The Structure and Dynamics of the Bacterial Replisome

Sarah M. Mangiameli

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

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Paul Wiggins, Chair

Houra Merrikh

Jason Detwiler

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

The Structure and Dynamics of the Bacterial Replisome

Sarah M. Mangiameli

Chair of the Supervisory Committee:
Professor Paul Wiggins
Physics Department

Cellular reproduction relies on DNA replication. Many bacteria, including *Bacillus subtilis* and *Escherichia coli*, possess only a single circular chromosome which is replicated bi-directionally from a single origin. It has remained controversial whether replication is achieved by mobile protein complexes that translocate along the DNA like a track, or by a stationary "replication factory" through which the DNA is pulled. In the first part of this thesis, we used time-lapse imaging to observe and quantify the position of fluorescently-labeled processivity-clamp (DnaN) complexes throughout the cell cycle. We find that the two replication forks maintain diffraction limited separation during the majority of the cell cycle, suggesting a factory-like model for replication. These data unify the models for replication across several domains of life.

The canonical model of DNA replication describes a highly-processive and largely continuous process by which the genome is duplicated. However, many simultaneously-occurring processes compete with replication for the DNA substrate, potentially causing disruptions that could decrease the timeliness of the replication process and leave DNA vulnerable to mutations. In the second part of this thesis, we characterize the dynamics of the replisome in individual cells using fluorescence microscopy with single-molecule sensitivity. We find that the replisome is disassembled multiple times per cell cycle. Furthermore, this instability is caused by transcription.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

BP: Base Pairs

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

DNA: Deoxyribonucleic Acid

GFP: Green Fluorescent Protein

IPTG: Isopropyl β-D-1-thiogalactopyranoside

KDE: Kernel Density Estimation

PSF: Point Spread Function

RNA: Ribonucleic Acid

RNAP: RNA Polymerase

SMFM: Single Molecule Fluorescence Microscopy

PUBLICATIONS

This thesis is based on the following publications:

(i) Mangiameli S.M., Cass, J.A., Merrikh, H., Wiggins, P.A. Bacterial replisomes show factory- like positioning. In preparation for *Current Genetics*.

(ii) Mangiameli S.M., Veit, B.T., Merrikh, H., Wiggins, P.A. (2017) The replisomes remain spatially proximal throughout the cell cycle in bacteria. *PLOS Genetics*. 11(6):e1005289.

(iii) Mangiameli S.M., Merrikh, C.N., Wiggins, P.A., Merrikh, H. (2017) Transcription leads to pervasive replisome instability in bacteria. *eLife*.

Reprints were made with permission from the respective publishers.

ACKNOWLEDGMENTS

First of all, I am forever grateful to my advisor, Paul Wiggins, for supporting me through four and a half years of grad school. Thank you for sharing your knowledge and ideas with me, and for teaching me to do research. Studying biological systems was not my intent when I signed up for a physics program. Thank you for ruining all my plans. At first, I was drawn in by your patience and enthusiasm, but now I can't imagine doing anything else. Thank you for everything.

To Houra Merrikh, who was like an unofficial second advisor to me. You were there for me in so many ways, and I would not have been able to do this without you. Thank you for forcing me to learn something about biology, for including me, and most importantly for letting me do cartwheels in your living room the day my first paper was accepted.

Chris Merrikh and Kevin Lang, thank you for excellent advice on science and spectacularly terrible advice on pretty much everything else. You guys taught me to clone, and now I have countless single-celled minions whose only job is to store my fancy DNA constructs. Life changing.

When I first started in the lab Nate Kuwada and Scot McBride patiently answered all my questions. Every five minutes. For weeks. Seriously.

To the members of my supervisory committee, Paul Wiggins, Houra Merrikh, Jason Detwiler, Andreas Karch, and Lutz Maibaum, thank you for your guidance. Special thanks to my reading committee (Paul Wiggins, Houra Merrikh, and Jason Detwiler), who got stuck reading this thesis in its entirety.

During my time in the Wiggins Lab, my labmates we always available for helpful

conversations. Colin, Jackie, Julie, Monica, and Stella, thank you for your insight and for helping preserve my sanity.

DEDICATION

To the countless bacteria who died in the making of this thesis. Thank you. Your sacrifice has allowed me to move forward.

Chapter 1
INTRODUCTION

Perhaps the most iconic molecule of all time is deoxyribonucleic acid (DNA), the genetic blueprint for all cellular organisms [1]. Instructions for all cellular processes are encoded by the sequence of four nucleotides along the well-recognized helical structure. Ultimately, these instructions must be duplicated and inherited by future generations.

Genome duplication is achieved by the coordinated action of multi-protein assemblies known as replisomes [1, 2]. Cell viability critically relies on the ability of the replisome to efficiently and accurately duplicate the genome, even in the hectic cellular environment where a number of simultaneously-occurring processes compete for the DNA substrate and impede the replisome [3–7]. Disruption of the replication process destabilizes the genome, leaving the DNA vulnerable to mutations [8–10]. Detrimental mutations in our own cells are responsible for many non-infectious diseases, such as cancers [9]. In pathogens, accumulation of mutations can lead to antibiotic resistance [6]. Despite the biological relevance of replication, much is unknown about the structure of the replisome, the replication mechanism, and how the replisome navigates competing processes.

The essential protein constituents of the replisome are highly conserved across all domains of life [2, 11]. Although an intact replisome has never been isolated from any organism, a significant body of genetic and biochemical research has identified these essential proteins and mapped the protein-protein interactions that combine them into a cohesive unit [12]. Functional replisomes can even be reconstituted *in vitro* from individual protein subunits purified from bacterial cells [13, 14]. Although these types of experiments have been instrumental in

[1] For some viruses, RNA, rather than DNA, serves as the genetic material

our understanding of replication, they hardly reproduce physiological conditions.

Until fairly recently, *in vivo* studies of replication have relied on ensemble methods. However, the stochastic nature of biological processes results in a high degree of cell-to-cell variation. Because of this heterogeneity, ensemble-averaged measurements mask the true molecular-scale dynamics. For example, emerging evidence suggests that the speed of an individual replisome varies significantly and stochastically during replication [15]. This dynamic is not detectable by ensemble methods where the signal is inherently the average replication speed and the contributions of individual replisomes are inseparable.

Recent developments in fluorescence microscopy provide sufficiently sensitive methods to study biological processes at the level of individual cells, or even individual proteins [16]. These methods facilitate *in vivo* work, while avoiding the pitfalls of ensemble methods. The two works presented in the subsequent chapters of this thesis [7, 17] rely on quantitative fluorescence microscopy to visualize the sub-cellular localization and dynamics of replication proteins in individual living cells as they duplicate the genome.

1.1 An Overview

The remainder of this chapter briefly introduces the tools used to study the replisome. In short, I will discuss the essential biology of two bacterial model organisms, *Bacillus subtilis* and *Escherichia coli*. Following an introduction to the relevant biological concepts, I will explain the significance of the problems investigated by the two published works [7, 17] comprising the subsequent chapters of this thesis. A short description of the imaging-based experimental techniques used in these studies concludes the chapter.

The first problem addressed by this thesis is sufficiently controversial to warrant a full-chapter of introduction (Chapter 2). The question is simple: Is it the DNA or the replisome that moves during replication? The answer is highly debated. Chapter 2 reviews the conflicting reports, and speculates about why the data have been so difficult to interpret. Briefly, previous studies have relied on the localization of the replisome on snap-shot images. Without knowledge of how the replisome positioning changes in time, the interpretation of

these images is unclear. My own work on this topic (Chapter 3), which was the first to visualize replisome positioning in time-lapse over complete cell cycles, strongly suggests that the DNA is pulled through a stationary replisome.

Chapter 4 argues against the common assumption that replication is a continuous process despite known obstacles. These obstacles include damage (double or single stranded breaks), DNA binding proteins, and concurrent cellular processes that also require DNA as a substrate. We find that the replisome disassembles multiple times per cell cycle. Furthermore, this instability is transcription-induced. Inhibition of transcription restores the stability of the replisome.

1.2 Bacterial model organisms

Experiments in this thesis were conducted in two bacterial model organisms: *Bacillus subtilis* and *Escherichia coli*. Due to the short doubling time, ease of genetic manipulations, and relatively simple chromosome structure, it is in these organisms that replication is currently best understood. Because replisome proteins are highly conserved [2, 11], the basic principles of replication uncovered in bacterial cells potentially have broad applicability to other organisms, including humans.

Both *B. subtilis* and *E. coli* are rod shaped bacteria, roughly two micrometers long and one micrometer in diameter. A multi-layered cell envelope encloses each bacterium, defining the cell boundary. The innermost layer of the cell envelope, the plasma membrane, surrounds the cytoplasm and mediates the exchange of molecules with environment. The cytoplasm itself is densely crowded, mainly with proteins that function to implement cellular processes. Bacterial cells lack the membrane-bound nucleus found in their eukaryotic counterparts, however, the chromosomal DNA localizes to a dense nucleoid region at the center of the cell (Fig. 1.1). Although the DNA is embedded directly in the cytoplasm, most proteins are unable to infiltrate the nucleoid region, and protein-DNA interactions usually occur at the cytoplasm-nucleoid interface [18].

Cellular reproduction occurs by binary fission. Exponentially growing cells elongate

according to

$$L = L_0 \exp(t/t_0), \tag{1.1}$$

where L_0 is the length of the cell at birth, t is the cell age, and t_0 is a constant that characterizes the growth rate. During steady-state growth, cell division occurs when the length has doubled after a time t_D given by:

$$t_D = t_0 \log(2). \tag{1.2}$$

Provided appropriate growth conditions, *B. subtilis* and *E. coli* cells are able to double in as little as 20 min. During this time, cells must double their biomass, replicate the chromosome, and separate the two copies into opposite cell halves prior to septum formation. A more detailed description of these processes follows in the next sections.

1.3 Central dogma

The central dogma of microbiology describes the flow of information during gene expression. In short, the genetic information from a specific DNA sequence is *transcribed* into an RNA message, and finally *translated* to build proteins [19]. Importantly, while some genes required for cellular maintenance are constitutively expressed, most genes are expressed conditionally. Regulation of a particular gene product may occur at any stage during the expression process. The subsequent sections describe each stage of gene expression in detail.

1.3.1 DNA

Nucleotides are the building blocks of DNA. Each nucleotide consists of a sugar, a phosphate, and one of four bases (adenine (A), thymine (T), cytosine (C), or guanine (G)). The sugar and phosphate groups form a structural backbone to support the bases. Bases pair specifically, A with T, and C with G. Double-stranded DNA is formed through hydrogen bonds between the bases of two complementary nucleotide chains. DNA strands are inherently polar. An exposed phosphate group defines the the 5' end while the 3' end has only a hydroxyl group.

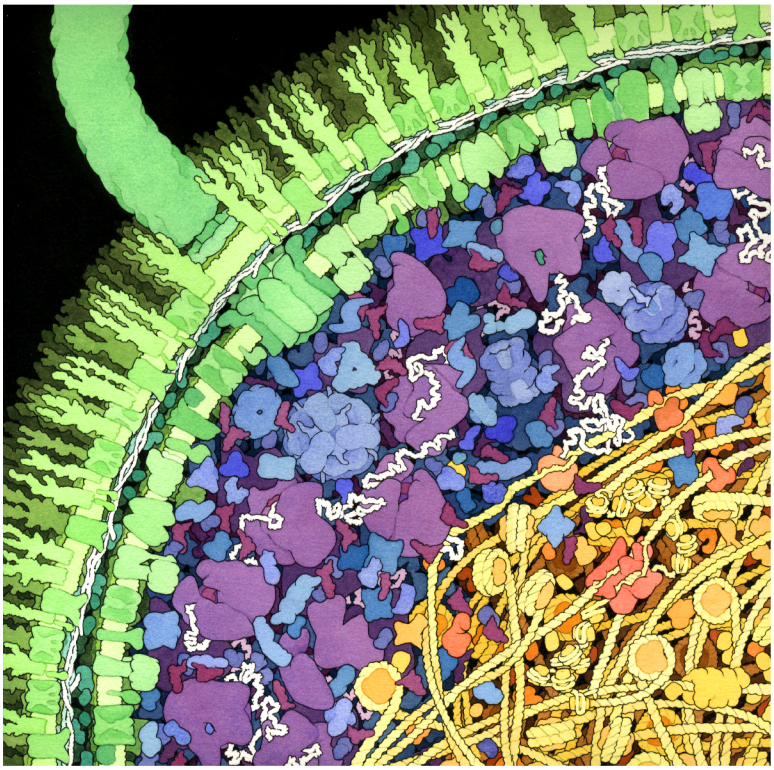

Figure 1.1: **The crowded cellular environment.** View of a small section of a *E. coli* cell. The cell envelope is shown in green. The dense nucleoid is shown in yellow. Many DNA-binding proteins coat the DNA. The remaining cytoplasmic space (blues and purples) is filled mainly with ribosomes and proteins proteins. In particular, the ribosomes (purple) can be seen translating mRNA (white strands) to build the necessary proteins for cellular functions. Illustration by David S. Goodsell, the Scripps Research Institute.

The strands in duplex DNA have opposite polarity. In the current context, polarity will be most important for orienting the DNA.

1.3.2 Transcription to RNA

Gene expression begins with transcription, the process of converting a particular DNA segment into RNA. For genes encoding proteins, messenger RNA (mRNA) will temporarily store a portable version of the DNA instructions for use in protein synthesis. RNA, like DNA, encodes information as a sequence of bases (although RNA is single-stranded). However, T is replaced with uracil (U) in the context of RNA.

RNA polymerase (RNAP) generates transcripts by reading one strand of the DNA duplex (the template strand) and synthesizing a complementary RNA strand. Polymerization by RNAP occurs exclusively in the 5'-3' direction. Highly expressed genes may be transcribed by several RNAP's in tandem, thereby increasing the RNA output. One such highly-transcribed region of the genome is the ribosomal DNA (rDNA). Transcripts generated from the rDNA (rRNA) are the main constituent of ribosomes[2], molecular machines that synthesize proteins, and one of the most numerous structures in the cell.

1.3.3 Translation to Proteins

For genes encoding proteins, the final step in expression is translation. Ribosomes construct proteins by assembling chains of amino acids as specified by the mRNA sequence. The mRNA is translated three nucleotides at a time, each triplet specifying a specific amino acid. Specific stop and start condons control which segments of the mRNA should be translated to form a protein product.

---

[2]Unlike messenger RNA, rRNA is not translated to form a protein product, but rather integrated directly into the structure of the ribosome.

1.4 DNA replication

Most bacteria have a single circular chromosome that is replicated bi-directionally from a unique origin sequence (*oriC*). Two replisomes initiate at the origin and work in opposite directions, each duplicating one arm of the chromosome before disassembling at the terminus (*ter*). This is shown schematically in Fig. 1.2. The *B. subtilis* and *E. coli* chromosomes are roughly 4.2 and 4.6 Mbp in length, respectively [20, 21]. Chromosome duplication has been reported to complete in as little as 40 min[3] *in vivo* [22], suggesting that the replisome is able to synthesize at least 1,000 bp/s. The replisome itself requires at least 11-13 distinct subunits to function, although the exact protein composition varies slightly between the two organisms [14, 23–25]. A selected list of these proteins and their functions are given in Table 1.1 (*E. coli*) and Table 1.2 (*B. subtilis*).

1.4.1 Structure of a replication fork

At the front of each replisome, a homohexameric helicase encircles one strand of the DNA duplex, serving to unwind and separate the template DNA ahead of the replisome [26, 27]. This results in a Y-shaped structure known as a replication fork (Fig. 1.3A and B). Separated template strands are quickly coated with single-stranded binding protein (SSB) to prevent re-formation of the DNA duplex and protect from damaging nucleolytic assaults [28]. The complementary strands to each of the separated templates are then synthesized simultaneously by the action of DNA polymerase.

Notably, DNA synthesis occurs only in the 5'-3' direction [19]. Due to the anti-parallel nature of the template DNA, only one of the templates (leading strand) may be replicated in the direction of replisome progression, while the other (lagging strand) must be replicated in reverse (Fig. 1.3B). Lagging-strand synthesis is discontinuous [29], with replication occurring

[3]I noted earlier that cells are able to double in as little as 20 minutes. This is possible because cells may initiate additional rounds of replication shortly after the origins have been copied. Cells are then born with partially duplicated chromosomes such that replication is able to keep up with division. To avoid the complication of multiple ongoing rounds of replication, the work here was done under slow growth conditions.

Component	**Protein Composition**	**Function**
Helicase	DnaB (6)	Separate DNA
Primase	DnaG (1)	RNA primer synthesis
Primary Polymerase	DnaE (1)	DNA syntehsis
	DnaQ (1)	Proofreading exonuclease
	HolE (1)	Unknown
Clamp loader	DnaX (3)	Close β-clamp
	HolA (1)	Open β-clamp
	HolB (1)	Structural element
	HolC (1)	Polymerase cycling
	HolD (1)	Structural element
β-clamp	DnaN (2)	Processivity clamp
Single-strand binding protein	SSB (4)	Protect ssDNA

Table 1.1: ***E. coli* replisome components.** Table lists core replisome components, and their constituent proteins. The number in parenthesis is the per-complex stoichiometry. For example, the helicase is composed of six identical DnaB proteins.

Component	**Protein Composition**	**Function**
Helicase	DnaC (6)	Separate DNA
Primase	DnaG (1)	RNA primer synthesis
Primary polymerase	PolC (1)	DNA synthesis, proofreading exonuclease
Accessory polymerase	DnaE (1)	DNA synthesis
Clamp loader	DnaX (3)	Close β-clamp
	HolA (1)	Open β-clamp
	HolB (1)	Structural element
β-clamp	DnaN (2)	Processivity clamp
Single-strand binding protein	SSB (4)	Protect ssDNA

Table 1.2: ***B. subtilis* replisome components.** Table lists core replisome components, and their constituent proteins. The number in parenthesis is the per-complex stoichiometry. For example, the helicase is composed of six identical DnaC proteins. Unlike *E. coli*, *B. subtilis* encodes distinct lagging- and leading-strand polymerases. The lagging-strand polymerase (DnaE) lacks proofreading capability and likely functions to begin Okazaki fragment replication before PolC takes over.

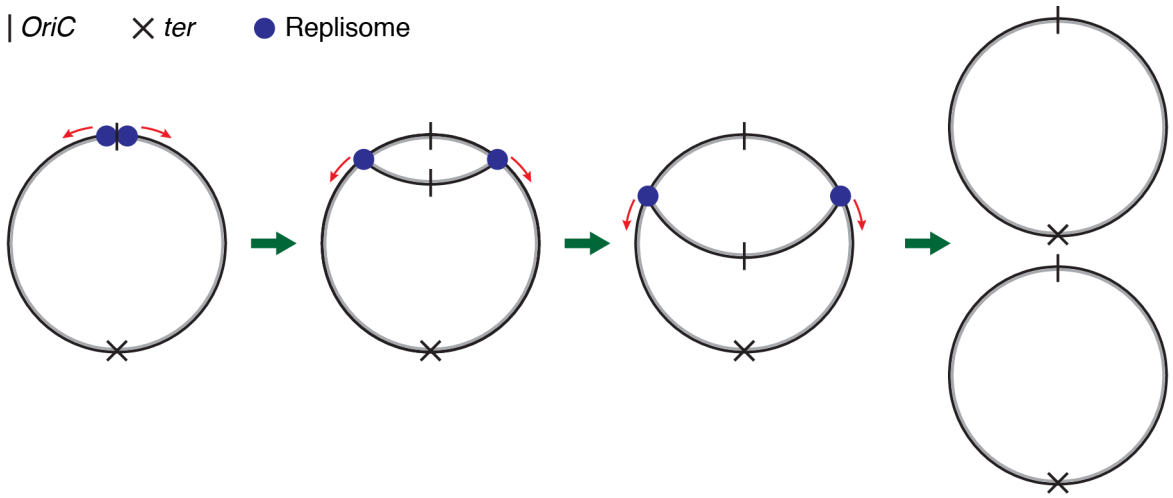

Figure 1.2: **Bi-directional replication of a circular chromosome shown schematically.** Two replisomes (blue dots) are assembled at the origin (*oriC*, vertical line) and proceed to duplicate the chromosome bi-directionally towards the terminus. (*ter*, cross). Red arrows indicate the direction of replication.

in short 1-2 kb segments [13], called Okazaki fragments. Replication of each segment is initiated with syntheses of an RNA primer (by primase, which associates transiently with the fork [30]), which is extended by DNA polymerase. Gaps between Okazaki fragments are later sealed by ligase [27].

In *E. coli* the leading- and lagging-strands are replicated by identical polymerases, which are composed of three protein subunits (a catalytic subunit, a proofreading exonuclease, and a subunit of unknown function) [27, 31]. Recent characterization of the stoichiometry of *E. coli* replisome components suggests that three polymerases associate with the replication fork [24], presumably one for leading-strand replication, and two that alternately replicate Okazaki fragments. *B. subtilis*, on the other hand, has two distinct polymerases [32, 33], each composed of a single subunit. One, DnaE, operates exclusively on the lagging strand [33], while the other (PolC) both replicates the leading strand, and possibly alternates with DnaE for lagging-strand synthesis [14, 22, 34].

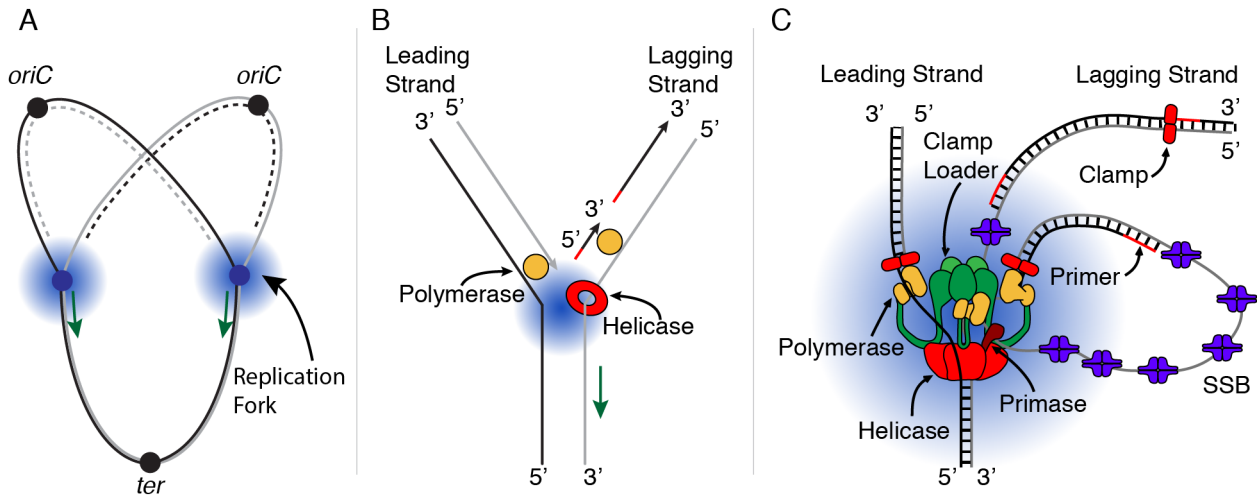

Figure 1.3: **Increasingly detailed views of the replication fork.** A) Partially-replicated chromosome. Blue shading represents the replication fork, and blue dots represent the replisomes. Newly synthesized DNA is indicated by broken lines. Green arrows show the direction of replication. B) Schematic of a single replication fork. The helicase (red) encircles the lagging strand template, unwinding and separating the DNA. The leading- and lagging-strand polymerases (yellow) simultaneously replicate both strands. DNA polymerase synthesizes 5'-3'. The leading strand may be replicated in the direction of fork progression (green arrow), but the lagging strand is replicated discontinuously in reverse. Red line segments represent RNA primers that initiate Okazaki fragments. C) Detailed view of the replication fork showing major replisome components.

Unaccompanied DNA polymerase is poorly processive [27], generally incorporating less than 10 bp before dissociating from the DNA. However, when linked to the DNA by the β-clamp, the polymerase is able to replicate ≥ 50 kb during a single binding event [12]. The β-clamp is pried open and assembled around the DNA by the clamp-loader complex [12, 26]. In addition to continually associating with the leading-strand polymerase, the β-clamp is also loaded at the beginning of each Okazaki fragment for lagging-strand synthesis. Although the β-clamp accumulates behind the replication fork [35, 36], it is eventually removed (likely by the clamp-loader) and recycled [37]. An additional function of the clamp loader is coupling of the leading- and lagging-strand templates through its interactions with the helicase and polymerases [14, 24, 38, 39] (Fig. 1.3C).

1.4.2 What moves during replication?

One could imagine two mechanisms for chromosome replication [40]: (i) The replisomes translocate along the DNA, like a train on a track (Track Model) or (ii) after assembly at the origin, the replisomes remain fixed in place while the DNA is pulled through (Factory Model). In principle, some combination of these two is also possible. Although conceptually simple, experimentally distinguishing between these options has proven quite difficult. A few decades worth of conflicting reports argue for both models (see Chapter 2). The Factory Model is most consistent with the author's own data (see Chapter 3 or Ref. [17]), and therefore assumed to be correct for the remainder of this thesis (note that Chapter 2 attempts to justify this assumption).

1.5 Chromosome structure

When visualizing DNA replication, it is important to remember that the chromosome is highly condensed in cells [18, 41, 42]. If the chromosome were a stretched out circle, the way we draw it schematically, the circumference would be over 1 mm. Cramming the chromosome into the roughly 1 µm-diameter compartment where it resides requires a significant amount of condensation [18]. To achieve this, the DNA is negatively super-coiled (underwound), causing

it to twist up on itself, and bent around numerous DNA-binding proteins that function to structure the chromosome [43–45]. This chromosome structure must be both intricately organized and fluid as cells must not only access specific genes for expression, but also replicate the entire chromosome and segregate the copies to opposite sides of the cells prior to division [18, 43, 44, 46].

1.6 Replication Conflicts

Replication forks must navigate a variety of obstacles as they progress through the chromosome. Damaged DNA, DNA-binding proteins, secondary structures in the DNA itself, and encounters with the transcription machinery all lead to *replication conflicts* [47–50]. Failure to resolve conflicts compromises genome integrity, and may reduce fitness [51, 52]. The work here focuses on *replication-transcription conflicts* in particular. In bacteria, the lack of spatial and temporal separation between the replication and transcription processes inevitably leads to clashes between the two machineries [3, 49, 50, 53].

1.6.1 Conflict orientation

Replication-transcription conflicts occur in two orientations: (i) head-on and (ii) co-directional [4]. The determining factor is which strand of the DNA serves as a template for transcription. Specifically, the replicative helicase encircles the lagging strand and processes DNA in the 5'-3' direction. Meanwhile, RNAP translocates along the template DNA strand in the 3'-5' direction. Therefore, RNA polymerases transcribing the lagging strand will encounter the replisome head-on (see Fig. 1.4, top). On-the-other hand, RNA polymerases transcribing the leading strand are co-oriented with the replisome (see Fig. 1.4, bottom). Importantly, transcription progresses at only 50-100 bp/s (compared to $\sim$1000 bp/s for replication), therefore the replisome will inevitably overtake RNAP [49].

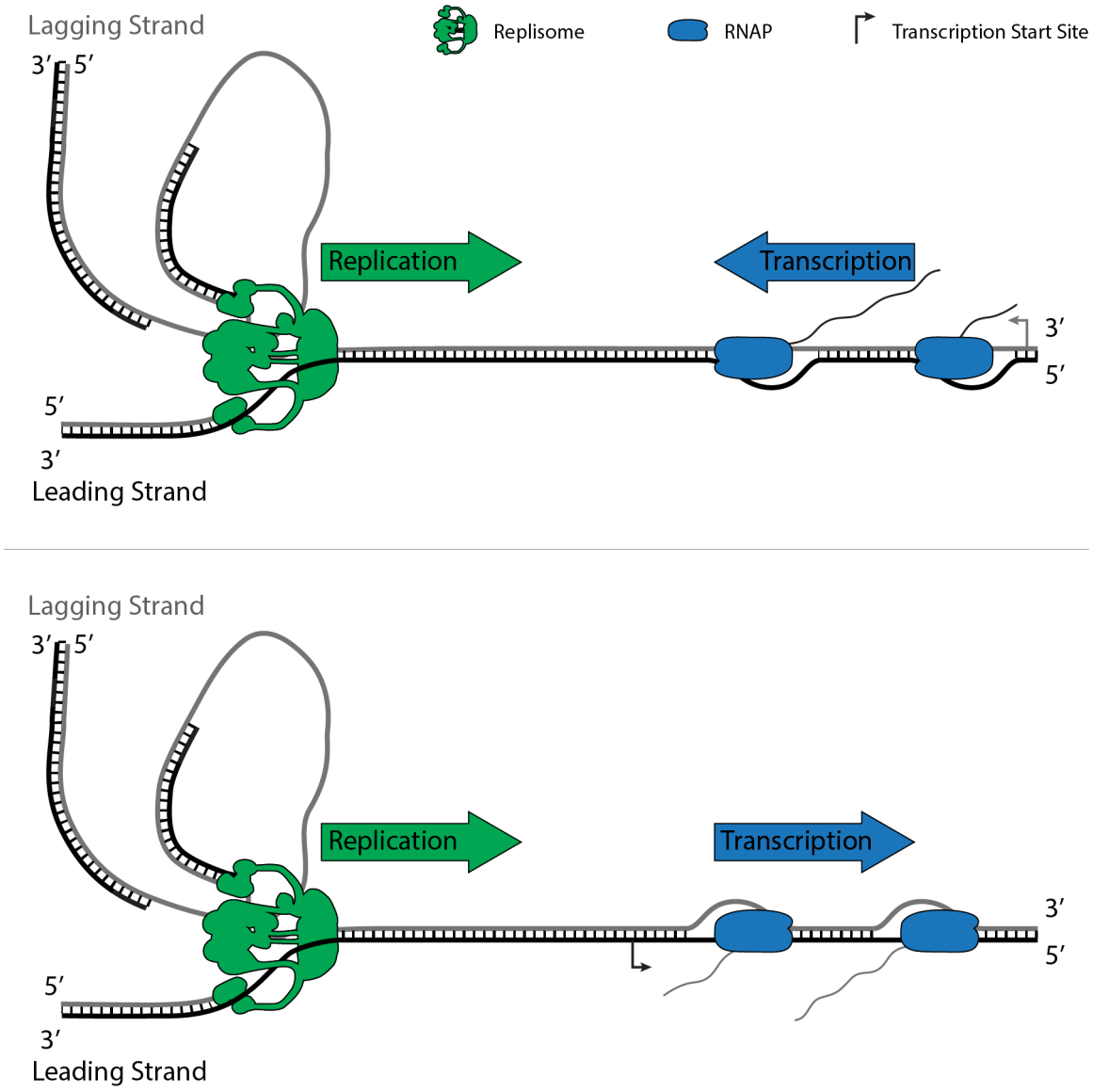

Figure 1.4: **Head-on and co-directional replication-transcription conflicts.** Top: For genes transcribed from the lagging strand (gray) the replisome encounters RNAP head-on. Bottom: For genes transcribed from the leading strand (black) the replisome encounters RNAP co-directionally.

1.6.2 Replisome disassembly as a result of conflicts

Head-on conflicts are more detrimental to the replication process than co-directional conflicts [50, 51, 54–62]. Replication fork stalling has been reported to occur at the sites of head-on conflicts, although the fate of the replisome itself during these stalling events has remained unclear [4]. However, full or partial disassembly of the replisome is suggested by (i) the essentiality of replication restart proteins [63, 64] and (ii) the accumulation of these restart proteins at the genomic locations of head-on conflicts [3, 5].

By chance, one might expect to find roughly equal numbers of genes encoded on the lagging and leading strands. This is not the case; bacterial genomes are organized so as to minimize head-on conflicts [4, 20, 65]. In particular, the highly-transcribed rDNA genes are encoded on the leading strand in virtually all bacteria, presumably as a mechanism for head-on conflict avoidance [66–70]. Interestingly, inversion of the rDNA (so that it will be transcribed in the head-on orientation) results in DNA damage and reduced fitness [50, 55–57].

Although less severe, co-directional conflicts are not without consequence. A co-directional encounter with a single RNAP may not be sufficient to disrupt the replisome [71, 72]. However, during rapid growth as many 50-100 RNAP's may simultaneously transcribe each rDNA operon to support increased levels of proteins synthesis. In fact, it has recently been reported that replication-restart proteins associate with both head-on genes *and* highly-transcribed co-orientated genes such as the rDNA [3].

The work presented in Chapter 4 explicitly observes replisome disassembly and reassembly during replication-transcription conflicts. Surprisingly, we find that these disassembly events occur many times per cell cycle. This is in contrast with the canonical view that replication is a continuous process.

1.7 An introduction to microscopy

Much of the work here relies on phase-contrast microscopy to visualize cell morphology, and fluorescence microscopy to localize and track replication proteins. During microscopy, cells

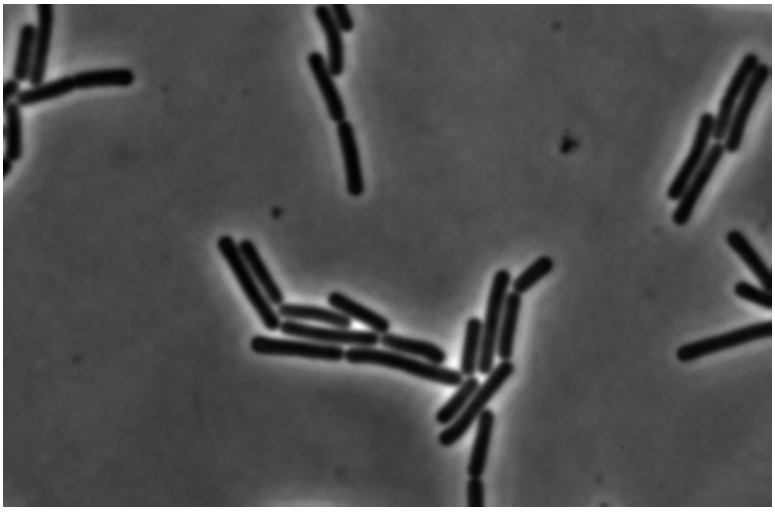

Figure 1.5: **Phase-contrast image of *B. subtilis* cells.** Phase-contrast is a technique for increasing contrast when observing a weakly absorbing specimen such as bacterial cells. Cells appear dark against a light background.

are mounted on a transparent gel made from growth medium that is supported by a glass slide. These conditions are sufficient to maintain exponential growth for several generations. By combining the methods described in the subsequent sections, we are able visualize the replication process in individual living cells as they elongate and divide.

1.7.1 Phase contrast microscopy

Because of their small size, bacterial cells are nearly transparent to incident light, and are therefore poorly visible against the background using basic light microscopy. Phase-contrast microscopy is a technique for enhancing contrast, first described in the 1930's by Dutch physicist Frits Zernike, who later won the Nobel prize for his invention [73]. Briefly, optical manipulations ensure that the small amount of light scattered by the sample is out of phase with undeviated light (in the image plane), causing cells to appear dark against a light background (Fig. 1.5). Although this technique is instrumental in locating cell boundaries, information about the cellular interior is unresolvable.

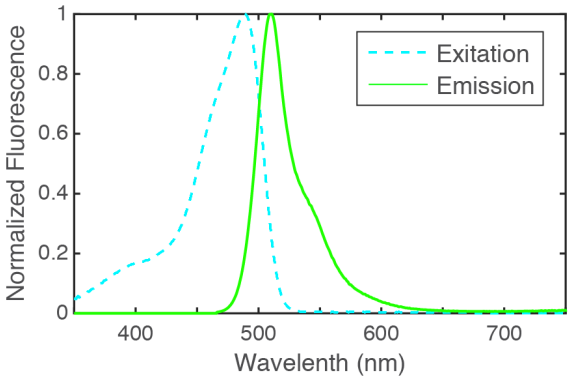

Figure 1.6: **Example excitation and emission spectrum.** A commonly used fluorophore, enhanced green fluorescent protein, is efficiently excited with a 488 nm laser line. The emitted light (spectrum shown in green) is separable from the excitation laser light due to its longer wavelength.

1.7.2 Fluorescence microscopy

Fluorescence microscopy is used to view and track specific molecules within a cell. In concept, the target molecule is conjugated to a fluorophore, a compound that absorbs and quickly re-emits light. The position of the target molecule is detected based on the emitted light from the fluorophore. Importantly, fluorophores re-radiate at a longer wavelength than they absorb, allowing the emitted light to be separated from the excitation light by selective filters (Fig. 1.6). Photodamage or chemical reactions with environmental molecules will eventually cause structural changes to the fluorophore that induce a permanent dark state. At this point, the fluorophore is said to have become photobleached and can no longer be tracked. Although generally undesirable, the work in Chapter 4 uses this effect to count the number of proteins in a complex.

1.7.3 Fluorescent fusions

Some fluorophores (fluorescent proteins) may be built from amino acids, and therefore manufactured by bacterial cells, provided the appropriate DNA instructions. The gene for the first fluorescent protein (green fluorescent protein, or GFP) was isolated from a bioluminescent jellyfish, *Aequorea victoria* [74]. GFP derivatives have since been expressed in many organisms, including bacterial cells.

Simply producing fluorescent protein isn't sufficient for many applications, in fact, it is often desirable to produce a genetic fusion between the target and fluorescent protein. In short, the gene encoding the fluorescent protein is introduced adjacent to (or occasionally in the middle of) the gene encoding the target proteins. Any intermediary translational stop condons are removed, so that expression of the modified gene will produce a fusion protein. The target protein (hopefully) retains its functionality but can be tracked using the fluorescent label. A variety of spectrally-separable fluorescent proteins are now available making it possible simultaneously image several target proteins [75]. The work here uses genetically encoded fusions as the main labeling strategy.

1.7.4 Diffusion

Fluorescence from bound protein appears structured, whereas fluorescence from cytoplasmic protein appear diffuse. The mechanism for this is as follows. Dynamics in the cytoplasm are governed by thermal forces, resulting in diffusive motion of the constituent proteins. The diffusion coefficient (D) for a spherical particle is given by the Stokes-Einstein relation,

$$D = \frac{kT}{6\pi r\eta} \tag{1.3}$$

and depends on the temperature (T), radius of the particle (r), and the viscosity of the surrounding medium (η). Boltzman's constant is indicated by k. An estimate of the diffusion constant for a standard globular protein in water is roughly 100 $\mu m^2/s$ [76]. By dimensional analysis, the characteristic time (t) for a particle to travel a distance d is roughly:

$$t \approx d^2/D \tag{1.4}$$

Taking the length of a typical bacterial cell as 2 µm, the typical time for a protein to diffuse across the cell is approximately 40 ms. Typical exposure times for this work were in the 300-600 ms range, more than enough for the fluorescence due to cytoplasmic protein to blur across the image.

In the context of this thesis, labeled replisome proteins that are bound to the DNA as part of an active replication complex appear as foci. However, in the event of replisome disassembly (either at termination or because of a conflict), the focus becomes diffuse. As a side note, the cell produces many more copies of each replication protein than are included in active replisomes at any given time. These delocalized proteins (an estimated 80-1300 copies, depending on the protein [24]) diffuse around the cytoplasm, contributing to the background fluorescence.

1.7.5 Diffraction limit

Although the sub-cellular locations of proteins may be visualized using fluorescence microscopy, the physical properties of light fundamentally limit the achievable spatial resolution. Briefly, light imaged through an aperture (such as an objective lens) is spread out due to diffraction [77]. The point spread function (PSF) is defined as the diffraction pattern produced by a point source when observed through an imaging system. The idealized PSF for a circular aperture is an Airy disk, which is well approximated by a Gaussian. Closely-spaced point sources in the sample plane will produce overlapping PSFs in the image plane, at some point causing the two sources to appear as a single emitter. The diffraction limit describes the minimum separation between two point sources such they can still be resolved in the image plane. For a typical setup, this distance is roughly 250 nm. Although many techniques for surpassing the diffraction limit have been reported [78], one cannot expect to resolve structural details below this length scale using basic fluorescence microscopy.

The size of a replication protein is on the order of a few nanometers. Fluorescently-labeled subunits within a single replisome will be diffraction limited and appear as a single focus. Interestingly, we find in Chapter 3 that the separation of *replisome pairs* is below the

diffraction limit during the majority of the replication cycle. Therefore, when visualizing replication proteins, a single diffraction-limited focus generally represents two replisomes.

Chapter 2

THE BACTERIAL REPLISOME HAS FACTORY-LIKE LOCALIZATION

2.1 Abstract

DNA replication is essential to cellular proliferation. The cellular-scale organization of the replication machinery (replisome) and the replicating chromosome has remained controversial. Two competing models describe the replication process: In the *track model*, the replisomes translocate along the DNA like a train on a track. Alternately, in the *factory model*, the replisomes form a stationary complex through which the DNA is pulled. We summarize the evidence for each model and discuss number of confounding aspects that complicate interpretation of the data. We advocate a factory-like model for bacterial replication where the replisomes form a relatively stationary and weakly-associated complex that can transiently separate.

2.2 Introduction

The essential process of DNA replication underlies cellular proliferation. In all organisms, multi-protein replication complexes assemble at specific DNA sequences (origins) and proceed to replicate the genome by a semi-conservative process [79]. Although functional replication complexes have been reconstituted *in vitro* from purified bacterial proteins [13, 14], many interesting questions remain about the replication process and its regulation in the context of the cell [17]. In particular, the *in vivo* cellular-scale organization of the replication machinery (replisome) and the replicating chromosome has remained controversial in bacteria. Do replisomes translocate along the DNA molecule, or is the DNA molecule pulled through the replisomes? We refer to the first model as a *track model*, and the latter as a *factory model*.

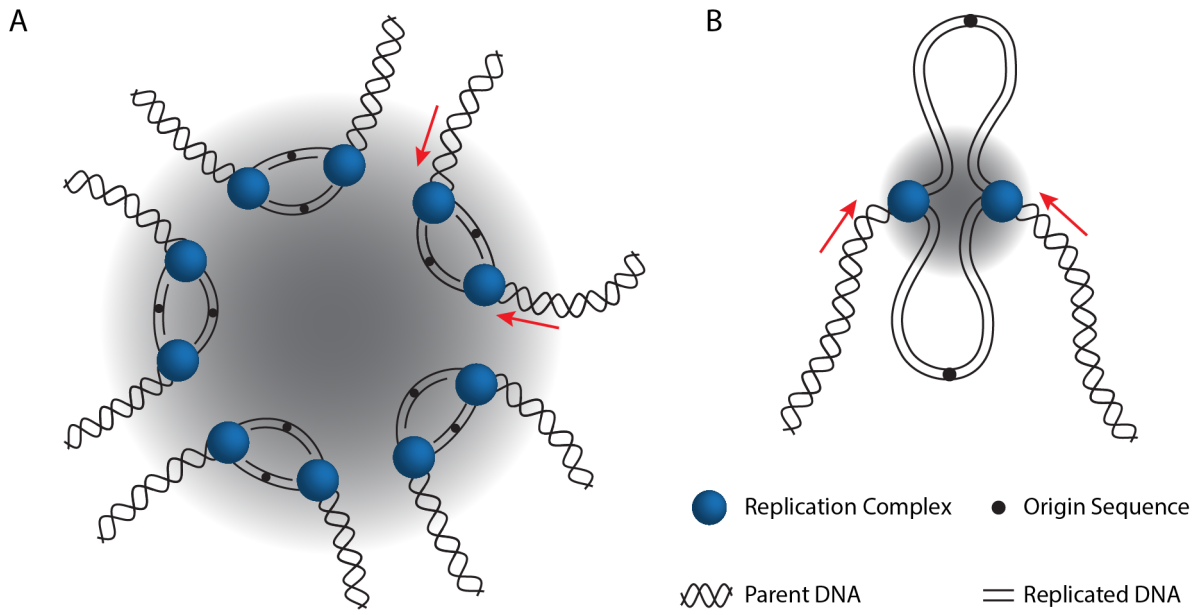

Figure 2.1: **Eukaryotic and Bacterial replication factories.** A) A Eukaryotic replication factory (gray) resulting from simultaneous activation of 5 origin sequences (black dots). Replication complexes (blue) assemble and disassemble, but no spatial migration is observed. Red arrows indicate the direction DNA is pulled into the complex. Adapted from Ref. [80]. B) A bacterial replisome factory would involve only two replication complexes working bi-directionally from from a single origin. Red arrows indicate the direction DNA is pulled into the complex.

Cell-biology experiments have provided evidence for both models in the bacterial cell.

The factory model was first established in eukaryotic cells [81], where DNA synthesis is localized to a number of puncta, called *replication factories*, distributed throughout the nucleus [80, 82]. The number of puncta is much smaller than the number of origins, implying that many replisomes co-localize to each factory. Each replication factory consists of roughly five clustered and synchronously-activated origins [83, 84] (Fig. 2.1). Early evidence for replication factories included the visualization of punctate foci formed by the co-localization

of newly synthesized DNA with replication proteins by immunofluorescence [85]. More recent imaging studies have demonstrated the spatiotemporal stability of replication factories in living cells based on the sub-nuclear positioning of a fluorescent fusion of the processivity clamp (PCNA), and have hypothesized the existence of a physical linker anchoring the factory in the nucleus [86]. However, the analysis of replication in eukaryotic cells is confounded by the large number of replication origins distributed throughout the genome, making it difficult to study the dynamics of individual replisomes.

Bacterial DNA replication is comparatively simple: Most bacteria possess only a single circular chromosome that is replicated bi-directionally from a single origin of replication. Two replisomes initiate at the origin and process DNA in opposite directions, each replicating one arm of the chromosome before meeting in the terminus region. Because replication proteins are highly conserved across all organisms [11], sequence conservation might be expected to result in a common organization of the replication process between bacteria and eukaryotes. However, it has remained unclear whether factory-like organization is conserved in bacteria.

Although published evidence generally supports the factory model in the model organisms *Caulobacter crescentus* [87] and *Bacillus subtilis* [7, 40, 88, 89], a number of recent reports suggest prolonged separation of replication fork pairs *Escherichia coli* [90–93]. We have recently published a comparative study of the organization of replication process in both *E. coli* and *B. subtilis* [17] where we observed unexpected similarities between the organization in these two highly-divergent species [7]. The study, which used fluorescence microscopy to track replisome components over complete cell cycles, provided a natural explanation for the contradictory reports: There are two different stages of the cell cycle that can result in two optically-resolvable replication foci: (i) track-like organization consisting of individually resolvable replisomes (ii) factory-like organization consisting unresolved replisome pairs at the quarter cell position after replication re-initiation. As a result, analysis dependent on counting the number of observed replication foci using single images cannot distinguish between these two populations and therefore is *not* a reliable method for determining replisome organization [7].

2.3 Models for the organization of replication

To discuss the evidence for the competing models, it is necessary to describe these models precisely. In all models, replisome pairs are coincident during and shortly after initiation. Subsequently, in the *track model*, the replisomes translocate in opposite directions along a (relatively) stationary DNA track, leading to sustained separation of replisome pairs as they follow chromosomal structure (Fig. 2.2). Since the replisomes translocate along the chromosome, the specific replisome localization predicted by the track model is dependent on nucleoid structure. For example, in *C. crescentus* the chromosome is organized such that *oriC* and *ter* are positioned at opposite cell poles [87] (see Fig. 2.2B), therefore the track model would predict end-to-end displacement of the replisomes along the DNA molecule. Due to the circular topology of the chromosome, replisome pairs may co-localize again at the terminus, but this is not necessary. If one member of a replisome pair arrives in the terminus region earlier, it may release the template DNA before other arrives.

In the factory model, DNA is pulled or spooled through the replication complex. From a mechanistic perspective, this model *may* include a factor that anchors the replisomes to the cell, or to one another, but it *may not*. Generically speaking, the sub-cellular localization of the replisomes may change, as long as the movement is driven by chromosomal re-arangements rather than motion of the replisomes along the DNA. One attractive feature of the factory model is that it may facilitate the process of chromosome segregation. For instance, an anchoring mechanism could prevent the newly replicated DNA from becoming concatenated. Furthermore, extrusion of the newly replicated DNA towards opposite poles of the cell could preventing mixing, further facilitating segregation [94, 95].

2.4 The visualization of replisome dynamics

The most direct approach to characterizing the cellular organization of the replication process is by time-lapse imaging throughout the cell cycle. We have characterized the full-cell-cycle dynamics of the replisome in both *B. subtilis* and *E. coli* cells [7]. The first challenge in these

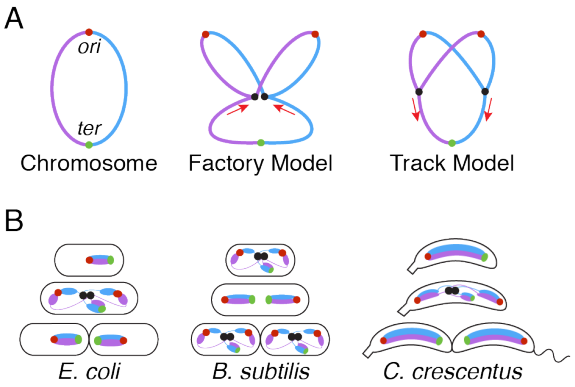

Figure 2.2: **The factory and track models.** A) Schematic diagrams for the factory and track models. Most bacteria have a single circular chromosome (left), with a single origin (red dot), positioned roughly opposite the terminus (green dot). The left and right arms of the chromosome are colored pink and green, respectively. In the factory model (center), DNA is pulled through the replisomes (black dots) in the direction indicated by the red arrows. In the track model (right), replisomes translocate along the template DNA. Red arrows indicate the direction of replisome motion. B) The factory model is shown for the chromosomal organizations of three model organisms. Decondensed DNA is represented by lines. Note that the *E. coli* chromosomal structure reflects experimental observations in Ref. [96].

experiments is finding a replisome-associated protein with sufficiently high stoichiometry to observe throughout the cell cycle. In both organisms we imaged a fluorescent fusion to the processivity clamp (DnaN), expressed from the endogenous locus, as a proxy for the replisome localization. These fusions do not lead to a detectable defect in replication or growth. Furthermore, cells were propagated using a slow growth rate to avoid the complications of multi-fork replication.

Contrary to previous reports, *E. coli* and *B. subtilis* showed virtually indistinguishable replisome dynamics, summarized as follows: (i) A single midcell focus appears shortly before or after the beginning of the cell cycle. Cells born with a focus initiated replication before the beginning of the cell cycle. We infer that this single focus represents a pair of replisomes that cannot be optically resolved. (ii) The mid-cell focus subsequently exhibits confined random motion (as observed previously [97]), and is occasionally observed to reversibly separate such that individual replisomes are optically resolvable. Importantly, the focus separation is both small (a fifth of a cell length on average) and transient. (iii) The midcell focus disappears before the end of the cell cycle. We infer that this replisome disassembly corresponds to the termination of replication. (iv) In roughly 45% of cells, foci re-appear at the quarter-cell positions. We infer that these quarter-cell foci result from re-initiation of replication prior to cell division as they persist through cell division and are unable to form in an initiation-deficient conditional mutant. Therefore these observed quarter-cell loci represent replisome pairs. In agreement with one of the oldest studies of replisome localization in *B. subtilis* [40], we find that replisome pair separation is diffraction-limited roughly 80% of the time [7] (See Fig. 2.3).

Throughout the majority of the replication cycle, a factory-like organization is observed in both *B. subtilis* and *E. coli*. Consistent with the factory model, replisome pairs operate in close proximity for the majority ($\approx 80\%$) of the replication cycle. Although the two replisomes are usually not optically resolvable, they do not appear to be strongly associated since transient separation events *do occur*. Although these transient separation events could be consistent with the track model, this is unlikely the case because (i) this behavior is not

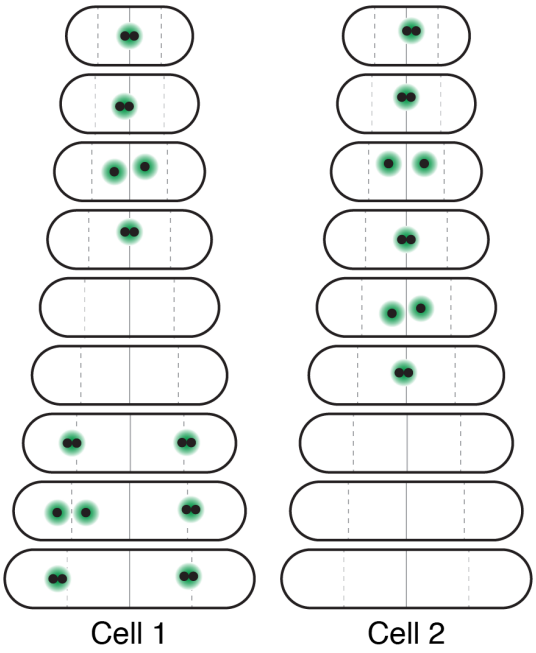

Figure 2.3: **Replisome dynamics in *B. subtilis* and *E. coli* based on snapshot imaging.** Schematic diagram showing full replication cycles in three individual example cells. Green spots represent diffraction-limited replisome foci, while black dots indicate the inferred number of constituent replisomes. Gray vertical lines indicate mid- and quarter-cell positions

.

representative of the replication process as a whole and (ii) these separation events have been previously reported in *B. subtilis* [89, 97] and are replication independent [89]. The structural and functional significance of these transient separations is unclear.

Even the time-lapse approach to studying replisome organization suffers from a number of potential shortcomings: Time-lapse imaging is especially challenging in the current context due to the low stoichiometry of many replisome components [24] which severely limits the number of images that can be captured before protein bleaching. We have therefore focused on characterizing DnaN as a proxy for the localization, rather than characterizing each replisome protein in time-lapse image analysis. Snap-shot images of other components are consistent with our model, but the replisome has only been extensively visualized one component at a time. Studies generally find self-consistency when using different markers for the replisome, however, from existing data it is hard to exclude the possibility that replisome components may not be co-localized throughout the replication process. Finally, fluorescent proteins are known to form aggregates which could lead to anomalous protein localization [98, 99].

2.5 The interpretation of snap-shot images

Although the time-lapse data from DnaN seems to strongly support a stationary factory model, there are reports supporting both the factory and track models in *E. coli* [7, 90–93, 100–103]. Many of these investigations have analyzed the organization of the replisome or the newly replicated DNA (by fluorescence microscopy) using snap-shot imaging (*i.e.* single images). To reconstruct replisome dynamics from the snap-shot data, asynchronous populations of cells are analyzed by length, a proxy for cell age. Alternatively, a sub-culture of a synchronous population can be fixed and visualized at many time points throughout the cell cycle [90]. Changes in the sub-cellular focus localization organization are analyzed relative to cell length. In sufficiently slow growth conditions such that only a single ongoing round of replication is expected, younger cells have a single focus while older cells often have pairs of foci localized to the quarter-cell positions. Two narrowly separated foci are also reported to occur in intermediate aged cells. In retrospect, it is clear that these data are consistent with both

the factory and track model since this approach cannot differentiate replisome splitting from re-initiation, as observed in the time-lapse data.

Snap-shot imaging studies have another shortcoming: Populations of genetically identical cells have a wide distribution of cell cycle lengths, even under the same growth conditions [104]. Furthermore, there is significant variation in cell length at birth and division. As a result, it is impossible to use the cell length as a precise measure of cell age [105], further complicating the snap shot analysis. Although it is more challenging to interpret, we have demonstrated that the factory-like localization pattern can be inferred directly from snap-shot data [7], providing additional support to the factory model from protein foci that are too dim to track throughout the cell cycle.

2.6 A translocating factory model?

A number of more complicated models have been proposed for replisome localization. For instance, in the *translocating factory model*, replisome pairs are localized to midcell for roughly half the replication cycle, before abruptly transitioning to the quarter-cell positions where they remain for the rest of the cell cycle [92, 106–108]. We note that studies citing results consistent with the translocating factory model argue that quarter-cell foci represent single replication forks, however our time-lapse imaging and experiments using the *dnaC2* allele strongly argue that these foci correspond to re-initiated pairs of replisomes [7].

2.7 Coupling or anchoring of the replisomes?

The observed transient separation of replisome pairs suggests that the forks are only weakly associated. If an anchor between replication fork pairs existed, it would need to frequently disassemble, or be of sufficient length to accommodate significant separations. Although it has not been excluded that the replisomes are linked (to each other or the cell), emerging evidence suggests that no functional dependency between the replisomes exists [91, 109]. In *B. subtilis* and *E. coli* there is no evidence excluding the possibility that the replisomes are anchored within the cell. However, in *C. crescentus*, this is clearly not the case since the

location of the replication factory moves during the cell cycle [87].

2.8 Chromosome structure and the factory model

Recent studies of chromosome structure and dynamics also have important implications for the organization of the replication process. We recently analyzed the cell-cycle of dynamics loci from different regions of the *E. coli* chromosome [96]. The factory model makes two closely-related predictions about the locus dynamics: (i) loci should move towards midcell (replisome) before replication and (ii) loci should begin segregation from midcell (replisome). In the track model, loci are stationary and then segregate from their pre-replication positions. Both factory-model predictions (i) and (ii) are strongly supported by a statistical analysis of locus trajectories from hundreds of cells. On-the-other-hand, the predicted locus translocation is *not* obvious from the inspection of a single cell due to fluctuations in the nucleoid structure [96]. Under rapid growth conditions, the structure of the newly replicated nucleoids also is consistent with a centrally located factory complex [110].

Recent work on the mechanism of chromosome structure and segregation in *B. subtilis* is also difficult to reconcile with the track model. It has been proposed that ring-shaped assemblies of SMC are loaded at the origin and slide down the left and right arms of the chromosome to *ter*, drawing the arms together [111]. After replication initiation, this model would act to pull the replisomes together as a consequence of the close proximity between chromosome arms. Since this mechanism of colocalizing the replisome is indirect, it could give rise to the weak association observed in experiments visualizing the replisome.

Although *C. crescentus* replication is also factory-like, the replisome dynamics are altered due to differences in nucleoid structure. Particularly, the origin of replication is proximal to the pole, rather than mid-cell, at the start of replication [87]. The replication factory is displaced towards mid-cell during replication, consistent with the motion being driven by the buildup of newly-replicated DNA at the origin-proximal pole.

2.9 Concluding Remarks

Although analysis of the replisome dynamics is complicated by differences in the chromosome structure between organisms, almost all observations appear to be consistent with DNA loci moving to, and splitting from, a pair of co-localized replisomes, as described by the replication factory model. Although replisome pairs are typically co-localized to the diffraction limit (250 nm), they appear to be weakly associated since they can transiently separate. We propose that the conflicting reports on the structure are the result of the misinterpretation of pairs of replisomes that re-initiate at the quarter cell position. The inconsistencies between reports speaks to the generic importance of both performing time-lapse imaging and the analysis of a large number of cells over complete cell cycles.

Although much has been learned about *in vivo* replisome structure, many important questions remain. For instance, it is unclear whether there is any direct interaction between replisomes when they appear co-localized to the diffraction limit. Super-resolution imaging could provide interesting insights into this question. Furthermore, although it has long been assumed that core replisome components remain stably-bound and co-localized throughout the replication process, this model is being challenged experimentally [17, 34, 112, 113]. There is great potential for future work to offer new and fundamental insights into the replication process in the bacterial cell.

2.10 Acknowledgements

The authors are grateful to Colin LaMont for helpful comments on the manuscript.

Chapter 3

THE REPLISOMES REMAIN SPATIALLY PROXIMAL THROUGHOUT THE CELL CYCLE IN BACTERIA

This chapter is a reproduction of Ref. [7]

3.1 Abstract

The positioning of the DNA replication machinery (replisome) has been the subject of several studies. Two conflicting models for replisome localization have been proposed: In the *Factory Model*, sister replisomes remain spatially co-localized as the replicating DNA is translocated through a stationary *replication factory*. In the *Track Model*, sister replisomes translocate independently along a stationary DNA track and the replisomes are spatially separated for the majority of the cell cycle. Here, we used time-lapse imaging to observe and quantify the position of fluorescently labeled processivity-clamp (DnaN) complexes throughout the cell cycle in two highly-divergent bacterial model organisms: *Bacillus subtilis* and *Escherichia coli*. Because DnaN is a core component of the replication machinery, its localization patterns should be an appropriate proxy for replisome positioning in general. We present automated statistical analysis of DnaN positioning in large populations, which is essential due to the high degree of cell-to-cell variation. We find that both bacteria show remarkably similar DnaN positioning, where any potential separation of the two replication forks remains below the diffraction limit throughout the majority of the replication cycle. Additionally, the localization pattern of several other core replisome components is consistent with that of DnaN. These data altogether indicate that the two replication forks remain spatially co-localized and mostly function in close proximity throughout the replication cycle. The conservation of the observed localization patterns in these highly divergent species suggests that the subcellular

positioning of the replisome is a functionally critical feature of DNA replication.

3.2 Author Summary

Cell proliferation depends on efficient replication of the genome. Bacteria typically have a single origin of replication on a circular chromosome. After replication initiation, two replisomes assemble at the origin and each copy one of the two arms of the chromosome until they reach the terminus. There have been conflicting reports about the subcellular positioning and putative co-localization of the two replication forks during this process. It has remained controversial whether the two replisomes remain relatively close to each other with the DNA being pulled through, or separate as they translocate along the DNA like a track. Existing studies have relied heavily on snapshot images and these experiments cannot unambiguously distinguish between these two models: i.e. two resolvable forks versus two pairs of co-localized forks. The ability of replication to re-initiate before cell division in bacterial cells further complicates the interpretation of these types of imaging studies. In this paper, we use a combination of snapshot imaging, time-lapse imaging, and quantitative analysis to measure the fraction of time forks are co-localized during each cell cycle. We find that the forks are co-localized for the majority (80%) of the replication cycle in two highly-divergent model organisms: *B. subtilis* and *E. coli*. Our observations are consistent with proximal localization of the two forks, but also some transient separations of sister forks during replication. The conserved behavior of sub-cellular positioning of the replisomes in these two highly divergent species implies a potential functional relevance of this feature.

3.3 Introduction

Rapid and faithful replication of the chromosome is essential for the proliferation of all living cells. Many bacteria possess a single circular chromosome. Replication is initiated at a single origin and two multicomponent protein complexes (replisomes) replicate bidirectionally around the chromosome, meeting at the terminus. Previous investigations of replisome localization have advocated two distinct models: In the *Factory Model*, the replication machineries at

sister forks remain relatively stationary and spatially proximal forming a factory through which the replicating DNA is pulled [40, 87, 88, 96, 102, 103]. Alternatively, in the *Track Model*, the replication machineries translocate along a stationary DNA track, resulting in significant separations of the forks as replication progresses to sequences genomically distal from the origin of replication (*oriC*), but re-localize as the forks converge at the terminus [90, 91].

Although previous fluorescence microscopy studies have already reported on replisome localization, a significant weakness of these studies is that the statistical analysis relied heavily on still images (snapshot data). As we will describe in this paper, without direct knowledge of the replication dynamics, it is difficult to differentiate between two pairs of co-localized replisomes that form as the result of re-initiation of replication and spatially-separated sister replisomes. In fact, both structures are observed in both *E. coli* and *B. subtilis*. Furthermore, the lengths of the replication and division cycles are highly variable in individual cells [104], creating the need for large-scale automated analysis which produces consistent results between time-lapse and snapshot data.

Our method of analysis is based on the localization of the beta-clamp (DnaN) in both *E. coli* and *B. subtilis*. DnaN is a particularly informative proxy for the replisome complex because it is localized to the replisome [114] at sufficiently high copy-number that its position can be observed clearly throughout the cell cycle without significant photobleaching of the fluorescent label. Using time-lapse data, we tracked the progress of the replisome in individual cells over multiple cell cycles. Under slow growth conditions, the initiation and termination of replication can be observed explicitly by the assembly and disassembly of the DnaN foci.

Analysis of the time-lapse data reveals a significant confounding feature in the analysis of snapshot images. For most of the cycle, sister replication forks maintain a sufficiently small separation such that the two foci cannot be resolved, but two sister forks are transiently resolvable in some cells, consistent with previous reports [89, 97]. Irrespective of whether forks are co-localized or spatially resolved, they typically remain localized around midcell. This story is complicated by the phenomenon of re-initiation of replication before the end of

the cell cycle [103]. Even under slow growth conditions, re-initiation pre-division is observed in some cells. This uncoupling between the replication and division cycles leads to the appearance of DnaN foci localized to the quarter-cell positions. Although this phenomenon is clearly observable in the time-lapse analysis, these foci can easily be misinterpreted as sister replication forks in snapshot analysis. However, careful statistical analysis of the snap-shot data clearly resolves two distinct subpopulations (resolved-sisters and colocolized replisome pairs). Our analysis of snapshot images produces results consistent with the replisome dynamics observed in time-lapse imaging, and can be applied to lower stoichiometry replisome components, for which we find similar results. Therefore the analysis of both time-lapse and snapshot images supports a model for replisome positioning in *E. coli* and *B. subtilis* where the sister replisomes localize with diffraction-limited separation for the majority of the cell cycle.

3.4 Results

3.4.1 Time-lapse imaging of the replisome reveals proximal positioning

Replisome positioning was observed using time-lapse fluorescence microscopy by imaging fluorescent fusions to DnaN in both *E. coli* and *B. subtilis*. Cells were elongating exponentially with a doubling time of roughly 3 hours (Fig. 3.1), and multiple complete cell cycles were tracked under these relatively slow growth conditions. It is widely accepted that the DnaN focus is localized to the replisome [114] and a good proxy for replication since: the beta-clamp is essential for replication, the focus is observed with midcell positioning (consistent with other replisome components), and the timing of assembly and disassembly is consistent with the known duration of replication.

A qualitative summary of the observed dynamics (see Fig. 3.2A) is as follows: In the absence of replication, there is no DnaN focus whereas actively replicating cells generally have between zero and four DnaN foci. DnaN foci tend to be localized either to midcell or the quarter-cell positions, as has been reported previously [114, 115] and is consistent with

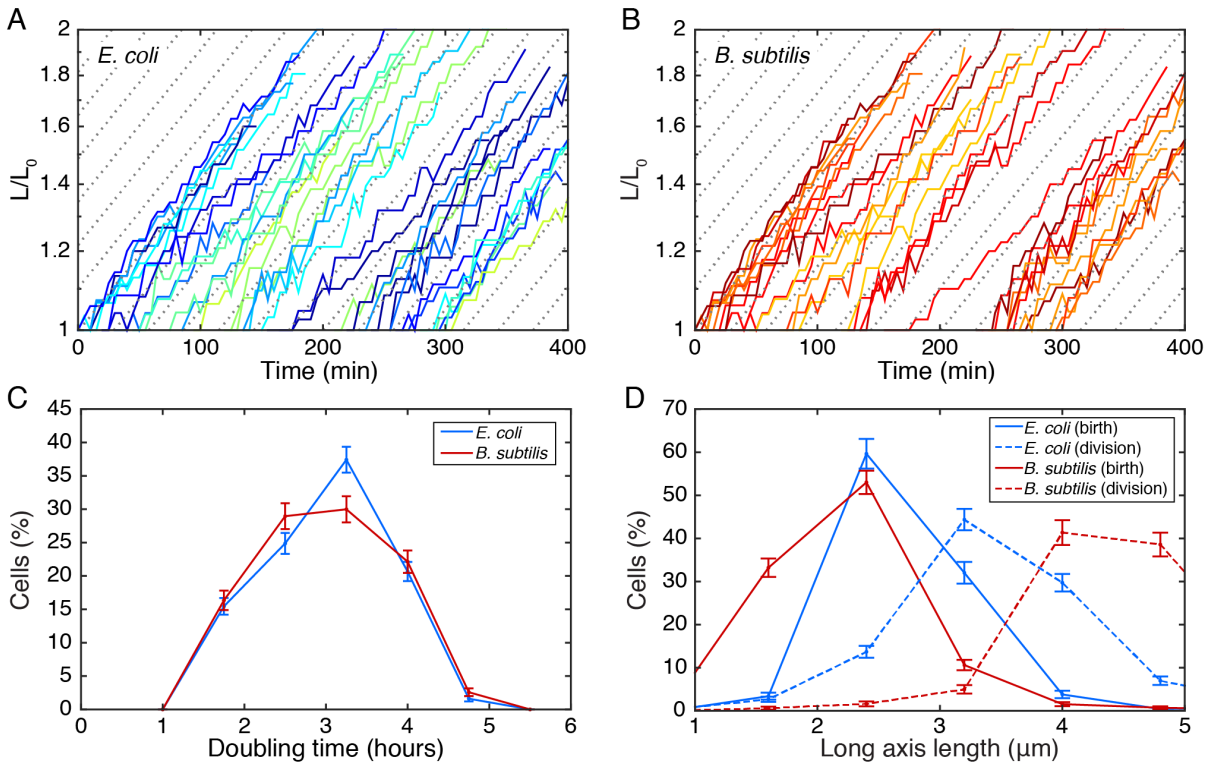

Figure 3.1: **Cell growth.** A-B: Typical length vs. time curves in *E. coli* and *B. subtilis*. Cell growth rate does not slow down significantly over time indicating that exposure to the laser is not damaging the cells. C: Distribution of doubling times on microscope slide for *E. coli* and *B. subtilis*. D: Distribution of cell lengths at birth (immediately following division of parent cell) and division.

the localization of other replisome components. There is typically no focus or one focus at the midcell position. The typical temporal history of the position of a DnaN focus under slow growth conditions is concisely summarized by the kymograph in Panel B of Fig. 3.2: A focus appears. The position of the focus executes confined random motion. In some cells, the focus fissions into two dim foci which then fuse to re-form a single focus (of the initial intensity). The focus is then observed to disassemble. Typically there is a short period between the disappearance of the focus at midcell and either the roughly synchronous appearance of new foci at the quarter cell positions or cell division. DnaN foci were not observed to be intermittent: Once DnaN foci assembled they did not disappear and re-appear at the same cellular location.

It is important to note the following qualifications about the number of DnaN foci: When the separation of foci are below the diffraction limit of our system (< 250 nm), only one focus will be resolved. Even when forks transiently separate enough to be resolvable, they remain well within the quarter cell positions: The average separation (when two foci are observed) is 0.2 cell lengths versus 0.45 cell lengths for foci near the quarter-cell positions.

3.4.2 Quarter cell foci are re-initiated replisome pairs

Single DnaN foci positioned at midcell are seen to fission and fuse (e.g. Fig. 3.2A, arrow 1). These midcell-positioned foci are always seen to disappear before cell division. On-the-other-hand, foci observed at the quarter-cell positions can persist through a cell division, consistent with these foci representing re-initiation of replication. If instead these foci were separated pairs of sister forks, they would be expected to co-localize at the terminus and disassemble before the end of the cell cycle. Therefore qualitative analysis of the kymographs strongly supports a model where quarter-cell-localized foci each include pairs of re-initiated sister replisomes. In fact, these quarter-cell foci are also seen to fission and fuse, occasionally allowing resolution of the individual sister replisomes (e.g. Fig. 3.2A, arrow 2).

To further test this model, we tracked DnaN foci in *E. coli* cells blocked for restart via a temperature sensitive version of the helicase loader protein, DnaC (*dnaC2* allele) [91].

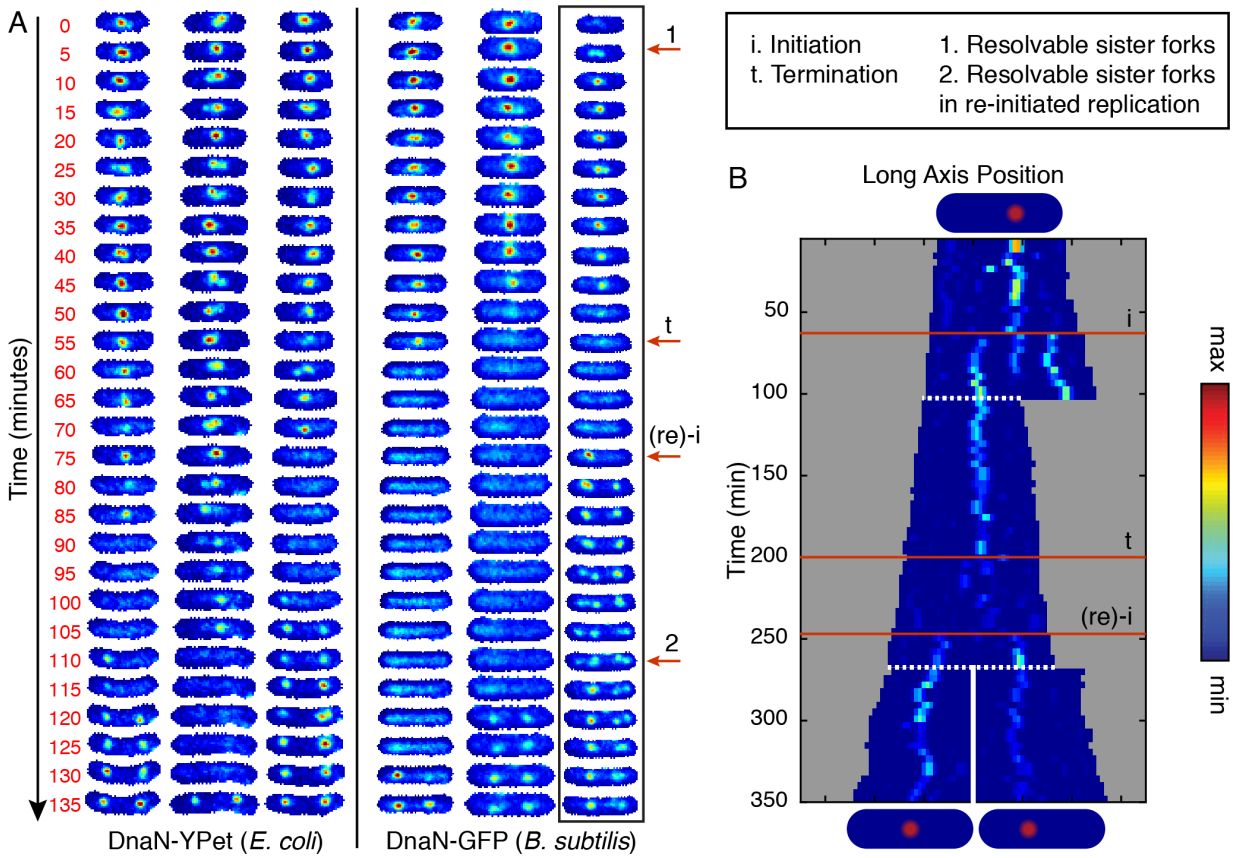

Figure 3.2: **Visualizing replication forks in individual cells through complete cell cycles.** A: Six representative examples of individual cells imaged at 5 minute intervals for DnaN-YPet in *E. coli* (left) DnaN-GFP in *B. subtilis* (right). Cells are tracked for complete cell cycles although images were cropped by up to a few frames to make the lengths consistent. Labeled red arrows point to example features in the boxed image strip. Starting at the beginning of the cell cycle, there is generally a single midcell focus representing both replication forks. However, occasionally sister forks can be resolved separately (e.g. arrow 1) but co-localize before termination of replication (e.g. arrow t). For a period of time, which varies cell to cell, no foci are observed until re-initiation on the newly replicated sister chromosomes (e.g. arrow (re)-i), an event which often happens before cell division. These new foci appearing at the quarter cell positions are consistent with replication factories since

they can occasionally be resolved into sister replication forks (e.g. arrow 2). See also Fig. 3.3 for additional full cell cycle images. B: Example single-cell kymograph spanning multiple cell divisions for DnaN-YPet in *E. coli*. Cell images are projected along the long axis of the cell and stacked in sequence. White dashed lines indicate division events. Numbered red lines indicate initiation, termination, and re-initiation, respectively.

Under the non-permissive conditions for the temperature sensitive mutant, the wild type cells were able to form quarter-cell-localized foci, however, the cells blocked for initiation were not (compare Fig. 3.4 panels A and C). To extend this analysis to many cells, we show conditional probability distributions of focus position given cell length in both the wild type and *dnaC2* mutants. The absence of localizations near the quarter-cell positions is clearly seen by comparison of Fig. 3.4, panels B and D. These data support our model that quarter-cell foci represent re-initiated replication fork pairs.

3.4.3 Replication and division timing is asynchronous

In the event that re-initiation of the sister chromosomes happens before cell division (about 45% of the time under our conditions), we can only observe complete replication cycles if we analyze overlapping cell cycles. We visualize entire replication cycles using kymographs, where we project the cell images onto the long axis of the cell, and align the projections in sequence (See Fig. 3.2, Panel B). This representation confirms that for the majority of the replication cycle, the sister forks remain near mid-cell and usually cannot be resolved separately.

Since the timing of division is inferred from the analysis of the phase-contrast image of the cell, some of the observed asynchrony could be accounted for by a failure to correctly segment the septum. However two lines of evidence refute this hypothesis: (i) Our previous work analyzing the cell-cycle dependent localization of FtsZ suggests that the timing of division

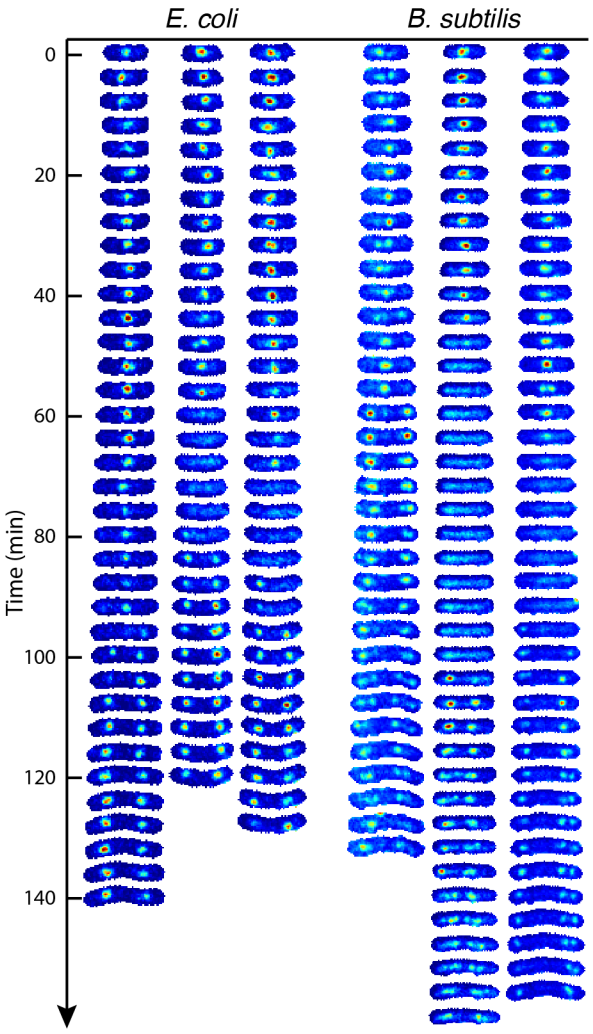

Figure 3.3: **Additional cell towers.** Additional example cell towers for both *B. subtilis* and *E. coli*.

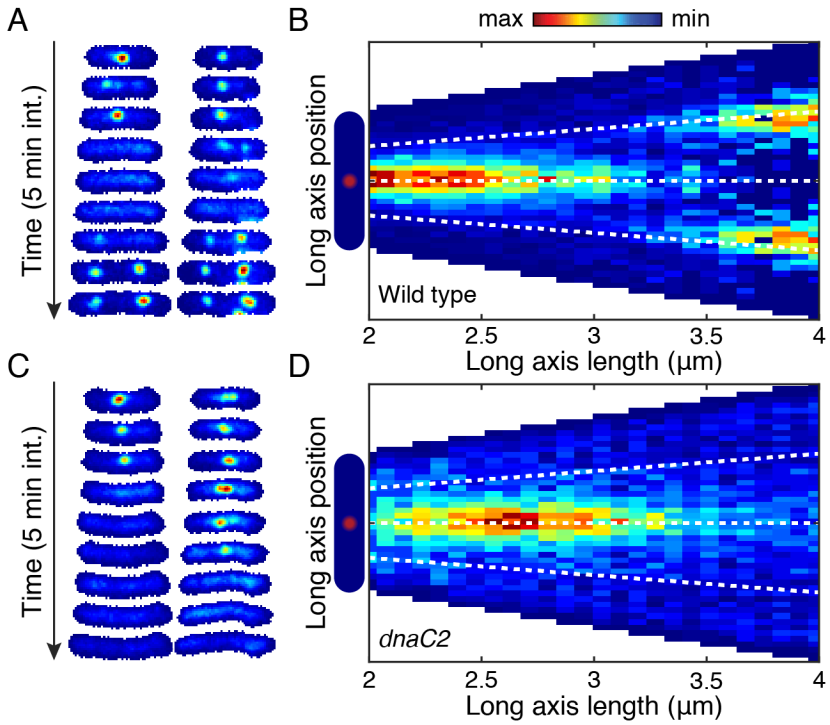

Figure 3.4: **Blocked initiation leads to loss of quarter cell foci.** DnaN-YPet (in *E. coli*) is imaged in both in wild type and cells containing the *dnaC2* allele at 37°C. Under these conditions, cells containing the mutant allele will be blocked from initiating new rounds of replication. A) Example wild type cell towers showing the the disappearance of the midcell focus may be followed appearance of a pair of foci near the quarter-cell positions. B) Conditional probability distribution (N = 4837 time points) shows localizations near the quarter cell positions in the wild type near the end of the cell cycle. C) Example cell towers for cells with blocked initiation do not show foci at the quarter-cell positions after disappearance of the mid-cell focus. D) In cells blocked for initiation, conditional probability no longer shows a significant number of localization at the quarter-cell positions (N = 1758 time points).

is determined to a precision better than ±10% of the cell cycle in *E. coli* [116]. (ii) In this study, we never observed the midcell DnaN focus persist through cell division, consistent with accurate determination of cell division.

3.4.4 Proximal replisome positioning is observed in snapshot analysis

Due to significant cell-to-cell variation, it is essential to present statistical evidence for the classification of each focus as an individual replisome or co-localized replisome pair. We apply a fully automated analysis to characterize localization patterns for over 10,000 time points (detailed description included in the materials and methods section). For this analysis, foci are identified in the fluorescence image (Fig. 3.5) and precisely located within the cell. It is important to note that there was no hand selection of data. The dataset consists of all cells observed that were elongating and segmented without errors and therefore contains no investigator-based cell-selection bias.

We first investigate the focus positioning relative to cell length. In this analysis, cell length was used as a proxy for cell age because of *B. subtilis* chaining which makes detection of septum formation in phase images difficult. Cell length provided a consistent and reliable proxy for cell age in our analyses across both species and cell-length based analysis can be applied in snapshot analysis where the cell age is unknown. The conditional probability density of focus position given cell length is shown in Fig. 3.6.

The Factory Model predicts that foci are localized at midcell throughout the replication cycle and that re-initiation occurs at the quarter cell positions (if it occurs). In terms of the conditional probability, this model would predict a density blob at midcell which persists from early until late in the cell cycle. A second pair of blobs are expected to form at the quarter-cell positions corresponding to re-initiation in some cells. This second pair of blobs is expected to persist into the next cell cycle, each ending up in a new cell without the other. In contrast, in the Track Model, a blob is expected to begin the cell cycle at midcell, before splitting into two blobs (if the separation is high enough) before merging into a single blob again at midcell. The conditional probability data alone favor a factory-like model, but do

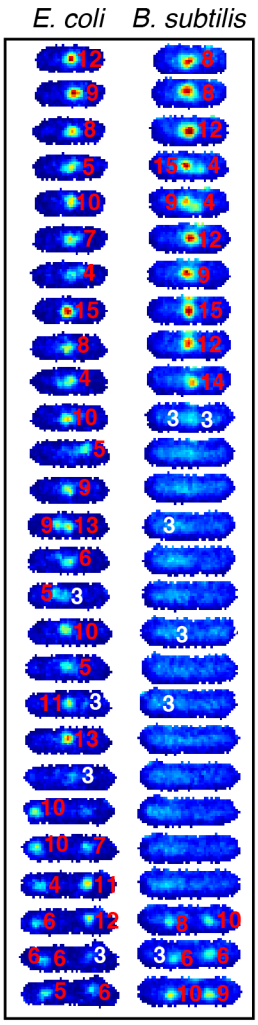

Figure 3.5: **Automatic focus identification using scoring.** Example scored foci in partial cell towers. Frame delay is five minutes. Scores are printed in red for foci scoring high enough to be included in analysis. Foci scoring 3 or lower (white) appeared randomly throughout the cell and were excluded.

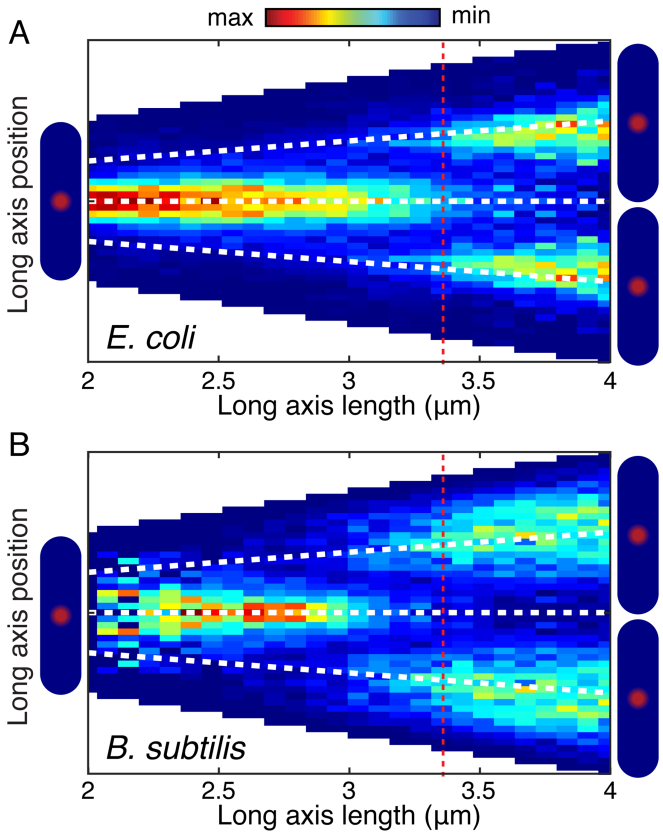

Figure 3.6: **Focus localization pattern is similar in *E. coli* and *B. subtilis*.** Probability of focus localization at a relative long axis position is shown as a function of cell length (analysis for over 10,000 foci in each organism). Dashed red line indicates mean length at re-initiation. A: DnaN-YPet in *E. coli* most probably localizes midcell until re-initiation when foci appear at the quarter-cell positions. Dashed white lines indicate mid- and quarter-cell positions. B: DnaN-GFP in *B. subtilis* shows a similar localization pattern to DnaN-GFP in *E. coli* (Panel A).

not exclude the possibility of a Track Model, provided the separation of the forks remains very small as the replisomes translocate along the DNA.

Further analysis is informed by the typical cellular focus localization patterns shown in Fig. 3.7A. The relative frequencies of these localization patterns were quantified both overall, and by cell length. We find that short/young cells are generally observed to have a single focus. As the cells grow, the probability of observing two foci monotonically increases. In the Track Model, we expect the total number of two-foci cells to grow significantly and then shrink as the forks first diverge from the origin and then reconverge at the terminus. This trend is not observed. Furthermore, before cells grow to a length where two foci is most probable, there is an increase in the probability of zero-foci cells, consistent with the Factory Model where two-foci cells have re-initiated but inconsistent with the Track Model where cells should transition from one to two foci without foci disassembling. Again, a Factory-like Model best summarizes the observed data.

To automatically distinguish between the observed cellular localization patterns, it is essential to make a distinction between co-localized replication fork pairs and resolvable sister replication forks. Motivated by the qualitative observation that resolved sister forks rarely separate more than 0.2 cell lengths, we examine the joint probability distribution between focus separation (relative to cell length) and cell length shown in Fig. 3.7, Panel B. The joint distribution reveals that the population consists of two distinct sub-populations: In short/young cells, resolved foci are on average 0.2 cell-lengths in separation whereas in long/old cells, resolved foci are typically 0.45 cell-lengths in separation, consistent with quarter-cell positioning. The subpopulation model facilitates the identification of resolvable sister replication fork pairs. All pairs of foci whose separation is consistent with the lower-separation population are counted as resolvable sister replication forks, whereas all other foci (including foci in single-focus cells) are inferred to represent a pair of co-localized replication forks (See Fig. 3.9 and the associated methods section for more detail).

Considering cells of all lengths, we find that replication forks co-localize about 82% of the time for DnaN in *E. coli* (Fig. 3.7C). Statistical analysis was performed similarly in *B. subtilis*,

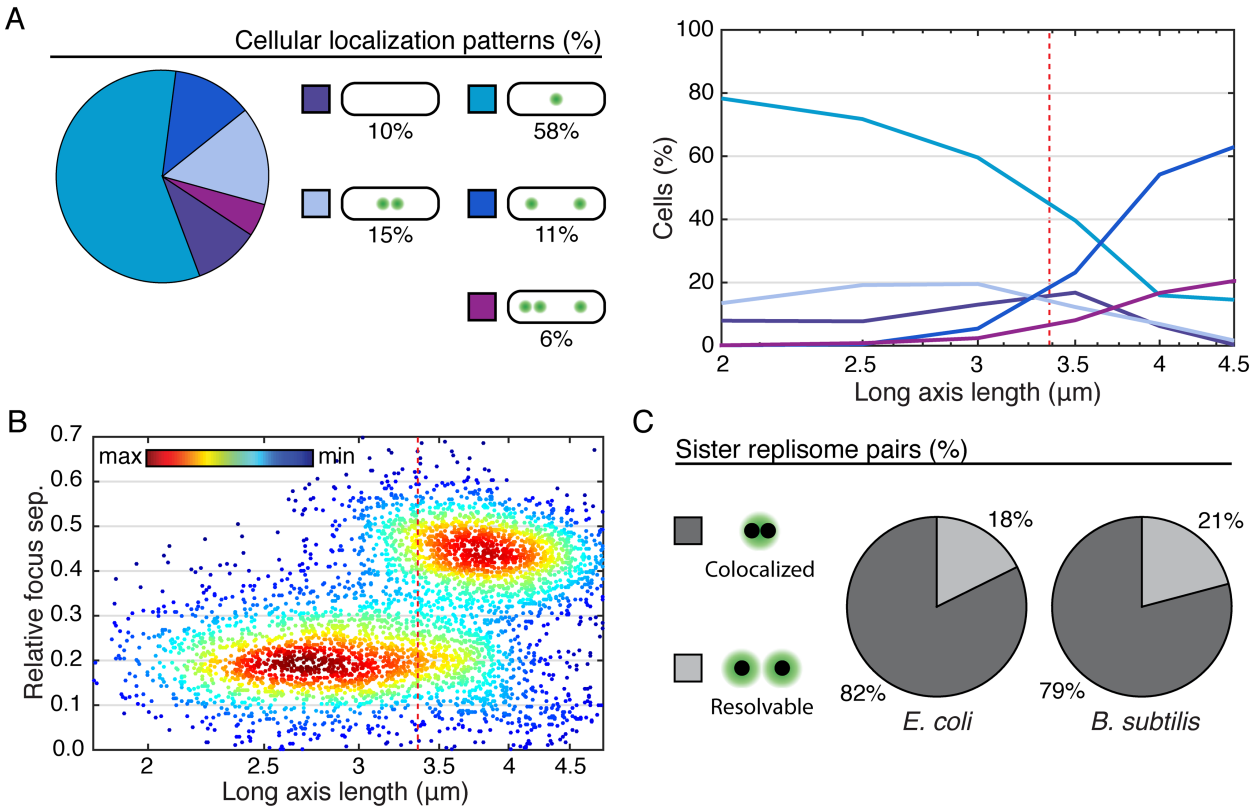

Figure 3.7: **Characterizing cellular and subcellular localization patterns in large populations (at least 10,000 time points).** A) Left: Typical cellular localization patterns and their relative frequency of occurrence (in *E. coli*, see Fig. 3.8 for corresponding *B. subtilis* data) are shown for all cells regardless of length. Extremely rare localization patterns (4+ foci) are not shown. Right: Relative frequency of the same typical cellular localization patterns as a function of cell length (age). Dashed red line represents mean cell length both at re-initiation and division. B) Multiple focus cellular localization patterns are distinguishable automatically based on inter-focus separations. Scatter plot shows relative focus separation as a function of cell length for multiple focus cells. Two well resolved populations are observed. Pairs of foci whose separation falls into the lower population are counted as individual members of a sister fork pair, whereas all other foci are each inferred to represent a co-localized replication fork pair. Dashed red line represents mean cell length at re-initiation and division. Example

shown is for DnaN-YPet foci in *E. coli*. See also Fig. 3.8 for corresponding B. subtilis data. C) Pie charts showing the fraction of sister replication fork pairs that are co-localized (replisomes have diffraction-limited separation) versus resolvable (each replisome appears as a member of a narrowly separated focus pair). In both organisms, replication fork pairs remain co-localized roughly 80% of the time, consistent with a factory-like model.

but because *B. subtilis* has a tendency to chain, division events are not always observable, at least not at the time of their occurrence. Since the division is not detected, more long cells that have re-initiated are observed. The exaggerated size of the high-separation population (Fig. 3.8) relative to the corresponding *E. coli* data is consistent with this known artifact. We also find that there is a longer time between termination of replication and re-initiation of replication on the sister chromosomes (D phase/G1 and G2 phase), increasing the fraction of zero focus cells. Strikingly, we find that the forks co-localize with 79% probability, roughly the same as observed in *E. coli* (Fig. 3.7C).

We applied the same statistical analysis used to quantify DnaN dynamics to snapshot images of three independent markers for the replisome (Fig. 3.10 and Fig. 3.11). Although DnaN is the only replisome component present at sufficiently high copy number to be imaged throughout entire cell cycles, the analysis we developed can be used to infer the dynamics of lower stoichiometry proteins based on snapshot images. We use SSB (single-stranded binding protein) and DnaQ (PolIII subunit) in *E. coli* and DnaX (clamp loader) in *B. subtilis* as additional markers for the replisome. We again find that separated foci can be classified into two populations, one representing separated sister replisomes, and the other representing pairs of co-localized forks (Fig. 3.11, panel A). We count the number of foci in each cell (Fig. 3.11, panel B) and classify each focus as an individual replisome or replisome pair. It is important to note that the number of zero-focus cells is not particularly meaningful in this context as it depends on the growth rate of individual cells, and in particular, non-replicating

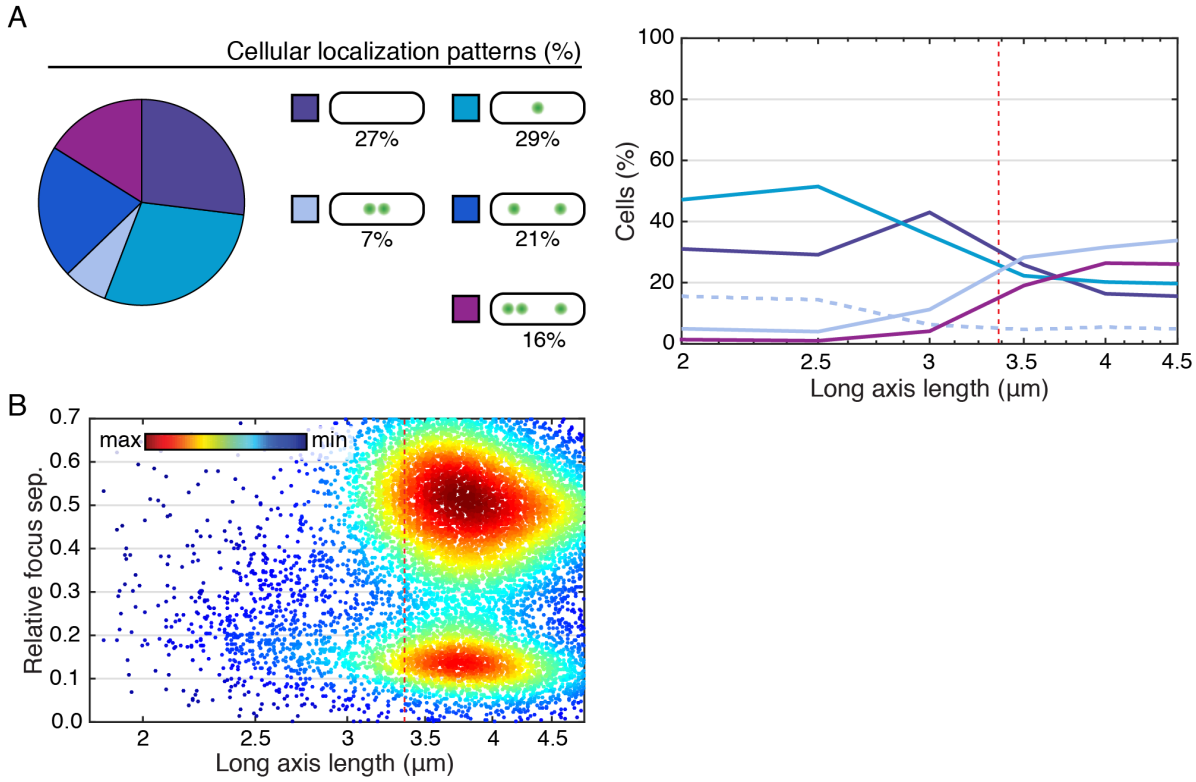

Figure 3.8: **Supporting *B. subtilis* data.** A) Relative frequencies of typical localization patterns overall (left) and separated by cell length (right). Dashed red line represents mean length at re-initiation B) Joint probability distribution for interfocus separation is used to automatically distinguish between cellular localization patterns. Dashed red line represents mean length at re-initiation.

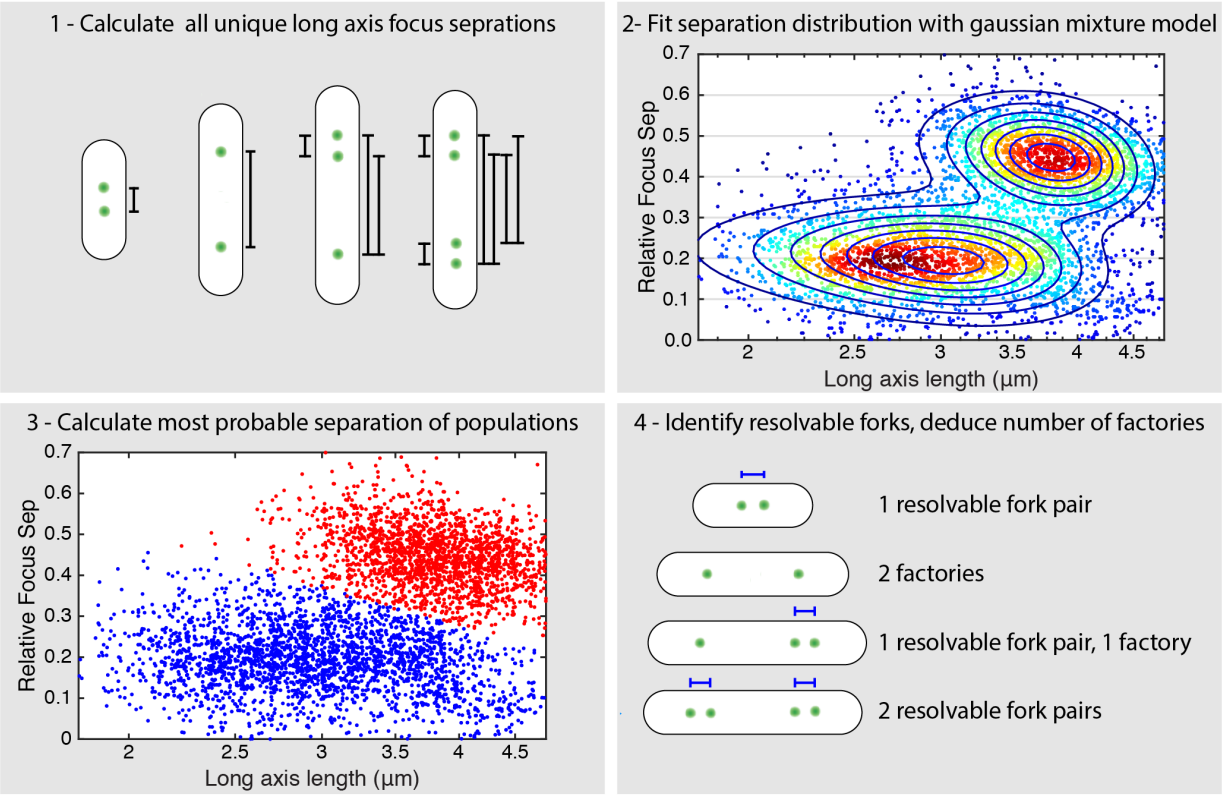

Figure 3.9: **Process for distinguishing resolvable forks from pairs of factories in cells with multiple foci.** 1) All possible long axis separations are calculated (black brackets). 2) Distribution of focus separation as a function of cell length (dots) is fit with a two Gaussian-mixture model (dark blue contours) using maximum likelihood. We note that the fit to the lower population is shifted slightly due to the tail, however the populations separate visually correctly. 3) Using the Gaussian mixture model obtained in step 2, each focus pair is classified as a member of the high (red) or low (blue) separation population. 4) Focus pairs that are determined to be members of the low-separation population (blue brackets) are classified as each representing an individual replisome. All other foci are inferred to be a co-localized replisome pair.

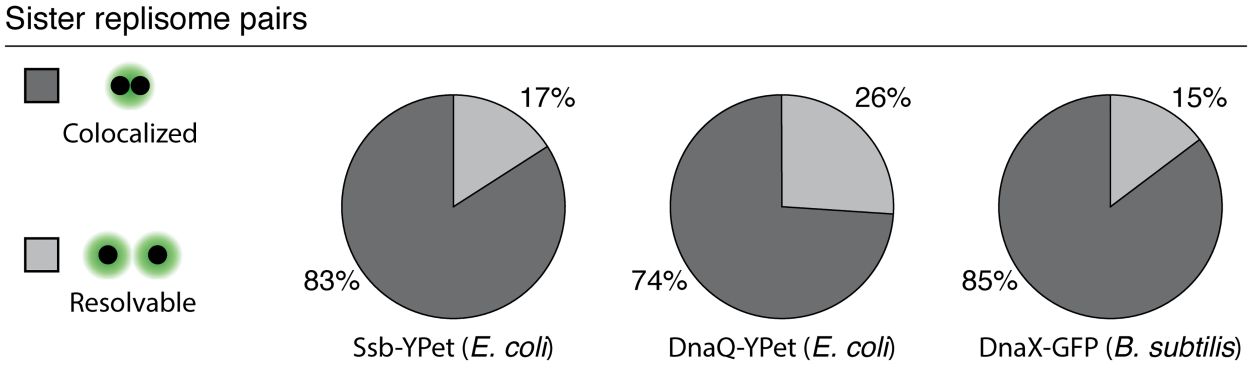

Figure 3.10: **Additional markers for the replisome show consistency with DnaN.** Pie charts showing the fraction of sister replication fork pairs that are co-localized (replisomes have diffraction-limited separation) versus resolvable (each replisome appears as a member of a narrowly separated focus pair) in SSB-YPet (N = 6493 Cells), DnaQ-YPet (N = 2187 cells), and DnaX-GFP (N = 10573 cells). Results are consistent with DnaN time-lapse imaging. See also Fig. 3.11 for complete analysis.

cells cannot be excluded based on snapshot images. We find that sister replisomes co-localize with 74-85% probability (Fig. 3.10), consistent with our observations based on DnaN.

3.5 Discussion

3.5.1 Sister replisomes have proximal positioning

Both time-lapse and snapshot analysis strongly support a model where the two replication forks remain proximal throughout the cell cycle. At the beginning of the replication cycle, we always observe a single focus. We associate this event with co-localization of the forks at or near *oriC*. After initiation, the replisomes remain well within the quarter cell positions, usually co-localizing, but separating enough to be individually resolvable approximately 20% of the time. The observed midcell positioning of the replisome is consistent with previous reports from our lab suggesting that chromosomal loci in *E. coli* are localized near midcell immediately prior to duplicating [96]. Together this study and the aforementioned previous

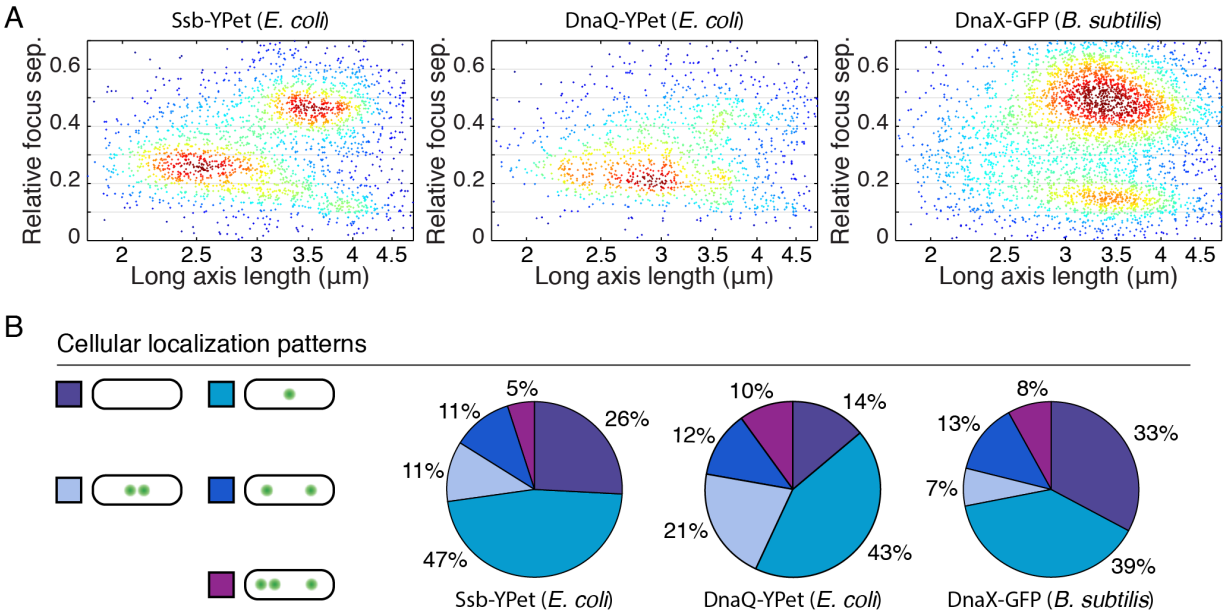

Figure 3.11: **Snapshot analysis of several replisome proteins implies dynamics consistent with observations in time-lapse imaging.** A) Joint probability distributions for interfocus separation are used to automatically distinguish between cellular localization patterns in SSB-YPet (N = 6493 Cells), DnaQ-YPet (N = 2187 cells), and DnaX-GFP (N = 10573 cells). B) Relative frequencies of typical localization patterns.

report suggest movement of the DNA through a relatively stationary replisome.

The occasional separation of sister replisomes suggests that, while the Factory Model correctly predicts the cellular-scale positioning of forks in both *E. coli* and *B. subtilis*, some elements traditionally associated with the Track Model are also correct in the sense that the sister forks need not be continuously co-localized. However, this must be qualified by noting that the vast majority of sister replisomes do appear to be co-localized by conventional fluorescence microscopy, and it is unlikely that this would be the case if the replisome translocating along the DNA were the complete model. Under our conditions, at least 80% of observed foci are in fact pairs rather than individual replisomes. This quantitative picture will be essential in interpreting short-time scale experiments which do not capture the cell cycle dynamics.

Our data strongly support a factory-like model, but do not address whether there is indeed a "factory." In the strictest definition of a factory, sister replisomes are coupled by a physical linker. While we do not exclude the possibility of a physical linker, such a linker would need to allow the replisomes to occasionally uncouple, or be sufficiently long to account for the observed separation events. The physical mechanism by which sister replisomes remain spatially proximal is an intriguing problem for future studies.

3.5.2 Two types of focus separation and dynamics

We see two distinct types of separation and dynamics. In the first class of separation, the relative distance between the foci is under a third of the cell length, and is about a fifth of the cell length on average. Foci with these small separations always merge. In contrast, if re-initiation precedes division, we observe half-cell-length focus separation. In addition to having a larger separation, these foci are never observed to merge and ultimately end up in different cells. Furthermore, widely-separated foci are unable to form in conditional mutants where replication restart is prevented. This localization behavior indicates that the half-cell-length foci we observe pre-division are new pairs rather than individual replisomes. These observations are summarized schematically in the model cell tower and kymograph

shown in Fig. 3.12.

3.5.3 Replisome positioning dynamics are conserved in Gram-positive and Gram-negative bacteria

Factory-like models where the replisomes remain relatively confined (but are not necessarily constantly co-localized or physically coupled) have been previously suggested in both *B. subtilis* [34, 40, 88, 89] and *C. crescentus* [87], while the Track Model was formulated based predominantly on data from *E. coli* [91, 104]. The reason for the existence of the outlying model in *E. coli* was often attributed to differences between the species. However, our comparison of replisome dynamics in *E. coli* and *B. subtilis* suggests that the conflicting models likely originated from differences in analysis and interpretation of data rather than species-specific replisome dynamics. The evidence presented here unifies the models for DNA replication in all three of these bacterial organisms.

In every stage of our analysis, the replisome positioning data reveals striking similarities between *E. coli* and *B. subtilis*. Gram-negative and Gram-positive bacteria are highly divergent. Although the basic principles of replication are conserved, many differences exist between *E. coli* and *B. subtilis* replication, including regulation, the leading and lagging strand polymerases, and the loading of the replicative helicase. The data presented here reveal that replisome positioning is one of the conserved features, suggesting that there may be fundamental mechanistic reasons for precise subcellular localization of this complex. Future studies should reveal the mechanism by which the two forks remain proximal throughout the replication cycle and identify the underlying reasons (if any) for this conserved localization pattern.

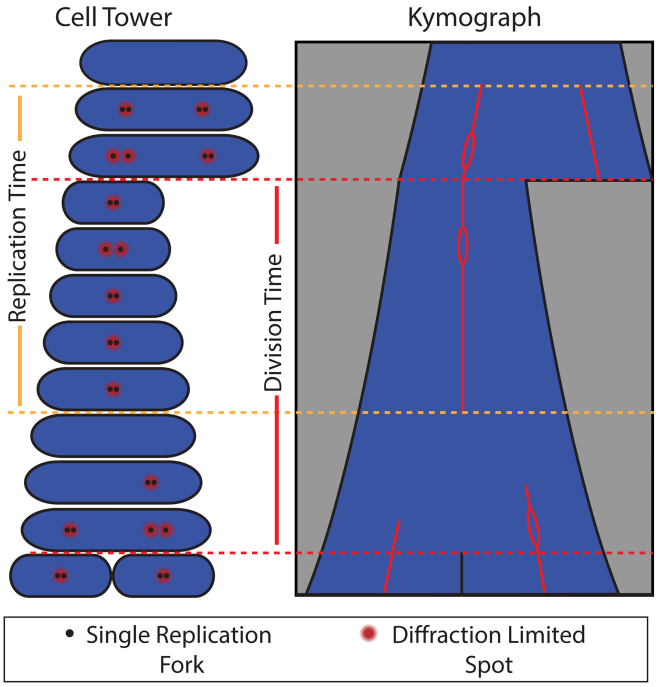

Figure 3.12: **A schematic model for replication fork localization during complete cell cycles.** Both the kymograph and cell tower representations are shown. Our model for replication fork localization is as follows: (i) The replication and division cycles are generally out of sync, so that full replication cycles occur across multiple division cycles. (ii) During replication, forks co-localize to a diffraction-limited spot about 80% of the time, consistent with factory-like behavior. (iii) When the forks are individually resolvable, separation is small (< 0.33 cell lengths) and transient. (iv) Termination of replication is indicated by disappearance of foci. (v) Re-initiation often occurs before cell division near the quarter-cell positions. Because these foci occasionally split into pairs of resolvable foci, they are consistent with factories containing two replication forks.

Table 3.1: **Strain List.**

Strain No.	Genotype	Species	Reference
PAW914	*ypet-dnaN*	*E. coli*	[24]
PAW885	*dnaN-gfp*	*B. subtilis*	[114]
PAW542	*dnaC2 thrA::Tn10*	*E. coli*	[91]
PAW1181	*dnaC2 ypet-dnaN*	*E. coli*	This study
PAW843	*ssb-ypet*	*E. coli*	[24]
PAW913	*dnaQ-ypet*	*E. coli*	[24]
PAW884	*dnaX-gfp*	*B. subtilis*	[114]

3.6 Materials and Methods

3.6.1 Strain construction and growth

See Table 3.1 for strain list. PAW1181 was produced by P1 transduction of the mutation from PAW542 into PAW914. Cells were cultured overnight at 30°C in minimal medium with shaking. Prior to imaging, cells were set back to OD_{600} 0.1 and allowed to grow to OD_{600} 0.3. To ensure sufficiently slow growth that initiation would happen only once per-division cycle, we use minimal medium supplemented with only the essential nutrients. For *E. coli*, cells were cultured in M9-minimal medium (1X M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% Glycerol, 100 µg/ml each Arginine, Histidine, Leucine, Threonine and Proline and 10 µg/ml thiamine hydrochloride). *B. subtilis* was cultured in Minimal Arabinose Medium (1x Spitzizens salts (3 mM $(NH_4)_2SO_4$, 17 mM K_2HPO_4, 8 mM KH_2PO_4, 1.2 mM $Na_3C_6H_5O_7$, 0.16 mM $MgSO_4$-$(7H_2O)$, pH 7.0), 1x metals (2 mM $MgCl_2$, 0.7 mM $CaCl_2$, 0.05 mM $MnCl_2$, 1 µM $ZnCl_2$, 5 µM $FeCl_2$, 1 µg/ml thymine-HCl), 1% arabinose, 0.1% glutamic acid, 0.04 mg/ml phenylalanine, 0.04 mg/ml tryptophan, and as needed 0.12 mg/ml threonine).

3.6.2 Microscopy slide preparation

For imaging, we gently heat the appropriate growth medium with 2% by-weight low melt agarose (Fisher: 16520050). The agarose mixture is then molded into a thin rectangular strip (2 mm × 4 mm × 0.05 mm) and allowed to dry. One micro-liter of OD_{600} 0.3 liquid culture is spotted centrally on the pad. Once the spot has dried, a cover glass is placed over the pad and sealed around the edges with VaLP (1:1:1 Vaseline, Lanolin, and paraffin mixture). This leaves a small channel of air around the pad, particularly important for the growth of *B. subtilis*.

3.6.3 Microscope configuration

Imaging was performed on our lab-built inverted fluorescence microscope. Cells were imaged through a Nikon CFI Plan Apo VC 100x 1.4 NA objective. A retractable external phase plate (Ti-C CLWD Ph3 Annulus Module) was inserted into the light path during phase-contrast imaging but removed for fluorescence imaging to avoid decreased signal due to the neutral density annulus on the phase plate.

For fluorescence imaging, we excite GFP and YPet proteins using a Coherent Sapphire 50 mW 488 nm or 150 mW 514 nm CW laser (see Table 3.2). The beam diameter is expanded, providing uniform illumination over the field of view. An Acousto-Optic Tunable Filter (AOTF, AA Opto-Electronic AOTFnC-400.650) controls the laser excitation intensity. Images were collected on an iXon Ultra 897 512×512 pixel EMCCD camera. The microscope system is controlled by Micro-Manager.

Table 3.2: **Microscope Settings.**

Protein	**Laser Power (mw)**	**Exposure (ms)**
GFP	1.1	600
YPet	0.93	600

3.6.4 Imaging

Cells are imaged at about 25°C in both phase contrast (to determine cell boundaries) and fluorescence (to measure replisome position) at five minute intervals. The built-in software autofocus is used to ensure focus at each time point. Laser intensity is maintained as low as possible to avoid bleaching the fluorescent protein and damaging the cells. We are generally able to image cells for about four hours before photobleaching becomes too significant to reliably track the forks. For lower stoichiometry replisome components, snapshot imaging was used where one phase-contrast and one fluorescence image was taken at each field of view.

Non-permissive conditions for the temperature sensitive initiation mutant were achieved using an objective heater (Bioptechs, 38°C set point). Cells were placed on the heated objective about 10 minutes prior to the start of imaging. Imaging continued with frames at 5 minute intervals for several hours, long enough to capture the beginnings of subsequent replication cycles in wild-type cells.

3.6.5 Characterizing cell elongation during microscopy

Using the cell boundaries determined from segmentation of the phase image, we track the long-axis length over time. Fig. 3.1A and B show typical length vs. time curves for E. coli and B. subtilis. Importantly, the cells were elongating exponentially (curves appear linear on a log scale) and the growth rate (slope) appears constant over time, indicating cell growth was not affected by prolonged exposure to the laser. We determine the doubling time for each cell by least squares fitting of the length vs. time curve with an exponential of the form:

$$L = L_0 \exp\left(t/t_0\right) \tag{3.1}$$

Where the parameter t_0 was allowed to vary and L_0 represents the length of the cell immediately following division. The doubling time was taken as:

$$t_D = t_0 \log\left(2\right) \tag{3.2}$$

The distribution of cell doubling times for individual cells is shown Fig. 3.1C.

3.6.6 Image processing

Cells were imaged in both phase-contrast and fluorescence at five minute intervals. Higher time resolution led to significant photobleaching and/or slowed cell elongation. Imaging continued sufficiently long to capture at least one full cell cycle, about 3.5-4 hours. For snapshot imaging, only one phase-contrast and fluorescence image were taken at each field of view. The process used to analyze the phase and fluorescence images is outlined below.

Cell segmentation from phase images

At each time point, we capture a phase-contrast image for the purpose of segmentation, the process of analytically identifying cell boundaries. Using our labs previously described custom segmentation tool [117, 118] (avalible at http://mtshasta.phys.washington.edu/website/ssodownload.php), superSegger, we generate cell masks in each frame for further analysis and track cells frame-to-frame for entire cell cycles. In order to ensure complete cell cycles in individual cells, we must observe two division events. Cells are tracked starting from their birth (division of the parent cell) until division. The lengths of cells at birth and division are highly variable, and the distributions are shown in Fig. 3.1. We note that the distribution for *B. subtilis* is artificially shifted towards longer lengths because chaining often prevents visual identification of a septum until significantly later than the cell division event.

Locating foci in fluorescence images

To manage xy drift, each phase image was aligned against the previous frame. The corrected alignment of each phase image (and corresponding fluorescence image) was retained for further analysis. Foci are then identified in each fluorescence image by the process described below. A one-pixel-radius Gaussian blur is applied to each frame, after which we apply a watershed process to the conjugate image to identify regions around each intensity maximum. Regions external to the cell masks are excluded from further analysis. Regions internal to the cell masks will be called intensity regions. In order to precisely locate foci within each intensity region, we form a 3 pixel radius circular region at the location of the maximum-intensity

pixel. Within the circular region, we model the intensity profile as a Gaussian distribution with the following form:

$$\mathrm{Ga}\left(\vec{x}\right) = I_g \exp\left(-\tfrac{1}{2b^2}\left(\vec{x} - \vec{x}_0\right)^2\right) + I_0 \tag{3.3}$$

Here, the parameter I_g defines the Gaussian peak amplitude and I0 represents the background intensity. The calculated focus position is and the parameter b characterizes the focus width.

The focus position obtained from the Gaussian fit is used for all localization calculations. In addition, quantifying focus intensity is important for avoiding false detection events. As a measure of focus intensity, we sum the intensity in a three-pixel-radius disk centered at the focus location obtained from the Gaussian fit. The background intensity (I_b) is defined as the minimum pixel value within the circular disk multiplied by the area of the disk. The standard deviation of intensity (δI respectively) is taken over the entire cell mask.

Focus scoring

The statistical significance of a focus is measured by its score σ. The integrated background subtracted intensity (I_a) inside a three-pixel-radius region is computed. The score is defined as:

$$\sigma = I_a / \delta I \sqrt{A} \tag{3.4}$$

Where A is the area of the circular region over which the intensity was integrated. The factor $\delta I \sqrt{A}$ is the expected standard deviation in the integrated intensity over the integration area, assuming the noise at each pixel is uncorrelated. Foci scoring 4 or higher were retained. Lower-scoring foci localized randomly throughout the cell and were consistent with stochastic fluctuation in the intensity (Fig. 3.5).

Data selection

Data processing was completely automated, with no selection by the investigator. Cells were selected for analysis based on the following three criteria:

- No segmentation errors

- Cells were growing exponentially with a doubling time between 1.5 and 4.5 hours.
- The cell length at birth was between 1-4 µm.

The second criterion did not apply to snapshot images where the doubling time could not be calculated, and the last criterion was included mainly to prevent the analysis of long chains of *B. subtilis* cells. Additionally, each experiment was repeated on at least two separate occasions, and data was pooled for multiple experiments.

3.6.7 Statistical analysis of focus localization

Calculation of focus separation

All possible unique distances between the foci in each cell were calculated, and the long-axis components were retained. Relative separation is the long-axis component of the separation between foci divided by the long-axis length of the cell. Visualization of relative focus separation as a function of cell length resulted in two roughly Gaussian populations. We interpret the low separation population (P_L) with transiently separated sister replication forks. The higher-separation population (P_H) is less straight-forward to interpret since these separations may be between pairs of re-initiated factories, a re-initiated factory and a member of a resolvable fork pair, or members of resolvable fork pairs on opposite sides of the cell. Additionally, this higher-separation population is exaggerated in *B. subtilis* since chaining often prevents visualization of the septum. We fit a two-Gaussian mixture model to the observed distribution of separations using a maximum likelihood process where the Gaussian means, variances, and mixing fractions were allowed to vary.

Calculating probability of a replication factory

All cells with a single focus are interpreted to have a co-localized replisome pair. For multiple-focus cells, we identify resolvable sister replisomes by identifying pairs of foci whose separation is most probably associated with P_L (as determined by the posterior probability distribution resulting from the Gaussian mixture model). For each cell, we then define the number of

co-localized replisome pairs as the total number of foci minus twice the number of resolvable replisomes. This process is illustrated in Fig. 3.9.

Chapter 4

THE REPLISOME DISASSEMBLES MULTIPLE TIMES PER CELL CYCLE

This chapter is a reproduction of Ref. [17]

4.1 Introduction

The rapid and faithful replication of the genome is essential to cell proliferation. Although the replisome, the cellular machinery responsible for DNA replication, has been extensively studied both *in vitro* and *in vivo* [119], fundamental questions remain about the dynamics and stability of the replication complex in the context of the living cell, where replication is one of a number of essential cellular processes competing for the genetic material as a template. This competition results in *replication conflicts*, the stalling or pausing of the replication process in the face of obstacles, including transcription and tightly-bound DNA-binding proteins [3, 4, 49, 53, 54, 56, 120–124]. Genetic evidence from bacterial studies suggests these replication conflicts can routinely necessitate a replication-restart process at highly-transcribed regions and genes transcribed in the opposite orientation to replication [3, 5, 125]. However, whether the replisome disassembles in response to conflicts, the rapidity of the restart process, and the frequency of such events per cell cycle are unknown [122, 126–129]. The current estimates for the number of replication restart events per cell cycle vary greatly, with numbers ranging from one per cell cycle to one in seven generations [26, 63, 130, 131]. However, estimates of less than one event per cell cycle are in conflict with the essentiality of the restart protein PriA in (*Bacillus subtilis* essentiality is demonstrated in rapid growth [132]), and the synthetic lethality of PriB and PriC proteins in *Escherichia coli* (PriA mutants as well as PriB PriC double mutants are largely unviable [127, 133, 134]). These observations are consistent with a more frequent requirement for replisome reactivation after conflicts

[127, 132, 134] and provide indirect evidence against the canonical model that replication is continuous *in vivo*.

Our previous investigations using chromatin immunoprecipitation (ChIP) of replication restart proteins determined the chromosomal locations of conflicts in very large ensembles of cells (in the population average) [3, 5]. However, due to population averaging over cells in various stages of the replication conflict and restart process, ChIP experiments are poor reporters of potential conflict-induced changes to the structure of the replisome complex, the frequency of conflicts in a single cell, and the rapidity of the replication restart process. Therefore understanding the fundamental character of DNA replication and conflicts necessitates a single-cell approach in which conflicts can be observed and quantitatively characterized one event at a time. The molecular-scale stoichiometry of the replisome further necessitates experiments with single-molecule sensitivity to detect any potential changes to replisome structure.

We visualized the replication process in single cells by *in vivo* Single-Molecule Fluorescence Microscopy (SMFM). We characterized the stoichiometry and lifetimes of the replicative helicase complexes (and other replication proteins) in growing *B. subtilis* and *E. coli* cells. These measurements revealed that a significant percentage of cells only have a single helicase complex and that many of the complexes are short-lived. These results are consistent with pervasive disassembly of replisomes. We find that transcription inhibition both increases the lifetimes and stoichiometry of several core replisome components, suggesting that endogenous replication-transcription conflicts frequently lead to disassembly of replisomes, potentially every cell cycle. The replication-conflict induced disassembly model suggests that conflicts may limit the rate of replication. Consistent with this model, we find that the inhibition of transcription, and the amelioration of conflicts, increases the replication rate as measured by thymidine incorporation assays.

4.2 Results

4.2.1 Replicative helicase and DNA polymerase stoichiometries are consistent with a single active complex in a large population of cells

To probe replisome stoichiometry in single cells with single-molecule sensitivity, we employ SMFM. In short, the discrete transitions in fluorophore intensity due to bleaching can be detected and analyzed to deduce the stoichiometry of localized fluorophores with single-molecule resolution. The quantitative characterization of the molecular stoichiometry of the replisome in living *E. coli* cells was recently realized by SMFM [24], and this SMFM analysis has been applied in many other contexts (e.g.[135] and [136]). However, SMFM has not been exploited to determine the impact of conflicts on the replisome, the continuity of the replication process, or frequency of disruptions to the replisome within living cells.

We analyzed replisome stoichiometry of the replicative helicase DnaC in *B. subtilis*. DnaC was chosen due to its essential role in the replication process, extensive biochemical characterization, its relatively large and well-accepted stoichiometry in the replisome, and because it is the first replisome protein reloaded onto the DNA during PriA-dependent replication restart [122, 137]. The replicative helicase is responsible for facilitating replication by unzipping the two strands of DNA ahead of the replication forks. *In vitro* biochemical studies, including X-ray crystallography, reveal that the helicase forms a homo-hexameric ring encircling the lagging strand of the DNA template [138–140]. *In vivo* measurements of stoichiometry in *E. coli* further support this model in the context of the living cell [24]. For our studies, we used a DnaC-GFP fusion (Fig. 4.1), which was expressed from its endogenous promoter, at its endogenous locus. The fusion protein localized to midcell in a replication-dependent manner, consistent with association with the replisome [40]. Under our experimental conditions (minimal arabinose medium), the growth rate (and the replication ratesee below) of the DnaC-GFP strain was indistinguishable from that in wild-type cells (During rapid growth in Luria-Bertani medium, DnaC-GFP strain has a minor growth defect [Fig. 4.1A and B]).

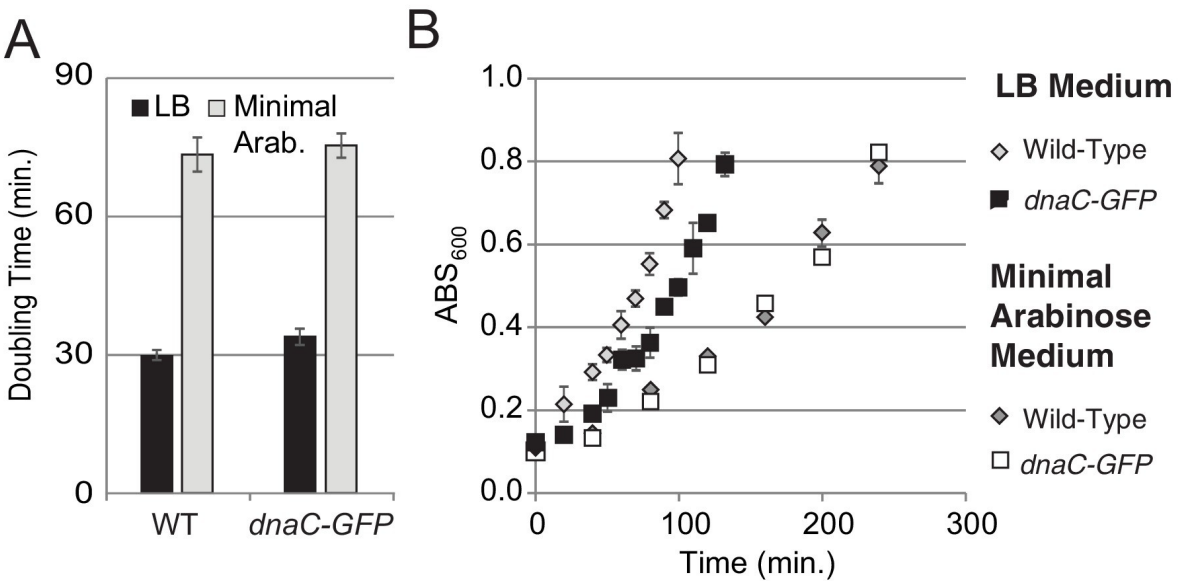

Figure 4.1: **Growth curves for DnaC-GFP.** For growth curves, the optical density of wild-type and *dnaC-gfp* strains growing in minimal arabinose medium at 30°C were monitored for 5 hr. Linear regression to OD_{600} readings were used to determine doubling time. (A) The *dnaC-gfp* allele does not confer a detectable growth defect in minimal medium. OD_{600} readings for a representative culture of wild-type and *dnaC-gfp* cells. (B) Calculated doubling times and OD600 readings show a small growth defect for the *dnaC-gfp* strain relative to wild-type in LB medium.

To measure the *in vivo* stoichiometry of the replisome proteins, we performed SMFM bleaching analysis (Fig. 4.2A and B, and Fig. 4.3 and 4.4). Most bacteria have a circular chromosome and a single origin of replication. After initiation, DNA replication progresses bi-directionally around the chromosome, with two active replisomes in each cell. The two forks in *B. subtilis* often localize to a single replication factory [40](Fig. 4.2A). The small fraction of cells (16%) having focus localization inconsistent with a replication factory were excluded from analysis. It is expected that in cells where the replication forks are co-localized, two replicative helicase complexes, and therefore 12 molecules of DnaC, will be localized to the factory (Fig. 4.2C). However, stoichiometry analysis of DnaC at the replication factory in cells undergoing active replication reveals that just under half the cells (41%) have a factory with only 6 DnaC proteins, corresponding to a single helicase complex (Fig. 4.2D and E). The rest of the population (59%) has 12 DnaC proteins, corresponding to two active helicases. Note that incomplete protein labeling cannot account for the low helicase stoichiometry since two sub-populations with an integer multiple of fundamental hexameric stoichiometry are observed, as has also been previously reported in *E. coli* [24]. Western-blot analyses further confirm that >98% of DnaC protein in the cell is indeed labeled with GFP (Fig. 4.5). The incorrect determination of the bleaching step size also cannot account for these observations since results can be reproduced by using an *in vitro* measure of the fluorophore step size (Fig. 4.6). These results are consistent with a model where elongating replisomes are frequently disrupted and disassembled.

Additionally, we analyzed the stoichiometry of PolC-YPet, a fluorescent fusion to the leading strand polymerase in *B. subtilis*. Even in rapid growth, the PolC-YPet fusion confers no growth defect relative to wild type cells (Fig. 4.7). We again find that the stoichiometry distribution for co-localized replication forks consists of two sub-populations, with the second (4 copies) centered at roughly twice the stoichiometry of the first (2 copies) (Fig. 4.2F). Importantly, these sub-populations were similarly proportioned to those observed for DnaC, with roughly 47% of factories having stoichiometries consistent with localization of PolC to only a single replication fork (Fig. 4.2G). This implied disassembly of the DNA polymerase, in

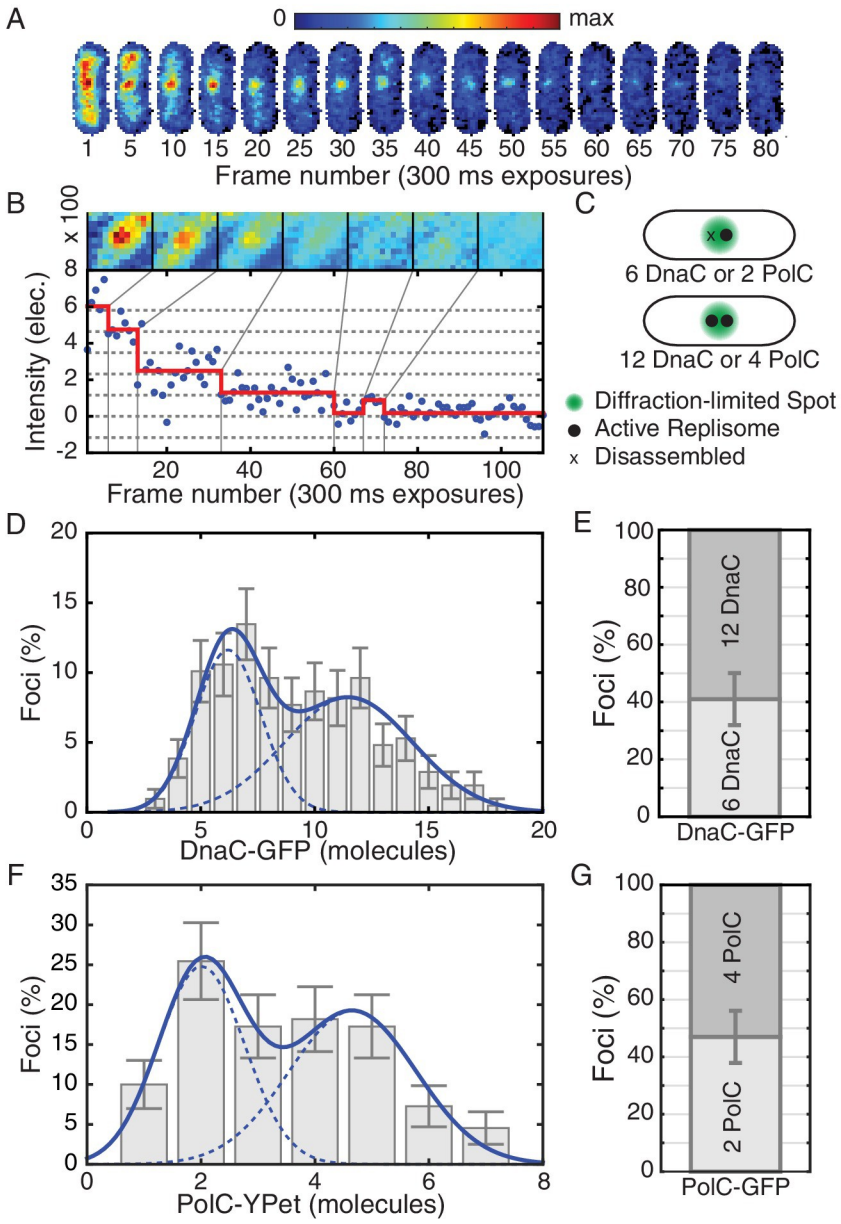

Figure 4.2: **Estimated stoichiometry distributions for core replisome proteins in *B. subtilis*.**(A) Photobleaching of DnaC-GFP in a replication factory. (B) A typical intensity trace (blue) is shown for a DnaC-GFP focus. Stepwise transitions are observed as the fluorescent protein bleaches. The intensity is filtered using Change-Point analysis (red) which determines the intensity step-size corresponding to the bleaching of single fluorophores (complete stoichiometry calculation is demonstrated in Figure 4.3 and 4.4, and detailed in the materials and methods section). The image mosaic above shows the time-averaged image

of the focus over each intensity level. (C) A schematic of the replication factory consisting of either one or two assembled replisomes (black dots) in a diffraction-limited spot (green). (D) Histogram of estimated factory DnaC-GFP stoichiometry in *B. subtilis*. Error bars represent counting error. The observed distribution is well fit by a two Gaussian model (solid blue), representing a mixed population of single-helicase (6 DnaC molecules) and two-helicase (12 DnaC molecules) factories. (Analysis for N = 213 factories.) (E) Relative abundance of factories with one and two helicases. (F) Estimated stoichiometry distribution for PolC-YPet in *B. subtilis* shows two populations (N = 125). Peak stoichiometries for each population (dashed blue) were determined by maximum likelihood fitting with a two Gaussian model (solid blue) to be 2 and 4 copies. (Note: the distribution included a small fraction (5%) of factories with stoichiometries greater than 10 copies which were removed for the purpose of fitting.) (G) Relative abundance of factories with 2 and 4 copies of PolC-GFP.

addition to the replicative helicase, further suggests that the replisome is frequently disrupted.

4.2.2 Replicative helicase complexes are short-lived

To test the frequent-disruption model, we measured the lifetimes (replisome dynamics) of replicative helicase complexes. This model predicts that any given replisome complex is shorter lived than the time required to traverse each arm of the chromosome. We visualized the helicase complexes over twenty-two minutes to estimate their lifetimes. In the replisome dynamics measurements, cells were imaged at two-minute intervals such that the helicase complexes can be tracked (Fig. 4.8A, left image strip; Fig. 4.9) over the twenty-two minute time frame without significant bleaching (Fig. 4.10). Disassembly events correspond to the cooperative loss of all fluorophores simultaneously. Although photobleaching affects our measurement only minimally over the course of the experiment, it would not be possible to visualize the replisome at this frame rate for the entire cell cycle (Fig. 4.10). Using the complex lifetimes observed during these short time courses, we estimate the disassembly rate, predicting that roughly five disassembly events occur during each cell cycle.

We found that the majority of foci are short lived with a mean lifetime of roughly 8 min. The observed distribution of lifetimes (Fig. 4.8B) appears roughly exponential, suggesting that the occurrence of disassembly events in individual cells may be approximated as a Poisson

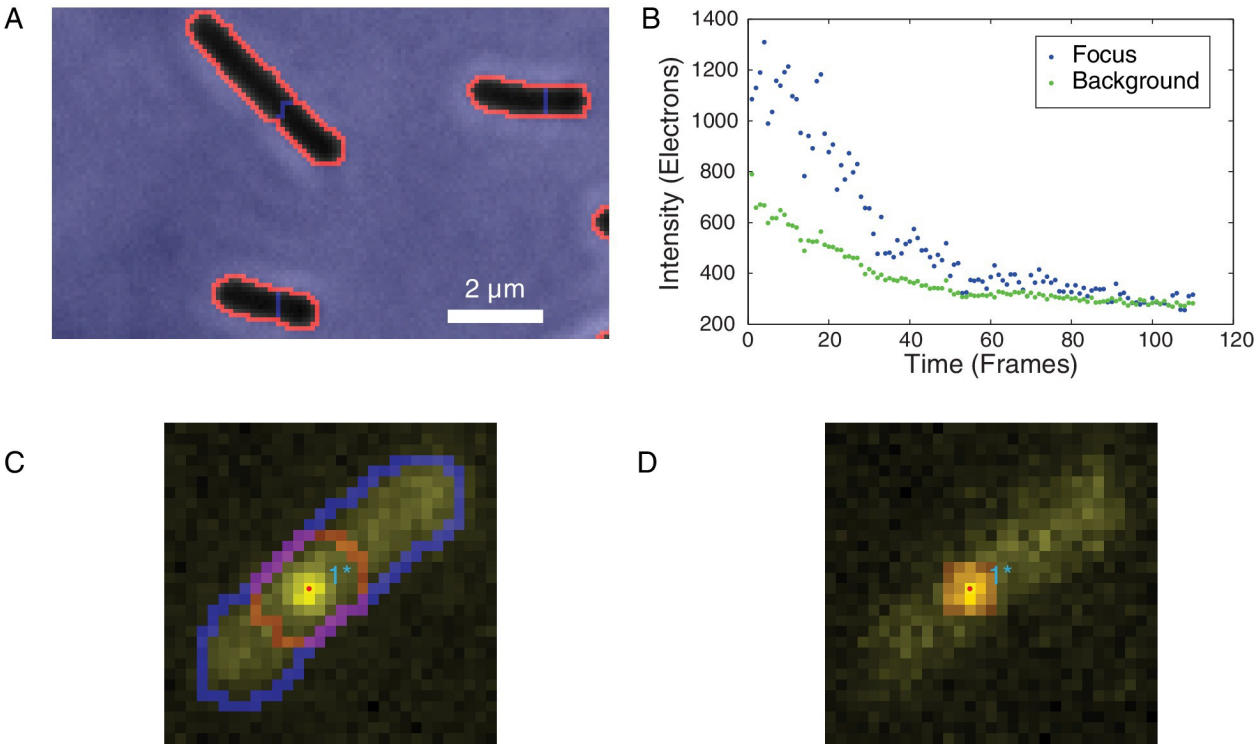

Figure 4.3: **Identification of foci and determination of raw focus intensity.**(A) The phase-contrast image of the cells was segmented to identify the cell boundaries (orange) surrounding each cell mask (black). (B) The integrated raw focus intensity (blue) and scaled background intensity (green) are plotted for a bleaching experiment. The intensity trace is computed by subtracting the background from the raw intensity. (C) The summed image is shown in yellow. The outline of the cell mask is shown in blue. The intensity regions (red outline) are determined by watersheding the summed intensity to divide the image into local maxima. The focus positions (points) are determined by fitting a Gaussian distribution in the intensity regions. (D) The raw focus intensity is determined by summing the intensity in a mask centered on the locus position. The background is computed by calculating the average intensity inside the cell (blue outline) but outside the intensity region (red).

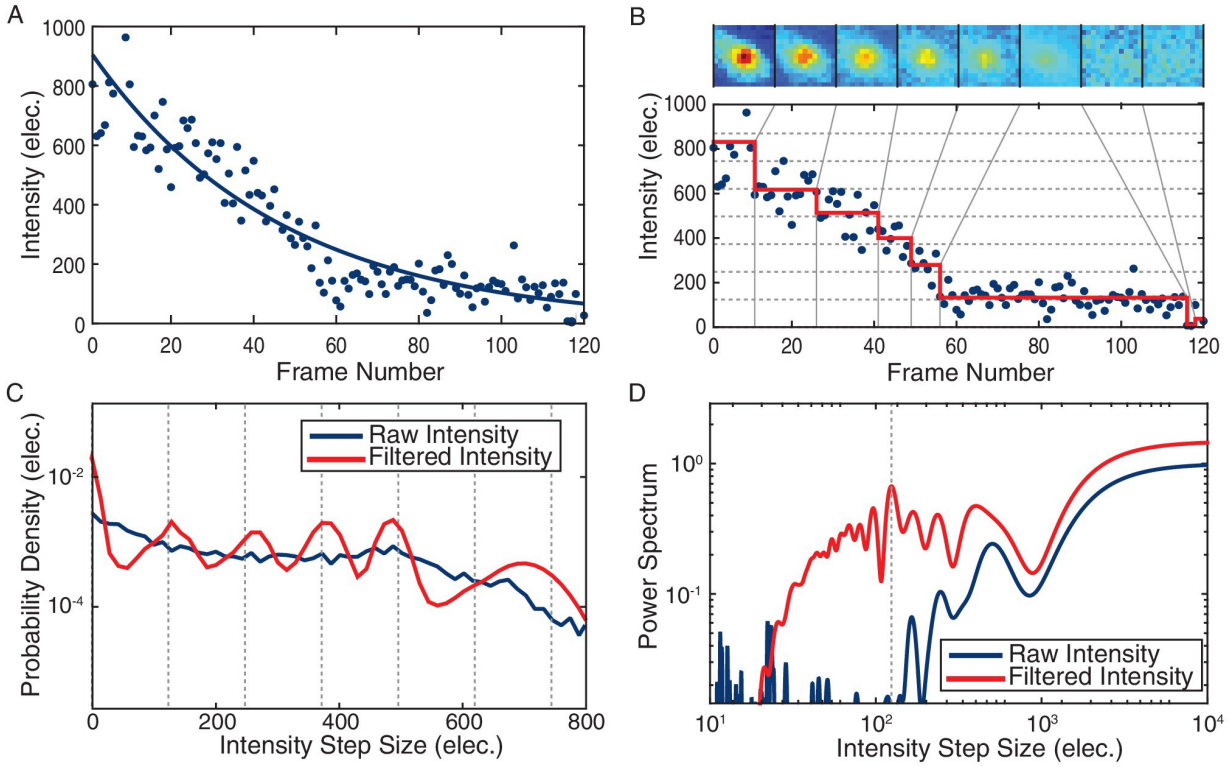

Figure 4.4: **Stoichiometry calculation demonstrated for a DnaC-GFP focus.** (A) Intensity tracking of a DnaC-GFP focus in *B. subtilis*. A decaying exponential (solid blue) is fit to the background subtracted raw intensities (blue points) traced over a series of 120 fluorescence images. The value of the exponential fit in the first frame is taken as the initial intensity of the DnaC-GFP complex (see methods). (B) The raw intensities (solid blue) are filtered (mean value in red) to reveal stepwise transitions. Summed fluorescence images of the DnaC-GFP complex are shown corresponding to each level detected by the filter in the image strip above the plot area. (C) The PPDD of the filtered intensities (red) reveals peaks at integral multiples of the unitary intensity step that would not be detectable in PPDD of the raw intensities (blue). (D) The power spectrum is used to identify the location of the first order peak in the PPDD which corresponds to the highest peak in the power spectrum (dashed gray). The unitary intensity step determined by the power spectrum corresponds to the spacing of the dashed gray lines in panels B and C.

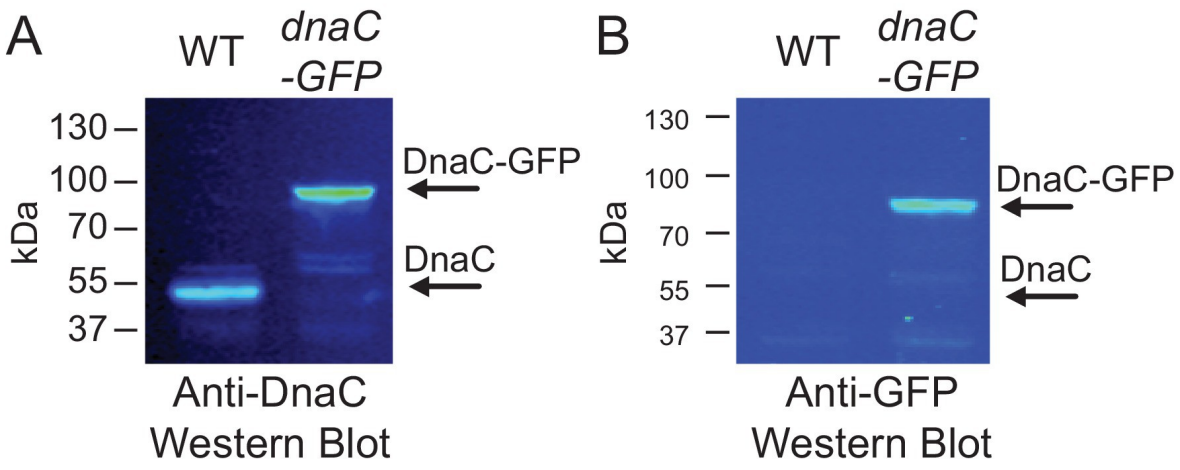

Figure 4.5: **Western blots for DnaC-GFP.** Cells produce complete DnaC-GFP fusion protein. Western blot analysis indicates that DnaC-GFP is fully synthesized, and that potentially truncated DnaC proteins lacking GFP are not detected. Individual lanes were normalized by total protein, and separate western blots were probed with either (A) anti-DnaC polyclonal antibodies or (B) anti-GFP polyclonal antibodies.

	T (min)	$N_{\tau<T}$	$N_{\tau\geq T}$	τ_{min} (min)	$\bar{\tau}$ (min)	$\hat{k}$ (min^{-1})	$\bar{\tau}_{calc}$ (min)	N_C
Wild-Type	22	296	31	4	10.4	0.12	8.3	4.8
Rif-Treated	22	86	97	4	12.9	0.03	29.2	1.4
*rpoB**	22	93	72	4	12.5	0.04	22.4	1.8

Table 4.1: **Parameters used in *B. subtilis* complex lifetime calculation.** The parameters summarized above are defined as follows: T is the duration of the experiment, $N_{\tau<T}$ is the number of complexes interpreted to disassemble before the end of the experiment, $N_{\tau\geq T}$ is the number of complexes surviving the length of the experiment, τ_{min} is the minimum observable complex lifetime, $\bar{\tau}$ is the empirical mean of the $N_{\tau<T}$ observable lifetimes, $\hat{k}$ is the calculated disassembly rate, $\bar{\tau}_{calc}$ is the calculated mean lifetime (i.e. $1/\hat{k}$), and N_c is the calculated number of conflicts per 40 min of replication.

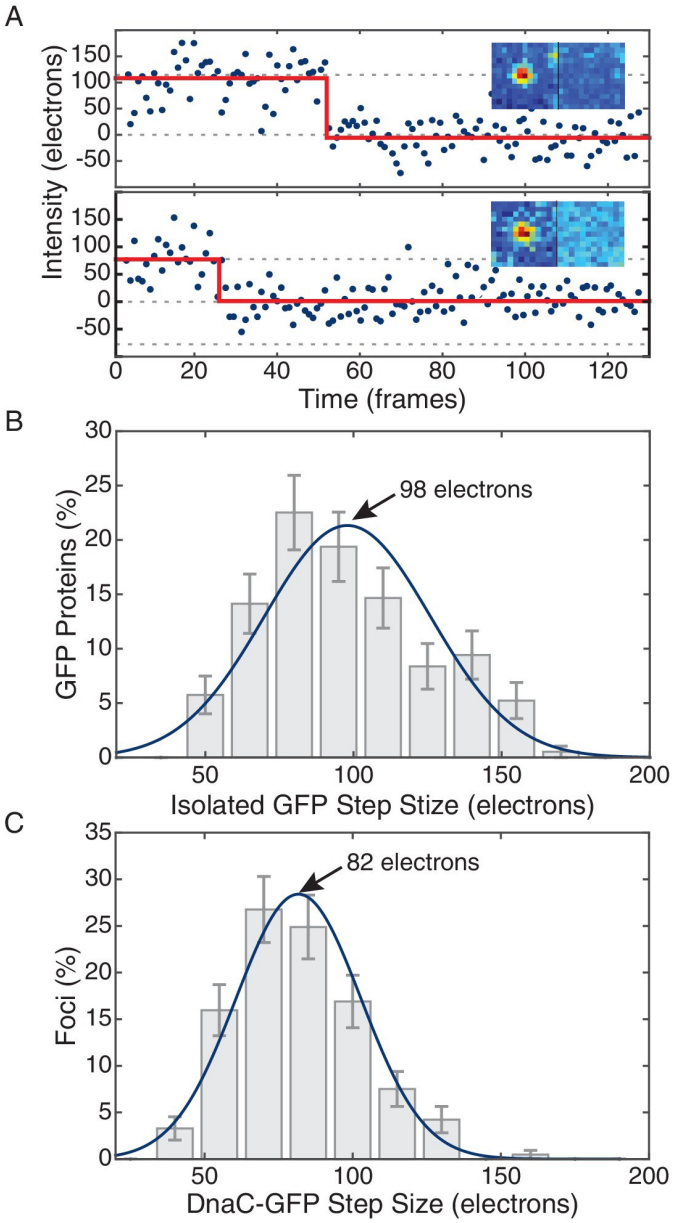

Figure 4.6: **Comparison of *in vivo* and *in vitro* step sizes for GFP.** (A) Bleaching traces for two different surface immobilized GFP proteins. Inserts show the mean fluorescence images in the bleached and unbleached states. (B) Maximum likelihood fits to unitary intensity step distributions for isolated GFP *in vitro*. (C) Example *in vivo* unitary step distribution for DnaC-GFP in *B. subtilis* and its maximum likelihood fit. For all experimental conditions in *B. subtilis*, unitary step distributions were peaked within 19% found *in vitro* value.

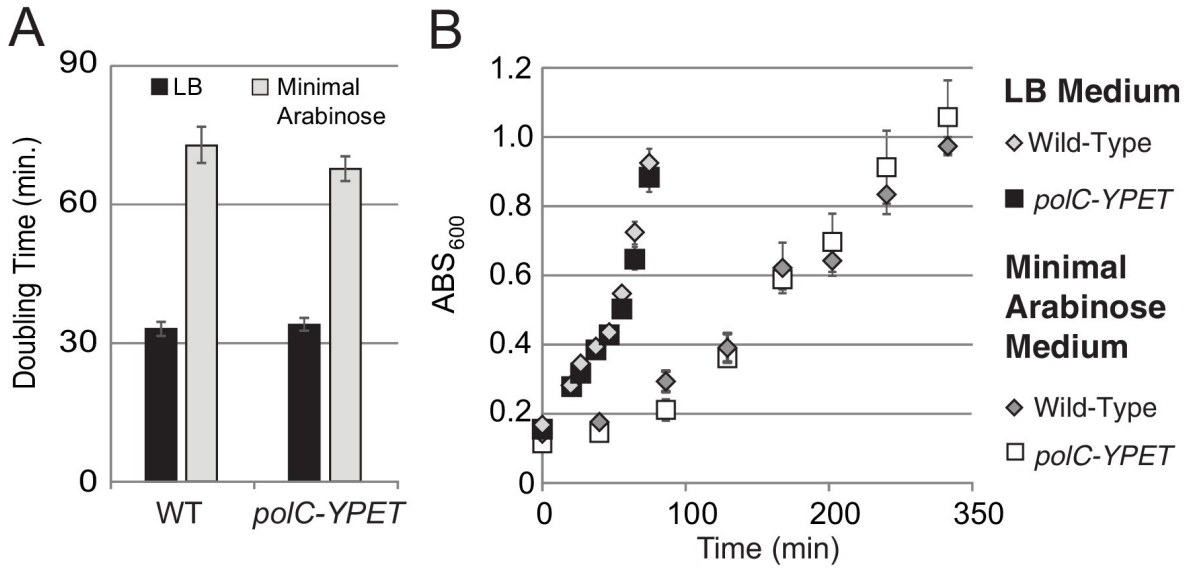

Figure 4.7: **Growth curves for PolC-YPet.** For growth curves, the optical density of wild type and *polC-ypet* strains growing in minimal arabinose medium at 30°C were monitored for 5 hr. Linear regression to OD_{600} readings were used to determine doubling time. (AB) No growth defect is observed for *polC-ypet* in either minimal or LB growth medium.

process, providing a method for roughly estimating the number disassembly events per cell cycle. (For simplicity, we will ignore the distinction between focus loss and fork disassembly. We present a more complicated model, that treats the number of forks per focus explicitly (Appendix A), but this model depends on a number of untested assumptions.) Figure 4.8C shows the model (dashed gray) and empirical (solid gray) survival curves for the helicase complex as a function of lifetime, where the model survival curve was fit using by maximum likelihood estimation. Multiplying the calculated rate of conflicts by the accepted 40 min long replication cycle predicts five conflicts per cell cycle (Fig. 4.8D). It is important to note that this estimate has significant systematic uncertainty since it depends on the length of the replication cycle (which depends on growth conditions [141] as well as the assumption made in the modeling (see Supplementary file 1). A summary of the calculated parameters for the dynamics experiment is included in Table 4.1. Disassembly of the replisome on the time scale of single replication cycle is consistent with the instability inferred from the analysis of the

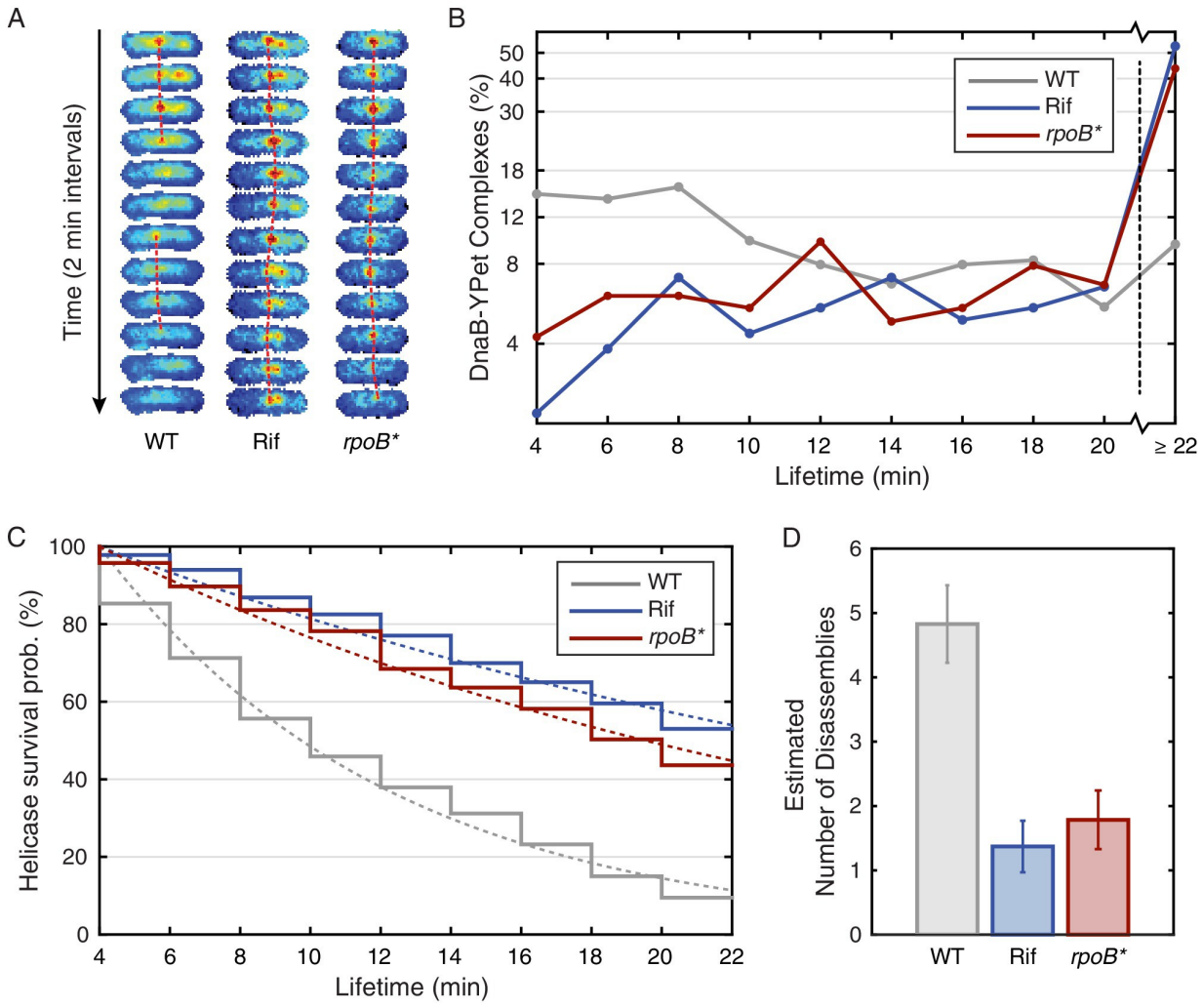

Figure 4.8: **Helicase complex dynamics in *B. subtilis* captured by time-lapse imaging.** (A) Typical frame mosaics for dynamics of the helicase in three conditions: Untreated (WT), rifampicin-treated (Rif) and *rpoB** cells (see also Fig. 4.9). (In this context, WT refers to cells carrying the *dnaC-gfp* allele but not the *rpoB** allele.) The helicase complexes in WT cells were observed to be intermittent: assembling and disassembling on the timescale of minutes. The helicase complexes in rifampicin-treated and *rpoB** cells were observed to be more persistent. The complexes were tracked by an automated algorithm (red, detailed description in material and methods). (B) Distribution of helicase complex lifetime in WT (gray, N = 327 complexes), rifampicin-treated (blue, N = 183 complexes) and *rpoB** (red, N = 165 complexes) cells. Data collection is limited to 22 min. (C) Probability of helicase

survival as a function of helicase complex lifetime. Solid lines represent the empirical survival curves. Dashed lines show fits determined by maximum likelihood estimation. (D) Estimated number of disassembly events per 40 min of replication using the Poisson process model (see also Table 4.1). Error bars were generated by simulating 100,000 distributions with the same rate parameter and number of complexes as the observed distribution. Simulated distributions were then fit, and the width of the rate parameter distribution was used to quantify the error.

estimated factory stoichiometry. Furthermore, the inferred frequency of disassembly events is significantly higher than expected from the continuous replication model.

The frequent disassembly model further predicts that disrupting the restart protein PriA would prevent replisome reassembly, leaving disassembled replisomes in the majority of cells during a single cell cycle. To test this, we depleted PriA using CRISPR interference and visualized the number of cells with DnaC-GFP foci using snapshot imaging. Consistent with the frequent disassembly hypothesis, after inducing PriA depletion in liquid culture for roughly one doubling time, only 13% of cells had foci compared with the roughly 43% of cells with foci in the precursor strain to the PriA CRISPR, which lacks the sgRNA (Fig. 4.11).

4.2.3 Transcription inhibition increases the lifetimes of DnaC complexes

The results described support a model in which replication is discontinuous, with pervasive replisome disassembly and assembly dynamics. Our lab and others have shown that transcription, especially at the highly-transcribed rDNA or head-on genes (those transcribed in the opposite orientation relative to replication), results in potentially severe replication-transcription conflicts [3, 55, 142, 143]. Our previous work measuring chromosomal regions where replication conflicts were most prevalent suggested that transcription is the main obstacle to replication [3]. Therefore, if transcription-induced conflicts were the principal mechanism responsible for replisome disassembly, we would predict that inhibition of transcription would result in a significant increase in replisome stability, either as assayed by stoichiometry or direct characterization of complex stability. To test this hypothesis, we

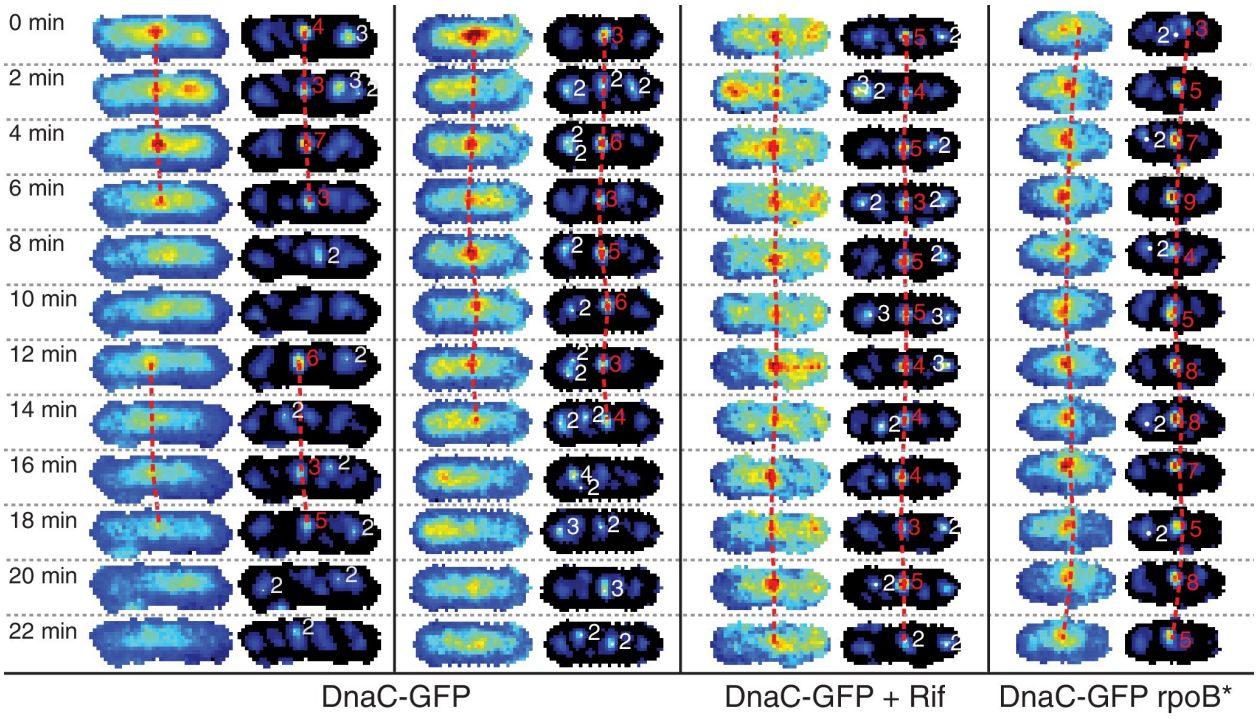

Figure 4.9: **Automatic tracking of helicase complexes using focus scoring.** Under normal growth conditions, a typical trajectory resulting from dynamics analysis shows rapid appearance and disappearance of DnaC-GFP foci. Both rifampicin treatment and a *rpoB** mutation results in longer-lived foci. Image strips to the left show raw images with dashed red lines indicating automatically detected trajectories. Corresponding filtered images to the right additionally show focus scores for all automatically detected foci. Scores printed in red indicate that the focus is included in a trajectory. Low scoring foci not selected for a trajectory are indicated in white. A detailed description of scoring and automated focus tracking is included in the materials and methods section.

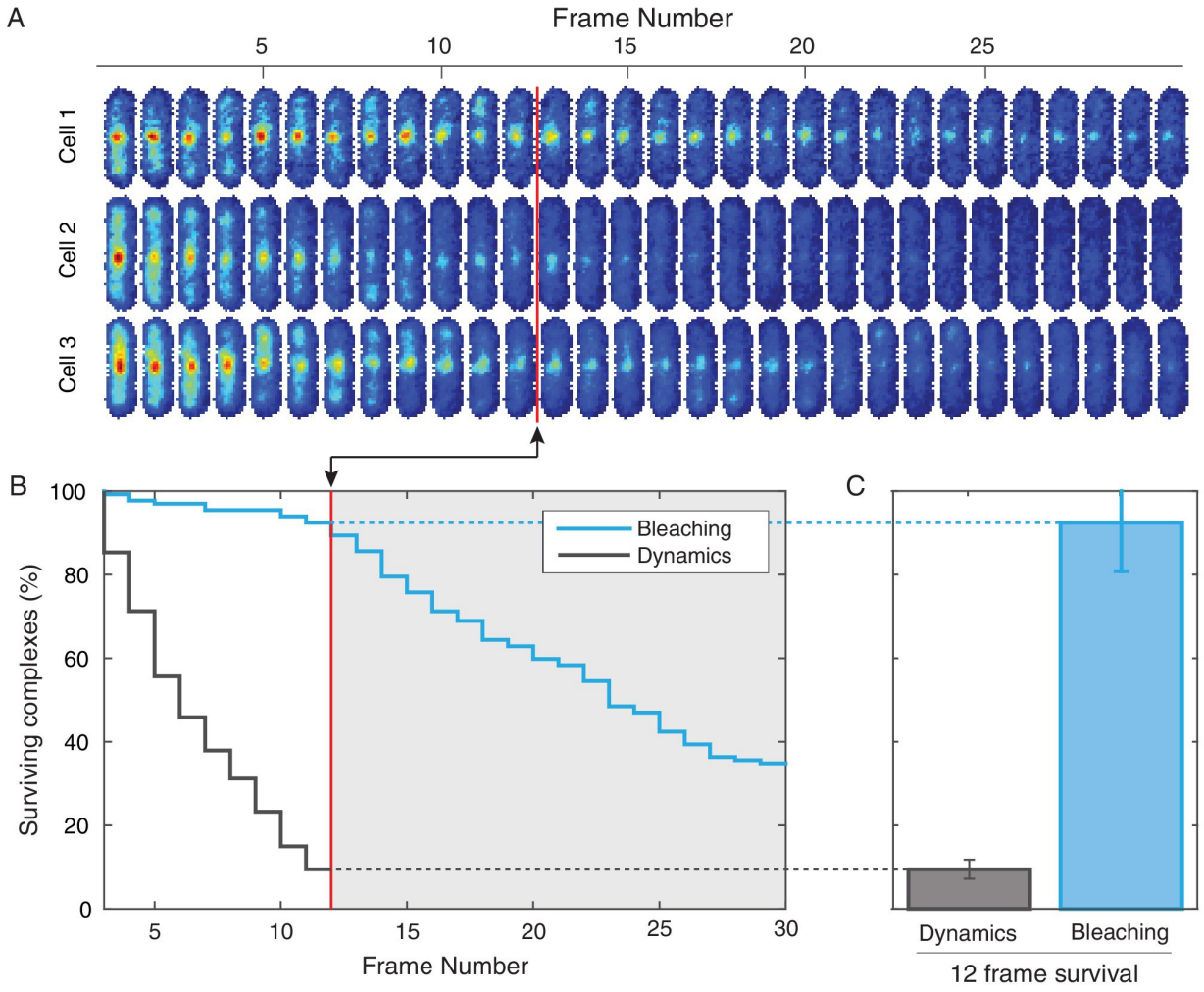

Figure 4.10: **Photobleaching minimally affects the complex lifetime experiments in *B. subtilis*.** (A) To isolate the effects of photobleaching, cells are imaged with the same intensity and exposure as in the dynamics measurement, but with no delay between frames. Red line indicates equivalent of the 12 frame time course used in the dynamics measurement for *B. subtilis*. (B) Automated tracking used for the dynamics measurement was applied to the bleaching data. Survival curve (blue) shows the fraction of (N = 132) complexes that were successfully tracked through a given number of frames. Survival curve for dynamics experiment (gray) is shown for comparison. (C) 92% of complexes were traceable for 12 frames. In contrast, only 9% of complexes survive the duration of the dynamics experiment.

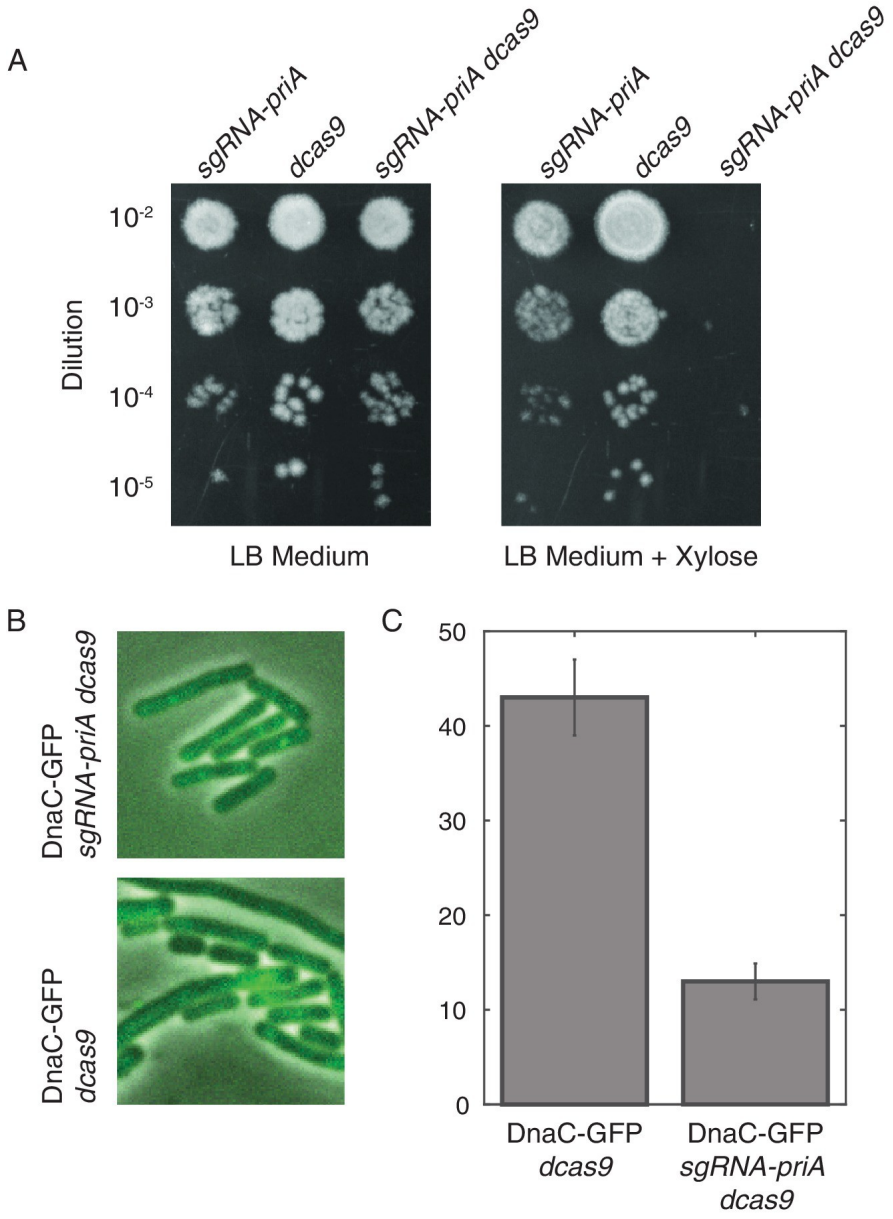

Figure 4.11: **Disruption of PriA leads to loss of DnaC-GFP foci.** (A) Survival assays show that the depletion of PriA is lethal on LB medium, but the control strains (which contain either the *sgRNA-priA* or dcas9) survive under both conditions. We find the same result using minimal arabinose medium. (B) Sample snapshots showing the full PriA CRISPR strain, and the control (no sgRNA), both after induction with 1% xylose for approximately two hours. (C) Disruption of PriA results in a decrease in the number of cells with DnaC-GFP foci from 43% in the control strain (N = 518 cells) to 13% in the CRISPR strain (N = 413 cells). Error bars represent counting error.

perturbed transcription by (i) the treatment of cells with rifampicin (Rif), an antibiotic that directly inhibits transcription initiation [144, 145] and (ii) *rpoB** mutants which destabilize RNA Polymerase-DNA association and/or rDNA expression [146–148]. It has been previously shown that *rpoB** mutations reduce severity of conflicts significantly [56, 70, 120]. Here, we observed that rifampicin treatment dramatically increases the lifetime of helicase complexes. For instance, the number of complexes that persisted longer than 20 min increased from 9% in untreated cells to 53% in rifampicin treated cells and the calculated number of conflicts per cell cycle decreased by a factor of four (Fig. 4.8D; Table 4.1). We also isolated and examined an *rpoB** mutant strain and found results consistent with those obtained from rifampicin-treated cells (Fig. 4.8). This second transcription perturbation ensures that the observed increase in replisome stability is not an artifact of rifampicin treatment. In general, the increased stability of the helicase complex is consistent with the increased stability of the replisome.

4.2.4 More cells have helicase and DNA polymerase stoichiometries consistent with two active replisome complexes upon transcription inhibition

To provide additional support for the transcription-dependence of the observed instability, we measured the stoichiometry of the helicase complexes in cells post transcription inhibition. An increase in replisome stability would predict an increase in the ratio of cells with two helicase complexes (12 DnaC molecules) relative to the number of cells with one helicase complex (6 DnaC molecules). Figure 4.12A shows the distribution of protein stoichiometry at the factory in cells with transcription inhibited by rifampicin. The data are clearly consistent with the majority of cells having two active replisomes. There was a significant increase in the percentage of cells containing two stable helicase complexes relative to cells containing one helicase complex in the *rpoB** mutant backgrounds as well (Fig. 4.12A and D, Table 4.2). A similar shift towards the higher order stoichiometry is observed for PolC after transcription inhibition by rifampicin (Fig. 4.12B and D). Increased stability of the polymerase was also observed in the *rpoB** mutant background. These results strongly support the model that

	N	μ_1, μ_2 (copies)	F_L (error)
DnaC-GFP	213	6.2, 11.2	41 (9.1)%
DnaC-GFP + Rif	69	N/A, 12.1	0 (14)%
DnaC-GFP *rpoB**	60	N/A, 11.1	0 (10.0)%
DnaC-GFP *lacZ* off	108	6.6, 11.5	65 (5.6)%
DnaC-GFP *lacZ* on	174	6.0, 10.9	88 (3.2)%
DnaC-GFP CD *lacZ*	185	5.8, 10.2	59 (9)%
DnaC-GFP + IPTG	216	6.3, 11.2	54 (11)%
PolC-YPet	125	2.2, 4.9	47 (13)%
PolC-YPet + Rif	117	N/A, 4.4	0 (14)%
PolC-YPet *rpoB**	81	N/A, 4.7	0 (9%)

Table 4.2: **Maximum likelihood fit parameters for *B. subtilis* count distributions.**

the frequent dissociation of the replicative helicase complex during each cell cycle results from replication-transcription conflicts.

4.2.5 A severe head-on conflict disassembles helicase complexes

Our data imply that replication-transcription conflicts increase the instability of replisome complexes. To test this model, we introduced an IPTG inducible *lacZ* gene ($P_{spank(hy)}$-*lacZ*) onto the chromosome in the head-on orientation (strain *lacZ*). We hypothesized that the induction of this ectopic conflict would further destabilize the replisome, leading to a reduction in the fraction of cells containing two intact helicase complexes relative to the control cells. As predicted, analysis of the single-molecule stoichiometry of DnaC in cells experiencing the additional severe engineered conflict (after addition of IPTG to induce *lacZ* expression) decreases the number of cells containing two intact helicase complexes (Fig. 4.12C and D; >80% of cells contain only 6 rather than 12 molecules of DnaC in the factory), demonstrating that replication-transcription conflicts indeed lead to the disassembly of the replicative helicase complexes. We note that insertion of the *lacZ* gene in the co-directional orientation leads to a small, but detectable decrease in the number of cells with two assembled helicase complexes under induced conditions (Fig. 4.13). This is consistent with previous reports from our

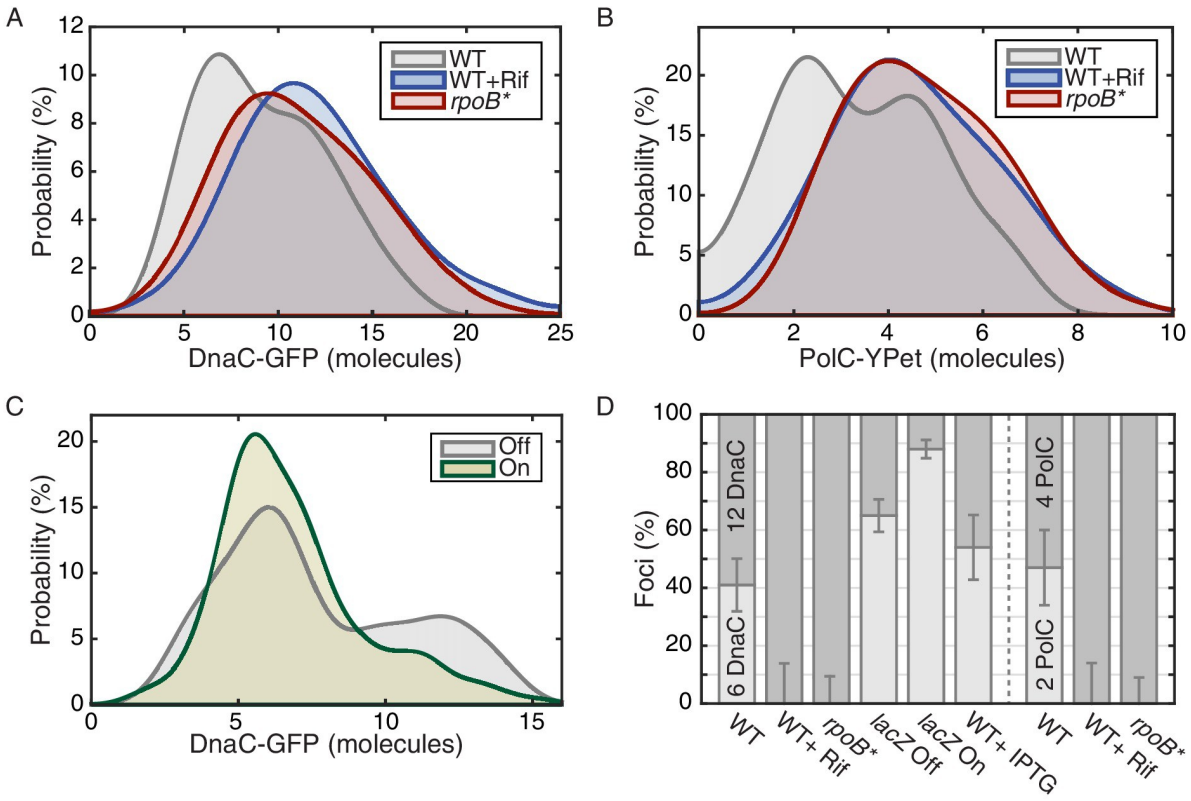

Figure 4.12: **The effect of perturbations to transcription on stoichiometry distributions.** (A) Helicase complex stoichiometry under three conditions: Untreated (WT), rifampicin-treated (Rif) and *rpoB** cells. Probability densities are represented as Kernel Density Estimates (KDEs). In contrast to WT (gray), in both rifampicin-treated (blue) and *rpoB** (red) cells, a significant fraction of single-helicase complexes (6 DnaC molecules) are lost with all observations being consistent with two-helicase complexes (12 DnaC molecules). (N = 70117) (B) Estimated PolC stoichiometry in untreated (gray), rifampicin-treated (blue), and *rpoB** (red) cells (N = 81125). The low stoichiometry peak is no longer resolvable after rifampicin treatment or in the *rpoB** mutant background, implying increased stability of the polymerase. (Note: rifampicin treatment also increased the exceedingly high fluorescence population which was again removed for the purpose of fitting.) (C) The relative abundance of helicase complex stoichiometries in cells with an ectopic inducible head-on replication-transcription conflict (*lacZ*). An IPTG (induction)-dependent increase in the single-helicase stoichiometry was observed, consistent with the reduction in factory stoichiometry being conflict-induced. (N = 108174) (D) Summary of stoichiometry for transcription-inhibition and ectopic-conflict experiments. Estimates for the relative abundance of first and second order stoichiometry sub-populations are determined by fitting the distributions of estimated stoichiometries (Fig. 4.14, Table 4.2).

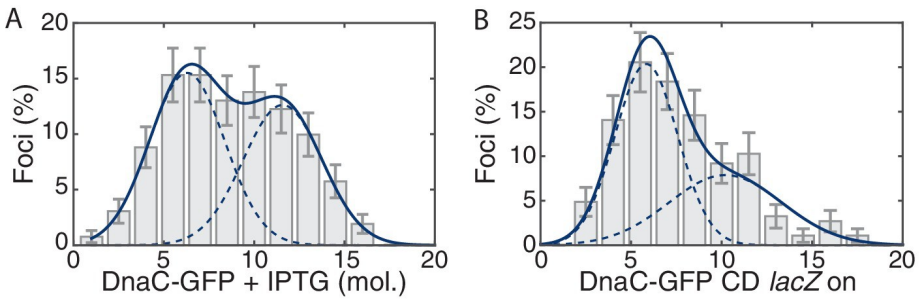

Figure 4.13: **Control for head-on conflict experiment.** (A) The observed stoichiometry of DnaC-GFP was not significantly altered by the addition of IPTG. Compare to Figure 4.2D. (B) Addition of *lacZ* (transcription on) in the co-directional orientation causes a small decrease to replisome stability.

lab showing that highly-transcribed co-directional genes can lead to conflicts as assayed by ChIP-qPCR experiments [5]. Control experiments showed that the addition of IPTG to cells without the *lacZ* construct did not cause a significant shift in the stoichiometry distribution (Fig. 4.13). These results (summarized in Fig. 4.12D, see also Fig. 4.14) altogether reveal that the inferred instability and disassembly of the helicase complexes, and potentially the replisome, is transcription-dependent. This is illustrated schematically in Figure 4.15.

4.2.6 E. coli replisome stoichiometry and dynamics corroborate transcription-dependent instability

Given the universality of conflicts, we hypothesized that other bacteria, outside of *B. subtilis*, should experience this phenomenon. We therefore investigated the stoichiometry and lifetimes of three different replisome proteins in a second bacterial model organism, *E. coli*. The replisome localization patterns are similar in *E. coli* and *B. subtilis*, where both replication forks typically co-localize to a single diffraction limited focus [7, 100, 149].

We measured stoichiometries of three different *E. coli* replication proteins in the replisome complexes: the replicative helicase (DnaB), clamp loader (DnaX), and DNA polymerase (DnaE), both before and after transcription inhibition with rifampicin. For the *E. coli* experiments, the fluorescent fusions were constructed with distinct linkers to the fluorescent

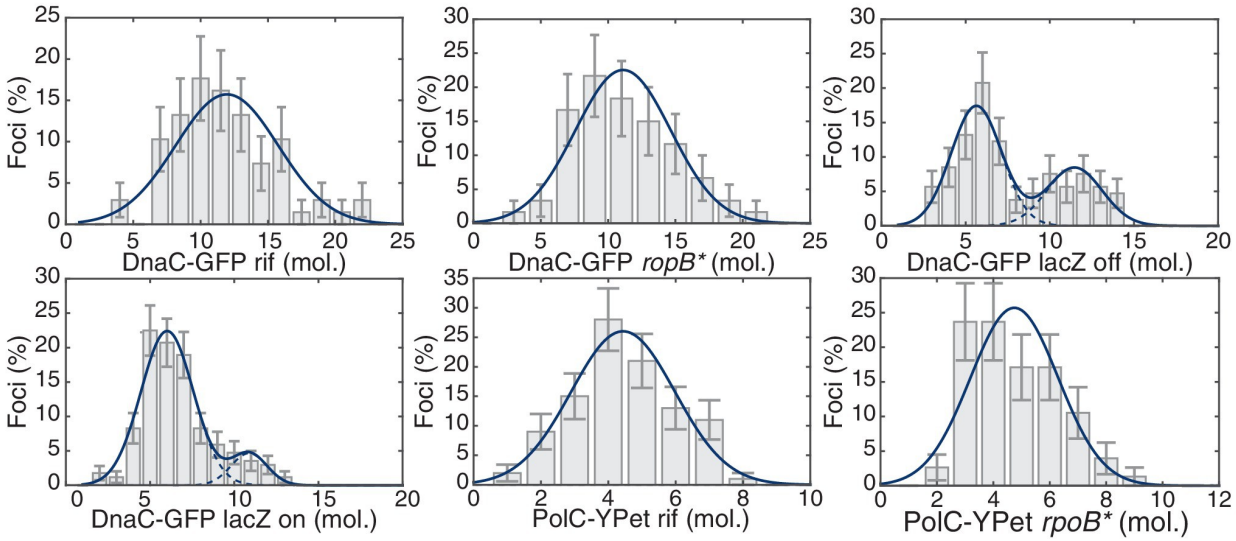

Figure 4.14: **Maximum likelihood fits to *B. subtilis* stoichiometry distributions.** The most likely two Gaussian model is selected varying the Gaussian means, widths and peak intensities. Maximum likelihood models (blue) and comprising Gaussians (dashed blue) are shown over histogram distributions. Error bars represent counting error. Calculated maximum likelihood fit parameters are summarized in Table 4.2

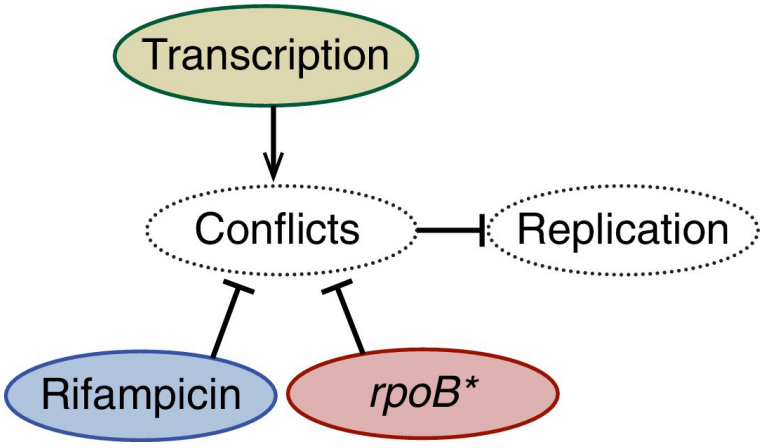

Figure 4.15: **A schematic model illustrating the effects of perturbations to transcription on replisome stability.** A schematic model illustrating the effects of perturbations to transcription on replisome stability. Amelioration of conflicts both by rifampicin or a *rpoB** mutant increases the stability of the replisome. Conversely, addition of a highly transcribed gene in the head-on orientation decreases replisome stability.

	N	μ_1, μ_2 (copies)	F_L (error)
DnaB-YPet	146	6.6, 11.6	59 (15)%
DnaB-YPet + Rif	53	N/A, 12.1	0 (13)%
DnaX-YPet	125	2.8, 5.8	53 (13)%
DnaX-YPet + Rif	154	2.8, 5.7	10.6 (17)%
DnaE-YPet	178	2.8, 5.8	46 (15)%
DnaE-YPet + Rif	144	2.8, 5.6	9.8 (9.3)%

Table 4.3: **Maximum likelihood fit parameters for *E. coli* count distributions.**

protein (YPet) [24]. As observed previously, there are 6, 3, and 3 molecules of DnaB, DnaX, and DnaE respectively in a significant proportion of the localized replisome foci in growing cells, corresponding to single a replisome complex (Fig. 4.16). However, as predicted, rifampicin treatment results in the low stoichiometry population shifting towards the higher stoichiometry peak in the majority of these replisome foci (Fig. 4.16D; Table 4.3; Fig. 4.17). This suggests that there are indeed two co-localized replication forks in the observed foci in *E. coli*, but that one replisome is often at least partially dissociated from the DNA due to replication-transcription conflicts.

Our replisome-stoichiometry measurements in *E. coli* predict that there is pervasive disassembly of replisomes in this species as well and therefore these protein complexes should also have short lifetimes. Because there are only 3 molecules of DnaE or DnaX, lifetimes of the localization of these components to the replisome foci cannot be reliably measured due to bleaching of a significant fraction of the fluorophores. However, because of the higher number of DnaB molecules in each complex, the helicase lifetimes can be measured over a twenty-minute time series without bleaching significantly affecting the results (Fig. 4.19). The *E. coli* helicase complex lifetime measurements show that the majority of DnaB complexes are indeed short lived with a mean lifetime of about 9 min (Fig. 4.20). Furthermore, treatment with rifampicin extends the lifetime of these complexes, increasing the number of complexes that persist for 20 min (or longer) from 21% to 46% and the predicted number of conflicts per (40 min) replication cycle decreases from roughly 4 to 2 (Figure 5D; Table 4.4).

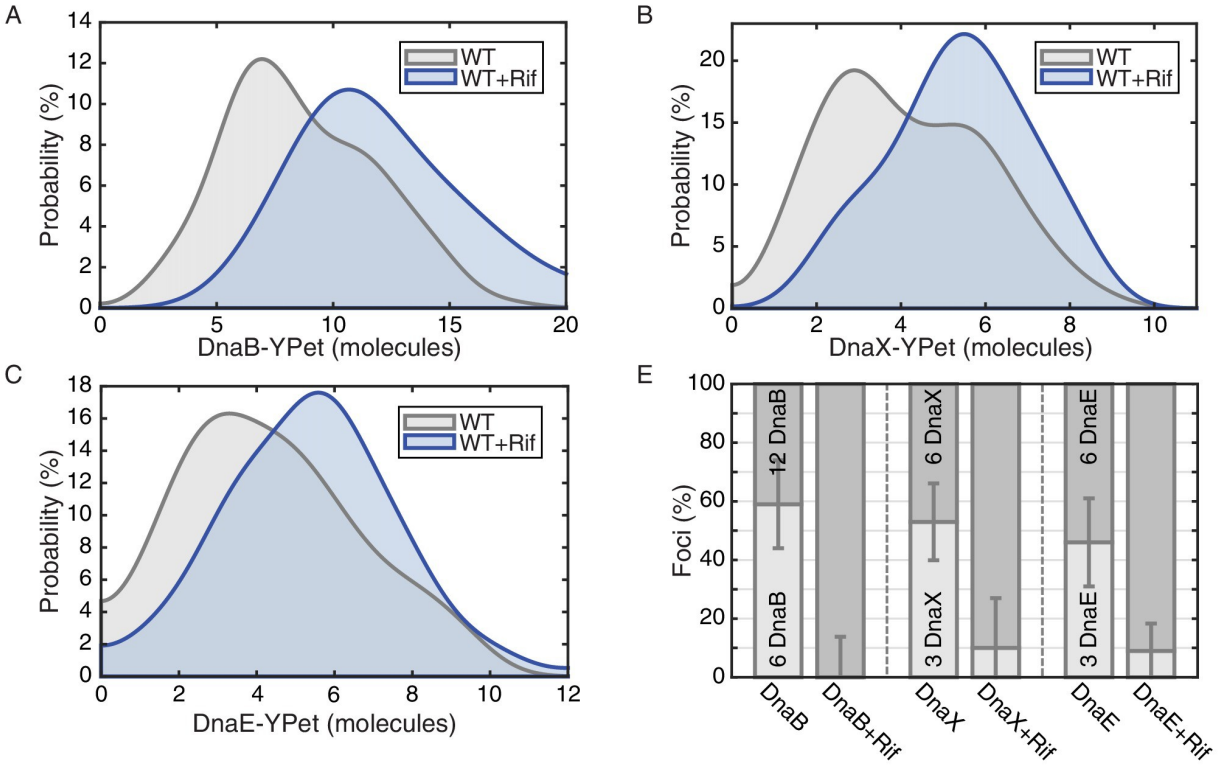

Figure 4.16: ***E. coli* stoichiometry distributions shift similarly to those in *B. subtilis* under transcription inhibition.** Stoichiometry distributions are represented using kernel density estimation (N = 53178 factories). (A) Estimated DnaB stoichiometry suggests that two hexameric helicases are present in most replication factories in the absence of transcription (blue). However, under normal conditions (gray) roughly half of factories consist of only a single helicase. (B) Transcription-inhibition increases the number of factories having higher DnaX stoichiometry. (C) Transcription-inhibition increases the number of factories having higher DnaE stoichiometry. (D) Estimates for the relative abundance of first and second order stoichiometry sub-populations are determined by fitting the distributions of estimated stoichiometries (Fig. 4.17 and Table 4.3).

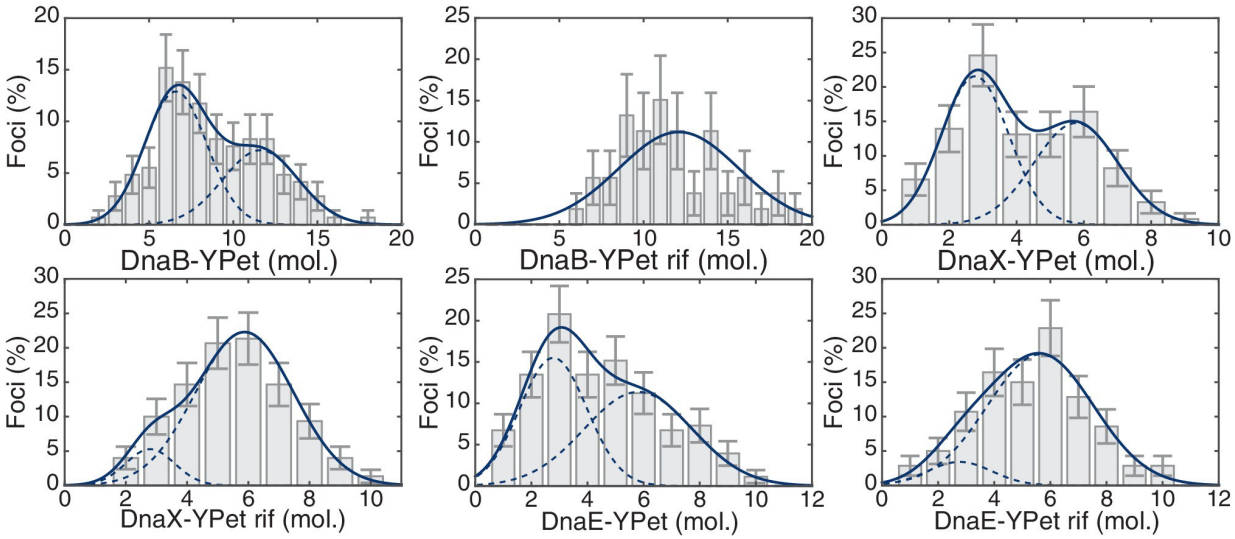

Figure 4.17: **Maximum likelihood fits to E. coli stoichiometry distributions.** The most likely two Gaussian model is selected varying the Gaussian means, widths and peak intensities. Maximum likelihood models (blue) and comprising Gaussians (dashed blue) are shown over histogram distributions. Error bars represent counting error. Error bars represent counting error. Calculated maximum likelihood fit parameters are summarized in Table 4.3.

	T (min)	$N_{\tau<T}$	$N_{\tau\geq T}$	τ_{min} (min)	$\bar{\tau}$ (min)	$\hat{k}$ (min^{-1})	$\bar{\tau}_{calc}$ (min)	N_C
Wild-Type	20	171	46	4	9.1	0.12	9.4	4.2
Rif-Treated	20	42	35	4	12.3	0.05	21.6	1.8

Table 4.4: **Parameters used in *E. coli* complex lifetime calculation.** The parameters summarized above are defined as follows: T is the duration of the experiment, $N_{\tau<T}$ is the number of complexes interpreted to disassemble before the end of the experiment, $N_{\tau\geq T}$ is the number of complexes surviving the length of the experiment, τ_{min} is the minimum observable complex lifetime, $\bar{\tau}$ is the empirical mean of the $N_{\tau<T}$ observable lifetimes, $\hat{k}$ is the calculated disassembly rate, $\bar{\tau}_{calc}$ is the calculated mean lifetime (i.e. $1/\hat{k}$), and N_C is the calculated number of conflicts per 40 min of replication.

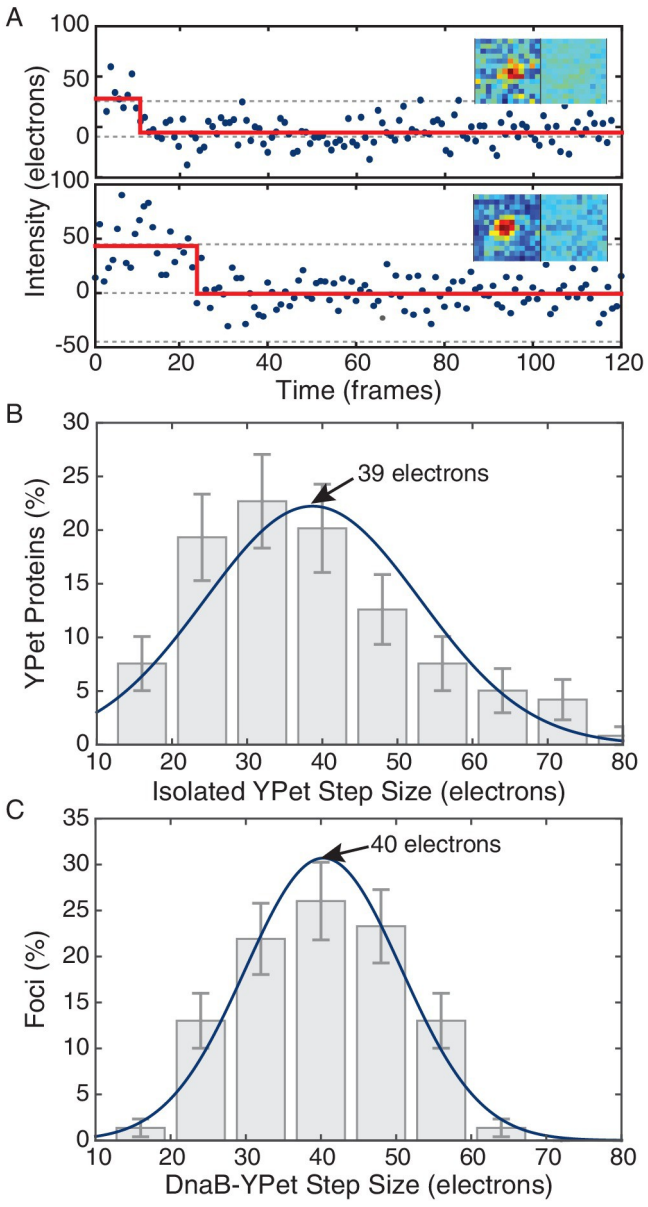

Figure 4.18: **Comparison of *in vivo* and *in vitro* step sizes for YPet.** (A) Bleaching traces for two different surface immobilized YPet proteins. Inserts show the mean fluorescence images in the bleached and unbleached states. (B) Maximum likelihood fits to unitary intensity step distributions for isolated YPet *in vitro*. (C) Example *in vivo* unitary step distribution for DnaB-YPet in *E. coli* and its maximum likelihood fit. For all experimental conditions in *E. coli*, unitary step distributions were peaked within 18% found in vitro value.

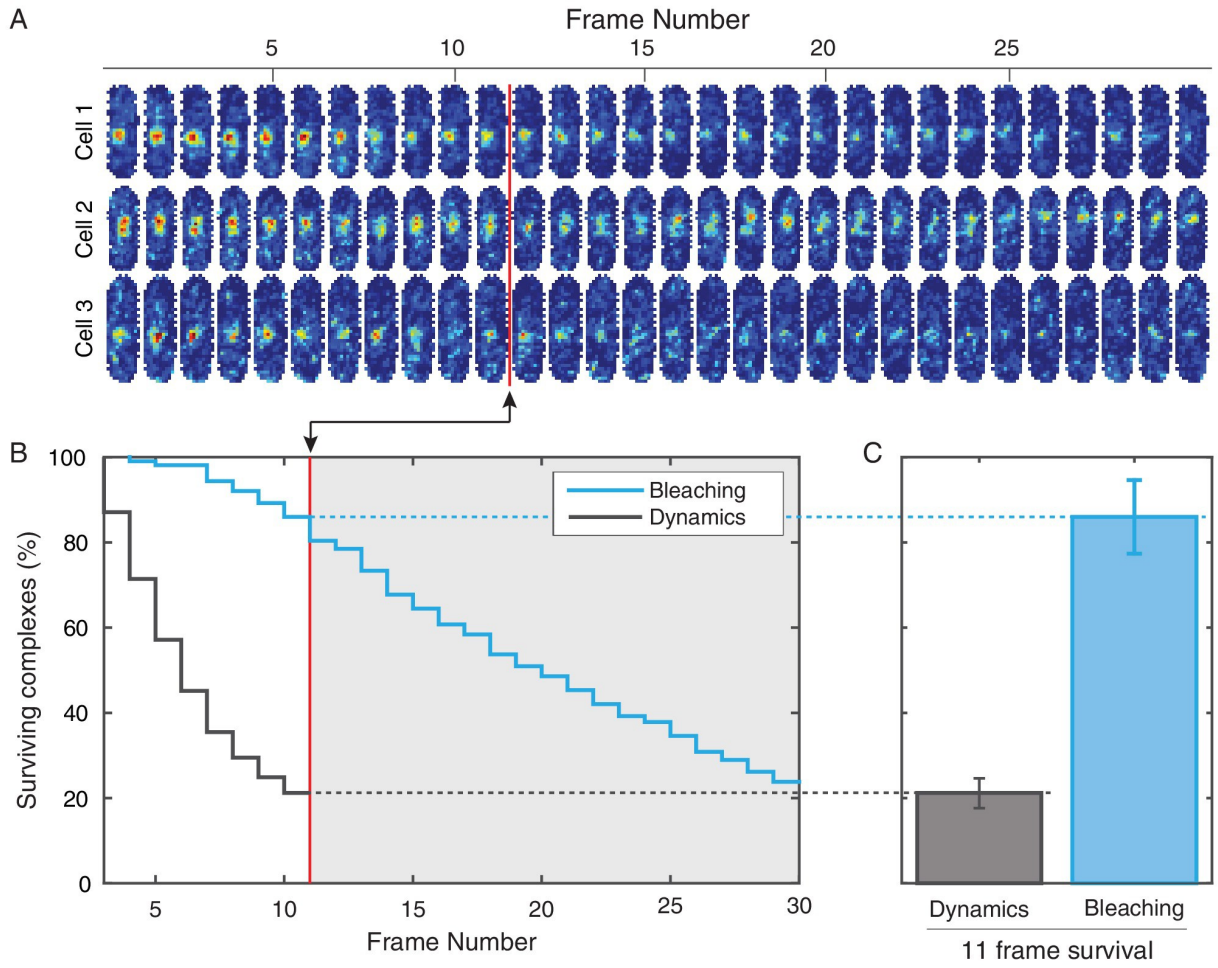

Figure 4.19: **Photobleaching minimally affects the complex lifetime experiments in *E. coli*.** (A) To isolate the effects of photobleaching, cells are imaged with the same intensity and exposure as in the dynamics measurement, but with no delay between frames. Red line indicates equivalent of the 11 frame time course used in the dynamics measurement. (B) Automated tracking used for the dynamics measurement was applied to the bleaching data. Survival curve (blue) shows the fraction of complexes that were successfully tracked through a given number of frames (N = 214). Survival curve for the dynamics experiment (gray) is shown for comparison (C) 86% of complexes were traceable for 11 frames. In the dynamics measurement, only 21% of complexes survive the duration of the experiment.

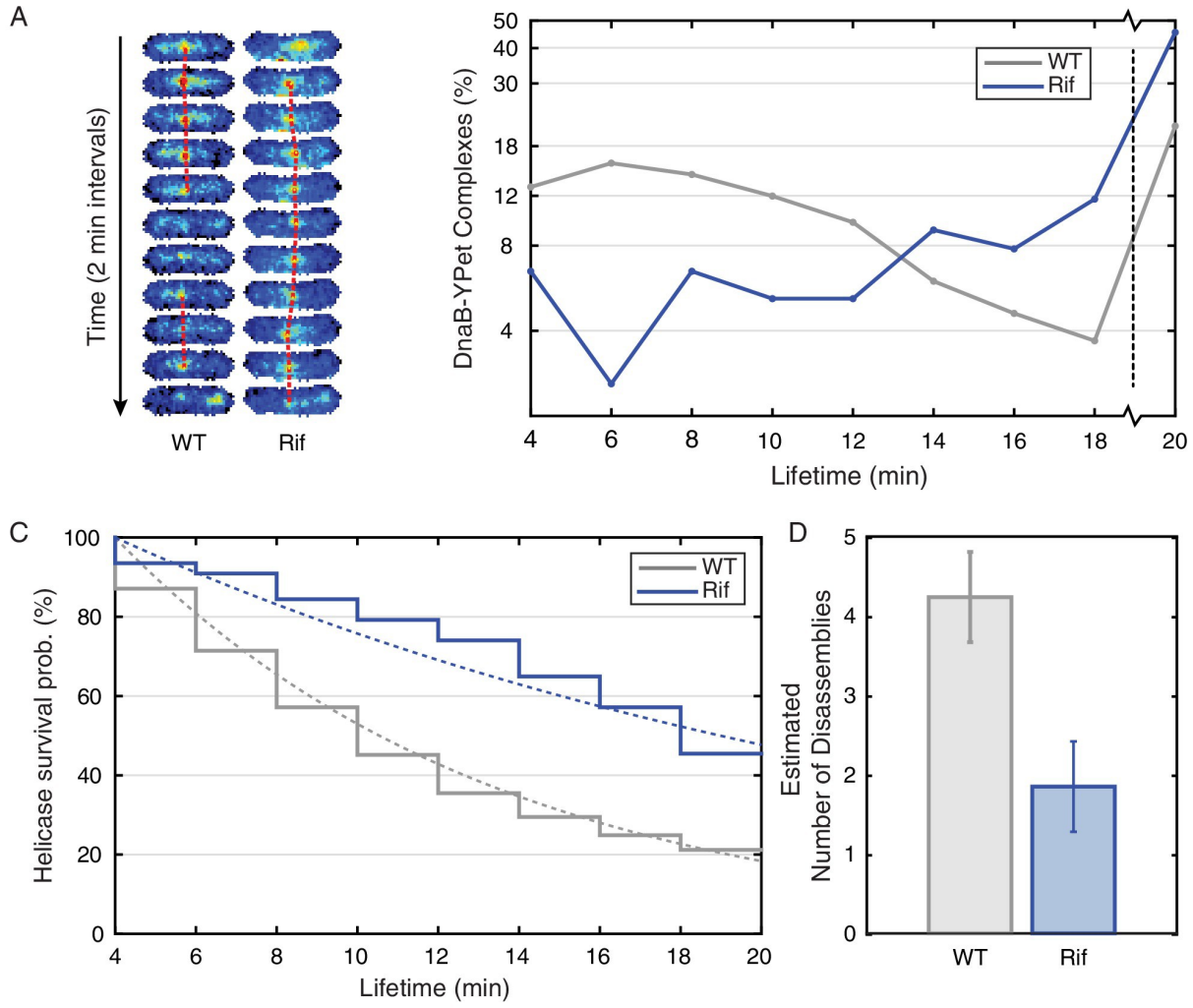

Figure 4.20: **Helicase complex dynamics in *E. coli* captured by time-lapse imaging.** (A) Typical frame mosaics for dynamics of the helicase in two conditions: Untreated (WT), rifampicin-treated (Rif). The helicase complexes in WT cells were observed to be intermittent: assembling and disassembling on the timescale of minutes. The helicase complexes in rifampicin-treated cells were observed to be more persistent. The complexes were tracked by an automated algorithm (red). (B) Distribution of helicase complex lifetime in WT (gray, N = 217 complexes), rifampicin-treated (blue, N = 77 complexes). Data collection is limited to 20 min. (C) Probability of helicase survival as a function of helicase complex lifetime. Solid lines represent the empirical survival curves. Dashed lines show fits determined by maximum likelihood estimation. (D) Estimated number of disassembly events per cell cycle using the Poisson process model (see also Table 4.4). Simulating 100,000 distributions with the same

rate parameter and number of complexes as the observed distribution generated error bars. Simulated distributions were then fit, and the width of the rate parameter distribution was used to quantify the error.

We further show that disruption of the restart process leads to permanent disappearance of the helicase complex. DnaC is required for the loading of the replicative helicase (DnaB), and is recruited to rescue stalled replication forks [134]. Here, we disrupted the restart process using a temperature-sensitive version of the helicase loader protein (*dnaC2* allele) [91]. We visualized DnaB-YPet in the *dnaC2* mutant under non-permissive conditions (37°C) using the same technique as in the dynamics experiments (Fig. 4.21A). Helicase complexes observed in the first frame disassembled before the end of the time course in 86% of cases. Furthermore, new complexes rarely developed (15% of the time). To further quantitate this data, we calculate the probability of observing a focus as a function of time in cells that have at least one focus. In the strains containing the *dnaC2* allele, this probability decreases throughout the time series, consistent with failure to restart (Fig. 4.21B). This observed loss of foci is too rapid to be explained by replication termination (see Fig. 4.21B, null hypothesis). In contrast, the probability of observing a focus in the wild type strain (at 37C) does not depend strongly on time, indicating that the replisome is in steady state due to its ability to restart. These data are consistent with reports suggesting that *dnaCts* alleles are ineffective for synchronization because replication does not proceed to completion [130, 150]. Overall, these findings are consistent with both our observations in *B. subtilis* and our interpretations of the results of the *E. coli* stoichiometry measurements, suggesting that *E. coli* cells also experience pervasive replisome instability due to transcription.

4.2.7 Transcription inhibition increases the rate of replication

To test the replisome-instability model, we measured replication rates *in vivo* by radioactive thymidine incorporation assays in *B. subtilis*, in both wild-type (without DnaC-GFP) and DnaC-GFP strains. It takes roughly 40 min to replicate 4.2 and 4.6 Mb of DNA in *B. subtilis*

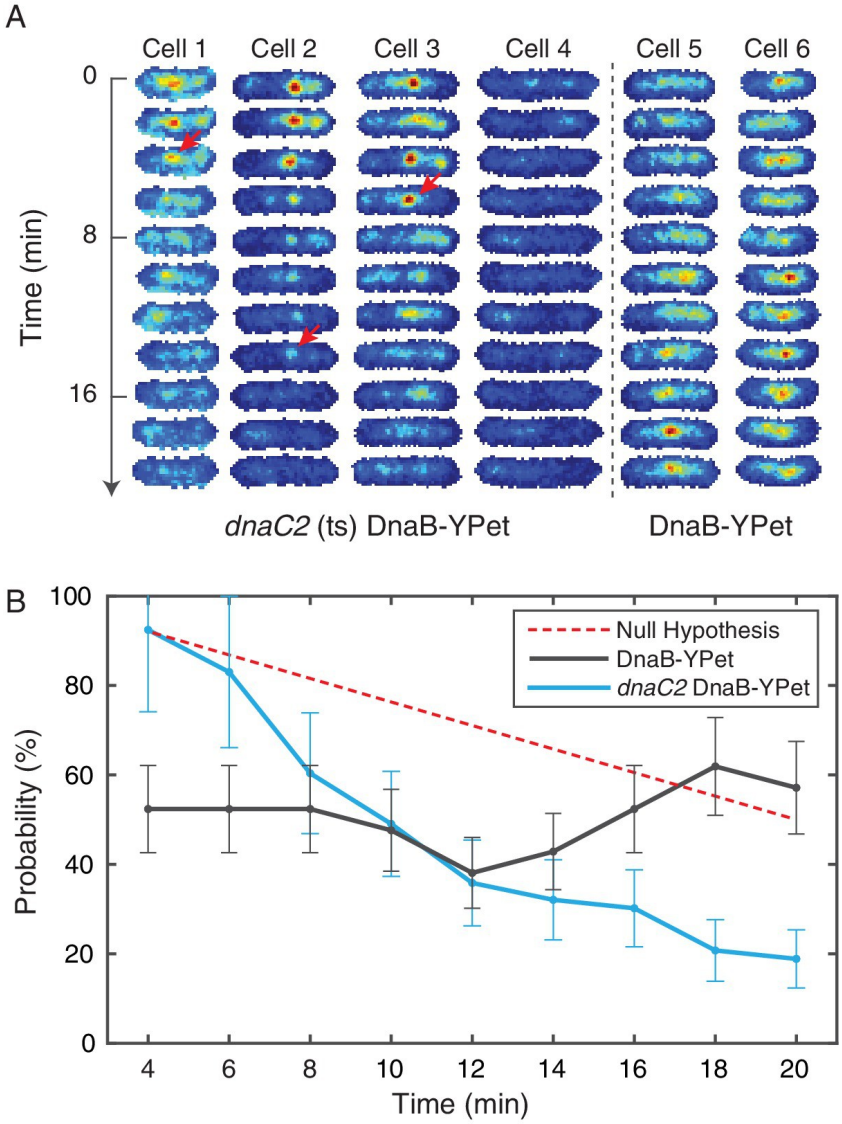

Figure 4.21: **Disruption of restart using temperature sensitive DnaC.** (A) DnaB-YPet imaged at 2 min intervals in strains with and without the *dnaC2* allele at 37°C (non-permissive for the temperature sensitive strain). In the temperature sensitive strain, either a complex is observed in the first frame (e. g. cells 13) or the cell generally does not develop complexes during the time course (e. g. cell 4). When a complex is observed, it most commonly disappears before the end of the time course. Once disappearance has occurred (red arrows), no new stable complexes are observed in the majority of cells (85%). In contrast, the wild-type strain (e. g. cells 56) shows intermittent foci (as observed in the dynamics experiments at 30°C), and complexes may be assembled later in the time series. (B) Probability of observing a focus as a function of time in cells where at least one trajectory

is observed. DnaC is involved both in initiation of replication at the origin and the rescue of stalled forks. The theoretical Null hypothesis curve (red) assumes a continuous 40 min replication cycle imaged for a random 20 min window. However, the probability of observing a focus in the *dnaC2* allele (blue, N = 72) decreases more quickly suggesting that the helicase is disassembled prior to completion of replication. Additionally, in the wild-type strain, the probability of observing a focus is roughly independent of time, indicating that both disassembly and re-assembly events are present.

and *E. coli*, respectively; at the 500 bp/s - 1 kb/s measurements from *in vitro* and *in vivo* experiments [151], which assume replisome stability during a replication cycle. However, our observation of frequent conflicts (during which the replisome stalls) predicts that the *in vivo* rates should increase if conflicts are reduced. To test the prediction of our model, we measured the rate of thymidine incorporation with and without rifampicin treatment, as well as in the *rpoB** strains - two different conditions where we observed the stabilization of the replisome. As predicted, both transcription perturbations (rifampicin and *rpoB**) increase the replication rate: Thymidine incorporation rates in rifampicin-treated cells are roughly 60 - 65% higher than that measured for untreated cells (Fig. 4.22A). The consistency of the incorporation rate in the *rpoB** strain with rifampicin treatment suggests that the observed increase in rates is probably not an artifact of changes in nucleotide uptake in rifampicin-treated cells [152]. This interpretation is also supported by the insensitivity of *rpoB** strains to rifampicin treatment with regards to thymidine incorporation rates. In addition, measurements of the thymidine incorporation rates in DnaC-GFP cells provide further evidence that DnaC-GFP is functional during slow growth: We did not detect any difference in thymidine incorporation rates between the wild-type and the DnaC-GFP strain under these conditions (Fig. 4.22A). These results provide further evidence that the rate of replication is severely inhibited by transcription in wild-type cells.

The observed transcription-induced replication conflicts should be at least partially due to the ribosomal RNA and protein genes. We previously showed that replication restart proteins preferentially associate with these regions during fast growth [3]. Although somewhat

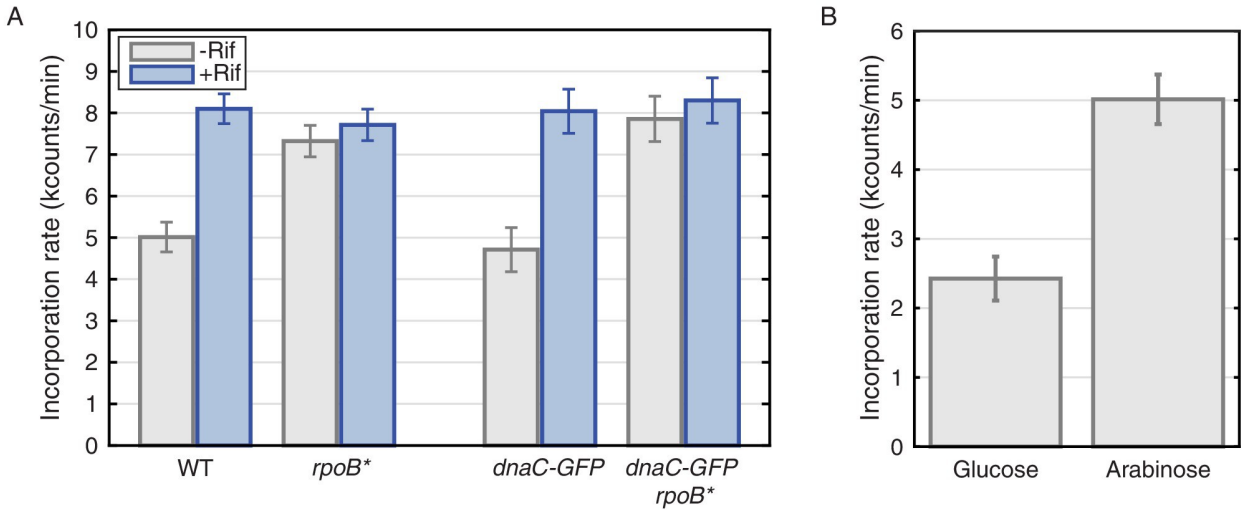

Figure 4.22: **Thymidine incorporation assays determine the effect of transcription on replication rate.** (A) DNA replication rates increase upon transcription-inhibition and in *rpoB** strains. Thymidine incorporation assays were used to measure the relative rates of DNA replication in WT and DnaC-GFP strains, with (blue bars) and without (gray bars) perturbations to transcription by rifampicin treatment (Rif) or *rpoB** cells. Note that *rpoB** strains are resistant to rifampicin and therefore show no additional increase in replication rate after rifampicin treatment. (B) DNA replication rates increase in cells grown in minimal arabinose medium, relative to glucose, where transcription of rRNA and other ribosomal protein genes is higher.

counterintuitive, our model predicts that replication rates should be slower during fast growth due to conflicts with transcription. To test this prediction, we measured replication-rates in slow and fast growth conditions. Despite the faster growth rate of *B. subtilis* on minimal glucose versus arabinose medium, we observe significantly higher thymidine incorporation rates in arabinose compared to glucose, where ribosomal RNA and protein genes are much more highly expressed and are a major source of conflicts (Fig. 4.22B, and [153]).

4.3 Discussion

The basic principles underlying the mechanism of genome replication are thought to be well established, especially in bacteria. Yet, a great part of this work has relied on a combination of *in vitro* reconstitution, from which potentially essential cellular factors and/or processes are absent, or *in vivo* ensemble measurements which are sensitive only to ensemble-averaged cellular behavior. Based primarily on these results and the absence of significant evidence to the contrary, DNA replication was thought to be largely a continuous process that initiates at the origin and moves processively to the terminus. However, there is a contradiction to this widely-accepted model: Our laboratory and others have demonstrated that the replication process can be disrupted from a number of concurrent processes including transcription. The resulting replication conflicts have dire consequences in the absence of rapid and efficient replication restart. If transcription is both an obstacle to the replisome and active throughout the cell cycle, how can the replication process be continuous? If conflicts can occur, what is their frequency? And if these events are frequent, what are the consequences for the replication machinery?

Using *in vivo* single-molecule fluorescence microscopy techniques in combination with a genetic and cell-biology toolkit developed for studying replication conflicts, we measured the stoichiometry and lifetimes of replication complexes in single cells with single-molecule resolution. The results of the replicative helicase complex stoichiometry experiments could be explained by a number of models. It is conceivable that the replisome traverses one of the two arms of chromosome much more rapidly than the other, reaching the terminus and

disassembling much earlier than the second replisome complex. However, genomic DNA analyses have already demonstrated that in growing cells, the copy number of the two arms are roughly equal, implying that the average rate of replisome movement on the two arms is about the same [50]. Another model for the low observed replisome stoichiometry is the observation of individual forks, as has previously been reported in *E. coli*. But, our own quantitative characterization of replisome localization reveals factory-like (co-localized) replisome positioning for 80% of the cell cycle in both *E. coli* and *B. subtilis*, which cannot account for the low observed stoichiometry in our own experiments [7] or previous work [24]. The data presented here is most consistent with the model that elongating replisomes are frequently disrupted (roughly 50% of the time) and disassembled during a single replication cycle.

The direct visualization of the replisome with SMFM in two different bacterial species strongly suggests that the replisome is frequently destabilized by transcription. The observation that these inferred disassembly events are transcription-dependent not only suggests that these events correspond to replication-transcription conflicts previously observed via biochemical means, but also that vast majority of replication conflicts are the result of transcription.

Our findings suggest that replicating cells often possess only a single active replisome complex. Together, the stoichiometry and dynamics of the replicative helicase and polymerase complexes as well as the replication rate measurements suggest that the replisome is subject to pervasive disassembly and reassembly events. These observations provide a potential explanation for the essentiality of replication restart proteins at least under growth conditions where rDNA and ribosomal protein genes are relatively highly expressed and PriA is essential.

Due to the low intensity of single fluorescent molecules, both the stoichiometry and dynamics experiments are technically challenging. In addition, the bacteria themselves are potentially subject to photo damage at the intensities required to resolve single molecules, and this photo damage itself affects cellular processes and replication in particular. Therefore, it is essential to consider these measurements in the context of additional corroboratory

biochemical and genetic evidence. The essentiality of restart proteins and the measurement of the replication rates provide independent lines of evidence for the pervasive disassembly model.

We propose that the number of replication-transcription conflicts is significantly greater than previously suggested: far from being rare, we infer from the data presented here that conflicts are generic and occur multiple times per cell cycle. If so, then these encounters frequently compromise the integrity of the replisome, leading to discontinuity in the process of DNA replication. These observations therefore present a new model for the canonical states of DNA replication, arguing against the well-accepted model that the process of replication is highly continuous *in vivo*.

4.4 Materials and Methods

4.4.1 Growth curves

Cell cultures were grown for 16 hours in minimal arabinose (or LB) medium supplemented with threonine and tryptophan until cells reached log phase growth. Cells were then diluted back to an OD_{600} of 0.1 and monitored at 25 min intervals.

4.4.2 Western blots

B. subtilis cells harboring the *dnaC-gfp* allele as well as the isogenic wild-type were grown to an OD_{600} of 0.3 and then harvested. Total protein was prepared in SDS loading buffer, and the same volume was loaded on each of the lanes shown. Western blots were probed in Odyssey blocking buffer (Li-Cor P/N 92740100) with either rabbit anti-DnaC polyclonal antibodies (1:5000 dilution), or rabbit anti-GFP polyclonal antibodies (1:7500 dilution). Blots were then probed with Li-Cor secondary anti-rabbit IR800 antibody at 1:15000 dilution. Blots were scanned and quantified on a Li-Cor Odyssey.

4.4.3 Strain list and strain construction

All strains used in this study are listed in Table 4.5. The strain containing the IPTG-inducible *lacZ* ($P_{spank(hy)}$-*lacZ*) construct (HM1316), was built by transformation of the plasmid (pHM149) encoding *thrC::*$P_{spank(hy)}$-*lacZ* into strain HM262. The *polC-ypet* strain (HM604) was constructed by transforming pHM93 (3‘ *polC*-LEGSG-*ypet* A206K) into wild-type JH642 *B. subtilis* cells. *rpoB** mutants were isolated by plating 3 ml of saturated *B. subtilis* cells harboring *dnaC-gfp*, or *E. coli* cells harboring *dnaB-ypet*, on LB+0.3 µg/ml rifampicin plates. Revertant colonies were isolated and their *rpoB* gene was sequenced. Rifampicin revertant strains having mutations within *rpoB* were considered *rpoB** strains. We examined mutants harboring the H482Y mutation which is in cluster 1 of the *rpoB* gene. The strains containing this mutation do not display growth defects on LB or LB supplemented with rifampicin, as was also previously reported [147]. HM2475 (PriA CRISPR precursor) was constructed by transforming the plasmid pHM273 into HM262 to incorporate the sgRNA at *amyE*. The full PriA CRISPR resulted from transformation of gDNA from HM1500 into HM2475. Additional control strain HM2387 containing only dcas9 resulted from transformation of gDNA from HM1500 into HM1. The *E. coli* strain containing the *dnaC2* allele was constructed by P1 transduction of *dnaB-ypet* from HM1391 into PAW542.

4.4.4 Localization of replisome components

We ensure that fluorescent foci localize near midcell during active replication, consistent with localization to the replication factory. Stationary phase cells harboring replisome fusions do not have foci indicating that replication is required for localization and that the observed foci are not a result of protein aggregates.

4.4.5 Cell preparation for microscopy

The following protocol was used to prepare cells for microscopy: (i) Cultures were grown overnight in a shaking incubator at 30°C. (ii) *E. coli* was cultured in M9-minimal media (1X M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% Glycerol, 100 µg/ml each Arginine, Histidine,

Strain #	Genotype	Species	Reference
HM1	*trpC2 pheA1*; wild-type	*B. subtilis*	[154]
HM262	*dnaC-gfp-spec*	*B. subtilis*	[40]
HM1312	*dnaC-gfp rpoB* (H482Y)	*B. subtilis*	This study
HM1316	*dnaC-gfp-spec thrC::*$P_{spank(hy)}$*-lacZ::erm*	*B. subtilis*	This study
HM604	*polC-ypet-erm*	*B. subtilis*	This study
HM1275	*polC-ypet rpoB* (H482Y)	*B. subtilis*	This study
HM1500	*lacA::*P_{xl}*-dcas9-erm*	*B. subtilis*	[155]
HM2387	HM1 *lacA::*P_{xl}*-dcas9-erm*	*B. subtilis*	This study
HM2475	*dnaC-gfp-spec amyE::*P_{veg}*-sgRNA-priA-cat*	*B. subtilis*	This study
HM2486	*dnaC-gfp-spec amyE::*P_{veg}*-sgRNA-priA-cat lacA::*P_{xl}*-dcas9-mls*	*B. subtilis*	This study
HM1318	wild-type (AB1157)	*E. coli*	[156]
HM1319	*dnaB-ypet-kan*	*E. coli*	[24]
PAW912	*dnaX-ypet-kan*	*E. coli*	[91]
PAW909	*dnaE-ypet-kan*	*E. coli*	[91]
PAW542	*dnaC2 thrA::tn10-tet*	*E. coli*	[91]
PAW1182	*dnaB-ypet-km dnaC2 thrA::tn10-tet*	*E. coli*	This study

Table 4.5: **Strains used in this study.**

Leucine, Threonine and Proline and 10 µg/ml Thiamine hydrochloride). (iii) *B. subtilis* was cultured in minimal arabinose media (1x Spitzizens salts (3 mM $(NH_4)_2SO_4$, 17 mM K_2HPO_4, 8 mM KH_2PO_4, 1.2 mM $Na_3C_6H_5O_7$, 0.16 mM $MgSO_4$-$(7H_2O)$, pH 7.0), 1x metals (2 mM $MgCl_2$, 0.7 mM $CaCl_2$,. 05 mM $MnCl_2$, 1 µM $ZnCl_2$, 5 µM $FeCl_2$, 1 µg/ml thymine-HCl) 1% arabinose, 0.1% glutamic acid, 0.04 mg/ml phenylalanine, 0.04 mg/ml tryptophan, and as needed 0.12 mg/ml tryptophan.) (iv) Overnight cultures at an OD_{600} of 0.3-0.9 were diluted back to an OD_{600} of 0.2 and incubated again for about 2 hr until they reached approximately OD_{600} 0.4. (v) Cells were then concentrated by a factor of 10 immediately before imaging by centrifugation. (vi) 1 µL of concentrated cell culture was spotted on a thin 2% (by weight) low-melt agarose pad (invitrogen UltraPure LMP Agarose, Cat. no. 16520050). (vii) The sample was sealed on the slide under a glass cover slip using VaLP (a 1:1:1 Vaseline, lanolin, and paraffin mixture.) (viii) For rifampicin treatment, a final concentration of 100 µg/mL was added to both the agarose pad and the liquid culture immediately prior to imaging. (ix) Cells were imaged at 30°C.

4.4.6 Microscope configuration

Imaging was performed on a custom-built inverted fluorescence microscope. To avoid light loss at the phase plate, cells were imaged through a Nikon CFI Plan Apo VC 100 1.4 NA objective. An external phase plate (Ti-C CLWD Ph3 Annulus Module) was inserted into the beam path during phase-contrast imaging and retracted for fluorescence imaging to avoiding decreased signal due to the neutral density annulus on the phase plate. The microscope focus was controlled by an IR-autofocus system. In short, an infrared beam was reflected off the coverslip-media interface and the back reflection is detected by a position-sensitive detector (PSD). The displacement was then processed using a PID feedback system to control the z-height of the piezo stage.

Fluorophores were excited by laser illumination. A Coherent Sapphire 50 mW 488 nm or 150 mW 514 nm CW laser is used to excite GFP and YPet, respectively. We expand the beam diameter to provide even illumination over the field of view. The excitation intensity

	Laser power at objective (mW)	Exposure time (ms)	Med. background 1st frame (electrons/pixel)
Bleaching (GFP)	1.3	300	87
Bleaching (YPet)	0.4	300	21.4
Lifetime (GFP)	1.1	600	76
Lifetime(YPet)	0.1	600	4.5

Table 4.6: **Microscopy parameters.**

is controlled via an Acousto-Optic Tunable Filters (AOTF, AA Opto-Electronic AOTFnC-400.650). Images were collected on an iXon Ultra 897 512 512 pixel EMCCD camera. The microscope system is controlled by Micro-Manager.

4.4.7 Imaging protocol for bleaching analysis

A single phase contrast image is followed by a stack of 300 ms fluorescence images (settings summarized in Table 4.22). Fluorescence images are continued until all foci become photobleached (∼120 frames). The imaging time is sufficiently fast that photodamage to the cells is not an issue. Because YPet is brighter, and the 514 nm laser tends to excite lower cellular background intensities (see methods for calculation of background), we were able to image at lower laser power.

4.4.8 Bleaching analysis

The method for determining the in vivo stoichiometry of fluorescent-fusion proteins in active replisomes is outlined below. Our protocol is based-upon the method described recently by Reyes-Lamothe et al. [24] with a number of modifications described in detail below.

4.4.9 Segmentation of the cells from the phase-contrast image

Cells are initially imaged in phase contrast for the purpose of segmentation, the computational process of determining cell boundaries from an image (Fig. 4.3A). Imaging for bleaching analysis is sufficiently fast (36 s) that the cells do not grow appreciably during

the process, allowing only a single phase image to be collected at the beginning of data acquisition. We then used the Wiggins Labs custom segmentation tool (superSegger) on the phase image to generate cell masks for analysis [117]. The Wiggins Lab's custom image processing software (superSegger) and documentation is available at the following link: http://mtshasta.phys.washington.edu/website/SuperSegger.php

4.4.10 Location and scoring of foci in the summed image

Firstly, to remove xy sample drift, all fluorescence images in the stack are aligned against the first. This corrected alignment is retained for the rest of analysis. Analysis of cell fluorescence begins by computing the summed image, which constitutes of summing the intensity values over all frames in the image stack and then applying a one pixel radius Gaussian blur. We then watershed the conjugate of the summed image to generate sub-regions around each intensity maximum (Fig. 4.3). Taking the union of these regions with the cell masks excludes regions external to the cells from analysis. These sub-regions will be called intensity regions. The intensity profile in each intensity region is modeled by a Gaussian distribution:

$$G\left(\overrightarrow{x}\right) = G_G \exp\left[-\frac{\left(\overrightarrow{x} - \overrightarrow{x_0}\right)^2}{2b^2}\right] + G_0, \tag{4.1}$$

where, the peak amplitude Gaussian intensity is defined by parameter G_G and the parameter G_0 defines the background intensity. The focus position is parameterized by $\overrightarrow{x_0}$ and the focus width by parameter b. Within each intensity region, a three pixel radius circular region is centered on the position of the maximum-intensity pixel. The Gaussian intensity model is then fit inside the union of each intensity region and the corresponding circular region. The resulting focus position from the fit ($\overrightarrow{x_0}$) is then retained throughout the analysis.

Once the fit has been performed in each region in the cell, these foci are excised from the cell image using a mask radius of 3 pixels. We compute the per-pixel mean and standard deviation of the intensity (I_B and δI, respectively) in the remaining cell area. We define the background-subtracted intensity:

$$\Delta I \equiv I - I_B. \tag{4.2}$$

In general, we found that the background-subtracted intensity had better statistical properties than the Gaussian fit parameters.

4.4.11 Focus scores

The focus score σ is a measure of the statistical significance of a focus. The integrated background-subtracted intensity I_A in a region (radius three pixels and area $A_M = \pi r^2$) is computed. We define the score:

$$\sigma = \frac{I_A}{\delta I \sqrt{A_M}} \tag{4.3}$$

where the factor of $\sqrt{A_M}$ in the denominator accounts for extended area A_M over which the intensity is integrated. The intensity noise at each pixel is assumed to be uncorrelated and therefore the expected standard deviation in the integrated intensity in area A_M is $\delta I \sqrt{A_M}$.

All foci scoring two or smaller tended to be located randomly throughout the cell, inconsistent with localization to the replication factory. We therefor retain foci with scores greater than two for further analysis.

4.4.12 Determination of the focus and background intensities in individual frames

Returning to the stack of fluorescence images, we compute the raw focus intensity (I_R) in each frame by summing the intensity within a disk (radius three pixels) centered on the position of the focus ($\overrightarrow{x_0}$, determined in the previous step). The per-pixel background intensity was again computed by excising each locus (using a mask radius three pixels) and then computing the mean intensity over the remaining cell area (I_B). We define the focus intensity (I_F) as the difference between the raw focus intensity and the total background intensity within the area of the 3-pixel-radius mask (A_M):

$$I_F = I_R - A_M I_B \tag{4.4}$$

This background subtraction method is illustrated for a focus in Figure 4.4, Panel B. The intensity, I_F, is interpreted as the intensity of the fluorophores at the focus for each frame and plotted to form bleaching traces.

4.4.13 Analysis of bleaching traces

A key step in the determination of the single-molecule protein stoichiometry is the determination of the fluorescence intensity of a single fluorescent molecule, which M. Leake has called the unitary step [135]. Two independent methods were used to determine this fluorescence intensity: (i) The intensity of GFP and YPet molecules were measured *in vitro* and (ii) the fluorophore intensity was inferred for each complex *in vivo* from the analysis of the bleaching curves [135].

In both cases, the intensity must be inferred from a noisy intensity trace. The method for trace analysis of Leake et al. [24, 135] was:

1. Compute the smoothed intensity trace $I'(t)$
2. Compute the Pairwise Probability Density Distribution (PPDD) of $I'(t)$
3. Compute the Power Spectrum of the PPDD
4. Infer the unitary intensity step (ΔI) from the power spectrum
5. Fit the intensity trace to an exponential to determine the initial intensity I_0
6. The inferred stoichiometry is $n = I_0/\delta I$

4.4.14 Smoothing the intensity trace

Because the raw intensity data has too large a variance to directly detect steps by analysis of the pairwise differences, Leake et al. applied an Edge-Preserving Chung-Kennedy (SED) Filter to smooth the intensity data [157]. Instead, we use a parameter-free change-point (CP) analysis to idealize the bleaching curves [158]. The analysis of simulated intensity traces (Fig. 4.23) with similar statistical properties to the observed data illustrates our method for calculating stoichiometry, and the performance of our CP idealization compared to the SED filter. We provide a more detailed discussion of the comparison of the CP method to competing methods in the following sections.

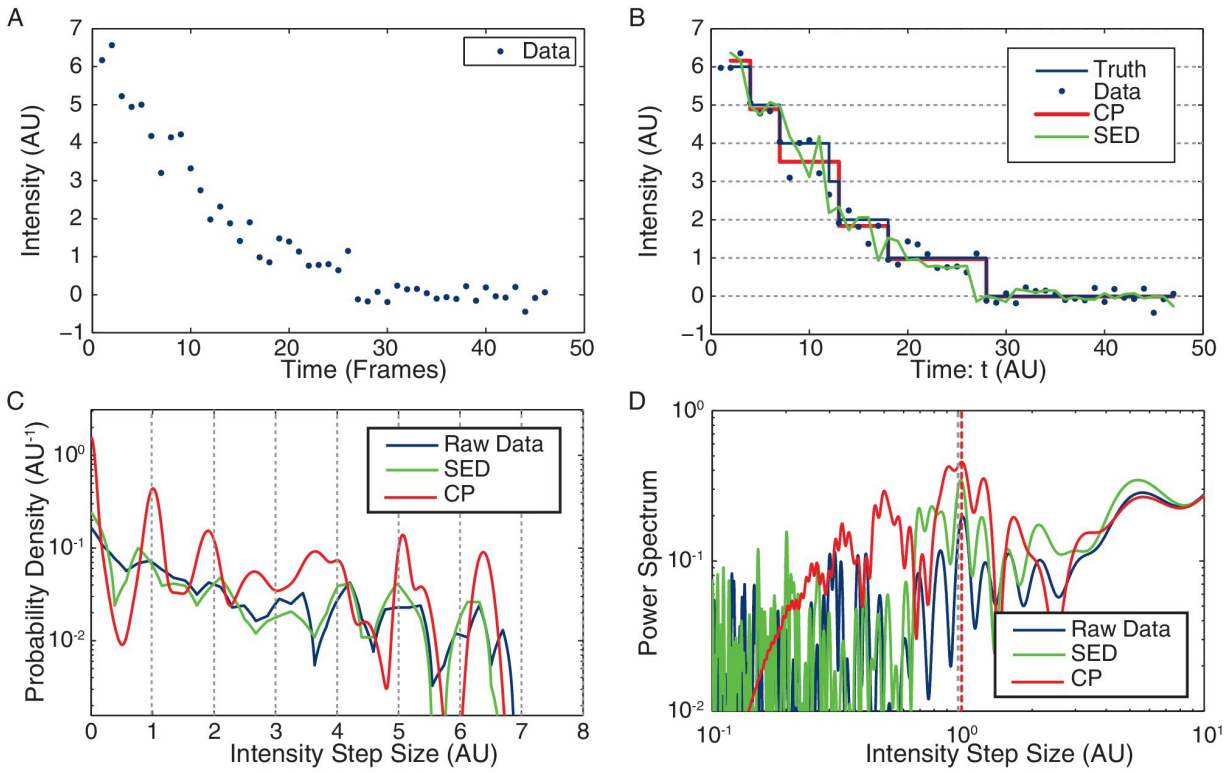

Figure 4.23: **Stoichiometry calculation applied to simulated data.** (A) Simulated intensity data. A bleaching experiment was simulated to demonstrate the performance of the CP Algorithm against the SED filter. The noise and trace length were chosen to closely approximate the observed data. (B) Idealized intensity traces. The data (blue dots) is identical to Panel A. True simulated mean (Truth) is shown in blue. There is excellent agreement between the truth (blue) and the CP idealization (red). The SED filtered trace is shown in green. (C) Pairwise intensity difference probability distribution (PDPD). The black dotted lines represent multiples of the true step-size 1 AU. (D) Power spectrum of the PDPD. The largest peak in the power spectrum is selected as the unitary step (red dotted). The true step-size is also shown (black dotted).

4.4.15 Pairwise distribution

Figure 4.23C shows the PPDD for the raw simulated data (blue), CP idealization (red) and SED filtered data (green). The black dotted lines are placed at integral multiples of the true step size (1 AU). Note that the peak widths in the CP idealization are computed from the Fisher Information [158]. The first peak in the pairwise distribution of the CP idealization is centered around the true value (1 AU), showing excellent agreement between the calculated and the true step sizes.

4.4.16 Power spectrum

Figure 4.23D shows the Power Spectrum of the PPDD. Again, the CP idealization shows a stronger enrichment in power at the true intensity step-size (1 AU). Note that the power spectrum is not an FFT, but rather a continuous representation of the Fourier transform of discrete data that is commonly applied in this context (e.g. [159]).

4.4.17 Determination of the unitary step

Leake et al. report that they select the first peak (with respect to increasing intensity) in power spectrum that is four standard deviations above the noise [135]. Clearly this method is ad hoc since lower noise would resolve higher harmonics of the unitary step, as are easily visible in the power spectrum for the CP idealization. Instead, we selected the largest peak in the power spectrum, as illustrated by the dotted red line (Fig. 4.23D). This approach resulted in the reliable determination of the unitary step in simulations. We show a histogram estimating unitary step size (relative to the true step size) in Figure 4.24. In our simulations, the Leake et al. method (using the SED filter) resulted in a greater number of failures to correctly identify the unitary step. Most worrying of all, the SED filter lead to a fairly frequent identification of the unitary step as roughly twice the true unitary step, leading to anomalous counts of complexes with half the true stoichiometry.

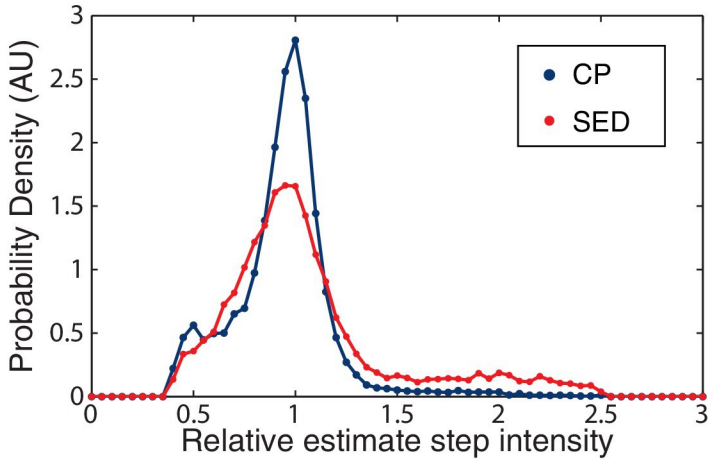

Figure 4.24: **CP versus SED filters in unity step determination.** The performance of the CP versus the SED filter is measured by a histogram of the relative size of the unitary step in simulated data. The CP idealization clearly results in a sharper distribution about the true value (unity).

4.4.18 Determination of the initial focus intensity

Because the probability of a fluorescent protein bleaching is proportional to the number of unbleached proteins, we expect the raw intensities follow a decaying exponential profile. To determine the initial focus intensity (I_0), we fit the background subtracted raw intensities to a model of the form:

$$I(t) = I_0 \exp(-t/t_b), \tag{4.5}$$

optimizing both I_0 and the characteristic bleaching time (t_b). We found that this method was more reliable than simply taking I_0 as the highest level of the filter based on simulated intensity traces (see Fig. 4.25). Although the highest level is less noisy than the exponential fit, we find highest level to be biased from below.

The mechanism for the bias generation for the highest-level method is as follows: The initial decay in intensity is too rapid to resolve all bleaching steps at high stoichiometry and therefore bleaching steps are merged at the beginning of the bleaching process. As a result the highest-level method typically averages over the first few true levels in the bleaching

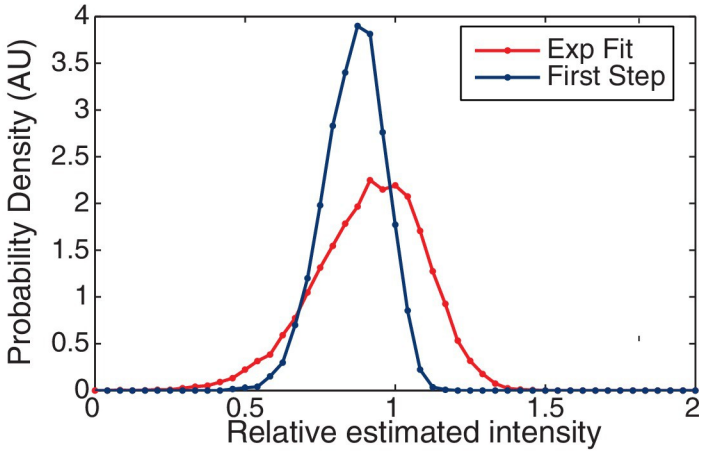

Figure 4.25: **Finding the initial intensity.** We simulated two approaches for determining the initial intensity of the trace: Exponential Fit and Highest-Level. The probability of the estimated intensity is shown relative to the true initial intensity. The Highest Level method is clearly biased from below relative to the Exponential Fit method which is centered around the true value (unity).

process. We note that, in general, both the highest-level and exponential fitting techniques produced acceptable results.

4.4.19 Data selection

After automatic processing, all traces are refereed by hand. We remove cells where:

1. There were segmentation errors.

2. Replisome foci have strong ellipticity (ellipticity >1.2).

3. The unitary intensity step size was not within a factor of 2.5 of the *in vitro* value.

4. Unitary peaks in the PPDD and its power spectrum not >10% larger than competing peaks.

5. Exponential fit to determine the initial intensity is >30% from the initial intensity measurement.

6. Focus localization was inconsistent with a replication factory.

While applying these data-selection rules had no qualitative effect on our results and conclusions, the selection rules did remove data points that were clearly aberrant from the analysis. All datasets were repeated on at least two occasions, with similar results.

4.4.20 Analysis of final count distributions

We divide the initial focus intensity by the unitary step size for a large number of cells, obtaining final count distributions for each strain. Many of these distributions were bi-modal. We fit all count distributions using a two-Gaussian mixture model where the Gaussian widths, means (μ_1, μ_2) and peak amplitudes were optimized. To avoid binning the data, we use a maximum likelihood process to determine the best-fit parameters. The fractional population of factories in the low and high stoichiometry populations (F_L and F_H) is proportional to the area of the model Gaussians peaked near the appropriate values. See Figure 4.14 and Figure 4.17 for maximum likelihood fits to all stoichiometry distributions not shown in the main figures. The fit parameters are represented graphically in main Figures 4.12D and 4.20D, and numerically in Table 4.2 and Table 4.3 for *B. subtilis* and *E. coli*, respectively. Note that the error in F_H and F_L was calculated by simulating 10,000 stoichiometry distributions with the same statistical properties as the corresponding empirical distribution. The width of the distribution of F_L (or F_H) based on the simulated distributions was taken as the error.

4.4.21 Kernel density estimates

For plotting the probability densities for protein stoichiometry, we used a Kernel Density Estimate (KDE) in order to avoid binning the data (e.g. [24]). A Gaussian kernel was used, and the optimal bandwidth was selected by minimization of the asymptotic mean integrated squared error.

4.4.22 Discussion of filtering algorithms

Because the raw intensity data has too large a variance to directly detect steps by analysis of the pairwise differences, Leake et al. applied an Edge-Preserving Chung-Kennedy (SED) Filter to smooth the intensity data [157], this approach appeared to have two important shortcomings: (i) The filter had what we considered poor performance under many scenarios (described below) and (ii) the filter uses two ad hoc filtering parameters which must be optimized. In our hands, the results of the SED filter were not particularly robust to parameter choice.

To eliminate the need to specify ad hoc parameters for the data analysis, we attempted to apply an objective parameter-free step-detection algorithm developed by Kalafut and Visscher (KV) [160]. The KV filter uses a Change-Point Algorithm for determining steps in a signal. The KV filter, as described, has two important shortcomings with respect to the current application: (i) The signal is assumed to have constant variance for all states (false) and (ii) the statistical test for step determination uses the Bayes Information Criterion (BIC), which we have recently demonstrated does not result in optimal performance [158], resulting in either over or underfitting depending on the application.

We describe the adaptation of Change-Point Methods to biophysical applications elsewhere [158]. In short, our CP algorithm is parameter free. The performance of the CP Algorithm and the SED filter applied to simulated intensity data is shown in Figure 4.23A shows the raw (unfiltered) data. Panel B shows the truth (blue, true mean intensity simulated), the simulated data (blue dots), CP idealization (red) and the SED filter (green). The performance of the CP Algorithm correctly determines the position of all steps except one and is qualitatively superior to the performance of the SED filter.

Note that in general the step determination at high intensity is imprecise due to the large variance ($\delta I^2 \propto I$). For higher intensities than those shown in the simulation, the CP algorithm cannot reliably determine the steps, although the CP idealization represents an optimal guess [158]. One might worry that these imprecisely-determined high-intensity steps

could result in a significant degradation in the pairwise distribution function, but there are two natural mechanisms for the suppression of their contribution: (i) The short duration of these steps in frames and the large std both result in a small Fisher Information. (ii) Furthermore the weighting in the PPDD is proportional to the lifetime of the step. Therefore these short steps have a much weaker role in determining the single-fluorophore intensity than the long-lived and less noisy steps at the end of the bleaching trace [158].

4.4.23 The KV versus CP filter

We discovered that the KV filter does not apply a statistical test with suitable frequentist statistical performance to evaluate the existence of new intensity levels. To compare the performance of the KV and CP filters, we simulated intensity data with no transitions. For simplicity, we used a Gaussian process with unit variance for 120 frames for 10,000 independent simulated datasets. We then idealized this data using both the KV and CP filters. See Figure 4.26 for the analysis of 10 typical simulated traces and their idealization using both the KV and CP filters. The KV algorithm has a Type I error (finding at least one false transition) 45% of the time! (See the bar plot of algorithm performance in Figure 4.26B) This result implies that data analysis is performed with a 55% confidence level in a Neyman-Pearson Hypothesis Test. Such a small confidence level is clearly unacceptable from a canonical frequentist perspective. In contrast, the Type I error rate for the CP filter is 5%, corresponding to a 95% confidence level. (Although the above simulation describes a scenario under which the KV filter leads to significant over-fitting, in other circumstances, the KV filter can result in significant under-fitting and therefore should not be used for quantitative applications.)

4.4.24 Simulation: Step detection efficiency using the CP filter as a function of step size

There are a number of important factors that influence the resolution of the CP filter. The ability of the CP algorithm to resolve steps depends principally on the lifetime of the states and the step size between states. To estimate the resolution of the CP filter in step determination, we first simulated transitions from one to zero flours (the final bleaching step)

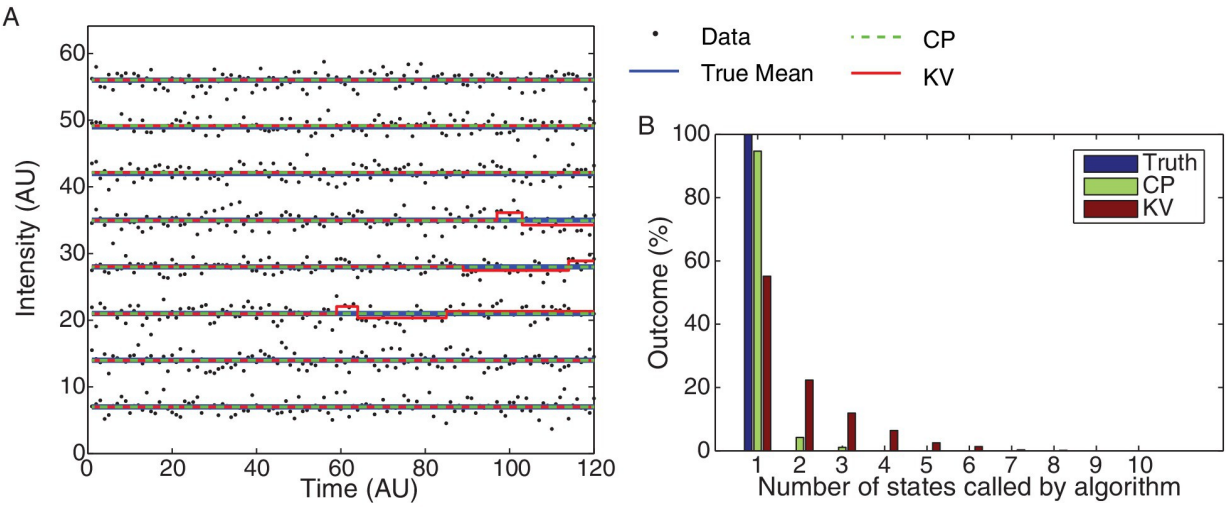

Figure 4.26: **CP vs KV filtering algorithm.** (A) Eight typical datasets (black dots) are plotted with the true mean (blue) and idealizations using both the KV (red) and CP (green) algorithms. The simulated data consisted of 10,000 Gaussian processes with unit variance simulated for 120 frames each. The over segmentation generated by the KV algorithm is clearly visible in the (false) transitions shown in the third, fourth, and fifth traces. No false steps were observed in these eight traces using the CP algorithm. (B) The total number of states for the 10,000 simulated datasets is shown for the true mean and the CP and KV idealization. All datasets consisted of a single true level. The true number of states (one) was found 95% of the time using the CP algorithm and only 55% of the time using the KV algorithm.

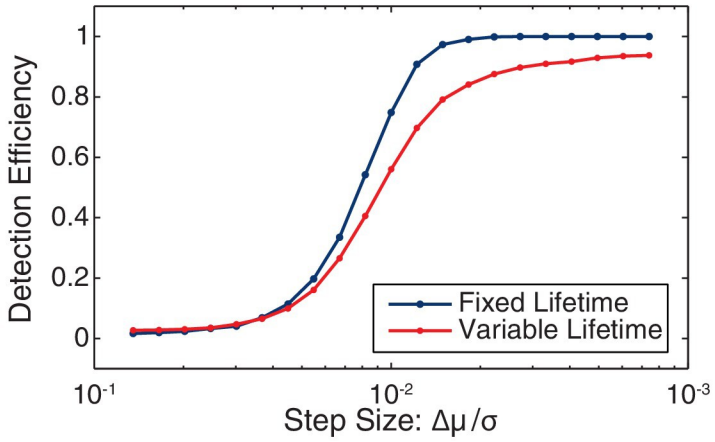

Figure 4.27: **Change point detection efficiency.** The efficiency of change-point detection was measured for fixed-lifetime and variable-lifetime (constant decay rate) steps. In both cases the mean lifetime was equal to the inverse observed decay rate (41 frames). For fixed-lifetime steps (blue), the detection efficiency is essentially unity for relative step-size greater than two. (We define the relative step size as the mean intensity difference divided by the standard deviation: $\Delta\mu/\sigma$.) For variable-lifetime steps (red), the detection efficiency is reduced by the existence of a small fraction of short-lived steps.

using a constant lifetime equal to the inverse observed intensity decay rate ($t_b = 41$ frames). The Fixed-Lifetime curve in Figure 4.27 shows the detection efficiency for the bleaching step as a function of step size. In our experiments, the step-size to the standard deviation ratio ($\Delta\mu/\sigma$) is between 2 and 3. Under these conditions, the step detection efficiency is essentially unity. Of course, the last bleaching step has a distribution of lifetimes, rather than a fixed lifetime equal to the mean lifetime. Next we simulated photobleaching events with variable lifetimes (i.e. stochastic lifetimes with a bleaching rate of $k = 1/41$ frames). The variable-lifetime curve shows the detection efficiency for the bleaching step as a function of step size. Unlike the fixed-lifetime curve, the variable-lifetime curve never reaches unity due to the existence of a small subset of events that are not long enough to resolve. In spite of the inability to resolve very short-lived states, the detection efficiency is still roughly 90% in the simulation at the observed signal to noise ratio ($\Delta\mu/\sigma \approx 2.5$).

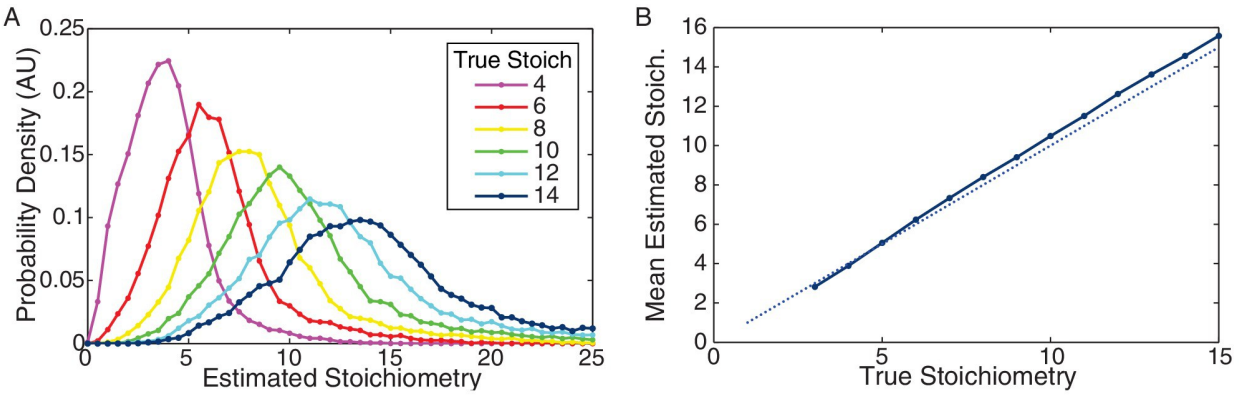

Figure 4.28: **Performance of CP algorithm on simulated data.** (A) Distribution of estimated stoichiometry for simulated data with true stoichiometries from 414 proteins. (Integer stoichiometries between 3 and 15 were simulated. For clarity only the distributions for even stoichiometries are plotted.) (B) Mean estimated stoichiometry as a function of true stoichiometry. The estimated stoichiometry slightly overestimates the true stoichiometry (by a fraction of a protein).

4.4.25 Simulation: Data generated by different protein stoichiometries

To test the overall consistency of the analysis approach, we simulated intensity traces using the observed ratio of step size to standard deviation ($\Delta\mu/\sigma \approx 2.5$ and assuming linear scaling of the variance with intensity) and state lifetime ($k = 1/41$ frames^{-1}) for different stoichiometries between 3 and 15 proteins. The distribution of estimated stoichiometries for selected simulated stoichiometries is shown in Figure 4.28A. The mean estimated stoichiometry as a function of the true stoichiometry is shown in Figure 4.28B. Our simulated analysis results in a very small (<0.5) bias in the mean estimated stoichiometry and a mono-modal distribution of stoichiometry around the true value. The widths of the distributions are roughly consistent with those observed in experiment.

4.4.26 Imaging protocol for replication-complex lifetime

Cells are imaged at two-minute intervals, taking both a phase and fluorescence image at each time point. For lifetime measurements, where the foci must be tracked over a longer time scale,

bleaching is undesirable. In order to minimize both bleaching and possible photodamage to the cells (as indicated by abnormally slowed elongation), we find a longer (600 ms) exposure at lower laser intensity to be optimal. Imaging may continue for about 20 min before focus visibility is significantly impaired by photobleaching, and we note that cells were elongating exponentially regardless of laser use (data not shown).

4.4.27 Analysis of replisome-complex lifetime

Replisome foci were observed to undergo step-like transitions between on and off states. We attempted to use the same analysis used for the bleaching experiments. But, this analysis is designed around the assumption that the replisome undergoes minimal motion. Over the 20 min timescale, the replisome can undergo significant movement and therefore we needed to find an alternative approach.

The phase contrast image for each time point is segmented to determine cell boundaries. For analysis of the fluorescence images, we used a locus-tracking engine that we have described previously to track and quantify loci [117]. In short, foci are detected and fit to a Gaussian point-spread function in each frame. Up to four foci are identified per cell. Each focus is assigned a score (as described in the section titled 'Focus scores' above). The larger the score the more confidence the algorithm has identified a true focus (versus a stochastic fluctuation is fluorescence intensity). Trajectories of replication complexes are then constructed by grouping foci based on the following rules: (i) No foci in the trajectory may score lower than 3. (ii) A focus cannot move more than 350 nm between frames. (iii) The mean of all scores in the trajectory must be a minimum of 4. (iv) At least one focus in the trajectory must score five or higher. (v) Trajectories may continue through a single frame with no (or score ≤ 3) focus provided that all of the above conditions are still met. (vi) Trajectories must last more than three frames. (vii) Included foci must show localization consistent with the replisome.

We have experimented with various rules for grouping foci, but the above best reproduced trajectories qualitatively consistent with the raw images. Foci scoring three or lower appeared randomly throughout the cell, inconsistent with protein bound to the replication factory.

Higher cutoffs were found to lead to gaps in the locus trajectories that lasted for a single frame, consistent with stochastic fluctuations in intensity. We have also required that trajectories last a minimum of three frames. Shorter-lived events are consistent with events observed in cells without active replication, and therefore we believe that these events are also predominantly the result of stochastic fluctuations in fluorescence intensity. Note that, regardless of our choice of grouping, trajectories were always longer lived (on average) for rifampicin treatment and the *rpoB** mutation.

For a disassembly event, foci must disappear for more than two minutes to be counted as disassembly events. (i.e. we require the off state to last for more than one frame to remove most intensity-fluctuation induced false negative events.) A typical trajectory corresponds to a cell containing a focus track with foci scores 3-7 transitioning to a cell with no foci or low-scoring foci (≤ 2) appearing in inconsistent locations in the cell from frame to frame, consistent with false positive focus identification due to photon shot noise (see Figure 4.9 for examples of scored trajectories).

4.4.28 Controlling for bleaching in replisome-complex lifetime experiments

To exclude the possibility that disappearance events are due to photobleaching, we image the cells with the same settings as in the replisome-complex lifetime experiment, but remove the delay between frames. Under these conditions the cells will be dosed with the same amount of light, but over the shortened time scale we would expect bleaching to occur before disassembly due to a conflict. The results confirm that minimal bleaching occurs over the time course. The algorithm successfully tracks 92% and 86% of complexes for the duration of the lifetime experiments in *B. subtilis* and *E. coli*, respectively.

4.4.29 Estimation of conflict number based on replisome complex lifetime

Modeling replisome disassembly events as a Poisson process, the distribution of focus lifetimes was fit using an exponential distribution. The likelihood for the focus lifetime is:

$$\mathrm{p}\left(\tau|k\right) = k\exp\left(-k\left(\tau - \tau_{min}\right)\right), \tag{4.6}$$

where k is the disassembly rate, τ_{min} is the shortest observable lifetime (in our case, 4 min since we disregard events lasting less than three consecutive frames). Since some foci persist throughout the experiment, duration T, we must also compute the survival probability. The survival probability is (one minus the cumulative probability):

$$\Pr\{\tau < t\} = 1 - \mathrm{P}\left(t|k\right) = \exp\left(-k\left(t - \tau_{min}\right)\right) \tag{4.7}$$

The disassembly rate k was estimated using *Maximum Likelihood Estimation*. The sum of the log-likelihood for the observed lifetimes is:

$$\sum_i \log\mathscr{L}\left(\tau_i|k\right) = \sum_{\tau_i<T} \mathrm{logp}\left(\tau_i|k\right) + \sum_{\tau_i\geq T} \log\left(1 - \mathrm{P}\left(T|k\right)\right) \tag{4.8}$$

Note that the sum in the first term is taken only over the observable lifetimes while the second term accounts for long-lived states. Maximizing the likelihood leads to the following expression for the maximum likelihood estimate of k:

$$\hat{k} = N_{\tau<T} / \left[N_{\tau<T}\left(\bar{\tau} - \tau_{min}\right) + N_{\tau\geq T}\left(T - \tau_{min}\right)\right], \tag{4.9}$$

where $N_{\tau\geq T}$ is the number of lifetimes that were at least the duration of the experiment and $N_{\tau<T}$ is the number of observed lifetimes and $\bar{\tau}$ is their empirical mean. A summary of the parameters used in the replisome-complex lifetime is provided in Table 4.1 (*B. subtilis*) and Table 4.4 (*E. coli*).

4.4.30 Protocol for temperature-sensitive DnaC experiment

Cells were prepared for microscopy as described above, and imaged under non-permissive conditions using an objective heater (Bioptechs). Imaging started roughly 10 min after cells were placed on the heated objective. Trajectories were generated using the algorithm developed to measure complex lifetime, and foci included in these trajectories counted towards the probability of observing a focus as a function of time. The theoretical 'null-hypothesis' curve is generated by assuming a random segment of continuous (no disassemblies due to conflicts) 40 min replication cycle is visualized for 10 frames at 2 min intervals. We include all possible outcomes where a focus is visible in the first frame.

4.4.31 Protocol for PriA CRISPR experiment

Due to the leakiness of the CRISPR system, PriA is already depleted three-fold before induction [155], and we note that the strain grew unusually slowly in minimal arabinose (however, the precursor strain without the sgRNA grew normally). The CRISPR system was fully induced by the addition of 1% xylose in liquid culture roughly 2 hr before imaging to allow the remaining PriA to be diluted out. DnaC-GFP foci were then identified from snapshot images (one phase contrast and one fluorescence image at each field of view). The number of cells with foci was quantified using the following rules: (i) the focus must score three or higher, (ii) the elipticity of the focus must be smaller than 1.2, and (iii) the focus localization must be consistent with the replisome. Because of the leakiness of the CRISPR, we compare to the number of foci in the precursor strain (does not contain sgRNA), also with the addition of 1% xylose. Time-lase microscopy was not productive because the PriA CRISPR strain formed few DnaC-GFP foci even without induction.

4.4.32 Protein purification

Purified GFP was gifted to us by the Asbury Lab at the University of Washington. For purification of YPet, DH5α*E. coli* cells were transformed with the plasmid ROD49 carrying an arabinose-inducible his-tagged mYPet with the monomeric A206K mutation. (This plasmid was the gift of R. Reyes-Lamothe.) The expression was induced at an optical density (OD600) of 0.1 with 0.2% L-arabinose for 1 hr at 37C. Cells resuspended in 20 mM HEPES pH 7.5, 0.5 M NaCl, 25 mM imidazole were lysed by sonication. Lysate was cleared by centrifugation for 1 hr at 18,000 x g, and proteins were purified by fast protein liquid chromatography (KTA System, GE Healthcare) using a metal-chelating affinity column (HisTrap HP, GE Healthcare). YPet-containing fractions eluted at high imidazole concentrations were identified by absorbance at 280 nm and confirmed using a microplate photometer to verify the correct excitation and emission spectra.

4.4.33 Preparation and imaging of surface immobilized protein

For imaging, we dilute purified protein in PBS to the point where individual molecules are visible during fluorescence microscopy. We fill a 10 µL flow cell (constructed by sticking a KOH cleaned coverslip to a base slide with two strips of two sided tape) with the PBS-fluorescent protein solution and allow it to sit upside down for 10 min, binding the proteins to the coverslip. The channel is then rinsed with 400 µL of PBS to clear any remaining fluorescent protein from the background.

We image and the isolated protein using the same settings as the *in vivo* bleaching experiments, and analysis proceeds as for live cell bleaching experiments (the only difference being that we do not take a phase contrast image and intensity regions are identified from the fluorescent images alone). See Figure 4.6A and Figure 4.18A for example isolated protein bleaching traces. We arrive at the unitary intensity step distribution for known single fluorophores. As a consistency check, we confirm that the in vivo unitary intensity step distributions for all strains are highly similar to those found in vitro. We find agreement of the peak values for in vivo and in vitro distributions to within 19% for GFP and 18% for YPet.

4.4.34 Thymidine incorporation assays

Exponentially growing cells raised in minimal MOPS medium [152] supplemented with either 1% arabinose or glucose, at 30°C, were split at OD_{600} 0.2 into equal 1.2 ml cultures. Cells continued to grow until they reached OD_{600} 0.3, at which point 30 µg/ml rifampicin was added to one of the cultures for 2 min. Next, 38 µCi ^{3}H-thymidine (Perkin Elmer 70-90 Ci/mMol) was added to both cultures, and timepoints were taken at 2 min intervals by pipetting 200 l of cells into 3 ml of ice-cold 10% TCA. Samples were collected on glass microfiber filters (GE Healtthcare #1825025), and washed 3x with 5% TCA prior to detection on a liquid scintillation counter.

4.5 Acknowledgements

We would like to acknowledge Dr. Jason Peters for generously sharing the strain containing the *dcas9* allele (CAG74209, or HM1500) and sgRNA expression vector (pJMP4) prior to publication of the *B. subtilis* CRISPRi system, which we used to construct the *priA* knockdown strains published in our study. We are grateful to Seemay Chou for purification of the YPet protein and Patrick Nugent for building of the $P_{spank(hy)}$-*lacZ* construct. This work was funded by the National Science Foundation grant MCB1243492.

Chapter 5
CONCLUDING REMARKS

Over the last ten years or so, experimental methods have become sufficiently advanced to study living cells at the level of individual molecules. These new technologies are allowing us to uncover previously overlooked features of biological processes. In this chapter, I summarize the main findings of the work presented in this thesis, and discusses yet unanswered questions which could be the focus of future work.

5.1 Summary of findings

In most bacteria, including *B. subtilis* and *E. coli*, two replisomes assemble at a single origin and work bi-directionally to duplicate the chromosome. During this time, the two replisomes are closely spaced (generally < 250 nm) and form a relatively stationary complex through which the DNA is pulled [7]. However, if you look inside diffraction-limited complex, one of the replisomes is often disassembled. Furthermore, if you watch the replication complex over time, disassociation from the DNA is observed multiple times every single replication cycle. Transcription inhibition restores stability of the replication complex. Therefore the replisome is frequently disrupted by pervasive replication-transcription conflicts [17].

Altogether, these findings represent a significant deviation from the textbook model of DNA replication. Until recently, replication was assumed to highly processive and continuous, with both replisomes remaining stably bound from initiation to termination. My own work as well as a number of other recent reports have used single-molecule techniques to reveal a much more dynamic picture of the replication process. Many core replisome components have been shown to turn over on the time scale of seconds [34, 112, 113]. My own work suggests collapse of the entire replisome many times per replication cycle [17]. These findings together suggest a need to revise the canonical model of replication.

5.2 Future studies

5.2.1 Replisome localization

Although the work here shows that replisome pairs operate in close proximity throughout the cell cycle, it is unknown how this localization is maintained. The occasional transient separation of replisome pairs implies that they are only weakly coupled and held together by a non-specific interaction. Future work may use high-throughput imaging-based screening techniques to visualize replisome localization in a library of conditional mutants. This would identify factors that directly or indirectly affect replisome structure.

Another loose end is that is that we have assumed that single replication focus represents a co-localized replisome pair. Future work will visualize the structure of the replication factory beyond the diffraction-limited focus, potentially resolving the two replisome separately. This is achievable by existing super-resolution techniques. Single-molecule localization based reconstruction could provide deeper structural insight into the spacing between the replisomes in a replication factory.

Most studies of the replisome, including the ones here, were based on imaging of the replisome one component at a time. An intriguing possibility is that replisome components are not continually co-localized throughout the cell cycle. Two-color imaging of the replisome could provide deeper insight into replisome structure in living cells.

5.2.2 Replisome dynamics

We have shown here that the replisome undergoes multiple rounds of disassembly and restart every cell cycle. The replication-restart pathway is fairly well understood, however, precisely how the replisome is disassembled remains unknown. It is possible that cellular machinery is needed to remove at least some of the replisome components from the DNA. A possible way to investigate this would be to study the disassembly of replisomes at an engineered barrier (e.g. an array of tightly bound repressor proteins). By simultaneous visualization of two fluorescently-labeled replisome components at a time, it would be possible to determine if

replisome disassembly at a barrier is ordered (i.e. occurs the same way every time). This would imply a specific mechanism for replisome disassembly rather than spontaneous disassociation from the DNA.

BIBLIOGRAPHY

[1] Benkovic SJ, Valentine AM, Salinas F (2001) Replisome-mediated DNA Replication. *Annu. Rev. Biochem.* 70:181–208.

[2] Yao NY, O'Donnell M (2010) SnapShot: The replisome. *Cell* 141(6):1088, 1088.e1.

[3] Merrikh H, Machón C, Grainger WH, Grossman AD, Soultanas P (2011) Co-directional replication-transcription conflicts lead to replication restart. *Nature* 470(7335):554–557.

[4] Merrikh H, Zhang Y, Grossman AD, Wang JD (2012) Replication-transcription conflicts in bacteria. *Nature Reviews Microbiology* 10(7):449–58.

[5] Merrikh CN, Brewer BJ, Merrikh H (2015) The *B. subtilis* Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. *PLOS Genetics* 11(6):e1005289.

[6] Lambert S, Carr AM (2013) Impediments to replication fork movement: stabilisation, reactivation and genome instability. *Chromosoma* 122(1-2):33–45.

[7] Mangiameli SM, Veit BT, Merrikh H, Wiggins PA (2017) The Replisomes Remain Spatially Proximal throughout the Cell Cycle in Bacteria. *PLOS Genetics* 13(1):e1006582.

[8] Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H (2013) Accelerated gene evolution through replication-transcription conflicts. *Nature* 495(7442):512–515.

[9] Brambati A, Colosio A, Zardoni L, Galanti L, Liberi G (2015) Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Frontiers in Genetics* 6:166.

[10] Gaillard H, Aguilera A (2016) Transcription as a Threat to Genome Integrity. *Annual Review of Biochemistry* 85(1):291–317.

[11] Baker TA, Bell SP (1998) Polymerases and the replisome: machines within machines. *Cell* 92(3):295–305.

[12] Johnson A, O'Donnell M (2005) CELLULAR DNA REPLICASES: Components and Dynamics at the Replication Fork. *Annual Review of Biochemistry* 74(1):283–315.

[13] Zechner EL, Wu CA, Marians KJ (1992) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. *The Journal of biological chemistry* 267(6):4045–53.

[14] Sanders GM, Dallmann HG, McHenry CS (2010) Reconstitution of the *B. subtilis* Replisome with 13 Proteins Including Two Distinct Replicases. *Molecular Cell* 37(2):273–281.

[15] Graham JE et al. (2017) Independent and Stochastic Action of DNA Polymerases in the Replisome. *Cell* 169(7):1201–1213.e17.

[16] Tinoco I, Gonzalez RL, Gonzalez RL, Jr (2011) Biological mechanisms, one molecule at a time. *Genes & development* 25(12):1205–31.

[17] Mangiameli SM, Merrikh CN, Wiggins PA, Merrikh H (2017) Transcription leads to pervasive replisome instability in bacteria. *eLife* 6.

[18] Wang X, Llopis PM, Rudner DZ (2013) Organization and segregation of bacterial chromosomes. *Nature Reviews Genetics* 14(3):191–203.

[19] Snyder L, Henkin TM, Peters JE, Champness W (2013) *Molecular Genetics of Bacteria, 4th Edition*. (American Society of Microbiology).

[20] Kunst F et al. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390(6657):249–256.

[21] Keseler IM et al. (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Research* 45(D1):D543–D550.

[22] Cooper S, Helmstetter CE (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *Journal of molecular biology* 31(3):519–40.

[23] O'Donnell M (2006) Replisome architecture and dynamics in *Escherichia coli*. *The Journal of biological chemistry* 281(16):10653–6.

[24] Reyes-Lamothe R, Sherratt DJ, Leake MC (2010) Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science* 328(5977):498–501.

[25] Duderstadt KE, Reyes-Lamothe R, van Oijen AM, Sherratt DJ (2014) Replication-Fork Dynamics. *Cold Spring Harbor Perspectives in Biology* 6(1):a010157–a010157.

[26] Beattie TR, Reyes-Lamothe R (2015) A Replisome's journey through the bacterial chromosome. *Frontiers in Microbiology* 6:562.

[27] McHenry CS (2011) Bacterial replicases and related polymerases. *Current Opinion in Chemical Biology* 15(5):587–594.

[28] Marceau AH (2012) Functions of Single-Strand DNA-Binding Proteins in DNA Replication, Recombination, and Repair in *Single-Stranded DNA Binding Proteins*. (Humana Press, Totowa, NJ) Vol. 922, pp. 1–21.

[29] Ogawa T, Okazaki T (1980) Discontinuous DNA Replication. *Annual Review of Biochemistry* 49(1):421–457.

[30] Wu CA, Zechner EL, Marians KJ (1992) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. *The Journal of biological chemistry* 267(6):4030–44.

[31] Timinskas K, Balvočit M, Timinskas A, Venclovas Č (2014) Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. *Nucleic acids research* 42(3):1393–413.

[32] Barnes MH, Miller SD, Brown NC (2002) DNA polymerases of low-GC gram-positive eubacteria: identification of the replication-specific enzyme encoded by *dnaE*. *Journal of bacteriology* 184(14):3834–8.

[33] Dervyn E et al. (2001) Two Essential DNA Polymerases at the Bacterial Replication Fork. *Science* 294(5547):1716–1719.

[34] Liao Y, Li Y, Schroeder JW, Simmons LA, Biteen JS (2016) Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells. *Biophysical Journal* 111(12):2562–2569.

[35] Su'etsugu M, Errington J (2011) The Replicase Sliding Clamp Dynamically Accumulates behind Progressing Replication Forks in *Bacillus subtilis* Cells. *Molecular Cell* 41(6):720–732.

[36] Moolman MC et al. (2014) Slow unloading leads to DNA-bound β_2-sliding clamp accumulation in live *Escherichia coli* cells. *Nature Communications* 5:5820.

[37] Leu FP, Hingorani MM, Turner J, O'Donnell M (2000) The δ Subunit of DNA Polymerase III Holoenzyme Serves as a Sliding Clamp Unloader in *Escherichia coli*. *Journal of Biological Chemistry* 275(44):34609–34618.

[38] Park AY et al. (2010) A Single Subunit Directs the Assembly of the *Escherichia coli* DNA Sliding Clamp Loader. *Structure* 18(3):285–292.

[39] Tsuchihashi Z, Kornberg A (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences of the United States of America* 87(7):2516–20.

[40] Lemon KP, Grossman AD (1998) Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science* 282(5393):1516–9.

[41] Holmes VF, Cozzarelli NR (2000) Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proceedings of the National Academy of Sciences of the United States of America* 97(4):1322–4.

[42] Trun NJ, Marko JF (1998) Architecture of a Bacterial Chromosome. *Am. Soc. Microbiol. News* 64(5):276–283.

[43] Postow L, Hardy CD, Arsuaga J, Cozzarelli NR (2004) Topological domain structure of the *Escherichia coli* chromosome. *Genes & Development* 18(14):1766–1779.

[44] Deng S, Stein RA, Higgins NP (2005) Organization of supercoil domains and their reorganization by transcription. *Molecular microbiology* 57(6):1511–21.

[45] Cui Y, Petrushenko ZM, Rybenkov VV (2008) MukB acts as a macromolecular clamp in DNA condensation. *Nature Structural & Molecular Biology* 15(4):411–418.

[46] Higgins NP, Yang X, Fu Q, Roth JR (1996) Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *Journal of bacteriology* 178(10):2825–35.

[47] Trautinger BW, Jaktaji RP, Rusakova E, Lloyd RG (2005) RNA Polymerase Modulators and DNA Repair Activities Resolve Conflicts between DNA Replication and Transcription. *Molecular Cell* 19(2):247–258.

[48] Tehranchi AK et al. (2010) The Transcription Factor DksA Prevents Conflicts between DNA Replication and Transcription Machinery. *Cell* 141(4):595–605.

[49] Brewer BJ, Shyy S, Wang J, Liu L, Patte J (1988) When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* 53(5):679–86.

[50] Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD (2010) Co-Orientation of Replication and Transcription Preserves Genome Integrity. *PLoS Genetics* 6(1):e1000810.

[51] Vilette D, Ehrlich SD, Michel B (1995) Transcription-induced deletions in *Escherichia coli* plasmids. *Molecular microbiology* 17(3):493–504.

[52] Gan W et al. (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development* 25(19):2041–56.

[53] Dutta D et al. (2011) Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* 146(4):533–43.

[54] Mirkin EV, Mirkin SM (2007) Replication Fork Stalling at Natural Impediments. *Microbiology and Molecular Biology Reviews* 71(1):13–35.

[55] Wang JD, Berkmen MB, Grossman AD (2007) Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* 104(13):5608–5613.

[56] Boubakri H, de Septenville AL, Viguera E, Michel B (2010) The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *The EMBO Journal* 29(1):145–157.

[57] French S (1992) Consequences of replication fork movement through transcription units *in vivo*. *Science (New York, N.Y.)* 258(5086):1362–5.

[58] Liu B, Alberts BM (1995) Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science (New York, N.Y.)* 267(5201):1131–7.

[59] Hill CW, Gray JA (1988) Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* 119(4):771–8.

[60] Deshpande AM, Newlon CS (1996) DNA replication fork pause sites dependent on transcription. *Science (New York, N.Y.)* 272(5264):1030–3.

[61] Mirkin EV, Castro Roa D, Nudler E, Mirkin SM (2006) Transcription regulatory elements are punctuation marks for DNA replication. *Proceedings of the National Academy of Sciences* 103(19):7276–7281.

[62] RUDOLPH C, DHILLON P, MOORE T, LLOYD R (2007) Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair* 6(7):981–993.

[63] Cox MM et al. (2000) The importance of repairing stalled replication forks. *Nature* 404(6773):37–41.

[64] Cox MM (2001) Historical overview: Searching for replication help in all of the rec places. *Proceedings of the National Academy of Sciences* 98(15):8173–8180.

[65] Blattner FR et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)* 277(5331):1453–62.

[66] McLean MJ, Wolfe KH, Devine KM (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *Journal of molecular evolution* 47(6):691–6.

[67] Rocha EPC, Danchin A (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic acids research* 31(22):6570–7.

[68] Rocha EPC, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nature Genetics* 34(4):377–378.

[69] Guy L, Roten CAH (2004) Genometric analyses of the organization of circular chromosomes: a universal pressure determines the direction of ribosomal RNA genes transcription relative to chromosome replication. *Gene* 340(1):45–52.

[70] Guy CP et al. (2009) Rep Provides a Second Motor at the Replisome to Promote Duplication of Protein-Bound DNA. *Molecular Cell* 36(4):654–666.

[71] Pomerantz RT, O'Donnell M (2008) The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature* 456(7223):762–766.

[72] Pomerantz RT, O'Donnell M (2010) Direct Restart of a Replication Fork Stalled by a Head-On RNA Polymerase. *Science* 327(5965):590–592.

[73] Zernike F (1942) Phase contrast, a new method for the microscopic observation of transparent objects. *Physica* 9(10):974–986.

[74] Shimomura O, Johnson FH, Saiga Y (1962) Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *Journal of Cellular and Comparative Physiology* 59(3):223–239.

[75] Heim R, Tsien RY (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current biology : CB* 6(2):178–82.

[76] Phillips R, Kondev J, Theriot J (2009) *Physical biology of the cell.* (Garland Science), p. 807.

[77] Inoue S, Spring KR (1997) *Video microscopy : the fundamentals.* (Plenum Press), p. 741.

[78] Huang B, Babcock H, Zhuang X (2010) Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* 143(7):1047–58.

[79] Meselson M, Stahl FW (1958) The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 44(7):671–82.

[80] Frouin I, Montecucco A, Spadari S, Maga G (2003) DNA replication: a complex matter. *EMBO Reports* 4(7):666–670.

[81] Dingman CW (1974) Bidirectional chromosome replication: some topological considerations. *Journal of theoretical biology* 43(1):187–95.

[82] Newport J, Yan H (1996) Organization of DNA into foci during replication. *Current opinion in cell biology* 8(3):365–8.

[83] Ma H et al. (1998) Spatial and temporal dynamics of DNA replication sites in mammalian cells. *The Journal of cell biology* 143(6):1415–25.

[84] Jackson DA, Pombo A (1998) Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *The Journal of cell biology* 140(6):1285–95.

[85] Hozák P, Hassan AB, Jackson DA, Cook PR (1993) Visualization of replication factories attached to nucleoskeleton. *Cell* 73(2):361–73.

[86] Leonhardt H et al. (2000) Dynamics of DNA replication factories in living cells. *The Journal of cell biology* 149(2):271–80.

[87] Jensen RB, Wang SC, Shapiro L (2001) A moving DNA replication factory in *Caulobacter crescentus*. *The EMBO journal* 20(17):4952–63.

[88] Lemon KP et al. (2000) Movement of replicating DNA through a stationary replisome. *Molecular cell* 6(6):1321–30.

[89] Berkmen MB, Grossman AD (2006) Spatial and temporal organization of the *Bacillus subtilis* replication cycle. *Molecular Microbiology* 62(1):57–71.

[90] Bates D et al. (2005) Chromosome and replisome dynamics in *E. coli*: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell* 121(6):899–911.

[91] Reyes-Lamothe R, Possoz C, Danilova O, Sherratt DJ (2008) Independent Positioning and Action of *Escherichia coli* Replisomes in Live Cells. *Cell* 133(1):90–102.

[92] Hiraga S, Ichinose C, Onogi T, Niki H, Yamazoe M (2000) Bidirectional migration of SeqA-bound hemimethylated DNA clusters and pairing of oriC copies in *Escherichia coli*. *Genes to cells : devoted to molecular & cellular mechanisms* 5(5):327–41.

[93] Kongsuwan K, Dalrymple BP, Wijffels G, Jennings PA (2002) Cellular localisation of the clamp protein during DNA replication. *FEMS microbiology letters* 216(2):255–62.

[94] Lemon KP, Grossman AD (2001) The extrusion-capture model for chromosome partitioning in bacteria. *Genes & development* 15(16):2031–41.

[95] Sawitzke J, Austin S (2001) An analysis of the factory model for chromosome replication and segregation in bacteria. *Molecular microbiology* 40(4):786–94.

[96] Cass JA, Kuwada NJ, Traxler B, Wiggins PA (2016) *Escherichia coli* Chromosomal Loci Segregate from Midcell with Universal Dynamics. *Biophysical Journal* 110(12):2597–2609.

[97] Migocki MD, Lewis PJ, Wake RG, Harry EJ (2004) The midcell replication factory in *Bacillus subtilis* is highly mobile: implications for coordinating chromosome replication with other cell cycle events. *Molecular Microbiology* 54(2):452–463.

[98] Swulius MT, Jensen GJ (2012) The helical MreB cytoskeleton in *Escherichia coli* MC1000/pLE7 is an artifact of the N-Terminal yellow fluorescent protein tag. *Journal of bacteriology* 194(23):6382–6.

[99] Landgraf D, Okumus B, Chien P, Baker TA, Paulsson J (2012) Segregation of molecules at cell division reveals native protein localization. *Nature Methods* 9(5):480–482.

[100] Koppes LJ, Woldringh CL, Nanninga N (1999) *Escherichia coli* contains a DNA replication compartment in the cell center. *Biochimie* 81(8-9):803–10.

[101] Molina F, Skarstad K (2004) Replication fork and SeqA focus distributions in *Escherichia*

coli suggest a replication hyperstructure dependent on nucleotide metabolism. *Molecular Microbiology* 52(6):1597–1612.

[102] Adachi S, Kohiyama M, Onogi T, Hiraga S (2005) Localization of replication forks in wild-type and mukB mutant cells of *Escherichia coli*. *Molecular Genetics and Genomics* 274(3):264–271.

[103] Den Blaauwen T, Aarsman MEG, Wheeler LJ, Nanninga N (2006) Pre-replication assembly of *E. coli* replisome components. *Molecular Microbiology* 62(3):695–708.

[104] Wallden M et al. (2016) The Synchronization of Replication and Division Cycles in Individual *E. coli* Cells. *Cell* 166(3):729–739.

[105] Cass JA, Stylianidou S, Kuwada NJ, Traxler B, Wiggins PA (2017) Probing bacterial cell biology using image cytometry. *Molecular Microbiology* 103(5):818–828.

[106] Yamazoe M, Adachi S, Kanaya S, Ohsumi K, Hiraga S (2004) Sequential binding of SeqA protein to nascent DNA segments at replication forks in synchronized cultures of *Escherichia coli*. *Molecular Microbiology* 55(1):289–298.

[107] Onogi T, Ohsumi K, Katayama T, Hiraga S (2002) Replication-dependent recruitment of the beta-subunit of DNA polymerase III from cytosolic spaces to replication forks in *Escherichia coli*. *Journal of bacteriology* 184(3):867–70.

[108] Sunako Y, Onogi T, Hiraga S (2002) Sister chromosome cohesion of *Escherichia coli*. *Molecular Microbiology* 42(5):1233–1241.

[109] Breier AM, Weier HUG, Cozzarelli NR (2005) Independence of replisomes in *Escherichia coli* chromosomal replication. *Proceedings of the National Academy of Sciences* 102(11):3942–3947.

[110] Youngren B, Nielsen HJ, Jun S, Austin S (2014) The multifork *Escherichia coli* chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes & Development* 28(1):71–84.

[111] Wang X, Brandão HB, Le TBK, Laub MT, Rudner DZ (2017) *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science* 355(6324).

[112] Beattie TR et al. (2017) Frequent exchange of the DNA polymerase during bacterial chromosome replication. *eLife* 6.

[113] Lewis JS et al. (2017) Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *eLife* 6.

[114] Goranov AI, Breier AM, Merrikh H, Grossman AD (2009) YabA of *Bacillus subtilis* controls DnaA-mediated replication initiation but not the transcriptional response to replication stress. *Molecular Microbiology* 74(2):454–466.

[115] Soufo CD et al. (2008) Cell-cycle-dependent spatial sequestration of the DnaA replication initiator protein in *Bacillus subtilis*. *Developmental cell* 15(6):935–41.

[116] Kuwada NJ, Traxler B, Wiggins PA (2015) Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. *Molecular Microbiology* 95(1):64–79.

[117] Kuwada NJ, Cheveralls KC, Traxler B, Wiggins PA (2013) Mapping the driving forces of chromosome structure and segregation in *Escherichia coli*. *Nucleic Acids Research* 41(15):7370–7377.

[118] Stylianidou S, Brennan C, Nissen SB, Kuwada NJ, Wiggins PA (2016) *SuperSegger*: robust image segmentation, analysis and lineage tracking of bacterial cells. *Molecular Microbiology* 102(4):690–700.

[119] O'Donnell M, Langston L, Stillman B (2013) Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor perspectives in biology* 5(7):a010108.

[120] Baharoglu Z, Lestini R, Duigou S, Michel B (2010) RNA polymerase mutations that facilitate replication progression in the rep uvrD recF mutant lacking two accessory replicative helicases. *Molecular Microbiology* 77(2):324–336.

[121] Bidnenko V, Lestini R, Michel B (2006) The *Escherichia coli* UvrD helicase is essential for Tus removal during recombination-dependent replication restart from *Ter* sites. *Molecular Microbiology* 62(2):382–396.

[122] Marsin S, McGovern S, Ehrlich SD, Bruand C, Polard P (2001) Early steps of *Bacillus subtilis* primosome assembly. *The Journal of biological chemistry* 276(49):45818–25.

[123] Mirkin EV, Mirkin SM (2005) Mechanisms of Transcription-Replication Collisions in Bacteria. *Molecular and Cellular Biology* 25(3):888–895.

[124] Soultanas P (2011) The replication-transcription conflict. *Transcription* 2(3):140–144.

[125] Million-Weaver S, Samadpour AN, Merrikh H (2015) Replication Restart after Replication-Transcription Conflicts Requires RecA in *Bacillus subtilis*. *Journal of Bacteriology* 197(14):2374–2382.

[126] Bruand C et al. (2005) Functional interplay between the *Bacillus subtilis* DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Molecular Microbiology* 55(4):1138–1150.

[127] Gabbai CB, Marians KJ (2010) Recruitment to stalled replication forks of the PriA DNA helicase and replisome-loading activities is essential for survival. *DNA Repair* 9(3):202–209.

[128] Lovett ST (2005) Filling the Gaps in Replication Restart Pathways. *Molecular Cell* 17(6):751–752.

[129] Zavitz KH, Marians KJ (1992) ATPase-deficient mutants of the *Escherichia coli* DNA replication protein PriA are capable of catalyzing the assembly of active primosomes. *The Journal of biological chemistry* 267(10):6933–40.

[130] Maisnier-Patin S, Nordström K, Dasgupta S (2001) Replication arrests during a single round of replication of the *Escherichia coli* chromosome in the absence of DnaC activity. *Molecular microbiology* 42(5):1371–82.

[131] McGlynn P, Lloyd RG (2002) Recombinational repair and restart of damaged replication forks. *Nature Reviews Molecular Cell Biology* 3(11):859–870.

[132] Polard P et al. (2002) Restart of DNA replication in Gram-positive bacteria: functional characterisation of the *Bacillus subtilis* PriA initiator. *Nucleic acids research* 30(7):1593–605.

[133] Sandler SJ et al. (1999) *dnaC* mutations suppress defects in DNA replication- and recombination-associated functions in *priB* and *priC* double mutants in *Escherichia coli* K-12. *Molecular microbiology* 34(1):91–101.

[134] Sandler SJ, Marians KJ (2000) Role of PriA in replication fork reactivation in *Escherichia coli*. *Journal of bacteriology* 182(1):9–13.

[135] Leake MC et al. (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* 443(7109):355–358.

[136] Ulbrich MH, Isacoff EY (2007) Subunit counting in membrane-bound proteins. *Nature Methods* 4(4):319–21.

[137] Bruand C, Farache M, McGovern S, Ehrlich SD, Polard P (2001) DnaB, DnaD and

DnaI proteins are components of the *Bacillus subtilis* replication restart primosome. *Molecular microbiology* 42(1):245–55.

[138] Bailey S, Eliason WK, Steitz TA (2007) The crystal structure of the *Thermus aquaticus* DnaB helicase monomer. *Nucleic Acids Research* 35(14):4728–4736.

[139] Fass D, Bogden CE, Berger JM (1999) Crystal structure of the N-terminal domain of the DnaB hexameric helicase. *Structure (London, England : 1993)* 7(6):691–8.

[140] Kaplan DL, Saleh OA, Ribeck N (2013) Single-molecule and bulk approaches to the DnaB replication fork helicase. *Frontiers in Bioscience* 18:224–240.

[141] Helmstetter CE, Cooper S (1968) DNA synthesis during the division cycle of rapidly growing *Escherichia coli*. *Journal of Molecular Biology* 31(3):507–518.

[142] De Septenville AL, Duigou S, Boubakri H, Michel B (2012) Replication Fork Reversal after ReplicationTranscription Collision. *PLoS Genetics* 8(4):e1002622.

[143] Million-Weaver S et al. (2015) An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* 112(10):E1096–E1105.

[144] Sensi P (1983) History of the Development of Rifampin. *Clinical Infectious Diseases* 5(Supplement 3):S402–S406.

[145] Sensi P (1969) Structure-activity relationships of rifamycin derivatives. *Acta tuberculosea et pneumologica Belgica* 60(3):258–65.

[146] Bartlett MS, Gaal T, Ross W, Gourse RL (1998) RNA polymerase mutants that destabilize RNA Polymerase-Promoter complexes alter NTP-sensing by *rrn* P1 promoters. *Journal of Molecular Biology* 279(2):331–345.

[147] Maughan H, Galeano B, Nicholson WL (2004) Novel *rpoB* mutations conferring rifampin resistance on *Bacillus subtilis*: global effects on growth, competence, sporulation, and germination. *Journal of bacteriology* 186(8):2481–6.

[148] Zhou YN, Jin DJ (1998) The *rpoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 95(6):2908–13.

[149] Lau IF et al. (2003) Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Molecular microbiology* 49(3):731–43.

[150] Ferullo DJ, Cooper DL, Moore HR, Lovett ST (2009) Cell cycle synchronization of *Escherichia coli* using the stringent response, with fluorescence labeling assays for DNA content and replication. *Methods* 48(1):8–13.

[151] Kornberg A, Baker TA (1992) *DNA replication*. (Wh Freeman New York:).

[152] Wang JD, Sanders GM, Grossman AD (2007) Nutritional Control of Elongation of DNA Replication by (p)ppGpp. *Cell* 128(5):865–875.

[153] Mirouze N, Prepiak P, Dubnau D (2011) Fluctuations in spo0A Transcription Control Rare Developmental Transitions in *Bacillus subtilis*. *PLoS Genetics* 7(4):e1002048.

[154] Smith JL, Goldberg JM, Grossman AD (2014) Complete Genome Sequences of *Bacillus subtilis* subsp. subtilis Laboratory Strains JH642 (AG174) and AG1839. *Genome Announcements* 2(4):e00663–14–e00663–14.

[155] Peters J et al. (2016) A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* 165(6):1493–1506.

[156] DeWitt SK, Adelberg EA (1962) The occurrence of a genetic transposition in a strain of *Escherichia coli*. *Genetics* 47(5).

[157] Smith DA (1998) A quantitative method for the detection of edges in noisy time-series. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 353(1378).

[158] Wiggins P (2015) An Information-Based Approach to Change-Point Analysis with Applications to Biophysics and Cell Biology. *Biophysical Journal* 109(2):346–354.

[159] Little MA, Jones NS (2012) Signal processing for molecular and cellular biological physics: an emerging field. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(1984):20110546–20110546.

[160] Kalafut B, Visscher K (2008) An objective, model-independent method for detection of non-uniform steps in noisy signals. *Computer Physics Communications* 179(10):716–723.

Appendix A

INDEPENDENT-FORK MODEL

In our analysis of the focus dynamics, we have not distinguished between one and two fork replisomes and interpreted the measured rate of focus loss as the rate of single fork disassembly to generate a crude estimate of the number of conflicts per cell cycle. We will call this qualitative model the *cooperative-fork model* since it implicitly postulates that a conflict at a single fork results is loss of the replication complex at both forks. While this model is clearly a useful qualitative device for making a crude estimate of the number of conflicts per cell cycle, it is inconsistent with the stoichiometry data which demonstrates that replisomes frequently have a stoichiometry consistent with a single fork. We now briefly explore a more self-consistent model which distinguishes between one and two-fork complexes. We will assume that the replication forks move and act independently and assume that the conflict-generation and replication restart are both Poisson processes with a constant rate:

$$\text{active} \underset{k_+}{\overset{k_-}{\rightleftharpoons}} \text{stalled}, \tag{A.1}$$

for a single fork. Since the forks in this model act independently, we will call this model the *independent-fork model*. This more complicated model changes none of the qualitative insights from the cooperative fork model.

A.1 Model and rates.

Since our experiments are sensitive only to the number of assembled fork complexes, we define the states in terms of the number of complexes in the focus: 0, 1, or 2. The rate

equation describing the number of two, one and zero active complexes is therefore:

$$\begin{pmatrix} \dot{N}_2 \\ \dot{N}_1 \\ \dot{N}_0 \end{pmatrix} = \begin{pmatrix} -2k_- & k_+ & 0 \\ 2k_- & -k_+ - k_- & 2k_+ \\ 0 & k_- & -2k_+ \end{pmatrix} \begin{pmatrix} N_2 \\ N_1 \\ N_0 \end{pmatrix}, \tag{A.2}$$

where the factors of two in the rate equation account for the fact that either of the two forks can disassemble. We analyze the linear model in the canonical way: we find the eigenvalues and vectors.

A.2 Steady-state solution.

In the current context, we only care about the steady-state solution with eigenvalue zero:

$$\boldsymbol{\psi}_0^T = \begin{pmatrix} k_+^2/k_-^2 & 2k_+/k_- & 1 \end{pmatrix}, \tag{A.3}$$

which determines the ratios between two and one fork replisomes is steady state:

$$R \equiv N_2(\infty)/N_1(\infty) = k_+/2k_-. \tag{A.4}$$

where the factor of two in the denominator is the result of the ability of either fork to transition to the stalled state. R is fit from the stoichiometry measurements and is therefore observable.

The second observable is the lifetime of a focus which is measured in the replisome dynamics experiments. As we demonstrate in the next section, the average lifetime is:

$$\bar{t} = \frac{4k_-^2 + 5k_-k_+ + k_+^2}{2k_-^2(2k_- + k_+)}, \tag{A.5}$$

as a function of the independent-fork model rates $k_\pm$. We will discuss the effective rate of focus loss $k_{\text{eff}} \equiv \bar{t}^{-1}$. An intuitive limit for k_{eff} is the large k_+, where the effective rate of focus loss is:

$$k_{\text{eff}} \approx \frac{2k_-^2}{k_+}. \tag{A.6}$$

The rate has a squared dependence on k_- and inverse dependence on k_+ since two successive conflict transitions are required without restart to lose a focus. This limit is *not* relevant in the current context since $k_- \approx k_+$.

A.3 Technical details: The first-passage-time calculation.

To analyze the lifetime, we will solve the first passage time problem using the canonical method: We make the zero-focus state absorbing, preventing transitions from 0 to 1 forks. The average lifetime for system is defined:

$$\bar{t} \equiv \int_0^\infty \mathrm{d}t \; t \; p_{1\to 0}(t), \tag{A.7}$$

where $p_{1\to 0}(t)$ is the probability density of transitions from the one-fork to the zero fork state. Number conservation implies:

$$p_{1\to 0}(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} P_2(t) \\ P_1(t) \end{pmatrix}, \tag{A.8}$$

where P_i is the probability of occupancy for the i fork state. To perform this calculation, we project onto the N_2-N_1 plane:

$$\boldsymbol{K}_* = \begin{pmatrix} -2k_- & k_+ \\ 2k_- & -k_+ - k_- \end{pmatrix}. \tag{A.9}$$

We integrate Eqn. A.7 by parts. We then use the solution to the matrix linear differential equation for the probabilities, given the initial conditions:

$$\begin{pmatrix} P_2(t) \\ P_1(t) \end{pmatrix} = \exp\left[\boldsymbol{K}_* t\right] \begin{pmatrix} P_2(0) \\ P_1(0) \end{pmatrix}. \tag{A.10}$$

The resulting integral can be easily evaluated:

$$\bar{t} = \int_0^\infty \mathrm{d}t \; \begin{pmatrix} 1 & 1 \end{pmatrix} \exp\left[\boldsymbol{K}_* t\right] \begin{pmatrix} P_2(0) \\ P_1(0) \end{pmatrix}, \tag{A.11}$$

$$= -\begin{pmatrix} 1 & 1 \end{pmatrix} \boldsymbol{K}_*^{-1} \begin{pmatrix} P_2(0) \\ P_1(0) \end{pmatrix}. \tag{A.12}$$

The inverse rate matrix $\boldsymbol{K}_*^{-1}$ can be computed using the well-known closed-form expression for a 2-by-2 matrix inverse. We initialize the system in a steady-state mix of two and one

fork states (Eqn. A.4):

$$\begin{pmatrix} P_2(0) \\ P_1(0) \end{pmatrix} = \frac{1}{2k_- + k_+} \begin{pmatrix} k_+ \\ 2k_- \end{pmatrix}. \tag{A.13}$$

The average lifetime is therefore:

$$\bar{t} = \frac{4k_-^2 + 5k_-k_+ + k_+^2}{2k_-^2(2k_- + k_+)}, \tag{A.14}$$

which is measured in the replisome dynamics experiments.

Aside: Note that the independent-fork model does not appreciably affect the lifetime distribution function (Eqn. A.8) for $R \approx 1$ relative to the cooperative model and therefore a new maximum likelihood fit need not be performed.

A.4 The underlying rates: Estimating the independent-fork model parameters.

R is fit from the stoichiometry measurements and is therefore observable. We estimate the ratio is

$$R = 1.4 \pm 0.4 \tag{A.15}$$

from the focus stoichiometry measurements. We estimate the effective focus loss rate k_{eff}:

$$k_{\text{eff}} = 0.12 \pm 0.01 \text{ min}^{-1}, \tag{A.16}$$

from the replisome dynamics experiments.

Even though the ratio of one to zero fork replisomes cannot be directly measured, since non-replicating cells anomalously increase the number of zero-fork cells, we can still estimate this quantity in the model:

$$R_{10} \equiv N_1(\infty)/N_0(\infty) = 2k_+/k_- = 4R, \tag{A.17}$$

which can be expressed in terms of the observable ratio R. We therefore expect:

$$R_{10} = 5.8 \pm 1.6, \tag{A.18}$$

The probability of a conflict in a single fork is

$$P_c = \frac{k_-}{k_- + k_+} = \frac{1}{1+2R} = 0.26 \pm 0.05. \tag{A.19}$$

We now explicitly estimate the conflict and restart rates from the independent-fork model. The conflict rate per fork is:

$$k_- = k_{\text{eff}} \frac{1 + \frac{5}{2}R + R^2}{(1+R)} = 0.33 \ \pm 0.09 \ \text{min}^{-1}, \tag{A.20}$$

and the restart rate per stalled fork is:

$$k_+ = 2k_- R = 0.95 \pm 0.71 \ \text{min}^{-1}. \tag{A.21}$$

We now estimate the number of conflicts per cell cycle. Assuming a replication time of T, we would expect:

$$N_c = k_-(1 - P_c)T = \frac{k_- k_+}{k_- + k_+} T, \tag{A.22}$$

conflicts per fork per round of replication. Our naïve estimate for the cooperative model is:

$$N_c' \equiv k_{\text{eff}} T. \tag{A.23}$$

The following table summarizes the conflict number predictions:

Replication time: T	Independent conflict number: N_c	Cooperative conflict number: N_c'
40 min	9.5 ± 0.8	4.8 ± 0.4
80 min	19.5 ± 1.6	9.7 ± 0.8
120 min	28.7 ± 2.4	14.5 ± 1.2

From the visualization of DnaN throughout the cell cycle, we estimate that the replication time under these growth conditions is 80-100 minutes (unpublished).

A.5 Discussion.

It is important to note that both the cooperative and independent-fork models make a number of simplifying assumptions. For instance, these models do not include the known position of

conflicts at the rDNA and other loci. In the model, we treat the generation of conflicts as a Poisson process with a constant rate. The existence of these sequence-specific hot spots could act to over-estimate the rate of conflicts if these forks undergo multiple rounds of abortive restart. Our dynamics experiments also do not detect short-lived conflicts that are resolved quickly. The independent-fork model also assumes that there is no interaction between the forks. Our data suggest that the forks are frequently co-localized, which is consistent with a role for functional coupling between forks.