Single-Molecule Localization, Dynamics and Interactions of DNA Replication and Repair Proteins Revealed by Live-Cell Super-Resolution Microscopy

by

Yi Liao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2015

Doctoral Committee:

Assistant Professor Julie S. Biteen, Chair
Professor Zhan Chen
Associate Professor Jennifer P. Ogilvie
Associate Professor Lyle A. Simmons
Assistant Professor Sarah L. Veatch

Acknowledgements

I was not expecting graduate school to be this rewarding, fun, and challenging. Now, as I get ready to start the next stage of my academic pursuit, I would like to express my gratitude towards the people who have guided me along the way, and to those who have made my stay here at Michigan a truly amazing experience.

I would like to start by thanking my advisor Professor Julie Biteen for taking me into the field of biophysics, for her excellent mentorship, and for her consistent trust, patience and encouragement. As a student, it is incredibly reassuring that whenever I need scientific insights Julie is there to help, even on her busiest days. And when things become difficult, I can always hear uplifting words from her as well. Over the course of my Ph.D. training, Julie has also encouraged me to attend science workshops to expand my skill set, and has granted me a number of opportunities to present my work at renowned science conferences within US and abroad. In addition to helping me navigate the science world, Julie has always been a tremendous role model whom I look up to, and hope one day I will be able to emulate her work ethic, scientific acumen and positivity.

I am also indebted to my collaborators from the *Bacillus subtilis* projects, Professor Lyle Simmons and Jeremy Schroeder. I have truly enjoyed working with Lyle and Jeremy, who are very knowledgeable biologists with brilliant insights and a great sense of humor. I want to thank them for their significant contribution to these projects, and also for sharing their knowledge and time in helping me to become a competent researcher in biology.

I would like to thank my committee members, Professors Zhan Chen, Jennifer Ogilvie and Sarah Veatch, for coming to my candidacy exam, data meeting and thesis defense, and for offering their wisdom and feedback for my research projects. I also want to thank Professor Qiong Yang for her valuable career advice.

My sincere gratitude also goes to my collaborators from the MCP project, Professor Adam Matzger and Dr. Kyoungmoo Koh. Without their valuable input and suggestions our work on this project would not be this fruitful.

As difficult as things sometimes become, I was able to keep my sanity thanks to the fantastic members of the Biteen Lab. Dr. Mou-Chi Cheng and Beth Haas were the first two lab members who welcomed me into the group, then patiently taught me everything I needed to know and really helped me to quickly adapt the graduate school research environment. I want to thank Dr. Esther Wertz for teaching me how to play lasers like a physicist and in a safe manner, and I wish her all the best in her career as a Professor. My thanks also go to Dr. Krishanthi Karunatilaka, for providing helpful research advice from a biochemist's perspective, as well as for her genuine and consistent encouragements. I especially want to thank Dr. Justin Lenhart, Chanrith Siv and Bing Fu for patiently teaching me biochemistry techniques step-by-step, and for just being wonderful people to have around. Also, I have to say it was really fun to be a member of our lab trivia team "the Nutators," which also happens to be the morning workout team of our lab: Dr. Hannah Tuson, Jessica Donehue Flynn, David Rowland and Ben Isaacoff. Thank you for all the fun we have had (and all the prizes we have won!) at trivia games, thank you for spending those nice summer mornings running with me in the Gallup Park, and thank you for sharing the office with me and bearing my (brilliantly witty but occasionally inappropriate) jokes. As well, I would like to thank the rest of the crew: Yilai Li, Yue Xie,

Stephen Lee, Josh Karslake and Sarah Hanner for all the good times in the lab and for being part of my wonderful memories here in the Biteen Lab.

Lastly and most importantly, I want to thank my parents far away back in China, for their unconditional love and support along the way. It has been 9 years since I came to the US, and being the only kid in the family, I understand this is anything but easy for them, and I have my deepest gratitude and respect for their sacrifice.

Yi Liao

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

The error-free progression of DNA replication is essential for all organisms. Approximately 80 known human diseases are caused by malfunction in DNA replication, and deficiencies in DNA repair mechanisms can also lead to severe consequences such as increased antibiotic resistance in bacteria and cancers in humans. A better understanding of DNA replication and repair requires knowledge of the key players along relevant pathways at the molecular level and in the cellular context. This characterization calls for a technique with superior sensitivity, accuracy and biocompatibility. In this thesis, I integrate single-molecule super-resolution microscopy and single-particle tracking with genetic and genomic approaches to study two proteins that play a pivotal role in maintaining genomic integrity: MutS and PolC.

From prokaryotes to human cells, homologs of the highly conserved mismatch repair (MMR) protein MutS recognize mispaired nucleotides and recruit the proteins responsible for downstream repair. Although the structure and function of MutS have been extensively characterized in biochemical isolation, it remains unclear how MutS efficiently identifies, among millions of correctly paired bases, a single mismatch in the complex and crowded cellular environment, where DNA is supercoiled, compacted and bound by hundreds of proteins. To obtain mechanistic insight into MMR initiation from an *in vivo* perspective, I applied super-resolution imaging in live *Bacillus subtilis* cells to follow the motion of single MutS proteins in real time, and we monitored how MutS behavior is affected by sequentially blocking critical steps along the MMR pathway. Results from imaging and genetics experiments together demonstrate that (1) a population of MutS is recruited to the site of active DNA replication (the replisome) prior to mismatch recognition to scan newly

synthesized DNA in anticipation of errors, while the other MutS population diffuses throughout the cell, (2) the recruitment of MutS to the replisome via interactions between MutS and replisome subunits is independent of mismatch formation, (3) mismatch binding must take place at the replisome and MutS cannot associate with mismatches distal to the replisome, and (4) MutS is physically associated with the site of ongoing DNA synthesis and is rapidly loaded at *oriC* upon chromosomal replication initiation. These results present the first real-time, super-resolution investigation of any dedicated DNA repair pathway in live cells, and provide quantitative characterizations of the intimate and dynamic association between DNA replication and mismatch detection in real time.

I then turn my focus to DNA replication itself. Specifically, I examined PolC, one of the two essential DNA polymerases in *B. subtilis*. The *B. subtilis* model organism presents an intriguing system in which to study DNA polymerase dynamics because many DNA replication mechanisms are conserved between *B. subtilis* and eukaryotic cells. Based on photobleaching-assisted microscopy and three-dimensional super-resolution microscopy, we quantified the stoichiometry and intracellular locations of PolC as well as the rate of polymerase exchange.

Finally, we extended the application of super-resolution microscopy to the field of renewable energy by tracking single molecules and visualizing guest-host interactions in microporous coordination polymers (MCPs). Imaging diffusion in these solid state materials demonstrates the broad applicability of single-molecule fluorescence imaging, and drove the development of new methods to analyze motion in multiple dimensions.

The results presented in this thesis will improve our current understanding of the crucial processes of DNA replication and repair, and also demonstrate the power and versatility of super-resolution microscopy to address fundamental questions across disciplines.

Chapter 1 Introduction

1.1 Optical microscopy and biology

Advances in our understanding of biology and medicine are inextricably linked to the development of optical microscopy techniques. Four centuries after its first appearance, the optical microscope has now become an indispensable tool in almost all areas of biomedical research, allowing us to see organisms previously unseen and to perform experiments on small scales previously inaccessible. As one major application, optical microscopy has been extensively used to visualize and study bacteria cells, which, due to their small size that is generally on the scale of a few micrometers (μm, or 10^{-6} m), are not visible by the naked eye. With a resolution around 300 nanometers (nm, or 10^{-9} m), a standard optical microscope is more than sufficient to resolve microbial behaviors at the single-cell and population levels. However, in order to dissect biochemical mechanisms underlying many of their important behaviors and properties, it is often necessary to peer into these tiny cells to visualize how relevant biomolecules function and interact with each other. Due to the much smaller size of biomolecules, typically on the scale of a few nanometers, such an endeavor ideally requires a two orders of magnitude improvement in spatial resolution over that offered by conventional optical microscopes, motivating what was a long-standing challenge in the field until the emergence of the single-molecule super-resolution microscopy two decades ago[1].

Similar to conventional fluorescence microscopy, single-molecule super-resolution microscopy also relies on fluorescent labels to visualize biomolecules of interest; however, it offers much improved spatial and temporal resolutions, enabling visualization of single

biomolecules in real time, while preserving many advantages of conventional microscopy such as high specificity and minimal invasiveness. Therefore, single-molecule microscopy opens up many new possibilities to visualize and examine the intricate inner working of cells. This thesis describes the application of this state-of-art technique to study how DNA is replicated and repaired in the live bacterium *Bacillus subtilis*. Combining experimental results obtained from super-resolution imaging and genomic approaches, I describe the crucial processes of DNA replication and repair from a single-molecule perspective in a quantitative manner that aims to uncover hidden insights and refine existing models. In additional, I provide one unique application of super-resolution microscopy in investigating guest-host interactions in mesoporous crystal structures. Imaging diffusion in solid-state materials demonstrates the broad applicability of single-molecule fluorescence imaging, and permits the development of new methods to analyze motion in multiple dimensions.

1.2 Fluorescence microscopy

Compared to other types of high-resolution imaging techniques such as electron microscopy (EM) (Figure 1.1), fluorescence microscopy is nondestructive and is compatible with live cell imaging because it uses visible light to probe samples. By labelling target molecules with a fluorescent tag, usually a fluorescent protein, a dye molecule or a quantum dot, fluorescence microscopy offers a high specificity that allows us to visualize only molecules of interest. Also, a variety of fluorophores with a wide range of colors have been developed and optimized for *in vivo* imaging, making it possible to carry out multi-color imaging experiments to visualize interactions between different molecules.

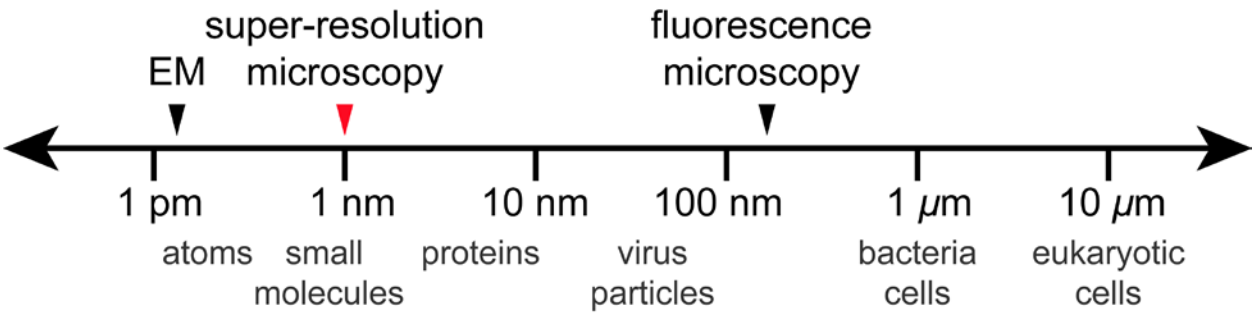


Figure 1.1 Scales of conventional fluorescence microscopy, super-resolution microscopy and electron microscopy compared to typical sizes of atoms, biological structures and organisms.

One major disadvantage of using light to visualize extremely small features such as biomolecules is the limited spatial resolution imposed by the fundamental, wave-like property of light. Because light has a finite wavelength on the scale of a few hundred nanometers, it is not possible to focus it to an infinitesimal point. Therefore, under a microscope, even an infinitesimally small object will appear to have a finite size that depends on the specific optical set-up. The diffraction pattern or equivalently, the apparent spreading of the shape of an infinitesimal object is described by the point spread function (PSF) (Figure 1.2A). In the lateral (x, y) plane, the PSF is an Airy function[2], and its width imposes an upper limit for the resolution of the corresponding set-up according to the following relation first discovered by Ernst Abbe[3]:

$$d = \frac{\lambda}{2NA} = \frac{\lambda}{2(n \cdot sin\theta)} \tag{1.1}$$

where λ is the excitation wavelength of the fluorophore, NA is the objective numerical aperture, n is the refractive index of the imaging medium, θ is the collection angle of the objective lens, and d is the minimal separation distance in the lateral plane between two objects that can still be resolved as two separate entities, This distance d is typically 200 nm – 300 nm for conventional

fluorescence microscopes, meaning that features separated from each other by shorter distances will appear as a single object. This resolution limit is not only about two orders of magnitude larger than the typical size of biomolecules, but it is also comparable to the size of a typical bacterium (~1 μm in width), which can be especially problematic because only a few molecules in a bacterial cell will be resolvable, given that they are physically well separated from each other.

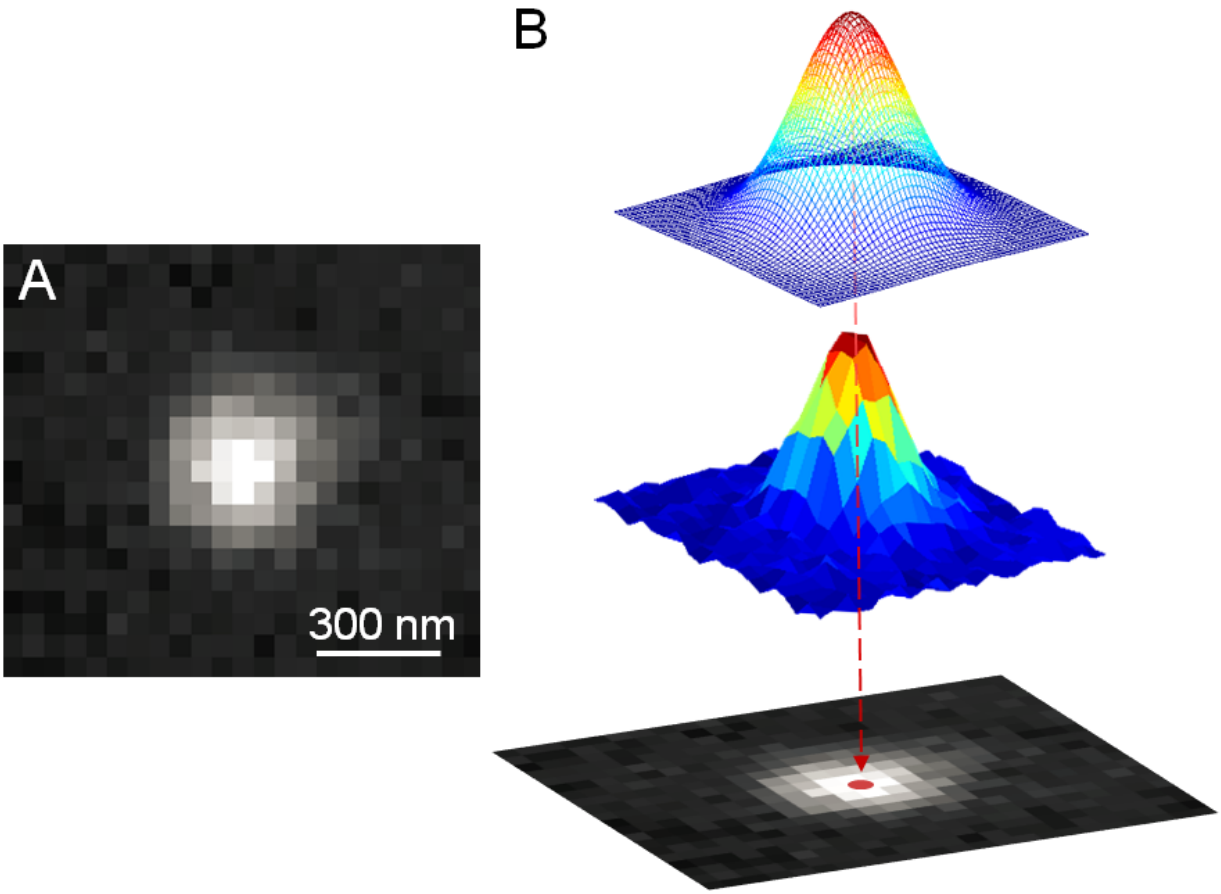


Figure 1.2 (A) The point spread function (PSF) of a single-molecule emitter. (B) Localizing the single-molecule position from its PSF through Gaussian-function fitting[4].

Another drawback of conventional fluorescence microscopy is the lack of sensitivity: only proteins that form significant agglomerates ("foci") are visible while individual copies are not. As a result, conventional microscope images depict the ensemble average of all the events that are simultaneously going on inside a cell, while masking valuable information regarding population heterogeneity and protein dynamics.

Considering the nature of the major outstanding problems in molecular biology and microbiology, and the limitations of conventional fluorescence microscopes, an imaging technique that retains the benefits of conventional fluorescence microscopy while offering better resolution as well as better sensitivity will be invaluable in tackling challenging questions in the biomedical field.

1.3 Optical microscopy beyond the diffraction limit

Since its advent in the late 1980s, Single-Molecule Fluorescence (SMF) Microscopy has enabled scientists to look at single fluorescent molecules one at a time[5]. Furthermore, by circumventing the diffraction limit that restricts conventional light microscopy, this technique allows us to visualize subcellular events *in vivo* with nanometer precision in real-time, and SMF Microscopy has helped to unravel the intricate architecture inside cells through a non-averaging, less ambiguous approach.

In order to detect fluorescence signals from a single-molecule emitter, the optical setup has to be carefully designed such that the majority of background noise[4], including reflected and scattered source light, as well as Raman scattering and fluorescent impurities from the solvent, can be rejected. A schematic representation of a typical SMF microscope set-up is depicted in

Figure 1.3. The laser beam passing through the excitation filter is circularly polarized by a λ/4 wave plate so that excitation is not biased toward any specific chromophore orientation. Adjustable mirrors and a periscope are used to direct the beams into a standard widefield epifluorescence microscope. Because both the resolution of the set-up and the precision with which an emitter can be localized improve with the number of photons detected, a high numerical-aperture (NA = 1.40) oil-immersion objective is used for photon collection. A dichroic mirror is incorporated into the excitation and detection pathways to separate the emission of the molecules from scattered light from the excitation beam; a long-pass emission filter further blocks any background noise. Finally, the emitted light passes through a 3x beam expander and reaches an electron-multiplying charge-coupled device (EMCCD) that detects and digitizes the incoming photons before sending the signals to a computer.

Depending on the application, the optical set-up can be modified to meet specific needs. For instance in two-color imaging, a beam splitter is inserted between the microscope and the 3x beam expander to allow for simultaneous acquisition of signals of two distinct emission colors, and other components along the optical path are modified accordingly to allow two-color excitation and emission lights to pass. If photoactivatable or photoswitchable fluorophores are used in the experiment[6-9], it is necessary to alternate the incoming laser between the activation beam (usually 406-nm) and the excitation beam (561-nm for example) such that photoactivation or photoswitching events can take place before fluorophore excitation. In this case, a second laser is added to the set-up and a shutter is placed in the optical path of each laser. Additionally, a weak cylindrical lens can be inserted between the microscope and the camera to allow for astigmatism-based three-dimensional imaging (Chapter 4).

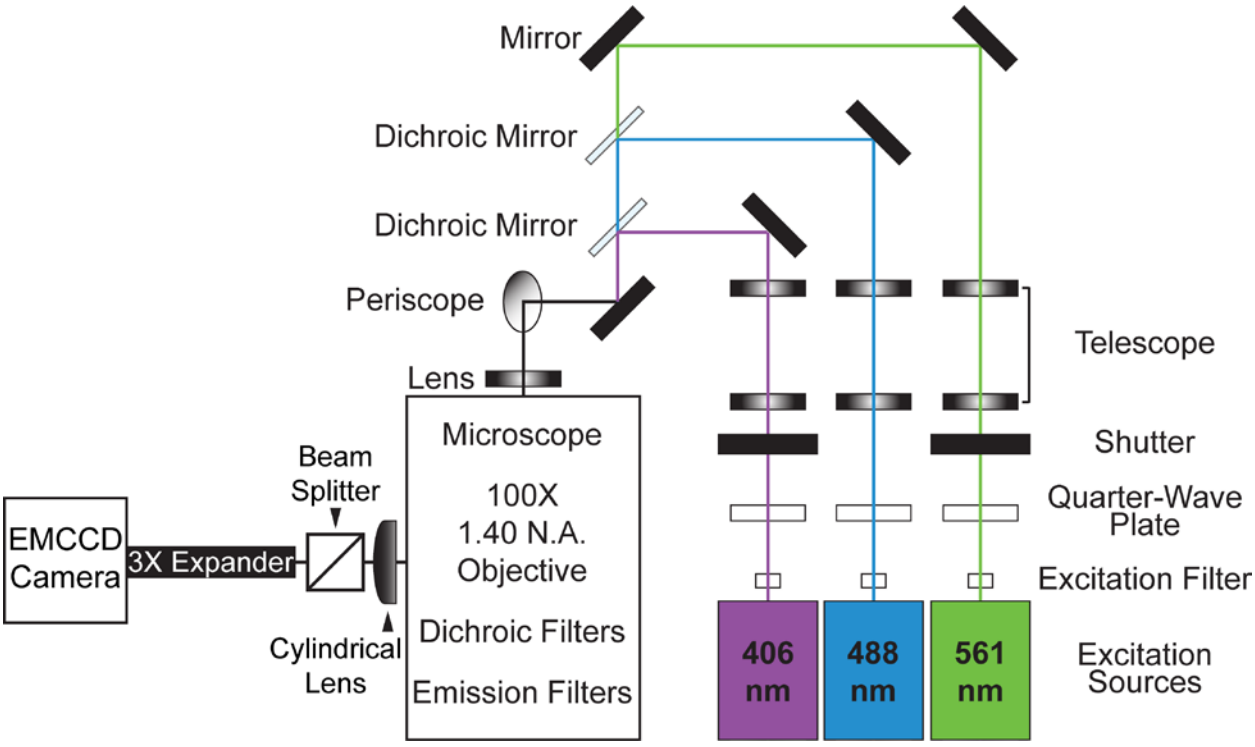


Figure 1.3 Optical set-up for single-molecule fluorescence imaging. Image is contributed by Hannah Tuson[4] and modified according to specific configurations used in studies from this thesis.

Single-molecule signals collected using the set-up above are still diffraction-limited, but they offer a starting point from which sub-diffraction-limit localization can be carried out. In his review article in Nature Methods[10], W. E. Moerner, the first person to optically detect single molecules in the condensed phase, used the analogy of finding the peak of a mountain to demonstrate the idea of localizing a molecule from its emission pattern: even though the mountain is many kilometers in diameter, its peak can be identified with reasonable certainty. The spatial distribution of fluorescence intensity from an isolated single-molecule emitter, which is most accurately described by an Airy function, very much resembles the shape of an isolated mountain and can be well approximated by a two-dimensional Gaussian function (Figure 1.2B),

$$f(x, y, A, \sigma_{xy}) = I_{bg} + A \cdot e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma_{xy}^2}} \tag{1.2}$$

where x_0 and y_0 are the coordinates of the Gaussian peak center and A is the signal amplitude above the background intensity I_{bg}. σ_{xy} is related to the full-width-at-half-maximum (FWHM) of the function according to:

$$\text{FWHM} = 2\sqrt{2ln2} \cdot \sigma_{xy}^2 \tag{1.3}$$

The Gaussian approximation of an Airy function is valid since the Airy function minor rings are almost always buried by background noise. Even so, this approximation usually yields localization precisions that are at least ten times smaller than the FWHM of the corresponding signal, but eventually how well single-molecule emitters can be localized is strongly affected by the signal-to-noise ratio (SNR) of raw data, which is defined as[11]:

$$\text{SNR} = \frac{A}{\sqrt{\sigma_{bg}^2 + \sigma_A^2}} \tag{1.4}$$

where σ_{bg} and σ_A respectively denote the standard deviations of the background intensity and the real intensity above the background. The higher the SNR (i.e., the steeper the "mountain"), the better the emitter can be located, and 1-nm localization precision has been reported, which is more than sufficient to study the position and dynamics of biomolecules in most experiments. Experimentally, two sources contribute to localization errors: shot noise associated with photon-counting and background noise associated with a number of factors such as out-of-focus fluorescence and camera readout noise. Shot noise causes the localization precision to scale with the inverse square root of number of photons ($N^{-1/2}$), whereas the background noise causes localization precision to deteriorate as a function of N. Combining these two, the localization error can be mathematically described as:[12]

$$\Delta x = \sqrt{\frac{s^2 + \frac{a^2}{12}}{N} + \frac{8\pi s^4 b^2}{a^2 N^2}} \tag{1.5}$$

where *s*, *a* and *b* respectively denote the standard deviation of the PSF, the pixel size of detector and the background noise.

1.4 Single-molecule microscopy modalities for resolving structures and motions

With spatial resolutions at the nanometer scale, single-molecule super-resolution microscopy has been most extensively used to study two aspects of biological macromolecules: structure and dynamics.

Photoactivated Localization Microscopy (PALM)[7], Stochastic Optical Reconstruction Microscopy (STORM)[6] and Fluorescence Photoactivation Localization Microscopy (FPALM)[8] are essentially identical methods independently developed by three research groups to visualize fine subcellular structures through the single-molecule approach. These methods and other variants are the most prominent super-resolution microscopy techniques. Because spatially overlapping PSFs would prevent the precise localization of point emitters, these methods circumvent the high-desnity problem by utilizing photoactivatable or photoswitchable fluorophores to separate overlapping PSFs temporally. By optically controlling the "on" and "off' states of photoactivatable or photoswitchable fluorophores, it is possible to visualize and localize only a subset of them that are sparsely located within a cell such that no spatial overlapping of PSFs occurs. Due to the stochastic nature of photoactivation and photoswitching, each round a different subset of fluorophores will be turned on, and through many interactions of

this photoactivation/photoswitching – imaging – localization – photobleaching cycle, the majority of molecules constituting a larger biological superstructure can be resolved with nanometer scale precision, producing a much refined image of the superstructure compared to conventional diffraction-limited microscopy.

Combing high spatial resolution and great temporal resolution in the millisecond (ms) range, single-molecule microscopy also provides a valuable tool to study the motion and dynamics of molecules of interest. As it is possible to visualize only a few molecules per cell at a time, it is possible to track the motion of fluorophores with reasonable photostability over a prolonged period of time to watch how they travel and navigate inside the cells[13,14]. The single-molecule trajectory and the rate of diffusion can provide important information regarding the function of the molecules. Although the irreversible photobleaching of fluorophores will eventually limit the length of trajectories that can be obtained, together many trajectories can provide valuable information about the molecules such as the presence of heterogeneous subpopulations of within a cell or across cells. Also, with time-lapse imaging, it is possible to observe the behavior of single molecules over a much longer period of time[9,15]. In addition, as biomolecules do not exist isolated from other molecules in the cell, two-color imaging has been implemented to study interactions between two different molecules in real time[16], providing direct evidence depicting the functional and dynamic relationship between multiple entities.

1.5 Single-molecule fluorescent probes

As biological molecules of interest cannot generally be directly detected in a fluorescence microscope, labelling these molecules with fluorescent tags is the standard way for visualization. Various labeling schemes involving quantum dots and small organic dyes have been implemented, but these fluorescent tags are not the optimal for live cell imaging due to difficulties in transporting them into the cells and targeting them to molecules of interest with high specificity. Therefore, the primary method to achieve high specificity *in vivo* imaging has been genetically expressing fluorescent proteins along with the targeted molecules.

Several properties have to be considered when choosing fluorescent proteins for single-molecule imaging[17]. First, because the localization precision depends strongly on the brightness of the fluorophore, fluorescent proteins with high quantum yield are greatly preferred over ones with poor quantum yield. Second, in PALM and similar experiments, the on-off-switching rate ratio limits how densely the sample can be labelled. If too many fluorophores are turned on spontaneously or by the imaging light, one has to lower the labelling density to avoid overlapping PSFs, which according to the Nyquist criterion will in turn adversely affect the resolution. Third, in single-particle tracking experiments, fluorophores that are more photostable are preferred because they allow for more time points to be acquired for each trajectory before undergoing photobleaching. Fourth, the tendency of dimerization, as exhibited by many known fluorophores, has to be carefully evaluated to determine if they are suitable for the specific types of applications[17]. As a general rule, for a given fluorophore, a higher copy number in the cell increases the likelihood that dimerization will cause artificial aggregates which do not reflect the true behavior of the tagged molecules. Fifth, the emission wavelengths ($\lambda_{emission}$) of fluorophores significantly affect the signal to noise ratio of images in single-molecule live cell microscopy. As

cellular autofluorescence is typically much stronger in the shorter wavelength range[18], fluorophores with longer emission wavelengths, such as those in the red and far-red spectral regions are more ideal because of better spectral separation between the signal and the background light. Finally, counting molecule copy numbers using fluorescence microscopy has remained to be a challenge in the field[19], and it is even more difficult to count the number of fluorescent molecules when the signaling efficiency, defined as the ratio of detectable molecules per cell and the number of fluorophores that are expressed, of the fluorophore used is low or unknown. The signaling efficiency is greatly affected by a number of processes such as protein folding, protein maturation and photoactivation/photoswitching efficiency. The names and properties of fluorophores used in this thesis are presented in Table 1.1 below.

Table 1.1 Properties of fluorophores used in this study.

Values may vary depending on specific experimental conditions.

Fluorophore	**$\lambda_{activation}$ (nm)**	**$\lambda_{excitation}$ (nm)**	**$\lambda_{emission}$ (nm)**	**Quantum Yield**	**Switching Rate (ms^{-1})**	**Signaling Efficiency**
Nile red[20]	N/A	540-580	615-665[a]	>0.7[a]	N/A	N/A
mCherry[18]	N/A	587	610	0.22	*unknown*	*unknown*
PAmCherry1[17,21,22]	405	564	595	0.46	5.13×10^{-7}	3.6-80%
mCitrine[18]	N/A	516	529	0.76	*unknown*	*unknown*

[a] The photophysical properties of Nile red are highly solvent-dependent. For instance, the maximum emission wavelength of Nile red is significantly blue-shifted in hydrophobic environment.

1.6 *Bacillus subtilis*

Bacillus subtilis is the model organism that I study in Chapters 3 and 4 of this thesis. *B. subtilis* cells are rod-shaped, Gram-positive bacteria generally found in the soil. They have a circular chromosome with a size around 4.2 Mbp[23]. Being a highly competent species, *B. subtilis* cells can readily take up foreign DNA molecules and incorporate them into their own genome through recombination, which not only helps them to adapt to various harsh environments, but also make it relatively easy for researchers to manipulate their genetics and to create mutant types for dissecting mechanisms of various biochemical pathways, as will be demonstrated in this thesis. Given its powerful genetics, high homology with eukaryotic DNA replication and repair pathways, and fast growth rate (doubling time in rich growth media is about 20 minutes), *B. subtilis* provides an ideal platform to systematically study the DNA polymerase[24,25] and repair proteins *in vivo*[26].

In this thesis, I carried out single-molecule imaging experiments in *B. subtilis* to address two open questions central to genome maintenance: (1) how is a DNA mismatch detected efficiently in a live cell? And (2) how does DNA polymerase behave during DNA replication?

1.7 Thesis outline

In this thesis, I combine live-cell single-molecule super-resolution microscopy, single-particle tracking and genomics to answer important biological questions about DNA replication and repair.

In Chapter 2, I outline several image and data analysis techniques that are extensively applied in the subsequent chapters, including cell segmentation, image registration, trajectory construction and diffusion analysis. Additional analysis methods specifically applied in each project will be covered in the relevant chapter. In Chapter 3, I peer into live *B. subtilis* cells to investigate the *in vivo* behavior of the DNA mismatch repair protein MutS. I quantitatively characterize the distribution and motion of MutS with respect to the replisome, and demonstrate an intimate, dynamic and highly conserved coupling between MutS and the site of DNA replication, which helps MutS to target mismatched base pairs in a highly efficient manner. In Chapter 4, I focus on DNA replication itself. I study one of the two essential DNA polymerases found in *B. subtilis*, PolC. I use PolC to map the sites of active DNA replication in live cells, and I employ photobleaching-assisted microscopy and time-lapse imaging to uncover the stoichiometry and dynamics of PolC at each replication site. In Chapter 5, I extend super-resolution microscopy to visualize nanometer-scale dynamics in nonliving nanomaterials. I track and quantify the motion of individual organic dyes diffusing inside meso- and nanoporous crystalline materials designed for energy and molecular storage and transport. We end this thesis in Chapter 6 with a summary of our findings from previous chapters as well as a discussion about the future research directions that our current research results may lead to.

References

1 Moerner, W. E. & Kador, L. Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.* **62**, 2535-2538 (1989).

2 Airy, G. B. On the diffraction of an object-glass with circular aperture. *Transactions of the Cambridge Philosophical Society* **5**, 283-291 (1835).

3 Abbe, E. Beiträge zur Theori des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für Mikroskopische Anatomie* **9**, 413-468 (1873).

4 Tuson, H. H. & Biteen, J. S. Unveiling the inner workings of live bacteria using super-resolution microscopy. *Anal. Chem.* **87**, 42-63 (2015).

5 Moerner, W. E. & Kador, L. Optical-Detection and Spectroscopy of Single Molecules in a Solid. *Phys. Rev. Lett.* **62**, 2535-2538 (1989).

6 Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods* **3**, 793-795 (2006).

7 Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645 (2006).

8 Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258-4272 (2006).

9 Biteen, J. S. *et al.* Super-resolution imaging in live Caulobacter crescentus cells using photoswitchable EYFP. *Nature methods* **5**, 947-949 (2008).

10 Moerner, W. E. Single-molecule mountains yield nanoscale cell images. *Nature methods* **3**, 781-782 (2006).

11 Kubitscheck, U., Kuckmann, O., Kues, T. & Peters, R. Imaging and tracking of single GFP molecules in solution. *Biophys. J.* **78**, 2170-2179 (2000).

12 Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775-2783 (2002).

13 Manley, S. *et al.* High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature methods* **5**, 155-157 (2008).

14 Kusumi, A., Tsunoyama, T. A., Hirosawa, K. M., Kasai, R. S. & Fujiwara, T. K. Tracking single molecules at work in living cells. *Nat. Chem. Biol.* **10**, 524-532 (2014).

15 Gebhardt, J. C. *et al.* Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature methods* **10**, 421-426 (2013).

16 Veatch, S. L. *et al.* Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PloS one* **7**, e31457 (2012).

17 Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8452-8457 (2014).

18 Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nature methods* **2**, 905-909 (2005).

19 Rollins, G. C., Shin, J. Y., Bustamante, C. & Presse, S. Stochastic approach to the molecular counting problem in superresolution microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E110-118 (2015).

20 Sackett, D. L. & Wolff, J. Nile red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces. *Anal. Biochem.* **167**, 228-234 (1987).

21 Durisic, N., Laparra-Cuervo, L., Sandoval-Alvarez, A., Borbely, J. S. & Lakadamyali, M. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nature methods* **11**, 156-162 (2014).

22 Subach, F. V. *et al.* Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nature methods* **6**, 153-159 (2009).
23 Kunst, F. *et al.* The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**, 249-256 (1997).
24 Lemon, K. P. Localization of Bacterial DNA Polymerase: Evidence for a Factory Model of Replication. *Science* **282**, 1516-1519 (1998).
25 Lemon, K. P. & Grossman, A. D. The extrusion-capture model for chromosome partitioning in bacteria. *Genes Dev.* **15**, 2031-2041 (2001).
26 Smith, B. T., Grossman, A. D. & Walker, G. C. Visualization of mismatch repair in bacterial cells. *Mol. Cell* **8**, 1197-1206 (2001).

Chapter 2 Obtaining Information from Single-Molecule Fluorescence Images

2.1 Introduction

After single-molecule fluorescence movie frames have been recorded, a two-stage post-processing routine, involving an intensity peak guessing step and a point spread function (PSF) fitting step, is carried out to extract single-molecule positions from the raw images that are inevitably noisy. Then, single-particle tracking (SPT) algorithms are implemented to construct trajectories from an array of single-molecule positions identified during the previous step. For cellular imaging, an extra white-light transmission image or a phase-contrast image of the corresponding cells can be tremendously valuable in rejecting spurious single-molecule peaks and trajectories. Finally, a variety of analyses can be done to obtain biologically relevant information from single-molecule trajectories.

There exist many different ways to approach each step of the post-processing routines mentioned above, some work better in a specific scenario than others. In this chapter, we introduce several image and trajectory analysis techniques that are extensively and specifically implemented for the work described in this thesis.

2.2 Point spread function peak guessing and fitting

To identify putative single-molecule signals from raw fluorescence images, we first band pass filter[1] the raw images to produce identify pixels likely corresponding to the peaks of single-molecule PSFs (Figure 2.1). First, the raw image (Figure 2.1 ①) is convolved with a low pass filter constructed from a Gaussian kernel whose frequency response rejects noise of higher frequencies such as the pixel noise (②). In addition, the image is convolved with a second low pass filter made from a boxcar function (③). The band pass filtering is achieved by taking the difference between the two low-pass-filtered images, leaving pixels constituting putative PSFs much brighter than the rest of the image (④). Pixels with intensities less than a certain threshold are set to zero, and the rest of nonzero regions are shrunk to single pixels that will serve as the center for the subsequent PSF fitting (⑤).

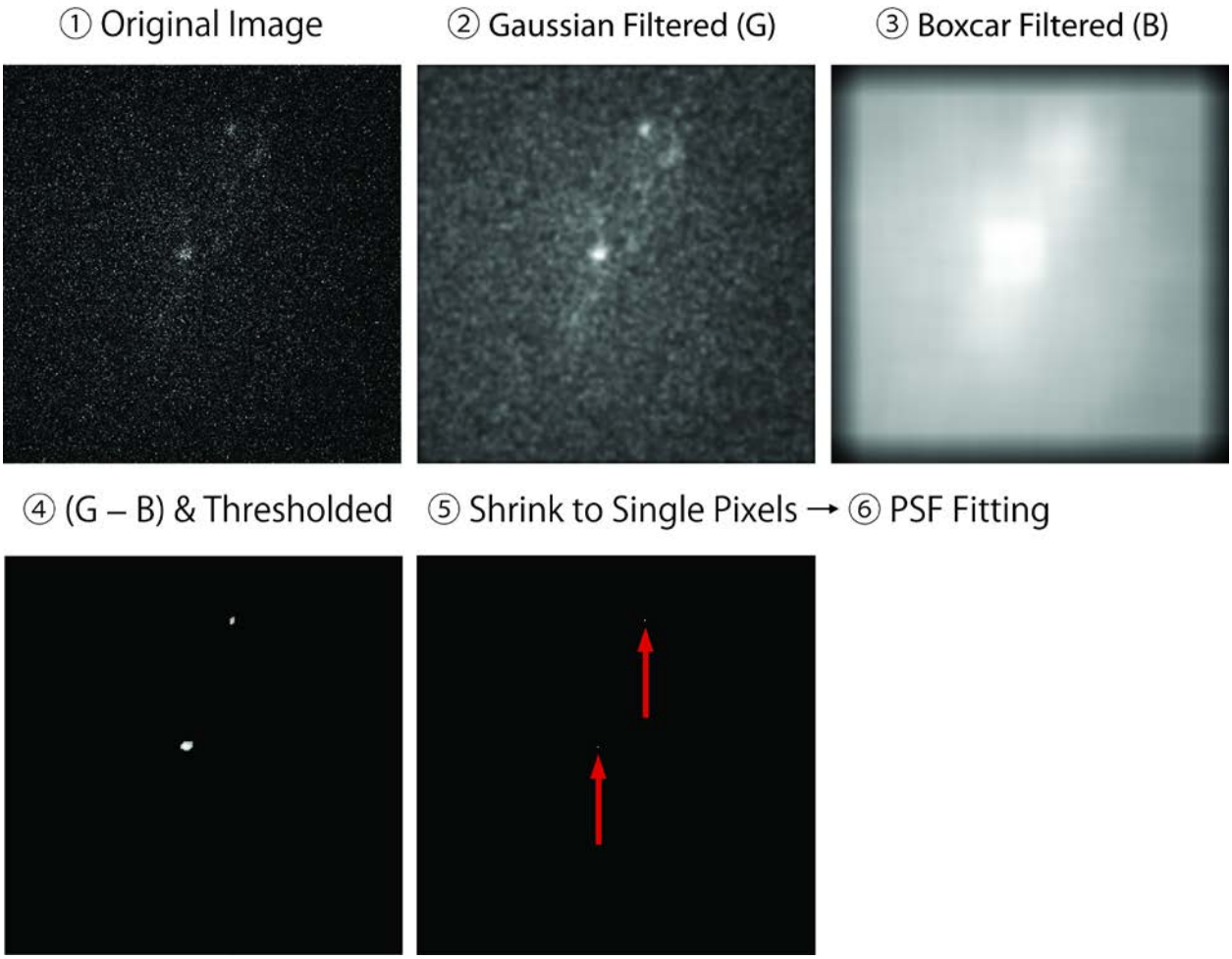


Figure 2.1 Data analysis procedures for peak guessing. Red arrows point to single pixels corresponding to guessed single-molecule locations. Starting from the locations in ⑤, PSF fitting will be performed over the original unfiltered intensities of these pixels and their surrounding pixels to determine whether the PSF shape and size are reasonable and to localize the position of single-molecule emitters with nanometer precision.

Once these candidate molecule center positions have been identified, PSF fitting will be performed over specified regions of the raw data centered on the candidate pixels to obtain the precise and sub-pixel single-molecule positions[2]. Based on fitting results, only PSFs with a positive amplitude and a fit width smaller than the diffraction limit will be considered as single molecules and recorded for further analysis.

2.3 Single-particle tracking

Single-molecule positions can be linked temporally to produce single-molecule trajectories, which reveal important information regarding the dynamics of the molecule under investigation. A number of factors affect the complexity and the amount of computation needed to perform single-particle tracking. For example, the stochastic nature of photoactivation, photoswitching, photobleaching and blinking of fluorophores can give rise to imaging frames with different numbers of molecules, and blinking and photobleaching mean that every molecule from one frame cannot necessarily be paired with another in the subsequent frame. Furthermore, when multiple molecules come close to each other (i.e., within the same diffraction-limited area), assigning trajectories to these "crossing" molecules can be especially challenging. Various algorithms, from the simple nearest-neighbor method to the more involved Bayesian approach[3], have been implemented to accommodate scenarios of different levels of complexity. Here, we construct single-molecule trajectories by globally and simultaneously optimizing all possible pairings of molecules between consecutive frames[4], rather than simply sequentially pairing molecules based on spatial proximity. For every possible pairing, we assign a merit value *m* to describe the likelihood that the two particles are actually the same molecule observed at different time points. The *m* is defined as:

$$m = e^{-\alpha\Delta d} \cdot e^{-\beta\Delta I} \cdot e^{-\gamma\Delta t} \quad (2.1)$$

where Δd, ΔI and Δt are respectively the spatial separation, the intensity difference and the temporal separation between the two particles, and α, β and γ are coefficients specifying how much penalty is imposed by each of the displacement, the change in intensity and the temporary disappearance of the molecule (blinking) when the two particles are paired to form a segment of a trajectory. Candidate segments having merit values smaller than a specified minimum m value will be rejected. Among the remaining pairs, the set that collectively maximizes the sum of merits will be chosen, and these segments will either be appended to existing trajectories or start new ones. This global optimization is in essence a classical assignment problem and is carried out using the Hungarian algorithm[5,6], which maximizes the sum of merits in a single step while guarantying a particle is only used once at most. Unpaired molecules will be kept in a pool of candidates for the next round of pairing. The maximization of sum of merits is done for each set of adjacent frames and is the repeated until all frames have been considered.

2.4 Cell segmentation

Knowing where cell boundaries are can be advantageous when analyzing intracellular fluorescence images and single-molecule data. For example, spurious single-molecule localizations that appear outside cell boundaries can be readily rejected, and implementation of single-particle tracking can be constrained within each cell to prevent single-molecule trajectories from unphysical crossing into neighboring cells. We used a custom-built cell segmentation program partially based on the "valley" detection algorithm described by Sliusarenko et al[7] to generate a digital "phase mask" for each phase contrast cell image. These phase masks have the same dimensions as the phase contrast microscopy images from which

they are produced, but only pixels corresponding to cell bodieshave nonzero intensity values. Because each cell-containing pixel region has a unique intensity value, cells can be easily indexed and processed on the single-cell basis. The MATLAB script for this program is in Appendix A.2 of this thesis, and below we outline the major steps of code; sample outputs from these steps are shown in Figure 2.2.

First, the phase contrast image of cells (Figure 2.2a) is inverted such that cells appear brighter and the background appears darker (Figure 2.2b). Due to a number of complications, such as poor image contrast, cell crowding and uneven illumination, it is challenging to perform cell segmentation relying on a single type of algorithm. Therefore, the inverted phase contrast image is subjected to a three-way post-processing, each based on a different algorithm, and the results are combined to generate the final phase mask:

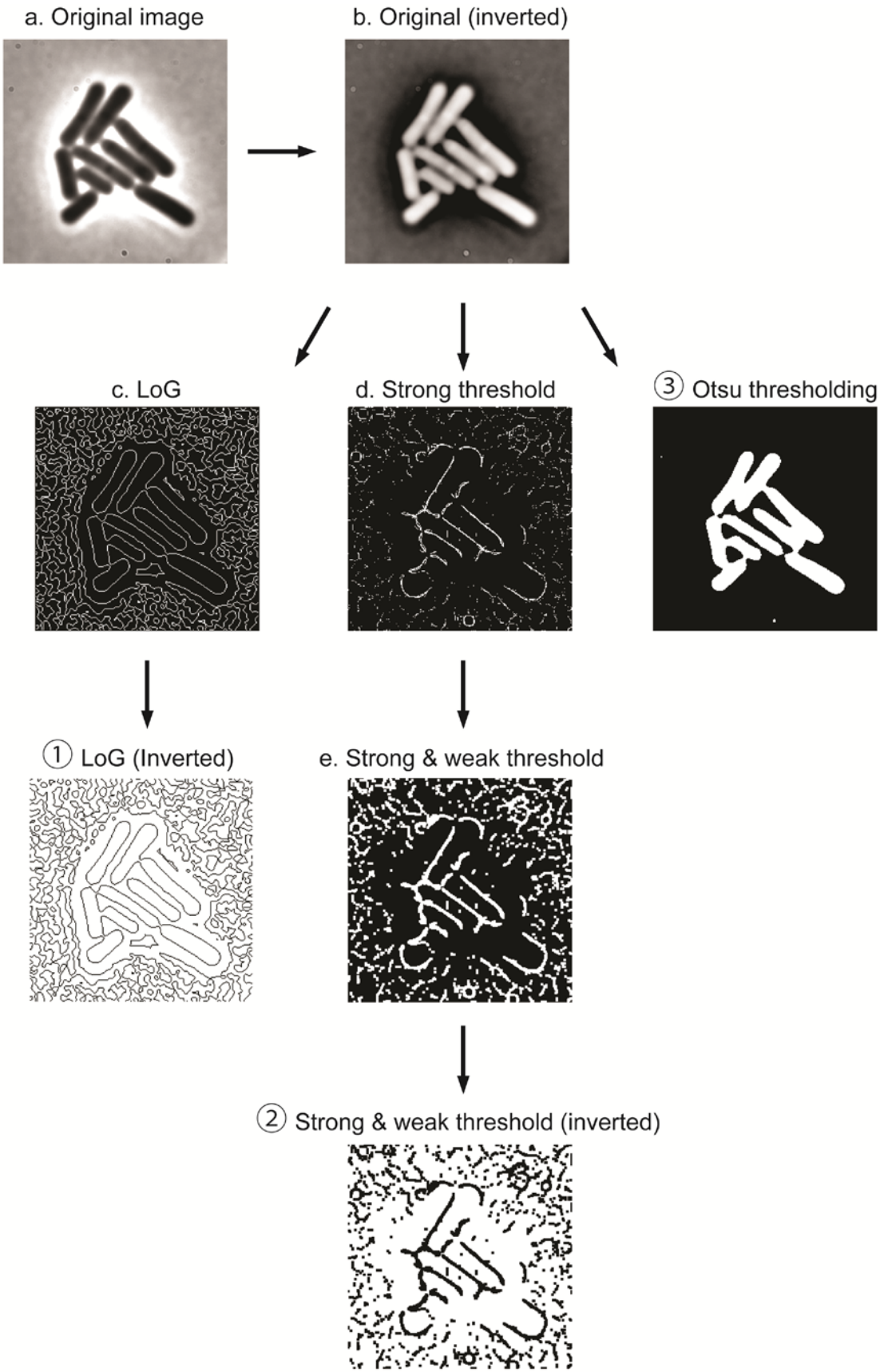
Algorithm ① (Laplacian of Gaussian algorithm): The Laplacian of Gaussian (LoG) algorithm is applied to identify cell edges from the inverted phase contrast image (Figure 2.2c). The LoG operation is a two-step process where the image is first smoothed with a Gaussian filter before being convolved with a Laplacian filter, which is a derivative filter and is therefore highly sensitive to local intensity changes such as cell edges. The Gaussian filtering step is necessary to prevent noise, which produces rapid and localized intensity changes, from being picked up as cell edges by the Laplacian filter.

Algorithm ② (valley algorithm[7]): The inverted phase contrast image is smoothed with a mean filter, and the valley algorithm is then applied to the resulting image to determine the valley depth for each pixel in a three-step procedure: first, the intensity change between the

center pixel and each adjacent pixel is calculated. This produces 8 intensity difference values (from 8 neighboring pixels) along 4 directions (horizontal, vertical and the two diagonals). Second, the smallest intensity difference value along each direction is chosen to represent the intensity change in that particular direction. Third, the valley depth for the center pixel is determined as the maximal intensity change amongst all 4 possible directions. Pixels with valley depths above a specified strong threshold are considered as valley pixels (Figure 2.2d), and pixels with valley depths above the weak threshold but below the strong threshold are treated as valley pixels only if they are 8-connected to at least one pixel with valley depth above the strong threshold. If a pixel satisfies either criterion, it is regarded as a valley pixel (Figure 2.2e).

Algorithm ③ (Otsu's method[8]): The intensity distribution of an inverted phase contrast image generally has bi-modal characteristics. Otsu's method examines the intensity histogram and determines a single best intensity value that minimizes the intra-class variance of the foreground (cells) and background pixel intensities. This value will serve as an intensity threshold to separate cells from the background pixels.

Each of the post-processing algorithms described above independently produces one binary image where the foreground pixels either correspond to cell edges (LoG algorithm), cell junction (valley algorithm) or cell bodies (Otsu's method). The phase mask is obtained by combining these three binary images and assigning an index number to each foreground object. A few additional manipulations are usually done to achieve a more accurate representation of the cells. For example, foreground objects with areas too small or too big compared to the typical cell size are excluded. Image closing is performed to fill smaller holes (background pixels) appearing inside cells and to further smooth cell contours. Lastly, image dilation (typically for 1-3 pixels) is performed to produce the final phase mask.


a. Original image
b. Original (inverted)
c. LoG
d. Strong threshold
③ Otsu thresholding
① LoG (Inverted)
e. Strong & weak threshold
② Strong & weak threshold (inverted)

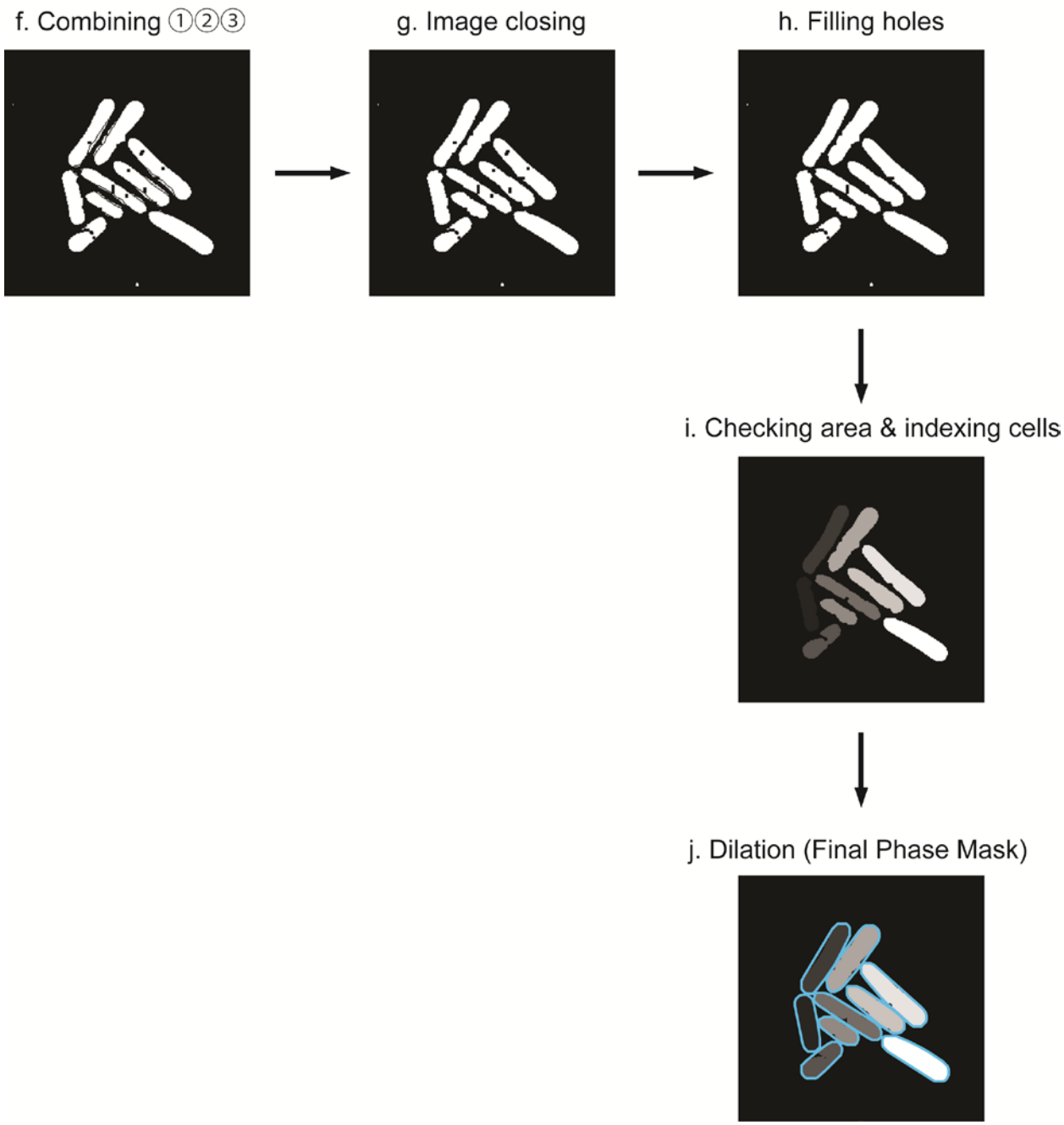


Figure 2.2 Procedures for cell segmentation. The original image is independently processed using the Laplacian of Gaussian algorithm, the valley algorithm and Otsu's method, and results from three different algorithms are combined to attain the final phase mask. Image size = 12.54 μm × 12.54 μm.

2.5 Image registration for two-color microscopy[a]

Simultaneously imaging two fluorophores of different emission colors can be achieved by placing a beam splitter between the microscope and the detector (Figure 1.3). The core components of the beam splitter (Figure 2.3) are a dichroic filter and several mirrors that split and direct light of different wavelengths onto different areas of the detector, generating two separate color channels on the detector array corresponding to the same observation region (Figure 2.4). In this thesis, we use the Cairn Research Optosplit II, but this procedure is generalizable to any setup. For an ideal optical configuration, the image from one color channel can be mapped onto the other in a linear manner involving translation, rotation and scaling only. In reality however, optical aberrations cause localized image distortions in a nonlinear fashion. We therefore employed a local weighted mean (LWM) transformation approach proposed by Churchman et al[9], which assigns localized second-order polynomials to describe the positional relations between the two color channels and achieves image registration with nanometer accuracy.

[a] Thanks to Matthew B. Stone and Professor Sarah L. Veatch for their assistance in implementing image registration.

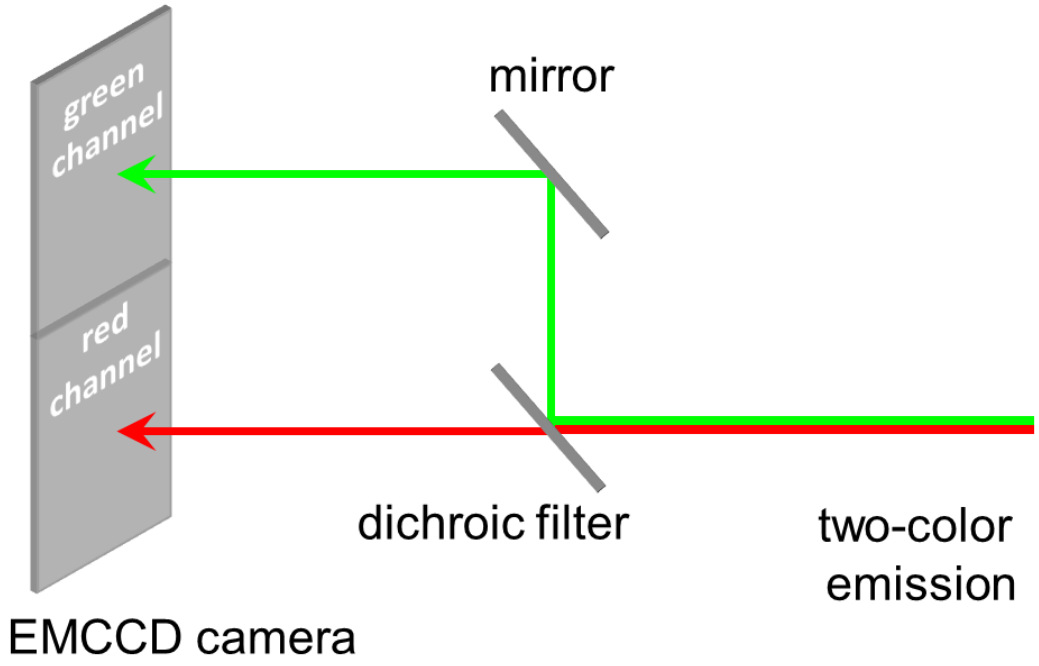


Figure 2.3 Beam splitter set-up for two-color imaging.

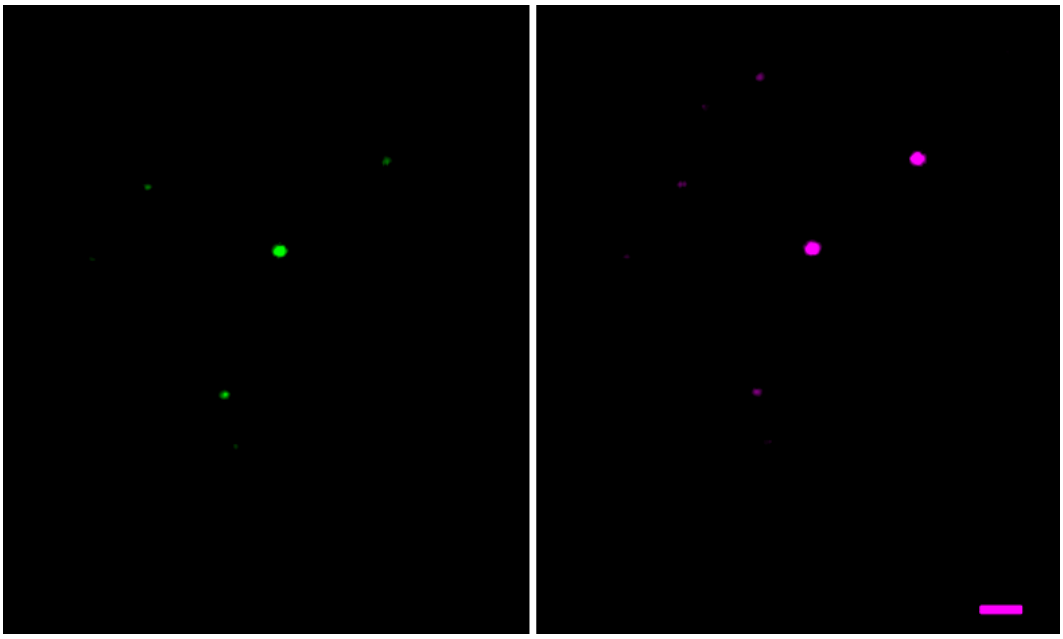


Figure 2.4 Two-color emission is projected onto two areas of the same camera. This image is taken with the 0.1 μm TetraSpeck™ microspheres that emit both green and red lights upon excitation by the 488-nm and the 561-nm lasers. Although both channels capture the same physical sample region, signals of different colors have different coordinates on the image, and nonlinear image registration is necessary to overlay the two channels with nanometer accuracy. Scale bar = 1 μm.

Specifically, for each two-color cell imaging experiment, we first generate a LWM mapping function for the set-up by imaging and localizing fiducial markers (0.1 μm TetraSpeck™ microspheres) that are sparsely dispersed within the field of view and can be simultaneously imaged using the blue (488 nm) and the green (561 nm) lasers (Figure 2.4). The commercially available fiducial markers are diluted 50x or 100x, attached to a microscope slide negatively charged through oxygen plasma etching, then immersed in glycerol and finally sandwiched with another coverslip.

The mapping function for transforming coordinates from the right channel (x, y) to the left channel (X, Y) is specified by two second-order polynomials, one for the $x \rightarrow X$ transformation and the other for the $y \rightarrow Y$ transformation (Figure 2.5):

$$[X \quad Y] = [1 \quad x \quad y \quad xy \quad x^2 \quad y^2] \begin{bmatrix} A & G \\ B & H \\ C & I \\ D & J \\ E & K \\ F & L \end{bmatrix} \tag{2.2}$$

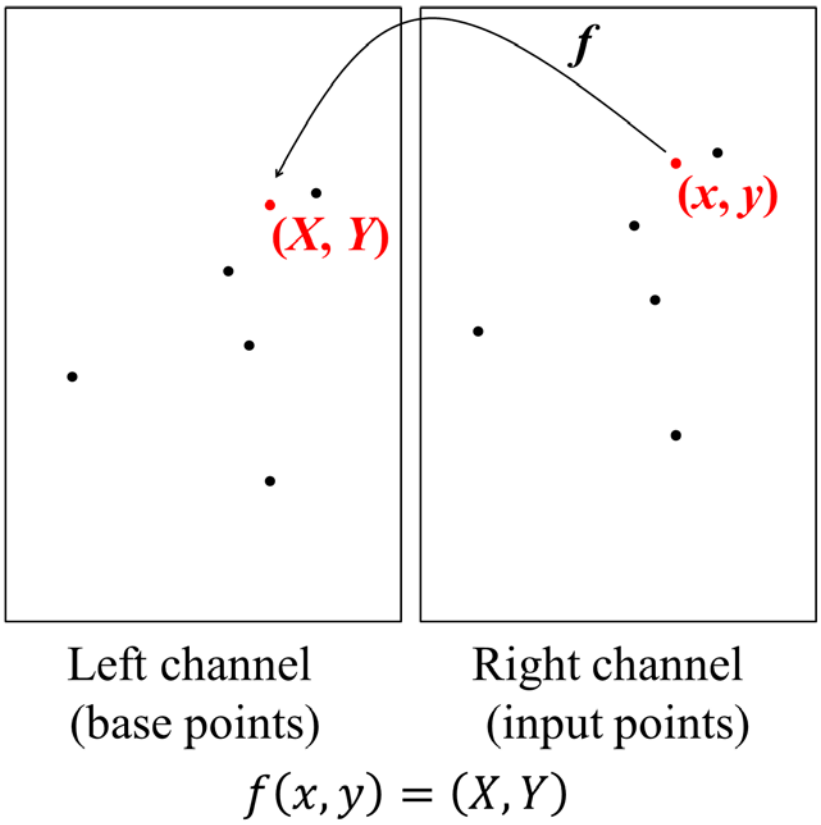


Figure 2.5 The mapping function *f* transforms coordinates from one color channel to the other. The red dot shows a control point to be mapped from the right channel to the left channel.

Because we solve for 12 mapping parameters (*A* to *L*) to be solved, at least 6 control points (pairs of fiducial marker positions from both channels) need to be imaged in each calibration frame. For each control point, these 12 mapping parameters characterize the local geometric transformation that is applicable only within a limited region termed the radius of influence, *r* (Figure 2.6):

$$r = \sqrt{(x - x_{5th})^2 + (y - y_{5th})^2} \tag{2.3}$$

where the x_{5th} and y_{5th} are the coordinates of the fifth neighbor (by ascending separation distance) of the control point under consideration.

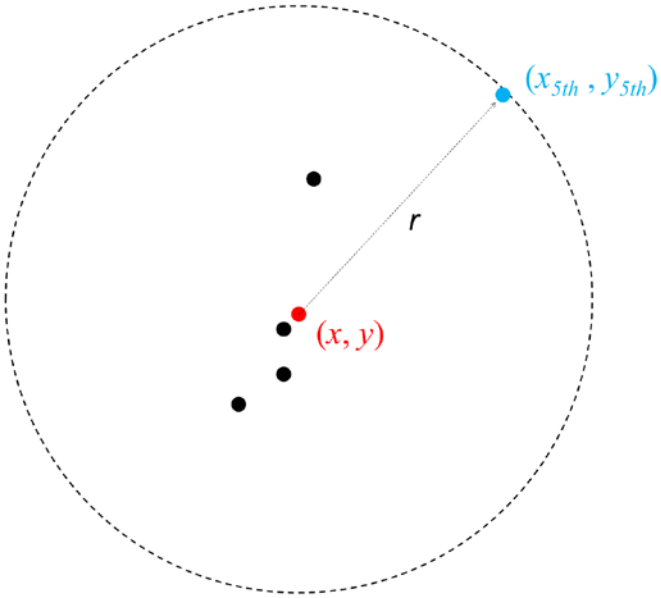


Figure 2.6 The radius of influence, *r*, dictates the size of the region over which a particular set of mapping parameters derived from one pair of control points contributes to the final mapping function *f*.

Having obtained the mapping parameters for all control points, we can calculate the mapping function *f* for any arbitrary point with a coordinate (*x*, *y*) in the right channel and a corresponding position (*X*, *Y*) in the left channel such that *f*(*x*, *y*) = (*X*, *Y*) (Figure 2.5). *f* is a weighted combination of the mapping functions of all control points for which the radius of influence encompasses the point (*x*, *y*):

$$f(x,y) = \frac{\sum_{i=1}^{N} W_i f_i(x,y)}{\sum_{i=1}^{N} W_i} \tag{2.4}$$

where the weight of each neighbor W_i at (x_i, y_i) is inversely proportional to the separation distance between the control point and the point at (*x*, *y*):

$$W_i = \begin{cases} 1 - 3R^2 + 2R^3, if 0 \le R \le 1 \\ 0, if\ R > 1 \end{cases} \tag{2.5}$$

where the normalized distance, *R*, is defined as

$$R = \frac{\sqrt{(x - x_i)^2 - (y - y_i)^2}}{R_n} \tag{2.6}$$

and R_n is the distance between the current point at (*x*, *y*) and its farthest neighboring control point that contributes a nonzero weight (Figure 2.7).

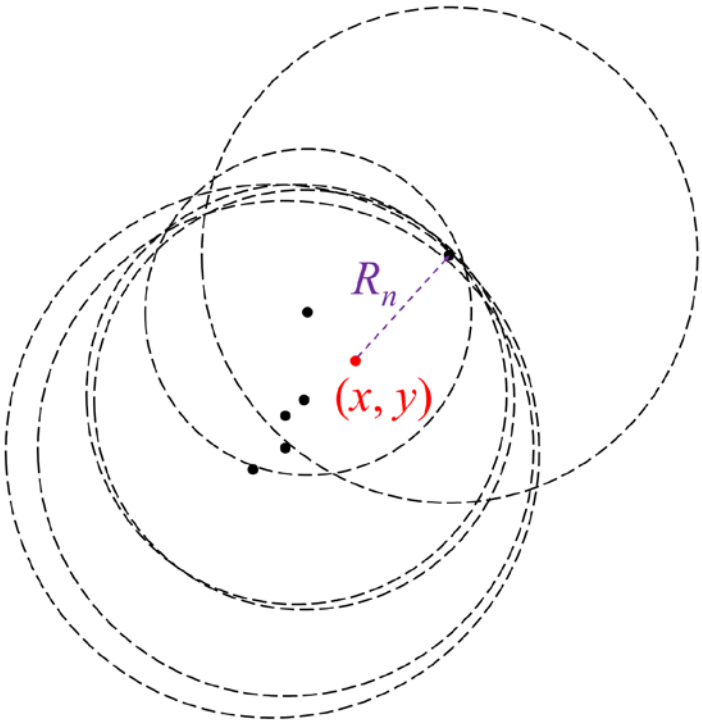


Figure 2.7 Each pair of control points has its own set of mapping parameters and radius of influence. To transform a coordinate at an arbitrary location from one color channel to the other, the mapping parameters from all neighboring control points are taken into account in a weighted manner (Equation 2.4). The weight contributed by a particular pair of control points depends on their distance from the location under consideration.

To estimate the error associated with the transformation, one can apply the mapping function, $f(x,y)$, to map control points from the right channel to the left channel and calculate the deviation of remapped coordinates from the actual coordinates in the left channel. The error is called the fiducial registration error (FRE):

$$\mathrm{FRE} = \left(\frac{1}{N} \sum_{i=1}^{N} [X_i - f(x)]^2 \right)^{1/2} \tag{2.7}$$

Another more involved but more accurate way to estimate the transformation error is to calculate the target registration error (TRE). Similar to the FRE, each control point is mapped from channel to the other. However, the LWM mapping function applied to each control point is calculated by considering all neighboring control points but itself, that is, the control point to be mapped does not contribute to its own mapping function. Repeating the transformation this way for all control points, the distance between the remapped coordinates and corresponding coordinates in the other channel is the TRE:

$$\mathrm{TRE} = \left(\frac{1}{N} \sum_{i=1}^{N} [X_i - f_i(x)]^2 \right)^{1/2} \tag{2.8}$$

In our two-color experiments the TRE is 12 nm.

2.6 Diffusion analysis based on mean squared displacement

Single-molecule trajectories can reveal behaviors of a molecule that are not directly observable using conventional fluorescence microscopy. For instance, by tracking single molecules, one can identify free diffusion, confined diffusion (e.g., transient dwelling) and stationary molecules, each characterized by its own range of diffusion coefficients. The diffusion coefficient characterizes the rate of diffusion for a particle in a solvent, and depends on the shape

and size of the particle, the viscosity of the solvent, as well as the interaction between the particle and the solvent. According to Albert Einstein, the diffusion coefficient *D* for a molecule follows the relationship:

$$D = \frac{kT}{f} \tag{2.9}$$

where *k* is the Boltzmann constant, *T* is the temperature, and *f* is the friction coefficient. In the simple case where a spherical molecule with radius of *a* diffuses in a solvent with viscosity η without any non-elastic interactions, *f* is described by Stoke's law:

$$f = 6\pi\eta a \tag{2.10}$$

The friction coefficient for molecules diffusing in live cells is difficult to obtain. Nonetheless, the diffusion coefficient can still be extracted from single-molecule trajectories. According to the Einstein-Smoluchowski relation, for a molecule engaging in Brownian motion (random walk), the molecule's mean squared displacement (MSD or $\langle r^2 \rangle$) (Figure 2.8) during a given time lapse τ is directly proportional to its diffusion coefficient[10]:

$$\langle r^2(\tau) \rangle = 2nD\tau \tag{2.11}$$

where *n* denotes the dimensionality of motion.

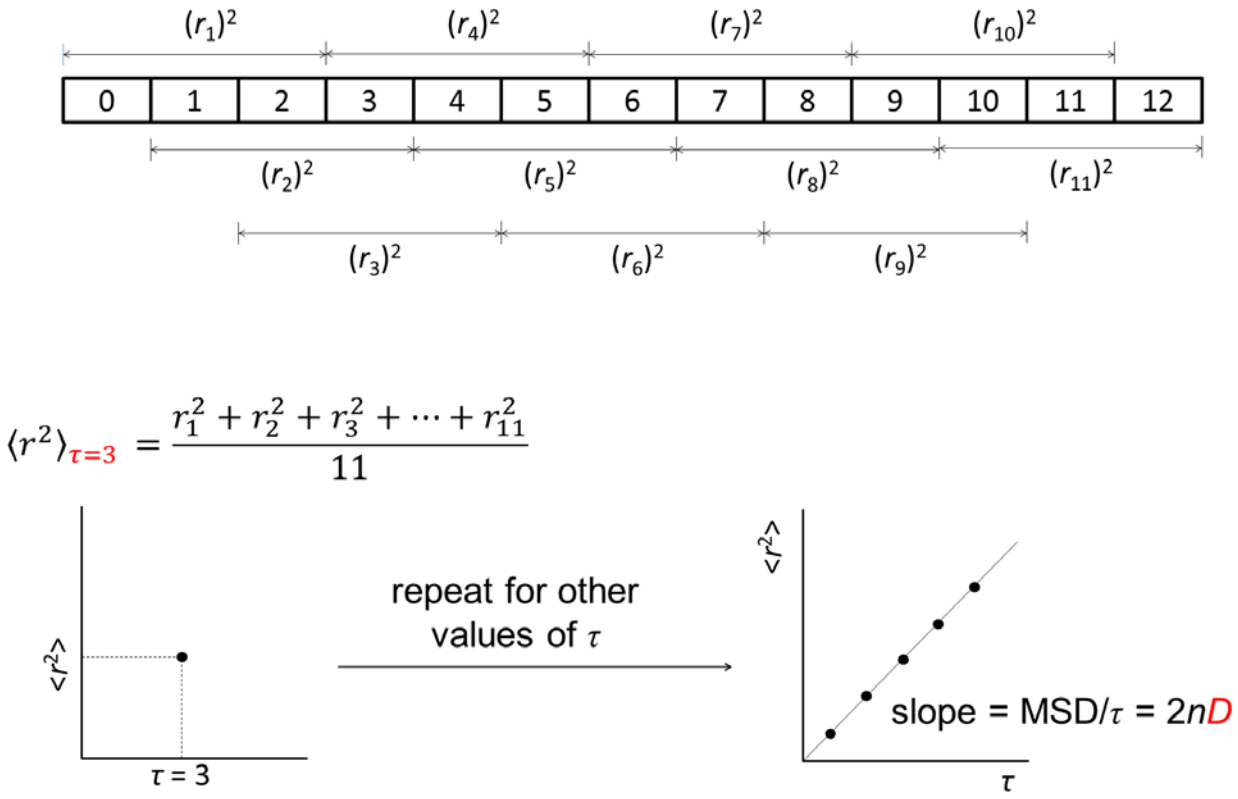


Figure 2.8 Schematic representation of the procedures for calculating MSD from a thirteen-frame movie and a time lag of τ = 3 as an example. The diffusion coefficient is obtained from the MSD vs. τ plot.

In fitting MSD vs. τ using Equation (2.11), it is recommended to use only the first quarter of the data points[11], as MSDs calculated at higher values of τ are averaged over only a few overlapping displacements and are thus more error-prone. The use of the first quarter data points is not an absolute requirement, however. A more rigorous threshold based on localization precision, estimated diffusion coefficient and camera integration time can be used to determine the number of points to be used[12,13].

The MSD vs. τ profile for a trajectory is linear only when the molecule engages in perfect Brownian motion. Anomalous diffusion such as super-diffusion resulting from active transport and sub-diffusion from macromolecular crowding will cause the MSD vs. τ relation to follow a power law[14]. Additionally, when a molecule binds and interacts with another molecule, it can undergo confined motion, in which case the MSD will plateau at larger τ values and the shape of the curve can be approximated as[15,16]:

$$\langle r^2(\tau)\rangle = \frac{L^2}{3} \cdot \left(1 - \mathrm{e}^{\frac{-12D_0\tau}{L^2}}\right) \tag{2.12}$$

where D_0 is the initial diffusion coefficient of the confined molecule and L is the size of the confinement domain.

2.7 Diffusion analysis based on cumulative probability distribution

Due to complex interactions in live cells, a molecule can exhibit different modes of motion with different diffusion coefficients. Thus, extracting a single D directly from the slope of a MSD versus τ plot is an oversimplification which might lose valuable data regarding the

heterogeneity along each trajectory[17]. To gain insight into these heterogeneous motions, we analyze the squared displacements of each particle and describe their distributions at various time lags using a corresponding diffusion model, from which multiple diffusion coefficients, each characterizing a specific mode of motion, can be obtained.

For ideal one-dimensional (1D) diffusion, the cumulative probability distribution (CPD) of the squared displacements for a given time lag τ is described by[18]:

$$P_{1D}(U,\tau) = \mathrm{erf}\left(\sqrt{\frac{U}{2(\langle r^2(\tau)\rangle + \sigma^2)}}\right) \tag{2.13}$$

where $P_{1D}(U,\tau)$ is the probability that the squared displacement r^2 during a time interval τ does not exceed some value U during the time τ. The localization accuracy σ can be experimentally determined. By fitting the CPD of the squared displacements for various τ to Equation (2.13), the MSD ($<r^2(\tau)>$) can be obtained, and the diffusion coefficient can be extracted via equation (2.11). Similarly, the CPD function for perfect 2D diffusion can be described by[19]:

$$P_{2D}(U,\tau) = 1 - e^{\frac{-U}{\langle r^2(\tau)\rangle}} \tag{1.14}$$

Beyond these cases of ideal, homogeneous motion, the power of this CPD analysis method lies in the fact that it can accommodate heterogeneous motion involving multiple diffusion coefficients by expanding Equations (2.13) and (2.14) to include several terms, each of which describes one type of motion with its own characteristic MSD and diffusion coefficient. On the other hand, because this analysis is not carried out on a trajectory-by-trajectory basis but rather the CPD is compiled by simultaneously considering all step sizes from all trajectories, the

major drawback is that spatial information is lost and it is not possible to pinpoint the location where a particular diffusion mode occurs in a cell.

Reference

References

1 Crocker, J. C. & Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.* **179**, 298-310 (1996).

2 Biteen, J. S. *et al.* Super-resolution imaging in live Caulobacter crescentus cells using photoswitchable EYFP. *Nature methods* **5**, 947-949 (2008).

3 Persson, F., Linden, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature methods* **10**, 265-269 (2013).

4 Lakadamyali, M., Rust, M. J., Babcock, H. P. & Zhuang, X. Visualizing infection of individual influenza viruses. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9280-9285 (2003).

5 Kuhn, H. W. The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly* **2**, 83-97 (1955).

6 Kuhn, H. W. Variants of the Hungarian method for assignment problems. *Naval Research Logistics Quarterly* **3**, 253-258 (1956).

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

8 Otsu, N. Threshold Selection Method from Gray-Level Histograms. *Ieee T Syst Man Cyb* **9**, 62-66 (1979).

9 Churchman, L. S., Okten, Z., Rock, R. S., Dawson, J. F. & Spudich, J. A. Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1419-1423 (2005).

10 Qian, H., Sheetz, M. P. & Elson, E. L. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical journal* **60**, 910-921 (1991).

11 Saxton, M. J. Single-particle tracking: the distribution of diffusion coefficients. *Biophys. J.* **72**, 1744-1753 (1997).

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

13 Michalet, X. & Berglund, A. J. Optimal diffusion coefficient estimation in single-particle tracking. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **85**, 061916 (2012).

14 Saxton, M. J. Anomalous diffusion due to binding: a Monte Carlo study. *Biophys. J.* **70**, 1250-1262 (1996).

15 Kusumi, A., Sako, Y. & Yamamoto, M. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophysical journal* **65**, 2021-2040 (1993).

16 Lommerse, P. H. *et al.* Single-molecule imaging of the H-ras membrane-anchor reveals domains in the cytoplasmic leaflet of the cell membrane. *Biophys. J.* **86**, 609-616 (2004).

17 McCain, K. S., Hanley, D. C. & Harris, J. M. Single-molecule fluorescence trajectories for investigating molecular transport in thin silica sol-gel films. *Anal. Chem.* **75**, 4351-4359 (2003).

18 Kirstein, J. *et al.* Exploration of nanostructured channel systems with single-molecule probes. *Nat Mater* **6**, 303-310 (2007).

19 Sonnleitner, A., Schutz, G. J. & Schmidt, T. Free Brownian motion of individual lipid molecules in biomembranes. *Biophys. J.* **77**, 2638-2642 (1999).

Chapter 3 Capturing the Dynamic Search for DNA Mismatches in Replicating Cells at the Single-Molecule Level[a, b]

3.1 Introduction: DNA mismatch repair and MutS

DNA mismatch repair (MMR) is the highly conserved process responsible for correcting DNA replication errors[1] (Figure 3.1). Although replication errors occur infrequently (~1 error per 33,000,000 base pairs)[2], the consequences of MMR failure on human health are severe[3]. The first protein involved in the MMR pathway, MutS, is responsible for detecting rare base-pairing errors and subsequently recruiting MutL, an endonuclease in *Bacillus subtilis* and eukaryotic organisms[4,5]. After MutL incision, the error-containing strand is removed and the DNA resynthesized[6].

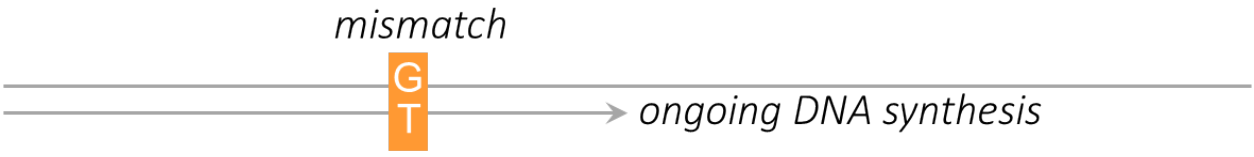


Figure 3.1 During DNA replication, the DNA polymerase can mistakenly incorporate the wrong nucleotide, causing G/T or A/C mismatches. If left uncorrected, such mistakes can lead to mutations.

[a] In collaboration with Jeremy W. Schroeder and Lyle A. Simmons. J.W.S. and L.A.S. constructed bacterial strains and performed biochemistry, molecular biology and genetic analyses. The work presented in this chapter has been submitted for publication.

[b] A List of genetic constructs used in this study can be found in Appendix 1 of this thesis.

The mechanism by which MutS homologs locate a single mismatch among millions of correctly paired nucleotides has been studied extensively. Biochemical studies have shown that MutS homologs can target a mismatch through one-dimensional scanning and sliding along DNA strands[7]. This work suggests that MutS is sufficient to identify base-pairing errors. Analysis of MutS in live cells over the past fourteen years using bulk fluorescence imaging approaches has shown that MutS forms foci that colocalize with the DNA replication machinery (replisome), and that MutS can interact with replisome processivity clamps[8-13] (Figure 3.2). However, the colocalization between MutS and the replisome is far from absolute. For example, in the absence of an exogenous mutagen, bulk microscopy detects fluorescent MutS foci in only ~9% of *B. subtilis* cells. Furthermore, only about half of these 9% of cells with a MutS focus have MutS and replisome foci that colocalize[10]. Therefore, only about 4.5% of cells under normal growth conditions in *B. subtilis* show colocalization between MutS and the replisome. Whether these foci colocalize due to active recruitment of MutS to the replisome or because mismatches are formed by the replisome and are therefore only found near the replisome has been unclear. Furthermore, the behavior of MutS in the remaining 91% of cells without a focus has been impossible to determine using ensemble fluorescence techniques.

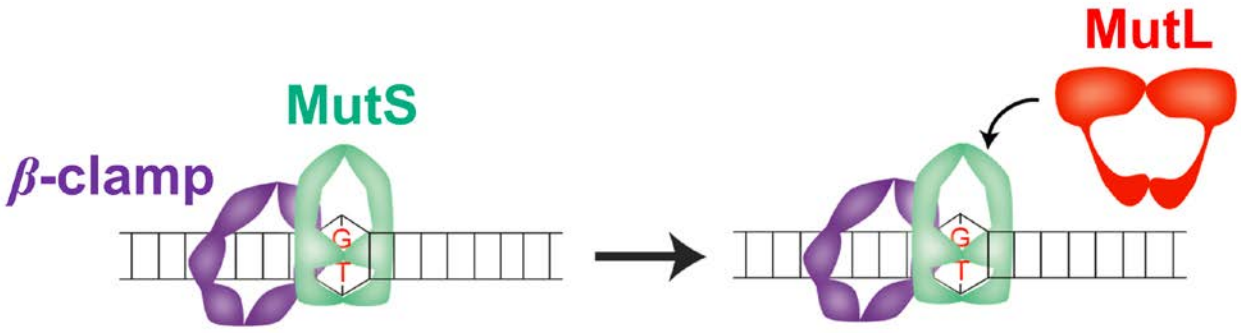


Figure 3.2 *In vivo* imaging in *Bacillus subtilis* using conventional fluorescence microscopy has suggested that MutS interacts with the replisome processivity clamps (the *β*-clamp) in some

cells. The significance of this interaction remains unclear. After MutS bind to a mismatch, MutL is recruited to form the MutS/MutL complex and initiate the downstream repair processes.

In *Saccharomyces cerevisiae*, fluorescent foci of the MutSα subunit Msh6 are observed more frequently (~80% of nuclei in cells with small buds), although a significant fluorescent signal from diffuse Msh6 is also detected. These data suggest that MutSα could be identifying mismatches through a combination of replisome-associated and replisome-independent mechanisms. In nuclei with both Msh6 and replisome foci, the foci colocalize in about 80% of cases[11,14]. In *S. cerevisiae* there exists an Exo1-dependent MMR pathway that appears only to depend weakly on replisome/MutSα interaction[11]. Mismatch repair and DNA replication are also temporally correlated in *S. cerevisiae*, as MutS homologs must be expressed during S-phase for mismatch repair to occur[14], but whether this is due to a fundamental inability of MutSα to bind mismatches in the absence of ongoing DNA replication or to other cell cycle-dependent factors such as post-translational modification of Exo1 is unknown[15].

Though ensemble fluorescence imaging has yielded important information regarding the positioning of MutS foci *in vivo* in both *B. subtilis* and *S. cerevisiae*, only more sensitive methods with higher spatial and temporal resolutions can unambiguously determine the extent to which MutS is enriched near the replisome, and answer two open questions: is this enrichment independent of mismatch binding and can MutS bind mismatches distal to the replisome *in vivo.* Ensemble fluorescence methods do not allow investigation of the behavior of individual molecules of MutS and its homologs throughout MMR, and as such, differing models have been proposed for the temporal dependence of replisome/MutSα interaction in *S. cerevisiae*. One model proposes that MutSα is positioned to detect mismatches by interactions with the replisome, while a more recent model suggests that MutSα is itself responsible for recruitment of

replisome components to the site where mismatched DNA has been detected[11,16]. A third possibility is that MutS could engage in both mechanisms under different circumstances. While most of these studies suggest that MutS is able to detect mismatches near the replisome, some suggest MutS can also detect mismatches away from the replisome[11,16]. Therefore, the timing of MMR and whether MutS must by necessity bind mismatches near the replisome still remain unclear in both *B. subtilis* and *S. cerevisiae*.

Another MMR step that is poorly understood *in vivo* is the role of MutS ATPase activity, and in particular how the ATPase cycle contributes to the movement and localization of MutS *in vivo*[17-19]. Abundant *in vitro* evidence shows that MutS carries out an ATPase cycle in which it is able to engage in mismatch searching while bound to ADP, and upon mismatch detection, MutS binds ATP and undergoes a conformational change to become a stable clamp capable of sliding in one dimension along the DNA[19-21]. In the presence of homoduplex DNA, MutS has a basal ATPase activity, which increases in response to mismatch binding[22]. Although ATPase activity is important to MutS function, its effect on the location and dynamics of MutS *in vivo* has not been studied.

In the complex cellular environment, obstacles from molecular crowding, chromosome packing and numerous DNA-binding proteins could easily impede MutS in its mismatch-searching process. To further understand the process of MMR inside cells and to gain new insight into the dynamics and location of MutS molecules *in vivo* throughout MMR, we probed the spatial and genomic distribution and dynamics of MutS in live bacteria with super-resolution fluorescence imaging[23-25], single-molecule tracking[26-28] and genomics. We investigated the effect of replisome association, active DNA synthesis, mismatch detection, ATPase activity and MutL action on MutS location and dynamics *in vivo.* In this work, we use *Bacillus subtilis* as a model

bacterial system for studying mismatch repair dynamics because, like most bacteria and all eukaryotes, *B. subtilis* lacks the *Escherichia coli* type Dam-methylation and MutH-dependent MMR system, making our results broadly applicable to MMR systems throughout biology. This is the first study to apply single molecule imaging to a dedicated DNA repair pathway in any organism. A high degree of spatial and temporal resolution allowed us to clarify the order of events during MMR and to determine the location and dynamics of single MutS proteins *in vivo*. We find that MutS is enriched near the replisome even under normal growth conditions. MutS is highly dynamic throughout the cell, with a diffusion rate that increases with separation distance from the replisome and following mismatch binding. We demonstrate that mismatch detection by MutS must occur near the site of DNA replication and that MutS remains associated with mismatches until MutL is recruited, when the DNA moves away from the replisome as DNA synthesis continues. Finally, we show that MutS physically associates with newly replicated DNA and we uncover interactions between MutS and the two essential DNA polymerase catalytic subunits PolC and DnaE, providing evidence that MutS interaction with the replisome is not restricted to processivity clamps.

3.2 Localization and dynamics of MutS in live *B. subtilis*

We constructed *B. subtilis* strains natively expressing MutS fused to the photoactivatable fluorescent protein PAmCherry1[29] as the sole source of MutS (Figure 3.3). MutS-PAmCherry retains MMR activity and the fusions are stable against proteolytic loss of PAmCherry (Figure 3.4).

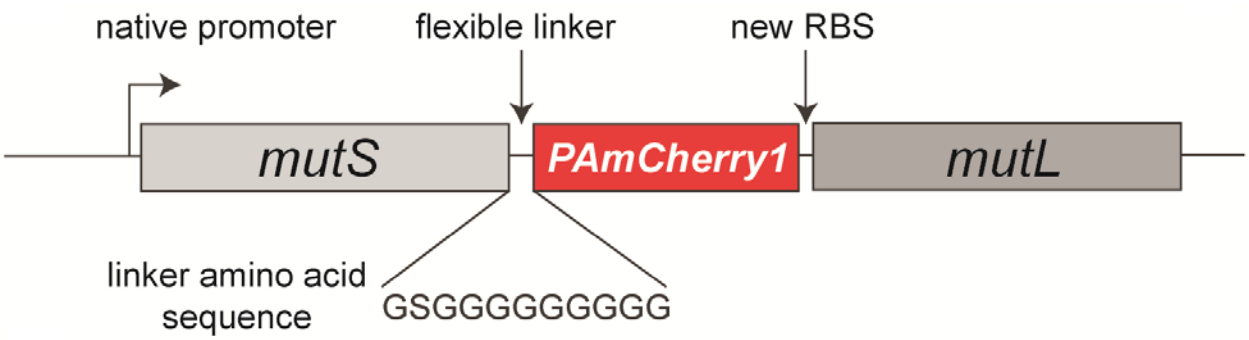


Figure 3.3 Labeling scheme for MutS-PAmCherry. RBS: ribosome binding site.

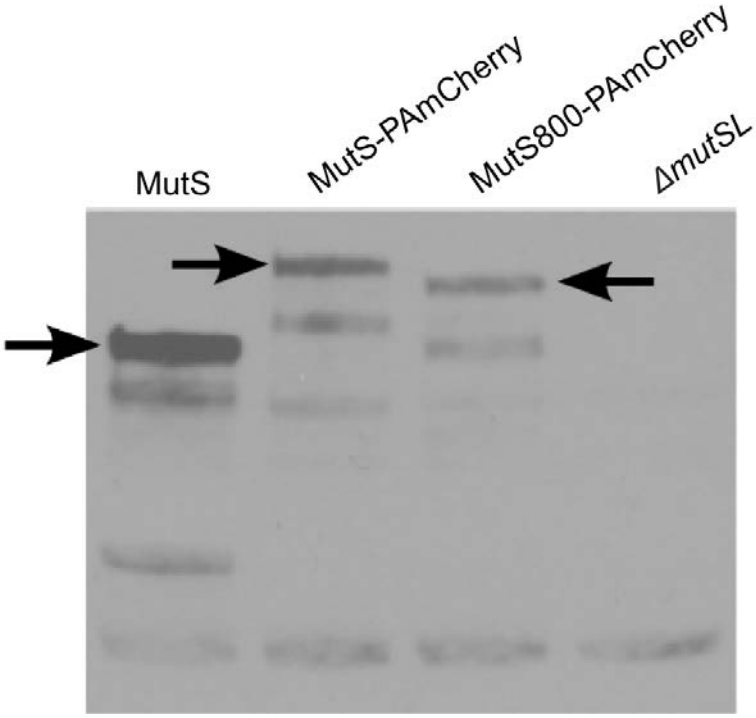


Figure 3.4 Western blot using antiserum directed against MutS. Full-length proteins are indicated by arrows. Truncated protein fragments are due to nonspecific proteolysis, and not due to cleavage of the PAmCherry tag from MutS, as they appear in the wild-type MutS lane as well as the MutS-PAmCherry and MutS800-PAmCherry lanes.

For microscopy, cells were grown at 30 ^{o}C in $S7_{50}$ minimal medium with starting $OD_{600\ nm}$ ~0.1, and were harvested during exponential phase when OD reached ~0.55 – 0.65. 2 μL of cell culture was pipetted onto a 1% agarose in S750 pad, which was sandwiched between two

coverslips that had been cleaned by oxygen plasma for 20 minutes (Figure 3.5). The sample was then mounted on the microscope objective for imaging. The widefield single-molecule epifluorescence microscopy was performed on an Olympus IX71 inverted microscope. Fluorescence emission was collected by a 1.40-NA 100x oil-immersion phase-contrast objective and detected on a 512x512 pixel Photometrics Evolve EMCCD at a rate of 25 Hz.

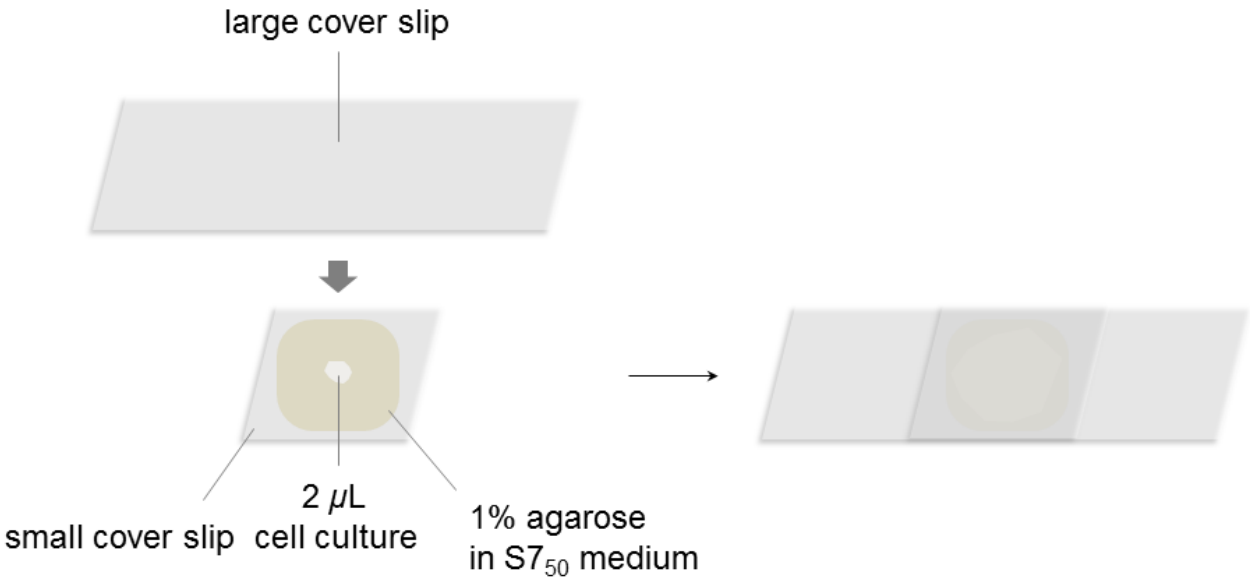


Figure 3.5 Sample preparation for single-molecule microscopy in live cells.

To investigate the dynamics of each MutS copy with sub-diffraction resolution, a 405-nm laser was used for 200 ms at low power (6.6×10^{-5} mW/cm^2) to produce 1–3 copies of emissive PAmCherry per cell at a time (Figure 3.6). The MutS-PAmCherry copies in this photoactivated subset were imaged and tracked in real time using a 561-nm laser at 1.0×10^{-4} mW/cm^2 until the PAmCherry molecules photobleached. Through 20–30 iterations of this photoactivated localization microscopy (PALM) activation/imaging/photobleaching cycle, super-resolution images were constructed, and multiple single-molecule trajectories were recorded for each cell[30-33]. We determined the localization precision for MutS-PAmCherry by imaging individual

PAmCherry1 proteins immobilized on a coverslip with polymethyl methacrylate (PMMA) under the same imaging conditions as in live-cell experiments[34].The full-width-at-half-maximum (FWHM) of the localization distribution gives a localization precision of 35 nm.

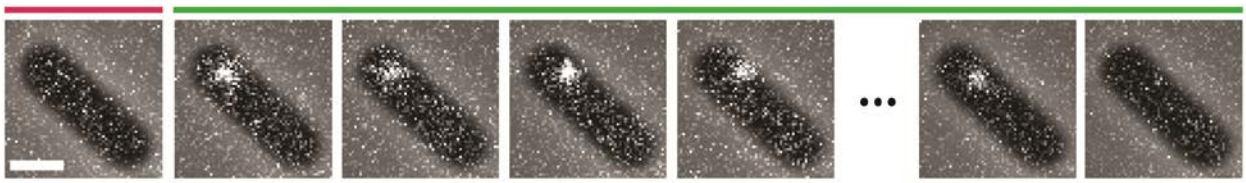


Figure 3.6 Representative frames showing the photo-activation of a single copy of MutS-PAmCherry in a cell. Purple and green lines above the frames correspond to the photo-activation laser pulse and the imaging laser. Scale bar: 1 μm.

In live *B. subtilis* cells expressing MutS-PAmCherry, we observe that MutS explores the entire cell while also experiencing significant enrichment and confinement near the cell center or cell quarter positions (Figure 3.7), where the replisome was suggested to reside by previous studies[35,36].

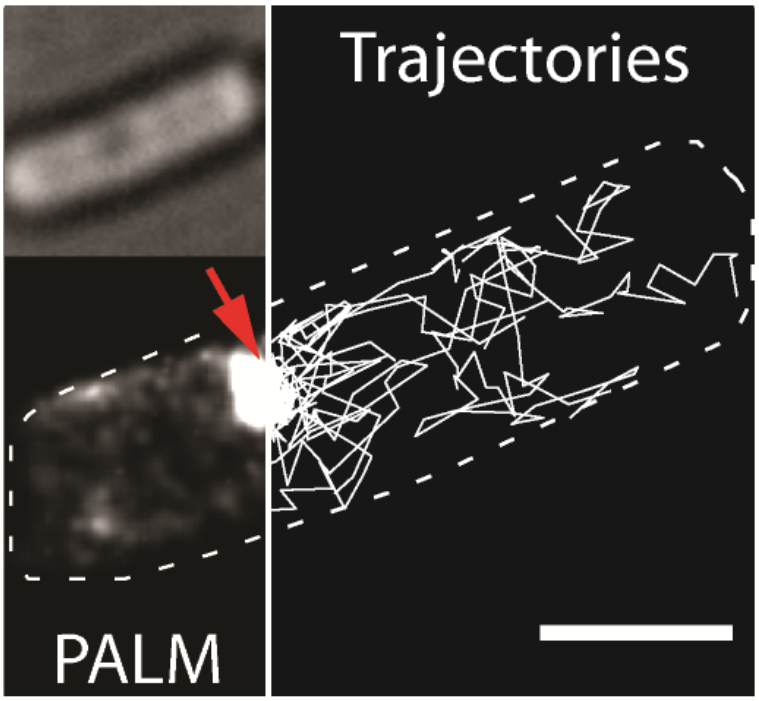


Figure 3.7 PALM reconstruction (lower left) and single-molecule trajectories (right) of MutS-PAmCherry in a live *B. subtilis* cell (upper left). Each sub-diffraction-limited coordinate of MutS-PAmCherry is plotted in the PALM image as a Gaussian blur with width equal to its localization uncertainty. Red arrow: region of MutS accumulation. White dashed lines: computer-detected cell boundary. Scale bar: 1 μm.

3.3 Localization and dynamics of the DNA replication machinery in live *B. subtilis*

To visualize the replisome position as a control, and to probe whether the sites of MMR and of DNA replication coincide, we labeled DnaX, part of the processivity clamp loader and a proxy for the replisome, with the yellow fluorescent protein mCitrine (Figure 3.8).

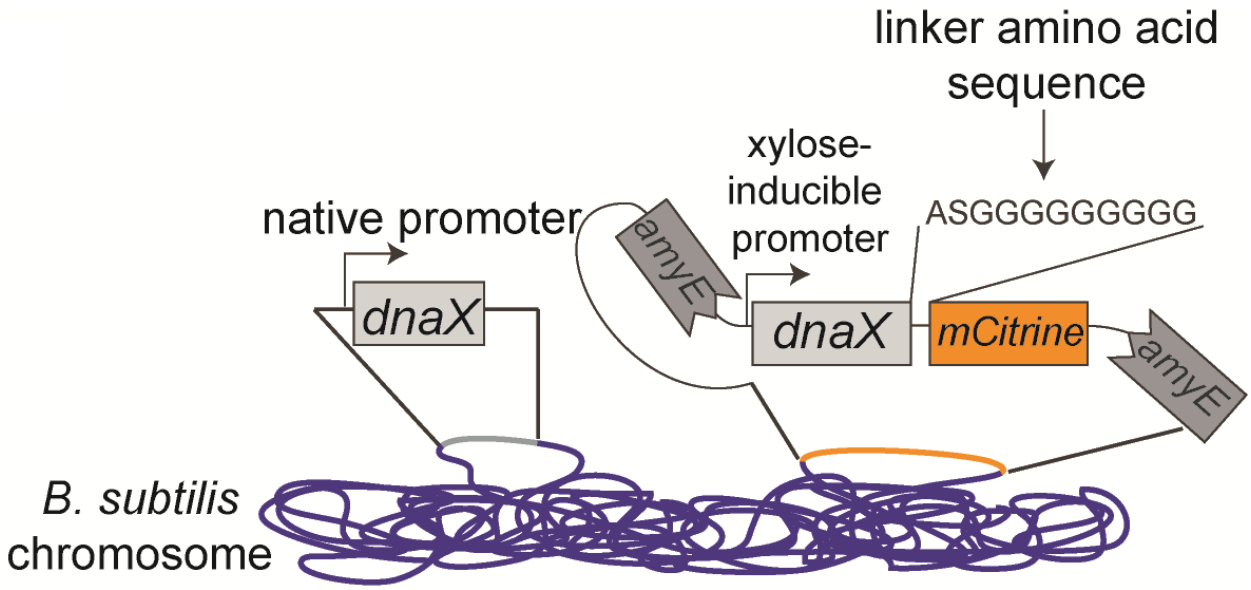


Figure 3.8 Labeling scheme for DnaX-mCitrine.

DnaX-mCitrine was imaged under 2.1×10^{-4} mW/cm^2 488-nm laser illumination with a localization precision of 41 nm. Consistent with earlier studies, we found that DnaX forms clusters either at the mid-cell or at the quarter-cell positions (Figure 3.9).

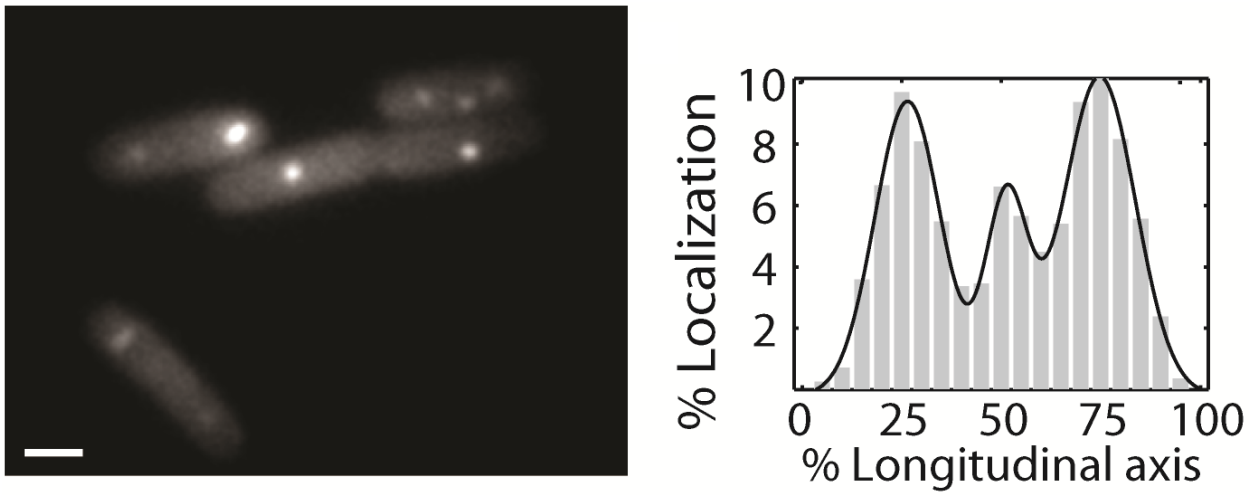


Figure 3.9 Sample fluorescence image of DnaX-mCitrine in five cells (left) and the localization probability of DnaX in 161 cells along the longitudinal cell axis (right). The replisome appears most frequently at the quarter positions in log-phase cells. Scale bar: 1 μm.

Because each replisome contains multiple copies[37] of DnaX-mCitrine with overlapping fluorescence signals (Figure 3.9), and because existing algorithms designed to extract single-molecule positions from images of densely populated fluorophores are not capable of accurately recovering positions if multiple emitters are separated from each other by less than 100 nm[38,39], we used a photobleaching-assisted approach[40,41] to achieve single-molecule localization of DnaX (Figure 3.10): upon identifying mCitrine molecules that photobleach from one imaging frame to the next[42], the mean of the frames after photobleaching is subtracted from the mean of the frames before photobleaching to produce the point spread function (PSF) of the photobleached molecule. The position of this PSF is subsequently determined from a fit to a 2D Gaussian function. Photobleaching-assisted microscopy was performed with a higher power density of 2.0×10^{-3} mW/cm^2 and a faster frame rate of 50 Hz. We found that within a cluster, DnaX molecules are on average separated from each other by 54 nm (early on in the replication cycle) or 119 nm (during the later stage of the replication cycle) (Figure 3.10)[36].

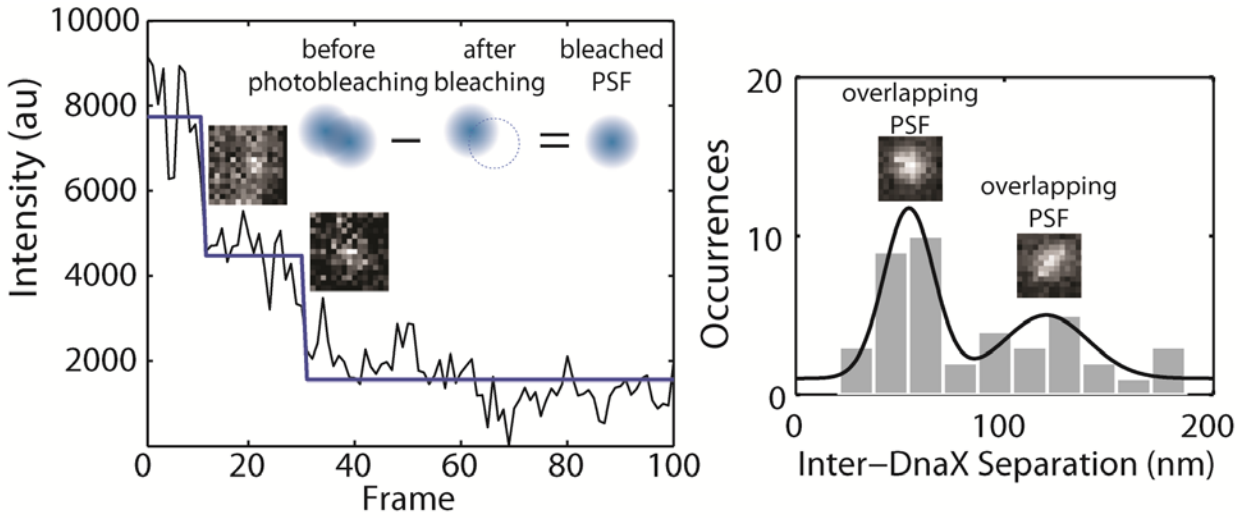


Figure 3.10 (left) Photobleaching-assisted localization of single DnaX-mCitrine molecules within a cluster. The intensity of a cluster is plotted against time (20 ms/frame) and photobleaching events are identified by maximum likelihood estimation[42]. The PSFs of photobleached single molecules can then be obtained by subtracting the average intensity of the frames following the photobleaching from that of preceding frames (schematic representation shown in inset), allowing the precise location of the photobleached molecule to be determined through 2D Gaussian fitting. Representative PSFs for two photobleached mCitrine molecules are shown above the cluster intensity trajectory. (right) Distribution of separation distance between DnaX-mCitrine within a replisome as determined from photobleaching-assisted localization, with two sample overlapping PSFs shown.

In addition, from time-lapse imaging without photobleaching, we also observed that DnaX clusters are neither mobile nor strictly stationary; instead, they engage in subtle motions exploring a small domain of size 84 ± 20 (s.d.) nm as measured by the radius of gyration of the centroid position (Figure 3.11):

$$R_g = \left[\frac{1}{N}\sum_{k=1}^{N}(r_k - r_{mean})\right]^{1/2} \qquad (3.1)$$

where *N* and *r* respectively denote the total number of localizations belonging the cluster and the position of each localization.

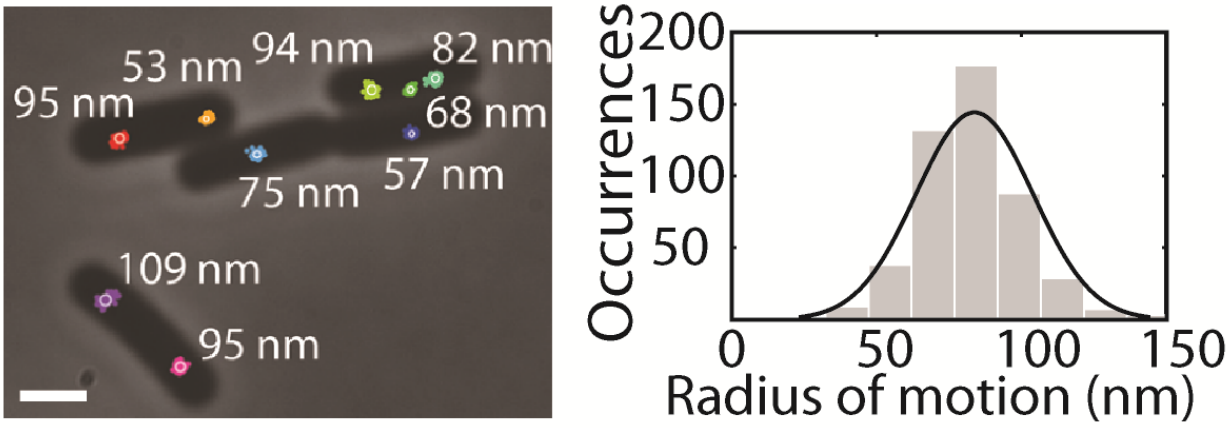


Figure 3.11 The radii explored by DnaX-mCitrine as calculated by tracking the motion of each cluster centroid using low-power time-lapse imaging (left) and the size distribution of the domain explored by each cluster (right), illustrating the subtle replisome motion that explores small domains of ~84 nm in radius on average. Scale bar: 1 μm.

To determine that DnaX-mCitrine localizes in the same way as other replisome subunits, we imaged the processivity clamp protein DnaN-mCitrine in live *B. subtilis* cells (Figure 3.12). Similar to DnaX-mCitrine, DnaN-mCitrine forms compact clusters at quarter positions as well as at cell centers. These results also demonstrate the highly confined localization for both DnaX and DnaN. Because the DnaN-mCitrine fusion partially compromises the MMR activity *in vivo*, we used DnaX-mCitrine as a proxy to the replisome throughout the study.

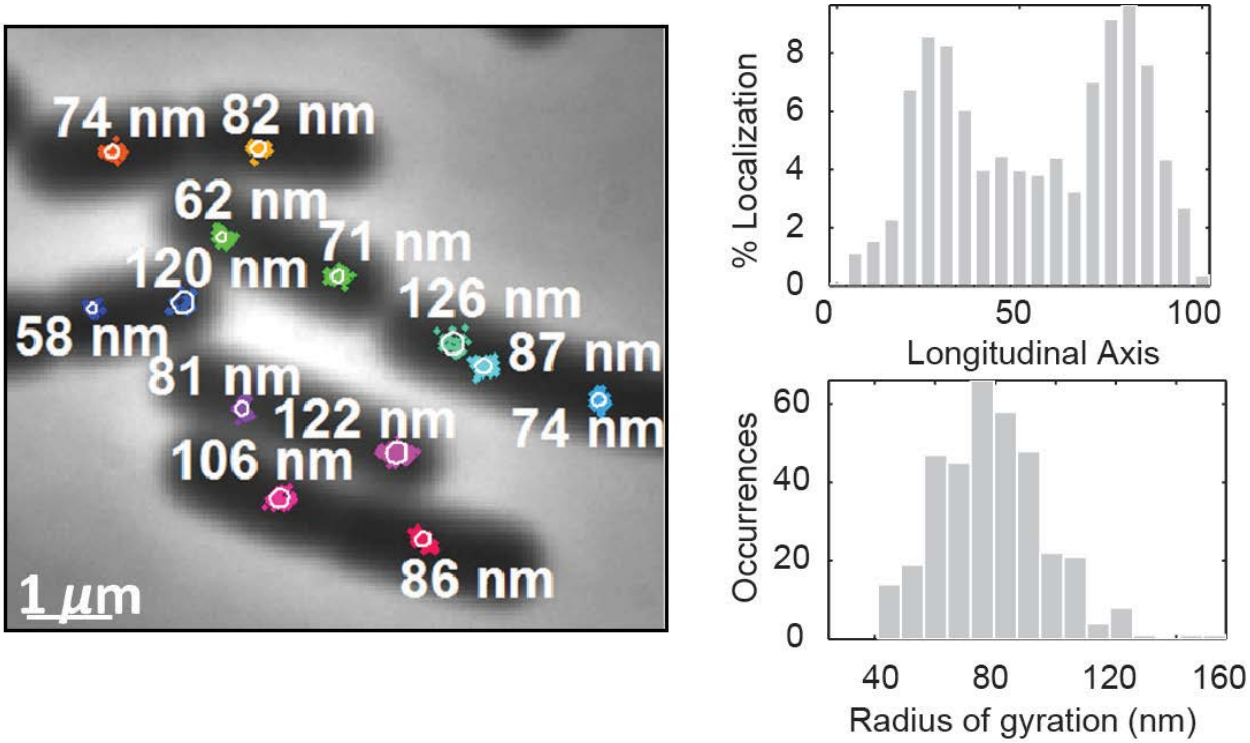


Figure 3.12 Localization of the processivity clamp DnaN-mCitrine in over 100 *B. subtilis* cells during log phase. To determine that DnaX-mCitrine localizes in the same way as other replisome subunits, we imaged the processivity clamp protein DnaN-mCitrine in live *B. subtilis* cells. Similar to DnaX-mCitrine, DnaN-mCitrine forms compact clusters with an average radius of motion ~79 ± 18 (s.d.) nm, and these clusters are most frequently found at quarter positions or, to a lesser degree, cell centers.

Our super-resolution images of DnaX-mCitrine and DnaN-mCitrine clusters support models describing the *B. subtilis* replisome as a confined assembly where template DNA is pulled in and newly synthesized DNA is extruded[36,43], as opposed to the model of a mobile replisome complex that tracks on chromosomal domains that has been described for *E. coli*.[44].

3.4 Relative positions and dynamics of DnaX-mCitrine and MutS-PAmCherry

Imaging cells expressing both DnaX-mCitrine and MutS-PAmCherry, we found that MutS accumulates near the replisome despite being overall more mobile than DnaX (Figure 3.13a). The instantaneous speed from single-molecule MutS tracks[45] also shows dependence on the separation distance between MutS and the replisome: upon entering the replisome region (separation distance < 100 nm), MutS slows down to match the average speed of DnaX clusters (Figure 3.13b), likely as a result of the known direct interaction between MutS and various replisomal components[10,46], or due to MutS engaging in mismatch searching and binding on replisome-proximal DNA[10,46], or a combination of both possibilities.

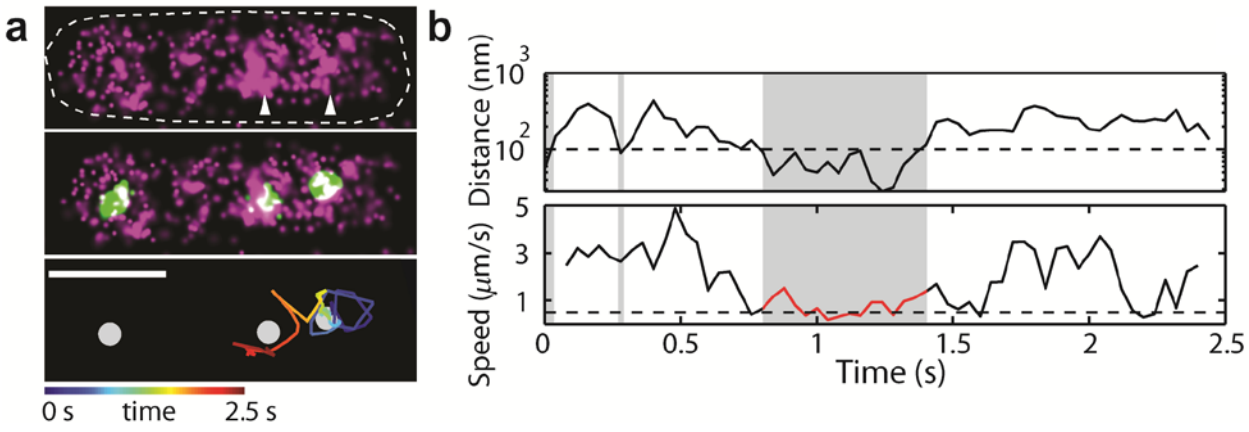


Figure 3.13 (a) PALM reconstruction of MutS-PAmCherry (magenta) in a cell with MutS-enriched regions indicated with white arrowheads (upper), and overlaid with DnaX-mCitrine clusters (green) (middle). A representative time-coded trajectory showing MutS entering, dwelling at, and leaving one of the replisome regions is shown in the lower panel. Scale bars: 1 μm. (b) Separation distance from the replisome (upper) and instantaneous speed (lower) as a function of time for the MutS trajectory shown in (a). Grey windows: time spent in the replisome region. Red curve: a prolonged period of decreased MutS speed. Black dashed lines: 100-nm MutS-DnaX separation distance (upper) and average DnaX speed (lower; 0.5 μm/s, as measured by tracking cluster centroids). Scale bar = 1 μm.

Cross-correlation analysis shows that MutS instantaneous speed does indeed positively correlate with the separation distance from the replisome (Figure 3.14). However, the relatively low correlation amplitude of ~0.2 indicates heterogeneity among MutS subpopulations, that is, some MutS do not slow down or slow down only transiently when passing by the replisome.

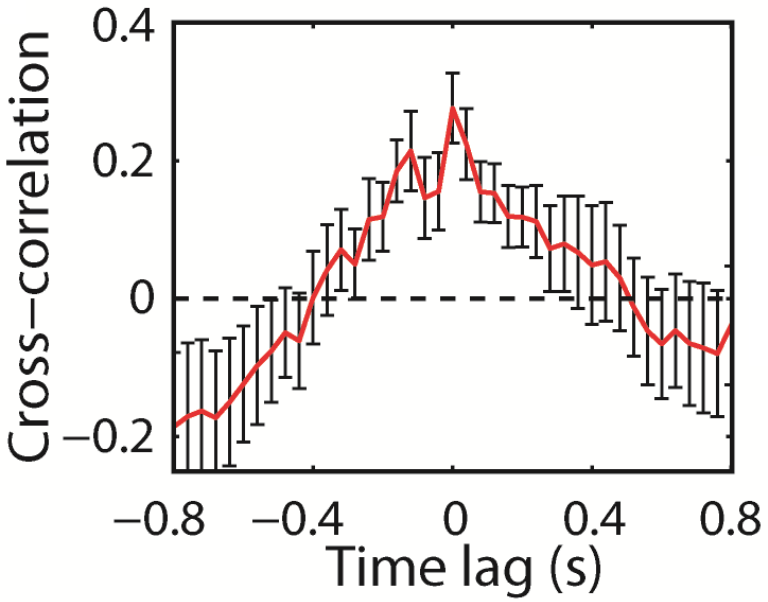


Figure 3.14 Cross-correlation between the separation between MutS and the center of DnaX cluster and the instantaneous speed of MutS from 11 cells, normalized from −1 to 1 [47]. Error bars represent the standard error of the mean.

To quantify how much time a MutS protein spends within the replisome region, we fit the cumulative probability of the dwell time of MutS in the replisome region (P(t) > t) with a two-term exponential decay function (Figure 3.15) and obtained two dwell time constants of 25 ms (42%) and 188 ms (58%). These dwell time constants represent a lower bound, as we only analyzed single-MutS trajectories that start outside the replisome, remain trackable within the replisome, and end outside the replisome (Figure 3.15 inset). As a result, photobleaching and blinking of the fluorophore do not affect the dwell time analysis, but only MutS trajectories that start and end outside the replisome are taken into account.

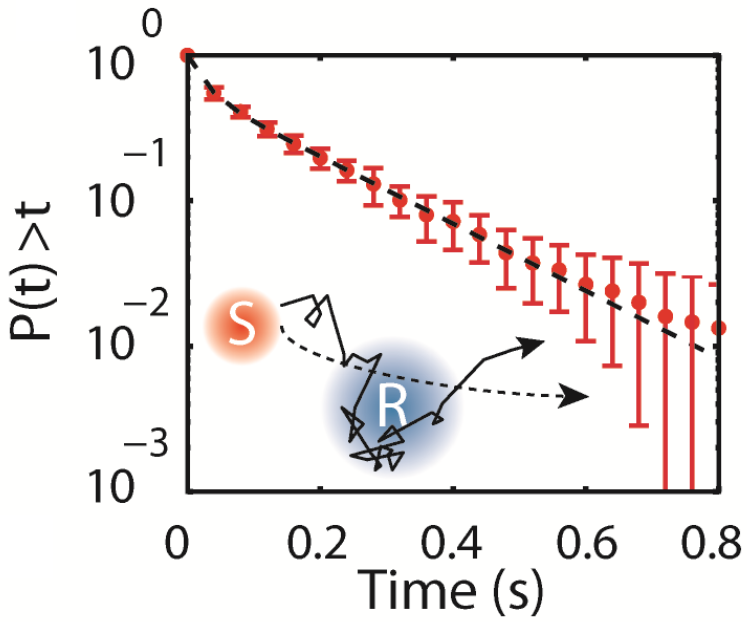


Figure 3.15 Cumulative probability distribution of time period MutS (red) spends within the same replisome region (blue), fit to a two-term exponential decay function (dashed line) $P = A_1\exp(-t/\tau_1)+A_2\exp(-t/\tau_2)$, where $A_1 = 0.42$, $A_2 = 0.58$, $\tau_1 = 25$ ms and $\tau_2 = 188$ ms. The error bars are standard deviations from 7 rounds of bootstrapping. The dwell time distribution is constructed using trajectories at least 10 frames long for molecules that can be tracked from the time they enter the replisome region until they leave the replisome (inset).

3.5 Quantifying the occurrence of mismatches in live cells

Because mismatches are produced at an extremely low rate in cells, we relied on the mismatch-inducing drug 2-aminopurine (2-AP) to look at the response of MutS to mismatches in real time in our experiments. 2-AP is an adenine-analogue capable of being paired with thymine during DNA replication. In the round of DNA replication subsequent to this substitution, a C may be placed across from 2-AP, ultimately leading to an AT-to-GC mutation.

At a concentration of 7.4 mM 2-AP, the probability that an AT base pair will be mutated to GC is a 1.5×10^{-5} [48,49]. Under our experimental conditions, with 600 μg/mL (4.44 mM) 2-AP added to the growth medium at OD ~ 0.35, we therefore expect a 9×10^{-6} probability that a given AT base pair will be converted to GC with 2-AP treatment. The probability of observing a cell which accumulates a given number of mismatches in its DNA during a specified window of time is a Poisson distribution:

$$P[N(t) = k] = \frac{(Rt)^k * e^{-Rt}}{k!}, \quad (3.2)$$

Where k is the number of mutations observed, R is the mutation rate in terms of time, and t is the duration of time the cell is observed.

To determine the probability that a single cell has at least one mismatch during time, t, we calculate the complement of $P[N(t) = 0]$:

$$P[N(t) = 0] = e^{-Rt} \quad (3.3)$$

$$P[N(t) > 0] = 1 - e^{-Rt} \quad (3.4)$$

In *E. coli*, the base pair substitution (BPS) rate in the absence of MMR is 273×10^{-10} BPS/base replicated[2]. We assume this approximates the mismatch incorporation rate of the replicating DNA polymerase, and applying this mismatch rate in the absence of 2-AP to *B. subtilis*, and using an overall replication rate of 1000 nt/s (combined rate, i.e., two forks)[50], we calculate a mismatch rate of 2.73×10^{-11} mismatches per second. This calculation ignores insertion/deletion rate, which will not significantly affect the resulting probability of observing a mismatch in a cell during a 210 s observation. We determine that during a 210 s window of time, which is the approximate time each cell is observed in our PALM experiments, we will observe a cell with a mismatch with a probability of 5.7×10^{-9}. This indicates that in untreated conditions, we effectively never observe a cell in which a mismatch arose during our observation period (Figure 3.16).

Under conditions of 2-AP treatment during the time of our observation the probability of an observed cell obtaining a mismatch is quite high. To calculate *R* under conditions of 2-AP treatment, we consider that with 2265514 AT pairs in the *B. subtilis* PY79 genome, 2-AP treatment causes approximately 20.4 ($9 \times 2265514 \times 10^{-6}$) AT to GC transitions per round of replication[51]. With 4033459 total base pairs in the genome, there are about 197719 (4033459/20.4) base pairs on average between mismatches caused by 2-AP. Two replication forks replicate the genome, each polymerizing DNA at a rate of 500 nucleotides per second, for a combined rate of 1000 nt/s[50]. This means there is one 2-AP insertion resulting in an AT to GC transition every 197.7 s, or 5.05×10^{-3} mutations caused by 2-AP per second. We have excluded the natural rate of mismatch formation from this calculation, because it is so low that it is negligible. Using equation 3.4, we find there is a probability of 0.65 that a 2-AP-treated cell will have at least one mismatch during our 210 s observation (Figure 3.16).

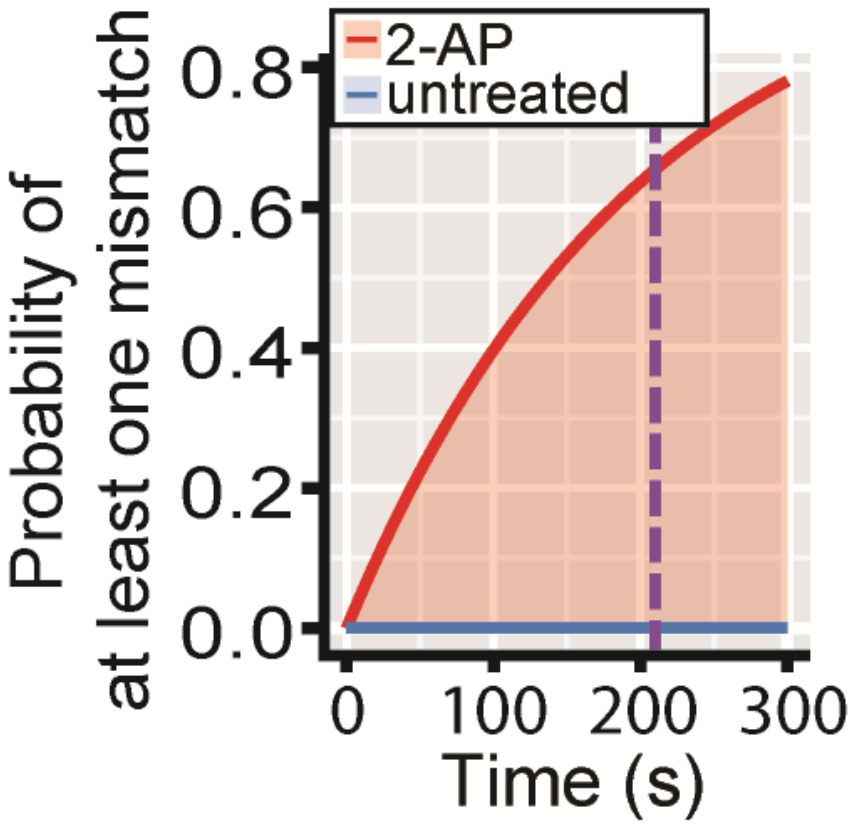


Figure 3.16 Distribution of the probability of a mismatch occurring in an observed cell under normal growth condition (blue) and under 2-AP treatment (red) over time. The vertical purple dashed line indicates the average duration (210 s) of observation for each cell in PALM experiments.

3.6 MutS accumulates at the replisome regardless of mutagen treatment

It is necessary to sample a large number of cells to make meaningful conclusions since there are only approximately 160 copies of MutS-PAmCherry in each cell and PAmCherry photoactivation efficiency is 4-50% [52,53]. Furthermore, only some fraction of the imaged cells has

mismatches to which MutS is able to respond at all (Figure 3.16). Therefore, to compare intracellular DnaX-mCitrine and MutS-PAmCherry positions across many cells, we plotted the percentage of DnaX and MutS localizations at corresponding positions inside a normalized cell as probability density maps (Figure 3.17)[54].

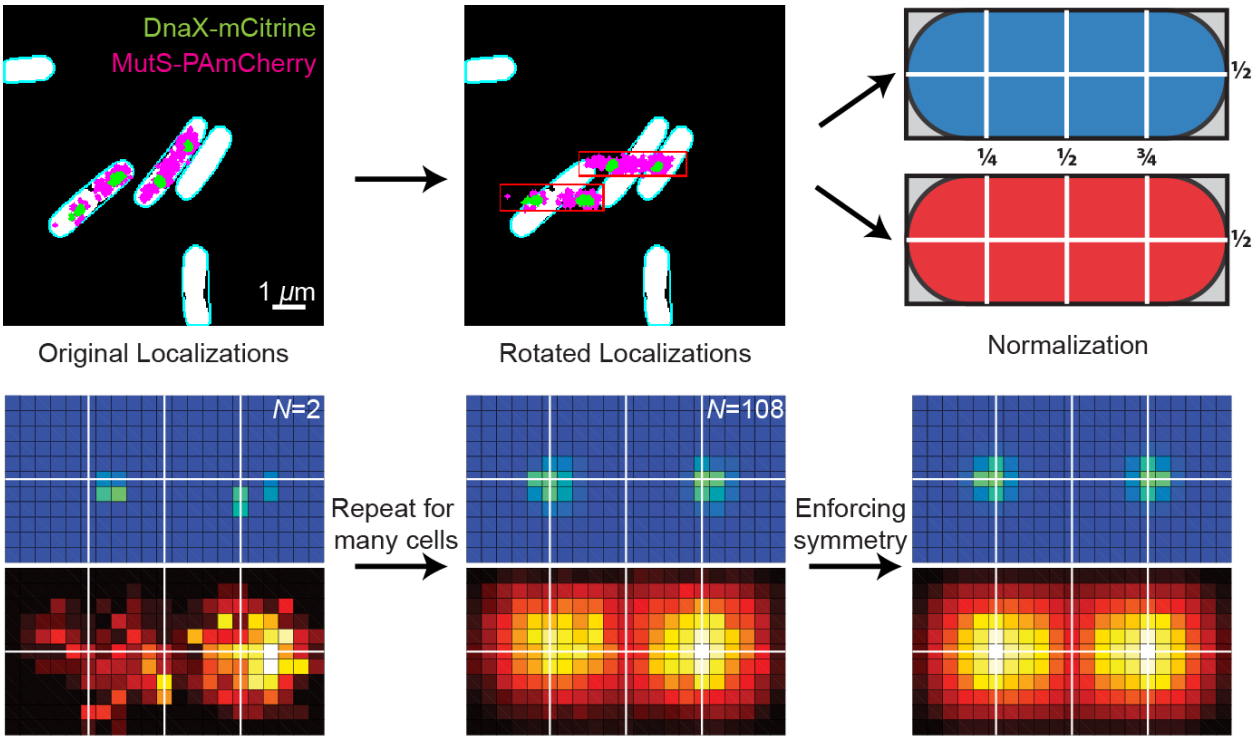


Figure 3.17 Six steps in the procedure for generating localization probability density maps, using data from WT− cells as an example. All cells with two DnaX-mCitrine clusters and having > 100 localizations in both color channels are analyzed.

A density map of wild-type (WT) cells treated with the mismatch-forming drug 2-aminopurine (2-AP) (WT+) shows that regions with the highest DnaX densities are also those most populated by MutS (Figure 3.18c). Importantly, the same pattern was also found in WT cells without 2-AP (WT–) (Figure 3.18a and compare Figure 3.18b to 3.18c). Although previous bulk fluorescence studies of *B. subtilis* have shown that < 10% of WT– cells form MutS foci near the replisome[8,55], here at the single-molecule level with improved sensitivity to capture transient

dwelling behaviors (Figure 3.13b), we reveal that the enrichment of MutS near the replisome is much more prevalent than previously suggested, even in cells with only the very low natural mismatch formation rate (i.e. no 2-AP addition). Furthermore, this mismatch-independent recruitment of MutS in *B. subtilis* resembles the behavior of MutSα in *S. cerevisiae* cells observed using bulk fluroescence[11], suggesting that mismatch recognition is an integral component of replisome function and has been conserved from bacteria to eukaryotes. However, we show that a difference between bacteria and yeast is that bacterial MutS is highly dynamic despite its recruitment to the replisome.

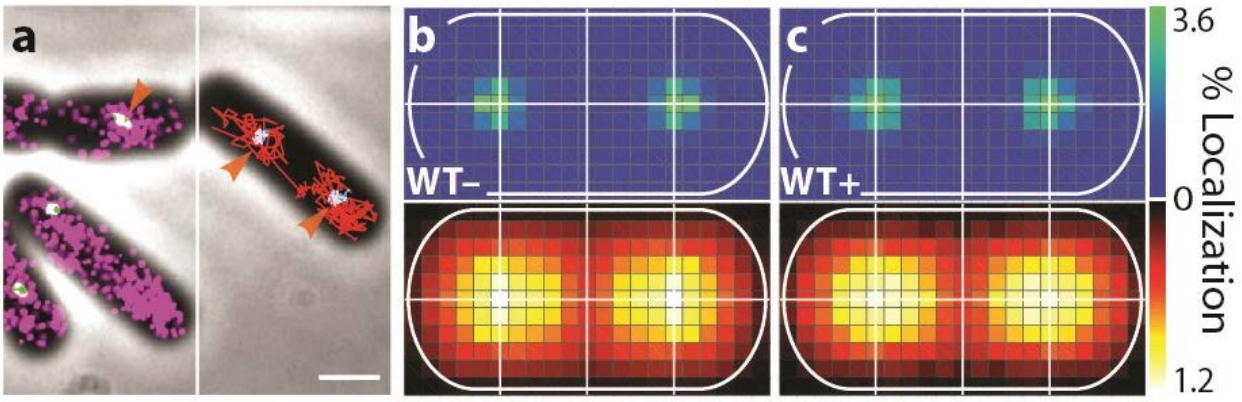


Figure 3.18 (a) PALM reconstruction (magenta) and single-molecule trajectories (red) of MutS-PAmCherry, overlaid with DnaX-mCitrine (green and blue) and phase-contrast cell images. Overlapping signals are colored in white. Orange arrows: replisome regions at which preferential MutS enrichment or dwelling is observed. Scale bar: 1 μm; (b, c) Localization probability density maps of DnaX-mCitrine (top; blue-green) and MutS-PAmCherry (bottom; red-yellow) within a normalized cell. White lines designate the ¼, ½ and ¾ positions along the cell long axis and the ½ position in the transverse direction. 108 WT– cells (b) and 91 WT+ cells (c) with two replisome clusters were used to generate the corresponding density maps. To allow for quantitative comparison of colocalization between different cases, the Pearson correlation coefficients between each pair of DnaX/MutS density maps are calculated. The correlation coefficients for WT– and WT+ cells are 0.83 and 0.81 respectively. Grid pixel size: 100~200 nm.

3.7 MutS speeds up after mutagen treatment

To quantify the motion of MutS, we analyzed the distribution of squared displacements (r^2) of MutS during a given time lag τ, and constructed the corresponding cumulative probability distribution (CPD) $P(r^2, \tau)$ (Figure 3.19):

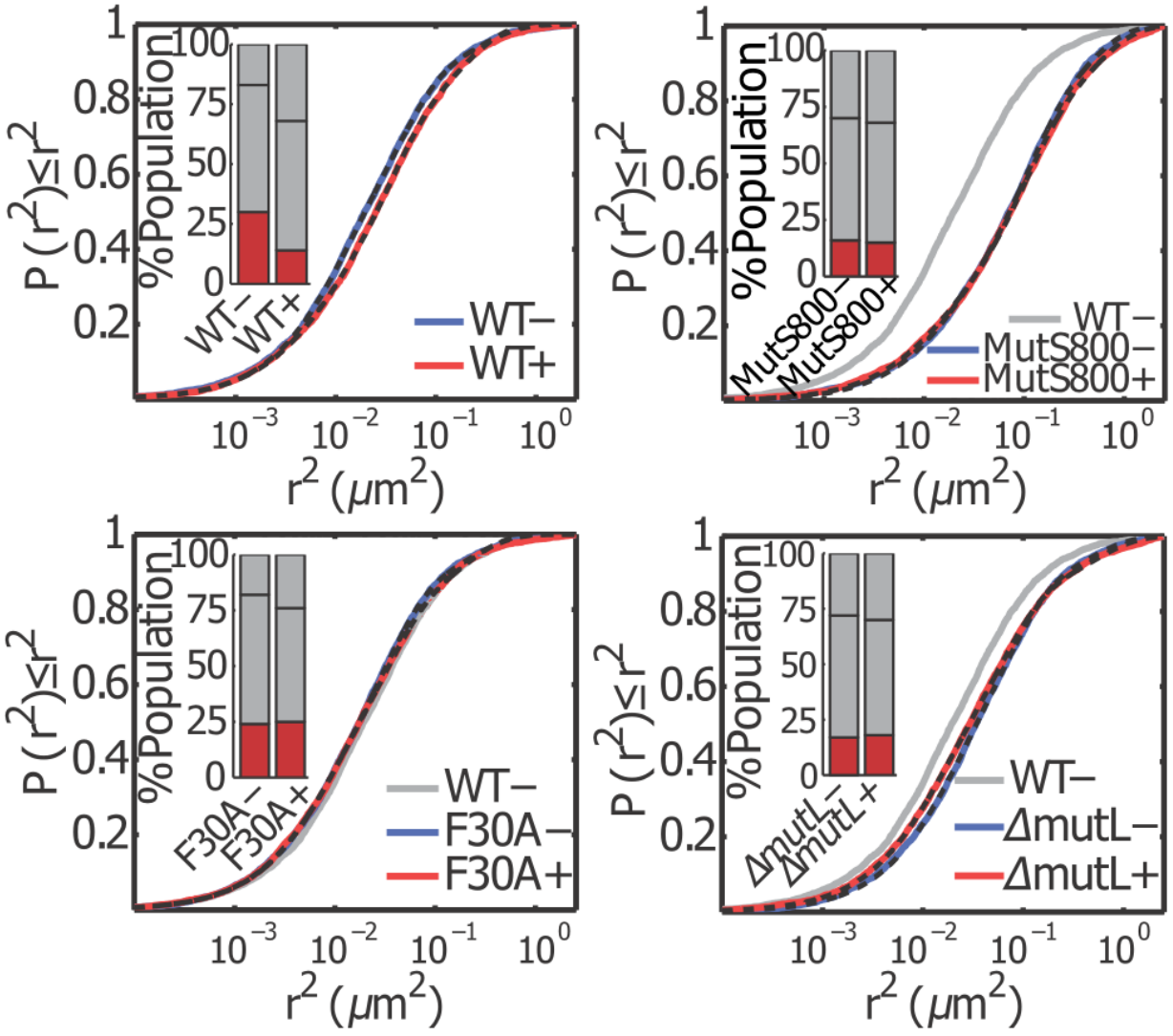


Figure 3.19 Cumulative probability distributions (CPDs) of squared displacements of MutS-PAmCherry (colored lines) from MutS mutant strains investigated in this study. The CPDs are fitted by a three-term diffusion model[56-58] (black dashed lines), with the percentages of two fast populations (grey) and a slow population (red) given in the insets.

To accommodate the heterogeneity of MutS motion, equation 2.14 is expanded to include three terms, each describing one population with a specific diffusivity:

$$P(r^2,\tau) = 1 - \left[\alpha \cdot e^{\frac{-r^2}{\langle r_1^2(\tau)\rangle+\sigma^2}} + \beta \cdot e^{\frac{-r^2}{\langle r_2^2(\tau)\rangle+\sigma^2}} + (1-\alpha-\beta)\cdot e^{\frac{-r^2}{\langle r_3^2(\tau)\rangle+\sigma^2}}\right] \quad (3.5)$$

In estimating the percentage of each population (α, β and $1-\alpha-\beta$), only the first step of each trajectory was used when fitting this equation to experimental data to minimize bias towards the slower population, which usually remain visible for a longer period of time[57] (Figure 3.19). To obtain the diffusion coefficient of each population, the fitting was then repeated for larger values of τ. Because on average a trajectory lasts ~15 frames, the maximum τ used was 7 frames (0.28 s). Generally, D can be calculated from the slope of the mean-squared-displacement (MSD, or $\langle r^2(\tau)\rangle$) vs. τ as described by equation 2.11. However, in the current case, the MSD vs. τ plots obtained from CPD analysis did not have the strictly linear profiles suggested by the Einstein-Smoluchowski relation, instead, spatial confinement, presumably resulting from interactions with other species, causes the MSD of the slow MutS population to reach a plateau at larger values of τ, and we address this deviation by invoking equation 2.12, which fits the MSD vs. τ relation of the slow population with a model assuming the particle has an initial diffusion coefficient D_0 and is confined within a region of length L. The values of D_0 and L for various strains investigated throughout the study are tabulated in Table 3.1.

Table 3.1 Initial diffusion coefficient D_0 and length of confinement L for the slow MutS population across various strains and conditions of 2-AP treatment

Strain	D_0 (μm/s^2)	L (nm)
WT−	0.029 ± 0.010	194 ± 21
WT+	0.030 ± 0.010	202 ± 23
MutS800−	0.065 ± 0.038	205 ± 18
MutS800+	0.045 ± 0.025	189 ± 18
MutS[F30A]−	0.032 ± 0.016	157 ± 14
MutS[F30A]+	0.031 ± 0.016	197 ± 23
Δ*mutL*−	0.048 ± 0.160	184 ±10
Δ*mutL*+	0.030 ± 0.010	182 ± 11

Despite the nearly identical net localization patterns for MutS under normal and mutagenic growth conditions (Figure 3.18b-c), on the single protein level, we found that MutS exhibits an overall faster motion in WT+ cells compared to WT– (Figure 3.19). This *in vivo* difference in speed is consistent with *in vitro* observations that MutS switches from a rotation-coupled sliding configuration to a faster and rotation-free sliding after mismatch binding[7]. To gain insight into the spatial dependence of MutS motion, we analyzed the average diffusion coefficient of MutS as a function of separation distance from the nearest replisome (Figure 3.20). The diffusion coefficient, *D*, is calculated from the mean square displacement, and only data from the first quarter of the time lags for each of over 1000 trajectories longer than 10 frames were analyzed to minimize errors associated with higher time lag values[59]. Regardless of whether 2-AP is present, we observe that MutS slows as it approaches the replisome in both WT+ and WT– cells, suggesting that error searching, and possibly subsequent binding events, are restricted

to nascent DNA in the neighborhood of DNA replication sites, and are likely initiated by interactions between MutS and the replisome.

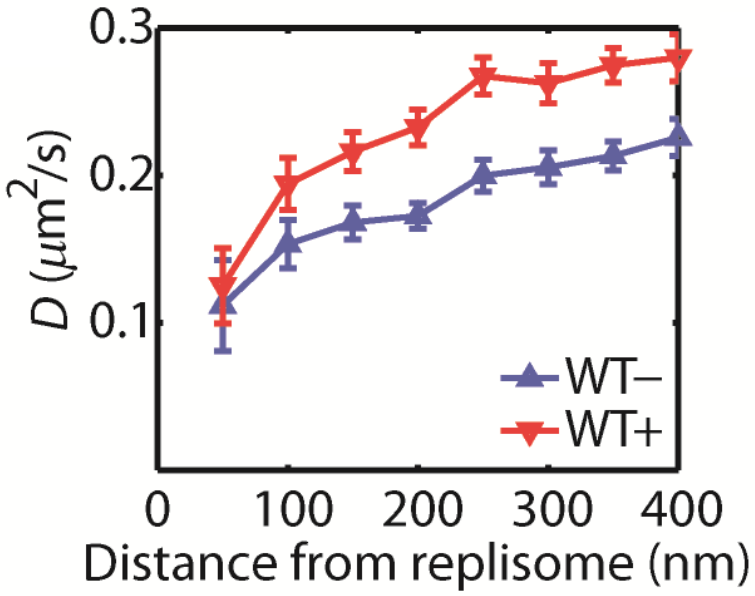


Figure 3.20 Diffusion coefficients of MutS-PAmCherry as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval.

3.8 The MutS/replisome interaction is necessary for MutS recruitment and MMR *in vivo*

To further understand the relationship between DNA replication and the position and dynamics of single MutS molecules, we constructed strains each designed to block one of four MMR steps: 1) MutS binding to β-clamp, 2) mismatch recognition, 3) MutS ATPase activity, and 4) subsequent MutL recruitment (Figure 3.21).

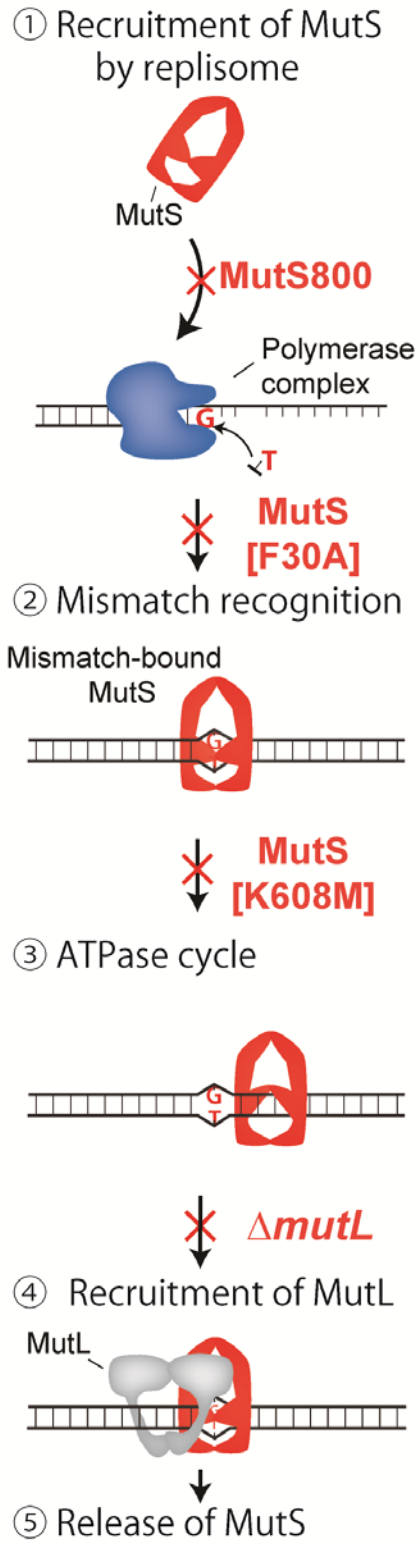
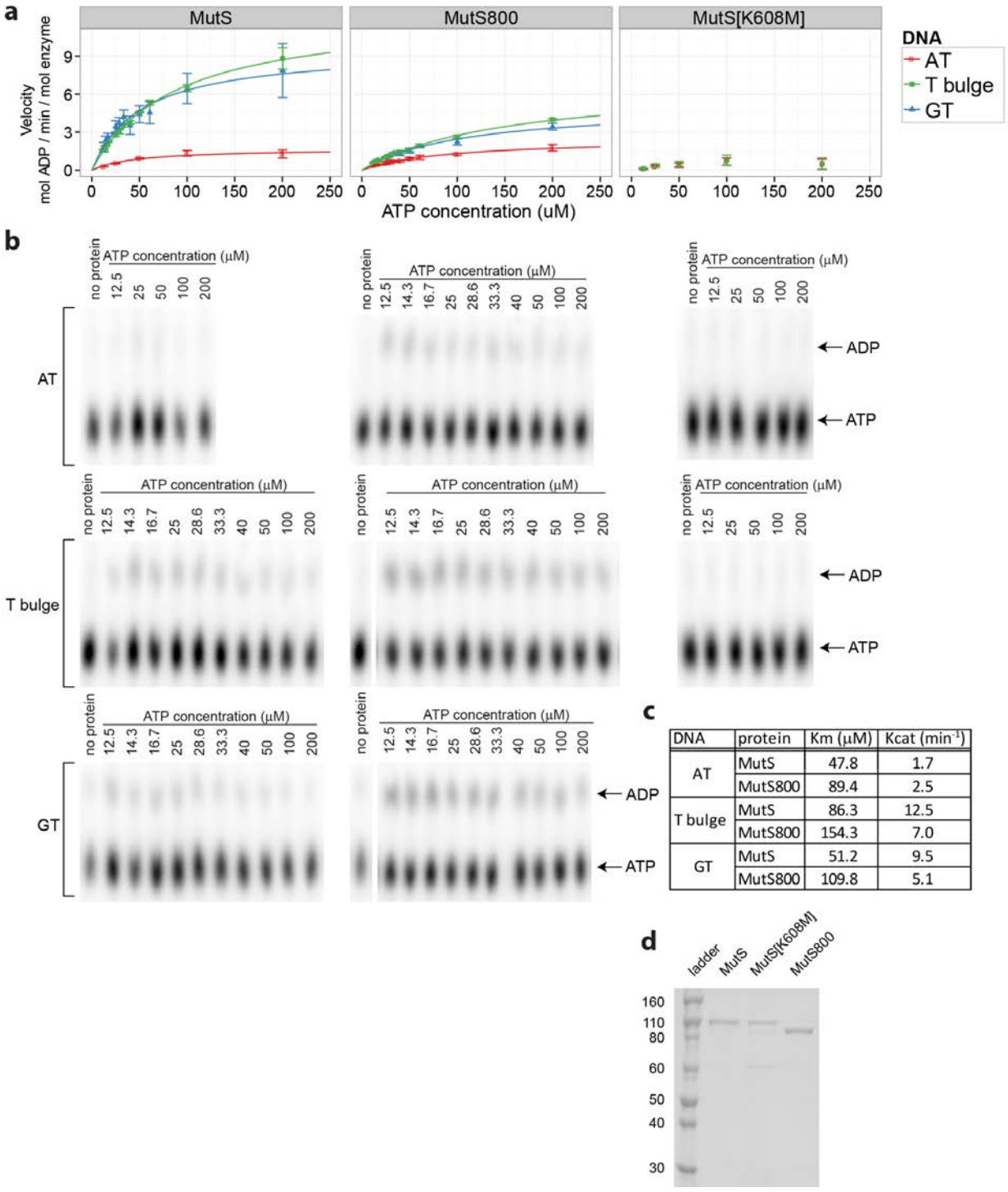


Figure 3.21 Schematic diagrams showing the first four steps of MMR, including replisome binding, mismatch recognition, ATPase activity and MutL recruitment, each of which is blocked in one of four mutant strains.

First, we tested the effect of replisome interaction perturbation on MutS motion and location using a two-pronged approach: we first examined MutS800 (Figure 3.21), a MutS variant with the domain that has β-clamp affinity removed. MutS800 is able to bind mismatches[46] and maintains similar ATPase activity to full-length MutS (Figure 3.22).



DNA	protein	Km (μM)	Kcat (min^{-1})
AT	MutS	47.8	1.7
	MutS800	89.4	2.5
T bulge	MutS	86.3	12.5
	MutS800	154.3	7.0
GT	MutS	51.2	9.5
	MutS800	109.8	5.1

Figure 3.22 ATPase activity of MutS, MutS800 and MutS[K608M]. ATPase assays were carried out using MutS or MutS800 in the presence of either homoduplex (AT) DNA, DNA with a GT mismatch, or DNA with a T bulge. In the case of MutS[K608M], ATPase assays were carried out with either AT DNA, or T bulge DNA. (a) Plots of ATPase activity for the indicated protein and DNA. Points and error bars represent the mean ± SEM of three independent experiments.

Lines represent the Michaelis-Menten equation fit to the data. (b) Representative phosphorimaging results of TLC plates used for each protein/DNA combination. The positions of ADP and ATP on the plates are indicated on the right. (c) Parameters derived from fitting the Michaelis-Menten equation to the ATPase assay data for each protein/DNA combination. Km and Kcat were not determined for MutS[K608M]. MutS and MutS800 hydrolyze ATP similarly in the presence of homoduplex DNA, and the ATPase activity of both proteins are stimulated by the presence of mismatched DNA, but wild type MutS ATPase activity responds more strongly to mismatched DNA. (d) A coomassie-stained gel of purified MutS, MutS[K608M], and MutS800. 0.5 μg of each protein was subject to 7.5% SDS-PAGE.

Second, we complemented the MutS800 investigations with studies of the β-clamp allele *dnaN5*, which is compromised for interaction with MutS[46]. Relative to the WT– and WT+ density maps, the MutS localization pattern is drastically changed in untreated MutS800 (MutS800–) cells (compare Figure 3.23b to Figure 3.18b-c). While DnaX locations remained largely the same, MutS800 was uniformly distributed throughout the cell, lowering the Pearson correlation between DnaX and MutS800 from 0.83 in WT– cells to 0.50 in MutS800–.

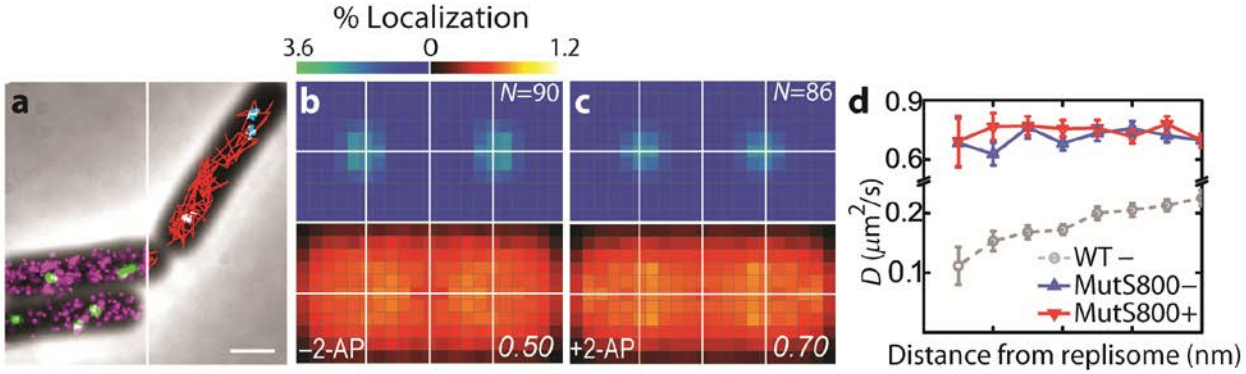


Figure 3.23 Localization and dynamics of MutS800 with respect to the replisome. (a) Two-color PALM (left) and trajectory (right) image of representative cells from the MutS800 strain. Scale bar = 1 μm. (b, c) Localization probability density maps of untreated cells (b) and 2-AP treated cells (c), generated from *N* cells with two replisome clusters. Pearson correlation coefficients between DnaX and MutS densities are listed at the lower right corner of corresponding MutS density maps. Note that MutS density maps may exhibit similar intensity levels but have different correlation coefficients due to differences in the positioning of corresponding DnaX density maps. (d) Diffusion coefficients of MutS800 as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval.

The significant decrease in preferential enrichment of MutS near the replisome confirms that the recruitment of MutS to the replisome observed in WT cells depends in part on interactions between MutS and the β-clamp. The recruitment of MutS to the replisome observed in WT cells depends in part on interactions between MutS and the β-clamp. The density maps after 2-AP treatment (Figure 3.23c) suggest that MutS800 still partially responds to mismatch incorporation (the correlation coefficient increases from 0.50 to 0.70 with addition of 2-AP). Although MutS800 is still capable of mismatch binding *in vitro*[46], only 10-12% of MMR activity is retained in MutS800 cells (Figure 3.24 and Table 3.2)[55], demonstrating the importance of replisomal interactions for efficient mismatch recognition *in vivo*, and we interpret the slightly higher densities near the quarter positions of the MutS800+ density map (Figure 3.23c) as MutS800 responsible for the remaining MMR activity.

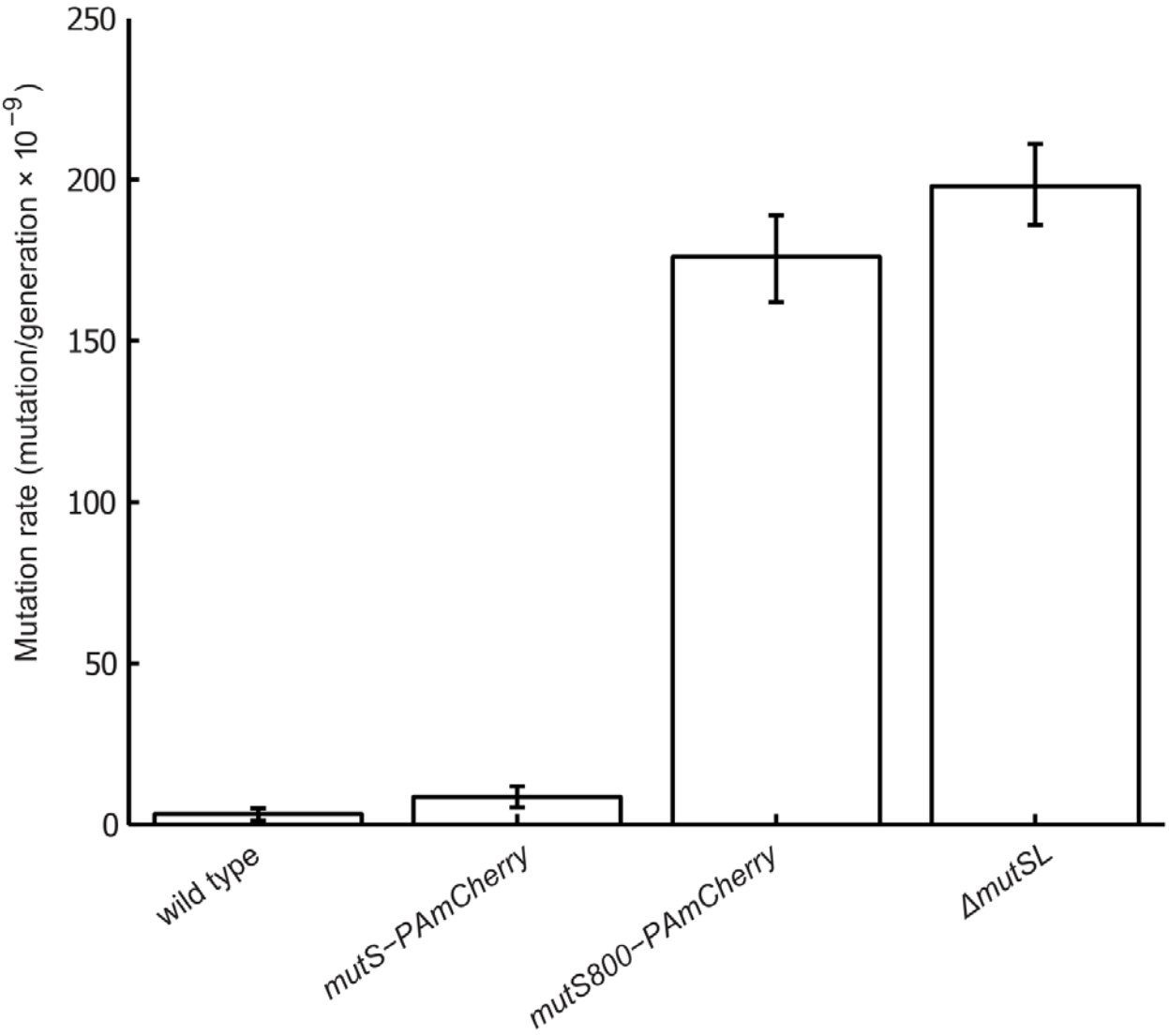


Figure 3.24 Mutation rates of the indicated strains. Rates were determined by maximum likelihood estimation implemented in the online tool FALCOR[60]. The rates indicate that the fusion is functional in mismatch repair for the mutS-PAmCherry strain, whereas the MutS800 mutant is significantly compromised in mismatch repair activity.

Table 3.2 Mutation rate[a] and percent mismatch repair of various strains

Strain	Mutation rate	95% CI	% MMR activity
Wild type	3.29	1.16 — 4.94	100
muts-PAmCherry	8.60	5.31—11.8	97
muts800-PAmCherry	176	162 — 189	12
ΔmutsSL	198	186 — 211	0

[a] expressed as mutations per generation × 10^9

In addition, the diffusion rates for MutS800 were significantly increased relative to WT– and WT+ (Figure 3.19 and Figure 3.23d). One possible contributing factor to the increase in diffusion rate is that mismatch-stimulated ATPase activity of MutS800 is 2-fold diminished relative to that of the wild type MutS, although the basal ATPase activity between the two proteins in the presence of homoduplex DNA is the same (Figure 3.22). However, our results for the complementary experiment in the DnaN5 mutant strain also show a partially compromised colocalization pattern between MutS and DnaX and a faster diffusion profile for MutS in the DnaN5 strain compared to WT cells (Figure 3.25), demonstrating that the increase in MutS800 diffusion should not be solely attributed to a slight decrease in stimulation of ATPase activity in response to a mismatch. One potential explanation for this change in dynamics is that the MutS/*β*-clamp interaction facilitates MutS binding to DNA strands and in the absence of such an interaction, MutS cannot efficiently engage in slow 1D searching motion.

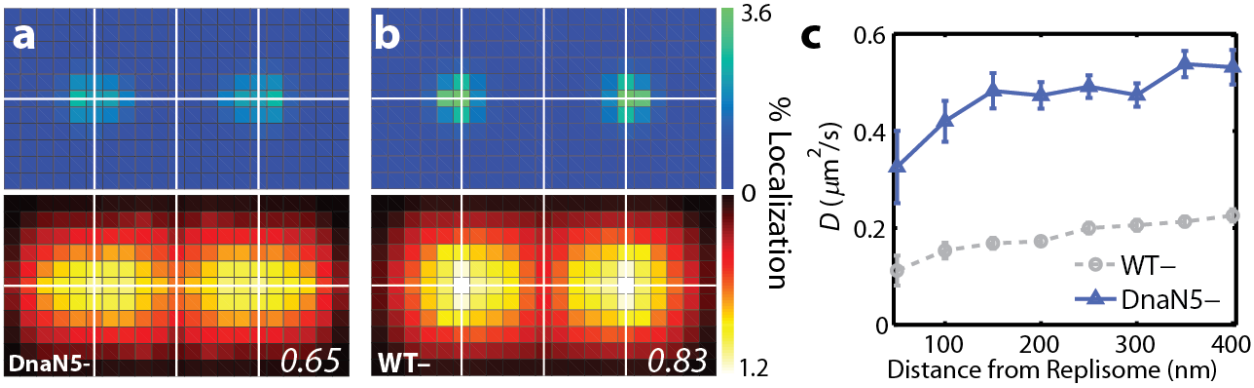


Figure 3.25 Localization and diffusion of MutS in the DnaN5 mutant strain compared to MutS from WT– cells. DnaN5 is a strain that contains a point mutation that renders the β-clamp compromised for binding MutS. Here we use this strain to probe whether MutS/ β-clamp interaction affects the recruitment of MutS to the replisome from a perspective complementary to MutS800. We observed a noticeable decrease in the correlation between MutS and DnaX from the DnaN5 strain (a) compared to the WT– case (b), although the attenuation in colocalization is not as severe as that observed in the MutS800– cells, presumably because DnaN5 retains some residual MutS binding. The diffusion rate of MutS is also increased in the DnaN5 strain (c).

3.9 MutS recruitment to the replisome occurs independently of mismatch recognition

As the MutS positioning did not change upon 2-AP treatment in WT cells, we tested if the recruitment of MutS to the replisome is contingent upon mismatch binding. We analyzed the distribution and motion of MutS[F30A] (Figure 3.21), which is unable to recognize mismatches[61]. Both with and without 2-AP, this mutant preserved the elevated MutS density around the replisome seen in WT cells (compare Figure 3.26b-c to Figure 3.18b-c). Also, the distribution of MutS step sizes, population composition (Figure 3.19), and dependence of diffusion rates on separation distance (Figure 3.26d) for both MutS[F30A]– and MutS[F30A]+ were virtually identical to those of WT–. Thus, as expected, because MutS[F30A] cannot bind mismatches[10], all aspects of MutS[F30A] localization and motion remain unresponsive to mismatches caused by 2-AP (Figure 3.26 and Figure 3.19). The highly similar positioning and dynamics of MutS[F30A] – and MutS in WT– cells supports the notion that the recruitment of

MutS by the replisome not only precedes mismatch recognition, but also occurs independently of it. This dynamic is consistent with the ability of MutS to efficiently respond to very rare mismatches[2]. Because MutS is enriched at the replisome, in close proximity to potential mismatches and prior to errors occurring, MutS has access to newly replicated, "naked" DNA strands along which prolonged sliding would be possible.

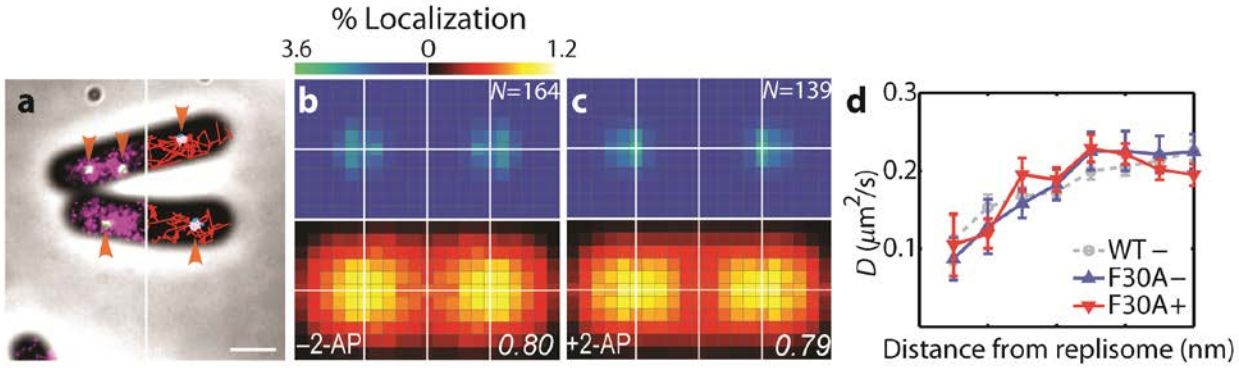


Figure 3.26 Localization and dynamics of MutS[F30A] with respect to the replisome. (a) Two-color PALM (left) and trajectory (right) image of representative cells from the MutS[F30A] strain. Scale bar = 1 μm. (b, c) Localization probability density maps of untreated cells (b) and 2-AP treated cells (c). (d) Diffusion coefficients of MutS[F30A] as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval.

3.10 The ATPase cycle is necessary for MutS recruitment

MutS engages in an ATPase cycle during MMR wherein MutS searches for mismatches in an ADP-bound state, and upon mismatch detection, MutS binds ATP and switches into a stable sliding clamp on DNA[19-21]. To better understand how the ATPase cycle affects MutS localization and dynamics *in vivo*, we constructed a strain harboring MutS[K608M]-PAmCherry as its only source of MutS (Figure 3.21). MutS[K608M] in *B. subtilis* is the homologous substitution to MutS[K620M] in *E. coli*[62]. This substitution changes the highly conserved Walker

A motif in MutS, and purified MutS[K608M] has no measurable ATPase activity (Figure 3.22)[62]. Strikingly, MutS[K608M] displayed highly diffusive behavior both with (MutS[K608M]+) and without (MutS[K608M]-) 2-AP treatment (Figure 3.27a-c). MutS[K608M] motion was also unresponsive to 2-AP treatment, and similar to MutS800, the diffusion of MutS[K608M] did not depend on its separation distance from the replisome (Figure 3.27d). However, in contrast to MutS800 motion, which was significantly faster, this constant MutS[K608M] diffusion rate throughout the cell closely resembled the rate of WT+ MutS at positions > 250 nm from the replisome (compare Figure 3.27d to Figure 3.20). These results suggest that, whereas the processivity clamp interaction alone is not entirely sufficient to recruit MutS to the site of DNA synthesis, the ATPase cycle of MutS is necessary for proper positioning of MutS within the cell, possibly to return MutS to its mismatch searching state near the replisome.

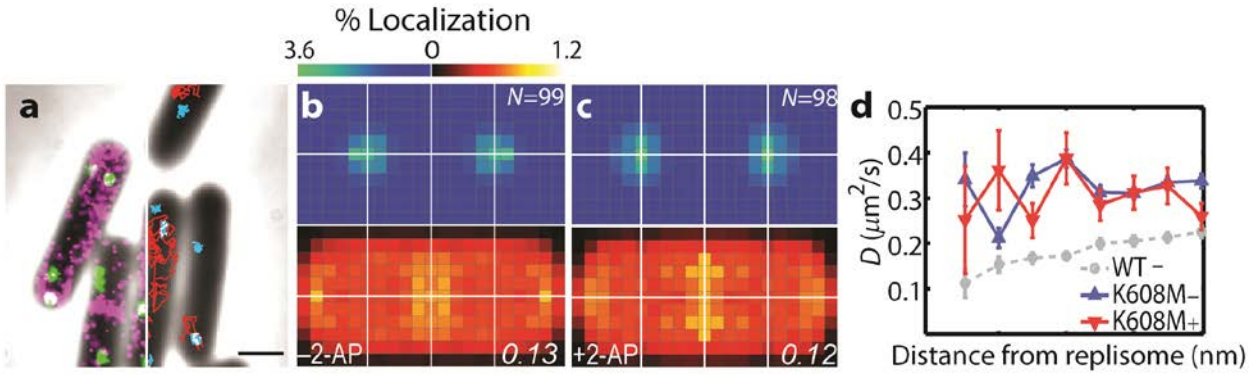


Figure 3.27 Localization and dynamics of MutS[K608M] with respect to the replisome. (a) Two-color PALM (left) and trajectory (right) image of representative cells from the MutS[K608M] strain. Scale bar = 1 μm. (b, c) Localization probability density maps of untreated cells (b) and 2-AP treated cells (c). (d) Diffusion coefficients of MutS[K608M] as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval.

3.11 MutS only recognizes mismatches spatially close to the replisome

Following mismatch detection, MutS binds MutL to form a complex[61], which is proposed to then slide away from the mismatch in search of strand-discrimination signals along the DNA[7,8,46]. To further test whether MutS binds mismatches near or distal to the replisome, we probed the effect of MutL binding on MutS dynamics by imaging DnaX and MutS in a *ΔmutL* strain having no MMR activity (Figure 3.21). Without 2-AP (*ΔmutL*–), the localization of MutS in *ΔmutL* largely resembles that observed in the WT and MutS[F30A] strains (compare Figure 3.28a to Figure 3.18b-c and Figure 3.26b-c), indicating that the pre-loading of MutS at the replisome is unaffected by the absence of MutL. However, it is notable that when mismatches were induced in *ΔmutL* cells by 2-AP treatment (*ΔmutL*+), the density of MutS in replisome-proximal regions diminished (compare Figure 3.28b to Figure 3.18c and Figure 3.26c). We postulate that the decline in MutS enrichment at the replisome is caused by MutS remaining mismatch-bound in the absence of MutL, and thus being carried away from the replisome with newly synthesized mismatch-containing DNA as DNA replication proceeds. We tested this hypothesis by imaging MutS in the *ΔmutL* strain treated with 2-AP for one hour, followed by treatment with HPUra, which blocks DNA replication. HPUra was added to the cell culture at a final concentration of 162 μM prior to imaging. Consistent with our hypothesis, enrichment of MutS at the replisome was restored after DNA replication was arrested (compare Figure 3.28c to Figure 3.28b), demonstrating that, in *ΔmutL*, MutS is a marker of mismatch positions on the newly replicated DNA and in the absence of HPUra, the attenuated MutS accumulation observed in *ΔmutL*+ (Figure 3.28b) is caused by the mismatch-bound MutS being carried away from the replisome during ongoing DNA synthesis. The process of MMR, from mismatch detection by MutS through ultimate replacement of the error-containing strand of DNA must occur quickly, as

deletion of MutL was necessary to observe an effect of 2-AP treatment on MutS position (Figure 3.28a-b).

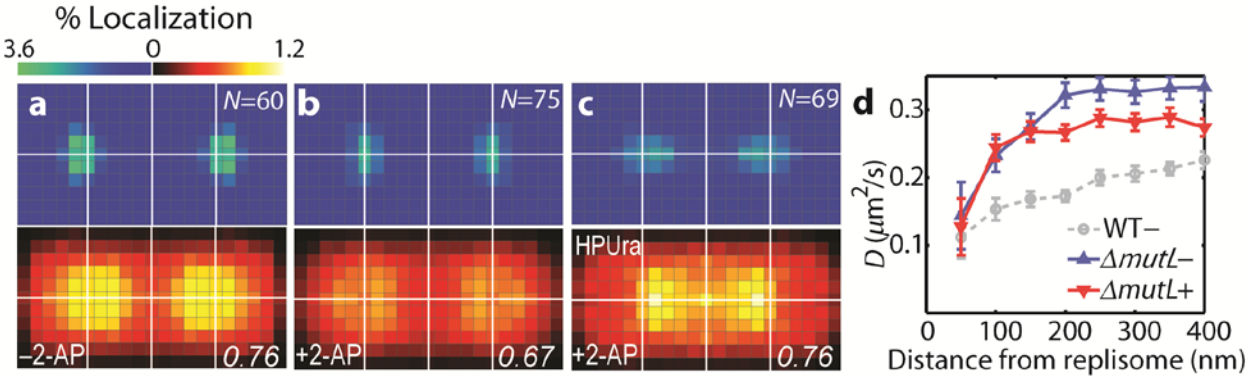


Figure 3.28 Localization and dynamics of MutS with respect to the replisome in the absence of MutL. (a-c) Localization probability density maps of untreated cells (a), 2-AP treated cells (b) and 2-AP/HPUra double treated cells (c). (d) Diffusion coefficients of MutS in the Δ*mutL* strain as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval.

Overall, the dependence of mismatch recognition on replisome coupling, as observed in MutS800 and DnaN5 strains, is further supported by the MutS localization in *ΔmutL*. Here, DNA mismatches will have accumulated throughout the genomic DNA of the cells in Figure 3.28b prior to HPUra treatment due to prolonged 2-AP treatment without MutL. Therefore, if MutS could bind mismatches away from the replisome, the MutS distribution would resemble the diffuse pattern for *ΔmutL*+ cells (Figure 3.28b). Rather, HPUra causes MutS to resume its normal enrichment at the replisome in Figure 3.28c. This, together with the loss of MMR activity in MutS800 and partial loss of MMR in DnaN5[46] (Figure 3.24 and Table 3.2), indicates that proximity to the replisome is critical for MutS to efficiently locate mismatches *in vivo*, and also that the proximity of mismatched DNA to the replisome plays a fundamental role in MMR

initiation, as mismatches distal to the replisome are not efficiently targeted by MutS. In contrast to WT cells, MutS from the *ΔmutL* strain slows down upon 2-AP treatment (Figure 3.28d), which presumably results from the increased number of mismatch-bound MutS molecules, as the difference in diffusion coefficients between *ΔmutL*– and *ΔmutL*+ is largest outside the replisome region. In contrast to perturbations caused by 2-AP, which has no effect on replisome localization, blocking DNA replication with HPUra changed the locations of both the replisome and MutS. For pre-divisional cells with two DNA replication sites, the replisome and MutS were both shifted inward from quarter positions toward the cell center (Figure 3.28c), likely due to the fact that, with DNA replication paused, cells continue to grow but fail to partition the replisome to daughter cells. The simultaneously shifted colocalization pattern observed in HPUra-treated cells again highlights the functional correlation between DNA replication and repair.

3.12 MutS interacts with essential DNA polymerases both *in vivo* and *in vitro*

To biochemically test whether MutS physically associates with replicating DNA, we synchronized cells for replication initiation using a temperature-sensitive allele of the replicative helicase loader DnaB (DnaB134), followed by ChIP-sequencing (ChIP-seq) of MutS and the essential DNA polymerase subunits PolC and DnaE to determine the location of each protein on the chromosome[63,64]. The *B. subtilis* genome consists of a single circular chromosome with a single origin of replication (*oriC*). DNA replication commences at *oriC* and proceeds bidirectionally toward *terC*. In synchronized culture during pre-initiation, we observed little to no enrichment of MutS, DnaE or PolC on the chromosome (Figure 3.29). In contrast, 10 minutes after replication has commenced, PolC, DnaE and MutS are co-enriched at the site of replication

initiation (*oriC*), each displaying about 2-fold enrichment (Figure 3.29). Importantly, there was little enrichment in ChIP-seq of MutS800 at 10 minutes after replication initiation, consistent with MutS800 being deficient in recruitment to the replisome, and ChIP-seq using antibodies directed against MutS in a strain lacking *mutS* yielded no enrichment at all (Figure 3.29). Given these controls, the less than 80% synchrony within our system[65], and the very similar patterns of enrichment we observed in our independent ChIP-seq of two essential components of the *B. subtilis* DNA polymerase, we are confident that the observed enrichment, although only 2-fold, represents the *bona fide* location of the DNA polymerase and MutS on the chromosome in the plurality of the cells in our culture synchronized for replication initiation. We conclude that MutS is physically associated with the site of ongoing DNA synthesis and is rapidly loaded at *oriC* upon chromosomal replication initiation.

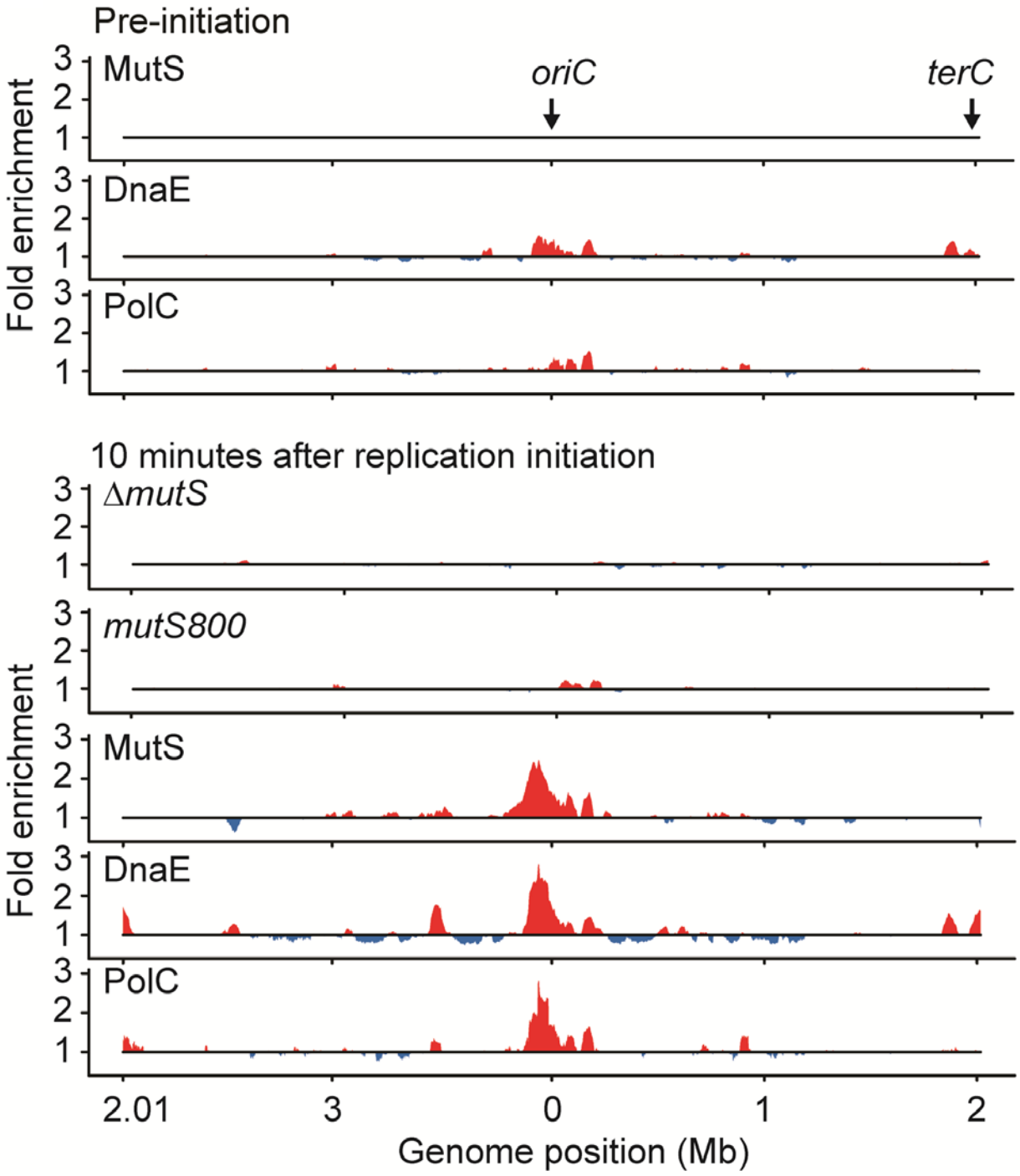


Figure 3.29 Analysis of pooled ChIP-seq data from two independent experiments showing the enrichment levels of MutS and the polymerases DnaE and PolC along the chromosome prior to (upper) and 10 minutes after (lower) DNA replication initiation.

Furthermore, we observed from co-immunoprecipitation (co-IP) that MutS binds both PolC and DnaE *in vivo* before mismatch formation (Figure 3.30). Using a far western blot, we also detected direct *in vitro* interaction of MutS with PolC and DnaE (Figure 3.31). Because we carried out the co-IP with the reversible membrane-permeable crosslinker DSP, it is possible that the co-IP of PolC and DnaE with MutS occurs indirectly via MutS interaction with β-clamp. Taken together with our imaging and ChIP-seq results, these data suggest that whether the interaction of MutS with PolC and DnaE *in vivo* is direct or indirect, in the absence of 2-AP, MutS is capable of searching DNA in extremely close proximity (DSP has a spacer arm length of 12.0 Å) to the actively replicating DNA polymerase complex, and with 2-AP in the growth medium, MutS scans newly replicated DNA in clamp zones trailing the replisome.

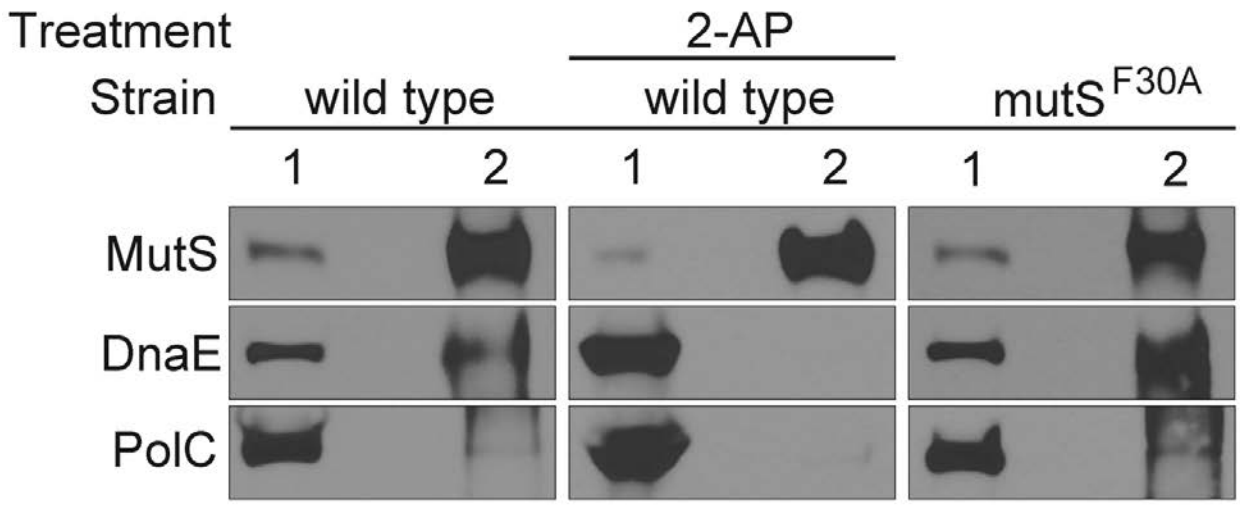


Figure 3.30 Co-IP of DnaE and PolC with MutS using affinity-purified antiserum directed against MutS. Lane 1: 5% input. Lane 2: anti-MutS immunoprecipitation**.**

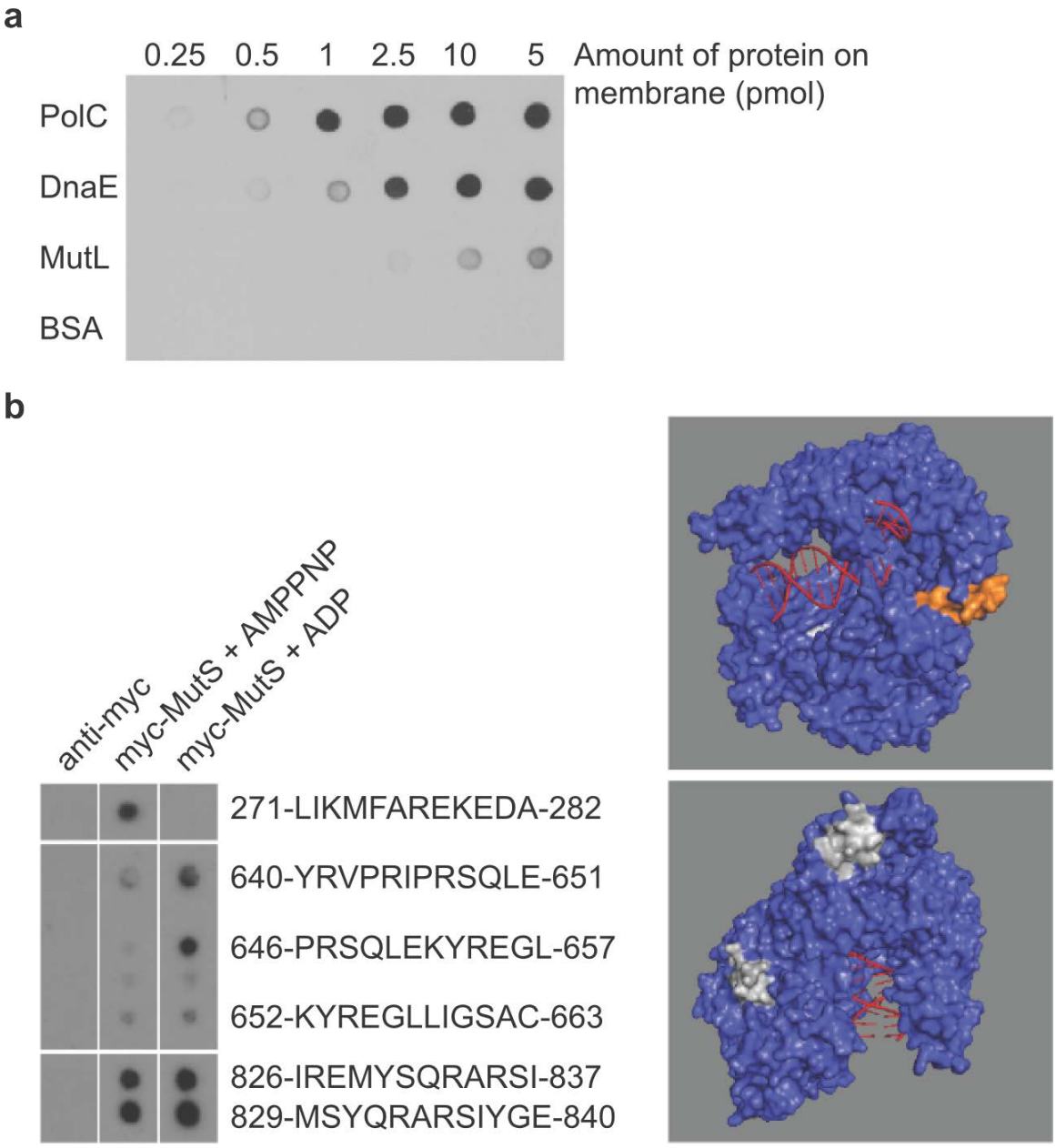


Figure 3.31 Interactions between MutS and PolC, DnaE and MutL *in vitro*. (a) Far-western blot with the indicated proteins bound to the membrane. BSA serves as a negative control. MutS bound to proteins on the membrane was detected using affinity-purified antisera. (b) Peptide array analysis of purified myc-MutS retained on a PolC peptide array. The peptides retaining myc-MutS are mapped to a model of *B. subtilis* PolC based on the structure of *Geobacillus kaustophilus* PolC (PMID: 3F2B)[66] using pymol. Regions of PolC corresponding to the peptides that retained myc-MutS are shaded gray (amino acids 640-663 and 826-840) and orange (amino acids 271-282). Myc-MutS retained by peptides was detected with monoclonal anti-myc antibodies.

3.13 Discussion

In the crowded cellular environment, 3D diffusion alone is too slow to allow for efficient detection of base-pairing mistakes[67]. Previous *in vitro* studies have demonstrated that 1D sliding is another mechanism by which MutS can locate DNA mismatches[7]. To further our understanding of the corresponding process from the *in vivo* perspective, and to complement and clarify our existing knowledge of MMR across species, we performed single-molecule super-resolution microscopy in live *Bacillus subtilis*. With nanometer-scale spatial resolution and millisecond-scale temporal resolution, we directly visualized and quantified the behavior of MutS *in vivo* in real time. Our results reveal that the heterogeneous behavior of MutS depends on mismatch binding state as well as intracellular location and ATPase activity. We have demonstrated that ATPase activity is necessary for the MutS cycle within the cell, and that if MutS cannot engage in nucleotide binding and hydrolysis, it is unable to properly localize to the replisome region *in vivo.* We show that there is a highly dynamic and transient interplay between MutS and the replisome which positions MutS to sites of ongoing DNA replication before mistakes occur such that MutS can constantly monitor the newly synthesized DNA. Such behavior is similar to that observed for the MutS homolog MutSα in eukaryotic cells[11], suggesting that replisome-association is a highly conserved mechanism for mismatch detection across species. On the other hand, unlike MutSα which may be able to bind mismatches independent of the replisome by engaging in Exo1-dependent MMR[11], our results obtained from *B. subtilis* suggest that mismatch binding by MutS must occur at the replisome. One possible explanation for this difference in mechanism between bacteria and *S. cerevisiae* is that bacterial DNA replication occurs continuously, each replication fork moves at 500 nt/sec, and in rich growth conditions, multiple rounds of replication initiation can occur prior to a single cell

division. On the other hand, in eukaryotic cells, DNA replication occurs only in S-phase, the replication fork moves at 27 nt/sec, and the behavior of biochemical pathways can be regulated in a cell cycle-dependent manner[15,50,68]. Therefore, mismatches must be detected very quickly in a rapidly proliferating bacterium like *B. subtilis* or they will become mutations within mere minutes when the next round of DNA replication duplicates the mismatched DNA. The large volumes of eukaryotic nuclei compared to typical bacterial cell volumes could also contribute to the emergence of this second mechanism to catch replication errors that have escaped initial proofreading at the replisome in a much larger cell volume. In contrast, in small bacterial cells with confined replisomes such as *B. subtilis*, near the replisome is arguably the best place for MutS to scan DNA for mismatches barrier-free. At the replisome, MutS has access to newly synthesized DNA largely free of proteins, and is also able to recognize rare replication errors as they are produced due to the spatial proximity between the two. Therefore, the replisome provides a scaffold that allows MutS to target a single mistake among tens of millions of correctly paired nucleotides in a timely manner, guarding the bacterial genome against mutations that could otherwise have deleterious effects on bacterial growth and fitness.

References

1 Jiricny, J. Postreplicative mismatch repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012633 (2013).

2 Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2774-2783 (2012).

3 Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **77**, 1 p following 166 (1994).

4 Pillon, M. C. *et al.* Structure of the endonuclease domain of MutL: unlicensed to cut. *Molecular cell* **39**, 145-151 (2010).

5 Kadyrov, F. A., Dzantiev, L., Constantin, N. & Modrich, P. Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* **126**, 297-308 (2006).

6 Tran, H. T., Gordenin, D. A. & Resnick, M. A. The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae. *Molecular and cellular biology* **19**, 2000-2007 (1999).

7 Gorman, J. *et al.* Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3074-3083 (2012).

8 Smith, B. T., Grossman, A. D. & Walker, G. C. Visualization of mismatch repair in bacterial cells. *Mol. Cell* **8**, 1197-1206 (2001).

9 Kleczkowska, H. E., Marra, G., Lettieri, T. & Jiricny, J. hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes & development* **15**, 724-736 (2001).

10 Lenhart, J. S., Sharma, A., Hingorani, M. M. & Simmons, L. A. DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. *Mol Microbiol* **87**, 553-568 (2013).

11 Hombauer, H., Campbell, C. S., Smith, C. E., Desai, A. & Kolodner, R. D. Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell* **147**, 1040-1053 (2011).

12 Klocko, A. D. *et al.* Mismatch repair causes the dynamic release of an essential DNA polymerase from the replication fork. *Mol. Microbiol.* **82**, 648-663 (2011).

13 Pluciennik, A., Burdett, V., Lukianova, O., O'Donnell, M. & Modrich, P. Involvement of the beta clamp in methyl-directed mismatch repair in vitro. *J. Biol. Chem.* **284**, 32782-32791 (2009).

14 Hombauer, H., Srivatsan, A., Putnam, C. D. & Kolodner, R. D. Mismatch repair, but not heteroduplex rejection, is temporally coupled to DNA replication. *Science* **334**, 1713-1716 (2011).

15 Tomimatsu, N. *et al.* Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nature communications* **5**, 3561 (2014).

16 Goellner, E. M. *et al.* PCNA and Msh2-Msh6 Activate an Mlh1-Pms1 Endonuclease Pathway Required for Exo1-Independent Mismatch Repair. *Molecular cell* (2014).

17 Alani, E., Sokolsky, T., Studamire, B., Miret, J. J. & Lahue, R. S. Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Molecular and cellular biology* **17**, 2436-2447 (1997).

18 Haber, L. T. & Walker, G. C. Altering the conserved nucleotide binding motif in the Salmonella typhimurium MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *The EMBO journal* **10**, 2707-2715 (1991).
19 Sharma, A., Doucette, C., Biro, F. N. & Hingorani, M. M. Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair. *Journal of molecular biology* **425**, 4192-4205 (2013).
20 Qiu, R. *et al.* Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. *The EMBO journal* **31**, 2528-2540 (2012).
21 Jeong, C. *et al.* MutS switches between two fundamentally distinct clamps during mismatch repair. *Nat. Struct. Mol. Biol.* **18**, 379-385 (2011).
22 Acharya, S. Mutations in the signature motif in MutS affect ATP-induced clamp formation and mismatch repair. *Mol Microbiol* **69**, 1544-1559 (2008).
23 Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods* **3**, 793-795 (2006).
24 Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645 (2006).
25 Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258-4272 (2006).
26 Yildiz, A. *et al.* Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061-2065 (2003).
27 Elmore, S., Muller, M., Vischer, N., Odijk, T. & Woldringh, C. L. Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in Escherichia coli. *J. Struct. Biol.* **151**, 275-287 (2005).
28 Biteen, J. S. & Moerner, W. E. Single-molecule and superresolution imaging in live bacteria cells. *Cold Spring Harb Perspect Biol* **2**, a000448 (2010).
29 Subach, F. V. *et al.* Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nature methods* **6**, 153-159 (2009).
30 Biteen, J. S. *et al.* Super-resolution imaging in live Caulobacter crescentus cells using photoswitchable EYFP. *Nature methods* **5**, 947-949 (2008).
31 Manley, S. *et al.* High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature methods* **5**, 155-157 (2008).
32 Haas, B. L., Matson, J. S., DiRita, V. J. & Biteen, J. S. Single-molecule tracking in live Vibrio cholerae reveals that ToxR recruits the membrane-bound virulence regulator TcpP to the toxT promoter. *Mol. Microbiol.* (2014).
33 Karunatilaka, K. S., Cameron, E. A., Martens, E. C., Koropatkin, N. M. & Biteen, J. S. Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts. *mBio* **5**, e02172-02114 (2014).
34 Pavani, S. R. *et al.* Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2995-2999 (2009).
35 Berkmen, M. B. & Grossman, A. D. Spatial and temporal organization of the Bacillus subtilis replication cycle. *Mol. Microbiol.* **62**, 57-71 (2006).
36 Lemon, K. P. Localization of Bacterial DNA Polymerase: Evidence for a Factory Model of Replication. *Science* **282**, 1516-1519 (1998).
37 Reyes-Lamothe, R., Sherratt, D. J. & Leake, M. C. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science* **328**, 498-501 (2010).

38 Holden, S. J., Uphoff, S. & Kapanidis, A. N. DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nature methods* **8**, 279-280 (2011).
39 Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster STORM using compressed sensing. *Nature methods* **9**, 721-723 (2012).
40 Burnette, D. T., Sengupta, P., Dai, Y., Lippincott-Schwartz, J. & Kachar, B. Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21081-21086 (2011).
41 Simonson, P. D., Rothenberg, E. & Selvin, P. R. Single-molecule-based super-resolution images in the presence of multiple fluorophores. *Nano letters* **11**, 5090-5096 (2011).
42 Watkins, L. P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* **109**, 617-628 (2005).
43 Imai, Y. *et al.* Subcellular localization of Dna-initiation proteins of Bacillus subtilis: evidence that chromosome replication begins at either edge of the nucleoids. *Mol. Microbiol.* **36**, 1037-1048 (2000).
44 Reyes-Lamothe, R., Possoz, C., Danilova, O. & Sherratt, D. J. Independent positioning and action of Escherichia coli replisomes in live cells. *Cell* **133**, 90-102 (2008).
45 Lakadamyali, M., Rust, M. J., Babcock, H. P. & Zhuang, X. Visualizing infection of individual influenza viruses. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9280-9285 (2003).
46 Simmons, L. A., Davies, B. W., Grossman, A. D. & Walker, G. C. Beta clamp directs localization of mismatch repair in Bacillus subtilis. *Mol. Cell* **29**, 291-301 (2008).
47 Lewis, J. P. Fast Normalized Cross-Correlation. *Vision interface* **10**, 120-123 (1995).
48 Goodman, M. F., Hopkins, R. & Gore, W. C. 2-Aminopurine-induced mutagenesis in T4 bacteriophage: a model relating mutation frequency to 2-aminopurine incorporation in DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4806-4810 (1977).
49 Ronen, A. 2-Aminopurine. *Mutat. Res.* **75**, 1-47 (1979).
50 Wang, J. D., Sanders, G. M. & Grossman, A. D. Nutritional control of elongation of DNA replication by (p)ppGpp. *Cell* **128**, 865-875 (2007).
51 Schroeder, J. W. & Simmons, L. A. Complete Genome Sequence of Bacillus subtilis Strain PY79. *Genome Announc* **1** (2013).
52 Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8452-8457 (2014).
53 Durisic, N., Laparra-Cuervo, L., Sandoval-Alvarez, A., Borbely, J. S. & Lakadamyali, M. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nature methods* **11**, 156-162 (2014).
54 Wang, W., Li, G. W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445-1449 (2011).
55 Lenhart, J. S., Sharma, A., Hingorani, M. M. & Simmons, L. A. DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. *Mol. Microbiol.* **87**, 553-568 (2013).
56 Schutz, G. J., Schindler, H. & Schmidt, T. Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophys. J.* **73**, 1073-1080 (1997).
57 Gebhardt, J. C. *et al.* Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature methods* **10**, 421-426 (2013).

58 Liao, Y., Yang, S. K., Koh, K., Matzger, A. J. & Biteen, J. S. Heterogeneous single-molecule diffusion in one-, two-, and three-dimensional microporous coordination polymers: directional, trapped, and immobile guests. *Nano letters* **12**, 3080-3085 (2012).
59 Saxton, M. J. Single-particle tracking: the distribution of diffusion coefficients. *Biophys. J.* **72**, 1744-1753 (1997).
60 Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564-1565 (2009).
61 Lenhart, J. S., Pillon, M. C., Guarne, A. & Simmons, L. A. Trapping and visualizing intermediate steps in the mismatch repair pathway in vivo. *Mol. Microbiol.* **90**, 680-698 (2013).
62 Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P. & Yang, W. Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Molecular cell* **7**, 1-12 (2001).
63 Dervyn, E. *et al.* Two essential DNA polymerases at the bacterial replication fork. *Science* **294**, 1716-1719 (2001).
64 Sanders, G. M., Dallmann, H. G. & McHenry, C. S. Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases. *Mol. Cell* **37**, 273-281 (2010).
65 Walsh, B. W. *et al.* RecD2 helicase limits replication fork stress in Bacillus subtilis. *J. Bacteriol.* (2014).
66 Evans, R. J. *et al.* Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20695-20700 (2008).
67 von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675-678 (1989).
68 Sekedat, M. D. *et al.* GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Molecular systems biology* **6**, 353 (2010).

Chapter 4 Three-Dimensional Visualization of DNA Polymerase Dynamics in Live Cells[a]

4.1 Introduction: replisome structure and dynamics in *Escherichia coli* and *Bacillus subtilis*

In bacterial cells, replisomes are multiprotein assemblies that include the DNA polymerases and various associated proteins that function together during DNA replication. Both the architecture and dynamics of the *E. coli* replisome have been well characterized *in vitro* and *in vivo*[1-3].

There are five types of DNA polymerases in *E. coli*: DNA polymerase I is the best understood DNA polymerase of any kind and has both DNA replication and repair activities. DNA polymerases II, IV and V are SOS induced, and DNA polymerase III (Pol III) is the primary enzyme responsible for chromosomal replication and also has 3' to 5' exonuclease activity for proofreading[4]. *E. coli* Pol III is a part of the ten-protein Pol III holoenzyme assembly that can be divided into three functional subunits: the Pol III core enzymes, the β sliding clamp and the clamp loader. The Pol III core enzymes themselves consist of three subunits: α (encoded by *dnaE* and has DNA polymerease activity), ε (encoded by *dnaQ* and has exonuclease activity) and θ (encoded by *holE* and is involved in proofreading). The β sliding clamp is encoded by *dnaN* and binds DNA polymerase to prevent it from disassociating from the template strand during DNA synthesis. The presence of the β sliding clamp significantly enhances the processivity of the DNA polymerase and increases the rate of DNA synthesis up to a thousand fold[5]. The sliding clamp does not assemble onto DNA spontaneously, however; rather, it requires

[a] In collaboration with Jeremy W. Schroeder and Lyle A. Simmons. Manuscript in preparation.

the clamp loader, a multiprotein complex, to open and close sliding clamps around DNA by using energy from ATP. Also, the Pol III holoenzymes are connected to the DNA helicase DnaB through the C-terminal region of the τ subunit of the clamp loader (τ_c). The physical connections between both Pol III holoenzymes and DnaB suggest that, although lagging strand synthesis proceeds in the opposite direction from leading strand synthesis, the Pol III on the lagging strand must move together with the Pol III on the leading strand and the helicase, and therefore the lagging strand must form a loop near the replication fork to accommodate this reverse motion[3].

Before DNA replication starts, and in the presence of the replication initiation proteins DnaA and DnaC, the *E. coli* replisome can assemble at the origin of replication *oriC* regardless of where *oriC* is located in the cell. In other words, the only deciding factor for replisome positioning in *E. coli* is *oriC*, and the replisome is not attached to any other cellular structure. As a direct consequence, DNA replication can initiate at any *oriC* in a cell, whether it is on a chromosome or a plasmid. Approximately 5 minutes later, sister replication forks are partitioned into different cell halves and continue to synthesize DNA independently by tracking along DNA until the end of the replication, when the two sister forks come back to each other at mid-cell again[1].

Using slimfield fluorescence microscopy and photobleaching analysis, Reyes-Lamothe et al determined the stoichiometry of the *E. coli* replisome components[2]. They found three polymerases—as opposed to the traditionally accepted two—positioned at each sister replication fork. Three dimers of the β sliding clamp are also present at each replication fork. However, in most cases, only two of the three polymerases are associated with a sliding clamp, suggesting that the third polymerase waits to be loaded onto the lagging strand to synthesize the next

Okazaki fragment. Three τ subunits of the clamp loader are also found at each replication fork, and they bind and trimerize Pol III.

Although *E. coli* has served as a prototype for understanding DNA synthesis *in vivo*, and although some *E. coli* DNA replication features are conserved across species, the replisomes of many other species have distinct organizations and modes of operation. The *B. subtilis* replisome, introduced in Chapter 3, does not replicate DNA by actively tracking along DNA as the *E. coli* replisome. Rather, the *B. subtilis* replisome behaves like a stationary spooling machine through which template DNA is pulled in and newly synthesized DNA is extruded[6-8]. In addition, the Gram-positive bacterium *B. subtilis* utilizes two distinct types of essential DNA polymerases PolC and DnaE for genome replication, as opposed to one in *E. coli*. *In vitro* reconstitution of the *B. subtilis* replisome has demonstrated that PolC is responsible for all leading strand synthesis as well as most of the lagging strand synthesis[9], whereas the more error prone and much slower DnaE (75 nt/s for DnaE compared to ~500 nt/s for PolC) is crucial for lagging strand synthesis: DnaE extends the lagging strand RNA primer before handing off to PolC, which finishes replicating the remaining segment of the Okazaki fragment[9]. The synergistic relation between different polymerases in the *B. subtilis* replisome resembles that found in eukaryotic systems. In eukaryotes including human cells[10], two essential replicases, Pol ε and Pol δ are respectively responsible for synthesizing the leading and the lagging strands. In analogy with DnaE in *B. subtilis*, the human polymerase Pol α extends RNA primers for a short segment before handing off to Pol δ.[11]

Given these observations, it is clear that *E. coli*-type DNA replication is not universally conserved, and the less well-understood *B. subtilis* replisome appears to exhibit greater similarity with replisomes found in eukaryotic systems. Considering the crucial role DNA replication plays

for the survival of species, and the fact that about 80 human diseases and genetic disorders are linked to malfunctions in DNA replication[12], a deeper understanding of how DNA replication functions in cells will have profound implications for human health. In this chapter, we employ super-resolution microscopy to investigate the *in vivo* behavior of one of the two essential DNA polymerases in *B. subtilis*, PolC. We fused PolC at its native genetic locus to the red fluorescent protein mCherry to learn about the stoichiometry of PolC at the replisome as well as within the whole cell. We also natively tagged PolC with PAmCherry to probe the *in vivo* localization and intracellular dynamics in three-dimensions (3D) of PolC. Our experiments highlight the dynamical behavior of PolC as it engages in DNA replication.

4.2 Stoichiometry of PolC

The red fluorescent protein mCherry is fused to PolC using the Gibson assembly method[13] for integration at the native genetic locus, and the copy number of PolC in each cell was estimated from the average fluorescence intensity of each cell. To correct for fluorescence intensity contribution from cellular background fluorescence, the average autofluorescence intensity obtained from wild type PY79 cells is subtracted from the average fluorescence intensity collected from cells with PolC-mCherry; this difference gives the fluorescence intensity of PolC-mCherry. The number of PolC molecules per cell is then obtained by dividing this background-subtracted PolC-mCherry fluorescence by the intensity of a single PolC-mCherry molecule, which is calculated by averaging the intensities of the last 1-5 copies of PolC-mCherry left after photobleaching. Our initial estimation from 19 cells indicates on average ~101 copies of PolC are present within each cell.

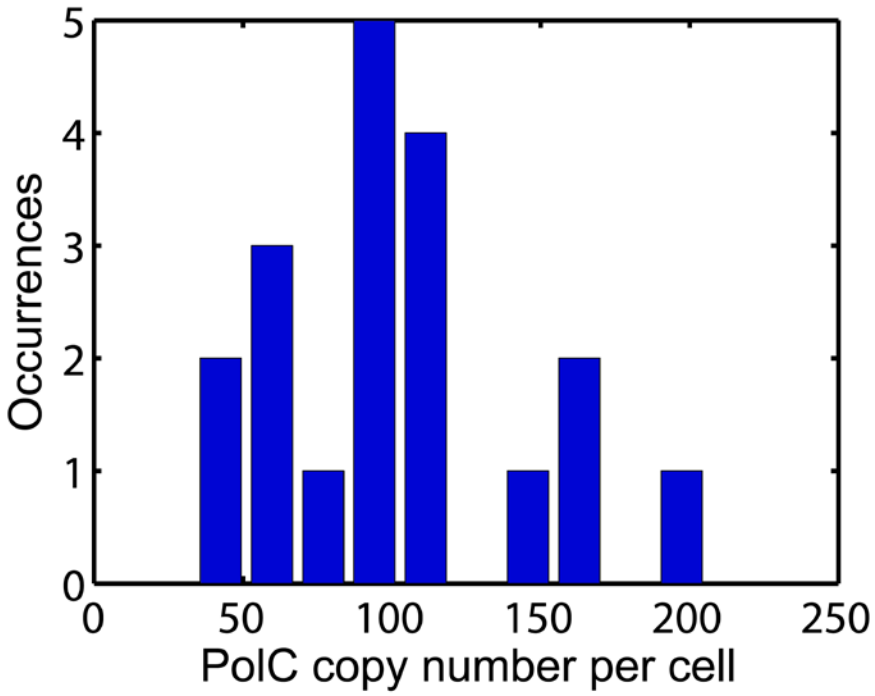


Figure 4.1 Distribution of PolC copy number within a whole cell.

As expected for a stationary replisome, and similar to the clamp loader protein DnaX described in Chapter 2, PolC-mCherry forms distinct foci at the mid-cell position or near the quarter-cell positions. These foci represent the active site(s) of DNA replication. Recent photobleaching experiments in *E. coli* have challenged the traditional view and pointed out that three, instead of two, copies of Pol III are positioned at each replication fork[2]. It is therefore equally interesting to probe the stoichiometry of the essential DNA polymerase PolC in *B. subtilis*. With PolC-mCherry, we applied the same approach of photobleaching-assisted microscopy[14,15] as we did for DnaX-mCitrine in Chapter 3, but instead of learning about the inter-molecular separation distance, this time we focus on the stoichiometry of PolC at each replication fork (Figure 4.2). The distribution of PolC-mCherry copy number indicates 3 copies of PolC are present at the replication fork. In the case of two sister replication forks spatially overlap, 6 copies of PolC might be found at each site of replication. However, presumably because cells are harvested for imaging during exponential phase and the sister forks are separated at each quarter cell position, in most cases three copies of PolC are found at each fork as opposed to six. Together with our estimation of about a hundred copies of PolC present within each cell, these findings pose an intriguing question of what is the function of those PolC that are

not associated with the replication fork.

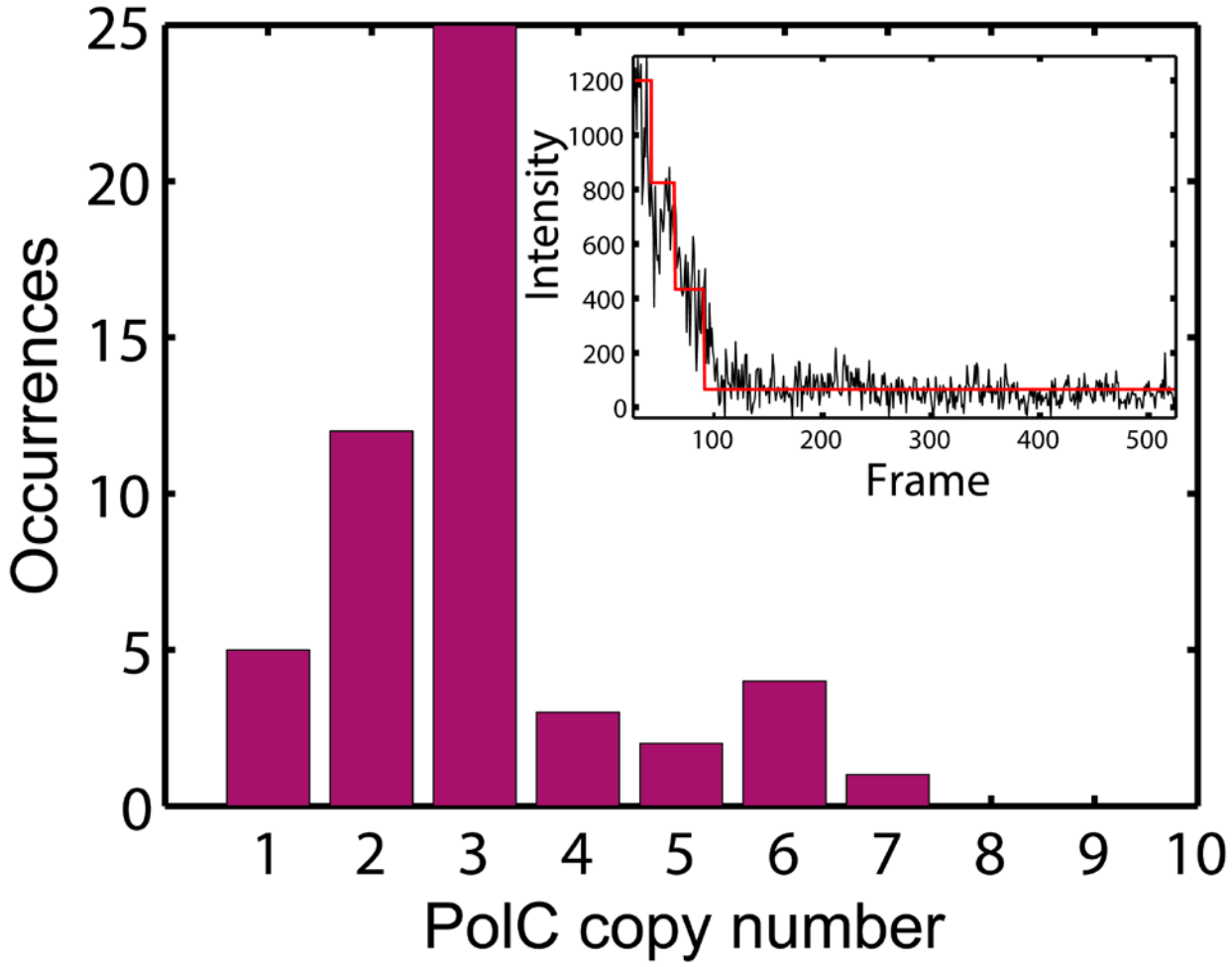


Figure 4.2 Stoichiometry of PolC at replication fork. A sample photobleaching intensity trace of PolC-mCherry is shown in the inset.

4.3 Three-dimensional super-resolution microscopy

To visualize PolC at the single-molecule level, the PolC-mCherry fusions in Section 4.2 were replaced by PolC-PAmCherry. Furthermore, to capture the *in vivo* positioning and dynamics of single PolC molecules unambiguously, we employed 3D super-resolution microscopy to simultaneously resolve the lateral (*x* and *y*) as well as the axial (*z*) position of single PolC molecules.

To extract the *z*-position of a molecule from a 2D image, the microscope PSF must be sensitive to changes in the axial position of a fluorescent point source. Several optical configurations can be used to achieve this axial sensitivity. For example, the biplane (BP) FPALM set-up splits the single-molecule emission signal into two equal halves, which are projected onto separate regions of the camera corresponding to different object planes[16]. Because the split signals travel different paths until reaching the camera, the combined PSF from the two halves depends on the axial location of the molecule with respect to the real object plane. Alternatively, using a spatial light modulator with a specifically designed phase-mask that multiplies the emission signal at the Fourier plane with a double-helix point spread function (DH-PSF), the Moerner group has engineered an imaging configuration that renders single-molecule emission as two rotating lobes, which encode the 3D position of the molecule in the relative orientation of these two lobes[17]. Recently, using this similar approach of manipulating PSFs at the Fourier plane, the Zhuang group has reported a self-bending point spread function (SB-PSF) that is suitable for 3D imaging over a depth of several micrometers and with comparable localization precisions in the lateral and axial planes[18].

Here, we implemented the astigmatism-based[19] 3D super-resolution imaging to visualize PolC. By inserting a weak cylindrical lens into the emission pathway (Figure 1.3), the emission rays are focused by the cylindrical lens only in one direction (but not the other). This introduces two slightly different focal planes for the two orthogonal image components and creates a PSF with a shape that depends on the axial-position of the fluorophore. When the fluorophore is at the average focal plane, that is, halfway between the two focal planes, its image is a circular spot similar to the PSF observed in typical 2D single-molecule fluorescence imaging experiments. The images of molecules residing above the average focal plane becomes elliptical along the x-

axis because the focal plane for the x-direction lies in front of the detector. In contrast, the images of fluorophores below the average focal plane are elongated along the y-direction. Thus, the direction and extent of PSF ellipticity caused by astigmatism depend on the axial location of the fluorophore and are characterized by σ_x and σ_y in a 2D elliptical Gaussian function:

$$f = I_{bg} + A \cdot e^{-\left[\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right]} \tag{4.1}$$

where I_{bg}, A, x_0 and y_0 respectively denote the background intensity, the amplitude of emission and the x- and y-center of the molecule. By fitting a single-molecule intensity profile to this equation, the width parameters σ_x and σ_y are obtained. To determine the axial position of the molecule, we compare σ_x and σ_y to a width vs. z-location profile pre-calibrated using 0.1 μm TetraSpeck™ beads (Figure 4.3). Specifically, we search along the calibration curves for a z location that minimizes the distance D between (σ_x, σ_y) and ($\sigma_{x\text{-calibration}}$, $\sigma_{y\text{-calibration}}$)[19,20]:

$$D = \sqrt{(\sigma_x^{0.5} - \sigma_{x-\text{calibration}}^{0.5})^2 + (\sigma_y^{0.5} - \sigma_{y-\text{calibration}}^{0.5})^2} \tag{4.2}$$

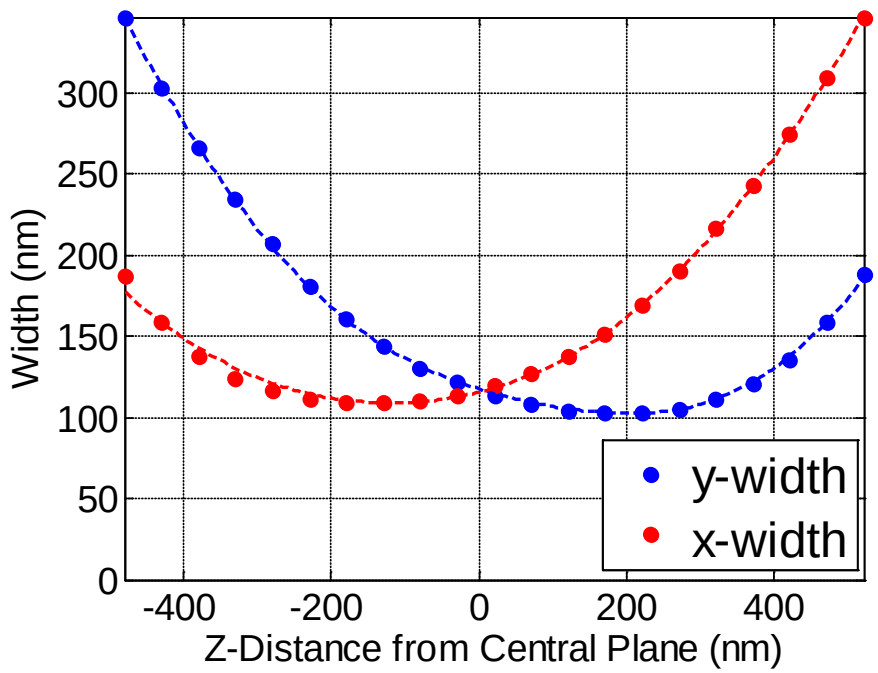


Figure 4.3 *x*- and *y*-direction fit widths vs. *z*-position calibration curves for 3D super-resolution microscopy. The *z*-position of a single molecule can be determined by comparing the widths of its PSF with this calibration profile.

In cellular imaging, the fluorophores are located in an aqueous environment, which has a different refractive index from the immersion oil and the glass coverslip over which the cells are kept. This refractive index mismatch at the glass-water interface induces a focal shift and causes the apparent *z*-position of the molecule to be overestimated. Within several micrometers from the glass surface, the discrepancy between the apparent location *d'* and the real location of a molecule *d* can be approximated as a linear relation, that is, $d = m \times d'$, where *m* is a correction constant. To determine *m*, we used the diffraction integral[21-23] to numerically simulate the 3D PSF of a single molecule immersed in $S7_{50}$ medium with a refractive index $n_{S750} = 1.35$. Figure 4.4 illustrates the PSF intensity on the *x*-*z* plane as the light prorogates from the glass-water

interface (bottom) to deep into the aqueous environment (top). The left graph shows a symmetric PSF in an ideal case where there is no refractive index mismatch. However, as shown to the right, when the molecule is imaged in a typical configuration: through an oil-immersion objective with $n_{oil} = 1.515$, the PSF becomes asymmetric and appears to be farther from the glass-water interface. From the simulation, we calculated $m = 0.79$, similar to previously reported values from experiments performed in aqueous environments[19,20,24].

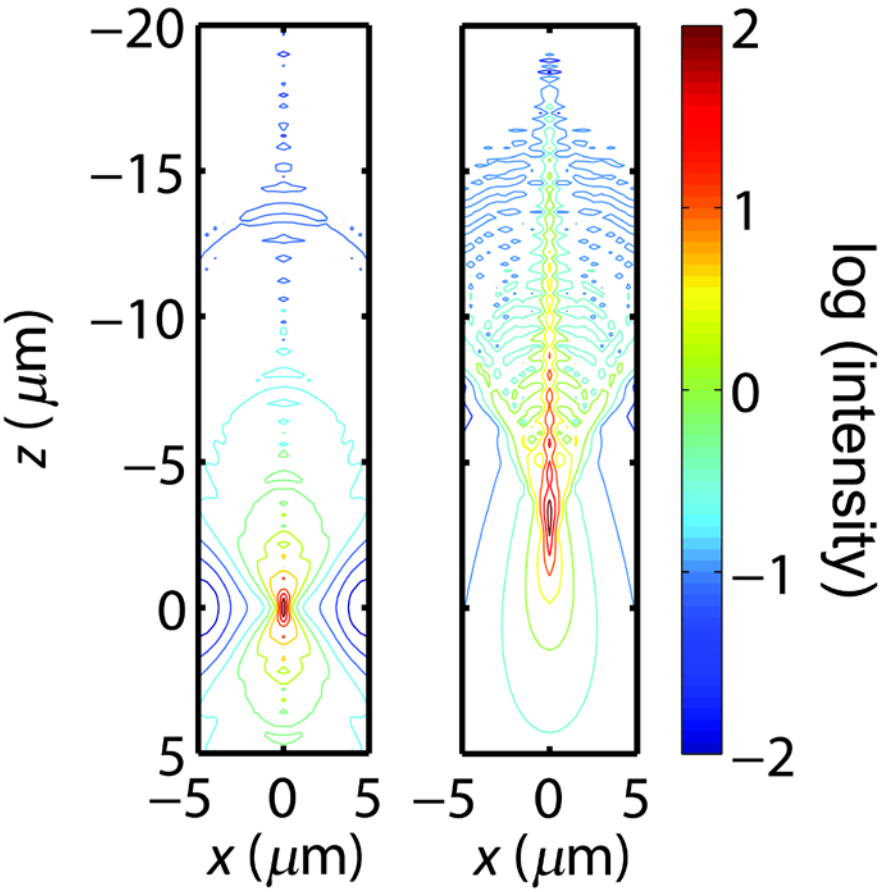


Figure 4.4 Contour plots of 3D single-molecule PSF along the *x*-*z* plane, without (left) and with (right) refractive index mismatch.

4.4 Localization and dynamics of PolC in live *B. subtilis* cells

From 3D super-resolution microscopy, we are able to localize individual PolC-PAmCherry molecules with nanometer precision both on the lateral plane as well as along the *z*-axis. Similar to the MutS behavior described in Chapter 3, a population of PolC aggregates near the mid-cell or quarter-cell positions, whereas the rest diffuses throughout the cell (Figure 4.5). Furthermore, 3D single-particle tracking can capture the dwelling and leaving behavior of PolC.

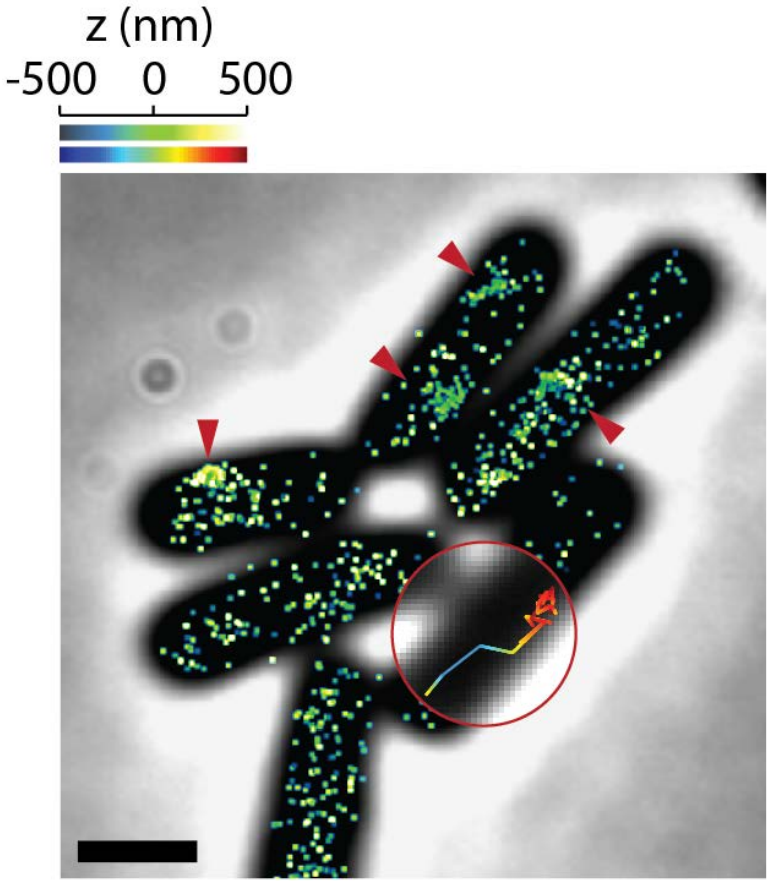


Figure 4.5 3D PALM reconstruction image of PolC-PAmCherry in live *B. subtilis*, overlaid on a phase contrast image of the *B. subtilis* cells. The position of each localization is indicated by a single dot with a width corresponding to the localization precision on the lateral plane (50 nm). The axial (*z*) position is color-coded according to the color bar shown above. Red arrows highlight the aggregation of PolC in the cell. A color-coded 3D single-molecule trajectory is also shown to illustrate the dwelling (replicating) and diffusing (leaving) behavior of a PolC-PAmCherry molecule, presumably while it undergoes polymerase exchange. Scale bar = 1 μm.

During DNA synthesis, polymerases including PolC undergo polymerase exchange: they continuously associate with and dissociate from the replication fork. Because lagging strands are synthesized in short Okazaki Fragments, polymerases on the lagging strands are expected to undergo polymerase exchange more frequently than those on the leading strands, and in *B. subtilis* the rate of polymerase exchange may be affected by the concentration of primase at the replication fork[9]. The dynamics of polymerase at the replication fork remains a long-standing

question in biology. Equipped with the ability to capture the motion of single PolC-PAmCherry molecules, we were able to investigate the rate of polymerase exchange for PolC in live *B. subtilis*. Due to the limited photostability of fluorescent proteins, it is not possible to capture long-lasting dwelling behavior of PolC merely relying on continuous laser illumination and image acquisition, therefore we turned to time-lapse imaging to measure dwell times. In time-lapse imaging mode, every frame is still captured with a 50-ms integration time (τ_{int}), but a time delay (τ_{delay}) is introduced between each pair of consecutive frames. The time-lapse period τ_{tl} is defined as the sum of τ_{int} and τ_{delay}. We define a step size in three dimensions equal to or smaller than 100 nm between consecutive frames as dwelling, and plot the probability of finding a dwelling PolC molecule at any position within a normalized cell as symmetrized probability density maps (Section 3.6) in Figure 4.6. Because the stages of DNA replication and cell division generally correlate well with cell length[1], we analyzed the position of PolC dwelling as a function of the aspect ratio of cell shapes. In shorter cells, most PolC dwells at the cell center. In midsized cells that are in the pre-divisional stage, PolC most likely dwells at the quarter cell positions, consistent with our previous studies of DnaX (Chapter 3). In longer cells, multiple PolC dwelling sites are found throughout the cells. This final localization pattern is likely caused by initiation of new rounds of DNA replication before cell division occurs, a scenario typically found in fast growing cells.

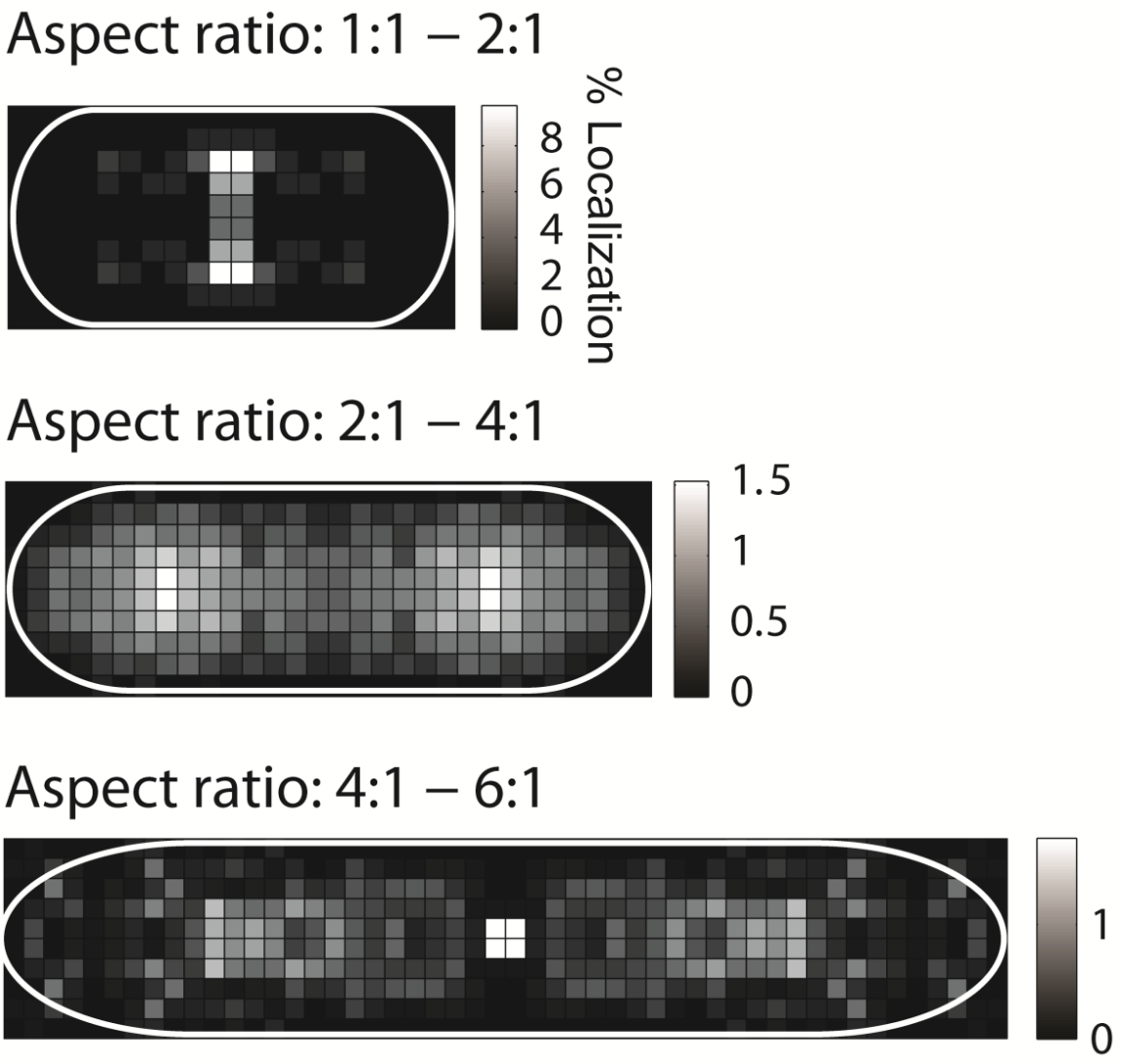


Figure 4.6 Probability density maps for dwelling PolC (step size ≤ 100 nm), analyzed separately for shorter, medium and longer cells. The density maps are symmetrized and normalized such that the total probability adds up to 100%. Over 1000 dwelling events are used to construct the density maps of the medium and the longer cells, and 30 dwelling events are used for the shorter cells. The dwelling probability for the shorter cells does not peak at the midpoint along the transverse cellular axis, presumably due to the limited amount of data available.

To probe the time scale of the dwelling behavior of PolC, we plot the distributions of PolC dwell time (τ, the length of time during which the molecule continuously takes steps of ≤ 100 nm) in five histograms (Figure 4.7), each of which corresponds to one time-lapse imaging

experiment with a different τ_{tl}. In each case, the distribution follows an exponential decay function:

$$f = Ae^{\frac{-\tau}{k_{eff}}} \tag{4.3}$$

where k_{eff} is the effective off-rate of PolC. k_{eff} characterizes the rate of apparent PolC dissociation, which includes contributions from two independent Poisson processes: photobleaching of PAmCherry with a rate constant of k_b, and PolC dissociation with a rate constant of k_{off} [25]:

$$k_{eff} \cdot \tau_{tl} = k_b \cdot \tau_{int} + k_{off} \cdot \tau_{tl} \tag{4.4}$$

As indicated by Equation (4.4), the relation between $k_{eff} \cdot \tau_{tl}$ and τ_{tl} is linear. The dwell time constant τ_{dwell} of PolC can thus be extracted by plotting $k_{eff} \cdot \tau_{tl}$ vs. τ_{tl}: the *y*-intercept of this equation yields the photobleaching rate constant k_b, and the slope corresponds to the real dissociation rate constant k_{off}, which is the reciprocal of the dwell time constant of PolC.

Currently we are measuring dwell times on the order of a second, though more data collected with longer τ_{tl} are needed in order to accurately quantify the time scale of PolC dwelling. However, considering that DNA polymerases including PolC undergo polymerase exchange during DNA synthesis, we hypothesize that the measured dwell time constant of PolC should be a function of either (1) the amount of time required to synthesize a single Okazaki fragment on the lagging strand (1-2 kbp synthesized ~500 bp/s = 2-4 s) or (2) the time duration over which the leading-strand-synthesizing PolC remains associated with the DNA in the presence of processivity clamp (predicted to be significantly longer).

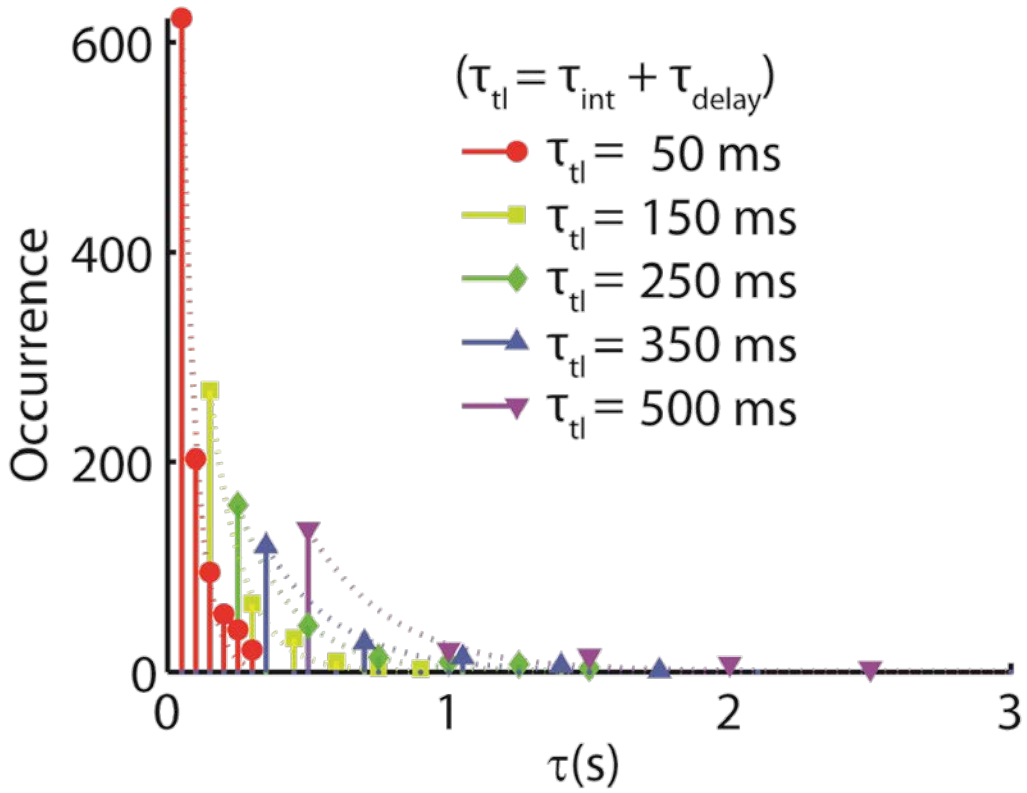


Figure 4.7 Dwell time distributions for PolC. For clarity, the distributions are shown as stem plots, and a color is assigned to each time-lapse image experiment with the respective τ_{tl} value listed in the legend. The dotted lines are exponential fits according to Equation 4.3.

4.5 Discussion

B. subtilis is an ideal system in which to study the DNA replication machinery, in part because the *B. subtilis* organization and lagging-strand synthesis mechanism differ significantly from those of the well-studied *E. coli*, but are similar to those found in higher organisms. In this chapter we employed single-molecule photobleaching-assisted microscopy and 3D PALM experiments to characterize the stoichiometry and localization of the essential DNA replicase PolC in *B. subtilis*. The copy number of PolC at each replisome is determined to be 3, which is different from the traditionally accepted 2 but which agrees with recent findings in *E. coli*[2]. The positioning of PolC aggregates in cells agrees well with the positioning of the clamp loader protein DnaX studied in Chapter 3, and again supports the notion that the *B. subtilis* replisome remains largely confined during DNA replication. However, PALM experiments also elucidated a significant population of PolC that is highly diffusive outside the replisome region; this high copy number suggests that the replisome functions as a scaffold where constituent protein subunits continuously undergo association and disassociation. We also showed with time-lapse imaging that it is possible to quantify the PolC dwelling time at the replisome, and we postulate that the dynamic association and disassociation observed for PolC at the replisome likely correspond to polymerases engaged in synthesizing Okazaki fragments on the lagging strand and/or the leading strand. Because Okazaki fragment synthesis by PolC is dependent on the presence of the other essential polymerase DnaE at the RNA primer, we hypothesize that the measured PolC dwelling time constant will be similar to that of the DnaE, and simultaneous two-color time-lapse imaging of PolC and DnaE will help to address this question unambiguously.

References

1 Reyes-Lamothe, R., Possoz, C., Danilova, O. & Sherratt, D. J. Independent positioning and action of Escherichia coli replisomes in live cells. *Cell* **133**, 90-102 (2008).

2 Reyes-Lamothe, R., Sherratt, D. J. & Leake, M. C. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science* **328**, 498-501 (2010).

3 O'Donnell, M. Replisome architecture and dynamics in Escherichia coli. *J. Biol. Chem.* **281**, 10653-10656 (2006).

4 Kelman, Z. & O'Donnell, M. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* **64**, 171-200 (1995).

5 Mizrahi, V., Henrie, R. N., Marlier, J. F., Johnson, K. A. & Benkovic, S. J. Rate-limiting steps in the DNA polymerase I reaction pathway. *Biochemistry* **24**, 4010-4018 (1985).

6 Sawitzke, J. & Austin, S. An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol. Microbiol.* **40**, 786-794 (2001).

7 Lemon, K. P. & Grossman, A. D. Movement of replicating DNA through a stationary replisome. *Mol. Cell* **6**, 1321-1330 (2000).

8 Meile, J. C., Wu, L. J., Ehrlich, S. D., Errington, J. & Noirot, P. Systematic localisation of proteins fused to the green fluorescent protein in Bacillus subtilis: identification of new proteins at the DNA replication factory. *Proteomics* **6**, 2135-2146 (2006).

9 Sanders, G. M., Dallmann, H. G. & McHenry, C. S. Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases. *Mol. Cell* **37**, 273-281 (2010).

10 Kunkel, T. A. Balancing eukaryotic replication asymmetry with replication fidelity. *Curr. Opin. Chem. Biol.* **15**, 620-626 (2011).

11 Kunkel, T. A. & Burgers, P. M. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol.* **18**, 521-527 (2008).

12 DePamphilis. *DNA Replication and Human Disease*. (Cold Spring Harbor Laboratory Press, 2006).

13 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345 (2009).

14 Burnette, D. T., Sengupta, P., Dai, Y., Lippincott-Schwartz, J. & Kachar, B. Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21081-21086 (2011).

15 Simonson, P. D., Rothenberg, E. & Selvin, P. R. Single-molecule-based super-resolution images in the presence of multiple fluorophores. *Nano letters* **11**, 5090-5096 (2011).

16 Juette, M. F. *et al.* Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nature methods* **5**, 527-529 (2008).

17 Pavani, S. R. *et al.* Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2995-2999 (2009).

18 Jia, S., Vaughan, J. C. & Zhuang, X. Isotropic 3D Super-resolution Imaging with a Self-bending Point Spread Function. *Nature photonics* **8**, 302-306 (2014).

19 Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813 (2008).

20 Huang, B., Jones, S. A., Brandenburg, B. & Zhuang, X. Whole-cell 3D STORM reveals interactions between cellular structures with nanometer-scale resolution. *Nature methods* **5**, 1047-1052 (2008).
21 Hell, S., Reiner, G., Cremer, C. & Stelzer, E. H. K. Aberrations in Confocal Fluorescence Microscopy Induced by Mismatches in Refractive-Index. *J. Microsc.* **169**, 391-405 (1993).
22 Torok, P., Varga, P., Laczik, Z. & Booker, G. R. Electromagnetic Diffraction of Light Focused through a Planar Interface between Materials of Mismatched Refractive-Indexes - an Integral-Representation. *Journal of the Optical Society of America a-Optics Image Science and Vision* **12**, 325-332 (1995).
23 Torok, P., Varga, P. & Nemeth, G. Analytical Solution of the Diffraction Integrals and Interpretation of Wave-Front Distortion When Light Is Focused through a Planar Interface between Materials of Mismatched Refractive-Indexes. *Journal of the Optical Society of America a-Optics Image Science and Vision* **12**, 2660-2671 (1995).
24 Biteen, J. S., Goley, E. D., Shapiro, L. & Moerner, W. E. Three-dimensional super-resolution imaging of the midplane protein FtsZ in live Caulobacter crescentus cells using astigmatism. *Chemphyschem* **13**, 1007-1012 (2012).
25 Gebhardt, J. C. *et al.* Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature methods* **10**, 421-426 (2013).

Chapter 5 Heterogeneous Single-Molecule Diffusion in Crystalline Microporous Coordination Polymers[1]

5.1 Introduction: Microporous Coordination Polymers (MCPs)

Crystalline microporous coordination polymers (MCPs), a member of the superclass of materials known as metal-organic frameworks (MOFs), are being developed at a rapid pace due to their unprecedented properties. MCPs are built with organic linkers that assemble metals or metal clusters in a periodic fashion to form a crystalline host structure. The structural regularity is particularly important for applications such as liquid phase separation[1] and gas storage.[2] Compared with other conventional porous solids including zeolites, metal oxides, or activated carbons, MCPs have a much broader and more controllable accessible pore size range (~0.35 – 3.4 nm), tremendously high pore volumes, and myriad pore geometries. MCPs have exceptional potential for conducting difficult separations of molecular and polymeric species if the transport, and in particular the anisotropic character of molecular diffusion, in these materials can be understood and harnessed. Interactions between MCP hosts and introduced guests have mainly been inferred from static X-ray diffraction experiments.[3,4] Probing dynamic movement of guest molecules in porous structures thus remains an important challenge as the various diffusion modes of guests are influenced by pore morphology and will profoundly affect separation ability.

Here we illuminate the diffusion behavior of the organic probe Nile red in MCPs with various pore geometries. We investigate three hosts from the University of Michigan Crystalline Material (UMCM) family that are built from $Zn_4O(CO_2R)_6$ clusters: UMCM-1,[5] UMCM-2,[6] and

[1] This work was published on May 2, 2012 in Nano Letters. DOI: 10.1021/nl300971t

UMCM-4.[7] UMCM-1 consists of Zn_4O clusters linked together by the organic linkers 1,4-benzenedicarboxylate (BDC) and benzene-1,3,5-tribenzoate (BTB). In the structure of UMCM-1, microporous cages assemble to build a mesoporous channel which should enable facile one-dimensional transport of guests (Figure 5.1a). In contrast, the organic linkers thieno[*3,2-b*]thiophene-2,5-dicarboxylate (T^2DC) and BTB in UMCM-2 form three types of cages with restricted pore apertures (Figure 5.1b); this structure is expected to allow restricted three dimensional diffusion of guests. Finally, UMCM-4 is prepared from BDC and 4,4',4''-nitrilotribenzoate (NTB) in the presence of Zn^{2+} and has a layered structure.[7] These layers are connected by BDC linker pillars, and the space between the layers is expected to have an aperture large enough to allow two-dimensional guest motion (Figure 5.1c).

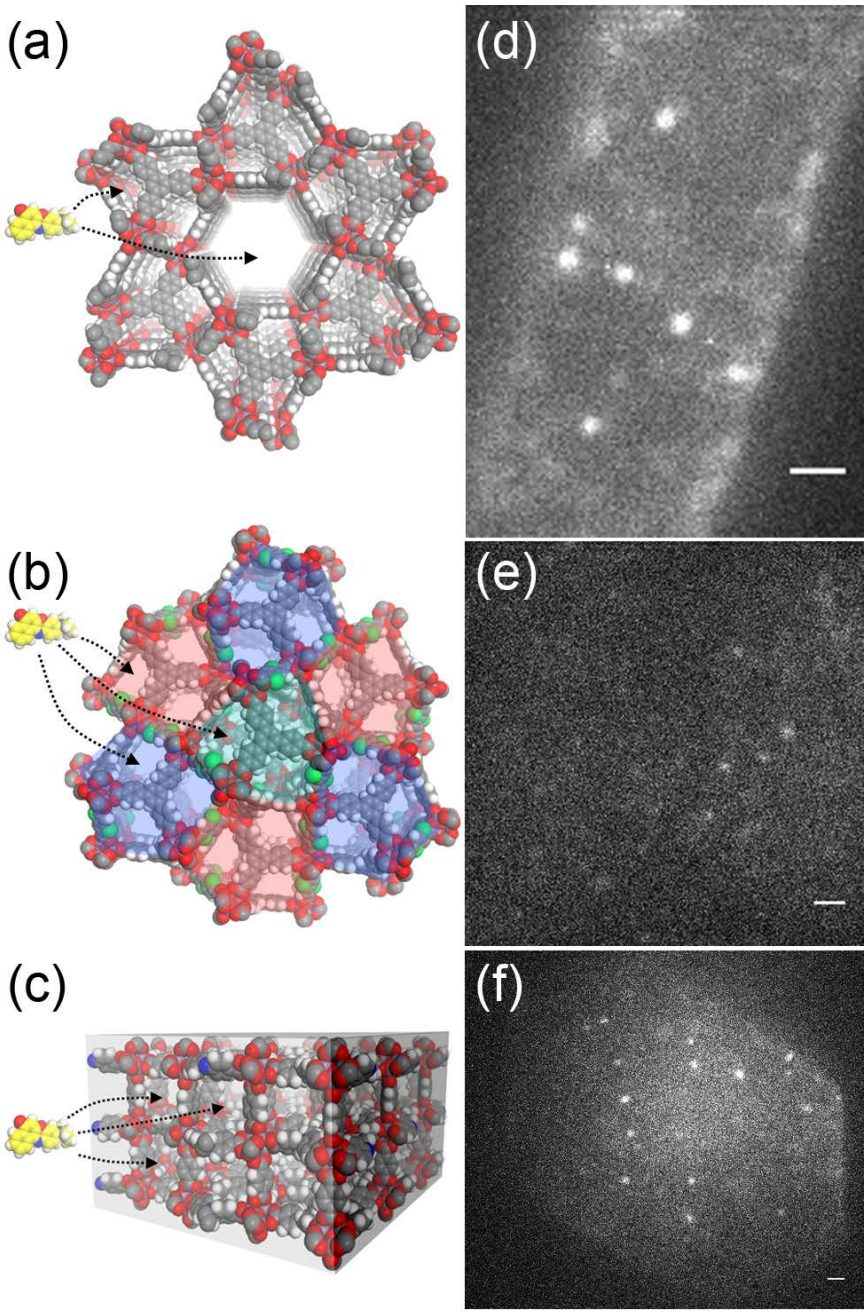


Figure 5.1 Single-molecule imaging in microporous coordination polymers. (a-c) Three-dimensional structures of Nile red in (a) UMCM-1 (b) UMCM-2 and (c) UMCM-4. Different dye molecule diffusivities are expected for each of the three crystals, and multiple modes of motion are anticipated within each crystal. UMCM organic framework: hydrogen (white), carbon (grey), nitrogen (blue), oxygen (red), sulfur (green). For clarity, the Nile red carbons are rendered in yellow. (d-f) Representative fluorescent images of individual Nile red dye molecules in (d) UMCM-1 (e) UMCM-2 and (f) UMCM-4. Scale bar = 1 μm.

5.2 Imaging single molecules diffusing in MCPs

A recent study of MOF crystals with fluorescence correlation microscopy revealed that molecular transport in MOF-5 is heterogeneous, and that the behavior of guest molecules is strongly affected by guest-host interactions[8]. Here for the first time, we use super-resolution fluorescence imaging to investigate single-molecule diffusion in MOF crystals in real time. By circumventing ensemble averaging, we isolate several modes of motion: directional diffusion, confined movement, and immobilization, each characterized by a different diffusion coefficient. By recording and analyzing molecular trajectories using the full information provided by imaging single molecules, we report on trajectory shapes and diffusion coefficients for each type of motion, and also quantify the distribution of dye molecules engaged in each sort of motion within a single crystal. Several single-molecule imaging experiments performed on porous materials have been previously reported[9-12] where the fluorescent probes were incorporated into the host during synthesis. In the present study however, Nile red probes were introduced from outside after native UMCM crystals were synthesized. This more closely mimics the diffusion process that naturally takes place when these host materials are in contact with guest molecules, and also limits the probes to regions that are accessible to the outside. Specifically, 5 pM Nile red in 1-butanol was added to UMCM crystals immediately before imaging. In the case of UMCM-1, diffusion along the mesoporous channels is rapid, so 10% mineral oil was added to the imaging mixture to slow the molecules to a temporally resolvable level.

The imaging experiment was performed using a standard widefield epifluorescence microscope. This allows simultaneous high-speed (29 Hz) acquisition of emission data from many isolated molecules. For a better signal-to-noise ratio in the acquired images, a high

numerical aperture (NA = 1.40) oil-immersion objective was used, and appropriate filters were incorporated into the excitation and detection pathways. To observe individual molecules unambiguously, the majority of the emitters within the illuminated area were first turned off by photobleaching using a 488-nm laser. After a 2 – 5 min pause, during which time a small amount of fresh dye molecules diffused back to the pre-bleached area, this sparse set of emissive Nile red molecules were imaged under 561-nm laser excitation with powers of 44 kW/cm^2, 2.2 kW/cm^2 and 5.2 kW/cm^2 for UMCM-1, UMCM-2 and UMCM-4, respectively.

Movies up to 42 s (1200 imaging frames) in length were collected, and a representative fluorescence image of isolated Nile red molecules in each of UMCM-1, UMCM-2 and UMCM-4 is shown in Figures 5.1d, 5.1e and 5.1f, respectively. The fluorescent signal from each individual emitter is a diffraction-limited Gaussian-shaped point-spread function, and the precise location of the molecules were obtained by fitting this emission intensity profile to a symmetric 2D Gaussian function using a home-built MATLAB code based on the nlinfit algorithm.[13] This fitting procedure achieved localization accuracies as good as 10 nm. A custom-made MATLAB tracking program identified trajectories by recognizing successive fits of the same molecule based on spatial proximity. This iterative process was continued until the dye molecule being tracked photobleached or diffused out of the focal plane (~1 μm). Only tracks longer than 0.5 s (15 frames) are discussed and analyzed below.

5.3 Probing local structure and geometry on the nanometer scale in MCPs

Figure 5.2a shows trajectories of dye molecules diffusing in a UMCM-1 crystal. Here, trajectories are distinguished from one another with false color, and overlaid on a white-light

transmission image of the UMCM-1 host. Three of the trajectories (indicated with arrows) are magnified in Figures 5.2b-d. A significant number of the trajectories in Figure 5.2a are qualitatively 1D and are aligned along the longitudinal axis of the crystal (e.g., the track highlighted in Figure 5.2b), suggesting that many molecules are diffusing inside the large linear channels formed by mesopores. Still, a few molecules appear to be spatially confined, such as the one tracked in Figure 5.2c, which covers an area of only 63 × 107 nm2 in 2.31 s. A detailed inspection of the trajectories indicates that the motion is not uniform even within each individual trajectory. For example, the trajectory in Figure 5.2b spans a width of 690 nm, which far exceeds that of a single mesoporous channel (3.2 nm), consistent with a molecule frequently hopping from one channel to another through micropores as it moves along the length of the crystal. Additionally, a molecule will sometimes hit “dead ends” and have to spend extra time bouncing back and forth before finding its way out. For example, the molecule in Figure 5.2b mainly shows fast motion, but is restricted for 1.05 s to the area indicated by the dashed-line square in Figure 5.2b. This behavior is especially evident in Figure 5.2d, where the molecule frequently hits “dead ends” and switches channels. The different types of motion illustrated by the trajectories in Figures 5.2b, c and d reflect the different physical and chemical environments with which these molecules interact. Nile red molecules exhibit drastically different behaviors in UMCM-2 crystals (Figure 5.2e), where most of trajectories are confined. For example, the 0.735-s track in Figure 5.2f is restricted to 37 × 62 nm2. This molecule is either trapped inside a microporous cage or else immobilized at an adsorption site on the host. A small number of molecules do diffuse freely (e.g., Figure 5.2g), presumably within defective voids. Finally, the majority of Nile red molecules in UMCM-4 crystals diffuse freely in two dimensions (Figure 5.2h). This planar exploration of the host is consistent with the structure of UMCM-4: it does not

possess any channel system within the 2D layers that would direct the motion of guest molecules along a specific route. Still, our imaging uncovers a heterogeneous environment within the UMCM-4 planes. The molecule in Figure 5.2i covers an area of 1.56 × 1.86 μm2 in the 3.465 s trajectory, while other molecules are constrained the entire time (e.g., Figure 5.2j). Constrained motion in all cases can be attributed to confinement at defect sites or to interaction between the guest and the host.

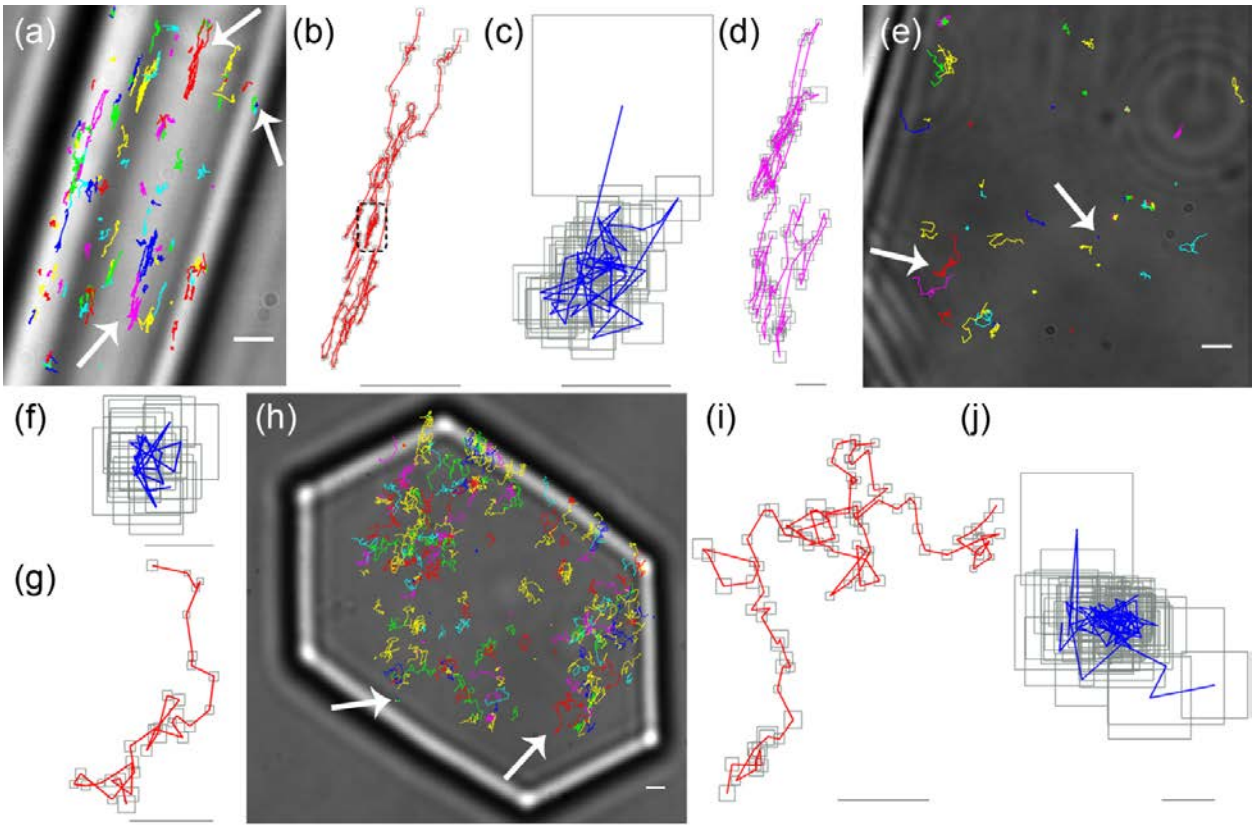


Figure 5.2 Single-molecule trajectories in MCPs. (a, e, h) All trajectories lasting longer than 0.5 s superimposed with false color on the white-light image of the corresponding host: UMCM-1, UMCM-2 and UMCM-4, respectively. Here, scale bar = 1 μm and the arrows point to trajectories highlighted in (b-d), (f-g) and (i-j). (b) Predominately 1D diffusion in UMCM-1. Grey squares represent the 95% confidence intervals for the localization of each fluorophore. Scale bar = 500 nm. (c) Confinement in UMCM-1. Scale bar = 50 nm. (d) Hopping motion in UMCM-1. Scale bar = 100 nm. (f) Confined motion in UMCM-2. Scale bar = 50 nm. (g) 2D diffusion in UMCM-2. Scale bar = 500 nm. (i) 2D diffusion in UMCM-4. Scale bar = 500 nm. (j) Confinement in UMCM-4. Scale bar = 50 nm. The dashed-line square in (b) indicates a region where the molecular motion is temporarily restricted for 1.05 s.

From the random-walk model, the mean square displacement (MSD) of a randomly diffusing particle increases linearly with time, and the diffusion coefficient, *D*, can be calculated from the Einstein-Smoluchowski equation:[14]

$$\langle r^2(\tau)\rangle = 2nD\tau \tag{5.1}$$

where $<r^2>$, n and τ denote MSD, dimensionality of motion and time lag, respectively. MSD versus τ for all tracks lasting at least 15 frames is plotted for Nile red in UMCM-1, UMCM-2, and UMCM-4, respectively, in Figures 5.3a-c. It is evident that in all three cases, the diffusion processes do not strictly follow the random-walk model of equation (5.1); instead, physical confinement, guest-host interactions and adsorption cause the plots to deviate from an ideal linear relationship. For many of the molecular trajectories in Figures 5.3a-c, MSD saturates at intermediate or large values of τ, confirming the existence of confinement and possibly immobilization in addition to normal diffusion within single trajectories.[15] This heterogeneity is consistent with the qualitative discussion of trajectories above. The slopes of each curve, as determined from a linear fit to the first 50% of the data points, are used to roughly divide the molecules in all three crystals into two categories: molecules that predominately diffuse freely (positive slope) and those that are either highly confined or completely immobile throughout the entire trajectory (flat slope). The MSDs of these two sub-populations are plotted in red and blue, respectively, in Figures 5.3a-c. These plots show that guest molecule behavior is significantly different in each of the three hosts. 127 of the 136 molecules tracked in UMCM-1 (Figure 5.3a) and 200 of the 229 molecules tracked in UMCM-4 (Figure 5.3c) diffuse freely (i.e., the curves in these panels are predominantly red), whereas more than half (46 of 79) of the molecules tracked in UMCM-2 show no apparent diffusion (i.e., the curves in Figure 5.3b are mostly blue). As well, the mobile sub-population in UMCM-1 has a broader distribution of initial MSDs (τ = 35 ms) than in UMCM-2 or UMCM-4.

Although MSD versus τ gives a general and qualitative description of the behavior of molecules in these crystals, extracting D directly from the slopes in Figures 5.3a-c is an oversimplification in this complex environment: calculating D directly from Equation (5.1) provides average information, and masks valuable data regarding the heterogeneity along each trajectory.[16] Therefore, instead of performing an analysis based on the average slope, step sizes are analyzed based on their distribution and described by a corresponding diffusion model, from which multiple diffusion coefficients, each characterizing a specific type of trajectory (mobile or immobile) can be obtained.[17] Choosing the appropriate diffusion model requires first categorizing the molecules according to the dimensionality of their motion, and this can be accomplished by analyzing the distribution of angles, θ, between each consecutive pair of displacement vectors (Figure 5.3d inset).

The molecules in UMCM-1 show predominantly 1D motion, since the angles between successive steps accumulate around 0^o and $\pm 180^o$ (Figure 5.3d), i.e., the molecules have the largest probability of moving directly forward or directly backward. On the other hand, the molecules in UMCM-4 show no preference for any particular direction within the 2D field of view (Figure 5.3f), a characteristic of 2D motion. The U-shaped angle distribution observed for molecules in UMCM-2 (Figure 5.3e) is expected for immobile molecules in the presence of Gaussian fitting noise: the likelihood of localization is a Gaussian function about the exact molecular location (Figure 5.4). This distribution indicates that most molecules in UMCM-2 are immobile or confined to very small volumes. The following analysis of molecular motions will therefore focus on molecules diffusing in UMCM-1 and UMCM-4.

Though the distribution of angles θ for Nile red in UMCM-1 (Figure 5.3d) suggests 1D diffusion, the observed trajectories (Figure 5.2a-d) and the presence of intermediate angles other than 0^o and $\pm 180^o$ imply a more complex scenario. This deviation from an ideal 1D random walk can be attributed to a variety of causes, including the fact that, in addition to facile diffusion along mesoporous channels, molecules can diffuse in microporous cages and within defective voids. The localization errors, which are two-dimensional, will also contribute to this deviation. As was pointed out by Kirstein,[18] it is therefore more appropriate to reduce this quasi-1D motion to true 1D motion by projecting the displacements onto a backbone along the overall trajectory. In the present study, the backbone of a given trajectory is defined as the line obtained from least-square linear regression over all positions of this trajectory (Figure 5.5).

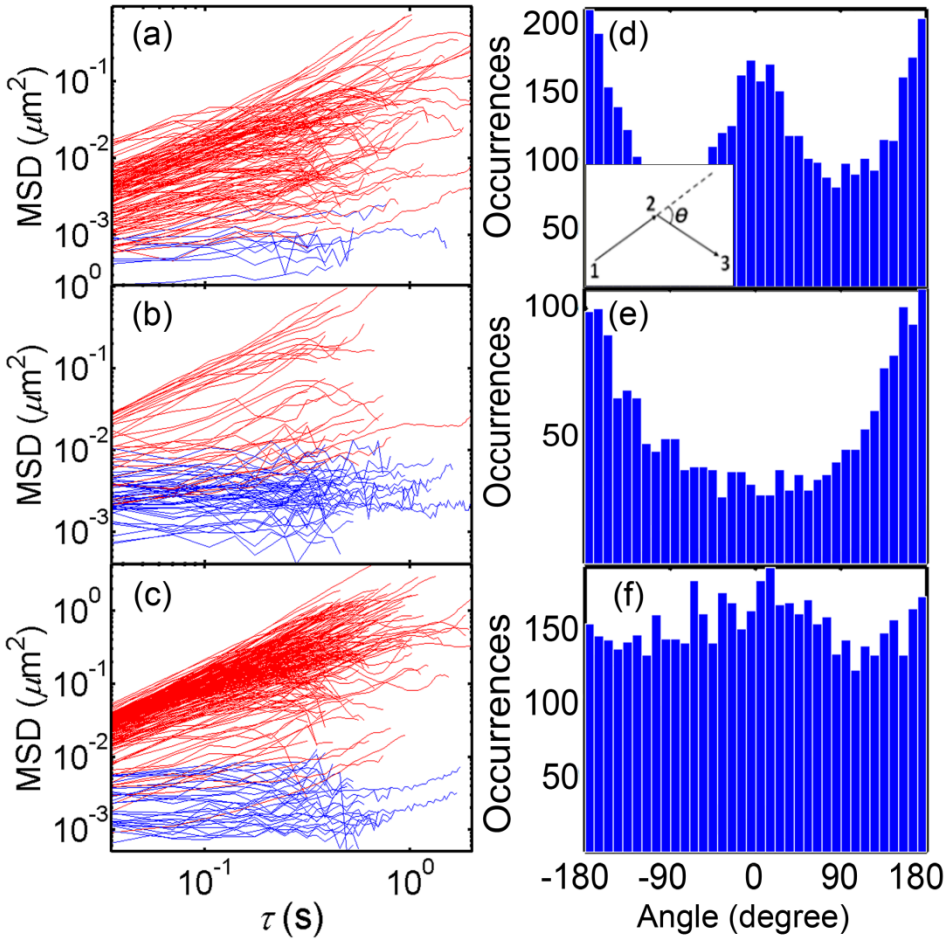


Figure 5.3 (a-c) MSD vs. time lag, τ, for all molecules tracked longer than 0.5 s in (a) UMCM-1 (b) UMCM-2 and (c) UMCM-4. Because of the decreasing reliability of MSD values at larger time lags, the curve for each molecule is cut off at three quarters of its trajectory length, and the slope is calculated for the first 50% of data points along each trajectory. Based on the slope, each trajectory can be categorized into one of two sub-populations: mobile (red) and confined (blue). (d-f) Distribution of angles between two steps. The θ is defined as the angle between two consecutive displacement vectors (1-2 and 2-3) (inset d). The distribution of θ in (d) UMCM-1, (e) UMCM-2 and (f) UMCM-4 have distinct shapes characteristic of a 1D random walk, immobile molecules, and a 2D random walk, respectively.

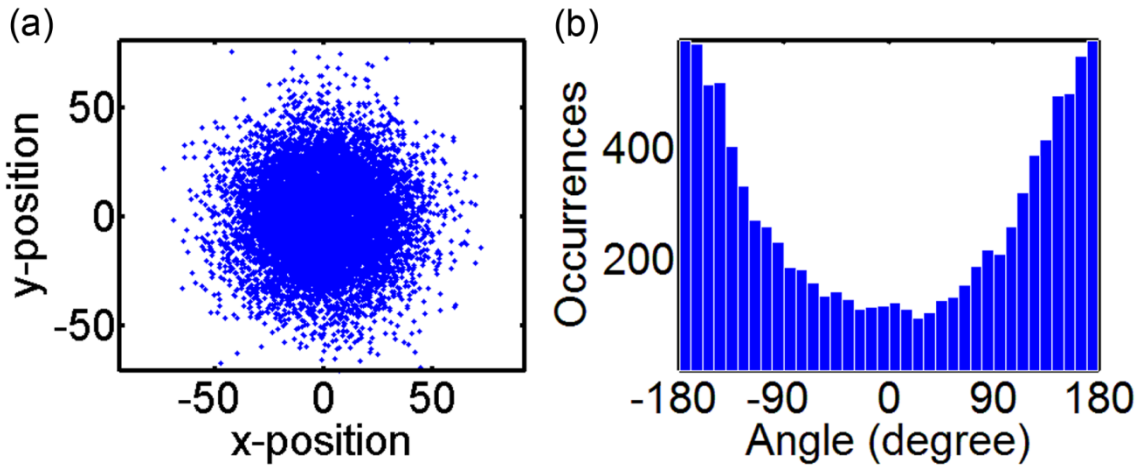


Figure 5.4 Angle Distribution of an Immobile Molecule in the Presence of Gaussian Fitting Noise. Due to the limited localization accuracy and the associated Gaussian fitting noise, an immobile molecule will appear as moving within a small region centered on its true position, which gives rise to an artificial trajectory. The angles between consecutive displacement vectors for such trajectory follow a U-shaped distribution with maxima at $\pm 180^{o}$. Here we use simulation to show the angle distribution from a simulated trajectory constructed from 10,000 data points centered about (0,0) with added Gaussian fitting noise. (a) 10,000 simulated data points are distributed around the origin in a Gaussian pattern and are connected sequentially to generate a trajectory similar to that obtained from tracking an immobile molecule in the presence of Gaussian fitting noise. For clarity, the connections between data points are not shown. (b) Angle distribution of the simulated trajectory.

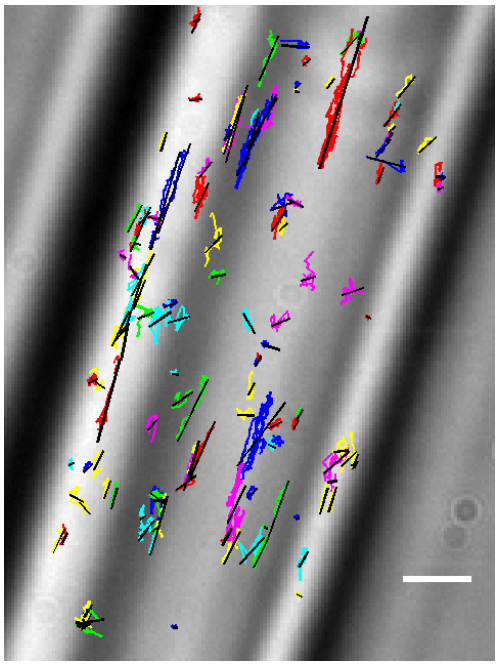


Figure 5.5 Backbone projections for single-molecule trajectories. To reduce the quasi-1D motion observed in UMCM-1 to ideal 1D motion for subsequent 1D cumulative probability distribution (CPD) analysis, all displacements in UMCM-1 are projected onto the track backbone prior to fitting the data to a one-dimensional diffusion model. The backbone (solid black line) for each trajectory (colored line) is obtained from a least-square linear regression over all positions along the track. Scale bar: 1 μm.

5.4 Characterizing heterogeneous single-molecule diffusion in MCPs

For normal 1D diffusion, the distribution of the squared step sizes for a given time lag τ is described by the 1D cumulative probability distribution (CPD):[19]

$$P_{1D}(U,\tau) = erf\left(\sqrt{\frac{U}{2(\langle r^2(\tau)\rangle + \sigma^2)}}\right) \tag{5.2}$$

where $P_{1D}(U,\tau)$ is the probability that the squared displacement, r^2, during time τ does not exceed some value U. The localization accuracy, σ, is experimentally determined from the standard deviation of fitted positions of immobile molecules in a calibration sample imaged under conditions identical to the tracking experiments. This σ is estimated to be 10 nm for UMCM-1 and 20 nm for UMCM-4 (which was imaged at lower excitation intensity). By fitting squared displacements from various τ to this model, the MSD (i.e., $<r^2(\tau)>$) can be obtained from equation (5.2), and then the diffusion coefficient can be extracted via equation (5.1).

To accommodate heterogeneous motion involving multiple diffusion coefficients, equation (5.2) is expanded to include several terms, each of which describes one type of motion with its own characteristic D.[20] In the case of UMCM-1, the observed motion is best described by a three-term CPD function consisting of two mobile terms and one immobile term:

$$P_{UMCM-1}(U,\tau) = \alpha \cdot erf\left(\sqrt{\frac{U}{2(\langle r_1^2(\tau)\rangle + \sigma^2)}}\right) + \tag{5.3}$$

$$\beta \cdot erf\left(\sqrt{\frac{U}{2(\langle r_2^2(\tau)\rangle + \sigma^2)}}\right) + (1-\alpha-\beta)\cdot erf\left(\sqrt{\frac{U}{2(\sigma^2)}}\right)$$

Here, α and β denote the fraction of time the molecules spend in a faster diffusion mode (term 1) and a slower one (term 2), respectively, and the molecules are immobile within the localization accuracy for the remaining time (term 3). Raw data (colored lines) and fits to Equation (5.3) (black lines) for representative short (τ = 0.035 s), intermediate (τ = 0.140 s) and long (τ = 0.280 s) time lags are presented in Figure 5.6a. Compared to a two-term CPD function, the fit of the data to this model decreases the reduced χ^2 from 1.89×10^{-4} to 1.55×10^{-5}. From this fitting result, two distinct MSDs are obtained at each τ (Figure 5.6b), and thus two diffusion coefficients were calculated from equation (5.1). We calculate D = 0.0695 μm^2/s for the fast motion and D = 0.0046 μm^2/s for the slow motion. On average, the whole population spent α = 64.5% of the time in the fast diffusion mode, β = 30.5% of the time in the slow mode, and only 5% of the time immobile.

The distribution of squared displacements can be analyzed in a similar way for 2D motion using the following expression:[21]

$$P_{2D}(U,\tau) = 1 - e^{\frac{-U}{\langle r^2(\tau)\rangle}} \tag{5.4}$$

and the behavior of molecules in UMCM-4 is best described using a three mobile term expansion of this relation:

$$P_{UMCM-4}(U,\tau) = 1 - \left[\alpha \cdot e^{\frac{-U}{\langle r_1^2(\tau)\rangle}} + \beta \cdot e^{\frac{-U}{\langle r_2^2(\tau)\rangle}} + (1-\alpha-\beta)\cdot e^{\frac{-U}{\langle r_3^2(\tau)\rangle}}\right] \tag{5.5}$$

which gives three independent MSD values at each τ, corresponding to a fast, a slow and a slowest motion type. This expansion decreases the reduced χ^2 of fit from 1.05×10^{-4} to 6.61×10^{-6} compared to a two-component CPD function. The average percentage of time molecules spent in the fast, slow and slowest modes were 73.3%, 13.7% and 13.0%, respectively. The fitting results at three time lags are shown in Figure 5.6c. Similar to the 1D case, the diffusion coefficients for the fast and slow motions are found from the slopes in Figure 5.6d to be D = 0.2761 μm^2/s and D = 0.0595 μm^2/s, respectively. However, the MSD of the slowest mode is not linear with τ, but rather reaches a plateau at larger values of τ (Figure 5.6d inset), which indicates that the motion is confined and can be explained by a diffusion model for particle motion restricted within a square with side length L:[22, 23]

$$\langle r^2(\tau)\rangle = \frac{L^2}{3}\cdot\left(1 - e^{\frac{-12D_0\tau}{L^2}}\right) \tag{5.6}$$

where D_0 is the initial diffusion constant. Fitting the MSD of the slowest mode to equation (5.6), D_0 was determined to be 0.0057 μm^2/s and the average side length of confinement was 68 nm. The apparent diffusion due to the 20-nm localization accuracy is reflected by a 0.4×10^{-3} μm^2 offset of this plot.

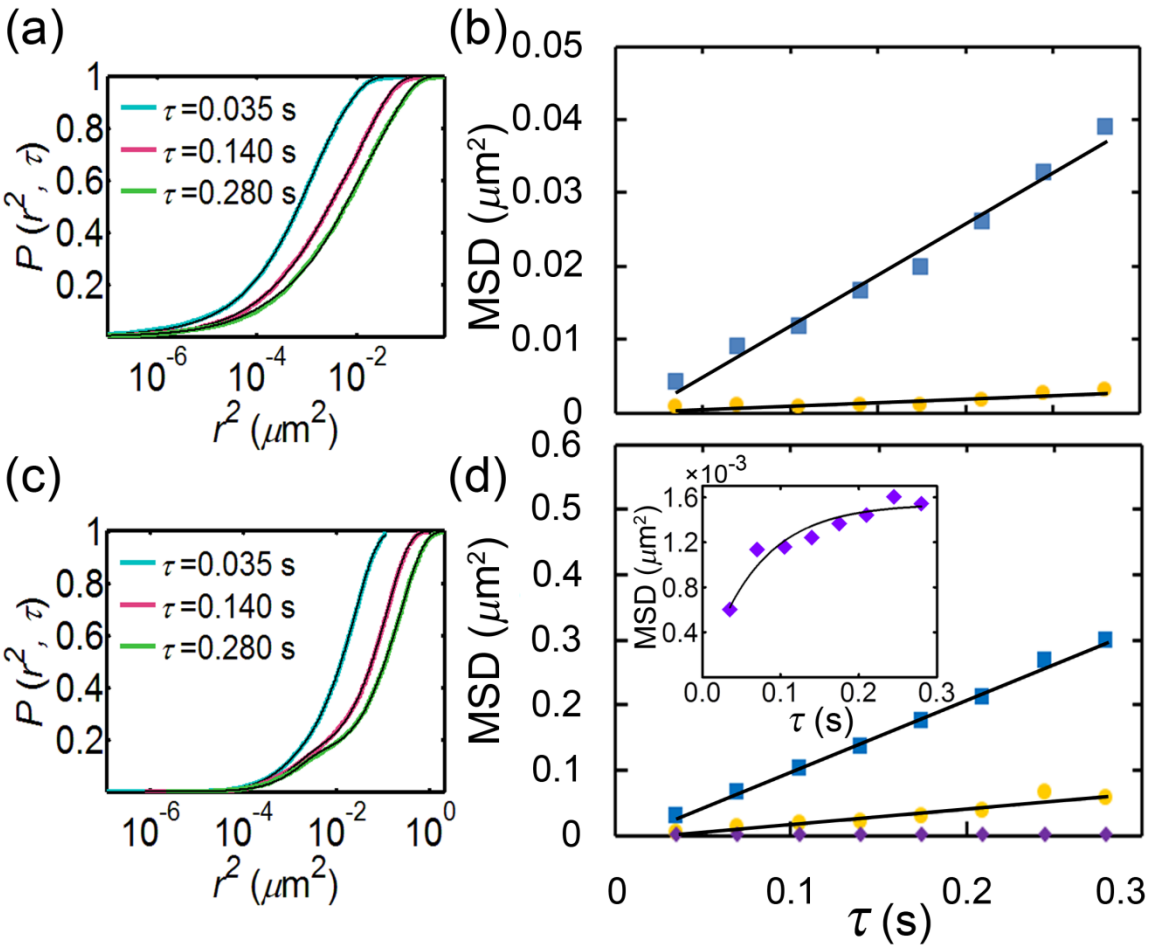


Figure 5.6 (a) Squared step sizes of molecules in UMCM-1 fit to the three-component 1D CPD function in Equation (5.3). Colored lines and black lines are the data and the fits, respectively. (b) MSD versus τ for the fast (blue squares) and the slow (yellow circles) sub-populations obtained from fitting results like those in (a). (c) Squared step sizes of molecules in UMCM-4 fit to the three-component 2D CPD function in Equation (5.5). (d) MSD versus τ for the fast (blue squares), the slow (yellow circles) and the slowest (purple diamonds) sub-populations, all obtained from the fitting results. The linear slopes in (b) and (d) give the diffusion coefficient characterizing each component, and in the inset of (d), the slowest sub-population is described by the 2D confinement model of Equation (5.6).

5.5 Discussion

In conclusion, the diffusion of individual Nile red guest molecules inside three different crystalline microporous coordination polymer (MCP) hosts was visualized in real time using single-molecule fluorescence microscopy. Trajectory analysis reveals that most molecules in UMCM-2 crystals are immobilized while molecules in UMCM-1 and UMCM-4 crystals moved according to a 1D random walk, and a 2D random walk, respectively. Nevertheless, guest motion was not homogenous in any of the crystals, and several sub-populations of molecules, each with its own mobility range, were identified in each crystal. In addition, due to structural heterogeneities, molecules that are reversibly trapped or adsorbed at a single site switch between different modes of motion along the way, depending on the physical and chemical properties of the local environments in which they reside and the features with which they interact. The dynamic information obtained with this technique can provide valuable insights for accessing and characterizing the guest-host interactions within crystalline MCP structures. This finding further highlights one major advantage of the single-molecule imaging technique, the ability to probe heterogeneous processes and properties that are often overlooked in ensemble analyses. The present study provides a framework for investigating and understanding the nanoscale properties of heterogeneous materials that are masked by conventional approaches.

References

1 Cychosz, K. A.; Ahmad, R.; Matzger, A. J. *Chem. Sci.* **2010**, *1*, 293-302.
2 Ma, S.; Zhou, H. C. *Chem. Commun.* **2010**, *46*, 44-53.
3 Alaerts, L.; Maes, M.; Giebeler, L.; Jacobs, P. A.; Martens, J. A.; Denayer, J. F. M.; Kirschhock, C. E. A.; De Vos, D. E. *J. Am. Chem. Soc.* **2008**, *130*, 14170-14178.
4 Peterson, V. K.; Liu, Y.; Brown, C. M.; Kepert, C. J. *J. Am. Chem. Soc.*, **2006**, *128*, 15578-15579
5 Koh, K.; Wong-Foy, A.; Matzger, A. J. *Angew. Chem. Int. Ed.* **2008**, *47*, 677-680.
6 Koh, K.; Wong-Foy, A.; Matzger, A. J. *J. Am. Chem. Soc.* **2009**, *131*, 4184-4185.
7 Koh, K.; Wong-Foy, A.; Matzger, A. J. *J. Am. Chem. Soc.* **2010**, *132*, 15005-15010.
8 Han, S.; Hermans, T. M.; Fuller, P. E.; Wei, Y.; Grzybowski, B. A. *Angew. Chem. Int. Ed.* **2012**, *51*, 2662-2666.
9 Kirstein, J.; Platschek, B.; Jung, C.; Brown, R.; Bein, T.; Bräuchle, C. *Nature Mater.* **2007**, *6*, 303-310.
10 Zürner, A.; Kirstein, J.; Döblinger, M.; Bräuchle, C.; Bein, T. *Nature* **2007**, *450*, 705-708.
11 Fu, Y.; Ye, F.; Sanders, W. G.; Collinson, M. M.; Higgins, D. A. *J Phys Chem B* **2006**, *110*, 9164-9170.
12 Hellriegel, C.; Kirstein, J.; Bräuchle, C.; Latour, V.; Pigot, T.; Olivier, R.; Lacombe, S.; Brown, R.; Guieu, V.; Payrastre, C.; Izquierdo, A.; Mocho, P. *J Phys Chem B* **2004**, *108*, 14699-14709.
13 Biteen, J. S.; Thompson, M. A.; Tselentis, N. K.; Bowman, G. R.; Shapiro, L.; Moerner, W. E. *Nat. Methods* **2008**, *5*, 947-949.
14 Qian, H.; Sheetz, M. P.; Elson, E. L. *Biophys. J.* **1991**, *60*, 910-920.
15 Saxton, M. J. *Biophys J* **1997**, *72*, 1744-1753.
16 McCain, K. S.; Hanley, D. C.; Harris, J. M. *Anal. Chem.* **2003**, *75*, 4351-4359.
17 Hellriegel, C.; Kirstein, J.; Bräuchle, C. *New J. Phys.* **2005**, *7*, 1-14.
18 Kirstein, J. U. Diffusion of single molecules in nanoporous mesostructured materials, Ph.D. Dissertation, Ludwig Maximilian University of Munich, Munich, Germany, 2007.
19 Jung, C.; Kirstein, J.; Platschek, B.; Bein, T.; Budde, M.; Frank, I.; Müllen, K.; Michaelis, J.; Bräuchle, C. *J. Am. Chem. Soc.* **2008**, *130*, 1638-1648.
20 Schütz, G. J.; Schindler, H.; Schmidt, T. *Biophys. J.* **1997**, *73*, 1073-1080.
21 Sonnleitner, A.; Schütz, G. J.; Schmidt, T. *Biophys. J.* **1999**, *77*, 2638-2642.
22 Kusumi, A.; Sako, Y.; Yamamoto, M. *Biophys. J.* **1993**, *65*, 2021-2040.
23 Lommerse, P. H. M.; Blab, G. A.; Cognet, L.; Harms, G. S.; Snaar-Jagalska, B.; Spaink, H. P.; Schmidt, T. *Biophys. J.* **2004**, *86*, 609-616.

Chapter 6 Conclusions and Perspectives

In this thesis, I developed super-resolution microscopy to visualize and study three important and fascinating processes occurring at the nanometer scale: DNA mismatch repair (MMR) in live *Bacillus subtilis* cells, polymerase exchange during *B. subtilis* DNA replication, and molecular transport in microporous coordination polymers (MCPs). The ability to capture, measure and analyze the motion of single molecules allowed us to uncover within each system heterogeneities that were previously inaccessible by conventional microscopy techniques, and to provide direct, unambiguous mechanistic insights into molecular motion and function.

Toward the goal of understanding how the crowded cellular environment and the tightly packed chromosome allows for rare DNA mismatch detection with high efficiency, I discovered that the DNA repair protein MutS is recruited to the replication fork to scan locally for replication errors prior to mismatch binding, and that this mismatch binding by MutS can only take place on DNA proximal to the replication fork. Pre-staging of MutS at the site of DNA replication not only positions MutS in close proximity to the newly synthesized DNA (and thus near newly produced mismatches), but also grants MutS access to nascent DNA strands largely free of the DNA-binding proteins which would otherwise occlude the scanning pathway. In addition, super-resolution microscopy is also a valuable tool for quantitative characterization of biophysical processes, and I have harnessed this power to quantify the diffusivity of MutS and the dynamics of MutS/replisome interactions in various mutant strains and under different drug treatment conditions, shifting our understanding of MMR from qualitative descriptions to a more

quantitative level.

Based on photobleaching-assisted microscopy and 3D single-particle tracking, I have studied the stoichiometry, positioning and dynamics of the PolC protein, one of the two essential DNA polymerases in *B. subtilis*. Our results shed light on the heterogeneous distribution of PolC within a cell as well as its highly dynamic dwelling and departing behavior at the replication fork. The ability to directly visualize polymerase exchange during DNA replication at the single-molecule level opens up a number of directions for further exploration. For example, it will be interesting to probe how the dwell time and thus the polymerase exchange rate of PolC is affected by arresting DNA replication using drugs such as HPUra, by collapsing the replisome structure by mitomycin C treatment, and by manipulating the primase concentration at the replication fork.

Unlike *Eschericia coli*, which uses a single type of essential DNA polymerase for the synthesis of both the leading and the lagging strands, *B. subtilis* employs two different DNA polymerases PolC and DnaE to accomplish DNA replication, in a manner similar to Pol ε and Pol δ found in eukaryotic cells. Some important and long-standing questions are: how do PolC and DnaE come together and cooperate to replicate DNA, what is their relative stoichiometry at the replication fork, and how do the dynamics of one affect the other? Since DnaE only synthesizes a small DNA segment on the lagging strand, we expect the dwell time of DnaE to be shorter than that observed for PolC. Future simultaneous two-color super-resolution microscopy will help to provide answers to these fundamental questions.

It is important to note that, although PolC and DnaE are responsible for synthesizing DNA, DNA replication would not occur in the absence of the other proteins of the replisome. The

replisome is a multi-protein complex with a very intricate organization, with each subunit having its own specialized function (Figure 6.1). Therefore, the characterization of DNA polymerases is only a starting point for understanding the structure and dynamics of the replisome, and many additional questions still remain, such as how does the DNA clamp protein DnaN interact with the DNA polymerase during DNA synthesis? And how does the single-stranded DNA binding protein (SSB) coordinate the recruitment of various replisome subunits and repair proteins in both untreated cells and cells challenged with DNA-damaging agents?

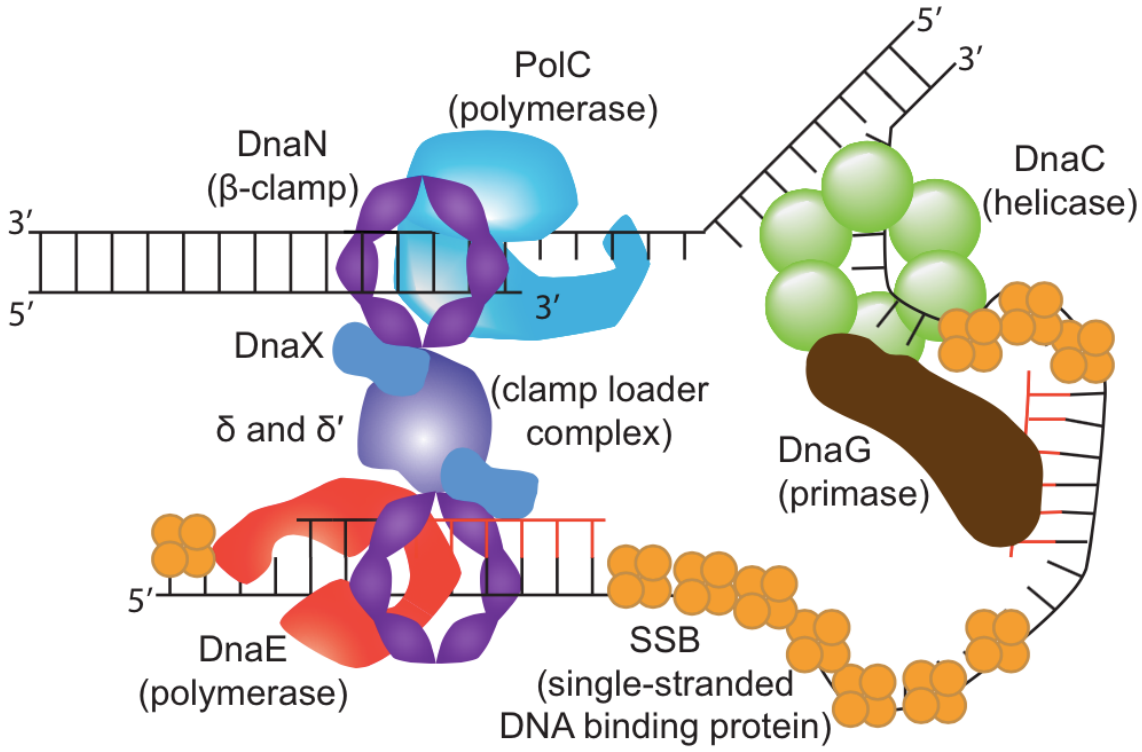


Figure 6.1 Schematic representation of the *B. subtilis* replisome. Figure adapted from Jeremy Schroeder.

One exciting aspect of the burgeoning field of super-resolution microscopy is the tremendous freedom it offers to researchers to explore every possible direction, and to even step into uncharted realms, which can usually lead to fruitful discoveries or contributions in unconventional ways. Although this emerging technique has primarily found applications in studying biology, Chapter 5 demonstrates that it is also possible to expand this technique into fields like renewable energy that are of equal importance to social well-being. The approach I developed in Chapter 5 to probe molecular transport and guest-host interactions in MCPs can be further extended. In particular, investigations with different dyes and solvents would allow us to distinguish guest and host properties. Additionally, one would expect some distribution of diffusion coefficients as a function of position, and this spatial distribution could be probed with scanning fluorescence correlation spectroscopy (FCS).

Overall, as super-resolution microscopy and single-molecule techniques are continuously being improved, we expect their applications in biomedicine and nanotechnology will become even broader in the future and their contributions across many disciplines more profound. With the development of better imaging configurations and detection devices, brighter and more photostable fluorophores, and more versatile labelling schemes, it will become increasingly feasible to visualize and track, not just proteins, but also other biological macromolecules such as RNA and sugar molecules *in vivo* to understand how they are organized, how they interact with each other, and how they are regulated and targeted by various biochemical pathways and external factors.

Appendix 1. List of strains

Strain	Relevant genotype	Reference
PY79	Wild type prototroph, SPβ°	[1]
JWS108	Δ*mutSL*	[2]
JWS121	*mutS*-PAmCherry-RBS-*mutL*	This study
JWS134	*amyE*::Pxyl:*dnaX*-mCitrine	This study
JWS154	*amyE*::Pxyl:*dnaN*-mCitrine	This study
JWS161	*mutS*[F30A]-PAmCherry1-RBS-*mutL*	This study
JWS162	*dnaB134*(ts) *zhb83*::Tn917 (tet) x PY79	[3]
JWS170	Δ*mutS*, *dnaB134*(ts) *zhb83*::Tn917 (tet)	This study
JWS176	*mutS*[F30A]-PAmCherry1-RBS-*mutL*, *amyE*::Pxyl:*dnaX*-mCitrine	This study
JWS185	*mutS*[K608M]-PAmCherry-RBS-*mutL*	This study
JWS194	*mutS*[K608M]-PAmCherry-RBS-*mutL*, *amyE*::Pxyl:*dnaX*-mCitrine	This study
JWS220	*mutS*-PAmCherry-RBS-*mutL*, *dnaN5*, *spoIIIJ*::kan	This study
JWS221	*mutS800*-PAmCherry-RBS-*mutL*	This study
JWS222	*mutS*-PAmCherry-RBS, Δ*mutL*	This study
JWS225	*mutS*-PAmCherry-RBS, Δ*mutL*, amyE::Pxyl:*dnaX*-mCitrine	This study
JWS243	*mutS800*-PAmCherry, *amyE*::Pxyl:*dnaX*-mCitrine	This study
JWS258	*mutS::mutS800*, *amyE*::*mutL*, *dnaB134*(ts) zhb83::Tn917 (tet)	This study
JWS297	*dnaN5*, *spoIIIJ*::kan; *amyE*::Pxyl:*dnaX*-mCitrine; *mutS*-PAmCherry-RBS-*mutL*	This study

References

1 Youngman, P., Perkins, J. B. & Losick, R. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene. *Plasmid* **12**, 1-9 (1984).

2 Yao, N. Y., Schroeder, J. W., Yurieva, O., Simmons, L. A. & O'Donnell, M. E. Cost of rNTP/dNTP pool imbalance at the replication fork. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 12942-12947 (2013).

3 Burnett, L. & Wake, R. G. Initiation and termination of chromosome replication at 45 degree C in a temperature-sensitive deoxyribonucleic acid initiation mutant of Bacillus subtilis 168, TsB134. *J Bacteriol* **130**, 538-539 (1977).

Appendix 2. MATLAB code for cell segmentation

```
function valley(final_dilate_factor, lower_thresh, higher_thresh,
manual_selection)
close all;

% Update History:

% 5/8/2015: YL: Manual selection enabled.

% 6/172013 YL: Updated Comments.
%% ------------------------------------------------------------------------
%  User-Defined Parameters:
%  ------------------------------------------------------------------------
% lower_thresh = -0.0;
% Can be negative, but usually use [-1 0.5].

% higher_thresh = 1.9;
% Usually use positive values [0 3] , should be larger than
% lower_threshold. *Decrease* this value first when cells are
% UNDER-segmented (e.g., 2 or more cells grouped as 1 region).

% Note: If cells are under-segmented, increase the lower_thresh and/or
% decrease the higher_thresh (recommended); If cells are over-segmented,
% decrease the lower_thresh and increase the higher_thresh.

% final_dilate_factor = 3;
% Factor of dilation for the final phase mask (after segmentation).
% Increase this value such that membrane bound peaks won't be excluded due
% to smaller phase masks.

min_area = 200;
% minimum allowable area (in px) for a single segmented region. Regions
% with area smaller than this will be discarded.
max_area = 5000;
% maximum allowable area (in px) for a single segmented region. Regions
% with area larger than this will be discarded.

%%


[wlimg_name, wlimg_path, ~] = uigetfile({'*.tif','*.tiff'}, ...
    'Select a white-light image');
if wlimg_path == 0
    display('No white light image selected. Aborting program.')
    return
end

frame=imread([wlimg_path wlimg_name]);
f = frame;
% figure, imshow(f, []); title('orginal')
```

```
f=imcomplement(f); % Invert intensity

% Laplacian of Gaussian filtering
[g,~]=edge(f,'log', 0);
% figure,imshow(~g,[])

% h=fspecial('sobel');
% fd=double(f);
% g=sqrt(imfilter(fd,h,'replicate').^2+imfilter(fd,h','replicate').^2);
% figure, imshow(g,[]); title('sobel');

f = imfilter(f, fspecial('average', 5), 'replicate');
% figure, imshow(f, []); title('average- or gaussian- filtered image');


% figure, imshow(f, []); title('inverted')

%-------------------------------------------------------------------------%
% Define Valley filters
V = zeros(3,3); V(2,2) = -1;
A1 = V; A1(3,1) = 1;
A2 = V; A2(2,1) = 1;
A3 = V; A3(1,1) = 1;
B1 = V; B1(3,2) = 1;
B3 = V; B3(1,2) = 1;
C1 = V; C1(3,3) = 1;
C2 = V; C2(2,3) = 1;
C3 = V; C3(1,3) = 1;
%-------------------------------------------------------------------------%

A1 = imfilter(f, A1, 'corr', 'replicate', 'same');
C3 = imfilter(f, C3, 'corr', 'replicate', 'same');
A2 = imfilter(f, A2, 'corr', 'replicate', 'same');
C2 = imfilter(f, C2, 'corr', 'replicate', 'same');
A3 = imfilter(f, A3, 'corr', 'replicate', 'same');
C1 = imfilter(f, C1, 'corr', 'replicate', 'same');
B3 = imfilter(f, B3, 'corr', 'replicate', 'same');
B1 = imfilter(f, B1, 'corr', 'replicate', 'same');

V1 = min(A1, C3); V2 = min(A2, C2); V3 = min(A3, C1); V4 = min(B3, B1);
V1 = max(V1, V2); V2 = max(V3, V4); V = max(V1, V2);

% figure, imshow(V,[]), title('valley values')
% figure, imshow(V~=0,[]), title('non-zero valley values')

% 8-connectivity correlation
mean_V = mean(V(:)); std_V = std2(V);
T = [mean_V+lower_thresh*std_V, mean_V + higher_thresh*std_V];
V2 = uint8(V>= T(2))*10; V2(V2~=10)=1; % Strong threshold
V1 = uint8(V>= T(1)); % Weak threshold
V1 = imfilter(V1.*V2, [1 1 1; 1 1 1; 1 1 1], 'corr', 'replicate',
'same')>=10;

% figure, imshow(V2,[]); title('strong threshold')
% figure, imshow(~V1,[]); title('strong and weak threshold')
```

```
% Thresholding
f2 = im2bw(f, graythresh(f));
% figure, imshow(f2, []); title('thresholding')

f3 = f2 & ~V1 & ~g;
% figure, imshow(f3, []); title('threshold and edge combined')

f4 = bwmorph(f3,'close',inf);
% figure, imshow(f4,[]); title('close')

f5 = imfill(f4,'holes');
% figure, imshow(f5,[]); title('imfill')

f6 = bwmorph(f5, 'majority', inf');
% figure,imshow(f6,[]); title('majority')

f7 = bwlabel(f6, 4);
% figure, imshow(f7, []); title('bwlabel')

seg_area = regionprops(f7, 'area');

for i = 1:length(seg_area)
    if seg_area(i,1).Area < min_area || seg_area(i,1).Area > max_area
        f7(f7 == i) = 0; % Get rid of small regions
    end
end

% figure, imshow(f7, []); title('final dilation by 1')
% f7 = bwlabel(f7 ~= 0, 4);
% figure, imshow(f7, []); title('bwlabel without small regions and dilated by
1')
% % Reassign region label after getting rid of small regions

f7 = bwlabel(~bwmorph(~f7,'diag',5));

if final_dilate_factor~=0
PhaseMask = imdilate(f7, strel('rectangle', [final_dilate_factor
final_dilate_factor]));

end


imshow(PhaseMask, []); title('Unfilled Phase Mask')

% dah= bwconvhull(f7~=0,'objects',4);figure,imshow(dah)

if manual_selection == 1
% Let user manually pick cells. Unchosen cells will have their
% corresponding pixels turn to background.

sel_done = 0;
click_xy = [];
```

```
use_all_cells = 0;

while sel_done ~= 1

[click_x, click_y, button] = jsbginput(1);

if button == 32 && ~isempty(click_xy) %% If SPACE key is pressed, finish
inputing cells
sel_done = 1;
elseif button == 32 && isempty(click_xy)
    % If SPACE key is pressed without selecting any cell, all cells will
    % then be kept.
    use_all_cells = 1;
    sel_done = 1;
else
   hold all, plot(click_x, click_y, 'r*', 'MarkerSize', 12)
click_xy = vertcat(click_xy, round([click_x, click_y]));  %#ok<AGROW>
end
end % While not done selecting cells

if use_all_cells == 0
chosen_cell_ind = PhaseMask(sub2ind(size(PhaseMask), click_xy(:,2),
click_xy(:,1)));
chosen_cell_ind = unique(chosen_cell_ind);
chosen_cell_ind(chosen_cell_ind==0) = [];

PhaseMask(~ismember(PhaseMask,chosen_cell_ind)) = 0;
end

end % Manually select cells

close all;
figure, imshow(frame, []); title('orginal')
figure, imshow(PhaseMask, []); title('Unfilled Phase Mask')
conv_hull = regionprops(PhaseMask, 'Convexhull');
for i = 1:length(conv_hull)
    hold all,
    plot(conv_hull(i,1).ConvexHull(:,1),conv_hull(i,1).ConvexHull(:,2), 'r-'
,'linewidth', 2)
    axis ij equal
end


% Save matrix to file
save([wlimg_path, wlimg_name(1:end-4), '_PhaseMask.mat'], 'PhaseMask')
fprintf(['''', wlimg_name, ''' segmented.\n'])
end % End of function 'valley.m'
```