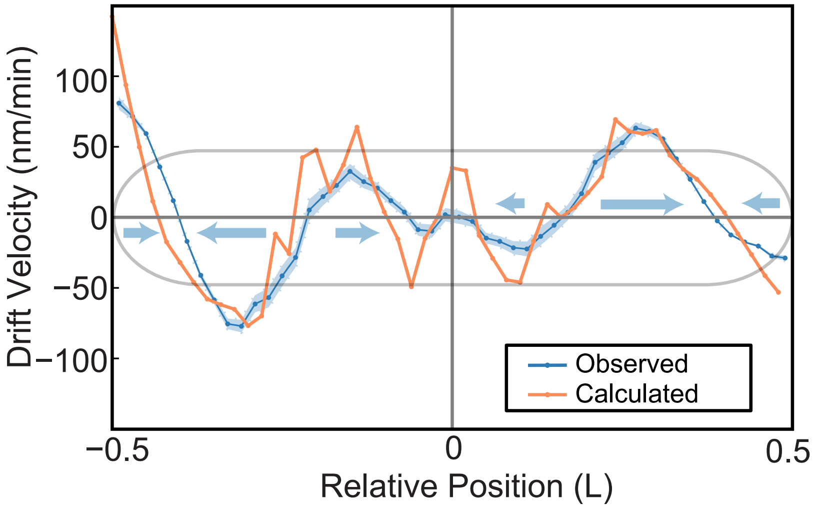


Figure C.2: **Agreement with Fick's law.** The calculated drift velocity profile using Eq. C.1 assuming Fick's Law shows remarkable agreement with the observed drift velocity profile in the cell.

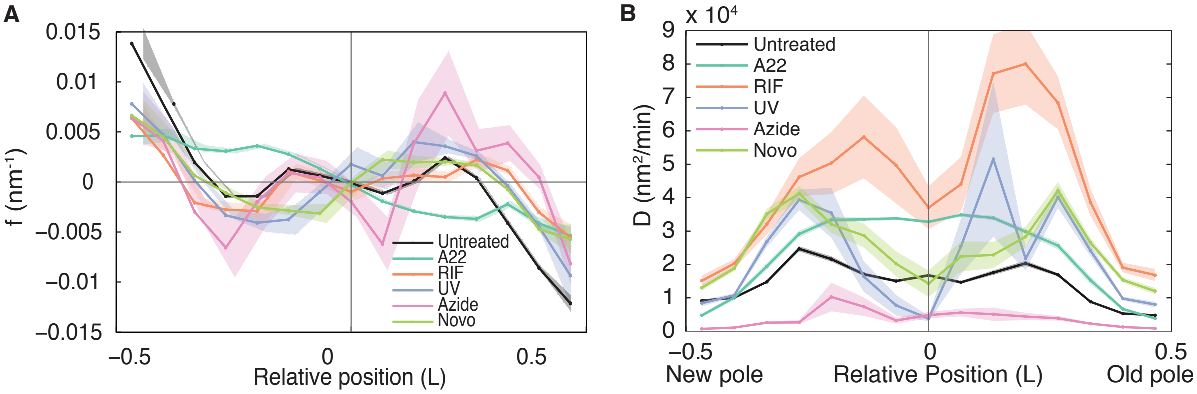


Figure C.3: **Forcing ratio and Diffusion Coefficient of treated and untreated cells. Panel A:** Forcing ratio of untreated and treated cells. Forcing ratio (Drift Velocity / Diffusion Coefficient) of MS2-mRNA complexes with respect to the relative position in the cell in untreated, ATP-depleted (Azide), DNA-damaged (UV), gyrase inhibited (Novobiocin), transcription inhibited (Rifampicin) and spherical cells (A22). Inhibiting ATP hydrolysis and metabolic activity does not affect the shape of the forcing ratio curve. **Panel B:** Spatial Dependence of diffusion coefficient of untreated and treated cells. Diffusion Coefficient of MS2-mRNA complexes with respect to the relative position in the cell in untreated, ATP-depleted (Azide), DNA-damaged (UV), gyrase inhibited (Novobiocin), transcription inhibited (Rifampicin) and spherical cells (A22). Inhibiting ATP hydrolysis and metabolic activity decreases the complexes mobility whereas in all other cases the diffusion coefficient dramatically increased. The highest diffusion coefficient appears to be in the nucleoid occupied regions, especially at 1/4 cell length.

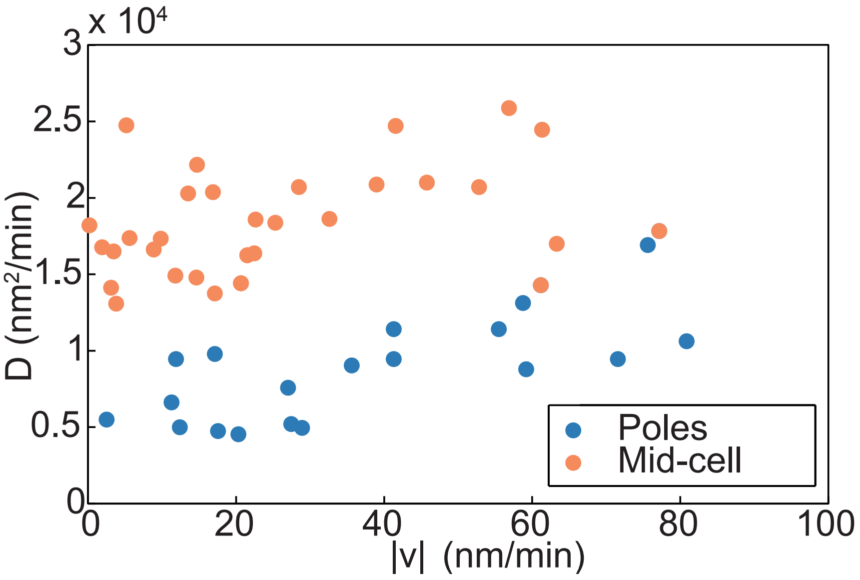


Figure C.4: **Scatter plot of diffusion coefficient versus drift velocity.** The diffusion coefficient plotted against the magnitude of the drift velocity for each bin of sub-cellular position shows two subpopulations of particles, with fast and slow dynamics, according to their cellular position. Both regions show a similar distribution of drift velocities, but particles at the poles generally exhibit lower diffusion coefficients whereas particles at middle and quarter cell regions (Midcell) exhibit higher diffusion coefficients.

C.2 Agreement between observed drift velocity and drift velocity from Fick's law support that the system is in steady state

As a test whether or not steady state statistical mechanics is an appropriate approximation for the dynamics of molecules in the cytoplasm, we calculated a theoretical drift velocity profile using Fick's Law. Under steady state conditions, the net the flux of complexes should be zero, thus

$$J = 0 = n_{\mathrm{mRNA}} v_{\mathrm{drift}} - D_{\mathrm{mRNA}} \boldsymbol{\nabla} n_{\mathrm{mRNA}} \tag{C.1}$$

where n_{mRNA} is the number density and DmRNA is the diffusion coefficient of MS2mRNA molecules, respectively. The number density was calculated from the full spatial distribution of complexes and a spatially dependent diffusion coefficient was estimated using the local mobility of complexes. The calculated drift velocity profile from Eq. C.1 shows remarkable overlap with experimental data, supporting the use of the Langevin approximation.

C.3 Derivation of the excluded volume of the nucleoid model

The cytoplasm is modeled as a lattice where nucleoid DNA and MS2-mRNA cannot occupy the same site. We treat the nucleoid as a gas of independent blobs with local number density n_N . The relative reduction in the number of nucleoid configurations as a results of opening up volume ΔV for a protein complex is

$$Z(V) \approx Z(0) \exp(-n_N \Delta V) \tag{C.2}$$

where Z(0) is the partition function in the absence of a complex volume ΔV . (Note that this can be understood as a consequence of the p dV term from the Helmholtz Free Energy.) Therefore, given an external number density n_N , we expect the number density of the mRNA complexes to be

$$n_{\mathrm{mRNA}} \propto Z(\Delta V) \propto \exp(-n_N \Delta V) \tag{C.3}$$

For convenience, we normalize the observed DNA number density to one and we will assume the nucleoid blob density is related to the DNA density,

$$n_N = N_N n_{\mathrm{DNA}} \tag{C.4}$$

where the constant of proportionality is the number of blobs N_N . We now define a unitless exclusion fraction:

$$\Phi = \frac{N_N \Delta V}{V_{\mathrm{cell}}} \tag{C.5}$$

where is the number of nucleoid blobs multiplied by the fraction of the cell volume excluded by the protein complex. To understand the meaning of Φ it is useful to express it in terms of the total volume excluded by the crowding agent (DNA). We rewrite the volume of the MS2 complex (ΔV) in terms of the volume of an individual crowding agent δV where

$$\Delta V = r\delta V \tag{C.6}$$

and r is the number of statistically independent crowing agents excluded by an MS2-mRNA complex. Φ can then be written in terms of the total excluded volume of the nucleoid ($V_N = N_N \delta V$) :

$$\Phi = \frac{V_N}{V_{\mathrm{cell}}} r \tag{C.7}$$

where Φ depends on both the fraction of volume excluded and the number of statistically independent crowding agents excluded by the MS2 complex. A closely related result is derived in more detail in Ref [16], Chapter 14 by an alternative method. We note that r depends on the structure of the nucleoid and is therefore not know a priori even though the volume excluded by the chromosome is widely believed to be roughly 10 - 20% of the cell [57]. Therefore the unitless constant must be fit to the data. We can now rewrite the number density of the mRNA

$$n_{\mathrm{mRNA}} \propto Z(\Delta V) \propto \exp(-\Phi V_{\mathrm{cell}} n_{\mathrm{DNA}}(x)) \tag{C.8}$$

where the model for the mRNA number density is parameterized by a single parameter, the exclusion fraction Φ. We can use this relationship between the number density of mRNA and DNA to obtain a value for the unitless parameter $\Phi = 1.9$. Since the factor r can be quite large (e.g. [16]), this value appears to be consistent with our proposed model. It follows that the free energy associated with the excluded volume of the chromosome regions can be approximated as

$$G = -k_bT \log Z = -k_bT \log \exp(-\Phi V_{\text{cell}} n_{\text{DNA}}) \tag{C.9}$$

We can estimate the effective external force caused by the chromosome unattainable regions to be

$$F = -\boldsymbol{\nabla} G = -k_bT \Phi V_{\text{cell}} \boldsymbol{\nabla} n_{\text{DNA}} \tag{C.10}$$

Finally, the contribution of the DNA excluded volume to the drift velocity can be approximated to be

$$\langle v \rangle (x,t) = \frac{\langle F \rangle}{\gamma} = -\Phi V_{\text{cell}} D(x,t) \boldsymbol{\nabla} n_{\text{DNA}}(x,t) \tag{C.11}$$

C.4 Derivation of membrane exclusion model.

To estimate the exclusion forces from membrane confinement at the cell poles, we treat the cell poles as hardwalls and explicitly solve the diffusion equation as a function of distance from the cell wall. Note that the solutions for each cell pole are symmetric, therefore we only show the solution for a single hardwall but that the complete solution is a combination of the symmetric solutions. We begin by calculating the expectation value of the step-size distribution of a diffusive particle a distance δ from the wall:

$$\langle \Delta x - \delta \rangle = \int_0^\infty d\Delta x (\Delta x - \delta) \rho(\Delta x, \delta) \tag{C.12}$$

where $\rho(x, \delta)$ is a probability density function. The hardwall condition can be represented using the method of images, where the probability density function is the sum of gaussians at $+\delta$ and $-\delta$:

$$\langle x - \delta \rangle = \frac{1}{\sqrt{2\sigma^2 \pi}} \int_0^\infty (\Delta x - \delta)(e^{-\frac{(\Delta x - \delta)^2}{2\sigma^2}} + e^{-\frac{(\Delta x + \delta)^2}{2\sigma^2}}) d\Delta x \tag{C.13}$$

In the case of 1D - diffusive motion with diffusion coefficient D, the standard deviation of the gaussians are defined as $\sigma^2 = 2D\delta t$, where δt is the lag time between successive measurements. This integral can then be solved exactly:

$$\langle \Delta x - \delta \rangle = \sqrt{\frac{4D\delta t}{\pi}} \exp\left(-\frac{\delta^2}{4D\delta t} \right) - \delta(1 - \mathrm{erf}(\frac{\delta}{\sqrt{4D\delta t}})) \tag{C.14}$$

To estimate the drift velocity profile from membrane confinement, we use the same convention that we use for experimental data and divide the step size distribution by the lag time between successive measurements:

$$\langle v \rangle = \frac{1}{\delta t} \sqrt{\frac{4D\delta t}{\pi}} \exp\left(-\frac{\delta^2}{4D\delta t} \right) - \frac{\delta}{\delta t}(1 - \mathrm{erf}(\frac{\delta}{\sqrt{4D\delta t}})) \tag{C.15}$$

C.5 Biased diffusive motion

A standard approach to characterize the stochastic motion of a diffusive particle is to calculate the mean squared displacement (MSD) (Eq. 1.1). In an ideal Newtonian fluid, the MSD of an object undergoing diffusive motion with a constant biasing force takes the following form:

$$MSD(\delta t) \equiv 2D\ \delta t + (v\ \delta t)^2 \tag{C.16}$$

where D is the diffusion coefficient, v is the bias velocity, and δt is the observation time. At short times the MSD is dominated by diffusive motion (first term) and at long times the motion is dominated by the bias velocity (second term). The bias velocity explicitly depends on the model of motion (absolute versus relative) while the diffusion coefficient is independent of the model. Because each term in Eq. C.16 scales differently with time, the differences between the two velocity models are irrelevant at short times whereas at long times the models are expected to be divergent, according to their bias velocity. We therefore propose to determine which model best represents the behavior of MS2-mRNA complexes by identifying the velocity model that has the smallest MSD and therefore the smallest bias.

We can also estimate a timescale at which the bias velocity dominates over the diffusive motion by setting the two terms in Eq. C.16 equal to each other. The timescale at which the bias velocity dominates the motion, $\delta t \approx 2D/v^2$, for our experimental values of $D \approx 104\ nm^2/min$ and $v \approx 50\ nm/min$ is approximately 10 minutes, which is consistent with the timescale that absolute model diverges from the relative model in Fig 4.4).

Appendix D

EXTREME VALUE THEORY APPENDIX

D.1 Figures

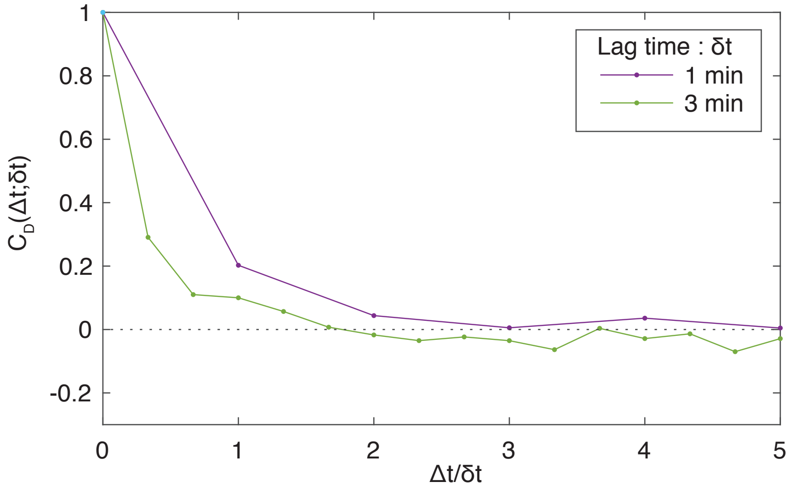


Figure D.1: **Diffusion Coefficient Autocorrelation.** The diffusion coefficient autocorrelation function (as defined in [83]) for the MS2-mRNA particles decays quickly to zero, showing very weak memory.

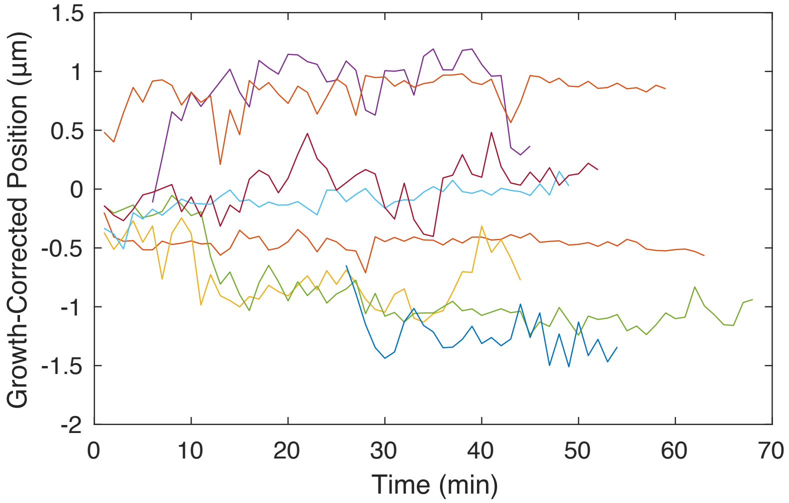


Figure D.2: **Observed trajectories of MS2-mRNA particles.** Sample experimental trajectories of MS2-mRNA particles with starting positions near the center of the cell (position = 0).

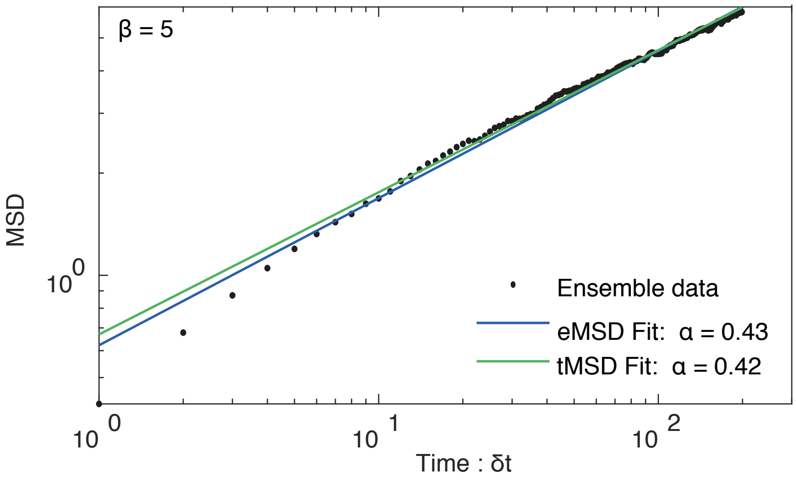


Figure D.3: **Ergodicity in simulations.** Ensemble-averaged MSD (eMSD) and time-and-then-ensemble-averaged MSD (tMSD) for simulated data with disorder strength $\beta = 5$ show that systems with strong disorder are ergodic.

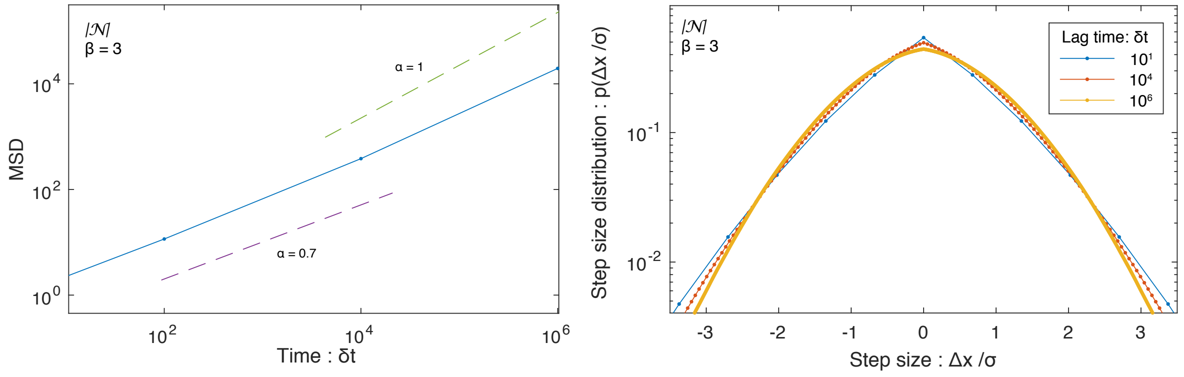


Figure D.4: **Normal free energy distribution and disorder strength.** The absolute value of the normal distribution does not satisfy the strong disorder definition for any variance. **Panel A:** At intermediate times sub-diffusive motion is observed, but at longer times the motion is diffusive. **Panel B:** At intermediate times the step-size distribution best fits a Laplace-like distribution, but at long times it best fits a Gaussian-like distribution.

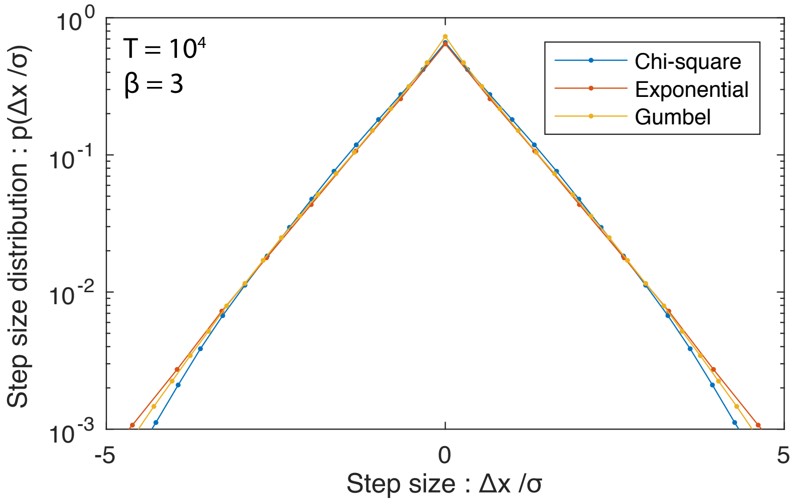


Figure D.5: **Step-size distribution for different free energy distributions.** Different distributions of free energy barriers result in self-similar Laplace-like step-size distributions.

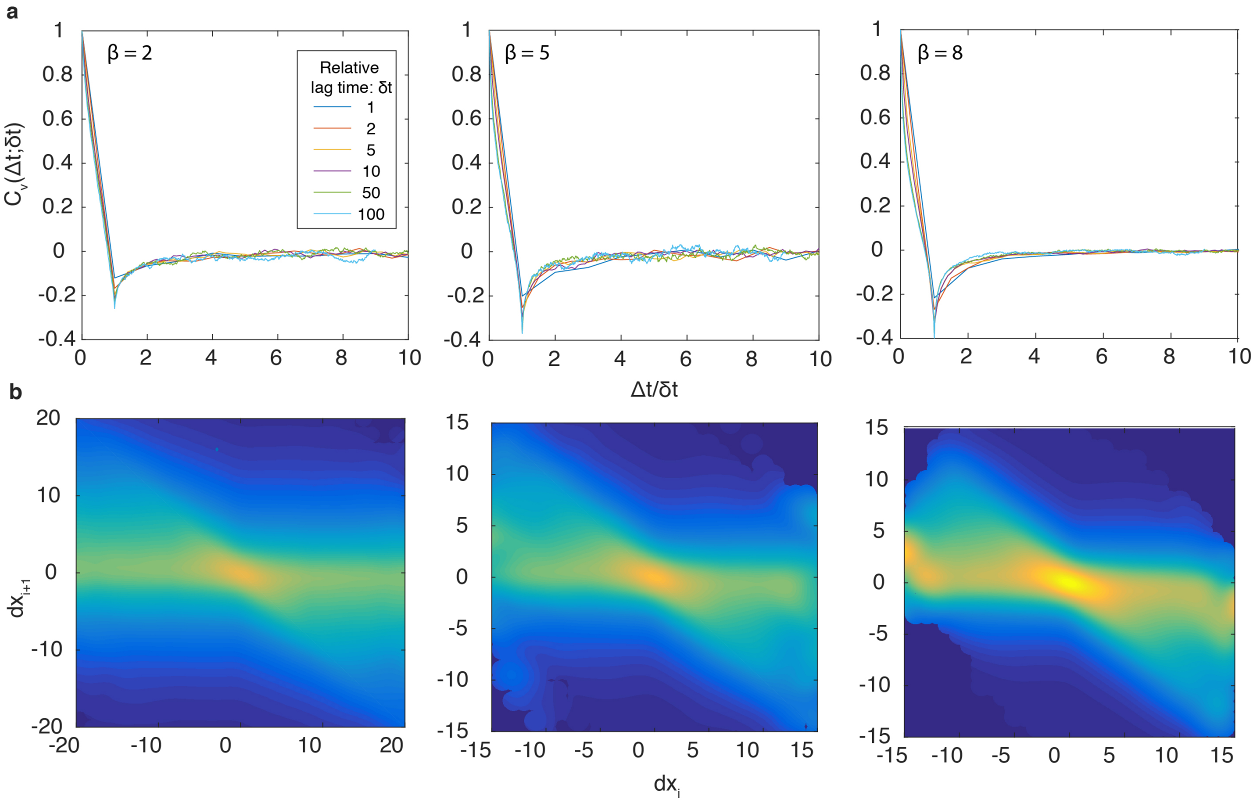


Figure D.6: **Comparison of different disorder strengths. Panel A:** Velocity Autocorrelation for different disorder strength for the barrier model. **Panel B:** Conditional Probability for sequential step size distribution

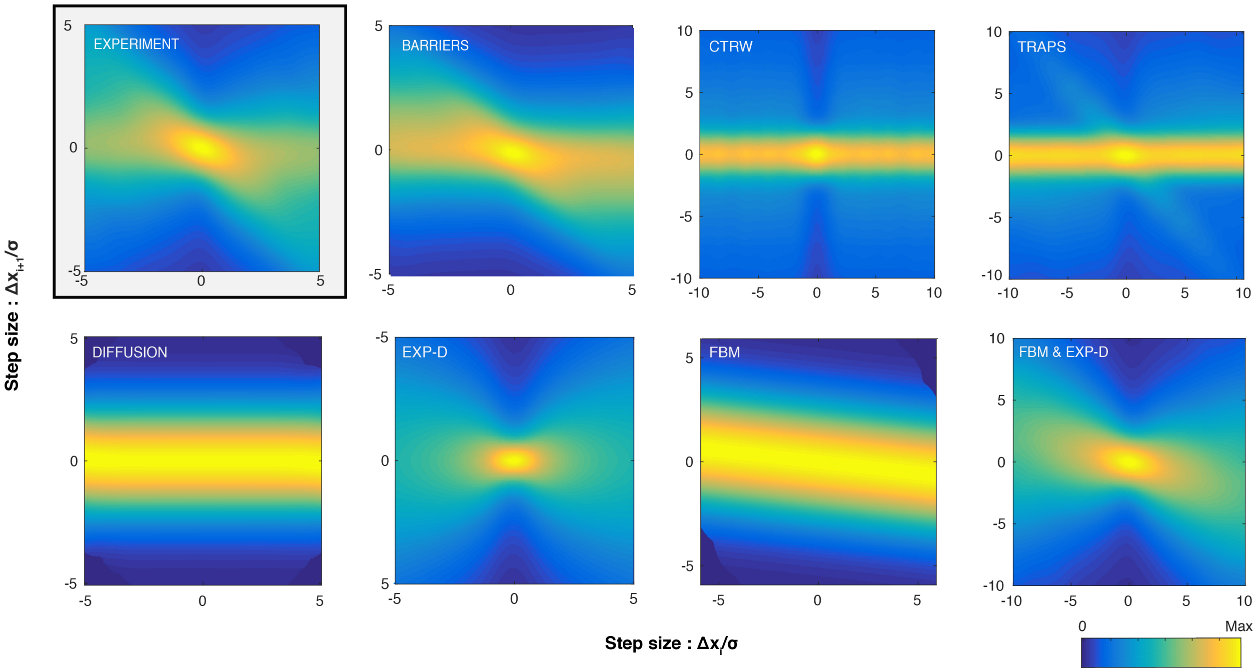


Figure D.7: **Conditional probability of sequential step displacements for different models.** Square root of conditional probability of sequential step displacements for barriers ($\beta = 2$), CTRW, traps ($\beta = 2$), diffusion, Exp-D (exponential distribution of static particle-specific diffusion coefficients), fBm ($\alpha = 0.65$) and fBm ($\alpha = 0.65$) combined with Exp-D.

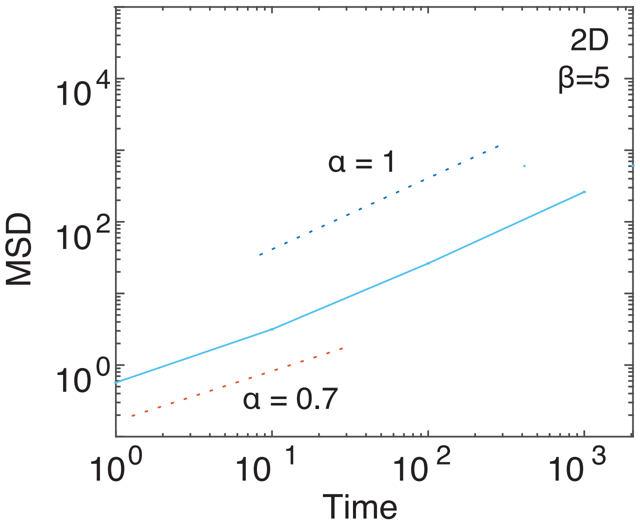


Figure D.8: **MSD for two dimensions.** For more than one, the MSD transitions from diffusion to sub-diffusion. The large obstacles can be avoided by traveling a more circuitous path, independent of the strength of the disorder.

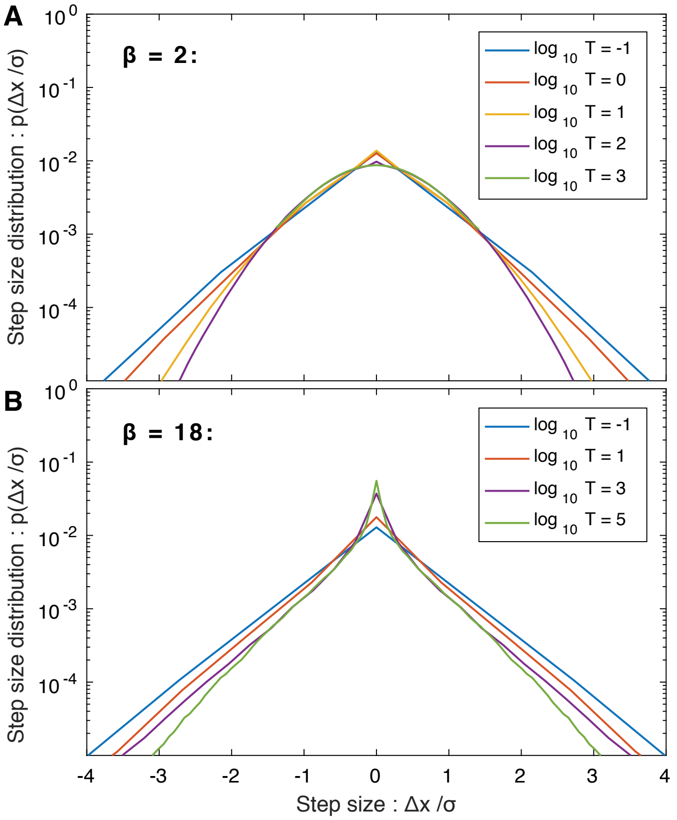


Figure D.9: **Barrier models in higher dimension.** For $\dim > 1$, barrier models are all diffusive at long lag times since large barriers can be avoid by circuitous paths. **Panel A:** In one dimension, $\beta = 2$ is sub-diffusive, but in two dimensions, it is diffusive at long lag times. The step-size distribution can be observed to rapidly transition to a Gaussian distribution (dark blue) at long lag times. **Panel B:** Even though motion is diffusive at sufficiently long times, the crossover time between sub-diffusion and regular diffusion may be extremely long for strong disorder. The step-size distribution for very strong chi-squared disorder is shown above. The EVT-shape is preserved even in the longest lag times shown.

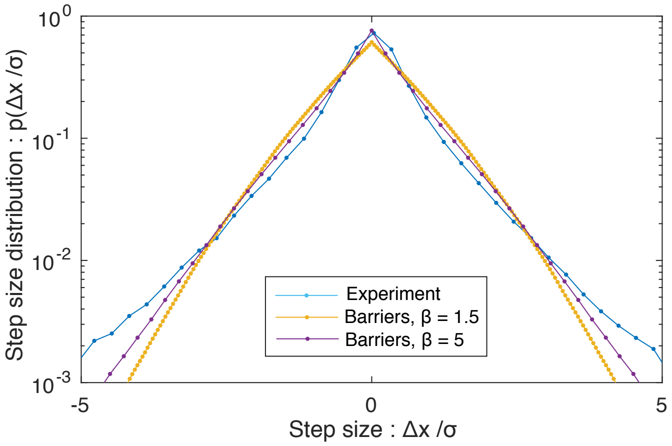


Figure D.10: **Step-size probability distribution for barrier model versus simulation.** Comparison of step-size probability distribution of barrier model and experimental data. The probability distribution for the barrier model is shown for disorder strength $\beta = 1.5$ (which has scaling factor for the MSD matching the experimental data) and $\beta = 5$, to show a stronger disorder.

D.2 EVT and strong disorder

The Gumbel distribution is defined as :

$$F_G(x;\mu,\sigma,0) = \mathrm{Gu}(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \tag{D.1}$$

Let $X_1, X_2 \ldots, X_N \ldots$ be a sequence of independent and identically-distributed random variables, and $M_N = \max\{X_1, \ldots, X_N\}$. Let $F_1(x)$ be the cumulative distribution for a single random variable. The cumulative distribution for M_n is

$$F_N(x) = (F_G(x))^N. \tag{D.2}$$

For N barriers the Gumbel distribution takes the following form :

$$\mathrm{Gu_N}(x) = \exp\left\{-N\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \tag{D.3}$$

We can write the Gumbel distribution in terms of the largest free energy barrier between the starting point 0 and final displacement Δx: $G_{(N)} = \max_{i=1..n} G_i$, where G is the free energy of the barrier. The final displacement of the particles, Δx, can also be related to the number of barriers, N, jumped: $|\Delta x| = N\,x_0$, where x_0 is the barrier spacing. For a large number of steps, N, the Gumbel for the maximum free energy barrier $G_{(N)} \leq g$ is :

$$\Pr\left(G_{(N)} \leq g\right) = \exp\left\{-\frac{|\Delta x|}{x_0}\exp\left(-\frac{g-\mu}{\sigma}\right)\right\}, \tag{D.4}$$

where the two constants, location parameter mu and the space parameter σ, depend on the underlying distribution of random potential G.

We define the Kramers time, the rate at which a particle escapes from a potential well, as the inverse hop rate:

$$\delta t \equiv \frac{1}{k} \equiv e^g, \tag{D.5}$$

and we shall also make the following redefinitions:

$$t_0' \equiv e^{\mu}, \tag{D.6}$$

and substitute into the cumulative distribution:

$$\Pr\left(G_{(n)} \leq g\right) = \exp\left\{-\frac{|\Delta x|}{x_0}\left(\frac{\delta t}{t_0'}\right)^{-1/\sigma}\right\}. \tag{D.7}$$

Now we make a critical reinterpretation of the cumulative distribution: For a particle to have displacement $|\Delta X| \geq |\Delta x|$ at the Kramers time δt, we shall make the assumption that all barriers must be height $G_{(n)} \leq g$. Therefore the probability of $|\Delta X| \geq |\Delta x|$ is approximately equal to the probability of $G_{(N)} \leq g$. We can write the complementary statement:

$$\Pr\left(|\Delta X| \geq |\Delta x|\right) \approx \Pr\left(G_{(N)} \leq g\right), \tag{D.8}$$

and therefore the cumulative distribution function for displacement $|X|$ is:

$$\begin{aligned} \Pr(|\Delta x|; \delta t) &= \exp(-\lambda_{\delta t}|\Delta x|), && \text{(D.9)} \\ \lambda_{\delta t} &\equiv \frac{1}{x_0}\left(\frac{\delta t}{t_0'}\right)^{-1/\sigma}. && \text{(D.10)} \end{aligned}$$

The corresponding PDF is :

$$\mathrm{p}(\Delta x|\delta t) = \frac{\lambda_{\delta t}}{2}\exp\left(-\lambda_{\delta t}|\Delta x|\right), \tag{D.11}$$

and the first and second moments of the distribution are:

$$\begin{aligned} \mathbb{E}_{\Delta X} &= 0, && \text{(D.12)} \\ \mathbb{E}_{\Delta X^2} &= \frac{2}{\lambda_{\delta t}^2} = 2x_0^2\left(\tfrac{\delta t}{t_0}\right)^{2/\sigma}, && \text{(D.13)} \end{aligned}$$

which is a statement for the average displacement and the mean square displacement (MSD) with $\alpha = 2/\sigma$. It is important to note that the expectations over random variable ΔX are understood to be expectation over quenched disorder in this context, rather than thermal fluctuations.

D.3 Details for chi-squared distribution

We need to compute the scale and location parameters for the chi-squared distribution to compare numerical results with analytic results. The cumulative distribution for a chi-squared is:

$$F_{\chi_1^2}(x) = \Gamma^{-1}(\tfrac{1}{2})\,\gamma(\tfrac{1}{2},\tfrac{x}{2}), \tag{D.14}$$

where Γ is the gamma function and γ is the lower incomplete gamma function:

$$\gamma(s,z) = \int_0^z t^{s-1}\,e^{-t}\,\mathrm{d}t. \tag{D.15}$$

To find the Gumbel location and scale parameters, we solve the following approximate equality for large x and N:

$$\exp\left[-N\exp\left(-\tfrac{x-\mu}{\sigma}\right)\right] = \exp\left[-\exp\left(-\tfrac{x-\mu_N}{\sigma_N}\right)\right] \approx F^N_{\chi_1^2}(x). \tag{D.16}$$

After some algebra, we have (given $z \equiv x/2$ and $s \equiv 1/2$):

$$\begin{aligned}
\tfrac{x-\mu_N}{\sigma_N} &\approx -\log N - \log\left[-\log\left(1 - \tfrac{1}{\Gamma(s)}\int_z^\infty t^{s-1}\,e^{-t}\,\mathrm{d}t\right)\right], && \text{(D.17)}\\
&\approx -\log N - \log\left[\tfrac{1}{\Gamma(s)}\int_z^\infty t^{s-1}\,e^{-t}\,\mathrm{d}t\right] + ... && \text{(D.18)}\\
&\approx -\log N - \log\left[\tfrac{z^{s-1}}{\Gamma(s)}\,e^{-z}\int_0^\infty (1+\tfrac{t}{z})^{s-1}\,e^{-t}\,\mathrm{d}t\right] + ... && \text{(D.19)}\\
&\approx -\log N + z - \log\tfrac{z^{s-1}}{\Gamma(s)} + ... && \text{(D.20)}\\
&\approx -\log N + \tfrac{x}{2} + \log\Gamma(s) - (s-1)\log\tfrac{x}{2} + ... && \text{(D.21)}\\
&\approx -\log N + \tfrac{x}{2} + \log\Gamma(s) - (s-1)\log\log N + ... && \text{(D.22)}
\end{aligned}$$

Matching up terms on the right and left, we have :

$$\begin{aligned}
\sigma_N &\equiv 2, && \text{(D.23)}\\
\mu_N &\equiv 2\log N - \log\log N - 2\log\Gamma(\tfrac{1}{2}) + ... && \text{(D.24)}
\end{aligned}$$

We now convert to the free energy random variable $G \equiv \beta X$, which β is the variance. The re-parameterization results in

$$\begin{aligned}
\delta g &\equiv 2\beta, && \text{(D.25)}\\
g_0 &\equiv 2\beta\left[-\tfrac{1}{2}\log\log N - \log\Gamma(\tfrac{1}{2})\right] + ... && \text{(D.26)}
\end{aligned}$$

which leads to a simple result for the scaling exponent:

$$\alpha = \beta^{-1}, \tag{D.27}$$

i.e., just the inverse of the disorder strength. Note that this is only true for sufficiently large β and N.

VITA

Stella Stylianidou was born in Nicosia, Cyprus, on September 3rd, 1988. She received in 2009 a BA in Physics from the University of Oxford, UK and in 2011 a BA in Primary School Education from the Kapodistriakon University of Athens, Greece. She started her PhD in September 2011 at the University of Washington in the United States and received her Masters in Physics in 2013. She has received scholarships from the Cyprus State Scholarship foundation and Fulbright Foundation and several awards in art, poetry and highest middle and high school grades.