Functional mapping of the components of the *E. coli* translational machinery

By

Sonisilpa Mohapatra

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 07/10/2018

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

James C. Weisshaar, Professor, Chemistry
Randall Goldsmith, Professor, Chemistry
Aaron Hoskins, Professor, Biochemistry
Tom Record, Professor, Chemistry
Jade Wang, Professor, Bacteriology

Functional mapping of the components of the *E. coli* translational machinery

Under the supervision of Professor James C. Weisshaar

At the University of Wisconsin-Madison

Summary

Bacterial cells generally lack membrane-bound organelles. Therefore, traditionally they were viewed as bags of freely diffusing macromolecules devoid of any internal organization. The development in imaging techniques in the last two decades has enabled better visualization of bacterial cells and led to careful investigation of their internal organization. Consequently, they are now shown to have an intricate internal organization of various proteins in their cytoplasmic space despite the absence of membrane boundaries. The spatiotemporal control of various cellular processes in bacteria is possibly regulated by this internal organization.

Most cellular processes are regulated by complex networks of intermolecular interactions. A systematic study quantifying the correlation in spatial organization of various biomolecules might help to unravel their underlying interactions. A positive correlation might suggest binding to a common site or common sites of production, degradation or action. A negative correlation might suggest a mechanism that sequesters the molecules from each other. A commonly used method to quantify linear correlation between images of two different species is the Pearson Correlation Coefficient (PCC). PCC however fails both qualitatively and

quantitatively in spherocylindrical cells such as *E. coli*. A modified PCC (MPCC) is proposed that corrects this issue by using a proper reference matrix for spherocylindrical cells (Chapter 2). The application of MPCC to experimental spatial distributions of RNA Polymerase and HU is demonstrated. A significantly positive correlation coefficient was obtained for the spatial distribution of the two DNA binding proteins. The MPCC concept could be generalizable to other cell shapes and can be used for both widefield and superresolution images.

Translation is one of the most important cellular process. Ribosomes are the central component of translation engaged in protein synthesis from mRNA. Optimal usage of ribosomes is dependent on their spatial organization. In this dissertation, the spatial and temporal organization of translation and a translation elongation protein, Elongation Factor P (EF-P) was investigated in live *E. coli* cells using superresolution fluorescence microscopy.

In rapidly growing *E. coli* and *B. subtilis*, strong nucleoid – ribosome segregation is observed. The picture of spatial organization in slowly growing *C. crescentus* is however completely different. The effect of growth rate on the segregation of ribosomes and nucleoids in *E. coli* is therefore investigated using two color superresolution fluorescence imaging in Chapter 3. MPCC was used to quantify the degree of segregation between spatial distribution of ribosomes and DNA in *E. coli* in the two different growth conditions. Slowly growing *E. coli* showed significant segregation of ribosomes from DNA, unlike slowly growing *C. crescentus*. Imaged ribosomes were classified as translating and non-translating based on their diffusive trajectories. The fraction of ribosomes engaged in translation in the two growth conditions was quantified. A small but significant reduction of translating ribosomes was observed in slowly growing cells. Simulated spatial distributions in spherocylinders of dimensions same as *E. coli* that best fit experimental spatial distributions of translating and non-translating ribosomes were

used to obtain a cellular map of ribosomes based on their biological functions. Based on the functional map of translating and non-translating ribosomes, it was concluded that a "circulation model" of ribosomes evidently applies to *E. coli* in all growth conditions.

In cells, the translation elongation is discontinuous, often suffering from periods of pauses. The cells have mechanisms to alleviate these pauses, irrespective of the reason for pausing. EF-P is a translation factor that alleviates pausing during translation of polyproline motifs. The spatial distribution and ribosome binding dynamics of EF-P were investigated in Chapter 4 to understand details of mechanism of interaction of EF-P with ribosomes during the translation elongation process. By localizing individual EF-P molecules, the spatial distribution of EF-P was shown to mimic the "three peaked" ribosome distribution in the cells. Photoactivated Localization Microscopy combined with single particle tracking (sptPALM) was employed to distinguish and quantify the different diffusive states of EF-P. The "slow" diffusing EF-P molecules are associated with translating ribosomes whereas the "fast" diffusing ones are EF-P copies searching for a potential binding site on ribosomes. Nearly 30% of EF-Ps are associated with translating ribosomes, implying 1500-6000 EF-P/ribosome complexes in a cell at any given time. This is significantly larger than the estimated 280 pausing events due to polyproline motifs being translated in the cell, suggesting interrogation of ribosomes by EF-P more frequently than the number of pausing events. In addition, faster imaging rates enabled measurement of binding/unbinding timescales of EF-P interactions with ribosomes (Chapter 4).

Overall, the reported works explore the synergism of interactions between different components of translation and their spatiotemporal organization. The famous biologist Jacques Monod noted "What was true for the *E. coli* would be true for the elephant". A better

understanding of the intracellular spatial organization of bacterial cells could elucidate the mechanisms that underlie spatial order in higher organisms.

Published work and work in preparation:

1 Mohapatra S, Choi H, Ge X, Sanyal S, Weisshaar JC. Spatial Distribution and Ribosome-Binding Dynamics of EF-P in Live *Escherichia coli*. mBio. 2017;8(3).

2 Mohapatra S, Weisshaar JC. Modified Pearson Correlation Coefficient for Two-color Imaging in Spherocylindrical Cells. bioRxiv. 2018.

3 Mohapatra S, Weisshaar JC. Functional mapping of the *E. coli* translational machinery using single-molecule tracking. *Submitted*

Acknowledgments

The past years in Weisshaar lab and Madison have been an enjoyable experience. I enjoyed it because of the time I shared with my amazing friends, co-operative lab mates and an extremely supportive advisor. This work is the result of collaborative effort, whether directly or indirectly, by a lot of people. And while there are far too many people to name individually, please understand that I truly appreciate your help and guidance in making the journey memorable.

Foremost, I would like to thank my advisor, Prof. James C. Weisshaar. His enthusiasm for science and meticulous attention to details are some of the qualities I hope to carry with me in my scientific journey. He provided me with academic freedom that allowed me to learn better. PhD can be an arduous journey at times. There have been moments when I have doubted myself or the outcome of a project, yet Jim would calmly say, “Hope springs eternal”. His calm demeanor in the face of difficulty is something I aspire to learn. In times of personal crisis, Jim has been very understanding. For all this support and much more, I am eternally grateful.

The Weisshaar lab is filled with wonderful people, some of whom are friends for life. I have had the good fortune of overlapping with and learning a great deal from: Somenath and Heejun. They continue to serve as my academic mentors away from the lab. I admire Somenath for his unwavering passion for science and optimism in life. I have learnt something new every time I conversed with him; be it about science or places to travel. I am especially thankful to Heejun for being a patient teacher, explaining intricate details of genetic engineering of *E. coli*. His dynamic attitude, temperament to learn and going all out to solve a scientific problem will inspire me. He has often had to lend an ear to my unnecessary ramblings at several points of

graduate school. I would like to thank Wenting for making me feel at home when I joined the lab and making sure I never went hungry. I thank Nikolai, Mainak and Yanyu for being excellent residents of 4225. We have shared uncountable discussions on science, movies, anime, books, politics, sports, Panda and CS, to name a few. It was a pleasure sharing my workspace with them. Zhilin and I joined the lab together. We have shared many adventures together which will be part of cherished memories. I am grateful to Ranga, Ken and Anurag for their camaraderie.

A vote of thanks to the members of Cavagnero, Goldsmith, Gourse, Hoskins, Record, Weibel, and Wang labs who have been kind enough to spare their time, loan me their equipments and chemicals in times of need. I would like to thank all members of the Chemistry Department of UW Madison who ensured that I had a smooth graduate school experience. A special thanks to our collaborator, Prof. Suparna Sanyal (Uppsala University) for providing strains necessary to proceed with my work. I would also like to thank Allen Buskirk (Johns Hopkins School of Medicine) and Sayar Karmakar (University of Chicago) for their insightful discussions that helped improve the impact of our work.

Others to whom I owe a great deal of thanks are Prof. Tom Record and Prof. Qiang Cui, both of whom have been members of my thesis committee through years of my graduate school and have provided helpful suggestions and constructive criticisms to enable me to be a better scientist. I am thankful to Prof. Randall Goldsmith, Prof. Aaron Hoskins, and Prof. Jade Wang for graciously agreeing to join my thesis committee for my final defense. I look forward to hearing their review and knowledgeable inputs on my work.

I am grateful to my undergraduate advisor, Prof. Arindam Ghosh and all other teachers at NISER, Bhubaneswar who introduced me to research and believed in me. Without their blessings and support, I would not have made it this far.

I came to USA about five years ago, not knowing a single soul in this foreign land. I would like to extend my sincere thanks to my roommates and friends, who have made Madison a home away from home. These five years would not have been nearly as memorable if not for the several trips to farmers market, drinks at the terrace, midnight gossips and long walks in the middle of the night. There are some special people who cheered me on throughout this journey across the oceans. I am blessed to have friends like, Aakrutee and Dibyanshu whose unconditional support and humorous antics kept me sane during my darkest days.

With due respect to everything and everyone in life, I think this journey would not be the same without Ved in it. He has been my best travel buddy, my punching bag, my toughest critic yet my biggest cheerleader. I feel grateful to have him in my life.

Lastly, I would like to thank my family for their unconditional support in all my life choices. My separation from my family is the highest cost I have paid for this education. This thesis is as much a result of their sacrifices as my hard work. My parents made sure I had access to the best of the facilities in all steps of my life. My father has always been my role model whose hard work, honesty, resilience, helpfulness, are qualities I hope to emulate. My mother is my best friend. Almost nothing in my life, would be possible without her. My sister has been my biggest source of strength, joy and happiness. She is the person who made sure everything was okay at home while I focus on my work here. I cannot thank her enough for always having my back.

Abbreviations

B. subtilis	*Bacillus subtilis*
C. crescentus	*Caulobacter crescentus*
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EF-P	Elongation Factor P
FISH	Fluorescence *in situ* hybridization
HU	Heat unstable protein
MPCC	Modified Pearson Correlation Coefficient
mRNA	Messenger ribonucleoic acid
PALM	Photoactivated localization microscopy
PAmcherrry	Photoactivatable mcherry
PCC	Pearson correlation coefficient
RNAP	RNA Polymerase
Spt	Single particle tracking
STORM	Stochastic optical reconstruction microscopy
YFP	Yellow fluorescent protein
3D	Three dimensional
2D	Two dimensional

Table of Contents

List of Figures

List of Supporting Figures

List of Tables

List of Supporting Tables

1

Introduction

Many fundamental biological processes in cells depend on complex networks of interactions between individual components (proteins, lipids, nucleic acids). These interactions provide spatial cues for the development of internal organization in compartments lacking bacteria. Spatial proximity of functionally related molecules in the crowded cell cytoplasm could contribute to the optimal efficiency of cellular processes (1). Quantitative investigations of the spatial organization of different cellular processes could therefore provide important insights into interactions between different components. In addition, spatial organization may regulate temporal dynamics of such interactions.

Much of our understanding of these interactions comes from *in vitro* experiments performed with purified proteins or cell lysates. Such experiments often poorly represent the physiological interactions that take place in the complex bacterial cell cytoplasm, crowded with macromolecules. Macromolecular crowding in the viscous cytoplasm can severely alter the rates and equilibria of intermolecular interactions. Fluorescence imaging methods have enabled direct, non-invasive imaging of dynamic processes in live cells. However, traditional fluorescence microscopy has a diffraction limited resolution of ~ 250 nm. This makes it ill-suited for studying subcellular protein localizations in tiny *E. coli* cells, typically 3-4 μm long and 0.8 -1 μm wide. Recently developed superresolution fluorescence imaging methods (2-4) have allowed localizations of individual proteins in single cells with a resolution of 30 – 50 nm, beyond the diffraction limit. This led to careful study of spatial organization of various molecules in the cytoplasmic space and their functional implications.

Comparing the spatial distribution of two molecules in the cytoplasm could provide information about their mutual interactions or their physicochemical similarities: hydrophobic molecules will partition into membranes, hydrophilic molecules to cytoplasm and ampiphilic

molecules are mostly found at interfaces. There are several methods to evaluate the spatial distribution of two acquired images (5-15). PCC (Pearson correlation coefficient) is commonly used to quantify the linear correlation between two spatial distributions; where +1 indicates perfect positive correlation, -1 indicates perfect anti-correlation and 0 reflects no linear correlation (8, 16). In Chapter 2, we have shown that in the case of spherocylindrical cells such as *E. coli*, PCC fails both qualitatively and quantitatively. In PCC, the two acquired images are compared with the constant mean value of respective images. The issue is that the projection of a randomly filled 3D spherocylinder is non-random in 2D. A modified version of PCC (MPCC) for spherocylindrical cells is proposed that corrects this issue by using the proper reference matrix for comparison with the two acquired images. The procedure should prove useful for both widefield and superresolution fluorescence microscopy. An application of MPCC to experimental spatial distributions of HU and RNAP in live *E.coli* is demonstrated.

The movement of molecules in the cytoplasmic space is a fundamental feature of biological processes. The observation and detailed analysis of molecular diffusion in live cells can add significant insight into our understanding of the nanoscale environment in which the molecules diffuse. Single particle tracking (spt) approaches are powerful tools to track individual protein copies in live cells and acquire their diffusive trajectories (17). These trajectories can be used to determine diffusion coefficient of individual molecules. The diffusive behavior of the imaged molecules is typically representative of their biological roles and binding state. Traditionally, analysis of diffusive trajectories is performed using mean square displacements of tracked particles (18). The mean diffusion coefficient of a population of tracked particles is estimated using the slope of the plot of mean square displacement, MSD (τ) vs. lag time (τ). This however does not provide any information about the heterogeneity of the population.

Distributions of single step displacements, *P(r)* from trajectories of multiple molecules acquired over several cells can reveal the heterogeneity and the dynamics of the population. Many previous studies of single-molecule diffusion fit the experimental *P*(*r*) distribution to a sum of analytical functions, with each component describing diffusion of the species in free space (19-22). However, diffusing molecules in *E. coli* often suffer from confinement due to the cell boundaries. Therefore, we relied on simulated random walk trajectories in model spherocylinders of dimensions the same as that of the imaged *E. coli* to account for confinement effects. The experimental distributions of single step displacements accompanied with simulated random walk trajectories in confinement help in quantifying the fraction of molecules in each diffusive state.

Translation is an essential cellular process through which cells synthesize proteins for different cellular functions. Bacteria devote a large fraction of their resources in synthesizing these proteins. Efficient regulation of translation is therefore crucial to cell growth. The spatial organization of translation in live bacteria plays an important role in the efficiency of the process. In this thesis, we answered key questions concerning the intracellular spatial and temporal organization of different components of translation in live *E. coli* using superresolution fluorescence microscopy. Antibiotic treatments and mutations targeting specific biochemical processes were employed to understand how different biochemical interactions shape the spatial organization of translational components in *E. coli*. The work presented in this dissertation provides an understanding of the role of spatial localization in bacteria, providing potential insights on how such localization mechanisms may be useful for efficient translation.

In *E. coli*, the DNA is compacted into a nucleoid (23). Ribosomes are shown to be excluded from the nucleoid (24). This picture of spatial segregation of translation from the

nucleoid region is seen in several other species of bacteria such as *B. subtilis* (25), that are phylogenetically far apart from *E. coli*. The high degree of DNA-ribosome and RNAP- ribosome segregation exhibited in the cells strongly suggests that most protein is translated from free mRNA messages that have diffused away from the nucleoids, indicating that most translation is uncoupled from transcription. In contrast, the spatial biology of *C. crescentus*, which grows very slowly, is strikingly different. The DNA and ribosomes are spread throughout the cell (26). To better understand the cause of this discrepancy, the spatial segregation of ribosomes and nucleoid was investigated in slow growing *E. coli* in Chapter 3. The doubling time, cell volume and DNA content of the studied slow growing *E. coli* was similar to that of *C. crescentus* to establish direct comparison. Two color superresolution fluorescence imaging of ribosomes and the nucleoid in the same cell provide their spatial organization in live *E. coli*. MPCC was used to quantify the degree of segregation in both growth conditions. Unlike *C. crescentus*, slow growing *E. coli* showed spatial segregation of ribosomes from the nucleoid region. The fraction of ribosomes engaged in active translation in both growth conditions is quantified. The fraction of translating ribosomes is reduced from 80% to 65% from fast to slow growing cells. In both growth conditions, the non-translating ribosomes can access the nucleoid while the translating ribosomes are mostly excluded from the nucleoid region.

Efficient regulation of ribosome usage is important for cell growth. It is functionally important that 30S and 50S subunits be able to penetrate the nucleoid meshwork rapidly to enable co-transcriptional translation within the nucleoid. Co-transcriptional translation has been shown to be important to prevent backtracking and premature termination of transcription by RNAP (27-29). Strong segregation of most translating 70S ribosomes from the DNA combined with the occurrence of co-transcriptional translation within the nucleoid requires that 30S and

50S ribosomal subunits circulate within the cell. The ribosomal subunit circulation model including co-transcriptional translation appears to be operative over the two different growth rates studied in Chapter 3.

In the cells, translation elongation is discontinuous, with periods of rapid peptide synthesis separated by pauses. The stalling of ribosomes during translation occurs for a variety of reasons and the cells have mechanisms to alleviate these pauses. Ribosome pausing can be mediated by translation of peptide moieties that contain polyproline motifs (30). EF-P is the only known translation factor that facilitates translation of polyproline motifs (31, 32). In Chapter 4, the role of EF-P in peptide bond formation using PALM and sptPALM was investigated. We compared the spatial distribution of EF-P and ribosomes in live *E. coli* under normal growth conditions as well as under the influence of translation and transcription halting drugs such as chloramphenicol and rifampicin respectively. Under normal growth conditions, EF-P and ribosomes share a similar characteristic three peaked distribution, with enrichment in the end-caps and the central region of the cell between the two nucleoid lobes. Under drug treatments, the distribution of EF-P also strikingly mimics that of ribosomes. The spatial distribution studies are suggestive of EF-P association to ribosomes. In this chapter, we also addressed whether the EF-P associations to ribosomes occurs selectively during translation of polyproline motifs. In *E. coli*, EF-P primarily exists in two diffusive states: slow diffusing, ribosome associated state and fast diffusing, searching state. We estimated that 30% of EF-Ps are associated to translating ribosomes in the cell. In the studied growth condition, the copy number of EF-P ranges from 5000-20,000 in the literature (33, 34). Thus, 1500-6000 EF-P/ribosome complexes were estimated to be present in the cell. This is much higher than the number of polyproline motifs being translated in the cell. Mechanistically, EF-P functions through binding to vacant ribosomal

E-sites (35). The frequent associations of EF-P with ribosomes suggests that E-site vacancies occur even when ribosomes are not paused. Frequent E-site vacancies provide evidence that translation elongation is a dynamic process in which translocation of P-site tRNA is not required for departure of E-site tRNA. Overall, the findings enable a better understanding of EF-P functionalities and binding mechanism with ribosomes.

References

1. Zimmerman SB, Trach SO. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. Journal of Molecular Biology. 1991;222(3):599-620.
2. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science. 2006;313(5793):1642.
3. Hess ST, Girirajan TPK, Mason MD. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. Biophys. J. 2006;91(11):4258-72.
4. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature Methods. 2006;3:793.
5. Andronov L, Orlov I, Lutz Y, Vonesch J-L, Klaholz BP. ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. Scientific Reports. 2016;6:24084.
6. Levet F, Hosy E, Kechkar A, Butler C, Beghin A, Choquet D, et al. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. Nature Methods. 2015;12:1065.
7. Malkusch S, Endesfelder U, Mondry J, Gelléri M, Verveer PJ, Heilemann M. Coordinate-based colocalization analysis of single-molecule localization microscopy data. Histochemistry and Cell Biology. 2012;137(1):1-10.
8. Manders EM, Stap J, Brakenhoff GJ, van Driel R, Aten JA. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. J Cell Sci. 1992;103(3):857.
9. Owen Dylan M, Rentero C, Rossy J, Magenau A, Williamson D, Rodriguez M, et al. PALM imaging and cluster analysis of protein heterogeneity at the cell surface. Journal of Biophotonics. 2010;3(7):446-54.
10. Pageon SV, Nicovich PR, Mollazade M, Tabarin T, Gaus K. Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data. Molecular Biology of the Cell. 2016;27(22):3627-36.

11. Ripley BD. Tests of `Randomness' for Spatial Point Patterns. Journal of the Royal Statistical Society Series B (Methodological). 1979;41(3):368-74.

12. Rossy J, Cohen E, Gaus K, Owen DM. Method for co-cluster analysis in multichannel single-molecule localisation data. Histochemistry and Cell Biology. 2014;141(6):605-12.

13. Sengupta P, Jovanovic-Talisman T, Skoko D, Renz M, Veatch SL, Lippincott-Schwartz J. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. Nature Methods. 2011;8(11):969-75.

14. Tarancón Díez L, Bönsch C, Malkusch S, Truan Z, Munteanu M, Heilemann M, et al. Coordinate-based co-localization-mediated analysis of arrestin clustering upon stimulation of the C–C chemokine receptor 5 with RANTES/CCL5 analogues. Histochemistry and Cell Biology. 2014;142(1):69-77.

15. Veatch SL, Machta BB, Shelby SA, Chiang EN, Holowka DA, Baird BA. Correlation Functions Quantify Super-Resolution Images and Estimate Apparent Clustering Due to Over-Counting. PLOS ONE. 2012;7(2):e31457.

16. Pearson K. Mathematical contributions to the theory of evolution III. Regression, heredity, and panmixia. Philos Trans R Soc Lond B Biol Sci. 1896;187:253.

17. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nature Methods. 2008;5:155.

18. Michalet X. Mean Square Displacement Analysis of Single-Particle Trajectories with Localization Error: Brownian Motion in Isotropic Medium. Physical review E, Statistical, nonlinear, and soft matter physics. 2010;82(4 Pt 1):041914-.

19. Bakshi S, Dalrymple RM, Li W, Choi H, Weisshaar JC. Partitioning of RNA polymerase activity in live *Escherichia coli* from analysis of single-molecule diffusive trajectories. Biophys. J. 2013;105(12):2676-86.

20. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. Mol Microbiol. 2012;85(1):21-38.

21. Stracy M, Jaciuk M, Uphoff S, Kapanidis AN, Nowotny M, Sherratt DJ, et al. Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living *Escherichia coli*. Nature Communications. 2016;7:12568.

22. Chen T-Y, Santiago AG, Jung W, Krzemiński Ł, Yang F, Martell DJ, et al. Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. Nature Communications. 2015;6:7445.

23. Kellenberger E. Functional consequences of improved structural information on bacterial nucleods. Research in Microbiology. 1991;142(2):229-38.

24. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution Imaging of Ribosomes and RNA Polymerase in Live *Escherichia coli* Cells. Molecular Microbiology. 2012;85(1):21-38.

25. Lewis PJ, Thaker SD, Errington J. Compartmentalization of transcription and translation in *Bacillus subtilis*. The EMBO Journal. 2000;19(4):710.

26. Montero Llopis P, Jackson AF, Sliusarenko O, Surovtsev I, Heinritz J, Emonet T, et al. Spatial organization of the flow of genetic information in bacteria. Nature. 2010;466:77.

27. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, et al. A NusE:NusG Complex Links Transcription and Translation. Science. 2010;328(5977):501.

28. McGary K, Nudler E. RNA polymerase and the ribosome: the close relationship. Current Opinion in Microbiology. 2013;16(2):112-7.

29. Proshkin S, Rahmouni AR, Mironov A, Nudler E. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. Science. 2010;328(5977):504.

30. Mohammad F, Woolstenhulme CJ, Green R, Buskirk AR. Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. Cell Reports. 2016;14(4):686-94.

31. Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV. EF-P Is Essential for Rapid Synthesis of Proteins Containing Consecutive Proline Residues. Science. 2013;339(6115):85.

32. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. Translation Elongation Factor EF-P Alleviates Ribosome Stalling at Polyproline Stretches. Science. 2013;339(6115):82.

33. An G, Glick BR, Friesen JD, Ganoza MC. Identification and quantitation of elongation factor EF-P in *Escherichia coli* cell-free extracts. Canadian Journal of Biochemistry. 1980;58(11):1312-4.

34. Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, et al. The quantitative and condition-dependent *Escherichia coli* proteome. Nature Biotechnology. 2015;34:104.

35. Doerfel Lili K, Rodnina Marina V. Elongation factor P: Function and effects on bacterial fitness. Biopolymers. 2013;99(11):837-45.

2

Modified Pearson Correlation Coefficient for Two-color Imaging in Spherocylindrical Cells

Reproduced verbatim from:

Mohapatra S, Weisshaar JC. Modified Pearson Correlation Coefficient for Two-color Imaging in Spherocylindrical Cells. bioRxiv. 2018.

Abstract

The revolution in fluorescence microscopy enables sub-diffraction-limit ("superresolution") localization of hundreds or thousands of copies of two differently labeled proteins in the same live cell. In typical experiments, fluorescence from the entire three-dimensional (3D) cell body is projected along the *z*-axis of the microscope to form a 2D image at the camera plane. For imaging of two different species, here denoted "red" and "green", a significant biological question is the extent to which the red and green spatial distributions are positively correlated, anti-correlated, or uncorrelated. A commonly used statistic for assessing the degree of linear correlation between two image matrices **R** and **G** is the Pearson Correlation Coefficient (PCC). PCC should vary from –1 (perfect anti-correlation) to 0 (no linear correlation) to +1 (perfect positive correlation). However, in the special case of spherocylindrical bacterial cells such as *E. coli* or *B. subtilis*, we show that the PCC fails both qualitatively and quantitatively. PCC returns the same +1 value for 2D projections of distributions that are either perfectly correlated in 3D or completely uncorrelated in 3D. The PCC also systematically underestimates the degree of anti-correlation between the projections of two perfectly anti-correlated 3D distributions. The problem is that the projection of a random spatial distribution within the 3D spherocylinder is non-random in 2D, whereas PCC compares every matrix element of **R** or **G** with the constant mean value $\bar{R}$ or $\bar{G}$. We propose a modified Pearson Correlation Coefficient (MPCC) that corrects this problem for spherocylindrical cell geometry by using the proper reference matrix for comparison with **R** and **G**. Correct behavior of MPCC is confirmed for a variety of numerical simulations and on experimental distributions of HU and RNA polymerase in live *E. coli* cells. The MPCC concept should be generalizable to other cell shapes.

Introduction

In widefield and superresolution fluorescence microscopy of eukaryotic and prokaryotic cells, the fluorescent species occupy a three-dimensional (3D) volume. In typical usage, the laser illuminates the entire thickness of the cell ("epi illumination"). The microscope then projects fluorescence from a 3D source along the *z* axis to form a two-dimensional (2D) image at the *xy* camera plane. For two-color imaging of two different species, herein called the "red species" and the "green species", an important biological question is the degree to which the red and green spatial distributions are positively correlated, anti-correlated, or uncorrelated with each other. Positive correlation may suggest binding to a common cytoplasmic element such as a membrane or the chromosomal DNA or common sites of production, action, or degradation. Negative correlation may suggest a physical or biochemical mechanism that sequesters red and green species from each other (1, 2).

The Pearson correlation coefficient (PCC) (3, 4) is one of the most commonly used statistical tools to measure the degree of linear correlation between two data sets X and Y:

$$\mathrm{PCC} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}. \qquad (2.1)$$

Here (x_i, y_i) are individual paired samples from the data sets X and Y and *n* is the total number of pairs; $\bar{x}$ and $\bar{y}$ are the mean values of the samples in data sets X and Y. With the advent of two-color superresolution fluorescence microscopy, the PCC is increasingly used as a statistic for quantifying the degree of correlation between the subcellular distributions of two distinguishable species.

For image matrices **R** (red channel) and **G** (green channel), the formula for PCC becomes:

$$\mathrm{PCC} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\bar{R})(G_{ij}-\bar{G})}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\bar{R})^2}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(G_{ij}-\bar{G})^2}}. \qquad (2.2)$$

Here *m* and *n* are the number of rows and columns in the image matrices; there are *m* x *n* total pixels in each image. The R_{ij} and G_{ij} are the corresponding intensities of pixel *ij* in **R** and **G**; for superresolution images these are integers (counts/pixel). $\bar{R}$ and $\bar{G}$ are the mean pixel intensities of **R** and **G**. In the PCC formula, all elements of the reference matrix with which **R** or **G** is compared have the same value. The value $\bar{R}$ (or $\bar{G}$) is subtracted from each individual pixel intensity R_{ij} (or G_{ij}), yielding both positive and negative difference intensities $(R_{ij} - \bar{R})$ and $(G_{ij} - \bar{G})$. Thus the product in the PCC numerator provides information about the correlation between deviations of R_{ij} from $\bar{R}$ and deviations of G_{ij} from $\bar{G}$. The denominator normalizes PCC so that it always lies in the range –1 to +1. Ideally, PCC = 1 indicates two perfectly linearly correlated images for which each red pixel *ij* deviates from the red mean in direct proportion to the deviation of the corresponding green pixel *ij* from the green mean. PCC = 0 indicates two linearly uncorrelated images. PCC = –1 indicates two perfectly anti-correlated images (red and green deviations of equal magnitude but of opposite sign). A PCC value significantly different from zero is a measure of the degree to which two distributions are correlated or anti-correlated as compared with the null hypothesis of PCC = 0, corresponding to two uncorrelated, random distributions.

The ImageJ software (5) extensively used for image analysis in the field of fluorescence microscopy provides Coloc2 and JaCoP plugins (6) that enable the user to calculate PCC between two images. In the recent literature, PCC has been used to characterize the correlation in 2D spatial distributions of two fluorescently labeled proteins in both bacterial cells (7-9) and

eukaryotic cells (10-16). McDonald and co-workers recently catalogued some common pitfalls in the use of PCC on eukaryotic cells (12).

We are particularly interested in small, rod-shaped, approximately spherocylindrical bacterial cells such as *E. coli* and *B. subtilis*, whose typical length is $L_{cell} \sim 4$ μm and whose diameter is $2r \sim 1$ μm. For the most common shapes of bacteria (spherical, rod-shaped and spiral), the standard PCC procedure fails both qualitatively and quantitatively. Spherocylinders have strong curvature at the two endcaps and in the cylindrical region. As a result, the projection of molecules randomly distributed in a 3D spherocylindrical volume does not form a random distribution in 2D. In Fig. 2.1, we illustrate the 2D projection of 5000 molecules that are distributed randomly in a 3D spherocylinder with dimensions similar to that of an *E. coli* cell in good growth conditions. The endcap regions and the edges of the spherocylinder project a smaller volume onto the camera plane, and thus have fewer counts/pixel in the 2D image than the central cylindrical region. This effect is clear in the pixelated 2D localization density maps shown in Fig. 2.1*C-E*. Pixels in the 2D projection of a random 3D distribution vary in intensity by a factor of five or more, depending on the chosen pixel size. The variations are highly systematic.

Consequently, the PCC reference matrix used for comparison with **R** and **G** is inappropriate. The PCC difference intensities $(R_{ij} - \bar{R})$ and $(G_{ij} - \bar{G})$ for pixels at the edges and end caps are both systematically negative, *i.e.*, strongly biased towards having fewer molecules/pixel than the mean value in a 2D projection of a 3D random distribution. In those regions, the products $(R_{ij} - \bar{R})(G_{ij} - \bar{G})$ are systematically positive. Similarly, the difference

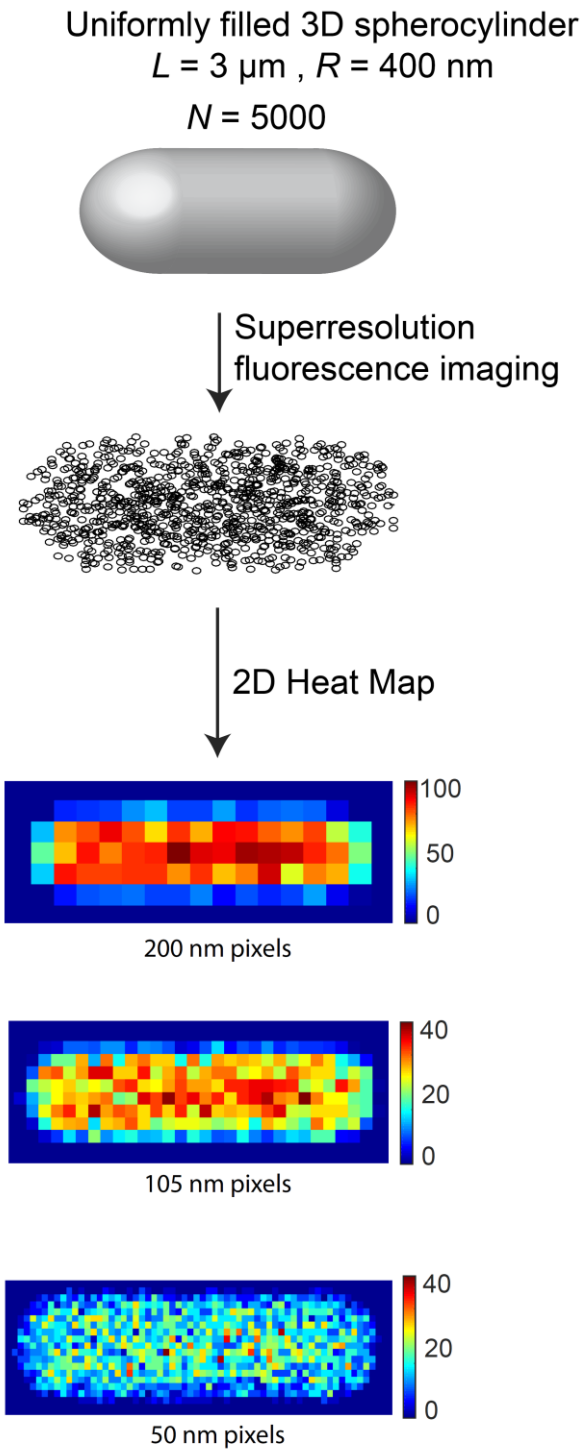


Fig. 2.1 Schematic of method for obtaining a 2D pixelated image from 3D distribution of molecules within a spherocylinder. (**A**) Uniformly filled spherocylinder representing a bacterial cell cytoplasm. (**B**) 2D projection of 5000 molecules distributed randomly in the 3D spherocylinder obtained by superresolution fluorescence imaging. (**C**) – (**E**) 2D localization probability density heat maps of imaged molecules with individual pixel sizes of 200 nm, 105 nm, and 50 nm.

intensities of the pixels in the central region of the spherocylinder are systematically positive, strongly biased towards having more molecules/pixel than the mean of a projection of a 3D random distribution. In that region, the products $(R_{ij} - \bar{R})(G_{ij} - \bar{G})$ are again systematically positive. For two uncorrelated, random distributions in 3D, this causes the traditional PCC of the 2D projection to incorrectly approach +1, not the desired result of zero. The same systematic positive bias causes the traditional PCC to underestimate the degree of anti-correlation between two perfectly anti-correlated images, as we will show.

In the following sections, we describe a procedure for calculating what we call the modified Pearson correlation coefficient (MPCC) in the special case of interest, spherocylindrical bacterial cells like *E. coli* and *B. subtilis*. The procedure should prove useful for both widefield and superresolution images, and in principle it could be adapted to other cell shapes. We use numerical simulations to show that MPCC properly becomes zero for two uncorrelated, random distributions, approaches –1 for two perfectly anti-correlated images, and approaches +1 for two perfectly correlated images. We also provide guidance for pixelation of superresolution images and show how to determine the probability *p* that a measured non-zero MPCC did not arise from two uncorrelated, random 3D distributions. We conclude with an experimental example of a significantly positive MPCC between superresolution images of RNA polymerase and of the DNA-binding protein HU in live *E. coli*. The package of MATLAB codes required for calculating MPCC between two different molecules imaged in rod shaped cells such as *E. coli* and *B. subtilis* is available on GitHub: *https://github.com/SoniMohapatra/MPCC*.

Results

The modified Pearson correlation coefficient MPCC

The MPCC of two images **R** and **G** is evaluated as follows:

$$\mathrm{MPCC} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\tilde{U}_{ij}^{R})(G_{ij}-\tilde{U}_{ij}^{G})}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\tilde{U}_{ij}^{R})^{2}}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(G_{ij}-\tilde{U}_{ij}^{G})^{2}}}. \qquad (2.3)$$

Here we have replaced $\bar{R}$ and $\bar{G}$ in Eq. 2.2 with the modified reference matrices $\tilde{U}_{ij}^{R}$ and $\tilde{U}_{ij}^{G}$, respectively. $\tilde{U}_{ij}^{R}$ and $\tilde{U}_{ij}^{G}$ denote the intensity of pixel *ij* in the 2D projection of a large set of molecules distributed randomly in a 3D spherocylinder. The total number of molecules in $\tilde{\mathbf{U}}^{R}$ and $\tilde{\mathbf{U}}^{G}$ has been scaled to be the same as the total number of molecules in **R** and **G**, respectively.

In favorable conditions, superresolution imaging provides (*x*,*y*) spatial localization of hundreds or thousands of molecules per cell with spatial resolution of $\sigma_{x,y} \sim$ 20–50 nm. Conversion of these single molecule locations into 2D probability density maps requires selection of a pixel size; several examples are shown in Fig. 2.1*C*-*E*. The intensity in each pixel equals the total number of molecules assigned to it. The dependence of the calculated MPCC on the chosen pixel size and the number of imaged molecules is described later. These pixelated 2D maps for the red and green channels are denoted by **R** and **G**, the image matrices in Eq. 2.3.

To form the numerator of Eq. 2.3, we then subtract $\tilde{\mathbf{U}}^{\mathbf{R}}$ and $\tilde{\mathbf{U}}^{\mathbf{G}}$ from the corresponding image matrix in the red and green channels (**R** and **G**, respectively) to obtain the (unnormalized) difference matrices $\mathbf{\Delta}^{\mathbf{R}}$ and $\mathbf{\Delta}^{\mathbf{G}}$. The resultant difference matrices have pixels with positive and negative values. Finally, to constrain MPCC to lie in the range +1 to -1, we normalize $\mathbf{\Delta}^{\mathbf{R}}$ and $\mathbf{\Delta}^{\mathbf{G}}$ so that the sum of the squares of individual pixel values in the difference matrix is 1. The resultant normalized 2D difference matrices are called $\hat{\mathbf{\Delta}}^{\mathbf{R}}$ and $\hat{\mathbf{\Delta}}^{\mathbf{G}}$ respectively. MPCC is

obtained by taking the Frobenius inner product of the two normalized matrices $\widehat{\boldsymbol{\Delta}}^{\mathbf{R}}$ and $\widehat{\boldsymbol{\Delta}}^{\mathbf{G}}$ (Eq. 2.6 in Methods). A detailed step-by-step description of the methodology for obtaining MPCC is presented in the Methods section.

The MPCC ranges from +1 to -1, as does standard PCC. The MPCC for two images is +1 when the normalized difference matrices are perfectly linearly related, *i.e.*, when $\hat{\Delta}_{ij}^{R} = \hat{\Delta}_{ij}^{G}$ for every pixel *ij*. As a result, $\mathrm{MPCC} = \sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\Delta}_{ij}^{R}\hat{\Delta}_{ij}^{G} = \sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\Delta}_{ij}^{R\,2} = +1$. The MPCC is -1 when the normalized difference matrices are perfectly inversely related to each other, *i.e.*, $\hat{\Delta}_{ij}^{R} = -\hat{\Delta}_{ij}^{G}$ for every pixel. As a result, $\mathrm{MPCC} = \sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\Delta}_{ij}^{R}\hat{\Delta}_{ij}^{G} = -\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\Delta}_{ij}^{R\,2} = -1$ .When the normalized difference matrices of two images are uncorrelated with each other, the MPCC is 0.

Next we carry out numerical simulations comparing MPCC with PCC for 2D projections from 3D spherocylinders in three cases: perfect 3D correlation that projects into perfect 2D correlation, perfect 3D anti-correlation that projects into perfect 2D anti-correlation, and uncorrelated, random 3D distributions. For all these examples, the **R** and **G** image matrices have 10,000 molecules each. The spherocylinder has tip-to-tip length L_{cell} = 3.5 μm and diameter 2r = 0.82 μm. The 2D pixel size in the image matrices **R** and **G** is chosen to be 200 nm in both dimensions, so that 75 pixels cover the 2D projection.

Perfect anti-correlation in 3D

To examine the case of two perfectly anti-correlated images, we have simulated 3D random distributions of 20,000 molecules confined to the spherocylindrical volume. The ~10,000 molecules located in the left half of the spherocylinder are designated red; the ~10,000 molecules located in the right half are designated green. This ensures that there is no spatial overlap of molecules in the red and green channels. We call this anti-

correlation Case I. For such strong spatial anti-correlation, we should expect MPCC = -1. An example of the corresponding 2D image matrices **R** and **G** is shown in Fig. 2.2*A*. In Fig. 2.2*B, C*, we have compared the reference matrices and the key normalized difference matrices the products of whose corresponding elements enter the traditional PCC (Eq. 2.2) and the new MPCC (Eq. 2.3).

For the traditional PCC (Fig. 2.2*B*), there are ~10,000 molecules of each color distributed in a cell area covering 75 pixels. As in Eq. 2.2, we subtract the mean pixel intensity $\bar{R}$ = 133.3 and $\bar{G}$ = 133.3 from each individual pixel intensities R_{ij} and G_{ij}. The resulting normalized difference matrices, $\frac{R_{ij}-\bar{R}}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\bar{R})^2}}$ and $\frac{G_{ij}-\bar{G}}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(G_{ij}-\bar{G})^2}}$, are depicted as heat maps labeled $\sim(\mathbf{R}-\bar{\mathbf{R}})$ and $\sim(\mathbf{G}-\bar{\mathbf{G}})$ in Fig. 2.2*B*. These are the PCC analogues of $\hat{\Delta}_{ij}^{R}$ and $\hat{\Delta}_{ij}^{G}$ in the MPCC equation. In the left half of the spherocylinder, the red difference matrix has a thin shell of systematically negative values (endcap and edge pixels) and a central core of systematically positive values. When multiplied by the corresponding elements of the left half of the green difference matrix, which contains all equal negative elements, the contributions to PCC will be positive and negative, respectively. The same type of systematically positive and negative contributions will arise from the right half of the spherocylinder. The resulting red and green contributions to PCC are not linearly anti-correlated. This is seen clearly in Fig. 2.2*D*, where we show a scatter plot of the individual red normalized differences vs the corresponding green normalized differences. The net result is PCC = –0.47, suggesting only partial anti-correlation of the two spatial distributions even though they are completely anti-correlated in both 3D and 2D.

In contrast, the MPCC formula of Eq. 2.3 subtracts from each pixel the proper 2D contribution of the projection of a smooth 3D random distribution (Fig 2.2*C*). The resulting

normalized difference matrices $\widehat{\boldsymbol{\Delta}}^{\mathbf{R}}$ and $\widehat{\boldsymbol{\Delta}}^{\mathbf{G}}$ are also depicted in Fig. 2.2*C*. The scatter plot of individual difference matrix elements $\hat{\Delta}_{ij}^{R}$ vs. $\hat{\Delta}_{ij}^{G}$ in Fig. 2.2*D* shows the expected strong linear anti-correlation for all pixels. The resulting MPCC is -0.99, very close to the expected value of –1.

In Appendix A2.1, we examine two additional examples of perfect anti-correlation. In anti-correlation Case II shown in Fig. A2.1 in the Appendix, the two endcap regions are occupied by ~10,000 red molecules and the central region is occupied by ~10,000 green molecules. Again, the normalized difference matrix elements are linearly anti-correlated and the calculated MPCC is –0.99. In anti-correlation Case III (Fig. A2.1), the ~10,000 red molecules occupy the leftmost 2/3 of the spherocylinder volume while the ~10,000 green molecules occupy the rightmost 1/3. The result is the same. The advantages of MPCC vs traditional PCC are apparent.

Perfect Positive Correlation in 3D and 2D

When the red and green 3D spatial distributions are perfectly positively correlated, so will be their 2D projections. As described before, an MPCC value of +1 is expected for a case of perfect correlation in the 2D projections. The same is true of the traditional PCC. To examine the case of two perfectly correlated images, we have simulated 3D random distributions of 20,000 molecules confined to the spherocylindrical volume. The ~10,000 molecules located in the left half of the spherocylinder are designated red; the molecules in the right half are deleted. We then independently simulated another 20,000 molecules distributed randomly in a 3D spherocylinder. The ~10,000 molecules located in the left half of the spherocylinder are designated green; the molecules in the right half are again deleted. The resulting 3D distributions are projected

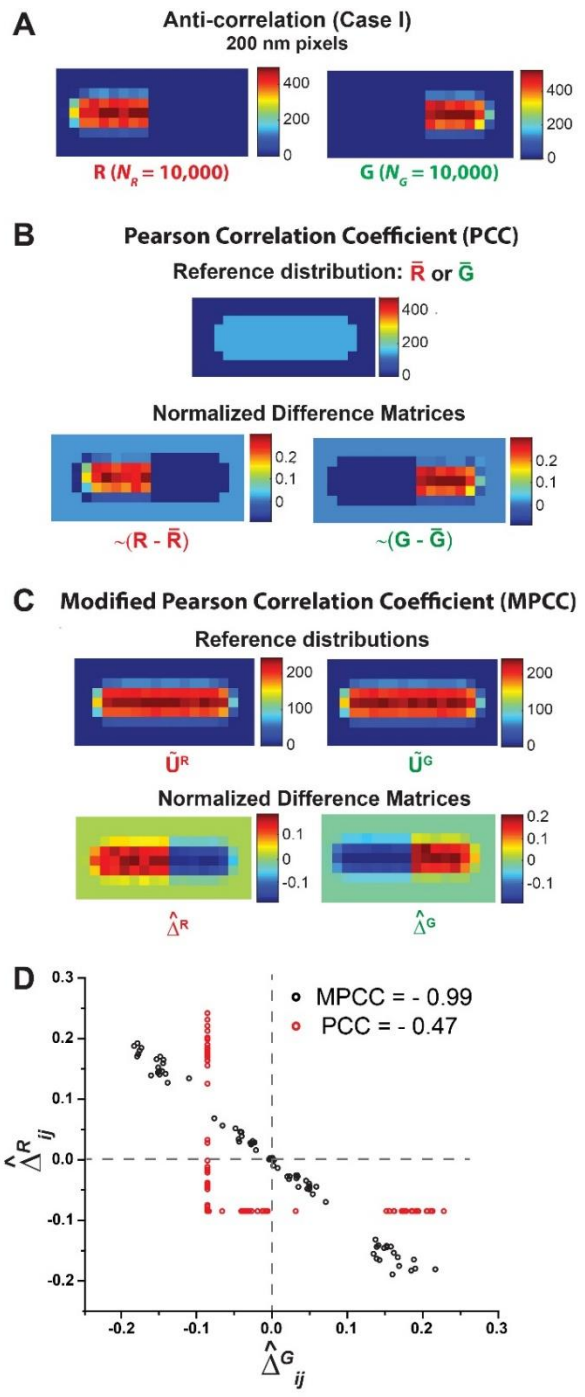


Fig. 2.2 Scheme for calculating PCC and MPCC for two representative images **R** and **G** that are perfectly anti-correlated in both 3D and 2D. (**A**) Heat maps of **R** and **G** with 200 nm pixels. Each image comprises ~10,000 molecules. Color scale indicates the number of molecules in each pixel. (**B**) Standard PCC calculation. *Top*: The 2D uniform reference distribution $\bar{\mathbf{R}}$ or $\bar{\mathbf{G}}$ that is subtracted from images **R** or **G**. *Bottom*: Normalized difference matrices $\sim(\mathbf{R} - \bar{\mathbf{R}})$ and $\sim(\mathbf{G} - \bar{\mathbf{G}})$ obtained after subtraction. The Frobenius inner product of these two difference matrices gives the PCC. (**C**) Modified PCC calculation. *Top*: Reference distribution $\tilde{\mathbf{U}}^{\mathbf{R}}$ and $\tilde{\mathbf{U}}^{\mathbf{G}}$, which are 2D projections of 3D random distributions of 100,000 molecules within the spherocylinder and normalized to have a total of 10,000 molecules. These are subtracted from images **R** and **G**, respectively. *Bottom*: Normalized difference matrices $\hat{\Delta}^{\mathbf{R}}$and $\hat{\Delta}^{\mathbf{G}}$ obtained after subtraction. The Frobenius inner product of these two difference matrices gives the MPCC. (**D**) Scatter plot of individual normalized difference matrix elements for PCC (*Red*) and for MPCC (*Black*). The MPCC elements are negatively correlated within the noise level, while the PCC elements are not. The resulting MPCC and PCC values are –0.99 and –0.47, respectively.

into 2D and pixelated to yield the image matrices depicted in Fig. A2.2*A* in the Appendix. We calculate the MPCC = +0.99 between these two distributions, very close to the anticipated value of 1. The resulting normalized difference matrices $\widehat{\mathbf{\Delta}}^{\mathbf{R}}$ and $\widehat{\mathbf{\Delta}}^{\mathbf{G}}$ obtained during evaluation of MPCC are depicted in Fig. A2.2*C*. The scatter plot of individual matrix elements $\hat{\Delta}_{ij}^{R}$ vs. $\hat{\Delta}_{ij}^{G}$ in Fig. A2.2*D* shows the expected strong linear correlation for all pixels. Similarly, the scatter plot of individual normalized difference matrix elements analogous to $\hat{\Delta}_{ij}^{R}$ vs. $\hat{\Delta}_{ij}^{G}$ for PCC in Fig. A2.2*D* shows the expected strong linear correlation for all pixels. If $R_{ij} = G_{ij}$ and $\bar{R} = \bar{G}$, then PCC = 1. Therefore, for two fluorescence images that are perfectly correlated in 3D, both the MPCC and the PCC will be +1 within the statistical noise.

Random Distributions in 3D

Two independent, uncorrelated, random distributions should have a Pearson correlation coefficient of 0 within the statistical noise. In the numerical tests, we have randomly distributed 10,000 red molecules and 10,000 green molecules in 3D within the spherocylinder. The two random distributions are generated independently, so we expect them to be uncorrelated with each other. We add appropriate localization errors σ_R = 50 nm and σ_G = 50 nm and then project the "measured" positions into the *xy*-plane. PCC and MPCC between the two 2D projection matrices (Fig. 2.3*A*) will be compared.

The resulting reference matrices and normalized difference matrices for PCC and for MPCC are depicted in Fig. 2.3*B* and *C* respectively. The scatter plots of $\hat{\Delta}_{ij}^{R}$ vs $\hat{\Delta}_{ij}^{G}$ for MPCC and of their analogues for PCC are shown in Fig. 2.3*D*. The data indeed appear uncorrelated for MPCC, but they are strongly positively correlated for PCC. The resulting calculated coefficients are MPCC = +0.10 and PCC = +0.98. The cause of the large, positive PCC value between two random 3D distributions was described in the

Introduction. The 2D projections have matching regions of systematically positive and systematically negative deviations from the 2D mean values.

Finally, we tested whether the distribution of calculated MPCC outcomes for two independent random distributions is appropriately centered at zero and unbiased towards positive or negative values. For 200 trials, we calculated MPCC values between two 2D projections of 3D independent, random distributions of 10,000 red and 10,000 green molecules using the same 200 nm pixel size. We fit the resulting distribution (Fig. A2.3 in the Appendix) to a Gaussian function. The mean of the best-fit Gaussian distribution is <MPCC> = +0.0041 and the standard error is σ_{MPCC} = 0.13. The mean is close to zero and the distribution is symmetric about zero, as hoped for. The probability that a particular trial would yield an MPCC of magnitude 0.10 or larger on either side of the Gaussian distribution is p = 0.44. The "measured" example MPCC of +0.10 (Fig. 2.3*D*) lies within 1σ of the mean; it was not a particularly unusual event.

Dependence of MPCC and its uncertainty on pixel size and total number of imaged molecules

Before evaluating MPCC between two superresolution images, the pixel size in the 2D localization density maps must be chosen. For a fixed cell size, the smaller the pixel size, the greater will be the total number of pixels N_p. We have shown in Appendix (Fig. A2.4 in the Appendix) that for a fixed number of localizations $N_R = N_G$ = 10,000 distributed randomly in 3D, as the pixel size decreases (and N_p increases) the width of the distribution of MPCC values becomes narrower. All the MPCC distributions for uncorrelated images are symmetric and centered about 0 and well fit by a Gaussian function. For these random, uncorrelated 3D distributions, the standard deviation of the Gaussian MPCC distributions scales as $N_p^{-1/2}$. This scaling holds even for N_R and N_G as small as 500.

Narrower widths of the MPCC distribution from random 3D distributions generally provide greater statistical confidence that a non-zero measured value of MPCC is significantly different from zero. This argues for fine pixelation. In practice, we suggest simulating the distribution of MPCC values between the 2D projections of 3D random distributions using the same number of molecules as were imaged in the red and green channels and the same pixel size chosen for **R** and **G**. This enables assignment of a probability *p* that the measured MPCC arose from two random 3D distributions. If *p* is unacceptably large, finer pixelation of both experimental and simulated locations may decrease *p*.

However, for non-random 3D distributions such as the completely anti-correlated distribution of Fig. 2.2 or the positively correlated distribution of Fig. A2.2, it is important not to pixelate so finely that the matrices **R** and **G** become too sparse. In the case of anti-correlated **R** and **G**, this leads to false positive linear correlations between $\hat{\Delta}_{ij}^{R}$ and $\hat{\Delta}_{ij}^{G}$. The zeroes appearing in the left-hand region of **R** positively correlate with the empty regions of **G**. Similarly, the zeroes arising due to sparseness in the right-hand region of **G** positively correlate with the empty regions of **R**. These systematically bias the MPCC for truly anti-correlated distributions towards more positive values, underestimating the degree of linear anti-correlation. We explore this effect numerically in Fig. A2.5 in the Appendix. For a given pixel size, the mean MPCC moves closer to the expected value of -1 for two anti-correlated images as the number of imaged molecules increases. In practice, we suggest carrying out numerical simulations of perfectly anti-correlated images using values of N_R and N_G that match experiment. The pixel size chosen for analysis of the experimental data should be the smallest pixel size for which

the mean MPCC for perfectly anti-correlated images is acceptably close to -1. In the numerical example of Fig. 2.2, with 10,000 molecules distributed over 75 pixels, the mean occupancy was 133 molecules/pixel, which yielded MPCC = –0.99. As a rule of thumb, it appears that if the mean occupancy is ~7 copies/pixel (~14 copies per pixel in the occupied halves of the case in Fig. 2.2), then the MPCC will be about –0.9.

For similar reasons, for two positively correlated images we expect that MPCC will systematically underestimate the degree of positive correlation as the red and green matrices become sparse. In the case of positively correlated **R** and **G** (Fig. A2.2), the zeroes appearing in the images due to sparseness are not positively correlated. The sparseness in number of molecules due to finer pixelation leads to false negative linear correlations between $\hat{\Delta}_{ij}^{R}$ and $\hat{\Delta}_{ij}^{G}$. This leads to systematic negative deviations of the calculated MPCC from the expected value of +1. We investigated the mean occupancy/pixel that is required for the calculated MPCC between strongly positively correlated images to be ~0.9, close to the expected value of +1. As shown in Fig. A2.5, a mean occupancy of ~7 copies/pixel yields MPCC values of about +0.9.

Based on our investigations on anti-correlated and positively correlated images described above, we estimate that the minimum mean occupancy should be ~7 copies/pixel for MPCC to give reasonably accurate results. The pixel size should be chosen keeping the minimum mean occupancy/pixel in mind.

Experimental example of MPCC from superresolution images of RNAP and HU in *E. coli*

To test our MPCC concept on real experimental data, we performed two-color superresolution fluorescence imaging of RNA polymerase and HU in live *E. coli* cells. RNAP is primarily located in the nucleoid region because of its frequent specific and

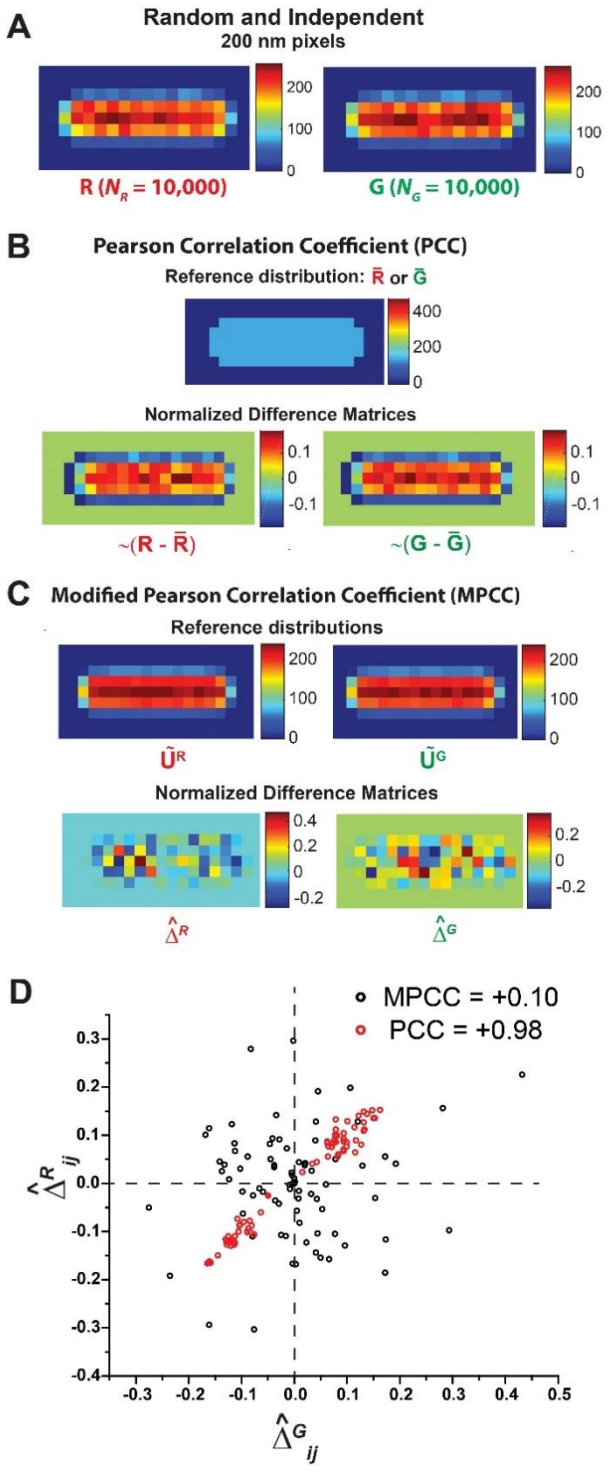


Fig. 2.3 Scheme for calculating PCC and MPCC for two representative projected images **R** and **G** arising from two random and independent distributions in 3D. (**A**) Heat maps of **R** and **G** with 200 nm pixels. Each image comprises ~10,000 molecules. Color scale indicates the number of molecules in each pixel. (**B**) Standard PCC calculation. *Top*: The 2D uniform reference distribution $\bar{\mathbf{R}}$ or $\bar{\mathbf{G}}$ that is subtracted from images R or G. *Bottom*: Normalized difference matrices $\sim(\mathbf{R} - \bar{\mathbf{R}})$ and $\sim(\mathbf{G} - \bar{\mathbf{G}})$ obtained after subtraction. (**C**) Modified PCC calculation. *Top*: Reference distribution $\tilde{\mathbf{U}}^{\mathbf{R}}$ and $\tilde{\mathbf{U}}^{\mathbf{G}}$, which are 2D projections of 3D random distributions of 100,000 molecules within the spherocylinder and normalized to have a total of 10,000 molecules. These are subtracted from images **R** and **G**, respectively. *Bottom*: Normalized difference matrices $\hat{\Delta}^{\mathbf{R}}$and $\hat{\Delta}^{\mathbf{G}}$ obtained after subtraction. (**D**) Scatter plot of individual normalized difference matrix elements for PCC (*Red*) and for MPCC (*Black*). The MPCC elements are randomly distributed, while the PCC elements are positively correlated. The resulting MPCC and PCC values are +0.10 and +0.98, respectively.

non-specific interactions with chromosomal DNA (17). HU is a DNA binding protein that should also localize within the nucleoids (18, 19). We expect significant positive correlation between the spatial distributions of RNAP and HU and therefore a positive value of MPCC.

For superresolution co-imaging of RNAP and HU in live *E. coli* cells, we constructed a strain where the gene coding for the fluorescent protein YFP (observed in the green channel) (20) is fused to the C terminus of the endogenous *rpoC* gene in *E. coli* VH1000. Single copies are imaged using the reversible photobleaching method described earlier (21). An inducible plasmid that expresses HU labeled with the photoactivatable fluorescent protein PAmcherry (22) (observed in the red channel) was introduced into the same strain. The cells were grown in EZ rich defined medium at 30°C, plated on a glass coverslip, and imaged with 30 ms exposure time. The details of strain construction, growth conditions, and imaging conditions are described in A2.3 in the Appendix.

To obtain a useful number of imaged copies, we combine locations of red HU and green RNAP molecules from different cells of essentially the same length. The imaged cells were sorted by tip-to-tip length based on phase contrast images in order to avoid broadening of spatial distribution of molecules due to the range of cell lengths. For the analysis, we chose cells of length 3.6 to 3.8 μm, the bin with the highest number of imaged cells. The resulting composite distribution of spatial localizations of N_G = 6570 RNAP-YFP and N_R = 8436 HU–PAmcherry molecules from 11 cells pixelated to 105 nm (279 total pixels) is illustrated in Fig. 2.4*A*. The mean number of molecules per pixel is ~25 and ~30 for the RNAP and HU channels respectively. The corresponding 1D projected axial distributions are compared in Fig. 2.4*B*. The raw data indeed suggest significant positive correlation between the two distributions.

For evaluation of MPCC we simulated two random distributions of 100,000 molecules each, corresponding to the RNAP (green) and HU (red) channels, using a spherocylinder whose dimensions match those of the chosen cells. The resulting reference images are normalized to have same number of molecules as imaged RNAP and HU. For accurate estimation of the cytoplasmic radius *r* of the imaged cells in the chosen length bin, we also imaged photoactivable Kaede molecules (23, 24), believed to distribute homogenously in the cytoplasmic volume (25). The detailed procedure is described in A2.4 in the Appendix. The resulting cell length is L_{cell} = 3.74 μm; the diameter is 2*r* = 0.82 μm (Fig. A2.6 in the Appendix). The two simulated 3D random distributions incorporated localization errors σ_{RNAP} = 38 nm and σ_{HU} = 60 nm, determined by the intercepts of MSD plots (Fig. A2.7 in the Appendix). We followed the procedure described above to calculate MPCC = +0.39. The scatter plot of $\hat{\Delta}_{ij}^{R}$ vs $\hat{\Delta}_{ij}^{G}$ (Fig. 2.4*C*) also indicates significant positive correlation.

The final step estimates the probability *p* that a value of MPCC = +0.39 or larger would be obtained from two random 3D distributions with the same number of imaged molecules and the same pixel size used for the experimental data. In Fig. 2.4*D*, we show a histogram of the outcomes of 200 such simulations. The best-fit Gaussian distribution has a mean value <MPCC> = -0.0030 and standard error σ_{MPCC} = 0.061. The measured MPCC value lies 6.4σ_{MPCC} away from zero. Under the assumption that the statistics of the simulated MPCC trials are Gaussian, the probability that two random 3D distributions would produce an MPCC value of magnitude 0.39 or larger on either side of the Gaussian curve is $p \sim 1.6 \times 10^{-10}$. Thus we can reject the null hypothesis that MPCC = +0.39 arose

from two random, uncorrelated 3D distributions and assert significant positive correlation between the RNAP and HU distributions with very high confidence.

The choice of pixel size does affect the calculated MPCC. For 200 nm pixels (N_p = 77 total pixels), the experimental MPCC is +0.46. The corresponding simulations of two random distributions gave <MPCC> = 0.0082 and σ_{MPCC} = 0.12. In this case, the probability that two 3D random distributions would produce an MPCC value of magnitude 0.46 or higher on either side of mean of Gaussian curve is $p \sim 1.3 \times 10^{-4}$. For 50 nm pixels ($N_p$ = 1178 total pixels), the experimental MPCC is +0.25. The corresponding simulations of two random distributions gave <MPCC> = 0.0027 and σ_{MPCC} = 0.033. In this case, the probability that two 3D random distributions would produce an MPCC value of magnitude 0.25 or higher on either side of mean of Gaussian curve is $p \sim 3.6 \times 10^{-14}$. The estimated experimental MPCC decreases systematically as N_p increases and the same data set is pixelated more finely, but the simulated σ_{MPCC} decreases more rapidly. As described previously, finer pixelation of both experimental and simulated locations helps to decrease p and reject the null hypothesis with greater confidence. In all three cases, the calculated MPCC lies many standard deviations away from zero. Thus we deem the conclusion of significant positive correlation between the RNAP and HU experimental distributions to be robust.

As suggested by the projected axial distribution of RNAP and HU (Fig. 2.4*B*), the two species are not completely correlated in space. There are several factors that may explain why the MPCC is significantly smaller than 1. We have averaged the data over 11 cells whose nucleoids have irregular shapes in 3D that are not axially symmetric and that vary from cell to cell. The limited number of molecules lends statistical fluctuations to each distribution. In addition, while RNAP and HU both bind to the DNA, they have

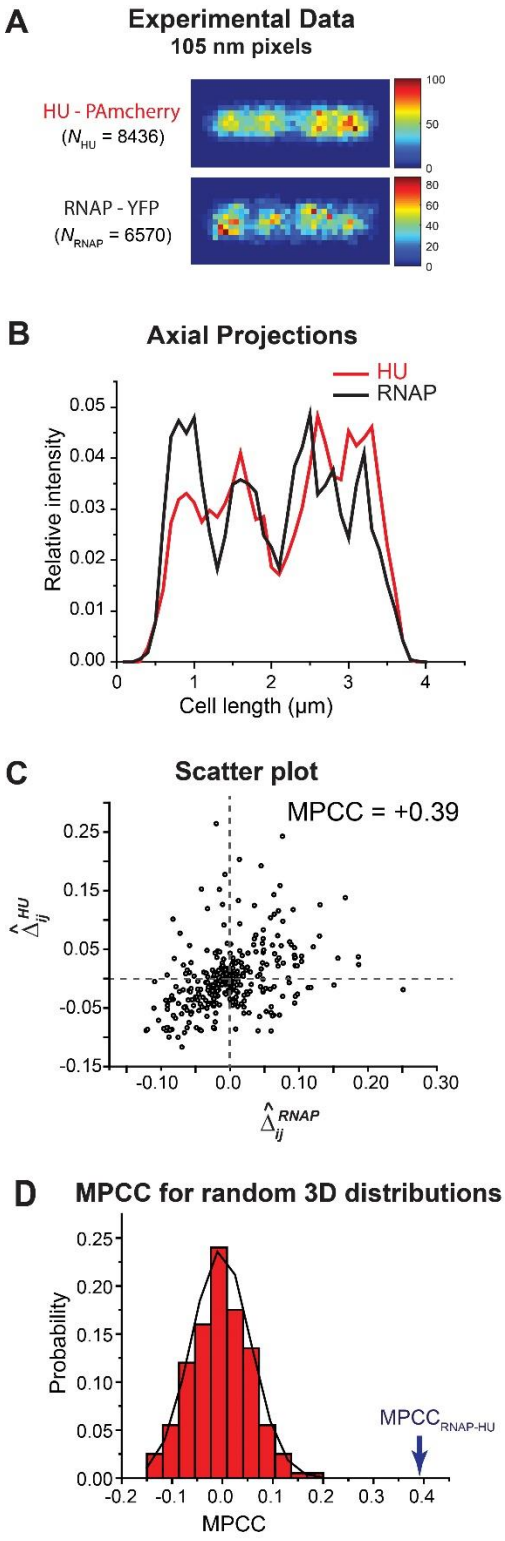


Fig. 2.4 (**A**) Experimental 2D localization probability density maps of 8436 HU–PAmcherry molecules (*Top*) and 6570 RNAP–YFP molecules (*Bottom*). Composite of data from 11 cells of tip-to-tip length L_{cell} in the range 3.6 to 3.8 μm. The color scale indicates the number of molecules in each pixel. (**B**) Axial probability density distributions of the imaged molecules. (**C**) Scatter plot of individual normalized difference matrix elements for MPCC, $\hat{\Delta}_{ij}^{HU}$ vs. $\hat{\Delta}_{ij}^{RNAP}$. Plot shows significant visual evidence of positive correlation; the calculated MPCC is +0.39. (**D**) Histogram of 200 MPCC values calculated for pairs of independent, random 3D distributions using the same number of HU and RNAP copies and the same pixelation as the experimental data. Best fit to a Gaussian curve has <MPCC> = –0.0030 and σ = 0.061 (*Black* curve). The experimental MPCC (arrow) lies at +6.4σ, making it highly improbable that two random distributions would produce such a large, positive result.

different biological functions and should not be expected to have spatial distributions that correlate perfectly.

Discussion

The Pearson correlation coefficient is one of the statistics commonly used for quantifying the degree of linear correlation in pixel-by-pixel intensity between two different images (3, 26-28). Owing to simplicity of usage and availability in most image analysis software packages (ImageJ, Colocalizer Pro), PCC is used increasingly in the literature of two-color fluorescence microscopy. Because it is pixel-based, PCC can be applied to both widefield and superresolution images. For sub-pixel resolution images, point-location methods such as Ripley's K (29-31) or pair correlation methods (32, 33) provide a global overview of the spatial distribution of points. These methods allow determination of whether the proteins are dispersed, clustered, or randomly distributed within the region of interest. All these methods have limitations for 2D projections of 3D locations and for small bacterial cells whose boundaries distort the meaning of random distributions. Moerner and co-workers have recently applied Ripley's K to the case of HU proteins in *C. crescentus* and corrected the reference random distribution by methods similar to those we employ here (34).

For two-color, three-dimensional fluorescence microscopy (35, 36), the PCC can provide an accurate measure of linear correlation, assuming the 3D image matrices are sufficiently populated. However, by far the more common case of two-color microscopy projects the 3D spatial distributions onto the 2D camera plane.

The central point of this work is simple. For most cell shapes, random 3D spatial distributions (no spatial correlations) do not make random 2D projections. In the particular case

of spherocylindrical cells, projections of random 3D distributions are skewed to have more molecules/pixel in the central region compared to the edges and the endcap regions (Fig. 2.1). This renders the standard PCC reference matrices (Eq. 2.2), whose elements are the constant values $\bar{R}$ and $\bar{G}$, highly inappropriate. As a result, the standard PCC fails both qualitatively and quantitatively to describe the nature and degree of the spatial correlation.

A calculated PCC value of +1 could equally well arise from perfectly correlated 3D images (Fig. A2.2) or from completely random 3D images (Fig. 2.3). For strongly anti-correlated images, the degree of anti-correlation will be systematically underestimated (Fig. 2.2).

In the special case of spherocylindrical cells, we have described a method for calculating a modified Pearson correlation coefficient (MPCC) that uses the 2D projection matrices $\widetilde{U}_{ij}^{R}$ and $\widetilde{U}_{ij}^{G}$ derived from independent 3D random distributions as the reference matrices with which the 2D image matrices **R** and **G** are compared (Eq. 2.3). The resulting MPCC is normalized to lie in the range –1 to +1. Within noise limitations, MPCC approaches 0 for the projections of two distributions that are independent and random in 3D, approaches –1 for two distributions that are perfectly anti-correlated in both 3D and 2D, and approaches +1 for two distributions that are perfectly positively correlated in both 3D and 2D. Additionally, we have used the new procedure to calculate a positive MPCC = +0.39 between experimentally obtained spatial localizations of individual RNAP and HU molecules in live *E. coli* cells (Fig. 2.4). Both RNAP and HU bind the chromosomal DNA, which occupies a subset of the cytoplasmic volume called the nucleoid. As expected, we obtain positive correlation that is significantly outside the range of model MPCC values computed for two uncorrelated distributions using the same pixel and copy number parameters as the experimental data.

While MPCC corrects a significant flaw in the standard PCC, it is important to note that for two images that are correlated or anti-correlated in 3D (*i.e.*, not random), the MPCC applied to the 2D projections will typically underestimate the degree of correlation or anti-correlation in 3D. Projection from 3D to 2D always involves a loss of information. If the two 3D distributions are correlated or anti-correlated, their 2D projections will typically be less so. Our model correlated images (Fig. A2.2) and anti-correlated images (Figs. 2.2 and A2.1) are special cases in that they preserve perfect correlation or anti-correlation when projected from 3D to 2D. More irregular, less symmetric 3D distributions generally will not. This means that a 2D MPCC value that is not significantly different from zero does not imply the absence of 3D spatial correlations.

We have also shown how a small average number of molecules per pixel leads to undesirable zeroes in the **R** and **G** matrices, causing systematic errors in MPCC values (Fig. A2.5). For both truly anti-correlated images and truly correlated images, this effect diminishes the magnitude of MPCC (biasing it towards zero). The MPCC user needs to measure sufficient numbers of localizations and to make a knowledgeable choice of pixel size. As a first estimate, we prescribe a minimum of 7 molecules per pixels to obtain a trustworthy MPCC.

In earlier work applying PCC to eukaryotic cells, Dunn, *et. al* (12) warned against inclusion of empty extracellular regions in the image matrices **R** and **G**. Such extra zeroes alter the mean value in the references matrices and also artificially inflate the calculated PCC due to positive correlations between the empty regions in both matrices. They suggested carefully outlining only the regions of space that are occupied by the cells of interest. However, the MPCC is impervious to such extra zeroes. The mean pixel intensity over the region of interest does not participate in the calculation of MPCC. The empty regions of the image outside the cell boundary cause corresponding zeroes in the 2D projected reference matrix. They affect neither the normalization

condition (Eq. 2.4 in Methods) nor the calculated MPCC (Eq. 2.3). In the MPCC procedure, one need not worry about empty regions of the image matrices that lie outside the projected cell boundaries.

The caveats outlined here apply to essentially all types of cells or organelles. In principle, the central concept of MPCC can be generalized to other cell geometries, including irregular eukaryotic shapes. However, the MPCC will probably find its greatest use in bacterial cells, whose shapes are often quite uniform for given growth conditions. It is relatively straightforward to simulate appropriate 2D reference distributions for rod-shaped bacteria like *E. coli* and *B. subtilis*, using a spherocylinder as the simplified model. The problem becomes more difficult for other shapes, such as the spiral shaped *H. pylori*.

One possible purely experimental solution would be to co-image a large population of freely diffusing fluorophores that presumably map out the 2D projection of a 3D random distribution in the cell volume of interest. To test this concept on *E. coli*, we have imaged Kaede under the same growth conditions used to image the RNAP and HU spatial distributions of Fig. 2.4. Kaede is a non-native tetrameric fluorescent protein that diffuses freely in *E. coli* and appears to fill the cytoplasm uniformly (25). We imaged Kaede in 15 cells of length 3.6 to 3.8 μm, the same length bin used for RNAP and HU. The composite distribution of 54,719 spatial localizations from 8 of the 15 cells was pixelated to give an experimental estimate of $\widetilde{\mathbf{U}}^{\mathbf{R}}$. An estimate of $\widetilde{\mathbf{U}}^{\mathbf{G}}$ was generated from the pixelated 2D projection of 66,301 Kaede copies from the other 7 cells. Using these experimentally generated matrices $\widetilde{\mathbf{U}}^{\mathbf{R}}$ and $\widetilde{\mathbf{U}}^{\mathbf{G}}$, we calculated MPCC for the same RNAP and HU spatial distributions to be 0.56, 0.42 and 0.32 for chosen pixel sizes of 200 nm,

105 nm and 50 nm respectively. These completely experimentally derived MPCC values are similar to the MPCC values of 0.46, 0.39 and 0.25 obtained from simulation of $\widetilde{\mathbf{U}}^{\mathbf{R}}$ and $\widetilde{\mathbf{U}}^{\mathbf{G}}$ for the same respective pixel sizes. For cases in which it is difficult to simulate the 3D cell geometry, the experimental approach to generation of the reference matrices may prove useful.

Methods

As a first step towards generation of the matrices $\widetilde{\mathbf{U}}^{R}$and $\widetilde{\mathbf{U}}^{G}$in Eq. 2.3, a large number of molecules (here 100,000) are randomly distributed in a spherocylinder whose dimensions match those of the cells being imaged. We want $\widetilde{\mathbf{U}}^{R}$and $\widetilde{\mathbf{U}}^{G}$to have high signal-to-noise in each pixel. For a cell of length 3.5 μm and width of 0.82 μm, the choice of 200 nm for the pixel size results in 75 pixels in the cell. 100,000 molecules makes the mean occupancy 1333 molecules/pixel. An appropriate localization error σ is applied to each particle location in both *x* and *y* coordinates by sampling a Gaussian distribution with standard deviation σ, yielding the model "measured" location of each molecule, which is binned appropriately. For generating 3D random distribution of molecules corresponding to the red and green channels, the localization error applied is the same as that measured upon imaging fluorescent molecules in red (σ_R) and green channels (σ_G) respectively. The 2D projections along the *z* axis of these two 3D random distributions give the matrices $\mathbf{U}^{\mathbf{R}}$ and $\mathbf{U}^{\mathbf{G}}$. The elements U_{ij}^{R} and U_{ij}^{G} are positive integers.

Next the counts in individual pixels of $\mathbf{U}^{\mathbf{R}}$ and $\mathbf{U}^{\mathbf{G}}$ are normalized so that the total number of red and green molecules is equal to N_R and N_G, the total number of molecules imaged in each channel. This yields the normalized matrix $\widetilde{\mathbf{U}}^{\mathbf{R}}$:

$$\widetilde{U}_{ij}^{R} = U_{ij}^{R} \times N_R/100{,}000 \qquad (2.4)$$

so that $\sum_{i=1}^{m}\sum_{j=1}^{n}\tilde{U}_{ij}^{R} = N_R$. Similarly, $\mathbf{U^G}$ is normalized so that the sum of all elements of $\tilde{\mathbf{U}}^{\mathbf{G}}$ is N_G. We then subtracted $\tilde{\mathbf{U}}^{R}$and $\tilde{\mathbf{U}}^{G}$ from the corresponding image matrix under analysis, **R** and **G** respectively to obtain the unnormalized difference matrices $\mathbf{\Delta^R}$ and $\mathbf{\Delta^G}$. We normalized $\mathbf{\Delta^R}$ and $\mathbf{\Delta^G}$ so that the sum of the squares of individual pixel values in the difference matrix is 1:

$$\hat{\Delta}_{ij}^{R} = \frac{\Delta_{ij}^{R}}{\|\mathbf{\Delta^R}\|},$$

(2.5)

where $\|\mathbf{\Delta^R}\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}{\Delta_{ij}^{R}}^{2}}$. The resultant normalized 2D difference matrix $\widehat{\mathbf{\Delta}}^{\mathbf{R}}$ has $\left\|\widehat{\mathbf{\Delta}}^{\mathbf{R}}\right\| = 1$. The difference matrix $\mathbf{\Delta^G}$ in the green channel is similarly normalized to obtain $\widehat{\mathbf{\Delta}}^{\mathbf{G}}$ such that $\left\|\widehat{\mathbf{\Delta}}^{\mathbf{G}}\right\| = 1$. MPCC is obtained by taking the Frobenius inner product of the two normalized matrices $\widehat{\mathbf{\Delta}}^{\mathbf{R}}$ and $\widehat{\mathbf{\Delta}}^{\mathbf{G}}$:

$$\text{MPCC} = \sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\Delta}_{ij}^{R}\hat{\Delta}_{ij}^{G}.$$

(2.6)

References

1. Mondal J, Bratton BP, Li Y, Yethiraj A, Weisshaar James C. Entropy-Based Mechanism of Ribosome-Nucleoid Segregation in *E. coli* Cells. Biophys. J. 2011;100(11):2605-13.
2. Neeli-Venkata R, Martikainen A, Gupta A, Gonçalves N, Fonseca J, Ribeiro AS. Robustness of the Process of Nucleoid Exclusion of Protein Aggregates in *Escherichia coli*. J Bacteriol. 2016;198(6):898-906.
3. Manders EM, Stap J, Brakenhoff GJ, van Driel R, Aten JA. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. J Cell Sci. 1992;103(3):857.
4. Pearson K. Mathematical contributions to the theory of evolution III. Regression, heredity, and panmixia. Philos Trans R Soc Lond B Biol Sci. 1896;187:253.

5. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5.

6. Bolte S, CordeliÈRes FP. A guided tour into subcellular colocalization analysis in light microscopy. J Microsc. 2006;224(3):213-32.

7. Karunatilaka KS, Cameron EA, Martens EC, Koropatkin NM, Biteen JS. Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts. MBio. 2014;5(6):e02172-14.

8. Männik J, Wu F, Hol FJH, Bisicchia P, Sherratt DJ, Keymer JE, et al. Robustness and accuracy of cell division in *Escherichia coli* in diverse cell shapes. Proc Natl Acad Sci U S A. 2012;109(18):6957-62.

9. Strahl H, Bürmann F, Hamoen LW. The actin homologue MreB organizes the bacterial cell membrane. Nat Commun. 2014;5:3442.

10. Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. Biophys. J. 2004;86(6):3993-4003.

11. Dedecker P, Mo GCH, Dertinger T, Zhang J. Widely accessible method for superresolution fluorescence imaging of living systems. Proc Natl Acad Sci U S A. 2012;109(27):10909-14.

12. Dunn KW, Kamocka MM, McDonald JH. A practical guide to evaluating colocalization in biological microscopy. Am J Physiol Cell Physiol. 2011; 300(4):C723-C742

13. Earle Kristen A, Billings G, Sigal M, Lichtman Joshua S, Hansson Gunnar C, Elias Joshua E, et al. Quantitative Imaging of Gut Microbiota Spatial Organization. Cell Host & Microbe. 2015;18(4):478-88.

14. Skočaj M, Resnik N, Grundner M, Ota K, Rojko N, Hodnik V, et al. Tracking Cholesterol/Sphingomyelin-Rich Membrane Domains with the Ostreolysin A-mCherry Protein. PLOS ONE. 2014;9(3):e92783.

15. Wu Z, Tang M, Tian T, Wu J, Deng Y, Dong X, et al. A specific probe for two-photon fluorescence lysosomal imaging. Talanta. 2011;87:216-21.

16. George TC, Fanning SL, Fitzgeral-Bocarsly P, Medeiros RB, Highfill S, Shimizu Y, et al. Quantitative measurement of nuclear translocation events using similarity analysis of multispectral cellular images obtained in flow. J Immunol Methods. 2006;311(1):117-29.

17. Cabrera JE, Jin DJ. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Mol. Microbiol. 2003;50(5):1493-505.

18. Castaing B, Zelwer C, Laval J, Boiteux S. HU Protein of *Escherichia coli* Binds Specifically to DNA That Contains Single-strand Breaks or Gaps. J Biol Chem. 1995;270(17):10291-6.

19. Wang W, Li G-W, Chen C, Xie XS, Zhuang X. Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. Science. 2011;333(6048):1445.

20. Nielsen HJ, Ottesen JR, Youngren B, Austin SJ, Hansen FG. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. Mol. Microbiol. 2006;62(2):331-8.

21. Biteen JS, Thompson MA, Tselentis NK, Bowman GR, Shapiro L, Moerner WE. Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable EYFP. Nat Methods. 2008;5(11):947-9.

22. Subach FV, Patterson GH, Manley S, Gillette JM, Lippincott-Schwartz J, Verkhusha VV. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. Nat Methods. 2009;6(2):153-9.

23. Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. Proc Natl Acad Sci U S A. 2002;99(20):12651-6.

24. Hayashi I, Mizuno H, Tong KI, Furuta T, Tanaka F, Yoshimura M, et al. Crystallographic Evidence for Water-assisted Photo-induced Peptide Cleavage in the Stony Coral Fluorescent Protein Kaede. J Mol Biol. 2007;372(4):918-26.

25. Bakshi S, Bratton Benjamin P, Weisshaar James C. Subdiffraction-Limit Study of Kaede Diffusion and Spatial Distribution in Live *Escherichia coli*. Biophys. J. 2011;101(10):2535-44.

26. Adler J, Parmryd I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry Part A. 2010;77A(8):733-42.

27. McDonald JH, Dunn KW. Statistical tests for measures of colocalization in biological microscopy. J Microsc. 2013;252(3):295-302.

28. Lagache T, Sauvonnet N, Danglot L, Olivo-Marin J-C. Statistical analysis of molecule colocalization in bioimaging. Cytometry Part A. 2015;87(6):568-79.

29. Owen DM, Rentero C, Rossy J, Magenau A, Williamson D, Rodriguez M, et al. PALM imaging and cluster analysis of protein heterogeneity at the cell surface. Journal of Biophotonics. 2010;3(7):446-54.

30. Perry GLW. SpPack: spatial point pattern analysis in Excel using Visual Basic for Applications (VBA). Environmental Modelling & Software. 2004;19(6):559-69.

31. Ripley BD. Tests of `Randomness' for Spatial Point Patterns. Journal of the Royal Statistical Society Series B (Methodological). 1979;41(3):368-74.

32. Sengupta P, Jovanovic-Talisman T, Skoko D, Renz M, Veatch SL, Lippincott-Schwartz J. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. Nat Methods. 2011;8:969.

33. Veatch SL, Machta BB, Shelby SA, Chiang EN, Holowka DA, Baird BA. Correlation Functions Quantify Super-Resolution Images and Estimate Apparent Clustering Due to Over-Counting. PLOS ONE. 2012;7(2):e31457.

34. Lee Steven F, Thompson Michael A, Schwartz MA, Shapiro L, Moerner WE. Super-Resolution Imaging of the Nucleoid-Associated Protein HU in *Caulobacter crescentus*. Biophys. J. 2011;100(7):L31-L3.

35. Fiolka R, Shao L, Rego EH, Davidson MW, Gustafsson MGL. Time-lapse two-color 3D imaging of live cells with doubled resolution using structured illumination. Proc Natl Acad Sci U S A. 2012;109(14):5311-5.

36. Jones SA, Shim S-H, He J, Zhuang X. Fast, three-dimensional super-resolution imaging of live cells. Nat Meth. 2011;8(6):499-505.

Appendix

A2.1. Two additional cases of anti-correlation in 3D and in 2D

We tested two additional cases (Case II and III) of perfectly anti-correlated distributions in both 3D and 2D, shown in Fig. A2.1. In these two cases, both the red and green channels contain 10,000 molecules. The molecules constituting **R** and **G** are a subset of a larger number of molecules distributed randomly in a spherocylinder. In Case II, the two end cap regions are occupied by red molecules and the central region of the spherocylinder is occupied by green molecules. The volume occupied by red molecules is greater than that occupied by green molecules. In Case III, the red molecules are confined to the leftmost 2/3 of the spherocylinder and the rest of the volume is occupied by green molecules. The volume occupied by red molecules is greater than that of the green molecules. In both Cases II and III, the red and green molecules never occupy the same region of space in 3D or 2D, the hallmark of perfect anti-correlation. For all the cases including case I described in the main text, the linear anti-correlation relationship is demonstrated by the scatter plot of $\hat{\Delta}_{ij}^{R}$ vs $\hat{\Delta}_{ij}^{G}$ (Fig. A2.1*B*). The calculated MPCC in these cases is –0.99, approaching the expected value of –1 within the limits of noise. However, the calculated PCC is –0.49, significantly underestimating the degree of anti-correlation.

A2.2. Example of positive correlation

An example of perfect positive correlation is presented in the main text. A detailed comparison of MPCC and PCC applied to this case is provided in Fig. A2.2.

A2.3. Materials and Methods

Strain construction

For determining cell radius and length, we imaged the photoactivatable fluorescent protein Kaede (1, 2), which was expressed from a tetracycline inducible plasmid pASK IBA3 plus (IBA). *E. coli* DH5α strain expressing Kaede was constructed as described by Bakshi *et al* (3). This vector was then transformed into VH1000 using the TSS protocol and selected against ampicillin resistance. For imaging RNAP, we used an *E. coli* VH1000 strain that endogenously expresses the RNAP β' subunit translationally fused with yGFP. Construction of the *rpoc*::*ygfp* strain was described before by Bakshi *et al*. The fluorophore yGFP is a variant of GFP (4). It absorbs green light (λ_{max} = 514 nm) and fluoresces green-yellow (λ_{max} = 533 nm) (5).

The DNA binding protein HU – PAmcherry used for imaging the nucleoid was expressed from a plasmid. The gene *hupA* (HU) was amplified from VH1000 using the primers (primers, 5'-ATCGGATCCCGCCGCCGCCTTAACTGCGTC-3'and 5'-GATCGAATTCTA GGAGGTATTCACATGAACAAGACTC-3'). The amplified *hupA* gene with restriction sites for EcoRI and BamHI was inserted into pASK IBA3 plus (IBA). Then *PAmcherry* was amplified using primers (primers, 5' -GATCAAGCTTTTACTTGTACAGCTCGTCCATGCC-3' and 5'-GATCGGATCCGTGAGCAAGGGCGAGG-3'). The amplified *PAmcherry* with restriction sites for BamHI and HindIII was inserted into the *hupA* vector. This vector was transformed into VH1000 using the TSS protocol and selected against ampicillin resistance.

Cell growth and preparation for imaging

Bulk cultures from frozen glycerol stock solution and subcultures for imaging were grown at 30°C with continuous shaking in 100 µg/mL ampicillin containing EZ rich, defined medium ("EZRDM"), which is a MOPS-buffered solution with supplemental metal ions (M2137;

Teknova), glucose (2 mg/mL), supplemental amino acids and vitamins (M2104; Teknova), nitrogenous bases (M2103; Teknova), 1.32 mM K_2HPO4, and 76 mM NaCl.

When cells had grown to midlog phase, anhydrotetracycline was added to a final concentration of 25 nM to induce the expression of the desired protein. After 5 min of induction, the cells were centrifuged and resuspended in fresh EZRDM with 100 μg/mL ampicillin to remove the inducer. The cells were then incubated again in growth medium for 15 min at 30°C to enable maturation of the respective fluorescent protein prior to imaging.

Superresolution imaging of live *E. coli* cells

Fluorophore-labeled cells were grown overnight with shaking at 30°C in EZRDM. Subcultures were made by diluting the stationary phase culture at least 1:100 into 2 mL of fresh EZRDM. Subcultures were grown to exponential phase (OD = 0.2–0.6 at 600 nm), upon which the culture was placed in CoverWell Perfusion Chamber Gaskets (Invitrogen, Carlsbad, CA) on a polylysine-coated coverslip. The volume of the closed chamber is ~ 150 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. Live cell imaging was carried out at 30°C for no longer than 30 min after plating. During that time, cells continue to grow.

Cells were imaged using an Eclipse Ti inverted microscope (Nikon Instruments, Melville, NY) equipped with an oil immersion objective (CFI Plan Apo Lambda DM 100x Oil, 1.45 NA; Nikon Instruments), a 1.5x tube lens, and the Perfect Focus System (Nikon Instruments, Melville, NY).

For imaging Kaede, we used a fast back-illuminated electron-multiplying charge-coupled device (EMCCD) camera with 128 x 128 pixels of 24 x 24 μm each (iXon DV-860; Andor Technology, South Windsor, CT). Each pixel corresponds to 160 x 160 nm^2 at the sample (150x

overall magnification). For acquisition of fast diffusing Kaede, all data were collected at a frame rate of 485 Hz, with exposure time within each frame of 2 ms. The Kaede was photoactivated with a 405-nm diode laser (CrystaLaser, Reno, Nevada) and subsequently imaged with a 561-nm laser (Sapphire 561 CW lasers; Coherent, Bloomfield, CT). The 405-nm power density at the sample was 4-15 W/cm^2. The photoactivation laser remained on throughout imaging. Power density of the 561 nm laser was kept at ~2 kW/cm^2. The probe laser was on continuously in the 2 ms/frame tracking experiments. To minimize the phototoxic effect of the laser, we collected data for <5 s per cell. The Kaede emission was collected through a 617/73 bandpass filter (bright line 617/73-25; Semrock, Rochester, NY).

Images of cells that express RNAP β' – yGFP along with HU – PAmcherry were recorded by a back – illuminated EMCCD camera with 512 x 512 pixels of 16 μm x 16 μm pixels (iXon DV – 887; Andor Technology, CT, USA). Each pixel corresponds to 105 x 105 nm^2 at the sample with an overall magnification of 150x. All images were collected at a frame rate of 31.2 Hz, with an exposure time within each frame of 30 ms. The RNAP β'–yGFP molecules were illuminated with the 514 nm laser with a power density of ~2 kW/cm^2. Images were acquired only when the rate of return of molecules to the fluorescent state becomes small enough that at most 3-4 copies per camera frame are fluorescent in each cell, enabling single molecules to be distinguished from each other. To minimize the phototoxic effect of the laser, we collected data for <35 s per cell. The HU–PAmcherry molecules were photoactivated with 405 nm diode laser and subsequently imaged with the 561 nm laser. The 405-nm power density at the sample was 6 - 10 W/cm^2 to ensure only 3 - 4 molecules of HU–PAmcherry are fluorescent per camera frame per cell. The power density of the 561 nm laser was kept at ~ 2 kW/ cm^2 and the laser was continuously on during the 30 ms/frame imaging experiments. Similarly, the cells were imaged

for HU–PAmcherry for <35 s per cell to ensure minimum cell damage and normal cell growth during the imaging period. We used polychroic ZT405-514-561rpc (Chroma, VT, USA). The fluorescence of RNAP β'–yGFP and HU–PAmcherry were isolated using triple band pass filter ZET442/514/561m (Chroma, VT, USA). We also added a 525 nm long pass filter (ET525lp, Chroma, VT, USA) to block the shorter wavelength band allowed by the used polychroic. To minimize leakage of emission of yGFP into the red channel, we first imaged S2–yGFP which results in photobleaching of majority of yGFP molecules.

Single molecule image analysis

Images were analyzed using a MATLAB GUI developed in our lab (3). Images were smoothed and filtered to obtain a zero-based image. Bright spots were located with pixel level accuracy by a peak-finding algorithm that finds local intensity maxima within an image. A user-defined intensity threshold was used as the minimum brightness of a pixel from single molecule trajectories. It is carefully set by the user so that it will not be so high as to cut a long trajectory short or so low as to include background noise.

Centroids of the bright spots of RNAP β'-yGFP and HU-PAmcherry were calculated from a 7 x 7 pixel square centered on the local maxima determined by the peak-finding algorithm. Since fast diffusing Kaede are blurred extensively, we use an 11 x 11 pixel square centered on the local maxima to calculate the centroid. As the images are asymmetrically blurred due to diffusion during the frame time, we calculate the centroid of the bright spots instead of using Gaussian fitting. This centroid analysis is fast and easily implemented for analysis of Monte Carlo modeling results as well. The (x, y) positions of the centroid are utilized to generate 2D localization probability density maps. For single-molecule tracking analysis, the (x, y) positions of the centroid are stored and connected into trajectories using a modified MATLAB version of

the tracking program written by Crocker and Grier (6). These trajectories will be used in determination of mean diffusion coefficient.

Determination of localization error of imaged molecules

The MSD as a function of lag time τ is given by $\mathrm{MSD}(\tau) = < (\boldsymbol{r}(t+\tau) - \boldsymbol{r}(t))^2 >$, where $\boldsymbol{r}(t)$ is the two-dimensional location of the particle at time t, τ is the lag time, and the average is taken over all times t and over many trajectories. We chose trajectories that lasted for at least seven frames in order to minimize studying of spurious molecules. Tracks longer than seven frames were truncated to seven frames. Suppose the best fit to the first two experimental points of a two dimensional mean-square displacement plot is given by the equation $\mathrm{MSD}(\tau) = a + b\tau$, with b the slope and a the extrapolated intercept at lag time $\tau = 0$. Then the best estimate of the diffusion coefficient is $D = b/4$ and the best estimate of the dynamic localization error is $\sigma = \frac{1}{2}(a + 4Dt_E/3)^{1/2}$, where t_E is the exposure time per camera frame (7).

For RNAP β'-yGFP, using 685 trajectories we determined the diffusion coefficient to be 0.064 ± 0.004 $\mu m^2\text{-}s^{-1}$ and $\sigma_{RNAP} = 38$ nm (Fig. A2.7). For HU-PAmcherry, using 1180 trajectories we determined the diffusion coefficient to be 0.26 ± 0.04 $\mu m^2\text{-}s^{-1}$ and $\sigma_{HU} = 60$ nm in (Fig. A2.7).

A2.4. Determining cell width and length

The cytoplasmic radius varies little from cell to cell To determine the cell radius r, we imaged photoactivable Kaede molecules, which are believed to spatially distribute homogenously in cytoplasmic volume (3). We used data from 15 cells under same growth conditions used for imaging cells expressing RNAP β'-yGFP and HU-PAmcherry. MSD(τ) plot for Kaede imaged with an exposure time of 2 ms is shown in Fig. A2.6*A*. We chose 709 trajectories that lasted for at least seven frames. Tracks longer than seven frames were truncated

to seven frames to avoid overcounting. The mean diffusion coefficient for Kaede was measured to be $D_{Kaede} = 4.77 \pm 0.21$ μm^2-s^{-1}. For $a = 0.025$ μm^2, $D = 4.77$ μm^2-s^{-1}, and $t_E = 2$ ms, the dynamic localization error is estimated to be $\sigma_{Kaede} = 97$ nm.

To estimate the cytoplasmic radius r, we then simulated Kaede diffusion using $D_{Kaede} = 5.0$ μm^2-s^{-1} and $\sigma_{Kaede} = 97$ nm in uniformly filled spherocylinders of varying radii, seeking a match to the experimental transverse spatial distribution. The model spherocylinders had cylindrical length of 3.5 μm. At $t = 0$, 15,000 particles were randomly distributed within the cell volume. Each particle undergoes a random walk independent of other particle positions. To model each 2-ms camera image, three-dimensional microtrajectories (1000 steps of 2 μs each) were generated. At each time step, each particle chooses a displacement in each of three Cartesian directions. These displacements are chosen from a Gaussian distribution whose standard deviation corresponds to the three-dimensional diffusion coefficient $D_{Kaede} = 5$ μm^2-s^{-1}. In the rare event that a particle attempts to step outside of the cell boundaries, the displacement for that microstep is taken to be zero. The location of each particle during each camera frame is obtained as the centroid of the model microtrajectories in order to mimic the analysis procedure used for the experimental images. The appropriate dynamic localization error σ_{Kaede} was then applied to each centroid location in both x and y coordinates by sampling a Gaussian distribution with standard deviation $\sigma = 97$ nm. By adding the error vector to the centroid position we obtain the model "measured" location for each 2 ms camera frame. The x and y coordinates of each measured location are stored for further analysis. For the next model camera frame, each particle continues to make microsteps in 3D starting from the endpoint of the previous camera frame. Only the particles in the central cylindrical region are used to generate the simulated radial distribution; the endcaps are deleted.

For each of 15 cells, we found the best fit of the experimental distribution to the spherocylindrical model in a least-squares sense by varying the model radius in increments of 10 nm. A single-cell experimental radial distribution of 1725 Kaede locations, excluding the endcap regions of the cell, is shown in Fig. A2.6*B*. The best-fit simulated radial distribution had r = 390 nm (Fig. A2.6*B*). Averaged over 15 cells, the mean cytoplasmic radius is r = 412 ± 22 nm (± 1σ).

For measuring each cell length L_{cell}, we scattered the localizations of the molecules imaged in the cells in the chosen length bin as shown in Fig. A2.6*C*. The points at which the scattered distributions begin to curve mark the beginnings of the hemispherical endcap regions and the ends of the cylindrical region. The error in determination of the inflection point is in the orders of ~50 nm which is significantly smaller than the measured cell length. The distance between these two inflection points is the length of the cylindrical region of the cell (L_{cyl}). The measured tip-to-tip cell length is then $L_{cell} = L_{cyl} + 2r$.

A2.5 Supporting Figures and Captions

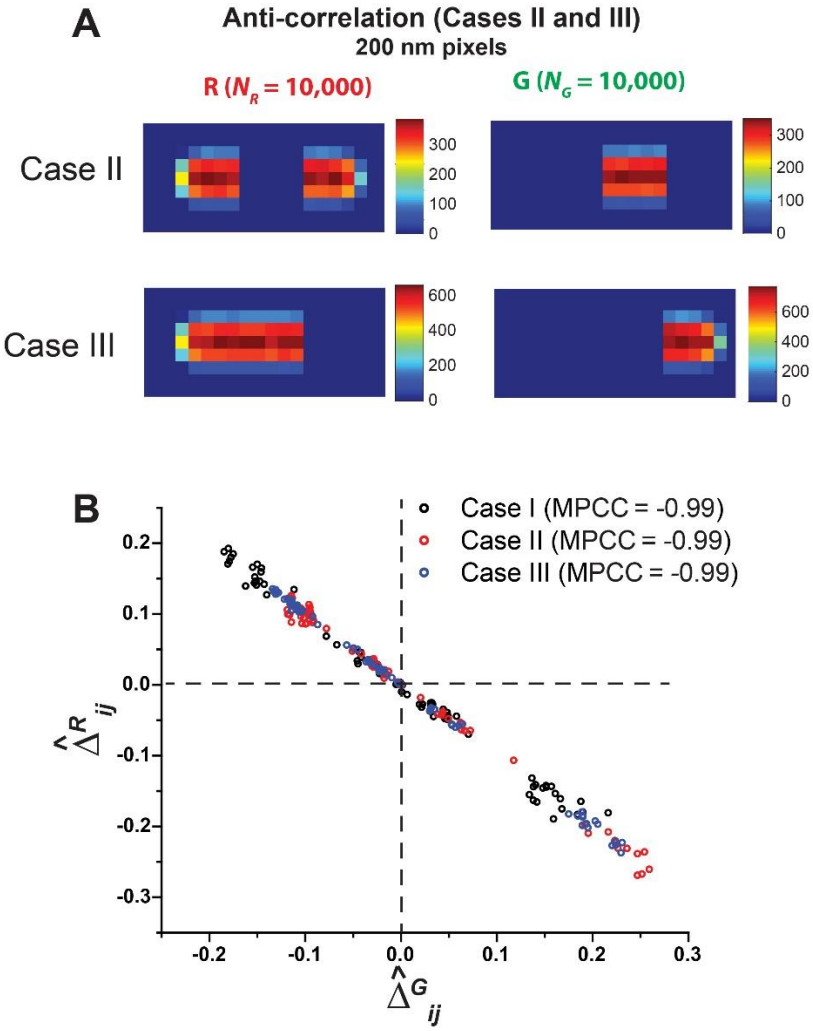


Fig. A2.1 Additional cases of images perfectly anti-correlated in 3D and 2D. In each case, there are ~10,000 red and ~10,000 green molecules, with 200 nm pixels. **(A)** *Top*: Case II. *Bottom*: Case III. The color of each pixel indicates the number of molecules as shown in the color bar. **(B)** Scatter plot of individual difference elements, $\hat{\Delta}^R_{ij}$ vs. $\hat{\Delta}^G_{ij}$ for Case I described in main text (*Black*), Case II (*Red*), and Case III (*Blue*). The calculated MPCC is –0.99 in all three cases.

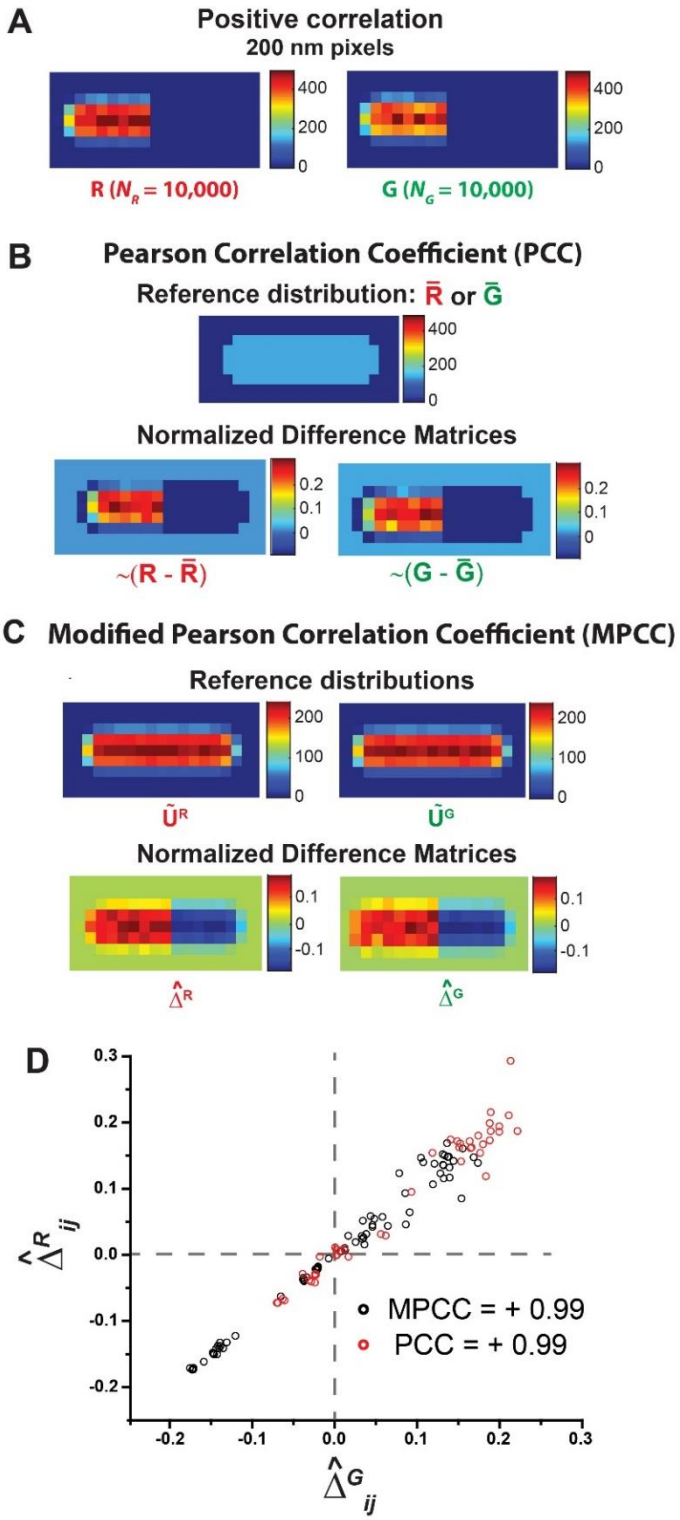


Fig. A2.2 Scheme for calculating PCC and MPCC for two representative images **R** and **G** that are positively correlated in both 3D and 2D. (**A**) Images **R** and **G** with pixels of size 200 nm, each containing ~10,000 molecules. The color of each pixel is indicative of the number of molecules as shown in the adjacent color bar. (**B**) Standard PCC calculation. *Top*: Reference distribution $\bar{\mathbf{R}}$ or $\bar{\mathbf{G}}$, with same individual pixel counts, that is subtracted from images **R** or **G**. *Bottom*: Normalized difference matrices $\sim(\mathbf{R} - \bar{\mathbf{R}})$ and $\sim(\mathbf{G} - \bar{\mathbf{G}})$ obtained after subtraction. The Frobenius inner product of these two difference matrices gives the PCC. (**C**) Modified PCC calculation. *Top*: Reference distribution $\tilde{\mathbf{U}}^{\mathbf{R}}$ or $\tilde{\mathbf{U}}^{\mathbf{G}}$, which are 2D projections of 3D random distribution in a spherocylinder, that is subtracted from images **R** or **G**. *Bottom*: Normalized difference matrices $\hat{\Delta}^{\mathbf{R}}$ and $\hat{\Delta}^{\mathbf{G}}$ obtained after subtraction. The Frobenius inner product of these two difference matrices gives the MPCC. (**D**) Scatter plot of individual difference elements, $\hat{\Delta}^{R}_{ij}$ vs. $\hat{\Delta}^{G}_{ij}$, for PCC (*Red*) and MPCC (*Black*). The resulting MPCC and PCC values are both +0.99.

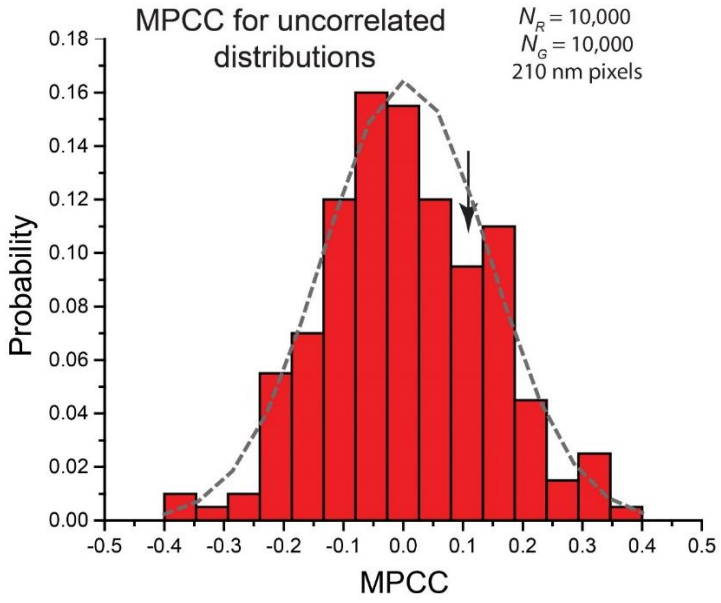


Fig. A2.3 Probability distribution of 200 MPCC calculated for sets of **R** and **G** matrices formed from 2D projections of uncorrelated, random 3D distributions. Each **R** and **G** matrix contains 10,000 molecules and uses 210 nm pixels. The best-fit Gaussian curve (*dashed grey line*) has mean <MPCC> = 0.0041 and standard deviation σ_{MPCC} = 0.13. The arrow shows the bin for the value MPCC = +0.14, the result for the sample calculation in the main text (Fig. 2.3).

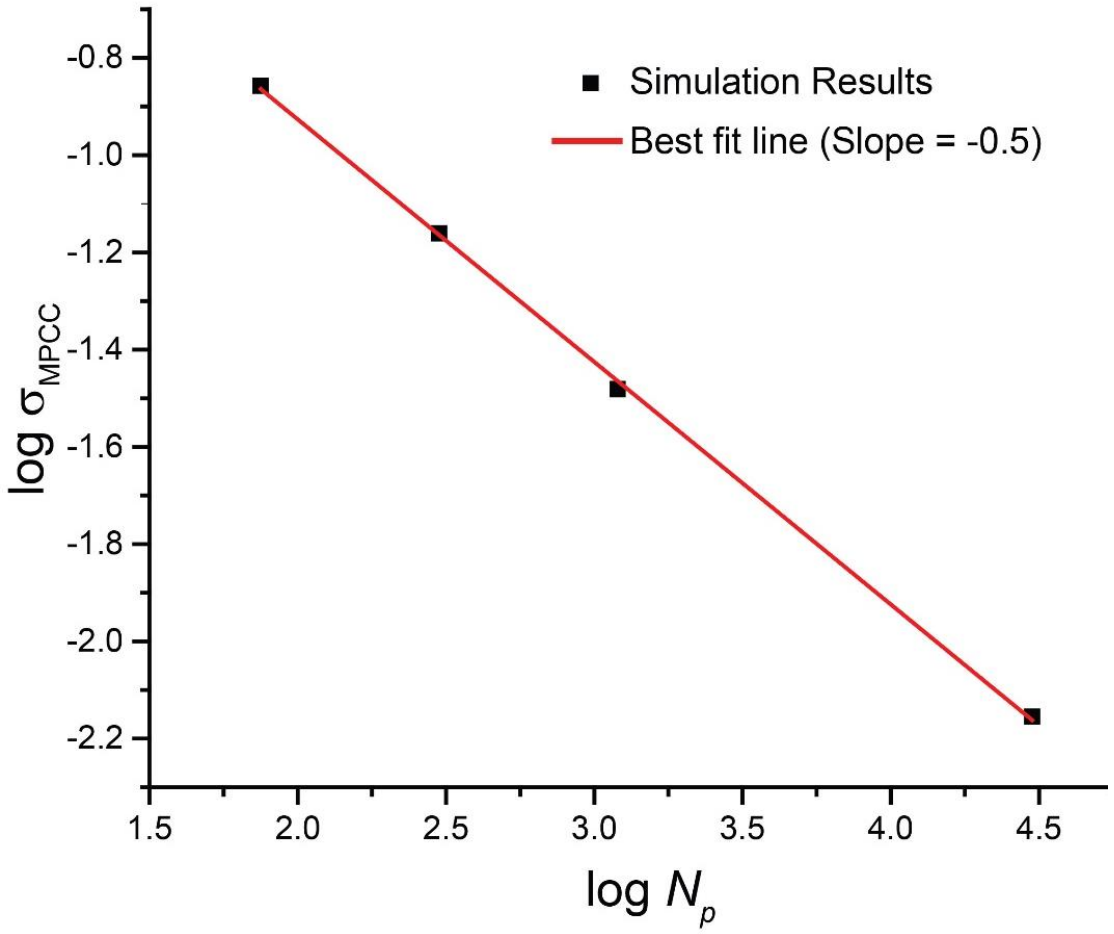


Fig. A2.4 Dependence of σ_{MPCC} (Fig. A2.3) on the total number of pixels (N_p) for uncorrelated **R** and **G** matrices. The number of red and green molecules is held constant at ~10,000 each. Best-fit line to log σ_{MPCC} vs. log N_p has slope –0.49.

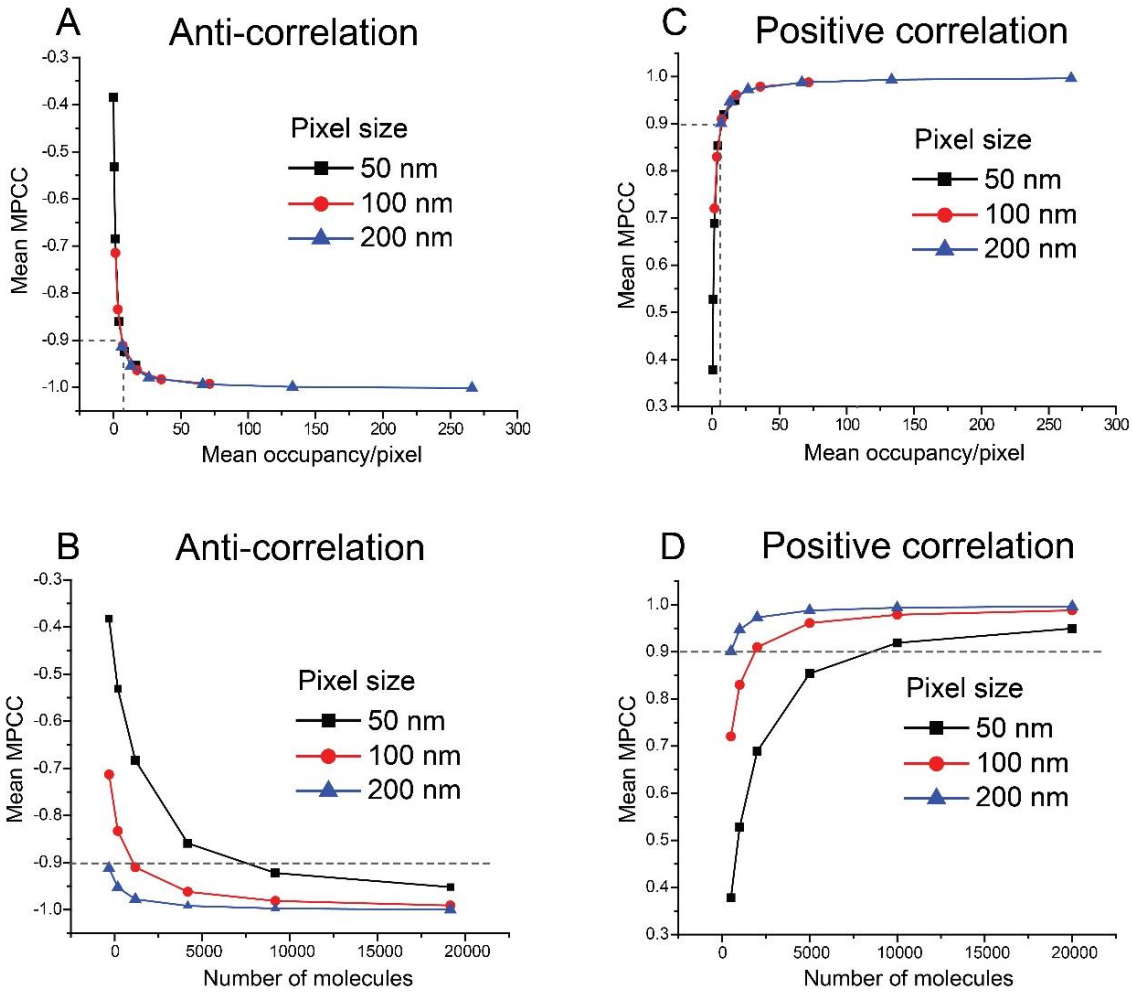


Fig. A2.5 Variation of mean MPCC (calculated from 200 trials as in Fig. A2.3) with number of pixels N_p and with total number of red and green molecules in the image matrices **R** and **G**. Models of perfect anti-correlation (as in Fig. 2.2) and perfect positive correlation (as in Fig. A2.2) were tested. (**A**) Anti-correlated images **R** and **G**. <MPCC> vs mean occupancy/pixel for three different pixel sizes as shown: 50 nm (N_p = 1178), 100 nm (N_p = 279), and 200 nm (N_p = 77). (**B**) Same data as in (A), now plotted vs the number of red and green molecules. (**C**) Positively correlated images **R** and **G**. <MPCC> vs mean occupancy/pixel for three different pixel sizes as shown: 50 nm (N_p = 1178), 100 nm (N_p = 279), and 200 nm (N_p = 77). (**D**) Same data as in (C), now plotted vs the number of red and green molecules.

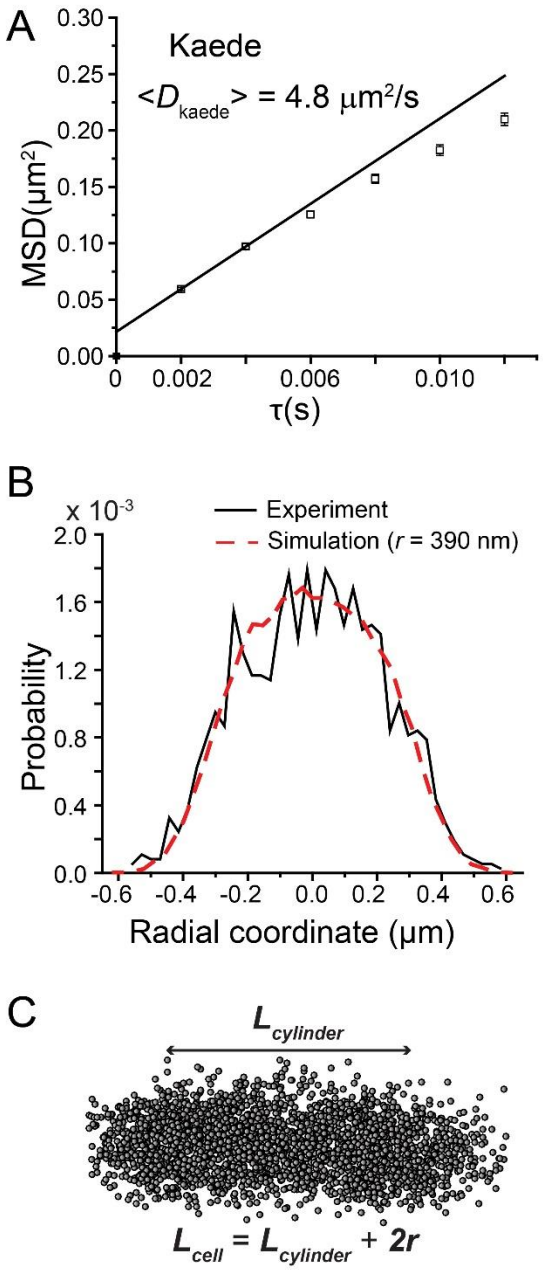


Fig. A2.6 (**A**) Mean-square displacement plot, MSD(τ), for Kaede under imaging conditions described in A3. Trajectories are truncated to six steps; error bars are ± one standard error of the mean. The curvature is likely due to confinement. Line drawn through first two data points yields the estimated mean diffusion coefficient $\langle D_{Kaede} \rangle$ = 4.77 μm^2/s. (**B**) (*Black*) An example experimental single-cell radial distribution of Kaede molecules within the straight cylindrical region of the cell (endcaps excluded). (*Dashed red*) Simulated radial distribution of homogenously distributed molecules in a 3D spherocylinder. Best-fit radius is r = 390 nm. (**C**) Scatter plot of imaged Kaede molecules in a single cell. $L_{cylinder}$ is taken as the length of the cylindrical region, as judged by eye. Tip-to-tip length $L_{cell} = L_{cylinder} + 2r$.

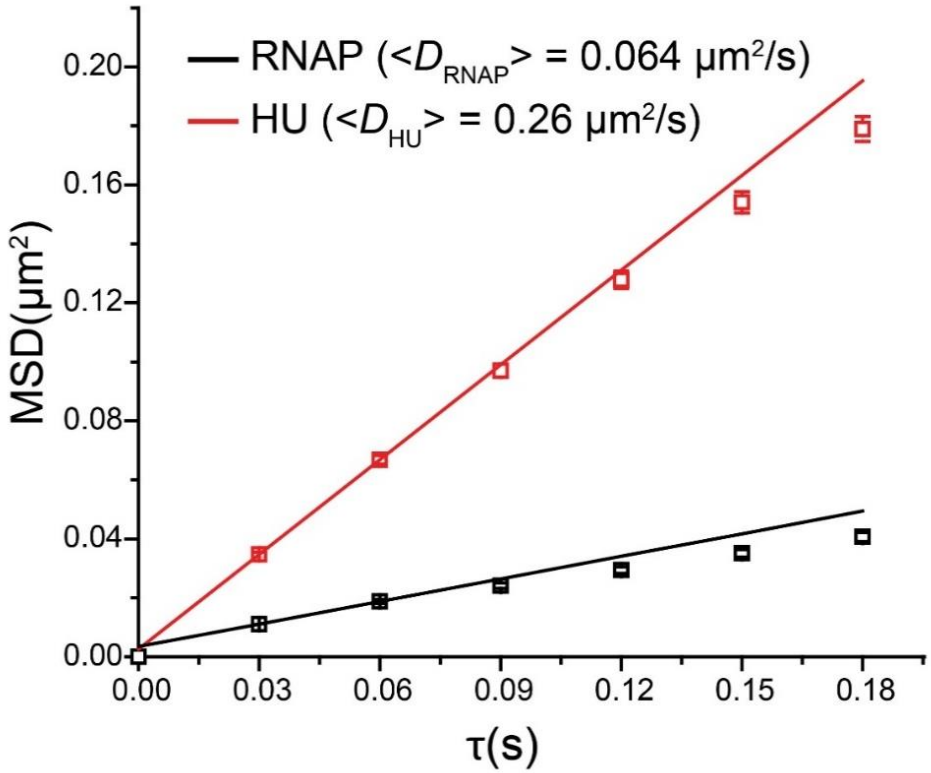


Fig. A2.7 Mean-square displacement plots, MSD(τ), for HU–PAmcherry (*red squares*) and for RNAP–YFP (*black squares*). Imaging conditions described in A3. Trajectories are truncated to six steps; error estimates are ± one standard error of the mean. Lines drawn through first two data points yield estimated mean diffusion coefficients $\langle D_{HU} \rangle$ = 0.26 μm^2/s and $\langle D_{RNAP} \rangle$ = 0.064 μm^2/s.

A2.6 Supporting References

1. Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. Proc Natl Acad Sci U S A. 2002;99(20):12651-6.
2. Hayashi I, Mizuno H, Tong KI, Furuta T, Tanaka F, Yoshimura M, et al. Crystallographic Evidence for Water-assisted Photo-induced Peptide Cleavage in the Stony Coral Fluorescent Protein Kaede. J Mol Biol. 2007;372(4):918-26.
3. Bakshi S, Bratton Benjamin P, Weisshaar James C. Subdiffraction-Limit Study of Kaede Diffusion and Spatial Distribution in Live *Escherichia coli*. Biophys. J. 2011;101(10):2535-44.
4. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution Imaging of Ribosomes and RNA Polymerase in Live *Escherichia coli* Cells. Mol Microbiol. 2012;85(1):21-38.
5. Nielsen HJ, Ottesen JR, Youngren B, Austin SJ, Hansen FG. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. Mol Microbiol. 2006;62(2):331-8.
6. Crocker JC, Grier DG. Methods of Digital Video Microscopy for Colloidal Studies. J Colloid Interface Sci. 1996;179(1):298-310.
7. Michalet X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Phy Rev E. 2010;82(4):041914.

3

Spatial organization of ribosomes across different growth conditions in *E. coli* supports a ribosomal circulation model

Reproduced from:

Mohapatra S, Weisshaar JC. Functional mapping of the *E. coli* translational machinery using single-molecule tracking. *Submitted*

Abstract

The organization of the chromosomal DNA and ribosomes in living *E. coli* is compared under two growth conditions: “fast” (50-min doubling time) and “slow” (147-min doubling time). Superresolution fluorescence microscopy reveals strong DNA-ribosome segregation in both cases. In both fast and slow growth, free ribosomal subunits evidently must circulate between the nucleoid (where they initiate co-transcriptional translation) and ribosome-rich regions (where most translation occurs). Single-molecule diffusive behavior dissects the ribosome copies into translating 70S polysomes and free 30S subunits, providing separate spatial distributions for each. In slow growth, ~21,000 total 30S copies/cell comprise ~65% translating 70S ribosomes and ~35% free 30S subunits. The ratio of 70S ribosomes to free 30S subunits is ~2.5 outside the nucleoid and ~0.50 inside the nucleoid. This new level of quantitative detail may motivate development of comprehensive, three-dimensional reaction-diffusion models of ribosome, DNA, mRNA, and RNAP spatial distributions and dynamics within the *E. coli* cytoplasm.

Introduction

Within the cytoplasm of the bacterial species *E. coli* and *B. subtilis*, the ribosomes exhibit strong segregation from the chromosomal DNA, despite the absence of membrane-bounded organelles (1, 2). The DNA occupies a region of space called the nucleoid, which is substantially smaller in volume than the complete cytoplasm (3). Apparently most transcription of protein genes takes place separately from the ribosome-rich regions, where the bulk of translation must occur. In sharp contrast, in *C. crescentus* the chromosomal DNA occupies the entire volume of the cytoplasm and the DNA and ribosomes are thoroughly mixed (4). Thus far, strong DNA-ribosome segregation was observed in rapidly growing *E. coli* and *B. subtilis* (1, 2), which are relatively large cells containing multiple chromosome equivalents of DNA within their highly compacted nucleoids. Since *C. crescentus* is smaller and grows very slowly, we decided to extend our previous work to slowly growing *E. coli* cells that are comparable to *C. crescentus* in volume, doubling time, and DNA content. We use superresolution fluorescence microscopy (5-7) to locate and track 30S ribosomal subunits and the non-specific DNA binding protein HU, a proxy for the spatial distribution of DNA, in two growth conditions: “fast” (50-min doubling time at 30°C) and “slow” (147-min doubling time at 30°C).

In earlier work on fast growing *E. coli* (1), detailed single-molecule tracking experiments on ribosomal subunits and RNA polymerase (RNAP) supported a circulation model for the movement of translating 70S ribosomes and of free 30S and 50S ribosomal subunits between nucleoid and ribosome-rich regions (Fig. 3.1). Formation of poly-ribosomes (polysomes) begins within the nucleoid while a message is still being transcribed (co-transcriptional translation). Polysomes can diffuse to a ribosome-rich region in several seconds or less, a short timescale compared with the typical mRNA lifetime of several minutes (8). When each 70S ribosome

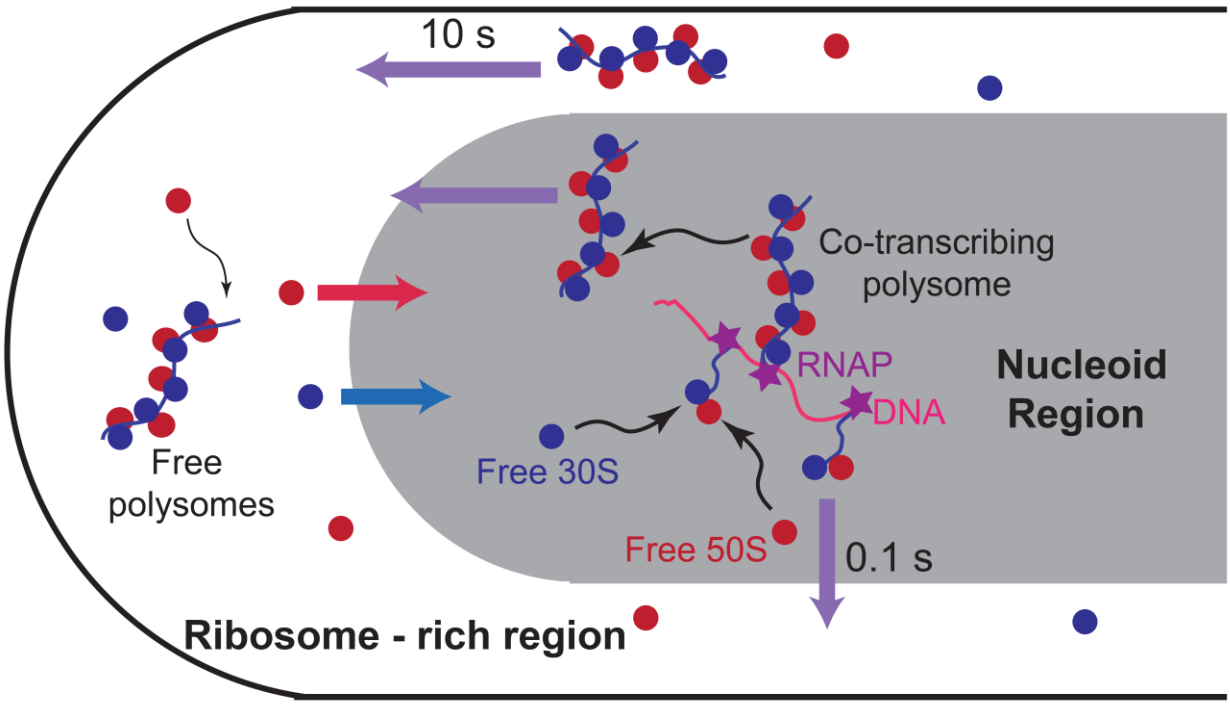


Fig. 3.1 Schematic showing the suggested circulation model for movement of ribosomal subunits into and out of the nucleoids.

completes a translation event within a ribosome-rich region, its 30S and 50S subunits are released. Many of these free subunits will find a new translation initiation site on an mRNA within the ribosome-rich region, but some must escape and return to the nucleoid where they can initiate further rounds of co-transcriptional translation (9).

The new results provide a quantitative, comprehensive comparison of the organization of the *E. coli* cytoplasm in fast vs slow growth. We find that *E. coli* exhibits similarly strong DNA-ribosome segregation in fast and slow growth, as quantified by a modified version of the Pearson Correlation Coefficient introduced here. We provide quantitative estimates of the peak DNA number density on a ~100 nm scale within the *E. coli* nucleoids of both fast and slow growing cells. The peak density is ~8-fold higher in fast growth. The slow-growth nucleoid is "fluffier", perhaps because it is being driven by transcriptional and gyrase activity that is more widely distributed throughout the nucleoid (10). In addition, we use single-molecule diffusive behavior to dissect the ribosome copies into slowly diffusing, translating 70S ribosomes (mostly polysomes) and more rapidly diffusing, free 30S subunits and provide separate spatial distributions for each species. Free 30S and 50S subunits can readily penetrate, and 70S polysomes are strongly excluded from, both the high density nucleoid in fast growth and the low density nucleoid in slow growth. The circulation model evidently applies to *E. coli* at all growth rates. In slow growth, a simple model of the nucleoid as a spherocylinder enables us to provide quantitative estimates of the partitioning of 70S and 30S species between the nucleoid and the ribosome-rich regions. We hope this new body of information will stimulate development of a quantitative, fully 3D, comprehensive reaction-diffusion model of ribosome, DNA, mRNA, and RNAP spatial distributions and circulation in slow growing *E. coli* (11).

The underlying causes of the segregation of ribosomes from the chromosomal DNA in *E. coli* and *B. subtilis* have been the subject of intense speculation. As a bare, charged polymer in salt solution, the chromosomal DNA would occupy ~1000 times the volume of the *E. coli* cytoplasm (12), yet in fact it occupies only about half the cytoplasm in the living cell. This "compaction" of the DNA into such a small volume involves both the physics of charged polymers confined to a small space and the complex biochemistry of the bacterial cytoplasm. Configurational and translational entropy (13, 14), excluded volume effects (11, 13), crowding (15), repulsion between polyanionic DNA and polyanionic ribosomes (16), supercoiling of DNA (17), and binding of DNA to nucleoid-associated proteins (18-21) may all contribute to DNA compaction and DNA-ribosome segregation.

The same factors should be in play in the smaller, more slowly growing species *Caulobacter crescentus*. Our original hypothesis was that the *C. crescentus* chromosome occupies the entire cytoplasm because the cytoplasmic volume is so small. Perhaps the combined compaction forces are too weak to force the chromosome to be smaller still. Yet our slow-growing *E. coli*, comparable to *C. crescentus* in DNA content and cytoplasmic volume, exhibit strong DNA-ribosome segregation. Having eliminated one possible cause of extensive DNA-ribosome mixing in *C. crescentus*, we speculate on other possible explanations and suggest new studies.

Results

Spatial distributions of ribosomes and DNA in fast and slow growth conditions

To compare the spatial distributions of ribosomes and DNA in different growth conditions, we performed two-color superresolution fluorescence imaging of ribosomes (30S subunits via labeling of the S2 protein) and the non-specific DNA binding protein HU in live *E.*

coli cells. In what we term fast growth, the cells are growing in EZRDM (EZ rich defined media) (22) with glucose as carbon source at 30°C; the doubling time is 50 min. In what we term slow growth, the cells are grown in MBM minimal media with glycerol as carbon source at 30°C; the doubling time is 147 min. We began with an *E. coli* strain VH1000 in which the gene coding for the fluorescent protein YFP (observed in the green channel) (23) is fused to the C-terminus of the endogenous *rpsB* gene, which codes for the S2 ribosomal protein. On the timescale of YFP maturation (24), S2 is efficiently incorporated into 30S ribosomal subunits (25). In effect we are imaging 30S subunits, which may themselves be freely diffusing (searching for translation initiation sites on mRNA) or incorporated into translating 70S ribosomes (primarily polysomes). Single 30S ribosome copies are imaged using the reversible photobleaching method described earlier (26). An inducible plasmid that expresses HU labeled with the photoactivatable fluorescent protein PAmcherry (27) (observed in the red channel) was introduced into the same strain. HU binds non-specifically to the chromosomal DNA; we take its spatial distribution to be a proxy for that of the chromosomal DNA. In each growth condition, cells were grown at 30°C, plated on a glass coverslip, and imaged with 30 ms/frame exposure time. For 30S-YFP, the localization error estimated from MSD plots (Fig. A3.1) is σ = 36 nm and 27 nm in fast and slow growth, respectively. For the more rapidly diffusing HU-PAmcherry, σ = 60 nm and 55 nm in fast and slow growth, respectively (Fig. A3.1). Details of strain construction, growth conditions, and imaging conditions are provided in Methods.

Typical single-molecule images are shown for 30S-YFP and for HU-PAmcherry in Fig. 3.2*A*. The imaged cells were sorted by tip-to-tip length based on phase contrast images in order to minimize broadening of the spatial distributions due to the range of cell lengths.

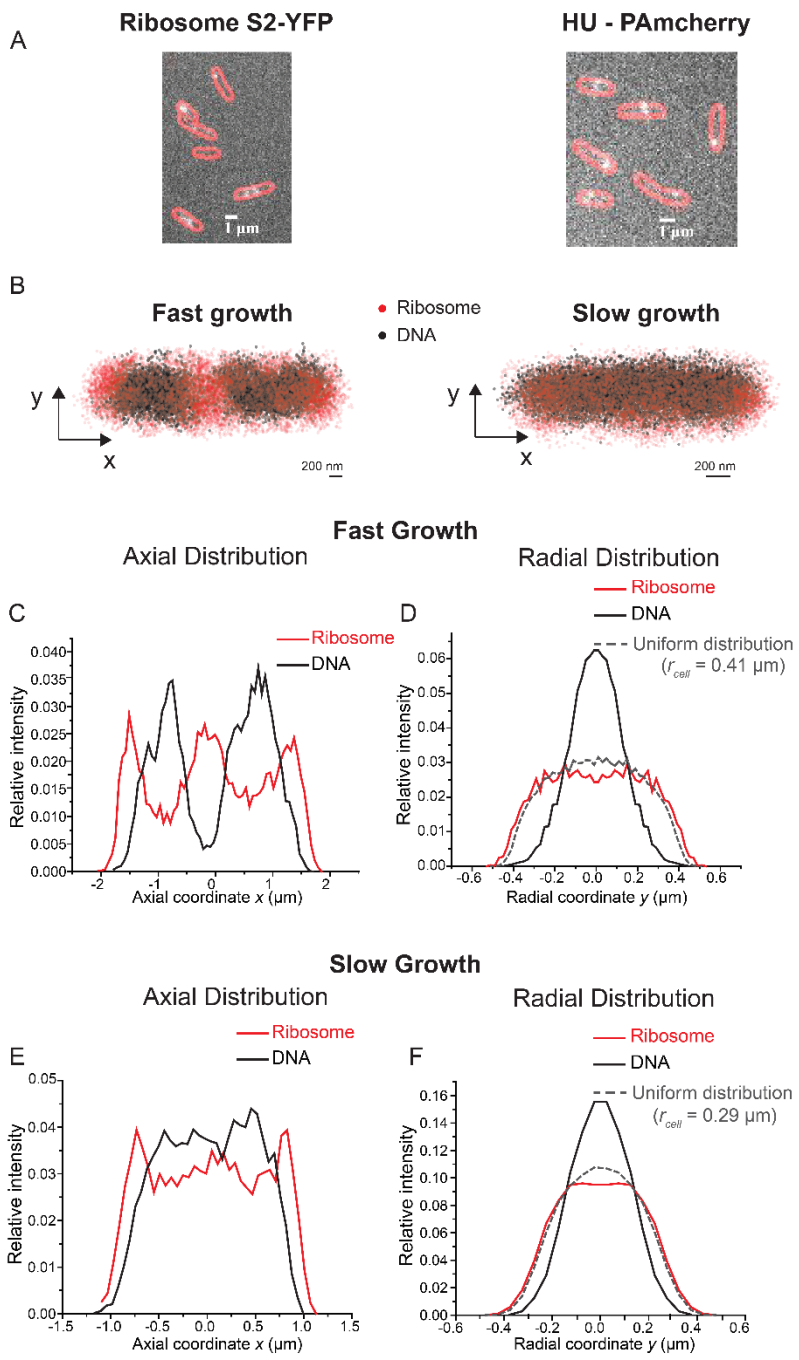


Fig. 3.2 (**A**) Representative examples of single-molecule images of ribosome S2-YFP (*Left*) and HU–PAmcherry (*Right*) with cell outlines shown in red. (**B**) *Left*: Composite of 2D spatial distributions of 9839 ribosomes (*Red*) and 8762 HU copies (*Black*) imaged in 8 fast growing cells of length in the range 3.8-4.2 μm. *Right*: Composite of 2D spatial distributions of 10,251 ribosomes (*Red*) and 13,071 HU copies (*Black*) imaged in 27 slow growing cells of length in the range 2.1-2.5 μm. Note different scales. (**C**) Axial (*x*-axis) probability density distribution of ribosomes (*Red*) and HU (*Black*) in fast growing cells. (**D**) Radial (*y*-axis) probability density distribution of ribosomes (*Red*) and HU (*Black*) in fast growing cells. Radial distribution for uniformly filled cylinder of radius r_{cell} = 0.41 μm is shown for comparison (*Dashed line*). (**E**) Axial (*x*-axis) probability density distribution of ribosomes (*Red*) and HU (*Black*) in slow growing cells. (**F**) Radial (*y*-axis) probability density distribution of ribosomes (*Red*) and HU (*Black*) in slow growing cells. Radial distribution for uniformly filled cylinder of radius r_{cell} = 0.29 μm is shown for comparison (*Dashed line*).

We combine all locations of red HU and green 30S molecules imaged in cells of essentially the same length. For the fast growth condition, we chose cells whose length ranges from 3.8 to 4.2 μm, the bin with the highest number of imaged cells. For fast growing cells, the resulting composite 2D distributions of 9839 ribosomal S2-YFP copies and 8762 HU–PAmcherry copies from 8 cells are shown in Fig. 3.2*B* (Left). In fast growth, the projected 1D axial distributions of DNA and ribosomes are strongly anti-correlated with each other (Fig. 3.2*C*), in agreement with earlier work by Bakshi *et al* using widefield imaging of ribosome S2-YFP and DRAQ5 staining the DNA (1). The axial distribution for all 30S copies shows a peak-to-trough ratio of 3:1 between the ribosome-rich regions and the densest part of the nucleoids. In the chosen length bin, the DNA has segregated into two nucleoid lobes (Fig. 3.2*C*). The axial distribution of HU–PAmcherry molecules shows two peaks with a peak-to-trough ratio of 7:1. Evidently there are some ribosomes within the nucleoids, but little or no DNA in the three ribosome-rich regions (1).

The projected 1D radial (short, *y*-axis) distributions of ribosomes and DNA in fast growing cells are shown in Fig. 3.2*D*. These radial profiles exclude the end cap regions as well as the ribosome-rich region in the central part of the cell in order to get a clearer picture of the degree of radial DNA-ribosome segregation. To minimize noise in this smaller sample, we symmetrized the radial ribosomal profile about the *x*-axis (long cell axis). The ribosome distribution extends outward to the cytoplasmic membrane. The dashed line shows a projected radial distribution for a model of a uniform distribution in a cylinder with radius r_{cell} = 0.41 μm, chosen from modeling of the spatial distribution of Kaede (Fig. A3.2). In comparison with the uniform distribution, the ribosome concentration is slightly depleted near the *x*-axis (where $y = 0$). The symmetrized radial distribution of the DNA marker HU is much narrower and strongly depleted in the annular regions near the cytoplasmic membrane.

For the slow growing cells, the length bin with the highest number of imaged cells is 2.1 to 2.5 μm. This length bin primarily samples cells before DNA segregation (28); there is a single nucleoid lobe. For 10,251 S2-YFP copies and 13,071 HU-PAmcherry copies imaged in 27 slow growing cells, the corresponding 2D scatter plot of locations is shown in Fig. 3.2*B* (Right). The projected axial distributions are shown in Fig. 3.2*E*. The peak-to-trough ratio in the axial distribution of ribosomes in slow growth conditions is only ~1.3:1. The axial distribution of HU molecules shows a single broad peak; DNA is depleted near the endcaps. The symmetrized radial distributions excluding the two endcap regions (Fig. 3.2*F*) show that the DNA radial distribution is narrower than that of ribosomes. Comparison with the model radial distribution for a uniform distribution (cylinder with r_{cell} = 0.29 μm; see Methods and Fig. A3.2) shows that ribosomes are somewhat depleted near the central *x*-axis, *i.e.*, within the nucleoid region. DNA is completely excluded from the annular region near the cytoplasmic membrane.

Degree of segregation of ribosomes and DNA in fast and slow growth conditions

Superficially, the 1D projections suggest that the ribosome-DNA segregation is stronger in fast growth than in slow growth. Next we return to the measured 2D projected locations (Fig. 3.2*B*, *C*) seeking to quantify the degree of segregation of ribosomes and DNA. Our metric is a modification of the commonly used Pearson Correlation Coefficient (PCC) (29, 30), adapted for 2D projections of 3D distributions from the *E. coli* spherocylindrical geometry. The standard PCC measures the degree of linear correlation between two pixelated image matrices **R** (red channel, here HU) and **G** (green channel, here ribosomes), on a scale from –1 (perfect anti-correlation) to +1 (perfect positive correlation) (29, 30). Standard PCC compares every matrix element of **R** and **G** with the constant mean value $\bar{R}$ or $\bar{G}$ and calculates the normalized correlation coefficient between these positive and negative deviations. However, the standard

PCC procedure fails both quantitatively and qualitatively for spatial distributions of molecules imaged in spherocylindrical cells such as *E. coli* (31). The projection of molecules randomly distributed in a 3D spherocylindrical volume is non-uniform in 2D, owing to strong curvature at the two endcaps and in the cylindrical region. As a result, standard PCC returns the same +1 value for 2D projections of distributions that are either perfectly correlated in 3D or completely uncorrelated in 3D. The standard PCC also systematically underestimates the degree of anti-correlation between the projections of two perfectly anti-correlated 3D distributions. The reasons for these failures of standard PCC are described in detail elsewhere (31).

We therefore applied a modified PCC procedure (called MPCC) that compares individual elements of the image matrices **R** and **G** with the matrix elements of the 2D projection of a large set of molecules distributed randomly within a 3D spherocylinder (Eq. 3.1 in Methods). For the composite images from fast growing cells (Fig. 3.2*B*, Left), we pixelated the spatial distribution of N_G = 9839 ribosomes and N_R = 8762 HU molecules into 50 nm x 50 nm bins to yield the experimental image matrices **R** and **G** (Fig. 3.3*A*). The mean number of molecules per pixel is 8.8 and 7.9 for the ribosome and HU channels respectively, which is greater than the recommended mean occupancy of at least 7 molecules per pixel for accurate estimation of MPCC values (31). To form the MPCC reference distributions (Fig. 3.3*B*), we simulated two random 3D distributions of 100,000 molecules each filling a spherocylinder, corresponding to the ribosome (green) and HU (red) channels. The dimensions of the randomly filled spherocylinder are the same as those of the imaged cells, length L_{cell} = 3.82 μm and diameter $2r_{cell}$ = 0.82 μm. The resulting reference images are normalized to have the same number of molecules as were imaged for ribosomes and HU, generating the matrices $\widetilde{\mathbf{U}}^{\mathbf{G}}$ and $\widetilde{\mathbf{U}}^{\mathbf{R}}$ shown in Fig. 3.3*B*. We then subtracted $\widetilde{\mathbf{U}}^{\mathbf{R}}$ and $\widetilde{\mathbf{U}}^{\mathbf{G}}$ from the corresponding image matrices **R** and **G** and

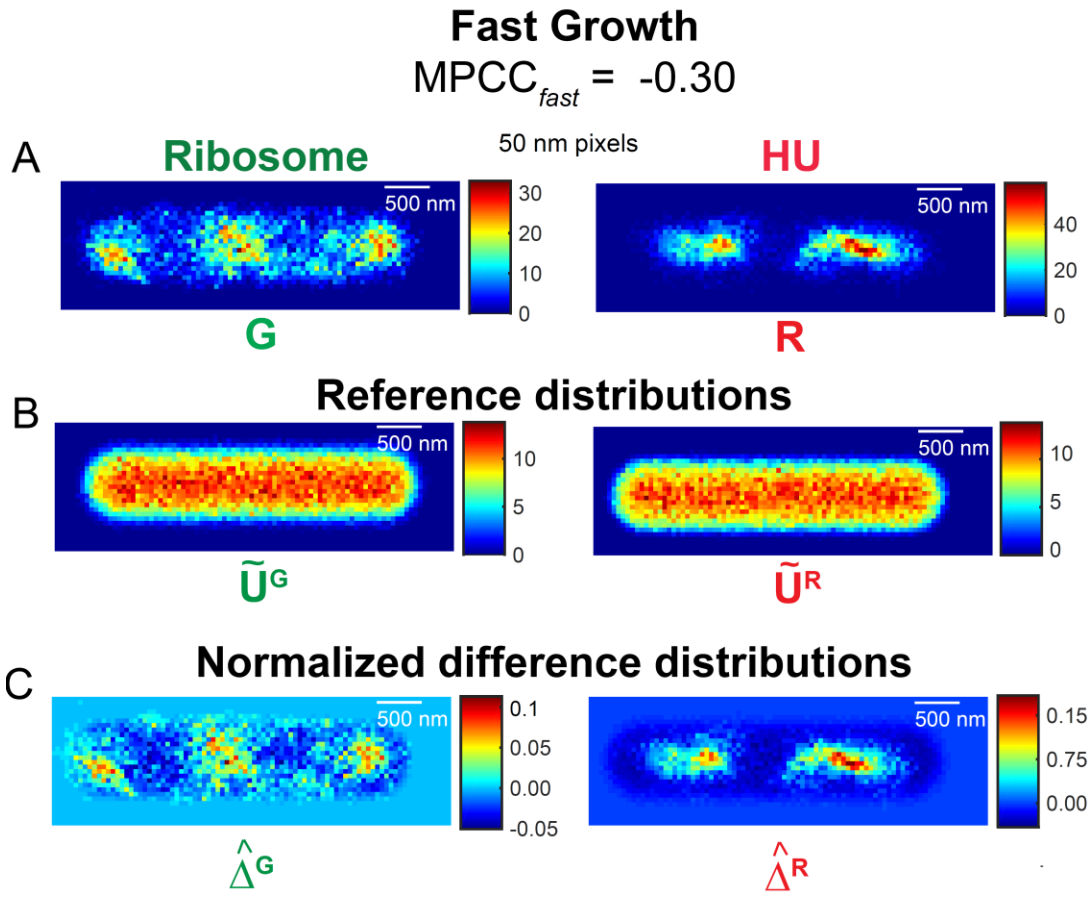


Fig. 3.3 Calculation of modified Pearson Correlation Coefficient (MPCC) between ribosome and DNA (HU) spatial distributions in fast growing cells. (**A**) 2D localization probability density maps of 9839 ribosomes (**G**, *Left*) and 8762 HU (**R**, *Right*) molecules imaged in fast growing cells. Color scale indicates number of molecules in each 50 nm pixel. (**B**) Reference distributions $\tilde{\mathbf{U}}^{\mathbf{G}}$ (*Left*) and $\tilde{\mathbf{U}}^{\mathbf{R}}$ (*Right*), which are 2D projections of 3D random distributions of 100,000 molecules within the spherocylinder (L_{cell} = 3.82 μm, r_{cell} = 0.41 μm) and normalized to have same total number of molecules as **G** and **R**, respectively. $\tilde{\mathbf{U}}^{\mathbf{G}}$ and $\tilde{\mathbf{U}}^{\mathbf{R}}$ are subtracted from images **G** and **R**, respectively. (**C**) Normalized difference matrices $\hat{\Delta}^{\mathbf{G}}$(*Left*) and $\hat{\Delta}^{\mathbf{R}}$ (*Right*) obtained after subtraction. The Frobenius inner product of these two difference matrices gives the value MPCC$_{fast}$ = –0.30 for fast growth conditions.

normalized the difference matrices to generate $\widehat{\boldsymbol{\Delta}}^{\mathbf{R}}$ and $\widehat{\boldsymbol{\Delta}}^{\mathbf{G}}$ (Fig. 3.3*C*), so that the sum of the squares of individual pixel values in each difference matrix is 1. The Frobenius inner product of the two normalized matrices $\widehat{\boldsymbol{\Delta}}^{\mathbf{R}}$ and $\widehat{\boldsymbol{\Delta}}^{\mathbf{G}}$ (Eq. 3.1) gives the result $MPCC_{fast} = –0.30$, indicating spatial anti-correlation.

Finally, we estimated the probability p that a value of $MPCC_{fast} = –0.30$ would be obtained from two random 3D distributions in spherocylinders of the same dimensions as the imaged fast growing cells, with the same number of imaged molecules and the same pixel size used for the experimental data. In Fig. A3.3*A*, we show a histogram of the outcomes of 200 such simulations. The best-fit Gaussian distribution has a mean value $\langle MPCC \rangle = –0.0017$ and standard error $\sigma_{MPCC} = 0.0309$. The measured $MPCC_{fast}$ value lies 9.6 σ_{MPCC} away from zero. Under the assumption that the statistics of the simulated MPCC trials are Gaussian, the probability that two random 3D distributions would produce an MPCC value of magnitude 0.30 or larger on either side of the Gaussian curve is $p \sim 10^{-21}$. Thus we reject the null hypothesis that $MPCC_{fast} = –0.30$ arose from two random, uncorrelated 3D distributions and assert significant negative correlation between the ribosome and HU spatial distributions in fast growth with high confidence.

In the same way, we used the MPCC procedure to quantify the degree of segregation between ribosomes and DNA in slow growing cells. The composite spatial distribution of $N_G = 10{,}251$ ribosome and $N_R = 13{,}071$ HU molecules from 27 slow growing cells is pixelated into 50 nm x 50 nm bins to obtain **R** and **G** (Fig. 3.4*A*). The mean number of molecules per pixel is 16.2 and 20.7 for the ribosome and HU channels. $\widetilde{\mathbf{U}}^{\mathbf{G}}$ and $\widetilde{\mathbf{U}}^{\mathbf{R}}$ shown in Fig. 3.4*B* are generated by normalizing two simulated random distributions of 100,000 molecules in a spherocylinder whose dimensions match those of

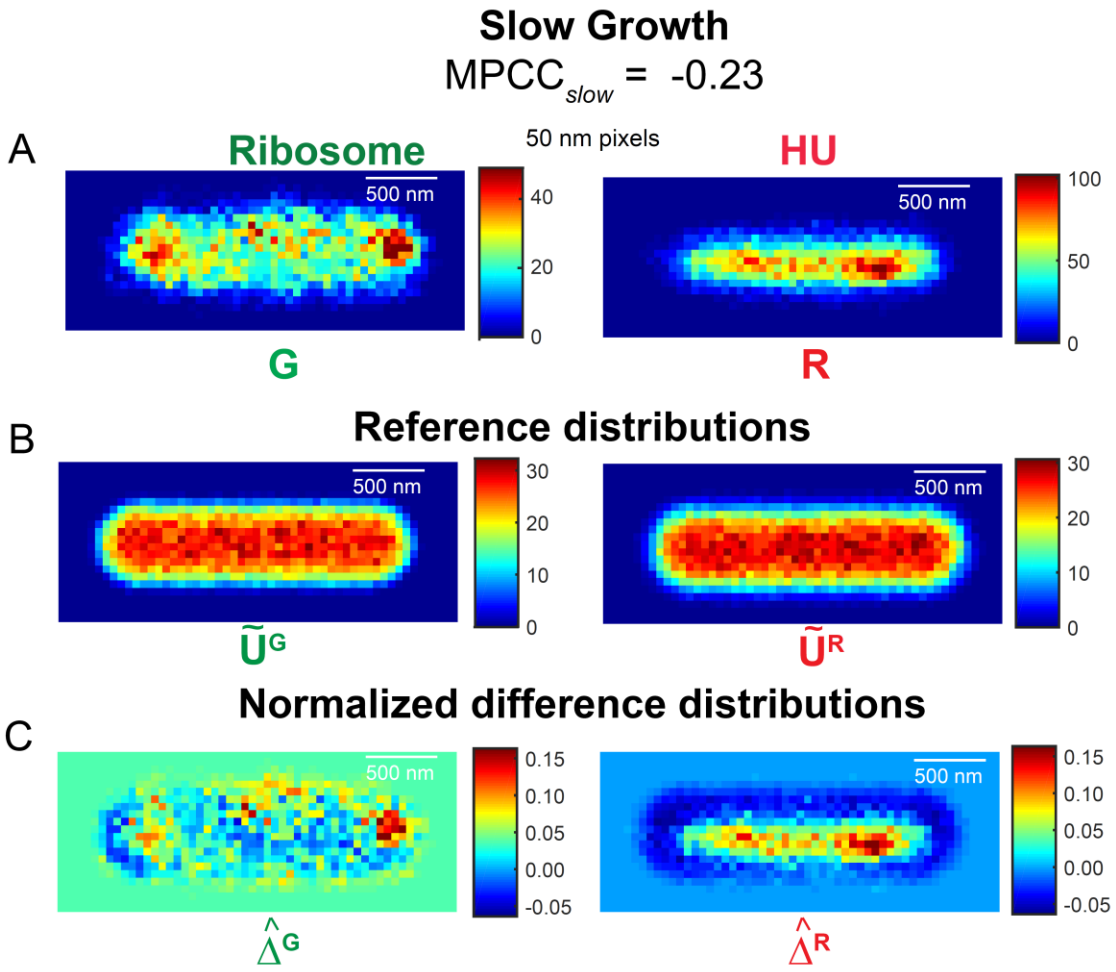


Fig. 3.4 Calculation of MPCC between ribosome and DNA spatial distributions in slow growing cells. (**A**) 2D localization probability density maps of 10,251 ribosomes (**G**, *Left*) and 13,071 HU (**R**, *Right*) molecules imaged in slow growing cells. Color scale indicates number of molecules in each 50 nm pixel. (**B**) Reference distributions $\tilde{\mathbf{U}}^{\mathbf{G}}$ (*Left*) and $\tilde{\mathbf{U}}^{\mathbf{R}}$ (*Right*), which are 2D projections of 3D random distributions of 100,000 molecules within the spherocylinder (L_{cell} = 2.14 μm, r_{cell} = 0.29 μm) and normalized to have same total number of molecules as **G** and **R**, respectively. $\tilde{\mathbf{U}}^{\mathbf{G}}$ and $\tilde{\mathbf{U}}^{\mathbf{R}}$ are subtracted from images **G** and **R**, respectively. (**C**) Normalized difference matrices $\hat{\Delta}^{\mathbf{G}}$(*Left*) and $\hat{\Delta}^{\mathbf{R}}$ (*Right*) obtained after subtraction. The Frobenius inner product of these two difference matrices gives the value $MPCC_{slow}$ = –0.23 for slow growth conditions.

chosen slow growing cells, length L_{cell} = 2.14 μm and diameter $2r_{cell}$ = 0.58 μm. We followed the procedure described above to calculate $MPCC_{slow}$ = –0.23 and $p \sim 10^{-6}$. Since the p value is small, we can again assert significant negative correlation between ribosomes and HU in slow growth with high confidence.

The estimated values $MPCC_{fast}$ = –0.30 and $MPCC_{slow}$ = –0.23 for ribosome and HU spatial distributions in the two growth conditions are both negative and significantly different from the value MPCC = 0 for random, uncorrelated distributions. In both growth conditions, the ribosomes and DNA are segregated from each other. The apparent degree of segregation of ribosomes from DNA in slow growth conditions is only slightly smaller than that for fast growing cells. These 2D MPCC values surely underestimate the degree of segregation in the actual 3D *E. coli* cells. Ribosomes at the top and bottom of the 3D annular region near the cytoplasmic membrane are projected into the nucleoid region, making a strong positive contribution to the 2D MPCC even though they are segregated from the DNA in 3D.

Classification of translating 70S ribosomes and free 30S subunits in fast and slow growth

We also carried out single-30S tracking experiments in both fast and slow growth conditions in an attempt to distinguish free 30S subunits from 30S subunits that are incorporated into translating 70S ribosomes. The smaller free 30S subunits should diffuse more rapidly than the larger translating 70S copies, most of which are incorporated into polysomes. By modeling the distribution of single-step displacements as arising from two populations that do not exchange on the 30-ms imaging timescale, we can provide an estimate of the fraction of translating 70S ribosomes in both fast and slow growth conditions. The two-state analysis will also enable estimation of separate spatial distributions for free 30S subunits and for translating

70S ribosomes in both growth conditions. The photoswitchable fluorescent protein mEos2 (32) gives us better control over the average number of fluorescent copies in each camera frame than the YFP label. For the ribosome tracking experiments, we fused the gene coding for mEos2 to the C-terminus of the endogenous *rpsB* gene (which encodes S2). Details regarding the construction of this strain were given previously (14). The imaging conditions and tracking analysis are described in Methods. Only 0, 1, or 2 30S-mEos2 molecules per cell were activated in each frame to enable accurate tracking of each copy over successive frames.

We selected trajectories that last at least seven frames; longer trajectories were truncated to seven frames. Examples are shown in Fig. A3.4. The chosen trajectories were divided into individual steps and single-step displacements were calculated and binned to form a distribution $P_{ribo}(r)$ in both fast and slow growth (Fig. 3.5). The static two-state model assumes that all translating 70S/polysomes have diffusion coefficient D_{slow} and all non-translating, free 30S subunits have diffusion coefficient D_{fast}. Each distribution of experimental single-step displacements $P_{ribo}(r)$ was fit in a least-squares sense to a numerical model function $P_{model}(r)$, which includes two populations diffusing under confinement within a spherocylinder of appropriate length and diameter. The three fitting parameters are D_{slow}, D_{fast} and the fractional population of slow, 70S ribosomes f_{slow}, which in turn fixes the fractional population of fast, free 30S copies, $f_{fast} = (1 - f_{slow})$. The fitting procedure is outlined in Methods and was described in detail earlier (33).

For fast growth (Fig. 3.5*A*), the experimental $P_{ribo}(r)$ distribution comprises 88,506 individual steps from 14,751 trajectories truncated to six steps, if they lasted longer.
We collected these trajectories from 271 different cells of varying lengths. Since the ribosomes are slowly diffusing molecules that seldom suffer from confinement by the cell boundary, we

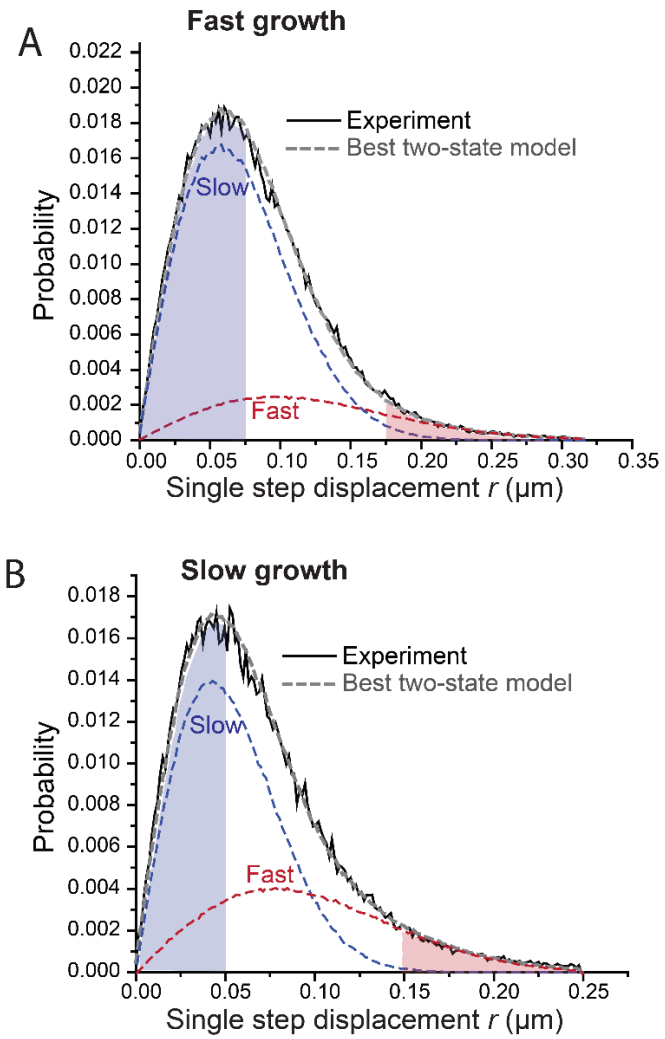


Fig. 3.5 Distinguishing translating 70S and free 30S ribosomes in fast and slow growth conditions from single step displacements. (**A**) Black: Probability distribution of experimental single step displacements taken by ribosome S2–mEos2 copies in fast growing cells during 30 ms frame interval. Dashed Line: Best fit to static two–state model distribution with parameters f_{slow} = 0.8, D_{slow} = 0.02 μm^2/s and D_{fast} = 0.2 μm^2/s. (**B**) Black: Probability distribution of experimental single step displacements taken by ribosome S2–mEos2 molecules in slow growing cells during 30 ms frame interval. Dashed Line: Best fit static two–state model distribution with parameters f_{slow} = 0.65, D_{slow} = 0.009 μm^2/s and D_{fast} = 0.13 μm^2/s. In both (A) and (B), the slow (Blue) and fast (Red) components of the best fit model distribution are shown as dashed lines as labeled. The blue and red shaded regions denote single step displacements that were classified as taken primarily by translating 70S ribosomes and by free 30S ribosomes, respectively.

include trajectories from all imaged cells. The best fit $P_{model}(r)$ shown in Fig. 3.5*A* used parameters f_{slow} = 0.80 ±0.07, D_{slow} = 0.020 ± 0.007 μm^2/s and D_{fast} = 0.20 ± 0.03 μm^2/s (χ_v^2 = 1.46). For slow growth (Fig. 3.5*B*), the experimental $P_{ribo}(r)$ distribution comprises 112,146 individual steps from 378 different cells. The best fit $P_{model}(r)$ shown in Fig. 3.5*B* used parameters f_{slow} = 0.65 ±0.05, D_{slow} = 0.009 ± 0.003 μm^2/s and D_{fast} = 0.13 ± 0.02 μm^2/s (χ_v^2= 1.55). Bootstrapping the data enabled us to obtain error estimates (± 1 standard deviation) for each of the fitted parameters.

We infer a moderate decrease in the fraction of 30S subunits incorporated into translating 70S ribosomes from 0.80 ± 0.07 in fast growth to 0.65 ± 0.05 in slow growth conditions. This trend is consistent with recent biochemical measurements of the fraction of translating ribosomes vs growth rate by Hwa and co-workers (34). According to the best-fit values of D_{slow}, in both growth conditions the root-mean-square single-step displacement of a translating ribosome is comparable to the localization error (σ ~ 30 nm). The difference in the fitted values of D_{slow} between the two growth conditions may not be significant.

Spatial distributions of free 30S subunits and translating 70S ribosomes

As shown in Fig. 3.5, the slow and fast components of the experimental distribution of single-step displacements $P_{ribo}(r)$ from the 30S-mEos2 tracking experiments show significant overlap for both growth conditions. This makes it difficult to determine whether any particular copy is fast or slow. To characterize the spatial distributions of slow and fast copies, we defined two cutoff values for each distribution $P_{ribo}(r)$. Steps shorter than the lower cutoff value r_{slow} are likely to come from the translating, 70S population. Steps longer than the higher cutoff value r_{fast} are likely to come from the searching, free 30S population. For fast growing cells, we chose r_{slow} = 0.075 μm and r_{fast} = 0.175 μm (Fig. 3.5*A*).

For slow growing cells we chose r_{slow} = 0.05 μm and r_{fast} = 0.15 μm (Fig. 3.5*B*). The blue and red shaded regions show the range of single-step displacements that are classified as translating 70S and free 30S copies and suggest in each case the purity of the assignments, according to the two-state model.

For fast growth conditions, we analyzed 13,716 displacements from 52 cells whose tip-to-tip length varied from 3.8 to 4.2 μm based on phase contrast images. Trajectories that lasted longer than seven frames (six single steps) were selected. We calculated single-step displacements *r* for all individual steps of the trajectories. The midpoint of the two endpoints (x_1, y_1) and (x_2, y_2) for each single step was defined as the location of that step. Each step was classified as coming from translating 70S or free 30S according to the cutoff prescription described above. In Fig. 3.6*A*, *B*, we show 2D probability density heat maps of the spatial locations of short steps (translating 70S) and of long steps (free 30S). Both translating 70S and free 30S are concentrated at the endcap regions and in the central region between the two nucleoid lobes, as seen in the axial distribution symmetrized about *y*-axis (Fig. 3.6*C*). The radial distributions are obtained only from molecules imaged in the two nucleoid lobes; they exclude the endcap regions and the central ribosome–rich region and are symmetrized about the *x*-axis (Fig. 3.6*D*). The radial distributions of translating 70S and free 30S are very different from each other. The translating 70S copies concentrate in the annular regions near the cell membrane and avoid the nucleoids. The free 30S copies readily access the nucleoids but are sparse near the cell membrane.

For slow growing cells (Fig. 3.7), 40,854 displacements from 87 cells in the length bin 2.1–2.5 μm were classified as translating 70S or free 30S based on the cutoffs r_{slow} and r_{fast}

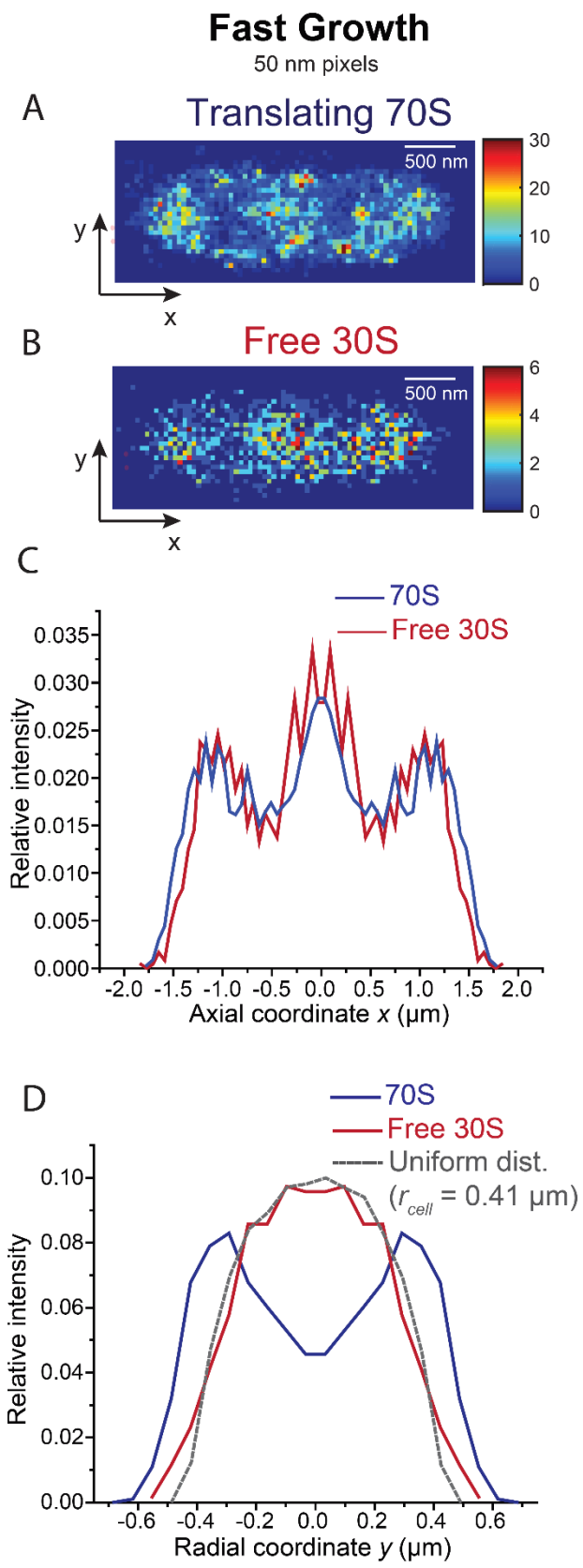


Fig. 3.6 Spatial organization of translating 70S and free 30S ribosomes in fast growing cells. (**A**) 2D localization probability density maps of translating 70S ribosomes (blue swath in Fig. 3.5*A*). Color scale indicates number of molecules in each 50 nm pixel. (**B**) 2D localization probability density maps of free 30S ribosomes (red swath in Fig. 3.5*A*). Color scale indicates number of molecules in each 50 nm pixel. (**C**) Axial probability density distribution of translating 70S (*Blue*) and free 30S (*Red*) ribosomes. (**D**) Radial probability density distribution of translating 70S (*Blue*) and free 30S (*Red*) ribosomes. Radial distribution of uniformly filled cylinder (r_{cell} = 0.41 μm) is shown for comparison (*Dashed Line*).

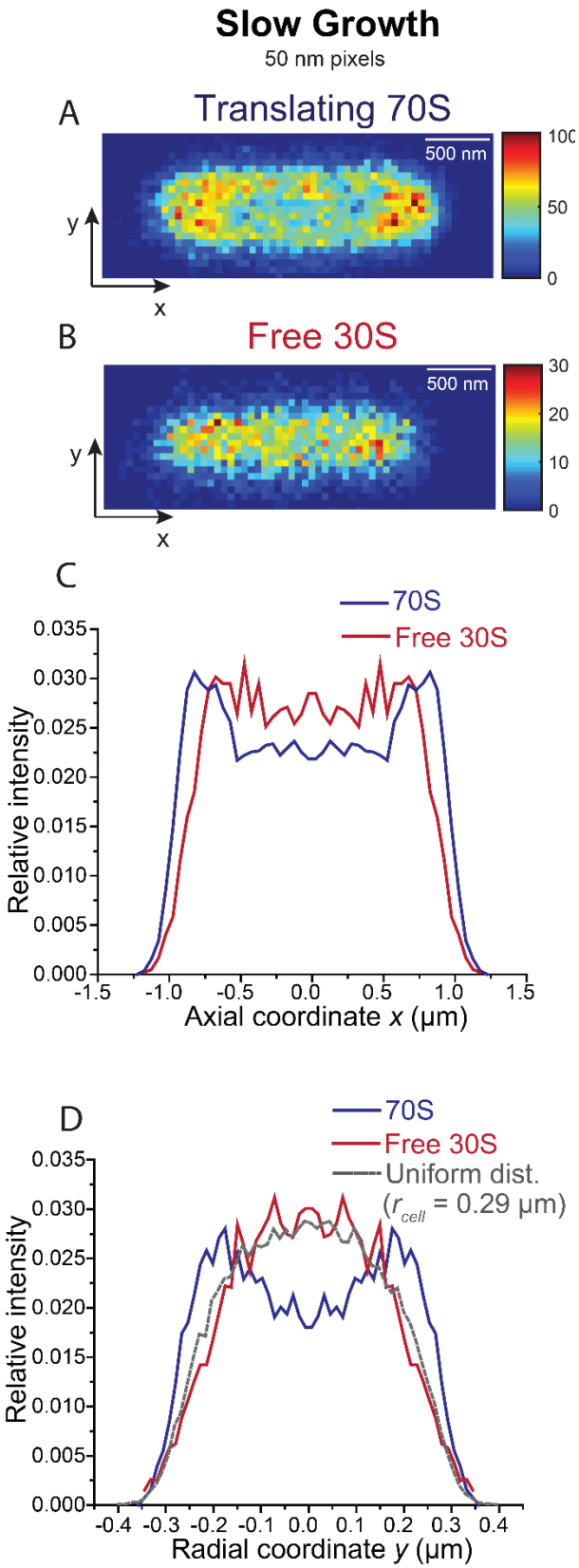


Fig. 3.7 Spatial organization of translating 70S and free 30S ribosomes in slow growing cells. (**A**) 2D localization probability density maps of translating 70S ribosomes (blue swath in Fig. 3.5*B*). Color scale indicates number of molecules in each 50 nm pixel. (**B**) 2D localization probability density maps of free 30S ribosomes (red swath in Fig. 3.5*B*). Color scale indicates number of molecules in each 50 nm pixel. (**C**) Axial probability density distribution of translating 70S (*Blue*) and free 30S (*Red*) ribosomes. (**D**) Radial probability density distribution of translating 70S (*Blue*) and free 30S (*Red*) ribosomes. Radial distribution of uniformly filled cylinder (r_{cell} = 0.29 μm) is shown for comparison (*Dashed Line*).

(Fig. 3.5*B*) and located at the midpoint of each step. The location heat map suggests that slowly diffusing ribosomes concentrate in the endcaps (Fig. 3.7*A*). The symmetrized axial distribution of translating 70S ribosomes is indeed concentrated in the endcaps and somewhat depleted in the central part of the cell where the DNA resides (Fig. 3.7*C*). In contrast, the symmetrized axial distribution of free 30S shows low concentration in the endcap tips and higher concentration in the nucleoid region. The radial distributions (Fig. 3.7*D*) exclude molecules in the endcap regions and are symmetrized about *x* axis. The radial distribution of the translating 70S ribosomes peaks away from the central cell axis and extends all the way to the cytoplasmic membrane. In contrast, the radial distribution of the free 30S subunits is essentially uniform (Fig. 3.7*D*), as shown by comparison with a uniformly filled model cylinder of radius r_{cell} = 0.29 μm.

Slow growth: Partitioning of 70S and free 30S species inside and outside of nucleoid

In our slow growth conditions, we can use the measured axial and radial spatial distributions of fast and slow diffusing copies (free 30S subunits and translating 70S ribosomes, respectively) and a simple geometric model of the nucleoid region to estimate how many free 30S and translating 70S copies reside inside and outside of the nucleoid. We have used the 30S-YFP construct to estimate the mean total 30S copy number to be ~21,000 per cell for slow growing cells in the 2.1–2.5 μm length bin (Methods, Fig. A3.5). According to the analysis of Fig. 3.5*B*, this includes ~7350 free 30S subunits (35%) and ~13,650 70S copies having 30S incorporated into translating 70S ribosomes (65%). In the simplified 3D geometric model depicted at the top of Fig. 3.8, the full cell volume is represented by a spherocylinder of tip-to-tip length L_{cell} = 2.14 μm, radius r_{cell} = 0.29 μm, and volume V_{cell} = 0.51 μm^3. These dimensions are based on model fits to the measured distribution of Kaede copies in slow growing cells in the 2.1-2.5 μm length bin, as estimated by phase contrast (Methods, Fig. A3.2). Slow growing cells

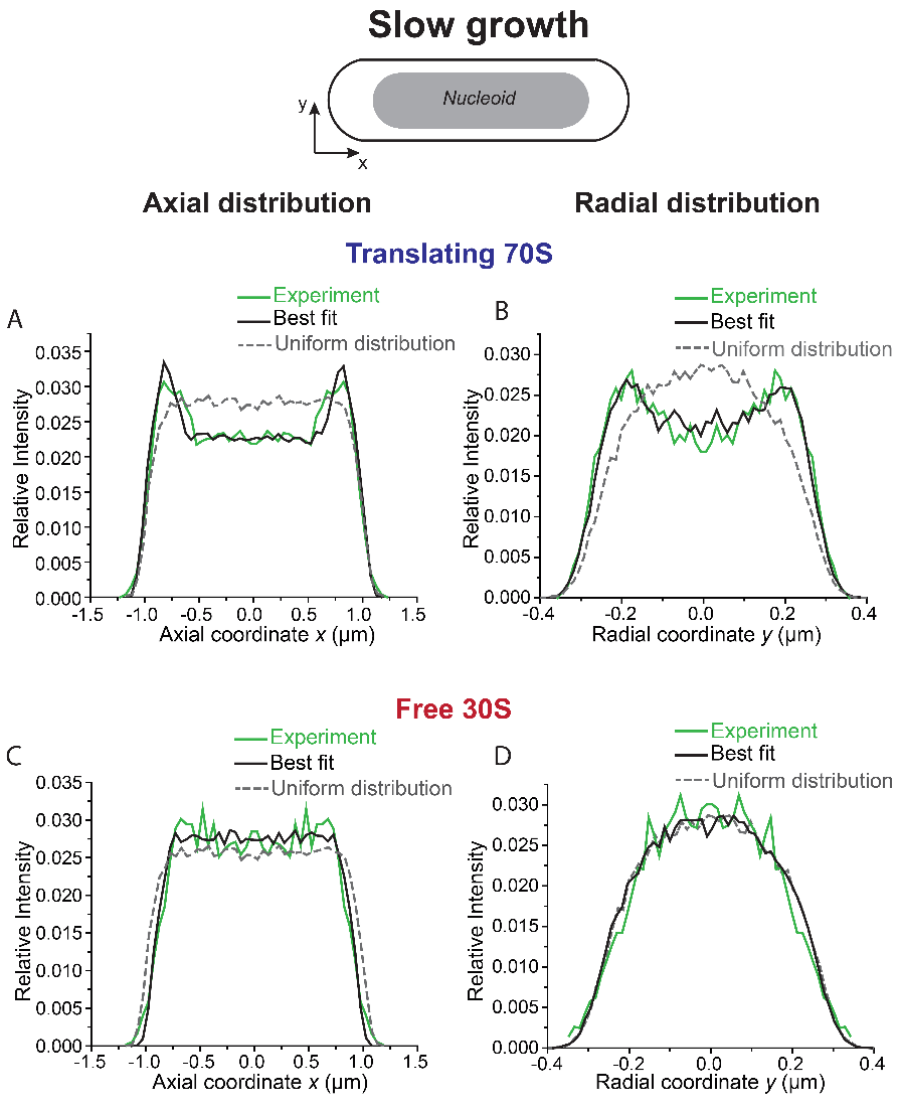


Fig. 3.8 Estimation of the spatial partitioning of translating 70S and free 30S ribosomes inside and outside of the nucleoid of slow growing cells. Nucleoid is modeled as a spherocylinder (*Top*) whose radius $r_{nucleoid}$ = 0.19 μm and length = 1.55 μm were chosen to fit the measured HU distribution (Appendix). (**A**) *Green*: Experimental axial distribution of translating 70S ribosomes. *Black*: Best fit model axial distribution (80% deletion of points inside nucleoid). *Dashed Line*: Axial distribution of uniformly filled spherocylinder (L_{cell} = 2.14 μm, r_{cell} = 0.29 μm) is shown for comparison. (**B**) *Green*: Experimental radial distribution of translating 70S ribosomes. *Black*: Best fit model radial distribution (80% deletion within nucleoid). *Dashed Line*: Radial distribution of uniformly filled cylinder (r_{cell} = 0.29 μm,) is shown for comparison. (**C**) *Green*: Experimental axial distribution of free 30S ribosomes. *Black*: Best fit model axial distribution (uniformly filled spherocylinder of L = 2 μm, r = 0.29 μm). *Dashed Line*: Axial distribution of uniformly filled spherocylinder (L_{cell} = 2.14 μm, r_{cell} = 0.29 μm) is shown for comparison. (**D**) *Green*: Experimental radial distribution of free 30S ribosomes. *Black*: Best fit model radial distribution (uniformly filled spherocylinder of L = 2 μm, r = 0.29 μm). *Dashed Line*: Radial distribution of uniformly filled cylinder (r_{cell} = 0.29 μm) is shown for comparison.

have a single nucleoid lobe and a fairly smooth distribution of DNA density, as shown in Fig. A3.6*A*. The nucleoid is represented by a shorter, narrower spherocylinder of dimensions $L_{nucleoid}$ = 1.55 μm, $r_{nucleoid}$ = 0.19 μm, and $V_{nucleoid}$ = 0.16 μm^3 = 0.31 V_{cell}. These were determined from spatial distributions of HU copies in slow growing cells within the same length bin, as detailed in Methods (Fig. A3.6).

Then we determined the fraction of translating 70S copies residing inside the nucleoid that best fits the experimental axial and radial distributions of the slowly diffusing copies (Fig. 3.7*C*, *D*). We randomly filled the model cell volume with 100,000 mathematical points. This places 31,353 points inside the nucleoid and 68,647 points outside the nucleoid. Then we randomly deleted 0%, 10%, 20%, …, and 100% of the 31,353 points within the nucleoid. For each deletion percentage, we broadened the spatial distribution of the remaining points according to the experimental localization error (σ = 27 nm, determined from MSD plot for S2-mEos2 in Fig. A3.11) and then computed a model axial and radial spatial distribution for comparison with experiment. The best fit in a least-squares sense (black lines in Fig. 3.8*A*, *B*) has deleted 80% = 25,082 points from the nucleoid region, leaving 6271 points inside the nucleoid and the original 68,647 points outside. In Fig. A3.7, we show that the fits to radial and axial distributions taken together becomes significantly worse when 70% or 90% of the points inside the nucleoid are deleted, *i.e.*, the procedure is sensitive to ±10% changes in the percentage deleted. A total of 74,918 points survive in the best-fit model, 8.4% inside and 91.6% outside the nucleoid. In terms of the estimated translating 70S ribosome copy number of 0.65 x 21,000 =13,650, the model estimates that ~1150 70S copies reside inside the nucleoid of slow growing cells, and ~12,500 70S copies reside outside the nucleoid.

For the free 30S copies, we found that the experimental axial and radial distributions are well matched by a uniform distribution of points within a spherocylinder of tip-to-tip length L_{cell} = 2.00 μm (140 nm shorter than the estimate from Kaede localizations) and radius r_{cell} = 0.29 μm (matching the Kaede estimate, Fig. A3.2). The need to shorten the model cell length by 70 nm at each tip is an artifact of our choice of very long steps (r > 150 nm in length; Fig. 3.5*B*) to obtain a fairly pure set of rapidly diffusing 30S copies. The midpoints of such long steps are necessarily depleted at the very tips of the cell. Detailed comparison of the axial distributions of static molecules and molecules with single step displacements longer than 150 nm (Fig. A3.8) shows that indeed the chosen long steps are depleted from the tips of the cell. This effect is less pronounced for the radial distribution. Since a uniform distribution of free 30S copies fits the axial and radial distributions of fast diffusing copies quite well (Fig. 3.8*C*, *D*) and the ratio $V_{nucleoid}/V_{cell}$ = 0.31, we estimate that 31% of the free 30S copies reside inside the nucleoid region while 69% reside outside. In terms of the estimated copy number of free 30S subunits (0.35 x 21,000 = 7350 per cell), this indicates ~2280 free 30S copies inside the nucleoid and ~5070 free 30S copies outside the nucleoid.

Combining the two results yields the estimate of ~3430 free, non-translating 30S plus translating 70S copies inside the nucleoid (16.3% of all ribosome species) and 17,570 30S plus 70S copies outside the nucleoid (83.7% of all ribosome species). As a cross-check on the modeling procedure, we carried out an analogous inside/outside fitting procedure for the axial and radial distributions of total 30S copies, regardless of step length (Fig. A3.9). Again we distributed 100,000 points uniformly within V_{cell} and deleted a varying percentage of points from $V_{nucleoid}$. The best fit to the distributions of total 30S copies (black lines in Fig. A3.9) deleted 65%

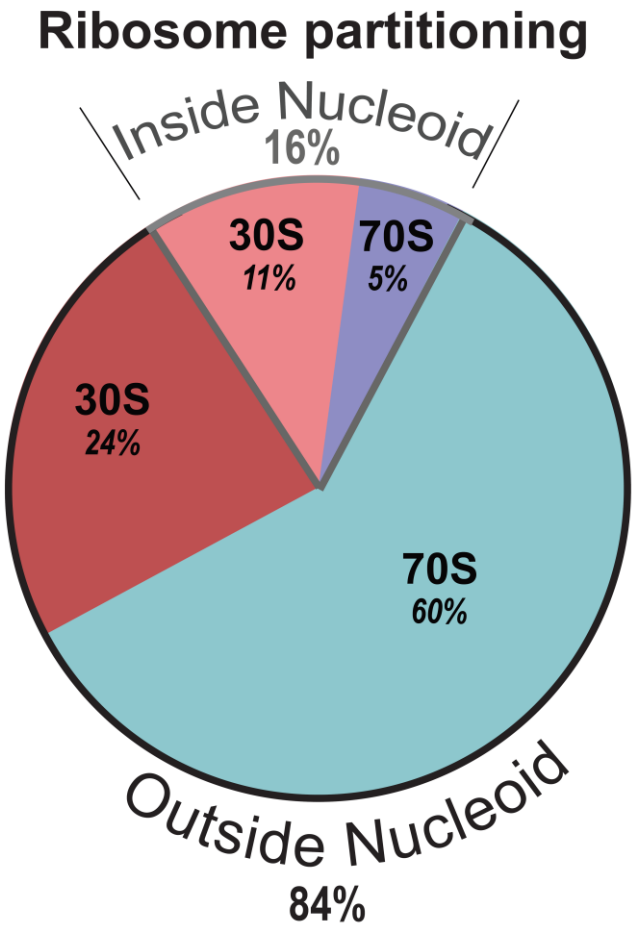


Fig. 3.9 Pie chart showing estimated partitioning of ribosomes in slow growing cells into translating 70S (*Blue*) and free 30S (*Red*) ribosomes overall, inside the nucleoid, and outside the nucleoid as indicated.

of the original 31,353 points within the nucleoid, leaving 10,974 points inside and 68,647 points outside. This yields the independent estimate of 13.8% of the combined 30S plus 70S copies inside the nucleoid and 86.2% outside. This result, which does not rely on two-state fitting of trajectory step lengths, agrees quite well with the result of 16.3% inside and 83.7% outside from the combined results from the separate models for free 30S and translating 70S copies. This agreement lends credence to our effort to quantitatively partition ribosome roles inside and outside the nucleoid.

As summarized in the pie chart of Fig. 3.9, we estimate that the average slowly growing cell has ~21,000 total 30S copies, ~3430 (16% of the total) inside the nucleoid and ~17,570 (84%) outside. Within the nucleoid, ~2280 copies (11% of the total) are free 30S subunits and ~1150 copies (5%) are translating 70S ribosomes, likely incorporated into co-transcriptionally translating polysomes. Outside the nucleoid, ~5070 copies (24% of the total) are free 30S subunits and ~12,500 copies (60%) are incorporated into translating 70S ribosomes

We refrained from attempting similar calculations for fast growing cells due to increased cell to cell heterogeneity in nucleoid size, shape, location, and density. This renders the composite HU distribution both wider and longer than typical single-cell distributions (Fig. A3.10*C*, *D*). A spherocylindrical model of the nucleoid volume is inappropriate in fast growth.

Discussion

Comparison of *E. coli* and *C. crescentus* at similar total DNA content and cell volume

Our original motivation for the present work was the observation of very different spatial organization within the cytoplasm of fast growing *E. coli* and *B. subtilis* compared with

relatively slow growing *C. crescentus*. Strong segregation of the chromosomal DNA (the nucleoid) from the ribosomes was first observed in rapidly growing *B. subtilis* (2). A similar degree of axial and radial segregation has now been observed for *E. coli* in fast (1), moderate (9), and slow growth (present work). Evidently in those two species, most transcription of protein genes occurs in different regions of space than the bulk of translation. In sharp contrast, in the slow-growing species *C. crescentus*, whose doubling time lies in the range ~120–150 min (35), the chromosomal DNA and the ribosomes are apparently thoroughly mixed and occupy the entire cytoplasmic volume (4).

A variety of mechanisms have been proposed to explain why the *E. coli* chromosome occupies only a fraction of the total cytoplasmic volume and why the DNA and ribosomes are strongly segregated from each other. Joyeux provides an excellent summary of possible physical and biochemical contributions to the compaction of the chromosomal DNA (16). These include segregative phase separation due to repulsive forces between polyanionic DNA and the highly negatively charged 70S ribosomes (–4500*e* each); DNA condensation arising from binding to small cations; macromolecular crowding ("depletion forces"); supercoiling; and binding to nucleoid associated proteins. He argues that the strongest forces arise from segregative phase separation. On the other hand, a minimalist model of DNA as a hyperbranched chain of beads and polysomes as freely jointed chains of beads of appropriate size found that the DNA polymer when confined by itself within a spherocylindrical volume already avoids the walls to enhance its configurational entropy (13). Entropic and hard-sphere excluded volume effects alone were sufficient to cause strong DNA-polysome segregation in that very simple model. No electrostatic repulsion was necessary.

Regardless of the underlying causes of segregation in *E. coli*, we wondered if the *C. crescentus* chromosome fills the entire cytoplasm because the cell volume is so small that the combined nucleoid compaction forces are too weak to cause the DNA to occupy a still smaller volume. Ribosomes and DNA would then necessarily mix. An estimate of the *C. crescentus* cytoplasmic volume as that of two truncated cones of radius 250 nm and height 1.5 μm each yields a lower bound on $V_{crescentus}$ of 0.20 μm^3; a spherocylindrical model of radius 250 nm and length 3 μm yields an upper bound of 0.55 μm^3. We take $V_{crescentus}$ = 0.4 μm^3 as a sensible estimate. Our slow growth condition yields an *E. coli* doubling time of 147 min and median cell volume V_{cell} = 0.51 μm^3, both comparable to those of *C. crescentus*. This is nearly four times less volume than the median V_{cell} = 1.88 μm^3 for *E. coli* in our fast growth conditions (doubling time 50 min). Both the slow growing *E. coli* and *C. crescentus* harbor ~1 chromosome per cell (36, 37) and the size of the chromosome is similar, 4.63 Mb and 4.01 Mb (38), respectively. The comparison in terms of cell volume and amount of DNA seems apt; nevertheless, we find substantial nucleoid-ribosome segregation in slow growing *E. coli*. The MPCC metric shows a comparable degree of anti-correlation between 30S ribosomes and the DNA marker HU for *E. coli* in our fast and slow growth conditions. The degree of radial segregation is strong in both growth conditions, and the endcaps concentrate ribosomes and exclude DNA in both cases.

Thus the underlying cause of the different spatial distributions in *E. coli* and *C. crescentus* remains unclear. Perhaps the compacting forces are somehow weaker in *C. crescentus.* There may be significant morphological differences between the *E. coli* chromosome and the *C. crescentus* chromosome. HiC analysis of the latter revealed a “bottlebrush” structure (39); a study of the *E. coli* chromosomal structure at comparable resolution might prove informative. Finally, a recent 3D superresolution study in *C. crescentus* discovered spatial clustering of

ribosomes, suggesting the possibility of localized pockets of segregation of ribosomes from DNA with the pockets widely dispersed throughout the cytoplasm (40). A similar study investigating the spatial distribution of HU in *C. crescentus* suggested clustering of HU in different stages of the cell cycle (41). It might prove informative to extend that work to 3D superresolution imaging of HU and ribosomes in the same cell to test for "patchy" segregation. 3D experiments on *E. coli* would also be informative; with enough data, they would yield 3D correlation coefficients undistorted by the "squashing" of the image from 3D to 2D inherent in the present study.

Degree of nucleoid compaction in fast and slow growth

With one exception (37), there is widespread agreement in the literature that the *E. coli* chromosome is denser ("more compacted") in fast growth than in slow growth (10, 11, 16, 28, 42). That conclusion is sometimes reached based on visual inspection of fluorescent images of nucleoids in various growth conditions (28, 42). Our new data corroborate this idea. In addition, we now attempt quantitative estimates of the peak DNA density on the ~100 nm length scale in our slow and fast growth conditions. In our slow growth condition (147 min doubling time), there is one nucleoid lobe containing ~1 chromosome = 4.6 Mb of DNA (43). The volume of our cylindrical model nucleoid is $V_{nucleoid}$ = 0.16 μm^3, yielding an estimated mean DNA density within the nucleoid of ~2.9 x 10^7 bp/μm^3. The DNA spatial distribution appears smooth on the 100 nm length scale, so the peak and average density are similar. In our fast growth condition (50 min doubling time), there are two nucleoid lobes containing a total of ~2.2 chromosome equivalents = 10.1 Mb of DNA (43). The HU distribution within each nucleoid lobe is typically crescent shaped, with much higher density near the center of the crescent (Fig. A3.10*A*). A spherocylindrical model of the composite HU distribution averaged across cells (Fig. A3.10*C*)

yields $V_{nucleoid}$ = 0.21 μm^3 for the total volume of the two lobes, providing an upper bound on the mean DNA density within each lobe of ~4.8 x 10^7 bp/μm^3. However, as shown in Fig. A3.10*D*, the nucleoid volume in a single cell is perhaps two-fold smaller than the estimate derived from the multi-cell average distribution. In addition, the density in the center of each lobe is perhaps 2.5 times higher than the average (Fig. A3.10*A*). A rough estimate of the peak DNA density on a ~100 nm scale in our fast growth conditions is thus ~2.4 x 10^8 bp/μm^3, some 8 times larger than in slow growth.

Why is the nucleoid "fluffier" in slow growth than in fast growth? One possible explanation derives from the very different spatial distributions of actively transcribing RNA polymerase copies in the two conditions. In fast growth, most transcription occurs on the seven *rrn* operons, six of which have been shown to cluster with each other (44). These "transcription foci" (28) tend to locate at the periphery of the nucleoid (10, 28), perhaps due to excluded volume effects (13). In slow growth, transcription of *rrn* operons is less important, many more protein genes are expressed (45), and transcription of protein genes is broadly distributed throughout the nucleoid (10). Perhaps the widely distributed activity of transcription, which causes local supercoiling and necessitates DNA gyrase activity, somehow causes expansion of the overall DNA volume. The detailed mechanism remains unclear. Perhaps the nucleoid should be considered a form of "active matter", whose overall volume is controlled by the density and distribution of energy dissipation by RNAP and other force-producing proteins (46).

Circulation model of movement of 70S and 30S ribosomal species

Co-transcriptional translation has been shown to be important to prevent backtracking of RNAP, premature termination of transcription, and premature degradation of mRNA by ribonucleases (47-49). Indeed we find clear evidence of translating 70S ribosomes in the core of

the nucleoid in both fast and slow growth (Figs. 3.6 and 3.7). The strong segregation of most translating 70S ribosomes outside the nucleoid combined with the occurrence of co-transcriptional translation within the nucleoid requires that 30S and 50S ribosomal subunits be able to circulate throughout the cell (Fig. 3.1). When a transcript is complete, it is already presumably decorated with 70S ribosomes and diffuses as a polysome, with $D_{slow} \sim 0.01\ \mu m^2/s$. For $r_{nucleoid} \sim 0.19\ \mu m$, the typical nascent mRNA will escape the nucleoid and find the annular ribosome-rich region surrounding the nucleoid very quickly, in ~0.1 s (A3.1). Polysomes in the ribosome-rich annular region will exchange with polysomes in the ribosome-rich endcaps on a timescale of ~10 s (A3.2). The same forces that exclude polysomes from the nucleoid will tend to prevent their re-entry into the nucleoid after escape. As each 70S ribosome within the ribosome-rich regions completes translation of a protein, 30S and 50S dissociate. These free ribosomal subunits will either find a new mRNA within the ribosome-rich regions or return to the nucleoid to initiate another co-transcriptional translation event.

This means that it is functionally important that 30S and 50S subunits be able to penetrate the nucleoid meshwork rapidly to enable co-transcriptional translation within the nucleoid. One might worry that in fast growth, the nucleoid is so dense that it precludes re-entry of ribosomal subunits into the nucleoid. This in fact occurs for the highly condensed nucleoids of erythromycin-treated cells (9). Elf and co-workers earlier showed that for *E. coli* in "medium growth" conditions (doubling time = 120 min at 25°C), free 50S and 30S subunits distribute uniformly across the cytoplasm, indicating ready access of subunits to the bulk of the nucleoid (9). The present work shows (Figs. 3.6 and 3.7) that free 30S subunits distribute uniformly across the cytoplasm for both our fast and slow growth conditions (doubling times of 50 min and 147 min). Although the peak DNA density is ~8 times greater in fast growth than in slow growth, it is

evidently not so high as to exclude ribosomal subunits. A model of ribosomal subunit circulation including co-transcriptional translation (Fig. 3.1) appears to be applicable over the entire range of *E. coli* growth rates studied in this fashion.

Connection with 1D global reaction-diffusion model of transcription and translation

Recently Wingreen and co-workers developed a quantitative numerical model of the *E. coli* transcription-translation machinery based on available experimental input for cells growing with doubling time of 60 min (11). Their steady-state reaction-diffusion model includes many details of the dynamics and spatial distributions of 50S ribosomal subunits, 70S ribosomes, mRNA, and chromosomal DNA. For numerical simplicity, the model is one-dimensional, with all concentrations expressed as molecules/length. The single nucleoid lobe is represented as a distribution of DNA density centered at mid-cell with FWHM of ~60% of the total L_{cell} = 3 μm, or ~1.8 μm. In the 1D model, the ribosome-rich endcaps map onto line segments at either end. An important prediction of the model is that ~90% of mRNAs are segregated within the endcaps, where most translation occurs.

Next we attempt to compare our estimated free 30S and translating 70S partition fractions (Fig. 3.9) with the 1D model results. We use our slow growth data (147 min doubling time), for which we were able to estimate free 30S and translating 70S copy numbers and percentages both inside and outside the nucleoid. The comparison is admittedly imperfect. The 1D model is built for a 60 min doubling time, but like our slow growth condition (147 min), it includes only one nucleoid lobe. The FWHM of 1.8 μm in the 1D nucleoid is similar to the length of our 3D nucleoid model in slow growth, $L_{nucleoid}$ = 1.55 μm. Of necessity, there is no ribosome-rich annular region surrounding the nucleoid in the 1D model.

Within the dense part of the nucleoid region of the 1D model (11), only ~34% of ribosomal subunits are carrying out co-transcriptional translation; ~29% are free subunits and the remaining ~37% are in short, free polysomes whose message was previously transcribed. The only way a free polysome born in the dense nucleoid region can escape to a ribosome-rich endcap is by 1D axial diffusion over an average distance of ~450 nm (A3.3). In our 3D model of the experimental spatial distributions in slow growth (Fig. 3.8), most free polysomes born in the nucleoid region will escape by 2D radial diffusion to the annular shell of translating ribosomes surrounding the nucleoid. In Appendix, A3.3 and Fig. A3.12, we show that ~90% of free polysomes initially uniformly distributed within our 3D model nucleoid volume would escape radially by 2D diffusion rather than axially by 1D diffusion. The average distance traveled to escape is only ~60 nm (A3.1), ~7 times less than in the 1D model. Since the root-mean square distance diffusively traveled in time t is MSD = $(2Dt)^{1/2}$ in 1D (linear diffusion) and MSD = $(4Dt)^{1/2}$ in 2D (radial diffusion), for the same diffusion coefficient the free polysomes in the 1D model will remain in the dense nucleoid region ~100 times longer than they would in our 3D model.

Therefore we feel justified in "re-assigning" the ~37% free polysomes in the 1D nucleoid as the free polysomes in the annular region surrounding the nucleoid in our 3D model. This means that 34% + 37% = 71% of ribosomes within the nucleoid of the 1D model are slow-moving and would be assigned as translating, 70S ribosomes in our experiments. Stated the other way around, for comparison with the 1D model we discard the two endcaps and calculate the fraction of free 30S and translating 70S ribosomes in the remaining volume, which includes both the nucleoid spherocylinder and the surrounding tube of ribosomes. In that volume, our 3D partitioning model finds 63% 70S translating ribosomes and 37% free 30S subunits, in remarkably good agreement with the re-assigned 1D model results. This agreement is

encouraging but it may be fortuitous, given the many assumptions involved in both models. We hope our new data provide impetus for development of a full, three-dimensional reaction-diffusion model of the transcription/translation machinery in slow-growing *E. coli*.

Methods

Strain construction

All experiments were performed on modified *E. coli* strains with a VH1000 background. For superresolution co-imaging of ribosomes and DNA binding protein HU in the same cell, we constructed a strain SM6 in which the S2 ribosomal protein was labelled with eYFP as before (1) and HU was labelled with photoactivatable fluorescent protein PAmcherry (27). The HU encoding gene (*hupA*) was amplified from MG1655 using the primers 5'-ATCGGATCCCGCCGCCGCCTTAACTGCGTC-3' and 5'-GATCGAATTCTAGGAGGTATTCACATGAACAAGACTC-3' and then subcloned into the plasmid pASK–IBA3plus between EcoRI and BamHI sites to generate plasmid pSM5. The PAmcherry encoding gene was amplified with forward primer 5'-GATCGGATCCATGGTGAGCAAGGGCGAGG-3' and reverse primer 5'-GATCAAGCTTTTACTTGTACAGCTCGTCCATGCC-3' and then subcloned into the above mentioned plasmid pSM5 expressing HU between the BamHI and HindIII sites. This resulted in plasmid pSM6, which expresses HU–PAmcherry upon induction with tetracycline. The construct MSG192 described in the work of Bakshi *et al* (1) already contained a translational fusion of *yfp* to the C–terminus of *rpsB* (the gene encoding the S2 ribosomal protein). We transformed MSG192 with the plasmid pSM6, which expresses HU–PAmcherry. The transformants were selected on an ampicillin containing plate. A successful transformant was named strain

SM6 (S2-YFP, HU–PAmcherry). The doubling time of this strain was measured to be 50 ± 7 min in our fast growth conditions (EZRDM media with glucose at 30°C) and 147 ± 18 min without induction in our slow growth conditions (MBM minimal media with glycerol at 30°C).

For single-molecule tracking of trajectories of diffusing ribosomes, we imaged S2 ribosomal proteins labelled with mEos2 (32), a photoconvertible fluorescent protein in the strain MSG196. The construction of this strain is described in an earlier work (14).

Cell growth and preparation for imaging

For imaging cells under fast and slow growth conditions, 2 mL cultures from frozen glycerol stock were grown overnight with continuous shaking in EZRDM or MBM media respectively, in a 30°C water bath. We subsequently made subcultures by diluting the stationary phase culture at least 1:100 into 2 mL of fresh EZRDM or MBM as required. EZRDM is a morpholinepropanesulfonic acid (MOPS)-buffered solution with supplemental metal ions (M2130; Teknova), glucose (2 mg/ml), supplemental amino acids and vitamins (M2104; Teknova), nitrogenous bases (M2103; Teknova), 1.32 mM K_2HPO_4, and 76 mM NaCl. Slow growth media MBM is also a MOPS based medium with supplemental metal ions (M2130; Teknova), 1.32 mM K_2HPO_4, 76 mM NaCl and 0.2 % glycerol as the carbon source.

The strain SM6 that endogenously expresses S2–YFP and includes the plasmid expressing HU–PAmcherry was grown with addition of 100 μg/mL ampicillin. When the cells reach mid-log phase, anhydrotetracycline was added to a final concentration of 45 nM to induce the expression of HU–Pamcherry. After 10 min of induction, the cells were centrifuged and resuspended in fresh growth media with 100 μg/mL ampicillin to remove the inducer. The cells were then incubated again in growth media for 15 min at 30^{o}C to enable maturation of PAmcherry (27) prior to imaging.

After the subcultures had grown to midlog phase (OD_{600} = 0.3 – 0.6), ~150 μL of culture was placed in a CoverWell perfusion chamber (70326-56, Electron Microscopy Sciences, Radnor, PA) on a polylysine-coated clean coverslip to fill the entire chamber volume. We allowed ~2 min for the cells to adhere to the coverslip. The plated cells were then rinsed with fresh, warm, aerated growth media to wash away the non-adhered cells. The cells are maintained at 30^{o}C throughout the imaging and imaged for no longer than 30 min after plating.

Superresolution fluorescence imaging of ribosomes and DNA

The cells were imaged using an Eclipse Ti inverted microscope (Nikon, Melville, NY) equipped with an oil immersion objective (CFI Plan Apo Lambda DM 100X oil, 1.45 N.A; Nikon Instruments), a 1.5X intermediate magnification lens, and the Perfect Focus system (Nikon, Melville, NY). Fast shutters (Uniblitz LS2; Vincent Associates, NY) 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, CT, USA). Each pixel corresponds to 105 x 105 nm^2 at the sample with an overall magnification of 150x. All images were collected at a frame rate of 31.2 Hz, with an exposure time within each frame of 30 ms.

For superresolution co-imaging of ribosomes and HU in the strain SM6, first the S2-YFP molecules were illuminated with the 514 nm laser with a power density of 2 kW/cm^2. Images were acquired only when the rate of return of molecules to the fluorescent state became small enough that at most 3-4 copies per camera frame are fluorescent in each cell, enabling single molecules to be distinguished from each other. Immediately after imaging of single S2-YFP copies at 514 nm, the HU–PAmcherry molecules were photoactivated with 405 nm diode laser (CrystaLaser, Reno, NV) and subsequently imaged with the 561 nm laser. The 405 nm power

density at the sample was 4 to 15 W/cm^2 to ensure only 3–4 molecules of HU–PAmcherry are fluorescent per camera frame per cell. The power density of the 561 nm laser was kept at ~2 kW/ cm^2. We used dichroic ZT405-514-561rpc (Chroma, Bellows Falls, VT). The fluorescence from S2-YFP and HU-PAmcherry was isolated using triple band pass filter ZET442/514/561m (Chroma, Bellows Falls, VT). A 525 nm long pass filter (ET525lp, Chroma, Bellows Falls, VT) was added to block the shorter wavelength band allowed by the triple band pass filter. To minimize leakage of emission of YFP into the red channel, we first imaged S2–YFP, which results in photobleaching of YFP molecules and enhances the contrast in subsequently obtained HU-PAmcherry images.

For single-particle tracking of ribosomes, we imaged ribosomal S2-mEos2 molecules in the MSG196 strain. The S2–mEos2 molecules were photoconverted with the 405 nm diode laser (CrystaLaser, Reno, NV) and subsequently imaged with the 561 nm laser. The 405 nm power density at the sample was 1 to 10 W/cm^2 to ensure only 0–2 molecules of S2–mEos2 are fluorescent per camera frame per cell. The power density of the 561 nm laser was ~2 kW/ cm^2. Emission was collected through a 617/73 band pass filter (bright line 617/73, Semrock, Rochester, NY).

In all imaging modes, the required lasers illuminated the sample for the entire duration of image acquisition. To minimize the phototoxic effects of the lasers, we collected data for <35 s per cell for each channel. This necessitates formation of composite, multi-cell distributions for both HU and S2.

Single molecule image analysis

Images were analyzed using a MATLAB graphical user interface (GUI) developed in our lab (50). Images were smoothed and filtered to obtain a zero-based image. Bright spots were

located with pixel-level accuracy by a peak finding algorithm that detects the local intensity maxima within an image. A user defined intensity threshold was used as the minimum brightness of a pixel arising from a single molecule. The threshold is carefully set by the user so that it will not be so high as to reject a real single molecule in the raw images or so low as to include background noise.

A modified MATLAB version of the tracking program written by Crocker and Grier (51) was used. A centroid algorithm was used to locate the identified particles with subpixel resolution. Centroids of the bright spots were calculated from a 7 x 7 pixel square containing the entire bright spot, centered on the local maximum determined by the peak finding algorithm. The centroid positions from successive frames were connected to form a trajectory. For ribosomal and HU trajectories, the centroid positions are connected only if they lie within 3 px = 315 and 6 px = 630 nm of each other respectively.

To generate the spatial distribution of molecules from several cells such as the one shown in Fig. 3.2*C*-*F*, the camera based coordinates are reoriented so that the x axis and y axis correspond to the long and short cell axis and y axis respectively. For identifying spatial localizations within the cell of choice, cell outlines generated by Microbe Tracker (52) from phase contrast images are used.

Calculation of Modified Pearson Correlation Coefficient (MPCC)

The MPCC of two images **R** and **G** is evaluated as follows:

$$\mathrm{MPCC} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\tilde{U}_{ij}^{R})(G_{ij}-\tilde{U}_{ij}^{G})}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ij}-\tilde{U}_{ij}^{R})^{2}}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(G_{ij}-\tilde{U}_{ij}^{G})^{2}}}. \qquad (3.1)$$

Here *m* and *n* are the number of rows and columns in the image matrices **R** and **G**; there are *m* x *n* total pixels in each image. The R_{ij} and G_{ij} are the corresponding intensities of pixel *ij* in **R** and

G. $\widetilde{U}_{ij}^{R}$ and $\widetilde{U}_{ij}^{G}$ denote the intensity of pixel *ij* in the 2D projection of a large set of molecules distributed randomly in a 3D spherocylinder. The total number of molecules in $\widetilde{\mathbf{U}}^{R}$ and $\widetilde{\mathbf{U}}^{G}$ has been scaled to be the same as the total number of molecules in **R** and **G**, respectively.

R and **G** are generated by pixelating the composite distribution of spatial localizations of HU-PAmcherry and S2-YFP molecules from cells in the chosen length bin to 50 nm. We simulated two random distributions of 100,000 molecules each, corresponding to the ribosome (green) and HU (red) channels, using a spherocylinder whose dimensions match those of the chosen cells. The resulting reference images are normalized to give $\widetilde{\mathbf{U}}^{R}$ and $\widetilde{\mathbf{U}}^{G}$, which have same number of molecules as imaged ribosomes and HU. The two simulated 3D random distributions, $\widetilde{\mathbf{U}}^{R}$ and $\widetilde{\mathbf{U}}^{G}$ incorporated localization errors σ_{HU} and σ_{Ribo}, determined by the intercepts of MSD plots for respective growth condition (Fig. A3.1). The detailed procedure for generating $\widetilde{\mathbf{U}}^{R}$ and $\widetilde{\mathbf{U}}^{G}$ is described elsewhere (31).

Classifying the ribosome single-step displacements into two diffusive states

To investigate the diffusive behavior of ribosomes in live cells under different growth conditions, we chose trajectories that lasted at least seven frames. The chosen trajectories were divided into individual steps (two locations) with Δt = 30 ms. The resulting distribution $P_{ribo}(r)$ of experimental single-step displacements is modeled as a sum of two static (non-exchanging) populations: $P_{model}(r) = f_{slow}P(r;D_{slow}) + (1 - f_{slow})P(r;D_{fast})$. First a series of numerical basis functions $P(r;D)$ is generated for a series of diffusion coefficients D by simulating a large number of random walk trajectories that incorporate localization error σ and confinement within a spherocylinder that mimics the dimensions of an *E. coli* cell. The static localization error σ = 35 and 27 nm for fast and slow growth conditions, respectively, is estimated from the intercept of the MSD(τ) plots for ribosome S2-mEos2 molecules shown in Fig. A3.11. The

experimental $P_{ribo}(r)$ is fit to a weighted average of two static populations, $P_{model}(r)$, in a least-squares sense. The least-squares fitting procedure involved a numerical search for the lowest value of χ_v^2 on a 3D grid of combinations of the three independent adjustable parameters (D_{fast}, D_{slow}, f_{slow}). We varied D_{slow} from 0.001 μm^2/s to 0.05 μm^2/s in increments of 0.001 μm^2/s. Similarly, D_{fast} was varied from 0.1 μm^2/s to 1 μm^2/s in increments of 0.01 μm^2/s. The parameter f_{slow} was scanned from 0 to 1 in intervals of 0.05. Additional details of generation of $P_{model}(r)$ and fitting procedure are described in a previous work (33). Bootstrapping the data into 10 subsets enabled us to obtain error estimates for each of the fitted parameters.

Localization error of imaged molecules

The mean-square displacement MSD as a function of lag time τ is given by $\mathrm{MSD}(\tau) = < (\boldsymbol{r}(t+\tau) - \boldsymbol{r}(t))^2 >$, where $\boldsymbol{r}(t)$ is the two-dimensional location of the particle at time t, τ is the lag time, and the average is taken over all times t and over many trajectories. We chose trajectories that lasted at least seven frames (six steps) for analysis. Tracks longer than seven frames were truncated to seven frames. Suppose the best fit to the first two experimental points of a two dimensional mean-square displacement plot is given by the equation $\mathrm{MSD}(\tau) = a + b\tau$, with b the slope and a the extrapolated intercept at lag time $\tau = 0$. Then the best estimate of the diffusion coefficient is $D = b/4$ and the best estimate of the dynamic localization error is $\sigma = \frac{1}{2}(a + 4Dt_E/3)^{1/2}$, where t_E is the exposure time per camera frame (53).

In fast growing cells, we used 2117 trajectories to determine the mean diffusion coefficient of ribosome S2-YFP to be $<D> = 0.050 \pm 0.001$ μm^2-s^{-1} and the dynamic localization error to be $\sigma = 36$ nm (Fig. A3.1*A*, Red). For HU-PAmcherry, from 1180 trajectories we determined the mean diffusion coefficient $<D> = 0.25 \pm 0.04$ μm^2-s^{-1} and $\sigma = 60$ nm in (Fig. A3.1*A*, Black). In slow growing cells, from 2692 trajectories we determined the mean diffusion coefficient of

ribosome S2-YFP to be $\langle D \rangle = 0.05 \pm 0.0017\ \mu m^2\text{-}s^{-1}$ and $\sigma = 27$ nm (Fig. A3.1*B*, Red). For HU-PAmcherry, using 9922 trajectories we determined the diffusion coefficient to be $\langle D \rangle = 0.17 \pm 0.01\ \mu m^2\text{-}s^{-1}$ and $\sigma = 55$ nm (Fig. A3.1*B*, Black).

Cell radius and length in fast and slow growth

We model the *E. coli* cytoplasm as a spherocylinder. For a given growth condition, the cytoplasmic radius varies little from cell to cell. To determine the radius r_{cell}, we imaged photoactivable Kaede molecules under the assumption that they distribute homogenously within the cytoplasmic volume (50). The exposure time was 2 ms/frame. For fast growth, we used data from 15 cells whose length based on phase contrast images falls within the bin 3.8-4.2 μm, the same range used for detailed analysis in Figs. 3.2, 3.3 and 3.6. For slow growth, we chose 15 cells of length 2.1-2.5 μm, matching Figs. 3.2, 3.4 and 3.7.

In each case, we chose Kaede trajectories that lasted for at least seven frames. Tracks longer than seven frames were truncated to seven frames. The dynamic localization error σ_{Kaede} was estimated using the mean diffusion coefficient D_{Kaede} and intercept of the MSD(τ) plot. To estimate the cytoplasmic radius r_{cell}, we then simulated Kaede diffusion using D_{Kaede} and σ_{Kaede} in a series of uniformly filled spherocylinders of varying radii, seeking a match to the experimental transverse spatial distribution. At $t = 0$, 15,000 particles were randomly distributed within the cell volume. Each particle undergoes an independent random walk. To model each 2-ms camera image, three-dimensional microtrajectories (1000 steps of 2 μs each) were generated. At each time step, each particle chooses a displacement in each of three Cartesian directions. These displacements are chosen from a Gaussian distribution whose standard deviation corresponds to the three-dimensional diffusion coefficient D_{Kaede}. In the rare event that a particle attempts to step outside of the cell boundaries, the displacement for that microstep is taken to be zero. The

location of each particle during each camera frame is obtained as the centroid of the model microtrajectories in order to mimic the analysis procedure used for the experimental images. The appropriate dynamic localization error was then applied to each centroid location in both *x* and *y* coordinates by sampling a Gaussian distribution with standard deviation σ_{Kaede}. By adding the error vector to the centroid position we obtain the model "measured" location for each 2 ms camera frame. The *x* and *y* coordinates of each location are stored for further analysis. For the next model camera frame, each particle continues to make microsteps in 3D starting from the exact endpoint of the previous camera frame. Only the particles in the central cylindrical region are used to generate the simulated radial distribution; the endcaps are deleted.

For fast growing cells, we obtained $D_{Kaede} = 4.77 \pm 0.21$ μm^2-s^{-1} and and $\sigma_{Kaede} = 97$ nm using 709 trajectories. In slow growing cells, $D_{Kaede} = 5.23 \pm 0.16$ μm^2-s^{-1} and $\sigma_{Kaede} = 95$ nm using 2126 trajectories. The measured D_{Kaede} and σ_{Kaede} in each growth condition were used to generate simulated transverse distributions in spherocylinders of varying radii.

For each of 15 cells in both fast and slow growth conditions, we found the best fit of the experimental distribution to the spherocylindrical model in a least-squares sense by varying the model radius in increments of 10 nm. Single cell experimental radial distribution of Kaede locations, excluding the endcap regions of the cell, along with best fits are shown in Fig. A3.2*A* (Black) and A3.2*B* (Black) for fast and slow growth conditions. Averaged over 15 cells, the mean cytoplasmic radius for fast growing cells is $r_{cell} = 412 \pm 22$ nm (± 1 SE). The mean cytoplasmic radius for slow growing cells is $r_{cell} = 290 \pm 14$ nm (± 1 SE).

For both fast and slow growing cells, a 2D composite scatter plot of Kaede localizations obtained from cells in the chosen length bin is used to estimate the tip-to-tip cell length L_{cell}, as illustrated in Fig. A3.2*C*. The points at which the scattered distributions begin to curve mark the

beginnings of the hemispherical endcap regions and the ends of the cylindrical region. The error in determination of this boundary is ~50 nm, much smaller than the total cell length. The distance between these two boundary points becomes the length of the cylindrical region of the cell (L_{cyl}). The measured tip-to-tip cell length is then $L_{cell} = L_{cyl} + 2r_{cell}$. The result is L_{cell} = 3.82 μm for fast growing cells in the nominal length bin of 3.8–4.2 μm (based on rougher estimates from phase contrast). For slow growing cells of nominal length 2.1-2.5 μm, the result is L_{cell} = 2.14 μm.

Nucleoid radius and length

We also model the composite HU distributions in slow and fast growth conditions as spherocylinders. To determine the nucleoid radius and length in each growth condition, we followed a procedure similar to that used for determining cell width and length.

For fast growing cells, a 2D composite scatter plot of HU molecules in the two nucleoid lobes of 8 cells with nominal length in the range 3.8–4.2 μm is shown in Fig. A3.10*C*. The composite HU spatial distribution in each nucleoid lobe roughly resembles a 2D projection of a spherocylinder with hemispherical end caps. The experimental radial distribution of all HU localizations in the cylindrical region of each nucleoid lobe, symmetrized about the long axis of the cylinder, is generated (Fig. A3.10*B*, Black). To estimate the nucleoid radius $r_{nucleoid}$, we simulated projected 2D distributions from uniformly filled cylinders of varying radii in increments of 20 nm and broadened the distribution with σ = 60 nm, determined for HU molecules imaged in fast growing cells. The best-fit simulated radial distribution had $r_{nucleoid}$ = 190 nm (Fig. A3.10*B*, dashed line). The simulated radial distribution does not fit the tails on either end of the experimental composite radial distribution. This is likely due to cell-to-cell heterogeneity in nucleoid shape, size, and location relative to the cell body.

For measuring the nucleoid length $L_{nucleoid}$, the points at which the 2D composite scatter plot of individual nucleoids (Fig. A3.10*C*) begin to curve mark the beginnings of the hemispherical endcap regions and the ends of the cylindrical region. The distance between these two boundary points is the length of the cylindrical region of the nucleoid (L_{cyl}). The measured tip-to-tip cell length for each nucleoid lobe is then $L_{nucleoid} = L_{cyl} + 2r_{nucleoid}$. The reported $V_{nucleoid}$ for fast growing cells is the total volume of the two lobes.

For slow growing cells, we combined localizations of HU molecules imaged in cells with length in the length bin 2.1–2.5 μm. An experimental radial distribution of the HU molecules in the cylindrical region, symmetrized about the x-axis is generated (Fig. A3.6*B*, black). Similar to the method adopted for determining the nucleoid radius in fast growing cells, we generated a family of radial distributions of uniformly filled cylinders of varying radii broadened with σ = 55 nm, determined for HU molecules imaged in slow growing cells. The best-fit simulated radial distribution for slow growing cells had $r_{nucleoid}$ = 190 nm (Fig. A3.6*B*, dashed line). $L_{nucleoid}$ was determined to be 1.55 μm following similar procedure.

For slow growing cells, the fit of the composite radial HU distribution to the spherocylinder model is very good. Single-cell HU distributions in slow growth are smooth and there is little cell-to-cell heterogeneity. The composite distribution is much more representative of the single-cell distribution in slow growth than in fast growth.

Total number of ribosomes in fast and slow growth

A detailed description of the procedure employed for counting ribosomes in the cells is provided in a previous work by Bakshi *et al* (1). We determined the copy number of ribosomes (S2-YFP) in the two different growth conditions by dividing the mean total ribosome S2-YFP intensity per cell belonging to the chosen length bin by the mean intensity of a single ribosome

S2-YFP molecule. Intensities of fluorescent ribosome S2-YFP single molecules imaged over multiple cells are used to determine the mean intensity of a single ribosome S2-YFP molecule. The total intensity of YFP in individual cells is measured by averaging over the first 10 frames of imaging using a much smaller laser intensity to avoid photobleaching. A correction to the mean total YFP intensity per cell is applied to account for the difference in laser intensity used for imaging single molecule YFP fluorescence and total YFP fluorescence per cell. All other parameters of image acquisition for single molecule and wide field fluorescence imaging were kept constant. An additional small correction is applied to the total YFP intensity to account for the presence of immature YFP copies that do not contribute to the total cell fluorescence intensity, as per Eq. S8 of Ref. (1).

The mean copy number of ribosomes per cell was measured to be 56,000 in fast growing cells with length in the range of 3.8–4.2 μm and 21,000 in slow growing cells with length in the range of 2.1–2.5 μm. The number of ribosomes per cell correlates with the cell volume across the different growth conditions, as shown in Fig. A3.5.

References

1. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution Imaging of Ribosomes and RNA Polymerase in Live *Escherichia coli* Cells. Mol. Microbiol. 2012;85(1):21-38.
2. Lewis PJ, Thaker SD, Errington J. Compartmentalization of transcription and translation in *Bacillus subtilis*. The EMBO Journal. 2000;19(4):710.
3. Kellenberger E. Functional consequences of improved structural information on bacterial nucleods. Research in Microbiology. 1991;142(2):229-38.
4. Montero Llopis P, Jackson AF, Sliusarenko O, Surovtsev I, Heinritz J, Emonet T, et al. Spatial organization of the flow of genetic information in bacteria. Nature. 2010;466:77.
5. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science. 2006;313(5793):1642.

6. Hess ST, Girirajan TPK, Mason MD. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. Biophys. J. 2006;91(11):4258-72.

7. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature Methods. 2006;3:793.

8. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN. 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 2002;99(15):9697.

9. Sanamrad A, Persson F, Lundius EG, Fange D, Gynnå AH, Elf J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. Proc Natl Acad Sci U S A 2014;111(31):11413.

10. Stracy M, Lesterlin C, Garza de Leon F, Uphoff S, Zawadzki P, Kapanidis AN. Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. Proc Natl Acad Sci U S A 2015;112(32):E4390.

11. Castellana M, Hsin-Jung Li S, Wingreen NS. Spatial organization of bacterial transcription and translation. Proc Natl Acad Sci U S A 2016;113(33):9286.

12. Vendeville A, Larivière D, Fourmentin E. An inventory of the bacterial macromolecular components and their spatial organization. FEMS Microbiology Reviews. 2011;35(2):395-414.

13. Mondal J, Bratton BP, Li Y, Yethiraj A, Weisshaar James C. Entropy-Based Mechanism of Ribosome-Nucleoid Segregation in *E.coli* Cells. Biophys. J. 100(11):2605-13.

14. Bakshi S, Choi H, Mondal J, Weisshaar James C. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. Mol. Microbiol. 2014;94(4):871-87.

15. Zimmerman Steven B, Murphy Lizabeth D. Macromolecular crowding and the mandatory condensation of DNA in bacteria. FEBS Letters. 1996;390(3):245-8.

16. Joyeux M. In vivo compaction dynamics of bacterial DNA: A fingerprint of DNA/RNA demixing? Current Opinion in Colloid & Interface Science. 2016;26:17-27.

17. Woldringh CL, Jensen PR, Westerhoff HV. Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces? FEMS Microbiology Letters. 1995;131(3):235-42.

18. Dame Remus T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol. Microbiol. 2005;56(4):858-70.

19. Wang W, Li G-W, Chen C, Xie XS, Zhuang X. Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. Science. 2011;333(6048):1445.

20. Lin J, Chen H, Dröge P, Yan J. Physical Organization of DNA by Multiple Non-Specific DNA-Binding Modes of Integration Host Factor (IHF). PLOS ONE. 2012;7(11):e49885.

21. Sato YT, Watanabe S, Kenmotsu T, Ichikawa M, Yoshikawa Y, Teramoto J, et al. Structural change of DNA induced by nucleoid proteins: Growth phase-specific fis and stationary phase-specific dps. Biophys. J. 2013;105(4):1037-44.

22. Neidhardt FC, Bloch PL, Smith DF. Culture Medium for *Enterobacteria*. Journal of Bacteriology. 1974;119(3):736-47.

23. Nielsen HJ, Ottesen JR, Youngren B, Austin SJ, Hansen FG. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. Mol. Microbiol. 2006;62(2):331-8.

24. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nature Biotechnology. 2002;20:87.

25. Lindahl L. Intermediates and time kinetics of the in vivo assembly of *Escherichia coli* ribosomes. J Mol Biol. 1975;92(1):15-37.

26. Biteen JS, Thompson MA, Tselentis NK, Bowman GR, Shapiro L, Moerner WE. Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable EYFP. Nature methods. 2008;5(11):947-9.

27. Subach FV, Patterson GH, Manley S, Gillette JM, Lippincott-Schwartz J, Verkhusha VV. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. Nature methods. 2009;6(2):153-9.

28. Jin DJ, Cagliero C, Zhou YN. Role of RNA Polymerase and Transcription in the Organization of the Bacterial Nucleoid. Chemical Reviews. 2013;113(11):8662-82.

29. Pearson K. Mathematical contributions to the theory of evolution III. Regression, heredity, and panmixia. Philos Trans R Soc Lond B Biol Sci. 1896;187:253.

30. Manders EM, Stap J, Brakenhoff GJ, van Driel R, Aten JA. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. Journal of Cell Science. 1992;103(3):857.

31. Mohapatra S, Weisshaar JC. Modified Pearson Correlation Coefficient for Two-color Imaging in Spherocylindrical Cells. bioRxiv. 2018.

32. McKinney SA, Murphy CS, Hazelwood KL, Davidson MW, Looger LL. A bright and photostable photoconvertible fluorescent protein. Nature Methods. 2009;6:131.

33. Mohapatra S, Choi H, Ge X, Sanyal S, Weisshaar JC. Spatial Distribution and Ribosome-Binding Dynamics of EF-P in Live *Escherichia coli*. mBio. 2017;8(3).

34. Dai X, Zhu M, Warren M, Balakrishnan R, Patsalo V, Okano H, et al. Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. Nature microbiology. 2016;2:16231.

35. Campos M, Surovtsev IV, Kato S, Paintdakhi A, Beltran B, Ebmeier SE, et al. A constant size extension drives bacterial cell size homeostasis. Cell. 2014;159(6):1433-46.

36. Jensen RB. Analysis of the Terminus Region of the *Caulobacter crescentus* Chromosome and Identification of the dif Site. Journal of Bacteriology. 2006;188(16):6016-9.

37. Kuhlman TE, Cox EC. Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. Molecular Systems Biology. 2012;8(1).

38. Nierman WC, Feldblyum TV, Laub MT, Paulsen IT, Nelson KE, Eisen J, et al. Complete genome sequence of *Caulobacter crescentus*. Proc Natl Acad Sci U S A. 2001;98(7):4136.

39. Le TBK, Imakaev MV, Mirny LA, Laub MT. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. Science. 2013;342(6159):731.

40. Bayas CA, Wang J, Lee MK, Schrader JM, Shapiro L, Moerner WE. Spatial organization and dynamics of RNase E and ribosomes in *Caulobacter crescentus*. Proc Natl Acad Sci U S A. 2018.

41. Lee Steven F, Thompson Michael A, Schwartz MA, Shapiro L, Moerner WE. Super-Resolution Imaging of the Nucleoid-Associated Protein HU in *Caulobacter crescentus*. Biophys. J. 2011;100(7):L31-L3.

42. Hadizadeh Yazdi N, Guet Calin C, Johnson Reid C, Marko John F. Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions. Mol. Microbiol. 2012;86(6):1318-33.

43. Bremer H, Dennis P. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal Plus. 2008.

44. Gaal T, Bratton BP, Sanchez-Vazquez P, Sliwicki A, Sliwicki K, Vegel A, et al. Colocalization of distant chromosomal loci in space in *E. coli*: a bacterial nucleolus. Genes & Development. 2016;30(20):2272-85.

45. Tao H, Bausch C, Richmond C, Blattner FR, Conway T. Functional Genomics: Expression Analysis of *Escherichia coli* Growing on Minimal and Rich Media. Journal of Bacteriology. 1999;181(20):6425-40.

46. Needleman D, Dogic Z. Active matter at the interface between materials science and cell biology. Nature Reviews Materials. 2017;2:17048.

47. Proshkin S, Rahmouni AR, Mironov A, Nudler E. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. Science. 2010;328(5977):504.

48. McGary K, Nudler E. RNA polymerase and the ribosome: the close relationship. Current Opinion in Microbiology. 2013;16(2):112-7.

49. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, et al. A NusE:NusG Complex Links Transcription and Translation. Science. 2010;328(5977):501.

50. Bakshi S, Bratton Benjamin P, Weisshaar James C. Subdiffraction-Limit Study of Kaede Diffusion and Spatial Distribution in Live *Escherichia coli*. Biophysical Journal. 2011;101(10):2535-44.

51. Crocker JC, Grier DG. Methods of Digital Video Microscopy for Colloidal Studies. Journal of Colloid and Interface Science. 1996;179(1):298-310.

52. Sliusarenko O, Heinritz J, Emonet T, Jacobs-Wagner C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. Molecular Microbiology. 2011;80(3):612-27.

53. Michalet X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Physical Review E. 2010;82(4):041914.

Appendix

A3.1. Timescale of polysome escape from nucleoid into the ribosome-rich annular region

When a transcript is complete, it is presumably already decorated with 70S ribosomes and diffuses as a polysome, with D_{slow} ~ 0.01 μm^2/s. Polysomes and complete transcripts escape the nucleoid into the annular region by means of 2D radial diffusion, as described below. In slow growth, $r_{nucleoid}$ = 0.19 μm. If polysomes are released inside the nucleoid with a uniform spatial distribution, the average radial distance traveled to escape the nucleoid is Δr ~ 60 nm as follows. The average radial localization of polysomes in a cylinder of radius 0.19 μm is:

$$r_{avg} = \frac{\int_{r=0\ \mu m}^{r=0.19\ \mu m} 2\pi r^2 dr}{\int_{r=0\ \mu m}^{r=0.19\ \mu m} 2\pi r dr} = 0.127\ \mu\text{m}$$

Thus the average radial distance a polysome needs to travel to escape the nucleoid is: Δr = 0.19 μm – 0.127 μm = 0.63 μm ~ 60 nm. In 2D, the time required to diffuse a root-mean-square distance of 60 nm is $\tau_{escape} = (0.06\ \mu m)^2 / 4D_{slow}$ = 0.09 s ~ 0.1 s. In slow growth, this provided a rough estimate of the timescale of mRNA/polysome escape from the nucleoid to surrounding ribosome-rich annular region.

A3.2. Time scale of exchange of polysomes between ribosome-rich annular region and endcap

The polysomes in the annular region can exchange with polysomes in end cap region by essentially 1D diffusion along the long cell axis. In slow growth, the length of the straight cylindrical region is L_{cyl} = 1.56 μm. For polysomes distributed uniformly along this length, the

mean distance to the beginning of the nearest endcap is $L_{cyl}/4 = 0.39$ μm. For 1D diffusion, we estimate the exchange time as: $\tau_{exchange} = (0.39\ \mu m)^2/(2D_{slow}) = 7.6$ s ~ 10 s.

A3.3. Connection with 1D global reaction-diffusion model of transcription and translation

In the one-dimensional model of the cell by Wingreen and co-workers (1), the nucleoid of the cell is modeled as a distribution of DNA of FWHM length $L = 1.8$ μm. Free polysomes born in the dense part of the nucleoid can only escape by 1D axial diffusion. For free polysomes uniformly distributed on a line of length 1.8 μm, the mean distance to the nearest ribosome-rich end region is $L/4 = 450$ nm.

In our 3D nucleoid model in slow growth, most of the newly born free polysomes will escape the nucleoid by 2D radial diffusion to the surrounding annular region, not by 1D axial diffusion to reach an endcap. We demonstrate this with a simple "diffusion to capture" model. To easily distinguish numerically between radial and axial escape, the cell nucleoid volume is represented by a 3D cylinder (with flat endplates, not hemispherical endcaps). We simulated 10,000 uniformly distributed molecules within a 3D cylinder ($L = 1.55$ μm, $r = 0.19$ μm) of dimensions similar to that of a nucleoid of slow growing cells. We allowed the molecules to diffuse with $D = 0.01\ \mu m^2/s$, as estimated for polysomes in Fig. 3.5*B*. Each trajectory was halted as soon as it stepped outside the cylinder and characterized as escaping from the cylindrical boundary or one of the flat endplates. Of the 10,000 diffusing molecules, 8971 (~90%) escaped radially and 1029 (~10%) escaped the cylinder axially. The distribution of axial and radial escape times is given in Fig. A3.12. The mean time to radial escape was 0.20 s, similar to the rough estimate above. The mean time to axial escape was 0.15 s. Only those copies that originate very near the endplates escape axially.

A3.4 Supporting Figures and Captions

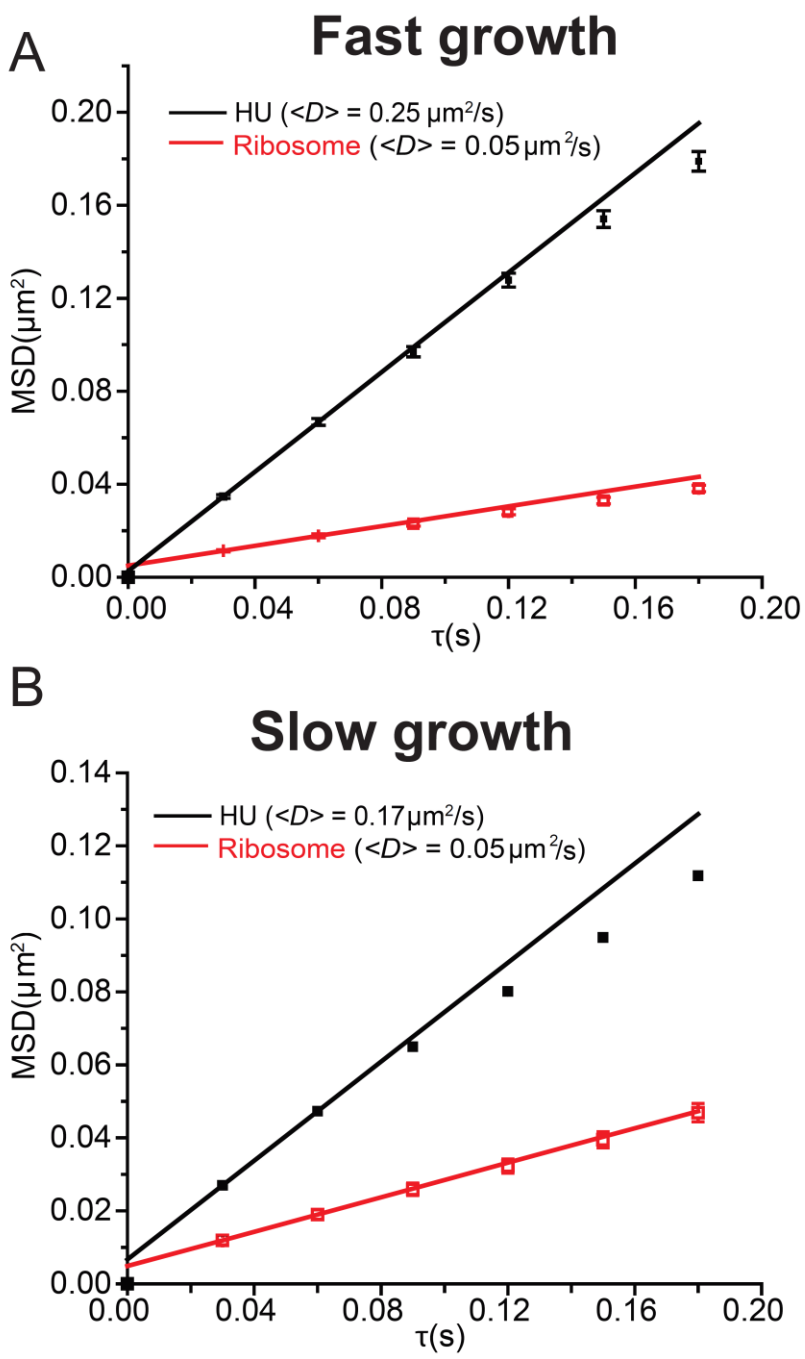


Fig. A3.1 Mean-square displacement plot vs lag time, MSD(τ), for ribosomes (*Red*) and HU (*Black*) in (**A**) fast growth conditions and (**B**) slow growth conditions. Trajectories are obtained at 30 ms/frame and truncated to six steps. Error bars are ± one standard error of the mean. Line drawn through first two data points yields the estimated mean diffusion coefficient <*D*>.

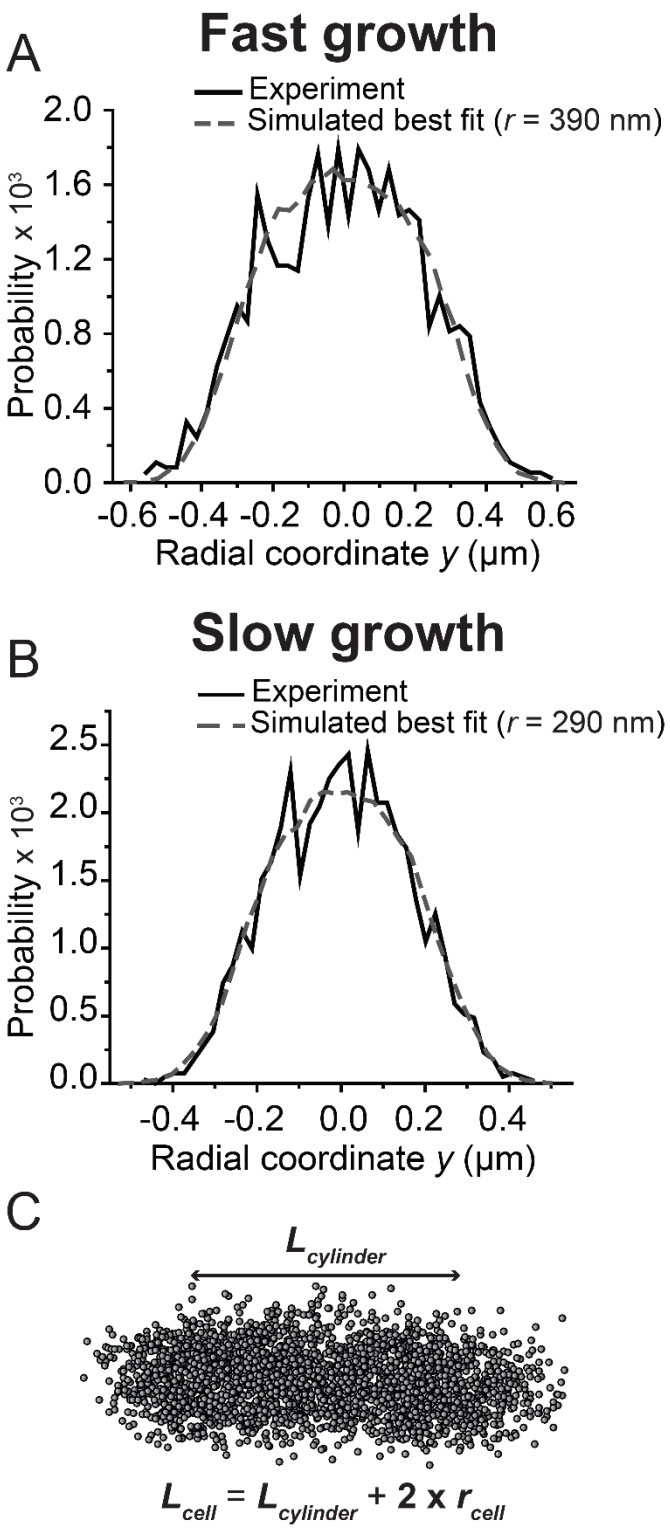


Fig. A3.2 Determination of cell radius and length using images of Kaede molecules. **(A)** *Fast growth condition*: Experimental single-cell radial distribution of 1725 Kaede molecules (black line) within the straight cylindrical region of the cell (endcaps excluded). For this particular cell, the best fit simulated radial distribution of homogenously distributed molecules in a 3D cylinder (dashed line) had r = 390 nm. Averaging over 15 fast growing cells yields r_{cell} = 412 ± 22 nm (± 1 SE). **(B)** *Slow growth condition*: Experimental single-cell radial distribution of 2328 Kaede molecules (Black Line) within the straight cylindrical region of the cell (endcaps excluded). For this particular cell, the best fit simulated radial distribution of homogenously distributed molecules in a 3D cylinder (Dashed Line) had r = 290 nm. Averaging over 15 slow growing cells yields r_{cell} = 290 ± 14 nm (± 1 SE). **(C)** Method for estimation of cell length from 2D composite scatter plot of imaged Kaede molecules. A representative scatter plot for Kaede localizations imaged in fast growing cells of length in the range of 3.8 μm – 4.2 μm is shown. $L_{cylinder}$ is taken as the length of the cylindrical region, as judged by eye. Tip-to-tip length is then $L_{cell} = L_{cylinder} + 2r_{cell}$. Same procedure is applied in fast and slow growing cells.

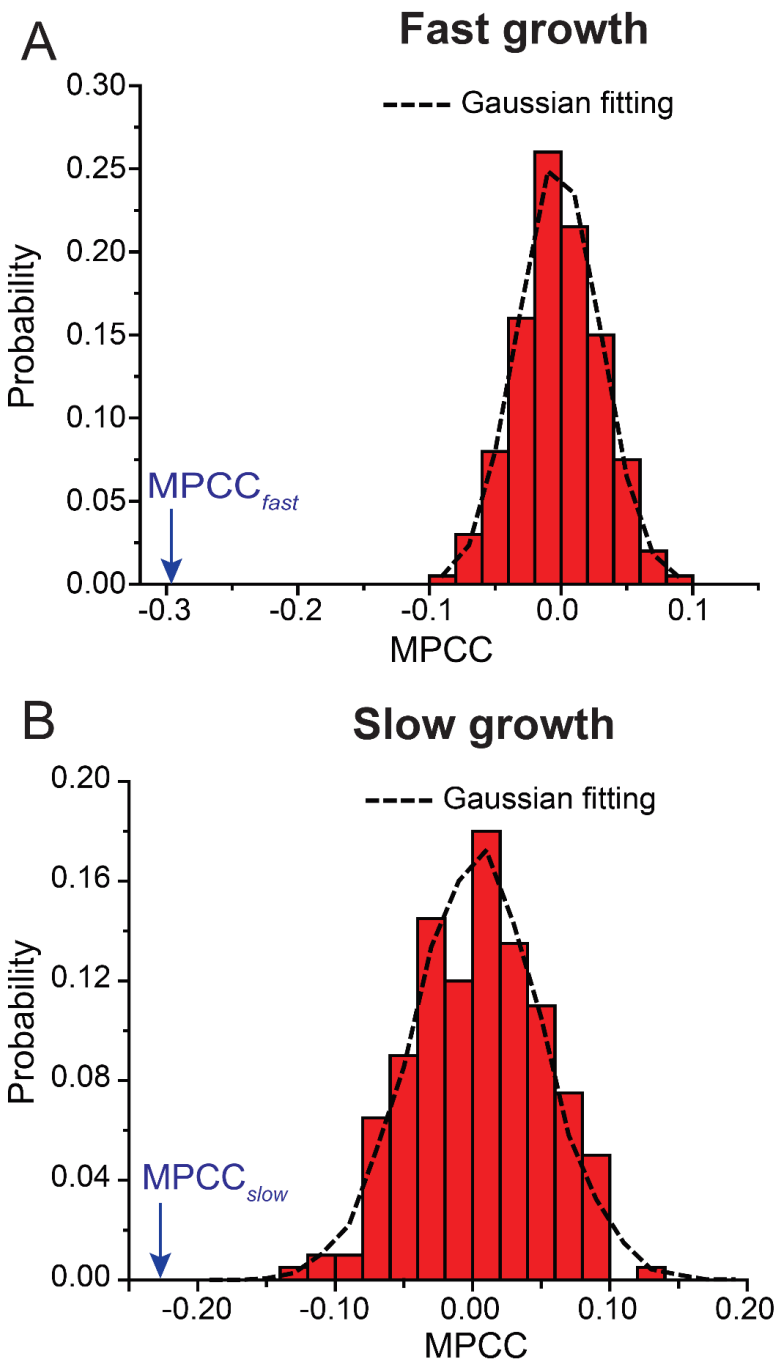


Fig. A3.3 Probability distribution of 200 independent MPCC calculations for sets of **R** and **G** matrices formed from 2D projections of uncorrelated, random 3D distributions in a 3D spherocylinder. Each simulated **R** and **G** matrix uses 50 nm pixels, same as the experimental **R** and **G** of Figs. 3.3 and 3.4. (**A**) Fast growing cells, with spherocylinder dimensions chosen accordingly. **R** and **G** have 8762 and 9839 molecules respectively, same as that of experimentally imaged HU and ribosome molecules. The best-fit Gaussian curve (Dashed Line) has mean <MPCC> = –0.0017 and standard error σ_{MPCC} = 0.0309. The arrow shows the experimental $MPCC_{fast}$ = –0.30, ~9.6 σ_{MPCC} away from zero. (**B**) Slow growing cells, with spherocylinder dimensions chosen accordingly. **R** and **G** have 13,071 and 10,251 molecules respectively, same as that of experimentally imaged HU and ribosome molecules. The best-fit Gaussian curve (Dashed Line) has mean <MPCC> = +0.0038 and standard error σ_{MPCC} = 0.0467. The arrow shows the experimental $MPCC_{slow}$ = –0.23, ~4.9 σ_{MPCC} away from zero.

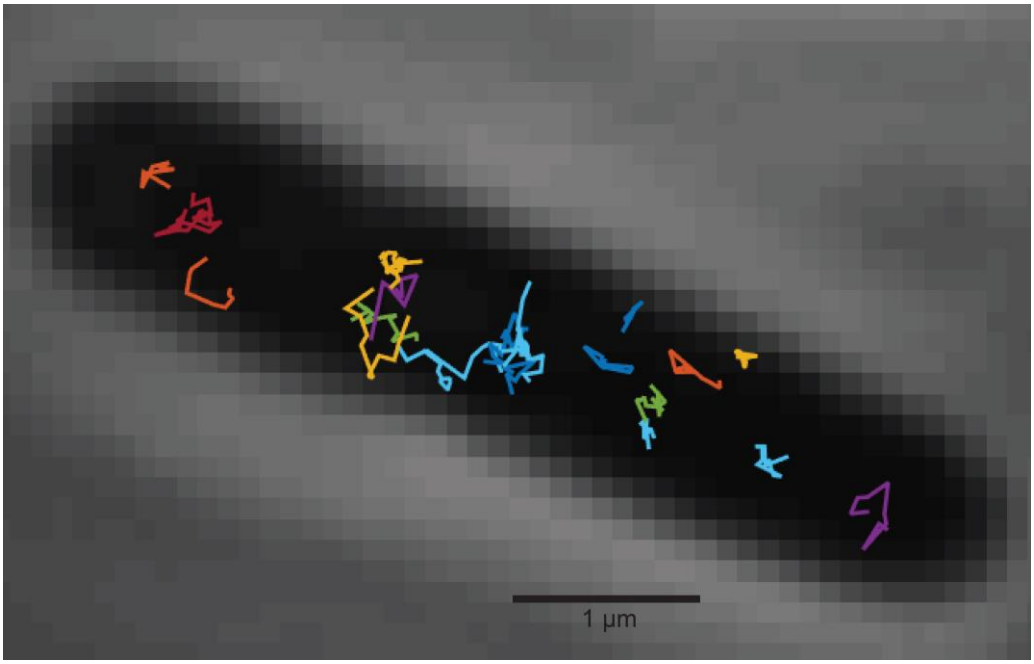


Fig. A3.4 Slow growing cell imaged by phase contrast overlaid with trajectories of single ribosome S2-mEos2 copies. Imaging at 30 ms/frame.

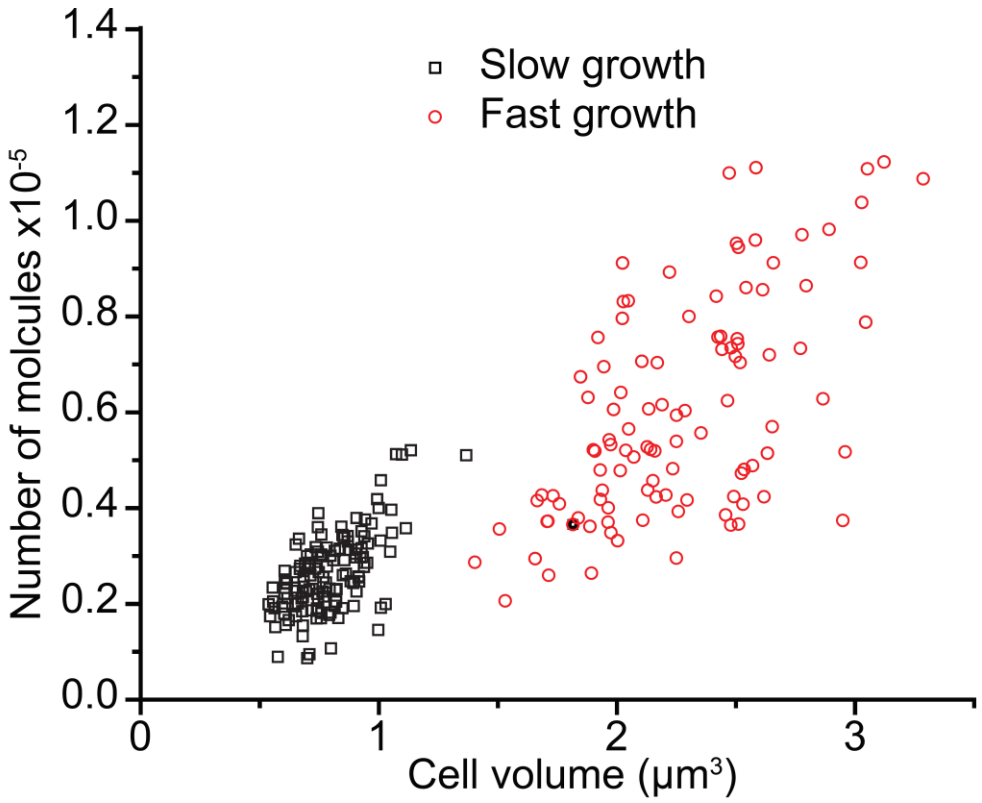


Fig. A3.5 Ribosome (total S2-YFP) copy number per calculated volume of slow (Black) and fast (Red) growing cells. Copy numbers from total YFP intensity per cell divided by mean single-molecule intensity, corrected for laser power differences and presence of immature YFP copies. Individual cell volumes calculated from cell length (estimate from phase contrast) and values of r_{cell} as in Fig. A3.2.

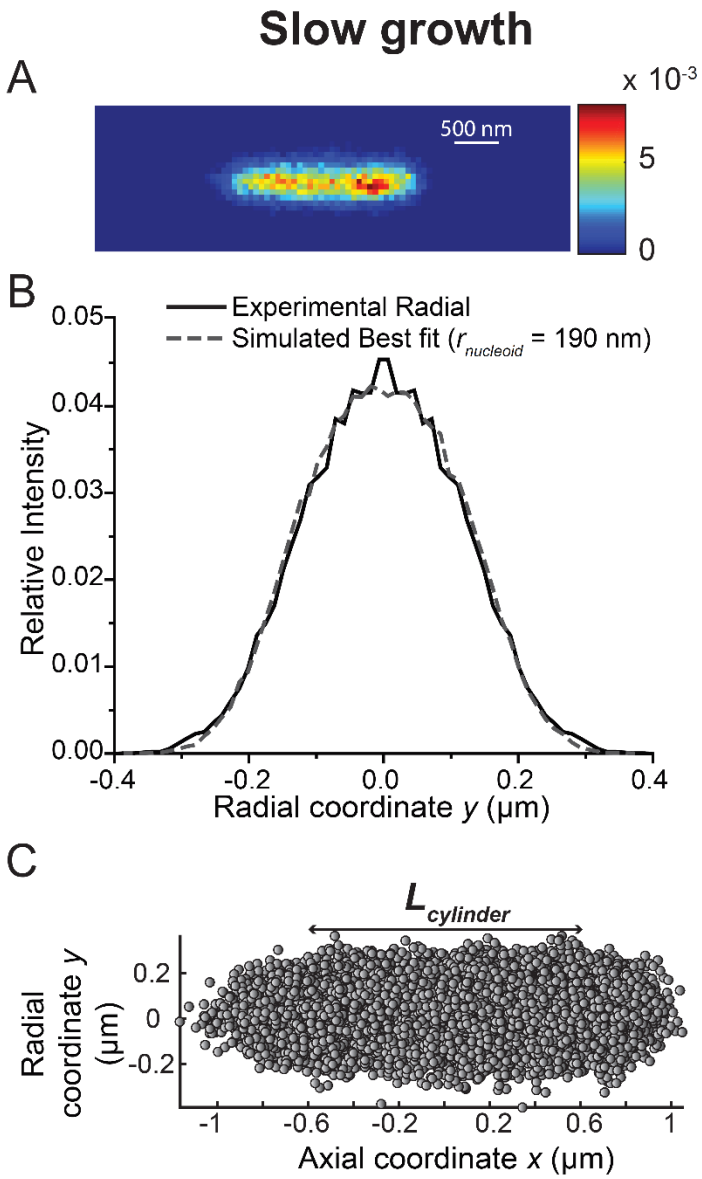


Fig. A3.6 Determination of nucleoid length and radius for slow growing cells. (**A**) 2D localization probability density heat maps of 13,071 imaged HU molecules in slow growing cells. Color scale indicates probability density of molecules in each 50 nm pixel. (**B**) Experimental radial distribution of HU molecules within the straight cylindrical region of the nucleoid (endcaps excluded) in cells of length in the range 2.1 μm – 2.5 μm. Dashed line*;* Simulated best fit radial distribution of homogenously distributed molecules in a 3D cylinder of $r_{nucleoid}$ = 190 nm is shown for comparison. (**C**) Composite 2D scatter plot of all imaged HU molecules in slow growing cells belonging to the chosen length bin. $L_{cylinder}$ is taken as the length of the cylindrical region, as judged by eye. Tip-to-tip length $L_{nucleoid} = L_{cylinder} + 2r_{nucleoid}$.

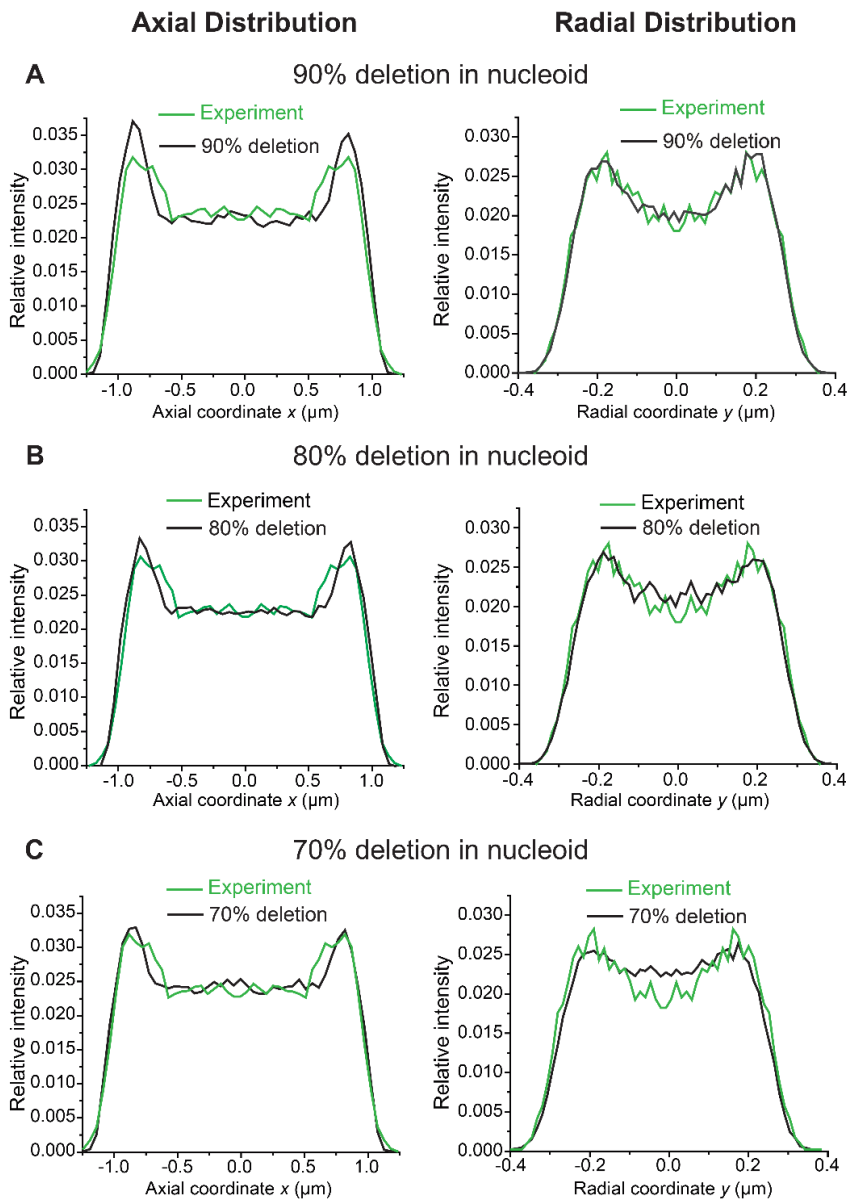


Fig. A3.7 For slow growth condition, comparison of experimental (Green) axial and radial distributions of 70S, translating ribosomes with 3D model results (Black). Entire cell is modeled as a spherocylinder (L_{cell} = 2.14 μm, r_{cell} = 0.29 μm). Nucleoid is modeled as a spherocylinder whose radius $r_{nucleoid}$ = 0.19 μm and $L_{nucleoid}$ = 1.55 μm is chosen to fit the measured HU distribution (Fig. A3.6). See text for details. In each case, we begin with 100,000 points distributed uniformly throughout the entire model cell, delete a variable percentage of those points lying within the model nucleoid volume, project the resulting 3D distribution into 2D, and add localization error to the result. Sum of square residuals (ssr) between the experimental and model distribution is calculated in each case. (**A**) Axial (ssr = 3.25 x 10^{-4}) and radial (ssr = 6.73 x 10^{-5}) distributions from deletion of 90% of points within nucleoid. (**B**) Axial (ssr = 1.77 x 10^{-4}) and radial (ssr = 7.31 x 10^{-5}) distributions from deletion of 80% of points within nucleoid. (**C**) Axial (ssr = 1.31 x 10^{-4}) and radial (ssr = 1.90 x 10^{-4}) distributions from deletion of 70% of points within nucleoid. The best fit deletes 80% of the points, as judged by the combined sum of squared residuals from the axial and radial fits.

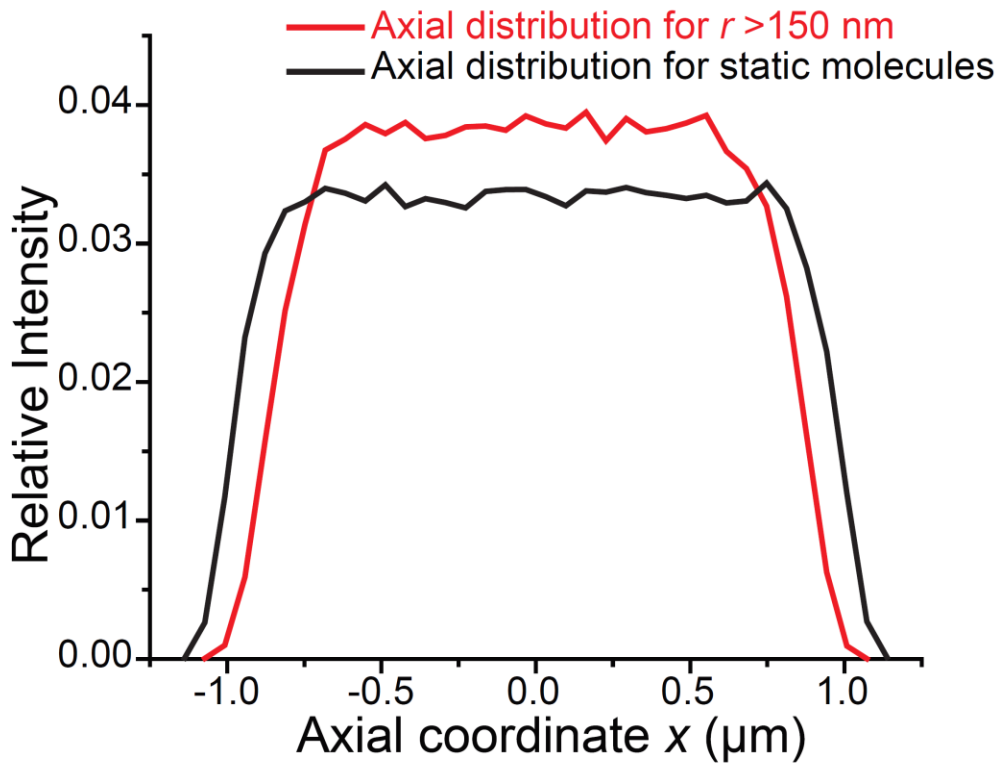


Fig. A3.8 Depletion of long steps from tips of endcaps. Axial projections from simulations within a 3D spherocylinder of length L_{cell} = 2.14 μm and radius r_{cell} = 0.29 μm, mimicking slow growth conditions. *Black:* Axial distribution of uniformly distributed stationary molecules simulated within the 3D spherocylinder. *Red*: From random walk simulation for uniformly distributed molecules with diffusion coefficient D_{fast} = 0.13 μm^2/s (taken from fit to single-step distribution of Fig. 3.5*B*) and 30 ms/frame. Axial distribution of mid-points of only those steps longer than r_{fast} = 0.15 μm, the cutoff used in Figs. 3.7 and 3.8. The axial distribution of such long steps is depleted near the cell tips.

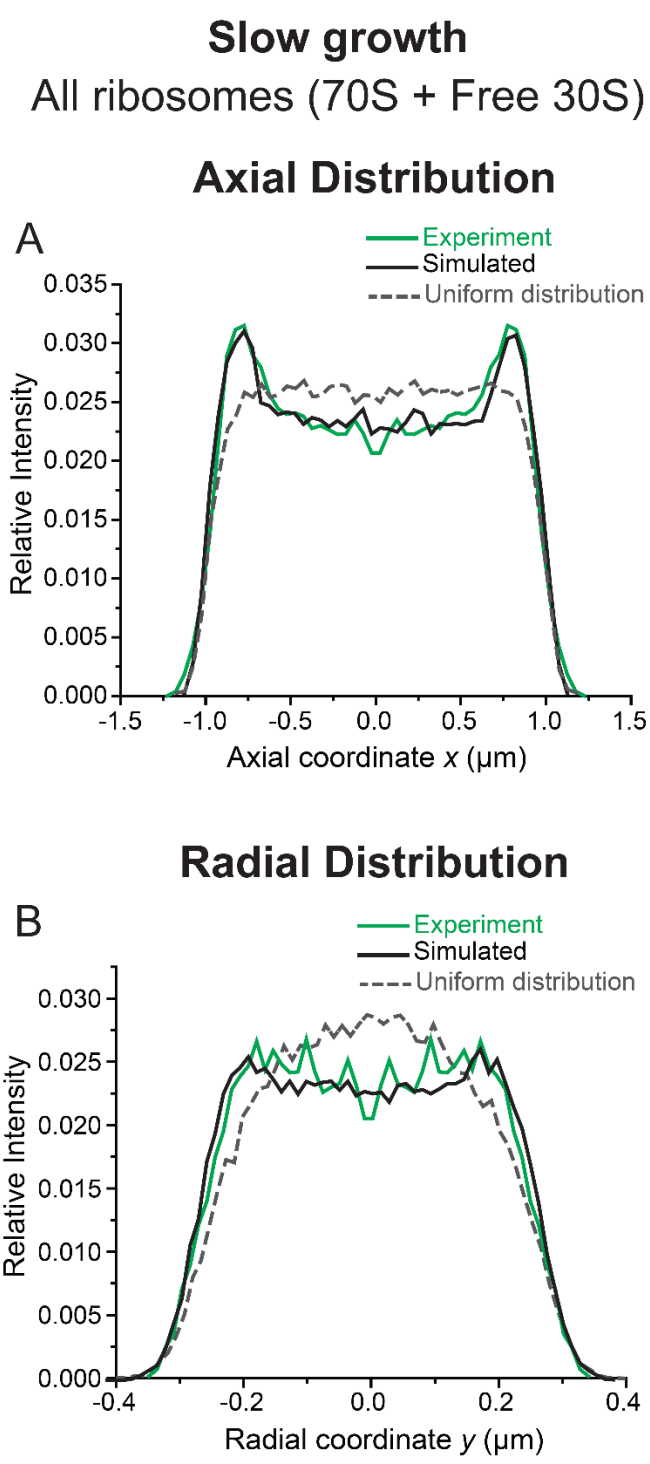


Fig. A3.9 For slow growth condition, comparison of experimental axial and radial distributions of all ribosomes, both translating 70S and free 30S copies, with 3D model results. Nucleoid is modeled as a spherocylinder whose radius $r_{nucleoid}$ = 0.19 μm and $L_{nucleoid}$ = 1.55 μm are chosen to fit the measured HU distribution (Fig. A3.6). Entire cell is modeled as a uniformly filled spherocylinder (L_{cell} = 2.14 μm, r_{cell} = 0.29 μm). Points are deleted from model nucleoid region as in Fig. A3.7. Best fit model deleted 65% of points within nucleoid. (**A**) *Green:* Experimental axial distribution of all ribosomes in slow growing cells of length in the range 2.1 μm – 2.5 μm. *Black*: Best fit model axial distribution (65% deletion). *Dashed Line*: Axial distribution of uniformly filled spherocylinder is shown for comparison. (**B**) *Green:* Experimental radial distribution of all ribosomes in slow growing cells of length in the range 2.1 μm – 2.5 μm. *Black*: Best fit model radial distribution (65% deletion). *Dashed Line*: Radial distribution of uniformly filled spherocylinder shown for comparison.

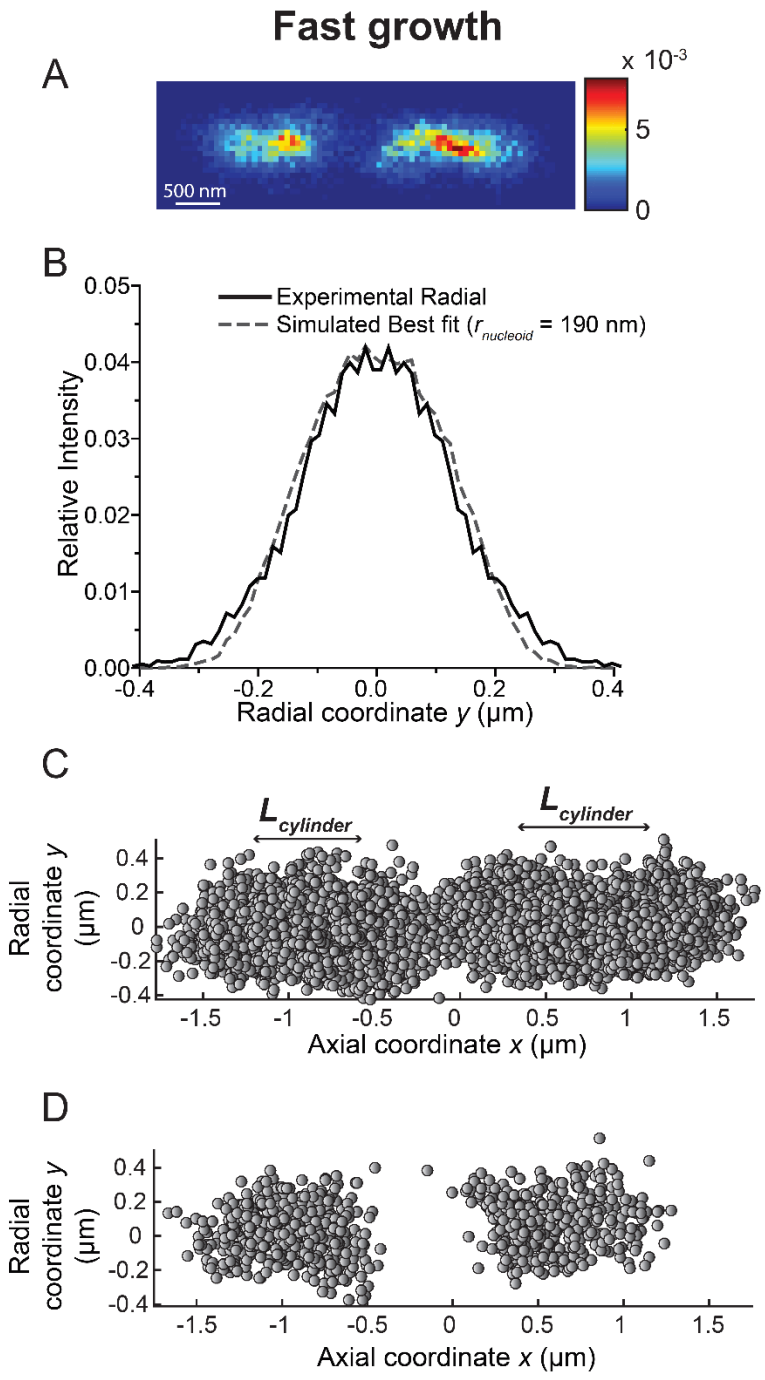


Fig. A3.10 Determination of nucleoid length and radius for fast growing cells. (**A**) 2D composite localization probability density heat maps of 8762 imaged HU molecules over two nucleoid lobes in fast growing cells. Color scale indicates probability density of molecules in each 50 nm pixel. (**B**) Experimental radial distribution of HU molecules within the straight cylindrical region of the two nucleoids (endcaps excluded) in cells of length in the range 3.8 μm – 4.2 μm. Dashed line; Simulated best fit radial distribution of homogenously distributed molecules in a 3D cylinder of r = 190 nm is shown for comparison. (**C**) Composite scatter plot of all imaged HU molecules in the two nucleoids of cells in the chosen length bin. $L_{cylinder}$ is taken as the length of the cylindrical region, as judged by eye. Tip-to-tip length $L_{nucleoid} = L_{cylinder} + 2r_{nucleoid}$. $V_{nucleoid}$ for fast growing cells is the sum of the nucleoid lobes. (**D**) Composite scatter plot of all imaged HU molecules in a single cell is shown for comparison. The nucleoids of a single cell are compact compared to the composite of nucleoids over multiple cells shown in (C).

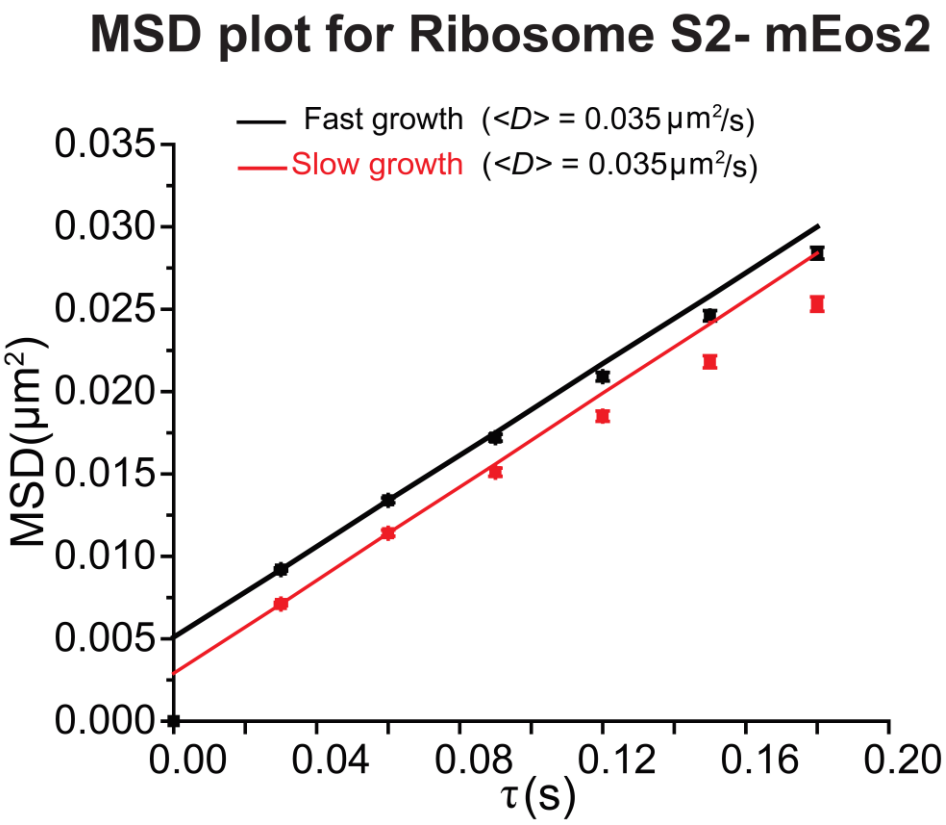


Fig. A3.11 Mean-square displacement plot MSD(τ) for imaged ribosome S2–mEos2 molecules in fast (*Black*) and slow (*Red*) growth conditions. Trajectories are truncated to six steps; error bars are ± one standard error of the mean. Line drawn through first two data points yields the estimated mean diffusion coefficient, $\langle D \rangle = 0.035\ \mu m^2/s$ in both cases.

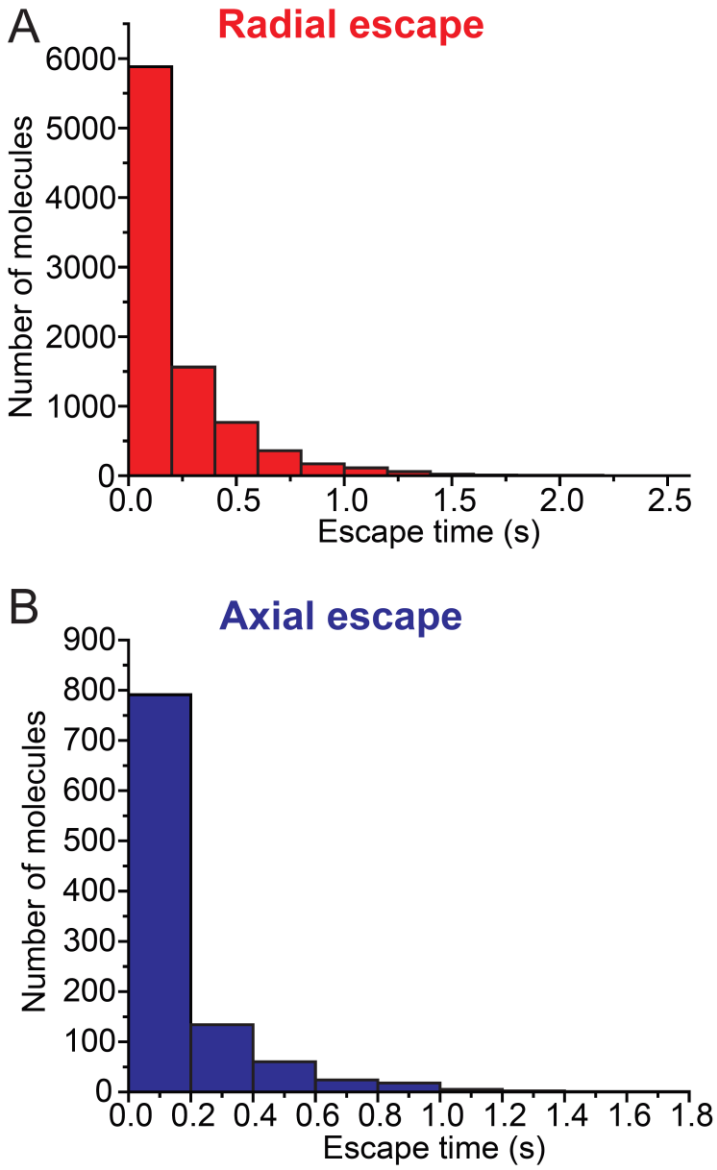


Fig. A3.12 Model distributions of times required to escape from the nucleoid in slow growing cells. 10,000 molecules were uniformly distributed in a 3D cylinder (L = 1.55 μm, r = 0.19 μm) representing the nucleoid. The molecules were allowed to diffuse with D = 0.01 μm^2/s, as estimated for polysomes in slow growth conditions in Fig. 3.5*B*. (**A**) Distribution of time taken by 8971 out of 10,000 molecules to escape radially. (**B**) Distribution of time taken by 1029 out of 10,000 molecules to escape axially.

A3.5 Supporting References

1. Castellana M, Hsin-Jung Li S, Wingreen NS. Spatial organization of bacterial transcription and translation. Proc Natl Acad Sci U S A. 2016;113(33):9286.

4

Spatial Distribution and Ribosome – Binding Dynamics of EF-P in Live *E. coli*

Reproduced verbatim from:

Mohapatra S, Choi H, Ge X, Sanyal S, Weisshaar JC. Spatial Distribution and Ribosome-Binding Dynamics of EF-P in Live *Escherichia coli*. mBio. 2017;8(3).

Abstract

In vitro assays find that ribosomes form peptide bonds to proline (Pro) residues more slowly than to other residues. Ribosome profiling shows that stalling at Pro-Pro-X triplets is especially severe, but is largely alleviated in *E. coli* by the action of elongation factor EF-P. EF-P and its eukaryotic/archaeal homolog IF5A enhance the peptidyl transfer step of elongation. Here a superresolution fluorescence localization and tracking study of EF-P–mEos2 in live *E. coli* provides the first *in vivo* information about the spatial distribution and on-off binding kinetics of EF-P. Fast imaging at 2 ms/frame helps to distinguish ribosome-bound (slowly diffusing) EF-P from free (rapidly diffusing) EF-P. Wild-type EF-P exhibits a three-peaked axial spatial distribution similar to that of ribosomes, indicating substantial binding. The mutant EF-P ^{K34A} exhibits a homogeneous distribution, indicating little or no binding. Some 30% of EF-P copies are bound to ribosomes at a given time. Two-state modeling and copy number estimates indicate that EF-P binds to 70S ribosomes during 25–100% of translation cycles. The timescale of the typical diffusive search by free EF-P for a ribosome binding site is $\tau_{free} \sim 16$ ms. The typical residence time of an EF-P on the ribosome is very short, $\tau_{bound} \sim 7$ ms. Evidently EF-P binds to ribosomes during many or most elongation cycles, much more often than the frequency of Pro-Pro motifs. Emptying of the E site during part of the cycle is consistent with recent *in vitro* experiments indicating dissociation of the deacylated tRNA upon translocation.

Introduction

During the elongation phase of protein synthesis in prokaryotes, ribosomes are assisted by at least two cofactors that bind and unbind during each translation cycle (1). EF-Tu mediates the proper binding of the correct aminoacyl-tRNA to the ribosomal A site. EF-G promotes the tRNA translocation step that follows formation of a new peptide bond between aminoacyl-tRNA in the A site and peptidyl-tRNA in the P site. *In vitro* translation assays have found that Pro is incorporated into peptides more slowly than other residues (2-4). This can lead to translational pausing (ribosome stalling), which in turn can cause premature termination of the elongation phase before a complete protein has been synthesized (2-5).

Ribosome stalling is known to occur especially frequently at consecutive Pro codons (Pro-Pro motifs), presumably due to geometric constraints imposed by the cyclic structure of proline. *In vitro* kinetics studies show that a third elongation factor EF-P helps to alleviate pausing at Pro-Pro motifs (4, 6-8). EF-P is the only known cofactor that accelerates the peptide bond formation step of the translation elongation cycle (9, 10). The size and shape of the 21 kDa EF-P mimics that of tRNA (11). A crystal structure reveals that *Thermus thermophilus* EF-P binds at the interface of the 30S and 50S subunits, between the E and P sites of the ribosome and in close proximity to the peptidyl transfer center; see Fig. 4.1 (12). Post-translational β-lysylation at residue Lys-34 and an empty ribosomal E site are evidently prerequisite to EF-P binding and activity (1, 4, 13).

In prokaryotes, EF-P is non-essential, but *efp*– strains exhibit pleiotropic phenotypes including slower growth, loss of membrane integrity, and increased sensitivity to antibiotics (1, 14, 15). Early kinetics and ribosome profiling studies have investigated the nature of stalling

sites that are alleviated by EF-P (4, 6-8). Distinct Z-Pro-Pro-X codon sequences have been shown to induce translational stalling in *efp*– mutant strains. The stalling strength varied substantially depending on the identity of the codon just downstream (X) or just upstream (Z) of Pro-Pro (6, 7). In addition, EF-P has been shown to alleviate stalling at certain sequences that do not contain a Pro-Pro duet (8, 16).

Here we use superresolution fluorescence microscopy (17-19) to locate and track single EF-P copies in live *E. coli* cells for the first time. The gene *efp* is replaced in the chromosome with the gene *efp–mEos2* for the fluorescent variant. A weak laser at 405 nm photoswitches a sparse set of EF-P–mEos2 molecules from a green fluorescent state to an orange fluorescent state. A probe laser at 561 nm is used to locate and track one or two copies at a time with ~60 nm spatial resolution and 2-ms time resolution. Ribosome-bound copies of EF-P generally exhibit slower diffusion than free copies, as depicted schematically in Fig. 4.1.

The new results provide the first *in vivo* information about the spatial distribution and ribosome-binding dynamics of EF-P. We provide estimates of the fraction of EF-P that is bound to ribosomes at a given time, the timescale of the diffusive search for empty ribosome E sites, and the duration of typical transient binding events. Our numerical estimates from two-state modeling indicate that EF-P binds to 70S ribosomes during at least one of every four elongation cycles, but typically for a very short time period (τ_{bound} ~ 7 ms). Such EF-P binding events evidently occur much more frequently than previously suspected. This suggests that the E site becomes empty and thus open to EF-P binding during part of each elongation cycle. That inference in turn is consistent with *in vitro* experiments indicating that the deacylated tRNA is released from the E site rapidly upon translocation (20-22). The new data are inconsistent with a concerted mechanism of tRNA progression through the A, P, and E sites, *i.e.*, a mechanism in

which the E site is never completely empty. However, our data do not resolve the question of whether or not accommodation of a new aa-tRNA at the A site is prerequisite to release of the deacylated tRNA (20-25).

Results

Comparison of axial spatial distributions of EF-P and ribosomes in normal growth

We have fused the gene coding for the photoswitchable fluorescent protein mEos2 to the C-terminus of the endogenous *efp* gene in *E. coli* MG1655 and then moved the fusion to the VH1000 background for further study (Fig. 4.1, Table A4.1) (26). In our growth conditions, the doubling time of the modified strain expressing EF-P–mEos2 is 1.3 times longer than that of the VH1000 background strain, suggesting that the labelling does not greatly affect the functionality of *efp* or the ribosome. For comparison, deletion of *efp* increases the doubling time by a factor of two, in both *E. coli* and in *Salmonella* (15). EF-P consists of three domains (Fig. 4.1). It binds to the ribosome between the P and the (otherwise empty) E site (4, 12). The Lys-34 residue in Domain I, which is responsible for the functionality of EF-P, lies near the N-terminus of the protein and contacts the P-site tRNA. The C terminal domain of EF-P interacts with the P-site tRNA and with the 30S ribosomal subunit (12). Based on the crystal structure (12), labeling of EF-P with mEos2 at the C-terminus may cause a steric clash between mEos2 and the 30S subunit. However, this evidently does not greatly perturb ribosome functionality, as judged by the moderate 30% increase in doubling time for the mutant strain compared with the WT VH1000 strain.

Our goal is to use diffusion data and spatial distributions to distinguish EF-P copies bound to 70S (translating) ribosomes from free EF-P copies. Free, monomeric EF-P–mEos2 has mass

Table 4.1 Summary of two-state, best-fit results for $P(r)$ for different species.[a]

Species (expression mode)	D_{slow} ($\mu m^2/s$)	D_{fast} ($\mu m^2/s$)	f_{slow}
Ribosome (chromosome; 30S)	0.2	0.8	0.65
EF-P (chromosome)	Constrained to 0.2	4.3 ± 1.0	0.30 ± 0.10
EF-P after Cam (chromosome)	Constrained to 0.2	1.2	0.45
EF-P after Rif (chromosome)	4.6	8	0.55
$EF\text{-}P^{K34A}$ [b] (plasmid)	3.2	9.7	0.65

[a] See Appendix for details of fitting procedure.

[b] Unlabeled WT EF-P is natively expressed from the chromosome.

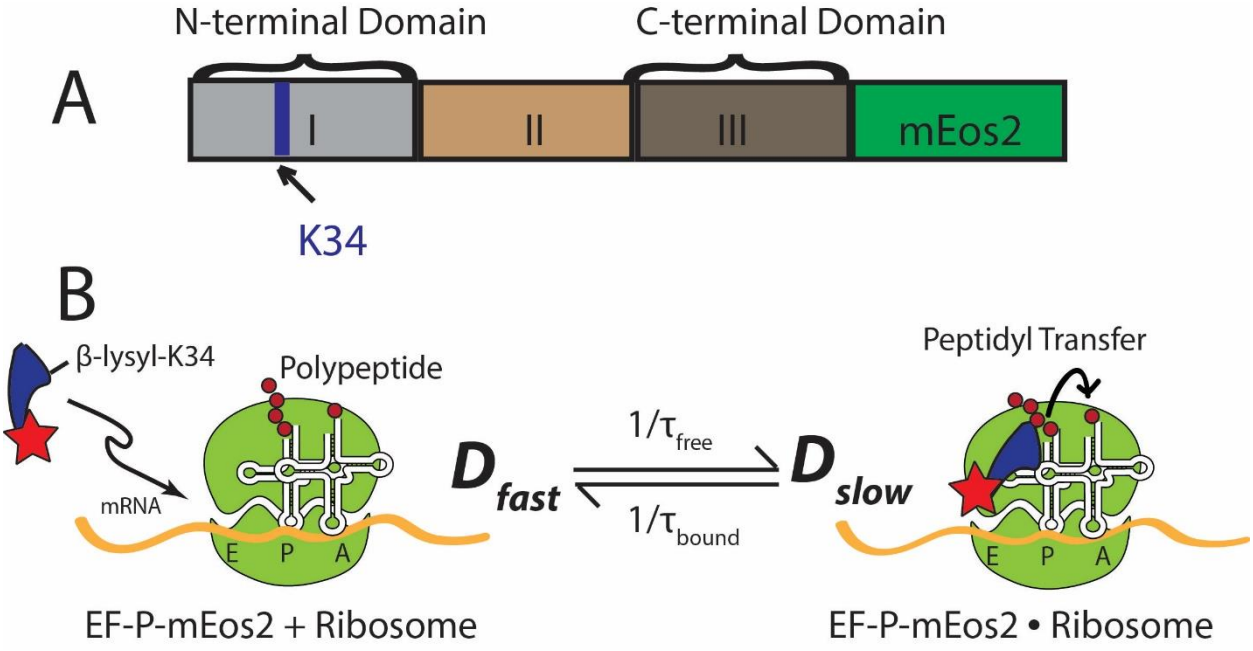


Fig. 4.1 (**A**) Organization of the different domains of the *E. coli* EF-P protein with C-terminal fusion to mEos2, a fluorophore. The key residue Lys-34 is marked with an arrow. (**B**) Simple two-state kinetics scheme for EF-P–mEos2 binding to and dissociating from a 70S ribosome with empty E site. Positions of Lys-34 at N-terminus and mEos2 (red star) at C-terminus are depicted schematically. The ribosome-bound species (D_{slow}) diffuses much more slowly than free EF-P (D_{fast}). Binding and dissociation rate constants are $1/\tau_{free}$ and $1/\tau_{bound}$, respectively.

of 47 kDa (11, 26). An unbound species of that mass would have a diffusion coefficient of roughly 4–10 $\mu m^2/s$ in the *E. coli* cytoplasm (27). In contrast, EF-P bound to 70S ribosomes would diffuse much more slowly, with diffusion coefficient of ~0.2 $\mu m^2\text{-}s^{-1}$ (28, 29). Bound EF-P should also exhibit a three-peaked, ribosome-like axial distribution of locations within the cytoplasm (30). We imaged EF-P–mEos2 molecules in live cells by photoactivating and locating fluorophores (17), connecting locations over multiple frames to form trajectories of individual molecules (31). Details are provided in Methods and in Appendix. To enable efficient superresolution imaging of rapidly diffusing molecules, the exposure time was 2 ms/frame with continuous laser illumination. The number of switched-on copies per cell was limited to 0-2 molecules per frame to avoid spatial overlap of the images.

We studied 120 cells whose tip-to-tip cell length varied from 3.2 to 8.4 μm. We sorted the cells by length into bins of 1 μm width. Some typical trajectories are shown in Fig. 4.2*A*. To obtain the axial distribution of EF-P (Fig. 4.2*B*), we combined axial locations of molecules from the 60 cells of length 3.5 – 4.5 μm. For comparison with the detailed trajectory analysis presented below, all locations measured in trajectories that last at least seven frames or longer contribute to Fig. 4.2*B*. The results are very similar when all locations from all trajectories are included. To account for cell length variability within the chosen bin width, the axial locations are scaled to the range –0.5 to +0.5, with zero representing the cell center and ±0.5 representing the locations of the two cell tips.

For direct comparison, we also performed single-molecule imaging of 30S subunits of ribosomes labeled by an S2-mEos2 construct expressed from the chromosome. Details regarding the construction of this strain were described previously (32). As inferred from the unchanged growth rate (33), labeling with mEos2 does not affect the functionality of S2, a small ribosomal

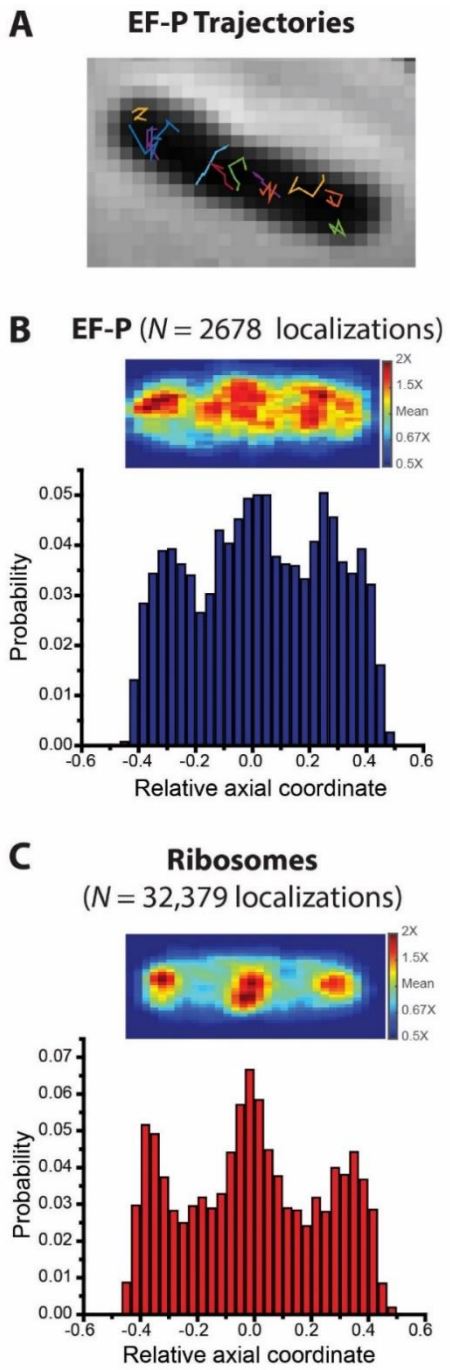


Fig. 4.2 (**A**) Phase contrast image of an example cell overlaid with trajectories of single EF-P–mEos2 copies. Imaging at 2 ms/frame. (**B**) *Top:* Localization probability density heat map of 2678 EF-P–mEos2 copies imaged at 2 ms/frame in different cells of length 3.5 – 4.5 µm. Each location is placed on a common scale of relative axial position. Only molecules that lasted at least 7 frames contribute to the distribution. *Bottom:* Distribution of axial projections on the same relative scale. (**C**) *Top:* Localization probability density map of 32,000 ribosome copies (30S labeled by mEos2). Same imaging conditions as for EF-P–mEos2. *Bottom:* Distribution of axial projections on the same relative scale.

protein that binds to the exterior of the 30S subunit. The spatial distribution of ribosomes (30S-mEos2 labeling) was obtained by imaging a combination of 30S subunits and translating 70S ribosomes at the same 2 ms intervals. The same scaling procedure was applied to localizations of ribosome trajectories that lasted at least seven frames in cells of length 3.5 – 4.5 μm (Fig. 4.2*C*).

The ribosome axial distribution exhibits three distinct peaks, consistent with our previous studies (30). As before, translating ribosomes are strongly segregated from the nucleoid regions. Comparing Figs. 4.2*B* and 4.2*C*, we see that the axial distribution of EF-P also exhibits three peaks, qualitatively similar to the ribosomes. This is consistent with some degree of binding of EF-P to ribosomes. However, the peak-to-trough ratio in the axial distribution of EF-P is not as large. Evidently only a fraction of the EF-P copies resides in the ribosome-rich regions at any given time. Detailed diffusion measurements described below will quantify the fraction of EF-P associated with ribosomes.

Drug effects on axial distributions: chloramphenicol and rifampicin

In vitro measurements have suggested that translational pausing and the resulting empty ribosome E site are prerequisites for binding of EF-P between the E and P sites of ribosomes (1). To further understand the binding of EF-P to ribosomes in live cells, we investigated the axial distribution of EF-P–mEos2 after treatment with the translation-halting drug chloramphenicol (Cam). Cam binds between the P and A sites of the 70S ribosome, inhibiting peptidyl transfer and stalling translating ribosomes in place (34-36). Binding of Cam presumably leaves the space between E and P sites more often available for binding by EF-P, due to dissociation of tRNA from the E site after translation is halted.

Our previous studies have shown that after treatment with Cam, very strong nucleoid–ribosome segregation occurs (37). We generated the axial distributions of EF-P mEos2 and of

30S-mEos2 at t = 30 min after treatment with 200 μg/mL of Cam (Fig. A4.1*A*, *B*). The average cell length decreases dramatically upon Cam treatment (37). To obtain sufficient data, cell lengths of 2.5–3.5 μm were selected. After Cam treatment, the ribosomes show a very clear two-peaked axial distribution, consistent with previous work (30). The EF-P axial distribution closely resembles that of ribosomes (Fig. A4.1*B*), indicating that EF-P strongly associates with 70S ribosomes whose translation has been halted by Cam.

Rifampicin (Rif) halts transcription and leads to depletion of mRNA-bound 70S ribosomes. After 3 hr of treatment with Rif, most ribosomal subunits are free (37). These free 30S and 50S subunits mix with the expanded nucleoids, and the 30S spatial distribution becomes essentially homogeneous throughout the cytoplasm (30, 38). We used the same imaging conditions for EF-P–mEos2 and 30S–mEos2 molecules at t = 3 hr after treatment with 300 μg/mL of Rif. Like 30S, the spatial distribution of EF-P molecules shows little or no axial segregation (Fig. A4.1*C*, *D*). However, the two rather homogeneous axial distributions do not determine whether EF-P is unbound or associates with 50S or 30S ribosomal subunits. Diffusion measurements after drug treatment will shed light on this question.

Axial distribution of EF-P^{K34A} expressed from a plasmid in normal growth

In vitro studies have shown that Lys-34 of EF-P is essential for functionality (4, 13, 15). Deletion of EF-P or expression of mutated, non-functional EF-P increases the doubling time by a factor of 2 and induces other growth defects as well (1, 14, 15). To study the importance of Lys-34 in live *E. coli*, we constructed a strain containing an inducible plasmid that expresses the mutant form EF-P^{K34A}–mEos2 (Table A4.1*A*). Fully functional, unlabeled EF-P continues to be expressed from the chromosome to ensure normal growth rate and cell functionality.

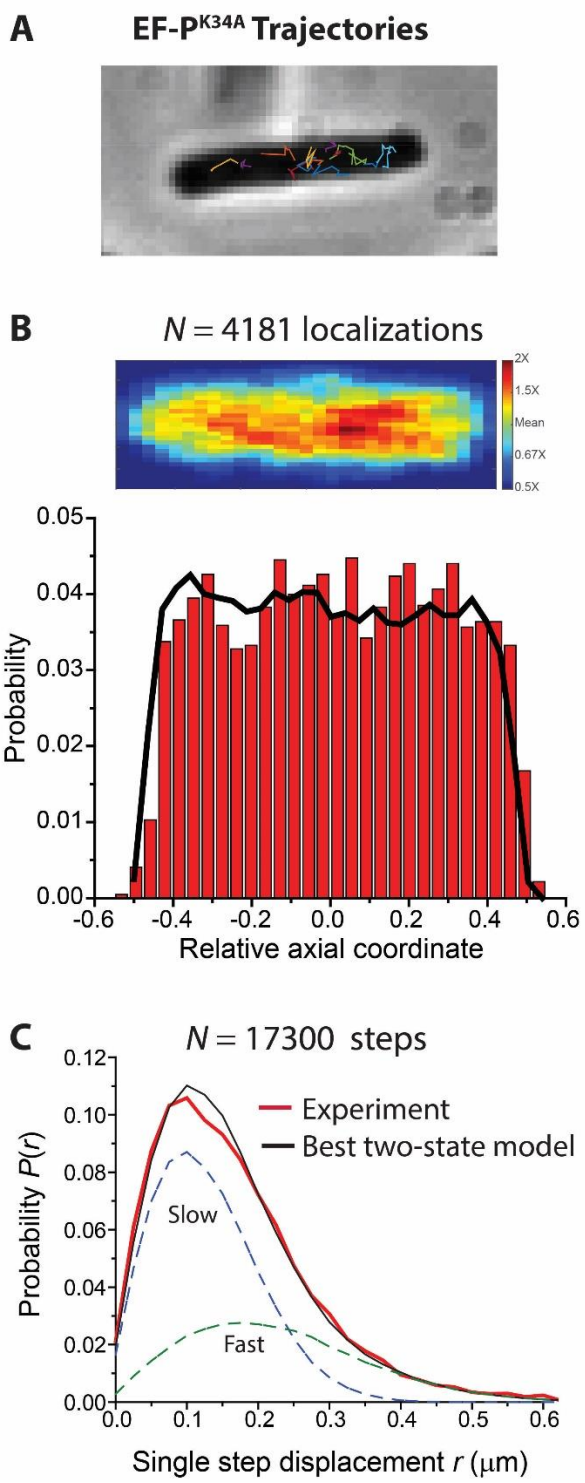


Fig. 4.3 (**A**) Phase contrast image of an example cell overlaid with trajectories of single mutant EF-P^{K34A}–mEos2 copies. Imaging at 2 ms/frame. (**B**) *Top:* Localization probability density heat map of 4180 EF-P^{K34A}–mEos2 copies imaged at 2 ms/frame in different cells of length 3.5 – 4.5 μm. Each location is placed on a common scale of relative axial position. Only molecules that lasted at least 7 frames contribute to the distribution. *Bottom:* Distribution of axial projections on the same relative scale. For comparison, axial distribution of molecules uniformly filling a spherocylinder of diameter 0.9 μm and length 4 μm is shown in black. (**C**) *Red:* Experimental probability distribution of 17,300 single-step displacements taken by EF-P^{K34A}–mEos2 molecules in 2 ms. *Black:* The best unconstrained fit to a static two-state model (without transitions). Model parameters: D_{slow} = 3.2 μm^2/s (σ_{slow} = 50 nm), f_{slow} = 0.65, D_{fast} = 9.7 μm^2/s (σ_{fast} = 90 nm), f_{fast} = 0.35, with χ_v^2 = 2.0. The slow and fast components are shown as dashed lines as labeled.

Accordingly, the doubling time for the mutant strain is almost identical to that of the VH-1000 parent strain (Table A4.1*B*).

We studied the axial distribution of EF-P^{K34A}–mEos2 after induction of the gene under normal growth conditions. The methodology was the same as for the EF-P-mEos2 imaging experiments. The EF-P^{K34A}- mEos2 trajectories (Fig. 4.3*A*) are generally more extended than those of EF-P-mEos2. The axial distribution of EF-P^{K34A}–mEos2 (Fig. 4.3*B*) is not ribosome-like, in sharp contrast to the three-peaked distribution of normal EF-P-mEos2 (Fig. 4.2*B*). Instead, the EF-P^{K34A}- mEos2 molecules are distributed essentially uniformly throughout the cell. For comparison, we simulated the axial distribution for a uniformly filled spherocylinder of 4 μm length and 0.9 μm diameter with the same number of molecules as the experimental axial distribution (solid black line in Fig. 4.3*B*). The experimental axial distribution of EF-P^{K34A}-mEos2 closely resembles the simulated homogeneous distribution; the wild type EF-P-mEos2 distribution (Fig. 4.2*B*) does not. This indicates that EF-P^{K34A} associates with translating ribosomes significantly less than with wild type EF-P, in agreement with previous binding studies *in vitro* (4).

To ensure that the homogeneous distribution of EF-P^{K34A}-mEos2 is not an artifact of overexpression from a plasmid, we constructed a control strain (SM4, Table A4.1*A*) in which the wild type EF-P–mEos2 is expressed from an analogous plasmid alongside the expression of unlabeled, endogenous EF-P from the chromosome. The gene induction conditions are the same in both experiments. The axial distribution of wild type EF-P–mEos2 expressed upon induction from the plasmid is again three-peaked, like that of ribosomes (Fig. A4.2). EF-P–mEos2 expressed from the plasmid also exhibits diffusive behavior very similar to that of EF-P

expressed endogenously from the chromosome. This indicates that the observed behavior of EF-P^{K34A}- mEos2 is not an artifact of expression from a plasmid.

Diffusion of EF-P expressed from the chromosome and of EF-P^{K34A} expressed from a plasmid

The axial distribution of EF-P–mEos2 gives only qualitative information about the fraction of EF-P that is associated with translating ribosomes *in vivo* and no information about the dynamics of the association process. To investigate the diffusive behavior of EF-P in live cells under normal growth conditions, we chose trajectories that lasted at least seven frames, similar to those used for generating the axial distributions. The mean diffusion coefficient $\langle D \rangle$ can be estimated from a plot of the two-dimensional mean-square displacement vs lag time, MSD(τ), using the slope of the first two data points (Fig. A4.4*A*). The MSD(τ) plot is a weighted average over different diffusive states of EF-P molecules. The result $\langle D \rangle$ = 3.4 μm^2/s for EF-P–mEos2 is somewhat smaller than our rough estimate for a freely diffusing molecule of similar size (4 – 10 μm^2/s), but much larger than $\langle D \rangle$ for ribosomes under similar fast imaging conditions (0.5 μm^2/s, Fig. A4.4*A*). Combined with the axial distribution results, this suggests that only a fraction of EF-P copies are associated with the translating ribosomes at a given time.

To quantify the fraction of ribosome-bound EF-P, the chosen trajectories were divided into individual steps with Δt = 2 ms between camera frames, in hopes of isolating short time intervals during which EF-P remains in one diffusive state. The resulting distribution of experimental single-step displacements, $P_{EF\text{-}P}(r)$, is shown for 5200 individual steps in Fig. 4.4*A*. The analogous plot for 30S-labeled ribosomes imaged under the same conditions, $P_{ribo}(r)$, is shown in Fig. 4.4*B*. We analyze $P(r)$ distributions by comparison to the results of a large number of

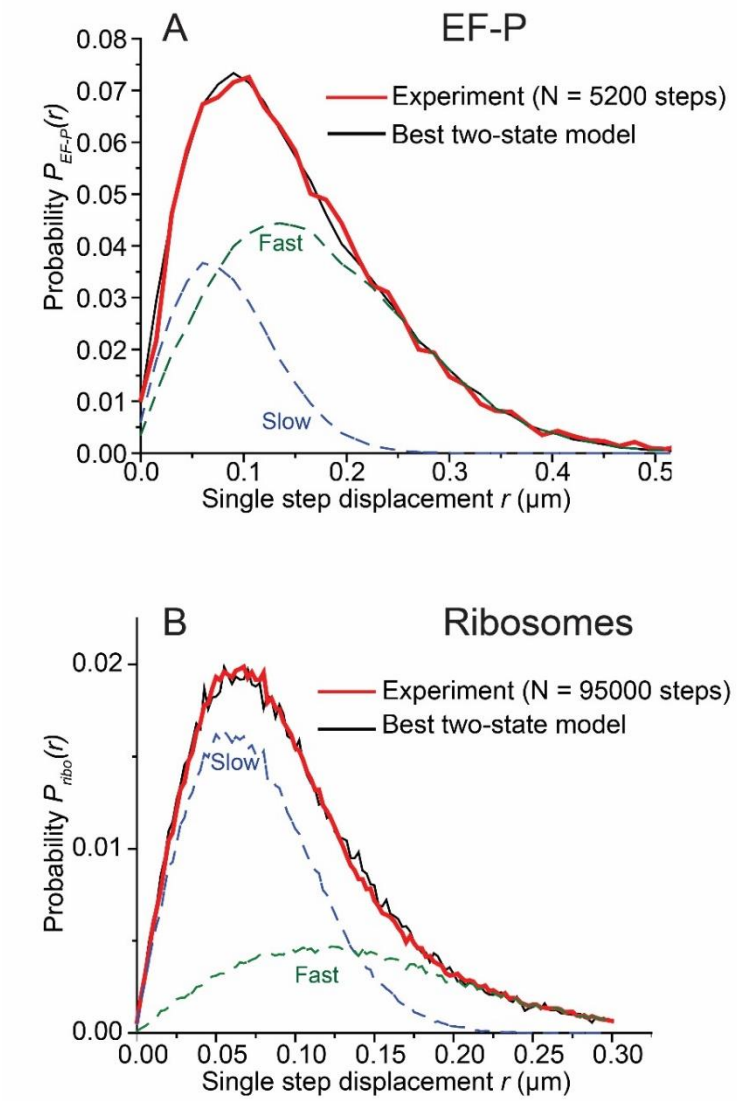


Fig. 4.4 (**A**) *Black:* Experimental probability distribution of single-step displacements taken by EF-P–mEos2 molecules in 2 ms. *Red:* The best fit to a static two-state model (without transitions) with D_{slow} constrained. Model parameters: D_{slow} = 0.2 μm^2/s (σ_{slow} = 50 nm), f_{slow} = 0.30, D_{fast} = 4.3 μm^2/s (σ_{fast} = 75 nm), f_{fast} = 0.7, with a χ_v^2 = 1.0. The individual slow and fast components are shown in dashed lines as labeled. (**B**) *Black:* Experimental probability distribution of single-step displacements taken by ribosomes (30S labeled with mEos2) in 2 ms. *Red:* The best unconstrained fit to a static two-state model (without transitions). Model parameters: D_{slow} = 0.20 μm^2/s (σ_{slow} = 40 nm), f_{slow} = 0.65, D_{fast} = 0.8 μm^2/s (σ_{fast} = 75 nm), f_{fast} = 0.35, with χ_v^2 = 2.3. The slow and fast components are shown as dashed lines as labeled.

simulated random walk trajectories that incorporate dynamic localization error σ and confinement within a spherocylinder that mimics the dimensions of an *E. coli* cell. Details are provided in SI. For each chosen diffusion coefficient D and measurement error σ, the simulations provide a numerical function we call $P_{model}(r;D)$. We attempt to fit the experimental distribution $P(r)$ in a least-squares sense to a single population or to a weighted average of two static populations. The goodness of fits was judged by the reduced chi-square statistic χ_v^2 (39). For a one-state model the only fitting parameter is D. For unconstrained models including two static (non-exchanging) states, the fitting function is the linear combination $P_{model}(r) = f_{slow}P(r;D_{slow}) + (1 - f_{slow})P(r;D_{fast})$. The three fitting parameters are D_{fast}, D_{slow}, and the fractional population f_{slow}, which in turn fixes $f_{fast} = (1 - f_{slow})$.

First we analyzed the single-step ribosome distribution $P_{ribo}(r)$ from 15,800 6-step trajectories (Fig. 4.4*B*, Table 4.1). All single-population fits were poor. The best fit to a sum of two static populations yielded $f_{slow} = 0.65$, $D_{slow} = 0.2$ μm^2/s (presumably the 70S ribosomes engaged in translation as polysomes of variable size), $f_{fast} = 0.35$, and $D_{fast} = 0.8$ μm^2/s (presumably the free 30S subunits). The minimum goodness-of-fit parameter (reduced chi-square) was $\chi_v^2 = 2.3$, indicating a reasonable fit given the noise in both experimental and model histograms; see Appendix for details. The MSD plot yielded a mean diffusion coefficient for ribosomes under imaging at 2 ms/frame of $D_{ribo} = 0.5$ μm^2/s (Fig. 4.4*A*), consistent with previous measurements by Elf, *et al*. under similarly fast imaging conditions (28). The value $D_{slow} = 0.2$ μm^2/s is similar to that used by Plochowietz *et al*. to identify tRNA associated with translating ribosomes (29).

For the experimental $P_{EF\text{-}P}(r)$ distribution of Fig. 4.4*A*, we again found no good one-state fits. Then we tested a two-state model of EF-P with no exchange between states on the 2-ms

timescale of the single steps. The two states were a slowly diffusing, ribosome-bound state (D_{slow}) and a rapidly diffusing, free state (D_{fast}), with population fractions f_{slow} and f_{fast}. The large measurement error precludes accurate determination of D_{slow}, as detailed in Appendix. The root-mean-square displacement of a ribosome bound (slow) EF-P copy is comparable to the typical static localization error of our experiments (~60 nm). We therefore constrained D_{slow} = 0.2 μm^2-s^{-1} (to match that of ribosomes) and optimized f_{slow} and D_{fast} to obtain the best fit to the experimental $P_{EF\text{-}P}(r)$. Our best parameter estimates are: f_{slow} = 0.30 ± 0.10, D_{fast} = 4.3 ± 1.0 μm^2-s^{-1}, and f_{fast} = 0.70 ± 0.10, with the best χ_v^2 = 1.0 indicating a good fit (Table 4.1). This fit is compared with the data in Fig. 4.4*A*. Error estimation for the parameters is described in Appendix. We assign the slowly diffusing EF-P fraction as the copies bound to translating 70S ribosomes, consistent with the axial spatial distribution. The rapidly diffusing fraction is attributed to free EF-P.

Under Cam treatment, we expect more E sites available for EF-P binding than in normal growth. The best fit two-state $P_{model}(r)$ was obtained for f_{slow} = 0.45 (with D_{slow} constrained to 0.2 μm^2/s) and f_{fast} = 0.55, D_{fast} = 1.2 μm^2/s, yielding χ_v^2 = 1.2 (Fig. 4.3*A*, Table 4.1). Cam treatment apparently increases the fraction of slowly diffusing EF-P and also decreases D_{fast} from 4.3 to 1.2 μm^2/s. The more rapidly diffusing component under Cam treatment is likely experiencing transient binding events to 70S ribosomes, consistent with the suggestion that Cam induces more binding opportunities.

In contrast, under rifampicin treatment, we obtained the best-fit $P_{model}(r)$ for two static populations with f_{slow} = 0.55, D_{slow} = 4.6 μm^2/s, f_{fast} = 0.45, and D_{fast} = 8 μm^2/s, with χ_v^2 = 1.5 (Fig. A4.3*B*, Table 4.1). Fits of the experimental $P(r)$ to a single population yielded much larger values of χ_v^2. Both diffusion coefficients are much faster than those of translating 70S ribosomes

and of free 30S and 50S subunits. The diffusion data shows no clear evidence of EF-P association with non-translating 30S or 50S ribosomal subunits. The corresponding homogenous axial distribution of EF-P after Rif treatment (Fig. 4.1*C*) is evidently due to essentially freely diffusing EF-P.

For EF-P^{K34A}–mEos2, a similar procedure resolved the experimental $P(r)$ into two static populations with f_{slow} = 0.65, D_{slow} = 3.2 μm^2/s, f_{fast} = 0.35, and D_{fast} = 9.7 μm^2/s, with χ_v^2 = 1.96 (Fig. 4.3*C*, Table 4.1). This is similar to the result for wild-type EF-P after Rif treatment, in that there is no clear evidence of a slowly diffusing component of EF-P^{K34A} that moves like translating ribosomes. Based on the spatial distribution as well as the diffusion measurements, we see almost no evidence of EF-P^{K34A} associating to the translating ribosomes. This is consistent with *in vitro* studies (4) that have shown a 30-fold reduction in the binding affinity of EF-P^{K34A} with ribosomes compared with wild type EF-P. With D_{slow} constrained to the ribosome-like value of 0.2 μm^2-s^{-1}, the best fit is qualitatively poor and the χ_v^2 value doubles. Evidently EF-P^{K34A} interacts much more weakly with translating ribosomes than does wild-type EF-P, consistent with the observed homogeneous spatial distribution. The value D_{slow} = 3.2 μm^2/s admits the possibility of weaker, shorter-lived interactions with 70S ribosomes.

Dynamics of EF-P association with ribosomes

The fit to the single-step $P_{EF\text{-}P}(r)$ distribution assumed two static populations that persist on the 2-ms timescale of single camera frames. Next we asked whether six-step trajectories (12 ms total duration) could provide information about the timescale of possible transitions between the slow, ribosome-associated state and the fast, freely diffusing state of EF-P. Accordingly, for 859 six-step trajectories of wild type EF-P–mEos2, we generated the distribution $P_{EF\text{-}P}(\langle r \rangle_{6\text{-step}})$, with $\langle r \rangle_{6\text{-step}}$ the mean displacement of the six steps. See Appendix for details. The experimental

distribution is shown in Fig. 4.5. Comparing Figs. 4.5 and 4.4, we see that $P_{EF\text{-}P}(\langle r\rangle_{6\text{-step}})$ is narrower than $P_{EF\text{-}P}(r)$ due to the averaging of six successive displacements.

We simulated $P_{model}(\langle r\rangle_{6\text{-step}})$ distributions for static populations by averaging the displacements from six-step simulated trajectories that included confinement and measurement error. If the EF-P molecules undergo transitions only rarely during a 12-ms trajectory, then the experimental $P(\langle r\rangle_{6\text{-step}})$ should match the $P_{model}(\langle r\rangle_{6\text{-step}})$ using the same input parameters that gave the best-fit $P_{model}(r)$ distribution. It does not. We simulated 15,000 six-step trajectories, 30% with D_{slow} = 0.2 μm^2/s (4500 trajectories) and 70% with D_{fast} = 4.3 μm^2/s (10,500 trajectories) in accord with the parameters of the best-fit $P_{model}(r)$ distribution of Fig. 4.4*A*. $P_{model}(\langle r\rangle_{6\text{-step}})$ for these static populations exhibits two partially resolved peaks (Fig. 4.5), unlike the experimental distribution. This suggests that the wild type EF-P indeed undergoes transitions during the six-step, 12-ms long trajectories.

In order to estimate the lifetime of the two diffusive states, we introduced two-state, binding/unbinding kinetics into our simulations, following Das *et al* (33, 40). Details are provided in Appendix. We generated large sets of six-step simulated trajectories with D_{slow} = 0.2 μm^2/s, D_{fast} = 4.3 μm^2/s and τ_{free}/τ_{bound} = 7/3; this lifetime ratio is fixed by the fractional populations from the best-fit $P_{model}(r)$ distribution. Here τ_{free} and τ_{bound} are the lifetimes of the free and ribosome-bound states of EF-P, respectively. While keeping the ratio τ_{free}/τ_{bound} constant, we varied τ_{free} from 0.1 ms to 100 ms, calculating 15,000 trajectories in each case. The best fit to the experimental $P_{EF\text{-}P}(\langle r\rangle_{6\text{-step}})$ was obtained for τ_{free} = 16 ms and τ_{bound} = 7 ms, with χ_v^2 = 1.4. This fit is compared with experiment in Fig. 4.5. Additional calculated distributions $P_{model}(\langle r\rangle_{6\text{-step}})$ for different assumed lifetimes are compared with experiment in Fig. A4.6. Based on the

quality of the different fits (detailed in Appendix), we take τ_{free} = 16 ± 5 ms and τ_{bound} = 7 ± 2 ms as best estimates of the average lifetime of free EF-P and of EF-P bound to ribosomes, respectively. The typical timescale for an EF-P/ribosome binding/unbinding cycle is then (τ_{bound} + τ_{free}) ~ 23 ms.

There are about 40,000 translating ribosomes in the average *E. coli* cell volume of ~3 μm^3, a concentration of ~2.2 x 10^{-5} M. The effective bimolecular rate constant for EF-P binding to translating ribosomes in the living cell can then be estimated as: $k_{bimol} = \tau_{free}^{-1}$ [ribosomes]$^{-1}$ ~ 2.9 x 10^6 $M^{-1}s^{-1}$. This is ~100 times slower than the diffusion limited rate constant for EF-P/ribosomes collisions, k_{diff} ~ 4 x 10^8 $M^{-1}s^{-1}$. The inefficiency is likely due to the small target site on the ribosome, the small fraction of an elongation cycle during which the E site is open for EF-P binding, and the steric requirements for binding.

Analogous lifetime analysis of the EF-P diffusion data after Cam treatment gave best-fit lifetimes of τ_{free} = 9 ms and τ_{bound} = 7 ms. The ribosome stalling under Cam treatment is presumably essentially permanent (41). There is no active translation that might push EF-P from its binding site as a deacylated tRNA translocates into the E site. This suggests that τ_{bound} ~ 7 ms is the natural lifetime of an EF-P/ribosome complex independent of whether or not elongation is actually occurring. The cycle time for binding/unbinding events under Cam treatment becomes (τ_{bound} + τ_{free}) ~ 16 ms, shorter than in normal growth because more binding sites are available for the search process and the binding sites presumably remain open.

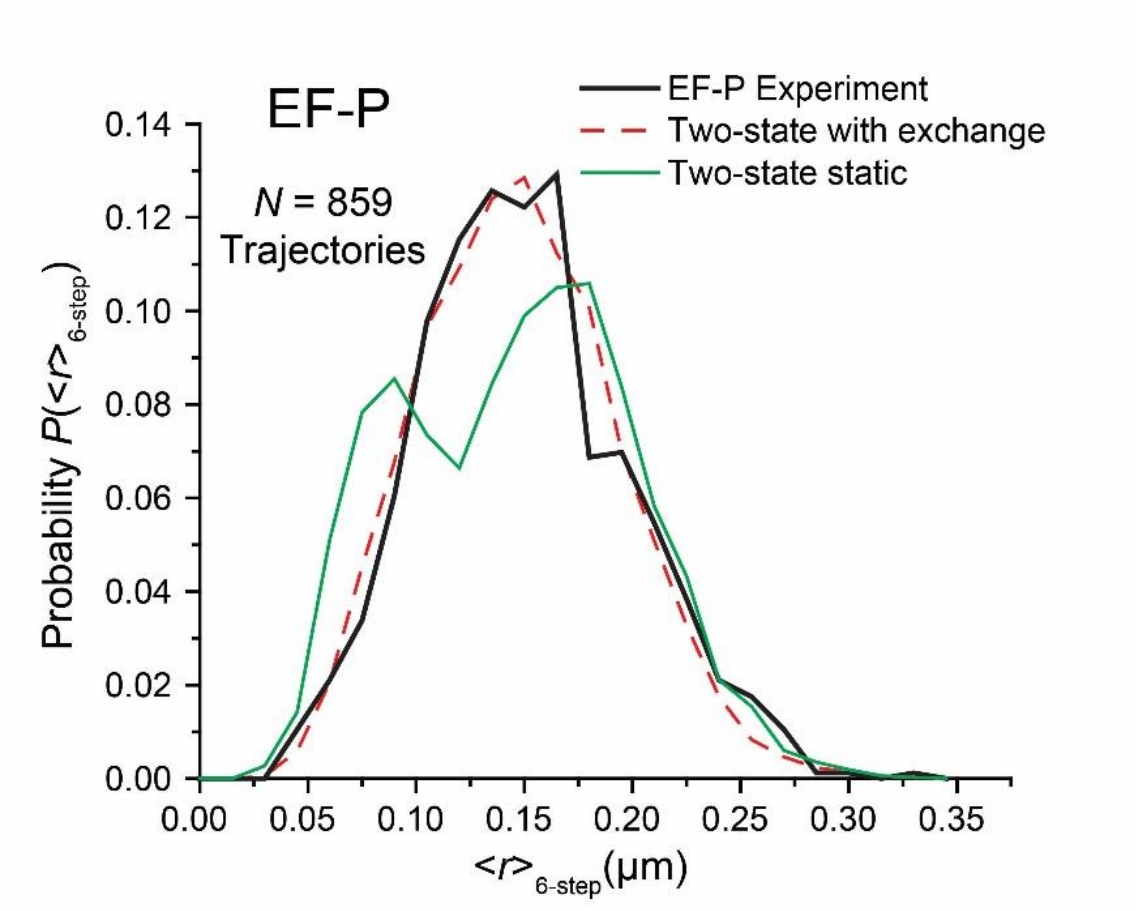


Fig. 4.5 *Black:* Experimental probability distribution of the mean of six consecutive steps from each trajectory for EF-P–mEos2 molecules. Longer trajectories truncated to 6 steps. *Red dashed line:* Best-fit two-state model with binding-unbinding kinetics. Model parameters: D_{slow} = 0.2 $\mu m^2/s$ (σ_{slow} = 50 nm), f_{slow} = 0.3, D_{fast} = 4.3 $\mu m^2/s$ (σ_{fast} = 75 nm), f_{fast} = 0.7, τ_{free} = 16 ms and τ_{bound} = 7 ms, with χ_v^2 = 1.4. *Green line:* For comparison, a simulated static two-state model (no transitions) using same diffusion coefficients and fractions gave χ_v^2 = 10.1.

Discussion

The crystal structure indicates that EF-P binds between the P and E sites of a translating 70S ribosome, *i.e.*, a binding event requires an empty E site (12). Studies of translation kinetics *in vitro* and *in vivo* found that wild-type EF-P binds to translating ribosomes and alleviates pausing at Pro-Pro motifs (4, 6-8). Those studies further suggested that pausing during a slow peptidyl transfer step may be a prerequisite to EF-P binding. For Pro-Pro motifs, the release of deacylated tRNA from the E site during the elongation round preceding the translational pause presumably renders the E site accessible sufficiently long for free EF-P to find its binding site. The bound EF-P then catalyzes more efficient incorporation of the P-site Pro into the nascent peptide chain. Kinetics data also showed that without the β-lysine modification at Lys-34, binding of EF-P to translating ribosomes diminished 30-fold (4).

Our results *in vivo* are consistent with the binding and mechanistic insights gleaned from the work *in vitro* and add new information about the timescale and frequency of typical EF-P/ribosome interactions. The axial spatial distribution of wild-type EF-P is three-peaked, qualitatively similar to that of the ribosomes (Fig. 4.2). This shows that EF-P concentrates in the ribosome-rich regions where most translation occurs. Based on the diffusion measurements, about 30% of EF-P copies are associated with translating ribosomes at a given time (Fig. 4.4*A*). The translation halting drug chloramphenicol is known to inhibit peptidyl bond formation and thereby stall translation. Cam treatment increases the fraction of EF-P associated with ribosomes to about 45% (Fig. A4.4*A*, Table 4.1), consistent with the idea that pausing and opening of the E site enable EF-P binding. The diffusion studies after Rif treatment (Fig. A4.3*B*, Table 4.1) show no clear evidence of binding of EF-P to free ribosomal 30S or 50S subunits.

The new data also indicate that there is little or no binding *in vivo* of the mutant form EF-P^{K34A} to 70S ribosomes. The spatial distribution of EF-P^{K34A} is homogenous (Fig. 4.3*B*), unlike the three-peaked axial distribution of wild-type EF-P and ribosomes. In addition, the diffusion measurements show no clear evidence of a fraction of the EF-P^{K34A} population that diffuses like the slow, translating ribosomes (Fig. 4.3*C*). However, the one-step $P(r)$ for EF-P^{K34A} is best fit to two populations, with D_{slow} = 3.2 μm^2/s and D_{fast} = 9.7 μm^2/s (Table 4.1). It is possible that the slower component results from very fast exchange between short-lived EF-P^{K34A}/ribosome complexes and freely diffusing EF-P^{K34A}.

For wild-type EF-P, two-state analysis of the diffusion data yielded rough estimates of the mean search time for EF-P to find a binding site (τ_{free} ~ 16 ms) and the mean duration of a typical binding event (τ_{bound} ~ 7 ms). The mean protein production rate in rapidly growing *E. coli* is ~20 amino acids per ribosome per sec. This sets an upper limit of ~50 ms on the average duration of a single elongation cycle, including aminoacyl-tRNA binding, peptide bond formation, and translocation. Our estimate of τ_{bound} is much shorter than the average elongation cycle time.

The detailed nature of these binding events is uncertain. *In vitro*, EF-P associates with translating, 70S ribosomes in a stoichiometric ratio of 1:1 (11). Quantitative interpretation of our results requires an estimate of the EF-P copy number. An early analysis using radioactive labeling and two-dimensional gels found that EF-P is present at about 0.1 copies per ribosome, independent of growth conditions (42). In our growth conditions, on average there are ~50,000 total 30S copies per cell (30), suggesting only ~5000 EF-P copies. However, a recent mass spectrometric study found ~20,000 EF-P copies per cell, a substantially larger number (43).

Some 80% of 30S copies are incorporated into 70S (translating) ribosomes, indicating 40,000 translating ribosomes in our conditions. The $P_{EF\text{-}P}(r)$ data indicate that roughly 30% of the EF-P copies are bound to a translating ribosome at a given moment. Depending on which copy number estimate we use, this implies that a snapshot would show 1500–6000 EF-P/ribosome complexes at a given moment in time.

For either copy number estimate, there are evidently too few Pro-Pro motifs in the single-cell transcriptome of *E. coli* to account for so many simultaneous EF-P/ribosome binding events. First, we use the most recent mass spectrometric data (43) and known gene sequences to estimate the fraction of Pro-Pro motifs within the set of transcripts that have been translated to produce the detailed set of proteins present per *E. coli* cell. Then we multiply the fraction of Pro-Pro motifs by the mean number of mRNA codons present in one cell to estimate the total number of Pro-Pro motifs present in the single-cell transcriptome at a given time. Our estimate assumes similar degradation rates for all proteins, so that the proteome provides a representative sample of the distribution of translated codon motifs.

A recent mass spectrometric study found 1611 proteins with abundances greater than 20 per cell for *E. coli* strain MG1655 growing exponentially in glucose minimal medium (43). The 1611-member sample had an average of 260 codons per protein. The total number of such protein copies was 4.8×10^6 per cell. Those proteins comprise 1.3×10^9 total amino acids, whose corresponding codons we assume to form a representative sample of the mRNA codons that are transcribed to make the proteome of one cell. For each of the 1611 genes in that codon sample, we multiplied the number of Pro-Pro motifs by the corresponding protein copy number to obtain the total number of Pro-Pro motifs in the sample, which is 6.6×10^5. Less than 3% of the total

number of Pro-Pro motifs comprise Pro-Pro-Pro triplets. The fractional occurrence of Pro-Pro motifs that were translated to form the representative protein sample is roughly $6.6 \times 10^5 / 1.3 \times 10^9 = 5.3 \times 10^{-4} \sim 0.05\%$.

How many total codons comprise the transcriptome (mRNA content) of the typical *E. coli* cell in exponential growth? The classic literature estimate for total mRNA copy number is 1200 (44) for the B/r strain growing in minimal medium at 37°C (doubling time 40 min, similar to our VH1000 strain in EZRDM at 30°C). We take 2000 mRNA/cell as an estimate for our conditions. The total number of Pro-Pro motifs present at a given moment in one *E. coli* cell is then ~2000 mRNA x 260 codons/mRNA x $5.3 \times 10^{-4} \sim 280$ Pro-Pro/cell. That is 5 times smaller than the lower estimate of 1500 simultaneous EF-P/ribosome binding events and 20 times smaller than the higher estimate of 6000 simultaneous binding events.

Thus it seems unlikely that the rare events in which EF-P is relieving a pause at a Pro-Pro motif cause a significant fraction of the binding events we have detected. Of course there are other motifs that cause pauses that are rescued by EF-P, but these are a minority of the pausing events detected in ribosome profiling studies on *efp*$^-$ strains. Instead, it seems plausible that the typical EF-P/ribosome binding event detected in our study is a kind of "housekeeping" event involving short-lived open E sites that may arise routinely when a wide variety of codons are present at the P site. When an EF-P finds an open E site, it typically binds very briefly ($\tau_{bound} \sim 7$ ms) before dissociating and beginning another search. The search lasts for only a short time ($\tau_{free} \sim 16$ ms). If the typical timescale of a searching/binding cycle is ($\tau_{free} + \tau_{bound}$) ~ 23 ms, then each EF-P copy would carry out some 40 cycles per sec.

According to these estimates, the 5000–20,000 EF-P copies combined would carry out ~200,000–800,000 “ribosome E-site interrogations” per sec. On average, each of the 40,000 translating ribosomes would experience an EF-P binding event about 5–20 times per sec. The 40,000 ribosomes translating at 20 aa per sec carry out a total of 800,000 elongation cycles per sec, which is comparable to the rate of interrogation by EF-P. This means that roughly 25-100% of all translation cycles leave the E site open for a sufficiently long period to enable EF-P to bind. Definitive determination of the EF-P copy number would sharpen that estimate. Regardless, the typical binding time of ~7 ms remains short compared with the overall elongation cycle time of ~50 ms. Our data do not constrain the binding time during those evidently rare events in which EF-P binds to a ribosome that is stalled at a Pro-Pro motif. The Pro-Pro rescue time could be similar to or substantially longer than 7 ms. However, we can estimate that each of the 40,000 translating ribosomes experiences at least 5 and perhaps as many as 20 of the brief EF-P visits per sec. Evidently the cell contains enough EF-P copies to ensure that long translational pausing events, such as those that might occur at Pro-Pro, would catch an EF-P on a timescale of ~200 ms or less.

It is possible that most of these brief EF-P binding events are non-functional, but essentially harmless, due to their short duration compared with an elongation cycle. Such frequent visits may at least serve the purpose of expediting the relief of pauses at Pro-Pro duets when they do arise. The stalled ribosome need not wait long for an EF-P arrival. Alternatively, the typical EF-P binding event help to alleviate brief pauses that are not yet detectable in ribosome profiling studies, because they are too short to be observed as a build-up of ribosome density. Or these events may serve some different purpose.

Finally, the occurrence of an EF-P binding event during many or most elongation cycles is consistent with recent single-molecule translation studies inferring that the deacylated tRNA dissociates from the E site after or during the translocation step (20-22). Our *in vivo* data are not consistent with a concerted mechanism of tRNA progression through the A, P, and E sites, *i.e.*, a mechanism in which the E site is never completely empty. Our results do not speak to the longstanding issue of whether or not dissociation of the deacylated tRNA from the E site occurs before or after binding of a new aa-tRNA at the A site (20-25).

Methods

Strain construction

All experiments were performed on *E. coli* of background strain VH1000. Wild-type and modified strains are described in Table A4.1. For imaging EF-P, we constructed an *E. coli* strain that expresses EF-P translationally fused with the photoswitchable protein mEos2 at the C-terminus.

The oligonucleotides used for this work are listed in Table A4.1. The mEos2-encoding gene was first amplified with primer *meos_F* and *meos_R* from plasmid pUC57-mEos2 and then sub-cloned into the plasmid pREST-B between *BamHI* and *EcoRI* sites to generate a tandemly placed mEos2 ORF and kanamycin resistance (kan^R) cassette separated by a linker containing independent ribosome binding site for kan^R. A linear DNA containing mEOS2- kan^R cassette was subsequently amplified with primer *efpmeos_F* and *efpmeos_R* which contained sequences homologous to immediate up- and down-stream region of the stop codon the *efp* gene, present at the b4147 locus in the genome of *E. coli* MG1655 (ECK4141). Next, the stop codon of the *efp* gene on the chromosome of *E. coli* MG1655 was replaced with mEOS2- kan^R cassette using λ-

Red recombineering (45-47). The resulting recombinants with EFP-mEos2 fusion were selected by screening for kanamycin resistance and further verified by DNA sequencing. One of the successful recombinants was named XG501 (MG1655 *efp-mEos2 Kan*), which was subjected to growth analysis and further strain preparation. The *efp-mEos2 Kan* fusion was then moved into a clean VH1000 background by P1 transduction, resulting in strain SM1.

The *efp-mEos2* cassette was PCR amplified by forward primer EM1 introducing BamHI and reverse primer EM2 introducing HindIII. The PCR amplified product was ligated into pASK-IBA3plus (Invitrogen, Carlsbad, CA) double digested with BamHI and HindIII. VH1000 cells were transformed with the resulting plasmid (pSM1) and transformants were selected on ampicillin-containing plates. This strain (SM4) was used for control experiments for imaging EF-P–mEos2 being expressed from the plasmid upon induction.

PCR amplification of pSM1 using oligonucleotides Mut1 and Mut2 and the commercially available Phusion Site-Directed Mutagenesis Kit was used to mutagenize the Lys-34 residue to Ala. This PCR amplified product was further ligated to generate pSM2. VH1000 cells were transformed with pSM2 and transformants (SM8) were selected on ampicillin-containing plates.

The doubling time of strain SM1 in bulk EZ rich defined medium (EZRDM) at 30°C (48) is 60 ± 3 min (Table A4.1*B*), a factor of 1.3 longer than the doubling time of 45 min of the wild type VH1000 under the same conditions (30). For the plasmid-containing strains SM4 and SM8, doubling times without induction were 51 ± 1 min and 50 ± 6 min, respectively (Table A4.1*B*).

Cell growth and preparation for imaging

Bulk cultures from frozen glycerol stock solution and subcultures for imaging were grown at 30°C with continuous shaking in EZRDM, which is a MOPS-buffered solution with supplemental metal ions (M2130; Teknova), glucose (2 mg/mL), supplemental amino acids and

vitamins (M2104; Teknova), nitrogenous bases (M2103; Teknova), 1.32 mM K_2HPO4, and 76 mM NaCl.

The strains that include the plasmid expressing EF-P–mEos2 or EF-P^{K34A}–mEos2 were grown with addition of 100 µg/mL ampicillin. When cells had grown to midlog phase, anhydrotetracycline was added to a final concentration of 45 nM to induce the expression of the desired protein. After 5 min of induction, the cells were centrifuged and resuspended in fresh EZRDM with 100 µg/mL ampicillin to remove the inducer. The cells were then incubated again in growth media for 30 – 45 min at 30ºC to enable maturation of the fluorescent protein prior to imaging.

Rifampicin and chloramphenicol stock solutions were prepared by dissolving 10 and 15 mg of the drugs in 0.5 mL of ethanol, respectively. Stock solutions of the drugs were added to the midlog cell culture to attain a final concentration of 300 µg/mL and 200 µg/mL respectively. For imaging of EFP under rifampicin and chloramphenicol treatment, the cultures were shaken for 3 hr and 30 min, respectively, before plating and imaging.

Superresolution imaging of live *E. coli* cells

Fluorophore-labeled cells were grown overnight with shaking at 30°C in EZRDM. Subcultures were made by diluting the stationary phase culture at least 1:100 into 2 mL of fresh EZRDM. Subcultures were grown to exponential phase (OD = 0.2–0.6 at 600 nm), upon which the culture was placed in CoverWell Perfusion Chamber Gaskets (Invitrogen, Carlsbad, CA) on a polylysine-coated coverslip. The volume of the closed chamber is 150 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. Live cell imaging was carried out at 30°C for no longer than 30 min after plating. During that time, cells continue to grow.

Cells were imaged using an Eclipse Ti inverted microscope (Nikon Instruments, Melville, NY) equipped with an oil immersion objective (CFI Plan Apo Lambda DM 100x Oil, 1.45 NA; Nikon Instruments), a 1.5x tube lens, and the Perfect Focus System (Nikon Instruments, Melville, NY). For time-lapse imaging, fast shutters (Uniblitz LS2; Vincent Associates, Rochester, NY) were used to synchronize illumination and image acquisition. Images were recorded by a fast back-illuminated electron-multiplying charge-coupled device (EMCCD) camera with 128 x 128 pixels of 24 x 24 μm each (iXon DV-860; Andor Technology, South Windsor, CT). Each pixel corresponds to 160 x 160 nm^2 at the sample (150x overall magnification). All data were collected at a frame rate of 485.4 Hz, with exposure time within each frame of 2 ms. The 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, Bloomfield, CT). The 405-nm power density at the sample was 15-25 W/cm^2. The photoactivation laser remained on throughout imaging. Power density of the 561 nm laser was kept at ~8 kW/cm^2. The probe laser was on continuously in the 2 ms/frame tracking experiments. To minimize the phototoxic effect of the laser, we collected data for <25 s per cell. The mEos2 emission was collected through a 617/73 bandpass filter (bright line 617/73-25; Semrock, Rochester, NY).

Single molecule image analysis

Images were analyzed using a MATLAB GUI developed in our lab (49). Images were smoothed and filtered to obtain a zero-based image. Bright spots were located with pixel level accuracy by a peak-finding algorithm that finds local intensity maxima within an image. A user-defined intensity threshold was used as the minimum brightness of a pixel from single molecule

trajectories. It is carefully set by the user so that it will not be so high as to cut a long trajectory short or so low as to include background noise.

Centroids of the bright spots were calculated from a 7x7 pixel square centered on the local maxima determined by the peak-finding algorithm. As the images are asymmetrically blurred due to diffusion during the frame time, we calculate the centroid of the bright spots instead of using Gaussian fitting. This centroid analysis is fast and easily implemented for analysis of Monte Carlo modeling results as well. For single-molecule tracking analysis, the (x, y) positions of the centroid are stored and connected into trajectories using a modified MATLAB version of the tracking program written by Crocker and Grier (50).

In order to generate the axial distribution of molecules from several cells, the camera-based coordinates are re-oriented so that the x-axis corresponds to the long cell axis. The axial positions are scaled from -0.5 to 0.5 using the cell length determined by MicrobeTracker (51) from phase contrast images.

Analysis of diffusive behavior

Details of mean-square displacement plots, trajectory simulations, two-state modeling of $P(r)$ and $P(\langle r \rangle_{6\text{-step}})$ distributions, and estimation of uncertainties in fitting parameters are provided in Appendix.

References

1. Doerfel LK, Rodnina MV. 2013. Elongation factor P: Function and effects on bacterial fitness. Biopolymers 99:837-45.
2. Johansson M, Ieong K-W, Trobro S, Strazewski P, Åqvist J, Pavlov MY, Ehrenberg M. 2011. pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. Proc Natl Acad Sci U S A 108:79-84.

3. Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC. 2009. Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc Natl Acad Sci U S A 106:50-54.

4. Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV. 2013. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science 339:85-8.

5. Tanner DR, Cariello DA, Woolstenhulme CJ, Broadbent MA, Buskirk AR. 2009. Genetic Identification of Nascent Peptides That Induce Ribosome Stalling. J. Biol. Chem. 284:34809-34818.

6. Peil L, Starosta AL, Lassak J, Atkinson GC, Virumae K, Spitzer M, Tenson T, Jung K, Remme J, Wilson DN. 2013. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. Proc Natl Acad Sci U S A 110:15265-70.

7. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR. 2015. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep 11:13-21.

8. Elgamal S, Katz A, Hersch SJ, Newsom D, White P, Navarre WW, Ibba M. 2014. EF-P Dependent Pauses Integrate Proximal and Distal Signals during Translation. PLOS Genetics 10:e1004553.

9. Hanawa-Suetsugu K, Sekine S, Sakai H, Hori-Takemoto C, Terada T, Unzai S, Tame JR, Kuramitsu S, Shirouzu M, Yokoyama S. 2004. Crystal structure of elongation factor P from *Thermus thermophilus* HB8. Proc Natl Acad Sci U S A 101:9595-600.

10. Choi S, Choe J. 2011. Crystal structure of elongation factor P from *Pseudomonas aeruginosa* at 1.75 å resolution. Proteins: Structure, Function, and Bioinformatics 79:1688-1693.

11. Aoki H, Xu J, Emili A, Chosay JG, Golshani A, Ganoza MC. 2008. Interactions of elongation factor EF-P with the *Escherichia coli* ribosome. FEBS J 275:671-81.

12. Blaha G, Stanley RE, Steitz TA. 2009. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. Science 325:966-70.

13. Park J-H, Johansson HE, Aoki H, Huang BX, Kim H-Y, Ganoza MC, Park MH. 2012. Post-translational Modification by β-Lysylation Is Required for Activity of *Escherichia coli* Elongation Factor P (EF-P). J. Biol. Chem. 287:2579-2590.

14. Keiler KC. 2015. Mechanisms of ribosome rescue in bacteria. Nat Rev Microbiol 13:285-97.

15. Zou SB, Hersch SJ, Roy H, Wiggers JB, Leung AS, Buranyi S, Xie JL, Dare K, Ibba M, Navarre WW. 2012. Loss of elongation factor P disrupts bacterial outer membrane integrity. J Bacteriol 194:413-25.

16. Hersch SJ, Wang M, Zou SB, Moon KM, Foster LJ, Ibba M, Navarre WW. 2013. Divergent protein motifs direct elongation factor P-mediated translational regulation in *Salmonella enterica* and *Escherichia coli*. mBio 4:e00180-13.

17. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642-5.

18. Hess ST, Girirajan TPK, Mason MD. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys J. 91:4258-4272.

19. Rust MJ, Bates M, Zhuang X. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods 3:793-5.

20. Adio S, Senyushkina T, Peske F, Fischer N, Wintermeyer W, Rodnina MV. 2015. Fluctuations between multiple EF-G-induced chimeric tRNA states during translocation on the ribosome. Nat Comm 6:7442.

21. Chen J, Petrov A, Tsai A, O'Leary SE, Puglisi JD. 2013. Coordinated conformational and compositional dynamics drive ribosome translocation. Nat Struct Mol Biol 20:718-727.

22. Wasserman MR, Alejo JL, Altman RB, Blanchard SC. 2016. Multi-perspective smFRET reveals rate-determining late intermediates of ribosomal translocation. Nat Struct Mol Biol 23:333-341.

23. Chen C, Stevens B, Kaur J, Smilansky Z, Cooperman BS, Goldman YE. 2011. Allosteric vs. spontaneous exit-site (E-site) tRNA dissociation early in protein synthesis. Proc Natl Acad Sci U S A 108:16980-16985.

24. Dinos G, Kalpaxis DL, Wilson DN, Nierhaus KH. 2005. Deacylated tRNA is released from the E site upon A site occupation but before GTP is hydrolyzed by EF-Tu. Nucleic Acids Research 33:5291-5296.

25. Wilson DN, Nierhaus KH. 2006. The E-site story: the importance of maintaining two tRNAs on the ribosome during protein synthesis. Cell Mol Life Sci 63:2725-2737.

26. McKinney SA, Murphy CS, Hazelwood KL, Davidson MW, Looger LL. 2009. A bright and photostable photoconvertible fluorescent protein. Nat Methods 6:131-133.

27. Kumar M, Mommer MS, Sourjik V. 2010. Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in *Escherichia coli*. Biophys J. 98:552-559.

28. English BP, Hauryliuk V, Sanamrad A, Tankov S, Dekker NH, Elf J. 2011. Single-molecule investigations of the stringent response machinery in living bacterial cells. Proc Natl Acad Sci U S A 108:E365-E373.

29. Plochowietz A, Farrell I, Smilansky Z, Cooperman BS, Kapanidis AN. 2016. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria. Nucleic Acids Research 45(2): 926-937.

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

31. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, Lippincott-Schwartz J. 2008. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat Methods 5:155-157.

32. Bakshi S, Choi H, Mondal J, Weisshaar JC. 2014. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. Mol Microbiol 94:871-887.

33. Li W, Bouveret E, Zhang Y, Liu K, Wang JD, Weisshaar JC. 2016. Effects of amino acid starvation on RelA diffusive behavior in live *Escherichia coli*. Mol Microbiol 99:571-85.

34. Wolfe AD, Hahn FE. 1965. Mode of action of chloramphenicol IX. Effects of chloramphenicol upon a ribosomal amino acid polymerization system and its binding to bacterial ribosome. Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis 95:146-155.

35. Hansen JL, Moore PB, Steitz TA. 2003. Structures of Five Antibiotics Bound at the Peptidyl Transferase Center of the Large Ribosomal Subunit. Journal of Molecular Biology 330:1061-1075.

36. Sohmen D, Harms JM, Schlünzen F, Wilson DN. 2009. SnapShot: Antibiotic Inhibition of Protein Synthesis I. Cell 138:1248-1248.e1.

37. Bakshi S, Choi H, Mondal J, Weisshaar JC. 2014. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *E. coli* chromosome and ribosomes. Mol Microbiol 94:871-887.

38. Sanamrad A, Persson F, Lundius EG, Fange D, Gynnå AH, Elf J. 2014. Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. Proc Natl Acad Sci U S A 111:11413-11418.

39. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press.

40. Das R, Cairo CW, Coombs D. 2009. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. PLOS Comput Biol 5:e1000556.

41. Harvey RJ, Koch AL. 1980. How partially inhibitory concentrations of chloramphenicol affect the growth of *Escherichia coli*. Antimicrobial Agents and Chemotherapy 18:323-337.

42. An G, Glick BR, Friesen JD, Ganoza MC. 1980. Identification and quantitation of elongation factor EF-P in *Escherichia coli* cell-free extracts. Canadian Journal of Biochemistry 58:1312-1314.

43. Schmidt A, Kochanowski K, Vedelaar S, Ahrne E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M. 2016. The quantitative and condition-dependent *Escherichia coli* proteome. Nat Biotech 34:104-110.

44. Bremer H, Dennis P. 2008. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal Plus

45. Chai Q, Singh B, Peisker K, Metzendorf N, Ge X, Dasgupta S, Sanyal S. 2014. Organization of Ribosomes and Nucleoids in *Escherichia coli* Cells during Growth and in Quiescence. J. Biol. Chem. 289:11342-11352.

46. Datta S, Costantino N, Court DL. 2006. A set of recombineering plasmids for gram-negative bacteria. Gene 379:109-115.

47. Ederth J, Mandava CS, Dasgupta S, Sanyal S. 2009. A single-step method for purification of active His-tagged ribosomes from a genetically engineered *Escherichia coli*. Nucleic Acids Research 37:e15-e15.

48. Neidhardt FC, Bloch PL, Smith DF. 1974. Culture medium for enterobacteria. J Bacteriol 119:736-47.

49. Bakshi S, Bratton BP, Weisshaar JC. 2011. Subdiffraction-limit study of Kaede diffusion and spatial distribution in live *Escherichia coli*. Biophys J 101:2535-44.

50. Crocker JC, Grier DG. 1996. Methods of digital video microscopy for colloidal studies. Journal of Colloid and Interface Science 179:298-310.

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

Appendix

Table A4.1 *E. coli* strains and oligonucleotides used in this work

Table A4.4A Oligonucleotides used in construction of strains described in the main text

Oligonucleotide	**Sequence 5′- 3′**
meos_F	CGC GGATCC ATGTCTGCTATTAAACCGG
meos_R	CCG GAATCC TTAGCGACGAGCATTATCCG
efpmeos_F	AGTGGATACCCGCTCTGGTGAATACGTCTCTCGCGTGAAGTCT GCTATTAAACCGGATATGAAGA
efpmeos_R	ATAAGTGATGGTGCAGCCTGCAGGCCGCACCACAACCGCATT AGAAAAACTCATCGAGCATC
EM1	GATCGGATCCATTTCAGAGGGCCTTATGGC
EM2	GATCAAGCTTTTAGCGACGAGCATTATCCGG
Mut1	AAAACCGGGT**GCA**GGCCAG
Mut2	ACGAATTCACTCGCTTCAACCG

Table A4.1B Strains used for imaging and doubling times at 30°C in EZ rich, defined medium (EZRDM).

Strain name	**Description**	**Doubling time (min)**
SM1	EF-P–mEos2 (chromosome)	60 ± 3
SM4	EF-P–mEos2 (plasmid)	51± 1
SM8	EF-P^{K34A} –mEos2 (plasmid)	50 ± 6
MSG196	Ribosome S2–mEos2 (chromosome)	51± 3
VH1000	Wild type	45 ± 1(1)

A4.2 Axial distributions under drug treatments

Figures 4.1*A* and 4.1*B* show 2D heat maps and 1D axial projections of the spatial distributions of EF-P–mEos2 and ribosomes (30S–mEos2 labeling) following treatment with chloramphenicol (Cam). Figures A4.1*C* and A4.1*D* show analogous distributions following treatment with rifampicin (Rif).

A4.3 Analysis of diffusive behavior

Mean-square displacement plots MSD(τ)

The MSD as a function of lag time τ is given by $MSD(\tau) = < (\boldsymbol{r}(t+\tau) - \boldsymbol{r}(t))^2 >$, where $\boldsymbol{r}(t)$ is the two-dimensional location of the particle at time t, τ is the lag time, and the average is taken over all times t and over many trajectories. MSD(τ) plots for 30S-labeled ribosomes and wild-type EF-P in normal growth conditions are shown in Fig. 4.4*A*. Plots for EF-P^{K34A} in normal growth and for wild-type EF-P after Rif treatment are shown in Fig. 4.4*B*.

The slope of the first two points of an MSD(τ) plot provides a first estimate of the true mean diffusion coefficient: D = slope/4. The MSD slope takes account of the dynamic localization error σ, but does not account for confinement effects. Even for 6 x 2 ms = 12 ms long trajectories, for rapidly diffusing species with $D \sim 5$ μm^2-s^{-1} confinement restricts diffusive trajectories and causes curvature of the MSD plot. This makes the diffusion coefficient from the two-point slope a lower bound on the true D. The trajectory analysis presented below is more accurate.

Estimation of dynamic localization error σ_{fast} and σ_{slow} for fast and slow EF-P molecules

Localization error in the single-step $P(r)$ distributions arises from the point-spread function (PSF) of the microscope, the finite number of photons detected per camera frame, and blurring of the images due to diffusive motion during each 2 ms camera frame (2, 3). While ribosomes and

ribosome-bound EF-P copies experience little diffusive blurring, the rapidly moving free EF-P copies are blurred substantially. We reasoned that model Monte Carlo diffusive trajectories should involve larger "dynamic localization error" for fast EF-P copies than for slow EF-P copies.

This problem has been addressed in detail by Michalet (2). Suppose the best fit to the first two experimental points of a two dimensional mean-square displacement plot is given by the equation $MSD(\tau) = a + b\tau$, with b the slope and a the extrapolated intercept at lag time $\tau = 0$. Then the best estimate of the diffusion coefficient is $D = b/4$ and the best estimate of the dynamic localization error is $\sigma = \frac{1}{2}\,(a + 4Dt_E/3)^{1/2}$, where t_E is the exposure time per camera frame. From the MSD plots for the total populations of EF-P and of ribosomes (Fig. 4.4*A*), we estimate that the diffusion coefficient of the free EF-P copies should be about 4 $\mu m^2\text{-}s^{-1}$ and that of the ribosome-bound copies should be about 0.5 $\mu m^2\text{-}s^{-1}$. If so, the root-mean-square displacement in two dimensions of a fast EF-P copy during the 2-ms camera frame is ~200 nm, substantially larger than typical static localization errors in live *E. coli* single-molecule studies of slowly moving species labeled with fluorescent proteins. The rms displacement of a slow EF-P copy is only ~60 nm, comparable to typical static localization errors.

To estimate the dynamic localization errors σ_{fast} and σ_{slow} for EF-P, we used six-step trajectories to form the distribution of the mean of six one-step estimates of D:

$< D >_{6-step} = \frac{1}{24\tau}\sum_{i=1}^{6}\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$, where (x_{i+1}, y_{i+1}) and (x_i, y_i) are the coordinates of final and initial positions of each step. This distribution is shown in Fig. 4.5*A*. Separate MSD plots for the slowest 10% of trajectories and the fastest 10% of trajectories are shown in Fig. A4.5*B*. The slope of the linear fit to the first two data points gives nominal slow and fast diffusion coefficients of 0.7 and 8.6 $\mu m^2\text{-}s^{-1}$, respectively. These results are combined

with the intercepts to give the estimates $\sigma_{slow} \sim 40$ nm and $\sigma_{fast} \sim 86$ nm. Because the analysis used the slowest and fastest copies, we chose $\sigma_{slow} = 50$ nm and $\sigma_{fast} = 75$ nm when simulating the diffusive trajectories used to fit $P_{EF\text{-}P}(r)$ in Fig. 4.3*A*. A similar method was used for the ribosome trajectories. It yielded the estimates $\sigma_{slow} = 40$ nm and $\sigma_{fast} = 75$ nm, which were used in fitting $P_{ribo}(r)$ in Fig. 4.3*B*.

A4.4 Monte Carlo simulations of diffusive trajectories

Many previous studies of single-molecule diffusion fit the experimental $P(r)$ distribution to a sum of analytical functions, with each component describing diffusion of the species in free space (4-7). However, molecules diffusing rapidly in the *E. coli* cytoplasm suffer from confinement due to the spherocylindrical cell boundaries, a problem for which there is no analytical solution. Therefore, we simulated the behavior of each diffusive component from a large number of random walk trajectories that incorporate the dynamic localization error σ and confinement effects within a model spherocylinder that mimics the dimensions of a typical *E. coli* cell in our growth conditions (tip-to-tip cell length = 4 μm and cell diameter = 0.9 μm). Each set of simulations models one diffusive state, with values of D and σ fixed. At $t = 0$, 15,000 particles were randomly distributed within the cell volume. Each particle undergoes a random walk independent of other particle positions. To model each 2-ms camera image, three-dimensional microtrajectories (1000 steps of 2 μs each) were generated. At each time step, each particle chooses a displacement in each of three Cartesian directions. These displacements are chosen from a Gaussian distribution whose standard deviation corresponds to the state's three-dimensional diffusion coefficient *D.* In the rare event that a particle attempts to step outside of the cell boundaries, the displacement for that microstep is taken to be zero. The location of each particle during each camera frame is obtained as the centroid of the model microtrajectories in

order to mimic the analysis procedure used for the experimental images. The appropriate dynamic localization error σ was then applied to each centroid location in both *x* and *y* coordinates by sampling a Gaussian distribution with standard deviation σ. By adding the error to the centroid position we obtain the model "measured" location for each 2 ms camera frame. The *x* and *y* coordinates of each measured location are stored for further analysis. For the next model camera frame, each particle continues to make microsteps in 3D starting from the endpoint of the previous camera frame. Model trajectories for EF-P and for ribosomes use the appropriate estimated value of σ_{slow} or σ_{fast}, determined as described above.

By connecting the sequence of "measured" simulation locations over seven frames, we form 15,000 model trajectories for each relevant value of *D*. These trajectories are used to compute model-based, numerical one-step probability distributions $P_{model}(r;D)$ that are the model functions for the least-squares analysis of the corresponding experimental distributions.

A4.5 Fitting of ribosomal distribution $P_{ribo}(r)$ to a static, two-state model

As shown previously (5), single-molecule methods with the 30S ribosomal subunit labeled can resolve two diffusive states of ribosome in live *E. coli*: slowly diffusing, translating 70S ribosomes (likely present as polysomes) and more rapidly diffusing, free 30S subunits. Ideally, if the transition time scale between these two states is much longer than the experimental observation time scale of single camera frames (2 ms), the magnitude of displacements for individual trajectory steps should indicate the diffusive state of the molecule. Nearly 15,000 experimental trajectories of 30S-labeled ribosomes that lasted for 6 steps or longer were selected for analysis. Longer trajectories were truncated to 6 steps. The 6-step trajectories were then sliced into individual steps. The single-step displacements

$r_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$ were pooled to form the distribution $P_{ribo}(r)$ in Fig. 4.3*B*.

We judge the goodness of fit by evaluating the reduced chi-square statistic: $\chi_\nu^2 = \frac{1}{(N-\alpha)} \sum_{j=1}^{N} \frac{(h_j - y_j)^2}{s_j^2}$. Here j labels the N bins in the (unnormalized) $P(r)$ and $P_{model}(r)$ histograms, h_j is the number of experimental counts in bin j, y_j is the number of counts in bin j of the simulated $P_{model}(D_i)$ histogram, s_j^2 is the variance of the value in bin j, and α is the number of fitted parameters. We take $s_j^2 = h_j$ as the estimate of the variance, assuming Poisson statistics. A good fit to an adequately flexible *analytical* model function should have $\chi_\nu^2 \sim 1$. Since our model functions are numerical, there is noise in both the experimental and simulated histograms. In some cases, the number of experimental and model input trajectories are comparable, in which case a good fit could have $\chi_\nu^2 \sim 2$.

As judged by the reduced chi-square statistic χ_ν^2, the distribution $P_{ribo}(r)$ is poorly fit by a one-state model, *i.e.*, to a single numerical function $P_{model}(r;D)$. Next we modeled $P_{ribo}(r)$ as a sum of two static (non-exchanging) populations: $P_{model}(r) = f_{slow}P(r;D_{slow}) + (1 - f_{slow})P(r;D_{fast})$. The appropriate values of σ_{fast} and σ_{slow} were used for each component. The least-squares fitting procedure involved a numerical search for the lowest value of χ_ν^2 on a 3D grid of combinations of the three independent adjustable parameters (D_{fast}, D_{slow}, f_{slow}). The best fit to this model function yielded $f_{slow} = (1 - f_{fast}) = 0.65$, $D_{slow} = 0.2$ μm^2/s (presumably the translating 70S ribosomes), $f_{fast} = 0.35$, and $D_{fast} = 0.8$ μm^2/s (presumably the free 30S subunits). The minimum goodness-of-fit parameter was $\chi_\nu^2 = 2.3$. Due to the large number of ribosome trajectories (15,000), we used the same number of model trajectories in the simulation. Both experimental data and the numerical model function have noise. The noise in the model function contributes to the value of χ_ν^2, so a value near one is not expected. The best fit is qualitatively good, as seen in Fig. 4.3*B*. Additionally, the best-fit combination of parameters gives the weighted average

$\langle D_{ribo} \rangle = f_{slow}D_{slow} + f_{fast}D_{fast} = 0.41\ \mu m^2\text{-}s^{-1}$. This agrees reasonably well with the mean value $\langle D_{ribo} \rangle = 0.49\ \mu m^2\text{-}s^{-1}$ from the initial slope of the MSD plot (Fig. A4.4*A*). We conclude that the static, two-state model is reasonable for fitting the $P_{ribo}(r)$ distribution.

For comparison, we also simulated $P_{ribo}(r)$ using a static, two-state model in which both the fast and slow constituent populations have the same $\sigma = 57$ nm, as estimated from the MSD including all ribosome trajectories. The resulting χ_v^2 were always very large (> 35).

A4.6 Fitting of EF-P distribution $P_{EF\text{-}P}(r)$ to a static, two-state model

The distribution of one-step displacements $P_{EF\text{-}P}(r)$ in Fig. 4.4*A* was generated from 1000 six-step trajectories for 2-ms camera frames. We first attempted to fit the distribution to a single diffusive state using 15,000 simulated EF-P trajectories of varying D and $\sigma = 61$ nm, obtained from the intercept of MSD(τ) plots for all EF-P trajectories (Fig. A4.4*A*). However, the best one-state fit had $D = 3.1\ \mu m^2\text{-}s^{-1}$ and an unacceptably large value $\chi_v^2 = 5.0$.

Then we tested two-state models of EF-P with no exchange between states on the 2-ms timescale. The two states were a slowly diffusing, ribosome-bound state (D_{slow}) and a rapidly diffusing, free state (D_{fast}), with population fractions f_{slow} and f_{fast}. Appropriate values of σ_{slow} and σ_{fast} were used, as described above. Trial and error showed that the large measurement error precludes accurate determination of D_{slow}. In most of the fitting effort, we therefore constrained $D_{slow} = 0.2\ \mu m^2\text{-}s^{-1}$ to match that of the best-fit slow component of $P_{ribo}(r)$ and optimized f_{slow} and D_{fast}. For each combination of input parameters selected from a 2D grid of values, we generated some 15,000 individual simulated trajectories, yielding a numerical model function $P_{model}(r) = f_{slow}P(r;D_{slow}) + f_{fast}P(r;D_{fast})$. The best fit was obtained for $f_{slow} = 0.30$, $D_{fast} = 4.3\ \mu m^2\text{-}s^{-1}$, and $f_{fast} = 0.70$. The corresponding goodness-of-fit parameter was $\chi_v^2 = 1.0$, indicating a good fit. We are now using 15,000 model trajectories to generate each model

function, while the experimental $P_{EF\text{-}P}(r)$ distribution is derived from only 1000 trajectories. The noise in the data is substantially larger than that in the numerical model function, so a value of χ_v^2 near one should be expected if the model function is adequate. This fit is compared with the data in Fig. 4.4*A*. We identify the slowly diffusing EF-P fraction as the copies bound to translating 70S ribosomes and the rapidly diffusing fraction as free EF-P, consistent with our original estimate of the diffusion coefficient for free copies. These parameters give a mean diffusion coefficient of $<D> = 3.1\ \mu m^2\text{-}s^{-1}$, quite similar to $<D> = 3.4\ \mu m^2\text{-}s^{-1}$ obtained from the MSD(τ) plot for all EF-P trajectories (Fig. 4.4*A*).

Upon optimizing D_{slow}, f_{slow}, D_{fast} and f_{fast} without any constraints on D_{slow}, the best fit was obtained for $D_{slow} = 0.5\ \mu m^2\text{-}s^{-1}$, $f_{slow} = 0.35$, $D_{fast} = 4.8\ \mu m^2\text{-}s^{-1}$, and $f_{fast} = (1 - f_{slow}) = 0.65$ with $\chi_v^2 = 0.78$ (slightly better than the best constrained fit). Thus the fractions and diffusion coefficients obtained from an unconstrained fit are quite similar to those from the constrained fit using $D_{slow} = 0.2\ \mu m^2\text{-}s^{-1}$.

To estimate the uncertainty in the model parameters, we examined the 3D grid of χ_v^2 values generated from the unconstrained fits that varied all three parameters D_{fast}, f_{slow}, and D_{slow}. Three two-dimensional planes passing through the values $D_{fast} = 4.3\ \mu m^2\text{-}s^{-1}$, $f_{slow} = 0.30$, and $D_{slow} = 0.2\ \mu m^2\text{-}s^{-1}$ are shown in Fig. 4.7. We judged the fits to be qualitatively poor whenever the value of χ_v^2 exceeded 1.5, which is 0.5 units larger than the value at our preferred parameter combination. These regions are surrounded by the blue boundaries in the figure. Our best estimates with error bars are $f_{slow} = 0.30 \pm 0.10$, $f_{fast} = 0.70 \pm 0.10$, and $D_{fast} = 4.3 \pm 1.0\ \mu m^2\text{-}s^{-1}$. The superimposed heat map colors shows that f_{slow} and D_{fast} are positively correlated, f_{slow} and D_{slow} are positively correlated, and D_{slow} and D_{fast} are essentially uncorrelated. Again, the value of

D_{slow} is not well determined due to the measurement error; reasonable fits can be obtained for D_{slow} in the range 0 -1 μm^2-s^{-1}.

Analogous fitting procedures were carried out for the $P(r)$ data for EF-P^{K34A} expressed from a plasmid and for wild-type EF-P after drug treatment with Cam and Rif. A summary of the fitting results is provided in Table A4.2.

Table A4.2 Summary of range of fitting searches for $P(r)$ for different species and imaging conditions.

Imaging conditions	**Range of D_{slow} (μm^2/s)**	**Range of D_{fast} (μm^2/s)**	**Best D_{slow} (μm^2/s)**	**Best D_{fast} (μm^2/s)**	**Best f_{slow}**
Ribosome	0-0.9 (σ_{slow} = 40 nm)	0-1 (σ_{fast} = 75 nm)	0.2	0.8	0.65
EF-P	Constrained to 0.2 (σ_{slow} = 50 nm)	1-15 (σ_{fast} = 75 nm)	Constrained to 0.2	4.3	0.30
EF-P under Cam	Constrained to 0.2 (σ_{slow} = 50 nm)	1-15 (σ_{fast} = 90 nm)	Constrained to 0.2	1.2	0.45
EF-P under Rif	1-10 (σ_{slow} = 125 nm)	1-10 (σ_{fast} = 150 nm)	4.6	8	0.55
EF-P^{K34A}	1-10 (σ_{slow} = 50 nm)	1-15 (σ_{fast} = 90 nm)	3.2	9.7	0.65

A4.7 Dynamics of EF-P association with ribosomes

The fit to the single-step $P_{EF\text{-}P}(r)$ distribution assumed two static populations that persist on the 2-ms timescale of a single camera frame. Next we asked whether six-step trajectories (12 ms total duration) can provide information about the timescale of possible transitions between the slow, ribosome-associated state and the fast, freely diffusing state of EF-P. Accordingly, for 859 six-step trajectories of wild type EF-P–mEos2, we generated the distribution $P_{EF\text{-}P}(\langle r \rangle_{6\text{-step}})$, the mean displacement of the six steps: $< r >_{6-step} = \frac{1}{6}\sum_{i=1}^{6}\sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2}$). This is shown in Fig. 4.5. Longer trajectories are truncated at six steps. Comparing Figs. 4.4*A* and 4.5,

we see that $P_{EF\text{-}P}(\langle r \rangle_{6\text{-step}})$ is substantially narrower than $P_{EF\text{-}P}(r)$ due to the averaging of six displacements, which decreases the relative contribution of measurement error.

We simulated $P_{model}(\langle r \rangle_{6\text{-step}})$ distributions for static populations by averaging the displacements from six-step simulated trajectories that included confinement and measurement error. If the EF-P molecules undergo transitions only rarely during a 12-ms trajectory, then the experimental $P(\langle r \rangle_{6\text{-step}})$ should match the $P_{model}(\langle r \rangle_{6\text{-step}})$ using the same input parameters that gave the best-fit $P_{model}(r)$ distribution. We simulated 15,000 six-step trajectories, 30% with D_{slow} = 0.2 μm^2/s (4500 trajectories) and 70% with D_{fast} = 4.3 μm^2/s (10,500 trajectories), in accord with the parameters of the best-fit model of the $P_{EF\text{-}P}(r)$ distribution of Fig. 4.4*A*. $P_{model}(\langle r \rangle_{6\text{-step}})$ for these static populations exhibits two partially resolved peaks (Fig. 4.5), unlike the experimental distribution. This suggests that the wild type EF-P indeed undergoes transitions during the six-step, 12-ms trajectories.

In order to estimate the lifetime of the two diffusive states, we introduced two-state, binding/unbinding kinetics into our simulations, following Das *et al* (8, 9). Particles initially occupy either of the two states randomly in space in proportion to the population fractions dictated by their lifetimes. When a particle is in a free or bound state, it diffuses with a mean diffusion coefficient of D_{fast} or D_{slow}, respectively. It is assumed that the transitions occur instantaneously between two 2-μs long microsteps. The parameters f_{slow}, D_{slow}, f_{fast}, and D_{fast} were fixed at their preferred values. There are two lifetimes in the problem, τ_{free} (mean search time for EF-P in the freely diffusing state) and τ_{bound} (mean lifetime of EF-P while ribosome-bound). Their ratio τ_{free}/τ_{bound} = 7/3 is fixed by the ratio of best-fit population fractions, so there is only one adjustable parameter.

We generated large sets of six-step simulated trajectories with D_{slow} = 0.2 μm^2/s, D_{fast} = 4.3 μm^2/s and τ_{free}/τ_{bound} = 7/3. While keeping τ_{free}/τ_{bound} constant, we varied τ_{free} from 0.1 ms to 100, calculating 15,000 trajectories in each case. The best fit to the experimental $P_{EF\text{-}P}(\langle r \rangle_{6\text{-step}})$ was obtained for τ_{free} = 16 ms and τ_{bound} = 7 ms, with χ_v^2 = 1.4. This fit is compared with experiment in Fig. 4.5. Additional calculated distributions $P_{model}(\langle r \rangle_{6\text{-step}})$ for different assumed lifetimes are compared with experiment in Fig. A4.6. Based on the range of τ_{free} that yields a value of χ_v^2 less than 2.0 (0.6 units larger than optimal), we take τ_{free} = 16 ± 5 ms and τ_{bound} = 7 ± 3 ms as our best estimates of the average lifetime of free EF-P and of EF-P bound to ribosomes, respectively. For χ_v^2 = 2.0 or larger, the fits are qualitatively poorer (Fig. A4.6). Our best estimate of the typical timescale for an EF-P/ribosome binding/unbinding cycle becomes (τ_{bound} + τ_{free}) ~ (23 ± 8) ms.

A4.8 Robustness of fitting procedure

We simulated numerical $P(r)$ distributions from model trajectories using known input values of the two-state model parameters f_{slow}, D_{slow}, f_{fast}, and D_{fast} and then tested how well our procedures recover the correct values. This tests the robustness of the procedure for various parameter combinations under the assumption that a two-state model without transitions on a 2-ms timescale is correct. For example, to mimic the experimental $P_{EF\text{-}P}(r)$, we simulated 300 trajectories with D_{slow} = 0.2 μm^2/s and σ_{slow} = 50 nm and 700 trajectories with D_{fast} = 4.3 μm^2-s^{-1} and σ_{fast} = 75 nm and combined them to generate a normalized $P_{model}(r)$ to mimic the experimental $P_{EF\text{-}P}(r)$. By constraining the model data and experimental data to include the same number of trajectories, they should also have similar noise levels.

Each synthetic model data set is then subjected to the same fitting procedure we used for the experimental data. We repeated the numerical experiment 20 times, producing 20 best fits to the 20 synthetic model data sets. For each fitting parameter, the 20 best-fit are used to produce a mean

and standard deviation. The same procedure was carried out for all the cases reported in our study. The table below depicts the reproducibility of the fits for each case. Given that a two-state system without exchange on the 2-ms timescale produced the data, the procedure recovers all parameters to good precision (±10% or better) with the exception of the slowest diffusion coefficients (±50 - 100%).

A4.9 Closing remarks on data analysis

It should be clear from the two-state decomposition of the single-step $P_{EF\text{-}P}(r)$ distribution (Fig. 4.3) that the present study approaches the limit of time-resolution of single-molecule tracking methods using fluorescent protein labels. The short frame time of 2 ms is dictated by the need to capture both fast and slow molecules in the same movie. However, the significant measurement error associated with such short observation times blends the apparent fast motion of free EF-P and the apparent slow motion of ribosome-bound EF-P together. They can barely be deconvolved, and the slow diffusion coefficient is only poorly determined. In contrast, the fast and slow fractions are returned with reasonable accuracy, which is important for the main biological conclusions of the Discussion. Finally, it was fortuitous that the bound ↔ free transition times fall within the range that can be unraveled in the short, 6 x 2 ms = 12 ms trajectories that are available experimentally (Fig. 4.5).

A4.10 Supporting Figures and Captions

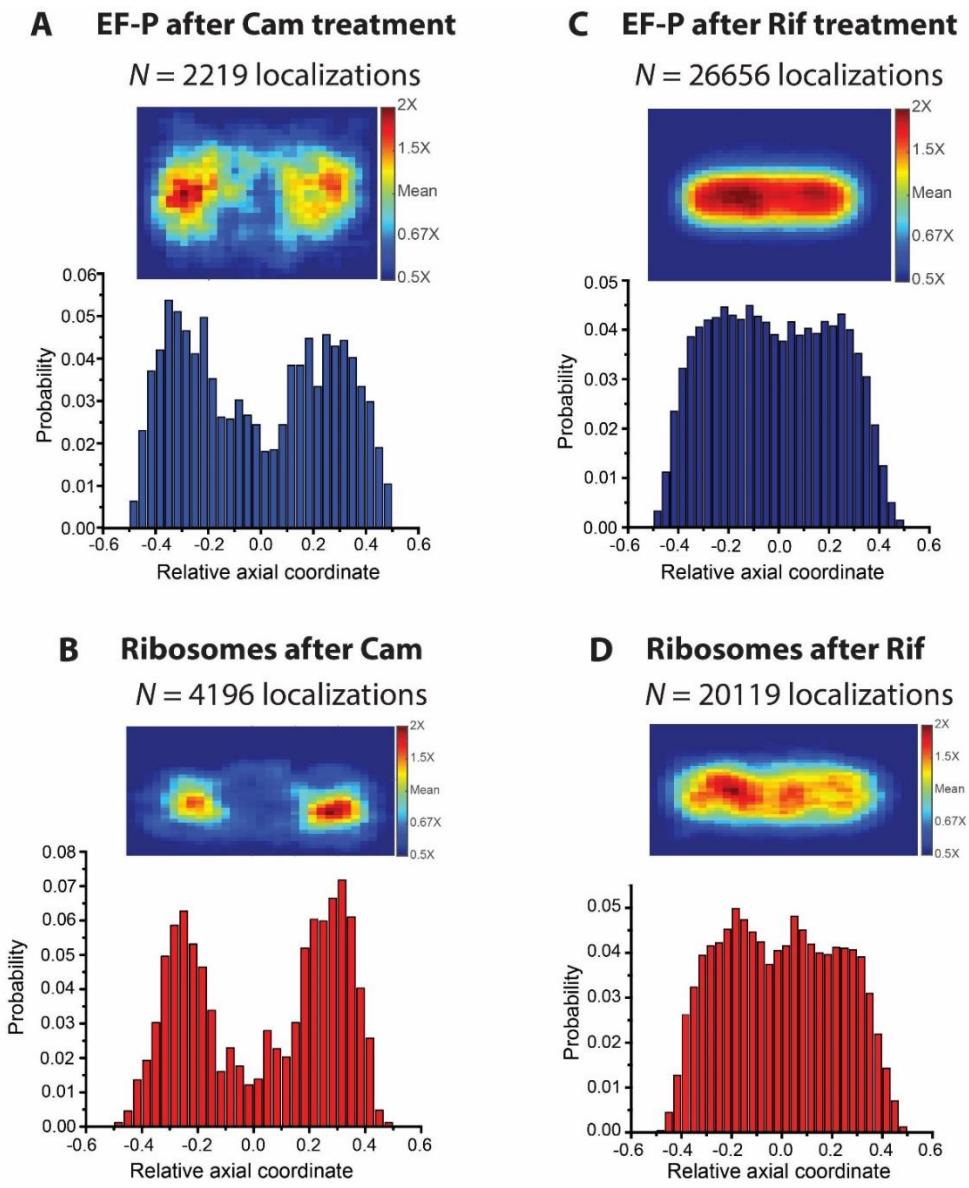


Fig. A4.1 (A) *Top:* Localization probability density heat map of 2220 EF-P–mEos2 copies imaged at 2 ms/frame in different cells of length 2.5–3.5 μm after 30 min of chloramphenicol treatment. Each location is placed on a common scale of relative axial position. Only molecules that lasted at least 7 frames contribute to the distribution. *Bottom:* Distribution of axial projections on the same relative scale. (**B**) Same as panel A, but for ribosomes imaged after Cam treatment. (**C**) Same as panel A, but for EF-P–mEos2 molecules imaged in cells of length 3.5–4.5 μm after 3 hr of rifampicin treatment. (**D**) Same as panel C, but for ribosomes imaged after Rif treatment.

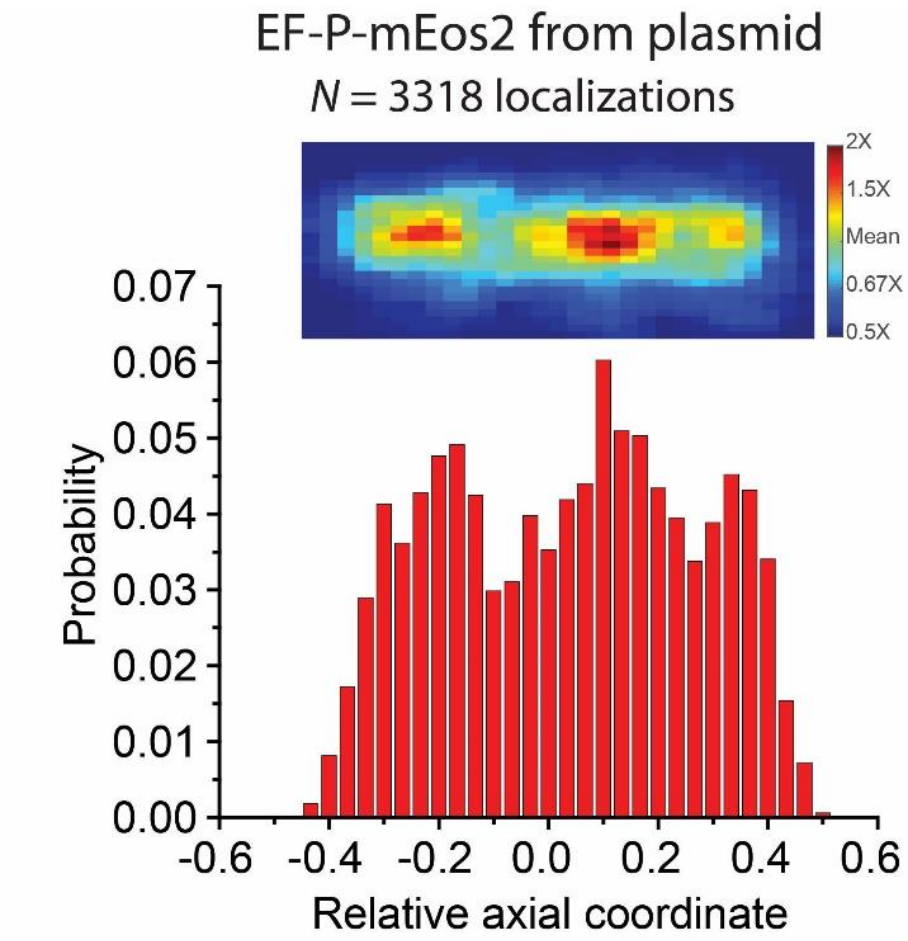


Fig. A4.2 Localization probability density maps of EF-P–mEos2 expressed from plasmid in normal growth conditions. A composite of images taken at 2 ms/frame from cells with lengths between 3.5–4.5 μm. Only molecules that lasted at least 7 frames contribute to the axial distribution. The axial distribution shows a three peak distribution similar to that of ribosomes under same imaging conditions.

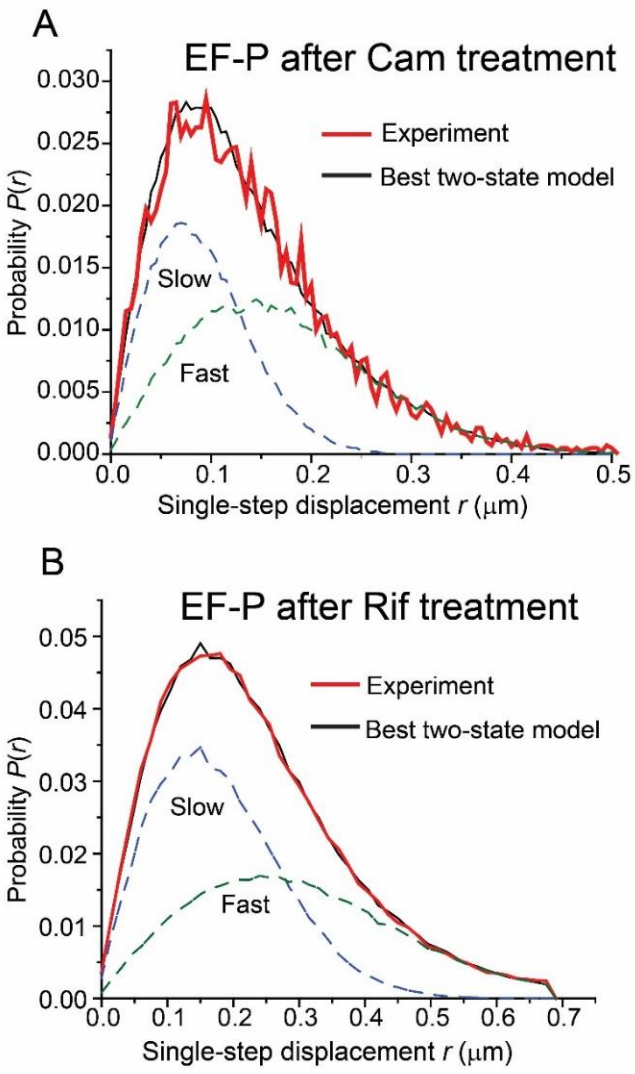


Fig. A4.3(**A**) *Red:* Experimental probability distribution of single-step displacements taken by EF-P–mEos2 molecules in 2 ms after 30 min of chloramphenicol treatment. *Black:* The best fit to a static two-state model (without transitions) with D_{slow} constrained. Model parameters: D_{slow} = 0.2 μm^2/s (σ_{slow} = 50 nm), f_{slow} = 0.45, D_{fast} = 1.2 μm^2/s (σ_{fast} = 90 nm), f_{fast} = 0.55, with χ_v^2 = 1.2. The slow and fast components are shown as dashed lines as labeled. (**B**) *Red:* Experimental probability distribution of single-step displacements taken by EF-P–mEos2 molecules in 2 ms after 3 hr of rifampicin treatment. *Black:* The best unconstrained two-state model fit. Model parameters: D_{slow} = 4.6 μm^2/s (σ_{slow} = 75 nm), f_{slow} = 0.55, D_{fast} = 8 μm^2/s (σ_{fast} = 150 nm), f_{fast} = 0.45, with χ_v^2 = 1.5. The individual slow and fast components are shown in dashed lines of blue and green respectively.

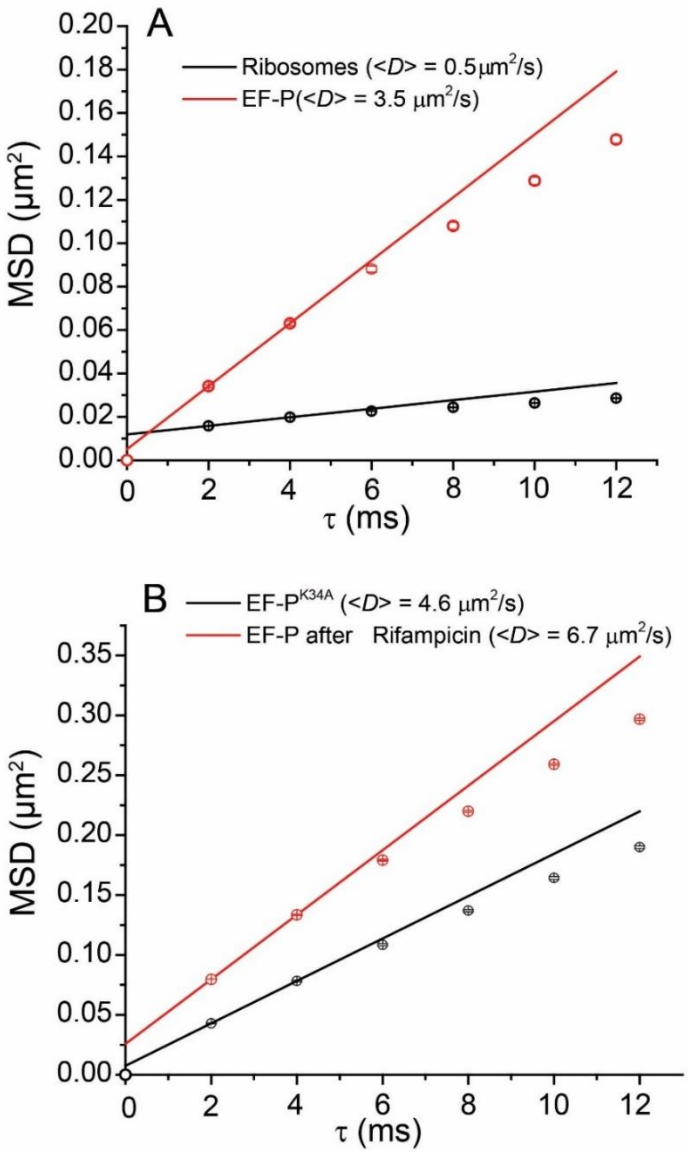


Fig. A4.4 (**A**) Mean square displacement plot, MSD(τ), for EF-P–mEos2 (red circles) and for ribosomes (30S–mEos2 labeling, black circles). Trajectories are truncated to six steps; error estimates are ± 1 σ of the MSD values. Lines are drawn through first two data, yielding diffusion coefficient estimates of 3.5 $\mu m^2/s$ and 0.5 $\mu m^2/s$, respectively. (**B**) Mean square displacement plot, MSD(τ), for EF-P–mEos2 after 3 hr of treatment with Rif (red circles) and the mutant form $EF\text{-}P^{K34A}$–mEos2 in normal growth conditions (black circles). Trajectories are truncated to six steps; error estimates are ± 1 σ of the MSD values. Lines are drawn through first two data, yielding diffusion coefficient estimates of 6.7 $\mu m^2/s$ and 4.6 $\mu m^2/s$, respectively.

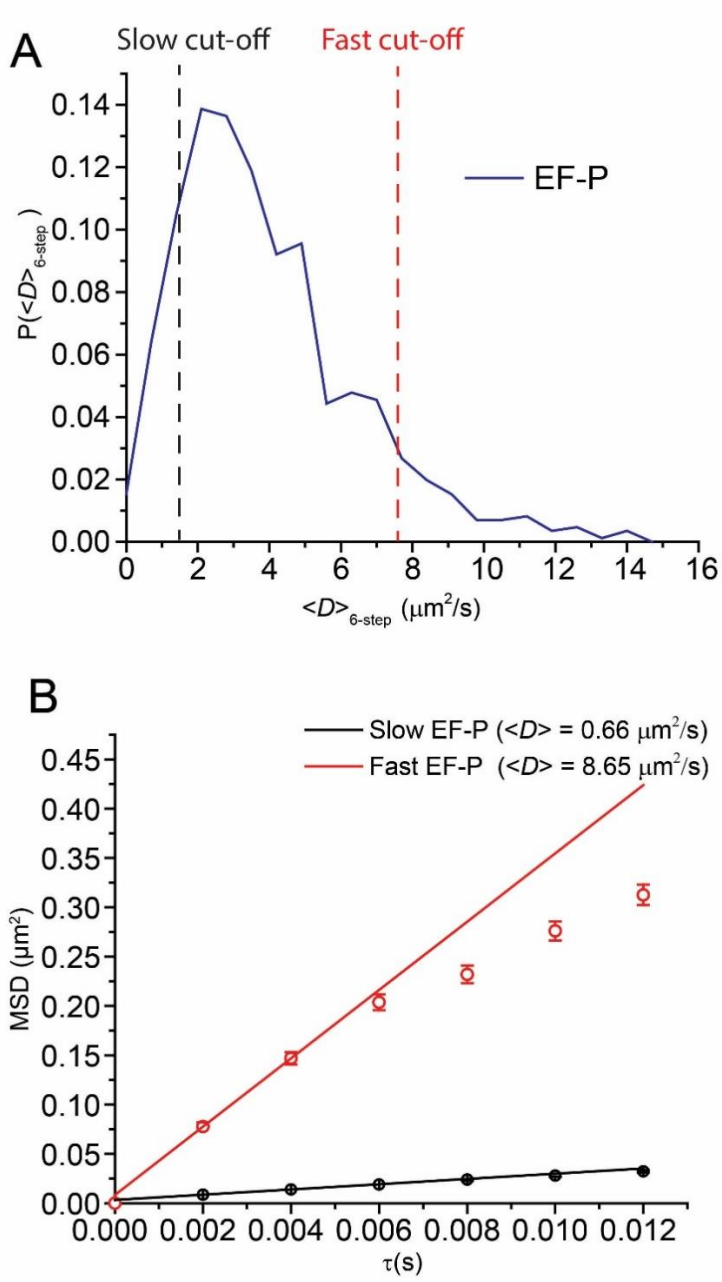


Fig. A4.5 (**A**) For EF-P–mEos2 in normal growth conditions, probability distribution of the mean of six successive one-step estimates of the diffusion coefficient, $P(\langle D \rangle_{6\text{-}step})$. Cut-off values of D = 1.56 μm^2/s and D = 7.8 μm^2/s select the 10% slowest and the 10% fastest mean values. (**B**) Mean square displacement plots, MSD(τ), for the fastest 10% of EF-P–mEos2 trajectories (red circles) and the slowest 10% (black circles). Error estimates are ± 1 σ of the MSD values. The slope of the first two data points yields diffusion coefficient estimates of 8.65 μm^2/s and 0.66 μm^2/s, respectively. Intercepts are used to estimate dynamic localization errors σ_{fast} and σ_{slow}.

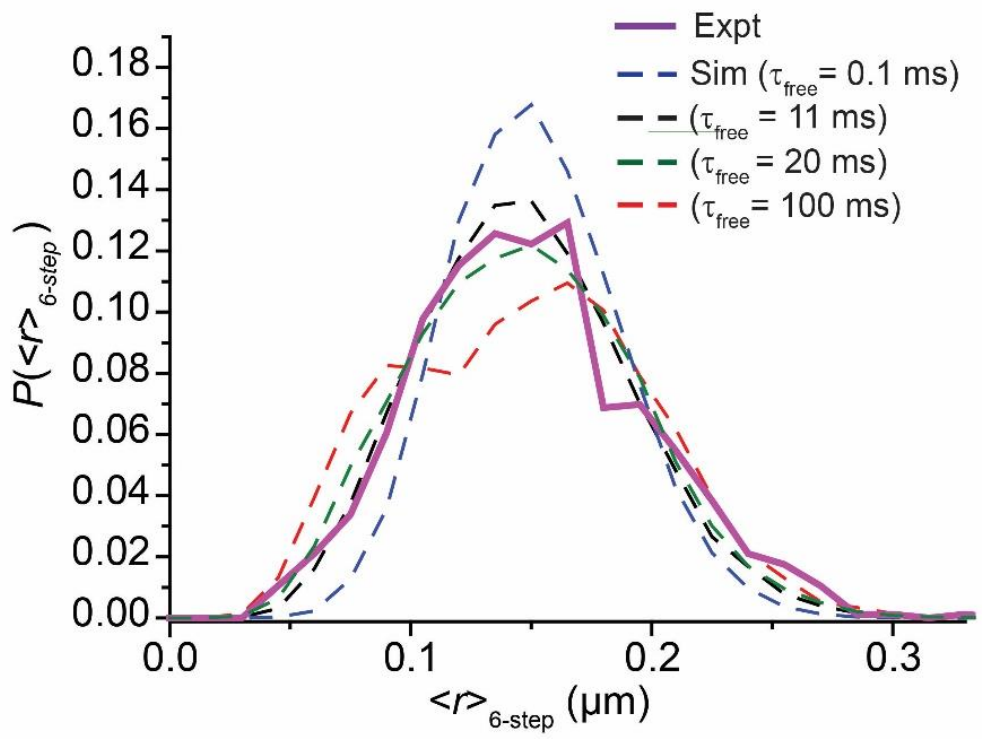


Fig. A4.6 Experimental probability distribution of the mean of six successive single-step displacements of EF-P–mEos2 trajectories truncated to 6 steps, $P(\langle r \rangle_6)$ (bold line). Dashed lines are simulated results for two-state models with binding-unbinding kinetics. All simulations use the model parameters D_{slow} = 0.2 μm^2/s (σ_{slow} = 50 nm), f_{slow} = 0.3, D_{fast} = 4.3 μm^2/s (σ_{fast} = 75 nm), f_{fast} = 0.7, taken from the best fit to the single-step $P_{EF\text{-}P}(r)$ as in Fig. 4.4*A*. The ratio τ_{free} / τ_{bound} = 7/3 is fixed by the best-fit fractions. The best fit has τ_{free} = 16 ms (and τ_{bound} = 7 ms), with χ_v^2 = 1.4, as shown in Fig. 4.4. The simulation results shown for τ_{free} = 11 ms (χ_v^2 = 1.95) and 20 ms (χ_v^2 = 2.12) set the error bars on τ_{free}. Also shown are simulations for τ_{free} = 0.1 ms (fast exchange limit compared with 2-ms camera frame time) and 100 ms (slow exchange limit), which give worse fits.

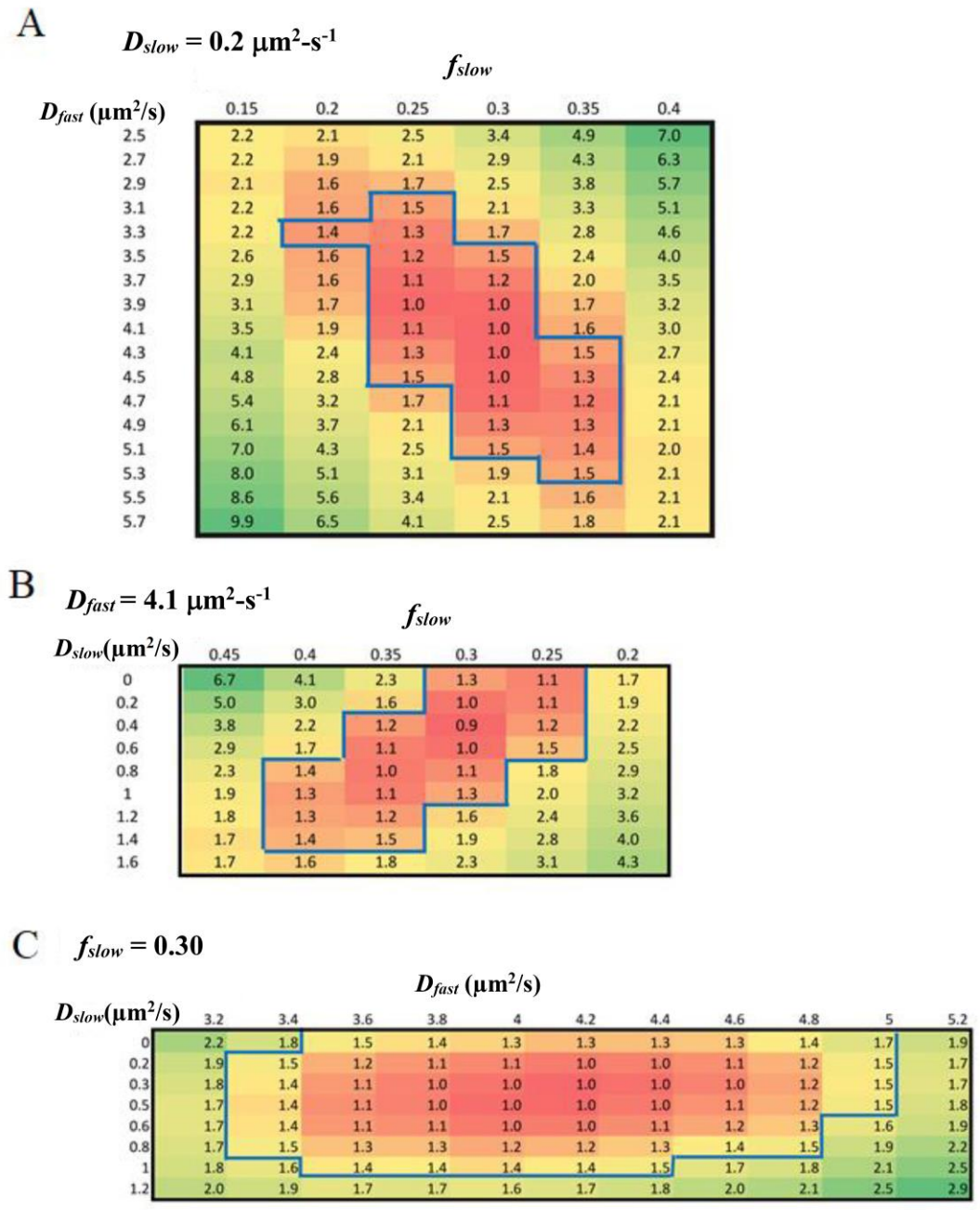


Fig. A4.7 Reduced chi-square values χ_v^2 for fits to $P_{EF\text{-}P}(r)$ using different parameter sets.Three slices through the 3D grid of χ_v^2 values (with fitting parameters D_{slow}, D_{fast}, and f_{slow}) obtained in modeling the experimental distribution of one-step displacements in Fig. 4.4*A*. (**A**) Slice through the plane with D_{slow} = 0.2 μm/s; D_{fast} and f_{slow} vary. (**B**) Slice through the plane with D_{fast} = 4.3 μm²/s; D_{slow} and f_{slow} vary. (**C**) Slice through the plane with f_{slow} = 0.30; D_{slow} and D_{fast} vary. Regions outlined in blue have $\chi_v^2 \leq 1.5$ and produce qualitatively poorer fits than those with $\chi_v^2 \sim 1$. Those regions were used for error estimates on best-fit parameters.

A4.11 Supporting References

1. Bakshi S, Bratton BP, Weisshaar JC. 2011. Subdiffraction-limit study of Kaede diffusion and spatial distribution in live Escherichia coli. Biophys J 101:2535-44.
2. Michalet X. 2010. Mean Square Displacement Analysis of Single-Particle Trajectories with Localization Error: Brownian Motion in Isotropic Medium. Physical review E, Statistical, nonlinear, and soft matter physics 82:041914-041914.
3. Thompson RE, Larson DR, Webb WW. 2002. Precise nanometer localization analysis for individual fluorescent probes. Biophys J 82:2775-2783.
4. Bakshi S, Dalrymple RM, Li W, Choi H, Weisshaar JC. 2013. Partitioning of RNA polymerase activity in live *Escherichia coli* from analysis of single-molecule diffusive trajectories. Biophys J 105:2676-2686.
5. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. 2012. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. Mol Microbiol 85:21-38.
6. Stracy M, Jaciuk M, Uphoff S, Kapanidis AN, Nowotny M, Sherratt DJ, Zawadzki P. 2016. Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli. Nature Comm 7:12568.
7. Chen T-Y, Santiago AG, Jung W, Krzemiński Ł, Yang F, Martell DJ, Helmann JD, Chen P. 2015. Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. Nature Comm 6:7445.
8. Li W, Bouveret E, Zhang Y, Liu K, Wang JD, Weisshaar JC. 2016. Effects of amino acid starvation on RelA diffusive behavior in live Escherichia coli. Mol Microbiol 99:571-85.
9. Das R, Cairo CW, Coombs D. 2009. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. PLoS Comput Biol 5:e1000556.

5

Future Directions

In this thesis, I have discussed how superresolution fluorescence microscopy was employed to investigate the spatial organization of different components of translational machinery. Single molecule tracking aided by computational models of diffusion within the confines of bacterial cell boundaries enabled us to quantify the binding/unbinding dynamics and time scales of diffusion of proteins in the crowded cell environment. These studies offer a flavor of how the complexities of protein diffusion in bacteria can be disentangled using single molecule tracking methods.

Our imaging method relies on 2D projections of 3D distributions of fluorescent proteins diffusing in the cell cytoplasm to quantify intracellular spatial distributions. 2D projections of 3D spatial distributions can often lead to loss of information. 3D PALM imaging and tracking (1, 2) that can rely on astigmatism, bifocal optics or point spread function engineering can resolve this issue and provide a more realistic picture of the internal cellular organization of proteins in bacteria. Recent studies on *C. crescentus* have employed 3D PALM to investigate the spatial organization of ribosomes (3) and HU (4). In our work, two color PALM imaging methods enabled us to visualize the relative spatial arrangement of ribosomes and DNA in *E. coli*. We relied on modified Pearson correlation coefficient (MPCC) to quantify the degree of anti-correlation of these two spatial distributions. The obtained 2D MPCC values surely underestimate the degree of segregation in the actual 3D *E. coli* cells. Two color 3D PALM experiments would be more informative; with enough data, they would yield 3D correlation coefficients undistorted by the “squashing” of the image from 3D to 2D inherent in our study. Such studies are usually complicated by the limited number of compatible fluorophore combinations available to perform two color PALM. Advancements in microscopy and

development of newer photoactivatable fluorescent proteins will provide more viable alternatives.

New insights into translation mechanisms *in vivo* will no doubt be obtained by labeling and tracking different components of translational machinery. Investigation of ribosomal binding/unbinding dynamics of other translation factors such as translation initiation factors, translation elongation factors like EF-G, release factors in live cells could provide us more information about mechanistic details of translation *in vivo*. This could also potentially motivate the design of newer translation halting antibiotics with different modes of action.

The ability to extend tracking of components of translational machinery to time scales comparable to those for the translation of an entire protein (~ 30 s) will be transformational. It will permit us to monitor ribosomes for the entire round of searching for initiation sites on mRNA and translation of a protein. Our studies on ribosomes in *E. coli* under different conditions are suggestive of existence of a "circulation model" of ribosomes across all growth conditions (5, 6). These long ribosome tracks should exhibit transitions from translating to non-translating state and give us more conclusive evidence of circulation of ribosomes. Careful regulation of ribosome usage is critical for proper cell growth. One of the ways the cell makes efficient usage of ribosomes is by recycling them after the termination of protein synthesis. Using these long ribosome trajectories, one could possibly characterize the movement of recycled ribosomes from ribosome-rich regions to the nucleoid region and the transfer of translating ribosomes from nucleoid to ribosome-rich regions.

Further improvement of fluorescent proteins that are more photostable than the ones in use currently (PAmcherry, mEos2, Dendra2 etc.) should enable extending tracks to longer time scales. Improving localization accuracy of imaging will make it easier to distinguish the different

diffusive states of our protein of interest. Organic fluorophores are much more photostable and have a better quantum yield compared to fluorescent proteins. They offer a promising future in terms of improving localization accuracy of imaging as well as increasing length of trajectories by orders of magnitude. Recently developed protein labelling strategies such as SNAPTag or HaloTags and dL5-MG have been demonstrated for their usefulness in obtaining long tracks of protein in bacteria (7-10). These tags bind tightly and form covalent bonds with a membrane permeable ligand, which can be modified with organic fluorescent dyes. Labelling of ribosomal subunits and/or translation factors with such tags in *E. coli* can add to our understanding of translation.

According to our model of *E. coli* spatial organization, most translation is physically separated from most transcription in all growth conditions. This is suggestive of the spatial distribution of total mRNA being similar to that of ribosomes. In the past, several studies have investigated the cellular location of mRNA using fluorescence microscopy. These studies have provided evidence for multiple, mRNA-specific localization patterns, such as localization within the cell cytoplasm (11), to the cell membrane or to the cell poles (12), and at the nascent septum separating daughter cells (13). However, such experiments are limited to specific mRNAs. A recent study (14) by Zhuang and co-workers employed a combination of STORM and FISH to reveal a dependence of mRNA localization on their function. Hopefully, in the future the study can be further extended to obtain a genome wide mRNA spatial distribution relative to that of ribosome spatial distribution. This will provide direct evidence supporting the picture of transcription and translation being segregated from each other in *E. coli*.

Single molecule imaging in live bacteria has already reported extensively on many properties of proteins and other biomolecules, such as their copy number, intracellular spatial

distribution, mobility and dynamics of interactions with other molecules. In the last decade, our understanding of bacterial cellular organization has benefited tremendously from the employment of superresolution fluorescence imaging methods. These imaging techniques are also applicable to address biological questions in medically relevant bacteria. Further developments of these imaging techniques and associated data analysis methods will allow deeper understanding of the dynamics and interactions of proteins inside living cells.

References

1. Holden SJ, Pengo T, Meibom KL, Fernandez Fernandez C, Collier J, Manley S. High throughput 3D super-resolution microscopy reveals *Caulobacter crescentus in vivo* Z-ring organization. Proceedings of the National Academy of Sciences. 2014;111(12):4566.
2. Huang B, Babcock H, Zhuang X. Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. Cell. 2010;143(7):1047-58.
3. Bayas CA, Wang J, Lee MK, Schrader JM, Shapiro L, Moerner WE. Spatial organization and dynamics of RNase E and ribosomes in *Caulobacter crescentus*. Proceedings of the National Academy of Sciences. 2018.
4. Lee Steven F, Thompson Michael A, Schwartz MA, Shapiro L, Moerner WE. Super-Resolution Imaging of the Nucleoid-Associated Protein HU in *Caulobacter crescentus*. Biophysical Journal. 2011;100(7):L31-L3.
5. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution Imaging of Ribosomes and RNA Polymerase in Live *Escherichia coli* Cells. Molecular Microbiology. 2012;85(1):21-38.
6. Sanamrad A, Persson F, Lundius EG, Fange D, Gynnå AH, Elf J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. Proceedings of the National Academy of Sciences. 2014;111(31):11413.
7. Fernández-Suárez M, Ting AY. Fluorescent probes for super-resolution imaging in living cells. Nature Reviews Molecular Cell Biology. 2008;9:929.
8. Hinner MJ, Johnsson K. How to obtain labeled proteins and what to do with them. Current Opinion in Biotechnology. 2010;21(6):766-76.
9. Ke N, Landgraf D, Paulsson J, Berkmen M. Visualization of Periplasmic and Cytoplasmic Proteins with a Self-Labeling Protein Tag. Journal of Bacteriology. 2016;198(7):1035-43.
10. Saurabh S, Perez AM, Comerci CJ, Shapiro L, Moerner WE. Super-resolution imaging of live bacteria cells using a genetically-directed, highly photostable fluoromodule. Journal of the American Chemical Society. 2016;138(33):10398-401.

11. Valencia-Burton M, McCullough RM, Cantor CR, Broude NE. RNA visualization in live bacterial cells using fluorescent protein complementation. Nature Methods. 2007;4:421.

12. Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O. Translation-Independent Localization of mRNA in *Escherichia coli*. Science. 2011;331(6020):1081.

13. dos Santos VT, Bisson-Filho AW, Gueiros-Filho FJ. DivIVA-Mediated Polar Localization of ComN, a Posttranscriptional Regulator of Bacillus subtilis. Journal of Bacteriology. 2012;194(14):3661-9.

14. Moffitt JR, Pandey S, Boettiger AN, Wang S, Zhuang X. Spatial organization shapes the turnover of a bacterial transcriptome. eLife. 2016;5:e13065

6

Spying on the bacteria

This chapter is written with an intent to communicate my research to general public

The Good, the Bad and the Ugly

The word “bacteria” often has a negative connotation. People generally think of bacteria in the context of infectious diseases and consider them to be villains that need to be eradicated from the face of earth. However, a world without microbes is not only uninhabitable but also unachievable. Most bacteria are harmless or even beneficial to us. Our human bodies host trillions of bacteria that not only help digest our foods but also help protect us from diseases. Some of our favorite foods such as cheese, wine, yoghurt, pickles are the result of hard work done by bacteria. Owing to the simple machinery of bacteria, scientists modified them to be used as “micro-factories” for production of insulin, vaccines, other life-saving drugs and even jet fuel.

Why study bacteria?

Bacteria are single celled organisms with simple internal structure. Their simplicity, compared to other organisms, makes them suitable study subjects to answer fundamental questions related to cell functioning. A study of various aspects of bacterial cell functioning can help us design antibiotics that target harmful, pathogenic bacteria. These bacterial studies also enable us to manipulate the bacterial cells to carry out functions that are beneficial to us.

One of the most well studied bacteria is the rapidly growing *E. coli*, which divides every 20-40 minutes. Every organism contains a genome, which is like an “instruction manual” for that organism. This genome contains information and instructions for the operation of every single component of the organism. This instruction manual is photocopied and passed down to the successive generations. The *E. coli* genome is nearly 1000 times smaller than the human genome, making it easier for scientists to analyze.

One reason we study *E. coli* so carefully is because the basic biological principles learned from experiments on *E. coli* are generally applicable to other organisms. French scientist Monod famously noted, "What is true for the *E. coli* is true for the elephant." Many of the genetic properties that govern *E. coli* are true for us.

Even though *E. coli* is a very well-studied organism, we still do not understand several aspects of its internal workings.

Understanding the inner workings of *E. coli*

If you were asked to analyze how a mechanical clock works, a reductionist would separate out its various components (gear trains, oscillators, springs) and then put them back one by one to understand how these different parts interact with one another. The overall behavior of the clock is determined by the arrangement of the different pieces of the clock as well as the interactions of the tiny pieces with each other.

Proteins are fundamental entities essential for the survival and growth of all biological cells, including *E. coli*. Their interactions result in a wide variety of protein complexes that serve as the basis for how cells function. Scientists extract proteins out of the cells and study protein interactions physically and biochemically *in vitro* (outside the cell). Most of these studies, however, cannot replicate the complexity proteins encounter in the cellular environment. Scientists find it extremely difficult to recreate the complicated cocktail of biomolecules that comprise the natural milieu of living cells.

E. coli is a much more complicated system than a clock. Its properties arise from a network of interactions between different components as well as with the surrounding environment. *E. coli* is more than the sum of its parts. The properties of any protein present in an

E. coli are not equivalent to the sum of properties of individual amino acids that constitute the protein. The usefulness of a protein depends not only on the sequence of amino acids, but also on its three dimensional structure and environment. A reductionist approach may therefore not properly account for properties of a complex system like *E. coli*. To understand the biological complexity of *E. coli*, we need to investigate both the individual components (proteins, lipids etc.) and how these components interact with one another on a system-wide level. Scientists are therefore making efforts to study protein interactions in the natural context of living cells.

The network of protein interactions govern the interior architecture of *E. coli*. Two proteins that are located near each other in a cell may be engaged in the same process, or bind to a common element, or share sites of production, degradation or action. On the other hand, if two proteins avoid each other in a cell, investigating their spatial localization can give us information about physical and biochemical processes that lead to their sequestration. We can look at how different components are organized inside *E. coli* and obtain information about communication between different components of the same process as well as communication between different processes. How instructions from the genome govern the internal architecture of the cell is one of the open questions in biology.

Peeking into the internal architecture of bacteria

Antonie van Leeuwenhoek, a Dutch scientist, was the first person to observe different kinds of bacteria under his handcrafted microscope. More improved microscopes enabled us to take a closer look into the bacteria. Microscopy images revealed that bacteria lack any internal organization or structural features. This is in stark contrast to other plant and animal cells where structures called organelles organize different cellular processes into different compartments of

the cell. Based on this major difference, scientists categorized organisms into prokaryotes (such as bacteria) and eukaryotes (such as fungi, algae, plants, and animals including humans).

Bacteria were viewed as bags of disorganized molecules devoid of any intracellular organization. On the other hand, eukaryotic cells could be compared to our houses. Our houses have areas designated for different activities: kitchen for cooking, bedroom for sleeping and so on. Similarly, eukaryotic cells have different cellular regions allocated for different functions.

The last few decades have seen tremendous developments in the field of microscopy and imaging. This has allowed us to study the interior of a bacteria in detail. Studies from the early 90s showed that different bacterial components exhibit unique spatial patterns. This was followed by a wealth of studies focused on spatial localization of various cellular processes in several different bacteria. Scientists realized that bacteria are not as simple as they previously believed.

Recent developments in microscopy enabled bacterial imaging

In the 1960s, Osamu Shimomura, who had an interest in studying bioluminescence (the process by which animals emit light, e.g. fireflies), isolated a bioluminescent protein from a jellyfish. This protein gives off green light when illuminated with UV (ultra-violet) light, by a process called fluorescence. He named this protein the green fluorescent protein (GFP). Since then, scientists have discovered and modified many variants of the members of the "GFP family" of proteins.

GFP and its other variants basically serve as "spies" that scientists could introduce into the cells. GFP could be fused to any protein of interest without affecting the function of the protein. The green fluorescent signal from a labelled protein maps its cellular location. These GFP-like proteins allow us to monitor in time and space an ever-increasing number of

phenomena like gene expression, protein localization and dynamics, protein-protein interactions, cell division, chromosome replication and organization and intracellular transport pathways in living cells and organisms.

GFP revolutionized our ability to dissect mechanisms in *E. coli* and investigate its internal structure using microscopes. The discovery of GFP and its modification to tag proteins of interest provided a non-invasive method of looking at proteins in their natural environment. A fundamental set of physical laws govern the highest achievable point-to-point resolution that can be obtained with an optical microscope. These resolution limitations are referred to as the "diffraction barrier**"**, which restricts the ability of optical instruments to distinguish between two objects separated by a lateral distance less than ~ 250 nm, nearly $1/400^{th}$ of human hair diameter. *E. coli* are rod-shaped cells that are 900 nm in diameter and 2-8 μm (~ $1/20^{th}$ of human hair diameter) in length. As a result, traditional microscopy methods with a resolution of ~ 250 nm were ill-suited for studying the inner workings of bacteria.

Scientists have modified some members of the GFP family to be "photoactivatable". We need two light sources to image these "photoactivatable" molecules. These molecules undergo chemical conversion when we shine a low power "activation" light at them. We then employ a high power laser to image only the activated molecules until they bleach. A picture of all of the photoactivatable proteins in the entire cell is developed by multiple rounds of imaging. This principle forms the basis of PALM (Photoactivated Localization Microscopy), whose discoverers were awarded the Nobel Prize in 2014. PALM has a "crowd control" feature that enabled us to image single molecules in live bacterial cells. Single molecule imaging in living cells provides us with a wealth of information. In addition to providing very precise information about the location of an individual protein in a cell, PALM allows us to follow the protein as it is moving. We can

think of PALM as a camera that not only provides static pictures of protein location but can also provide us with videos of that protein moving.

Importance of studying single molecules and their motion

Averages can be misleading at times. Just looking at averages, one could conclude that every person on this earth is half male and half female. Only by looking at individuals does it become obvious that they have different gender identities. In biological studies, different molecules behave differently and experience different local cell environments. Different cellular processes do not start and stop at the same time. In such situations, ensemble measurements can lead to the same sort of problem. Measurements that capture only the average of a population of molecules blur the differences between individual molecules, and important information is lost. One of the reasons for exploring single molecules is to discern whether there is different behavior from one copy of a molecule to the next. For some biological questions, single-molecule studies are a necessity. With the new PALM microscopy method, we can perform single molecule experiments on proteins in their natural environment.

We can use PALM to determine how fast a molecule is moving. This in turn can help us sort molecules into groups of different mobility. One can also identify in which areas of the cell a protein diffuses slowly and in which regions of the cell it diffuses faster. Correlating the motion of a protein with its intracellular location can give us an idea about the function of the protein. For example, if we knew GPS locations of all the members in the Chemistry department of UW Madison, we could potentially make educated guesses about the roles of each member in the department. If members spend most of their time in business offices, they are most likely employees of the business office. If some members spend time in different classrooms during the day, they are more likely to be students of the department. Similarly, if a protein localizes to the

region of the cell where DNA exists, it is probably a DNA binding protein. On the other hand, if a protein localizes to the membranes of the cell, it is likely to be a membrane protein engaged in transportation of molecules into and out of the cell, and communication between cells. This information on the motion of single proteins, in combination with the vast number of *in vitro* experiments done on proteins in *E. coli*, helps us determine the function of proteins in different regions of the cell.

A protein molecule could switch from very mobile → immobile →very mobile and vice-versa during the course of its motion. The mobility of a molecule is related to its role in a cellular process. When a protein is bound to its target site, it is immobile; and when it is searching for its target site, it is very mobile. We can measure the time a protein spends bound to its target by measuring the amount of time it stays immobile.

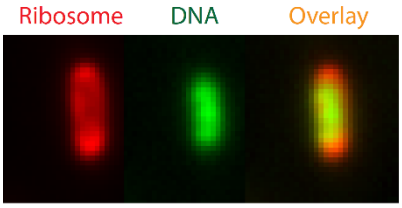
What did I work on during my PhD?

During my PhD, I relied on these single molecule tracking methods to quantitatively study the translational machinery of *E. coli*. There are many processes that take place in cells that are essential for life. Translation is one of the most important cellular process in all organisms. It is the process by which cells make proteins for their survival. The cells spend a lot of energy making these proteins. Cells employ ribosomes as the protein builders.

As a part of my thesis, I have worked with *E. coli* in two different growth conditions; fast and slow. In the fast growth conditions, I provided *E. coli* with nutritious food that allowed them to grow rapidly (dividing every 50 min). In the slow growth conditions, *E. coli* were starved which led them to grow slowly and divide every 150 min. We were curious if ribosomes would be distributed differently in a happily growing *E. coli* and a starved *E. coli*. I labelled ribosomes

with YFP (yellow fluorescent protein, a member of the GFP family). In addition, I labelled the DNA of the cell with a dye called Sytox Orange. This allowed me to see the location of the fluorescent ribosome molecules with respect to that of DNA. In Fig. 6.1, I have shown the location of ribosomes and DNA in the two differently grown *E. coli*. Similar to earlier studies, we observed that in fast growing *E. coli*, there are two regions (nucleoids) where most of the DNA is located. In contrast, in slow growing cells, there is a single nucleoid region. The ribosomes and DNA avoid each other despite the difference in growth conditions.

A.



B.

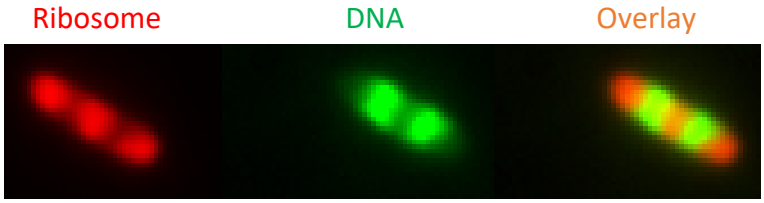


Fig. 6. 1 Images of ribosomes labelled with YFP and DNA labelled with Sytox Orange in two different growth conditions. The composite image shows that ribosomes and DNA avoid each other in *E. coli*. **(A)** Slowly growing *E. coli* **(B)** Rapidly growing *E. coli*

The ribosomes that are engaged in protein building (translating ribosomes) are less mobile compared to the ribosomes that are looking for sites to begin protein building (searching ribosomes). Using PALM, I was able to distinguish the translating ribosomes and the searching ribosomes. Our comparison of how many ribosomes are engaged in translation in these two different kinds of cells revealed that there is a slight decrease in the number of translating ribosomes in slow growing *E. coli*. One possible explanation for this is that the starving *E. coli* require fewer proteins to survive.

We were able to map ribosomes in different regions of the cell based on their functions. We found that the majority of translating ribosomes are located outside the nucleoid region, which could indicate that most of the translation in cells occurs in regions outside the nucleoid. The ribosomes that localized to the nucleoid region of the cell are searching for sites to initiate construction of a protein. We were able to estimate timescales of exchange of ribosomes between different cellular regions.

How can my research benefit the public?

In my graduate research, I have used cutting edge tools to obtain a detailed and quantitative understanding of the translational machinery in *E. coli*. We hope that my graduate research work using the model bacteria *E. coli* could open doors for possible development of new antibiotics targeting the translation process. Human biology encompasses much more than what goes on in human cells, as there are at least as many bacterial cells as human cells on and in our bodies. *E. coli* may not be the most prevalent bacteria in our body, but we hope that our work can be extended to other bacteria that are critical to human biological processes.