SPATIAL BIOLOGY OF TRANSCRIPTION AND TRANSLATION IN LIVE *E. COLI* USING SUPER-RESOLUTION FLUORESCENCE MICROSCOPY

By

Somenath Bakshi

A dissertation submitted in partial fulfillment of

The requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

THE UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 8/27/2013

The dissertation is approved by the following members of the Final Oral Committee:

James C. Weisshaar, Professor, Chemistry

Richard L. Gourse, Professor, Bacteriology

M. Thomas Record, Professor, Chemistry and Biochemistry

Silvia Cavagnero, Professor, Chemistry

Jade Wang, Associate Professor, Bacteriology

To my mother

SPATIAL BIOLOGY OF TRANSCRIPTION AND TRANSLATION IN LIVE *E. COLI* USING SUPER-RESOLUTION FLUORESCENCE MICROSCOPY

Under the supervision of Professor James C. Weisshaar

At the University of Wisconsin-Madison

Abstract

Due to the lack of membrane bound organelles in bacteria, both transcription and translation take place inside the same compartment. The objective of this thesis is to understand the fundamental, as yet unanswered question of where transcription and translation take place inside bacteria. I have focused on the bacterium *E. coli*, as it is the best characterized organism, both molecularly and physiologically. I have used superresolution microscopy techniques in live *E. coli* cells to study the spatio-temporal organization of transcription and translation machinery. By localizing individual molecules with 30 nm scale accuracy, I was able to resolve the spatial distribution and dynamics of RNA polymerase (RNAP) and of the ribosomes. The spatial distribution of RNAP mimicked the spatial distribution of the chromosomal DNA, imaged by DNA-stains and superresolution imaging of DNA-binding protein HU. Ribosomes on the other hand were found to be segregated from the chromosomal DNA. By relating the position, dynamics and copy numbers of ribosomes and RNAPs, I was able to conclude that most translation in *E. coli* is taking place away from the nucleoid in the ribosome-rich regions without transcriptional coupling. This finding is consistent with the timescale of transcription and life time of mRNA. Based on the comparison of spatio-temporal dynamics of chromosomal loci and RNA polymerase, I was able to infer that transcription of protein genes is scattered more or less uniformly throughout the nucleoid. Here the mRNA is probably translated co-transcriptionally. Once the transcription is terminated, the mRNA diffuses out into the ribosomes-rich regions of the end cap of

cytoplasm and is translated until it is degraded by the degradosome machinery. In sharp contrast, transcription of the ribosomal RNA seems to be localized to the periphery of the chromosomal DNA. Such placement of rRNA transcription at the interface between the dense nucleoids and the ribosome-rich regions may be functionally important for efficient ribosome assembly. I also found significant numbers of transcribing RNAPs and translating ribosomes near the cytoplasmic membrane. This provides circumstantial evidence for "transertion hypothesis", which posits the co-transcriptional translation of membrane proteins, which are evidently inserted into the membrane by the Sec-machinery.

Here is a list of the papers I wrote on my dissertation work.

1. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live *Escherichia coli*. Biophys. J. 101:2535-2544.

2. Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol. Microbiol. 85:21-38.

3. Bakshi, S., R. M. Dalrymple, W. Li, H. Choi and J. C. Weisshaar. 2013. Partitioning of RNA Polymerase Activity in Live *E. coli* from Analysis of Single-molecule Diffusive Trajectories. (Submitted)

Acknowledgments

First, and most importantly, I would like to thank my family. Almost nothing in my life would have been possible without my mother. Her sacrifices and support are the most important reasons for me to become who I am now, and who I aspire to be. It was her teachings and endless love that shaped my education and morality from my childhood. Now that she is not among us, she still continues to teach me more. I dedicate this thesis to her. My dad has always been my role model and a great friend, with whom I share a lot of common interests in literature, history and politics. He was my biggest supporter for moving out from my village Chakla, state West Bengal, and then my country India, for higher studies. I am thankful for his encouragements and sacrifices that have helped me throughout my graduate life. My family went through tough times of crisis during my graduate studies, and the person whom I could rely on and continue my studies during most of those days, is my sister. She is my best friend and I am really thankful to have a lovely and caring person like her in my life.

I would be eternally grateful to my advisor, Professor Jim Weisshaar. Throughout my graduate studies, Jim has always encouraged me in my research endeavors, even for non-productive and half complete ones, and provided me with all the freedom and time I needed to pursue them. Outside my experiments in the lab, Jim has always allowed me to attend several conferences and take numerous vacations that I think were the keys to keep my mind motivated, fresh and relaxed. I specially thank him for allowing me all the time I needed to be there with my family.

I am thankful to all the teachers and professors of my life. I am grateful to my uncle and my first mathematics teacher, late Arun Kumar Bakshi, for he was one of the first persons to encourage me for pursuing higher studies in science. I would like to express my gratitude to my teacher Debabrata Pathak and Ashim Mukherjee, who introduced me to scientific thinking in my high school. I am thankful to my former professors at Ramakrishna Mission Vidyamandira of Calcutta University, and IIT Kanpur for they

played a major role in strengthening my knowledge of chemistry and shaping my goals. I was also very fortunate to interact with some of the gifted scientists working at Indian Association for Cultivation of Science in Kolkata. A special thanks to Professor Debashis Mukherjee and Amitava Sarkar for their encouragements and teachings. I am equally grateful to the teachers of the University of Wisconsin Madison. After finishing my master's studies as an organic chemistry major, I was starting my graduate studies in UW-Madison as a physical chemistry major. The teachings of Professor Arun Yethiraj, Professor J.R. Scmidt, and Professor Jim Skinner made that transition smooth, interesting, and enriching.

I want to thank the present and past members of Weisshaar group. The Weisshaar group has always been populated with great scientists and friendly people. I learned a lot from Elizabeth Smith, *coli*n Ingram and Ben Bratton. My labmate Wenting Li is probably one the most caring persons in the world of science. I am thankful to Heejun Choi for all his help and our wonderful discussions. I had great time working with Renee Dalrymple, and was able to reduce the stress of research by chatting and sharing stress together. I am thankful to Ken Barns and Ranga Rajan for good chats and for their help with editing the chapters of my thesis. I would also like to thank the members of Gourse lab, especially Dr. Tamas Gaal, for providing experimental advices, and necessary strains. A special thanks to our collaborator Professor Mark Goulian (University of Pennsylvania) for providing strains and helpful suggestions to enhance the quality of my work.

I would like to thank all the people who have made my life in Madison an enriching and enjoyable experience. I am grateful to Dr. Jagannath Mondal, for his support, advices and encouragements for pursuing research career. He has been a great mentor and role model for my education, since my days as a bachelor student in Ramakrishna Mission Vidyamnadira. I really enjoyed my time with my roommate Sudipta Mukherjee. I will always cherish our discussions on various topics, funny and serious, that have enriched my knowledge and wisdom. I am also grateful to my friend Puja, Jayashree, and Arnab Ganguly for wonderful discussions and HUmorous moments. There are also few

people from outside Madison, who has contributed greatly to my life in Madison. Discussions with my friends Siddhartha, Shyamal, Abhiruk, Pritamda, and Atanu have always provided much needed reminders that there is more to life than microscopes and movies. They have kept my interest in photography and travelling alive, amidst the busy days of graduate life. I am eternally grateful to my friend Sasanka Dalapati for being there for me and my family. He did most of my duties to my family as a son, when I was away struggling between career and conscience.

Finally, I thank my thesis committee, Professor James C. Weisshaar, Professor Richard Gourse, Professor Tom Record, Professor Silvia Cavagnero, and Dr. Jade Wang, for coming together on short notice and for providing all the constructive comments on my research.

Table of contents

ABSTRACT ..... **I**

ACKNOWLEDGMENTS ..... **III**

CHAPTER 1 ..... **1**

INTRODUCTION 1
References *13*

CHAPTER 2 ..... **16**

SUPERRESOLUTION IMAGING OF DYNAMICS AND DISTRIBUTION OF A FLUORESCENT PROTEIN MARKER KAEDE IN LIVE *E. COLI* 16
Introduction *16*
Materials and Methods *17*
Bacterial strains and plasmids 17
Cell growth and preparation 18
Single Molecule Imaging by PALM 19
Single Particle Tracking 20
Fluorescence recovery after photobleaching (FRAP) 21
Monte Carlo Simulation 21
Results *22*
Spatial distribution of Kaede is uniform throughout the E. *coli* cytoplasm 22
Sub-linear MSD of diffusing Kaede 25
Determination of the diffusion coefficient from MSD plots 27
Tests for heterogeneity and sub-diffusion 31
Superfast tracking of Kaede 34
Relief of confinement with Cephalexin 34
Comparison of single-particle tracking data with FRAP data 35
Diffusion map of Kaede is inhomogeneous 36
Discussion *39*
Appendix A: Localization precision *43*
Appendix B: Depth of field *44*
Appendix C: Degree of chromosomal segregation with cell length *47*
Appendix D: Sensitivity of msd_r distribution to two populations with different diffusion coefficients *48*
References *50*

CHAPTER 3.....53

SUPERRESOLUTION IMAGING OF RIBOSOMES AND RNA POLYMERASE IN LIVE *ESCHERICHIA COLI* CELLS.....53
Introduction.....*53*
Materials and Methods.....*57*
Bacterial strains and plasmids.....57
Cell growth and preparation.....58
DNA staining with DRAQ5.....59
Widefield epifluorescence imaging.....60
Superresolution imaging of ribosome-YFP and RNAP-yGFP.....60
Single-molecule image analysis.....63
Monte Carlo simulations.....63
Results.....*64*
Widefield co-imaging of ribosomes (S2-YFP labeling) and nucleoid (DRAQ5 staining).....64
Widefield co-imaging of RNAP (β'-yGFP) and nucleoids (DRAQ5 staining).....68
Superresolution ribosome spatial distribution.....69
Superresolution RNA polymerase spatial distribution.....71
Copy numbers and number densities of ribosomes and RNAP.....73
Diffusion of ribosomes.....76
Effects of rifampicin and chloramphenicol on distribution and diffusion of ribosomes.....81
Discussion.....*84*
What S2-YFP-labeled species are located and tracked?.....84
Ribosome-nucleoid and protein-nucleoid segregation in bacterial cells.....86
Ribosome and RNA polymerase copy numbers.....88
Co-transcriptional translation in E. *coli*.....90
Transertion in E. *coli*.....93
Appendix 3A: Numerical test of the precision of centroid fitting.....*95*
Appendix 3B: Creating a binary mask and cell outline from the white light images.....*98*
Appendix 3C: Measuring cell doubling time for plated cells.....*100*
Appendix 3D: Effects of laser illumination on DNA and ribosome distribution.....*102*
Appendix 3E: Kinetics scheme for cell growth and YFP maturation.....*104*
Appendix 3F: Photon counts from revived YFPs.....*105*
Appendix 3G: Imaging Ribosome-DNA segregation with low-NA objective.....*107*
Appendix 3H: Ribosome-DNA segregation at different growth conditions.....*108*
Appendix 3I: Test for fast-diffusing free S2-YFPs.....*109*
Appendix 3J: Determining of cell dimension from superresolution of YFP.....*110*
Appendix 3K: Radial profile of DNA and RNAP.....*111*
Appendix 3L: Cell-to-cell variation of ribosome diffusion coefficient.....*112*
Appendix 3M: Photo-activation pulse chase experiments of ribosome-mEos2 dynamics in rifampicin treated cells.....*113*
Appendix 3N: Experimental set-up for Photo-activation pulse chase.....*115*

References ... *116*

CHAPTER 4 ... 120

PARTITIONING OF RNA POLYMERASE ACTIVITY IN LIVE *ESCHERICHIA COLI* FROM ANALYSIS OF SINGLE-MOLECULE DIFFUSIVE TRAJECTORIES ... 120

Introduction ... *120*

Materials and Methods ... *121*

Bacterial strains ... 121

Cell growth and sample preparation for microscopy ... 122

Single-molecule imaging ... 123

Analysis of single-molecule tracking data ... 124

Monte Carlo Simulations ... 125

Results ... *125*

Heterogeneous RNAP diffusion on 1-s timescale ... 125

Comparison of slow RNAP motion with fluorescently labeled chromosomal loci ... 129

Mixed-state motion modeled as a confined random walk ... 132

Transitions between slow and mixed diffusion populations on 7-s timescale ... 134

Direct observation of freely diffusing RNAP on 20-ms timescale ... 135

Quantitative partitioning of RNAP ... 137

Discussion ... *140*

What species are we tracking? ... 140

Comparison with earlier study of RNAP diffusion ... 142

Partitioning of RNA polymerase activity ... 143

Appendix 4A: Construction of strains ... *146*

Appendix 4B: Estimation of mean number of RNAPs per cell ... *148*

Appendix 4C: RNAP partitioning: test for cell-to-cell heterogeneity ... *149*

Appendix 4D: RNAP diffusion data for cells grown and imaged at 37ºC ... *150*

Appendix 4E: RNAP diffusion data for cells expressing β'-yGFP ... *151*

Appendix 4F: HU-*mEos2 spatial distribution and estimating nucleoid dimensions* ... *152*

Appendix 4G: Global fit of MSD(τ) of the mixed-state population to free diffusion with confinement ... *153*

Appendix 4H: Analysis of 2-ms/frame movies: $F_{DNA\text{-}bound}$ *and* $F_{unbound}$ ... *155*

Appendix 4I: Effects of limited depth of focus of the 1.45 NA objective ... *159*

Appendix 4J: Contribution of localization error to D_i *distribution for compact trajectories* ... *162*

References ... *164*

CHAPTER 5 ... 167

SUPERRESOLUTION IMAGING OF THE SPATIAL BIOLOGY OF TRANSCRIPTION IN LIVE *ESCHERICHIA COLI* CELLS ... 167

Introduction ... *167*

Materials and Methods ... *169*

Bacterial strains ... 169

Cell growth and sample preparation for microscopy: 170
Microscopy: 171
Results *171*
Spatial distribution of transcribing RNAP copies 171
Spatial distribution of RNAPs transcribing rrnD 174
Spatial distribution of transcription foci and rrn operons in LB 176
Slow RNAP and DNA near the membrane 177
Slow ribosomes near the membrane 178
Discussion *181*
Appendix 5A: Comparison of growth rates *186*
Appendix 5B: y' is not a true "peripheral coordinate" *187*
Appendix 5C: Making a sub-pixel outline of the cell from its phase contrast image *189*
Appendix 5D: Measurement of cell radius *191*
Appendix 5E: Comparing the centerlines of the cell outline imaged before and after the data acquisition *193*
Appendix 5F: Montage of superresolution image of locations of slow RNAP from multiple cells *194*
Appendix 5G: P(y') distribution of two rrn operons *195*
Appendix 5H: Diffusion analysis of ribosome (S2-mEos2) *196*
References *198*

CHAPTER 6 200

TIME-LAPSE IMAGING OF *E. COLI* NUCLEOID MORPHOLOGY AFTER TREATMENT WITH TRANSCRIPTION AND TRANSLATION HALTING DRUGS 200
Introduction *200*
Materials and Methods *202*
Bacterial strains 202
Sample preparation for microscopy 202
Time-lapse fluorescence microscopy 203
Image acquisition and data analysis 204
Results *205*
Parameters to describe the morphology of nucleoid 205
Effects of chloramphenicol on nucleoid morphology and cell growth 206
Effects of rifampicin on nucleoid morphology and cell growth 209
Effect of nucleoid compaction on the distribution of proteins inside the cytoplasm 211
Effects of rifampicin on the distribution of ribosomes 213
Discussion *214*
Appendix 6A: Time-lapse images of DNA with chloramphenicol injection *219*
Appendix 6B: Axial line scans of nucleoid after chloramphenicol injection *220*
Appendix 6C: Transverse profile of nucleoid, before and after chloramphenicol injection *221*
Appendix 6D: Time course of changes in radial dimension of nucleoids (W) from 11 cells after chloramphenicol addition *222*
Appendix 6E: Time-lapse images of DNA with rifampicin injection *223*

Appendix 6F: Time course of radial dimension (W) from 4 different cells after injection of rifampicin *224*
Appendix 6G: Phase contrast images of GFP-expressing cells before and after injection of chloramphenico *225*

CHAPTER 7 200

FUTURE DIRECTIONS 200
References *200*

Chapter 1

Introduction

In the post-genomic era, it is becoming increasingly obvious that the sequence of the entire genome and knowledge of the entire proteome is not sufficient to understand the complexities of how organisms develop and function (1-3). Biological functionalities don't arise from individual gene or their products, but from the network of interactions among them (4). Therefore, along with the knowledge of the individual components (nucleic acids, proteins, lipids etc.) involved in a process, a system level picture of mutual interactions among all the components is important for developing an insight about the biological complexity of an organism (2). The understanding of spatio-temporal organization of the different components involved in a biochemical process within a cell, termed "spatial biology", is the key to such insight (5, 6). Quantitative investigations of the spatial biology of different biochemical processes will not only provide information regarding how different components within the same process talk to each other, but will also give clues regarding the mutual communication between different processes.

The objective of this thesis is to develop a quantitative picture of the spatial biology of transcription and translation processes within bacterial cells. Transcription and translation are two major processes involved in the transfer of genetic information. In bacteria, unlike eukaryotes, due to the lack of membrane-enclosed organelles, both transcription and translation are happening inside the same compartment. Therefore, how the two processes communicate with each other in space/time has great significance for the co-regulation of the two processes and spatio-temporal control of genetic information transfer from the DNA to the translated proteins (7). One special case for transcription-translation

coupling is transertion (8). According to the "transertion hypothesis" transcription, translation, and the insertion of proteins into the membrane, are all taking place simultaneously (Fig. 1.2.C). This may tether genes coding for membrane proteins to the membrane by DNA–RNA polymerase–mRNA–ribosome–membrane protein chains and act as an important "expanding force" on the nucleoid (9). Evidence for co-localization of ribosomes and RNA polymerases involved in transertion can help the understanding of organization of the chromosomal DNA in a bacterium. Transcription and translation are also highly regulated processes targeted by many drug design strategies. A better fundamental understanding of how these systems function together in three dimensions in live bacteria cells might suggest new drug strategies.

Understanding the spatial biology of transcription and translation will require quantitative information about the copy numbers of the machineries involved in these processes, where they are localized within the cell, and their spatially resolved dynamics. The recent developments in the '-omics' technology based approaches is able to deliver high-throughput data for the amount of transcripts and proteins (10, 11). However, the '-omics' technologies have very low spatial and temporal resolution. Most of '-omics' based approaches require the cells to be lysed before the measurement, and therefore are unable to provide any information regarding the dynamics associated with the biomolecules of interest. Similarly, the cryo-EM and especially the new cryo-tomography techniques provide extraordinarily detailed views of the proteins inside a cell, but they are limited to fixed cells and cannot distinguish different proteins from each other (12). This is where live-cell based quantitative imaging technologies excel, as they are able to probe the spatial and dynamic properties with minimal perturbations.

Due to its noninvasive nature and high specificity, fluorescence microscopy has been the method of choice for imaging biological processes inside living cells. The noninvasive nature allows following temporal dynamics of biochemical molecules in a live cell, and the high specificity allows detection of one component at a time. However, due to the limited resolution (~250 nm), simple wide-field

fluorescence microscopy is unable to provide a quantitative high resolution picture of the inside of a cell. The new superresolution techniques of fluorescence microscopy (13-15) enable location and tracking of single molecules in live cells with a spatial resolution that is below the diffraction limit (~30 nm).

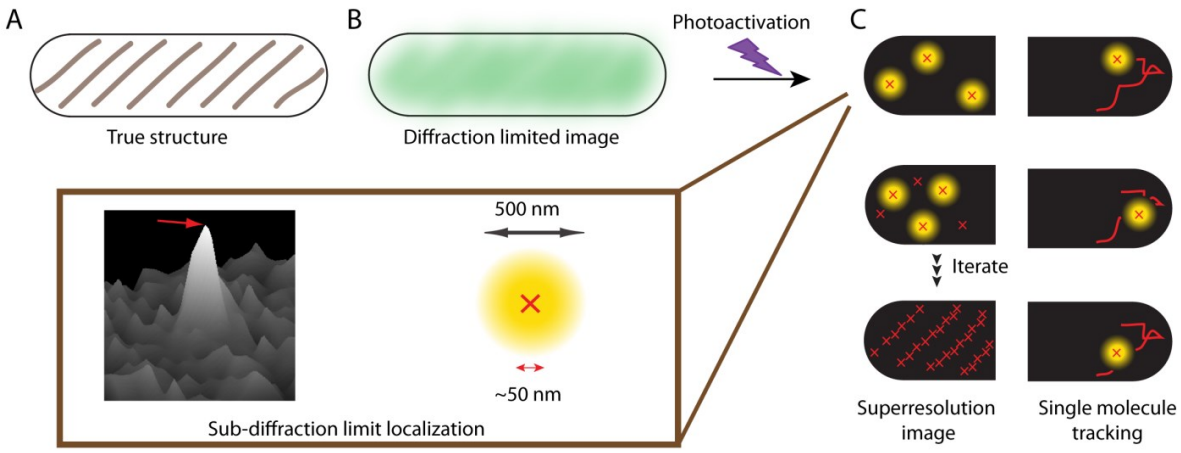

Figure 1.1: Principle of superresolution based single molecule imaging. Because of the diffraction associated with optical imaging, a wide field fluorescence image of the true structure (A) will appear blurred and unresolvable (B). (C) Superresolution microscopy based on photoactivation/photoswitching provides a solution to this problem. An activation pulse switches on a small subset of fluorophores such that their images (yellow spots) are resolvable from each other. Each spot appears to be ~500 nm wide (inset). However a numerical fitting algorithm can be used to determine the position of the peak (Inset: left) with much higher accuracy, shown as a red crosses. Once the positions of the molecules in a set is determined and stored, they are bleached. A new subset of molecules is activated and the same process localizes their positions. Multiple iterations of the activation-localization-bleaching process enable us to get the position information of a large number of molecules. Plotting the positions of the entire set of localized molecules creates the *superresolution image* of the true distribution. Due to the non-invasive nature of fluorescence imaging, one can follow the locations of individual molecules over time. This provides information regarding dynamics of molecules at very spatial and temporal resolution.

Sequential imaging and storing the sub-diffraction limit localization of a large number of single molecules in this way allow one to construct a very high-resolution picture of the spatial organization of the components (Fig. 1.1.C: left panel). On the other hand, following the movement of a single molecule in time provides information about the temporal dynamics of individual components (Fig. 1.1.C: right panel). These unique strengths make single molecule imaging using superresolution techniques ideally suited for the studies of spatial biology of different processes in bacterial cells (16).

This thesis aims to answer key questions about the organization and time scales of transcription and translation in live bacteria cells with unprecedented spatial and temporal resolution using the supereresolution based single molecule imaging techniques. We focused on the bacterium *E. coli,* which is both molecularly and physiologically the best characterized organism. The primary targets were the ribosomes, RNA polymerase (RNAP), and specific DNA loci including the *rrn* operons. We have used genetically encoded fluorescent proteins (GFP variants: YFP and mEos2) to label the specific protein of interest. Using either the reversible photobleaching method for YFP (17), or the photoswitching method for mEos2 (18), we were able to localize and track the positions of thousands of copies of single proteins in a single cell. To examine the spatio-temporal behavior of a specific position along the chromosome, we have imaged chromosomal foci labeled with the *parS*-ParB fluorescent labeling scheme that marks one specific DNA locus at a time (19). All these experiments were conducted in live, growing cells. This provides information about dynamics and also avoids common fixation artifacts such as changes in nucleoid morphology and poor single-molecule localization accuracy due to polarization (20). Spatial information from a large number of localized molecules was used to construct a sub-diffraction limit image. Trajectories of individual molecules were used to examine the differences in temporal dynamics among different biochemical states of a system. New automatic trajectory classification techniques were developed to characterize the biochemical state of the system of interest based on diffusivity. This

allowed us to assign different biochemical states associated with rapid or slow dynamics to their state-specific spatial organization pattern which gives a new level of detail into the picture. These experimental techniques were accompanied with computer simulation of reaction-diffusion processes to enhance the interpretation of the data.

It is important to segregate the underlying biological phenomena from the perturbations due to the fluorescent labeling strategies. To investigate the spatial biology of the fluorescent tag itself, we needed to examine the spatial distribution and dynamics of a photoactivatable fluorescent protein. In chapter 2, we have used two-dimensional photoactivation localization microscopy (PALM) to study the spatial distribution and diffusion of the protein Kaede, expressed from a plasmid in the cytoplasm of live *E. coli* under moderately rapid growth conditions. The spatial distribution of Kaede is uniform within the cytoplasm. Under these moderate growth conditions (67 min doubling time), the cytoplasmic radius was determined to be 380 ± 30 nm with little variation from cell to cell (21). Single-particle tracking (SPT) using 4 ms exposure times reveals negatively curved plots of mean-square displacement (MSD) vs time for the axial coordinate x, the transverse coordinate y, and the two-dimensional distance coordinate r. Detailed comparison with Monte Carlo simulations of free diffusion in a spherocylindrical volume shows that the curvature can be quantitatively understood in terms of free diffusion within a confining volume. A curve-fitting procedure takes account of confinement effects to yield diffusion coefficients more accurate than those from the initial slope of the MSD plots. The mean diffusion coefficient across cells is $\langle D_{Kaede} \rangle = 7.3 \pm 1.1\ \mu m^2\text{-}s^{-1}$, sensibly consistent with the likely homo-tetrameric form of Kaede. The moment scaling spectrum is linear up to the 10^{th} moment with slope 0.42 ± 0.04, again consistent with free diffusion (22). Within one cell, the distribution of squared displacements along x for individual Kaede molecules is consistent with homogeneous diffusion. Homogeneous Brownian diffusion, and uniform spatial distribution of a fluorescent tag makes it a good candidate as a label, as the deviations

from Brownian motion and/or uniform distribution of the fusion proteins can be associated with important biological properties of the protein of interest.

In the chapter 2, we also introduce the concept of the diffusion map. Diffusion map, a spatial distribution of locally averaged estimates of the diffusion constant can provide information about the variation of diffusivity of a protein across the cell. Such variations across subcellular locations within the cell can be used to understand the spatial variation in crowding and other physical/biological factors that affect the motion of a protein. In case of Kaede, even though the distribution of diffusion constants from different trajectories seems to follow a homogeneous distribution, the short time diffusion constant (1-step estimate) across the different regions of the cell was inhomogeneous (21). For longer cells a spatial map of one-step estimates of the diffusion coefficient along *x* shows that diffusion is some 20-40% faster within nucleoids than in the ribosome-rich region lying between nucleoid lobes at the cell mid-plane. We have measured the concentration profile of ribosomes across cell length, and find that a large volume fraction (~25%) of the ribosome-rich regions in the end cap and at the mid-plane of the cell is occupied by ribosomes (ribosome-rich regions, chapter 3). The ribosomes in these regions can act as crowders and impede the diffusion of the small fluorescent tag in the ribosome-rich regions to cause the inhomogeneous diffusion map.

Chapter 3 of this thesis examines the quantitative spatial distributions of ribosomes (labeled by S2-YFP) and RNA polymerase (labeled by β'-yGFP) in live *E. coli* measured by superresolution fluorescence microscopy (23). We find that in moderate growth conditions, nucleoid-ribosome segregation is strong, and RNAP localizes to the nucleoid lobes (Fig. 1.2). Only 10-15% of the ribosomes lie within the densest part of the nucleoid lobes, and at most 4% of the RNAPs lie in the two ribosome-rich endcaps (23). Using superresolution single molecule tracking of ribosomes, we find that ~80% of the ribosomes are diffusing with a diffusion coefficient of $D_{ribo} = 0.03\ \mu m^2/s$, attributed to the diffusion of free mRNA being translated by one or more 70S ribosomes. The rest of ribosomes (~20%) are diffusing

with a much faster diffusion coefficient (~0.15 $\mu m^2/s$). We assign the fast fraction to the 30S subunits that have disengaged from mRNA after they have finished translation and are searching to initiate another round of translation. We find no clear evidence of sub-diffusion, as would arise from tethering of ribosomes to the DNA via co-transcriptional translation. The degree of DNA-ribosome segregation argues persuasively against co-transcriptional translation as the primary means of protein expression (Fig. 1.2B). It seems that most newly synthesized mRNA copies, decorated with ribosomes and other protective species, freely diffuse to the ribosome-rich regions where the bulk of translation occurs (Fig. 1.2A). The measured diffusion constant of the slow ribosomes ($D_{ribo} = 0.03\ \mu m^2\text{-}s^{-1}$) then refers to mRNA copies decorated with multiple 70S ribosomes (polysomes) (24). If so, the estimated time for a completed message to diffuse from the nucleoid to a ribosome-rich region is < 1 s, sufficiently fast to enable many rounds of protein synthesis prior to mRNA degradation. Both RNAP and ribosome radial distributions extend to the cytoplasmic membrane, consistent with the transertion hypothesis (Fig. 1.2.C). However, few if any RNAP copies lie near the membrane of the endcaps. This suggests that if transertion occurs, it exerts a direct radially expanding force on the nucleoid, but not a direct axially expanding force (9).

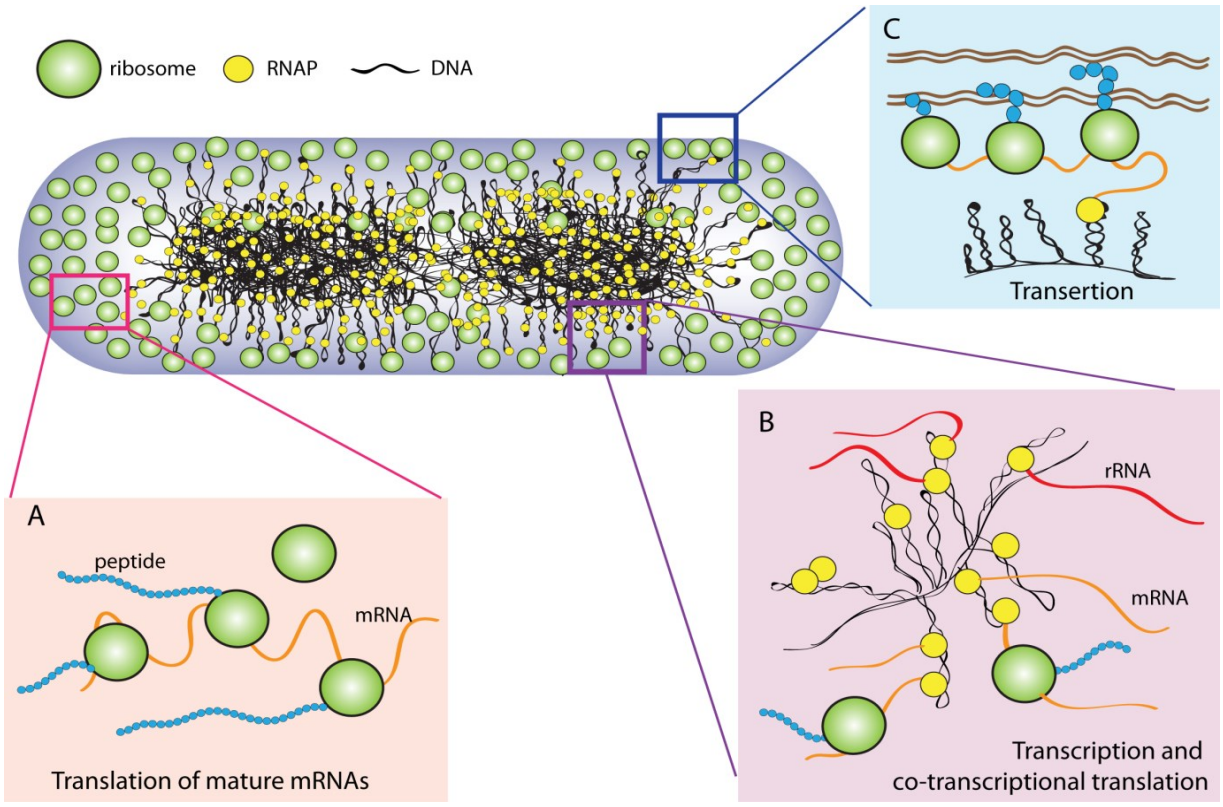

Figure 1.2: Spatial biology of transcription and translation in *E. coli*. In the growth conditions used in the experiments of this thesis, the chromosomal DNA (black curly lines) appears to be present in two or four lobes. Based superresolution imaging of ribosome in *E. coli,* we conclude that ribosomes (green spheres) are strongly segregated from the chromosomal DNA, and are highly concentrated near the end caps and near the mid-cell region between the two chromosome lobes. RNA polymerase (yellow spheres), on the other hand, are exclusively localized on the chromosomal DNA. (A) Majority of the ribosomes are possibly involved in translation of mature mRNA. (B) ~50% of the RNAPs are specifically bound to the DNA and are involved in transcription of mRNA and stable RNAs (rRNA and tRNA). Some fraction of ribosomes can be involved in co-transcriptional translation of the mRNAs as they come out of the RNA polymerase core. (C) For some membrane proteins, this co-transcriptional translation process can be coupled to insertion (transertion) via Sec-machinary.

RNA Polymerase (RNAP) can be freely diffusing in the cytoplasm, nonspecifically bound to DNA, or transcribing a gene. Partitioning of RNAPs among these states plays a key role in regulating the gene expression in bacteria. In chapter 4, we have used superresolution fluorescence microscopy to locate and track the diffusive motion of single copies of RNA polymerase (RNAP) in live *E. coli*. On a timescale of 0.1–1 s, most RNAP copies separate remarkably cleanly into two states with comparable populations. The "slow" RNAPs move indistinguishably from DNA loci. We assign them to specifically bound copies that are initiating transcription, elongating, pausing, or awaiting termination. The "mixed-state" RNAP copies act as a homogeneous population with $D_{mixed} = 0.2\ \mu m^2\text{-}s^{-1}$. These are assigned as a rapidly exchanging mixture of non-specifically bound copies and copies undergoing free, three-dimensional diffusion within the nucleoids. In longer trajectories of 7-s duration, we directly observe transitions between the slow and mixed states, corroborating the assignment of the slow state as specifically bound to DNA. Short trajectories of 20-ms duration (tracked at 500 Hz) enable direct observation of RNAP copies with $D_{free} = 0.7\ \mu m^2\text{-}s^{-1}$, which we attribute to unbound, freely diffusing RNAP. Analysis of single-particle trajectories provides quantitative estimates of the partitioning of RNAP into different states of activity: $f_{trx} = 0.54 \pm 0.07$, $f_{ns} = 0.28 \pm 0.05$, $f_{free} = 0.12 \pm 0.03$ and $f_{nb} = 0.06 \pm 0.05$ (fraction unable to bind to DNA on a 1-s timescale). These fractions disagree with earlier estimates that relied on measurements of the RNAP content of DNA-free mini-cells (25-27).

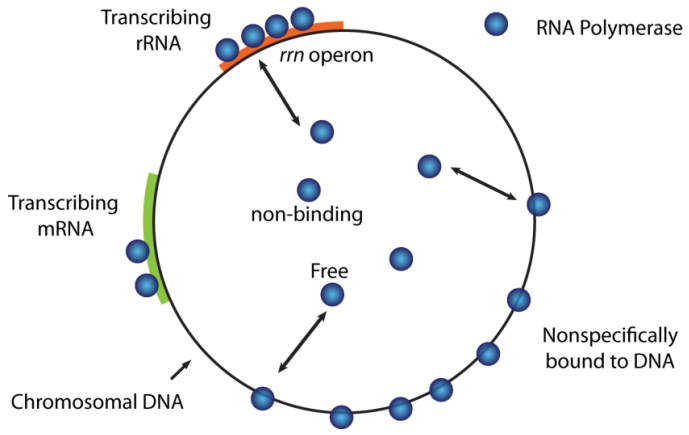

Figure 1.3: Schematic of RNAP Partitioning. During exponential growth there are about 4 major classes of RNAP in a cell, RNAPs involved in transcription (transcribing mRNA and rRNA), freely diffusing RNAPs, RNAPs nonspecifically bound to DNA, and non-binding RNAPs. The circle represents chromosomal DNA.

Adaptive bacteria needs to selectively and effectively transfer the genetic information stored in the chromosomal DNA. Spatio-temporal organization of transcription plays an important role in controlling the transfer of this genetic information. In chapter 5, we specifically investigate the spatial biology of transcription. Based on the findings of chapter 4, we are able to specifically detect RNAP copies that are actively involved in the transcription process using a new assay based on diffusion blur over long exposure time. Consistent with previous results of RNAP partitioning experiments, described in chapter 4, we find that ~45% of RNAP molecules diffuse very slowly, their fluorescence signal appears diffraction limited over a 500 ms long exposure. On the other hand, RNAP molecules exchanging between nonspecifically bound states and freely diffusing states appear blurred. We have used a simple thresholding method to selectively localize the sharply focused diffraction limited spots from the transcribing RNAPs. The spatial distributions of transcribing RNAP, *rrnG*, and transcription foci enable us to infer that in rapid growth conditions, transcription of *rrn* operons occurs preferentially at the periphery of the nucleoids. We suggest that it may be functionally important to synthesize long pre-rRNA transcripts in the region of space near the membrane, where they can be processed and assembled efficiently. Transcription of genes encoding proteins is evidently more uniformly distributed throughout the nucleoids.

The existence of a small population of transcribing RNAP and some DNA at or very near the cytoplasmic membrane lends support to the transertion hypothesis of co-transcriptional translation of membrane proteins (Fig. 1.2.C). If the slow RNAPs near the membrane are truly involved in "transertion" we should be able to find slowly diffusing ribosomes near the membrane as well. These ribosomes should be tethered to membrane and chromosomal DNA and therefore should display a very limited motion. We needed to extend the timescale of ribosome tracking to examine such low-dispersion phenomenon. Due to the reversible photobleaching based methodology used for the single molecule imaging used for ribosomes labeled by S2-YFP in chapter 2, we could not arbitrarily lengthen the time

between frames. However a photoactivatable/photoswitchable fluorescent protein tag makes it possible to space the camera frames widely in time and observe the dynamics of protein at much longer timescales. Therefore in chapter 5, we extend the ribosome tracking experiments to much longer times using the cell line expressing S2-mEos2 from the chromosome. The trajectories at longer timescale (5s) reveal a third, very slow ribosome population that diffuses even slower than the DNA loci. About 10-15% of the trajectories displayed such low diffusivity. These copies tend to locate at the membrane, making them candidates for direct insertion of membrane proteins via the Sec machinery.

Transcription and translation can have significant impact on the organization of the bacterial chromosome. Transcription/ translating halting drugs are used to study the effect of these processes on nucleoid organization. In chapter 6, we describe our findings from time lapse fluorescence imaging of DNA morphology using cell permeable DNA-binding chemical probes as the cells are treated with transcription/translation halting drugs. Historically the effects of such processes are examined with the observations made at only two time points (before and after the addition of the drug). Following the changes in the spatial distribution of the fluorescence signal from the DNA-binding probe with time enabled us to learn about new mechanistic signatures of the process affecting the DNA morphology. Addition of translation halting drug chloramphenicol caused significant radial compaction of the chromosomal DNA within 5 mins. Interestingly similar phenomena were observed with the treatment of the transcription halting drug rifampicin.

This observation is consistent with our hypothesis that transertion could act as a radially expanding force on the chromosome. Halting either translation or transcription will stop transertion. In the absence of this radially expanding force the DNA will collapse in the radial dimension. The distribution of the chromosomal DNA in turn affects the distribution of proteins within the bacterial cytoplasm. Compaction of the chromosome caused small proteins, which are otherwise uniformly distributed across the DNA and cytoplasm in an untreated cell, to segregate away from the nucleoid core and concentrate in

the end cap regions. Previous experiments on protein aggregates (28, 29) suggested macromolecular crowding can cause protein distribution patterning within bacteria cells. In line with their observations, our findings suggest that chromosomal DNA acts as a sieve in *E. coli*. Size dependent partitioning of proteins based on the compactness of this DNA-mesh may play an important role in determining protein distribution across the cytoplasm. This can also provide a rationale for the observed segregation of ribosomes from the chromosomal DNA (23).

Overall, the findings of this thesis enable us to construct a uniquely detailed picture of how the chromosomal DNA, RNAPs, and ribosomes work together inside a living, growing *E. coli* cells (Fig. 1.2). The inherent complexities associated with biological systems make it very difficult to have an intuitive understanding of the experimental data. Complete system level understanding of any such complex system can only be tested by mathematical modeling and computer simulations. Systems biology based computational modeling approaches to determine such holistic pictures needs quantitative experimental data. The type of quantitative, observational science described in this thesis will critically test and guide the refinement of such computational models of transcription, translation, and of key diffusive transport mechanisms in the *E. coli* cytoplasm (27, 30). This information will be able to answer a number of longstanding questions, and raise new questions about the underlying biochemical and biophysical forces that dictate the inner workings of the bacterial cytoplasm.

References

1. Chakravarti, A. 2011. Genomics is not enough. Science. 334:15-15.

2. Kitano, H. 2002. Systems biology: A brief overview. Science. 295:1662-1664.

3. James, P. 1997. Protein identification in the post-genome era: The rapid rise of proteomics. Q. Rev. Biophys. 30:279-331.

4. Verveer, P. J., and P. I. H. Bastiaens. 2008. Quantitative microscopy and systems biology: Seeing the whole picture. Histochem. Cell Biol. 130:833-843.

5. Shapiro, L., H. H. McAdams, and R. Losick. 2009. Why and how bacteria localize proteins. Science. 326:1225-1228.

6. Hurtley, S. 2009. Location, location, location introduction. Science. 326:1205-1205.

7. Campos, M., and C. Jacobs-Wagner. 2013. Cellular organization of the transfer of genetic information. Current Opinion in Microbiology. 16:171-176.

8. Norris, V., and M. S. Madsen. 1995. Autocatalytic gene expression occursviatransertion and membrane domain formation and underlies differentiation in bacteria: A model. Journal of Molecular Biology. 253:739-748.

9. Woldringh, C. L. 2002. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Mol. Microbiol. 45:17-29.

10. Anderson, N. L., and N. G. Anderson. 1998. Proteome and proteomics: New technologies, new concepts, and new words. Electrophoresis. 19:1853-1861.

11. Wang, Z., M. Gerstein, and M. Snyder. 2009. Rna-seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57-63.

12. Matias, V. R. F., A. Al-Amoudi, J. Dubochet, and T. J. Beveridge. 2003. Cryo-transmission electron microscopy of frozen-hydrated sections of *Escherichia coli* and pseudomonas aeruginosa. J. Bacteriol. 185:6112-6118.

13. Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science. 313:1642-1645.

14. Hess, S. T., T. P. K. Girirajan, and M. D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys. J. 91:4258-4272.

15. Rust, M. J., M. Bates, and X. W. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). Nat. Meth. 3:793-795.

16. Cattoni, D. I., J. B. Fiche, and M. Noellmann. 2012. Single-molecule super-resolution imaging in bacteria. Current Opinion in Microbiology. 15:758-763.

17. Deich, J., E. M. Judd, H. H. McAdams, and W. E. Moerner. 2004. Visualization of the movement of single histidine kinase molecules in live caulobacter cells. Proc. Natl. Acad. Sci. U. S. A. 101:15921-15926.

18. McKinney, S. A., C. S. Murphy, K. L. Hazelwood, M. W. Davidson, and L. L. Looger. 2009. A bright and photostable photoconvertible fluorescent protein. Nat. Meth. 6:131-133.

19. Nielsen, H. J., J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen. 2006. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. Mol. Microbiol. 62:331-338.

20. Backlund, M. P., M. D. Lew, A. S. Backer, S. J. Sahl, G. Grover, A. Agrawal, R. Piestun, and W. E. Moerner. 2012. Simultaneous, accurate measurement of the 3d position and orientation of single molecules. Proc. Natl. Acad. Sci. U. S. A. 109:19087-19092.

21. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live *Escherichia coli*. Biophys. J. 101:2535-2544.

22. Ferrari, R., A. J. Manfroi, and W. R. Young. 2001. Strongly and weakly self-similar diffusion. Physica D. 154:111-137.

23. Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol. Microbiol. 85:21-38.

24. Golding, I., and E. C. Cox. 2004. Rna dynamics in live *Escherichia coli* cells. Proc. Natl. Acad. Sci. U. S. A. 101:11310-11315.

25. Bremer, H., P. Dennis, and M. Ehrenberg. 2003. Free RNA polymerase and modeling global transcription in *Escherichia coli*. Biochimie. 85:597-609.

26. Shepherd, N., P. Dennis, and H. Bremer. 2001. Cytoplasmic RNA polymerase in *Escherichia coli*. J. Bacteriol. 183:2527-2534.

27. Klumpp, S., and T. Hwa. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proc. Natl. Acad. Sci. U. S. A. 105:20245-20250.

28. Coquel, A.-S., J.-P. Jacob, M. Primet, A. Demarez, M. Dimiccoli, T. Julou, L. Moisan, A. B. Lindner, and H. Berry. 2013. Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. Plos Computational Biology. 9.

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

30. Klumpp, S., Z. Zhang, and T. Hwa. 2009. Growth rate-dependent global effects on gene expression in bacteria. Cell. 139:1366-1375.

Chapter 2

Superresolution Imaging of Dynamics and Distribution of a Fluorescent Protein Marker Kaede in Live *E. coli*

Introduction

Sub-diffraction-limit fluorescence microscopy (1-4) in two dimensions enables location and tracking of single copies of specifically labeled proteins in live cells with spatial resolution on the order of 20-50 nm. Such high resolution is particularly advantageous in small bacterial cells (5-9) like the rod-shaped *E. coli*, which has cylindrical radius of only 400 nm and length of 2-4 μm. One recent study found clustering of the architectural protein HU in fixed *Caulobacter crescentus* cells, with the clustering particularly strong in pre-divisional cells (8). Another unraveled details of the stringent response mechanism in live *E. coli* (9). Photoactivation localization microscopy (PALM) promises a new level of quantitative insight into the spatial distribution of many types of specific components of live bacterial cells.

In this prototypical study of diffusion in live *E. coli* cells, we use PALM to determine the spatial distribution and diffusive properties of the photoactivatable fluorescent protein Kaede, expressed from a plasmid in the *E. coli* cytoplasm. Kaede is a 28 kDa β-barrel which probably exists as a homotetramer (10, 11). The spatial distribution of Kaede is uniform and fills the entire cytoplasm to high accuracy, in contrast to the results of theoretical models of globular protein segregation between nucleoids and ribosome-rich regions (12, 13). PALM also enables measurement of the radius and length of the cytoplasm of live *E. coli* to an accuracy of ~30 nm. For HUndreds of Kaede copies in each single, live cell, we measure mean-square displacement vs lag time ($MSD_i(\tau)$) in the axial coordinate x, the transverse

coordinate *y*, and the two-dimensional distance coordinate $r = \sqrt{x^2 + y^2}$. All three MSD plots are curved under our typical measurement conditions of 4 ms/frame. By comparison with Monte Carlo simulations of free diffusion in a spherocylindrical confining volume, we show that the curved MSD plots can be explained by confinement alone, without recourse to sub-diffusion. Here sub-diffusion refers to any mechanism other than the finite boundaries of the cell that causes MSD plots to increase sub-linearly in time (14). Local caging effects or strong transient binding are possible examples. The mean diffusion coefficient across cells is D_{Kaede} = 7.3 ± 1.1 μm^2-s^{-1}**,** which is a factor of 1.3 smaller than that of GFP under similar growth conditions.

Additional statistical tests based on higher moments of the displacements (15) are also consistent with free diffusion in a confining spherocylindrical volume. However, for long cells (~4 μm) the spatial distribution of one-step squared *x*-displacements, when averaged over the transverse coordinate *y* and plotted vs the axial coordinate *x*, shows a 20-40% dip at the cell mid-plane. The longest cells also show the strongest spatial segregation of the chromosomal DNA into two lobes. The results thus suggest that Kaede diffusion is more facile within the nucleoids than in the ribosome-rich region between nucleoids. The distribution across cells of D_{Kaede} from single-particle tracking (SPT) and from fluorescence recovery after photobleaching (FRAP) are similar. Finally, we describe relatively straightforward methods for extracting diffusion coefficients accurate to ±6**%** from curved $MSD_x(\tau)$ plots for spherocylindrical geometries, without the need for Monte Carlo simulation of confinement effects.

Materials and Methods

Bacterial strains and plasmids

All experiments were performed on the *E. coli* strain DH5α. Kaede was expressed from a tetracycline inducible plasmid. The Kaede ORF (a gift from D. Weibel, UW-Madison Biochemistry) was amplified by PCR (primers, 5'-TGACTCTGCAGAGACTTGACGTTGTCC-3' and 5'-

GTGACGTCTAGAATGAGTCTGATTAAACCAG-3') and digested by XbaI and PstI. This fragment was ligated to a similarly digested fragment of pASK-IBA3plus (Invitrogen, Carlsbad, CA). DH5α cells were transformed with the resulting plasmid (pJW2) and transformants were selected on ampicillin containing plates.

Cell growth and preparation

E. coli cells were grown overnight with shaking at 30°C in EZ rich defined medium (EZRDM) with 100 μg/mL ampicillin (16, 17). We subsequently made subcultures of these cells by diluting the stationary phase culture at least 1:250 into 3 mL of fresh EZRDM. When cells had grown to midlog phase, (OD_{600} = 0.4-0.6), anhydrotetracycline was added to a final concentration of 45 nM. After 6 min of induction the cells were centrifuged and resuspended twice in fresh EZRDM to remove the inducer. The cells were then incubated in growth medium for at least 30 min at 30°C to enable maturation of the fluorescent proteins. Following induction of Kaede and resuspension of the cells in fresh growth media, the doubling time was 67 min, longer than the 57 min doubling time before induction. From single-cell fluorescence intensities and volumes, we estimate the Kaede concentration in the cytoplasm is typically about 5 μM, comparable to GFP concentrations in earlier diffusion studies(16, 18). For the diffusion studies, 7 μL of cells were immobilized on poly-L-lysine-coated coverslips and sealed with nailpolish (CG Continuous Color 3-in-1, 024 Peek-a-Boo Pink). FRAP measurements and single-molecule tracking with photoactivation were carried out for no longer than 30 min after plating (16). During that time, cells continue to grow. For studies of elongated cells treated by cephalexin, 30 μL of an overnight culture was diluted 100-fold into EZRDM medium and shaken at 30°C for 1 h. Cephalexin (60 μg/mL) was then added, and the culture was grown with shaking to an OD of 0.3-0.8. Kaede was expressed as above except that the resuspension medium contained cephalexin (60 μg/mL) to continue to suppress septum formation.

Single Molecule Imaging by PALM

Fluorescence images of Kaede were acquired using widefield epifluorescence microscopy. Prior to photoactivation, Kaede absorbs blue-green light (λ_{max} = 508 nm in absorption) and fluoresces green (λ_{max} = 518 nm). This form was used in the FRAP studies. After photoactivation, Kaede absorbs green light (λ_{max} = 558 nm) and fluoresces yellow (λ_{max} = 582 nm) (10, 11, 19, 20). Photoactivation was carried out with a 405-nm diode laser (CrystaLaser, Reno, Nevada). The intensity of the activation laser varied from 1-50 W/cm^2 at the sample to control the number density of activated molecules. The yellow, photoactivated state was probed by excitation at 561 nm (diode laser, CrystaLaser) with 3-15 kW/cm^2 at the sample. At these intensities, the 561-nm laser also photoactivates Kaede, albeit with a much lower efficiency than the 405-nm laser. We often used the 561-nm laser for both photo-activation and localization in order to severely limit the number of activated molecules to less than one particle per cell per camera frame.

Images for single-particle tracking were acquired with a Nikon Eclipse Ti inverted microscope with oil immersion objectives (Nikon Plan Fluor 100X, NA 1.30 and Nikon APO TIRF 100X, NA 1.49) and a 1.5X tube lens. For time lapse imaging, fast shutters (Uniblitz LS2, Vincent Associates, New York) were used to synchronize illumination and image acquisition. Images were recorded by a back-illuminated EMCCD camera with 16 μm x 16 μm pixels (iXon DV-887, Andor Technology, Connecticut). Each pixel corresponds to 105×105 nm^2 at the sample with 150X magnification. A 64 by 64 pixel region of the EMCCD was used during data acquisition from each single cell. This enables imaging at a frame rate of 234 Hz with 4.0 ms exposure time in each camera cycle. The molecules diffuse in the three-dimensional cytoplasmic volume whose radius is about 400 nm. With the NA 1.30 objective, detection sensitivity apparently varies little with position along the focal direction *z*. We are able to track each molecule throughout the cell volume while maintaining good localization accuracy. With the NA 1.49 objective, molecules in the horizontal slabs at the top and bottom of the cell are not detected, and molecules that

move sufficiently far up or down may exhibit shortened trajectories. Details are provided below**.** To achieve faster imaging, an iXon DV860 EMCCD camera (128×128 nm^2 pixel equivalent) was used with 750.0 μs exposure time (1.30 kHz frame rate).

Single Particle Tracking

Images were analyzed using a MATLAB program adapted from earlier work on tracking colloids by Grier and Crocker (21) and ported to Matlab by Blair and Dufresne (http://physics.georgetown.edu/matlab). A custom GUI to these particle tracking routines with additional features and enhancements is maintained by our lab. In brief, each frame was convolved with an 11 x 11-pixel Gaussian kernel with $\sigma = 1$ pixel. This acts as a low-pass filter and yields an image called **gconv**. Next, the original image is convolved with a 15 x 15 pixel "boxcar" kernel, all of whose values are 1/225. This yields an image called **bconv**. Finally, **bconv** is subtracted frame by frame from **gconv** to yield the image used for particle localization. This acts as a high-pass filter, making the image zero-based far from particle locations. A thresholding algorithm then locates bright pixels due to single molecules. The mean number of signal photons detected per molecule per frame is 290 ± 60 (one standard deviation). After smoothing, the background is sufficiently flat that the threshold for a detected particle could be set at 3 photons in the brightest pixel. The centroid of each bright feature is calculated to sub-pixel accuracy from a 10 x 10 pixel square centered at the brightest pixel and stored as an (x,y) location. At 4 ms/frame, typical single molecules yield fluorescence spread over roughly a 4 px x 4 px region.

Following the work of Michalet (22), we estimate the dynamical localization accuracy to be $\sigma \sim 40$ nm in one dimension when using the 1.49 NA objective (Supplemental Material). This includes a static contribution $\sigma_0 \sim 17$ nm. Trajectories were collected from each cell over a period of 3-5 min (40,000-70,000 frames). In that time we typically observe 1500-2000 single molecules whose trajectories comprise some 10000-12000 localizations.

Some 28-40% of all trajectories are longer than four localizations (three successive steps). For these trajectories, the mean length varied from 10 to 14 steps (11 to 15 localizations), and the longest trajectory in a cell ranged from 30 to 45 steps. For MSD analyses, we chose to analyze only longer trajectories to improve accuracy over longer lag times. In practice we varied the cutoff length for each cell. The lower limit on trajectory length was chosen in the range 12-16 steps so that at least 80–100 trajectories are included for each cell. This produces high-quality MSD plots for the first several points, which are crucial to the analysis.

Fluorescence recovery after photobleaching (FRAP)

We have also used the green fluorescence of inactivated Kaede to measure single-cell, bulk diffusion coefficients on a 1-μm length scale and 200 ms time scale by fluorescence recovery after photobleaching (FRAP), as described previously (16, 18). These comparison measurements were carried out on a Nikon Eclipse TE3000 microscope with a Nikon Plan Fluor 100x oil immersion objective (NA = 1.30) and an iXon DV897 EMCCD camera. A 488-nm laser
(Ar^+ laser, 2.2 μm FWHM, 16 kW/cm^2 peak intensity, 50 ms in duration) focused on one end of the cell photobleaches Kaede molecules in that half of the cell. In subsequent frames we measured the axial distribution of Kaede molecules as a function of time by imaging with a broad, less intense 488-nm laser (18 μm FWHM, 150 W/cm^2 peak intensity). As before, we fit the normalized amplitude of the first Fourier component of the axial intensity distribution to an exponential decay in time (16, 18). The best-fit decay rate k and the axial FWHM length of the cell L_{FWHM} were used to estimate the diffusion constant from the relationship $D_{FRAP} = kL_{FWHM}^{\ 2} / \pi^2$.

Monte Carlo Simulation

In order to distinguish the effects of confinement from possible effects of sub-diffusion, we carried out an extensive battery of Monte Carlo simulations of free diffusion in spherocylindrical volumes that

closely mimic the shape of the *E. coli* cytoplasm. The cell volume is characterized by two parameters, the length L of the cylindrical portion and the common radius R of the cylindrical portion and the hemispherical endcaps. In the simulations, some 5000 particles were started at random locations within the cell volume. Each random walk is independent of other particle positions; there are no particle-particle interactions. For each time step i, each particle chooses a displacement in each of three Cartesian directions. These displacements are chosen from a normal distribution whose standard deviation is defined by the free, three-dimensional diffusion coefficient D. That is, the mean-square displacement after time lag τ is $6D\tau$ and the standard deviation of the corresponding Gaussian propagator is $\sqrt{2D\tau}$ (Methods in membrane lipids:Chapter: Modelling 2D and 3D diffusion, Saxton). Because of the fractal nature of diffusion, the product of $D\tau$ can be in any units, which allows easy scaling of units between the experimental and simulation results. The product $D\tau$ was always chosen so that $\sqrt{2D\tau}$ is less than $0.02R$. Test runs using $\sqrt{2D\tau} = 0.002R$ gave the same results. Any move that would cause a particle to go outside the cell volume is rejected and that particle waits in its original position for the next move (23, 24). Trajectories were typically calculated for 10^3 cycles of 100 steps each. Analysis of the trajectories from simulation was similar to analysis of the experimental trajectories. All simulations were performed in MATLAB.

Results

Spatial distribution of Kaede is uniform throughout the E. coli cytoplasm

For *E. coli* cells in our standard growth conditions, we imaged single photoactivated Kaede molecules in two dimensions with 5.0 ms exposure time and 190.0 Hz frame rate. In Fig. 2.1 B we show smoothed single molecule images of single Kaede molecules in *E. coli.* Centroids of the bright spots were calculated and the dynamic localization accuracy is estimated to be $\sigma \sim 40$ nm in each dimension x and y (see Appendix A). By plotting a point at the (x,y) coordinates of the centroid of each location, we obtain a

high-resolution spatial distribution averaged over the several-minute acquisition period. Superresolution distribution of Kaede within three *E. coli* cells is shown in Fig. 2.1 C. The spatial distribution of Kaede molecules is quite uniform throughout the cytoplasmic volume. However, a quantitative examination of the spatial distribution is difficult with the high numerical aperture objective. The shallow depth of focus of the high numerical aperture objective in combination with the threshold settings in the single-particle tracking routine imposes a significant limitation on the range in z within which molecules are detected within the cytoplasm of a plated cell. In all experiments, we focus the objective on the cell center. The NA = 1.49 objective then misses molecules near the top and bottom of the cytoplasm (see Appendix B). A lower numerical aperture objective (NA = 1.30) was used to obtain an essentially complete xy spatial distribution of Kaede in the cytoplasm. In Fig. 2.1F-G we compare the measured spatial distribution of Kaede y-locations and x-locations with distributions calculated for a uniformly filled spherocylinder with L = 3.48 μm and R = 360 nm. There is no readily discernible concentration variation along x. The good agreement between experiment and calculated distributions indicates an essentially homogeneous distribution of Kaede within the cytoplasm. For 12 cells, we compared the measured distribution of particles locations along the transverse coordinate y with a theoretical distribution from a uniformly filled spherocylinder. The plots include molecules in the entire cell, including endcaps. Experiment and theory match very well. The goodness of fit at the rising edges is very sensitive to the chosen value of R. We estimate the accuracy of R to be ±30 nm. The length L of the cylindrical segment can be obtained to similar accuracy from the distributions of locations along x. Although the length of cells varies substantially over the growth cycle, the cytoplasmic radius measured in this way is quite uniform across our unsynchronized cell population. For these growth conditions, 12 cells yielded a mean radius $\langle R \rangle$ = 380 ± 30 nm.

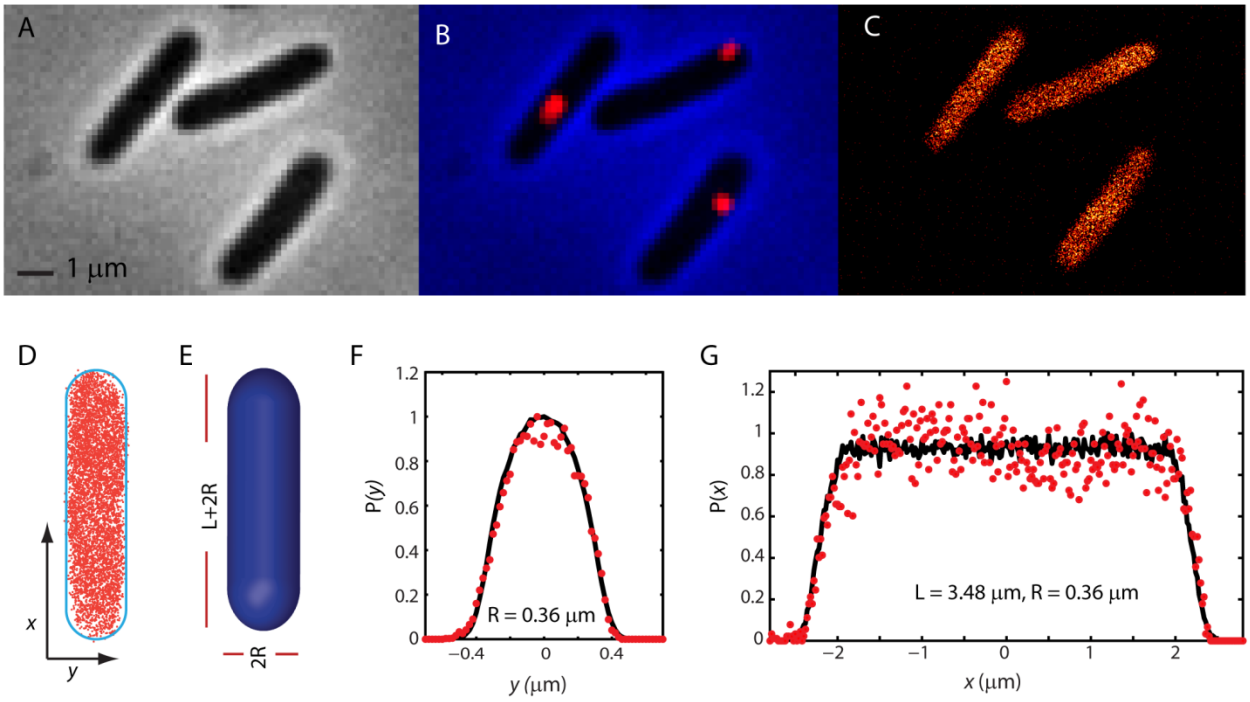

Figure 2.1: (A) Phase contrast image of three *E. coli* cells is shown. (B) An example frame with image of single Kaede molecules within those cells is shown. (C)Superresolution images of Kaede distribution within K-12 cells grown in EZRDM at 30ºC. Each localization is plotted as a point at the calculated centroid position. (D) All the positions detected from a single cell are shown. The distribution is then compared with a uniformly filled spherocylinder (E). (F-G) Experimental distribution of Kaede positions (red circles) along *x* and *y*, as obtained in a single cell by PALM using the 1.3 NA objective. Black lines are theoretical projections along *x* and *y* for a homogeneously filled spherocylinder whose dimensions *L* and *R* were chosen to match experiment.

Sub-linear MSD of diffusing Kaede

For 22 *E. coli* cells in our standard growth conditions, we tracked single photoactivated Kaede molecules in two dimensions with 4.0 ms exposure time and 234.0 Hz frame rate. Trajectory of a single Kaede trajectory is shown in Fig. 2.2 - inset. For one representative cell, plots of mean-square displacement (MSD) vs time for the coordinates x (axial dimension of the cell), y (transverse dimension), and $r = \sqrt{x^2 + y^2}$ are shown in Fig. 2.2A. The $<r^2>$ plot is scaled by a factor of ½ so that the limiting slope as $\tau \rightarrow 0$ should have the value $2D_{Kaede}$ for all three coordinates if free diffusion held. The plots include data from the 140 Kaede trajectories that were 15 steps or longer. Steps beyond the first 15 were discarded. We calculated the MSD using internal averaging over the entire 15-step trajectory, i.e., all pairs of points separated by time lag τ were used to calculate $MSD_i(\tau)$. All three MSD plots typically curve downward, with MSD_y showing non-linearity at the earliest time lag, followed by MSD_r, followed by MSD_x. This is qualitatively consistent with free diffusion confined in a spherocylinder. The MSD plots for different cells differ significantly from each other, consistent with the previously observed dispersion in effective diffusion coefficient of cytoplasmic GFP among identically treated cells. We carried out a variety of quantitative tests to learn how best to extract the mean diffusion coefficient for each cell and whether sub-diffusion may contribute to the curvature.

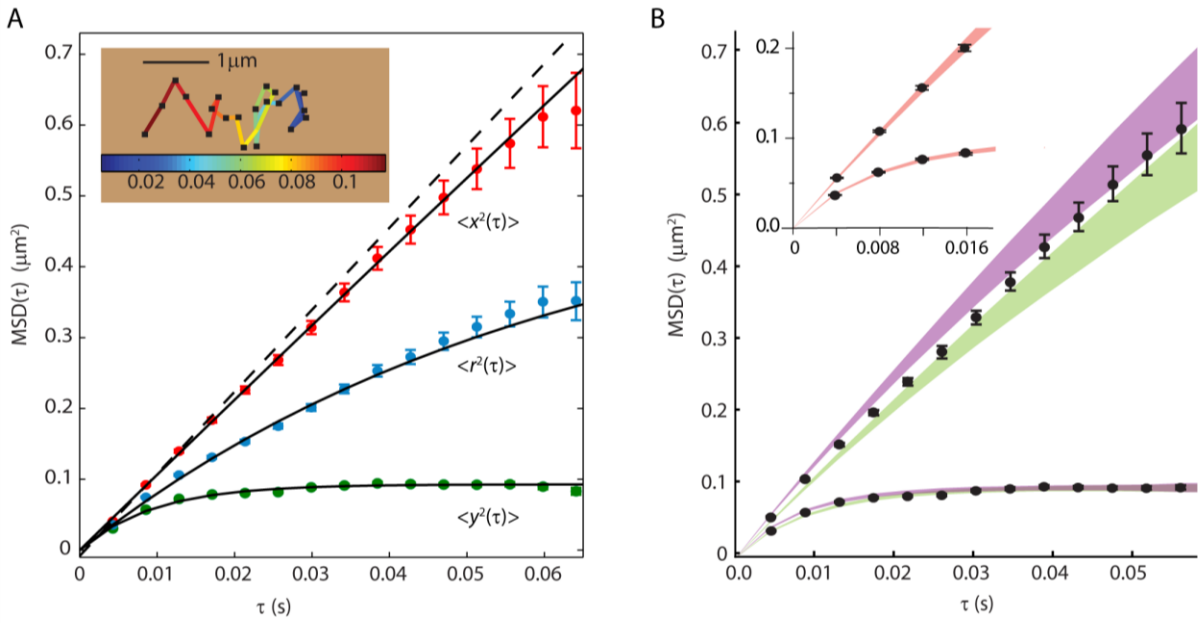

Figure 2.2: (A) Mean-square displacement (MSD) vs time lag τ for the coordinates *x* (axial dimension), *y* (transverse dimension), and $r = \sqrt{x^2 + y^2}$. Data include 160 trajectories of 15 steps from a single cell. Error bars are ± one standard deviation of the mean values. Solid lines are least-squares fits to the equation: $\mathrm{MSD}_i(\tau) = A_i(1 - \exp(-\tau/T_i))$. The dotted line is a linear fit to the first 3 points of MSD_x. *Inset:* Two-dimensional projection of trajectory with time encoded by color as shown. (B) Comparison of experimental MSD plots from (A) (circles) with those from Monte Carlo simulations of free diffusion in a spherocylinder of cylinder length L = 3.950 μm and radius R = 400 nm. Dimensions were chosen to match the spatial distribution of Kaede molecules in the cell of (A). Solid points are MSD along *x* and *y* from (A); error bars show one standard deviation of the mean value. Color swaths show the spread of MSD values (± one standard deviation of the mean) for 200 Monte Carlo runs using D = 6.8 μm^2-s^{-1} (green) and 8.4 μm^2-s^{-1} (purple). *Inset:* Comparison of the first 4 experimental data for MSD_x and MSD_y with the spread of MSD values from the Monte Carlo simulations using D = 7.6 μm^2-s^{-1} (red swath).

Determination of the diffusion coefficient from MSD plots

The simplest way to estimate D_{Kaede} is to measure the limiting slope of MSD plots as $\tau \rightarrow 0$. For the example cell of Fig. 2.2, the slope of the first three points yields the estimates 4.4 $\mu m^2 \cdot s^{-1}$ from MSD_r and 6.3 μm2·s−1 from MSD_x. This is a serious discrepancy. The limiting slope of MSD_y is not well defined for data taken at 4.3 ms per frame.

A more rigorous method for determining D_{Kaede} uses Monte Carlo simulations in a spherocylinder whose dimensions R and L are determined by the experimental spatial distribution of the Kaede locations measured by superresolution imaging. For the particular cell in Fig. 2.2, the best-fit values for the dimensions of spherocylinder gave a total length $L + 2R = 4350 \pm 100$ nm and a radius $R = 400 \pm 25$ nm, which yields the cylinder length $L = 3950 \pm 125$ nm by subtraction. With the geometry parameters R and L now fixed, the value of D_{Kaede} in the Monte Carlo simulations is adjusted until the initial rise of the MSD_x and MSD_y curves is well fit over the first four points (inset of Fig. 2.2B), as judged by eye. Fitting of MSD_x and MSD_y ensures a good fit to MSD_r. Repeated averaging of Monte Carlo runs of 160 trajectories of 15 steps each gives a sense of the accuracy of the fitting procedure. The statistical noise is modest over these first four points, but becomes substantial at longer lag times (Fig. 2.2B). To illustrate the statistical variability of the calculations and the sensitivity to the input value of D_{Kaede}, we show bands of outcomes (swaths in Fig. 2.2B) for the $MSD_x(\tau)$ and $MSD_y(\tau)$ for assumed diffusion coefficients of 6.8 and 8.4 $\mu m^2 \cdot s^{-1}$. Based on the Monte Carlo results, our best estimate for the cell in Fig. 2.2 is $D_{Kaede} = 7.6 \pm 0.8$ $\mu m^2 \cdot s^{-1}$. The uncertainty estimate takes approximate account of the sensitivity of the simulation results to all three inputs (D_{Kaede}, L, and R) plus the variability of MSD plots for short samples of confined random walks.

It is impractical to measure R and L and to run a series of Monte Carlo simulations for each cell, and in many cases experimental conditions may not permit fast enough frames with high enough signal/noise ratio to enable accurate extraction of D from the initial slope of MSD_x. Following Anderson and Cherry

(25), we have developed a third, phenomenological method grounded in the Monte Carlo simulations. There are analytical results for MSD(τ) for confined diffusion in the simple model problems of diffusion on a line segment (1D), within a circle (2D), and within a sphere (3D) (26). In all three cases, the functional form of $MSD(\tau)$ is a sum of multiple exponentials:

$$MSD(\tau) = A + B\sum_{n=0}^{\infty} k_n \exp\left(-\frac{j_n \cdot \tau}{T}\right) \tag{2.1}$$

As $\tau \to 0$, $MSD(\tau)$ reduces to $2dD\tau$, the result for free diffusion. Here d is the dimensionality. The linear approximation of $MSD(\tau) = 2dD\tau$ is a poor estimator of D when the timescale of confinement is similar to the timescale of the measurement, as occurs in our Kaede measurements. When truncated to the first term, the result for all three model problems has the generic form:

$$MSD(\tau) \approx A \cdot \left(1 - \alpha \cdot \exp\left(-\frac{D \cdot \beta \cdot \tau}{A}\right)\right) \tag{2.2}$$

Here A is the asymptotic amplitude of MSD(τ), which depends on the shape of the model problem. The parameters α and β also depend somewhat on shape. The values for A, α, and β are shown in Table 2.1.

	Line segment (1D)	**Circle (2D)**	**Sphere (3D)**
A	$L^2/6$	R^2	$6R^2/5$
α	0.986	0.997	0.998
β	1.65	1.69	1.73

Table 2.1. Leading terms in exact solutions to diffusion on a line segment, in a circle, and in a sphere.

In these model problems, it is only at moderate timescales that the first term dominates. At shorter and shorter times, the higher terms in the expansion contribute more. That is, as $\tau \rightarrow 0$, MSD(τ) $\rightarrow 2dD\tau$, but this is not the short-time limit of the truncation in Eq. 2.2.

For the spherocylindrical geometry of interest, we chose to fix α at 1 and fit β_x and β_y to the results of Monte Carlo simulations of MSD_x and MSD_y. Because MSD_r is the sum of MSD_x and MSD_y and each component has its own time constant, MSD_r is not expected to be well fit by a single rising exponential. As shown in Fig. 2.3 for fixed R = 400 nm, when the entire spherocylindrical volume is used in the simulations, the best value of β_x varies approximately linearly from about 1.72 to 1.86 as the cylinder length varies from 0.5 to 5.5 μm. The best value of β_y is almost constant at 1.69 over the same range. We recommend use of MSD_x for the most accurate determination of the diffusion coefficient because it better defines the ratio of asymptotic limit to time constants (A_x/T_x, Eq. 2.3), which in turn best evaluates D.

Monte Carlo simulations confined to the central slab detected by the NA = 1.49 objective show that β_y depends significantly on the depth of focus while β_x does not. This is in accord with the effects of depth of focus on the asymptotes of MSD_y and MSD_x described above. As shown in Fig. 2.3, the slab simulations change the best β_x very little, while the best β_y value changes to from 1.68 to 1.35. This provides yet another reason to use MSD_x to analyze the tracking data obtained with the high numerical aperture objective.

Based on this we approximate each $MSD_i(\tau)$ curve as a rising single exponential function:

$$\mathrm{MSD}_i(\tau) = A_i\left(1 - \exp(-\frac{\tau}{T_i})\right), \qquad (2.3)$$

with $i = x$, r, and y. The solid lines in Fig. 2.2A show the quality of the fits to these equations. Each dimension has its own asymptotic amplitude A_i and time constant T_i. For spherocylinders with L/R ratio

similar to that of *E. coli*, the diffusion coefficient D, the time constant T_i, and the asymptotic amplitude A_i are related approximately as follows:

$$D = A_i / (\beta_i d_i T_i). \quad (4)$$

Here d_i is the dimensionality (one for x or y, two for r). A simple procedure with good accuracy fits the entire MSD_x curve to Eq. 2.3 to obtain A_x and T_x and then uses the best value of β_x to extract D. For the cell in Fig. 2.2 the best-fit values are $A_x = 2.53$ μm^2 and $T_x = 0.198$ s. For the length of this cell, the prescribed value $\beta_x = 1.75$ then yields the diffusion constant $D_{Kaede} = 7.3$ μm^2-s^{-1}, in excellent agreement with the value 7.6 μm^2-s^{-1} from the more rigorous treatment using L and R from the PALM spatial distributions and varying D_{Kaede} in Monte Carlo simulations to match all three MSD curves. Use of MSD_y is much less accurate.

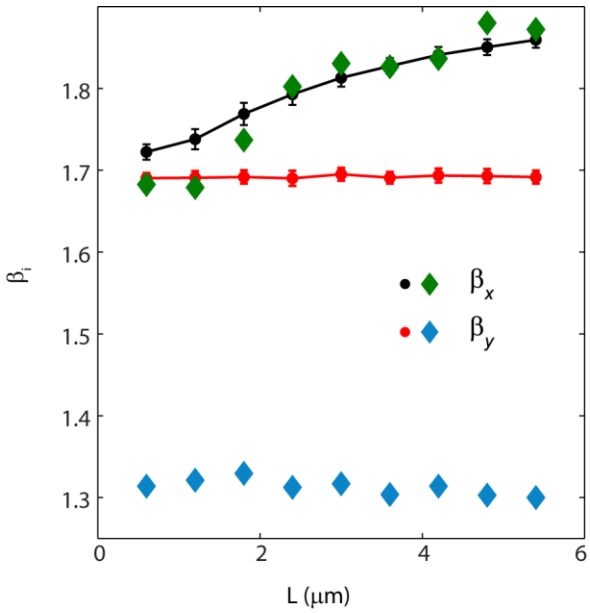

Figure 2.3: Best values of β_x and β_y from 15-step, 4 ms/step, $D = 7$ μm^2-s^{-1} Monte Carlo trajectories in model spherocylindrical volumes as a function of length L with radius fixed at $R = 400$ nm. Circles use all trajectories. Diamonds use only those trajectories remaining in the 500 nm central slab.

This A_x, T_x method is more accurate than the straight-line fit to the first three points of MSD_x (which yields 6.3 μm^2-s^{-1} in this example), but shares the advantage of not requiring accurate values of L and R

for each cell. For eight cells examined in detail, the A_x, T_x method gives the same diffusion constant as the Monte Carlo-based fitting method with a root- mean-square deviation of only 6%. This simplified method was used to generate the diffusion coefficients for the 22 cells studied in detail. The mean is $\langle D_{Kaede} \rangle = 7.3 \pm 1.1\ \mu m^2\text{-}s^{-1}$, where the ± value is one standard deviation of the distribution of values across cells. A histogram of measured values across cells is shown in Fig. 2.7D.

To summarize, for Kaede diffusion in the *E. coli* cytoplasm with our quality of data and frame rate, fitting the first three points of $MSD_x(\tau)$ to a straight line systematically underestimates D_{Kaede} by about 15%. Fitting the initial slope of $MSD_r(\tau)$ or $MSD_y(\tau)$ is highly inaccurate. The A_x, T_x method that fits $MSD_x(\tau)$ to Eq. 2.3 and extracts D from the ratio of A_y/T_y using $\beta_x = 1.70–1.85$ (depending on cell length) is accurate to about 6%. Here the benchmark for the most accurate value of D_{Kaede} is the fit based on Monte Carlo simulations in a spherocylindrical geometry determined from the PALM spatial distribution. We have not addressed possible systematic error arising from the modest tapering of the cell body.

Tests for heterogeneity and sub-diffusion

Thus far we have established that a model of free, homogeneous diffusion within the confining volume of the cytoplasm can quantitatively reproduce the nonlinear MSD plots from experiment. However, plots of $MSD(\tau)$ do not incisively distinguish confined, free diffusion from heterogeneous diffusion (different particles having different diffusion coefficients, perhaps subject to time-dependent exchange) or from sub-diffusion (sub-linear evolution of MSD).

For the set of trajectories from a single cell, the distribution of $\langle x^2(\tau) \rangle_j$ and of $\langle r^2(\tau) \rangle_j$ among different molecules provides a good test for heterogeneity (27). Here the brackets denote an average over all displacements of lag time τ within the trajectory of a single molecule j. There is no averaging over molecules, as was the case for the MSD plots. We examine the distribution of $\langle x^2(\tau) \rangle_j$, rather than $\langle r^2(\tau) \rangle_j$, in order to minimize confinement effects. As witness the limiting slope, MSD_r is distorted by

confinement effects even at the shortest time lags used in our measurements. For homogeneous diffusion in two dimensions, the distribution of $\langle r^2(\tau)\rangle_j$ has been given analytically (27). There is no analytical form for the case of free diffusion in one dimension. Therefore, we have compared the distribution of $D_{x,j} = \langle x^2(\tau)\rangle_j/2\tau$ from experiment with that from Monte Carlo simulations using the mean value $\langle D_{Kaede}\rangle$ = 7.3 ± 1.1 μm^2-s^{-1}. The comparison is shown in Fig. 2.4 for the particular time lag τ = 12.9 ms (three steps), which is only slightly perturbed by confinement. The close agreement provides further support for homogeneous, free diffusion on the time and length scales studied. However, we will demonstrate below clear evidence of mild *spatial heterogeneity* in D_{Kaede} which goes undetected in this three-step $D_{x,j}$ test.

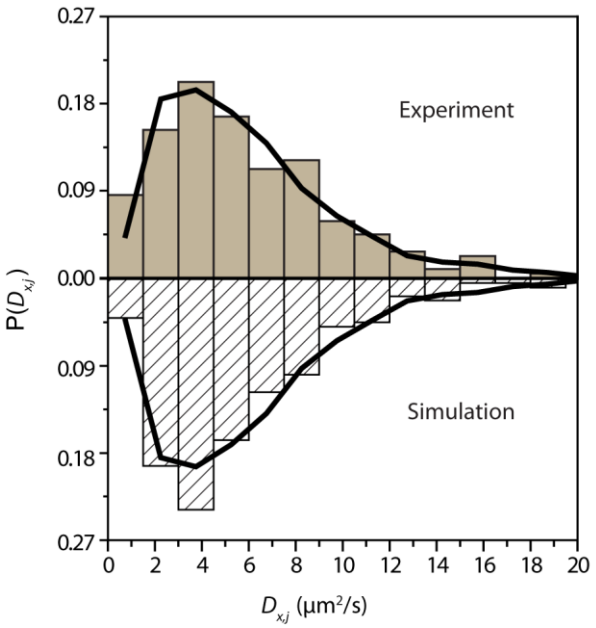

Figure 2.4: *Solid bars:* Experimental distribution of $D_{x,j} = \langle x^2(\tau)\rangle_j/2\tau$ (3-step estimate of D_x for molecule j, taken from 13-step trajectories of 180 molecules in a single cell). *Striped bars:* Same, for 180 13-step trajectories from Monte Carlo simulations with D = 7 μm^2-s^{-1}, chosen to match the experimental value for that cell. The solid curves show the smooth distribution of $D_{x,j}$ obtained from 2000 Monte Carlo trajectories.

To distinguish sub-diffusion from confined, free diffusion, Ferrari *et al*. (15) have devised statistical tests based on higher moments of the displacement vector. Accordingly, we have analyzed moments of the *x*-coordinate of the displacement vector up to the 10th moment. The analysis assumes that each moment depends on time as a power law, as in $\langle x^n(\tau)\rangle = k_n \tau^{\lambda n}$ with k_n a constant. The scaling coefficients λ_n are determined by a linear least square fit of log($\langle x^n\rangle$) vs log τ (Fig. 2.5). A plot of the λ_n vs *n* is called a moment scaling spectrum (15, 28, 29). For strongly self-similar processes such as free

diffusion, the plot is a straight line through the origin; λ_0 is always equal to 0. For Kaede diffusion data from a particular cell, Fig. 2.5 shows that the moment scaling spectrum is linear with a slope of 0.46. The average slope over eight cells analyzed in this way was 0.44 ± 0.06 (±1σ). This is smaller than the theoretical value of 0.50 for free diffusion in an infinite space. The difference is probably due to confinement. For comparison, our Monte Carlo simulations in spherocylinders of realistic aspect ratio with trajectory lengths mimicking the experiments yielded linear moment scaling spectra with a slope of 0.46 ± 0.05. Thus, the experimental deviation from a slope of 0.5 is well reproduced by the Monte Carlo results. In free space, the Monte Carlo result was 0.49 ± 0.05. The Ferrari test yields no evidence of Kaede subdiffusion on the timescale and length scale studied here (4 ms and 100 nm, respectively).

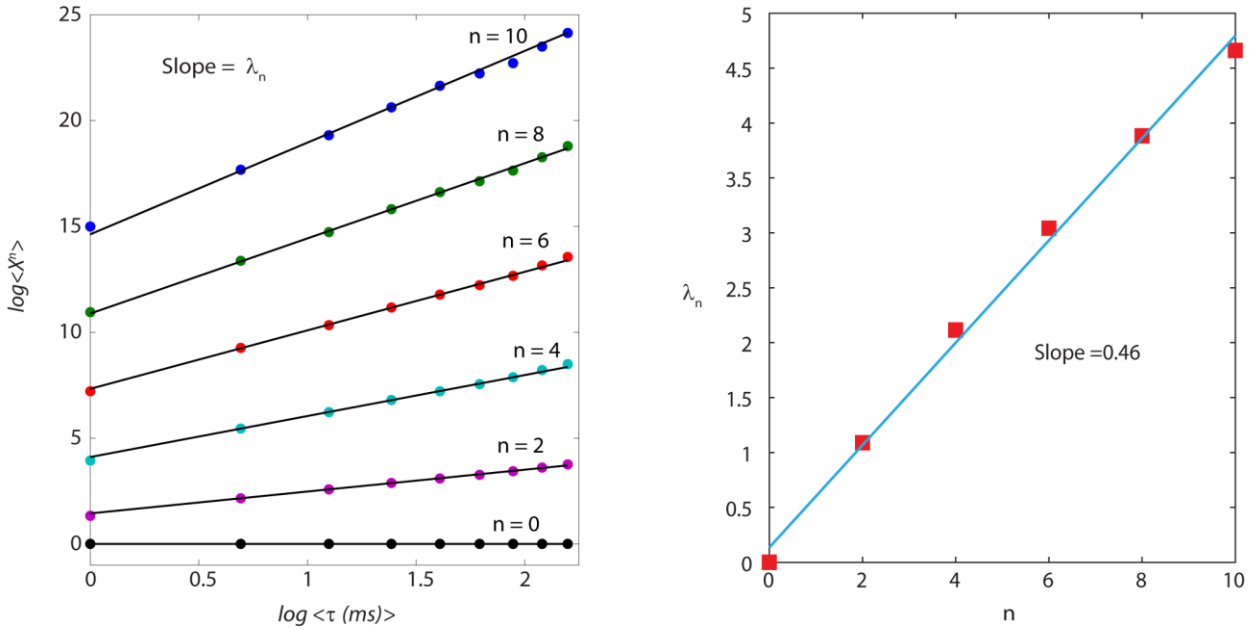

Figure 2.5: (A) For a single cell, plot of log of 0-10th moments of *x* displacements vs log τ.(B) The slope of each curve in 5A gives the scaling factor λ_n for nth moment. Moments of the scaling spectrum for moments 2-10 obtained from 160 15-step Kaede trajectories from a single cell. Linear least-squares fit (solid line) gives slope of 0.46.

Superfast tracking of Kaede

For five additional cells we tracked single molecules with 750.0 μs exposure time and 1.30 kHz frame rate, 5.5 times faster than the rate used for most of these studies. At this frame rate, the laser intensity must be increased to the point at which trajectories are much shorter, typically 4-6 steps. For these cells, MSD_x is linear over the time scale 0.75–5.00 ms (Fig. 2.6A). From the slope of MSD_x plots, we obtained $\langle D_{Kaede} \rangle = 7.2 \pm 3.3\ \mu m^2/s$ (mean of five cells ± one standard deviation), consistent with the more elaborate analysis of the data taken at the slower frame rate. There is no evidence of sub-diffusion on the time and length scales of 1 ms and 100 nm.

Relief of confinement with Cephalexin

As a final experimental test, we used the drug cephalexin to inhibit normal cell division and create cells of 8-12 μm length (30). The cells continue to elongate but are unable to divide and retain roughly the same diameter. For imaging of cephalexin-treated cells at 234 Hz, $MSD_x(\tau)$ remains linear for at least nine steps, while MSD_y reaches an asymptote similar to the unperturbed case (Fig. 2.6B). For 11 cells treated with cephalexin, the initial slope of $MSD_x(\tau)$ up to the 8^{th} data point was used to obtain the mean diffusion coefficient $\langle D_{Kaede} \rangle = 7.2 \pm 1.5\ \mu m^2/s$, quite similar to the value of $7.3 \pm 1.1\ \mu m^2\text{-}s^{-1}$ for untreated cells. This result again indicates that the nonlinearity in the MSD plots for untreated cells arises from confinement effects.

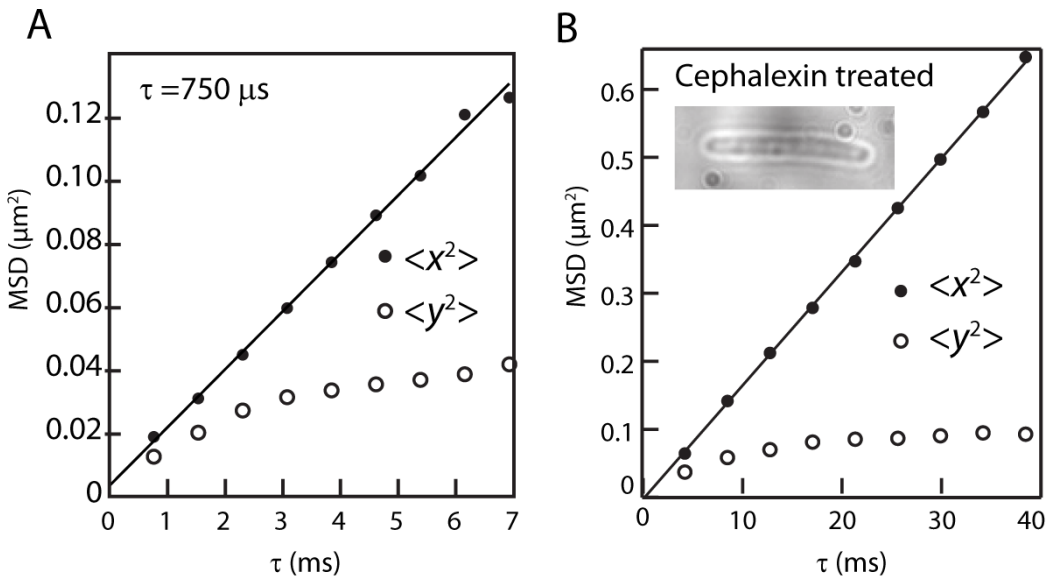

Figure 2.6: (A) MSD_x and MSD_y from tracking Kaede at 750 μs frame time. Solid line a linear fit to the first 6 points. (B) MSD_x and MSD_y from tracking Kaede in a cell treated with cephalexin to stop cell division. White light image of the cell is shown in inset. Solid line is linear fit to first eight data points.

Comparison of single-particle tracking data with FRAP data

The average diffusion constant for the entire population of Kaede molecules in a cell can also be obtained from fluorescence recovery after photobleaching (FRAP) (16) using green fluorescence from un-photoswitched Kaede excited at 488 nm. For 23 cells under the same growth conditions, the mean diffusion constant is $\langle D_{FRAP} \rangle = 8.3 \pm 1.6$ μm^2-s^{-1}, in reasonably agreement with the value $\langle D_{Kaede} \rangle = 7.3 \pm 1.1$ μm^2-s^{-1} from the 22 SPT measurements. The two distributions of measurements are compared in Fig. 2.7D. A two-sided Kolmogorov-Smirnov test indicates that the distributions are not significantly different ($p = 0.\mathbf{03}$). On the 50 ms time scale and 2 μm length scale of the FRAP experiment, all the Kaede molecules are mobile; the mobile fraction is $f_{mobile} = 0.98 \pm 0.05$.

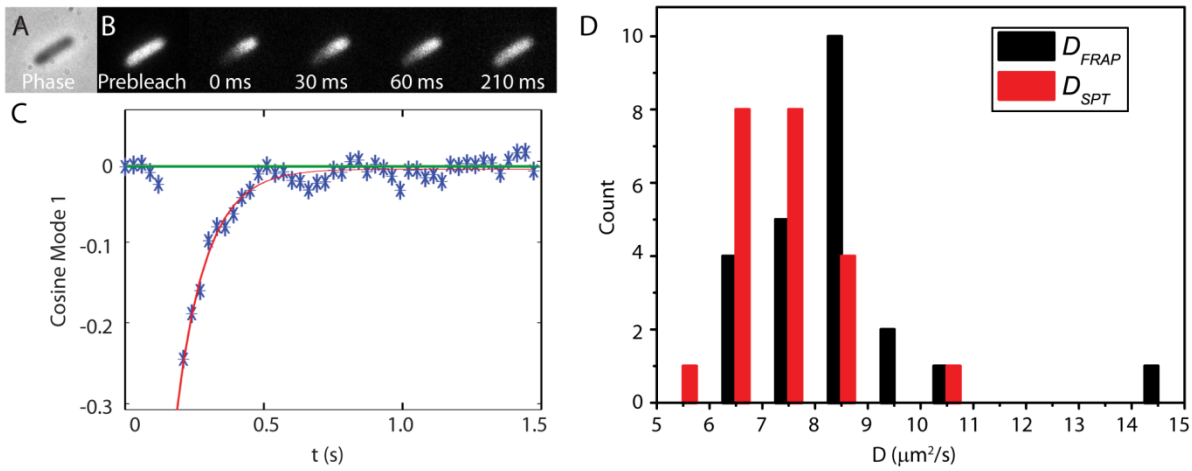

Figure 2.7: (A) Phase contrast image of the cell and (B) fluorescent image for one frame before and 4 frames after the bleach is shown. (C) Recovery of the signal and quality of single-exponential fit is shown. (D) Histograms of D_{SPT} (D_{Kaede} from single particle tracking) from 22 cells and D_{FRAP} (D_{Kaede} from FRAP) from 23 cells.

Diffusion map of Kaede is inhomogeneous

In *E. coli* cells growing moderately rapidly, the chromosomal DNA is partially segregated to form two nucleoid "lobes" near the central axis of the cell while the ribosomes segregate to the end-caps, and the region between nucleoid lobes. This axial segregation is shown clearly in images of the same cell with ribosomes labeled by the protein S2-eYFP and with DNA stained by the dye DRAQ5 (Fig. 2.8A-C). The DNA density typically peaks at the ¼ and ¾ axial positions. Such spatial heterogeneity within the cytoplasm opens the possibility that proteins may diffuse differently in different regions of space.

For the longer *E. coli* cells, where the extent of chromosomal segregation is high (see Appendix C), the localization accuracy of PALM indeed reveals mild spatial heterogeneity in the diffusion of Kaede. In Fig. 2.8D we show a false-color, two-dimensional diffusion map of $\langle D_x(x,y)\rangle = \langle \Delta x^2(x,y)/2\Delta t\rangle$. This is

the locally averaged, one-step estimate of the mean diffusion coefficient in the x direction at position (x,y). Here Δx^2 is the squared one-step displacement vector along x, (x,y) is the initial location, and Δt = 4.3 ms is the camera frame duration. Each step of each trajectory generates a value of D_x at a position (x,y). In order to form a smoothed two-dimensional diffusion map, we placed a 50 nm × 50 nm grid of points over the cell image. Each grid point is assigned the mean of all values of D_x for initial positions lying inside a 200 nm × 200 nm square centered at that grid point. Some 10,000 single-step values contribute to the map. The false-color image reports mean values at each grid point.

To form the one-dimensional axial diffusion map $\langle D_x(x)\rangle_y$ shown in Fig. 2.8E, we calculate the weighted mean of D_x over all y values at each value of x. This highly averaged local diffusion constant dips in the center of the cell. The error bars show one standard deviation of the mean of all values averaged at each axial position. In the same growth conditions, images of DRAQ5-stained DNA (Fig. 2.8B) show two nucleoid lobes, peaking at the ¼ and ¾ axial positions, for essentially all cells. Images of ribosomes (Fig. 2.8A) show three intensity peaks: two in the endcaps and one between the nucleoid lobes. For the particular cell of Fig. 2.8D, apparently on average Kaede diffuses some 20-30% more rapidly in the DNA-rich nucleoid regions than in the ribosome-rich space between nucleoid lobes. The end caps of the cells are also ribosome-rich and the mean local diffusion constant is also lower at the end caps. However, confinement also has a strong effect on $\langle D_x(x)\rangle_y$ in the end caps and we have not attempted to separate the two effects.We applied such analysis to eight cells varying in length from 2.9—4.4 μm. The three longest cells and two of the five moderate-length cells show a clearly discernible dip in $\langle D_x(x)\rangle_y$ at the mid-cell plane. Among these cells, the peak to valley ratio varied in the approximate range 1.3–1.8. Three of the five moderate-length cells did not show a clear dip. We suspect this variability arises from the variable degree of segregation of the chromosomal DNA among unphased cells. In the longer cells, which tend to be at a later stage of the cell cycle, DRAQ5 staining of the DNA shows a deeper minimum in DNA density at the mid-cell plane.

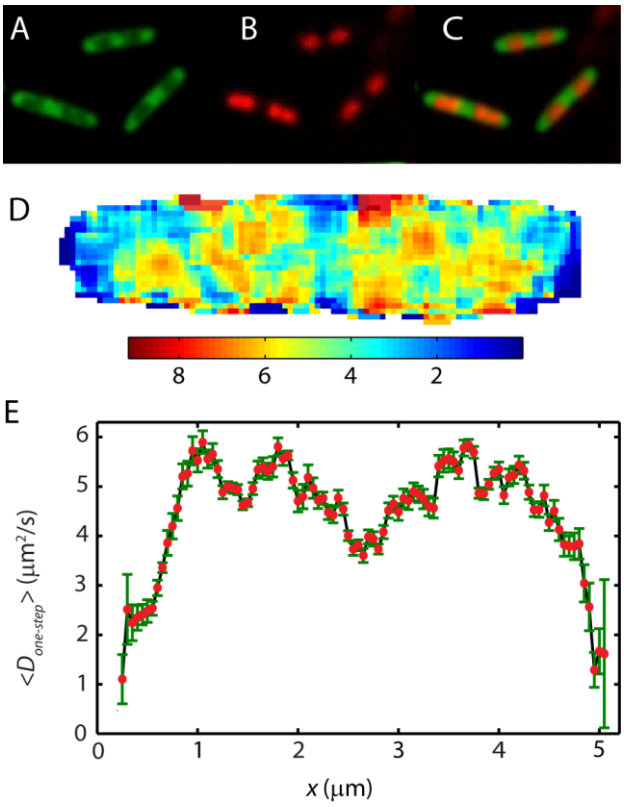

Figure 2.8: (A) Widefield image of ribosomes in three cells labeled by S2-eYFP construct. (B) Widefield image of chromosomal DNA in the same cells labeled by DRAQ5. (C) Overlay of the ribosome and DRAQ5 images, showing strong ribosome-nucleoid segregation. (D) Two-dimensional Kaede diffusion map in a different cell, with $\langle D_x(x,y)\rangle$ plotted as a false color map with scale as shown. See text. (E) One-dimensional Kaede diffusion map $\langle D_x(x,y)\rangle_y$ obtained as the weighted average of the 2D map over all y values at each x.

This mild spatial heterogeneity is not inconsistent with the apparent homogeneity inferred from analysis of the $MSD_i(\tau)$ plots and from the distribution of three-step mean-square displacements along x (Figs. 1 and 4). Both were derived from averages over all molecules within a cell. As is well known, short trajectories of diffusive motion are very noisy. In the Appendix D we present Monte Carlo simulation results showing that at short times, two equal populations with time-independent diffusion coefficients differing by as much as a factor of *two* will yield MSD plots and also three-step $D_x(\tau)$ distributions that are effectively indistinguishable from those of a homogeneous population of intermediate diffusion coefficient (Fig. 2.A.4). Evidently the spatial diffusion maps provide a particularly sensitive way to detect heterogeneity in diffusion behavior.

Discussion

Similar to GFP, Kaede is a β-barrel of mass 28 kDa. In a typical cell under our expression conditions, the PALM method can activate and detect some 8000 single Kaede copies. For a cell volume of ~3 μm^3, this corresponds to a *detected* Kaede concentration of ~5 μM. This should be considered a lower bound on the Kaede concentration because not all Kaede molecules will activate properly.

A crystal structure of Kaede found compact homo-tetramers in which one pair of parallel, side-by-side barrels is bound to another pair of parallel, side-by-side barrels whose long axes are rotated by 90º relative to the first pair (10, 11). Multi-angle light scattering measurements in buffer solution determined the mass of Kaede particles to be 116.0 kDa, 4.3 times the monomer mass. This indicates that Kaede forms stable, compact homo-tetramers in solution, much like the closely related DsRed protein (11). In the crowded conditions of the *E. coli* cytoplasm, the propensity to form tetramers will be stronger than in buffer.

In contrast, an FCS study of Kaede in buffer at low-nM concentrations (19) obtained diffusion coefficients of 17 μm^2-s^{-1} for the green-fluorescent form and 10 μm^2-s^{-1} for the yellow-fluorescent (photoswitched) form, compared with 67 μm^2-s^{-1} for monomeric GFP. The Stokes-Einstein relation for diffusion of spheres of radius r in solution of viscosity η is: $D = k_B T/6\pi\eta r$, where T is the temperature and k_B is the Boltzmann constant. If the hydrodynamic radius of a globular protein scales roughly as the cube root of its mass, then $Kaede_4$ should diffuse more slowly than GFP monomer by roughly a factor of the cube root of 4, which is 1.59. Measurements of the GFP diffusion coefficient in buffer vary from 67 to 87 μm^2-s^{-1}, so we would expect $D(Kaede_4)$ in buffer to be ~50 μm^2-s^{-1}. The Stokes relationship then suggests that the FCS results involve very large aggregates containing HU*ndreds* of Kaede molecules, even though the concentration is in the low-nM range. The FCS and light scattering results are not readily reconciled.

We proceed under the assumption that the species under study in the *E. coli* cytoplasm is $Kaede_4$. Cytoplasmic protein diffusion coefficients are evidently controlled by a variety of factors (16, 18, 31-34) including the "microviscosity" of the medium, crowding and hydrodynamic effects (35), transient binding to less mobile elements, and possible "sieving" effects due to the tortuosity of space in the presence of complex, branched biopolymers (DNA–RNA polymerase–mRNA–ribosome–polypeptide chains).

In related work on the DH5α strain of *E. coli* in LB media at 37°C, Mullineaux and co-workers studied diffusion of a series of head-to-tail GFP oligomers of stoichiometry TorA-$(GFP)_n$, with n = 2–5. The mean diffusion coefficient decreased monotonically with increasing oligomer size, from 9.1 μm^2-s^{-1} for the monomer to 5.5 μm^2-s^{-1} for the tetramer. This decrease by a factor of 1.65 is in good accord with Stokes-Einstein scaling. In the DH5α strain of *E. coli* cytoplasm studied here, the mean diffusion coefficient for Kaede is $\langle D_{Kaede} \rangle = 7.3 \pm 1.1$ μm^2-s^{-1}. This is a factor of 1.3 smaller than that of GFP monomers in an MG1655 strain under the same growth conditions ($\langle D_{GFP} \rangle = 9.5 \pm 2.3$ μm^2-s^{-1}) (18, 30). The result is roughly consistent with Stokes-Einstein scaling assuming a tetrameric Kaede. In particular, there is no evidence that cytoplasmic Kaede exists as large oligomers like those suggested by the FCS diffusion study in buffer (19).

The result is also in reasonable agreement with the predicted size dependence of diffusion coefficients in cytoplasm from the hard-particle simulation of Elcock and co-workers (36). The model treats the cytoplasm as a well-mixed collection of hard, realistic protein shapes, but does not include hydrodynamic effects. The strength of attractive forces between proteins is adjusted to fit the GFP diffusion coefficient. For crowding at the level of 300 mg/ml, the simulations would predict $D(Kaede_4)$ = 6.6 μm^2-s^{-1}. However, the degree of crowding in K12 *E. coli* in our growth conditions is not known.

Analysis of the MSD(τ) plots in terms of confinement and of the distribution of single-molecule D_x values estimated from three-step trajectories (Fig. 2.4) shows that the data are consistent with free,

homogeneous diffusion of $Kaede_4$ on the 1–200 ms time scale and the 100 nm–1 μm length scale probed here. Furthermore, the analysis of higher moments of the displacements compared with Monte Carlo results in model spherocylinders found no evidence of sub-diffusion (Fig. 2.5).

However, the diffusion map of Fig. 2.8 reveals moderate spatial heterogeneity in Kaede diffusion, especially for longer cells. A recent study of the monomeric GFP-like protein mEos2 did not report spatial heterogeneity (9). In longer cells the segregation of the two nucleoid lobes is more complete (see Appendix C), which correlates with greater enrichment of ribosomes in the region *between* nucleoid lobes (at the mid-cell plane). The correlation between the enrichment of ribosomes at mid-cell and the decrease of the Kaede diffusion constant at the mid-cell strongly suggests that Kaede diffuses slightly less rapidly in the ribosome rich regions than in the dense parts of the nucleoid.

The cause is uncertain. Although Kaede is uniformly distributed in the cytoplasm (Fig. 2.1 C -G), other cytoplasmic components need not be, as witness DNA and ribosomes. Spatial partitioning of other cytoplasmic components may underlie the spatial inhomogeneity in Kaede diffusion. One possibility is that the ribosome-rich regions are significantly more crowded than the nucleoid; they contain essentially all the ribosomes plus a substantial share of globular proteins as well. Larger globular proteins may preferentially partition to the ribosome-rich regions due to excluded volume effects of the DNA. Enhanced crowding within the ribosome-rich regions would be in accord with slower diffusion there. It is also possible that the oligomeric state of Kaede varies in space. Stronger crowding effects within the ribosome-rich region may enhance formation of Kaede tetramers or higher oligomers.

In contrast, possible sieving effects due to the structure of DNA in the nucleoid would work in the direction opposite the observations, causing slower diffusion within the nucleoid. Evidence for sieving by the branched polymeric structure within the nucleoids is mixed. In a study of diffusion vs protein mass, the fall-off in globular protein diffusion coefficient with size was much faster than predicted by the Stokes-Einstein equation, suggesting sieving (34). Enhanced hydrodynamic effects for larger particles

provides an alternative explanation for this fall-off (35). In contrast, the fall-off of diffusion coefficient with size for a sequence of end-to-end GFP oligomers (32) and also the diffusion coefficient of $Kaede_4$ reported here are in reasonable accord with Stokes-Einstein expectations. Comparisons of diffusion coefficients among different proteins may be complicated by differential transient binding to less mobile species in the ribosome-rich region vs the nucleoids. However, the preponderance of evidence indicates that sieving effects are unimportant for globular proteins of mass ~100 kDa or smaller.

The homogeneous spatial distribution of Kaede (uniform concentration) within the entire cytoplasm lies in sharp contrast with earlier work using immersive refractometry to estimate the volume fraction of protein inside vs outside the nucleoids (12). That study found substantially larger protein density in the ribosome-rich regions as compared with the nucleoids. Odijk (13) used this result to build a model of the cytoplasm in which the nucleoid and the surrounding cytoplasm are in osmotic equilibrium, with proteins 2.5 times more concentrated outside the nucleoid than inside. It remains possible that differential partitioning between the nucleoid and ribosome-rich regions indeed occurs for other proteins.

Interpretation of high-resolution single-particle tracking data in rod-shaped cells is always complicated by the possibility of confounding effects due to confinement and sub-diffusion. We have shown that separation of motion along the axial coordinate x from that along the transverse coordinate y is useful in minimizing confinement effects on MSD plots. Curve-fitting to the model exponential function of Eq. 2.1 with the parameter β_i set to 1.70-1.85 (depending on cell length) can further enhance the accuracy without requiring extensive Monte Carlo modeling for each cell. A long-term goal is to measure spatial distributions and diffusion maps for a variety of proteins of varying size and charge in order to provide a much more complete picture of cytoplasmic organization.

Appendix A: Localization precision

We have tracked static Kaede molecules dried on glass using similar laser power and other instrument settings as those in our live cell tracking experiments. The average number of signal photons detected per Kaede molecule is 323, which is similar to the value 306 in live cells. Based on repeated measurements of the same static molecule, the estimated standard deviation of the position localization is $\sigma_0 = 17 \pm 6$ nm in x and $\sigma_0 = 18 \pm 7$ nm in y dimension (85 particles).

However, in the live cells the particle diffuses during the camera frame. This causes the localization error to be higher than the standard deviation of the point spread function. A careful analysis (22) showed that the localization error of a diffusing particle can be estimated as:

$$\sigma = \sqrt{\sigma_0^{\,2} + D \cdot t_E / s_0^2} \qquad \text{(A1)}$$

Here σ_0 is the static-molecule error, D is the diffusion coefficient, t_E is the frame duration, and s_0 is the standard deviation of the point-spread function. For Kaede in live cells studied with the 1.49 NA objective, we have $\sigma_0 = 17$ nm, $D_{Kaede} = 7.3$ μm^2-s^{-1}, $t_E = 4$ ms, $\lambda = 580$ nm, and $s_0 = 0.2\lambda/NA = 78$ nm (37). This yields the estimated *dynamic* localization error of $\sigma = 40$ nm for tracking Kaede in live cells. This number is used in fitting the observed spatial distributions along x and y to obtain cell dimensions R and L; see below.

Appendix B: Depth of field

Two microscope objectives were used for tracking Kaede; the nominal numerical apertures (NA) are 1.30 and 1.49. In combination with the threshold settings in the single-particle tracking routine, the high NA objective imposes a significant limitation on the range in z within which molecules are detected within the 760-nm diameter cytoplasm of a plated cell. The most complete measurements of Kaede spatial distributions come from the NA = 1.30 objective due to its larger depth of focus. In all experiments, we focus the objective on the cell center. The NA = 1.49 objective then misses molecules near the top and bottom of the cytoplasm.

For the sake of better photon collection efficiency, most of the Kaede tracking experiments were carried out using the NA = 1.49 objective. An example of the corresponding distribution of detected particle locations along y is shown in Fig. 2.A.1B. For this cell, R = 400 nm best fits the rising edges of the distribution. We are clearly missing molecules near y = 0, which is where the molecules near the top and bottom of the cytoplasm would be positioned. We attribute this result to the limited depth of focus of the high-NA objective. The solid line through the data in Fig. 2.A.1B is the distribution along y of molecules lying in a central slab of the spherocylinder defined by $-250\text{ nm} \le z \le +250\text{ nm}$. This assumes a sharp cutoff in the detection efficiency at the slab boundary. That model fits the data quite well. These results are in reasonable agreement with the axial Rayleigh criterion for depth of focus. The "Rayleigh unit" in the z direction is $\pm\lambda/(\text{NA})^2$, where λ is the wavelength of emitted light. Using λ = 582 nm for photoswitched Kaede, this becomes ±262 nm for the NA = 1.49 objective and ±344 nm for the NA = 1.30 objective.

For five cells observed with the NA = 1.30 objective, we compare the measured distribution of particles locations along the transverse coordinate y with a theoretical distribution from a uniformly filled spherocylinder of radius R = 360 nm (Fig. 2.A.1A). The plots include molecules in the entire length of the

cell, including endcaps. Experiment and theory match very well. The goodness of fit at the rising edges is very sensitive to the chosen value of R. We estimate the accuracy of R to be ±30 nm. The best-fit radius is quite similar across cells (380 ± 30 nm, one standard deviation). The length L of the cylindrical segment can be obtained to similar accuracy from the distributions of locations along x. It is important to determine whether or not the limited depth of focus of the NA = 1.49 objective imposes bias in the trajectory measurements and in the analysis of the diffusion coefficient. As shown in Fig. 2.A.1, detection of molecules only from a 500-nm tall central slab does not affect determination of R or L from the spatial distributions along x and y. The reason is that R and L are obtained by fitting the rising edges of the distributions, and these are unperturbed.

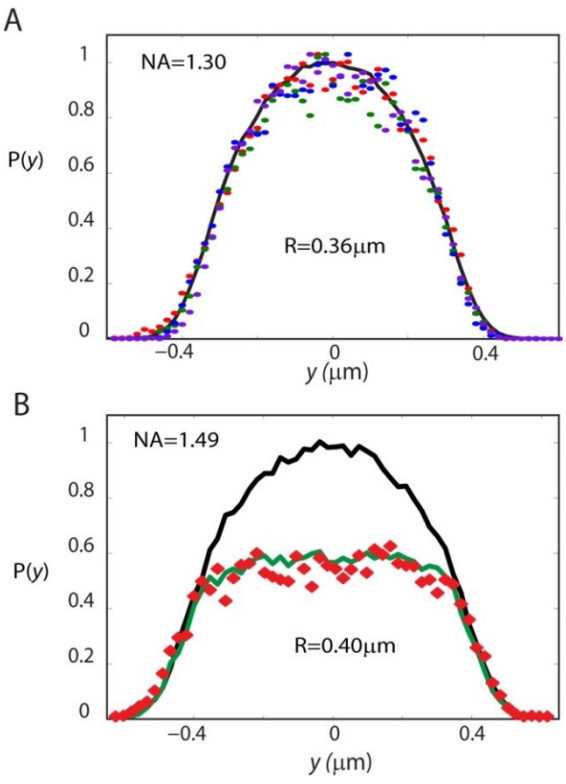

Figure 2.A.1: (A) Distribution of y locations of Kaede molecules from 5 cells (different colors) tracked with the NA = 1.30 objective is compared with simulated distribution of particles in a uniformly filled spherocylinder of R = 360 nm. (B) Distributions of y locations of Kaede molecules tracked with the NA = 1.49 objective are plotted with the simulated distribution from a spherocylinder whose radius is adjusted to match the edges (black line). The green line shows the distribution from a central, 500-nm thick slab in a cell with R = 400 nm and L = 3.5, chosen to match the cell of interest.

For a homogeneous sample (one common diffusion coefficient for all molecules) in infinite space, isotropic diffusion means that the x, y, and z components of diffusive motion are independent. In

this case, detection limited to a central slab parallel to the *xy* plane will not affect measurements along *x* or *y*. As shown in Fig. 2.A.2, restriction to a central slab slightly decreases the degree of confinement observed along *x* and *y*, but to different degrees. For a cell with R = 400 nm and L = 3.0 μm, if we detect molecules only from the central 500 nm slab of the cell then the asymptotic value of MSD_y is 15% higher than the true value. The asymptotic value of MSD_x is also higher, but by only 2%. The reason is that the upper and lower slices whose volumes are not observed are more confining along *x* and *y* than the central slab. The effect on MSD_x is minimal because of the large aspect ratio of the spherocylinder.

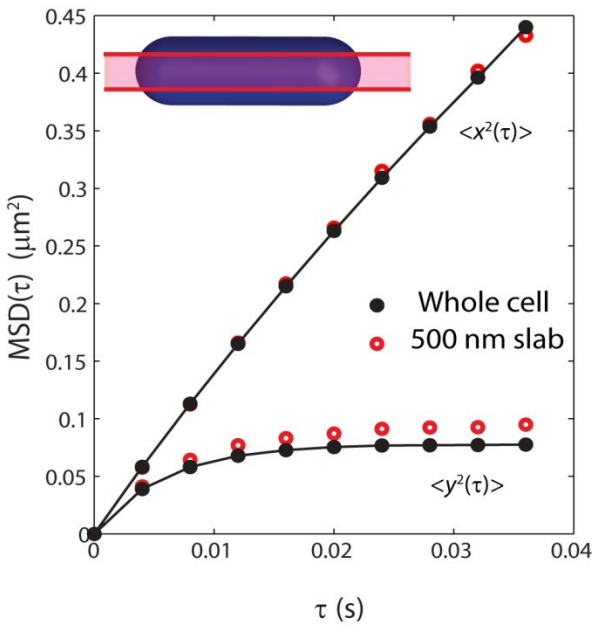

Figure 2.A.2. MSD for *x* and *y* from 10-step trajectories from Monte Carlo simulations with R = 400 nm, L = 3.0 μm, and D = 7 μm^2-s^{-1}. Black points come from the entire data set while open red circles use only those trajectories that remain within the central 500 nm slab (z = ±250 nm) to mimic the NA = 1.49 objective. Inset shows such a slab through a spherocylindrical volume.

For heterogeneous populations (two sets of molecules with different diffusion coefficients), the limited depth of focus of the high-NA objective and the imposition of a lower bound on trajectory lengths to be included in the analysis could introduce sampling bias (38) . Slower molecules tend to stay in the focal volume for longer time, leading to an overestimate of the population with slower diffusion coefficient. As long as diffusion is isotropic (*x*, *y*, and *z* components equal), the values of D are not affected.

Appendix C: Degree of chromosomal segregation with cell length

Using DRAQ5 signal as a marker for the distribution within the cell, we have examined the correlation of DNA distribution with the cell length. The axial intensity profile of DRAQ5-DNA is used to determine the degree of segregation of the two chromosomal lobes. The longer the cell, the greater the degree of DNA segregation away from the mid-plane.

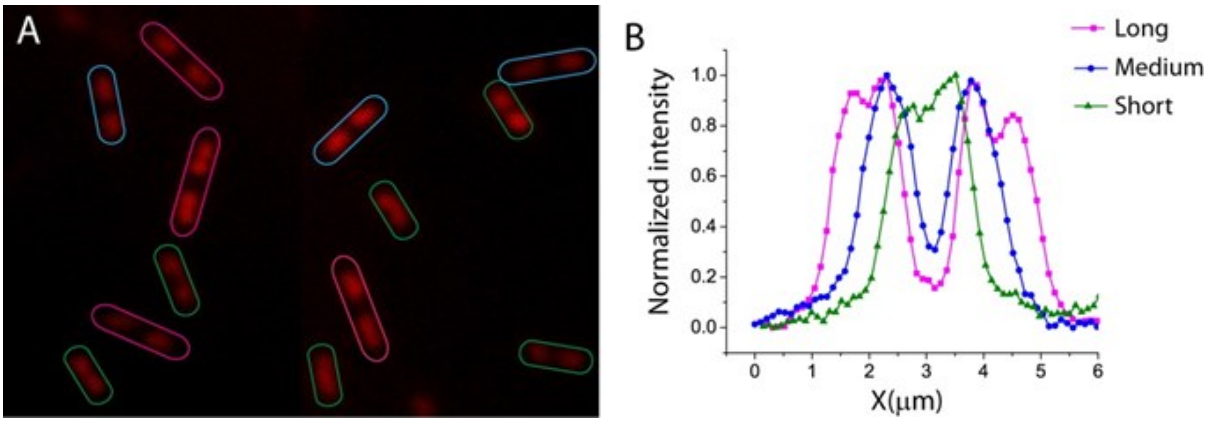

Figure 2.A.3: (A) Images of DNA stained with DRAQ5 for cells of different length. (B) Axial intensity distributions of DRAQ5 stain for three cells.

Appendix D: Sensitivity of msd_r distribution to two populations with different diffusion coefficients

The diffusion map of Fig. 2.8 observes significant spatial heterogeneity in D_{Kaede} for long cells, those in which segregation of the chromosomal DNA is the most complete. Yet the distribution of $MSD_x(\tau = 3$ steps) in Fig. 2.4 is well fit by a homogeneous distribution. As is well known, distinguishing two populations with different diffusion coefficients from the distribution of MSD values at fixed lag time τ is difficult or impossible unless the ratio of the two diffusion coefficients D_1/D_2 is substantial. Here we use the Monte Carlo simulations to illustrate how different the two diffusion coefficients must be to enable clear observation of two distinct components in $MSD(\tau)$.

We have simulated 13-step trajectories in free space for two models: (1) two sub-populations N_1 and N_2 which retain their identity, diffusing with D_1 and D_2 for the duration of the measurements, and (2) a homogeneous population with diffusion constant equal to the population-weighted mean of the two sub-populations. The case of two interconverting states is treated elsewhere (39, 40). Selection of two parameters, D_1/D_2 and $\%D_1 = N_1/(N_1 + N_2)$ determines both models. For both models, we constructed Monte-Carlo distributions of three-step squared displacements $MSD_x(\tau = 3$ steps) analogous to those in Fig. 2.4 using a sample size comparable to experiment. The Kolmogorov-Smirov procedure can then be used to test the statistical distinguishability of the two sample distributions, one from model (1) and the other from model (2). The null hypothesis that the two sample distributions came from the same underlying distribution is tested over a range of parameters D_1/D_2 and $\%D_1 = N_1/(N_1 + N_2)$. The calculations were carried out using the Matlab function "kstest2". The resulting Kolmogorov-Smirnov *P*-values are plotted on a log scale in Fig. 2.A.4. The contour having $P = 0.2$ is shown as a curve through parameter space. For parameters above and to the right of the curve, $P < 0.2$, meaning that two identical underlying distributions would give two such sample distributions less than 20% of the time. Conversely,

the two sample distributions would have been produced from two *different* underlying distributions at least 80% of the time. In other words, heterogeneity of diffusion at this confidence level can be inferred only when D_1/D_2 is larger than about 2 (and both sub-populations have substantial weight). The heterogeneity inferred for longer cells from the spatial diffusion map (Fig. 2.8) indicates $D_1/D_2 \sim 1.3$, explaining why the $MSD_x(\tau = 3 \text{ steps})$ distribution of Fig. 2.4 would not discern heterogeneity.

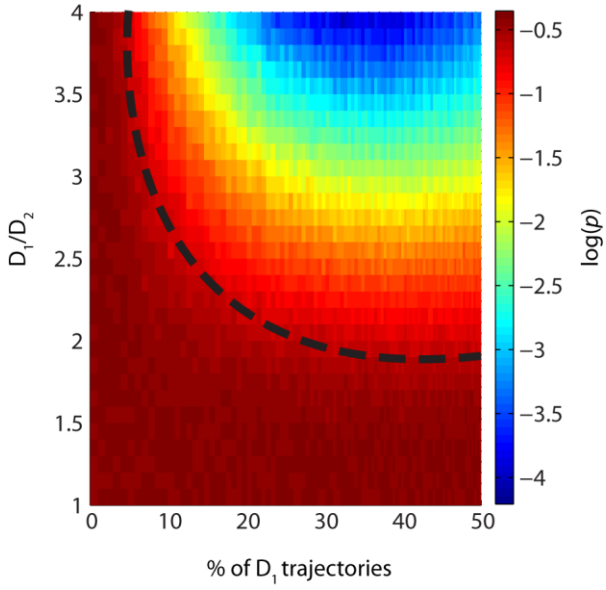

Figure 2.A.4: *P*-value results of Kolmogorov-Smirnov tests of the statistical distinguishability of two sample distributions of $MSD_x(\tau = 3 \text{ steps})$ generated from Monte Carlo simulations. The color scale shows the logararithm of the *P*-values from the test as a function of the two input parameters. The dashed curve shows the contour $P = 0.2$. For parameter sets above and to the right of the curve, there is less than 20% chance that the two samples were produced from the same distribution.

References

1. Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science. 313:1642-1645.

2. Hess, S. T., T. P. K. Girirajan, and M. D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys. J. 91:4258-4272.

3. Rust, M. J., M. Bates, and X. W. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). Nat. Meth. 3:793-795.

4. Manley, S., J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, and J. Lippincott-Schwartz. 2008. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat. Meth. 5:155-157.

5. Niu, L., and J. Yu. 2008. Investigating intracellular dynamics of ftsz cytoskeleton with photoactivation single-molecule tracking. Biophys J. 95:2009-2016.

6. Biteen, J. S., and W. E. Moerner. 2010. Single-molecule and superresolution imaging in live bacteria cells. Cold Spring Harbor Perspectives in Biology. 2.

7. Fu, G., T. HUang, J. Buss, C. Coltharp, Z. Hensel, and J. Xiao. 2010. In vivo structure of the e. *coli* ftsz-ring revealed by photoactivated localization microscopy (palm). PLoS One. 5:e12682.

8. Moerner, W. E., S. L. Lee, S. F., M. A. Thompson, M. A. Schwartz, and L. Shapiro. 2011. Super-resolution imaging of the nucleoid-associated protein HU in caulobacter crescentus. Biophysical Journal. 100:L31-L33.

9. English, B. P., V. Hauryliuk, A. Sanamrad, S. Tankov, N. H. Dekker, and J. Elf. 2011. Single-molecule investigations of the stringent response machinery in living bacterial cells. Proc Natl Acad Sci U S A. 108:E365-373.

10. Hayashi, I., H. Mizuno, K. I. Tong, T. Furuta, F. Tanaka, M. Yoshimura, A. Miyawaki, and M. Ikura. 2007. Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein kaede. Journal of Molecular Biology. 372:918-926.

11. Ando, R., H. Hama, M. Yamamoto-Hino, H. Mizuno, and A. Miyawaki. 2002. An optical marker based on the uv-induced green-to-red photoconversion of a fluorescent protein. Proc. Natl. Acad. Sci. U. S. A. 99:12651-12656.

12. Valkenburg, J. A., and C. L. Woldringh. 1984. Phase separation between nucleoid and cytoplasm in *Escherichia coli* as defined by immersive refractometry. J. Bacteriol. 160:1151-1157.

13. Odijk, T. 1998. Osmotic compaction of supercoiled DNA into a bacterial nucleoid. Biophys. Chem. 73:23-29.

14. Saxton, M. J. 2007. A biological interpretation of transient anomalous subdiffusion. I. Qualitative model. Biophys. J. 92:1178-1191.

15. Ferrari, R., A. J. Manfroi, and W. R. Young. 2001. Strongly and weakly self-similar diffusion. Physica D. 154:111-137.

16. Konopka, M. C., I. A. Shkel, S. Cayley, M. T. Record, and J. C. Weisshaar. 2006. Crowding and confinement effects on protein diffusion in vivo. J. Bacteriol. 188:6115-6123.

17. Neidhard.Fc, P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. J. Bacteriol. 119:736-747.

18. Sochacki, K. A., I. A. Shkel, M. T. Record, and J. C. Weisshaar. 2011. Protein diffusion in the periplasm of e-coli under osmotic stress. Biophys. J. 100:22-31.

19. Dittrich, P. S., S. P. Schafer, and P. Schwille. 2005. Characterization of the photoconversion on reaction of the fluorescent protein kaede on the single-molecule level. Biophys. J. 89:3446-3455.

20. Schafer, S. P., P. S. Dittrich, E. P. Petrov, and P. Schwille. 2006. Single molecule fluorescence imaging of the photoinduced conversion and bleaching behavior of the fluorescent protein kaede. Microsc. Res. Tech. 69:210-219.

21. Crocker, J. C., and D. G. Grier. 1996. Methods of digital video microscopy for colloidal studies. J. Colloid Interface Sci. 179:298-310.

22. Michalet, X. 2010. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Physical Review E. 82:041914.

23. Olveczky, B. P., and A. S. Verkman. 1998. Monte carlo analysis of obstructed diffusion in three dimensions: Application to molecular diffusion in organelles. Biophys. J. 74:2722-2730.

24. Jin, S., and A. S. Verkman. 2007. Single particle tracking of complex diffusion in membranes: Simulation and detection of barrier, raft, and interaction phenomena. J. Phys. Chem. B. 111:3625-3632.

25. Anderson, C. M., G. N. Georgiou, I. E. G. Morrison, G. V. W. Stevenson, and R. J. Cherry. 1992. Tracking of cell-surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera - low-density-lipoprotein and influenza-virus receptor mobility at 4-degrees-c. J. Cell Sci. 101:415-425.

26. Bickel, T. 2007. A note on confined diffusion. Physica A: Statistical Mechanics and its Applications. 377:24-32.

27. Vrljic, M., S. Y. Nishimura, S. Brasselet, W. E. Moerner, and H. M. McConnell. 2002. Translational diffusion of individual class ii mhc membrane proteins in cells. Biophys. J. 83:2681-2692.

28. Ewers, H., A. E. Smith, I. F. Sbalzarini, H. Lilie, P. Koumoutsakos, and A. Helenius. 2005. Single-particle tracking of murine polyoma virus-like particles on live cells and artificial membranes. Proc. Natl. Acad. Sci. U. S. A. 102:15110-15115.

29. Sbalzarini, I. F., and P. Koumoutsakos. 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. J. Struct. Biol. 151:182-195.

30. Elowitz, M. B., M. G. Surette, P. E. Wolf, J. B. Stock, and S. Leibler. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. J. Bacteriol. 181:197-203.

31. Konopka, M. C., K. A. Sochacki, B. P. Bratton, I. A. Shkel, M. T. Record, and J. C. Weisshaar. 2009. Cytoplasmic protein mobility in osmotically stressed *Escherichia coli*. J. Bacteriol. 191:231-237.

32. Nenninger, A., G. Mastroianni, and C. W. Mullineaux. 2010. Size dependence of protein diffusion in the cytoplasm of *Escherichia coli*. J. Bacteriol. 192:4535-4540.

33. Mika, J. T., G. Van Den Bogaart, L. Veenhoff, V. Krasnikov, and B. Poolman. 2010. Molecular sieving properties of the cytoplasm of *Escherichia coli* and consequences of osmotic stress. Mol. Microbiol. 77:200-207.

34. Kumar, M., M. S. Mommer, and V. Sourjik. 2010. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. Biophys. J. 98:552-559.

35. Ando, T., and J. Skolnick. 2010. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proc. Natl. Acad. Sci. U. S. A. 107:18457-18462.

36. McGuffee, S. R., and A. H. Elcock. 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput Biol. 6:e1000694.

37. Thompson, R. E., D. R. Larson, and W. W. Webb. 2002. Precise nanometer localization analysis for individual fluorescent probes. Biophys. J. 82:2775-2783

38. Goulian, M., and S. M. Simon. 2000. Tracking single proteins within cells. Biophys. J. 79:2188-2198.

39. Jung, S., and R. M. Dickson. 2009. Hidden markov analysis of short single molecule intensity trajectories. The journal of physical chemistry. B. 113:13886-13890.

40. Das, R., C. W. Cairo, and D. Coombs. 2009. A hidden markov model for single particle tracks quantifies dynamic interactions between lfa-1 and the actin cytoskeleton. Plos Computational Biology. 5.

Chapter 3

Superresolution Imaging of Ribosomes and RNA Polymerase in Live *Escherichia coli* Cells

Introduction

The cytoplasm of bacterial cells exhibits remarkable spatial organization whose underlying causes are uncertain (1) . In *E. coli*, the chromosomal DNA is "condensed" into an irregularly shaped, central region called the nucleoid (2). The nucleoid does not fill the entire cytoplasmic volume, in spite of severe compaction of ~1.5 mm of DNA contour length within a 3 μm long x 1 μm diameter cell. The chromosomal origin *oriC* is anchored near the cell center until replication and segregation, when the two copies migrate to the quarter-cell positions (3). Certain proteins and lipids are known to concentrate in the polar or septal regions during specific phases of the cell cycle (4). There is recent evidence that specific mRNA transcripts carry information that targets them to specific cellular locations for translation (5).

In good growth conditions, multiple ribosomes are translating mRNA strands while they are being synthesized by RNA polymerase ("coupled transcription and translation", or "co-transcriptional translation") (6). Poly-ribosomes ("polysomes") are necklaces of ribosomes serially translating the same mRNA copy (7). Polysomes tethered to DNA via RNAP linkages were observed in early EM studies of *E. coli* extracts (8). Polysomes of compact, quasi-helical structure have recently been directly observed in cryoelectron tomography images (7).

In some species, transcription and translation evidently occur predominantly in different sub-regions of the cytoplasm. In *B. subtilis,* the ribosomes are concentrated outside the nucleoid in ribosome-rich regions of the cytoplasm comprising the two polar end-caps, the space between nucleoid lobes, and a thin shell proximal to the cytoplasmic membrane (9). Because some 80% of ribosomes are believed to be actively translating proteins in all growth conditions, the ribosome-rich regions evidently form a localized "protein factory" occupying some 50% of the cytoplasmic volume (10). Similarly, in *E. coli* early EM studies of thin slices of fixed cells found ribosomes concentrated near the periphery of the cell with the nucleoid avoiding the near vicinity of the cytoplasmic membrane (2). More recently, widefield immuno-fluorescence microscopy found the L7/L12 subunits of ribosomes to be strongly localized outside the nucleoid (11). In essential agreement with these data, a simple physical model of plectonemic DNA and hard-sphere ribosomes found that at sufficiently high DNA density, configurational entropy and excluded volume effects alone can cause strong nucleoid-ribosome segregation (12). However, some species appear *not* to exhibit strong nucleoid-ribosome segregation. A recent study of *Caulobacter crescentus* found ribosomes and DNA to be spread throughout the cytoplasm (13).

At least in *B. subtilis* and *E. coli*, the spatial separation of most of the chromosomal DNA from most of the ribosomes raises the question of the overall percentage of translation that could be coupled to transcription (9, 14). Early EM evidence from fixed *E. coli* cells indicates that RNA polymerase (RNAP) localizes near the boundary between the nucleoid and the cytoplasmic periphery (15), but this has not been demonstrated in live cells. Comparison of the mean lifetime of mRNA (~5 min) with the mean time to transcribe a message (~20 s) suggests that most translation occurs on complete mRNA copies that have released from RNAP and diffused to the ribosome-rich regions (13, 16). Free messages coated with ribosomes and perhaps cold-shock proteins as well (17) might avoid degradation by ribonucleases long enough to reach the ribosome-rich regions, where efficient polysomic translation could continue to protect them from damage.

The diffusion of free mRNA in live bacterial cells is not well characterized. In *C. crescentus*, a recent fluorescence *in situ* hybridization study found that mRNA strands did not diffuse significantly from the location where they were synthesized (13). In sharp contrast, a recent study in *E. coli* found that free messages evidently diffuse to the specific regions of the cytoplasm where their protein products will be used, suggesting some type of location code within the message (5). A very long, 6000 kDa mRNA-MS2-GFP complex exhibited seemingly tethered diffusion with a short-time diffusion coefficient of 0.03 μm^2-s^{-1} (18).

The new superresolution fluorescence microscopy methods (PALM, (19)
F-PALM, (20) and STORM (21)) are poised to provide a new level of quantitative information in live bacterial cells (22-25). This includes protein copy numbers, spatial distributions with ~30-nm accuracy, and diffusion coefficients on sub-second time scales for a variety of key cytoplasmic components. Here we present a quantitative study of ribosomes in live *E. coli* labeled by the construct S2-YFP. Under moderate growth conditions, most cells exhibit two partially segregated nucleoid lobes and three "ribosome-rich regions", the two endcaps and the region between nucleoid lobes. We find strong nucleoid-ribosome segregation. On average, only ~10-15% of the ribosomes are found within the dense regions of the nucleoid. The single-particle tracking measurements are consistent with free diffusion of 70S monomers or polysomes under effective confinement in the ribosome-rich regions, with $D_{ribo} \sim 0.04$ μm^2-s^{-1}. This value presumably describes the effective diffusion coefficient of free mRNA decorated with translating ribosomes. It follows that free mRNA born in the dense nucleoid region will find a ribosome-rich region in ~1 s or less. This is much shorter than typical mRNA degradation times of ~3-8 min (16). Finally, the diffusion data do not rule out coupled transcription and translation, but neither do we find any clear evidence of sub-diffusion of ribosomes, which would arise from tethering of ribosomes to DNA.

In a different strain of cells, superrresolution images of RNA polymerase are obtained from a β'-yGFP construct in live *E. coli* under the same growth conditions. yGFP is a YFP-like construct engineered

from GFP (3). The spatial distribution of RNAP mimics that of DNA, as stained by the red dye DRAQ5. In moderate growth conditions, we do not find evidence for a high-density shell of RNAP surrounding the nucleoids. There are very few RNAP copies in the ribosome-rich regions.

Taken together, our results strongly suggest that at least in *E. coli*, most translation is *not* coupled with transcription (the "co-transcriptional translation" mechanism), counter to a view common in the literature. Instead, the data suggest that completed messages diffuse to find the ribosome-rich regions where the bulk of translation occurs. This is consistent with the fact that lifetime estimates of mRNA in live bacteria exceed typical times required for transcription of a message by a factor of 15 or so (16).

At the same time, we do observe a significant number of ribosome and RNAP copies lying within ~60 nm of the cytoplasmic membrane in close proximity to the dense nucleoid lobes. These copies are candidates for participation in "transertion", which is the co-transcriptional translation of membrane proteins (6, 26). Transertion has been proposed as an "expanding" force on the nucleoids (6). Accordingly, we observe *radial* compaction of the nucleoids after treatment with rifampicin or chloramphenicol. However, the data do not support transertion of membrane proteins at the endcaps; there are essentially no RNAP copies close to the endcap sections of the cytoplasmic membrane. Thus transertion seemingly cannot provide the direct *axial* expanding force on the nucleoid that has been inferred from axial contraction of nucleoid lobes after chloramphenicol treatment.

Throughout this paper the x coordinate runs parallel to the long axis of the cell and the y coordinate runs perpendicular to x and along the short axis (Fig. 3.1B-inset).

Materials and Methods

Bacterial strains and plasmids

The construct used for imaging ribosomes, AFS55, which contains a translational fusion of *yfp* to the C-terminus of *rpsB* (the gene encoding S2), was constructed by lambda-Red mediated recombination. The YFP variant here is GFPmut3.1, having the substitutions V68L Q69K Q80R T203Y. A DNA segment containing *yfp* and a kanamycin resistance gene was amplified by PCR from the plasmid pEB45 (27) with the primers 5'–
CCCAGGCGGAAGAAAGCTTCGTAGAAGCTGAGVNNVNNVNNVNNVNNVNNCGTAAAGGAG
AAGAAC-3' and 5'-
ACTCGAACTATTTTGGGGGAGTTATCAAGCGTGTAGGCTGGAGCTGCTTC-3'.The underlined portions in the first and second primers anneal to the C-terminus of *rpsB* just before the stop codon, and the region beginning five bases after the stop codon, respectively. The first primer encodes an 18 base randomized linker connecting S2 and YFP. The PCR product was electroporated into MG1655/pKD46 as described earlier (28). A kanamycin resistant colony was selected and verified to express YFP. The *rpsB-yfp* fusion was then moved to a clean MG1655 background by P1 transduction, resulting in AFS55. The linker between *rpsB* and *yfp* was sequenced and determined to be: 5'-
GAGCAGGAAAGGCGACAGGAGCGT-3'. The underlined portion is the linker and the first and last codons are the last codon of *rpsB* and the first codon of *yfp*, respectively.

For the RNAP studies, the rpoc::ygfp strain (RLG7470) was constructed in the same way as the rpoc::gfp strain as described elsewhere (29). The construct was transferred to a VH1000 background using P1 transduction. The protein yGFP is a variant of GFP developed by Nielsen et al (3); it has similar absorption and emission spectra to those of YFP, in spite of its name. The labeled entity is then called β'-yGFP. For the construct expressing bare YFP, we used colony PCR to extract the YFP gene from the S2-

YFP construct. This fragment was ligated to a similarly digested fragment of pASK-IBA3plus (Invitrogen, Carlsbad, CA). DH5α cells were transformed with the resulting plasmid (pJW2), and transformants were selected on ampicillin-containing plates. MG1655 cells were subsequently transformed with the same plasmid. For measurement of the radius of MG1655 and VH1000 cells, we transformed both cell lines with a pASK-IBA3plus a plasmid expressing Kaede.

Cell growth and preparation

In bulk EZ rich defined medium (EZRDM, (30)) at 30ºC with 50 μg/mL Kanamycin, the S2-YFP-expressing strain AFS55 grow in bulk medium with doubling times of 54 min. For wild-type MG1655 under the same growth conditions, the doubling time in bulk medium is 51 min. The β'-yGFP-expressing strain RLG7470 has a doubling time of 43 min. This is comparable to the 45 min doubling time of the background strain VH1000.

Cells were grown overnight with shaking in 30ºC water bath. We subsequently made subcultures by diluting the stationary phase culture at least 1:250 into 3 mL of fresh EZRDM. When cells had grown to midlog phase (OD_{600} = 0.4-0.6), 7 μL of culture was placed on poly-L-lysine-coated coverslips to immobilize cells for observation. The assembly was sealed with nail polish. Imaging by widefield epifluorescence or by single-molecule photoactivation was carried out at 30ºC for no longer than 40 min after plating. During that time, cells continue to grow. As a control, we have also imaged the ribosome distribution in cells plated on agar pads. The distribution for cells in agar is indistinguishable from that for cells plated on polylysine. For imaging the time course of cell growth and the action of antibiotics, cells were plated in a temperature-controlled sample chamber (RC-20H; Warner Instruments, Hamden, CT). The whole assembly was brought into contact with the microscope objective and warmed to 30°C.

Doubling times for cell growth in sealed coverslip-slide chambers and in the flow chamber were estimated by measuring cell length vs time using phase contrast microscopy (Fig. 3.A.6). The constructs AFS55 and RLG7470 grow with a doubling time of 75-80 min for coverslip-slide chambers and of 60-65

min in the flow chamber with aerated EZRDM growth medium flowing. A few experiments were carried out in the sealed chamber for cells grown in a chemically defined MOPS-buffered minimal medium (0.28 Osm MBM, (30, 31)) and in the undefined rich medium (LB).

To examine the distribution of the label by itself, the bare protein YFP was imaged using the strain JCW112, which contains the YFP expressing plasmid in the same MG1655 strain used for the ribosome studies. Cells were grown overnight with shaking at 30ºC in EZRDM with 100 μg/mL ampicillin. We subsequently made subcultures of these cells by diluting the stationary phase in fresh EZRDM. When cells had grown to midlog phase, anhydrotetracycline was added to a final concentration of 45 nM to induce YFP expression. After 6 min of induction the cells were centrifuged and resuspended in fresh EZRDM twice to remove the inducer. The cells were then incubated again in growth medium for at least 30 min at 30^{o}C to enable maturation of the fluorescent protein and then plated on poly-L-lysine coated coverslips for imaging.

Rifampicin and chloramphenicol stock solutions were prepared by dissolving 1 mg and 10 mg of the drugs in ethanol, respectively. Stock solutions of the drugs were added to the midlog cell culture to attain a final concentration of 200 μg/mL. For imaging of ribosome and DNA static spatial distributions, the cultures were shaken for 30 min before plating and imaging. For widefield imaging of the time course of drug action, we injected drug solution in EZRDM into the flow chamber already containing plated cells in growth medium. Superresolution imaging of ribosomes in the presence of drugs was carried out in sealed coverslip-slide chambers.

DNA staining with DRAQ5

To image the DNA distribution of the cells, we used the red fluorescent DNA stain DRAQ5 (Biostatus Limited). The MIC of DRAQ5 was determined to be 5 μM for MG1655 cells growing at 30ºC in EZRDM. For lower concentrations growth was maintained. A DRAQ5 concentration of 200 nM provided adequate fluorescence intensity. For these studies 2 μL of 100 μM stock solution of DRAQ5 was

added to a 1 mL culture of cells in midlog phase and the culture was shaken for 10 minutes, after which 7 μL of culture was plated on a cover slip for imaging.

Widefield epifluorescence imaging

The fluorophore in the S2-YFP constructs both absorb green light (λ_{max} = 514 nm in absorption) and fluoresce green-yellow (λ_{max} = 533 nm in emission). The β'-yGFP construct has similar spectra. A 514 nm Ar^+ laser (Melles Griot, Carlsbad, CA) was used for imaging. The yellow emission was collected using a 560/50 emission filter (Chroma). The DNA stain DRAQ5 absorbs red light (λ_{max} = 647 nm in absorption). When intercalated into double-strand DNA it fluoresces with emission maximum at 691 nm. A 633 nm laser (HeNe, Coherent, Santa Clara, CA) was used to image the DRAQ5-DNA distribution inside cells. Emission from the DRAQ5 excitation was collected using a 700/75 emission filter (Chroma). Widefield fluorescence images were acquired with a Nikon Eclipse Ti inverted microscope with oil immersion objectives (either Nikon Plan Fluor 100X, NA 1.30 or Nikon APO TIRF 100X, NA 1.49) and a 1.5X tube lens. For time-lapse imaging, fast shutters (Uniblitz LS2, Vincent Associates, New York) were used to synchronize illumination and image acquisition. Images were recorded by a back-illuminated EMCCD camera with 16 μm x 16 μm pixels (iXon DV-887, Andor Technology, Connecticut). Each pixel corresponds to 105×105 nm^2 at the sample with 150X overall magnification.

Superresolution imaging of ribosome-YFP and RNAP-yGFP

The initial concentration of the active YFP/yGFP fluorophores inside the cell is far too high for single molecule imaging. Moerner and coworkers (32) showed that YFP can be used for superresolution imaging by first photobleaching the entire sample, after which some 10-20% of the non-fluorescent molecules gradually return to a fluorescent state. The photobleaching pulse of 514 nm light (4 kW/cm^2 at the objective for 10-20 s) was applied once to each cell. Immediately after the pulse, the rate of return of YFP molecules to the fluorescent state is too high for single-molecule imaging, especially in the case of S2-YFP with its large copy number. Eventually, the cells exhibit only a few (0-3) fluorescent puncta.

Even long after the photobleaching pulse, new fluorescent copies continually appear during the observation time. This may be caused in part by gradual maturation of YFP-containing species as the cells grow and divide.

In the subsequent localization and tracking movies of S2-YFP, different laser intensity was used for measuring ribosome spatial distributions vs ribosome diffusion properties. In both cases, we used 30 ms camera frames (33 Hz frame rate) with 15 ms exposure time within each frame. To obtain superresolution images of the time-averaged ribosome spatial distribution, we used a stronger beam of 514 nm light (2 kW/cm^2 at the objective) to provide bright single-molecule images with fast bleaching. Under these conditions, we typically localized some 1200-3000 molecules per cell over some 7000-10,000 camera frames. We collect one image per 30 ms over an interval of 3-4 min, after which time the rate of appearance of new molecules has become very low. The molecules fluoresce for only 1-3 frames at this high intensity. Ribosomes move rather slowly. With 15 ms exposure times, typical single molecules yield fluorescence spread over roughly a 3 px x 3 px region, essentially a diffraction-limited spot.

Single-molecule tracking of ribosomes was carried out using a lower intensity of the 514 nm laser (0.5-1 kW/cm^2 at the objective). In these conditions, we collect some 100-200 photons per molecule per frame and the average trajectory length increases to 5-6 frames. Trajectories were collected from each cell over a period of 10-15 min (20,000-30,000 frames). In that time we typically observe 500-1500 single molecules whose trajectories comprise some 5000-7000 localizations. Typically particles in 1-4 cells were tracked simultaneously to enhance the efficiency of data collection. For the tracking analysis, we have separated single cells into set of movies for each cell. For mean-square displacement (MSD) analyses of ribosome diffusion, we chose to include only longer trajectories to improve accuracy over longer lag times. In practice we varied the cutoff time scale for each cell. The lower limit on trajectory length was chosen in the range 8-13 steps so that at least 40–60 trajectories are included for each cell. Trajectories longer than the cutoff for each cell were truncated at the cutoff value.

In principle it would be possible to space the camera frames more widely in time and observe the manner in which the MSD plots approach an asymptote at long times. In practice, we could not arbitrarily lengthen the time between frames due to the nature of the YFP imaging. The photobleaching laser pulse converts all molecules to a dark state. After conversion, a small fraction of the dark molecules become fluorescent again. However, we cannot control the rate at which they return. If frame spacing is chosen too long, too many molecules are fluorescent in each frame. Moreover, there are limits on the total laser dosage the cell can endure while maintaining an unperturbed DNA and ribosome spatial distribution; see below.

The yGFP label is used for superresolution imaging by reversible photobleaching here for the first time. The number of β'-yGFP copies per labeled-RNAP cell is a factor of 10 smaller than the number of S2-YFP copies per labeled-ribosome cell. For the RNAP-yGFP studies, we used 1-3 kW/cm^2 of 514 nm light at the objective both for bleaching (0.5-1.0 min) and for imaging (60-ms frames with 30 ms exposure time within each frame). We rejected frames with too many single molecule images (more than 4 particles per cell). We typically detect some ~400 photons per molecule per frame, and the mean trajectory length is 3-4 frames.

As a cautionary note, we have directly observed laser-induced changes in DNA morphology. In both wild-type and S2-YFP expressing cells, strong laser illumination at 514 nm or 561 nm for minutes expands the nucleoid spatial distribution, diminishes ribosome-DNA segregation, increases the ribosome diffusion coefficient (much like Rif treatment), and slows growth (Fig. 3.A.4). Laser-driven nucleoid expansion occurs even in wild-type (unlabeled) cells. Such effects were readily observed following continuous illumination of cells at 6-7 kW/cm^2 for 10-15 min. The mechanism is unclear; perhaps local heating or photodamage affects important cellular constituents which determine DNA morphology. Such high laser dosages were avoided in the study of single molecule imaging of both ribosome and RNAP molecules.

Single-molecule image analysis

Images were analyzed using a MATLAB GUI developed in our lab (33). Images were smoothed and filtered to obtain a zero-based image. Bright spots were located with pixel level accuracy by a peakfinding algorithm which finds local maxima in an image. A user defined intensity threshold was used as the minimum brightness of a pixel arising from a single molecule. Centroids of the bright spots were calculated from an 8 x 8 pixel square centered on the local maxima determined by the peak finding algorithm.

We prefer localization by centroid rather than fitting to model point-spread functions because the images are asymmetrically blurred by diffusion during each camera frame. Centroid is much faster than Gaussian fitting and is easily implemented for analysis of Monte Carlo results as well. Importantly, the smoothing and filtering places single-molecule images on a near-zero background, which is essential for accurate tracking by centroid. The results in Fig. 3.A.1 show that the centroid gives dynamic localization precision $\sigma_{x,y}$ varying from 15-30 nm in each dimension as the peak signal to noise ratio (PSNR) varies from 7 to 14, the relevant range in our experiments. This is only 10-15% larger localization standard deviation than that of Gaussian fitting under the same conditions.

For single-molecule tracking we store the (x,y) positions of the centroid. Using a modified MATLAB version of the tracking program written by Crocker and Grier (34) we connect the centroids of a bright feature from consecutive frames into a trajectory. Analysis of trajectories was carried as described before (33). To determine the distribution of molecule positions along the long and short axis of the cell, we have used a mask generated from the white light image (A2). Analyzed positions were rotated to project along the two axes using the orientation of the mask (Fig. 3.A.2).

Monte Carlo simulations

Monte Carlo simulations of free diffusion were performed in truncated spherocylindrical volumes that mimic the shape of the space occupied by ribosomes within the *E. coli* cytoplasm. Because there are

few ribosomes within the nucleoid, the dense region of nucleoid is represented by two impenetrable cylinders placed within the larger spherocylinder representing the entire cytoplasm (Fig. 3.5). The distribution of initial positions is chosen randomly within the available space. This model problem is characterized by four parameters, the length and radius of the outer spherocylinder (L_1 and R_1 = 380 nm, with L_1 the length of the straight, cylindrical part) and the length and radius of the inner spherocylinders (L_2 = 800 nm and R_2 = 380 nm). The end caps of each spherocylinder have the same radius as the cylindrical part. These parameters were varied to mimic the experimental ribosome spatial distribution. Random walk simulations were carried out in MATLAB as described before (33). The diffusion coefficient was varied until experimental and simulated MSD plots were similar.

Results

Widefield co-imaging of ribosomes (S2-YFP labeling) and nucleoid (DRAQ5 staining)

For live cells growing in the complete, defined medium EZRDM at 30ºC, Fig. 3.1A shows examples of widefield fluorescence images of ribosomes (labeled by the S2-YFP protein and observed in the yellow channel) and the chromosomal DNA (labeled by the permeable dye DRAQ5 and observed in the red channel). Ribosomes localize to the end caps and the central region between two segregated DNA lobes, the "ribosome-rich regions". The DRAQ5-stained images show the chromosomal DNA segregated to varying degree into two nucleoid lobes. Axial intensity distributions (intensity summed along *y* at each axial position *x*) show that the degree of DNA segregation consistently increases from short to medium to long cells (Fig. 3.1C and 1D). As cells approach the division stage of the cycle, each of the two primary lobes begins to split into two sub-lobes (Fig. 3.1C). The peak-to-valley intensity ratio in the S2-YFP axial intensity profiles is typically about 1.5:1.0, with ribosome peaks correlating with DNA valleys (Fig. 3.3.1B). However, the degree of ribosome segregation is underestimated from these axial fluorescence

profiles obtained with the 1.49 NA objective. The S2-YFP images in Fig. 3.1A are in principle a superposition of broadly distributed cell autofluorescence; rapidly-diffusing, free S2-YFP or perhaps YFP copies; and a mixture of more slowly diffusing 30S subunits, 70S mono-ribosomes, and polysomes. Based on the diffusion evidence presented below, we believe the contribution of free S2-YFP and YFP is negligible. The background autofluorescence contributes only 2% of the yellow intensity, as judged by tests of wild-type cells grown and imaged under the same conditions. The widefield ribosome images are somewhat broadened by diffraction. However, the primary broadening is due to out-of-focus light collected by the 1.49 NA microscope objective, whose depth-of-focus parameter is only 500 nm (±250 nm about the focal point) (35). Widefield images of the ribosome distribution taken with a low numerical aperture objective (NA = 1.25, depth-of-focus parameter 660 nm) show stronger axial segregation (Fig. 3.A.6; peak-to-valley intensity ratios of 2:1 or even 3:1). The red DRAQ5 fluorescence images taken with the 1.49 NA objective are narrower than the S2-YFP images because the nucleoids are concentrated near the cell axis; there is much less out-of-focus light. Importantly, there is no clear evidence that the DNA distribution extends all the way to the tips of the cells, as seen by comparing the widefield DNA vs ribosome axial distributions (Fig. 3.1B).

The peaks in the axial profiles of ribosome-YFP intensity distributions are consistently anti-correlated to the peaks in the DNA-DRAQ5 intensity profile (Fig. 3.1B and 1D). The three typical cells in Fig. 3.1B were selected based on length: a short cell (2.8 μm tip-to-tip), a medium-length cell (3.2 μm), and a long cell (4.4 μm). For most of the short cells (2.6-3.0 μm tip-to-tip) the DNA localizes to the middle of the cell and the peak in DNA intensity appears near the center of the cell. In these short cells, ribosomes concentrate only at the end caps. For mid-length cells (3.0-4.5 μm tip-to-tip) DNA segregates into two lobes and the ribosome distribution has a third peak at mid-cell, the region between the segregated DNA lobes. Just before the cells divide, they are very long (4.5-5.5 μm tip-to-tip) and each DNA lobe begins to split into two sub-lobes (Fig. 3.1B at right). The ribosome distribution now has 5

peaks; even the minor DNA dips correspond to ribosome peaks. These same plots show the ribosome distribution filling both endcaps, whereas there is little or no DNA near the tips of the cell.

In Fig. 3.1D we plot the axial location relative to cell center of the peaks in the DNA and ribosome axial distributions from 286 cells of varying length. The locations of the peaks from each cell are displaced vertically according to total cell length, a rough measure of the phase of the cell cycle. As cells become longer, both ribosome and DNA peaks move symmetrically outward and remain segregated. Evidently the ribosome distribution responds faithfully to the DNA distribution, always avoiding the densest regions of DNA.

The radial distribution of S2-YFP intensity from the widefield images is significantly wider in the DNA-rich regions than in the ribosome-rich regions (data not shown). This is consistent with the presence of a shell of higher ribosome concentration near the cytoplasmic membrane and lower ribosome concentration near the axis of the cell where the DNA is densest (12). This feature of the radial distribution of ribosomes becomes more obvious in the superresolution images presented below.

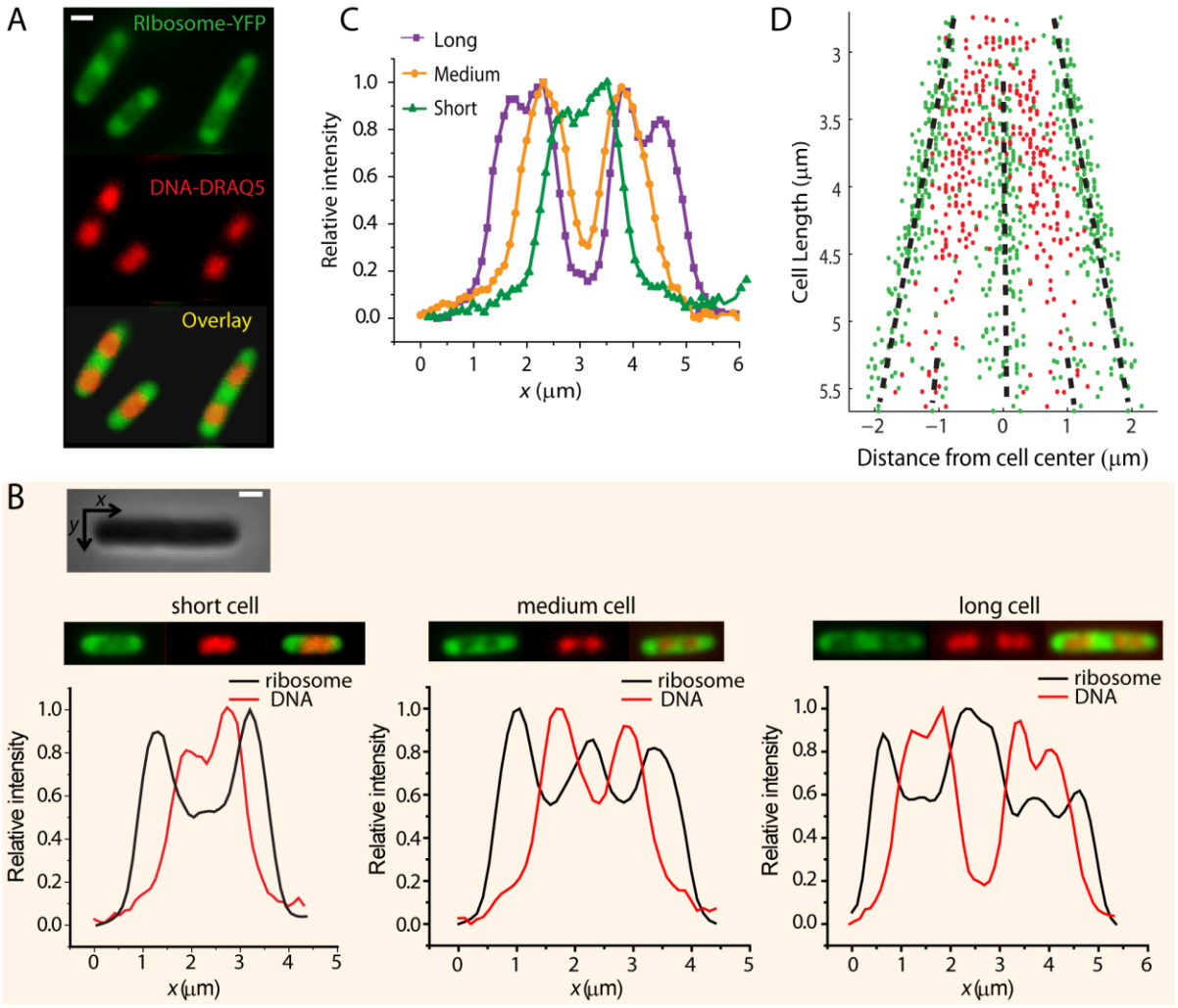

Figure 3.1. Widefield imaging of ribosomes (S2-YFP labels) and chromosomal DNA (DRAQ5 stain) for K-12 cells grown in EZRDM at 30ºC. (A) Ribosome distribution (green) and DNA distribution (red) for three typical cells. The composite image shows the anti-correlation between the two distributions. (B) Axial intensity distributions in the ribosome and DNA channels for a short, medium, and long cell. Intensity is summed along *y* (short axis coordinate) at each *x* (long axis coordinate). The two channels are strongly anti-correlated. (*Inset:* Phase contrast image of a cell showing *x* (along the long axis) and *y* (along short axis) coordinates.) (C) DNA axial intensity distribution for a short, medium, and long cell plotted on same axes to show progressive segregation as the cell elongates. (D) The distance from the cell center of the local maxima in ribosome distribution (green dots) and DNA distribution (red dots) plotted for 286 cells of different cell length. The black dashed lines guide the eye. Scale bar = 1 μm.

Widefield co-imaging of RNAP (β'-yGFP) and nucleoids (DRAQ5 staining)

Figure 2A shows examples of widefield fluorescence images of RNAP (labeled by the β'-yGFP protein and observed in the yellow channel) and DRAQ5-labeled chromosomal DNA (observed in the red channel). As described before (29, 36), within each cell the RNAP distribution and the chromosomal DNA distribution are very similar (Fig. 3.2B). Like the DNA, the RNAP distribution typically exhibits two lobes; in the longest cells, each lobe has begun to sub-divide. The three examples in Figs. 2C show a typical short cell (2.6 μm tip-to-tip), mid-length cell (3.9 μm) and long cell (5.1 μm). The RNAP distribution follows the progressive segregation of DNA as the cell elongates. As noted elsewhere, the RNAP distribution also exhibits narrow, bright features that are absent in the DNA images (29, 36). These have been attributed to transcription foci, locations of particularly dense transcription activity (9, 37).

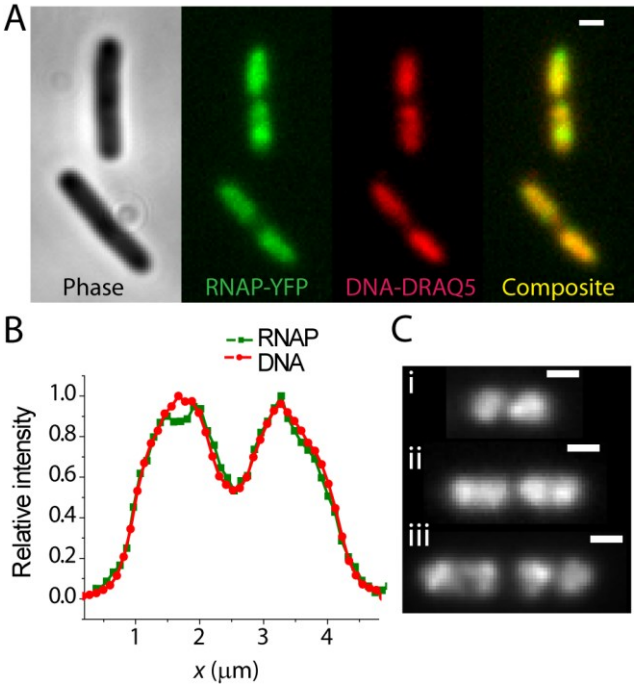

Figure 3.2. Widefield imaging of RNA polymerase (β'-yGFP labels) and chromosomal DNA (DRAQ5 stain) for K-12 cells grown in EZRDM at 30ºC. (A) Phase contrast image, RNAP image, and DNA image are shown for two different cells. The composite image shows RNAP co-localizes with DNA. Scale bar is 1 μm. (B) Axial intensity profile for RNAP-yGFP and DRAQ5-labeled DNA compared from a single cell. (C) RNAP images from three cells of length (i) 2.6 μm, (ii) 3.9 μm, and (iii) 5.1 μm. Scale bar is 1 μm. Progressive segregation of RNAP distribution is similar to DNA segregation shown in Fig. 3.1.

Superresolution ribosome spatial distribution

The superresolution images obtained here use the reversible photobleaching method (Biteen *et al.*, 2008). Initially all the YFP copies in a cell are photobleached using 514 nm light. Subsequently, some 10-20% of the photobleached population spontaneously and gradually returns to the fluorescent state. The low density of these "revived" YFP copies enables single-molecule imaging. By plotting a point at the (x,y) coordinates of each location, we obtain a high-resolution spatial distribution averaged over the several-minute acquisition period. The dynamic localization accuracy is estimated to be $\sigma \sim 30$ nm in each dimension x and y (Fig. 3.A.1). This superresolution method is impervious to background autofluorescence because almost no such fluorescence remains after the photobleaching step. In addition, use of the 1.49 NA objective plus the single-particle thresholding algorithm leads to significant sectioning of the cell in the x-y image. We are most sensitive to molecules in a horizontal, central slab of approximately 500-600 nm thickness (z in the range ±250-300 nm), which is substantially smaller than the 780 nm cell diameter (33). The top and bottom layers of the cytoplasm are not imaged). Unlike the widefield images, the superresolution images obtained with the 1.49 NA objective are not broadened by out-of-focus light from the top and bottom of the cell; those molecules are rejected by the threshold criterion.

Ribosome superresolution images for several cells are shown in Fig. 3.3A. At the 33 Hz frame rate, we were able to study some 1,000-3,000 distinct S2-YFP copies in the typical, 3.5 μm long cell over a period of 2-3 min. The mean trajectory length under these imaging conditions is only 3 frames, so we typically obtain about 5,000-10,000 locations over the same period. The significant improvement in resolution and increased contrast between the DNA-rich region and the ribosome-rich region are immediately obvious. In Fig. 3.3C we display the superresolution image for a mid-length cell (3.6 μm tip-to-tip), which evidently has fairly strong nucleoid lobe separation. The histogram of Fig. 3.3C gives $N_{ribo}(x)$, the number of ribosomes detected within each axial bin at position x, with bin widths chosen at

100 nm to provide good signal-to-noise ratio. There are very few ribosomes within the nucleoid lobes. The peak-to-valley concentration ratio is at least 5:1 between the ribosome-rich regions and the densest part of the nucleoids. This may well be an underestimate due to the presence of a thin shell of ribosomes around each nucleoid lobe (Fig 3.5B) and the irregular shape of the lobes themselves. This peak-to-valley ratio varies from cell to cell as illustrated in Fig. 3.3D.

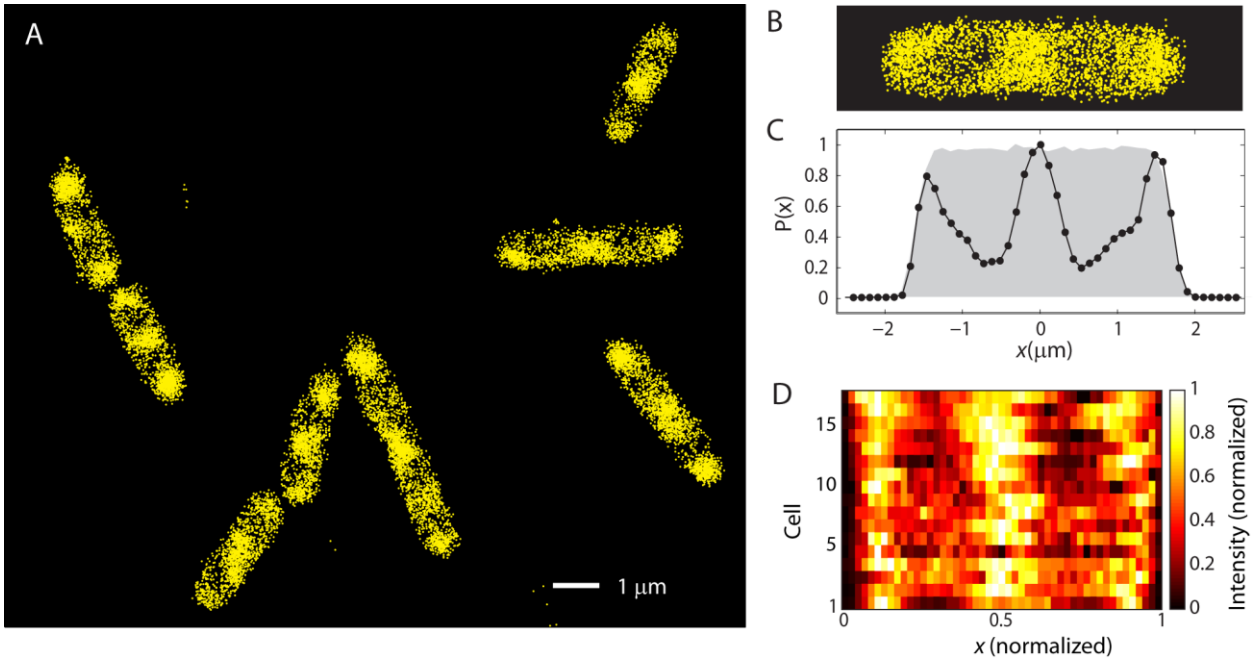

Figure 3.3. Superresolution images of ribosomes (S2-YFP) within K-12 cells grown in EZRDM at 30ºC. Each localization is plotted as a point at the calculated centroid position. (A) Nine representative cells. (B) Expanded view of superresolution image of ribosomes in a single cell. (C) Relative number of ribosomes at each axial position, with data along *y* at each *x* summed into 100 nm bins. The grey background shows the theoretical profile for a uniform distribution filling the model spherocylinder, taking account of measurement uncertainty and binning. (D) Distribution of ribosomes along the long axis of the cell for 17 cells is plotted as relative intensity vs relative position and color coded.

Evidence for the thin shell of ribosomes surrounding the nucleoid lobes is shown in Fig. 3.5B. Due to the sectioning by the 1.49 NA objective, the *y*-coordinate is a reasonable proxy for a radial coordinate. The *y*-axis profiles of the ribosome distribution limited to the regions containing the two nucleoid lobes for 12 cells are combined to make a smoother histogram. We often observe the axis of a nucleoid lobe to be off-center relative to the overall cell axis, meaning one edge of the thin ribosome shell is wider than the other. This causes broadening of features in the composite histogram. Nevertheless, the ribosome radial distribution clearly peaks away from the cell axis and extends all the way to the cytoplasmic membrane. As a control, we used cells expressing monomeric YFP from a plasmid to measure the *y*-distribution of free YFP molecules. The result is very similar to the theoretical distribution assuming YFP fills the cytoplasm uniformly and taking account of the sectioning by the 1.49 NA objective (Fig. 3.A.6).

Superresolution RNA polymerase spatial distribution

In the RNAP imaging experiments we obtain about 1000-3000 localizations of β'-yGFP from about 500 individual RNAP molecules. Fig. 3.4A shows images of four cells. The yellow dots are the centroid of the fluorescence image of single RNAP molecules in each camera frame. In Figs. 3.4B and 3.4C we show expanded superresolution images of the RNAP distribution in a medium-length cell (3.9 μm) and a long cell (5.2 μm tip-to-tip). The white outline is the mask created by edge detection from the white light image of the cell. For the longer cell, the RNAP axial distribution (Fig. 3.4D) closely mimics the four-peaked DNA distribution common among the longest cells (Fig. 3.1B), but with enhanced spatial resolution. The endcap region and the region at mid-cell between the two pairs of DNA lobes are essentially devoid of RNAP. The number density of RNAP in the ribosome-rich region between nucleoid lobes is <10% of that in the densest part of the nucleoid lobes.

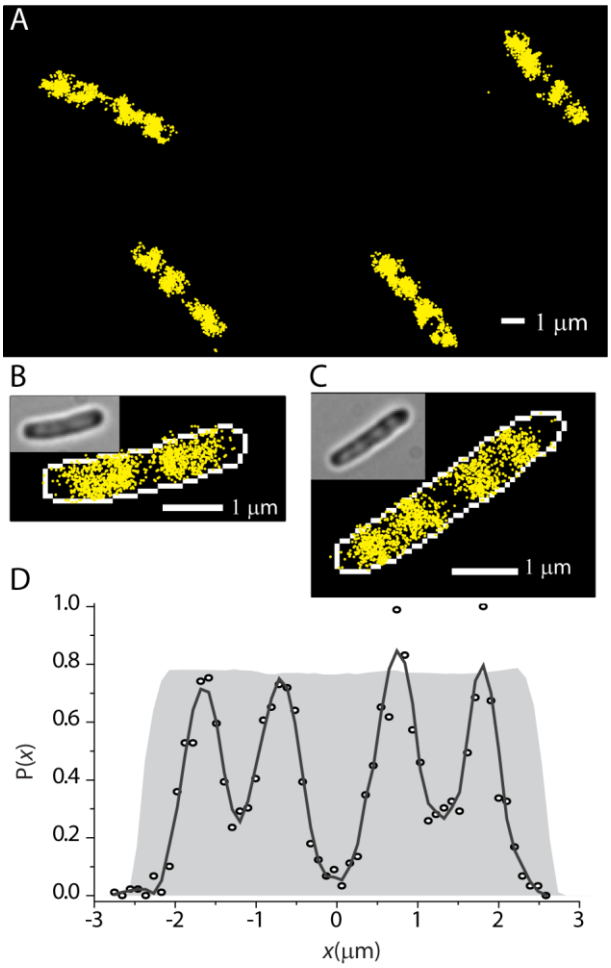

Figure 3.4. Superresolution images of RNA polymerase (β'-yGFP) within K-12 cells grown in EZRDM at 30ºC. (A) Four representative cells. (B) Expanded view of a medium-length cell (tip-to-tip length = 3.9 μm). (C) Expanded view of a long cell (tip-to-tip length = 5.2 μm). (D) Relative number of RNAP copies at each axial position *x* (100 nm bins). Solid line is the result of a Savitzky-Golay smoothing filter. The grey background shows the theoretical profile for a uniform distribution filling the model spherocylinder, as in Fig. 3.3.

Importantly, the *y*-coordinate profiles of the RNAP molecules, although noisy, clearly extend all the way to the cytoplasmic membrane (example in Fig. 3.5C). The presence of RNAP very near the cytoplasmic membrane suggests the presence of some DNA there as well, but only in the straight, cylindrical part of the cell. There is essentially no RNAP near the endcap portion of the cytoplasmic membrane (Fig. 3.4D). This has implications for the previously proposed transertion mechanism.

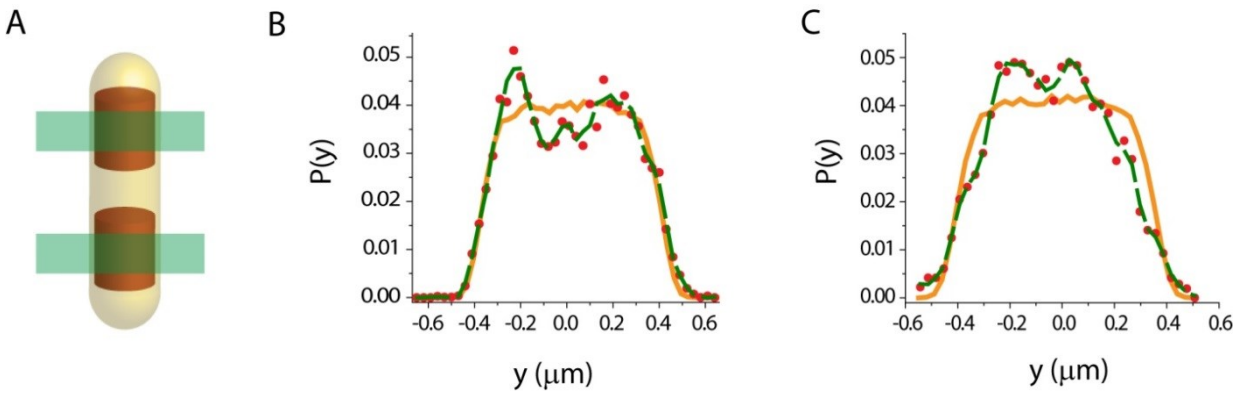

Figure 3.5: (A) Schematic of the two regions of *x* (green swaths) proximal to the dense nucleoid regions (brown cylinders) within which the *y*-distributions of ribosomes (S2-YFP) and RNAP (β'-yGFP) are analyzed. (B) Ribosome distribution P(*y*) projected onto the *y*-axis. Because there are few ribosomes proximal to the nucleoid, this is a composite of data from 12 cells. There is some broadening of the width due to slight misalignment of images of multiple cells. (C) RNAP distribution P(*y*) projected onto the *y* axis (Data from a single cell).

Copy numbers and number densities of ribosomes and RNAP

Because the S2-YFP gene replaces the normal S2 gene on the chromosome, all copies of S2 should carry the YFP label. We further assume that all S2-YFP copies are fluorescent and that essentially all S2 proteins are incorporated into 30S subunits. In Appendix G, we show that original and "revived" YFP copies have the same fluorescence intensity under single-molecule imaging conditions (Fig. 3.A6). We can then scale the total, pre-bleach YFP intensity to that of a single YFP molecule to estimate the copy numbers of ribosomes and RNAP in each cell (22). For ribosomes (meaning the sum of 30S subunits and complete 70S ribosomes), the resulting estimated copy number ranges from ~30,000 to ~70,000 ribosomes per cell (42 cells, Fig. 3.6C). The mean number of ribosomes per cell is ~55,000. The ribosome copy number increases monotonically with cell volume. The estimated copy number of β'-yGFP ranges

from ~2000 per cell to ~10,000 per cell. The average number of β'-yGFP per cell is 4,600 (31 cells, Fig. 3.6D). The number of β'-yGFP is correlated with cell volume.

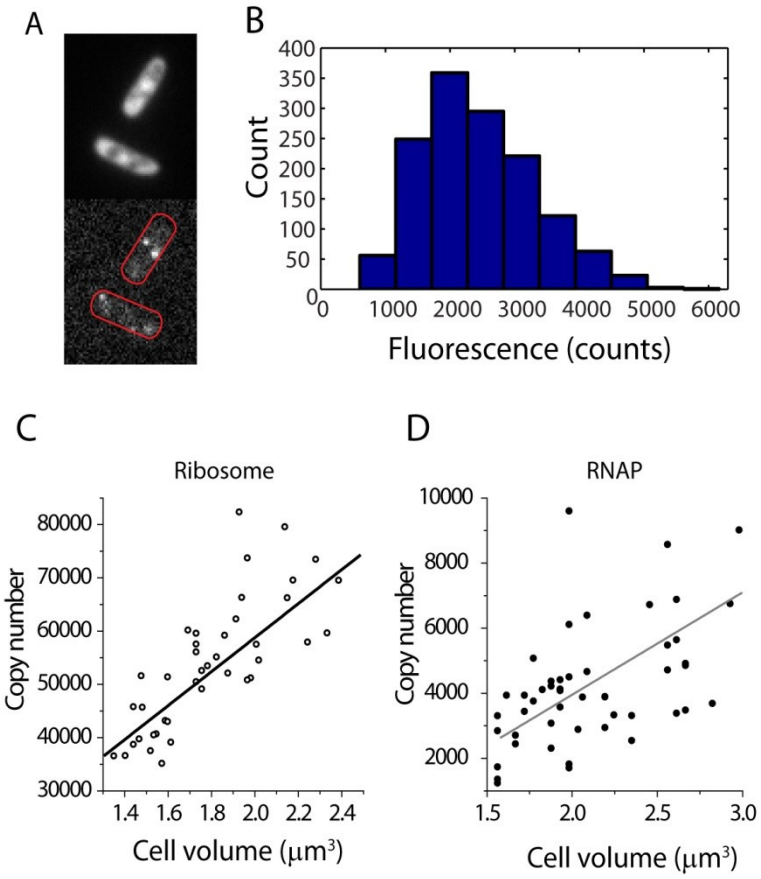

Figure 3.6. (A) Ribosome (S2-YFP) images by widefield (top) and superresolution method (bottom) of the type used to calibrate the single-molecule intensity. (B) Distribution of single-molecule intensities in camera counts. The mean of this distribution, when properly scaled for laser intensity and detector gain, is used to estimate the total number of fluorescent YFP copies per cell in the widefield images. (C) Ribosome copy number per calculated cell volume. (D) RNAP copy number per calculated cell volume. All data obtained in EZRDM at 30ºC.

To accurately determine cell geometry, we have also carried out superresolution imaging of free YFP expressed from a plasmid (Fig. 3.A8, (33)). The cell radius $R = 380 \pm 20$ nm is remarkably well conserved from cell to cell and is essentially independent of cell length. This enables calculation of the volume of each cell to about 10% accuracy. For each cell, the corrected fluorescence intensity and the known cell geometry then enables estimation of the average ribosome number density (*i.e.*, the average concentration over the entire cytoplasmic volume) and the peak ribosome number density in the ribosome-rich regions. The average ribosome number density varies from 25,000 to 31,000 ribosomes/μm^3 with a mean across cells of 27,000 ribosomes/μm^3. The peak number density within ribosome-rich regions varies from 56,000 to 73,000 ribosomes/μm^3, with a mean of 65,000 ribosomes/μm^3. These values should prove useful in modeling ribosome diffusion (38). As shown by the axial distributions, the peak ribosome density is usually very similar in the central region and in the endcaps. From axial distributions such as those in Fig. 3.3D, we estimate that about 10-15% of the total ribosome population lies within the DNA-rich nucleoid lobes.

Similarly, we use the total RNAP-yGFP intensity and the known geometry to calculate the average number density of RNAP within the cytoplasm. This varies across cells from 1800 to 3500 RNAP/μm^3, with a mean of 2200 RNAP/μm^3. The peak RNAP number density within the nucleoid is difficult to estimate because the nucleoid lacks a sharply defined volume.

Clearly under these moderate growth conditions at 30ºC, most of the 55,000 ribosomes and very few of the 4,600 RNAP molecules lie in the ribosome-rich regions of the cytoplasm. The cleanest quantitative estimates come from the endcaps, whose geometry is well defined; the central ribosome-rich region has poorly defined geometry. From composite data from five cells ranging in total length from 2.8 to 4.8 μm, we estimate that on average *at most* 180 copies (4%) of RNAP lie in the two endcaps combined. Essentially none of these few RNAP copies are proximal to the endcap membrane. The two endcap volumes contain some 17,000 ribosomes, either as 30S subunits or complete 70S ribosomes.

Diffusion of ribosomes

A detailed study of diffusion by tracking of single S2-YFP labels in two dimensions (using lower laser power) provides important clues to the nature of the species under study. The number of switched-on copies per cell was limited to 1-3 per frame, which usually enables accurate tracking of these slow-moving molecules. In practice, we were able to track 500-2000 molecules per cell over a period of 15-20 min. Reasonably good trajectories were obtained with 10 ms frames taken at 30 ms intervals. The distribution of trajectory lengths is roughly exponential, with a mean of 5-6 frames. In Fig. 3.7 we display those trajectories of 13 steps or longer for a single cell. There are typically ~50 such trajectories per cell. Trajectories longer than 13 steps are truncated at 13 for analysis. The trajectories are rather compact, and usually sample only one of the three ribosome-rich regions.

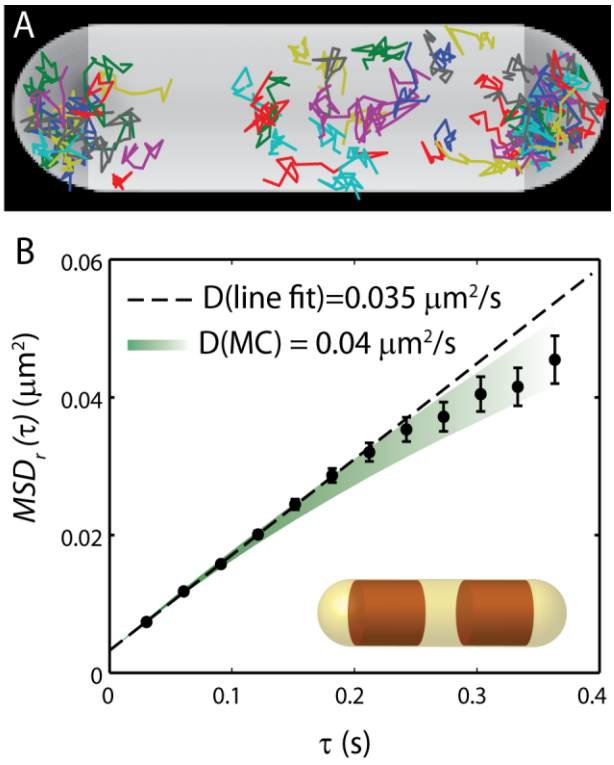

Figure 3.7. (A) Trajectories of single ribosomes (S2-YFP labels) within one cell, plotted within a spherocylinder chosen to match the phase contrast image. (B) Mean-square displacement vs lag time, averaged over 53 single-molecule trajectories from the same cell, each consisting of 13 steps. Error estimates are $\pm 1\sigma$ of the MSD values from single molecules. Dashed line is a linear fit to first three data points, yielding a diffusion coefficient estimate of $D = 0.035$ μm^2-s^{-1}. Colored swath represents the range of theoretical $MSD_r(\tau)$ plots ($\pm$ one standard deviation of the mean) obtained from averaging fifty 13-step Monte Carlo simulated trajectories using $D = 0.04$ μm^2-s^{-1}, the best-fit value as judged by eye. The simulations were run in a spherocylinder within which two truncated cylinders (representing the segregated DNA nucleoid lobes) block ribosome diffusion, as shown in the inset.

For a particular cell, the mean-square displacement vs lag time, $MSD_r(\tau)$, is computed as a running average (Fig. 3.7B). This is the average of over all molecules within one cell of the squared displacement from a starting location, with $r = (x^2 + y^2)^{1/2}$. Cells exhibit negative curvature of $MSD_r(\tau)$ after about five 30-ms steps. For free diffusion in an infinite volume, the mean-square displacement increases linearly in time. Finite volume in which to diffuse or sub-diffusion mechanisms (see below) could cause the MSD to rise more slowly than linearly in time.

We tested the effects of confinement in the ribosome-rich regions using Monte Carlo (random walk) simulations in a representative model geometry: a spherocylinder containing two impenetrable truncated cylinders to mimic the nucleoid lobes (Fig. 3.7B, inset). Initial particle positions were chosen randomly within the three regions available to ribosomes. The range of Monte Carlo outcomes for fixed D, each averaged over a number of trajectories similar to the amount of experimental data, is illustrated by the swath in Fig. 3.7. The simulations show that the curvature of $MSD_r(\tau)$ is reasonably well explained by a model of free diffusion of ribosomes in a confined space. After testing the data for five cells against Monte Carlo simulations, we found that simply fitting a straight line to the slope of the first four points of $MSD_r(\tau)$ underestimates D by a consistent factor of 1.15 ± 0.05. For consistency on 23 cells, we used the linear estimate with the correction factor to estimate D for each cell. The resulting distribution of D across cells is shown in Fig. 3.A10. The mean diffusion coefficient is $\langle D \rangle = 0.04 \pm 0.01$ $\mu m^2/s$ (± one standard deviation). The dispersion is likely due to a combination of errors in D estimation and real cell-to-cell variation in ribosome diffusion, as observed for other species (29).

The confined Monte Carlo simulations do not reproduce the curvature of the $MSD_r(\tau)$ plots within the experimental uncertainty. This may be due to the oversimplified model of nucleoid geometry—in practice the edges of the nucleoid are "soft" and likely do not completely exclude ribosomes. Monte Carlo simulations using $D = 0.04$ $\mu m^2\text{-}s^{-1}$ curve towards an asymptote of ~0.1 $\mu m^2\text{-}s^{-1}$, but much of the curvature occurs on a time scale longer than 1 s (beyond our experimental range). We cannot completely

rule out possible effects of sub-diffusion, defined here as a sub-linear increase in $MSD_r(\tau)$ due to causes other than geometric confinement. Two possible sub-diffusion mechanisms are: (1) monosomes or ribosomes translating mRNA that remains tethered to DNA by co-transcriptional translation (13, 23) and (2) entanglement of free monosomes or polysomes due to long, trailing segments of unfolded nascent polypeptide chains. While the curvature in $MSD_r(\tau)$ can be explained without recourse to sub-diffusion, the data would not be strongly affected by a 10% tethered component. The near linearity of MSD plots does not preclude the possibility that different labeled copies have different diffusion coefficients, as would be expected for 30S subunits vs polysomes. We tested for such heterogeneity of the diffusion by examining the distribution of *single-molecule* mean-square displacements at fixed time lag τ. We call this single-molecule quantity $msd_r(\tau)$; in contrast to $MSD_r(\tau)$, it is not averaged over all trajectories obtained for each cell. In Fig. 3.8A we used all trajectories of 10 steps or longer to form the distribution of $msd_r(\tau)$ within one particular cell. The experimental distribution is compared with that from analogous Monte Carlo simulations assuming free, homogeneous diffusion in the confined geometry of Fig. 3.7. The breadth of the model distribution illustrates how noisy short, diffusive trajectories are. When the simulation diffusion coefficient is chosen as $D = 0.04\ \mu m^2\text{-}s^{-1}$ to match the narrow peak of the experimental distribution, the experimental amplitude of the peak and the long tail on the data are not well fit. When the diffusion coefficient is chosen as $D = 0.05\ \mu m^2\text{-}s^{-1}$ to match the mean of the experimental data, the fit again fails. This indicates heterogeneity in the diffusive behavior, meaning at least two populations with diffusion coefficients sufficiently different that they are distinguishable over a lag time of 3 steps.

While the data almost surely contain contributions from *many* species with somewhat different diffusion coefficients, we have attempted to fit the data to the simplest heterogeneous model of only two non-exchanging populations. The fractional populations of the two components are f_1 and $f_2 = 1 - f_1$ and the different diffusion coefficients D_1 and D_2 are independent of time over the period of measurement. In

the present case, this might be slowly diffusing polysomes and faster diffusing monosomes or 30S subunits, for example. Composite data from three cells is reasonably well fit by a combination of two non-exchanging populations, ~80% of a slow component with $D_1 = 0.04\ \mu m^2\text{-}s^{-1}$ and ~20% of a faster component with $D_2 = 0.12\ \mu m^2\text{-}s^{-1}$, as shown in Fig. 3.8B. Even under the assumption of only two components, this combination of parameters is by no means unique. The percentage of the slow component can be varied from about 65-85% while still achieving a reasonable fit. The two-population model is almost surely too simple, but the resulting estimates of diffusion coefficients and populations help us interpret the underlying nature of the diffusing species (Discussion).

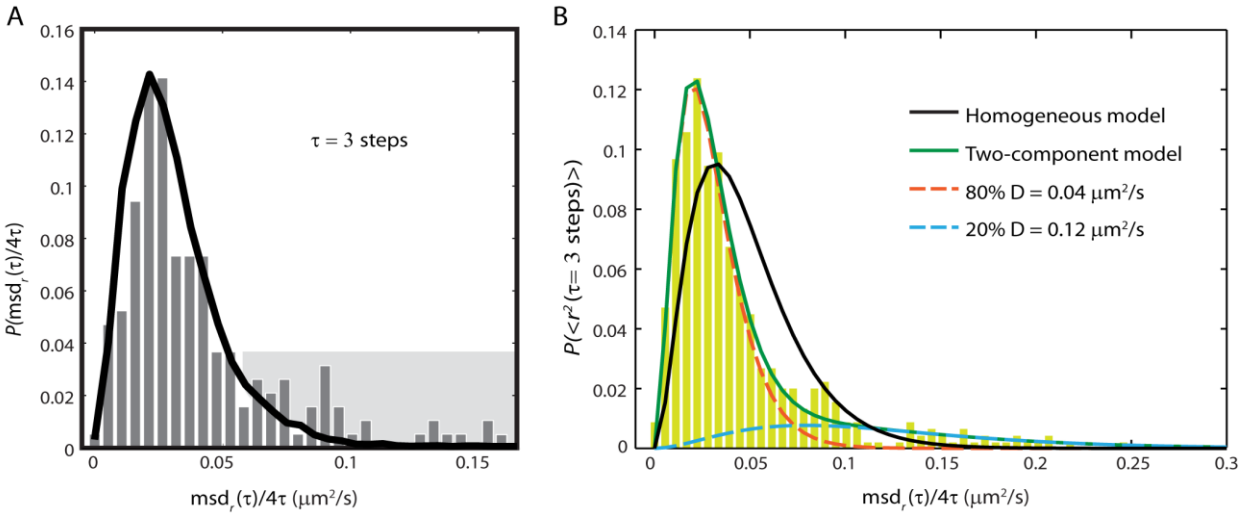

Figure 3.8. (A) Experimental distribution (bars) of best-estimate single-ribosome diffusion coefficients calculated from the three-step mean-square displacement. Each trajectory of ten steps or longer in a single cell is truncated at 10 steps. For each single-molecule j, $msd_{r,j}$ at lag time τ = 3 steps = 30 ms is calculated as a running average over all 10 steps. The single-molecule msd is estimated as: $msd_{r,j} = \langle r^2(\tau=3 \text{ steps})\rangle_j/4\tau$. The black line represents the distribution from a simulation with 10,000 trajectories of length 10 step, using a homogeneous diffusion coefficient of 0.04 $\mu m^2/s$ and normalized to match experiment. The geometric model includes two impenetrable cylinders to represent nucleoid lobes, as shown in Fig. 3.7. The peak and the shaded tail of the experimental distribution are not well fit by the model. (B) Histogram of 3-step $msd_{r,j}$ for combined ribosome data from 3 cells. Black line shows distribution for homogeneous free diffusion using D = 0.05 $\mu m^2\text{-}s^{-1}$, which corresponds to the mean of the experimental $msd_{r,j}$ values. Green line shows an example of a fit to a two-component distribution with 80% D = 0.04 $\mu m^2\text{-}s^{-1}$ and 20% D = 0.12 $\mu m^2\text{-}s^{-1}$. The fit is far from unique, but is useful for a rough description of the heterogeneity.

Effects of rifampicin and chloramphenicol on distribution and diffusion of ribosomes

As is well known, rifampicin (Rif) halts transcription initiation. Chloramphenicol (Chl) halts translation, evidently by preventing release of ribosomes. Both drugs are known to significantly affect the morphology of the nucleoid (36). On long time scales after drug treatment, Rif expands the nucleoid, while Chl further compacts the nucleoid (36). We have measured the time-dependent effects of these drugs on nucleoid-ribosome segregation and on ribosome diffusion. On a time scale of 8 min after treatment, we find that both Rif and Chl *contract* the nucleoid, both radially and axially (data not shown).

After 30 min of treatment with Rif at 200 μg/mL, the nucleoid has expanded axially, seemingly filling the cytoplasm uniformly (Fig. 3.9A). As measured by either widefield or superresolution microscopy, the ribosome distribution also expands (Fig. 3.9B). However, widefield intensity distributions along *y* show that the ribosome distribution remains somewhat broader than the DNA distribution (Fig. 3.9C). Evidently the DNA polymer continues to avoid the cytoplasmic boundary, in accord with a simple physical model of plectonemic DNA (12). At the same time, the diffusion coefficient of the S2-YFP labels is enhanced about 10-fold. To minimize confinement effects, here we choose *x* as the analysis coordinate. An example of an $MSD_x(\tau)$ plot is shown in Fig. 3.9D. The trajectories are short due to background introduced by Rif, which fluoresces weakly. In separate tests on wild-type cells, we have shown that the fluorescence quantum yield of Rif is too small to enable single-molecule detection. As estimated from the short-time slope, the diffusion coefficient after Rif treatment is $\langle D_{30S} \rangle \sim 0.6 \pm 0.2$ $\mu m^2/s$ (average over 9 cells). We tentatively attribute the much faster diffusion to 30S subunits. Using photo-activated pulse study, we have examined the long-time dynamics of the 30S subunits. As described in Appendix 3M, the 30S subunits are able to diffuse across the entire cytoplasm.

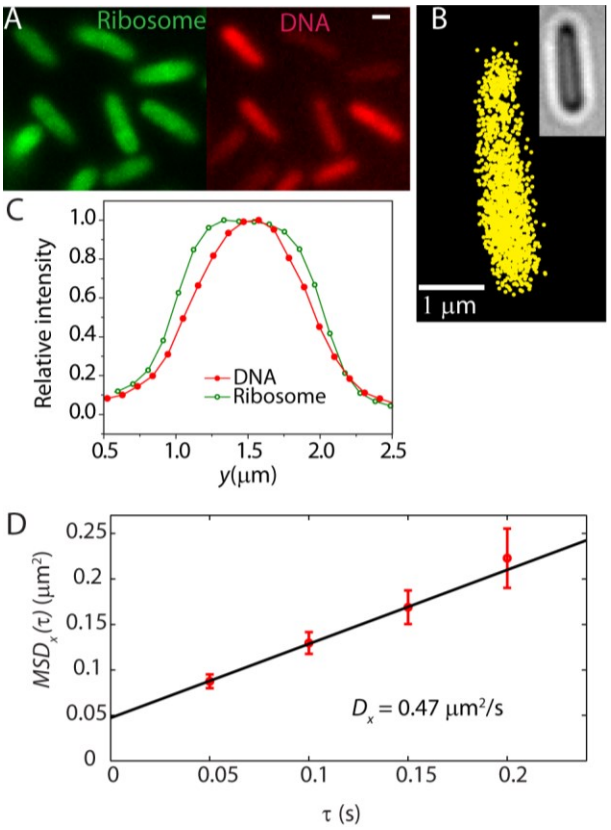

Figure 3.9. Effects of treatment with rifampicin. See text for details. (A) Widefield ribosome (S2-YFP) and DNA (DRAQ5) spatial distributions 30 min after rifampicin addition. Scale bar =1 μm. (B) Super-resolution image of ribosome distribution. See Fig. 3.3 for comparison with untreated cells. (C) Widefield intensity distributions for ribosomes and DNA along the short cell axis *y*. (D) $MSD_x(\tau)$ vs τ plot for single-ribosome diffusion in one cell after rifampicin treatment. Linear fit to the first 3 points yields the estimate $D = 0.47\ \mu m^2\text{-}s^{-1}$.

In sharp contrast, after 30 min of treatment with Chl at 200 μg/mL, the nucleoid has contracted both radially and axially, and very strong nucleoid-ribosome segregation is observed (Fig. 3.10A, C). The spatial distribution within a central 400-nm wide slice along the *x*-axis shows significantly greater ribosome-nucleoid segregation than in untreated cells, with peak-to-valley ratio of at least 10:1. From initial slopes of $MSD_x(\tau)$ plots (Fig. 3.10E), we obtain $\langle D \rangle = 0.05 \pm 0.02\ \mu m^2\text{-}s^{-1}$ (mean of 16 cells). Because Chl halts translation and does not permit release of the ribosomes, we tentatively attribute this value to a combination of ribosome monomers and polysomes similar to that in untreated cells.

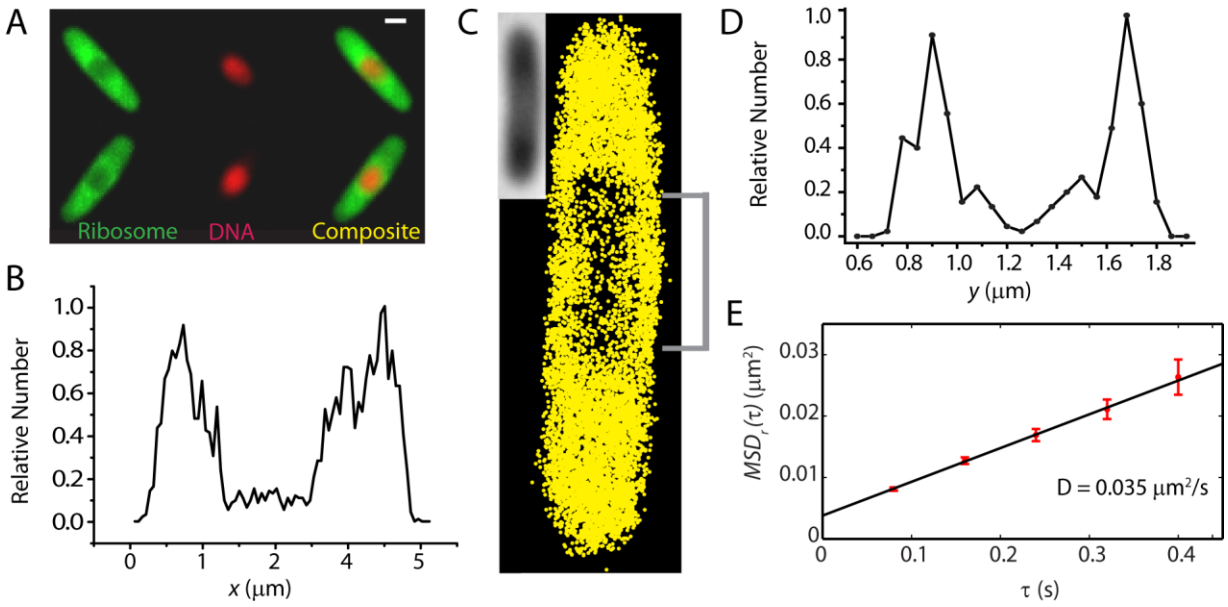

Figure 3.10. Effects of treatment with chloramphenicol. (A) Widefield ribosome (S2-YFP) and DNA (DRAQ5) spatial distributions 30 min after addition of chloramphenicol. Scale bar = 1 μm. (B) Axial ribosome profile from the super-resolution image shown in (C). Only the central 400 nm are included in the plot. (C) Superresolution image of ribosomes, 40 min after chloramphenicol treatment. (Inset = white light image of the cell). (D) Ribosome distribution along *y*, averaged over the 1 μm swath through the cell center as shown in (C). (E) $MSD_r(\tau)$ plot from single-particle tracking of ribosomes 40 minutes after chloramphenicol treatment. Linear fit yields the estimate D = 0.035 μm^2-s^{-1}.

Discussion

What S2-YFP-labeled species are located and tracked?

The first question is exactly what S2-YFP-containing species we are locating and tracking. Several time scales are important here. The maturation time of YFP *in vitro* is $\tau_{mat} \sim 10$ min; this is limited by the lag time between translation and folding of the protein and the oxidation step that renders it fluorescent (39). We assume a similar maturation time applies to the cytoplasm of *E. coli* (40). The rise time for exponential growth of *E. coli* in EZRDM at 30ºC is $\tau_g = 76$ min. This base-*e* rise time is 1.44 times the ~54-min doubling time $t_{1/2}$: $\tau_g = t_{1/2}/\ln 2$. As shown in the simple exponential growth model in Appendix E, the fraction of YFP that is mature is $f_{mat} = \tau_g/(\tau_{mat} + \tau_g) \sim 0.9$. The time for complete assembly of the 30S and 50S ribosomal subunits is $\tau_{ribo} \sim$ 2-3 min (41).

Probably most S2-YFP copies are maturing to become fluorescent *after* incorporation into a 30S subunit. This means that some 90% of total YFP copies should be fluorescent and essentially all the fluorescent copies we detect are incorporated into 30S subunits. There is no evidence of the small, free proteins S2-YFP or YFP, which are expected to diffuse very rapidly in cytoplasm (Appendix I). Based on the 55 kDa size of the S2-YFP construct, we estimate $D_{S2\text{-}YFP} \sim 5\ \mu m^2\text{-}s^{-1}$ (33), which is 100 times faster than the observed diffusion. If a significant fraction of the labels were free S2-YFP or YFP, the histogram of 3-step mean-square displacements (Fig. 3.8) would show a much longer tail. In Appendix I, we describe a rigorous test for rapidly diffusing molecules using faster frame times.

The species under study could be any combination of free 30S ribosomal subunits, translating 70S ribosome monomers, and polysomes (multiple 70S ribosomes simultaneously translating the same message). Classical data indicate that some 80% of ribosomes are translating at a given moment (42) (43), meaning that most 30S subunits are incorporated into translating 70S ribosomes. Accordingly, the diffusion of the labels is heterogeneous with two resolvable components (Figs. 8 b-C): roughly 80% slow,

with $D_{ribo} \sim 0.04\ \mu m^2\text{-}s^{-1}$ and roughly 20% faster, with $D_{30S} \sim 0.12\ \mu m^2\text{-}s^{-1}$. There is no obvious correlation of fast and slow diffusion with location in the different regions of space. Therefore, we infer that the faster component is due to free 30S subunits and the slower component is due to an undetermined mixture of ribosome monomers and polysomes. Because the 70S monosomes and polysomes are all translating mRNA, the slow component would then actually be attributed to diffusion of mRNA being translated by one or more ribosomes in live *E. coli.* Data from the Rif-treated cells confirms this inference. In Rif-treated cells, $D_{30S} \sim 0.6\ \mu m^2\text{-}s^{-1}$, some 15 times faster than the major component and five times faster than the minor component in untreated cells. This remains ten times too slow to arise from free S2-YFP copies. Instead, we suggest that the Rif-treated cells contain predominantly 30S subunits. Treatment with Rif is known to halt transcription (44). During the 30-min Rif treatment prior to observation, we expect all 70S ribosomes to complete translation of their messages and dissociate into free 30S and 50S subunits. Existing mRNA should degrade on a time scale of ~5 min. In the absence of new mRNA, the dominant YFP-labeled species should then be free 30S subunits. It is plausible that $D_{30S} \sim 0.6\ \mu m^2\text{-}s^{-1}$ in Rif-treated cells would be significantly faster than $D_{30S} \sim 0.12\ \mu m^2\text{-}s^{-1}$ in untreated cells because the 30S subunits are diffusing in a very different medium. The DNA meshwork has expanded, which would attenuate possible sieving effects on diffusion. The volume fraction occupied by ribosomal components (now 30S and 50S subunits) has roughly halved. As a result, crowding, sieving, and hydrodynamic effects should hinder diffusion significantly less for 30S subunits under Rif treatment than for 30S subunits in untreated cells.

A conceivable alternative explanation of our data posits that in untreated cells the faster component is due to 70S ribosome monomers translating mRNA and the slower component is dominated by polysomes. However, if that were so, we would have found no evidence of the ~20% 30S subunits expected in live *E. coli* (42). Furthermore, we expect a smooth distribution of the number of ribosomes translating mRNA copies: ribosome monomers, pairs of ribosomes translating the same message, trios,

and so forth up to larger polysomes. In all these cases, the diffusing entity is one or more ribosomes bound to a single mRNA message. Regardless of the details, the mean diffusion coefficient is expected to vary slowly with the number of ribosomes carried by a transcript. There is no reason for monomer diffusion to stand out from the other members of the distribution. In contrast, it seems sensible that diffusion of 30S might stand out from that of a smooth distribution of 70S monomers and polysomes. In addition to size and crowding considerations, free 30S is not bound to an mRNA message and is not trailing nascent polypeptide chains, both of which might impede diffusion of translating 70S ribosomes. We know little about the ribosome copy number per polysome.

Ribosome-nucleoid and protein-nucleoid segregation in bacterial cells

Our quantitative study of ribosome and DNA spatial distributions in live *E. coli* confirms previous reports of strong ribosome-DNA segregation in fixed cells. Early EM studies of thin slices of fixed *E. coli* cells by Robinow *et al* found ribosomes to be concentrated near the periphery of the cell and the nucleoid avoiding the near vicinity of the cytoplasmic membrane (2). From widefield immuno-fluorescence microscopy, the L7/L12 subunits of ribosomes were observed to localize outside the nucleoid (11). Our new live-cell results indicate that nucleoids of normal compaction strongly exclude ribosomes. We observe comparably strong ribosome-DNA segregation in the complete, chemically defined EZRDM, in minimal medium, and in the rich, undefined medium LB (Fig. 3.A.7).

Movies of widefield images with spatial resolution of ~250 nm show that strong nucleoid-ribosome segregation persists over the entire cell cycle. It is possible that the superresolution images acquired over several minutes underestimate the sharpness of the boundary between nucleoid and ribosome-rich regions due to time-dependent fluctuations in the underlying spatial distributions and also the possibility of complex, heterogeneous structure within dense nucleoid lobes on the scale of tens of nm. Regardless of these effects, the quantitative estimate of the fraction of ribosomes lying within the dense nucleoid regions in close proximity to DNA (10-15%) should be an accurate average over long periods of time.

In *E. coli* and in *B. subtilis*, we believe that strong exclusion of ribosomes from the dense portion of normal nucleoids determines the overall spatiobiology of transcription and translation. The present experimental results are consistent with a recent Monte Carlo simulation of ribosome-DNA segregation (12). The single DNA polymer lobe is modeled as a branched, freely jointed chain with persistence length chosen to mimic that of plectonemic DNA. The ribosomes are modeled as hard spheres, and polysomes as freely-jointed chains of 10 hard spheres. Both DNA and ribosome concentrations were taken from *E. coli* content measurements. This simplified model predicts strong polysome-DNA segregation while accounting only for the effects of excluded volume and plectoneme chain conformational entropy (Mondal *et al*., 2011). With only one nucleoid lobe, the model finds ribosomes concentrated in the two endcaps and in a thin cylindrical shell surrounding the nucleoid, much as observed here (Figs. 3.1B and 3.3B-D). The model does not include the transertion mechanism, which may anchor the DNA to the cell periphery by co-transcriptional translation of membranes proteins.

Two recent papers presented superresolution images of ribosomes in live *E. coli* under growth in minimal medium. Zhuang and co-workers (25) imaged S22-mEos2 and found ribosomes concentrated at the endcaps (their Fig. S3), much as observed here. Elf and co-workers show widefield images of S2-Dendra2 (23) that to our eye show strong ribosome segregation (their Fig. 3E). These copies of S2 were produced from the chromosome just like ours. However, the superresolution images of Dendra2-L25 (their Fig. 3D) seem to show less ribosome segregation. The initial slope of $MSD_r(\tau)$ yields a diffusion coefficient of 0.5 μm^2-s^{-1} for Dendra2-L25 single molecules (their Fig. 3A), which is much faster than our value of 0.04 μm^2-s^{-1} using S2-YFP. The Dendra2-L25 was expressed from a leaky plasmid, not from the chromosome. Evidently these plasmid-expressed copies are distributed more uniformly and diffuse much more rapidly than our S2-YFP copies expressed from the chromosome. Further comparison studies are needed to resolve this discrepancy.

In *C. crescentus*, the evidence indicates that both ribosomes and DNA are dispersed throughout the cytoplasm (13) in contrast to the results for *E. coli* and *B. subtilis* (5, 13). In *C. crescentus*, mRNA (possibly as a polysome) diffuses very little from the location of the gene from which it originated over some minutes (13). It is conceivable that the density or morphology or local dynamics of DNA is substantially different in *C. crescentus* in a way that simultaneously enables ribosome-DNA mixing and also entangles polysomes within the DNA meshwork, greatly slowing their diffusion.

Ribosome and RNA polymerase copy numbers

We estimate that the mean copy number of S2-YFP per cell is ~55,000 for a modified K-12 strain under moderate growth conditions at 30ºC. Based on heterogeneity in the distribution of single-molecule mean-square displacements (Figs. 3.8A and 3.8B) and on earlier content studies (45) (42), we estimate that some 80% are incorporated into complete 70S ribosomes and some 20% are in 30S subunits. As expected, the number of ribosomes per cell correlates with the cell volume (Fig. 3.6). Our estimate of mean copy number is about three times larger than the earlier value of 18,000 (interpolated between reported data for 40-min and 60-min doubling times) for *E. coli* B/r cells which were grown at 37ºC (45). Part of the difference may be due to different temperatures and growth media. In addition, the mean cell volume in our study is almost two times larger than the mean volume reported for *E. coli* B/r cells. Thus the number of ribosomes per unit volume is more similar, ~27,000/μm^3 in the K-12 strain vs ~18,000/μm^3 in the B/r strain.

Roughly 85% of the ribosomes lie in the ribosome-rich regions. What is the biopolymer volume fraction in these regions? In the endcaps, whose geometry is well defined, the average density of ribosomes is ~55,000 ribosomes/μm^3. Using an endcap radius of 400 nm, a ribosome radius of 10 nm, and assigning 30S one third the volume of a 70S ribosome, we estimate that ~0.20 of the volume in the endcaps is occupied by 30S and 70S copies. The ribosome volume fraction is evidently similar in the central ribosome-rich region, but quantitative estimates are more difficult for that less well defined

volume. In *E. coli*, the best quantitative estimates of mean, total biopolymer volume fraction (including protein, ribosomes, and nucleic acid) were obtained for K-12 cells in minimal growth conditions that do not match the current study (31). The result at normal growth osmolality was 0.16, about half of which was due to ribosomes. If we add the protein volume fraction of ~0.08 from minimal growth conditions to the ribosome volume fraction of ~0.20 from the present work, the total biopolymer volume fraction *in the ribosome-rich regions* could easily be ~0.28. This is large enough to exert a substantial excluded-volume effect on protein and ribosome diffusion (33).

These estimates enable rough comparison of our results to a recent model diffusion calculation treating cytoplasmic elements as appropriately sized hard spheres and including hydrodynamic interactions (38). The biopolymer mass density in the model cytoplasm varied from 250–350 mg/ml, with the mole fraction of each component (proteins and ribosomes) derived from content measurements and kept constant. This model obtained good agreement with experimental GFP diffusion coefficients without recourse to strong inter-particle attractions or unusually high viscosity of the cytoplasmic fluid. The estimate for single 70S ribosomes at 350 mg/ml (somewhat larger than our estimated volume fraction of ~0.30) is D_{ribo} = 0.2 μm^2-s^{-1}, which is five times larger than the experimental D_{ribo} ~ 0.04 μm^2-s^{-1} for our majority component. However, this is probably not an appropriate comparison. The model assumed a well-mixed cytoplasm of proteins and 70S ribosome monomers. That is, the model did not account for the strong segregation of ribosomes documented here or for the likely presence of polysomes. A more realistic model of the ribosome-rich regions would have a larger fraction of the biopolymer mass carried by ribosomes and would explore the effects of monosomes vs polysomes. Both factors would slow diffusion in the model and bring the calculations into closer agreement with our experimental results. As a final cautionary note, the diffusing entity in our study may not even be appropriately modeled as bare polysomes. Real polysomes are trailing long, nascent polypeptide chains. It is not known if these nascent

proteins are unfolded, partially folded, enveloped by chaperones, or fully folded. Long polypeptide chains and large chaperones might further hinder diffusion in the real experimental system.

The mean copy number of RNAP is estimated to be about 4600 per cell. An estimate from early work on B/r cells is about 4000, interpolated between results for 40-min and 60-min doubling times (45). The spatial distribution of RNA polymerase closely mimics that of the DNA itself (Figs. 3.2 and 3.4), consistent with strong non-specific and specific binding to DNA. Significantly, the widefield RNAP radial distribution in the vicinity of the nucleoids is slightly wider than the DRAQ5-DNA signal and extends all the way to the cytoplasmic membrane (Fig. 3.A9 and 5C). This suggests somewhat stronger association of RNAP to the outer regions of DNA lobes, in rough agreement with earlier images of fixed and immuno-stained cells (15). Transcription may occur preferentially near the interface between nucleoid and ribosome-rich regions. Finally, there is very little evidence of either RNAP (Fig. 3.4D) or DNA (Fig. 3.1B) proximal to the endcap membranes.

Co-transcriptional translation in E. coli

By co-transcriptional translation, we mean the translation of mRNA by ribosomes while the message is being transcribed, thus forming a large, branched biopolymer complex. The same phenomenon is often referred to as coupled transcription and translation. Early EM studies of *E. coli* cell lysate by electron microscopy observed DNA decorated with multiple RNAP molecules, each attached to a strand of mRNA which was in turn coated with a chain of ribosomes (8).

The high degree of DNA-ribosome and RNAP-ribosome segregation exhibited by our images does not prohibit co-transcriptional translation, but it strongly suggests that most protein is translated from free mRNA messages that have diffused into the ribosome-rich regions. That is, most translation is uncoupled from transcription (13). The reason is that we expect mRNA messages to be compact in space and thus unable to span typical distances of ~0.5 μm between DNA and the ribosome-rich regions. Polymers tend

to adopt compact configurations because they lose configurational entropy when stretched. There is direct evidence for compact polysome configurations from a recent cryoelectron tomography study; indeed, polysomes appear to adopt definite compact three-dimensional structures (7). It remains plausible that many of the ~10-15% of ribosomes located within the dense part of the nucleoid are carrying out co-transcriptional translation. These copies may also be undergoing ribosome assembly. However, the ~85-90% majority in the ribosome-rich regions are quite distant from the DNA and presumably unable to couple to DNA-tethered, partially synthesized transcripts.

Consistent with this argument, we found no clear evidence of tethered diffusion of ribosomes. The mild curvature in the MSD(τ) plots could be reasonably well explained by confinement within the ribosome-rich regions on the limited timescale of the tracking experiments. Furthermore, in these moderate growth conditions, the two endcaps contain very little of the DNA (as best we can discern from widefield images, Fig. 3.1B), some 17,000 of the 55,000 ribosomes, and at most 180 of the 4,600 RNAP copies (only 4%). Even in the unlikely event that all these RNAP copies were transcribing DNA, there would be at most 180 tethered partial messages in the endcaps at a given moment. That is far too little mRNA to engage the 17,000 ribosomes there. We suspect that the few "stray" RNAP copies in the endcaps are probably not transcribing. Instead, they are "lost" copies in search of a transcription initiation site. A future detailed study of RNAP diffusion will test this suggestion more critically.

The estimate of ~7000 ribosomes lying within the dense nucleoid regions is consistent with the suggestion that most of these are 70S ribosomes engaged in coupled transcription and translation. If ~1200 RNAP copies (25% of the total) are actively transcribing at a given moment, then some 1200 nascent mRNA copies are coupled to DNA. If the mean length of a complete transcript is ~1000 bp (contour length of 350 nm), then the mean contour lengths of these partially completed transcripts is ~175 nm. If 70S ribosomes of 20-nm diameter were close-packed along the nascent mRNA, there would be

sufficient total contour length to accommodate some 10,000 ribosomes. It is thus plausible that many of the copies within the dense nucleoid are 70S ribosomes actively translating.

Based on these arguments, we suggest that complete mRNA copies typically find their way to the ribosome-rich regions by diffusion. They are likely decorated by ribosomes and perhaps cold-shock proteins that protect them from degradation *en route* (7, 17). Based on $D_{ribo} = 0.04\ \mu m^2\text{-}s^{-1}$, a ribosome-decorated mRNA copy diffuses a root-mean-square distance of 0.5 μm (about half the size of a nucleoid lobe) in only 1 s. The time to transcribe an average-length protein message of 1000 nt is $\tau_{trx} \sim 20$ s (13, 45). The mean lifetime of an mRNA copy relative to degradation is $\tau_{mRNA} \sim 5$ min (16). After transcription and release a complete message will find the ribosome-rich region very quickly. The fraction of protein production coupled to transcription would be roughly $\tau_{trx}/\tau_{mRNA} \sim 0.07$. This is roughly comparable to our estimate of the fraction of ribosomes found within the nucleoids, ~0.10-0.15.

Our evidence that the majority of translation is uncoupled with transcription agrees with earlier conclusions based on the ratio of τ_{trx}/τ_{mRNA}. It stands in apparent conflict with the early EM study of *E. coli* cell extracts (8), in which no free polysomes were found. Essentially all the ribosome-decorated mRNAs were attached to chromosomal DNA.

The spatiobiology of transcription/translation in *C. crescentus* may be very different from that in *E. coli* and *B. subtilis*. Using both FISH (fluorescence *in situ* hybridization, which requires fixation and permeabilization) and live-cell imaging methods, Jacobs-Wagner and co-workers recently found that mRNA remains localized near the site of transcription for long periods (13). Images of mRNA were punctal and co-localized with markers of the corresponding gene. Most ribosomes did not diffuse in photobleaching/recovery experiments, as if they were translating mRNA that is somehow tethered to the DNA. In addition, the ribosomes and chromosomal DNA are distributed homogeneously in the *C. crescentus* cytoplasm. Even if messages do not diffuse long distances, they are already in close proximity

to ribosomes (13). We do not presently understand the underlying causes of the different DNA-ribosome morphologies in *C. crescentus* vs *E. coli* and *B. subtilis.*

Transertion in E. coli

A substantial body of evidence indicates that most membrane proteins are inserted co-translationally by the signal recognition particle SRP in conjunction with the Sec translocase (46). The transertion hypothesis refers to co-transcriptional translation and simultaneous insertion of membrane proteins (6). Because transertion would tether the chromosomal DNA to the cytoplasmic membrane, Woldringh and others have suggested that transertion acts as an "expanding force" on the DNA itself (6, 10, 36). In this view, the spatial extent of the chromosomal DNA is a compromise between expanding forces (primarily transertion) and compacting forces (supercoiling, inter-strand coupling by architectural proteins, and the tendency of branched polymers to avoid walls and maximize configurational entropy) (10). Woldringh further suggested that serial transertion of proteins into the crowded membrane might assist DNA segregation (6).

The present work adds circumstantial, quantitative support for the transertion mechanism. First, we have directly observed time-dependent radial shrinkage of the nucleoids at t ~ 8 min after treatment with either rifampicin or chloramphenicol (data not shown). Both treatments eventually halt translation, which will break the link between the chromosomal DNA and the cytoplasmic membrane. Second, in our moderate growth conditions we find clear evidence that the radial distributions of both ribosomes and RNAP extend all the way to the cytoplasmic membrane in close proximity to the nucleoid lobes (Figs. 3.5B and 3.5C). While these membrane-proximal copies are plausible candidates for the transertion mechanism, we have no evidence as yet that these RNAP copies are transcribing DNA or that these ribosomes are translating protein.

To the extent that our results support the feasibility of the transertion mechanism, they do so only at the cylindrical part of the cytoplasmic membrane near the flanks of the dense nucleoid lobes. All the

evidence argues against transertion at the endcaps. There are very few if any RNAP copies near the endcap membrane regions (Fig. 3.4D), and the (lower resolution) widefield DNA axial distribution dies off well before reaching the endcap membranes (Fig. 3.1B). Extensive coupling of the endcap membranes to DNA would evidently require stretching an mRNA message some 300-500 nm. In our view, transertion might well supply a direct radially expanding force on the nucleoids, but probably cannot supply a direct axially expanding force. Thus halting of transertion can readily explain radial compaction of the nucleoids (which occurs for both Rif and Chl treatment at t = 8 min), but not the observed axial contraction under Chl treatment. It remains possible, however, that direct insertion of proteins into a crowded cytoplasmic membrane in effect pulls the nucleoid outward as the cell grows (6). The correlation between the outward movement of segregated DNA lobes and the cell growth as observed in Fig. 3.1D provides some support for such a mechanism.

Appendix 3A: Numerical test of the precision of centroid fitting

Determination of particle position by calculating the centroid of the fluorescent spot rather than fitting the distribution to a Gaussian function requires justification by numerical simulation. This is especially true in view of an earlier numerical study that found large systematic error and poor localization accuracy from a centroid algorithm (47). However, the Cheezum study used an 80 x 80 pixel grid to calculate the centroid and did not apply a high-pass spatial filter to the image. This meant that the numerous pixels far from the image spot have tremendous leverage in the centroid. In contrast, our method smooths and filters the image first, yielding a fluorescent spot that lies in a roughly 5 x 5 grid of pixels. We use an 8 x 8 grid to calculate the centroid. Most of the pixels near the edge of the grid have values near zero, so they do not unduly bias the centroid value. We prefer centroid estimation to Gaussian fitting because diffusing molecules make non-circularly symmetric images, centroid calculation is very fast, and the results are easily compared with Monte Carlo calculations of diffusion.

To estimate the accuracy of centroid vs Gaussian fitting for ribosomes in our imaging conditions, we simulated random walks of a particle with $D = 0.04\ \mu m^2\text{-}s^{-1}$ on a square lattice with 5 nm spacing between points. At each time step of 0.171 ms, the particle moves 0 or ±1 step to left or right and 0 or ±1 step up or down. For each random walk, a 170-step trajectory was formed to match the experimental 30 ms/frame. Each position was blurred with a two-dimensional Gaussian of $\sigma_{x,y} = 75$ nm, closely approximating the point spread function using a NA = 1.49 objective at $\lambda = 560$ nm. The sum of the 170 broadened images was used to represent a 30 ms/frame image of a single, diffusing ribosome. The images were re-binned into a 20 x 20 pixel images with 100 nm pixels, matching experiment. Two noise components were also added: shot noise with $\sigma_{shot} = n^{1/2}$, where n is the number of photons in the pixel, and a background noise contribution chosen from a Gaussian distribution with $\sigma_{background}$ varied to achieve

different peak signal-to-noise ratio (PSNR). The PSNR is defined as the peak signal pixel value divided by the standard deviation of the background, $(\sigma_{background}^2 + \sigma_{shot}^2)^{1/2}$.

In Fig. 3.A.1, we show sample simulated images for PSNR = 6 and 11; the relevant range for our experiments is 8-15. The adjacent images have been smoothed and filtered as for experiment, removing much of the background. Analysis of the particle locations was performed in the same way as for the tracking data from experiment. The localization results for one particular random walk trajectory with PSNR = 6 and 11 are shown for the two algorithms, centroid and Gaussian fitting, in Fig. 3.A.1 A,B. Each *x*-*y* graph covers a 200 nm x 200 nm region at the sample. The true particle centroid of the random walk on the fine grid is shown as a '+' at the center of the region. For that one random walk, we carry out 100 realizations of the shot noise and background noise and locate the particle using the centroid calculation and the symmetric Gaussian fitting on the 8 x 8 pixel grid. The resulting localizations are shown as the dots. Systematic errors are modest for both methods, as evidenced by the symmetric placement of dots about the central +. For a diffusing particle, each trajectory has its own peculiar shape. To assess the effects of the variable shape of trajectories, we did the same analysis for 10 unique trajectories. The localization precision using Gaussian fitting vs centroid calculation are compared in Fig. 3.A.1D as a function of PSNR. For PSNR = 6-17, the precision (standard deviation) of the centroid algorithm varied from ~30 nm to ~10 nm. In the manuscript, we state this as σ = 30 nm in each dimension. The standard deviation is 10-15% higher for centroid calculation vs Gaussian fitting at each PSNR. Unlike the Cheezum result, the centroid algorithm does not fail catastrophically at modest PSNR. Presumably Gaussians are somewhat more precise because the diffusion coefficient is small enough to create fairly round images. We suspect, but have not shown, that centroid fitting may be preferred in cases of more severe blurring due to faster diffusion, where the images become more highly asymmetric.

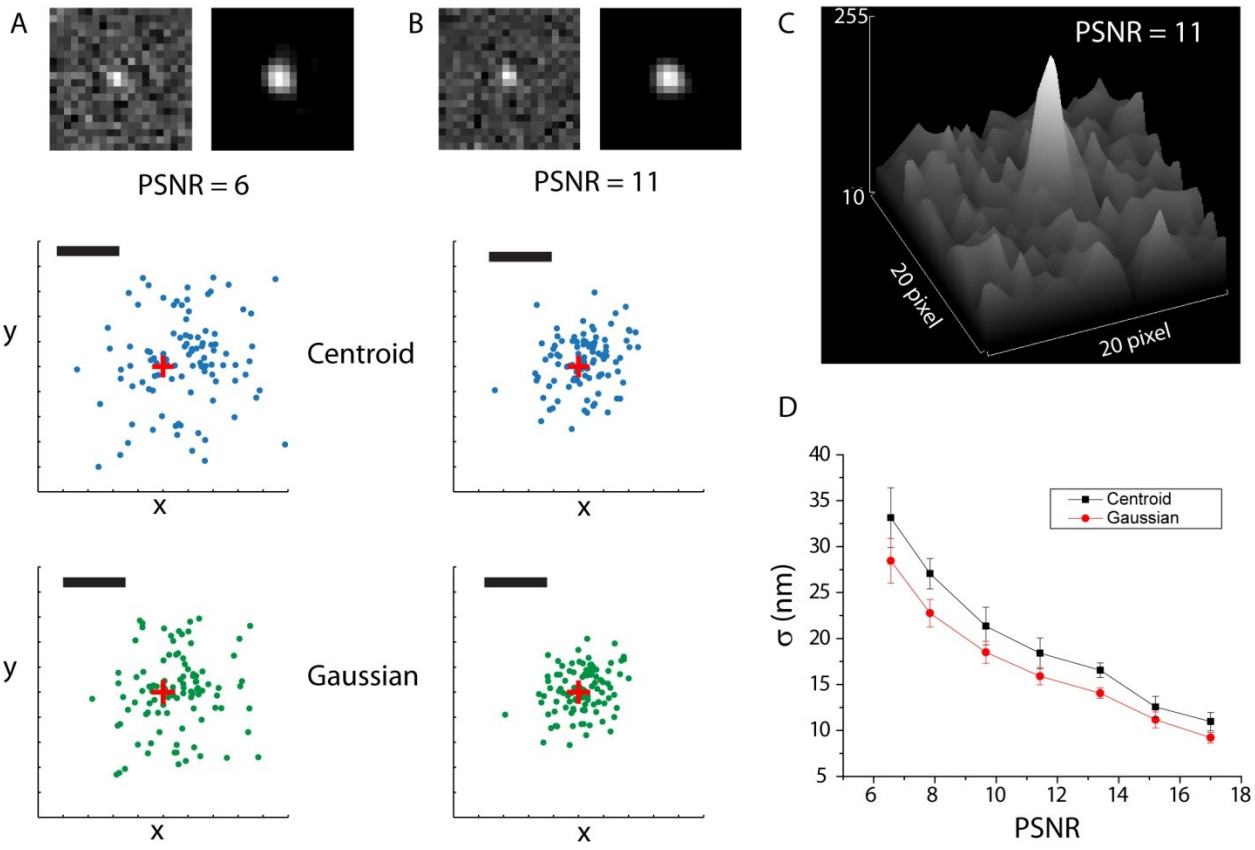

Figure 3A.1. Top panel shows two simulated images with PSNR 6 and 11 (left: raw, right: filtered). The bottom panels show the localization results for 100 realizations of noise for the same trajectory obtained by applying the centroid or Gaussian fitting to the resulting image. The red '+' is the exact centroid of the trajectory. Scale bar = 50 nm. (C) Intensity level across the pixels of an image with PSNR = 11 is shown for the 20 x 20 pixel region around the peak. (D) Mean localization precision (1 standard deviation) vs PSNR is shown from 10 simulated trajectories, using Gaussian and centroid fitting algorithm. Error bars represent one standard deviation of the distribution of localization precisions.

Appendix 3B: Creating a binary mask and cell outline from the white light images

Simple white light images obtained with the 1.49 NA objective are used for determining the cell outline. We have used available thresholding algorithms and developed our own algorithm to generate cells masks and outlines from the low contrast white light images. The principal axes of these images become the *x*, *y* coordinate system used in plotting spatial distributions.

Before image processing the white light images (Fig. 3.A.2A) are inverted by subtracting all the pixel values from the maximum pixel value. In this format the cell appears as a bright object on a dark background (Fig. 3.A.2B). Then we make a preliminary binary mask of the cell using a threshold. The threshold is determined by the Matlab function 'greythresh', which utilizes Otsu's method (48) of computing the threshold. We then use the Laplacian and Gaussian (LoG) algorithm of edge detection, as explained by Silusarenko, *et al.* (49). This image is saved as Image1 (Fig. 3.A.2C). The inverted image created from the original image is then used to create another filled mask. We use the Sobel method of edge detection by determining the locations of maximum gradient of intensity (default algorithm in MATLAB's "edge" function). Then we dilate the image using the Matlab function 'imdilate'. Finally the holes in this image are filled (a hole is set of isolated background pixels) to create the filled mask Image2 (Fig. 3.A.2D). By combining the Image1 (edge detection) and Image2 (filled mask) together using the AND operation, we generate the final mask (Fig. 3.A.2E). This mask is used for determining principal axes of the cell to compute the necessary rotation matrix to place images of different cells along the same axes. The outline of this mask can be simply by the outermost pixels of the mask (Fig. 3.A.2F).

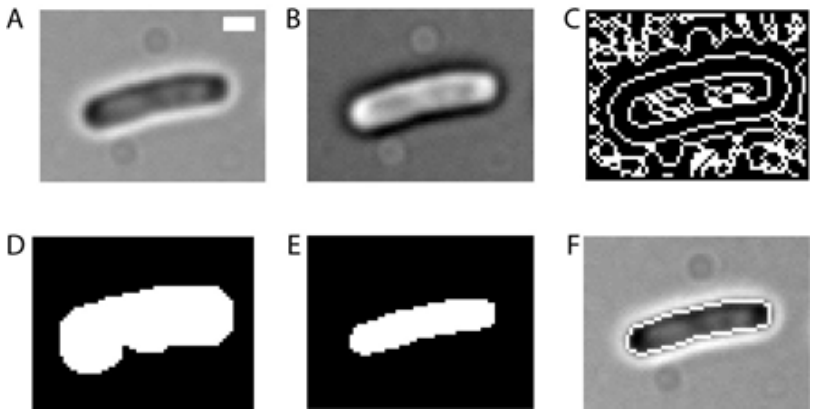

Figure 3A.2: Steps of image processing to create mask from a white light image of the cell.

Appendix 3C: Measuring cell doubling time for plated cells

In a bulk culture, the doubling time of a construct is obtained by measuring optical density (OD) vs time. To estimate doubling times for plated cells under our imaging conditions, we have used phase contrast images to measure cell length vs time, both in the flow chamber and in coverslip-slide sealed chambers. We used a phase contrast objective (NA = 1.30) to create time-lapse movies of 50-ms frames with 18-sec intervals for 200 frames (60 min). These images were analyzed using Matlab. The inverted phase contrast image was used to create a binary mask outlining each cell. Masks for many cells were created and separated with connected component analysis (50). For each binary mask, the major axis length and the centroid of the region were determined. The length of the major axis is taken as the cell length. For following the growth of individual cells in a sequence of frames, the centroids in successive frames are compared. Using the Delaunay triangulation method, we determine the nearest neighbors in successive frames and call them the same cell. This allows for movement of cells and microscope stage drift over 60 min. We fit plots of cell length vs time to the function:

$$L(t) = L_0 \cdot 2^{t/t_{1/2}} \qquad \text{(A1)}$$

Here the usual doubling time is $t_{1/2}$ and the initial cell length is L_0. The base-e time constant for growth used in the main text and in the kinetics scheme below is $\tau_g = t_{1/2}/\ln 2 = 1.44\ t_{1/2}$. An example data set is shown in Fig. 3.A.3. In chambers the strains grow 15% slower than the hot air shaker.

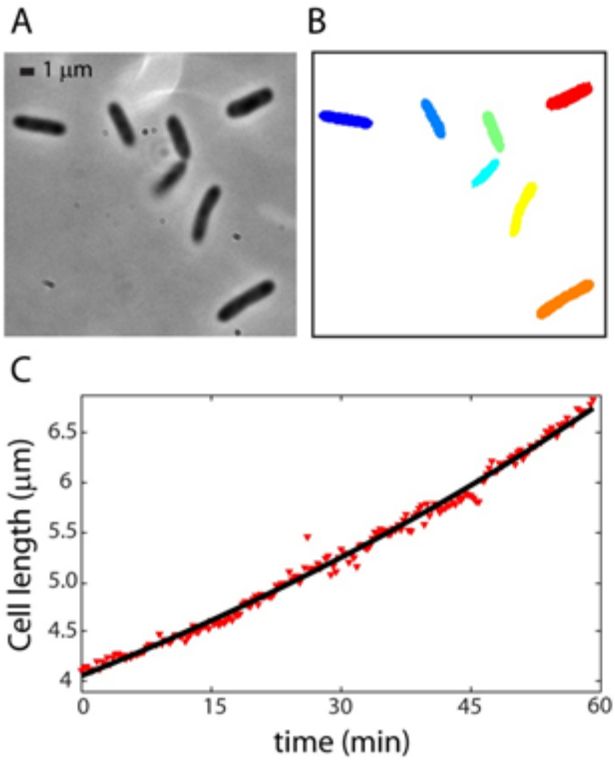

Figure 3A.3. (A, B) Phase contrast images of 7 cells (scale bar 1μm) and the corresponding binary mask used to define the (x,y) coordinate system and measure cell length. (C) The major axis length for the binary mask of the left-most cell (dark blue) is plotted over 200 frames taken at 18-s intervals. Solid line is a fit to Eq. A1 to obtain the doubling time $t_{1/2}$ = 75 min.

Appendix 3D: Effects of laser illumination on DNA and ribosome distribution

For the superresolution spatial distributions of S2-YFP and β'-yGFP, images were typically obtained at 3 kW/cm^2 of 514 nm light on the sample for periods of 2-3 min. For the single- molecule tracking experiments we used 0.5-2 kW/cm^2 over a longer period of 10-15 min. We inadvertently discovered that still higher laser exposure alters the spatial distribution of ribosomes and of DNA and also alters the diffusive properties of ribosomes. For example, after 10 min of 514 nm illumination at 6 kW/cm^2, both ribosome and DNA distributions become much more homogeneous, and the mean ribosome mobility appears faster than in unperturbed cells. Similar effects were observed at 561 nm.

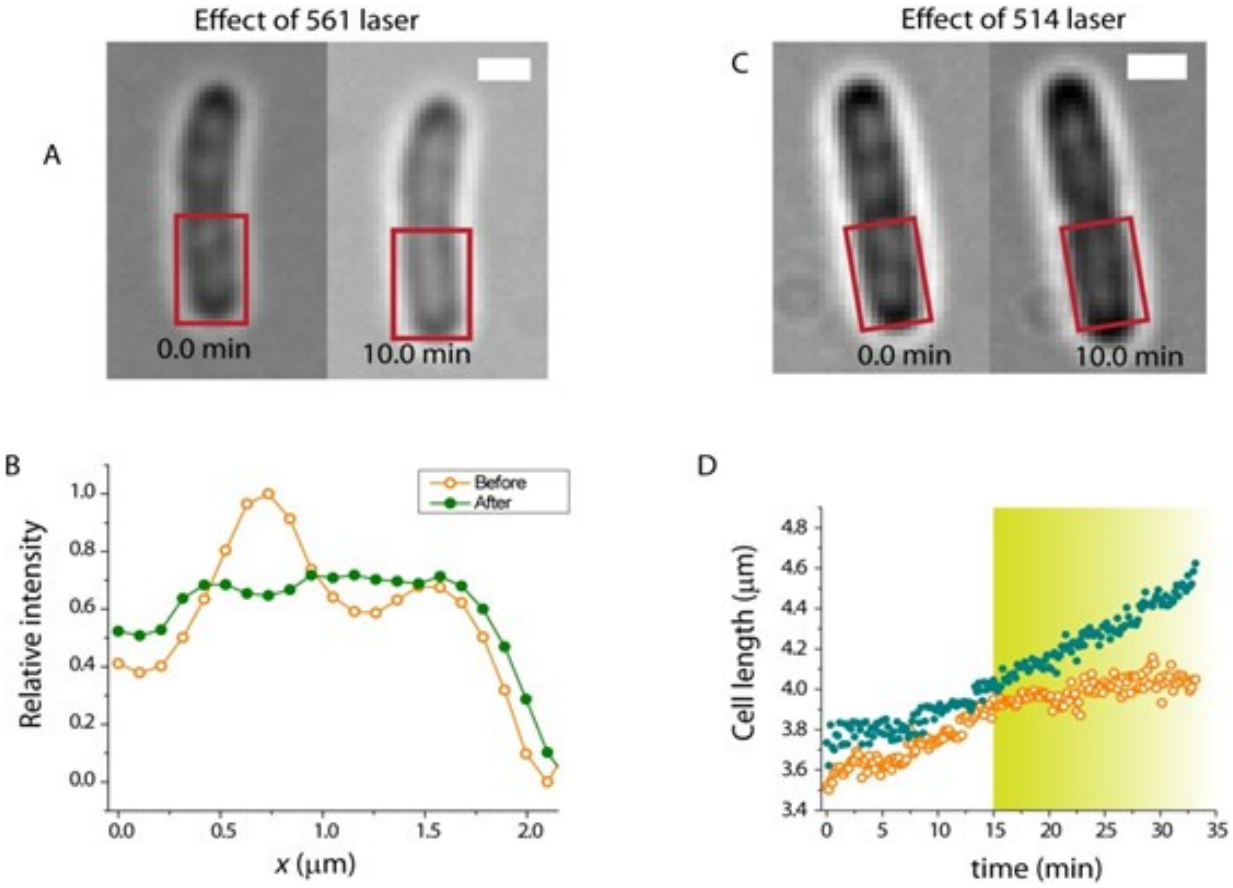

Figure 3A.4. Laser effects. White light image of a cell clearly indicates the distribution of nucleoid as the lighter sections in a cell image. (A) White light image of WT cells before and after illumination at 561 nm at 7 kW/cm^2 for 10 min. Rectangle highlights one nucleoid sub-lobe that seems to fuse due to the laser treatment. (B) Axial white light intensity within the rectangle before and after illumination. (C) White light image of WT cells before and after 10-min illumination at 514 nm (6 kW/cm^2). (D) Cell length vs time for two cells from the same imaging area. One cell (orange data) was illuminated by the 514 nm laser at 6 kW/cm^2 while the other cell was not. The laser was turned on at t = 15 min, at which time growth of the illuminated cell slows immediately. Scale bar =1 μm.

Appendix 3E: Kinetics scheme for cell growth and YFP maturation

Here we use Y to denote the number of immature, non-fluorescent copies of YFP and Y* to denote the number of mature, fluorescent copies. The overall rate constant for exponential growth of cells is k_g. The base-*e* rise time for exponential cell growth is $\tau_g = k_g^{-1}$; the more usual "doubling time" is $t_{1/2} = \tau_g/\ln 2 = 1.44\ \tau_g$. The unimolecular rate constant for YFP maturation is k_{mat}. The simplest kinetics scheme assumes that the total rate of production of new YFP copies in the culture is proportional to the current total population, (Y + Y*). This amounts to assuming that as cells grow exponentially, the total YFP population is proportional to total cellular mass. Production and maturation of YFP is then governed by the simple equations:

$$dY/dt = k_g(Y + Y^*) - k_{mat}Y \tag{A2}$$

$$dY^*/dt = k_{mat}Y \tag{A3}$$

In exponential growth, the ratio Y / Y* should be constant:

$$Y^* = cY \tag{A4}$$

Thus
$$dY^*/dt = c\,dY/dt \tag{A5}$$

Combining (A2), (A3), and (A5) yields the result:

$$k_g c^2 + (k_g - k_{mat})c - k_{mat} = 0 \tag{A6}$$

whose solution is:
$$c = k_{mat}/k_g \tag{A7}$$

Thus
$$Y^* = (k_{mat}/k_g)\,Y = (\tau_g/\tau_{mat})Y \tag{A8}$$

Appendix 3F: Photon counts from revived YFPs

We have tested that the YFP fluorescence intensity under single-molecule imaging conditions is the same for the original copies as for the "revived" copies that return to a fluorescent state over minutes after photobleaching. We needed very few original copies for this test to work, so we used a cell line expressing the fusion protein HU-YFP from a plasmid inducible by tetracycline. Due to strong binding to DNA, HU-YFP has a diffusion coefficient D_{HU} of about 0.05 μm^2-s^{-1}, comparable to the ribosomes under study. Our strategy was to create small copy numbers of original fluorescent molecules by imaging cells in the absence of inducer. In practice, cells exhibited roughly 200-500 original fluorescent copies, far fewer than the YFP-labeled ribosomes or RNAPs. When the 514 nm laser is turned on continuously, the original copies photobleach very rapidly. After several seconds, only a few or no fluorescent YFP copies are observed. Then some 20-50 photobleached copies return gradually, typically over a time period of several minutes. By measuring single-molecule intensities over the period 3-10 s after the laser is turned on, we obtain an intensity histogram that is due primarily to copies in the original fluorescent population. We also require these copies to be located within an earlier clump of fluorescent species that have mostly photobleached, i.e., not to appear suddenly on a dark background as revived copies would. Then we continue to measure single-molecule intensities over the subsequent two-minute period and make a second histogram, this time including only molecules that appear suddenly on a dark background.

The two histograms of integrated camera counts from original vs revived copies state are compared in Fig. 3.A.5. The distribution dominated by original copies has a mean of 104 ± 40 arbitrary intensity units (± 1σ). The distribution of revived copies has a mean of 106 ± 49 arbitrary intensity units. The good agreement strongly suggests that revived copies have the same photophysical properties in the cytoplasm as original copies.

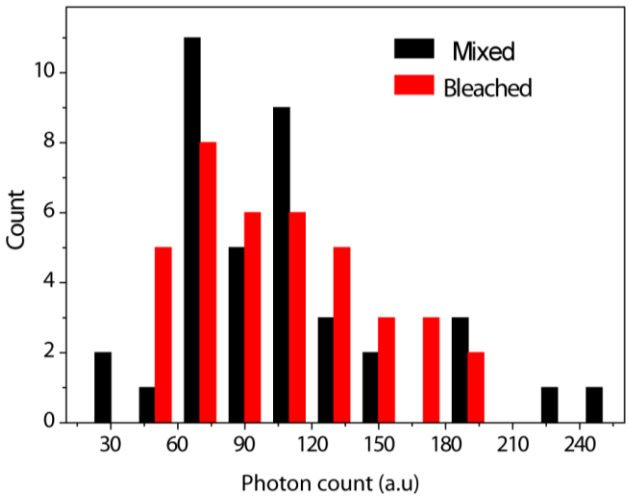

Figure 3A.5. Comparison of photon count from reversibly bleached ("revived") HU-YFP and a mixed population of mostly original, unbleached HU-YFP plus some revived HU-YFP.

Appendix 3G: Imaging Ribosome-DNA segregation with low-NA objective

The widefield ribosome images are somewhat broadened by diffraction. However, the primary broadening is due to out-of-focus light collected by the 1.49 NA microscope objective, whose depth-of-focus parameter is only 500 nm (±250 nm about the focal point) (35). Widefield images of the ribosome distribution taken with a low numerical aperture objective (NA = 1.25, depth-of-focus parameter 660 nm) show stronger axial segregation (Fig. 3.A.6; peak-to-valley intensity ratios of 2:1 or even 3:1).

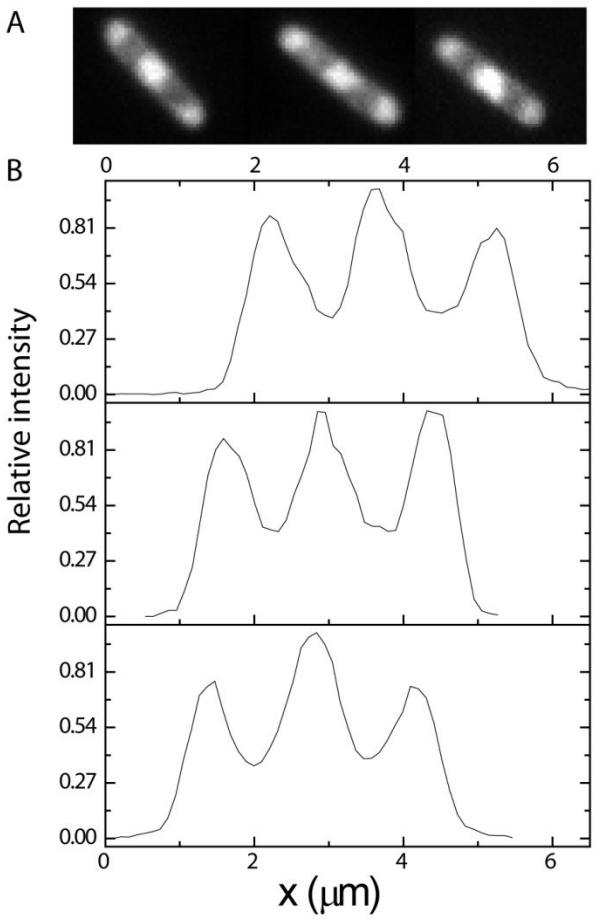

Figure 3.A.6. (A) Widefield images of ribosomes (S2-YFP) taken with 1.25 NA objective. (B) Axial intensity profiles of ribosome distributions showing larger peak-to-valley ratios (2.0 to 2.5) than those from the 1.49 NA objective (Fig. 3.1).

Appendix 3H: Ribosome-DNA segregation at different growth conditions

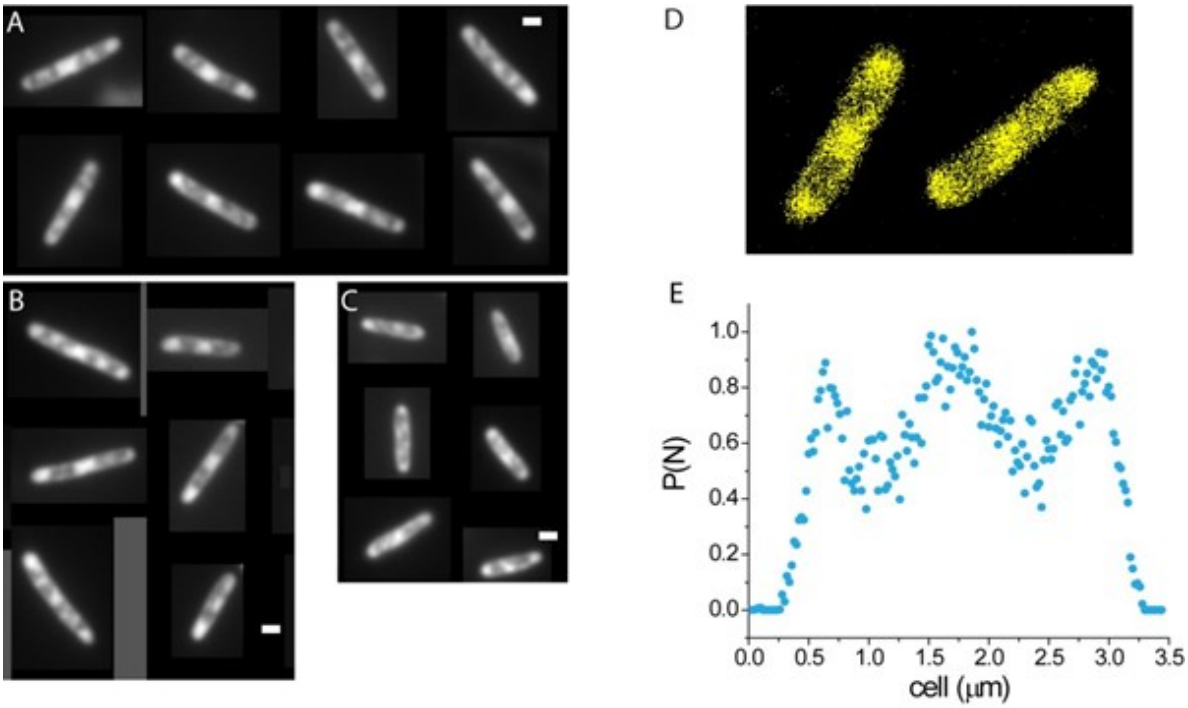

Figure 3A.7. Examples of widefield ribosome (S2-YFP) images for *E. coli* cells growing in three different media. All images are taken at 30ºC. (A) Neidhardt defined complete medium, EZRDM. (B) Luria broth, LB. (C) Minimal MOPS-buffered medium, MBM. Bulk culture doubling times are 30 min, 55 min, and 87 min, respectively. Scale bar =1 μm. (D) Superresolution ribosome (S2-YFP) image of two cells grown in minimal MBM. (E) Axial ribosome distribution for a cell grown in minimal MBM.

Appendix 3I: Test for fast-diffusing free S2-YFPs

In the usual ribosome tracking movies, we found no evidence of rapidly diffusing species such as S2-YFP or bare YFP. However, it might be argued that rapidly diffusing molecules would make blurred images that are not detected by the superresolution thresholding algorithm for 15-ms exposure times. Could the tail on the measured distribution be due to free S2-YFP or YFP copies that happen to diffuse slowly enough to be detected? To test this possibility, we plotted the spatial distribution of only the "faster" molecules in Fig. 3.6 and found it to be segregated into three ribosome-rich regions like the "slower" molecules. Based on Kaede and YFP superresolution images, we would expect free S2-YFP or YFP copies to be homogeneously distributed throughout the cytoplasm. In addition, we imaged S2-YFP using 4-ms exposure times appropriate for such rapid diffusion (33). There was no change in the ribosome spatial distributions. Finally, the distribution of 2-step mean-square displacements from the 4-ms movies remained consistent with $D \sim 0.04\ \mu m^2\text{-}s^{-1}$; there was no evidence of diffusion at $\sim 5\ \mu m^2\text{-}s^{-1}$. The absence of free S2-YFP suggests that the cell does not make a substantial excess of S2 relative to the number of 30S subunits.

Appendix 3J: Determining of cell dimension from superresolution of YFP

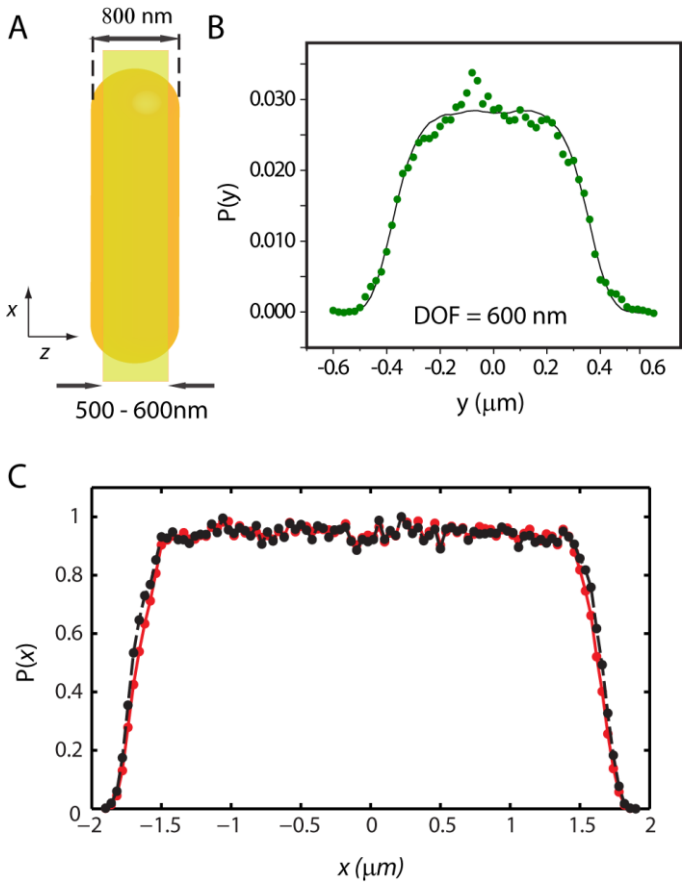

Figure 3A.8: (A) Diagram shows sectioning along *z* by 1.49 NA objective. Single-molecule detection is most sensitive to molecules in a central slab of 500-600 nm thickness. (B) Experimental P(*y*) for free YFP. Best-fit model curve uses the 600-nm thick slab and radius *R* = 380 nm. The thickness of the detection slab may vary with single-particle signal-to-noise ratio, diffusion coefficient, and tuning of the intensity threshold for localizing a single molecule. We conclude that for S2-YFP, detectability is best for molecules in a slab of 500-600 nm thickness. (C) Monte Carlo distribution of projections onto *x*-axis for randomly distributed locations within a spherocylinder of diameter 760 nm and cell length *L* = 3600 nm, including endcaps. Black: no sectioning. Red: Detection only in a 500 nm thick slab (*z* = ±250 nm). Sectioning has little effect on axial distribution.

Appendix 3K: Radial profile of DNA and RNAP

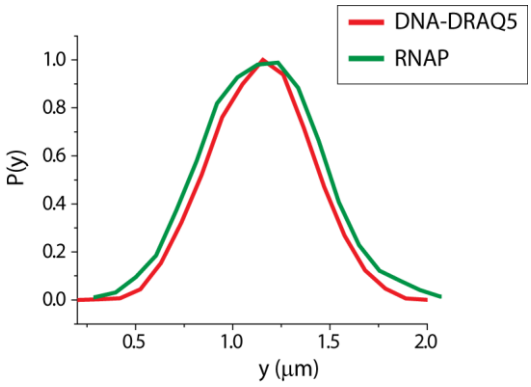

Figure 3A.9: Radial distribution of RNAP (green curve) is broader than radial distribution of DRAQ5-DNA (red curve) from the same cell.

Appendix 3L: Cell-to-cell variation of ribosome diffusion coefficient

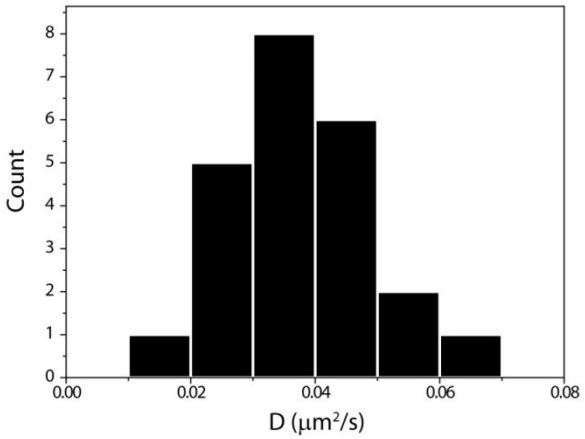

Figure 3A.10. Distribution of diffusion constant of ribosome from 23 cells.

Appendix 3M: Photo-activation pulse chase experiments of ribosome-mEos2 dynamics in rifampicin treated cells

To see if the low density of the chromosomal DNA after its expansion with rifampicin treatment allows ribosomes to diffuse throughout the cytoplasm, we have monitored ribosome dynamics after rifampicin treatment. To examine if ribosomes can diffuse across the entire cell, we have used photo-activation pulse chase studies to follow the spatial distribution of ribosome intensities over time. The experimental details of the photo-activated pulse chase method are explained in the Appendix 3N. According single particle tracking results, the 30S subunits diffuse quite fast in the rifampicin treated cell ($D_{30S} \sim 0.6\ \mu m^2\text{-}s^{-1}$). With our current time resolution of pulse chase experiments (1s), it is difficult to probe such fast diffusion in a 3-4 micron long cells as ribosomes diffuse across the cell within 2-3 seconds. Therefore we have treated cells with cephalexin (60 μg/mL) to halt septation. This causes cells to become longer as they can't divide and after 2 doubling times we obtained cells that have average length of ~10 μm. An example cell is shown in Fig. 3.A.11. A focused photo-activation beam was pointed at one tip of the cell. Before the activation the mEos2 labeled ribosomes of the cell don't emit any signal in the 561 excitation channel. After the photoactivation pulse, signal appears at the tip (red arrow in top-left image of Fig. 3.A.11). Within 30 seconds after the activation pulse the intensity of the activated ribosome-mEos2 spread across almost the entire length of the cell. Different time points along this diffusive redistribution of the ribosomes are shown in Fig. 3.A.11. This experiment clearly demonstrates that after the expansion of the chromosomal DNA in a rifampicin treated cell, 30S subunits can diffuse throughout the cell.

We performed Monte Carlo simulations to model the pulse chase data of ribosome diffusion in rifampicin treated cells. The observed dynamics of ribosomes is consistent with a diffusion constant of

0.5 $\mu m^2/s$ (data not shown). This is consistent with the estimated diffusion constants of 30S subunits from single molecule tracking (D_{30S} = 0.6 $\mu m^2/s$).

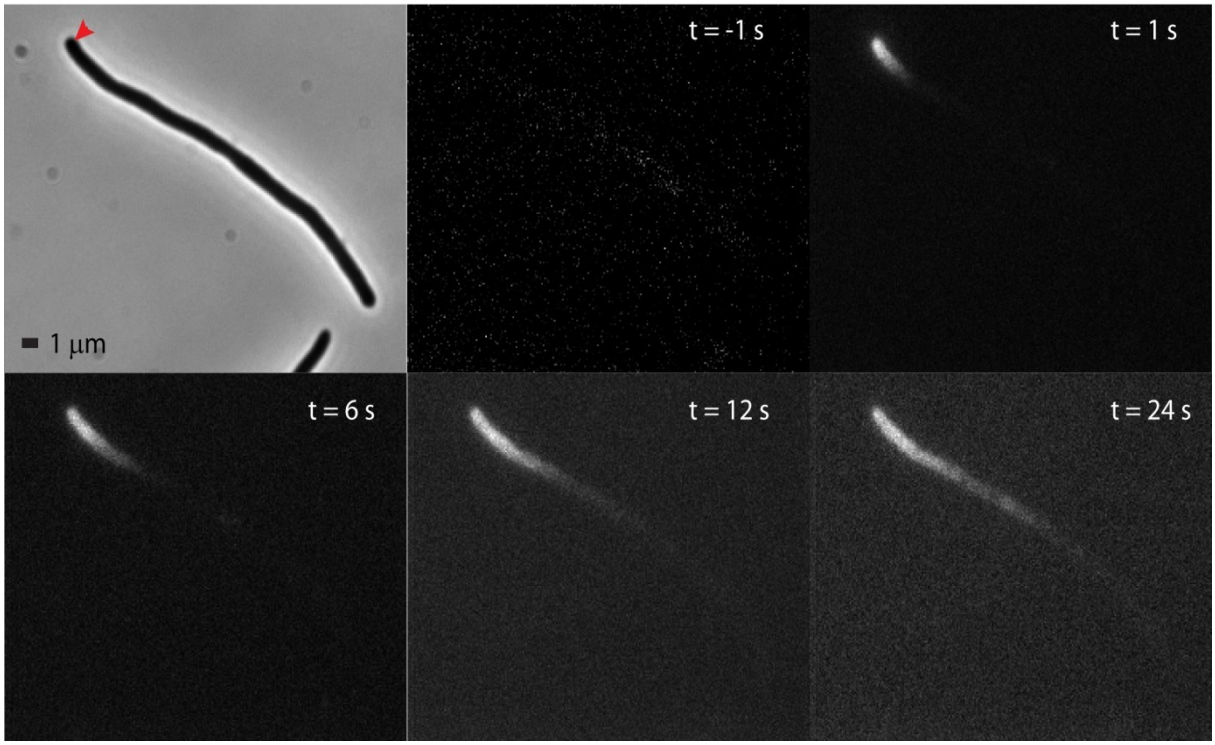

Figure 3.A.11: Photo-activated pulse chase study of ribosome diffusion 30 min after rifampicin addition. The top-left panel shows the phase contrast image of the cell (treated with cephalexin). Scale bar =1 μm. At t = 0 s the pulse of 405 nm beam is used to activate molecules at the tip of the cell (shown with the red arrow head in the top-left panel). Intensity distribution of ribosome-mEos2 immediately before (t = -1 s, top middle) and after (t = 1s, top right) the activation pulse are shown. Three more time-points (t = 6 s, 12 s, and 24 s) are shown in the bottom panel.

Appendix 3N: Experimental set-up for Photo-activation pulse chase

A schematic of the photo-activated pulse chase study is given in the Fig. 3.A.12. The intensity of the mEos2 labeled ribosomes molecules (strain details are available in chapter 5) is probed with a 561 nm beam. The beam is flat and it is pulsed to take an image every 1s with 50 ms exposure within each frame. Initially almost all the mEos2 molecules are in the inactive states and therefore they don't have any emission in the 561 excitation channel. At time t = 0 we activate a number of molecules at one tip of the cell using a focused 405 nm activation laser (shown as the purple spot in Fig. 3.A.12). The 561 nm laser channel then probes the spatial distribution of these activated molecules over time.

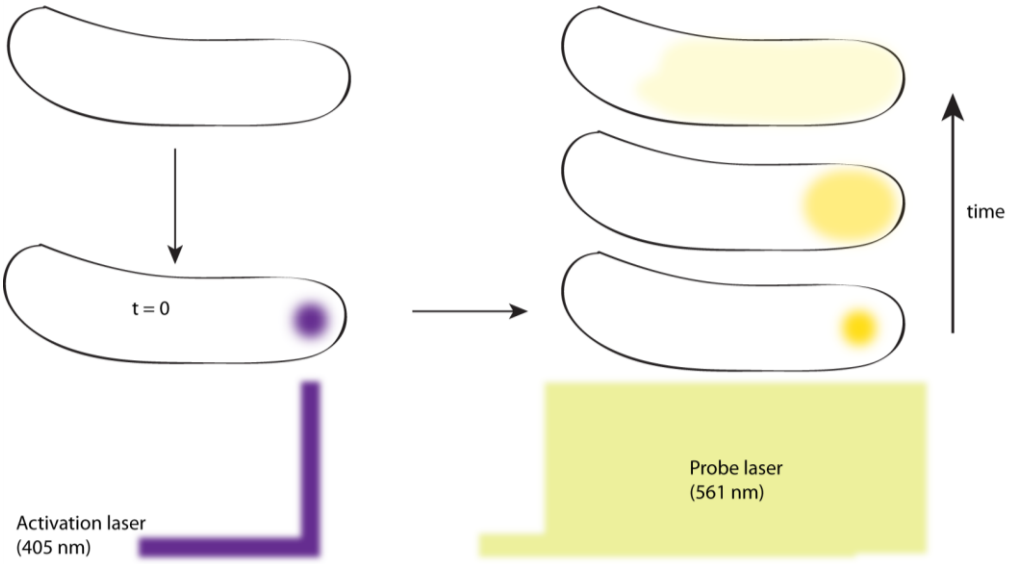

Figure 3.A.12: Schematic of photo-activated pulse chase experiment.

References

1. Lewis, P. J. 2004. Bacterial subcellular architecture: Recent advances and future prospects. Mol. Microbiol. 54:1135-1150.

2. Robinow, C., and E. Kellenberger. 1994. The bacterial nucleoid revisited. Microbiol. Rev. 58:211-232.

3. Nielsen, H. J., J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen. 2006. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. Mol. Microbiol. 62:331-338.

4. Shapiro, L., H. H. McAdams, and R. Losick. 2009. Why and how bacteria localize proteins. Science. 326:1225-1228.

5. Nevo-Dinur, K., A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder. 2011. Translation-independent localization of mrna in e. *coli*. Science. 331:1081-1084.

6. Woldringh, C. L. 2002. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Mol. Microbiol. 45:17-29.

7. Brandt, F., S. A. Etchells, J. O. Ortiz, A. H. Elcock, F. U. Hartl, and W. Baumeister. 2009. The native 3d organization of bacterial polysomes. Cell. 136:261-271.

8. Miller, O. L., B. A. Hamkalo, and C. A. Thomas. 1970. Visualization of bacterial genes in action. Science. 169:392-&.

9. Lewis, P. J., S. D. Thaker, and J. Errington. 2000. Compartmentalization of transcription and translation in bacillus subtilis. EMBO Journal. 19:710-718.

10. Woldringh, C. L., and N. Nanninga. 2006. Structural and physical aspects of bacterial chromosome segregation. J. Struct. Biol. 156:273-283.

11. Azam, T. A., S. Hiraga, and A. Ishihama. 2000. Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. Genes Cells. 5:613-626.

12. Mondal, J., B. P. Bratton, Y. Li, A. Yethiraj, and J. C. Weisshaar. 2011. Entropy-based mechanism of ribosome-nucleoid segregation in e. *coli* cells. Biophysical Journal. 100:2605-2613.

13. Llopis, P. M., A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. 2010. Spatial organization of the flow of genetic information in bacteria. Nature. 466:77-U90.

14. Mascarenhas, J., M. H. W. Weber, and P. L. Graumann. 2001. Specific polar localization of ribosomes in bacillus subtilis depends on active transcription. EMBO Rep. 2:685-689.

15. Durrenberger, M., M. A. Bjornsti, T. Uetz, J. A. Hobot, and E. Kellenberger. 1988. Intracellular location of the histone-like protein HU in *Escherichia coli* J. Bacteriol. 170:4757-4768.

16. Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. Global analysis of mrna decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proceedings of the National Academy of Sciences of the United States of America. 99:9697-9702.

17. El-Sharoud, W. M., and P. L. Graumann. 2007. Cold shock proteins aid coupling of transcription and translation in bacteria. Sci. Prog. 90:15-27.

18. Golding, I., and E. C. Cox. 2004. Rna dynamics in live *Escherichia coli* cells. Proc. Natl. Acad. Sci. U. S. A. 101:11310-11315.

19. Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science. 313:1642-1645.

20. Hess, S. T., T. P. K. Girirajan, and M. D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys. J. 91:4258-4272.

21. Rust, M. J., M. Bates, and X. W. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). Nat. Meth. 3:793-795.

22. Taniguchi, Y., P. J. Choi, G. W. Li, H. Y. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. Quantifying e-*coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science. 329:533-538.

23. English, B. P., V. Hauryliuk, A. Sanamrad, S. Tankov, N. H. Dekker, and J. Elf. 2011. Single-molecule investigations of the stringent response machinery in living bacterial cells. Proc. Natl. Acad. Sci. U. S. A. 108:E365-E373.

24. Lee, S. F., M. A. Thompson, M. A. Schwartz, L. Shapiro, and W. E. Moerner. 2011. Super-resolution imaging of the nucleoid-associated protein HU in caulobacter crescentus. Biophys. J. 100:L31-L33.

25. Wang, W. Q., G. W. Li, C. Y. Chen, X. S. Xie, and X. W. Zhuang. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. Science. 333:1445-1449.

26. Norris, V., and M. S. Madsen. 1995. Autocatalytic gene expression occursviatransertion and membrane domain formation and underlies differentiation in bacteria: A model. Journal of Molecular Biology. 253:739-748.

27. Batchelor, E., and M. Goulian. 2006. Imaging ompr localization in *Escherichia coli*. Mol. Microbiol. 59:1767-1778.

28. Datsenko, K. A., and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* k-12 using pcr products. Proceedings of the National Academy of Sciences of the United States of America. 97:6640-6645.

29. Bratton, B. P., R. A. Mooney, and J. C. Weisshaar. 2011. Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. J. Bacteriol. 193:5138-5146.

30. Neidhardt. F. C., P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. Journal of Bacteriology. 119:736-747.

31. Konopka, M. C., J. C. Weisshaar, and M. T. Record, Jr. 2007. Methods of changing biopolymer volume fraction and cytoplasmic solute concentrations for in vivo biophysical studies. Osmosensing and Osmosignaling. 428:487-504.

32. Biteen, J. S., M. A. Thompson, N. K. Tselentis, G. R. Bowman, L. Shapiro, and W. E. Moerner. 2008. Super-resolution imaging in live caulobacter crescentus cells using photoswitchable eyfp. Nature Methods. 5:947-949.

33. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live *Escherichia coli*. Biophysical Journal. 101:2535-2544.

34. Crocker, J. C., and D. G. Grier. 1996. Methods of digital video microscopy for colloidal studies. J. Colloid Interface Sci. 179:298-310.

35. Hall, D. 2008. Analysis and interpretation of two-dimensional single-particle tracking microscopy measurements: Effect of local surface roughness. Anal. Biochem. 377:24-32.

36. Cabrera, J. E., C. Cagliero, S. Quan, C. L. Squires, and D. J. Jin. 2009. Active transcription of rrna operons condenses the nucleoid in *Escherichia coli*: Examining the effect of transcription on nucleoid structure in the absence of transertion. J. Bacteriol. 191:4180-4185.

37. Cabrera, J. E., and D. J. Jin. 2003. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Molecular Microbiology. 50:1493-1505.

38. Ando, T., and J. Skolnick. 2010. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proc. Natl. Acad. Sci. U. S. A. 107:18457-18462.

39. Nagai, T., K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat. Biotechnol. 20:87-90.

40. Sniegowski, J. A., J. W. Lappe, H. N. Patel, H. A. HUffman, and R. M. Wachter. 2005. Base catalysis of chromophore formation in arg(96) and glu(222) variants of green fluorescent protein. Journal of Biological Chemistry. 280:26248-26255.

41. Lindahl, L. 1975. Intermediates and time kinetics of *in vivo* assembly of *Escherichia coli* ribosomes. Journal of Molecular Biology. 92:15-37.

42. Forschhamer, J., and L. Lindahl. 1971. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of *Escherichia coli* 15. Journal of Molecular Biology. 55:563-568.

43. Young, R., and H. Bremer. 1976. Polypeptide-chain-elongation rate in *Escherichia coli* b/r as a function of growth rate. Biochemical Journal. 160:185-194.

44. Campbell, E. A., N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb, and S. A. Darst. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell. 104:901-912.

45. Dennis, P. P., and H. Bremer. 1974. Macromolecular-composition during steady-state growth of *Escherichia coli* b/r. J. Bacteriol. 119:270-281.

46. Driessen, A. J. M., and N. Nouwen. 2008. Protein translocation across the bacterial cytoplasmic membrane. Annual Review of Biochemistry. 77:643-667.

47. Cheezum, M. K., W. F. Walker, and W. H. Guilford. 2001. Quantitative comparison of algorithms for tracking single fluorescent particles. Biophys. J. 81:2378-2388.

48. Otsu, N. 1979. Threshold selection method from gray-level histograms. Ieee Transactions on Systems Man and Cybernetics. 9:62-66.

49. Sliusarenko, O., J. Heinritz, T. Emonet, and C. Jacobs-Wagner. 2011. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. Mol. Microbiol. 80:612-627.

50. Rosenfel.A, and J. L. Pfaltz. 1966. Sequential operations in digital picture processing. Journal of the Acm. 13:471-&.

Chapter 4

Partitioning of RNA polymerase Activity in Live *Escherichia coli* from Analysis of Single-molecule Diffusive Trajectories

Introduction

The new super-resolution techniques of fluorescence microscopy (1-3) enable location and tracking of single molecules in live bacterial cells with spatial resolution of ~30 nm and temporal resolution as fast as several ms. For the first time, specific molecular processes occurring within the small volume of the bacterial cytoplasm can be dissected in space and time (4). In earlier work on *E. coli*, we used a YFP-based photobleaching/recovery method (5) to estimate ribosome and RNAP copy numbers, to demonstrate the degree of ribosome-RNAP spatial segregation, and to show that most 70S ribosomes diffuse freely within the ribosome-rich regions, presumably as a mixture of polysomes of different lengths (6).

Here we present a detailed study of RNAP diffusion in live *E. coli*, tracked by a β'-mEos2 fusion protein expressed from the chromosome. On a 1-s timescale, about half of the observed RNAP copies sub-diffuse very slowly, moving in the same fashion as DNA loci. We assign these as RNAP copies specifically bound to DNA, including transcription initiation, elongation, pausing, and termination. The other half of the observed copies are assigned to a rapidly equilibrating mixture of non-specifically bound copies and copies freely diffusing within the nucleoid regions. Longer 7-s trajectories obtained at

0.5s/frame reveal transitions between the slow state and the mixed state. Shorter 20-ms trajectories obtained at 2 ms/frame directly reveal unbound, freely diffusing RNAP copies.

A quantitative description of the partitioning of RNAP among its various roles in live *E. coli* provides an important test of global models of cellular processes (7-9). In particular, the concentration of free RNA polymerase, which has never before been measured directly, strongly affects the expression level of most "housekeeping" genes and of *rrn* operons. The new data enable us to partition RNAP activity into the fraction transcribing (f_{trx}, which includes initiation, elongation, pausing, and termination), the fraction of non-specifically bound to DNA (f_{ns}), the fraction able to bind DNA but freely diffusing in three dimensions (f_{free}), and a small fraction that evidently cannot bind to DNA on a 1-s timescale (f_{nb}). Our best estimates of these fractions disagree with earlier results based on mini-cell preparations (7,8).

Materials and Methods

Bacterial strains

Except as noted, all the strains are based on a VH1000 background strain. Strain details are given in Appendix 1. Most of the work used strain HC1, which we call RNAP-β'-mEos2. It has the gene for the photoswitchable protein mEos2 inserted into the chromosome at the C-terminus of *rpoC* (*rpoc::meos2*). Chromosomally expressed RNAP-β'-yGFP (RLG7470) contains a fusion of *yGFP* to the *rpoC* C-terminus (6). The DNA binding protein HU-mEos2, used for imaging the nucleoid, was expressed from a plasmid. For tracking specific DNA loci using the *parS*-ParB-GFP method, we used the Left-1 strain (RLG7455) and the NSL-2 strain (RLG7457). In both strains, ParB-GFP is expressed without any induction. These two strains are in an MG1655 background.

Doubling times of all the constructs and the background strain VH1000 in EZRDM at 37ºC were measured in a microplate reader (Tecan, Infinite M1000 PRO). Under these conditions, all doubling times lie in the narrow range 39 ± 2 min. The doubling time decreases to 30 min for cells grown at 37ºC in a hot-air shaker for the single-molecule studies.

Cell growth and sample preparation for microscopy

Cells were grown in EZ rich, defined medium ("EZRDM", 0.28 Osm) (10) overnight with shaking in a 37ºC hot air shaker. We subsequently made subcultures by diluting the stationary phase culture at least 1:100 into 2 mL of fresh medium. When cells had grown to midlog phase (OD_{600} = 0.4-0.6), culture was placed in Cover-Well Perfusion Chamber Gaskets on a polylysine-coated coverslip. The volume of the closed chamber is 140 μL. We allowed ~2 min for the cells to adhere to the coverslip, then replaced the liquid in the chamber with fresh, aerated medium to rinse away the non-adhered cells.

Except as noted, imaging by widefield epifluorescence or by single-molecule photoactivation was carried out at 25ºC within 10 min of plating. The photophysical properties of mEos2 are enhanced at 25ºC. After plating, cells continue to grow with a doubling time of 42 min, estimated by measuring cell length vs time using phase contrast microscopy (6). The longer doubling time under imaging conditions is likely due to the lower temperature and to incomplete oxygenation of cells in the sealed chamber. The average length of cells is ~4.8 μm, and most cells exhibit two primary nucleoid lobes, often segregated into two sub-lobes each.

For imaging of RNAP in fixed cells, 37% formaldehyde by weight was added to 1 mL of culture to a final concentration of 3.7%. The culture was kept at room temperature for 1 hour for fixation. Cells were then pelleted with a brief centrifugation and rinsed twice with 1X PBS and resuspended in 1 mL of 1X PBS prior to plating and imaging.

Single-molecule imaging

Cells were imaged using a Nikon Eclipse Ti inverted microscope equipped with an oil immersion objective (CFI Plan Apo Lambda DM 100X Oil, 1.45 NA) and a 1.5X tube lens. The focus was maintained with the Nikon perfect focus system. For time-lapse imaging, fast shutters (Uniblitz LS2, Vincent Associates, New York) were used to synchronize illumination and image acquisition. Images were recorded by a back-illuminated EMCCD camera with 256 x 256 pixels of 16 μm x 16 μm each (iXon DV-887, Andor Technology, Connecticut). Each pixel corresponds to 105×105 nm^2 at the sample (150X overall magnification). Most data were collected at a frame rate in the range 2–33 Hz, with exposure time within each frame of 15–30 ms. To image freely diffusing RNAP, some data was collected at a frame rate of 500 Hz using a fast EMCCD camera (ANDOR iXon 860, 128 x 128 pixels, 24 μm x 24 μm each).

RNAP-β'-mEos2 was photoactivated with a 405-nm diode laser (CrystaLaser, Reno, Nevada) and subsequently imaged with a 561 nm laser (Sapphire 561 CW lasers, Coherent). The 405 nm power density at the sample was 0.1-1.5 W/cm^2. The photoactivation laser remained on throughout imaging. Power density of the 561 nm laser was kept at approximately 1.5 kW/cm^2. In time-lapse mode (30 ms/frame, 100 ms/frame), the probe laser was pulsed; it was on continuously in the 2 ms/frame tracking experiments. The 561 nm laser itself causes only minor activation. At this power we achieved good signal-to-noise ratio for single molecules with minimal laser damage to the cell. To minimize the phototoxic effect of the laser we collected data for less than 6 min per cell. Cell growth slows significantly after ~9 min of 561 nm laser exposure. Emission was collected through a 617/73 bandpass filter (bright line 617/73-25, Semrock). The total number of photoactivatable RNAP-β'-mEos2 copies was ~700 per cell for cells grown at 37ºC and imaged at 25ºC. We estimate the mean total RNAP copy number to be at least ~6,000 per cell from RNAP-yGFP under these conditions (Appendix 4E). Only ~12% of the RNAP copies are observable at 25ºC. The mean trajectory length is 8-10 steps. The key

results were compared with data from cells grown and imaged at 37ºC to test for possible effects of imaging at the lower temperature (Fig. 4.A.1). In those experiments, only ~200 molecules per cell are fluorescent, and the mean trajectory length is only 5-7 steps.

Single-molecule tracking of RNAP-β'-YGFP was carried out using the reversible photobleaching method (5) as described earlier (6). A 514 nm Ar^+ laser (Melles Griot, Carlsbad, CA) was used for imaging. The yellow emission was collected using a 560/50 emission filter (Chroma). Superresolution imaging of HU-mEos2 was carried out with the same procedure used for RNAP-β'-mEos2. Images were acquired at a frequency of 12.4 Hz with 50 ms exposure time within each frame.

Fluorescently labeled (*parS*-ParB-GFP) chromosomal loci were tracked at 10 Hz with 20-ms exposure time per frame. Each set of cells was tracked for 100 frames before moving to a new set of cells. Data was collected on each sample for less than ten minutes. GFP was excited with a 488 nm Ar^+ laser (Melles Griot, Carlsbad, CA) and emission was collected with a 525/50 emission filter (ET 525/50M Chroma). The 488 nm power density at the sample was approximately 0.5 kW/cm^2.

Analysis of single-molecule tracking data

Single-molecule trajectories were obtained using a centroid algorithm to locate molecules, as described earlier (11,12). The population-averaged mean-square displacement vs lag time was calculated by pooling trajectories from multiple movies and from different sets of cells:

$$MSD_r(\tau) = \frac{1}{NM}\sum_i \sum_m \left[r_i(m\Delta t + \tau) - r_i(m\Delta t)\right]^2 \qquad (4.1)$$

Here x lies along the cell centerline (long axis), y lies along the transverse coordinate parallel to the camera plane, $r = (x^2 + y^2)^{1/2}$, $r_i(t)$ is the particle position in trajectory i at time t, τ is the lag time, N is the total number of trajectories, and Δt is the time between frames. The index m runs from 0 to the largest possible value M given the lag time and the trajectory length.

To test for heterogeneity of diffusion from molecule to molecule, we examined the distribution of nominal single-molecule diffusion constants (D_i) estimated from the time-averaged mean square displacement for each molecule (13):

$$D_i(\tau) = \frac{1}{4\tau M}\sum_m [r_i(m\Delta t + \tau) - r_i(m\Delta t)]^2 \quad (4.2)$$

We also find that scatter plots of the maximum excursion of a trajectory along x and y (ε_x vs ε_y) can be useful in dissecting heterogeneous diffusion data.

Monte Carlo Simulations

The effects of confinement on diffusion of RNAP are simulated as Monte Carlo random walks within a truncated cylinder representing one nucleoid sub-lobe. Simulations were performed in MATLAB, as described earlier (12). For comparison with experiment, we include the effects of the localization error σ by sampling from the appropriate two-dimensional Gaussian function. The one-dimensional localization error is estimated by extrapolating the $MSD_r(\tau)$ plot to $\tau = 0$ and equating the intercept with $4\sigma^2$ (14).

Results

Heterogeneous RNAP diffusion on 1-s timescale

First we present a detailed analysis of RNAP diffusion data combined from 128 cells grown in EZRDM at 37ºC and imaged at 25ºC. RNAP-β'-mEos2 was tracked at 10 Hz frame rate (100 ms/frame) with 30 ms laser exposure within each frame. These conditions provided good signal-to-noise and circularly symmetric single-molecule images. In Fig. 4.1A we show several typical 9-step RNAP trajectories taken at 100 ms/frame (10 localizations, 0.9 s total trajectory duration). Cursory inspection reveals two very different trajectory types. About half of the trajectories are compact ("slow" diffusion,

contained within an approximately 200 nm x 200 nm box), and about half are extended ("mixed-state" diffusion, as explained below; covering an area of about 1.5 μm x 0.7 μm, comparable to that of one nucleoid sub-lobe). Our working hypothesis, defended below, is that the compact (slow) trajectories are due to RNAP copies specifically bound to DNA and engaged in transcription initiation, actively transcribing, or pausing during elongation or during termination. The extended (mixed-state) trajectories are due to a rapidly interconverting mixture of two RNAP states: free copies diffusing in three dimensions within the nucleoids and copies non-specifically bound to DNA. For trajectory durations of ~1 s, we observe little evidence of "go-stop" or "stop-go" trajectories, i.e., RNAP copies that make a transition between the mixed state and the slow state. We also observe a small fraction (<10% as judged by eye) of fluorescent labels that move much faster than the mixed-state trajectories and are blurred and difficult to track at 10 Hz. Based on experience, their diffusion constant is ~1 μm^2-s^{-1}. They seem *not* to exchange with the non-specifically bound population. We call this non-binding fraction f_{nb}. These faster copies will be characterized accurately in movies taken at 2 ms/frame (500 Hz), described below.

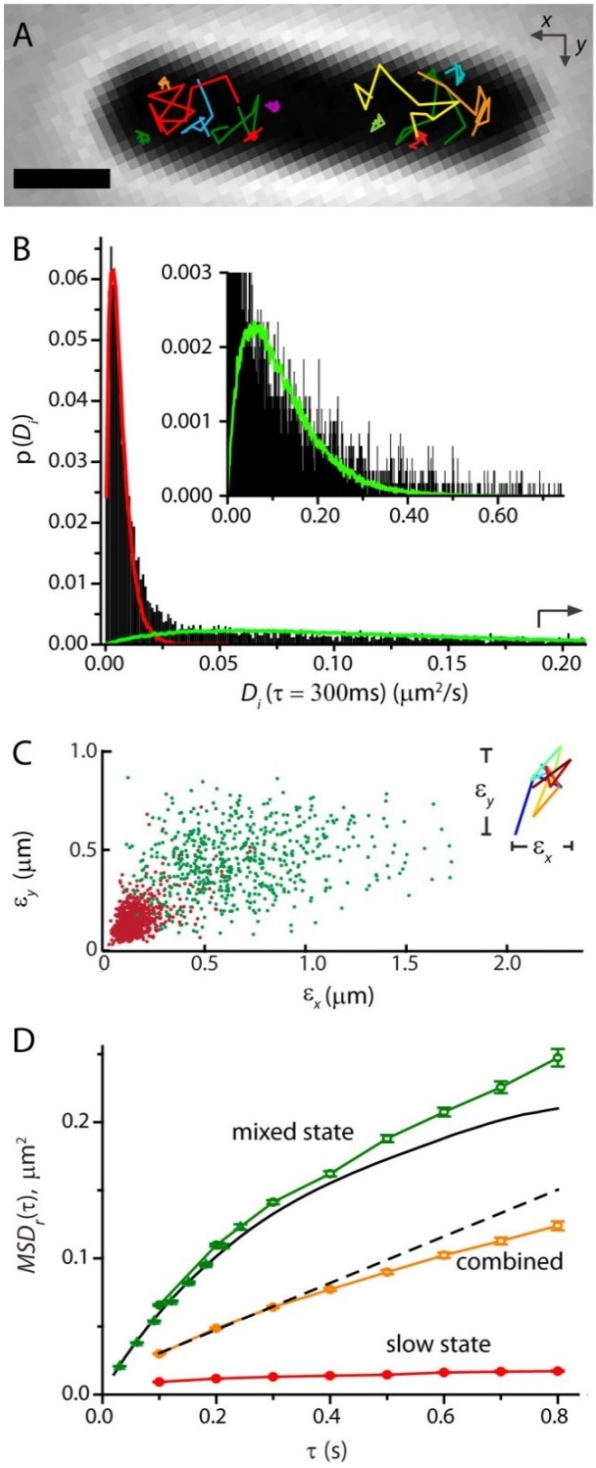

Figure 4.1: (**A**) Examples of RNAP trajectories from 10-Hz movies displayed atop the phase contrast image. Scale bar = 1 μm. (**B**) Distribution of single molecule diffusion constants D_i (Eq. 4.2) from 5996 trajectories truncated to 6 steps for the lag time $\tau = 300$ ms. The red and green lines are the two components of a two-state model that mimics the peak and the tail of the distribution closely. *Inset*: Expanded view of long tail of the distribution. The green line is the result of a model of free diffusion with $D_{mixed} = 0.21$ μm^2-s^{-1} including confinement within one nucleoid sub-lobe. (**C**) Scatter plot of maximum excursions ε_y vs ε_x for single trajectories (9-steps at 100 ms/frame) along the *x* (long axis) and *y* (short axis) directions. Red dots have $D_i < 0.03$ μm^2-s^{-1} and green dots have $D_i > 0.03$ μm^2-s^{-1}. The separation is quite clean. (**D**) Ensemble averaged $MSD_r(\tau)$ from trajectories truncated at 9 steps (Eq. 4.1). *Orange*: all RNAP copies; dashed line is linear fit to first three points, yielding $\langle D_{RNAP} \rangle = 0.03$ μm^2-s^{-1}. *Green:* "Mixed-state" RNAP population only (defined as $D_i > 0.03$ μm^2/s). Circles are data taken at 10 Hz; triangles at 33 Hz. The black solid line is from a Monte Carlo model for the mixed-state trajectories with $D_{mixed} = 0.21$ μm^2-s^{-1}, including confinement in one nucleoid sub-lobe. See text for details. *Red:* Slow RNAP population only (defined as $D_i < 0.03$ μm^2/s).

The distribution of estimated single-molecule diffusion constants D_i (Eq. 4.2) is shown in Fig. 4.1B (13). For 5996 trajectories of 6 steps or longer taken from 128 cells under identical conditions, we truncated the trajectories at 6 steps (seven locations, 0.6 s trajectory length) and formed a histogram of D_i for the lag time τ = 300 ms. The distribution is clearly heterogeneous. Each cell in the sample shows a comparable number of compact and extended trajectories. The heterogeneity in the experimental distribution of D_i is not primarily due to differences among cells (Supporting Material).

In Fig. 4.1D we plot the two-dimensional mean-square displacement $MSD_r(\tau)$ vs lag time (Eq. 4.1 below) averaged over 2665 trajectories truncated at 9 steps (10 localizations). The limiting slope at small τ corresponds to a mean diffusion constant $\langle D_{RNAP} \rangle$ = 0.03 μm^2-s^{-1}, but this number represents an average over a heterogeneous sample. The upper curve (green) shows $MSD_r(\tau)$ for only the "mixed-state" subset of trajectories, defined as those molecules with D_i > 0.03 μm^2-s^{-1}. This curve is a composite of 779 trajectories taken at 10 Hz with 30 ms exposure time (open circles in green) and 470 trajectories taken at 33 Hz with 15 ms exposure time (green triangles). The two data sets match well in the region of overlapping lag times (0.1-0.2 s). The mixed-state *MSD* curves can be understood in terms of an effective diffusion constant D_{mixed} = 0.21 μm^2-s^{-1} including the effects of confinement within the nucleoids, as explained below. The lower *MSD* plot in Fig. 4.1D is for the "slow" subset of trajectories, defined as D_i < 0.03 μm^2-s^{-1}. Its limiting slope corresponds to D_{slow} = 0.0032 μm^2-s^{-1} and is very similar to that of a DNA locus, as detailed below.

The homogeneous slow distribution that best fits the narrow peak in the D_i distribution has fractional amplitude 0.52. The mixed-state distribution that best fits the long tail on the D_i distribution has amplitude 0.40. The sum of these components is systematically too low at the "knee" of the total distribution, near D_i = 0.03 μm^2-s^{-1}. The extra amplitude at the knee comprises the remaining 8% of the detected molecules in 10 Hz movies. Examination of these specific trajectories shows that they are

typically mostly compact, with one or two long steps at the beginning or end of the trajectory. We attribute them to transitions between the slow and mixed states occurring on the 300-ms timescale.

Among all *successfully tracked* copies of RNAP in the 10 Hz movies, our best estimate of the fractions of slow and mixed-state copies becomes $F_{slow} = 0.58$ and $F_{mixed} = 0.42$. This is based on the 52% slow and 40% mixed-state amplitudes and dividing the 8% extra copies at the knee as 6% slow and 2% mixed-state. The unequal division is due to the nature of the extra copies at the knee of the overall distribution. The observed fractions F_{slow} and F_{mixed} will be corrected below for the <10% of very fast copies that cannot be tracked in 10 Hz movies.

For all the data thus far, cells were grown at 37ºC but imaged quickly at 25ºC to improve the photophysical properties of mEos2. As a control, we also obtained a set of data for cells grown at 37ºC and imaged at 37ºC. The distribution of single-molecule estimates of D_i is very similar (Fig. 4.A.1). To test whether the specific label mEos2 influences the observed diffusive behavior of RNAP (15), we also acquired analogous data for RNAP-β'-YGFP, expressed from the chromosome of cells grown at 37ºC and imaged at 25ºC using the photobleach/recovery method (5,6). As shown in Fig. 4.A.2, the YGFP-labeled D_i distribution is very similar to that of the mEos2-labeled RNAP.

Comparison of slow RNAP motion with fluorescently labeled chromosomal loci

To test the hypothesis that the slow RNAP fraction in 10 Hz movies remains bound to DNA on a time scale of ~1 s, we studied the diffusive motion of two specific DNA loci in the same growth conditions (EZRDM at 37ºC). Left-1 and NSL-2 (Fig. 4.2A) were tracked using the *parS*-ParB-GFP methodology (16). The ParB labeling scheme produces bright, diffraction-limited puncta that enable accurate tracking (σ ~ 30 nm) over 50 camera frames or more. For growth in EZRDM at 37ºC, each cell typically exhibits 2-4 bright puncta. We discarded movie segments in which two loci interfere with accurate tracking of each other. Movies at 10 Hz with 30 ms exposure time within each frame produced HUndreds of trajectories longer than 12 steps.

On a time scale of 1 s, all DNA loci trajectories are compact (contained within a 200 nm x 200 nm box, Fig. 4.2B). The ensemble-averaged $MSD_r(\tau)$ was calculated from 725 (for Left-1) and 968 (for NSL-2) trajectories longer than 13 localizations and truncated at 12 steps (Fig. 4.2C). Over the range $\tau = 0.1–1$ s, the MSD_r plots for both DNA loci curve mildly downward. The curvature indicates sub-diffusive movement of the labeled foci within the DNA polymer, as observed earlier (17). For comparison, in Fig. 4.2C we also plot $MSD_r(\tau)$ for the slow RNAP molecules, again defined as $D_i < 0.03$ μm^2-s^{-1}. This cutoff includes all of the molecules in the narrow, slow peak of Fig. 4.1B plus some faster copies at the knee of the curve. The *MSD* plot is constructed from 985 trajectories truncated to 12 steps each (10 Hz, 30 ms exposure time). To demonstrate that the slopes of the DNA and RNAP plots are due to real molecular motion rather than sample drift or localization error alone, we also show $MSD_r(\tau)$ for RNAP in fixed cells. The much larger slope of the *MSD* plots in live cells ensures that we are measuring real motion.

The *MSD* plots for the two DNA loci and for the slow RNAP population are similar in slope and curvature. The differences in the intercepts presumably arise from differences in measurement accuracy, with the brighter DNA foci localized more accurately. The initial slope of the *MSD* plots yields nominal short-time diffusion constants $D_{slow} \sim 0.0032$ μm^2-s^{-1} and $D_{DNA} \sim 0.0027$ μm^2-s^{-1} for NSL-2 and 0.0024 μm^2-s^{-1} for Left-1. These are only nominal diffusion constants; the motion of slow RNAP and of DNA loci is sub-diffusive. The slightly larger RNAP slope is due in part to the somewhat arbitrary cutoff $D_i <$ 0.03 μm^2-s^{-1} that includes some faster RNAP copies at the knee of the D_i distribution. In addition, the bulky size of the cluster of ParB labels probably renders their diffusion within the DNA polymer somewhat slower than that of a DNA locus alone (18).

In Fig. 4.A.8, we compare the distribution of single-molecule D_i estimates (Eq. 4.2) for the NSL-2 DNA locus and for the slow component of RNAP, both at a lag time $\tau = 300$ ms. The two distributions are similar, but as the *MSD* slopes would indicate, the RNAP distribution is somewhat broader. The mean

values of D_i (0.0055 μm^2-s^{-1} for NSL-2 and 0.007 μm^2-s^{-1} for RNAP) are 2–3 times larger than the estimates of the short-time diffusion constant from the limiting slope of $MSD_r(\tau)$ plots. The reason is that localization error makes a significant contribution to D_i estimates for compact motion on short timescales (Appendix 4J). This is readily seen in Fig. 4.2C, where the extrapolated intercepts contribute >50% of the apparent MSD_r for τ in the range 0.1-0.3 s. The limiting slope of an *MSD* plot avoids the effects of measurement error.

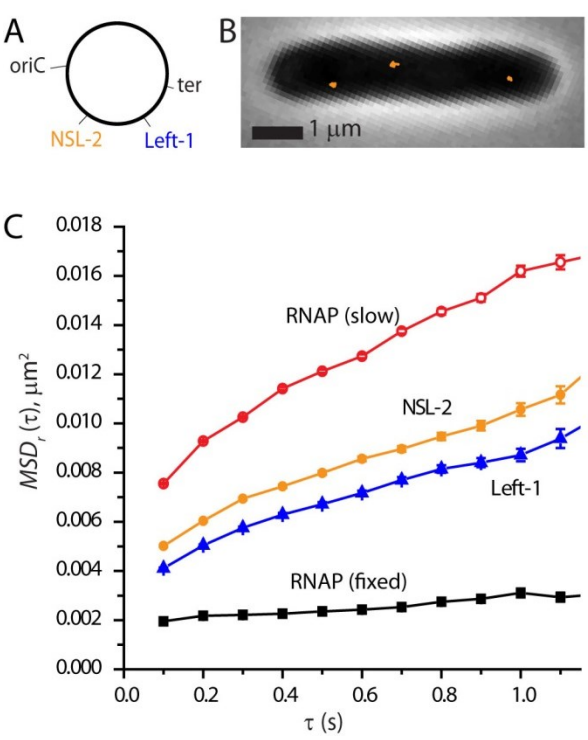

Figure 4.2: (**A**) Positions of the DNA loci NSL-2 and Left-1 on the chromosome. (**B**) Example 1-s long trajectories for Left-1 taken at 10 Hz and displayed atop the phase contrast image. (**C**) Ensemble-averaged $MSD_r(\tau)$ plots from 10 Hz trajectories truncated to 12 steps for the slow RNAP population (D_i < 0. 03 μm^2/s, 985 trajectories), for Left-1 (725 trajectories), and NSL2 (968 trajectories). MSD_r data for RNAP in cells fixed with formaldehyde is shown for comparison.

We have also compared the distribution of D_i for slow RNAPs with that of the two DNA foci for different combinations of trajectory lengths and lag times. The short-time motions of the two DNA loci match each other, and they match the slow component of RNAP reasonably well. We conclude that on the timescale of 0.1–1.0 s, the slowly diffusing RNAP population moves very much like a DNA locus.

Mixed-state motion modeled as a confined random walk

Assuming that HU labels the DNA fairly uniformly (19), we used the HU spatial distribution $P(x, y)$ to estimate appropriate model cylinder dimensions for one nucleoid sub-lobe (Appendix 4F). The measured HU distribution within each nucleoid sub-lobe is very similar to that of a uniformly filled cylinder of radius R = 390 nm and length L = 1.0-1.2 μm (Fig. 4.A.3). For the RNAP data taken at 33 Hz, the 300-ms long, 9-step trajectories almost always remain contained within one nucleoid sub-lobe, as shown in scatter plots of the maximal excursions along x and y (ε_x and ε_y) in Fig. 4.3A. For the 10 Hz data (Fig. 1C), the 0.9-s long, 9-step trajectories are usually confined within one nucleoid sub-lobe, but sometimes "leak out" to an adjacent sub-lobe (range of x > 1.2 μm).

We used the 33 Hz "mixed-state" data having D_i > 0.03 μm^2-s^{-1} to test whether a random walk model confined within one nucleoid lobe can explain the curvature of the *MSD* plot. Two-dimensional projections of random walks within a cylindrical volume exhibit a different degree of confinement along x (long axis), y (short, transverse axis), and r (total distance) (12). The most stringent test compares the model with *MSD* data for both x and y coordinates, $MSD_x(\tau)$ and $MSD_y(\tau)$. We carried out an extensive search of the L, R, and D_{mixed} parameter space (Appendix 4G) for the best-fit parameters for $MSD_x(\tau)$ and $MSD_y(\tau)$ in a least-squares sense. The results are: L = 1.10 ± 0.06 μm, R = 0.38 ± 0.02 μm, and D_{mixed} = 0.205 ± 0.005 μm^2-s^{-1}, with the uncertainties indicating 95% confidence intervals (Supporting Material and Fig. 4.A.4). The best estimates of L and R agree well with the estimated length and radius of one nucleoid sub-lobe from superresolution imaging of HU (Fig. 4.A.3). The same model fits $MSD_r(\tau)$ from the 30 ms/frame data (τ = 0.03–0.3 s), but at longer lag times of 0.5–0.9 s (Fig. 4.1.D), it progressively underestimates the 100 ms/frame data, as if the confining volume of the model is too small. This is presumably due to RNAP copies beginning to cross over and explore more than one nucleoid sub-lobe at longer time scales, as shown in the scatter plots of RNAP trajectory maximum spatial extents along x and y (Fig. 4.1.C). Finally, the same model diffusion constant D_{mixed} = 0.21 μm^2-s^{-1} with confinement effects

also fits the long tail on the single-molecule D_i distribution quite well (green curve in Fig. 4.1.B). We conclude that the *MSD* data for the mixed-state RNAP at 0.03–1 s can be well explained by Brownian diffusion (random walks) within the confined region defined by the nucleoid. Recourse to sub-diffusive mechanisms is not necessary.

It is important to realize that the confinement of mixed-state RNAP within the nucleoid lobes does not arise from any physical barrier. Instead, the density of DNA binding sites and the strength of non-specific binding combine to make it rare for RNAP copies to escape the nucleoid regions. Accordingly, earlier bulk FRAP measurements obtained an inter-lobe transit time on the order of 20 s (20). If the mixed-state RNAP motion were instead completely free diffusion without binding, a large fraction of RNAP would occupy the entire cytoplasm. In fact, there is very little RNAP in the ribosome-rich regions (6).

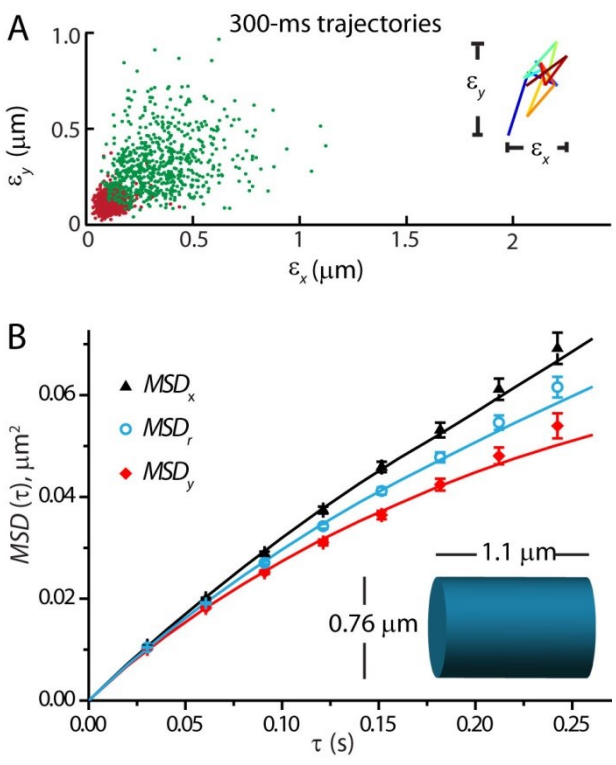

Figure 4.3: (**A**) For all detected RNAP copies, scatter plot of maximum excursion ε_y vs ε_x for 9-step trajectories taken at 33 Hz (300-ms trajectory length). Red dots denote slow RNAP trajectories (D_i < 0.03 μm^2-s^{-1}) and green dots denote mixed-state trajectories (D_i > 0.03 μm^2-s^{-1}). (**C**) *MSD*(τ) plots for the mixed-state RNAP molecules (D_e > 0.03μm^2/s) along the short axis *y*, the long axis *x* and for $r = (x^2 + y^2)^{1/2}$. $MSD_r(\tau)$ is divided by 2 to place it on same scale as *x* and *y*. Solid lines are simulated *MSD* plots from Monte Carlo model with D_{mixed} = 0.21 μm^2-s^{-1} and confinement within cylinder of R = 380 nm and L = 1.1 μm as shown.

Transitions between slow and mixed diffusion populations on 7-s timescale

If our assignment of the slowly diffusing population of RNAP to specifically bound copies and the rapidly diffusing RNAP to a mixed state of free/non-specifically bound copies is correct, it should be possible to observe transitions between these two states of motion on longer timescales. In a separate set of experiments, we slowed the frame rate to 2 Hz (0.5 s/frame with 30 ms exposure time) and collected trajectories of 14 steps or longer (7 s total duration). A sampling of these trajectories is shown in Fig. 4.4. Of 106 7-s trajectories studied across 22 cells, about 20% exhibit a clear transition from slow to mixed-state diffusion ("stop-go") or from mixed-state to slow diffusion ("go-stop"). The comparable numbers of transitions in either direction is consistent with a steady average population of RNAP copies in the slow and mixed states. In these 7-s long trajectories, we never observe clear evidence of *double* transitions ("stop-go-stop" or "go-stop-go"). In the Discussion, we show that the frequency of transitions on the 7-s timescale is a reasonable match for expectations based on known transcription rates.

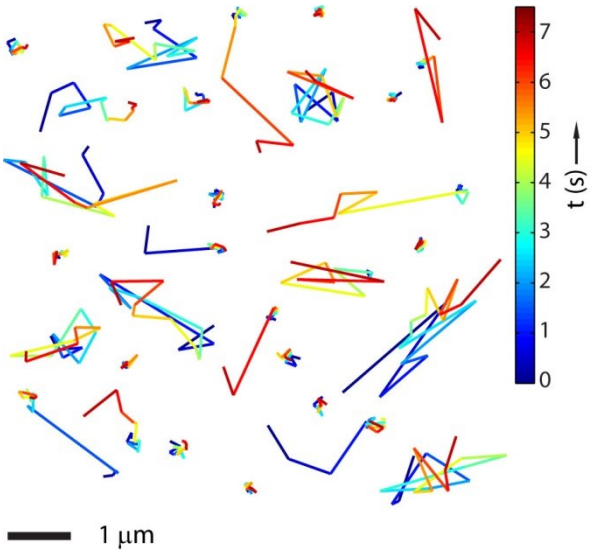

Figure 4.4: Examples of 14-step, 7-s long RNAP trajectories obtained at 2 Hz (0.5 s/frame) with advancing time color-coded. Note examples of "go-stop" and "stop-go" trajectories.

Direct observation of freely diffusing RNAP on 20-ms timescale

On the 0.1–1 s timescale, we visually observed <10% of RNAP copies moving very rapidly compared with the slow copies and the mixed-state copies. In order to accurately characterize the motion of this non-binding population and attempt to catch freely diffusing copies before they bind non-specifically, we imaged RNAP-β'-mEos2 at 500 Hz (2 ms/frame). Example trajectories, the apparent D_i distribution, the scatter plot of maximum excursions ε_y vs ε_x for each trajectory, and *MSD* plots are provided in Fig. 4.5.

Visual inspection of the trajectories hints at two types, compact and more extended. The D_i distribution is not well fit by a one-component model, but neither does it exhibit a clean separation of two components. We assume the two components present are a DNA-bound fraction (including *both* specific and non-specific binding on this short, 20-ms timescale) and an unbound, freely diffusing fraction, and that they do not exchange on such a short timescale. Our goal is to extract the fractions $F_{DNA\text{-}bound}$ and $F_{unbound}$ and the free RNAP diffusion constant $D_{unbound} = D_{free}$.

As described in Appendix 4H, the scatter plot of ε_y vs ε_x (Fig. 4.5C) proved much more useful than the D_i distribution in dissecting the two states. Our procedure for obtaining the diffusion constants $D_{DNA\text{-}bound}$ and $D_{unbound} = D_{free}$ and the fractions $F_{DNA\text{-}bound}$ and $F_{unbound}$ is complicated and thus relegated to Appendix. The results are: $F_{DNA\text{-}bound} = 0.82 \pm 0.03$, $F_{unbound} = 0.18 \pm 0.03$, $D_{DNA\text{-}bound} = 0.005 \pm 0.002$ μm^2-s^{-1}, and $D_{free} = 0.7 \pm 0.1$ μm^2-s^{-1}. In the partitioning analysis below, we will assume that $F_{DNA\text{-}bound} = f_{trx} + f_{ns}$ (the sum of specifically plus the non-specifically bound fractions) and that $F_{unbound} = f_{nb} + f_{free}$ (the sum of the fraction that cannot bind to DNA, which was the blurred fraction in the 100-ms movies, plus the freely diffusing RNAP that is exchanging with the non-specifically binding state to form the mixed fraction in the 100-ms movies).

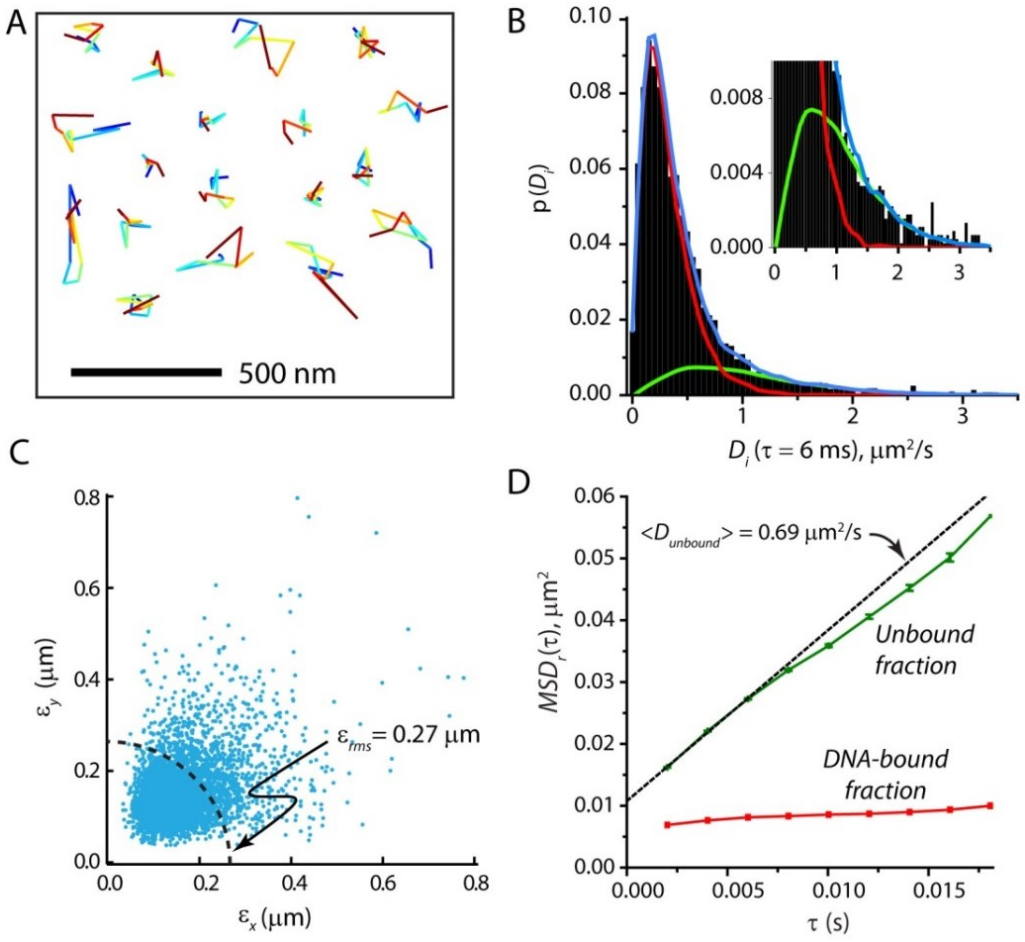

Figure 4.5: (**A**) Examples of 10-step, 20-ms long RNAP trajectories obtained at 500 Hz (2 ms/frame). Dark blue marks the beginning and dark red marks the end of each trajectory. (**B**) D_i distribution for 4551 trajectories obtained at 500 Hz; lag time is τ = 3 steps = 6 ms. Solid lines are model calculations for two non-interacting populations, one with $D_{DNA\text{-}bound}$ = 0.005 mm^2-s^{-1} and $F_{DNA\text{-}bound}$ = 0.81 (red curve) and the other with $D_{unbound} = D_{free}$ = 0.7 μm^2-s^{-1} and $F_{unbound}$ = 0.19 (green curve). Their sum (blue curve) fits the overall distribution well, indicating the model is adequate. (**C**) For 10-step trajectories obtained at 500 Hz, scatter plot of ε_y vs ε_x. Dashed circle denotes the value of ε_{rms} = 0.27, which most cleanly separates the DNA-bound from the unbound copies. See Supporting Material for details. (**D**) Ensemble-averaged $MSD_i(\tau)$ plots for the DNA-bound and unbound fractions from 10-step trajectories of 20-ms duration. The initial slope of the unbound *MSD* plot yields the best estimate D_{free} = 0.69 μm^2-s^{-1}.

Quantitative partitioning of RNAP

We use upper case F to denote measured fractions and lower-case f to denote the derived RNAP partitioning fractions of interest. To summarize, from the movies at 100 ms/frame, we obtained the fractions $F_{slow} = 0.58 \pm 0.04$ and $F_{mixed} = 0.42 \pm 0.04$ (expressed as fractions of the *successfully tracked* copies) and the diffusion constant $D_{mixed} = (0.21 \pm 0.02)$ μm^2-s^{-1}. From the faster movies at 2 ms/frame, we obtained the fractions $F_{DNA\text{-}bound} = 0.82 \pm 0.03$ and $F_{unbound} = 0.18 \pm 0.03$ and the diffusion constant $D_{unbound} = D_{free} = (0.7 \pm 0.1)$ μm^2-s^{-1}. The combination of all these results enables us to partition total RNAP into four steady-state fractional sub-populations: (1) f_{trx}, the fraction specifically bound to DNA (including transcription initiation, elongation, pausing, and transcription termination); (2) f_{ns}, the fraction non-specifically bound to DNA; (3) f_{free}, the fraction freely diffusing and exchanging with the non-specifically bound population; and (4) f_{nb}, the fraction freely diffusing but unable to bind to DNA.

We write the partitioning equations:

$$1 = f_{trx} + f_{ns} + f_{free} + f_{nb} \tag{4.3a}$$

$$F_{unbound} = f_{free} + f_{nb} \tag{4.3b}$$

$$F_{mixed}\,(1 - f_{nb}) = f_{ns} + f_{free} \tag{4.3c}$$

$$D_{mixed} = D_{free}\,[\,f_{free} / (f_{free} + f_{ns})]. \tag{4.3d}$$

Equation 4.3a is the normalization condition. Equation 4.3b from the fast movies at 2 ms/frame equates the unbound fraction with the two freely diffusing populations, those that can exchange with non-specifically bound copies and those that cannot bind to DNA. This assumes no exchange between the DNA-bound and unbound copies on the fast, 20-ms timescale, as justified by the good fit to the D_i distribution using two non-exchanging populations. Equation 4.3c equates F_{mixed} with the non-specifically bound plus free fractions. The factor $(1 - f_{nb})$ corrects the observed value of F_{mixed} for the very fast, untrackable copies in the 100-ms/frame movies. Finally, Eq. 4.3d describes D_{mixed} from the 100 ms/frame

movies as a population-weighted average of D_{free} and D_{DNA}. The non-specifically bound copies, which diffuse like D_{DNA}, make a negligible contribution to the observed mixed-state motion on the 1-s timescale.

The solution to these four equations in four unknown fractions is: $f_{trx} = 0.54 \pm 0.07$, $f_{ns} = 0.28 \pm 0.05$, $f_{free} = 0.12 \pm 0.03$, and $f_{nb} = 0.06 \pm 0.05$. The error estimates were obtained by propagating the estimated errors of the input parameters into the output parameters. They show that the solution set is not overly sensitive to small changes in the input parameters.

There remains the question of possible systematic errors. Our primary concern is the depth of focus of the 1.45 NA objective. De-focusing of RNAP copies that move near the top and bottom of the nucleoids lowers their peak intensity, causing the thresholding algorithm to discriminate against them and sometimes cut their trajectories into short segments. In the 100 ms/frame movies, "mixed-state" copies move further in a given period than "slow" copies and are thus more likely to go out of the detection range. In the 2 ms/frame movies, unbound copies move further than DNA-bound copies, but neither type moves very far.

In Appendix 4I, we estimate that the "depth of detection" is the range $z = \pm 340$ nm above and below the cell center axis. The HU distribution is well modeled as a uniformly filled cylinder of radius 390 nm. Thus depth-of-focus effects should be modest. If slow and mixed-state copies were similarly distributed throughout the nucleoids, we estimate that depth-of-focus effects would correct our partitioning estimates only moderately to $f_{trx} = 0.49$, $f_{ns} = 0.28$, $f_{free} = 0.12$, and $f_{nb} = 0.11$. However, we have preliminary evidence that in our growth conditions, the slow, transcribing RNAP copies tend to distribute towards the nucleoid *periphery*. In contrast, the mixed-state copies are distributed more centrally, more like HU. The effect is difficult to quantify from two-dimensional images, but it is clear that such a difference in radial distribution will work in the opposite direction from the depth-of-focus effect and push the best estimates back towards the uncorrected values above. Hence we prefer the

original estimates: $f_{trx} = 0.54 \pm 0.07, f_{ns} = 0.28 \pm 0.05, f_{free} = 0.12 \pm 0.03$, and $f_{nb} = 0.06 \pm 0.05$. We include both the corrected and uncorrected values in the bar graph of Fig. 4.6.

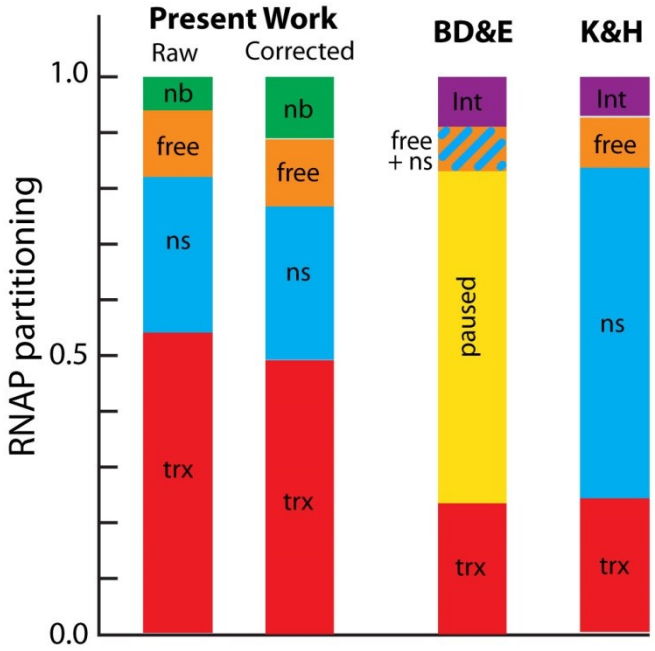

Figure 4.6: Comparison of estimates of RNAP partitioning fractions from the present work with earlier results from Bremer, Dennis, and Ehrenberg (BD&E, Ref. (8)) and from Klumpp and Hwa (K&H, Ref. (7)). We show our raw, uncorrected fractions and also our fractions corrected for depth of detection effects. The labels "nb", "free", "ns", and "trx" denote RNAP fractions that do not bind to DNA, that are freely diffusing within the nucleoid, that are binding non-specifically to DNA, and that are transcribing, respectively. "Int" denotes assembly intermediates. BD&E interpreted their minicell fractions in terms of what we call the sum of free plus non-specific binding. They postulated a "paused" state that stores a large number of non-elongating, specifically bound copies. K&H re-interpreted those copies as non-specifically bound.

Discussion

What species are we tracking?

The maturation time of mEos2 *in vitro* is ~175 min (21). This is much longer than the VH1000 doubling time of 30 min in EZRDM at 37ºC. Based on a simple exponential growth model (Supporting Material in Ref. (6)), we expect the fraction of mature mEos2 to be ~0.2 in steady state. This is similar to the estimated fraction of RNAP copies detected in our experiment: 700 detectable β'-mEos2 per cell out of a total of ~6000 RNAPs per cell is 0.12. The estimated time required for incorporation of the β' subunit into the RNAP core enzyme *in vivo* is ~2.5 min (22,23), much shorter than the maturation time of mEos2. Therefore, essentially all the fluorescent β'-mEos2 copies should be incorporated into the RNAP core enzyme. Any mEos2 copies that might have been cleaved from β' would diffuse much too fast (~10 μm^2-s^{-1}) to be detected in the 10 Hz imaging experiments. At 500 Hz, they would be detected but would exhibit spatial extents ε_x and ε_y much larger than observed. The β'-mEos2 fluorophores localized and tracked here should be almost entirely incorporated into RNAP core enzyme or holoenzyme.

On the time scale 0.1–1 s, the slow RNAP copies move like DNA loci (Fig. 4.2). We assign them to RNAP specifically bound to DNA, either engaged in transcription initiation, actively transcribing, or pausing during elongation or termination steps. The more mobile, mixed-state RNAP copies are assigned to a rapidly equilibrating mixture of non-specifically bound RNAP plus RNAP freely diffusing in 3D within the nucleoid, i.e., copies that are searching for a promoter sequence on DNA. The good fit of the fast population to a homogeneous model supports the assumption of rapid equilibration on the 300-ms timescale.

Essentially all holoenzymes in exponential growth conditions utilize σ^{70} (24). Our assignments are supported by a recent direct measurement *in vitro* of mean binding times of σ^{70}-RNAP to the non-promoter regions of λ DNA: $\tau_1 \sim 30$ ms for non-specific binding, $\tau_2 \sim 3.5$ s for binding to promoter DNA

to form closed complexes that dissociate, and $\tau_3 \sim 6000$ s for copies that proceed to form open complexes (but were prevented from elongation) (25). The value of τ_1 is consistent with our assumptions that ns-binding and free copies exchange rapidly on the 1-s timescale of the 100 ms/frame movies and that the DNA-bound and the unbound fractions do not exchange on the 20-ms timescale of the 2 ms/frame movies. However, the *in vitro* value of τ_1 was obtained in somewhat low salt conditions that lengthen the non-specific binding times. In more normal salt, the *in vitro* non-specific binding time would be shorter. The effects of crowding and the presence of other co-factors in the live cell are difficult to assess.

Additional support for the assignments comes from the direct observation of transitions between the slow state and the mixed state (Fig. 4.4). About 20% of the 7-s long RNAP trajectories (Fig. 4) show clear evidence of transitions, and we observe comparable numbers of "go-stop" and "stop-go" examples. The seven *rrn* operons comprise ~5500 nt each and are transcribed at ~100 nt/s, yielding a mean total transcription time τ_{rrn} ~55 s (26). The average length of an mRNA transcript is ~900 nt (27), and genes are transcribed more slowly, at a mean speed of ~40 nt/s (28), probably due to transcriptional pausing. The mean total time of transcription of a gene is thus ~20 s, but there is a broad distribution of gene lengths. The single-RNAP tracking experiment should sample the mixture of transcribing complexes uniformly at all stages of progress. If searching and transcribing copy numbers are in quasi-steady state, transitions from searching to transcribing should occur about as frequently as the reverse. Therefore, during a 7-s long trajectory, we would expect roughly 10% of RNAP copies transcribing an *rrn* operon to change from slow to fast and roughly 30% of copies transcribing a protein gene to change from slow to fast. At a doubling time of 30-40 min, we expect one-half to two-thirds of the transcribing RNAP copies to be transcribing *rrn* operons, with the remainder transcribing genes (7). The sensible agreement of the observed ~10% overall frequency of "stop-go" transitions with these estimates thus supports our assignment of the slow RNAP state to transcribing copies.

As a cautionary note, a recent *in vitro* study of σ^{54}-RNAP non-specific binding to non-promoter DNA found a lifetime of ~4 s under physiological salt conditions (29). Non-specific binding of σ^{54}-RNAP to DNA is apparently stronger than that of σ^{70}-RNAP. If the σ^{70}-RNAP non-specific binding time *in vivo* were as long as 4s, then our slow RNAP population on the 1-s timescale would include both specifically and non-specifically bound copies. That would force us to assign the faster RNAP population to freely diffusing (unbound) copies, with $D_{unbound} = 0.2\ \mu m^2\text{-}s^{-1}$. However, in the 2 ms/frame movies we directly observe "*very fast*" copies that diffuse with $D_{free} = 0.7\ \mu m^2\text{-}s^{-1}$. That is our preferred estimate for the diffusion constant of unbound RNAP within the nucleoids. It is sensibly consistent with variation of diffusion constant with protein size in the *E. coli* cytoplasm (30).

Finally, we speculate that the "non-binding" RNAP copies included as f_{nb} may be σ^{70}-RNAP that is bound to the small, non-coding 6S RNA (31). 6S RNA is known to abrogate transcription initiation, but it is not certain whether or not 6S-σ^{70}-RNAP can bind non-specifically to DNA. The non-binding fraction could also include misfolded RNAP copies.

Comparison with earlier study of RNAP diffusion

In earlier work we used fluorescence recovery after photobleaching (FRAP) to study the single-cell, micron-scale diffusive motion of RNAP labeled with β'-GFP in live *E. coli* (20). As in the present study, the chromosomal DNA exhibited two primary lobes. Labeled RNAP within one nucleoid lobe (half the cell) were photobleached at $t = 0$. A weak probe laser monitored the recovery of the axial profile of intensity as bleached and unbleached copies inter-diffused. Averaged over many cells, ~53% of the RNAP were mobile on the time scale of 10–20 s with an estimated diffusion constant of $\langle D_{RNAP} \rangle = 0.22 \pm 0.16\ \mu m^2\text{-}s^{-1}$. The remaining 47% of the population did not recover even on a 30-s timescale. We tentatively assigned the immobile fraction to RNAP copies specifically bound to DNA. The mobile fraction was assigned to jumping and sliding (non-specifically bound) copies.

The results of the present single-molecule tracking study of sub-micron motion are in basic agreement with the earlier work. In the FRAP study, recovery required movement of RNAP copies from one nucleoid lobe to the other. In the present study, single particle trajectories are sensitive to motion on the 0.05 μm length scale, much smaller than the size of a single nucleoid lobe. The effective diffusion constant of the mixed-state population $D_{mixed} = 0.21\ \mu m^2\text{-}s^{-1}$ is very similar to the longer-range diffusion constant gleaned from the earlier FRAP study (20). This suggests that the slow step in the inter-lobe transfer observed by FRAP is the migration of RNAP copies to the nucleoid periphery, where they can efficiently cross the central gap between the two primary nucleoid lobes. That gap is a ribosome-rich region (6) that evidently presents relatively little impedance to RNAP diffusion.

Partitioning of RNA polymerase activity

Estimation of the free RNAP concentration and the percentage of free RNAP in the *E. coli* cytoplasm has been highly indirect thus far. In 2001, Shepherd, Dennis and Bremer (SD&B) measured total amounts of β and β' relative to total protein in minicells and in unperturbed normal cells (32). They estimated that for a doubling time of 24 min, the combination of free β, free β', core and holoenzyme in the non-nucleoid part of the cytoplasm comprised 17% of total RNAP. The underlying assumption is that the unincorporated β and β', core enzyme, holoenzymes, *and all proteins* equilibrate their concentrations between the parent cell and the daughter minicell. They inferred that most of the β, β' detected in the minicells was in the form of free, unincorporated subunits and that only about 2% of total core and holoenzyme was present as free RNAP. A subsequent, more comprehensive modeling effort by Bremer, Dennis, and Ehrenberg (8) revised this estimate to 8% free RNAP at 24-min doubling time (5% at 60-min doubling time). Importantly, both SD&B and BD&E include the sum of our free and non-specific binding fractions in their definition of free RNAP. Our result $f_{free} + f_{ns} = 0.40 \pm 0.06$ is much larger than their estimates of 2–8%. Note that the long maturation time of mEos2 implies that our fractions exclude free, unincorporated β' and assembly intermediates.

Improved understanding of the spatial distribution of RNAP suggests a likely reason for this discrepancy. Our earlier single-molecule study of ribosomes and RNAP showed very strong segregation between ribosomes and DNA and close similarity between the RNAP and DNA spatial distributions (6). We estimated that at most 4% of total RNAP resides outside the nucleoids in the two ribosome-rich endcaps. We double this to estimate that less than 8% of total RNAP resides in all the ribosome-rich regions combined (two endcaps plus the central region between nucleoid lobes). In view of the new partitioning data, many of these RNAPs are likely the non-binding copies detected here for the first time. For those RNAP copies that *can* bind to DNA, the strength of DNA-RNAP non-specific binding and the sheer density of available ns-binding sites within the nucleoid region evidently cause the chromosomal DNA to tightly sequester RNAP within the nucleoids.

Turning to the minicells, the simplest model assumes that a minicell provides an additional volume of about 10% of the normal cell volume, but devoid of DNA. In terms of RNAP spatial partitioning, the minicell volume would effectively add about 20% to the volume of the ribosome-rich regions. This small additional volume will not greatly affect the ability of the chromosomal DNA to sequester the fractions f_{free}, f_{ns}, and f_{trx} within the nucleoids. We infer that SD&B/BD&M's 2–8% estimate is experimentally sensible, but it should be compared with our non-binding fraction $f_{nb} = 0.06 \pm 0.05$, not with our $f_{free} + f_{ns} = 0.40 \pm 0.06$.

Unfortunately, comprehensive models of RNAP partitioning by BD&M and by Klumpp and Hwa (here "K&H") relied heavily on the original interpretation of the minicell result (7). Their derived fractional populations are compared with our best estimates in the bar graph of Fig. 4.6. For example, the K&H model at 30-min doubling time found 24% of total RNAP actively transcribing, including initiation and pausing (16% rRNA and 8% mRNA); 60% non-specifically bound to DNA; 9% freely diffusing; and 8% "assembly intermediates". The minicell result constrained the sum of free and assembly intermediate fractions to be only 17%, which had a strong impact on all the other fractions. We also note that our

assignment of F_{mixed} as the sum of $f_{ns} + f_{free}$ is inconsistent with SD&B/BD&M's model, which includes a large fraction of RNAP sequestered in non-productive, "paused" states. It is consistent with the partitioning concept of K&H, which includes a large fraction of non-specifically bound RNAP.

In future work, we plan to extend these methods to *E. coli* cells to different growth conditions. The result will be a comprehensive picture of how RNAP partitions its activities in fast, medium, and slow growth.

Appendix 4A: Construction of strains

The strain HC1used for imaging chromosomally expressed β'-mEos2 contains a fusion of *meos2* to the C-terminus of *rpoC*. This strain was constructed by lambda-Red mediated recombination. The mEos2 used here is optimized for the *E. coli* codon set. A cassette containing optimized mEos2 and a kanamycin resistance gene was amplified by PCR with the 5'-CAGAACTGCTGAACGCAGGTCTGGGCGGTTCTGATAACGAGCTAGAAATAATGT CTGCTATTAAACCGGA-3' and 5'-CCCCCCATAAAAAAACCCGCCGAAGCGG GTTTTTACGTTATTTGCGGATTATGTAGGCTGGAGCTG-3'. The underlined portions in the first and second primers anneal to the C-terminus of *rpoC* just before the stop codon and the region immediately following the stop codon of *rpoC,* respectively. The PCR product was electroporated into DY330 as previously described (1) . A kanamycin-resistant colony was selected and verified to express mEos2. The *rpoC-mEos2* was then moved to a clean VH1000 (MG1655 *pyrE lacl lacZ*) background by P1 transduction, resulting in the strain HC1. The strain was then sequenced to verify the correct linker (CTAGAAATA) using two different sequencing primers, 5'-GGT ACC GGT TAC GCG TAC CAC CAG G-3' and 5'-GGC TCA GGA TTT CAA TAC AGT GAT CCA CG-3'.

We imaged two DNA loci, Left-1 and NSL-2, using the *parS*-ParB-GFP methodology. The plasmid pALA2705 P1 ParB-GFP (RLG7493) (2) has ampicillin resistance. It was transformed into the Left-1, and NSL-2 strains using TSS protocol (3). It was expressed in cells without being induced as was carried out by Nielsen, H. J., *et al*. (4).

To image the nucleoid, we have used HU-mEos2 expressed from a plasmid. The gene HU*pA* (HU) was amplified from MG1655 using the primers (primers, 5' - ATCGGATCCCGCCGCCGCCTTAACTGCGTC -3' and 5'- GATCGAATTCTA GGAGGTATTCACATGAACAAGACTC -3') .The amplified HU*pA* gene with restriction sites for EcoRI

and BamHI was inserted into pASK IBA3 plus (IBA). Then, codon optimized *meos2* was amplified using primers (primers, 5'- GATCGGATCCATGTCTGCTATTAAACCGG -3' and 5'- GATCAAGCTTTTAGCGACGAGCATTATCCG -3'). The amplified *meos2* with restriction sites for BamHI and HindIII was inserted into the HU*pA* vector with a linker of three alanines (GCGx3). This vector was transformed into VH1000 using the TSS protocol and selected against ampicillin resistance.

Appendix 4B: Estimation of mean number of RNAPs per cell

A lower bound on the number of RNAP copies per cell was determined by scaling the total intensity of all the fluorescent copies of YGFP-labeled RNAPs in a cell to the intensity of a single YGFP molecule, as described in our earlier work (5). This was carried out for cells grown at 37ºC and imaged at 25ºC, as described in the main text. The mean number of RNAPs per cell is at least 6,300 (39 cells). We take that as a lower bound due to the presence of misfolded and immature copies.

Appendix 4C: RNAP partitioning: test for cell-to-cell heterogeneity

The distribution of single-molecule diffusion constants D_i, constructed with data combined from 128 cells, is heterogeneous (Fig. 4.1). To test the possibility that the heterogeneity is due to differences among cells, we selected the 45 cells that yielded at least 80 trajectories. When more than 80 trajectories were obtained, 80 of the trajectories were selected randomly. Trajectories with $D_i > 0.03$ μm^2-s^{-1} were assigned as mixed-state and trajectories with $D_i < 0.03$ μm^2-s^{-1} were assigned as slow. Then we calculated the fraction of slow trajectories for each cell based on this criterion. The average fraction of slow trajectories per set was 0.54 ± 0.07. This is very similar to the fraction of slow RNAPs found from the composite distribution.

To test what part of the standard deviation can be attributed to finite sample size, we assume a homogeneous group of cells for which the true fraction of slow molecules is 0.54. Randomly sampled outcomes would then follow a binomial probability distribution with $p = 0.54$. The mean of 80 such measurements with $n = 80$ is $\langle np \rangle = 43.2$. The variance of outcomes from 80 such measurements (number of slow molecules in a sample of 80 RNAPs) will therefore be $80 \cdot p \cdot (1-p) = 19.872$. The standard deviation of the fraction estimate (proportion of slow molecules in a sample of 80 RNAPs) will be $p \cdot (1-p)/80 = 0.056$.This is only slightly smaller than the observed result of $\sigma = 0.07$, indicating that most of the 80-trajectory, cell-to-cell deviation in the slow fraction is due to sampling statistics, not cell-to-cell heterogeneity.

Appendix 4D: RNAP diffusion data for cells grown and imaged at 37ºC

For the strain expressing β'-mEos2, the primary data was obtained for cells grown at 37ºC but imaged within the first ten minutes at 25ºC to improve the photophysical properties of mEos2. We tested for possible effects of the temperature change by obtaining a more limited data set for cells grown and imaged at 37ºC. Both $MSD_r(\tau)$ and D_i distributions are very similar regardless of imaging temperature (Fig. 4.A.1 and Fig. 4.1.B).

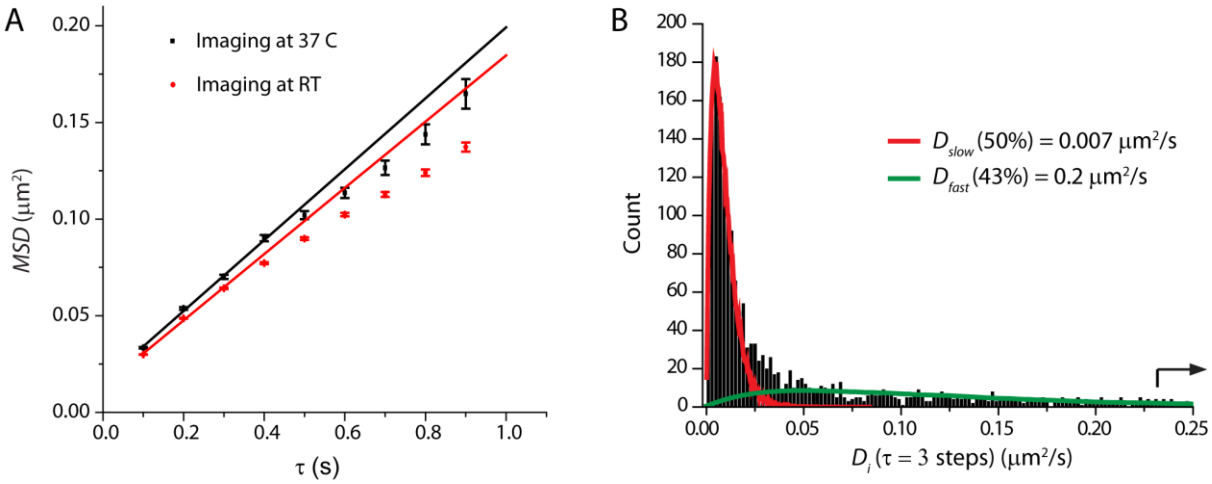

Figure 4.A.1. (**A**) $MSD_r(\tau)$ plots for all RNAP copies combined for cells expressing β'-mEos2. Cells were grown at 37ºC and imaged at either 37ºC or room temperature (RT, 25ºC) as indicated. Movies at 100 ms/frame. Slope of the *MSD* curves are very similar (~0.03 $\mu m^2/s$). (**B**) Distribution of the single-molecule diffusion constants D_i for cells grown at 37 ºC and imaged at 37ºC. The fractions of fast and slow copies are very similar to those obtained with imaging at room temperature (Fig. 1B).

Appendix 4E: RNAP diffusion data for cells expressing β'-yGFP

We tested for possible effects of the mEos2 label by studying the RNAP diffusive behavior for cells expressing β'-yGFP. Cells were grown at 37ºC and imaged at 37ºC in both cases. The $MSD_r(\tau)$ and D_i distributions are very similar regardless of label (Fig. 4.A.2).

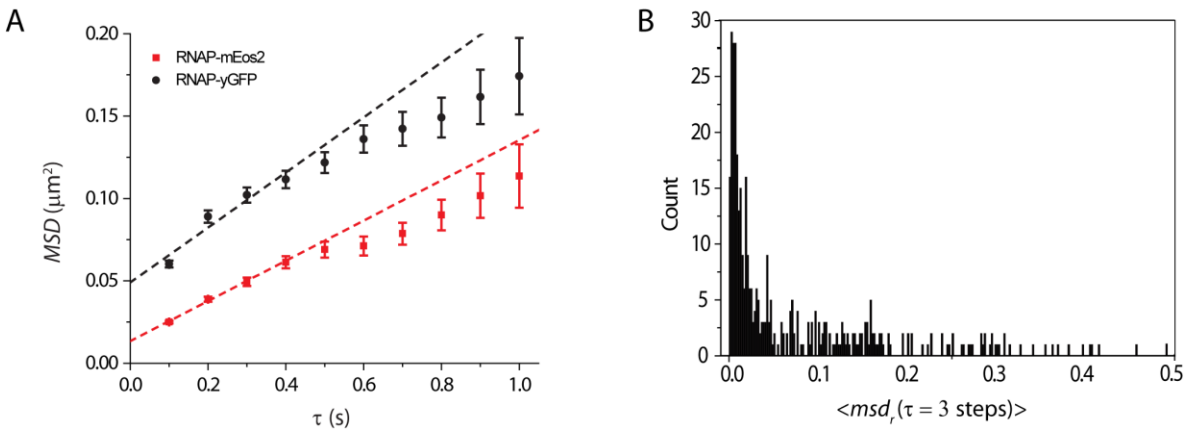

Figure 4.A.2. (**A**) Ensemble-averaged mean squared displacement vs lag time compared for tracking of RNAP-β-mEos2 and RNAP-β'-YGFP at 100 ms/frame. Measurement accuracy (τ = 0 intercept) differs but limiting slopes and curvature are similar. (**B**) Distribution of estimated single-molecule diffusion constants D_i for RNAP-β'-YGFP. The result is quite similar to that for RNAP-β'-mEos2 (Fig. 4.1C).

Appendix 4F: HU-mEos2 spatial distribution and estimating nucleoid dimensions

We use HU-mEos2 expressed from a plasmid as a proxy for total DNA density. A typical superresolution image shows the expected nucleoid structure of two primary lobes, each partially divided into sub-lobes. The projected distribution of locations along the short cell axis, $P(y)$, is well fit by a model of a uniformly filled spherocylinder of radius $R = 390 \pm 20$ nm (Fig. 4.A.3B). This radius matches the nucleoid radius that best fits the mixed-state diffusion data, as described below. For this cell, the projected distribution along the long cell axis, $P(x)$, has full width at half maximum height of ~2.30 μm for each nucleoid lobe. Half this value yields an estimate of the length of one nucleoid sub-lobe as $L = 1.15$ μm. A very similar model length best fits the mixed-state diffusion data taken at 33 Hz.

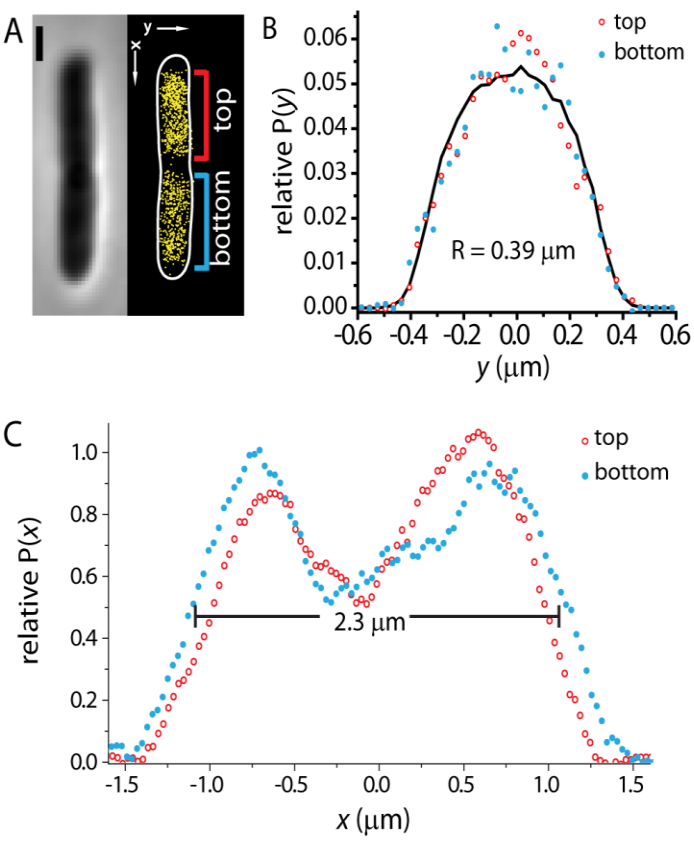

Figure 4.A.3. (**A**) Phase contrast image (left) and the superresolution image of HU-mEos2 distribution (right). (**B**) Distribution of the distances of the HU-mEos2 positions from the cell centerline. Red circles are for top nucleoid lobe and blue circles are for bottom lobe. Black line is a two-dimensional model distribution for uniformly distributed particles in a spherocylinder of length $L = 1.2$ μm and radius $R = 390$ nm, including localization error $\sigma = 40$ nm. (**C**) The distribution of HU-mEos2 locations along the long axis of the cell (red circles for the top lobe and blue circles for the bottom lobe).

Appendix 4G: Global fit of *MSD(τ)* of the mixed-state population to free diffusion with confinement

Monte Carlo (MC) simulations of the 33 Hz diffusion data for the mixed-state population (defined as $D_i > 0.03$ μm^2-s^{-1}, Fig. 4.1) were carried out in truncated cylinder geometries to extract the best model parameters (D_{mixed}, length of the cylinder L, and radius of the cylinder R). We conducted a grid-based search of the parameters to best fit the $MSD_x(\tau)$ and $MSD_y(\tau)$ over the range τ = 0.03-0.18 s in a least-squares sense. MSD_r was excluded as it leads to coupling between the parameters L and R. For every combination of the three parameters, we ran 770 MC trajectories (the same as in the experimental sample) from randomly selected positions within the cylinder. The confined random-walk simulation is described in earlier work (6) . Minimization of chi-square was used to determine the best-fit parameters. The contours of constant chi-square for different combinations of these three parameters are shown in Fig. 4.A.4. For fitting 6 data points on each *MSD* curve (MSD_x and MSD_y) with three parameters (degrees of freedom = 6 x 2 - 3 = 9), the 95% confidence interval lies within the chi-square contour 16.9. Therefore we use all combinations of parameters that result in chi-square less than 16.9 to define the 95% confidence interval as the range of values of each parameter within these combinations. The results are: $D_{mixed} = 0.205 \pm 0.005$ μm^2-s^{-1}, $L = 1.10 \pm 0.06$ μm, and $R = 0.38 \pm 0.02$ μm. The radius is a good match for fits to the transverse distribution of HU-mEos2 copies (0.39 μm).

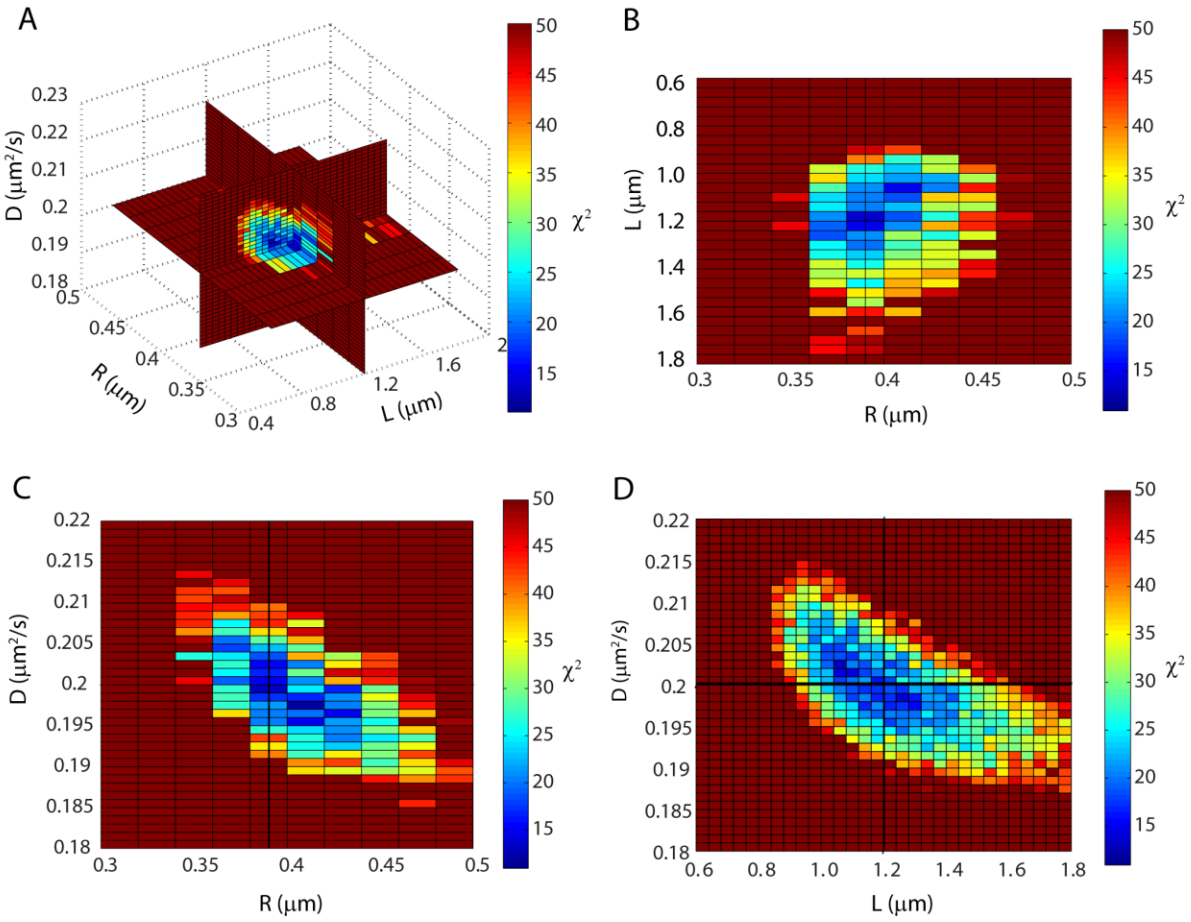

Figure 4.A.4. Contour plots of chi-squared vs values of D_{mixed}, nucleoid radius R, and nucleoid sub-lobe length L for global fits to MSD_x and MSD_y from 33 Hz data. (**A**) Three-dimensional plot. (**B, C, D**) Contours of chi-square for all three pairwise combinations of the three parameters.

Appendix 4H: Analysis of 2-ms/frame movies: $F_{DNA\text{-}bound}$ and $F_{unbound}$

We assume that the 10-step, 2 ms/frame trajectories arise from two non-exchanging populations (DNA bound and unbound). The D_i distribution (Fig. 4.5B) does not fit a single- population model, but neither does it separate cleanly into two populations. This is because the localization error in the 2 ms/frame movies is large. We found that the scatter plot of the maximum excursion of each trajectory (ε_y vs ε_x) is able to resolve the two fractions better (Fig. 4C). The trajectories of DNA-bound RNAPs remain compact because they are dominated by localization error, and the error vectors tend to cancel. In comparison, there is more real motion in the trajectories of unbound RNAPs, leading to significantly larger spatial extents.

To separate the two populations, we first obtain a good estimate of the diffusion constant of the unbound fraction from the long tail of the D_i distribution (Fig. 4.5B). The result is $D_{unbound} = D_{free} = 0.7 \pm 0.1$ μm^2-s^{-1}. Left-1 was tracked at 498 Hz in frame-transfer mode. We obtained 400 frames of data for each cell. The ensemble-averaged $MSD_x(\tau)$ was calculated from 132 trajectories longer than 49 localization and truncated at 50 steps (Fig. 4.A.5). The apparent diffusion constant estimated from the initial slope of the *MSD* plot is $D_{DNA} = 0.0047$ μm^2-s^{-1}. This value on a 20-ms timescale is about twice as large as the value 0.0024 μm^2-s^{-1} on a 1-s timescale, as estimated from 100 ms/frame movies. For sub-diffusion of a DNA locus within the DNA polymer network, an increase in the slope of the *MSD* plot is expected as the lag time decreases.

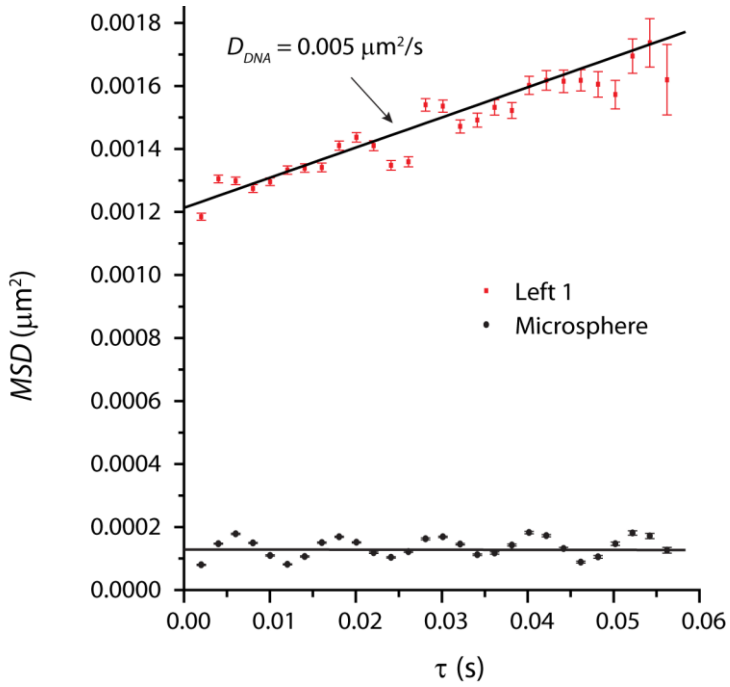

Figure 4.A.5. Ensemble averaged $MSD_x(\tau)$ plots from 123 trajectories taken at 500 MHz and truncated to 30 steps for Left-1 and for fluorescent microsphere beads adhered to glass. The oscillations are due to instrument vibration. The microsphere data shows that the slope of Left-1 MSD is due to real diffusive motion. The different intercepts are due to greater localization error for the Left-1 loci compared with the microspheres. We fit the first 20 points to estimate the short-time diffusion constant of DNA.

Next we used the two diffusion coefficients to simulate model 20-ms trajectories including localization error, estimated from the intercept of the *MSD* plot as $\sigma = 45$ nm. We generate 1000 simulated trajectories for each component and make an ε_y vs ε_x scatter plot for each (Fig. 4.A.6). We define the root-mean-square excursion of a trajectory as $\boldsymbol{\varepsilon_{rms}} = \sqrt{\boldsymbol{\varepsilon}_x^2 + \boldsymbol{\varepsilon}_y^2}$. Our goal is to determine the value of ε_{rms} which best discriminates DNA-bound from unbound trajectories. For the model trajectories, the upper bound $\varepsilon_{rms} < 0.27$ μm captures 90% of the DNA-bound trajectories (missing only 10%) and the lower bound $\varepsilon_{rms} > 0.27$ μm captures ~89% of the unbound trajectories (missing only 11%). Applying this same cutoff to the actual 2 ms/frame data (Fig. 4C), 25% of all the trajectories were classified as unbound RNAPs ($\varepsilon_{rms} > 0.27$ μm) and 75% were classified as DNA-bound RNAPs ($\varepsilon_{rms} > 0.27$ μm). The limiting slope of the $MSD_r(\tau)$ plot for the unbound trajectories (Fig. 4.5B) yielded $D_{unbound} = 0.69$ μm^2-s^{-1}, in good agreement with the original estimate from the long tail of the D_i distribution.

Finally, for a roughly 75:25 mixture of DNA-bound to unbound copies, the cutoff in ε_{rms} will incorrectly assign more DNA-bound copies than unbound copies. The final, best-estimate fractions $F_{DNA\text{-}bound} = 0.82 \pm 0.03$ and $F_{unbound} = 0.18 \pm 0.03$ properly correct the original estimate so that the same number of errors are made in both directions. The $MSD_r(\tau)$ plot for the unbound trajectories (Fig. 4.5B) has intercept of 0.011 μm^2. Setting this equal to $4\sigma^2$ yields an estimated localization accuracy $\sigma \sim 50$ nm for the rapidly moving copies. The $MSD_r(\tau)$ plot for the DNA-bound trajectories (Fig. 4B) has intercept of 0.0065 μm^2, corresponding to $\sigma \sim 40$ nm for the nearly stationary copies. We use these fractions and localization error estimates to simulate the D_i distribution (Fig. 4.5B). The result is in excellent agreement with experiment. We also model the distribution of ε_{rms} values (Fig. 4.A.6D) and again the fit is good. The quality of these fits indicates that the model of two non-exchanging populations (DNA-bound and unbound) is adequate to describe the data.

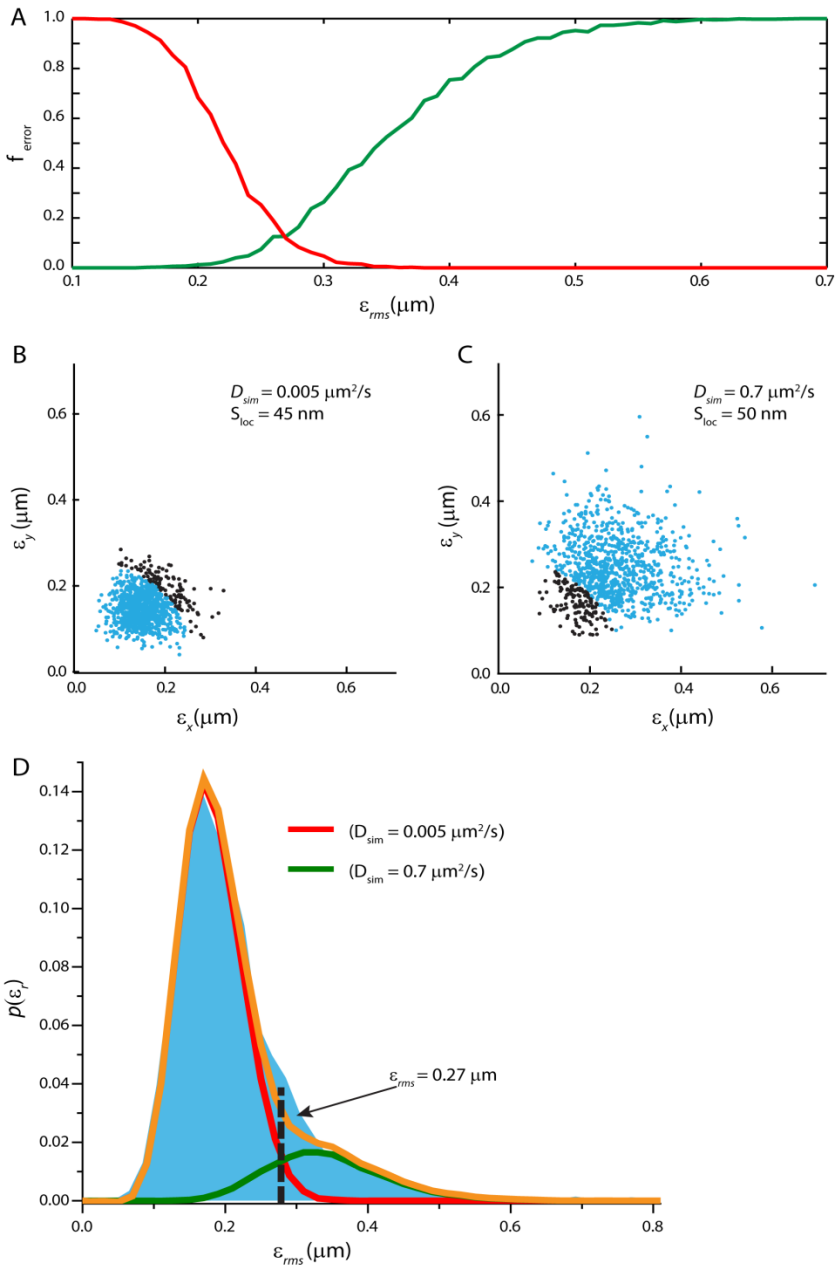

Figure 4.A.6. (A) fraction of DNA bound and unbound trajectories missing at different choice of the ε_{rms} cutoff. (**B**) ε_x vs ε_y plot for 1000 10-step trajectories with D = 0.005 μm^2-s^{-1} and 45 nm localization accuracy. Black dots are trajectories that are beyond the cutoff and represent the missing fraction. (**C**) ε_x vs ε_y plot for 1000 10-step trajectories with D = 0.7 μm^2-s^{-1} and 50 nm localization error. (**D**) Distribution of ε_{rms} from experiment in blue. Simulated excursion distribution from 10,000 trajectories (82% D = 0.005 μm^2-s^{-1} (in red) and 18% D = 0.7 μm^2-s^{-1} (green). The composite is plotted in orange and mimics the data from the experiment well.

Appendix 4I: Effects of limited depth of focus of the 1.45 NA objective

The limited depth of focus of the imaging system combined with the intensity threshold imposed by the tracking algorithm limits the range of detectability of a single molecule as it moves up and down in space (along the z coordinate, with $z = 0$ defined as the cell centerline). The peak intensity of the fluorescent spot decreases as the molecule moves along the microscope axis towards larger values of $|z|$ (Fig. 4.A.7A). If the intensity goes below the intensity threshold of the particle localization algorithm, the particle is not detected. The depth of focus decreases with increasing numerical aperture (NA) of the objective. In earlier work tracking rapidly moving single Kaede tetramers ($D_{Kaede} = 7.1\ \mu m^2\text{-}s^{-1}$), we estimated the range of single-molecule detection using a 1.49 NA objective to be ~600 nm (±300 nm) (5) . The present work tracking slower-moving RNAP copies using a 1.45 NA objective should have a somewhat larger "depth of detection" Δ.

Our goal is to estimate the effective depth of detection for RNAP under our imaging conditions. The typical timescale for which we can track a particle in the absence of photobleaching is determined by the diffusion constant D, the depth of field d, and the signal-to-noise ratio of the image, which interacts with the threshold set by the detection algorithm. The timescale should vary as d^2/D (7) . In a system consisting of particles with different diffusion constants, the faster particles will tend to move out of the depth of focus faster, resulting in truncated (shorter) trajectories. Selection of a given trajectory length for analysis thus imposes a bias towards detection of slower particles. In the 100 ms/frame movies, the slow RNAP are transcribing copies; in the 2 ms/frame movies, the slow RNAP are bound to DNA.

First we examined the experimental trajectory length distributions of slower and faster trajectories from the 10 Hz tracking data. For a total of 9515 trajectories longer than 5 steps, we determined the trajectory length distribution of 5788 trajectories with $D_i < 0.03\ \mu m^2 s^{-1}$ (red curve in Fig. 4.A.7B) and

3727 trajectories with $D_i > 0.03$ μm^{2}s^{-1} (green curve). As expected, shorter trajectories are slightly more likely for the more rapidly diffusing population.

We built a simple model for the depth of detection by assuming a sharp cutoff in detectability at $z = \pm\Delta$. The model assumes that for $|z| < \Delta$, molecules are detected with 100% efficiency and for $|z| > \Delta$, molecules are detected with 0% efficiency. We ran simulated random walk trajectories with random initial particle positions within a cylinder whose radius is 380 nm (taken from the measured HU distribution) for transcribing and mixed-state fractions in the 100 ms/frame movies and for the DNA-bound fraction in the 2 ms/frame movies. The radius was chosen as 450 nm for the non-binding fraction (based on the radius of the entire cytoplasm, as estimated from Kaede spatial distribution as before). We truncated trajectories based on the assumed sharp cutoff in detectability at $z = \pm\Delta$. For each value of Δ and for each chosen trajectory length, we thus estimated the fraction of particles whose trajectories will be rejected because they moved outside $|z| < \Delta$ before reaching the chosen trajectory length.

For a 7-step trajectory length taken at 100 ms/frame, the effects of different values of Δ on the fraction of molecules "missing" due to the cutoff is shown in Fig. 4.A.7C. The mixed state diffuses further and therefore a larger fraction of those copies is lost compared with the transcribing RNAPs. We carried out similar random walk simulations for the 2-ms movies and estimated the fraction of missed molecules for each type, DNA-bound vs unbound, for a 10-step trajectory length. Using these corrected values for the input fractions $F_{unbound}$ and F_{mixed} in Eq. 4.3, we solved for the corrected fractions f_{trx}, f_{ns}, f_{free}, and f_{nb} for assumed values of Δ ranging from 250 nm to 400 nm (Fig. 4.A.7D). As the depth of detection decreases, the main effect is a transfer of fractional population from f_{trx} to f_{ns}.

Finally, in order to estimate the actual depth of detection of our system, we compared the measured relative frequencies of trajectories of different lengths (Fig. 4.A.7B) with the simulation results. The relative frequencies of 5-step and 7-step trajectories from the data matched the simulated results well

for Δ in the range of 340 ± 25 nm. This is sensibly larger than the estimate of Δ = 300 nm in the previous study of Kaede using the 1.49 NA objective. The RNAP fractions for Δ = 340 nm are: $f_{trx} = 0.49$, $f_{ns} = 0.28$, $f_{free} = 0.12$, and $f_{nb} = 0.11$. These are compared with the uncorrected fractions in Fig. 4.6.

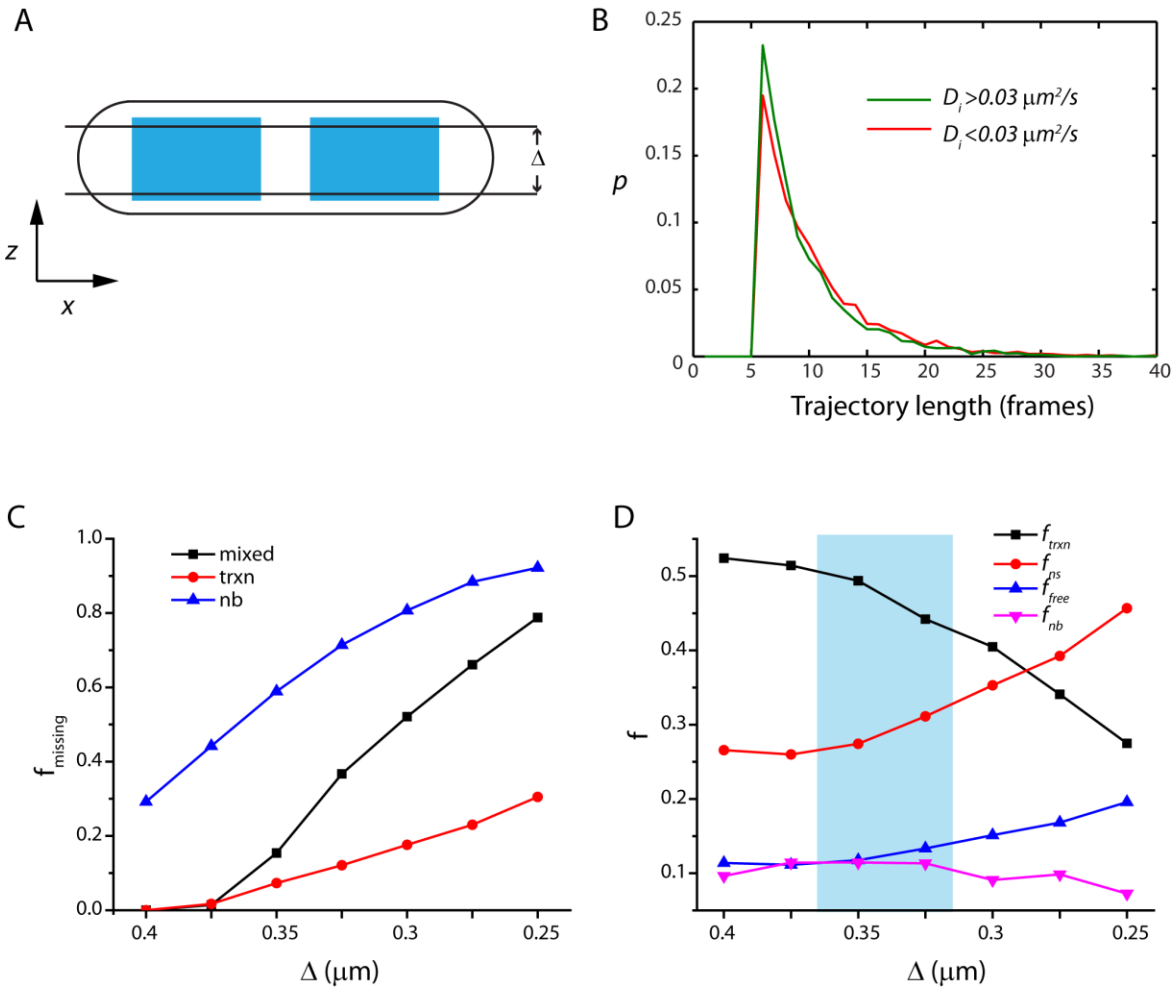

Figure 4.A.7. (**A**) Schematic of the depth of detection. Rectangles represent the 2-D projections of the nucleoid lobes. (**B**) Experimental distribution of trajectory lengths from 100 ms/frame movies for the mixed-state trajectories (green, $D_i > 0.03$ μm^{2}s^{-1}) and slow trajectories (red, $D_i < 0.03$ μm^{2}s^{-1}). (**C**) Fractions of 10-step trajectories that would be missed for different choices of the depth of detection Δ. (**D**) Effect of different depths of detection on the derived fractional partitioning of RNAPs states. The blue swath shows the range 340 ± 25 nm, which is our best estimate for the true depth of detection in our system.

Appendix 4J: Contribution of localization error to D_i distribution for compact trajectories

For compact trajectories, the localization error σ can make a significant contribution to the measured distribution of estimates of single-molecule diffusion coefficients D_i. To determine the magnitude of this effect for the slow RNAP fraction in the 100 ms/frame, 1-s long trajectories, we ran random-walk simulations with D_{DNA} = 0.003 μm^2-s^{-1}. That is the limiting slope of the $MSD_r(\tau)$ plot for τ in the range 0.1–0.3 s (Fig. 2C). Localization error was added to each simulated particle location by sampling from a two-dimensional Gaussian error distribution characterized by different standard deviations σ. The results are shown in Fig. 4.A.8A. For σ = 30 nm, the mean value of the apparent D_i distribution becomes 0.006 μm^2-s^{-1}, twice the true value of 0.003 μm^2-s^{-1}.

For comparison, the actual experimental D_i distributions for the DNA locus NSL-2 and for the slow fraction of the RNAP copies (having D_i < 0.03 μm^2-s^{-1}) are juxtaposed in Fig. 4.A.8B. The data are taken from 7-step trajectories at 100 ms/frame with τ = 3 steps = 300 ms, as in Fig. 4.2. The extrapolated $MSD_r(\tau)$ plots give estimated localization errors of σ = 30 nm for the DNA locus and σ = 40 nm for the slow RNAP copies. The D_i distributions are similar. The larger breadth of the RNAP distribution is due primarily to the larger localization error, but also to inclusion of some of the faster-diffusing RNAP population at the knee of the D_i distribution.

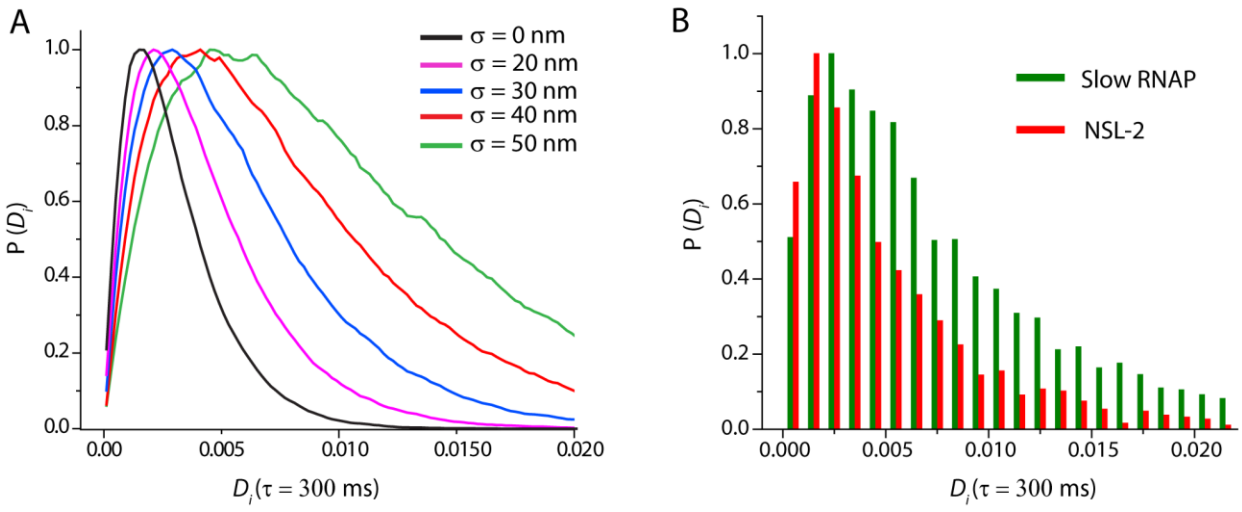

Figure 4.A.8. (**A**) Simulated D_i distributions for a true value $D = 0.003$ μm^2/s with effects of localization error of different magnitude as shown. In each case, 10^5 trajectories with 7 steps at 100 ms/frame were simulated. Localization errors σ = 0-50 nm added as indicated and D_i was calculated for a lag time of τ = 3 steps = 300 ms as in Fig. 4.2. (**B**) Experimental data for RNAP and for the DNA locus NSL-2 under the same conditions.

References

1. Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642-1645.

2. Hess, S. T., T. P. K. Girirajan, and M. D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophysical Journal 91:4258-4272.

3. Rust, M. J., M. Bates, and X. W. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Meth. 3:793-795.

4. Cattoni, D. I., J. B. Fiche, and M. Noellmann. 2012. Single-molecule super-resolution imaging in bacteria. Current Opinion in Microbiology 15:758-763.

5. Deich, J., E. M. Judd, H. H. McAdams, and W. E. Moerner. 2004. Visualization of the movement of single histidine kinase molecules in live caulobacter cells. Proc. Natl. Acad. Sci. U. S. A. 101:15921-15926.

6. Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Molecular Microbiology 85:21-38.

7. Klumpp, S., and T. Hwa. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proceedings of the National Academy of Sciences of the United States of America 105:20245-20250.

8. Bremer, H., P. Dennis, and M. Ehrenberg. 2003. Free RNA polymerase and modeling global transcription in *Escherichia coli*. Biochimie 85:597-609.

9. Tadmor, A. D., and T. Tlusty. 2008. A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rrna operon copy number. PLoS computational biology 4:e1000038.

10. Neidhardt. F. C., P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. Journal of Bacteriology 119:736-747.

11. Wang, T., C. Ingram, and J. C. Weisshaar. 2010. Model lipid bilayer with facile diffusion of lipids and integral membrane proteins. Langmuir 26:11157-11164.

12. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of Kaede diffusion and spatial distribution in live *Escherichia coli*. Biophysical Journal 101:2535-2544.

13. Vrljic, M., S. Y. Nishimura, S. Brasselet, W. E. Moerner, and H. M. McConnell. 2002. Translational diffusion of individual class II MHC membrane proteins in cells. Biophysical Journal 83:2681-2692.

14. Wieser, S., M. Moertelmaier, E. Fuertbauer, H. Stockinger, and G. J. Schuetz. 2007. (Un)confined diffusion of CD59 in the plasma membrane determined by high-resolution single molecule microscopy. Biophysical Journal 92:3719-3728.

15. Landgraf, D., B. Okumus, P. Chien, T. A. Baker, and J. Paulsson. 2012. Segregation of molecules at cell division reveals native protein localization. Nature Methods 9:480-U498.

16. Espeli, O., R. Mercier, and F. Boccard. 2008. DNA dynamics vary according to macrodomain topography in the e-*coli* chromosome. Molecular Microbiology 68:1418-1427.

17. Weber, S. C., A. J. Spakowitz, and J. A. Theriot. 2010. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. Physical Review Letters 104.

18. Javer, A., Z. Long, E. Nugent, M. Grisi, K. Siriwatwetchakul, K. D. Dorfman, P. Cicuta, and M. Cosentino Lagomarsino. 2013. Short-time movement of *E. coli* chromosomal loci depends on coordinate and subcellular localization. Nature communications 4:3003.

19. Wang, W. Q., G. W. Li, C. Y. Chen, X. S. Xie, and X. W. Zhuang. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. Science 333:1445-1449.

20. Bratton, B. P., R. A. Mooney, and J. C. Weisshaar. 2011. Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. Journal of Bacteriology 193:5138-5146.

21. McKinney, S. A., C. S. Murphy, K. L. Hazelwood, M. W. Davidson, and L. L. Looger. 2009. A bright and photostable photoconvertible fluorescent protein. Nature Methods 6:131-133.

22. Ito, K., Y. Iwakura, and A. Ishihama. 1975. Biosynthesis of rna-polymerase in *Escherichia-coli* 3. Identification of intermediates in assembly of rna-polymerase. Journal of Molecular Biology 96:257-271.

23. Iwakura, Y., and A. Ishihama. 1975. Biosynthesis of rna-polymerase in *Escherichia-coli* 2. Control of rna-polymerase synthesis during nutritional shift up and down. Mol. Gen. Genet. 142:67-84.

24. Maeda, H., N. Fujita, and A. Ishihama. 2000. Competition among seven *Escherichia coli* sigma subunits: Relative binding affinities to the core RNA polymerase. Nucleic Acids Research 28:3497-3503.

25. Wang, F., S. Redding, I. J. Finkelstein, J. Gorman, D. R. Reichman, and E. C. Greene. 2013. The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion. Nature Structural & Molecular Biology 20:174-181.

26. Dennis, P. P., M. Ehrenberg, D. Fange, and H. Bremer. 2009. Varying rate of RNA chain elongation during *rrn* transcription in *Escherichia coli*. Journal of Bacteriology 191:3740-3746.

27. Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. Global analysis of mrna decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proceedings of the National Academy of Sciences of the United States of America 99:9697-9702.

28. Proshkin, S., A. R. Rahmouni, A. Mironov, and E. Nudler. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science 328:504-508.

29. Friedman, L. J., J. P. Mumm, and J. Gelles. 2013. RNA polymerase approaches its promoter without long-range sliding along DNA. Proceedings of the National Academy of Sciences of the United States of America 110:9740-9745.

30. Kumar, M., M. S. Mommer, and V. Sourjik. 2010. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. Biophysical Journal 98:552-559.

31. Wassarman, K. M., and G. Storz. 2000. 6S RNA regulates *E-coli* RNA polymerase activity. Cell 101:613-623.

32. Shepherd, N., P. Dennis, and H. Bremer. 2001. Cytoplasmic RNA polymerase in *Escherichia coli*. Journal of Bacteriology 183:2527-2534.

33. Yu, D., J. A. Sawitzke, H. Ellis, and D. L. Court. 2003. Recombineering with overlapping single-stranded DNA oligonucleotides: Testing a recombination intermediate. Proc Natl Acad Sci U S A 100:7207-7212.

34. Li, Y. F., K. Sergueev, and S. Austin. 2002. The segregation of the *Escherichia coli* origin and terminus of replication. Molecular Microbiology 46:985-995.

35. Chung, C. T., S. L. Niemela, and R. H. Miller. 1989. One-step preparation of competent *Escherichia-coli* - transformation and storage of bacterial-cells in the same solution. Proceedings of the National Academy of Sciences of the United States of America 86:2172-2175.

36. Nielsen, H. J., J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen. 2006. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. Molecular Microbiology 62:331-338.

37. Goulian, M., and S. M. Simon. 2000. Tracking single proteins within cells. Biophys J 79:2188-2198.

Chapter 5

Superresolution Imaging of the Spatial Biology of Transcription in Live *Escherichia coli* Cells

Introduction

Adaptive bacteria need to selectively and efficiently transfer the genetic information stored in the chromosomal DNA. Spatio-temporal organization of transcription plays an important role in controlling the transfer of this genetic information (1). In rapidly growing bacteria, more than half of all transcription process result in stable RNA (rRNA and tRNA) (2, 3). This enables synthesis of ribosomes at a sufficient rate to maintain rapid growth. In the same conditions, single-cell, widefield fluorescence imaging of labeled RNA polymerase (RNAP) in both *E. coli* and *B. subtilis* reveals several bright puncta on a broader distribution of "background" RNAP (4, 5). These puncta, dubbed "transcription foci", very likely arise from clusters of several *rrn* operons heavily populated by RNAP copies involved in transcription of *rrn* operons (4). The underlying cause of such clustering is unknown, but it suggests the existence of "ribosome factories" where bacterial ribosome components may be synthesized and assembled. Such locations may be analogous to the nucleolus of eukaryotic cells. There are also hints of specific regions for transcription of specific types of genes. In recent work on *E. coli*, two genes placed within the chromosome and coding for membrane proteins were observed to migrate towards the cell membrane after induction, while genes coding for two cytoplasmic proteins were more uniformly distributed throughout the nucleoids (6).

In chapter 3, we have used YFP based photobleaching/recovery method to localize and count ribosomes and RNAPs inside living *E. coli* cells (7). Combining the spatial distribution information and

the total copy number of each protein, we are able to generate concentration profiles for ribosomes and RNAPs across the *E. coli* cytoplasm. Ribosomes were found to be strongly segregated from the nucleoid. RNAPs were exclusively localized on the nucleoid (7). However, this study of RNAP lacked the information regarding their biochemical states.

In a growing *E. coli* cell, RNAP can be freely diffusing in the cytoplasm, nonspecifically bound to DNA, or transcribing a gene. Actively transcribing RNAPs are expected to have lower mobility than RNAPs that are exchanging between freely diffusing and non-specifically bound state. In chapter 4, we have used photoactivation based single molecule tracking of RNAP (labeled by a β'-mEos2) to estimate the diffusion constant of each of these states and are able to partition the RNAPs accordingly. About half of the RNAP copies sub-diffuse very slowly, moving in the same fashion as DNA loci. We assigned these as RNAP copies specifically bound to DNA, including transcription initiation, elongation, pausing, and termination.

The diffusivity of the slow (transcribing) and fast (exchanging between nonspecifically bound states and the freely diffusing state) RNAP copies is so different that we can distinguish them in a single camera frame using diffusion blur. Using long exposure time we are able specifically locate the transcribing RNAPs. In this chapter we attempt to examine the spatial distribution of these transcribing RNAPs. The spatial distributions of transcribing RNAP, *rrnD*, and transcription foci enable us to infer that in rapid growth conditions, transcription of *rrn* operons occurs preferentially at the periphery of the nucleoids. We suggest that it may be functionally important to synthesize long pre-rRNA transcripts in the region of space near the membrane, where they can be processed and assembled efficiently. Transcription of genes encoding proteins is evidently more uniformly distributed throughout the nucleoids. However, we do find a significant number of slowly diffusing ribosomes near the nucleoid periphery, which suggests that co-transcriptional translation might be happening preferentially at the interface of the ribosome-rich regions and DNA rich nucleoid (7). It is possible that genes spend a

fraction of transcription time being translated and during this time they move towards the periphery. The existence of a small population of transcribing RNAP and of some DNA at or very near the cytoplasmic membrane lends support to the transertion hypothesis of co-transcriptional translation of membrane proteins (8, 9). In accord with this, we also find a significant population of slowly diffusing ribosomes near the inner membrane.

Materials and Methods

Bacterial strains

RNAP-β'-mEos2 (HC1) and RNAP-β'-yGFP (RLG7470) were used to image RNAP. The construction of these two strains is described in chapter 4 (Appendix 4.A). To image the nucleoid, we have used HU-mEos2 expressed from plasmid (construct described in chapter 4).

The *rrnD* (RLG10679) and *rrnG* (RLG11967) were constructed using the method described in Nielsen, H. J., J. R. Ottesen, et al. (2006). The plasmid pALA2705 P1 ParB-GFP (RLG7493) (Li, Y., K. Sergueev, et al. (2002)) has ampicillin resistance. It was transformed into the *rrnD*, and *rrnG* strains using TSS protocol (Chung, C.T. *et al*., 1989). It was expressed in cells without being induced as done in Nielsen, H. J., Y. Li, et al. (2006). Both *rrn* opeon strains are constructed by Tamas Gaal from R. L. Gourse laboratory and used with their permission.

For the ribosome studies, the rpsB::mEos2 strain (MDG 196) was constructed in the same way as the *rpsB::ygfp* strain as described elsewhere (7). The same codon-optimized mEos2 was used for both the RNAP and the ribosome strain. The strain was transferred into VH1000 background.

All the strains are in VH1000 background. Growth rates in EZRDM were similar for wild type and fusion constructs (39 ±2 min) (Appendix 5.A).

To measure the cytoplasmic radius of the strain under this growth condition, we used the plasmid strain JCW11 (described in chapter 2) that expresses Kaede when induced by tetracycline. The construct was transferred to the VH1000 background using TSS transformation protocol.

Cell growth and sample preparation for microscopy:

For single molecule imaging experiments, cells are grown in EZRDM at 37ºC. Average length of cells is ~4.8 μm. To check if different constructs have similar cell radius, we measured the radius of the strains using the outline constructed from the phase contrast image. In appendix 5C we describe the procedure for generating the sub-pixel outline of the cell from its phase contrast image. Appendix 5D describes the procedure of determining the cell radius from this sub-pixel outline of the cell. The measurements of radius are very precise with only 7 nm standard deviation. The radius of the cell determined from the outline was compared with the cytoplasmic radius determined from Kaede study on the wild type cell. The cytoplasmic radius determined from Kaede study was consistently smaller by 30 nm from the cell radius determined from phase contrast image (Fig. 5.A.4). Therefore we used the radius determined from the phase contrast images to compare the radius of different strains. Average cell radii of all the strains are comparable to the background strain VH1000 (475-483 nm: Fig. 5.A.4). Therefore we use the cytoplasmic radius of VH1000 (R = 445nm), determined from the spatial distribution of Kaede, as the cytoplasmic radius for all the strains.

For imaging the transcription foci, the cells are grown in LB at 37 ºC. The average cell radius determined from phase contrast image is slightly higher in this condition (490 nm). The average length of cells was about 5 μm.

Sample preparation for the microscopy is the same as described before (chapter 4).

Microscopy:

Long exposure images of single molecule RNAP-mEos2 was photoactivated with a 405-nm diode laser (CrystaLaser, Reno, Nevada) and subsequently imaged with a 561 nm laser (Sapphire 561 CW lasers, Coherent). Images were taken in time lapse mode (500 ms exposure, 1 s/ frame). Emission was collected through a 617/73 bandpass filter (bright line 617/73-25, Semrock).

Protocols for imaging ParB-GFP labeled chromosomal foci are same as chapter 4. Wide field imaging of RNAP-YGFP signal was is done as described in chapter 3.

Single molecules of ribosome-S2-mEos2 were imaged with the 561 nm laser. Power density of the 561 nm laser was kept at approximately 1.5 kW/cm^2. In time-lapse mode (30 ms/frame, 500 ms/frame), the probe laser was pulsed. At this power we achieved good signal-to-noise ratio for single molecules with minimal laser damage to the cell. The large number of ribosome per cell (~50,000) means that stochastic photo-activation caused by the probe laser itself (561 nm laser) is frequent enough for observing single ribosome-S2-mEos2 in good density. Therefore no photo-activation laser was necessary for this study. To minimize the phototoxic effect of the laser we collected data for less than 10 min per cell.

Results

Spatial distribution of transcribing RNAP copies

In spite of the intrinsic noisiness of short, diffusive trajectories, the behavior of the slow ("transcribing") and fast ("mixed state") RNAP copies is so different that we can distinguish them in a single camera frame using diffusion blur. In Fig. 5.1, we show several cell images taken with a 500-ms exposure time and the 561 nm laser on continuously at diminished power of only 100 W/cm^2 to minimize photobleaching. Even over 500 ms, the images of the slowly diffusing copies remain quite round with

FWHM ~340 nm, enabling location of the image centroid to an accuracy of ~40 nm. In contrast, the images of the "mixed" (jumping/ns-bound) copies become broad and diffuse. A simple thresholding algorithm can accurately separate the slow copies from the "mixed state" copies. In this way, over a long series of camera frames we can easily locate and classify about 150-100 individual slow RNAP copies per cell. By lowering the threshold of the single molecule detection, we can then detect both slow and fast copies in the same movie. For 1676 RNAP images from 37 cells, we found 44% slow copies and 56% "mixed state" copies. This is reasonably consistent with the previous partitioning estimate from the single-molecule diffusion (chapter 4).

After the series of 500-ms frames, we injected Sytox Orange into the sample chamber and obtained one image of the nucleoids for comparison with the spatial distribution of slow RNAP copies. The locations of slow RNAP copies (putative transcription events) are non-uniformly distributed within the nucleoids (Fig. 5.2. A, B, C). Slow RNAPs have a tendency to lie near the periphery of the nucleoid, i.e., at the interface between the nucleoid and the ribosome-rich regions (7). This tendency appears to be stronger in some cells than in others. Several additional single-cell examples are shown in Fig. 5.A.6.

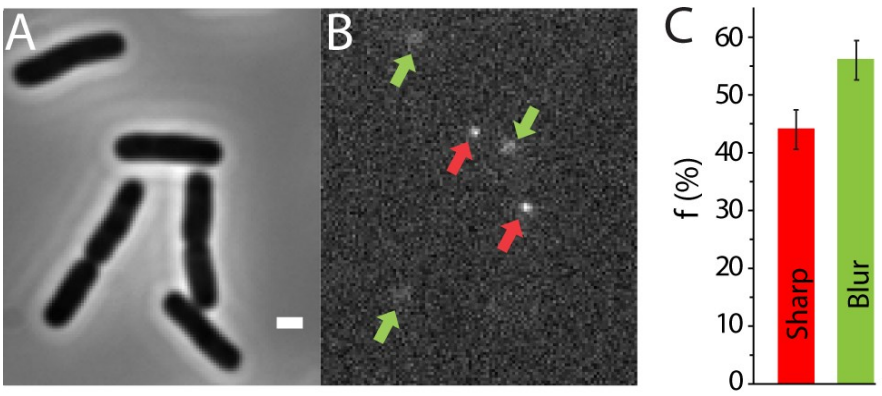

Figure 5.1: (A) Phase contrast image of 5 cells (3 non-septating and 2 septating); (B) one frame with 500 ms exposure showing 3 blurred spots (green arrows) and 2 sharp spots (red arrows). (C) Fraction of sharp and blurred spots from a total of 1676 spots detected in 200 single snapshots from different field of view.

Next we attempt to quantify the degree to which slow RNAP copies tend to preferentially locate towards the nucleoid periphery. This is non-trivial. First, we have only the two-dimensional projection of a three-dimensional spatial distribution. Second, in order to obtain enough data for careful comparisons, we must combine slow RNAP locations from many cells. A nucleoid-based coordinate system using principal component analysis of the nucleoid staining intensity was found to be problematic due to the variable shape of nucleoids from cell to cell. Instead, we focus on $P(y')$, the probability distribution of measured distances perpendicular to the cell body centerline. We determined a centerline coordinate system based on the cell body, as imaged by phase contrast. The smooth, mildly curved x' axis is determined as described in the Appendix 5C. The y' coordinate of any point is the perpendicular distance to the x' axis (the centerline of the cell body). An example of a centerline and the cell outline from which it was derived is shown in Fig. 5.2.C.

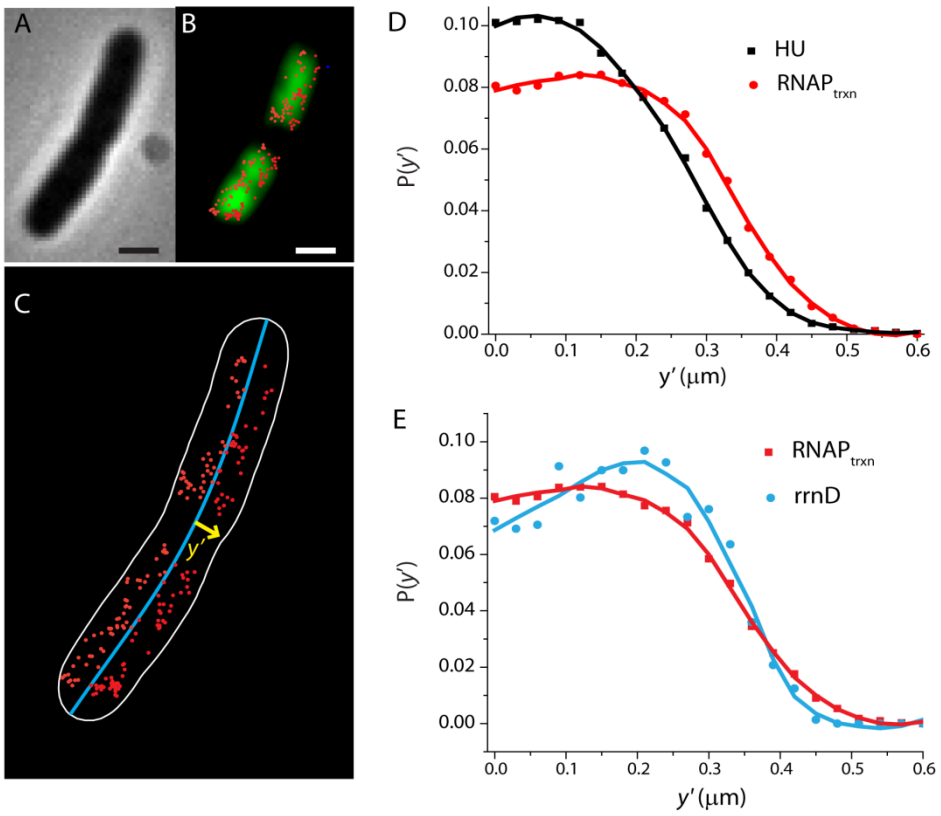

Figure 5.2: (A) Phase contrast image of a single cell is shown. Scale bar = 1 μm. (B) Red dots are positions of slow RNAPs ($RNAP_{slow}$) detected from the cell shown in (A). The Sytox orange labeled nucleoid is in the green channel. (C) Sub-pixel outline of the cell (white line) is constructed from the phase contrast image in (A). Blue line is the centerline of the outline. Radial coordinate (*y'*) of any position is its perpendicular distance from the centerline. (D) *P(y')* of slow RNAPs (red circles) and HU (black squares) is compared. Solid lines are smoothed interpolation of the data.(E) Comparison of *P(y')* of slow RNAPs (red squares) and *rrnD* (blue circles).

Spatial distribution of RNAPs transcribing rrnD

In what follows, we carefully compare the distribution *P*(*y'*) for slow RNAP, HU, *rrnG*, and transcription foci. As described in detail in Appendix 5B, *P*(*y'*) is a highly imperfect way to estimate the propensity of a particle to locate near the periphery of the nucleoid or near the cytoplasmic membrane.

Our inferences are necessarily qualitative. We obtained a large set of cell-body based (x', y') coordinates for slow RNAP (9106 copies over 260 cells), for HU-mEos2 (a proxy for the overall DNA distribution, 14151 copies over 70 cells), and for the DNA locus *rrnD* (3214 copies of each over ~800 cells). All three mutated VH1000 strains have very similar cell radii (Appendix 5D, Fig. 5A.4), meaning that close comparison of the transverse density profiles $P(y')$ for the three labeling schemes is appropriate. In each case, the total fluorescence image acquisition time for each cell is limited to less than 1.5 min. To determine possible effects of cell growth or thermal drift on the coordinate system, we compared the centerlines determined from phase contrast images taken before and after the fluorescence data acquisition (Appendix 5E). The average shift of the centerline along the transverse dimension was less than 30 nm (Fig. 5.A.4). Data from cells that displayed a shift larger than 100 nm were rejected.

In Fig. 5.2.D we compare the transverse probability distributions $P(y')$ for HU, and for slow RNAP all normalized to the same area. For HU, $P(y')$ peaks near the center of the cell and shows small but significant amplitude for $y' > 450$ nm, the mean cell radius under these conditions. We take this as a proxy for the total DNA distribution. In contrast, the distribution of slow RNAP is rather flat near the centerline ($y' = 0$), shows a mild peak at $y' \sim$ 150-200 nm, and has significantly greater amplitude beyond 450 nm, as described below. This suggests that transcribing RNAPs are preferentially localized towards the periphery of the nucleoid, as inferred from the comparison with wide field images of Sytox Orange stained DNA. Again, we emphasize that $P(y')$ grossly underestimates the propensity of slow RNAPs to localize near the periphery of the nucleoid (Appendix 5B).

In Fig. 5.2.E, we compare the $P(y')$ of slow RNAP with one of the 7 *rrn* operons, *rrnD*, in *E. coli* cell. For *rrnD*, $P(y')$ shows a marked minimum at $y' = 0$ and a strong peak at $y' \sim 200$ nm. The *rrnD* distribution terminates more abruptly near $y' = 450$ nm than the HU distribution. In Appendix 5G, we compare $P(y')$ distribution of *rrnD* with $P(y')$ distribution of *rrnG.* The distributions are qualitatively similar. $P(y')$ distribution of *rrnG* also peaks away from the center at $y' \sim 300$ nm (Fig. 5.A.7).

Spatial distribution of transcription foci and rrn operons in LB

The *P*(*y'*) distributions of *rrnD* operon suggest that transcription of *rrn* operons may occur preferentially at or near the nucleoid periphery. To test this inference, we sought to determine *P*(*y'*) for "transcription foci", locations at which a high concentration of RNAP is involved in transcription of rRNA. *E. coli* cells growing rapidly in rich medium show spatially narrow features in the diffraction-limited distribution of labeled RNAP (4, 5). If all or most of the *rrn* operons preferentially locate near the nucleoid periphery like *rrnG*, then we would expect the transcription foci to do the same. Obtaining sub-diffraction-limit images of transcription foci in live cells is difficult due to movement of DNA loci on the several-minute timescale of image acquisition. Because the transcription foci are not nearly as punctal and well defined in our standard conditions (EZRDM medium at 37ºC, 30-min doubling time) as they are in faster growth conditions (LB medium at 37ºC, 22 min doubling time), we obtained single-frame snapshots of the spatial distribution of RNAP-β'-yGFP for 50 cells growing in LB. An example image is shown in Fig. 5.3.A. We locate the bright puncta deemed transcription foci to ±100 nm by determining the pixel of local maximum intensity measured on the *y'* scale (centerline-based coordinate system). This enables formation of a histogram *P*(*y'*) for transcription foci, as shown in Fig. 5.3.B (orange line). In similar growth conditions, we also imaged *rrnD* operons (Fig. 5.3.B). The *P*(*y'*) of *rrnD* operons of cells grown in LB (blue line in Fig. 5.3C). The *P*(*y'*) distribution of transcription foci has a deep minimum at *y'* = 0 and a strong maximum at *y'* ~ 200 nm. The *P*(*y'*) of *rrnD* operons also peaked near 150-200nm.

Taken together, these data indicate that in medium/fast growth conditions, transcription of *rrn* operons occurs preferentially near the nucleoid periphery. The data speak less clearly about the distribution of actively transcribed genes expressing protein. However, if the slow RNAP distribution is a roughly equal mixture of *rrn* transcription and gene transcription, then the distributions in Figs. 5.2.E suggest that transcription of genes is more uniformly distributed throughout the nucleoid than transcription of *rrn* operons. To see this, mentally subtract half of the *rrnD* distribution (or the

transcription focus distribution) from the slow RNAP distribution. The remainder will have a peak at $y' = 0$ and roughly mimic the HU distribution.

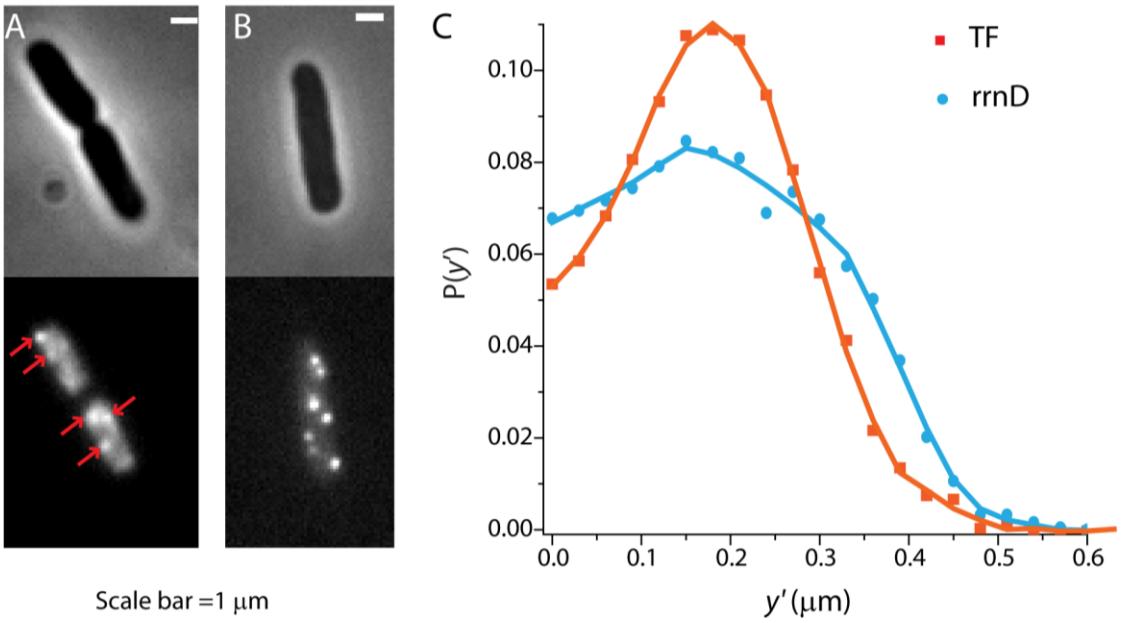

Figure 5.3: (**A**) Phase contrast and fluorescence image of a cell expressing RNAP-β'-yGFP and grown in LB medium at 37ºC. Red arrows point to bright, punctal "transcription foci". (**B**) Phase contrast and fluorescence image of a cell with *rrnD-pars*-ParB-GFP grown in LB medium at 37ºC. (**C**) Comparison of distribution $P(y')$ of transcription foci (combined data from ~60 cells xx check number) and $P(y')$ of *rrnD* (114 cells).

Slow RNAP and DNA near the membrane

Finally, we turn to the tails at large y' on $P(y')$ for HU and for slow RNAP. In Fig. 5.4, we compare the slow distribution with a model distribution from a uniformly filled cylinder of radius 450 nm, chosen to match the cell radius estimated from superresolution imaging of the spatial distribution of Kaede (appendix 5D). The model includes localization error of $\sigma = 40$ nm, chosen to match the localization

accuracy of sharp diffraction limited spots of slow RNAP. The tail of the slow RNAP distribution matches that of the cylinder model very well, strong evidence that there is significant slow RNAP at or very near the cytoplasmic membrane. The HU distribution also has a significant tail extending beyond the 380 nm cylinder radius (described in chapter 4) all the way out to 500 nm or more. This provides evidence of at least *some* HU/chromosomal DNA at or very near the cytoplasmic membrane. These data provide circumstantial evidence for the transertion hypothesis of co-transcriptional translation of membrane proteins, as discussed below.

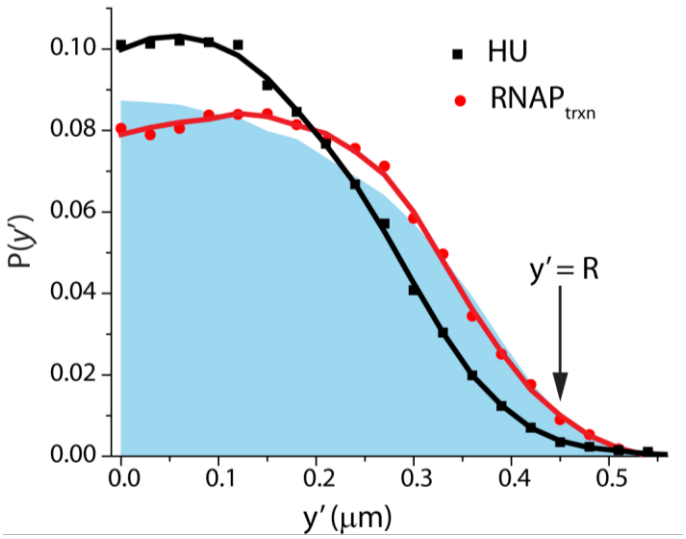

Figure 5.4: Comparison of distribution of $P(y')$ of slow RNAP and HU with uniformly filled cylinder. The blue area represents the $P(y')$ of simulated particles that are uniformly distributed within a cylinder of radius 450 nm (chosen to match the radius of the cytoplasm).

Slow ribosomes near the membrane

If the tail of the slow diffusing RNA polymerases truly represents RNAPs that are involved in "transertion", we should be able to find evidence for slowly diffusing ribosomes near the inner membrane as well. The ribosomes involved in "transertion" processes are expected to be tethered to both the inner

membrane and the nucleoid. Therefore, these ribosomes should display very limited dispersion about their position.

In chapter 2, we have tracked ribosomes at 33Hz frame rate with about 30-40 nm localization accuracy. The diffusion-based dispersion of all the translating ribosomes (~80% of all the ribosomes) at 33 Hz frame rate is very close to the localization accuracy. To separate the transerting ribosomes from rest of the translating ribosomes we needed to extend the timescale of tracking. Due to reversible photobleaching based methodology used for the single molecule imaging used for ribosomes labeled by S2-YFP we could not arbitrarily lengthen the time between frames. The photobleaching laser pulse converts all molecules to a dark state. After conversion, a small fraction of the dark molecules become fluorescent again. However, we cannot control the rate at which they return. If frame spacing is chosen too long, too many molecules are fluorescent in each frame. However a photoactivatable/photoswitchable fluorescent protein tag makes it possible to space the camera frames widely in time and observe the dynamics of protein at much longer timescales. Therefore in this chapter, we extend the ribosome tracking experiments to much longer times using the cell line expressing S2-mEos2 from the chromosome.

We show example ribosome trajectories of 5-s duration, tracked at 2 Hz rate, in Fig. 5.5A. The trajectories at longer timescale (5s) reveal a third, very slow ribosome population that diffuses even slower than the DNA loci (red trajectories in Fig. 5.5A). About 10-15% of the trajectories displayed such low diffusivity. In Appendix 5H, we describe the analysis of the excursion of the ribosome trajectories. We selectively examine the trajectories with very limited excursions ($\varepsilon_x, \varepsilon_y < 300$ nm) (Fig. 5.A.8). In Fig. 5.5.B we plot the $P(y')$ distribution of this population. The $P(y')$ of slow ribosomes peak near 300-350 nm. We compare the $P(y')$ of slow ribosomes with the simulated $P(y')$ distribution of simulated particles that are uniformly distributed within a cylinder of radius 450 nm (blue area in Fig. 5.5.B), and with a uniformly filled shell of thick ness 50 nm that is placed immediately inside the cytoplasmic membrane

(red line in Fig. 5.5.B). The $P(y')$ distribution of slow ribosomes appear to be similar to that of the 50 nm shell. The peak of the $P(y')$ distribution from slow ribosomes is pushed towards the center compared to the simulated $P(y')$ from the 50 nm shell at the membrane. In addition, the tail of the $P(y')$ of slow ribosomes has significantly higher amplitude near $y' > 450$ nm. We think that these discrepancies arise due to thermal drift during the data acquisition, which causes the fluorescence images to out of registry from the phase contrast images. The data also includes cells that are septating, which can cause the $P(y')$ of membrane bound copies to appear closer to the center. Therefore, we conclude that these slow ribosomes are distributed within 50 nm of the cytoplasmic membrane. While these copies are candidates for transertion, they might simply be ribosomes engaged with the Sec machinery.

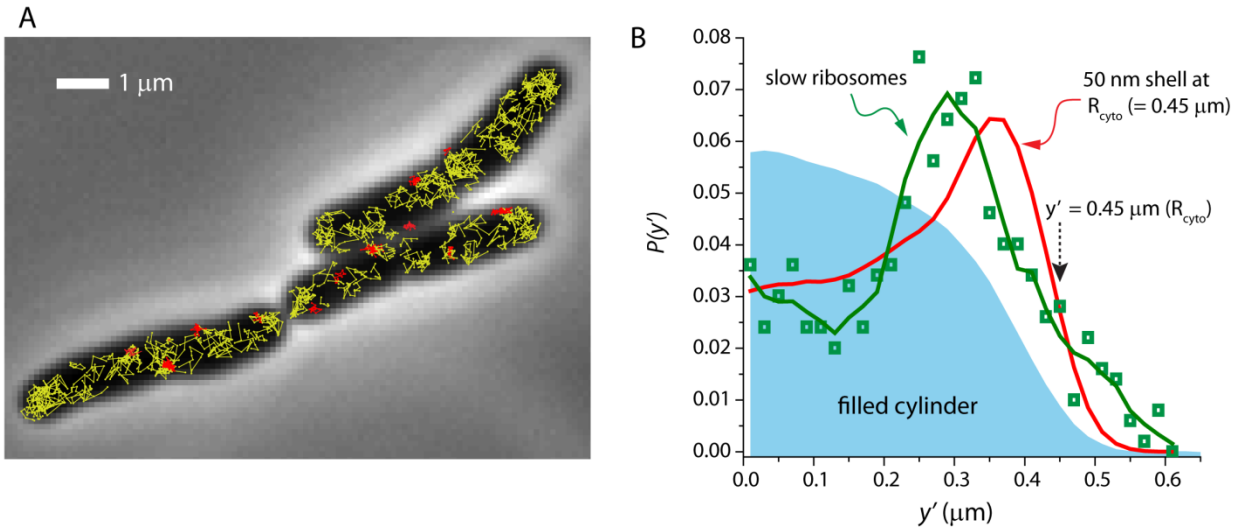

Figure 5.5: (A) Ribosome (S2-mEos2) trajectories of 5-s duration. Unusually compact trajectories (red) cluster near the cytoplasmic membrane. (B) Distribution of $P(y')$ of the compact trajectories ($\varepsilon_x, \varepsilon_y < 300$ nm, Fig. 5.A.8) as green sqaures. The green line is a smoothed interpolation of the data to guide the eye. We compare the $P(y')$ of compact ribosomes trajectories with a uniformly filled shell of thickness 50 nm inside the cytoplasmic radius (450 nm). The blue area represents the $P(y')$ distribution of simulated particles that are uniformly distributed within a cylinder of radius 450 nm (chosen to match the radius of the cytoplasm).

Discussion

The aim of this study was to investigate the spatial biology of transcription and get an insight about the organization of chromosome with respect to its genomic function. In the previous chapter we have determined the candidates of RNA polymerases that are involved in active transcription. This was done using the diffusivity of a RNAP molecule as a marker of its biochemical state and partitioning all the RNAPs inside a cell into different states based on their single molecule diffusion coefficient. In this chapter the subcellular locations of the RNAPs involved in active transcription were obtained using "localization enhancement"- based detection of the specifically bound RNAPs (10). These locations were obtained from a large number of cells and compared with the distribution of chromosomal DNA, measured by the distribution of DNA-binding protein HU. This provides a measure of the relative distribution of active genes with respect to the chromosomal DNA.

The spatial distribution of transcribing RNAPs is dynamic. In the chapter 4, we have described the sub-diffusive motion of chromosomal DNA. Such fluctuation of chromosomal DNA renders fluidity to the distribution of the bound RNAPs. Therefore the "transcription foci", which are composed of a large number of RNAPs engaged in transcribing *rrn* operons, appear mobile (5). The dynamic nature of the spatial distribution of the transcribing RNAPs has important consequences. A single position on the chromosome explores about 300 nm x 300 nm region in time scale of ~20s (average transcription time scale). This provides an estimate of the spatial dispersion of the transcribed mRNAs and thus sets a lower limit on the precision of proposed chromosome-centric organization of translation and mRNA decay (11). Spatial distribution of mRNAs produced over multiple rounds of transcription is expected to spread even more. Therefore it seems very unlikely that position of genes along the chromosome of *E. coli* can work as a template to spatially organize translation and mRNA decay (11). However, we note the possibility that mRNAs produced in bursts can have lower spreads (12).

The time averaged picture of the distribution of slowly diffusing, specifically bound RNAP copies show interesting organization. In the present work we have shown that the transverse distributions $P(y')$ of *rrnD* and of transcription foci both show a deep minimum at $y' = 0$ and a strong peak at $y' = 200\text{-}300$ nm. As outlined in Supporting Information, a true *radial* probability distribution $P(r')$ would have more intensity at large r' and less intensity near $r' = 0$ than is suggested by $P(y')$. In contrast to *rrnD* and the transcription foci, the HU distribution peaks near $y' = 50\text{-}100$ nm. If we take the HU distribution as a reasonable proxy for total chromosomal DNA density, it is clear that both *rrnD* and the transcription foci tend to lie much further from the cell axis than does DNA overall. For the doubling time of 30-40 min, Klumpp and Hwa estimate that half to two-thirds of the actively transcribing RNAP copies are transcribing *rrn* operons (2). To a first approximation, $P(y')$ for the slow (transcribing) RNAP looks like a combination of comparable amounts of an HU-like distribution and an *rrnG*/transcription foci distribution. We infer that transcription of *rrn* operons occurs near the nucleoid periphery, whereas the distribution of actively transcribed genes (encoding proteins) is more HU-like. As a cautionary note, the oligomer of ParB-GFP molecules bound to the *parS* site near the chromosomal locations can be a significantly large object. This can perturb the radial distribution of genes. In chapter 3 and 6, we describe the size dependent segregation of proteins from chromosomal DNA. It is possible that large ParB-GFP oligomers have a propensity to be segregated from the chrosomoal DNA and therefore to float to the periphery of the nucleoid. We plan to perform similar studies with *rrn* operons labeled with *tetO*-terR based labeling system, which can control the number of proteins bound to a certain site.

At the same time, we find convincing evidence of slow (transcribing) RNAP copies *at or near* the cytoplasmic membrane and some evidence of HU copies there as well. From previous work, we know there are ribosomes at or near the cytoplasmic membrane as well (7). In this work we selectively examine ribosomes with very low diffusivity. A significant fraction number of these ribosomes are found within 50 nm of the cytoplasmic membrane. These observations provide circumstantial evidence in favor of the

transertion hypothesis, which posits the co-transcriptional translation of membrane proteins, which are evidently inserted into the cytoplasmic membrane by the membrane-bound Sec-machinery (13).

The *E. coli* chromosome includes seven *rrn* operons, labeled A through G. In good growth conditions, there are some 30 total *rrn* operons per cell, far more than the apparent 3–6 transcription foci observed per cell. This indicates that multiple *rrn* operons must cluster together in space (4). The nature of the clustering mechanism is unknown. Our results raise the new question of what might cause clusters of *rrn* operons to locate at the nucleoid periphery. In rich growth conditions each *rrn* operon will be decorated with ~50 RNAP copies. A cluster of several *rrn* operons and associated RNAP copies will form a HUge biochemical object of mass ~20 MDa, ten times the mass of a 70S ribosome. We already know that ribosomes (perhaps in the form of polysomes) segregate strongly from the nucleoids (7). In earlier work (14), we built a toy model that provided a physical explanation for ribosome-nucleoid segregation in terms of entropic and excluded volume effects. Early on, Cook suggested a similar rationale for the clustering of highly active *rrn* operons with each other (15). Large clusters of *rrn* operons contain long segments of the chromosomal DNA itself, so they cannot segregate from the nucleoid. However, they might localize at the nucleoid periphery due to analogous entropic/excluded volume effects.

Finally, we suggest that placement of *rrn* operon clusters at the interface between the dense nucleoids and the ribosome-rich regions (7) may be functionally important for efficient ribosome assembly. Nascent pre-rRNA (~5500 nt) must be processed by RNase III, RNase E, and RNase P, among others, to form the 16S and 23S rRNA that are incorporated into the 30S and 50S ribosomal subunits, respectively. RNase III cleaves the pre-rRNA on either side of the 17S unit while transcription is occurring. RNase E then clips off the 5’ end of 17S to form pre-16S, which is then processed further. Both immuno-gold labeling and fluorescence studies have indicated that RNase E is anchored at the cytoplasmic membrane (16). Our suggestion is that placement of *rrn* clusters at the nucleoid periphery facilitates the diffusive search of 17S for the membrane-bound RNase E. If 17S were produced within the

dense nucleoid core and became entangled there, the search time to find the RNase E might be too long for efficient ribosome synthesis. In a possibly relevant experiment, Golding and Cox tracked long (4000 nt, comparable to the ~2000nt 17S), MS2-GFP tagged mRNA molecules that were transcribed from a plasmid in live *E. coli (17)*. Most copies exhibited sub-diffusion, as if tethered to DNA (or entangled there). They moved only ~100 nm over ten minutes. It is possible that all the enzymes that process pre-rRNA might distribute at the nucleoid-ribosome interface in a sort of localized ribosome factory. This hypothesis is testable using imaging methods analogous to those developed here.

While not yet conclusive, the following picture becomes plausible for *E. coli* in rapid growth conditions. Transcription of *rrn* operons occurs towards the periphery of the densest regions of nucleoids. Transcription of genes is more broadly distributed throughout the nucleoids. At least some membrane proteins are translated co-transcriptionally at the cytoplasmic membrane (the transertion mechanism). Accordingly, a less dense portion of the nucleoids extends radially all the way to the cytoplasmic membrane. Transcription of actively expressed genes for soluble proteins distribute widely throughout the nucleoids. There is evidence in favor of this picture from the Goulian lab (6). They have measured $P(y')$ for two specific genes that code for membrane proteins under conditions where transcription of these genes was induced or was not. The distributions for both membrane protein genes moved outward towards the membrane and away from the cell axis soon after induction (6). In contrast, two genes for soluble proteins showed no obvious change in distribution following induction (6). Future tracking studies of a variety of labeled genes might begin to discern how the membrane proteins genes find the cytoplasmic membrane.

In eukaryotes, spatial segregation of domains of the chromosome varies with transcription activity. Transcription activity dependent spatial segregation of domains of the chromosome is observed in eukaryotes which lead to the proposed inter-chromosome domain (ICD) model (18). Spatial segregation ofgenes, based on their transcriptional activity, can be an important regulator for the chromosomal

organization in *E. coli* also. As mentioned before, the two dimensional projection based analysis of the spatial distribution is very inaccurate in terms of quantifying the extent of segregation. Future work on understanding the three-dimensional topology of genes in bacteria will be important. Correlating the 3D position of a gene inside the cell with its transcriptional state will be able provide deeper insight regarding the chromosome organization principle in *E. coli*.

Appendix 5A: Comparison of growth rates

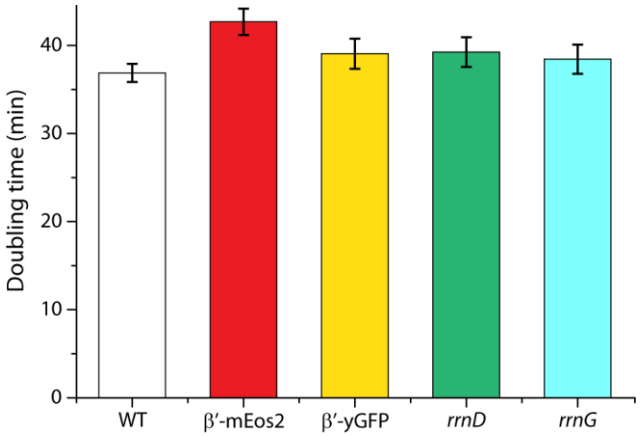

Figure 5.A.1: Growth of all the constructs used in this work are compared with the background strain (WT = VH1000). Doubling times for all the strains were comparable (39 ± 2 min).

Appendix 5B: *y'* is not a true "peripheral coordinate"

While $P(y')$ as defined can demonstrate that slow RNAP is more peripheral than HU, it gives a false impression of the *degree* to which this occurs. First, *y'* is a cell-body centerline coordinate, not a nucleoid centerline coordinate. The long axis of the nucleoid is sometimes poorly centered with respect to the cell-body. This will tend to push both the slow RNAP and the HU data towards larger *y'* than would occur for a nucleoid-based transverse coordinate, but we do not believe this is a large effect. More importantly, *y'* is not a true "peripheral coordinate". Inspection of Fig. 5.2.B shows that transcribing RNAPs can lie at the tips of the nucleoid periphery as well as the flanks. Such copies have $y' \sim 0$, so they will tend to fill in what might otherwise be a more pronounced dip in $P(y')$ near the centerline. Finally, and probably most importantly, *y'* is not a *radial* coordinate. A much more informative quantity is the radial distribution of slow RNAPs, which we call $P(\rho)$. Here $\rho = [(z')^2 + (y')^2]^{1/2}$ is a cylindrical coordinate defined in Fig. 5.A.2. Due to the volume element inherent in the "squashing" of the 3D image onto two dimensions, locations with large ρ (near the nucleoid periphery) are often mapped onto much smaller values of *y'*, as illustrated in Fig. 5.A.2. For these reasons, our $P(y')$ distribution badly *underestimates* the degree to which slow RNAPs and *rrnD* tend to distribute near the nucleoid periphery.

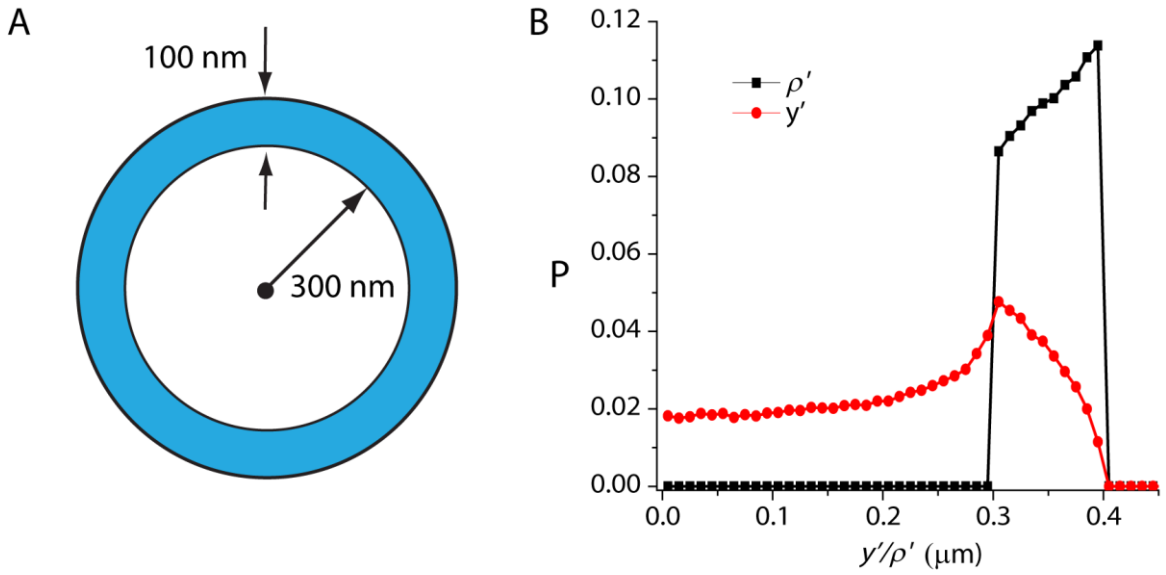

Figure 5.A.2: *P(y')* distribution badly *underestimates* the degree of peripheral distribution. (A) Schematic of a shell like distribution. We simulated uniform distribution of particles within a shell (Radius of the inner cylinder =300 nm and outer cylinder 400 nm, giving a shell thickness of 100 nm). The distribution of radial positions (*ρ'*) from 100,000 simulated positions is shown as the black line. The *P(ρ')* is nonzero in the range of *ρ'*= 300-400 and zero everywhere else. Within the range *ρ'*= 300-400, the distribution increases slightly as we move outwards to higher *ρ'* value. This is readily understood in terms of the increasing volume elements. Compared to the *P(ρ')* distribution, the *P(y')* distribution has significant amplitude at lower y' values. Also, the *P(y')* distribution has its peak near 300 nm.

Appendix 5C: Making a sub-pixel outline of the cell from its phase contrast image

The subpixel outline of the cell is created in a two-step process, similar to the work by Sliusarenko *et al* (19). The first step creates a pixel level mask of the cell from the phase contrast image. In the second step we use an active contouring algorithm described by Chan and Vese in “Active contours without edges" to get a finer subpixel outline (20). Before image processing the phase contrast images are inverted by subtracting all the pixel values from the maximum pixel value. In this format the cell appears as a bright object on a dark background. Then we clean up the image using simple morphological operations called "opening-by-reconstruction" and "closing-by-reconstruction". The crude pixel level mask is then generated by choosing a threshold (Fig. S3B). The threshold is determined by the Matlab function ‘graythresh’, which utilizes Otsu’s method (21)of computing the threshold from an image. Then we use the original phase contrast image with the binary mask created in the last step as an input to the “region_seg” function written by Shawn Lankton (available in Matlab file exchange). This computes the signed distance function (SDF) of the binary mask and evolves the contour of the SDF to minimize energy. These steps are off-lattice and therefore create a subpixel SDF. We generally apply 50 iterations per image after which the change in the SDF is minimal. The contour of the SDF at [0, 0] serves as the finer mask of the cell (Fig. S3C).

To get the cell based coordinate system we then proceed to determine the centerline of the cell using a Voronoi transformation (Matlab function “voronoi”) of the subpixel mask obtained in the previous section. First we select points of the Voronoi transformation that lie within the cell outline. Then we reject points at the end caps to avoid branching. Now a spline is fit to the points, and linear extrapolation of the spline using the Matlab function “fnxtr” to each pole gives a smooth subpixel

centerline of the cell (Fig. S3D). This centerline is then used to determine the coordinate for any particle position detected within the cell.

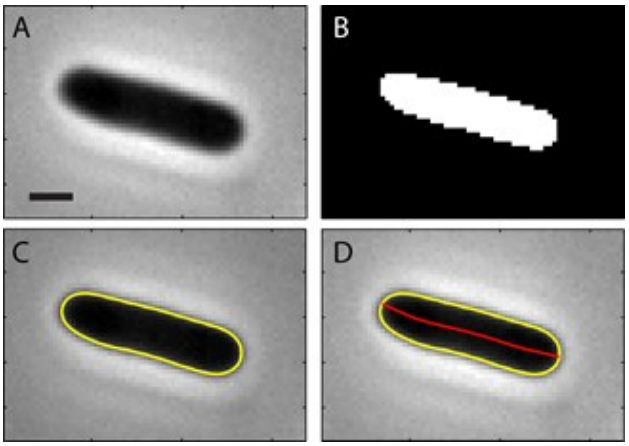

Figure 5.A.3: (A) Phase contrast image of *E. coli* cell is shown. (B) Binary mask from the same image after cleaning by "opening and closing by reconstruction" and thresholding. (C) Final result of the active contouring gives a smooth outline of the cell, shown as a yellow line. (D). The centerline of the cell, calculated from the smooth outline in (C), is plotted as a red line.

Appendix 5D: Measurement of cell radius

To get the cytoplasmic radius of the background strain VH1000, we have used the spatial distribution of Kaede expressed inside the cell from a plasmid as explained in Bakshi *et al*., 2011(22). Data for the superresolution image of Kaede was collected for 10,000 frames at 240 Hz. The centroids of the fluorescence images of single Kaede molecules are shown (Fig. 5.A.4.A: orange dots are positions that fall within the cylindrical portion of the cell; blue dots are position that fall in the end cap region). We then proceed to compare the distribution of the distance of the molecules within the cylindrical portion (orange dots in Fig. 5.A.4.A) from the centerline of the cell (red line in Fig. 5.A.4.B) with a model distribution expected from a homogeneously filled cylinder. The radius of the cylinder model gives the best estimate of the radius of the cytoplasm of the cell. For 36 cells, we have determined the mean radius to be 445 ± 25 nm (R-Kaede in fig. 5.4.A.C). A more simplistic approach is to get an estimate for the cell radius is using the phase contrast image itself. Using the contouring algorithm explained in section regarding the sub-pixel outline of the cell, we get points on the smooth outline of the cell. The mean perpendicular distances of the points within the cylindrical region (shown as blue dots in Fig. 5.A.4.B) from the centerline is assigned as the radius of the cell. Results of this analysis are compared with the rigorous approach explained previously (Fig. 5.A.4.C). The radius determined from the contour of the phase contrast image (mean radius = 480 ± 20 nm) is in good agreement with the superresolution based measurement (with a systematic overestimation by ~30 nm). Therefore we proceeded to compare the radius of the strains used in our study using the simple phase contrast imaging based approach. All the cells seem to have comparable radius (Fig. 5.A.4. D: R (RNAP-β'-mEos2) =476 ± 18 nm, R (*rrnD-pars*-ParB-GFP) = 481±20, R (*rrnG-pars*-ParB-GFP) = 486±24 and R (HU-mEos2) =482 ± 20 nm.)

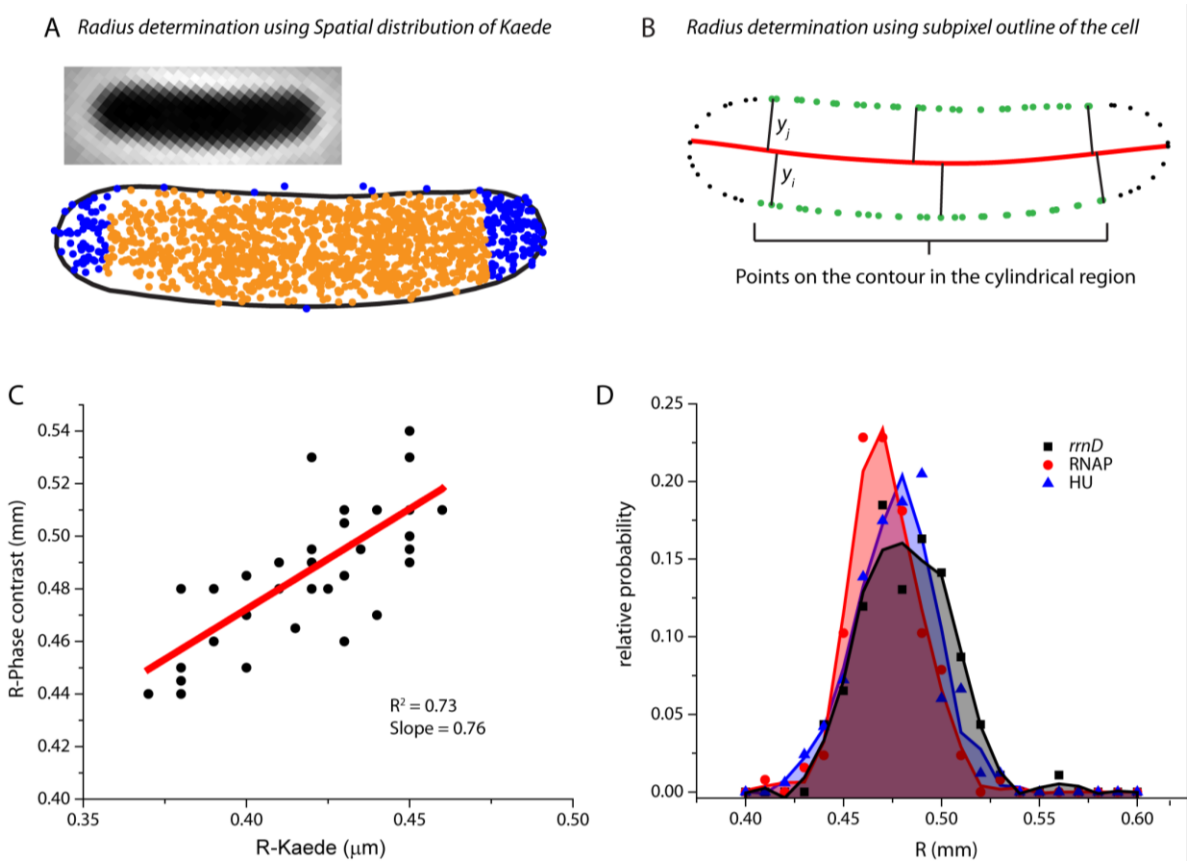

Figure 5.A.4: (A) Centroids of fluorescence images of single Kaede molecules detected from a cell are plotted over the cell outline determined from the phase contrast image of the cell shown in inset. Positions that lie within the end cap portion (chosen as any position within 500 nm of the cell pole) are plotted as blue dots and rest of the positions that fall within the cylindrical region are plotted as red dots. (B) Points on the subpixel outline of the cell are shown along with the centerline (blue dots are boundary points at the cylindrical portion and black dots are boundary points in the end cap). Black lines show the perpendicular lines drawn to the centerline from each point (y_i). Length of this line is the perpendicular distance from the cell centerline. Average of the absolute values of this perpendicular distances ($\sum_i y_i \Big/ N$) gives the radius of the cell. (C) Radius of the cell determined from the two methods is compared. (D) Distribution of cell radius from three strains used in the work (RNAP-β'-mEos2, *rrnD* and the HU-mEos2 are compared)

Appendix 5E: Comparing the centerlines of the cell outline imaged before and after the data acquisition

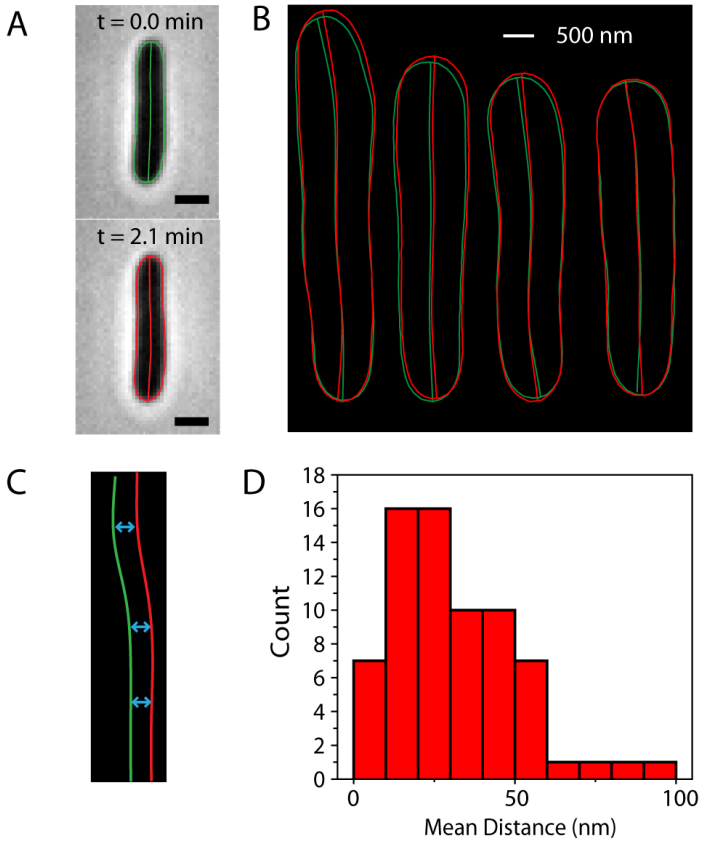

Figure 5.A.5: (**A**) Sub-pixel mask and centerlines are calculated from phase contrast image taken before and after the fluorescence data acquisition. (**B**) Sub-pixel mask and the centerline of the cell calculated from the phase contrast images taken before the fluorescence data acquisition (green) and after (red) are compared. (**C**) Perpendicular distance between the centerlines determined before and after the measurement gives an estimate of total shift along the short axis of the cell during the entire image acquisition time. (**D**) Histogram of the mean perpendicular distance between the centerlines is plotted. The mean is around 30 nm.

Appendix 5F: Montage of superresolution image of locations of slow RNAP from multiple cells

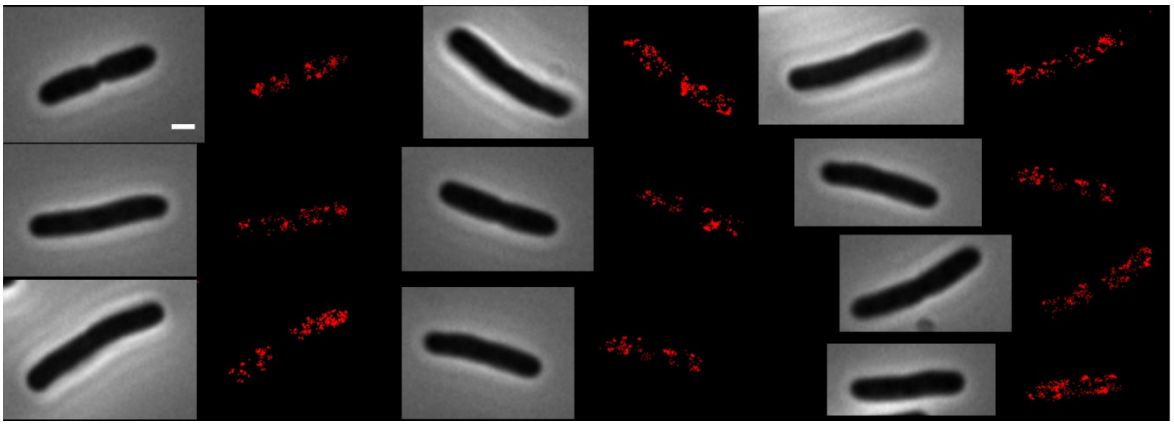

Figure 5.A.6: Montage of superresolution images of slow RNAP positions from multiple cells. Scale bar =1 μm.

Appendix 5G: *P(y')* distribution of two *rrn* operons

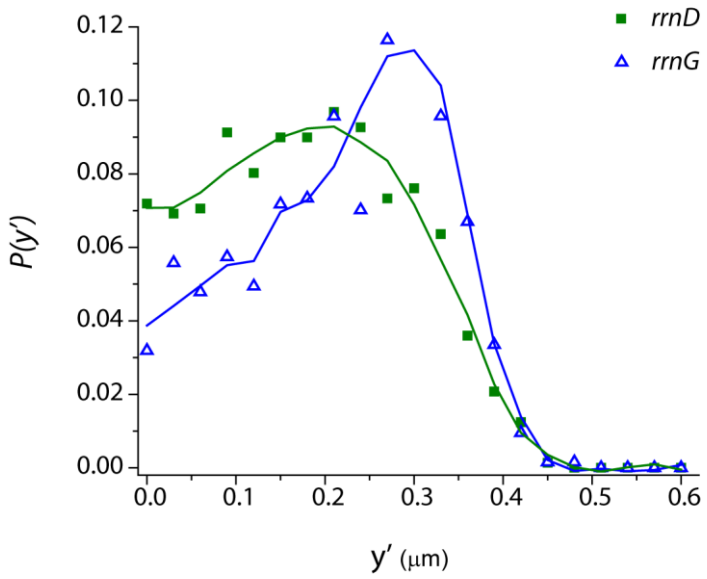

Figure 5.A.7: *P(y')* of *rrnD* (green squares) and *rrnG* (blue triangles). Solid lines are smoothed interpolation of the data to guide the eye.

Appendix 5H: Diffusion analysis of ribosome (S2-mEos2)

Our goal is to determine the value of $\varepsilon_x, \varepsilon_y$ which best discriminates trajectories of ribosomes translating untethered, mature mRNA. In previous work (chapter 3), we have estimated the diffusion constant of such ribosomes to be 0.03 $\mu m^2/s$. For the model trajectories with D = 0.03 $\mu m^2/s$ and 30 nm localization accuracy, the lower bound $\varepsilon_x, \varepsilon_y > 0.3$ μm captures 99% of the trajectories. Therefore we used that cutoff (black dotted line in Fig. 5.A.7 (A)) to reject trajectories of ribosomes translating mature mRNA.

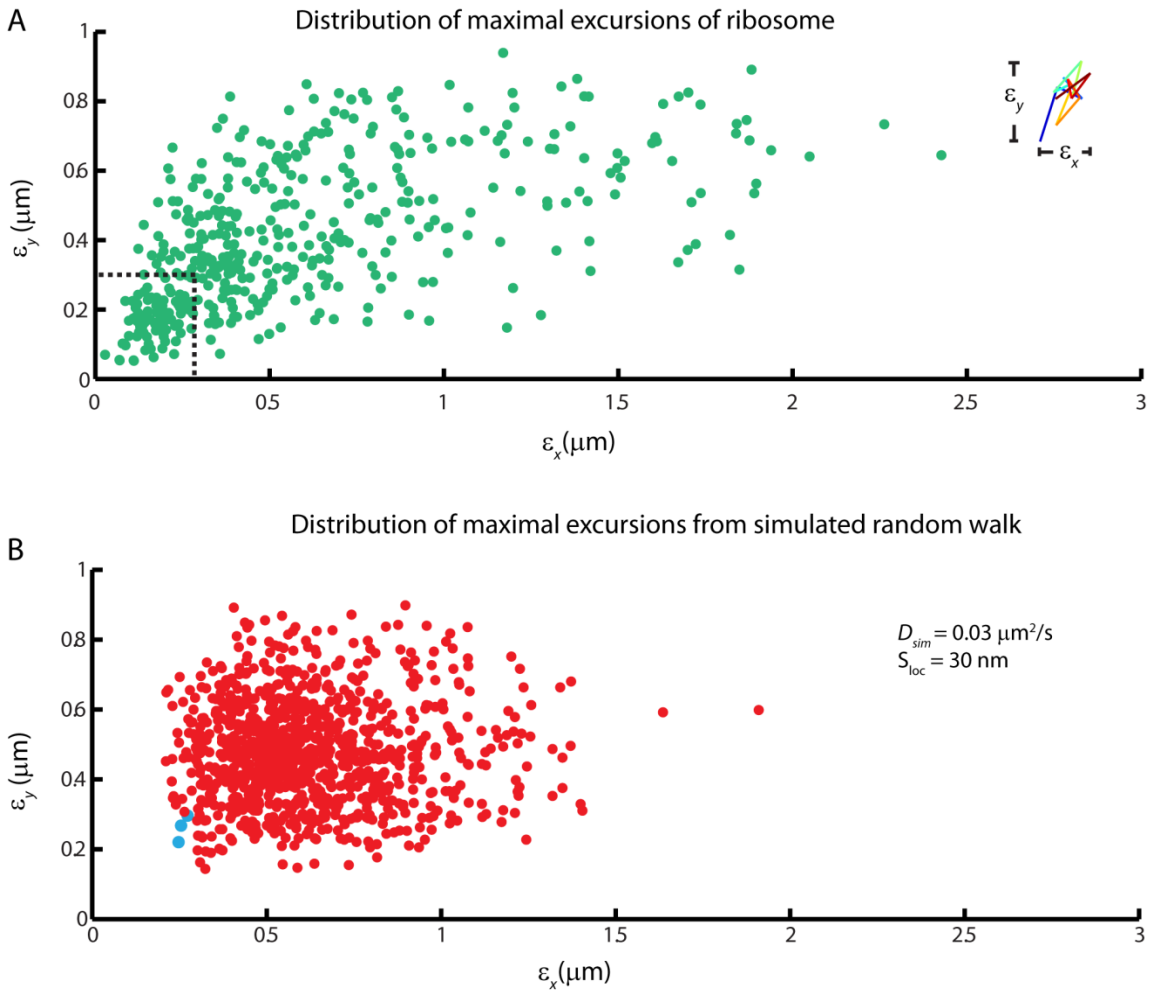

Figure 5.A.8: (**A**) For 10-step trajectories of ribosome obtained at 2 Hz, scatter plot of ε_y vs ε_x. Dashed line denotes the value of ε_x , ε_y = 0.3 μm, which most cleanly separates the slowly diffusing ribosomes. (B) Red dots are ε_x vs ε_y scatter plot for 1000 10-step simulated trajectories with D = 0.03 μm^2-s^{-1} and 30 nm localization accuracy. Blue dots are trajectories that are beyond the cutoff.

References

1. Campos, M., and C. Jacobs-Wagner. 2013. Cellular organization of the transfer of genetic information. Current Opinion in Microbiology. 16:171-176.

2. Klumpp, S., and T. Hwa. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proc. Natl. Acad. Sci. U. S. A. 105:20245-20250.

3. Bremer, H., P. Dennis, and M. Ehrenberg. 2003. Free RNA polymerase and modeling global transcription in *Escherichia coli*. Biochimie. 85:597-609.

4. Cabrera, J. E., and D. J. Jin. 2003. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Mol. Microbiol. 50:1493-1505.

5. Bratton, B. P., R. A. Mooney, and J. C. Weisshaar. 2011. Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. J. Bacteriol. 193:5138-5146.

6. Libby, E. A., M. Roggiani, and M. Goulian. 2012. Membrane protein expression triggers chromosomal locus repositioning in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109:7445-7450.

7. Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol. Microbiol. 85:21-38.

8. Woldringh, C. L. 2002. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Mol. Microbiol. 45:17-29.

9. Norris, V., and M. S. Madsen. 1995. Autocatalytic gene expression occursviatransertion and membrane domain formation and underlies differentiation in bacteria: A model. Journal of Molecular Biology. 253:739-748.

10. Elf, J., G.-W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. Science. 316:1191-1194.

11. Llopis, P. M., A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. 2010. Spatial organization of the flow of genetic information in bacteria. Nature. 466:77-U90.

12. Taniguchi, Y., P. J. Choi, G. W. Li, H. Y. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. Quantifying e-*coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science. 329:533-538.

13. Driessen, A. J. M., and N. Nouwen. 2008. Protein translocation across the bacterial cytoplasmic membrane. Annual Review of Biochemistry. 77:643-667.

14. Mondal, J., B. P. Bratton, Y. Li, A. Yethiraj, and J. C. Weisshaar. 2011. Entropy-based mechanism of ribosome-nucleoid segregation in e. *coli* cells. Biophys. J. 100:2605-2613.

15. Marenduzzo, D., K. Finan, and P. R. Cook. 2006. The depletion attraction: An underappreciated force driving cellular organization. J. Cell Biol. 175:681-686.

16. Murashko, O. N., V. R. Kaberdin, and S. Lin-Chao. 2012. Membrane binding of *Escherichia coli* rnase e catalytic domain stabilizes protein structure and increases rna substrate affinity. Proc. Natl. Acad. Sci. U. S. A. 109:7019-7024.

17. Golding, I., and E. C. Cox. 2004. Rna dynamics in live *Escherichia coli* cells. Proc. Natl. Acad. Sci. U. S. A. 101:11310-11315.

18. Dietzel, S., K. Schiebel, G. Little, P. Edelmann, G. A. Rappold, R. Eils, C. Cremer, and T. Cremer. 1999. The 3d positioning of ant2 and ant3 genes within female x chromosome territories correlates with gene activity. Exp. Cell Res. 252:363-375.

19. Sliusarenko, O., J. Heinritz, T. Emonet, and C. Jacobs-Wagner. 2011. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. Mol. Microbiol. 80:612-627.

20. Chan, T. F., and L. A. Vese. 2001. Active contours without edges. Ieee Transactions on Image Processing. 10:266-277.

21. Otsu, N. 1979. Threshold selection method from gray-level histograms. Ieee Transactions on Systems Man and Cybernetics. 9:62-66.

22. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live *Escherichia coli*. Biophys. J. 101:2535-2544.

Chapter 6

Time-lapse Imaging of *E. coli* Nucleoid Morphology after Treatment with Transcription and Translation Halting Drugs

Introduction

The genome of *E. coli* consists of 4.6 million base pairs. The contour length of a fully stretched genome is about 1.5 mm. In a rapidly growing *E. coli* cell, multiple genome equivalents of chromosomal DNA have to fit within the small cell volume of about 2 μm^3. However, as observed with electron microscopy and fluorescent dye based DNA-staining, the genomic DNA does not fill the entire volume of the cell (1-3) . This irregularly shaped compacted DNA inside bacterial cells is called the nucleoid. The shape and size of the nucleoid is presumably decided by a balance between compacting forces, like macromolecular crowding, DNA supercoiling, DNA binding proteins, and depletion entropy (4, 5), and expanding forces like "transertion". The transertion hypothesis refers to the process of co-transcriptional translation and simultaneous insertion of membrane proteins (6, 7). Since transertion would tether the chromosomal DNA to the cytoplasmic membrane, Woldringh and others have suggested that transertion acts as an "expanding force" on the DNA itself (6, 8, 9). It has also been proposed that transertion-mediated membrane attachment can play an important role in chromosome segregation (6). However, the possible effects of transertion on the shape and size of chromosomal DNA in bacteria remains largely untested.

A recent study found that the spatial distribution of genes coding for membrane proteins shifts outwards towards the cytoplasmic membrane when these genes are induced (10). Such preferential

movement of membrane proteins' genes towards the membrane provides circumstantial evidence for transertion-mediated membrane tethering of chromosomal DNA. Such membrane tethering of DNA can act as an expanding force on the nucleoid. The same study probed the locations of the genes along the radial dimension (short axis) of the cell. It was not clear whether membrane protein genes also move towards the membrane along the axial dimension (long axis of the cell) upon induction.

In earlier work, based on superresolution imaging of ribosomes and RNAPs in *E. coli,* we found clear evidence that radial distributions of both ribosomes and RNAP extend all the way to the cylindrical section of cytoplasmic membrane in close proximity to the nucleoid lobes (11). In chapter 5, using single molecule tracking experiments, we were able to establish that a significant number of these RNA polymerases and ribosomes are actively involved in transcription and translation, respectively. However, there were very few if any RNAP near the membrane in the end cap region. This suggested that transertion can only supply a direct radially expanding force on the nucleoid.

As is well known, rifampicin treatment halts transcription initiation by RNAP (12). On the other hand, chloramphenicol treatment halts translation, by preventing release of ribosomes that are already translating a message (13, 14). If the transertion hypothesis is correct, then halting of either transcription or translation should eventually cause radial contraction of the nucleoid, either by breaking the DNA–RNAP linkage (rifampicin) or by breaking the ribosome-membrane linkage (chloramphenicol). Based on typical transcription and translation time scales, we would expect the nucleoid to undergo radial compaction within ~ 5 min. However, static images of DNA stains obtained ~30 min after drug treatment show that while chloramphenicol further compacts the nucleoid, rifampicin actually expands the nucleoid both radially and axially (9, 15). This later result would seem to contradict the transertion hypothesis.

To investigate the time-course of the effect of these drugs on the morphology of *E. coli* nucleoid, we have performed time-lapse fluorescence microscopy of with the DNA stain Sytox Orange. Time-lapse

fluorescence microscopy of the action of drugs on the nucleoids of single, live bacterial cells provides a completely new level of real-time mechanistic detail that is not discernible from measurements at static images at two time points. Fluorescence imaging on live cells can also avoid possible artifacts of cell fixation (9). In the following sections we describe our method and give the results, ending with a discussion of the implications of our findings.

Materials and Methods

Bacterial strains

The wild type *E. coli* strain used for this work is VH1000. For widefield images of the ribosome distribution we used the ribosome-S2-YFP strain (AFS55) described elsewhere (11). The strain was transferred to VH1000 background using P1 transduction. For the photo-activated pulse chase studies of ribosome dynamics, we used the ribosome-S2-mEos2 described in chapter 5.

To image the distribution of Kaede we used the strain JCW11, described in chapter 2 and our previous work (16). A strain expressing GFP under the control of IPTG was used for the GFP imaging studies. Construction of this strain is described in our earlier work (17).

Sample preparation for microscopy

Cell cultures were grown in EZ Rich Defined Medium (EZRDM from Teknova) (18). Cells were grown overnight with shaking in 30ºC water bath. We subsequently made subcultures by diluting the stationary phase culture at least 1:100 into 3 mL of fresh EZRDM. When cells had grown to midlog phase (OD_{600} = 0.4-0.6), 7 µL of culture was injected into a pre-warmed flow chamber. The construction of the flow chamber has been explained (19). Cells were allowed to adhere to the poly-L-lysine–coated coverslip for 2 minutes and then the remaining cells were rinsed away with fresh EZRDM. Temperature

of the flow chamber was maintained at 30ºC throughout the experiment. The cells grow with a doubling time of 70 min in the flow chamber, longer than the 50 min doubling time in the water bath shaker.

Spatial distribution of chromosomal DNA inside *E. coli* cells are imaged with Sytox Orange. A 5 mM solution (in DMSO) of Sytox Orange was purchased from Molecular Probes (S7020). To image the nucleoid a 500 nM stock solution of Sytox Orange in EZRDM was injected into the chamber. We waited for 10 minutes for Sytox Orange to flow into the cells and stain the nucleoid. The cells continued to grow during this time with a doubling time of 75 min as measured by cell length calculated from the phase contrast images. This is comparable to the doubling time (70 min) without Sytox Orange staining.

Rifampicin and chloramphenicol stock solutions were prepared by dissolving 5 mg and 30 mg of the drugs in methanol, respectively. For widefield imaging of the time course of drug action, we injected drug solution in EZRDM (concentration: 200 μg/mL) into the flow chamber already containing plated cells in growth medium. For imaging of ribosome and DNA static spatial distributions, stock solutions of the drugs were added to the midlog cell culture to attain a final concentration of 200 μg/mL, and the cultures were shaken for 30 min before plating and imaging.

Time-lapse fluorescence microscopy

Cells were imaged using a Nikon Eclipse Ti inverted microscope equipped with an oil immersion objective (CFI Plan Apo Lambda DLL 100X Oil, 1.30 NA). A 514 nm Ar^+ laser (Melles Griot, Carlsbad, CA) was used for imaging ribosome-S2-YFP distribution. A 560/50 emission filter was used to collect the emission. Distribution of Sytox Orange stained DNA is monitored with a 561 nm laser (CrystaLaser, Reno, Nevada). A 617/73 bandpass filter (bright line 617/73-25, Semrock) was used to collect the Sytox Orange signal. To monitor cell length vs time (a proxy for growth) in the presence of transcription/translation halting drugs, the cells were alternately excited by 561 nm (for imaging DNA)/ 514 nm (imaging ribosome) and white light at a rate of 1 frame every 6 seconds with an exposure time of

50 ms/frame. For the wide field imaging of Kaede and GFP a 488 nm Ar^+ laser (Melles Griot, Carlsbad, CA) was used. Emission was collected through a 510/20 bandpass filter (Chroma).

Images were recorded by a back-illuminated EMCCD camera with 512 x 512 pixels of 16 μm x 16 μm each (iXon DV-897, Andor Technology, Connecticut). Each pixel corresponds to 107×107 nm^2 at the sample (150X overall magnification).

Image acquisition and data analysis

The interleaved fluorescence and phase contrast images were separated during image processing. Phase contrast images were segmented into binary masks of individual cell by simple thresholding techniques. An elliptical mask was made in MATLAB that matched the cell boundaries. The major axis length of this mask was measured to estimate the cell length. Cell lengths were calculated at each time-point for correlation with events in the same cell signaled by changes in fluorescence intensity profiles vs time. Images were analyzed using a MATLAB GUI developed in our lab. Briefly, the GUI proceeds in following steps to analyze cell growth from phase contrast image and morphology of the nucleoid from Sytox Orange images. First, an elliptical mask was made from the phase contrast image of each cell. The major axis length of this mask was measured to estimate the cell length vs time for correlation with events in the same cell signaled by changes in fluorescence intensity vs time. For analyzing the Sytox Orange stained DNA images, we first determined the two principal axis of the cell from the elliptical mask. The long axis is referred to x (axial dimension) and the short axis perpendicular to the x-axis is referred to y (radial dimension). We measure the intensity profiles of Sytox Orange along x and y axis for the nucleoid shape analysis. The intensity distribution of ribosome-S2-YFP was analyzed similarly.

Results

Parameters to describe the morphology of nucleoid

We have used Sytox Orange as a marker for the spatial distribution of chromosomal DNA within the cell. For a DNA stain to be useful for time-lapse imaging, it is important that the imaging condition has minimal perturbative effect. In our previous work, we have demonstrated that the combination of staining with DRAQ5 and illuminating the cell with the 633 nm laser greatly perturbs the structure of the DNA (unpublished work). The nucleoid shape of the DRAQ5 stained DNA expands from its usual bi-lobed structure to fill a large fraction of the cell within 15 seconds of illumination with a 633 nm laser. Cell growth also halted at similar time scale. Similar pertubative effects were observed with the DNA stain DAPI. In contrast, DNA stained by Sytox Orange did not show any such perturbations to the nucleoid shape and cells stained with Sytox Orange continue to grow with a doubling time of 75 min. This is comparable to the doubling time of the cells without any stain (70 min).

We have used two parameters to describe the size and shape of the nucleoid from the Sytox Orange intensity distribution. Without any drug treatment, these parameters follow the time course that is characteristic of normal growth (measured by looking at snapshots of Sytox Orange stained DNA from cells at different growth phase (unpublished data). The axial distribution was characterized with the length of the nucleoid, L, which is the full-width at half-maximum height (FWHM) of the axial fluorescence distribution. The radial width of the nucleoid is measured with a second parameter, W, the full width at half-maximum maximum intensity along the transverse dimension. In Figure 6.1 we show an illustration of these parameters for an example cell stained with Sytox Orange. L is not a good parameter for measuring the true axial morphology of the nucleoid. The nucleoid in a cell has two lobes and the sum of length of those individual lobes is much smaller than L. However, the long tails in the Sytox Orange images cause the intensity from two lobes of the nucleoid to overlap, and therefore we cannot examine the length of each lobe independently.

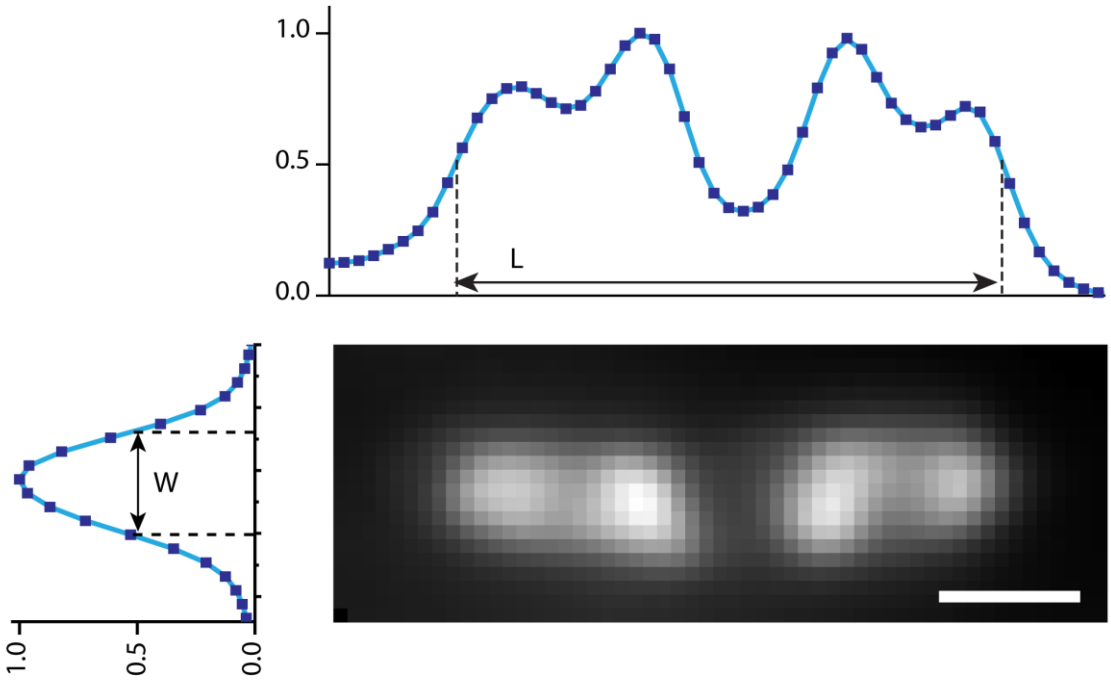

Figure 6.1: Image of a Sytox-orange stained nucleoid. Most cells exhibit two primary nucleoid lobes, often segregated into two sub-lobes each. The two parameters we use for characterizing the morphology of the nucleoid are the full width at half max intensity along the two principal axes of the cell. The length (L) of the nucleoid is the full width at half max along the long axis (x) of the cell and the width (W) of the nucleoid is full with at half max along the short axis (y) that runs perpendicular to the long axis.

Effects of chloramphenicol on nucleoid morphology and cell growth

The effect of halting translation with 200 µg/mL chloramphenicol on a single cell is illustrated in Fig. 6.2. Prior to the injection of chloramphenicol, the nucleoid of the cell displayed two major lobes that are divided into two sub-lobes. The cell displays two spherical shaped compacted nucleoid lobes 30 minutes after the chloramphenicol injection (Fig. 6.2A). This is consistent with previous studies on effect

of chloramphenicol (9). Prior to the injection (green swaths in Fig. 6.2 B-C-D), the cell length continues to increase gradually. The length continues to increase until about 5 minutes after the injection of chloramphenicol, then abruptly stops and plateaus for rest of the movie (30 minutes). The nucleoid undergoes dramatic morphological changes during this time (Appendix 6.A.1). The two sub-lobes within each nucleoid lobe merge into one elliptical lobe. The elliptical lobes shrink axially for the rest of the movie and become almost spherical (Appendix 6.A.1). In Fig. 6.A.2, we show the axial line scans of the Sytox Orange intensity distribution for this cell at different time-points during the movie. The plot of combined length of the two lobes, L, shows that axial dimension starts to shrink about 2 minutes after the injection and continues to gradually decrease with time (Fig. 6.2.D). The decrease in L badly underestimates the compaction of the nucleoid along the axial dimension. Currently we lack a method to probe the axial dimension of each of the lobes separately. The radial dimension of nucleoid shrinks within first five minutes of injection, as demonstrated from the ~20% drop in W (Fig. 6.2.C). Transverse line scans of the Sytox Orange intensity, for the cell used in Fig. 6.2, are shown in Fig. 6.A.3.

These morphological changes of the nucleoid after chloramphenicol injection are fairly general. In Appendix 6B we show time course of W after chloramphenicol injection for 12 more cells (Fig 6.A.4.A). In all cases, W drops by about 20% of its initial magnitude ($W_{t=0}$ = 750 nm) within the first five minutes of injection. We have not synchronized the growth of cells, and therefore different cells are in different growth phase during the experiment. This causes the axial dimension of the nucleoids to be very different from cell-to-cell. For this reason we could not perform a quantitative comparison of cell-to-cell variation of the effect of chloramphenicol on nucleoid morphology. However, the trends are very similar across different cells.

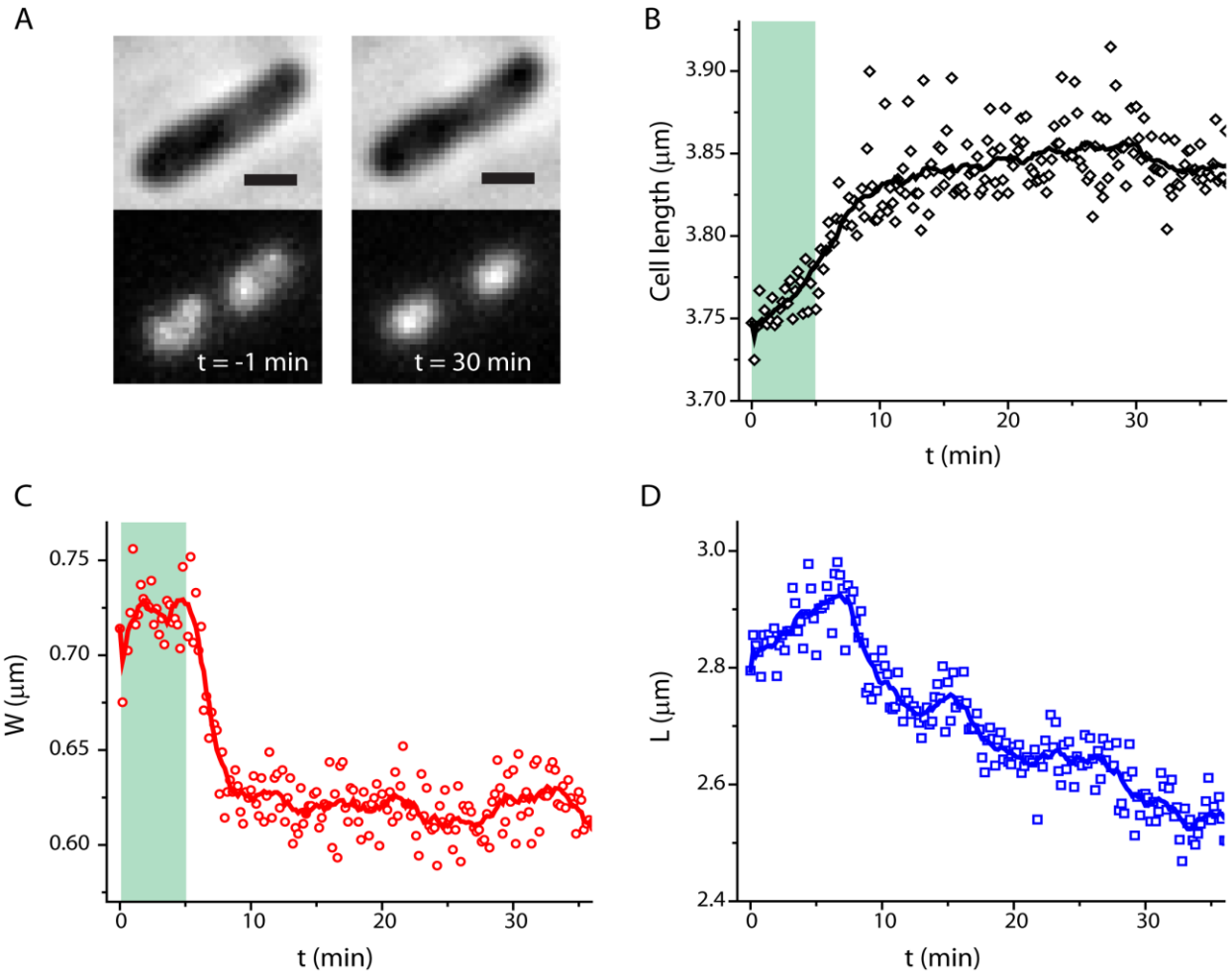

Figure 6.2: Effect of treatment with chloramphenicol. (A) Phase contrast image of the cell (top panel) and wide field images of Sytox Orange stained DNA before and 30 min after addition of 200 μg/mL chloramphenicol. Scale bar =1 μm. (B) The length of cell, measured from the phase contrast image, is plotted as a function of time. Chloramphenicol was added at 5 min. The cell length continues to increase up to 5 min after the addition of chloramphenicol and then slows down. Cell elongation completely halts 10 minutes after the addition of chloramphenicol. (C) The width of the nucleoid (W) is plotted as a function of time. The nucleoid shrinks by ~20% within five minutes after the addition of chloramphenicol along the radial dimension. No significant change in the width is observed for the rest of the movie (duration 40 min). (D) The length (L) of nucleoid continues to increase up to the first 2-3 minutes and then gradually shrinks for rest of the movie. The green swath marks the time before injection. The solid lines in B-D are smoothed interpolation of the data to guide the eye.

Effects of rifampicin on nucleoid morphology and cell growth

In Fig. 6.3 we illustrate the effect on a single cell of halting transcription with 200 μg/mL rifampicin. Prior to rifampicin addition the cell length increases gradually and the cell displays two lobes of nucleoid. The example cell is a short one (3.2 μm) and therefore does not have prominent sub-lobe structure. The image of the nucleoid at different time points after rifampicin addition is shown in Appendix E (Fig. 6.A.5). Within the first five minutes after rifampicin addition the cell elongation slows down, the two lobes of the nucleoid merge, and the nucleoid shape changes from a dumbbell shape to a filled ellipsoid from. During this morphological change, the radial dimension of the nucleoid shrinks significantly (~20% decrease in W, Fig. 6.3.C). As the two nucleoid lobes merge, they move towards the center of the cell causing the axial dimension of the nucleoid to shrink (decrease in L shown in Fig. 6.3.D). After 6 min the nucleoid begins to expand in both radial and axial dimension. In a normally growing cell, the axial dimension of the chromosomal DNA ends about 500 nm away from the tips of the cell; there is essentially no DNA in the end caps (11). On the other hand, the gap between the nucleoid and the inner membrane is about 60-80 nm (11, 20). Therefore the DNA has more room to expand axially than radially. The nucleoid radially expands back to the normal width during the next 20 minutes. The axial dimension expands quickly to fill almost the entire length of the cell. However, as we describe below, even in the maximally expanded state the bulk of the DNA avoids the inner membrane (Fig. 6.4). During the nucleoid expansion phase, the cell length also increases and then halts as the nucleoid elongation stops. The reason for this is not clear.

The effect of halting transcription with rifampicin addition on nucleoid morphology is also very similar from cell-to-cell. In appendix 6F we show the time course of radial dimension (W) from 4 different cells. The average effect of rifampicin on radial profile of nucleoid is shown in Fig. 6.A.6.B. As explained before, the different number of nucleoid sub-lobes in cells at different growth phase makes it difficult to compare the time course of axial profile changes from multiple cells.

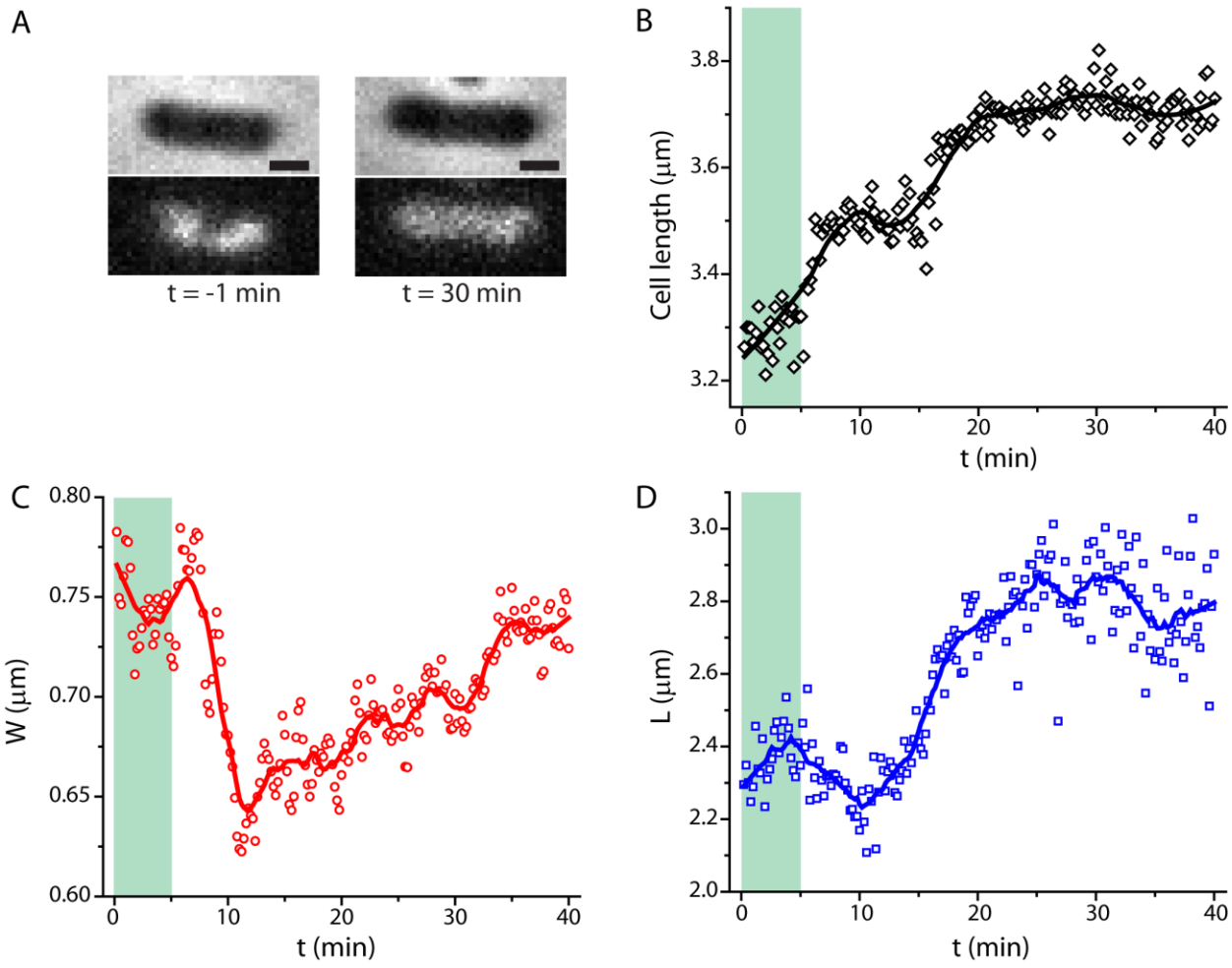

Figure 6.3: Effect of treatment with rifampicin. (A) Phase contrast image of a cell (top panel) and wide field images of Sytox Orange stained DNA before and 30 min after addition of 100 µg/mL rifampicin. Scale bar =1 µm. (B) The length of cell, measured from the phase contrast image, is plotted as a function of time. Rifampicin was added at 5 min. The cell length elongation stops after about 20 min after the injection. (C) Nucleoid width (W) vs time plot is shown. Within the first five minutes the nucleoid shrinks radially by ~ 20%, as observed in case of chloramphenicol treatment. After five minutes, the nucleoid starts to gradually expand back to the original width over the next 30 minutes. (D) The length (L) of nucleoid shrinks up to the first 5 minutes and then starts to increase until it almost fills up the entire cytoplasm. The green swath marks the time before injection. The smoothed interpolations are shown to guide the eyes.

Effect of nucleoid compaction on the distribution of proteins inside the cytoplasm

Compaction of the nucleoid after chloramphenicol addition should cause an increase in DNA density within the nucleoid. To see if the increased density of DNA can have an effect on the distribution of proteins, we examined the spatial distributions of different sized proteins after the action of chloramphenicol. The three proteins examined in this study are ribosomes (Mass = 2.5 MDa; radius ~20 nm), GFP (27 kDa; radius ~2.5 nm), and Kaede (110 kDa; radius ~4 nm) (21), which is a tetramer of GFP-like monomers. Wide field images of the ribosome and chromosomal DNA distributions in a cell, 30 minutes after chloramphenicol injection, are shown in Fig. 6.4.A. The DNA is compacted and appears as a sphere in the center of cell. The ribosomes occupy the space outside the DNA. Using superresolution microscopy of ribosome distribution after the injection of chloramphenicol, we have shown that the extent of segregation of ribosomes from the nucleoid increases after nucleoid compaction (11).

The axial intensity distributions of Kaede and GFP, before and after the injection of chloramphenicol, are shown in Fig. 6.4B and Fig. 6.4C. Prior to addition of chloramphenicol, the axial distribution of both Kaede and GFP appear to be consistent with the uniform distribution of the proteins within the cytoplasm. The axial intensity distribution of Kaede shows a dip near the mid-cell region after chloramphenicol addition. This suggests that Kaede is slightly excluded from the compacted nucleoid. The axial distribution of GFP after chloramphenicol addition stays qualitatively unchanged.

The experiments with Kaede and GFP were conducted without any DNA stain. To see if chloramphenicol has really caused compaction of nucleoid, we compared the phase contrast images of the cell taken before and after the injection. The phase contrast image of a cell with a DLL objective shows a distinctly lighter region in the middle representing the nucleoid and dark region for the ribosome-rich domains. Before the injection, the cells display two lighter regions, symbolizing two lobes of the nucleoid. 30 minutes after the injection of chloramphenicol, the phase contrast image shows only one lighter region in the middle, suggesting the drug did compact the nucleoid (Fig. 6.A.7.).

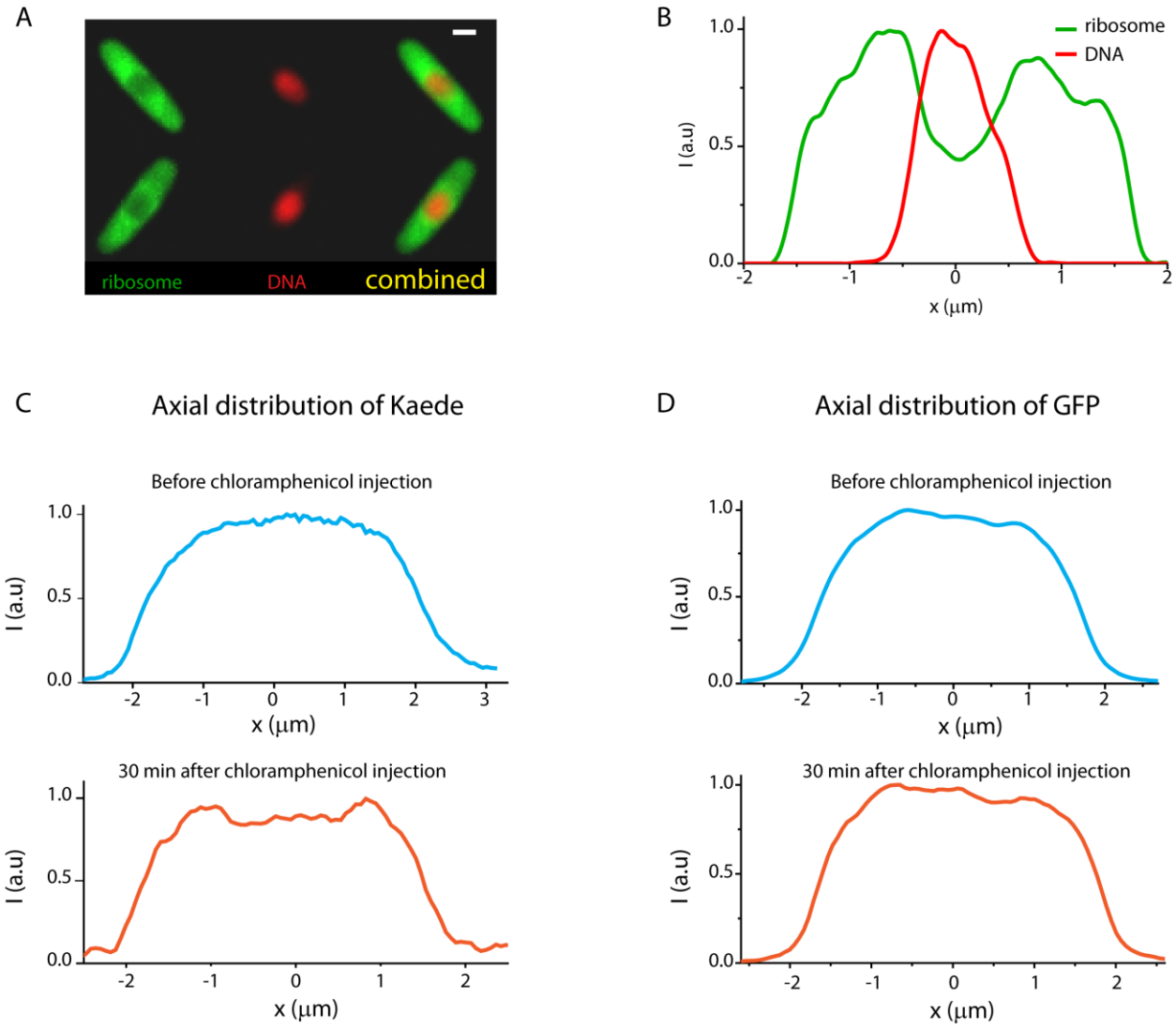

Figure 6.4: Effect of nucleoid compaction on the spatial distribution of proteins within the cytoplasm. (A) Wide field ribosome (S2-YFP) and DNA (DRAQ5) intensity distributions 30 min after addition of chloramphenicol. Scale bar = 1 μm. (B) Axial intensity distributions in the ribosome (green) and DNA channels (red) after 30 minutes of chloramphenicol treatment. (C) Axial intensity distribution of Kaede molecules before (top panel: blue) and 30 min after chloramphenicol addition (bottom panel: orange). (D) Axial intensity distribution of GFP before (top panel, blue) and after 30 min after chloramphenicol injection (bottom panel, orange).

Effects of rifampicin on the distribution of ribosomes

In our earlier work, we have demonstrated that in rifampicin-treated cells ribosomes exist predominantly as 30S subunits (11). Treatment with rifampicin is known to halt transcription. During the 30-min rifampicin treatment, we expect all 70S ribosomes to complete translation of their messages and dissociate into free 30S and 50S subunits. No new mRNA is produced during this time. Existing mRNA should degrade on a time scale of ~5 min. In the absence of any mRNA, the dominant YFP-labeled species should then be free 30S subunits. These 30S subunits are distributed homogeneously across the cytoplasm. Even though the DNA expands after rifampicin addition, comparison of axial and radial profiles of ribosome intensity distribution and DNA clearly demonstrates that the DNA stays away from the inner membrane (Fig. 6.5 B and C). As described before, the expansion of DNA meshwork allows ribosomes to diffuse throughout the cytoplasm (Chapter 3, Appendix 3M).

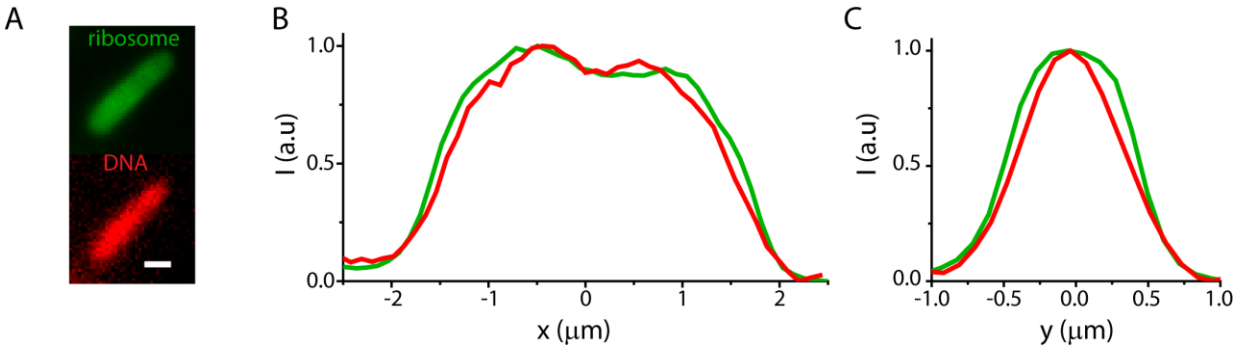

Figure 6.5: Effect of nucleoid expansion (after rifampicin addition) on the distribution and diffusion of ribosomes. (A) Wide-field ribosome (S2-YFP) and DNA (DRAQ5) spatial distributions 30 min after rifampicin addition. Scale bar =1 μm. (B) Axial intensity distributions for ribosomes and DNA after rifampicin injection. (C) Radial intensity distribution (along short axis, y) of ribosome and DNA after rifampicin addition.

Discussion

We know of no previous studies of the time-dependence of drug effects on nucleoid morphology. In this work, we have performed time-lapse microscopy of the chromosomal DNA, stained with DNA stain Sytox Orange at 6 sec time resolution. This allowed us to dissect the course of events during the action of chloramphenicol and rifampicin. We find that on a 5-7-min time scale both rifampicin and chloramphenicol lead nucleoid contraction, both radially and axially. The two sub-lobes of a nucleoid lobe also merge together during this time. Control experiments without drugs show that Sytox Orange alone does not affect growth rate or nucleoid morphology. We have also performed similar experiments with time-lapse imaging of HU-YFP, a protein that binds nonspecifically to the DNA randomly throughout the entire chromosome (22). The results of time-lapse imaging of spatial distribution of HU-YFP, as the cells are treated with chloramphenicol and rifampicin, are qualitatively similar to the results of the Sytox Orange study described here (data not shown).

Both drugs cause contraction of the width of the nucleoid by about 20% within 6 minutes of the drug injection, qualitatively consistent with the transertion hypothesis. Based on our superresolution studies of ribosome and RNAP imaging (11), and single molecule tracking (chapter 5), we have established that transertion should be able to provide a direct radially expanding force, but not a direct axially expanding force. Thus halting transertion can cause the observed dramatic radial shrinkage of the nucleoid. However, we also observe axial compaction of the nucleoid in similar timescale, for both of the drugs. Therefore it is also possible that transertion also provides an indirect force that stretches axially. In one hypothesis, transertion of proteins into the cytoplasmic membrane in effect pins the nucleoid to the membrane and pulls the nucleoid axially outward as the cell grows (6), allowing transertion to exert an indirect axially expanding force on the nucleoid. Our previously observed correlation of the axial distribution of nucleoid and the cell length is in accord with this hypothesis (11).

The effect of the drugs at long times (~30 minutes) can be specific to the processes involved in the drug action. During the 30 min rifampicin treatment, all the70S ribosomes complete the translation of their messages and dissociate into free 30S and 50S subunits. When transcription is halted with rifampicin, no new mRNA is produced, and all the existing mRNAs are degraded in the 30 minutes timescale (23). In the absence of any mRNA all the ribosomes exist as either 50S or 30S subunits. The conversion of ribosomes from polysomes (chains of ribosomes translating the same message) into monomeric subunits can in turn affect the morphology of the nucleoid. A recent Monte Carlo study modeled ribosomes as hard spheres and DNA polymers as branched freely jointed chains to understand ribosome DNA segregation in *E. coli* cells (24). The shape of the DNA was found to be elongated by about 30% of its original length when the polysomes were replaced with monomeric ribosomes. This is explained by the fact that polysomes prefer to stay near the end cap and monomers prefer to stay near the membrane along the radial dimension of the nucleoid.

This picture is qualitatively similar to the observed effect of rifampicin. The radial profile of the ribosomes in rifampicin treated cells indeed looks wider than that of DNA (Fig. 6.5.C). The axial profile of the nucleoid is also narrower than the axial profile of the ribosome distribution (Fig. 6.5.B). Evidently the DNA polymer continues to avoid the cytoplasmic boundary, in accordance with a simple physical model of plectonemic DNA (9). However, the degree of axial elongation of the nucleoid is much higher than what is predicted from the Monte Carlo study (24). It is possible that transcription has secondary effects such as DNA-supercoiling (9) that can also alter the nucleoid morphology.

The long term effect of chloramphenicol is further gradual compaction of the nucleoid along the axial dimension. The radial dimension of the nucleoid does not change much after the initial compaction. As chloramphenicol causes translating ribosomes to be unable to release from the mRNA, but does not inhibit a free 30S subunit to bind to an mRNA site and start translation, after sometime all the ribosome subunits of the cell should be in the form of polysomes. The tendency of the increased number of

polysomes to preferentially locate near the end cap can be responsible for the observed axial compaction (24).

We recognize that drugs seldom act on one biochemical element alone. It is possible that effects of chloramphenicol and rifampicin have unknown secondary components. In the future, we plan to perform similar studies with alternative drugs such as streptolydigin (a transcription elongation inhibitor) (25) and kasugamycin (a translation initiation inhibitor) (26) to corroborate inferences from the rifampicin and chloramphenicol results. It would be interesting to measure drug effect on nucleoid binding proteins such as HU, H-NS, and IHF, which are known to affect nucleoid morphology.

In a recent study the Goulian lab used a *tetO* array to place labels near specific genes for cytoplasmic proteins (mCherry and Adda) and membrane proteins (LacY and TetA-mCherry) under the control of the *lac* repressor.(10) Before induction, the distribution of punctal gene labels peaked near the cell center for all four genes. However, after induction of both membrane protein genes, the distribution of labels shifted significantly towards the cytoplasmic membrane. After induction of both cytoplasmic protein genes, the distribution of labels did not change. This behavior is qualitatively consistent with the transertion hypothesis. However, no attempt was made to track the movement of these loci before and after the induction and therefore the timescale of the reorganization of the chromosome with respect to the activation are not known. Following the position of the membrane gene loci as a function of time after their transcription or translation of the corresponding mRNA is inhibited by rifampicin or chloramphenicol, respectively, will allow us to correlate the spatio-temporal organization of individual gene locations with the overall morphology of the nucleoid.

Finally, we turn to the effect of nucleoid compaction on the protein distribution patterns within the cytoplasm of *E. coli*. The increased DNA density in the compacted nucleoid after chloramphenicol treatment caused proteins that are otherwise uniformly distributed to be segregated from the nucleoid and

localized preferentially towards the end cap. However, this effect is size dependent. Ribosomes that are well segregated from the nucleoid of a normal cell get even more pushed out of the nucleoid region after chloramphenicol mediated compaction (11). Kaede, a tetramer of GFP-like monomers (21), is uniformly distributed across the cytoplasm in a normal cell (16). However, after the nucleoid is compacted with chloramphenicol the distribution of Kaede seems to show some occlusion from the DNA rich region near the middle of the cell. GFP on the other hand stays uniform before and after the compaction of the nucleoid. Overall, the observed spatial distribution pattern of proteins of different sizes suggests a size dependent exclusion of proteins from the DNA-rich regions of the cell. Recent studies on the spatial distribution of protein aggregates in the *E. coli* cytoplasm suggested that nucleoid occlusion was the necessary and sufficient to explain the observed patterns of spatial distribution of protein aggregates (27). Our results are in accordance with the nucleoid occlusion theory of that study and provide evidence for the expected size dependence of the occlusion.

We also observe a minor effect of the nucleoid compaction on the shape of the cell. In chloramphenicol treated cells, the radial dimension of the cells look much wider than normal cells (Fig. 6.A.5). As membrane tethering via transertion can act as an expanding force on the nucleoid, it has been proposed that the DNA itself can exert force on the cytoplasmic membrane through transertion (28). This is Newton's second law. In the absence of transertion, and thereby absence of the pulling force exerted by the DNA, the membrane can expand. However, we need to quantify the fraction of membrane proteins that get exported via the transertion pathway, to understand the significance of this mechanism.

To see if an expanded nucleoid will have less occlusion effect, we also examined the spatial distribution of ribosomes after the cells were treated with rifampicin for 30 minutes. As expected, the ribosomes subunits are more or less uniformly distributed across the entire the cytoplasm (Fig.6.5.B-C) . Our previous pulse chase studies based on the photo-activation of a small subset of mEos-2 labeled

ribosomes suggested that the subunits are free to diffuse throughout the cytoplasm, and not localized to DNA pockets (Chapter 3, Appendix 3M).

Overall, the results of this study demonstrate that translation and transcription play an important role in controlling nucleoid morphology in *E. coli*. Transertion seems to be important in affecting the nucleoid segregation and radial expansion of the DNA. The net effect of transertion and other forces related/uncoupled to transcription and translation decides the shape and size of the nucleoid, which in turn affects the distribution of proteins in the cytoplasm.

Appendix 6A: Time-lapse images of DNA with chloramphenicol injection

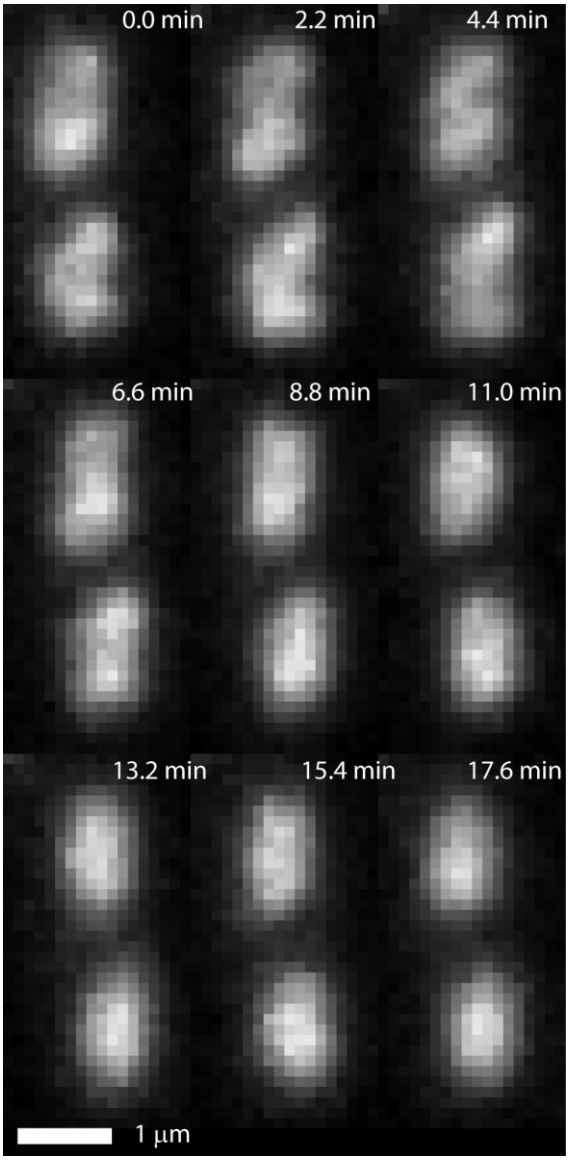

Figure 6.A.1: Montage of Sytox Orange stained nucleoid images from the cell described in Fig. 6.1 at different time points during the action of chloramphenicol.

Appendix 6B: Axial line scans of nucleoid after chloramphenicol injection

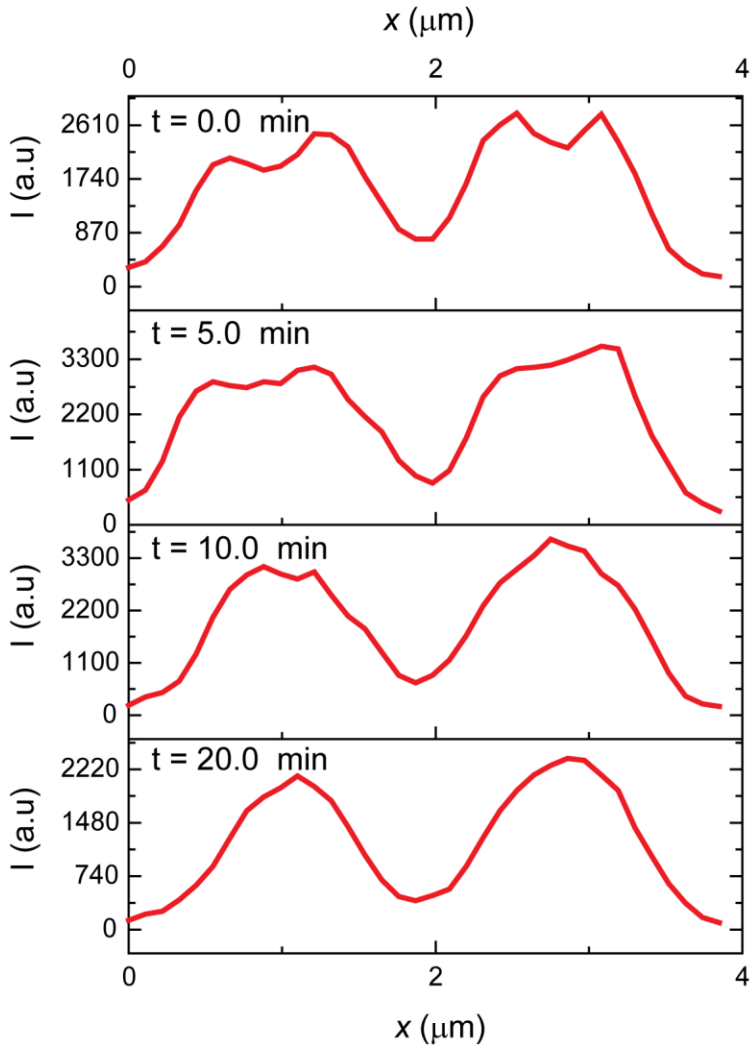

Figure 6.A.2: Axial line scans of Sytox Orange intensity at four different time-points during chloramphenicol effect. The axial profile in the top panel (t = 0 min) clearly demonstrates two sub-lobes within each nucleoid lobe. Within 5 min after chloramphenicol injection, the sub-lobes merge (t = 10 min) and they further compact at longer times (t = 20 min).

Appendix 6C: Transverse profile of nucleoid, before and after chloramphenicol injection

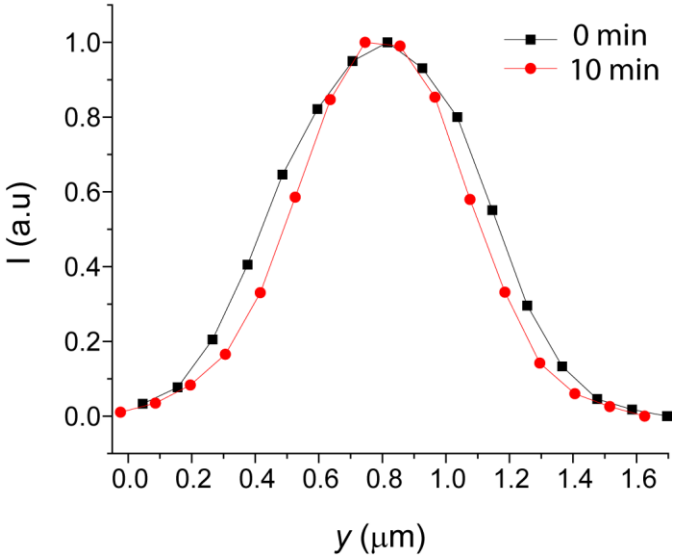

Fig. 6.A.3: Line scans of Sytox Orange intensity along the transverse axis of the nucleoid at two different time points.

Appendix 6D: Time course of changes in radial dimension of nucleoids (W) from 11 cells after chloramphenicol addition

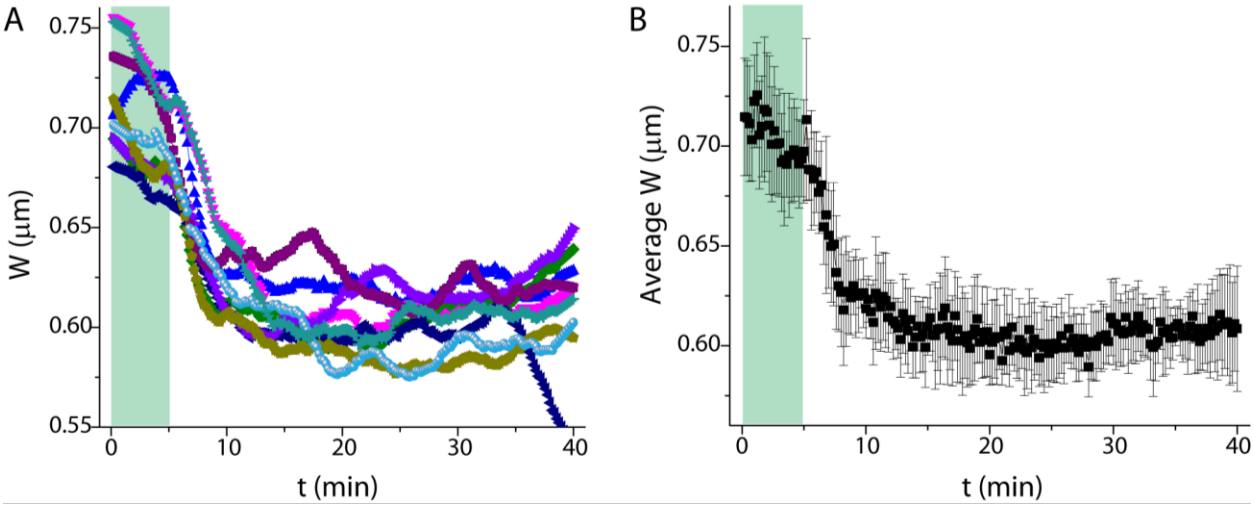

Figure 6.A.2: (A) Radial dimension (W) of the Sytox Orange signal from 11 different cells are plotted as function of time after chloramphenicol addition. The data is smoothed to see the trends for 11 cells clearly. (B) The average time course of those 11 cells (unsmoothed data) is shown. The error bars represent the standard deviation of W at each time point.

Appendix 6E: Time-lapse images of DNA with rifampicin injection

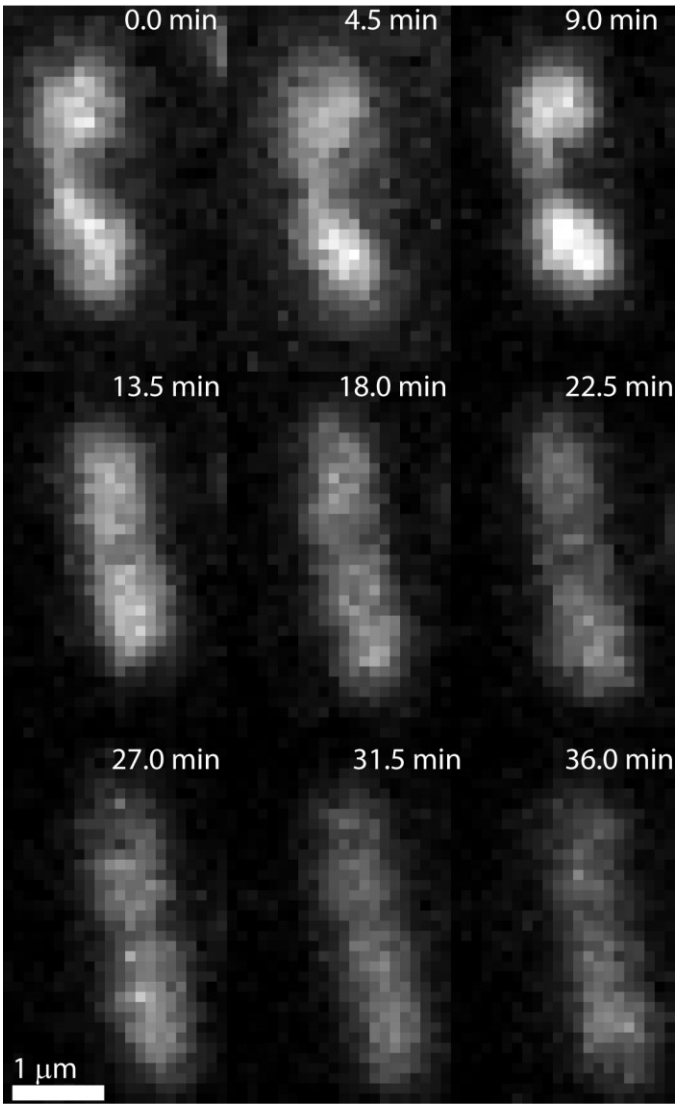

Figure 6.A.5: Montage of Sytox Orange stained nucleoid images from the cell described in Figure 6.2 at different time points after rifampicin injection.

Appendix 6F: Time course of radial dimension (W) from 4 different cells after injection of rifampicin

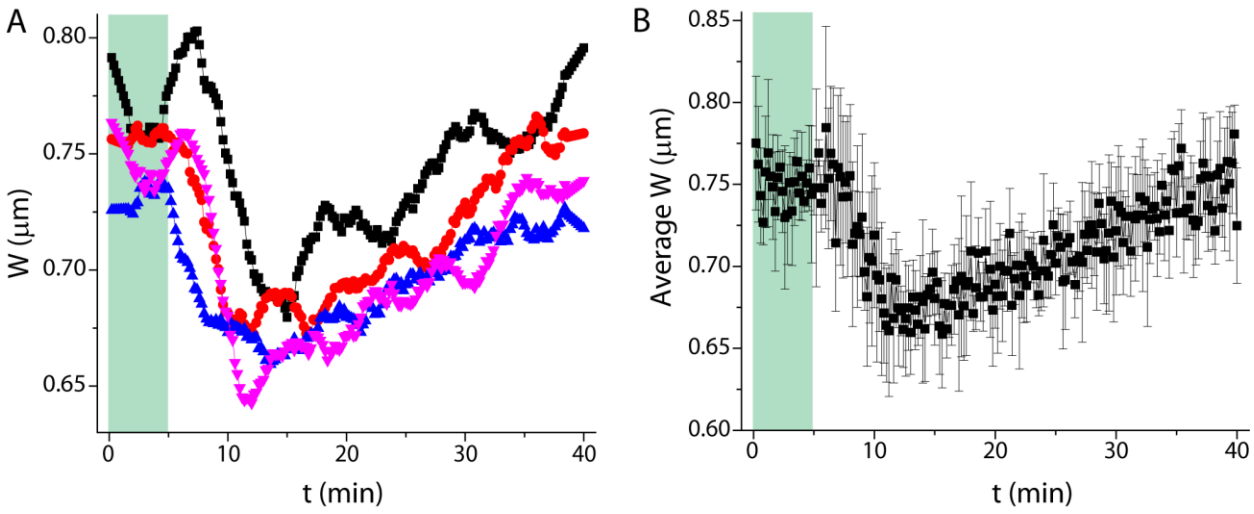

Figure 6.A.6: (A) Radial dimension (W) of the Sytox Orange signal from 4 different cells are plotted as function of time after rifampicin addition. (B) The average time course of those 4 cells is shown. The error bars represent the standard deviation of W at each time point. The Data is smoothed with adjacent averaging.

Appendix 6G: Phase contrast images of GFP-expressing cells before and after injection of chloramphenicol

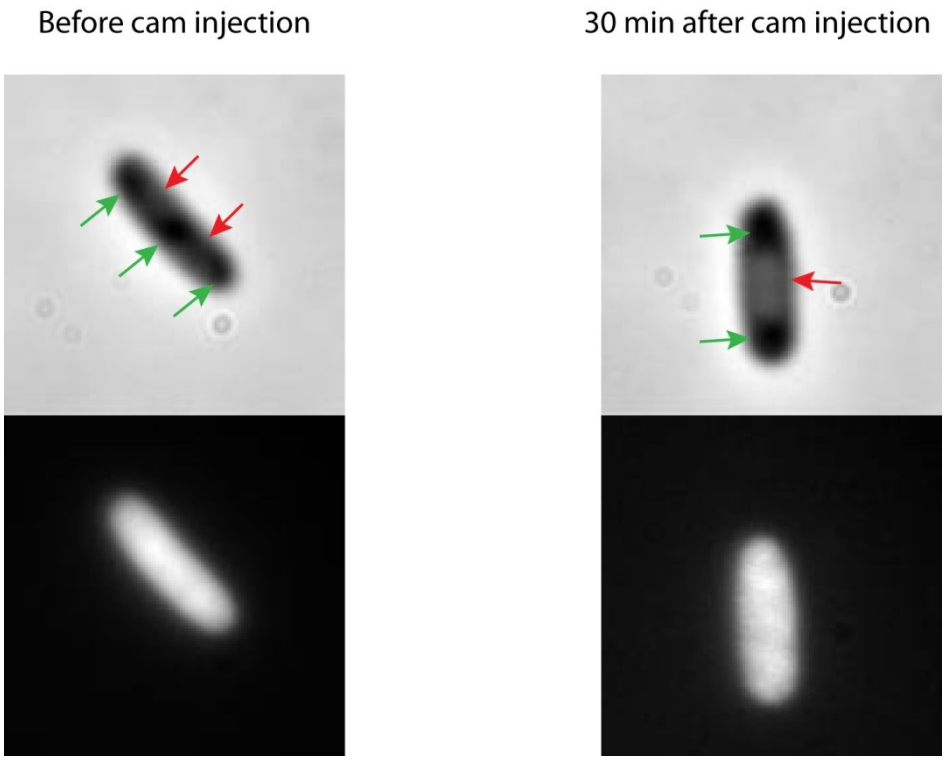

Figure 6.A.7: Phase contrast images of cells before and after chloramphenicol addition. The left panel shows the image of a normal GFP expressing cell. The top image is the phase contrast image of the cell (the red arrows indicate the two lighter region, representing the two nucleoid lobes, and the green arrows mark the darker regions corresponding to ribosome-rich regions), and the bottom image is GFP intensity distribution. In the right we show the phase contrast image and GFP intensity distribution of a cell treated with chloramphenicol for 30 minutes. The phase contrast image displays a single light region marked by the red arrows suggesting a single compacted nucleoid lobe.

References

1. Robinow, C., and E. Kellenberger. 1994. The bacterial nucleoid revisited. Microbiol. Rev. 58:211-232.

2. Wery, M., C. L. Woldringh, and J. Rouviere-Yaniv. 2001. HU-gfp and dapi co-localize on the *Escherichia coli* nucleoid. Biochimie. 83:193-200.

3. Valkenburg, J. A., and C. L. Woldringh. 1984. Phase separation between nucleoid and cytoplasm in *Escherichia coli* as defined by immersive refractometry. J. Bacteriol. 160:1151-1157.

4. Toan, N. M., D. Marenduzzo, P. R. Cook, and C. Micheletti. 2006. Depletion effects and loop formation in self-avoiding polymers. Physical Review Letters. 97.

5. Marenduzzo, D., K. Finan, and P. R. Cook. 2006. The depletion attraction: An underappreciated force driving cellular organization. J. Cell Biol. 175:681-686.

6. Woldringh, C. L. 2002. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Mol. Microbiol. 45:17-29.

7. Norris, V., and M. S. Madsen. 1995. Autocatalytic gene expression occursviatransertion and membrane domain formation and underlies differentiation in bacteria: A model. Journal of Molecular Biology. 253:739-748.

8. Woldringh, C. L., and N. Nanninga. 2006. Structural and physical aspects of bacterial chromosome segregation. J. Struct. Biol. 156:273-283.

9. Cabrera, J. E., C. Cagliero, S. Quan, C. L. Squires, and D. J. Jin. 2009. Active transcription of rrna operons condenses the nucleoid in *Escherichia coli*: Examining the effect of transcription on nucleoid structure in the absence of transertion. J. Bacteriol. 191:4180-4185.

10. Libby, E. A., M. Roggiani, and M. Goulian. 2012. Membrane protein expression triggers chromosomal locus repositioning in bacteria. Proc Natl Acad Sci U S A. 109:7445-7450.

11. Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol. Microbiol. 85:21-38.

12. Campbell, E. A., N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb, and S. A. Darst. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell. 104:901-912.

13. Rendi, R., and S. Ochoa. 1962. Effect of chloramphenicol on protein synthesis in cell-free preparations of *Escherichia coli*. The Journal of biological chemistry. 237:3711-3713.

14. Lovett, P. S. 1996. Translation attenuation regulation of chloramphenicol resistance in bacteria -- a review. Gene. 179:157-162

15. Binenbaum, Z., A. H. Parola, A. Zaritsky, and I. Fishov. 1999. Transcription- and translation-dependent changes in membrane dynamics in bacteria: Testing the transertion model for domain formation. Mol. Microbiol. 32:1173-1182.

16. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live *Escherichia coli*. Biophys. J. 101:2535-2544.

17. Konopka, M. C., I. A. Shkel, S. Cayley, M. T. Record, and J. C. Weisshaar. 2006. Crowding and confinement effects on protein diffusion in vivo. J. Bacteriol. 188:6115-6123.

18. Neidhard.Fc, P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. J. Bacteriol. 119:736-747.

19. Sochacki, K. A., K. J. Barns, R. Bucki, and J. C. Weisshaar. 2011. Real-time attack on single *Escherichia coli* cells by the HUman antimicrobial peptide ll-37. Proc. Natl. Acad. Sci. U. S. A. 108:E77-E81.

20. Bakshi, S., Renee M Dalrymple, Wenting Li, Heejun Choi, James C. Weisshaar 2013. Partitioning of RNA polymerase activity in live *Escherichia coli* from analysis of single-molecule diffusive trajectories.

21. Hayashi, I., H. Mizuno, K. I. Tong, T. Furuta, F. Tanaka, M. Yoshimura, A. Miyawaki, and M. Ikura. 2007. Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein kaede. Journal of Molecular Biology. 372:918-926.

22. Wang, W. Q., G. W. Li, C. Y. Chen, X. S. Xie, and X. W. Zhuang. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. Science. 333:1445-1449.

23. Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. Global analysis of mrna decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 99:9697-9702.

24. Mondal, J., B. P. Bratton, Y. Li, A. Yethiraj, and J. C. Weisshaar. 2011. Entropy-based mechanism of ribosome-nucleoid segregation in e. *coli* cells. Biophys. J. 100:2605-2613.

25. Pronin, S. V., and S. A. Kozmin. 2010. Synthesis of streptolydigin, a potent bacterial RNA polymerase inhibitor. Journal of the American Chemical Society. 132:14394-14396.

26. Tai, P.-C., B. J. Wallace, and B. D. Davis. 1973. Actions of aurintricarboxylate, kasugamycin, and pactamycin on *Escherichia coli* polysomes. Biochemistry. 12:616-620

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

28. Rabinovitch, A., A. Zaritsky, and M. Feingold. 2003. DNA-membrane interactions can localize bacterial cell center. Journal of Theoretical Biology. 225:493-496.

Chapter 7

Future directions

The measurements described in this thesis provide a new level of quantitative information on spatio-temporal organization of proteins that dictates important biological processes in live *E. coli* with a uniquely detailed and quantitative picture of how chromosomal DNA, RNAPs, and ribosomes work together in space and time (Fig 1A). Through our measurements, we answered a number of longstanding questions in the field. This in turns allowed asking for more detailed mechanistic picture of these processes. At the same time, we found some areas for improvement. In future, we will further improve our understanding of these biological processes using different labeling schemes and improved imaging and analytical tools. We plan to pursue the following lines of research in recent future:

1. The use of superresolution microscopy in exploring spatial biology inside bacteria is a relatively new field (1). As described in this dissertation and other published works (2-5), two-dimensional superresolution imaging of different cell components has significantly contributed to the improved understanding of organization within bacteria. However, biological systems are three-dimensional objects and therefore the two dimensional projection of the data has its limitations. As described in chapter 5, the radial distribution estimated by the distribution of *P(y')* severely underestimates the degree to which slow RNAPs tend to localize near the periphery of the nucleoid. The extent of segregation of ribosomes from chromosomal DNA is also strongly underestimated due to the same reason. Hence we will attempt to extend the measurement in three dimensions in future work. The new techniques of iPALM, biplane-FPALM and 3D-STORM (6-9) allow superresolution imaging in 3D with ~50 nm accuracy. 3D superresolution microscopy of transcription/translation machineries will provide better quantification of the three dimensional

distribution and dynamics. Implementing 3D techniques to localize and track chromosomal foci will improve the understanding of the 3D organization of the chromosome.

2. The relative distribution of transcribing RNAPs or translating ribosome with respect to the chromosomal DNA is estimated by comparing distributions of RNAP/ribosomes with HU distribution from different cells. We can improve such comparison using two-color 3D superresolution imaging (4). Localizing HU proteins and RNAP using two-color 3D PALM technique will be able give cell specific information regarding the relative distribution of transcribing RNAPs with respect to the chromosomal DNA. Similarly, comparison of the positions of slow ribosomes and that of a membrane stain from the same cell will allow quantitative examination of the membrane tethered ribosomes that are involved in transertion.
3. In our current study, we have emphasized on the average picture of distribution and dynamics of different cell components (chromosomal DNA, RNAPs, and ribosomes) to understand the features of spatial biology. However, recent experiments have provided evidence for significant cell-to-cell heterogeneity within a population of genetically identical cells. Examining the cell-to-cell variations in the spatio-temporal distributions of the chromosomal DNA, RNAPs, and ribosomes will provide a new level of detail.

4. The current chromosomal tracking studies use the ParB-*parS* labeling scheme to visualize the chromosomal foci. However, this labeling scheme could cause some artifacts to the motion or localization. ParB proteins are known to polymerize near the *parS* sequence, and therefore has limited control of the copy number of ParB bound to any *parS* site. Larger clusters of ParB labels are likely to slow down the motion of the DNA locus. Therefore, in future, we will examine different labeling strategies to measure similar phenomenon. One possible option is the use of fluorescent repressor operator system (FROS) to label the chromosomal foci (10). In this approach, a few (2-4) tandem copies of *tetO* operator sequences are inserted near a specific

chromosomal position. Controlled expression of GFP-labeled tetR from a plasmid has been used to visualize the positions of interest (5, 10). As only one tetR binds to a *tetO* with relatively strong binding constant (~ nM) we expect 2-4 tetR-GFP to bind the specific DNA locus. Therefore, this method can limit the artifacts associated with bulky cluster of labels at a DNA site.

5. In future, we can extend the timescale of observing the dynamics of RNAPs to 20-30 s using the diffusion blur based imaging method, described in chapter 5. This can allow us to make real time measurement of average search time and transcription time inside a live cell.

References

1. Cattoni, D. I., J. B. Fiche, and M. Noellmann. 2012. Single-molecule super-resolution imaging in bacteria. Current Opinion in Microbiology. 15:758-763.

2. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live escherichia coli. Biophys. J. 101:2535-2544.

3. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of kaede diffusion and spatial distribution in live escherichia coli. Biophys. J. 101:2535-2544.

4. Gahlmann, A., J. L. Ptacin, G. Grover, S. Quirin, A. R. S. von Diezmann, M. K. Lee, M. P. Backlund, L. Shapiro, R. Piestun, and W. E. Moerner. 2013. Quantitative multicolor subdiffraction imaging of bacterial protein ultrastructures in three dimensions. Nano Letters. 13:987-993.

5. Wang, W. Q., G. W. Li, C. Y. Chen, X. S. Xie, and X. W. Zhuang. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. Science. 333:1445-1449.

6. Shtengel, G., J. A. Galbraith, C. G. Galbraith, J. Lippincott-Schwartz, J. M. Gillette, S. Manley, R. Sougrat, C. M. Waterman, P. Kanchanawong, M. W. Davidson, R. D. Fetter, and H. F. Hess. 2009. Interferometric fluorescent super-resolution microscopy resolves 3d cellular ultrastructure. Proc. Natl. Acad. Sci. U. S. A. 106:3125-3130.

7. Juette, M. F., T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf. 2008. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. Nat. Meth. 5:527-529.

8. Huang, B., W. Wang, M. Bates, and X. Zhuang. 2008. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science. 319:810-813.

9. Pavani, S. R. P., M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. E. Moerner. 2009. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. Proc. Natl. Acad. Sci. U. S. A. 106:2995-2999.

10. Lau, I. F., S. R. Filipe, B. Soballe, O. A. Okstad, F. X. Barre, and D. J. Sherratt. 2003. Spatial and temporal organization of replicating escherichia coli chromosomes. Mol. Microbiol. 49:731-743.