Super-resolution Imaging of DNA Replisome Architecture and Dynamics in Live *Bacillus subtilis*

by

Yilai Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2018

Doctoral Committee:

      Associate Professor Julie S. Biteen, Chair
      Professor Ari Gafni
      Assistant Professor Johann Gagnon-Bartsch
      Associate Professor Lyle A. Simmons
      Assistant Professor Qiong Yang

Yilai Li

yilai@umich.edu

ORCID ID: 0000-0002-6280-8019

Acknowledgements

Writing this acknowledgement means that this journey has finally come to an end. I would like to express my gratitude to the people who have helped me along the way.

First, I would like to thank my advisor Professor Julie Biteen, for her help and wonderful mentorship, her understanding, encouragement, and infinite patience with me. It has been an amazing experience for me to be a part of the Biteen lab, and I feel very lucky to be her student. I would like to thank Professor Lyle Simmons for all the help on this project and letting me to be a part of his lab at one point. I could never finish the work in this thesis without his help. I would also like to thank my other committee members, Professor Ari Gafni, Qiong Yang, and Johann Gagnon-Bartsch, for attending my data meeting and thesis defense, and for the valuable suggestions and feedbacks on my research.

I would like to thank all my current and previous lab mates. I want to thank Dr.Yi Liao, Dr. Chanrith Siv, Dr. Bing Fu, and Dr. Hannah Tuson for all the help you gave me when I first joined the lab. I want to thank Dr. Jeremy Schroeder and Dr. Lindsay Matthews for the successful collaboration on this project. I want to thank Ziyuan Chen for helping me with my experiments. I would also like to thank the rest of the members in the Biteen lab and the Simmons lab, thank you for all the small chats and the good times.

I would also like to thank our graduate student coordinator, Sara Grosky, for answering all my questions and booking the rooms for me when I needed.

I would like to thank my friends, for being good listeners and always being there for me.

Finally, I would like to thank my family. I would like to thank my parents for their unconditional support, and for having faith in me at all times. I would like to thank my girlfriend Kexin Zhang, for always believing me, helping me to be stronger, and holding me up when I was depressed. I would like to thank the two lovely kittens Cypress and Maple for their company and consolation.

Thank you everyone for everything.

Yilai Li

Table of Contents

List of Figures

List of Tables

Abstract

DNA replication happens in all living organisms and assures that the genome is accurately copied and maintained. The replisome is the molecular machine in cells that replicates DNA, and this protein assembly includes DNA polymerases which directly synthesize DNA by adding nucleotides. Although the bacterial replisome has been studied extensively *in vitro*, single-molecule microscopy is now providing a new perspective on the dynamics and architecture of replisome components *in vivo*. In this thesis, I studied the architecture and dynamics of several highly conserved replisome components *in vivo* in the model organism *Bacillus subtilis*. Photoactivated localization microscopy (PALM) and single-molecule tracking enable us to localize and track every single protein molecule with a resolution of $20 - 40$ nm. I investigated the dynamics during normal DNA replication of a number of replisomal proteins, including the replicative DNA polymerases PolC and DnaE and the β-clamp loader DnaX. I then extended these investigations to replication arrest by two different pathways: PolC disruption by the damage-independent 6-hydroxy-phenylazo-uracil (HPUra) and cross-linking with mitomycin C (MMC), a damaging agent. I quantified the real-time behavior of different replisomal proteins during the DNA synthesis process. The results presented in this thesis show that all of these replisomal proteins are highly dynamic and exchange more rapidly than previously expected, and I characterized the molecular-scale

distribution of each replisome component as well as responses to cellular mutations and external stimuli with a combination of single-molecule tracking, time-lapse imaging, and spatiotemporal image correlation spectroscopy. Finally, I developed a new approach to characterize the stoichiometry of replisomal proteins by counting the photobleaching steps with Bayesian statistics. Overall, these new insights into DNA replication indicate that the activities of bacterial replisomal proteins may be regulated in cells by coordinating and modulating the dynamics of protein recruitment, binding, and unbinding at the site of DNA synthesis.

Chapter 1: Introduction

1.1 Optical microscopy and bacteria

Since the invention of the microscope, it has been used as a powerful tool in many fields of biology because of its ability to see small things that cannot be seen by the naked eye. In microbiology, optical microscopy is widely used to resolve many aspects from morphology to specific biochemical pathways. The low perturbation to the cells makes the optical microscopes a powerful tool to study the physiology of bacteria in living cells, and although many optical microscopes are not ideal for thick samples, the small size of most bacteria (on the scale of a few micrometers) makes the optical microscopes sufficient to study the behavior of bacteria at the single-cell level. Fluorescence microscopy, specifically, has become an important tool due to its ability to observe specific biomolecules in either fixed cells or live cells by tagging a fluorescent label (chemical dyes or fluorescent proteins) to the molecule of interest. Therefore, fluorescence microscopy can be used to study subcellular components with high specificity. However, fluorescence microscopy still suffers from the low resolution caused by the diffraction limit, making it difficult to study very small structures in bacteria.

1.2 Single-molecule super-resolution microscopy

The resolution of the conventional optical microscopy is limited because the ~400 – 700 nm wavelength of visible light is finite. The intensity profile of a point emitter through the microscope is described by the point spread function (PSF) (Figure 1.1A), which is an Airy function in the (*x, y*) plane. The best resolution attainable can thus be described by the equation below[1]:

$$d = \frac{\lambda}{2NA} \tag{1.1}$$

where *d* is the minimal separation distance between two objects that still can be resolved, λ is the emission wavelength of the fluorescent emitter, and *NA* is the numerical aperture of the objective, which is related to the amount of light that can be collected by the objective and is a property of the objective itself. The *NA* of the best available objectives is about 1.4~1.5, which means that the minimal *d* from the conventional optical microscopy is 200~300 nm. This limited resolution makes it impossible for conventional microscopy to visualize extremely small cellular structures. One way to improve the resolution is to decrease λ by using a beam of accelerated electrons (with a nearly negligible wavelength) instead of the visible light as a source of illumination, which is the essence of the electron microscope. However, although different kinds of electron microscopes have been widely used in physics, electron microscopy (EM) requires the fixation of the sample, which can introduce artifacts and which make EM unsuitable to probe live biological samples for cell biologists and microbiologists.

On the other hand, since the image of a single fluorescent emitter can be described by an Airy function, if one can isolate one single-molecule emitter experimentally, a much better resolution can be attained with an optical microscope by finding the peak of the PSF. In practice, the Airy function can be well approximated by a two-dimensional Gaussian function[2] (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}. The localization error is affected by many factors and can be described by the following equation[3]:

$$\Delta x = \sqrt{\frac{s^2 + \frac{a^2}{12}}{N} + \frac{8\pi s^4 b^2}{a^2 N^2}} \tag{1.3}$$

where s, a, and b, respectively, denote the standard deviation of the PSF, the pixel size of detector, and the background noise, and N denotes the number of photons collected. Therefore, it is easy to obtain nanometer-scale resolution with moderate background noise and a fluorescent probe with reasonably high quantum yield.
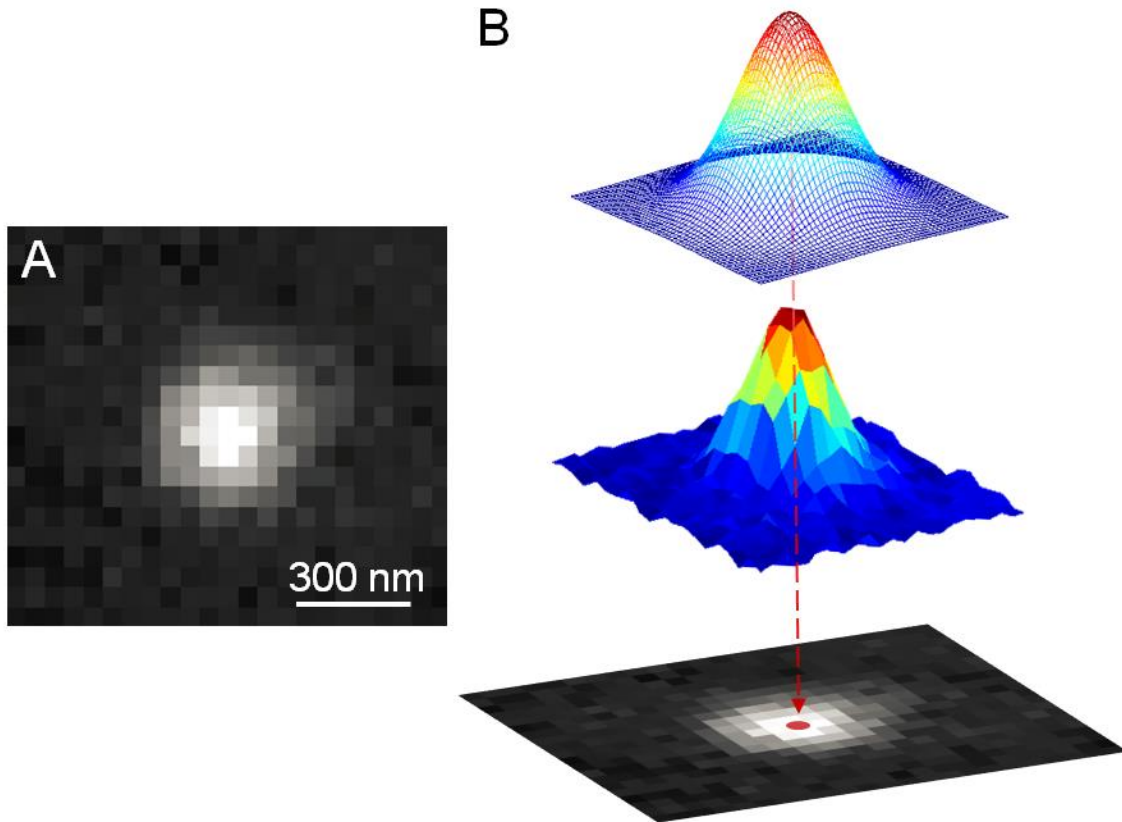
Figure 1.1 (A) The point spread function of a single-molecule emitter. (B) Using Gaussian-function fitting to resolve the localization of the single-molecule below the diffraction limit[4].

1.3 Fluorescent labeling in live bacteria

Since most subcellular components are not intrinsically fluorescent, to study the specific

biomolecules of interest in live bacteria cells with fluorescence microscopy, we must

attach a fluorescent label to the target biomolecules. Organic dyes have good quantum

yields and photostability, but they are not suitable for live-cell imaging because of the

difficulty of transporting and targeting these fluorescent probes into the cells. Fluorescent

proteins (FPs) can be genetically encoded and expressed easily, which makes them the

best candidates for live-cell imaging.

The key to beating the diffraction limit through Gaussian-function fitting described above is to isolate single-molecule fluorescent emitters. In live bacterial cell imaging, this isolation can be achieved by using photoactivatable FPs, which can be stochastically activated optically with a second, co-aligned laser source[5]. Therefore, we can activate, visualize, and localize only a sparse subset of the molecules until photobleaching, and through many iterations of this photoactivation process, a refined image with nanometer resolution can be resolved. An ideal photoactivatable fluorescent protein candidate usually has high quantum yield, good photostability, and a longer emission wavelength to avoid the cellular autofluorescence in the shorter wavelength range[6]. The names and the properties of fluorescent proteins used in this thesis are shown in Table 1.1.

Table 1.1 Properties of fluorescent proteins 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
mCherry[6]	N/A	587	610	0.22	*unknown*	*unknown*
PAmCherry1[7-9]	405	564	595	0.46	5.13×10^{-7}	3.6-80%
mCitrine[6]	N/A	516	529	0.76	*unknown*	*unknown*

The FPs, whether photoactivatable or not, must be genetically tagged to the protein of interest. In most bacteria, the easiest way to achieve this labeling is to express

the protein-FP fusion ectopically, and the expression level can be controlled by an inducible promoter if required. However, in this case, the cells will overexpress the protein (i.e., produce more proteins than needed), such that some subset of these proteins may display non-physiological behavior; this anomaly will bias the dynamics quantified, especially at the single molecule level. To avoid such artifacts, it is better to express the protein-FP fusion endogenously, such that all the protein molecules of interest are labeled and representative. However, molecular cloning of endogenous expression is usually more complicated than that of ectopic expression, and it is not always feasible. For example, adding a fluorescence tag to an essential protein might affect its normal function. An alternative method provides a compromise: knocking out the original gene in the chromosome, and ectopically expressing the protein-FP fusion as a compensation. In this case, the protein of interest cannot be essential (or else the knockdown step will lead to cell death), and the ectopic expression level needs to be tuned to the same level as in the wildtype strain.

Besides the considerations discussed above, a few other things should be noted when using fluorescent proteins. Many fluorescent proteins tend to dimerize[7], which is a property that needs to be evaluated carefully when designing experiments as it can lead to mislocalization artifacts. Also, it is possible that adding an FP tag to the protein of interest will change the original structures of either or both proteins, which might lead to the malfunction of the protein of interest, and/or change the intensity and photostability (fluorescence yield) of the FP. It is necessary to make sure that the fluorescent tag will not affect the function(s) of the protein. This functional consideration is not a problem in this thesis, since all proteins being investigated are essential proteins. Last, when using

6

most photoactivatable fluorescent proteins as fluorescent labels, even if the protein-FP

fusion is endogenously expressed and multiple activation rounds are performed, some

subset of the whole population will not be photoactivated[7] and thus will evade

characterization.

1.4 From single-molecule imaging to the characterization of dynamics

As discussed in the previous sections, single-molecule super-resolution microscopy can

be used to resolve biological superstructures, which is often performed in fixed cells so

that the structure being super-resoled remains static. In live-cell imaging, where proteins

are usually more dynamic, single-molecule super-resolution imaging can be used to

characterize the dynamics of these proteins of interest.

1.4.1 Diffusion analysis

Series of single-molecule localizations can generate single-molecule trajectories, which

can be further used to quantify the dynamics of the molecule. One important property is

the diffusion coefficient, *D*, which describes how fast the molecule moves. The *D* of a

molecule engaging in Brownian motion can be calculated with[10]:

$$\langle r^2(\tau) \rangle = 2nD\tau \tag{1.4}$$

where $\langle r^2(\tau) \rangle$ denotes the mean squared displacement (MSD) of the molecule during a

given time interval τ, and *n* denotes the dimensionality of the motion.

The correlation described in Equation 1.4 is valid when the molecule's diffusion is Brownian. The molecules being characterized in this thesis usually undergo confined motion, which can be described approximately with the following equation[11,12]:

$$\langle r^2(\tau) \rangle = \frac{L^2}{3} \cdot (1 - e^{\frac{-12 D_0 \tau}{L^2}})$$ (1.5)

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

Proteins in live cells can have multiple binding partners, undergo non-specific binding, and diffuse freely. Thus, these proteins often have different modes of motion with different diffusion coefficients. To account for this heterogeneity, we analyze each of the squared displacements and describe their distributions at various time lags using a corresponding diffusion model. The cumulative probability distribution (CPD) of square step sizes can incorporate multiple diffusion coefficient terms. For example, a two-term CPD model for 2D diffusion can be described by[13]:

$$P_{2D}(r^2, \tau) = 1 - \left[\alpha \cdot e^{\frac{-r^2}{\langle r_1^2(\tau) \rangle}} + (1 - \alpha) \cdot e^{\frac{-r^2}{\langle r_2^2(\tau) \rangle}} \right]$$ (1.6)

where $\langle r_i^2(\tau) \rangle$ is the MSD for each subpopulation $i = \{1,2\}$, and α and $(1 - \alpha)$ account for the relative abundance of each subpopulation.

Other models have been developed to study the heterogeneity of the diffusive behavior. For example, the Hidden Markov model (HMM) can account for transitions between different diffusion states of the molecule[14,15]. In HMM analysis, each step of the

8

molecule is described by its diffusion state, and such information is included in the model, where the probability of observing a certain diffusion state can be calculated generally by[16]:

$$P_{\mathrm{T}}(\mathrm{T}|\theta;\mathrm{K}) = \sum_S \prod_{j=1}^{m} P(S)\frac{1}{2\pi|\Sigma_{sj}|^{\frac{1}{2}}}\exp[-\frac{1}{2}\Delta\mathbf{x}_j^{\mathrm{T}}\Sigma_{sj}^{-1}\Delta\mathbf{x}_j] \qquad (1.7)$$

where $P(S)$ is the probability of a particular diffusion state, θ denotes all the parameters in the model, $\Delta\mathbf{x}$ is the lateral displacements, Σ is the covariance matrix describing the amplitudes of the displacements, and $\boldsymbol{\pi}$ is the relative abundance of each subpopulation. All the K possible diffusion states are summed to calculate the probability of a certain state. Such an HMM analysis is incorporated into a Gaussian mixture model (GMM), where the lateral displacements, $\Delta\mathbf{x}$, in the same diffusion subpopulation follow a multivariate normal distribution, and there are K diffusion coefficients in the model. This HMM analysis can provide not only the diffusion coefficients of the different subpopulations, but also the probabilities of the transitions between each different states, which can be used to characterize the kinetics of the chemical reactions between biomolecules with single-molecule microscopy *in vivo*.

Although not used in this thesis, in a single trajectory with different diffusion states, it is sometimes valuable to find the change points where the diffusive behavior starts to change. One method is to determine the parameters by the maximum likelihood estimators and use the likelihood-ratio test to detect such change points, so that no kinetic information is needed in the analysis[17].

Many models assume that the distribution of the molecule displacements is normal. While this assumption is true in a perfect world, in experimental data, the distribution is often skewed towards the left, making the shape of the distribution more like log-normal. Two major causes of this effect are the limited temporal resolution of the equipment and the confinement from the cell boundary. When a molecule hits the boundary of the cell, it will bounce back, but usually the camera captures only the start and the end points of the molecule, rather than capturing this whole process. Such long step sizes will be underestimated in the calculation. Therefore, the stronger the confinement and the faster the molecule, the more skewed the distribution will be. Also, many single-particle tracking algorithms favor shorter step sizes over longer ones to avoid ambiguity. Finally, many diffusion models are derived from physics, so when applying the models to the bioimaging data *in vivo*, the assumptions made by the models need to be carefully examined.

Single-particle tracking is difficult, especially in live-cell experiments when the diffusion coefficient is very high (> 1 μm^2/s), due to the limitation of the temporal resolution and the insufficient brightness of the molecule (the photons from a quickly moving molecule will be spread out over multiple imaging pixels). As an alternative, spatiotemporal image correlation spectroscopy (STICS)[18,19] can quantify the average diffusion coefficient of the molecules without tracking, which is used in Chapter 3. STICS computes the image correlation function of an entire fluorescence-imaging movie instead of using localization and tracking information, so it can resolve the dynamics of very fast diffusion that precludes fitting. Specifically, STICS computes the average of the spatial cross-correlations of all pairs of images separated by some time lag τ. This

correlation function is fit to a Gaussian function, and the long-axis variance of the Gaussian fits is called the image-mean-squared displacement (iMSD), which increases with time lag, τ. The iMSD vs. τ is plotted (Fig. 2B), and fit to a model for square-confined diffusion[11]:

$$\text{iMSD}(\tau) = \frac{L^2}{6}\left(1 - \frac{96}{\pi^4}\sum_{n\ odd}\frac{1}{n^2}\exp\left[-(\frac{n\pi}{L})^2 D\tau\right]\right) + C \qquad (1.8)$$

where D is the diffusion coefficient, L is the confinement length, and C is a constant offset.

The major drawback of STICS is that it can only calculate the average diffusion coefficient of all the molecules and it is difficult to retrieve the heterogeneous diffusion information if needed. Also, if the system has some very slow-moving or even immobile molecules, STICS will underestimate the diffusion coefficient of the molecule. Sometimes it is possible (for example, in Chapter 3) that the molecules are recruited to and released from a specific place at a very fast exchange rate. If all the molecules are illuminated, one will observe stationary foci, although the single molecules are very dynamic. In this case, when using STICS analysis, it is crucial to photoactivate an optimal number of molecules, so that on one hand, no foci can be observed, and on the other hand, the number of molecules activated is enough to calculate an average diffusion coefficient.

1.4.2 Dwell time analysis with time-lapse imaging

In live cells, some protein molecules will dwell at a specific position for a certain amount of time, indicating binding events. Such dwelling behaviors sometimes last for a few

11

seconds or even longer, and it is not possible to capture using continuous laser illumination and image acquisition, due to the limited photostability of the fluorescent proteins. In this thesis, I use time-lapse imaging to capture long dwelling behaviors at multiple timescales, and quantify the dwell time of the molecule to study the kinetics of the binding events. Specifically, after the photoactivation of the fluorescent proteins, every frame is captured with a 50-ms integration time (τ_{int}) followed by a time delay of 0 – 1.45 s (τ_{delay}). This time-lapse enables us to capture dwelling events that last for even a few seconds despite the finite (< 1 s) photobleaching lifetime of the fluorescent proteins. The total time-lapse period (τ_{tl}) is defined as the sum of τ_{int} and τ_{delay}.

The distribution of the dwell times at each τ_{tl} was plotted and fit to an exponential decay function, f:

$$f(\tau) = Ae^{\frac{-\tau}{k_{eff}}} \tag{1.9}$$

where k_{eff} is the effective off-rate of the molecule. Two independent factors contribute to this rate of apparent dissociation: the physical dissociation of the molecule, described by the rate constant k_{off}, and the photobleaching of the fluorescent protein, described by the rate constant k_b[20]:

$$k_{eff}\tau_{tl} = k_b\tau_{int} + k_{off}\tau_{tl} \tag{1.10}$$

We extract k_{off} by fitting $k_{eff}\tau_{tl}$ vs. τ_{tl} to a linear function, in which the slope corresponds to the real dissociation rate constant k_{off}. The dwell time (τ_{dwell}) is the reciprocal of k_{off}.

1.5 DNA replication in bacteria

The contents of this section include the following reference:

Li, Y., Schroeder, J.W., Simmons, L.A. and Biteen, J.S., 2018. Visualizing bacterial DNA replication and repair with molecular resolution. *Current opinion in microbiology*, 43, pp.38-45.

Understanding how genomes are accurately copied and maintained during the cell cycle is of fundamental importance to all of biology. When cells fail to replicate their DNA accurately, numerous disease states or even cell death can occur in multicellular organisms, while in bacteria, DNA replication mistakes affect cell fitness and viability[21]. Therefore, it is critical for cells to replicate their DNA accurately and efficiently. In all cells, this process is accomplished by replisomes, which are multi-protein machines containing all components necessary to duplicate genomic DNA[22]. The architecture and dynamics of the *Escherichia coli* bacterial replisome has been extensively characterized (Figure 1.2a)[23-25]. The *E. coli* replisome is the Pol III holoenzyme (HE), a molecular machine composed of ten different proteins grouped into several functional subassemblies[26]. The Pol III core is the catalytic subassembly important for replication; it has three subunits: α, which has DNA polymerase activity, ε, which has exonuclease activity, and θ, which is involved in proofreading[27-29]. The Pol III* subassembly is the core plus the protein τ and the γ complex, both involved in clamp loading [30,31]. Finally, the Pol III HE is Pol III* plus the β sliding clamp encoded by *dnaN* [32,33]. The β clamp enhances processivity and increases the DNA synthesis rate up to a thousand fold by holding the DNA polymerase on the template strand to increase the number of nucleotides synthesized per binding event [33,34]. The multiprotein clamp loader complex is

required to open and close the β clamps around DNA for loading and unloading during replication [30]. Apart from the Pol III HE, the DNA helicase (DnaB) is essential and required for DNA replication [35]. The helicase is linked to Pol III HE through τ [36]. With the help of the replication initiation proteins DnaA and DnaC, the *E. coli* replisome assembles at *oriC* [37] and then tracks along DNA until the end of replication[24].

A replisome that exhibits a different mechanism is found in *Bacillus subtilis*, a prototypical Gram-positive bacterium (Figure 1.2b). Rather than tracking along the DNA strands, the *B. subtilis* replisome is relatively stationary during replication, and the template DNA is pulled into the replisome for duplication [38,39]. Unlike *E. coli*, the *B. subtilis* replisome has two distinct replicative DNA polymerases: PolC and DnaE. Reconstitution of the *B. subtilis* replisome *in vitro* has shown that PolC is responsible for all leading strand synthesis and most of the lagging DNA strand synthesis, while DnaE is used to extend the lagging strand RNA primer before handing off to PolC [40]. The DNA replication mechanism of *B. subtilis* is more reminiscent of DNA replication in eukaryotic systems. For example in human lagging strand replication, polymerase Pol α extends RNA primers before handing off to Pol δ [41,42].
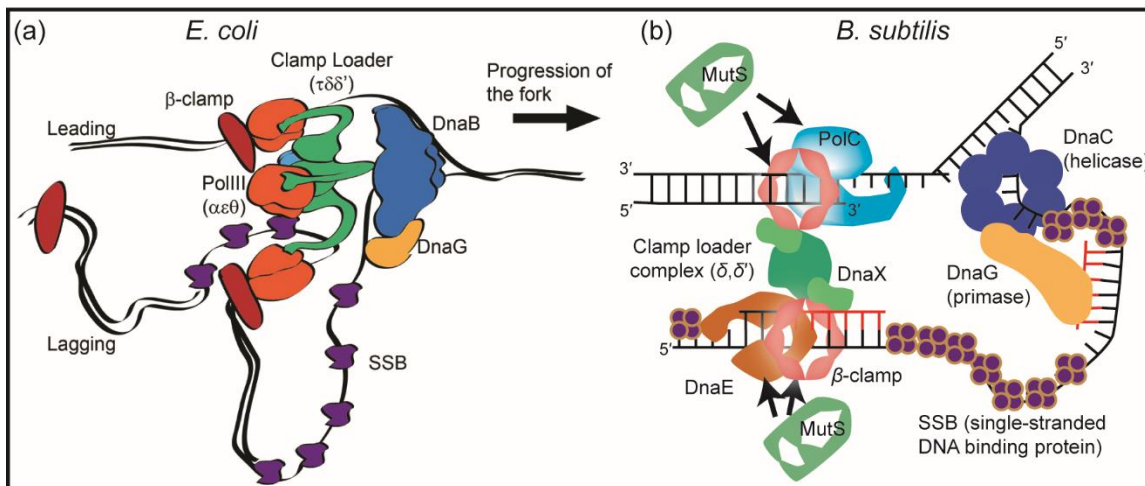
Figure 1.2 Molecular architecture of the DNA replication fork. (a) The architecture of the *E. coli* DNA replication fork. Reproduced from *eLife* 2017, 6:10.7554 43[43]. Creative commons. (b) The architecture of the *B. subtilis* DNA replication fork.

Biochemical methods and bulk fluorescence imaging have been used to understand how cells maintain their genomic integrity [24,40,44-51]. However, such methods generally lose the heterogeneity of a system by averaging over all observations. Understanding all the complex and highly dynamic aspects of DNA replication requires quantitative biophysical tools, and molecular resolution is necessary to characterize the dynamics of all subpopulations, to understand detailed kinetics, and to uncover intricate mechanisms that are relevant *in vivo*. Single-molecule microscopy is a great solution to this problem because it captures the different behaviors of every molecule [52] and provides nanometer-scale resolution [53], even in living cells [4]. By using photoactivatable or photoswitchable fluorescent labels, one can observe and image the fluorescence from individual single emitters, and each isolated molecule image can be analyzed to obtain the nanometer-scale locations and trajectory of that molecule [2,5,54]. Importantly, single-molecule microscopy

15

can be easily applied *in vivo* in bacteria [55-57], making it now possible to visualize DNA replication in real time in living bacterial cells.

1.6 Thesis outline

The limited previous *in vivo* studies of the *B. subtilis* replisome mainly suffer from the poor resolution from the conventional microscopy techniques, and consequently, the architecture of the replisome and the dynamics of each of the replisome components remain unclear. In this thesis, I will use super-resolution microscopy to investigate the architecture and dynamics of the replisome in live *B. subtilis* cells. Overall, we have learned that although bulk fluorescence data show that the replisome is stationary in the cell, the replisomal proteins are more dynamic than anticipated.

In Chapter 2, I use super-resolution imaging and single-molecule tracking to analyze the localization and quantify the dynamics of a replicative DNA polymerase in *B. subtilis*: PolC. To understand the mechanism of DNA replication more deeply, I use a drug to arrest DNA synthesis, analyze, and compare the effect on the dynamics of PolC. My findings show that the PolC is highly dynamic and undergoes fast exchanges, with rates that correspond to the specific functions.

In Chapter 3, I will continue my study of the replisome by characterizing the dynamics of three replisomal proteins, the two DNA polymerases PolC and DnaE, and a *β*-clamp loader subunit DnaX, in live *Bacillus subtilis* cells. I have made the DnaX-PAmCherry (endogenous expressed) strain, and the dynamics and dwell times are analyzed under three different conditions (untreated and treated by two DNA synthesis-inhibiting drugs). My results suggest a possible coupling between PolC and DnaX in the

DNA replication process, and I uncovered an unexpected phenotype of DnaX: molecules that dwell for a very long time at the replisome, under drug-induced DNA damage.

Apart from the dynamics of replisome components, the stoichiometry of each replisome component is of great interest. However, we lack an objective method that can estimate the stoichiometry without the bias introduced by subjective parameters. In Chapter 4, I develop an approach that counts the bleaching steps of a cluster of fluorescent molecules without using any subjective parameters or thresholds based on Bayesian statistics. I apply this method to determine the stoichiometry of DnaX in live *B. subtilis* cells.

Finally, in Chapter 5, I summarize the key findings of the previous chapters and discuss the future research directions that will be essential for us to better and more fully understand the architecture of the replisome and the mechanism of DNA replication in live bacteria cells.

References

1. Abbe E. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für Mikroskopische Anatomie*. 1873;9(1):413-468.

2. Moerner WE. Single-molecule mountains yield nanoscale images. *Nature Meth*. 2006;3:781-782.

3. Thompson RE, Larson DR, Webb WW. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J*. 2002;82:2775-2783.

4. Tuson HH, Biteen JS. Unveiling the inner workings of live bacteria using super-resolution microscopy. *Anal Chem*. 2015;87(1):42-63.

5. Betzig E, Patterson GH, Sougrat R, et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*. 2006;313(5793):1642-1645.

6. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nat Methods*. 2005;2(12):905-909.

7. Wang S, Moffitt JR, Dempsey GT, Xie XS, Zhuang X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc Natl Acad Sci U S A*. 2014;111(23):8452-8457.

8. Durisic N, Laparra-Cuervo L, Sandoval-Álvarez Á, Borbely JS, Lakadamyali M. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an *in vivo* nanotemplate. *Nat Methods*. 2014;11(2):156-162.

9. Subach FV, Patterson GH, Manley S, Gillette JM, Lippincott-Schwartz J, Verkhusha VV. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat Methods*. 2009;6(2):153-159.

10. Qian H, Sheetz MP, Elson EL. Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophys J*. 1991;60(4):910-921.

11. 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. *Biophys J*. 1993;65(5):2021-2040.

12. Lommerse PHM, Snaar-Jagalska BE, Spaink HP, Schmidt T. Single-molecule diffusion measurements of H-ras at the plasma membrane of live cells reveal microdomain localization upon activation. *J Cell Biol*. 2005;118:1799-1809.

13. Sonnleitner A, Schutz GJ, Schmidt T. Free brownian motion of individual lipid molecules in biomembranes. *Biophys J*. 1999;77(5):2638-2642.

14. Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, Mellman I. Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature*. 2010;464(7289):783-U163.

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

16. Yu J. Single-molecule studies in live cells. *Annual Review of Physical Chemistry, Vol 67*. 2016;67:565-585.

17. Yin S, Song N, Yang H. Detection of velocity and diffusion coefficient change points in single-particle trajectories, https://Doi.org/10.1016/j.bpj.2017.11.008, in press. .

18. Rowland DJ, Tuson HH, Biteen JS. Resolving fast, confined diffusion in bacteria with image correlation spectroscopy. *Biophys J*. 2016;110:2241-2251.

19. Hebert B, Costantino S, Wiseman PW. Spatiotemporal image correlation spectroscopy (STICS) theory, verification, and application to protein velocity mapping in living CHO cells. *Biophys J*. 2005;88(5):3601-3614.

20. Gebhardt JC, Suter DM, Roy R, et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature Meth*. 2013;10(5):421-426.

21. Lenhart JS, Schroeder JW, Walsh BW, Simmons LA. DNA repair and genome maintenance in bacillus subtilis. *Microbiol Mol Biol Rev*. 2012;76(3):530-564.

22. Johnson A, O'Donnell M. Cellular DNA replicases: Components and dynamics at the replication fork. *Annu Rev Biochem*. 2005;74:283-315.

23. O'Donnell M. Replisome architecture and dynamics in escherichia coli. *J Biol Chem*. 2006;281(16):10653-10656.

24. Reyes-Lamothe R, Possoz C, Danilova O, Sherratt DJ. Independent positioning and action of escherichia coli replisomes in live cells. *Cell*. 2008;133(1):90-102.

25. Reyes-Lamothe R, Sherratt DJ, Leake MC. Stoichiometry and architecture of active DNA replication machinery in escherichia coli. *Science*. 2010;328(5977):498-501.

26. Kelman Z, O'Donnell M. DNA polymerase III holoenzyme: Structure and function of a chromosomal replicating machine. *Annu Rev Biochem*. 1995;64:171-200.

27. Maki H, Kornberg A. The polymerase subunit of DNA polymerase III of escherichia coli. II. purification of the alpha subunit, devoid of nuclease activities. *J Biol Chem*. 1985;260(24):12987-12992.

28. Welch MM, McHenry CS. Cloning and identification of the product of the dnaE gene of escherichia coli. *J Bacteriol*. 1982;152(1):351-356.

29. Scheuermann R, Tam S, Burgers PM, Lu C, Echols H. Identification of the epsilon-subunit of escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: A fidelity subunit for DNA replication. *Proc Natl Acad Sci U S A*. 1983;80(23):7085-7089.

30. Jeruzalmi D, O'Donnell M, Kuriyan J. Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. *Cell*. 2001;106(4):429-441.

31. Onrust R, Finkelstein J, Naktinis V, Turner J, Fang L, O'Donnell M. Assembly of a chromosomal replication machine: Two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. I. organization of the clamp loader. *J Biol Chem*. 1995;270(22):13348-13357.

32. Stukenberg PT, Studwell-Vaughan PS, O'Donnell M. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J Biol Chem*. 1991;266(17):11328-11334.

33. Kong XP, Onrust R, O'Donnell M, Kuriyan J. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell*. 1992;69(3):425-437.

34. Mizrahi V, Henrie R, Marlier J, Johnson K, Benkovic S. Rate-limiting steps in the DNA polymerase I reaction pathway. *Biochemistry*. 1985;24(15):4010-4018.

35. Beyersmann D, Messer W, Schlicht M. Mutants of escherichia coli B-r defective in deoxyribonucleic acid initiation: dnaI, a new gene for replication. *J Bacteriol*. 1974;118(3):783-789.

36. Kim S, Dallmann HG, McHenry CS, Marians KJ. Coupling of a replicative polymerase and helicase: A tau-DnaB interaction mediates rapid replication fork movement. *Cell*. 1996;84(4):643-650.

37. Chodavarapu S, Kaguni JM. Replication initiation in bacteria. *The Enzymes*. 2016;39:1-30. doi: http://dx.doi.org/10.1016/bs.enz.2016.03.001.

38. Lemon KP, Grossman AD. Localization of bacterial DNA polymerase: Evidence for a factory model of replication. *Science*. 1998;282(5393):1516-1519.

39. Sawitzke J, Austin S. An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol Microbiol*. 2001;40(4):786-794.

40. Sanders GM, Dallmann HG, McHenry CS. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol Cell*. 2010;37(2):273-281.

41. Kunkel TA, Burgers PM. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol*. 2008;18(11):521-527.

42. Kunkel TA. Balancing eukaryotic replication asymmetry with replication fidelity. *Curr Opin Chem Biol*. 2011;15(5):620-626.

43. Beattie TR, Kapadia N, Nicolas E, et al. Frequent exchange of the DNA polymerase during bacterial chromosome replication. *eLife*. 2017;6:10.7554.

44. Smith B, Grossman A, Walker G. Visualization of mismatch repair in bacterial cells. *Mol Cell*. 2001;8(6):1197-1206.

45. Kleczkowska H, Marra G, Lettieri T, Jiricny J. hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev*. 2001;15(6):724-736.

46. Lenhart JS, Sharma A, Hingorani MM, Simmons LA. DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. *Mol Microbiol*. 2013;87(3):553-568.

47. Hombauer H, Campbell CS, Smith CE, Desai A, Kolodner RD. Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell*. 2011;147(5):1040-1053.

48. Klocko AD, Schroeder JW, Walsh BW, Lenhart JS, Evans ML, Simmons LA. Mismatch repair causes the dynamic release of an essential DNA polymerase from the replication fork. *Mol Microbiol*. 2011;82(3):648-663.

49. 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*. 2009;284(47):32782-32791.

50. Lemon K, Grossman A. Movement of replicating DNA through a stationary replisome. *Mol Cell*. 2000;6(6):1321-1330.

51. Meile J, Wu L, Ehrlich S, 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*. 2006;6(7):2135-2146.

52. Monachino E, Spenkelink LM, van Oijen AM. Watching cellular machinery in action, one molecule at a time. *J Cell Biol*. 2017;216(1):41-51.

53. Yildiz A, Selvin PR. Fluorescence imaging with one nanometer accuracy: Application to molecular motors. *Acc Chem Res*. 2005;38(7):574-582.

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

55. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*. 2007;316(5828):1191-1194.

56. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. *Science*. 2006;311:1600-1603.

57. Biteen JS, Thompson MA, Tselentis NK, Bowman GR, Shapiro L, Moerner WE. Super-resolution imaging in live *caulobacter crescentus* cells using photoswitchable EYFP. *Nat Methods*. 2008;5(11):947-949.

Chapter 2: Single-molecule DNA polymerase (PolC) dynamics at a bacterial replisome in live cells

The contents of this chapter were published in the following reference:

Liao, Yi, Yilai Li, Jeremy W. Schroeder, Lyle A. Simmons, and Julie S. Biteen. "Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells." Biophysical journal 111, no. 12 (2016): 2562-2569.

This Chapter refers to a collaborative project. I contributed the data in Figure 2.1 and 2.5, and assisted with data collection for Figure 2.3 and 2.4.

PolC is one of two essential replicative DNA polymerases found in the Gram-positive bacterium *Bacillus subtilis*. The *B. subtilis* replisome is eukaryotic-like in that it relies on more than one essential DNA polymerase catalytic subunit for chromosomal replication. To quantitatively describe how the replicative DNA polymerase PolC functions in *B. subtilis*, we applied photobleaching-assisted microscopy, three-dimensional super-resolution imaging, and single-particle tracking to examine the *in vivo* behavior of PolC at single-molecule resolution. We report the stoichiometry of PolC proteins within each cell and within each replisome, we elucidate the diffusion characteristics of individual PolC molecules, and we quantify the exchange dynamics for PolC engaged in lagging strand synthesis. We show that PolC is highly dynamic: this polymerase is constantly recruited to and released from a centrally located replisome, providing new insight into the architecture and organization of the replisome in bacterial cells.

23

2.1 Introduction

All cells must accurately duplicate and segregate their chromosomal DNA, and failure to do so is a critical underlying cause of over 80 different human diseases and genetic disorders[1]. Replisomes are multiprotein assemblies that are responsible for DNA replication in cells[2-4]. DNA polymerases are the essential replisome components that synthesize new DNA molecules from nucleotides to create new DNA strands with high fidelity[5,6]. It is therefore appropriate that great efforts have been made to better understand how DNA polymerases work both *in vitro* and *in vivo*. For example, the stoichiometry and architecture of the replicative DNA polymerase holoenzyme (Pol III) have been well characterized in the model organism *Escherichia coli*[7-9]. 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 bacterial species have distinct organizations and operate differently[10-12]. In particular, the replisome in the Gram-positive model bacterium *Bacillus subtilis* does not seem to replicate DNA by actively tracking along DNA, in contrast to the *E. coli* replisome[7,13,14]. Rather, the *B. subtilis* replisome has been shown to reside in a more restricted location through which template DNA is pulled in and newly synthesized DNA is extruded[15], although the replicative DNA polymerase in *B. subtilis* has not been analyzed at single-molecule resolution. In addition, *B. subtilis* utilizes two distinct types of essential DNA polymerases, PolC and DnaE, for genome replication[10,12], and *B. subtilis* uses two helicase loaders[16]. The use of two replicative polymerases and two helicase loaders in *B. subtilis* differs from *E. coli* and is more similar to the mechanism found in eukaryotes. Overall, the *in vivo* composition and architecture of the elongating replicative DNA

polymerase in *B. subtilis* is poorly understood.

     In vitro reconstitution of the *B. subtilis* replisome has demonstrated that PolC is responsible for all leading strand synthesis as well as most lagging strand synthesis, whereas the more error prone and much slower DNA replicase DnaE (25 nt/s for DnaE compared to ~500 nt/s for PolC) plays a crucial role in initiating lagging strand synthesis[10,17]. DnaE is important for extending the lagging strand RNA primer before handing off to PolC, which completes replication of the Okazaki fragment[10]. The synergistic relationship between two polymerases in the *B. subtilis* replisome resembles the synergy found in eukaryotic replication. For example, in *S. cerevisiae*, two essential replicases[18], Pol ε and Pol δ, are respectively responsible for synthesizing the leading and the lagging strands[19-21]. In analogy with DnaE in *B. subtilis*, the eukaryotic polymerase Pol α extends RNA primers for a short segment before handing off to Pol δ[22]. Given these observations, the less well-understood *B. subtilis* replisome appears more eukaryotic-like than the *E. coli* replisome, and a deeper *in vivo* understanding of how DNA polymerase recruitment and dynamics are performed in *B. subtilis* will provide valuable insight into how bacteria and eukaryotic organisms use a two DNA polymerase system for replication. Here, we apply photobleaching-assisted microscopy[23,24], three-dimensional (3D) single-molecule super-resolution microscopy[25], and single-particle tracking[26,27] to PolC in live *B. subtilis* cells to measure the stoichiometry, intracellular distribution, and dynamics of the essential DNA polymerase required for the majority of leading and lagging strand synthesis.

2.2 Copy number of PolC within a cell and the stoichiometry of PolC at replisome

To investigate the copy number of PolC in live *B. subtilis* cells, we constructed *B. subtilis* strains (YL001) natively expressing PolC fused to the red fluorescent protein mCherry[28] as the sole source of PolC. The PolC-mCherry fusion remained intact in the cells as shown by the Western blot (Figure 2.1). Under identical imaging conditions, the fluorescence intensity level in cells harboring PolC-mCherry was elevated compared to the intensity in unlabeled wild type (wt) PY79 cells (Figure 2.2a-b). This elevated fluorescence emission comes from all the PolC-mCherry molecules in the cytosol, and the average number of PolC molecules per cell can be approximated by subtracting the median background fluorescence of wt cells from the median fluorescence intensity of labeled cells and then dividing the background-subtracted labeled cell fluorescence intensity by the intensity of a single PolC-mCherry molecule. Our estimation from 136 cells indicates that on average 61 copies of PolC are present within each cell; this copy number is slightly higher than the 40 copies of pol III core enzymes quantified in *E. coli*[29].
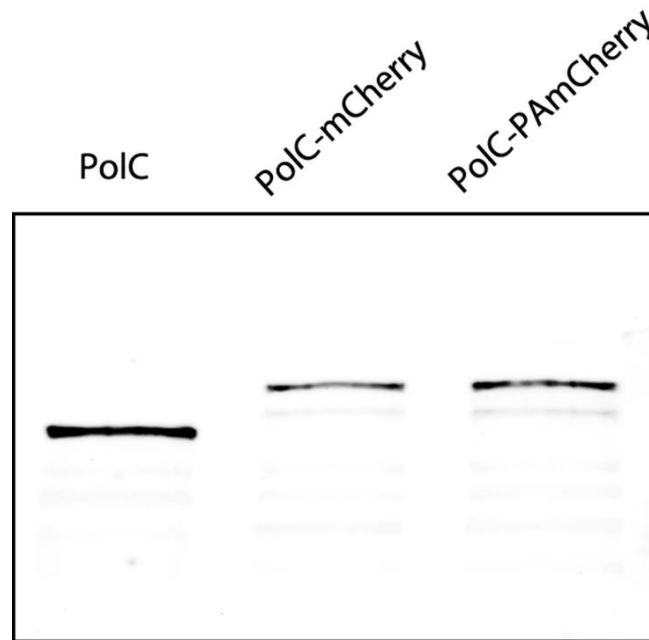
Figure 2.1 Western blots of PolC in (wt) PY79 cells, PolC-mCherry in YL001, and PolC-PAmCherry in JWS 213. Both PolC-mCherry and PolC-PAmCherry remain intact in the cells.

As expected for a spatially restricted replisome, and consistent with previous studies[2], PolC-mCherry forms distinct foci at the mid-cell position or near the quarter-cell positions (Figure 2.2a). The PolC-mCherry foci likely represent the site(s) of active DNA replication[2]. Photobleaching experiments and *in vitro* measurements of *E. coli* have found that three copies of DnaE are positioned at each replication fork[7,30], while other *in vitro* studies have found two copies[9]. It is therefore important to determine the stoichiometry of the essential DNA polymerase PolC in *B. subtilis*. Performing photobleaching-assisted microscopy on live cells with PolC-mCherry, we observe PolC-mCherry foci photobleaching in real time (Figure 2.2c). To determine the stoichiometry of PolC at the site of DNA replication, we divided the intensity of the detected state[31] corresponding to the highest copy number of PolC (I_{max}) by the intensity change corresponding to the

smallest jump between two states (ΔI_{min}). Compared to simply counting the number of

photobleaching steps, this method gives accurate stoichiometry information even when

multiple copies of molecules are photobleached within a single frame (Figure 2.2c, right

panel). The distribution of PolC-mCherry copy numbers (Figure 2.2d) indicates that 3—

or in the case where two sister replication forks spatially overlap, 6—copies of PolC are

present at each replication fork. If PolC were responsible *in vivo* for just leading strand

replication, we would have expected approximately one PolC molecule at each fork and

two at forks that spatially overlap. Our observation that three copies of PolC are present

at each fork is most supportive of the model that PolC is responsible for leading strand

replication and the majority of lagging strand replication *in vivo*. Taken together with our

estimation that about 61 total copies of PolC are present within each cell, these findings

pose an intriguing question: why do the cells express so many copies of PolC that do not

appear to be associated with the replisome?
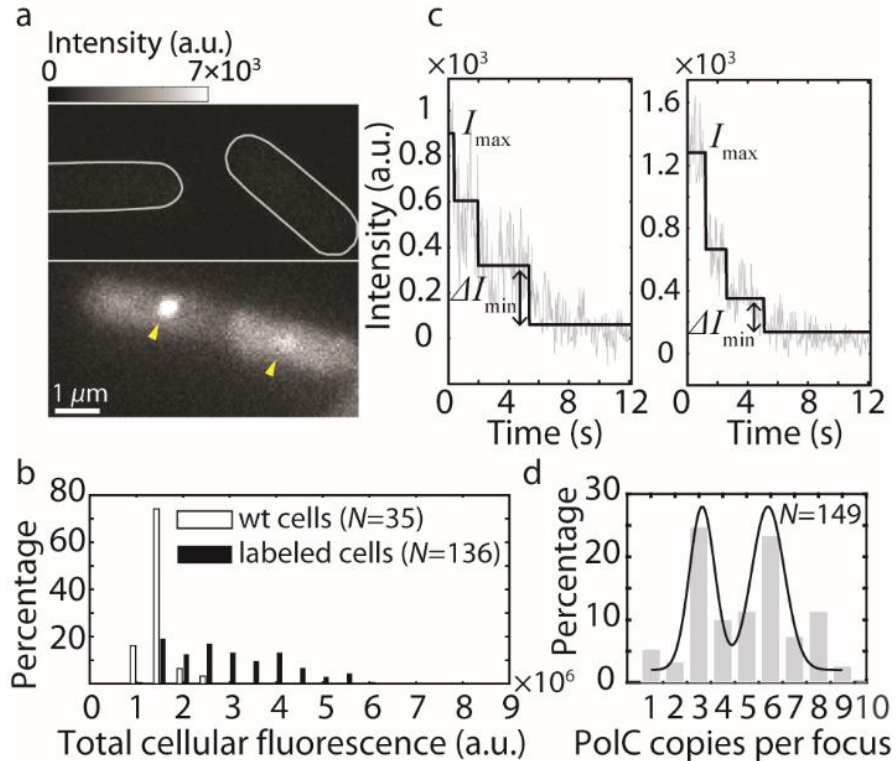
Figure 2.2 PolC stoichiometry within each cell and at each replisome. a) Representative fluorescence image of wild type (wt) *B. subtilis* cells (upper) and cells with PolC-mCherry (lower). PolC-mCherry foci are highlighted by yellow arrows. b) Distribution of initial total cellular fluorescence intensity for wt cells (median $= 1.43 \times 10^6$) and cells labeled with PolC-mCherry (median $= 2.77 \times 10^6$). The difference in fluorescence indicates the total intensity of mCherry in labeled cells. c) Representative background-subtracted intensity time traces (grey lines) of two different PolC-mCherry foci undergoing photobleaching, where different intensity states (black lines) are identified by maximum likelihood estimation. The left panel shows typical, fairly uniform steps, whereas the right panel shows a scenario where multiple photobleaching events occurred within a single frame. d) Distribution of PolC copy numbers at each PolC-mCherry focus. The black line is a guide to the eye.

2.3 *In vivo* localization and diffusion of PolC

To visualize both replisome-associated and free cytosolic PolC, we replaced the mCherry label on PolC with the photoactivatable red fluorescent protein PAmCherry[32]. The fusion remained intact in the cells as shown by the Western blot (Figure 2.1). This fusion allows us to visualize the dynamics of individual PolC proteins at the single-

molecule level by stochastically switching a small subset (1 – 3 molecules per cell) of the many PolC-PAmCherry molecules into a fluorescent state at a time. To follow the motion of single PolC molecules in these ~1 – 2 μm thick cells, we tracked each photoactivated molecule with astigmatism-based 3D super-resolution microscopy to simultaneously resolve the lateral (x, y) as well as the axial (z) position of single PolC molecules[25]. 3D tracking allowed us to unambiguously resolve PolC motion and dwelling within the 3D confines of a live bacterial cell.

Consistent with what was observed for PolC-mCherry (Figure 2.2a), we detect PolC-PAmCherry enrichment near the mid-cell or quarter-cell positions (red arrows in Figure 2.3a), whereas the rest diffuses throughout the cell more rapidly (Figure 2.3a). Furthermore, using single-particle tracking, we unambiguously detect the PolC-PAmCherry molecules in 3D, including occasional direct observations of a molecule changing from a fast-diffusing mode to a slow dwelling motion (Figure 2.3a, inset). Many PolC-PAmCherry molecules diffuse at a rate slower than 0.1 μm²/s (Figure 2.3b), but none were strictly stationary, as all molecules have diffusion coefficients larger than the apparent diffusion coefficient of truly stationary PolC-PAmCherry molecules (0.003 μm²/s) measured in fixed cells. Referring to other DNA-binding proteins of comparable sizes[33,34], the range of PolC-PAmCherry diffusion coefficients indicates that the slower moving PolC proteins are engaged in confined motion, and we interpret the slow moving PolC-PAmCherry as a subpopulation actively engaged in DNA replication.

Previously, we observed with fluorescence microscopy that two *B. subtilis* replisome subunits, the processivity clamp protein DnaN and the clamp loader protein

DnaX, both engage in subtle motions confined within a small domain of ~100 nm[35]; these movements likely represent the dynamics of the *B. subtilis* replisome in general. Here, defining consecutive steps with 3D displacements < 100 nm as dwelling, we next probe where such dwelling events occur for PolC, whether the PolC dwelling positions are consistent with the location of other replisomal subunits reported earlier, and how the dwelling positions depend on cell growth rate. We found that in most cells of typical lengths of $2 - 4\,\mu m$, PolC predominantly dwelled at the quarter cell positions (Figure 2.3c), consistent with diffraction-limited images of PolC, DnaX, and DnaN[2,36]. Thus, the slower moving PolC molecules appear to be interacting with the replication fork. In fast growing *B. subtilis* cells, new rounds of genome replication can start before cell division occurs, giving rise to 3 or even 4 replication forks per cell[15]. In our experiment, we did indeed observe that the spatial distribution of PolC dwelling events is more dispersed in longer cells with lengths of $4 - 6\,\mu m$ (Figure 2.3c). For example, in longer cells, PolC tends to dwell at the cell center in addition to at the quarter positions, indicating that PolC molecules are actively engaged in DNA replication at multiple sites in fast growing cells.

Notably, the change from fast to slow motion for PolC at the replisome is similar to the behavior we observed for the DNA mismatch repair protein MutS[35]. This observation suggests that the replisome may function as a scaffolding "hub" through which DNA polymerases and repair proteins are recruited to and released from active sites of DNA replication. Such an exchange could represent a convenient mechanism for the cell to place more confined scaffolding proteins to act as a recruitment hub to coordinate the more dynamic proteins, important for completing other critical tasks necessary for genome replication, maintenance and repair.

We asked if PolC would persist through interactions with other replication proteins in the absence of ongoing DNA replication. We hypothesized that rapid replication fork arrest would not disperse PolC, and that PolC would continue to be actively recruited to existing replisome regions, while new replisome complexes would not be assembled in the absence of ongoing DNA synthesis. To test the effect of rapid replication fork arrest on PolC localization and dynamics, we treated cells with the PolC-specific inhibitor HPUra[10,37]. HPUra immediately blocks DNA synthesis without collapsing the replisome complex, as previously determined through bulk fluorescence imaging[38]. 3D super-resolution images of HPUra-treated cells show that PolC still forms foci at the quarter-cell and mid-cell positions (Figure 2.3d), and the overall distribution of PolC diffusion coefficients remains similar to that of untreated cells (Figure 2.3e). However, the spatial distribution of dwelling events was affected by HPUra treatment: whereas PolC still preferentially slows down at quarter-cell positions in cells of average length, dwelling at other cellular positions, including the mid-cell, is much less frequent in longer cells (Figure 2.3f). These HPUra experiments demonstrate that pausing DNA replication will stop formation of new replisome complexes, but will not prevent PolC from interacting with existing replisomes. Our results further suggest that PolC is recruited only after the assembly of the replisome has begun.
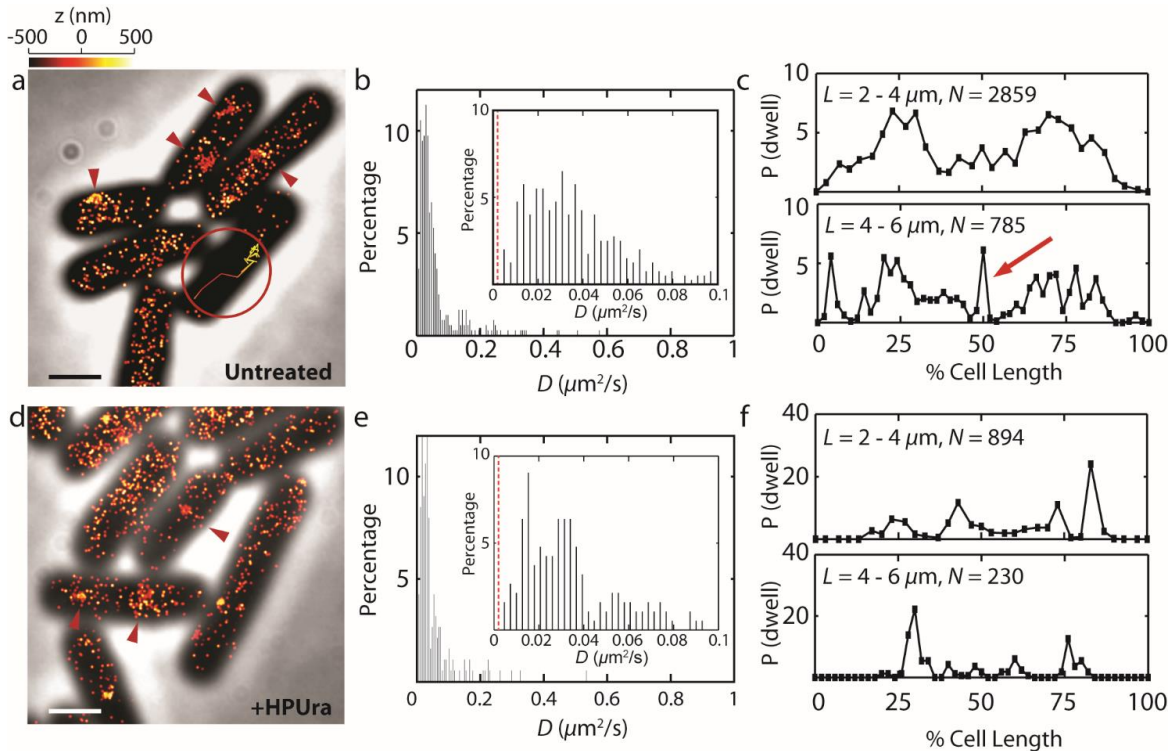
Figure 2.3 Diffusion and dwelling behavior of PolC-PAmCherry in live *B. subtilis* cells. a) 3D super-resolution reconstruction image of PolC-PAmCherry in untreated *B. subtilis*, overlaid on a phase contrast image of the cells. The position of each localization is indicated by a single dot with width corresponding to the average localization precision in the lateral plane (25 nm). The axial (*z*) position is color-coded according to the color bar above. Red arrows indicate regions of PolC enrichment in the cell. Inset: A representative, color-coded 3D single-molecule trajectory illustrates a PolC-PAmCherry molecule making a transition from diffusing (bottom) to dwelling (top). b) Distribution of PolC-PAmCherry diffusion coefficients, *D*, in untreated cells. The inset zooms in on the $0 - 0.1$ μm^2/s region of the original histogram. Red dashed line: average apparent diffusion coefficient (0.003 μm^2/s) for stationary PolC-PAmCherry molecules measured in fixed cells. c) Localization probabilities of dwelling events along the longitudinal cellular axis in untreated cells. *L*: cell length, *N*: total number of dwelling events analyzed. Red arrow: the mid-cell dwelling events that disappeared after HPUra treatment. d) 3D super-resolution reconstruction image of PolC-PAmCherry in HPUra-treated cells. e) Distribution of PolC-PAmCherry diffusion coefficients, *D*, in HPUra-treated cells. f) Localization probabilities of dwelling events in HPUra-treated cells. Scale bars = 1 μm.

2.4 Dwell time of PolC at the replisome

Under normal conditions, in addition to interacting with various replisomal subunits, PolC also interacts with DNA extensively as it replicates DNA. Thus, although PolC undergoes dynamic exchange at the replisome regardless of ongoing DNA replication, the extent to which PolC dwells at the replisome likely depends on whether it is engaged in DNA synthesis. To test the relationship between PolC dwell times and DNA synthesis, we quantified the dwell times for PolC during normal growth and following HPUra treatment.

Due to the limited photostability of fluorescent proteins, continuous illumination yields single-molecule PolC-PAmCherry trajectories that last on average for only about 15 imaging frames (~750 ms) before PAmCherry undergoes irreversible photobleaching. We therefore turn to time-lapse imaging to capture dwelling behavior of PolC-PAmCherry molecules at multiple timescales[39]. In our time-lapse imaging mode, every frame is still captured with a 50-ms image integration time (τ_{int}), but a time delay (τ_{delay}) of 0 – 1.45 s 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} (Figure 2.4a).

To measure the time scale of PolC dwelling events, we model the interaction between PolC and various replisomal subunits as a simple association/dissociation reaction:

$$\text{Replisome} \cdot \text{PolC} \xtofrom{k_{diss\_app}} \text{Replisome} + \text{PolC} \qquad (1)$$

This reaction has a forward reaction rate constant of $k_{\text{diss\_app}}$. This *apparent* dissociation rate of PolC is caused by PAmCherry photobleaching as well as PolC dissociating from the replisome. The $k_{\text{diss\_app}}$ can be obtained by plotting the distribution of PolC dwell times, τ_{measured}, for each of five time-lapse imaging experiments performed with $\tau_{\text{TL}} = 50$ ms, 500 ms, 700 ms, 1 s, and 1.5 s, respectively (Figure 2.4b). As above, two sequential localizations separated by < 100 nm in 3D are counted as one dwelling event. In each case, $\tau_{\text{measured}} = (n - 1) \times \tau_{\text{TL}}$, where n is the number of sequential frames where dwelling events occurred, and the distribution of τ_{measured} follows an exponential decay function:

$$f = Ae^{\frac{-\tau_{\text{measured}}}{k_{\text{diss\_app}}}} \tag{2}$$

During imaging, two Poisson processes contribute to the observed apparent rate of dissociation of PolC-PAmCherry: (i) photobleaching of PAmCherry with a rate constant of $k_{\text{bleaching}}$, and (ii) PolC dissociating from the replisome with a rate constant of k_{diss}. Photobleaching of PAmCherry occurs only when the laser illumination is on (i.e., only during τ_{int}), while the physical dissociation of PolC from the replisome can occur anytime during the time-lapse interval τ_{TL}. These two processes contribute independently to the apparent dissociation, and respectively produce each of the two terms on the right hand side of (40):

$$k_{\text{diss\_app}} \cdot \tau_{\text{TL}} = k_{\text{bleaching}} \cdot \tau_{\text{int}} + k_{\text{diss}} \cdot \tau_{\text{TL}} \tag{3}$$

Equation (3) indicates that the relation between $k_{\text{diss\_app}} \cdot \tau_{\text{TL}}$ and τ_{TL} is linear. The average dwell time, τ_{dwell}, of PolC can thus be extracted by plotting $k_{\text{diss\_app}} \cdot \tau_{\text{TL}}$ vs. τ_{TL}

(Figure 2.4b, inset): dividing the y-intercept of this equation by τ_{int} yields the

photobleaching rate constant $k_{bleaching} = 42.1 \pm 6.4$ s^{-1}, and the slope corresponds to the

real dissociation rate constant $k_{diss} = 0.36 \pm 0.09$ s^{-1}, which is the reciprocal of the PolC

dwell time constant $\tau_{dwell} = 2.79 \pm 0.47$ s. It should be noted that τ_{dwell} does not depend on

whether or how long PolC has been dwelling before PAmCherry was photoactivated, as

exponential distributions are by definition memoryless: the dwell time distribution

remains the same even if PolC-PAmCherry molecules have been bound to the replisome

for some amount of time before being photoactivated and imaged.

Because our measurements do not discriminate between PolC engaged in leading

or lagging strand synthesis, and because many more copies of PolC are required to

synthesize the lagging strands due to the shorter length of the Okazaki fragments, we

hypothesize that the majority of our measurements of PolC-PAmCherry correspond to

PolC synthesizing the lagging strands. In *B. subtilis*, the speed of the replication fork is

about 500 nt/s (10, 38). With our measured dwell time close to 3 seconds, these data

correspond to a PolC molecule synthesizing ~1500 nucleotides each time it is recruited to

the replisome, which is consistent with the typical length of ~1-2 kbp for a single

Okazaki fragment on the lagging strand for *B. subtilis* (10). Unlike lagging strands that

are synthesized in short segments, the leading strands are synthesized more continuously

by a single PolC protein for an extended period of time. Therefore, the dwell time for

PolC engaged in leading strand synthesis is likely to be significantly longer than that of

PolC synthesizing the lagging strand, and we do indeed occasionally detect PolC copies

dwelling for much longer; we attribute these longer (up to ~5-10 s) dwell times to PolC

synthesizing the leading strand. If these measurements do in fact represent PolC engaged

36

in leading strand synthesis, we would expect that 2,500-5,000 nucleotides are incorporated during the longer dwell times that we observe.

To determine whether the PolC dwell time is affected by arresting DNA replication, we calculated the PolC dwell time in the presence of HPUra, and we found that τ_{dwell} decreased to 0.97 ± 0.04 s (Figure 2.4c). Because no DNA replication takes place after HPUra is added, the residual dwelling behavior of PolC likely results from interactions between PolC and certain replisome proteins, such as the replication sliding clamp DnaN, as well as from transient, nonspecific interactions between PolC and DNA, and this decreased τ_{dwell} establishes a baseline for discriminating between replicating PolC and these transient interactions.
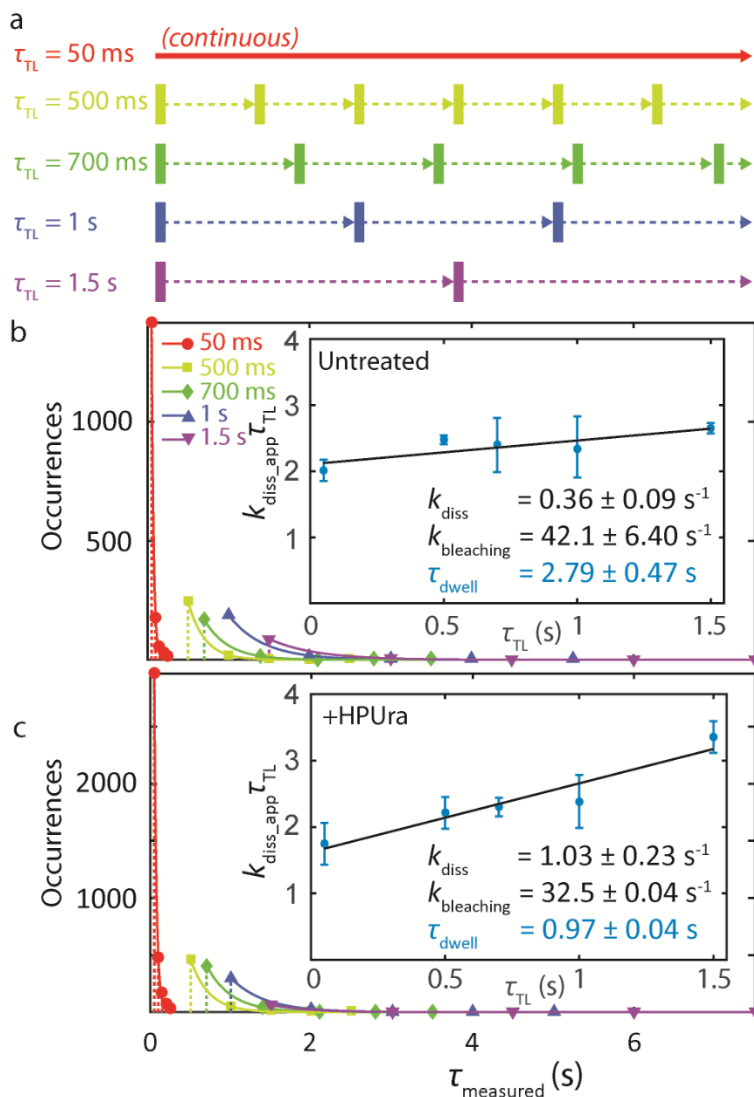
Figure 2.4 Dwell time analysis for PolC-PAmCherry. a) Scheme for time-lapse imaging. Every imaging frame is captured with a 50-ms integration time (solid rectangles), and a time delay (dashed arrows) is introduced between each pair of consecutive frames. The time-lapse period (τ_{TL}) is the sum of the integration time and the time delay. b) Dwell time distributions for PolC-PAmCherry in untreated cells. For clarity, the distributions are shown as stem plots, and the color corresponds to τ_{TL} as described in 'a'. The solid lines are fits to the exponential decay in Equation (2). Inset: linear fit (black line) of ($k_{diss\_app} \cdot \tau_{TL}$) vs. τ_{TL}, from which the dissociation rate constant, k_{diss}, the photobleaching rate constant, $k_{bleaching}$ and the dwell time constant, τ_{dwell} are obtained according to Equation (3). Errors bars are from 4 rounds of bootstrapping. c) Dwell time distributions and analysis for PolC-PAmCherry in HPUra-treated cells.

2.5 Conclusions

Here, we have applied photobleaching-assisted microscopy, 3D super-resolution microscopy and single-particle tracking to quantify the copy number, localization and dynamics of the essential DNA polymerase PolC in live *B. subtilis* cells. At the single-molecule level, the behavior of PolC is highly dynamic. We show that PolC is constantly recruited to and released from a centrally located replisome. Our results support the model that during genome replication, the relatively confined replisome is assembled and then functions as a central hub to coordinate the recruitment of PolC and other more dynamic proteins to the site of active DNA replication. Our stoichiometric analysis of PolC showing three and six copies in a focus supports *in vitro* models that PolC is responsible for leading strand and the majority of lagging strand replication (10). Furthermore, our imaging data and dwell time analysis suggests that PolC exchanges on a rapid timescale, leading to the possibility that the seemingly large number of PolC molecules in the cell is necessary to ensure that the probability of PolC recruitment to the replisome via PolC-DnaN (β-clamp) interactions is sufficiently high. Therefore, a large number of PolC molecules per cell would ensure that PolC is always positioned at the assembled replisome to engage in lagging strand synthesis. Such a mechanism would allow for efficient synthesis of each Okazaki fragment as PolC rapidly exchanges.

2.6 Materials and methods

2.6.1 Microscopy

B. subtilis cells from a plate were inoculated in $S7_{50}$ minimal medium at OD_{600} ~ 0.1, followed by growth with shaking at 30 ^{o}C for ~3.5 hours to OD_{600} ~0.5 – 0.6. When it was used, a final concentration of 162 μM HPUra was added to the cell culture immediately prior to imaging. Cells were deposited on pads of 1% agarose in $S7_{50}$, each of which was sandwiched between two coverslips. As discussed in our prior studies, *B. subtilis* is highly susceptible to fluorescent impurities present on the coverslips and in the growth medium[40]. The impurities produce single-molecule-like fluorescent signals that largely resemble those from PAmCherry in both brightness and photostability. We prevented such artifacts by cleaning the coverslips in an oxygen plasma (PE-50, Plasma Etch Inc.) at 200 mTorr for at least 20 min and by using freshly prepared growth medium and agarose pads each day to achieve nearly background-free imaging conditions (Figure 2.2a).

Following preparation, the sample was mounted on an Olympus IX71 widefield inverted microscope for imaging, with appropriate excitation, dichroic, and emission filters (Semrock) placed along the light path to achieve optimal signal-to-noise. Fluorescence emission was collected by a 1.40-NA 100× oil-immersion phase-contrast objective and detected on a 512×512 pixel electron multiplying CCD detector (Photometrics Evolve). Cells were still able to grow under the microscope, as shown in the growth curve in Figure 2.5.
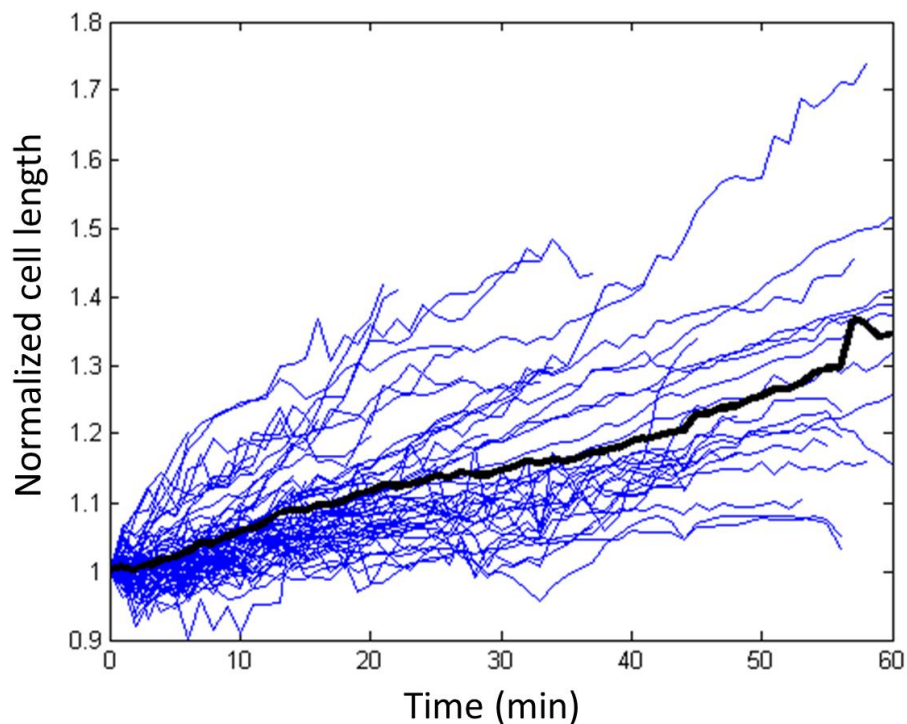
Figure 2.5 Normalized cell length over time under the microscope at 30 degrees for 61 cells. Black line shows the mean of the growth curves. When a cell divides, the growth curve stopped at that time point.

In our photobleaching-assisted microscopy, the cells were illuminated by a 561-nm laser (Coherent Sapphire 561-50) with a power density of 50 W/cm^2, and images were recorded at 30 ms/frame. In our 3D super-resolution imaging, a weak cylindrical lens in the emission pathway between the microscope and the camera induced astigmatism[25], such that the ellipticity of the microscope point spread function changed based on the axial (z) position of the emitting molecule. To visualize only 1 – 3 copies of PolC-PAmCherry at a time, a 200-ms 405-nm laser (Coherent 405-100) pulse with power density ~100 W/cm^2 was used to photoactivate PAmCherry, and PolC-PAmCherry

molecules were subsequently imaged by a 561-nm laser with a power density of 400 W/cm^2 at 50 ms/frame.

2.6.2 Image processing

In both photobleaching-assisted microscopy and 3D super-resolution microscopy, initial estimates for signal locations were made based on band-pass filtered images. For photobleaching-assisted microscopy (Figure 2.2), the center pixel of each mCherry focus was determined from the filtered image. In each raw data image, the average intensity of all pixels within 3 pixels of this center pixel was measured. Background fluorescence intensity for each cell in each imaging frame was determined from the average intensity value of all other pixels within the cell not belonging to any mCherry focus. This average background intensity of each cell was subtracted from the intensity of each mCherry focus. The resulting background-corrected focus intensity was recorded over time to monitor the photobleaching process for that mCherry focus and to generate the intensity time traces as in Figure 2.2c.

In the case of 3D super-resolution microscopy (Figure 2.3), unfiltered images were fit at the positions of the initial guesses with an asymmetric Gaussian function,

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

where I_{bg}, A, x_0, and y_0, respectively denote the background intensity, the amplitude of emission, and the x- and y-center positions of the molecule. The widths of the signal in the x-direction (σ_x) and the y-direction (σ_y) characterize the elliptical emission signal,

and we pre-calibrated the relationship between (σ_x, σ_y) and z-position using 0.1 μm

TetraSpeck™ beads and a PIFOC® piezo objective scanner. For each localization, we

then obtained the z-position of the molecule by finding the z value that minimizes the

difference between the measured (σ_x, σ_y) and $(\sigma_{x-\text{calibration}}, \sigma_{y-\text{calibration}})$.

Table 2.1 Bacterial strains used in this study.

Strain Name	Genotype	Source
PY79	Wild type, SPβ°	Youngman P., Perkins, J. B., and Losick, R. 1984. *Plasmid* 12:1-9.
YL001	*polC-mCherry*	This work.
JWS213	*polC*-PAmCherry	This work.

References

1. DePamphilis ML. *DNA replication and human disease.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006.

2. Lemon KP, Grossman AD. Localization of bacterial DNA polymerase: Evidence for a factory model of replication. *Science*. 1998;282(5393):1516-1519.

3. Migocki M, Lewis P, Wake R, Harry E. The midcell replication factory in bacillus subtilis is highly mobile: Implications for coordinating chromosome replication with other cell cycle events. *Mol Microbiol*. 2004;54(2):452-463.

4. Berkmen MB, Grossman AD. Spatial and temporal organization of the bacillus subtilis replication cycle. *Mol Microbiol*. 2006;62(1):57-71.

5. Johnson A, O'Donnell M. Cellular DNA replicases: Components and dynamics at the replication fork. *Annu Rev Biochem*. 2005;74:283-315.

6. McHenry CS. DNA replicases from a bacterial perspective. *Annu Rev Biochem*. 2011;80(1):403-436.

7. Reyes-Lamothe R, Sherratt DJ, Leake MC. Stoichiometry and architecture of active DNA replication machinery in escherichia coli. *Science*. 2010;328(5977):498-501.

8. Georgescu RE, Kurth I, O'Donnell ME. Single-molecule studies reveal the function of a third polymerase in the replisome. *Nature Structural & Molecular Biology*. 2012;19(1):113-116.

9. Dohrmann PR, Correa R, Frisch RL, Rosenberg SM, McHenry CS. The DNA polymerase III holoenzyme contains gamma and is not a trimeric polymerase. *Nucleic Acids Res*. 2016;44(3):1285-1297.

10. Sanders GM, Dallmann HG, McHenry CS. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol Cell*. 2010;37(2):273-281.

11. Sun J, Shi Y, Georgescu RE, et al. The architecture of a eukaryotic replisome. *Nature Structural & Molecular Biology*. 2015;22(12):976-982.

12. Dervyn E, Suski C, Daniel R, et al. Two essential DNA polymerases at the bacterial replication fork. *Science*. 2001;294(5547):1716-1719.

13. Joshi MC, Bourniquel A, Fisher J, et al. Escherichia coli sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc Natl Acad Sci U S A*. 2011;108(7):2765-2770.

14. Bates D, Kleckner N. Chromosome and replisome dynamics in E. coli: Loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*. 2005;121(6):899-911.

15. Lemon K, Grossman A. Movement of replicating DNA through a stationary replisome. *Mol Cell*. 2000;6(6):1321-1330.

16. Velten M, McGovern S, Marsin S, Ehrlich S, Noirot P, Polard P. A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol Cell*. 2003;11(4):1009-1020.

17. Le Chatelier E, Becherel O, d'Alencon E, et al. Involvement of DnaE, the second replicative DNA polymerase from bacillus subtilis, in DNA mutagenesis. *J Biol Chem*. 2004;279(3):1757-1767.

18. Kunkel TA. Balancing eukaryotic replication asymmetry with replication fidelity. *Curr Opin Chem Biol*. 2011;15(5):620-626.

19. Georgescu RE, Schauer GD, Yao NY, et al. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *Elife*. 2015;4:e04988.

20. McElhinny SAN, Gordenin DA, Stith CM, Burgers PMJ, Kunkel TA. Division of labor at the eukaryotic replication fork. *Mol Cell*. 2008;30(2):137-144.

21. Pursell ZF, Isoz I, Lundstroem E, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*. 2007;317(5834):127-130.

22. Kunkel TA, Burgers PM. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol*. 2008;18(11):521-527.

23. Burnette DT, Sengupta P, Dai Y, Lippincott-Schwartz J, Kachar B. Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proc Natl Acad Sci USA*. 2011;108(52):21081-21086.

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

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

26. Yildiz A, Forkey JN, McKinner SA, Ha T, Goldman YE, Selvin PR. Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science*. 2003;300:2061-2065.

27. Elmore S, Muller M, Vischer N, Odijk T, Woldringh C. Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in escherichia coli. *J Struct Biol*. 2005;151(3):275-287.

28. Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *discosoma* sp. red fluorescent protein. *Nat Biotechnol*. 2004;22(12):1567-1572.

29. Maki H, Kornberg A. The polymerase subunit of DNA polymerase III of escherichia coli. II. purification of the alpha subunit, devoid of nuclease activities. *J Biol Chem*. 1985;260(24):12987-12992.

30. McInerney P, Johnson A, Katz F, O'Donnell M. Characterization of a triple DNA polymerase replisome. *Mol Cell*. 2007;27(4):527-538.

31. Watkins LP, Yang H. Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B*. 2005;109:617-628.

32. Subach FV, Patterson GH, Manley S, Gillette JM, Lippincott-Schwartz J, Verkhusha VV. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat Methods*. 2009;6(2):153-159.

33. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*. 2007;316(5828):1191-1194.

34. Uphoff S, Reyes-Lamothe R, de Leon FG, Sherratt DJ, Kapanidis AN. Single-molecule DNA repair in live bacteria. *Proc Natl Acad Sci U S A*. 2013;110(20):8063-8068.

35. Liao Y, Schroeder JW, Gao B, Simmons LA, Biteen JS. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proc Natl Acad Sci U S A*. 2015;112(50):E6898-E6906.

36. Simmons LA, Davies BW, Grossman AD, Walker GC. Beta clamp directs localization of mismatch repair in bacillus subtilis. *Mol Cell*. 2008;29(3):291-301.

37. Wang JD, Sanders GM, Grossman AD. Nutritional control of elongation of DNA replication by (p)ppGpp. *Cell*. 2007;128(5):865-875.

38. Goranov A, Katz L, Breier A, Burge C, Grossman A. A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. *Proc Natl Acad Sci U S A*. 2005;102(36):12932-12937.

39. Gebhardt JC, Suter DM, Roy R, et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature Meth*. 2013;10(5):421-426.

40. Tuson HH, Aliaj A, Brandes ER, Simmons LA, Biteen JS. Addressing the requirements of high sensitivity single-molecule imaging of low-copy number proteins in bacteria. *ChemPhysChem*. 2016;17:1435-1440.

Chapter 3: Single-molecule dynamics of three replisomal proteins in live bacteria cells

The contents of this chapter will be included in the following reference:

The replisome is a multiprotein complex that replicates DNA in the cell. Using single-molecule super-resolution imaging, we characterized the dynamics of three replisomal proteins in live *Bacillus subtilis* cells: the two replicative DNA polymerases, PolC and DnaE, and a β-clamp loader subunit, DnaX. We quantified their mobility and dwell times during normal replication and upon replication arrest by two different pathways. We understand the mechanism of the highly dynamic and cooperative process of DNA replication based on changes in the measured diffusion coefficients and dwell times. These experiments reveal that the replisomal proteins are all highly dynamic and that the exchange rate depends on whether replication is active or arrested. Our results also suggest coupling between PolC and DnaX in the DNA replication process, and indicate that DnaX might play a role in replication repair.

3.1 Introduction

The replisome is composed of proteins that work together to replicate DNA; in bacteria, this molecular machine includes the DNA polymerase(s), the *β*-clamp and the *β*-clamp loader complex, the helicase, the primase, and the single-stranded DNA binding protein (SSB)[1]. Overall, the bacterial replisome has been investigated at length both *in vitro* and *in vivo*[1-6].

In particular, the replisome of *Bacillus subtilis*, a model organism for Gram-positive bacteria, has been studied extensively[1,5,6]. Unlike the replisome in Gram-negative bacteria such as *Escherichia coli*, the *B. subtilis* replisome is confined at a specific position in the cell[7], and the DNA templates are pulled into the replisome to be replicated[6,8,9]. Moreover, the *B. subtilis* replisome requires two different DNA polymerases, PolC and DnaE, for replication[10], which resembles the mechanism of eukaryotic replication[11,12]. *In vitro* study of the *B. subtilis* replisome suggests that PolC is responsible for all the leading strand synthesis and most of the lagging DNA strand synthesis, whereas DnaE, an error-prone polymerase, extends the lagging strand RNA primer before handing off to PolC[1]. However, the *in vitro* model of DNA replication is incomplete, and studies in living bacteria are furthering our understanding these processes. For instance, Seco et al. recently reported that DnaE might be involved in the *B. subtilis* leading strand synthesis as well[13]. Moreover, other replisomal proteins can modulate the activity and fidelity of DnaE polymerase, suggesting that DnaE is capable of replicating substantial amounts DNA in live *B. subtilis* cells[14].

Despite a multitude of *in vitro* and *in vivo* studies of the *B. subtilis* replisome, the architecture and dynamical interactions of the replisome components in live *B. subtilis* cells remain unclear and the functions of some replisomal proteins in the normal DNA replication process, for example DnaE, are still under debate. Recently, the development of single-molecule imaging has provided a new tool to study DNA replication in living cells with high sensitivity and spatial resolution[13,15-18]. In Chapter 2, we characterized the stoichiometry and the dynamics of one of the major DNA polymerases, PolC, in live *B. subtilis* cells[18]. Still, the replisome proteins work cooperatively to replicate DNA. Thus, more contextual information is needed to understand the whole picture of the DNA replication in *B. subtilis*. Here, we use single-molecule microscopy and single-particle tracking to determine the localization and dynamics of three replisomal proteins during normal DNA replication: the PolC and DnaE replicative DNA polymerases and the DnaX subunit of the *β*-clamp loader. Based on these measurements during normal DNA replication, we elucidate the DNA replication mechanism by examining the localization and dynamics of the replisomal proteins when the DNA replication is arrested via different mechanisms: PolC disruption with 6-hydroxy-phenylazo-uracil (HPUra)[19,20] or cross-linking with mitomycin C (MMC)[21]. Overall, the subcellular positioning, motion, and responses to DNA replication arrest indicate that the replisomal proteins exchange dynamically during DNA replication and that these dynamics change when DNA replication is arrested.

3.2 Dynamical positioning of replisome proteins in live *B. subtilis* cells during normal DNA replication

To study the localization and dynamics of the replisome proteins, we genetically engineered fusions of the photoactivatable red fluorescent protein PAmCherry to the C-terminal of PolC (JWS213), DnaE (LAM380.1), or DnaX (LYL001) as the sole source of these essential proteins at the native loci of the respective genes. The PolC-PAmCherry strain also includes an ectopically expressed DnaX-mCitrine fusion under control of a xylose promoter, as a marker of the replisome position[15]. We stochastically photoactivated 1 – 3 PAmCherry molecules per cell at a time, imaged this photoactivated subset until all PAmCherry was photobleached, and then photoactivated a new subset. Super-resolution images were constructed after 5 – 10 iterations of this photoactivation-imaging-photobleaching cycle. For the PolC-PAmCherry strain, DnaX-mCitrine foci were imaged before any photoactivation of PAmCherry. Consistent with earlier findings[15], DnaX-mCitrine forms clusters at the mid-cell or quarter-cell positions (Figure 3.1A). PolC-PAmCherry molecules are enriched around the replisome area, although a significant number of molecules are distributed distal to the replisome throughout the cell (Figure 3.1A).

To measure the heterogeneous dynamics of PolC molecules, we further tracked PolC-PAmCherry molecules using single-particle tracking (Figure 3.1B). We differentiated between two sorts of dynamical behaviors within the heterogeneous PolC molecule motion by fitting the cumulative probability distribution (CPD) of the squared step sizes to a diffusion model with two mobile terms, and we thus quantified the fast and

slow diffusion coefficients of the two PolC subpopulations. By analyzing 1230 tracks, we found (Figure 3.2, Figure 3.3A) that although about 73% PolC molecules move slowly in the cells (average diffusion coefficient $D_{PolC\text{-}slow} = 0.026 \pm 0.005$ μm^2/s), a significant amount (27%) of PolC molecules diffuse much more rapidly, with an average diffusion coefficient of $D_{PolC\text{-}fast} = 0.5 \pm 0.2$ μm^2/s. To understand how the motion varies with position within the cell, we mapped the step sizes of PolC as a function of distance from the replisome (Figure 3.4). We found that the step sizes of PolC decrease near the DnaX foci, implying that the slowly diffusing PolC-PAmCherry molecules correspond to PolC actively effecting replication in the replisome, whereas the fast population corresponds to PolC searching along the DNA strands away from the replisome.
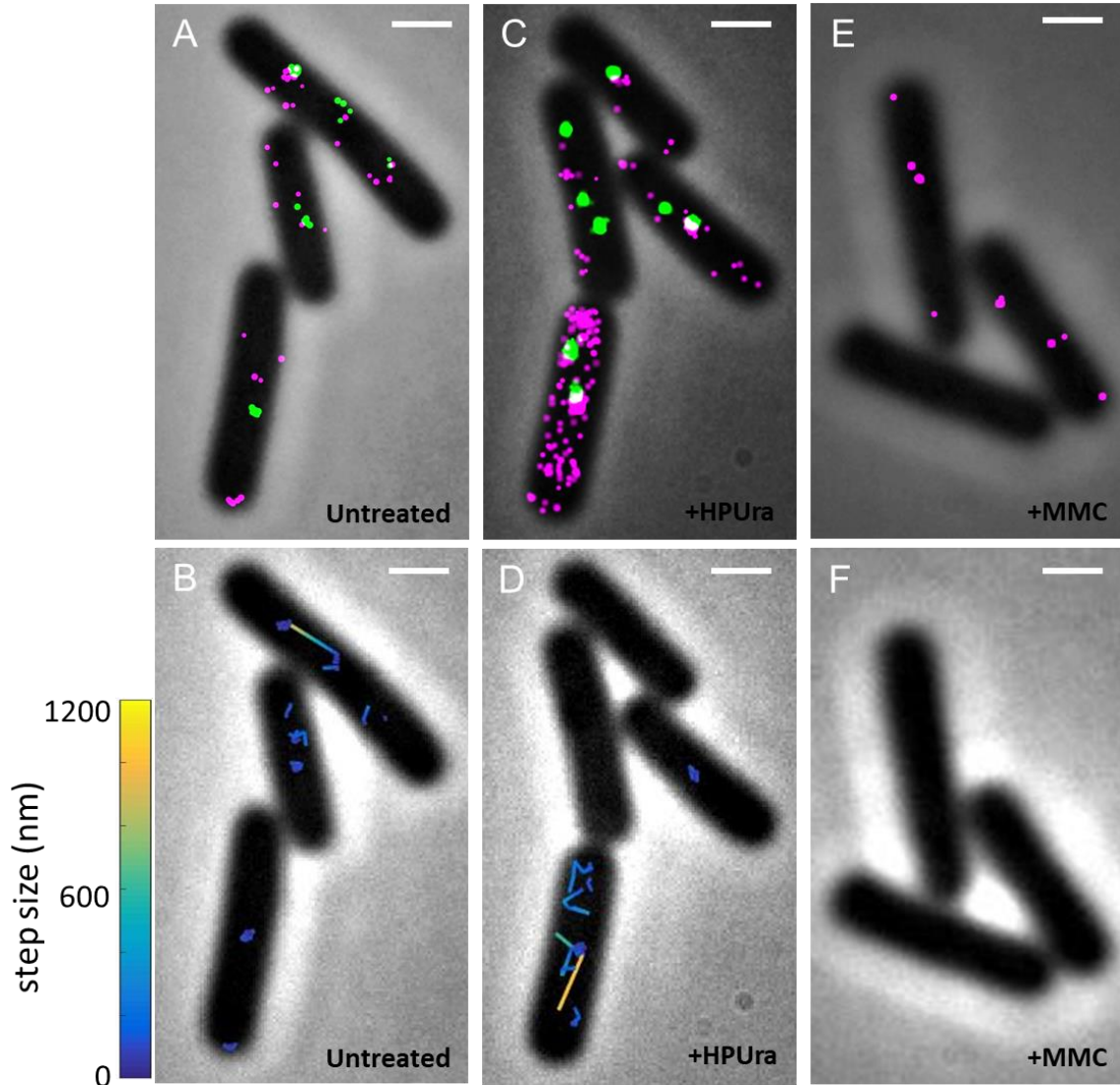
Figure 3.1 Location and dynamics of PolC in representative live *B. subtilis* cells. (A) Photoactivated localization microscopy (PALM) reconstructions of PolC-PAmCherry (magenta) overlaid with DnaX-mCitrine clusters (green) in untreated cells. (B) Representative trajectories of PolC-PAmCherry in untreated cells. The trajectories are color-coded by step size to distinguish the fastest motion (yellow) from the slowest motion (blue). (C) PALM reconstructions of PolC-PAmCherry (magenta) and DnaX-mCitrine (green) in HPUra-treated cells. (D) Representative step-size-coded trajectories of PolC-PAmCherry in HPUra-treated cells (D). (E) PALM reconstructions of PolC-PAmCherry (magenta) in MMC-treated cells. (F) PolC-PAmCherry molecules were un-trackable in MMC-treated cells. Scale bar: 1 μm.
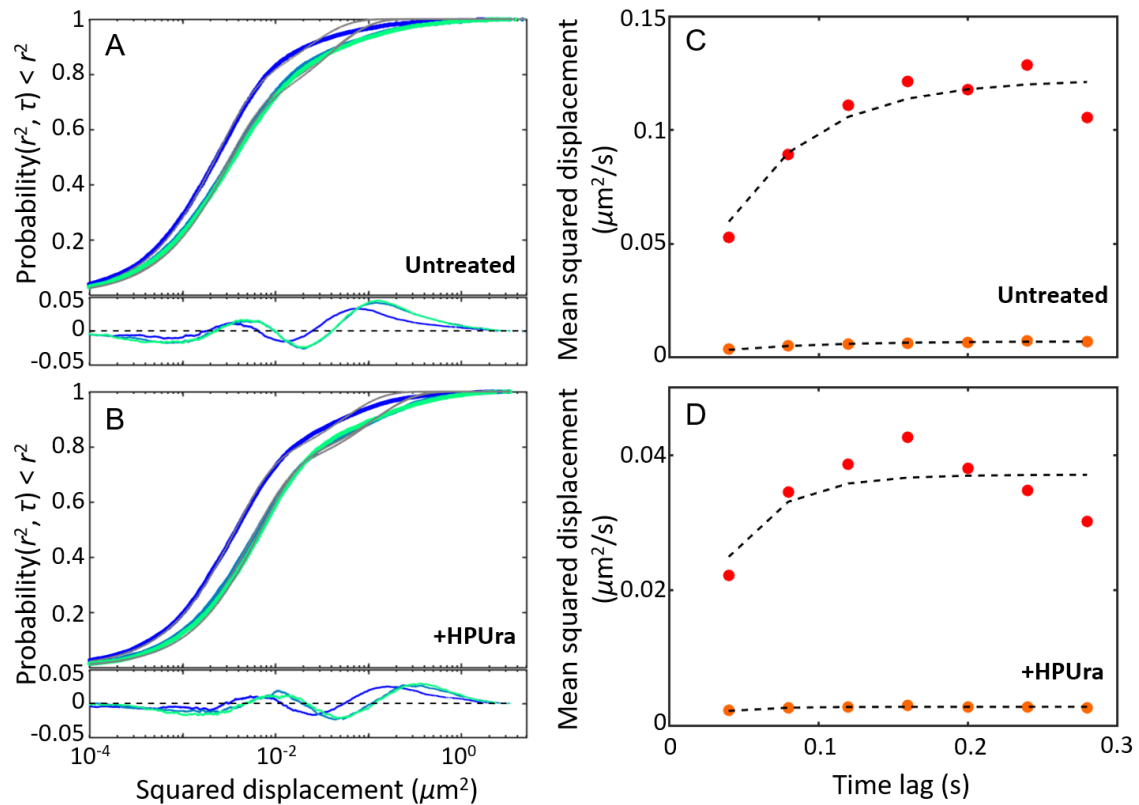
Figure 3.2 Fitting the cumulative probability distribution (CPD) of squared step sizes separates the heterogeneous PolC dynamics into the constitutive fast motion and slow motion. (A) – (B) CPD of squared displacements for each of the first three time lags, $\tau = \{40, 120, 200\ ms\}$ for (A) PolC-PAmCherry in untreated cells and (B) PolC-PAmCherry in HPUra-treated cells. These distributions are fit to our model (Eq. 2) for two mobile, diffusive populations (grey lines) and the fit residuals are plotted below. (C) – (D) The mean squared displacement (MSD) for each population (red: 'fast' and orange: 'slow') is extracted from the fits in (A) and (B) and plotted versus time lag, τ, for the first seven time lags for (C) PolC-PAmCherry in untreated cells and (D) PolC-PAmCherry in HPUra-treated cells. The MSD curves were fit to a model for square-confined diffusion (Eq. 3). The initial slopes of the curves describe the diffusion coefficients, $D_{PolC-slow}$ and $D_{PolC-fast}$. The saturation level of the curves is related to the confinement length, L, of the two populations.

DnaE, which is a smaller protein than PolC, is the other polymerase essential to DNA replication in *B. subtilis*. We photoactivated 1 – 3 DnaE-PAmCherry molecules per cell at a time and imaged under the microscope. Unlike PolC, DnaE does not show any measurable heterogeneous dynamics. Rather, although DnaE-PAmCherry molecules

were photoactivated (Figure 3.5), no DnaE-PAmCherry could be tracked at our imaging

rate of 50 ms/frame, and DnaE trajectories were not even measurable even upon imaging

at a faster rate of 20 ms/frame. This observation is consistent with extremely rapid

diffusion of these polymerase molecules, which means that DnaE molecules are freely

mobile in the cytoplasm most of the time, and the amount of time they must spend at the

replisome is far shorter than the imaging speed. This motion is not surprising, since DnaE

is responsible for extending the lagging strand DNA after the RNA primers[1] before

handing off to PolC, a process that should happen in less than 10 ms. To quantify the

dynamics of the un-trackable DnaE-PAmCherry, we applied the spatiotemporal image

correlation spectroscopy (STICS) algorithm to our data (Figure 3.3B); this analysis uses

image correlation instead of tracking information to calculate the average diffusion

coefficient in a cell. The DnaE-PAmCherry diffusion coefficient of $D_{DnaE,avg} = 1.9 \pm 0.2$

$\mu m^2/s$ was calculated by STICS (Figure 3.3D).

Because DNA replication requires the cooperation of a number of proteins[1], the

dynamics of all replication machinery proteins must be coordinated at the replisome. We

therefore also quantified the dynamics of the β-clamp loader component DnaX-

PAmCherry. We photoactivated 1 – 3 DnaX-PAmCherry molecules per cell at a time and

visualized the molecules in our microscope. Three types of DnaX-PAmCherry motions

are observed. The fast motion is too fast to be tracked ($D_{DnaX-fast} > 1 \mu m^2/s$) even when

imaging with a fast rate of 20 ms/frame, and STICS could not be used to quantify this fast

diffusion coefficient since it cannot differentiate between the fast and slow dynamics.

The medium population, representing about 33% of the trackable molecules, has a

diffusion coefficient the same order of magnitude of $D_{PolC-fast}$ ($D_{DnaX-med} = 0.2 \pm 0.1$

μm^2/s), which corresponds to the DnaX molecules searching along the DNA strands. The slow population demonstrates obvious dwelling, in which a single molecule was observed to remain at a sub-diffraction-limited position for a significant amount of time (> 100 ms). To understand this dwelling, we quantified the dwell time of DnaX-PAmCherry. Here, consecutive steps with displacements < 100 nm are considered to demonstrate dwelling, since the radius of gyration, which describes subtle motions of DnaX foci and thus the replisome, has been determined to be 84 ± 20 nm[15]. Because the photobleaching time of the PAmCherry tag occurs on the same seconds timescale as the DnaX dwell times, we uncovered the true dwell time with time-lapse imaging (Methods). We measured an average DnaX dwell time of $\tau_{DnaX-dwell}$ 2.63 ± 0.97 s (Figure 3.6A, 3.6B), which is similar to previous reports of the PolC dwell time in *B. subtilis*[15] ($\tau_{PolC-dwell} = 2.79 \pm 0.41$ s). These dwell times correspond to the time needed for the synthesis of ~1500 DNA nucleotides, which is within the typical length of a single Okazaki fragment (~1 – 2 kbp) synthesized on the lagging strand in *B. subtilis*[1].
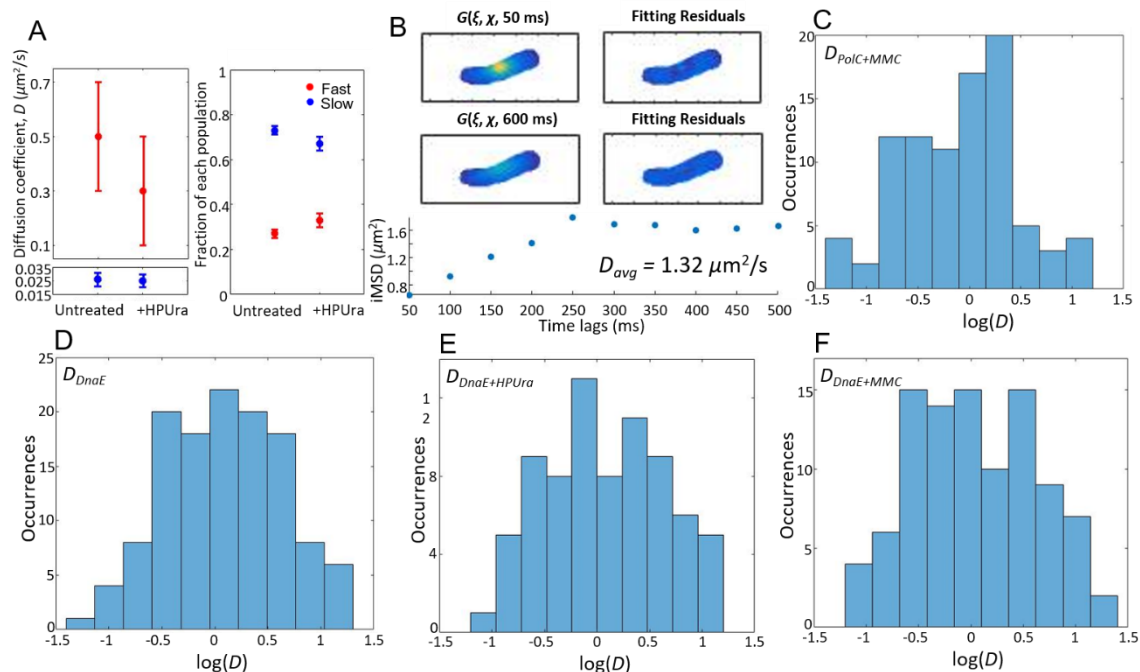
Figure 3.3 Quantification of the diffusion of PolC and DnaE in live *B. subtilis* cells. (A) Diffusion coefficients and fractional contributions from CPD analysis. Left: Diffusion coefficients, $D_{PolC\text{-}fast}$ and $D_{PolC\text{-}slow}$, for the 'fast' and 'slow' populations of PolC-PAmCherry in untreated and HPUra-treated cells, respectively calculated from the MSD curves in Figure 3.2. Error bars indicate 95% confidence interval. Right: Fraction of the 'fast' and 'slow' populations in the cells under both conditions. (B) The diffusion coefficient estimation of fast-moving molecules by STICS. The STICS correlation function is computed and then fit to a Gaussian function, *G*. The long-axis variances of the Gaussian fits to the correlation functions (*iMSD*) were plotted as a function of time lag, τ (dots), and this *iMSD* curve was fit to a model for square-confined diffusion (Eq. 6) to obtain a single average diffusion coefficient measurement for each cell. (C) Distribution of the average PolC diffusion coefficient measured in each of 88 different MMC-treated cells estimated by STICS. (D) – (F) Distributions of the average DnaE diffusion coefficients measured in (D) untreated, (E) HPUra-treated, and (F) MMC-treated cells estimated by STICS.
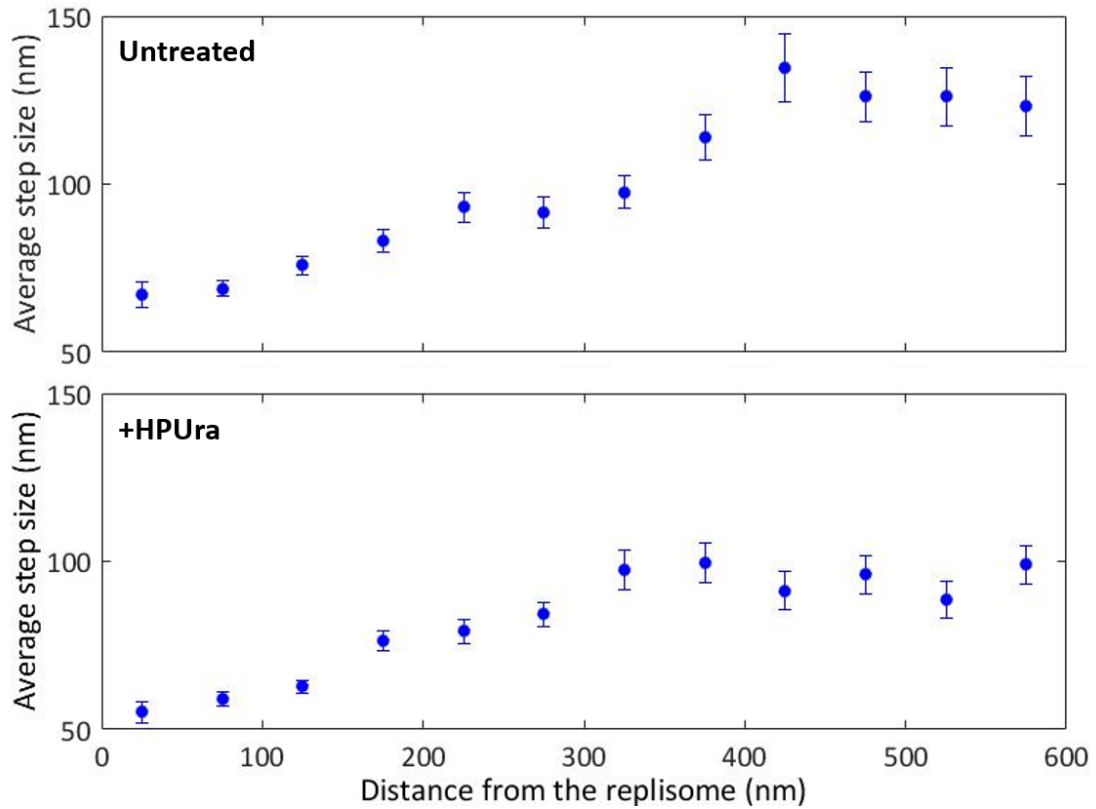
Figure 3.4 Measuring PolC dynamics as a function of subcellular positon. The mobility of PolC increases as a function of separation distance from the replisome. Here, the PolC/replisome separation distance is calculated between the beginning of the PolC step to the DnaX centroid in untreated (top) and HPUra-treated (bottom) cells. Error bars: standard error on the mean (s.e.).
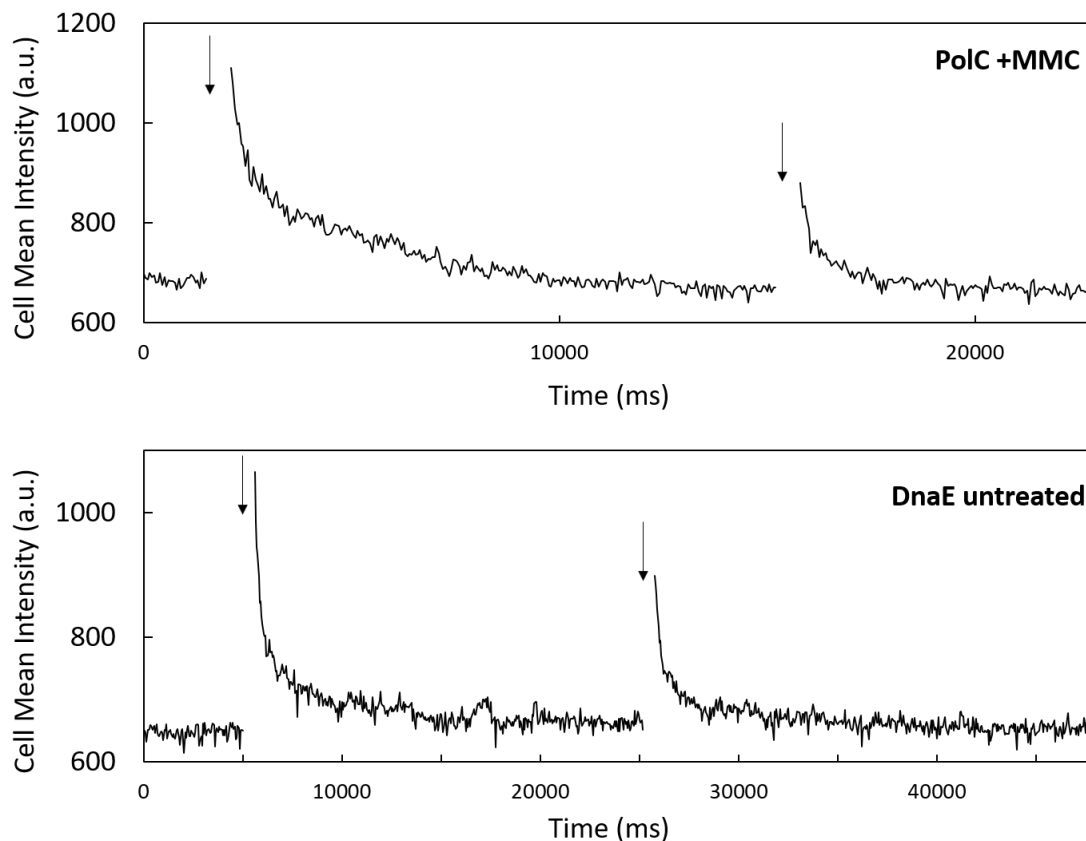
Figure 3.5 Evidence of PAmCherry photoactivation based on the time evolution of the single-cell intensity. Because PolC-PAmCherry or DnaE-PAmCherry diffusion was extremely rapid, very few PAmCherry single molecules were localized and no PAmCherry molecules were tracked in MMC-treated cells expressing PolC-PAmCherry or in cells expressing DnaE-PAmCherry (all conditions). Thus, we verified that PAmCherry photoactivation was in fact successful in these cells by measuring the mean cell intensity vs. time after 405-nm laser photoactivation (arrows). The increased intensity after each photoactivation in these representative data sets shows that PolC-PAmCherry (top) and DnaE-PAmCherry (bottom) were in fact successfully photoactivated. The cell intensity is calculated every 50 ms.

3.3 Dynamics of replisome proteins under replication arrest

To understand the mechanism that gives rise to these dynamics of motion and exchange at the *B. subtilis* replisome, we treated the cells with 6-hydroxy-phenylazo-uracil (HPUra), a drug that immediately arrests DNA synthesis by binding to the active site of PolC. Previously, we have shown that HPUra treatment arrests DNA replication

59

while still maintaining the replisome complex: DnaX-mCherry foci persist after HPUra treatment[15]. Here, we find that DNA replication arrest by HPUra does not arrest all replisome protein dynamics.

After HPUra treatment, we imaged PolC-PAmCherry, reconstructed PALM super-resolution images (Figure 3.1C), and tracked PolC-PAmCherry in living *B. subtilis* cells (Figure 3.1D). HPUra treatment did not significantly change the PolC dynamics: the motions are still separated into two distributions: fast and slow, and the CPD of the squared step sizes is fit by nearly identical diffusion coefficients and population weights (Figure 3.3A, Figure 3.4). Taking this unchanged motion in the context of our previous reports that PolC has a shorter dwell time (faster exchange rate) after HPUra treatment[22] implies that, after HPUra arrests replication, PolC can still bind to the replisome, but that PolC is more likely to dissociate because it cannot add new nucleotides to the DNA strain. Upon adding HPUra to arrest replication in *B. subtilis* cells expressing DnaE-PAmCherry, the DnaE was still un-trackable by single-molecule localization methods. We used STICS to measure a diffusion coefficient of $2.1 \pm 0.3 \ \mu m^2/s$ (Figure 3.3E). This essentially unchanged diffusion coefficient indicates free diffusion of DnaE in the cytoplasm, and indicates that replication arrest does not give rise to DnaE dwelling behavior.

Upon HPUra treatment of *B. subtilis* cells expressing DnaX-PAmCherry, two mobile DnaX populations were still observed. As in the untreated cells, the fast population remains too fast for the diffusion coefficient to be measured, while the slow population shows a dwelling behavior. However, after HPUra treatment, this dwelling

behavior is changed: some DnaX molecules dwell for a very long time ($\tau_{DnaX\text{-}dwell\text{-}1} > 8$ s) (Figure 3.6C); the full extent of these dwells is not quantifiable with our time-lapse imaging because the typical length of a movie is $2 - 4$ min and the longest practical time-lapse interval is 5 s. Sample drift or even cell growth could happen for longer (> 5 min) movie lengths, making the determination of dwelling events less certain. The average dwell time of the rest of the slow-moving DnaX, calculated using only tracks that dwells shorter than 8 seconds, is $\tau_{DnaX\text{-}dwell\text{-}2} = 1.61 \pm 0.10$ s (Figure 3.6D), which is shorter than the dwell time of DnaX in untreated cells and consistent with the increased PolC exchange rate after HPUra treatment: like PolC, most of the DnaX is more rapidly exchanged at the replisome upon replication arrest ($\tau_{PolC\text{-}dwell} = 0.97 \pm 0.10$ s[22]).

Unlike PolC however, the observation of two different DnaX dwell times indicates that DnaX has two different binding affinities in HPUra-treated cells: after HPUra causes replication arrest, DnaX might bind at two different binding sites to carry out two different functions. It is likely that the DnaX molecules with the 1.6 s dwell times are still binding to the original sites in the replisome and exchanging faster in the same manner as PolC. This accelerated exchange rate demonstrates that the DnaX and PolC exchange dynamics are coupled at the replisome. On the other hand, the longer DnaX dwell times are observed only after replication arrest and only for DnaX (not for either polymerase). We therefore hypothesize that these long dwells are related to DNA repair: upon replication arrest, some DnaX molecules are required to stabilize the replisome structure and fix the damage.

To further investigate the effect of stopping DNA synthesis, we treated the cells

with mitomycin C (MMC), a drug that also arrests DNA replication but via a different

mechanism. As opposed to HPUra, which arrests replication by binding to the PolC

active site, MMC arrests replication by inducing DNA crosslinks. It was previously

shown by bulk fluorescence imaging that MMC will not cause the disintegration of the

replisome[23]. Contrary to HPUra-induced replication arrest, MMC treatment leads to

PolC-PAmCherry molecules that are extremely fast-moving: the number of single PolC-

PAmCherry molecules that can be localized and fit is decreased (Figure 3.1E) and the

molecules are not trackable (Figure 3.1F). We therefore used STICS to quantify the

average diffusion coefficient of PolC, and measured to $D_{PolC,avg} = 1.4 \pm 0.2$ µm^2/s (Figure

3.3C). This large diffusion coefficient is far greater than even the fastest PolC-

PAmCherry diffusion coefficient in the untreated cells ($D_{PolC\text{-}fast} = 0.5 \pm 0.2$ µm^2/s),

which indicates that, upon MMC treatment, PolC-PAmCherry is mobile in the cytoplasm

and not strongly associated at the replisome. This effect of the MMC treatment shows

that MMC causes PolC to dissociate from the replisome. Moreover, in MMC-treated

cells, DnaE-PAmCherry still diffuses rapidly in the cytoplasm, with a diffusion

coefficient of $D_{DnaE,avg} = 2.0 \pm 0.2$ µm^2/s (Figure 3.3F). This diffusion coefficient is

bigger than $D_{PolC,avg}$ because PolC is bigger than DnaE in size. Therefore, neither of these

two DNA polymerases can associate with the replisome tightly after MMC induces DNA

cross-linking.

Finally, DnaX-PAmCherry in MMC-treated cells still exhibited three mobile

DnaX populations were still observed; one too fast to be tracked, $D_{DnaX\text{-}med} = 0.16 \pm 0.06$

µm^2/s, and $D_{DnaX\text{-}slow} = 0.025 \pm 0.006$ µm^2/s. The exchange dynamics of DnaX-

PAmCherry in MMC-treated cells were measured by single-molecule time-lapse tracking and found to be similar to the dynamics of DnaX-PAmCherry in HPUra-treated cells. As in the case of HPUra treatment, here we observed two different dwell time distributions for the slow-moving DnaX. The first one was too long to measure with our time-lapse imaging ($\tau_{DnaX-dwell-1} > 8$ s). The second dwell time was calculated by using only tracks that dwells shorter than 8 seconds ($\tau_{DnaX-dwell-2} = 3.12 \pm 0.78$ s) (Figure 3.6D), which is similar to the DnaX dwell time in untreated cells. This correspondence between dwell times before and after MMC treatment indicates that although MMC can cause PolC to dissociate from the replisome, the DnaX binding site in the replisome still remains, and the affinity of DnaX for this binding site is not changed by MMC treatment and the ensuing cross-linking.
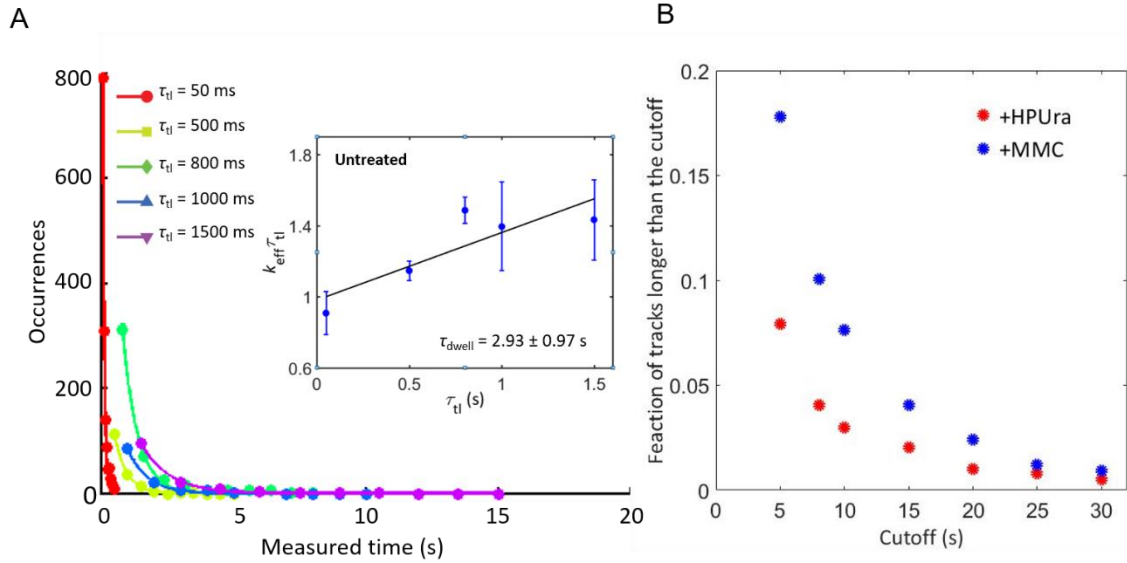
Figure 3.6 Characterization of the DnaX-PAmCherry dwelling behavior in live *B. subtilis* cells. (A) Dwell time distributions for DnaX-PAmCherry in untreated cells. The colors correspond to the time-lapse period (τ_{tl}), which is the sum of the integration time and the delay time. The distributions were fit to an exponential decay function (Eq. (4)) to calculate k_{eff}. (*Inset*) Linear fit (black line) of $k_{eff} \cdot \tau_{tl}$ versus τ_{tl}, from which the dissociation rate constant, k_{off}, the photobleaching rate constant k_b, and the dwell time constant, τ_{dwell}, are obtained (Eq. (5)). Errors bars are from four rounds of bootstrapping. (B) Fraction of long tracks in HPUra- and MMC-treated DnaX-PAmCherry cells. Each point shows the fraction of the tracks longer than the cutoff time when τ_{tl} = 1500 ms.

Table 3.1 Summary of measured diffusion coefficients (in $\mu m^2/s$) for PAmCherry-labeled proteins in untreated *B. subtilis* cells and in cells treated with HPUra or MMC. The fast and slow diffusion coefficients, D_{fast} and D_{slow}, respectively, are measured by single-molecule tracking and CPD analysis based on Eqs. (2) – (3). The average diffusion coefficients, D_{avg}, are measured by STICS based on Eq. (6). For cases where two populations could be measured by CPD analysis, the population fraction is indicated in parentheses. Error bars indicate 95% confidence interval (CPD analysis) or standard error on the mean (s.e.) (STICS analysis).

	Untreated	**+HPUra**	**+MMC**
PolC	$D_{PolC\text{-}fast} = 0.5 \pm 0.2$ (27%)	$D_{PolC\text{-}fast} = 0.3 \pm 0.2$ (33%)	$D_{PolC,avg} = 1.4 \pm 0.2$
	$D_{PolC\text{-}slow} = 0.026 \pm 0.005$ (73%)	$D_{PolC\text{-}slow} = 0.025 \pm 0.005$ (67%)	
DnaE	$D_{DnaE,avg} = 1.9 \pm 0.2$	$D_{DnaE,avg} = 2.1 \pm 0.3$	$D_{DnaE,avg} = 2.0 \pm 0.2$
	$D_{DnaX\text{-}fast} > 1$	$D_{DnaX\text{-}fast} > 1$	$D_{DnaX\text{-}fast} > 1$
DnaX	$D_{DnaX\text{-}med} = 0.2 \pm 0.1$ (33%)	$D_{DnaX\text{-}med} = 0.17 \pm 0.02$ (45%)	$D_{DnaX\text{-}med} = 0.16 \pm 0.06$ (36%)
	$D_{DnaX\text{-}slow} = 0.026 \pm 0.007$ (67%)	$D_{DnaX\text{-}slow} = 0.031 \pm 0.007$ (55%)	$D_{DnaX\text{-}slow} = 0.025 \pm 0.006$ (64%)

Table 3.2 Summary of measured dwell times for PAmCherry-labeled proteins in untreated *B. subtilis* cells and in cells treated with HPUra or MMC. Error bars indicate 95% confidence interval.

	Untreated	**+HPUra**	**+MMC**
PolC	$^{*}\tau_{PolC\text{-}dwell} = 2.79 \pm 0.41$ s	$^{*}\tau_{PolC\text{-}dwell} = 0.97 \pm 0.10$ s	*N/A*
DnaX	$\tau_{DnaX\text{-}dwell} = 2.63 \pm 0.97$ s	$\tau_{DnaX\text{-}dwell\text{-}1} > 8$ s	$\tau_{DnaX\text{-}dwell\text{-}1} > 8$ s
		$\tau_{DnaX\text{-}dwell\text{-}2} = 1.61 \pm 0.10$ s	$\tau_{DnaX\text{-}dwell\text{-}2} = 3.12 \pm 0.78$ s

*From ref. [22].

3.4 Conclusions

Here, we have paired single-molecule super-resolution microscopy with complementary data analysis techniques to characterize the dynamics of three replisomal proteins, including the two essential DNA polymerases, in live *B. subtilis* cells. Our results show that all three replisomal proteins undergo dynamic exchange: they are recruited to and released from the replisome. The similar kinetics of PolC and DnaX during normal replication imply a possible coupling between these two proteins in the DNA replication process. On the other hand, DnaE undergoes very fast exchange, indicating that a single DnaE protein only extends the lagging DNA strand for only a few ms after the RNA primer. When HPUra, a drug that only binds to the active site of PolC, arrests DNA replication, PolC and a portion of DnaX both have a faster exchange rate. Concurrently, another population of DnaX molecules has a very long dwell time, meaning that a new binding site with stronger affinity to the replisome is available for DnaX. When DNA replication is arrested by the crosslinking introduced with MMC, all PolC molecules dissociate from the replisome, but DnaX can still bind to the replisome, possibly to hold the core structure of the replisome and help to fix the damage.

3.5 Materials and methods

3.5.1 Sample preparation and single-molecule imaging

Bacillus subtilis cells were grown at 30 °C in S7$_{50}$ minimal medium with 1% arabinose starting from OD$_{600}$ ~0.1. To minimize the background fluorescence and the fluorescent impurities in *B. subtilis*, minimal medium was filtered with a 0.22 μm syringe filter ON

66

the day of imaging. For PolC-PAmCherry experiments, 0.125% xylose was added to the medium to induce the ectopically expressed DnaX-mCitrine. Cells were harvested upon reaching the exponential phase (OD_{600} ~0.55). When used, HPUra was added to the culture immediately before imaging to a final concentration of 162 μM, or MMC was added to the culture ~40 min before imaging to a final concentration of 200 ng/mL. A 2% agarose in $S7_{50}$ pad was sandwiched between two coverslips, which had been cleaned in an oxygen plasma (Plasma Etch PE50) for 20 minutes. 1 – 2 μL of cell culture was pipetted onto the agarose pad and then used for imaging.

For single-molecule imaging, a wide-field inverted microscope was used. The fluorescence emission was collected by a 1.40-N.A. 100× oil-immersion phase-contrast objective and detected on a 512×512-pixel electron-multiplying charge-coupled device detector camera (Photometrics Evolve, Princeton Instruments, Acton, MA) at frame rates of 40 ms/frame for the PolC imaging and 50 ms/frame for the DnaE and DnaX imaging. PAmCherry molecules were activated by a 200-ms exposure to a 405-nm laser with a power density of ~100 W/cm^2 (Coherent 405-100) and imaged with a 561-nm laser with a power density of ~200 W/cm^2 (Coherent Sapphire 561-50). For two-color experiments, DnaX-mCitrine was imaged before PolC-PAmCherry with a 488-nm laser with a power density of ~7 W/cm^2 (Coherent Sapphire 488-50).

3.5.2 Super-resolution imaging and single-molecule tracking

The precise localizations of single molecules were determined by fitting the point spread function (PSF) of each single molecule to a 2D Gaussian function using home-built MATLAB code[18]. We obtained the PALM super-resolution reconstruction images by

plotting the localization of each fit convolved with a Gaussian blur with width equal to its localization uncertainty.

Single-molecule tracking was performed by maximizing the likelihood that the two subsequent localizations are the same molecule at different time points[24]. A merit value, m, was assigned for every possible pairing:

$$m = e^{-\alpha\Delta d}e^{-\beta\Delta I}e^{-\gamma\Delta t} \tag{1}$$

where Δd, ΔI, and Δt correspond respectively to the spatial separation, the intensity difference, and the temporal separation between the two molecules, and α, β, and γ are coefficients that specify how much penalty is given to each factor.

The sum of the likelihood was maximized for each set of consecutive frames. The maximization was repeated until all frames were processed. Only tracks longer than 5 frames were saved for future analysis.

References

1. Sanders GM, Dallmann HG, McHenry CS. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol Cell*. 2010;37(2):273-281.

2. Reyes-Lamothe R, Possoz C, Danilova O, Sherratt DJ. Independent positioning and action of escherichia coli replisomes in live cells. *Cell*. 2008;133(1):90-102.

3. Reyes-Lamothe R, Sherratt DJ, Leake MC. Stoichiometry and architecture of active DNA replication machinery in escherichia coli. *Science*. 2010;328(5977):498-501.

4. O'Donnell M. Replisome architecture and dynamics in escherichia coli. *J Biol Chem*. 2006;281(16):10653-10656.

5. Lemon KP, Grossman AD. Localization of bacterial DNA polymerase: Evidence for a factory model of replication. *Science*. 1998;282(5393):1516-1519.

6. Meile J, Wu L, Ehrlich S, 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*. 2006;6(7):2135-2146.

7. Liao Y, Schroeder JW, Gao B, Simmons LA, Biteen JS. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proc Natl Acad Sci U S A*. 2015;112(50):E6898-E6906.

8. Lemon K, Grossman A. Movement of replicating DNA through a stationary replisome. *Mol Cell*. 2000;6(6):1321-1330.

9. Sawitzke J, Austin S. An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol Microbiol*. 2001;40(4):786-794.

10. Dervyn E, Suski C, Daniel R, et al. Two essential DNA polymerases at the bacterial replication fork. *Science*. 2001;294(5547):1716-1719.

11. Kunkel TA, Burgers PM. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol*. 2008;18(11):521-527.

12. Kunkel TA. Balancing eukaryotic replication asymmetry with replication fidelity. *Curr Opin Chem Biol*. 2011;15(5):620-626.

13. Seco EM, Ayora S. Bacillus subtilis DNA polymerases, PolC and DnaE, are required for both leading and lagging strand synthesis in SPP1 origin-dependent DNA replication. *Nucleic Acids Res*. 2017;45(14):8302-8313.

14. Paschalis V, Le Chatelier E, Green M, Kepes F, Soultanas P, Janniere L. Interactions of the bacillus subtilis DnaE polymerase with replisomal proteins modulate its activity and fidelity. *Open Biology*. 2017;7(9):170146.

15. Liao Y, Schroeder JW, Gao B, Simmons LA, Biteen JS. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proc Natl Acad Sci U S A*. 2015;112(50):E6898-E6906.

16. Lewis JS, Spenkelink LM, Jergic S, et al. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *eLife*. 2017;6:e23932.

17. Beattie TR, Kapadia N, Nicolas E, et al. Frequent exchange of the DNA polymerase during bacterial chromosome replication. *eLife*. 2017;6:10.7554.

18. Liao Y, Li Y, Schroeder JW, Simmons LA, Biteen JS. Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells. *Biophys J*. 2016;111(12):2562-2569.

19. Brown NC. 6-(p-hydroxyphenylazo)-uracil: A selective inhibitor of host DNA replication in phage-infected bacillus subtilis. *Proc Natl Acad Sci U S A*. 1970;67(3):1454-1461.

20. Brown NC, Wisseman C, Matsushita T. Inhibition of bacterial DNA replication by 6-(p-hydroxyphenylazo)-uracil. *Nature New Biol*. 1972;237(72):72-74.

21. Iyer VN, Szybalski W. A molecular mechanism of mitomycin action: Linking of complementary DNA strands. *Proc Natl Acad Sci U S A*. 1963;50(2):355-362.

22. Liao Y, Li Y, Schroeder JW, Simmons LA, Biteen JS. Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells. *Biophys J*. 2016;111(12):2562-2569.

23. Simmons LA, Grossman AD, Walker GC. Replication is required for the RecA localization response to DNA damage in bacillus subtilis. *Proc Natl Acad Sci U S A*. 2007;104(4):1360-1365.

24. Lakadamyali M, Rust MJ, Babcock HP, Zhuang X. Visualizing infection of individual influenza viruses. *Proc Natl Acad Sci U S A*. 2003;100(16):9280-9285.

Chapter 4: Determining the stoichiometry of replisome components with a Bayesian mixture model

4.1 Introduction

The results in Chapters 2 and 3 indicate that super-resolution microscopy can achieve great precision and dynamic information, but the stoichiometry of different proteins (the copy numbers of each different types of proteins) in the architecture of the replisome *in vivo* is still unknown and of great interest. The most common and straightforward way to measure the protein number at the replisome is to label the protein of interest with a fluorescent protein or dye, record the intensity of the focus over time (Figure 4.1), and then determine the number of bleaching steps in the time trace. This photobleaching step counting has been used to quantify the stoichiometry of a number of complexes[1-5]. The difficulty is that the photobleaching events are often buried under the high noise of these intensity time traces (c.f. frames 50 – 300 in Figure 4.1). To overcome this limitation, various methods have been developed to count the photobleaching steps of this time series data[6-15]. However, many contemporary methods for photobleaching event counting suffer from problems of model selection bias[9]. Model comparison via maximizing likelihood or penalized maximum likelihood, such as Akaike information criterion (AIC) or Bayesian information criterion (BIC), have been applied to help with the model

selection[8,10-13,15]. Although such methods can find the model that best explains the data while remaining the least complex, we are unable to evaluate rigorously the confidence in models relative to each other. Such methods also do not always yield physically meaningful results from experimental data. Furthermore, subjective thresholds or parameters are often used, making these algorithms sensitive to parameter changes. In this Chapter, I introduce a method in which the model selection is purely based on the properties of the fluorescent molecules rather than on any subjective numbers.
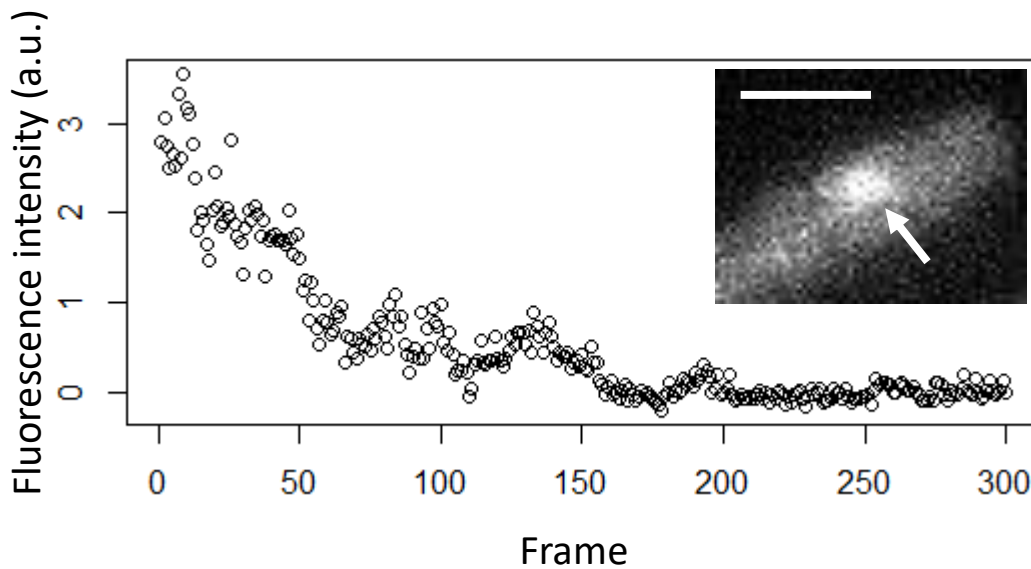
Figure 4.1 Representative background-subtracted fluorescence intensity time trace showing molecules bleaching at a focus. *Inset*: An example of a DnaX-mCherry focus (arrow) in a *B. subtilis* cell. Scale bar: 1 μm.

4.2 Model

The noise of the intensity time traces mainly comes from the shot noise of the fluorescent molecules themselves, which is associated with the particle nature of light. Shot noise is

72

also called Poisson noise, and it can be modeled by a Poisson process: if we consider the intensity observations (X_t) from different states $(k = 1, 2, \ldots, K)$ as random samples drawn from different Gaussians distributions, the means and the variances of the Gaussian distributions should be linear. $k = 1$ corresponds to the initial state where no molecules are photobleached yet, $k = 2$ corresponds to the state after the first photobleaching event(s) happens, and so on. We can then perform a cluster analysis of our data, and the final goal is to choose the best model and to determine K, the number of the photobleaching steps in the intensity time trace.

A schematic figure of the model when $K = 3$ is shown in Figure 4.2. The Z_t are latent variables that are related through a Markov process, which means that the probability of each photobleaching step only depends on the state attained in the previous step. In context of the photobleaching experiments, $k = 1, 2, 3$ each correspond to a different state with a different number of "on" (fluorescing) molecules in the focus. In state $Z_t = k$, the X_t are the observed random variables, which are the fluorescence intensities that follows a normal distribution with mean μ_k and variance σ_k^2. The transition between each different state can be described by the transition matrix Π:

$$\Pi = \begin{pmatrix} \pi_{11} & \cdots & \pi_{1k} \\ \vdots & \ddots & \vdots \\ \pi_{k1} & \cdots & \pi_{kk} \end{pmatrix}$$

where $\pi_{ij} = p(Z_{t+1} = i | Z_t = j)$.

The model is fit in a Bayesian framework (which derives the posterior probability as a consequence of a prior probability and a likelihood function derived from a statistical

model for the observed data), where (μ_k, σ_k^2, Π) are random variables with prior distributions described in the full model below:

$$Z_{t+1}|Z_t = k \sim Categorical\ (\pi_{k.})$$

$$X_t|Z_t = k \sim N(\mu_k,\ \sigma_k^2),\ k = 1,2,\dots,K$$

Priors:

$$\pi_{k.} \sim Dirichlet\ (a)$$

$$\mu_k \sim Uniform\ (min(X),\ max(X))$$

$$\sigma_k^{-2} \sim Gamma\ (\alpha,\ \beta)$$

In practice, $a = 1, \alpha = \beta = 0.001$. A high informative prior for μ_k is used because we know that the means of the Gaussian clusters should be within the range of the intensity data.

For $k = K$ specifically (the final state when all molecules are photobleached), the model is just a simple normal model with its conjugate prior:

$$X_t|Z_t = K \sim N(\mu_K,\ \sigma_K^2)$$

Priors:

$$\mu_K \sim N(\mu_0,\ \sigma_0^2)$$

$$\sigma_K^{-2} \sim Gamma\ (\alpha,\ \beta)$$

In practice, $\mu_0 = 0, \sigma_0^2 = 10000, \alpha = \beta = 0.001$.
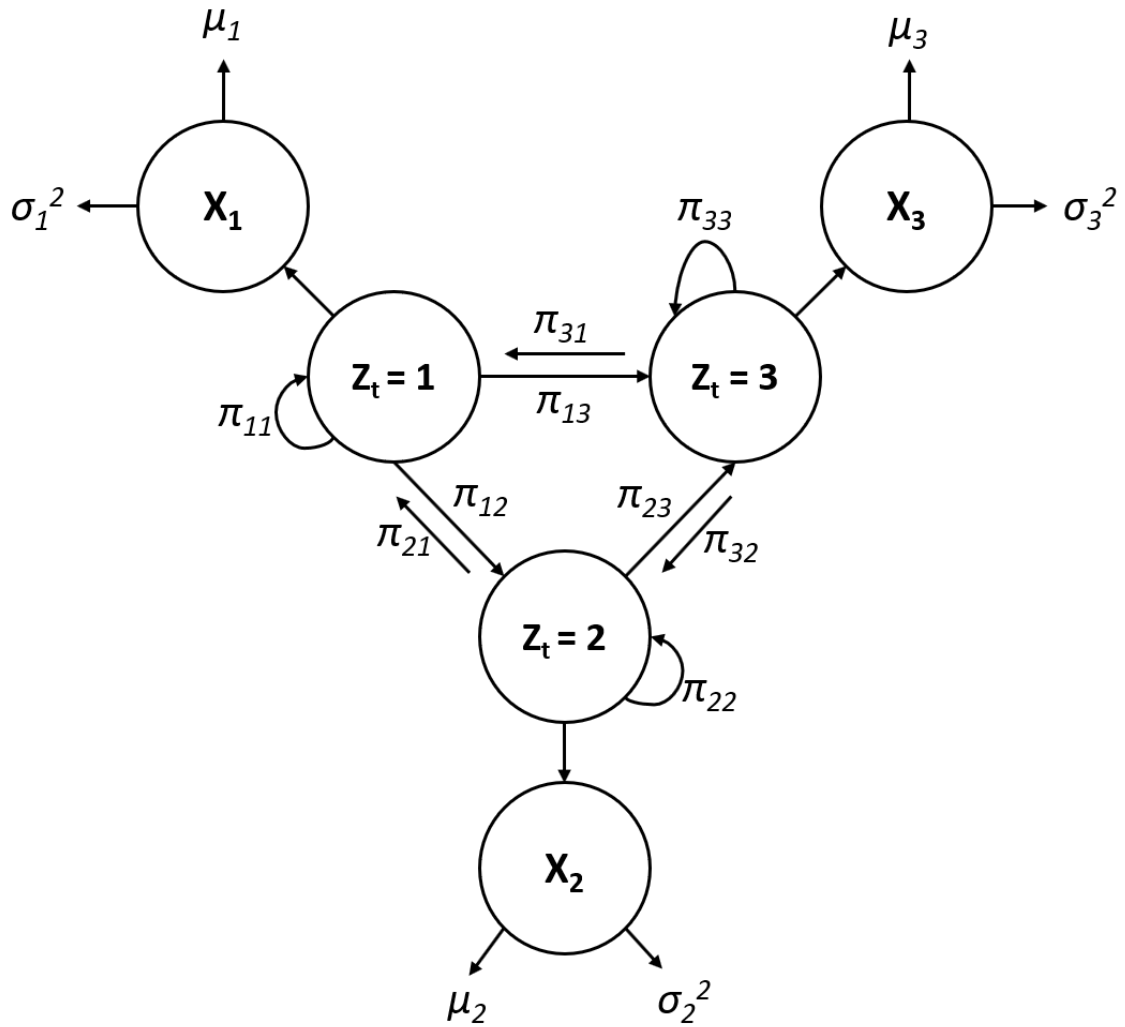
Figure 4.2 A schematic figure showing the model when $K = 3$. In the Bayesian framework, $(\mu_k,\ \sigma_k^2, \Pi)$ are treated as random variables with prior distributions, which are not shown here.

Since it is easy to obtain the data for the final step (i.e., the fluorescence intensity after photobleaching) experimentally, we use the final step data to obtain μ_K and σ_K^2 first, and then include μ_K and σ_K^2 as fixed values into the full model. All posterior distributions can be simulated with Markov chain Monte Carlo (MCMC) methods using JAGS (Just Another Gibbs Sampler)[16].

Since the variances, σ^2, should be proportional to the means, μ, we should be able to fit (μ_k, σ_k^2) to a straight line. The goal is to determine the best K. The best model should (1) fit the data very well, and (2) produce a set (μ_k, σ_k^2) that is linear. For requirement (1), the deviance information criteria (DIC) is used to evaluate the goodness-of-fit, and for requirement (2), the adjusted R-square is used.

The deviance information criteria (DIC) is a popular Bayesian analog of AIC and BIC[17]. It is defined as:

$$DIC = D + pD$$

where $D = E(-2 \log(p(Y|\theta))$ is the posterior mean of the deviance, and pD is the effective number of parameters.

The lower the DIC, the better the fit. However, DIC has been criticized to tend to select over-fit models, because the observed data is used both in constructing the posterior distribution and in the evaluation of the models. I therefore use the linear relationship of (μ_k, σ_k^2) to overcome this tendency. Overall, the best model should have a low DIC and a high adjusted R-square. The DIC and the R-square for each different k are scaled between 0 and 1, and the model is selected based on a self-defined score:

$$Score = \text{scaled adjusted } R \text{ square} - \text{scaled DIC}$$

The model with the highest score is chosen.

In an approach that explicitly considers the possibility that multiple photobleaching events can occur within a single imaging frame, the stoichiometry

(number of molecules in a focus) is determined by dividing the intensity of the detected state corresponding to the highest copy number of molecules by the intensity change corresponding to the smallest jump between two states[18].

4.3 Data preparation

A representative imaging frame showing a fluorescent focus in a cell is show in the inset of Figure 4.1. For each such frame of each movie, the center pixel of each focus was determined from the filtered image, which was used to differentiate the fluorescence molecules from the noise. In each raw data image, the average intensity of all pixels within 3 pixels of this center pixel (i.e, a 7×7 pixel box) was measured. In live-cell imaging, the background fluorescence intensity mainly came from the cell autofluorescence, which was not constant but rather photobleached over time. Though more sophisticated background-subtraction algorithms can be incorporated with our approach, here the approximate background intensity for each cell in each imaging frame is determined from the average intensity value of all other pixels within the cell not belonging to any focus, as in the determination of the stoichiometry of PolC in Chapter 2. In this way, the average background intensity of each pixel in the cell was subtracted from the intensity of each focus in that same cell. The resulting background-subtracted focus intensity was recorded over time to monitor the photobleaching process for that mCherry focus and to generate intensity time traces as in Figure 4.1.

4.4 Data analysis and results

4.4.1 Simulation results

The Bayesian algorithm was first tested on simulated data. An example of the six-state simulated time series is shown in Figure 4.3. Here the true means are $\mu = (1, 2, 3, 4, 6, 8)$, and the true variances are $\sigma^2 = (0.005, 0.055, 0.105, 0.155, 0.255, 0.355)$. It is obvious that the transition between different states is one-directional (the fluorescence intensity can only transit to a state with a smaller μ). The intensity decreases from 8 to 6 and 6 to 4 mimic two instances in which a pair of molecules bleaches at the same time, which often happens at the beginning of the movie. The variances imitate the variances in the experimental intensity time series collected in the following section. The true value of K is 6 (corresponds to 6 photobleaching steps), and the true stoichiometry is 7 (6 photobleaching steps mean 5 photobleaching events, and there are two events where two molecules photobleach at the same time, so overall there are 8 molecules in total).
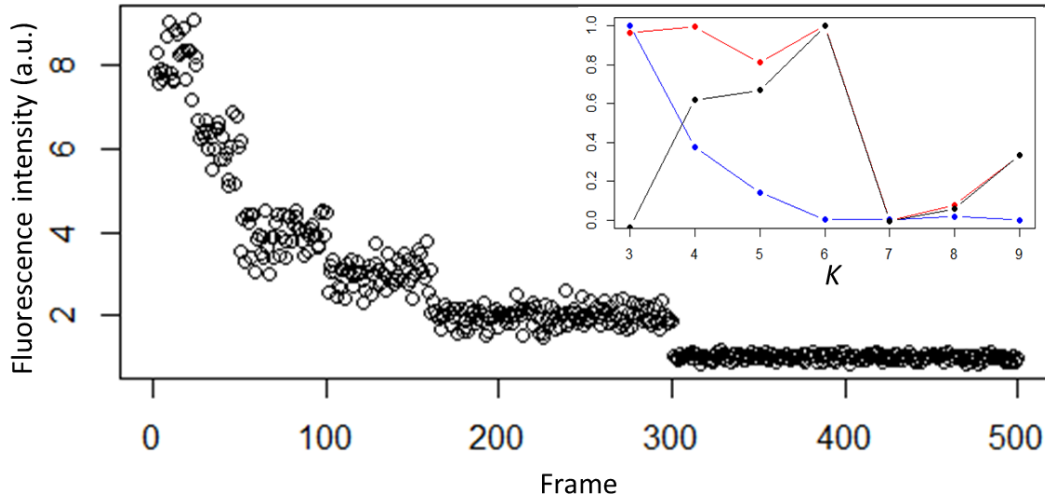
Figure 4.3 Simulated fluorescence intensity time trace from a six-state process. *Inset*: Plot of scaled adjusted R-square (red), scaled DIC (blue), and score (equals to scaled adjusted R-square - scaled DIC, black) at different values of K. The true $K = 6$. When K is lower than the true value, we get good adjusted R-square (red) but bad DIC (blue), meaning that the model did not fit the data well. On the other hand, when K is higher than the true value, we get good (low) DIC, but poor adjusted R-square, indicating over-fitting.

The last 80 points were first used to estimate the base mean and variance. Starting from $K = 3$, the data was then fit to the Bayesian model proposed above with 9000 iterations and 1000 burn-in. Three Markov chains were generated to fit the data, and the chain with the lowest DIC was selected. The same procedure was then performed for $K = 4, 5, ... , 9$. The scaled adjusted R-squares, scaled DICs, and the scores for each K are plotted in Figure 4.3 (inset). The adjusted R-squares are high when K is lower than the true value, which is anticipated since the algorithm will combine two clusters as one. When K is higher than the real value, the best model fits have poor adjusted R-squares, because the algorithm will tend to separate the cluster with higher mean to two clusters, but such separation will make the variance much smaller than the real value, thus destroyed the linear relationship. On the other hand, the DICs decrease sharply until $K =$

79

6, meaning that the goodness-of-fits are poor when K is lower than the real value. The best K should have both high adjusted R-square and low DIC, which is easily identified at $K = 6$.

The MCMC trace plot of μ and the plot of μ vs. σ^2 when $K = 6$ are shown in Figures 4.4 and 4.5. The μ is very well estimated and the linear correlation between μ vs. σ^2 is obvious when K is true.
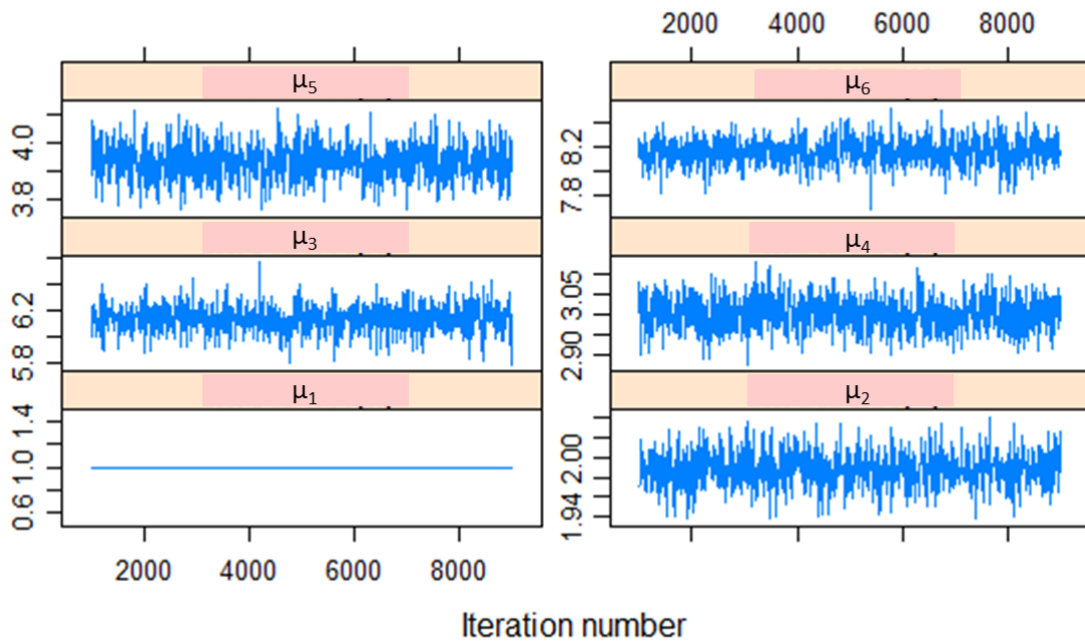
Figure 4.4 MCMC trace plot of μ when $K = 6$ (the true value). The sampled μ values agree with the true values in the simulation. μ_1 was already estimated in the previous step, so it is not a random variable here.
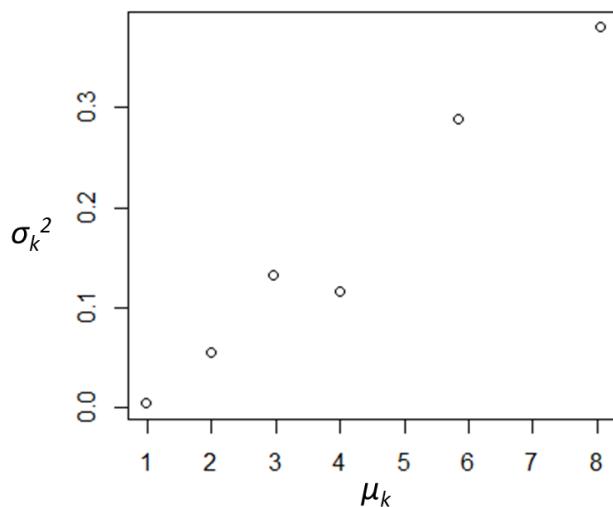
Figure 4.5 Plot of the fit μ vs. the fit σ^2 when $K = 6$ (the true value). The fit μ and σ^2 have a linear relationship, as set in the simulation.

For 50 simulated time traces ($K = 6$ and the stoichiometry is 7), the algorithm successfully estimated the stoichiometry in 48 cases. It should be noted that in some cases, the stoichiometry is still correct although the K is wrong (e.g., we found $K = 5$), because as long as we correctly estimate the minimum intensity difference and the maximum μ, the stoichiometry (which is given by the ratio of maximum μ/minimum intensity difference) will be true.

4.4.2 Test on experimental data

The algorithm was then tested on real experimental data to determine the stoichiometry of the *Bacillus subtilis* β-clamp loader component DnaX at the replisome. The intensity time trace of each replisomal focus in live cells was retrieved and background-subtracted. Figure 4.6 shows an example of a background-subtracted time trace of a DnaX-mCherry focus in live cells under constant illumination and undergoing

photobleaching. We can see the photobleaching events, but it is difficult to know the

number of bleaching steps due to the low signal to noise ratio. Using the method

introduced above, the best model selected is when $K = 5$, and the stoichiometry

determined is that, in these cells that express DnaX-mCherry at the native DnaX

promoter as the sole source of DnaX, there are 8 DnaX copies at each *B. subtilis*

replisome.
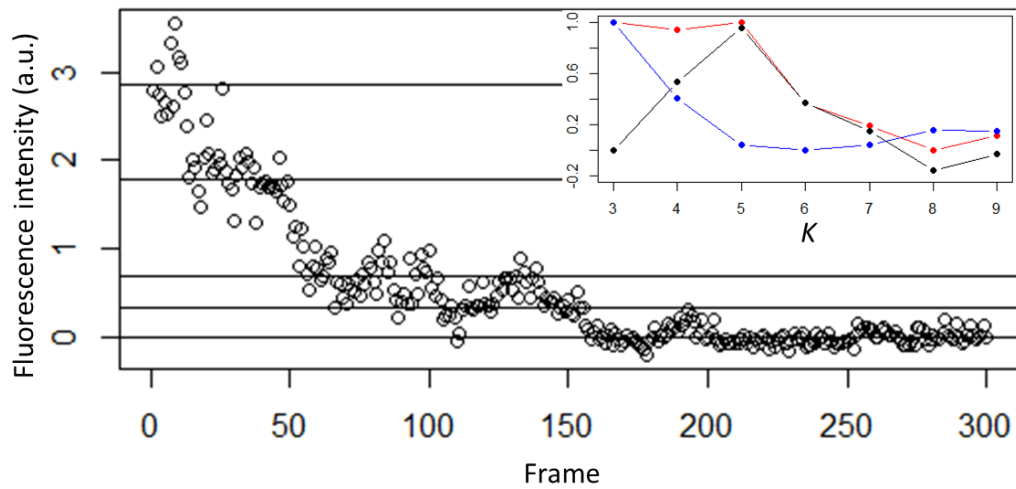
Figure 4.6 A representative background-subtracted fluorescence intensity time trace at a DnaX-mCherry focus undergoing photobleaching. Horizontal lines show the μ_k from the best model selected. The stoichiometry is determined to be 8. *Inset*: Plot of scaled adjusted R-square (red), scaled DIC (blue), and score (black) at different values of K from the real data estimation above. The best model selected is when $K = 5$.

A total of 34 time traces were analyzed, and we determined a stoichiometry of $7 \pm$

1 (standard error on the mean). More time traces need to be analyzed to plot the

distribution and a more accurate stoichiometry determination.

4.5 Discussion and perspectives

In this chapter, I developed and introduced a new method of counting the photobleaching steps using a Bayesian mixture model. This algorithm does not depend on subjective thresholds or parameters, which were often used in other methods. The prior distributions in the model are either flat or based on common knowledge, and the model selection is based on the property of the fluorescent molecules, keeping the method as objective as possible. This approach is widely applicable to many important problems in biophysics and cell biology.

The stoichiometry is determined not simply by the bleaching step count, to account for event(s) where more than one molecule photobleaches at the same time. Therefore, even if K is mistakenly predicted, the true stoichiometry can still be obtained if the minimum intensity difference and the mean of the first step (μ_1) are correct. However, this is a double-edged sword, because in this case, a small change in the minimum intensity difference might cause a drastic change in the final stoichiometry. This sensitivity makes algorithm heavily reliant on the correct estimation of the minimum intensity difference. To improve this caveat, fluorophores with better photostability can be used in the photobleaching experiment.

The major problem of counting the photobleaching steps in bacteria cells is the background autofluorescence intensity from the cells themselves, which can make the intensity readout very inaccurate. The background-subtraction procedure used in this chapter has some obvious problems. First, the average intensity of all other pixels in the cell not belonging to any focus is not an accurate measurement of the autofluorescence

background specifically at the location of the focus. For instance, the distribution of fluorescent molecules, even outside of the foci, is not homogenous. For example, it is known that when a molecule that can exchange at the replisome is near the replisome focus, it will tend to slow down (as shown in Chapter 3). This heterogeneity will lead to an underestimation of the background-subtracted intensity of the focus. Second, since the background fluorescence intensity will also decrease over time, the variance of the background intensity will decrease as well. By simply subtracting the average intensity from the intensity of the foci, each state will have a higher variance than the real value.

One way to avoid these problems is to use SMALL-LABS (Single-Molecule Accurate Localization by Local Background Subtraction)[19], which is a background-subtraction method that accurately locates and measures the intensity of single molecules, regardless of the shape or brightness of the background, based on temporal fluctuations in the signal. We have previously proven that SMALL-LABS works very well in biological imaging with high autofluorescence background (Figure 4.7). In general, accurate estimation of the fluorescence intensities will greatly improve the performance of the algorithm.
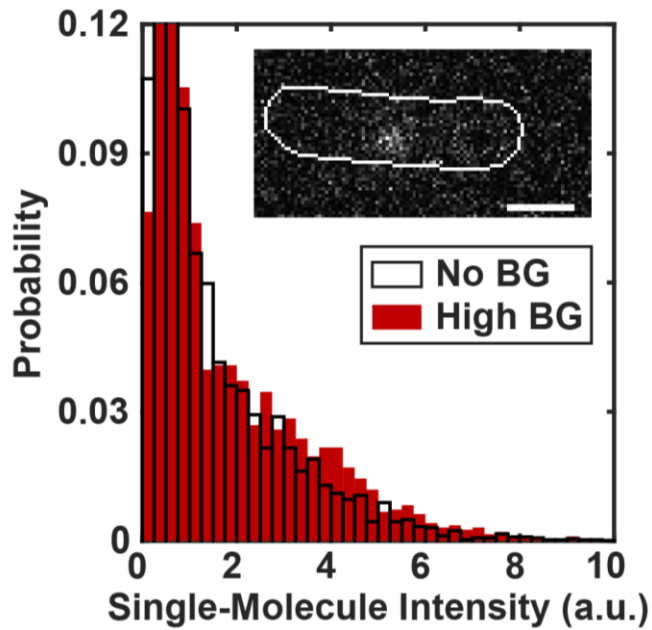
Figure 4.7 The distribution of fluorescent protein brightness in the high-background conditions (High BG; red) accurately reproduces the distribution measured in higher signal-to-noise experiments (Low BG; white). Scale bar: 1 μm.

Since the stoichiometry does not include any information of dynamics, fixed cells can be used in the experiment. This will simplify the model: unlike in living cells where fluorescence can increase due to dynamic exchange of proteins at the replisome, the Markov chain of different states in fixed cells will be strictly one-directional: once a fluorescent molecule photobleaches, there is no mechanism for the fluorescence intensity to increase again. By performing the experiment in fixed cells, using chemical dyes as the fluorophore is also an option if the specificity of the labelling is ensured. Thus, brighter fluorophores with better photostability can be used, so that better intensity time traces can be retrieved.

References

1. Delalez NJ, Wadhams GH, Rosser G, et al. Signal-dependent turnover of the bacterial flagellar switch protein FliM. *Proc Natl Acad Sci U S A*. 2010;107(25):11347-11351.

2. Engel BD, Ludington WB, Marshall WF. Intraflagellar transport particle size scales inversely with flagellar length: Revisiting the balance-point length control model. *J Cell Biol*. 2009;187(1):81-89.

3. Lawrimore J, Bloom KS, Salmon ED. Point centromeres contain more than a single centromere-specific Cse4 (CENP-A) nucleosome. *J Cell Biol*. 2011;195(4):573-582.

4. Coffman VC, Wu P, Parthun MR, Wu J. CENP-A exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast. *J Cell Biol*. 2011;195(4):563-572.

5. Mangiameli SM, Merrikh CN, Wiggins PA, Merrikh H. Transcription leads to pervasive replisome instability in bacteria. *eLife*. 2017;6:e19848.

6. Manzo C, Finzi L. Quantitative analysis of dna-looping kinetics from tethered particle motion experiments. *Methods in Enzymology, Vol 475: Single Molecule Tools, Pt B: Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*. 2010;475:199-220.

7. Watkins LP, Yang H. Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B*. 2005;109:617-628.

8. Kalafut B, Visscher K. An objective, model-independent method for detection of non-uniform steps in noisy signals. *Comput Phys Commun*. 2008;179(10):716-723.

9. Hines KE, Bankston JR, Aldrich RW. Analyzing single-molecule time series via nonparametric bayesian inference. *Biophys J*. 2015;108(3):540-556.

10. Csanady L. Statistical evaluation of ion-channel gating models based on distributions of log-likelihood ratios. *Biophys J*. 2006;90(10):3523-3545.

11. Wagner M, Timmer J. Model selection in non-nested hidden markov models for ion channel gating. *J Theor Biol*. 2001;208(4):439-450.

12. Ball FG, Sansom MSP. Ion-channel gating mechanisms:Model identification and parameter-estimation from single channel recordings. *Proc R Soc Lond B Biol Sci*. 1989;236(1285):385-416.

13. Horn R. Statistical methods for model discrimination. applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys J*. 1987;51(2):255-263.

14. Tsekouras K, Custer TC, Jashnsaz H, Walter NG, Presse S. A novel method to accurately locate and count large numbers of steps by photobleaching. *Mol Biol Cell*. 2016;27(22):3601-3615.

15. Wiggins PA. An information-based approach to change-point analysis with applications to biophysics and cell biology. *Biophys J*. 2015;109(2):346-354.

16. Http://Mcmc-jags.sourceforge.net/. .

17. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis.* CRC Press; 2004.

18. Liao Y, Li Y, Schroeder JW, Simmons LA, Biteen JS. Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells. *Biophys J*. 2016;111(12):2562-2569.

19. Isaacoff B, Li Y, Lee S, Biteen J. SMALL-LABS: An algorithm for localizing and measuring single molecules in the presence of obscuring backgrounds. submitted. .

Chapter 5: Conclusions and perspectives

In this thesis, I applied super-resolution microscopy to investigate the stoichiometry and the dynamics of the replisomal proteins in live *B. subtilis* cells. Various data analysis models and algorithms were used to characterize the diffusion, kinetics, and the stoichiometry under different conditions. The ability to visualize and quantify the motion of single protein molecules *in vivo* allowed us to uncover the architecture and the dynamics of the replisome in *B. subtilis*, which is inaccessible by conventional optical microscopy and poorly understood before.

By analyzing single-molecule trajectories and applying STICS analysis to our data, I have shown that the replisomal proteins undergo dynamic exchanges in the DNA replication process[1]. While the factory model[2]—in which the replisomes are relatively stationary and the DNA strands are pulled through the replisome—is true in *B. subtilis*, as supported by the bulk fluorescence imaging data, the replisomal proteins including the two polymerases are constantly recruited to and released from the replisomes in live cells. Heterogeneous dynamics were characterized using the cumulative probability distribution (CPD) model introduced in Chapter 1, revealing three different modes of motion: 1) the molecule binds to the replisome and has a diffusion coefficient, D, around 0.02 μm^2/s, for example, the slow population of PolC in untreated and HPUra-treated cells; 2) the

molecule searches along the DNA strands and has $D \sim 0.5 \; \mu m^2/s$, like the fast population of PolC in untreated and HPUra treated cells; 3) the molecule diffuses freely in the cytoplasm and has $D > 1 \; \mu m^2/s$, which cannot be quantified using single-particle tracking methods. The different motions of the replisomal proteins suggest that DNA replication is a very dynamic process, and although only a few molecules are needed at a time at the replisome, a large number of proteins in the cell is necessary to ensure the efficiency of replication due to the fast exchange rate of the replisomal proteins.

Besides the diffusion coefficients, dwell time analysis shed light on the kinetics of the interaction between the replisomal proteins and the replisome. Both PolC and DnaE have a dwell time of about 2.7 s, which is the time needed for the synthesis of an Okazaki fragment[3]. The similar dynamics also indicates a possible coupling between these two proteins. On the other hand, the fast diffusion of DnaE suggests that a single DnaE molecule will only bind to the replisome for a very short time (< 20 ms) before being released, meaning that DnaE is only responsible for the synthesis of a short DNA segment at a time.

How the dynamics of the replisomal proteins change when the replication is arrested is also of great interest. In this thesis, I arrested the DNA replication with two different mechanisms, and characterized the corresponding protein dynamics. HPUra, a drug that binds to the active site of PolC, did not change the diffusion coefficients and the corresponding fractions of PolC, but caused a shorter dwell time. This response suggests that the architecture of the replisome is not destroyed upon HPUra treatment, and that PolC can still bind to the replisome. However, the binding affinity does change, resulting

in a faster exchange rate. Similarly, a portion of DnaX molecules also had a shorter dwell time in HPUra treated cells, supporting the previous hypothesis that the two proteins are coupled in DNA replication. Surprisingly, another portion of DnaX molecules showed a very long dwell time (> 8 s), indicating that there is another binding site with a very strong affinity at the replisome available to DnaX in HPUra treated cells.

The changes in dynamics were different when the cell was treated with mitomycin C (MMC), a drug that causes DNA crosslinks and thus arrests replication. PolC became no longer trackable in the MMC-treated cells; we not only lose the slow-moving subpopulation of PolC that binds to the replisome, but we also lose the fast-moving subpopulation that searches along the DNA strands. All PolC molecules were freely diffusing in the cytoplasm with a diffusion coefficient of about 1.4 μm^2/s. However, this does not mean that MMC causes the collapse of the replisome structure. DnaX could still bind to the replisome in MMC-treated cells. Like in HPUra-treated cells, DnaX in MMC treated cells also has two subpopulations with two dwell times: one portion dwells for the time similar to that in untreated cells (~ 3 seconds), and the other portion dwells for a very long time at the replisome (> 8 seconds). The dynamics show that MMC did not affect the whole architecture of the replisome, and the original binding site for DnaX still remains with a similar affinity. Another binding site with a very strong affinity, possibly related to the DNA repair process, is available to DnaX just like in HPUra-treated cells.

We also investigated the stoichiometry of the replisomal proteins *in vivo*. In Chapter 2, the stoichiometry was determined by counting the photobleaching steps of the protein foci in live *B. subtilis*. In Chapter 4, I developed a new algorithm, which applies

90

Bayesian statistics to count the photobleaching steps with minimum arbitrary parameters and subjective decisions. The algorithm proves to give reasonable results in both simulated and experimental data. The results yield by this algorithm still have room for improvement; future work will retrieve the true fluorescence intensity of the molecules with a recently developed background subtraction methods called SMALL-LABS[4].

DNA replication is a complicated process, and the replisome is a multi-protein complex with an intricate organization. There still remain many questions that can be answered with our powerful microscopy tools. For example, given the dynamics of DnaX, the β-clamp loader, an obvious future plan is to characterize the dynamics of DnaN, the β-clamp itself. Another interesting protein to look at is the single-stranded DNA binding protein (SSB), which is involved in both DNA replication and DNA repair. Additionally, protein-protein interactions can be observed and analyzed with two-color imaging. On the other hand, the current data analysis tools can be improved as well to augment our image processing. For example, non-parametric Bayesian statistics can be applied to the diffusion model to quantify the diffusion coefficients more objectively (free of experimenter's bias)[5], and with more information, including the probabilities of the transitions between the diffusive modes.

Overall, with improvements to super-resolution microscopy and data analysis methods, their applications in biomedical research is becoming broader. On one hand, better imaging configurations and devices, brighter fluorescent labels, and versatile labelling methods have made it possible to achieve high localization precisions with high temporal resolutions *in vivo*. On the other hand, better statistical models and the

development of data mining and machine learning will enable us to characterize, model, and predict the behaviors of single-molecules and the interactions between molecules in a profound and efficient way. All of this information will greatly improve our understanding of many biological processes and pathways.

References

1. Liao Y, Li Y, Schroeder JW, Simmons LA, Biteen JS. Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells. *Biophys J*. 2016;111(12):2562-2569.

2. Sawitzke J, Austin S. An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol Microbiol*. 2001;40(4):786-794.

3. Sanders GM, Dallmann HG, McHenry CS. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol Cell*. 2010;37(2):273-281.

4. Isaacoff B, Li Y, Lee S, Biteen J. SMALL-LABS: An algorithm for localizing and measuring single molecules in the presence of obscuring backgrounds. submitted. .

5. Hines KE, Bankston JR, Aldrich RW. Analyzing single-molecule time series via nonparametric bayesian inference. *Biophys J*. 2015;108(3):540-556.