Illuminating the Replisome, One Molecule at a Time

MCGILL UNIVERSITY

DOCTORAL THESIS

Author	**Supervisor**
Nitin Kapadia	Dr. Rodrigo Reyes-Lamothe

April 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Department of Biology

I dedicate this thesis to all my family and friends in Toronto and Montreal, for all their support

Acknowledgements

First, I would like to thank my supervisor, Rodrigo Reyes-Lamothe, for making my experience in graduate school truly exceptional. During my time here, I was always excited to go to the lab, even when I had hit a roadblock, and that was in large part to the atmosphere Rodrigo created in the lab. He gave me the confidence and freedom to explore ideas I thought were interesting and would treat me, along with others in the lab, as his equals. Even in our moments of disagreement regarding the science, he would always keep an open mind and that meant a lot to me.

Second, I would like to give thanks to former and current members of the Reyes Lab. Their support and advice on the work presented in this thesis has been of great value. I would like to give special thanks to Thomas Beattie, a postdoctoral fellow who was there the moment I first arrived to the lab. Aside from his friendly demeanour and patience, making it very easy to work with him, his mentorship during my early years has left a lasting impact on how I approach scientific questions. His humility, combined with intellectual rigour, is something I hope to carry on with me in my future endeavours.

Next, I would like to thank my committee members, Prof. Adam Henricks and Prof. Jackie Vogel. Their feedback on projects have been instrumental on improving the quality of the work and direction. I would also like to thank Prof. Stephanie Weber, whose enthusiasm and insight during lab meetings has been incredibly helpful.

Much of the work on budding yeast would not have been possible without the help from Vogel lab members, including Eric Yen, Kristy Shulist, and Allen Leary.

I don’t think completing my PhD would have been possible without the support from friends, and two in particular I would like to thank are Daneck Lang-Ouellette and Ziad El-Hajj. I am truly grateful to have them as friends and we have shared many memories together, but what I will cherish the most is simply going for drinks and getting into discussions on esoteric subjects and the absurdities of the world around us (e.g. picturesque construction sites in Montreal, free will, Boris Johnson).

Finally, I would like to thank my family, especially my mother and my two older sisters, Neema and Seema. As with most family members of PhD students, they never fully understood what my

PhD was all about, but they were always there to encourage me because they knew I was passionate about my work and science in general, and what it meant to me. Although my mother didn’t study science, she did teach me one thing from a very young age that I’ve kept in mind throughout my PhD: whether it be drawing superheroes, building Lego structures, playing guitar, or solving math equations, put care in your work and be proud of your accomplishments.

Statement of Originality

I, Nitin Kapadia, claim the following parts of the thesis are distinct contributions of original knowledge and original scholarship.

- Observing fast *E.coli* polymerase dynamics and characterization of replisome proteins at the single molecule level in live cells, along with developing analysis methods to quantify residence times and rebinding (Chapter 3). Published in Beattie, T.R and Kapadia, N., et al. *eLife* 2017;6:e21763.
- Developed a machine learning method to classify DNA-bound proteins from single-molecule tracking experiments, in order to obtain accurate estimates of residence times. This method was shown to be accurate *in silico*, and in live *E.coli* and budding yeast (Chapter 4). Manuscript in preparation: Kapadia, N., et al. 2019.
- Developed experimental and analysis protocols to visualize DNA-bound proteins at the single-molecule level in live budding yeast and estimate their residence times. While a few single-molecule studies have been done in budding yeast, they have been mainly focused on fast capture data (Chapter 5). Manuscript submitted: Kapadia, N., et al. 2019.
- Characterization of binding kinetics of core budding yeast replisome proteins at the single-molecule level in live cells. (Chapter 5). Manuscript submitted: Kapadia, N., et al. 2019.
- Observation of stable Pol δ directly in live cells, suggesting a different eukaryotic replisome architecture (Chapter 5). Manuscript submitted: Kapadia, N., et al. 2019.

Contribution of Authors

This thesis is manuscript-based. Chapters 3-5 are based on manuscripts.

Chapter 3

I analyzed all the single-molecule tracking data and developed the image analysis methods. T.R. Beattie collected single-molecule tracking data. E. Nicholas assisted with FRAP data collection. S. Uphoff wrote the PALMtracker software. A.J.M Wollman helped with FRAP analysis. R. Reyes-Lamothe collected the FRAP data and supervised the project.

Chapter 4

I conceived the project, wrote the Python computer simulations of images, collected the data for both *E.coli* and budding yeast, wrote the Matlab code for machine learning and subsequent analysis, and wrote the manuscript. Z.W. El-Hajj constructed the Top2-Halo strain. R. Reyes-Lamothe supervised the project.

Chapter 5

I conducted the experiments and developed the protocol, analyzed the data and wrote Matlab code for image analysis, wrote the Python computer simulations of Pol δ binding, and wrote the manuscript. Z.W El-Hajj conducted the control experiments on strains (growth curves, western blots, and flow cytometry), and constructed strains. T.R Beattie helped construct strains. A.Yu conducted initial HaloTag experiments. R. Reyes-Lamothe supervised the project and wrote the manuscript.

Abstract

DNA replication is a crucial event in the life of a cell, and the cellular machinery responsible for replicating DNA is a multi-protein complex known as the replisome. The replisome needs to replicate DNA in an efficient manner, while ensuring high fidelity to preserve genome integrity, as errors can lead to mutations which can potentially result in cell death or cancer. One important question is do replisomes from a range of organisms behave and have structures similar to one another – including in cases where the proteins do not share common ancestry – implying that there is a general mechanism for how DNA replication is performed. To address this question, I used single-molecule tracking (SMT) to directly visualize and quantify the dynamics of replisomal proteins in live *Escherichia coli* (*E.coli*) and *Saccharomyces cerevisiae* (budding yeast). In the process, I also developed a novel method using machine-learning for classifying DNA-bound molecules from SMT experiments.

As a result of the work presented here it was revealed that the *E.coli* DNA polymerases exchange quite frequently over the course of replication. In contrast, budding yeast DNA polymerases remain stably associated to the replisome implying that the two organisms do not share similarities with respect to polymerase dynamics. However, the results suggested that both the leading and lagging strand polymerases in budding yeast are physically coupled to the replisome akin to the *E.coli* polymerases. Therefore, the eukaryotic and bacterial replisomes may have converged to similar architectures. This work reveals *in vivo* dynamics of the replisome, providing new insight into how replisomes function, along with contributing a technical framework for future questions.

Résumé

La réplication de l'ADN est un évènement crucial dans la vie d'une cellule. La machinerie cellulaire responsable de la réplication de l'ADN est composée d'un complexe de plusieurs protéines, mieux connu sous le nom du réplisome. Le réplisome a besoin de répliquer l'ADN de manière efficace, et aussi d'assurer une intégrité génomique de grande qualité, puisque des erreurs de réplication peuvent mener à des mutations et potentiellement à la mort cellulaire ou bien le cancer. Une question importante est : Est-ce que les réplisomes de différents organismes partagent des fonctions et des structures similaires entre eux, même dans les cas où ils ne partagent pas d'ancêtre commun? Ceci impliquerait ainsi d'un mécanisme général selon lequel la réplication de l'ADN pourrait être effectuée. Afin d'adresser cette question, j'ai utilisé la suivie de molécule unique afin de visualiser, et ainsi quantifier la dynamique des protéines du réplisome *In vivo* chez *Escherichia coli* (*E.coli*) et chez *Saccharomyces cerevisiae* (levure bourgeonnante). Au cours de ce processus, j'ai aussi développé une nouvelle technique utilisant l'apprentissage machine afin de classifier les molécules liées à l'ADN, à partir de nos expériences de molécules unique.

Les résultats du travail présenté ci-dessous démontrent que l'ADN polymérase d'*E.coli* est interchangé très rapidement au cours de la réplication de l'ADN. Contrairement à *E.coli*, l'ADN polymérase de la levure bourgeonnante demeure fermement associée au réplisome, ce qui implique que les deux organismes ne partagent pas de similarité entre la dynamique de leur polymérase respective. Par contre, les résultats suggèrent que les deux polymérase sur les brins d'ADN précoce et tardif chez la levure bourgeonnante, sont physiquement liées au réplisome, comme chez *E.coli*. Donc, les réplisomes eucaryote et bactérien semblent avoir convergé vers la même architecture. Ce travail a révélé la dynamique *in vivo* du réplisome, fournissant un nouvel aperçu du fonctionnement du réplisome, ainsi que contribuant un nouveau cadre technique afin de répondre aux questions futures.

Table of Contents

Chapter 1

Introduction

Deoxyribonucleic acid (DNA) is the molecule that stores the genetic information required by organisms to properly function and develop, and its inheritance is the basis by which evolution of life can occur [1]. It is composed of two strands arranged in an antiparallel manner to form a double helix, which needs to be unwound for DNA replication to occur [1]. DNA replication occurs in a semiconservative manner whereby one strand is synthesized continuously in the direction of DNA unwinding (termed the leading-strand), whereas the other strand is synthesized discontinuously as a result of synthesizing in the opposite direction of unwinding (lagging-strand) [2]. This results in fragments of DNA on the lagging-strand, called Okazaki fragments, which are subsequently ligated together to form a continuous strand [3]. The protein complex responsible for orchestrating these activities is known as the replisome, which needs to perform them in a tightly regulated manner to ensure the faithful transmission of genetic information to daughter cells.

The range and scale over which replisomes have been studied is quite expansive; from biochemical, ensemble studies to single-molecule fluorescence microscopy, and model systems ranging from bacteriophages to mammalian cells [4]. The most well-characterized systems are prokaryotes and phages, with the replication processes in eukaryotes comparatively less known [5, 6]. As a consequence of DNA replication arising at least twice over the course of evolution, the bacterial and eukaryotic replisome proteins do not share common ancestry, but the question remains as to whether their replisomes share any similarities in how they function and their structures [7].

While a stable replisome composition is thought to ensure high fidelity during DNA replication, it is conceivable that some form of dynamic behaviour and architectural change can also help in this regard, especially when encountering obstacles on DNA, for example: DNA lesions, secondary structures, protein-DNA roadblocks, and transcription machinery [4, 8]. Bacteria and eukaryotes may differ from one another in this regard, especially when considering their differences in chromosome structure [9-11]. For example, eukaryotes have nucleosomes – an octamer histone core wrapped with DNA – which may pose as potential roadblocks for the replisome. Histones also contain important epigenetic marks and it remains a question as to how the replisome coordinates replication with preserving these marks for epigenetic inheritance [12]. Therefore, one can hypothesize that differences in chromosome structure and requirements for coordinated activities may result in the evolution of two distinct replisomes.

To quote the famous physicist, Richard Feynman:

"It is very easy to answer many of these fundamental biological questions; you just look at the thing!" [13]

While Feynman subsequently admits that it is an exaggeration, the quote emphasizes the value of being able to visualize biological processes and structures.

Single-molecule tracking (SMT) is an advanced microscopy technique that allows for visualization and quantification of protein kinetics directly in live cells [14]. The aim of my thesis was to use SMT to determine if the eukaryotic and bacterial replisome converged to similar architectures and dynamics by characterizing the binding kinetics of replisomal subunits in *Escherichia coli* (*E.coli*) and *Saccharomyces cerevisiae* (budding yeast). In the process, I developed novel analysis methods for accurate estimation of binding kinetics, including the use of machine-learning for classification of DNA-bound proteins.

Chapter 2 will provide a review on DNA replication in *E.coli* and budding yeast, to provide context for the thesis. *E.coli* is a Gram-negative bacterium, while budding yeast is a unicellular fungi [5, 6]. They were used as model systems for their respective domains for the work presented in Chapters 3-5.

Chapter 3 will describe the use of single-molecule microscopy and developing analysis tools to probe the kinetics of replisomal components in live *E.coli* to determine if DNA replication proceeds in a kinetically continuous process. We found unexpectedly, that the polymerases exchanged quite frequently. This work was published in *Elife* (2017), where I am a co-first author [15].

Chapter 4 builds off the previous chapter and introduces the use of machine learning to improve the robustness of classifying DNA-bound molecules in single-molecule microscopy experiments and shows how it can be useful in experiments done in *E.coli* and budding yeast. This is currently a manuscript in preparation, of which I am first author.

Chapter 5 applies the techniques and analysis methods from Chapters 3 and 4 to test the binding kinetics of replisomal subunits in budding yeast, to ask whether the eukaryotic replisome functions in a similar manner to the *E.coli* replisome. We found that there might be some structural similarities to the *E.coli*, but with a more stable complex. This is a submitted manuscript of which I am first author.

Chapter 6 will summarize and discuss the results presented in the thesis, while also providing an outlook for future work.

Chapter 2

Introduction to DNA Replication

This literature review will focus on *E.coli* and budding yeast DNA replication, specifically focussing on their replisome structure and activities.

2.1 Chromosomal Structure and Organization

Escherichia coli

E.coli contains a single chromosome (4.6Mbps in size) that is packed ~1000 times – through DNA supercoiling, molecular crowding, and DNA-binding proteins – into a membrane-less nucleoid structure, which contains the DNA as well as DNA-binding proteins [9, 10] (**Figure 2.1)**. The nucleoid is separated from the inner membrane of the cell by the cytoplasm and polyribosomes. Its organization is sensitive to environmental conditions, such as starvation, where it has been proposed to form a phase-separated organelle that is still able to undergo transcription [10]. As part of the compaction process, DNA is organized into looped structures – which also occurs in eukaryotes – by structural maintenance of chromosomes (SMC) proteins. In addition, bacteria have histone-like nucleoid structure (H-NS) proteins which not only assist in the organization and compaction of DNA, but also can inhibit transcription in the case of DNA-H-NS-DNA bridges [10]. Despite their name, H-NS proteins do not share many similarities in protein function to histones found in eukaryotes [10]. There are many additional nucleoid organizing factors reviewed in [10].

The chromosome is organized in a defined manner in the nucleoid with the single origin of replication (*oriC*), located mid-cell prior to replication **(Figure 2.1)**. The left and right macrodomains of the chromosome are positioned on opposite sides from one another, while the terminus region opposite to *oriC*, spans the length of the cell thereby joining the two macrodomains. Genes are arranged in a defined manner along the chromosome to minimize replication and transcription conflicts along with regulating expressions levels [9].

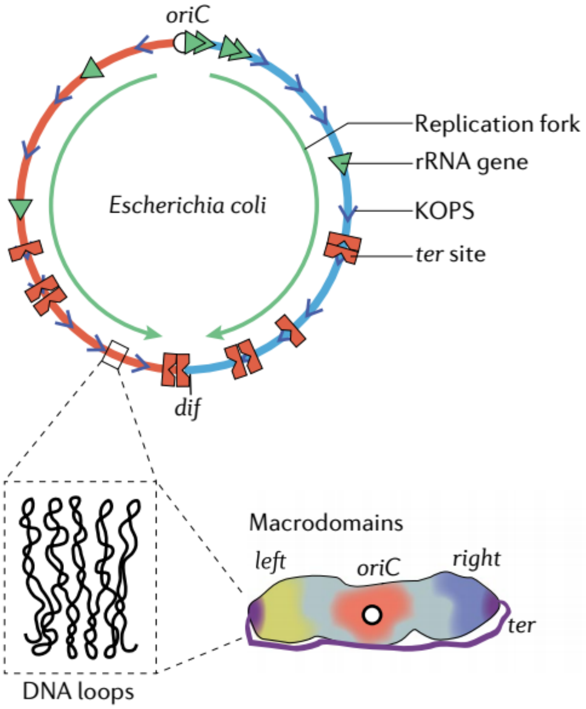


Figure 2.1: Diagram of *E.coli* genome, showing the formation of loops required for compaction as well as organization within the cell. Taken with permission from [9].

In contrast to eukaryotes, the different stages of the cell cycle in bacteria are not distinct; for example, DNA segregation occurs simultaneously with DNA replication. Initiation of replication begins at *oriC* which subsequently moves to the cell quarter positions in daughter cells during segregation [9].

Saccharomyces cerevisiae

Eukaryotes not only have multiple chromosomes, but also multiple origins of replication on each chromosome, possibly to accommodate a larger genome size, while still ensuring that DNA replication occurs in a short amount of time. In addition, another major difference with respect to bacteria is that DNA is organized into chromatin, through its interaction with histones, which leads to a greater level of compaction than in bacteria, as well as providing a platform for epigenetic inheritance, and gene expression [11]. The linear eukaryotic chromosomes are also characterised by the presence of telomeres, which are repetitive sequences at the ends of chromosomes that protect against chromosome shortening due to DNA replication [11]. Haploid budding yeast cells contain 16 chromosomes, encompassing 12Mbps of DNA, that are non-randomly positioned in the nucleus [16, 17]. For example, heterochromatin regions of DNA (transcriptionally inactive) are found around the nuclear periphery, and centromeres on yeast chromosomes are bound by the spindle pole body (SPB), proving a structural constraint to chromosome location and movement [11]. The nucleus also contains the nucleolus, which occupies approximately one third of the nuclear volume and primarily functions as a factory for ribosome biogenesis [11] (**Figure 2.2**).

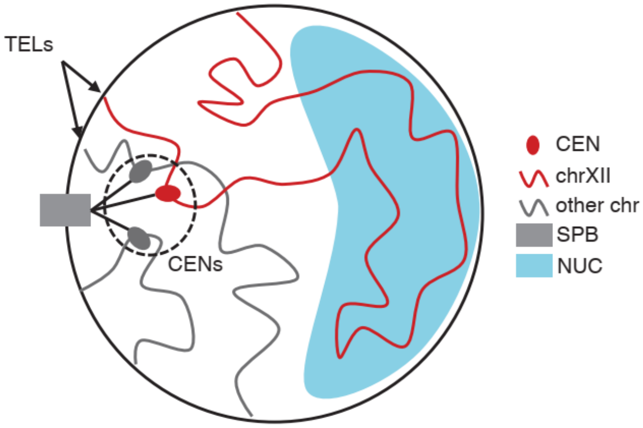


Figure 2.2: Diagram of the structure of the budding yeast nucleus during interphase, representing the RabI organization. Shown are the nucleolus (NUC), spindle pole body (SPB), telomeric regions (TEL), and centromere (CEN). Taken with permission from [18].

Budding yeast chromosomes also have a high gene density, with one gene every ~ 2kb, but have different mechanisms than *E.coli* to prevent conflicts with transcription and replication machineries: natural replication fork barriers exist for highly transcribed ribosomal DNA (rDNA) regions, and transcription and replication seem to be separated temporally [16, 19].

2.2 Initiation of Replication

Escherichia coli

OriC contains an AT-rich DNA unwinding element (DUE) and binding sites for DnaA, the protein responsible for chromosomal replication initiation in bacteria [20]. It is thought that upon binding to *oriC*, DnaA homo-oligomerizes, forming a helical filament on DNA, facilitating the melting of the DUE region, which allows for DnaB helicase – a homohexamer that unwinds DNA – to be loaded onto the lagging strand template [20] (**Figure 2.3**). Only DnaA bound to adenosine triphosphate (ATP) is capable of initiating replication. Given the major role DnaA plays in initiation, it is also the main target of regulation to prevent reinitiation. First, on newly replicated DNA, DNA adenine methylase (DAM) will methylate the A residues in GATC sequences, which exist in 11 copies in *oriC* [21]. Hemimethylated DNA are binding sites for SeqA; therefore, this not only prevents the binding of DnaA-ATP to its binding sites within *oriC*, but also the gene for DnaA (*dnaA*) is located near *oriC* and contains multiple GATC sequences, resulting in inhibition of transcription of DnaA by SeqA [21]. Second, DnaA regulates itself by binding to the promoter region for *dnaA* [21]. Third, the DnaA-ATP form is reduced by ATP hydrolysis through the regulatory inactivation of DnaA (RIDA) system. This system consists of the DNA-bound processivity clamp (β), DnaA homologue protein (Hda), and ADP. Hda is bound to β, and ADP activates Hda [21]. The interaction of Hda with DnaA-ATP stimulates the hydrolysis of the ATP. Therefore, this results in less DnaA-ATP available to bind to *oriC* [21]. Fourth, DnaA binding sites are also located in a region called the *datA* locus. This effectively titrates DnaA away from *oriC* [21].

DnaB is loaded by DnaC, a helicase loader that breaks open the ring structure of DnaB, allowing for DnaB to encircle single-stranded DNA (ssDNA) [22]. However, DnaC inhibits DnaB helicase activity implying that DnaC must dissociate before unwinding can begin. It has been suggested that DnaG (primase) binding to DnaB leads to a conformational change in DnaB that releases DnaC [23].

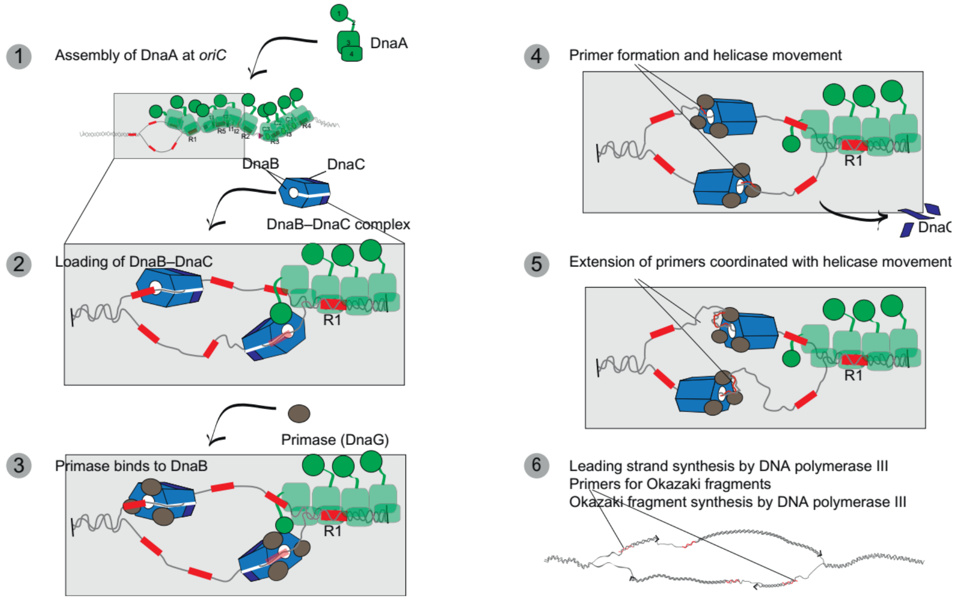


Figure 2.3: Diagram illustrating the steps to initiation and assembly of the replisome. Taken with permission from [24].

The other replisome components including the clamp loader complex, PolIII (polymerase), β clamp are subsequently incorporated into the replisome complex, with DnaB acting as a platform for recruitment of these factors [25] (**Figure 2.3**). The complex of PolIII, clamp loader, and β clamp is known as the PolIII holoenzyme; the complex without β clamp is known as PolIII*. After assembly at the origin, each sister replisome will move away from each other at an average rate of 600bp-1000bp/s @ 37 degrees Celsius (or 300bp/s @ 23 degrees Celsius), to complete 2.3Mbps which takes a little over two hours to complete at room temperature [26-29].

Saccharomyces cerevisiae

While *E.coli* has only one origin, eukaryotes have a much higher number of origins (> 700 potential origins), leading to a more complicated firing scheme [5]. In budding yeast, there are an estimated 300 replication forks active in mid S-phase [30]. The current model in eukaryotes is that replisomes form "replication factories" where sister replisomes remain in close proximity to one another (and possibly physically coupled), and pairs of sister replisomes are grouped together at discrete sites [30-32]. However, *in vitro* single-molecule assays using *Xenopus* egg extracts have indicated that sister replisomes act independently from one another and are not physically connected, and a significant number of factories in budding yeast only contain one pair of sister replisomes [30, 33]**.** Future advancements in high resolution microscopy and single-molecule techniques will hopefully provide more detail into the organization of sister replisomes.

Origin recognition complex (ORC) binds to specific sequences on DNA representing origins of replication, known as autonomously replicated sequences (ARS), throughout the cell cycle [5, 34, 35]. Within these regions, contain an AT-rich consensus sequence important for ORC binding and thus replication [5]. After ORC is bound, Cdc6 is recruited leading to the recruitment of the MCM 2-7 (MiniChromosome Maintenance) helicase/ Cdt1 complex [36]. MCM loading during G1 is referred to as licensing and it is tightly regulated; importantly, during S-phase MCM subunits are exported out of the nucleus to prevent relicensing [37]. Loaded MCMs are not active in G1; it is only upon phosphorylation by DDK (Dbf4-dependent kinase) leading to the recruitment of Cdc45, and CDK(Cyclin-dependent kinase) in early S-phase, leading to the recruitment GINS, that an active helicase is formed, known as the CMG helicase [38, 39]. CDK activities also helps recruit the leading strand polymerase, Pol ε [40, 41].

In contrast to origin firing in *E.coli* where a single origin is used every cell cycle, eukaryotic origin firing is stochastic [5, 42]. Although at the population level certain origins will have a higher probability or efficiency to fire, at the single cell level, the subset of activated origins will vary for each cell cycle [42]. Origins are also separated temporally with some firing early in S-phase (early origins) and others firing late (late origins). Early origins also cluster together – thought to happen in G1 – and are typically found in centromeric regions as well as highly-transcribed regions, where the euchromatin state is present, whereas late origins are found in telomeric and heterochromatin regions [42, 43]. A possible advantage for having many origins, and stochastic, temporally separated activation is flexibility; if some origins fail to activate or result in stalled replication forks, their regions can be passively replicated, through the activation of neighboring origins. Although the source of this stochasticity is unclear, one hypothesis is that it involves the recruitment of limiting initiating factors, which include Sld2, Sld3, Sld7, Dpb11, Dbf4, and Cdc45 [44, 45]. These are factors that are in low abundance in the cell but each of which is required for origin activation. This provides stringent control that only a subset of origins are activated since each origin has to recruit all of these limiting factors, and also provides an explanation for stochastic activation; recruitment of these factors is to a certain degree, stochastic, leading to quasi-random selection of origins.

Of course, limiting factors alone does not explain why certain origins tend to fire early while others do not. If the process were completely stochastic, one might not expect a temporal firing pattern. As mentioned previously, early origins cluster with one another and a possible mechanism for this is the binding of a transcription factor named Forkhead (Fkh) [43]. There is an enrichment of FkH binding sites in early origins while late origins are depleted of these sites [43]. It is known that the FkH protein dimerizes and it is thought that this dimerization allows FkH proteins bound to different origins, to cluster these origins. Recently, it was shown that mutants lacking the dimerization motif exhibit altered timing of origin firing (i.e. early origins begin to fire late) [46]. Further, some early origins (e.g. ARS607) will still retain their temporal firing pattern even if moved to a late origin locus, possibly by still being able to cluster with other early origins [47]. It is thought that clustering of origins results in increased local concentrations of the limiting factors [42]. If a few origins within the cluster are activated through the recruitment of the limiting factors, those factors will be recycled to the other origins found within the cluster. Therefore, it results in a cluster of origins with similar activations times (i.e. early origins) **(Figure 2.4)**.

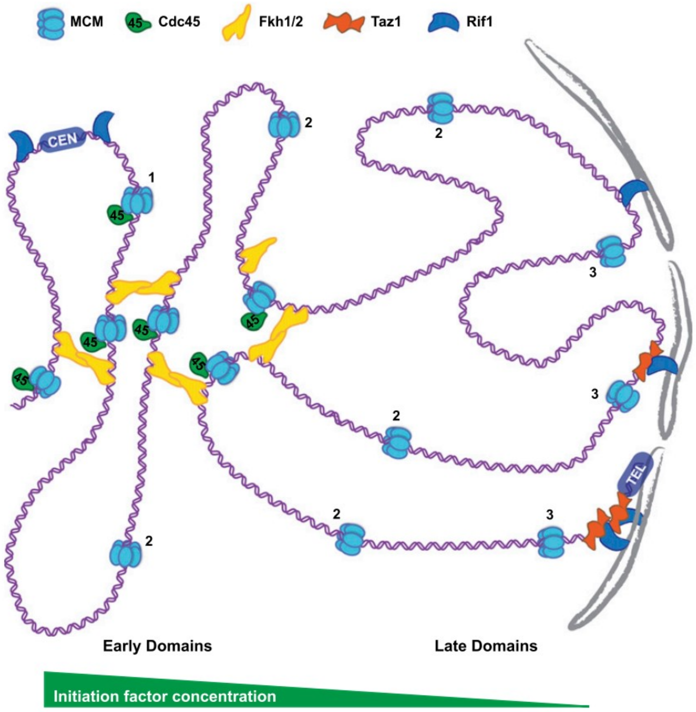


Figure 2.4: Model of the role of FkH proteins in clustering origins, and limiting initiation factors. Early origins cluster, possibly by FkH dimerization, which can increase the local concentration of limiting initiation factors. Taken with permission from [42].

Prevention of re-initiation (the use of an already duplicated replication origin) is crucial for genome integrity. Control mechanisms exist both during G1 and S-phase at both MCM loading and activation levels [5]. During G1, S-CDKs are not present, preventing activation of MCM. Dbf4, a subunit of DDK is degraded during the M-G1 transition, also preventing helicase activation during G1. Helicase loading is restricted to early-mid G1, while in late G1, G1-CDKs are present which phosphorylate Cdc6, leading to its degradation. During S-phase, S-CDK will phosphorylate multiple targets: MCM3 leading to the export of MCM2-7 subunits, Cdc6, and ORC, all of which leads to the inhibition of helicase loading [5].

2.3 DNA synthesis

Escherichia coli

The defining feature of the bacterial replisome is the trombone model: the architecture closely resembles a trombone – with a ssDNA loop on the lagging strand - and has both the leading and lagging strand polymerases tightly coupled to the clamp loader (through the τ subunit of the clamp loader) and facing in the direction of helicase [6, 48] (**Figure 2.5**). The feature of this model is that allows for coupled synthesis between leading and lagging polymerases, as well as to helicase unwinding, thereby preventing excess ssDNA exposure. In addition, since polymerases synthesize 5'- 3', this would pose a topological problem if the two polymerases were to translocate in opposite directions, but this configuration helps to solve that issue [25]. However, although ssDNA loops have been detected in T4 and T7 phage systems, they have yet to be detected in *E.coli* replication systems [25, 49, 50].

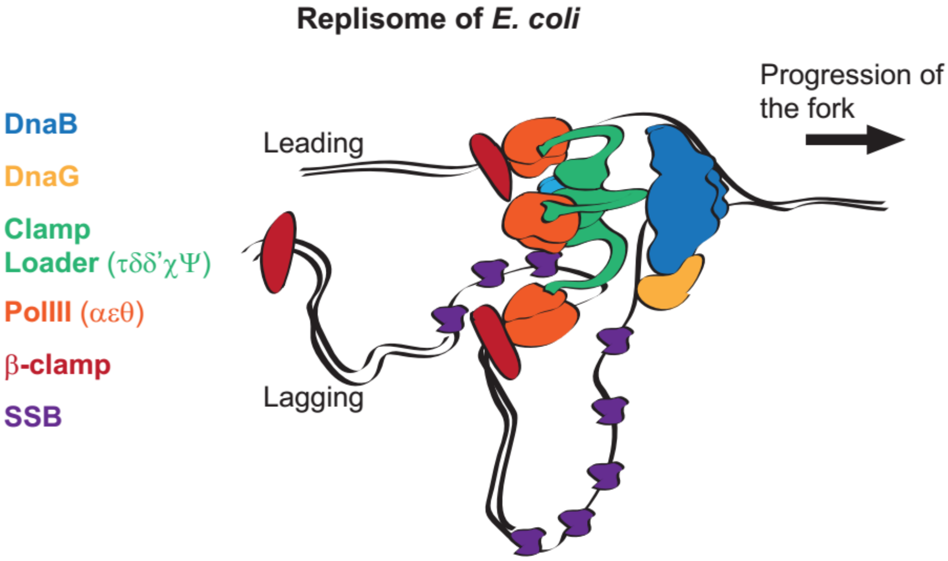


Figure 2.5: Model of *E.coli* replisome. Taken with permission from [4].

The dynamics of the bacterial replisome have been extensively studied *in vitro* through biochemical reconstitutions [27, 28, 51, 52]. The common assay used for *in vitro* studies is visualizing the synthesis of DNA – commonly with an intercalating dye that stains dsDNA – with a rolling circle template under flow conditions with varying concentrations of protein subunits (**Figure 2.6**). This elegant and tightly-controlled system allows one to probe the dynamics of the bacterial replisome - including its processivity and rate of synthesis - and the effects of varying replisomal protein concentration.

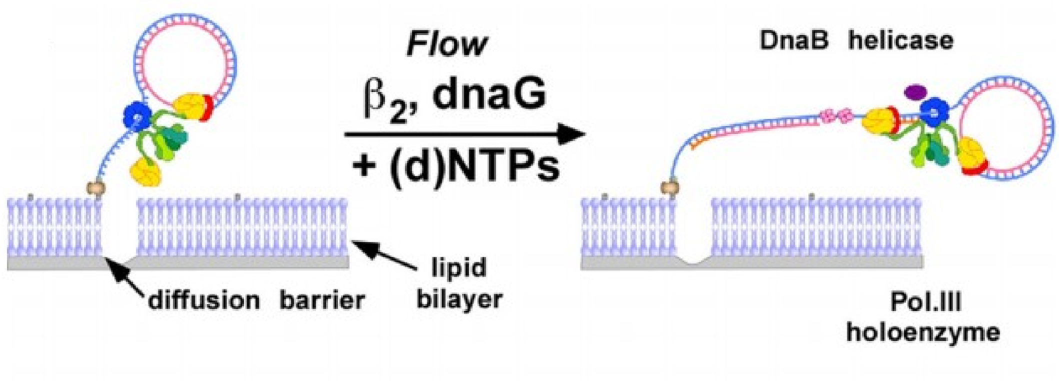


Figure 2.6: Diagram of rolling circle replication with the *E.coli* replisome. A DNA minicircle is attached to a lipid bilayer via biotin. A diffusion barrier is incorporated to align all templates after flow. Replisomes are preassembled and upon addition of additional replication proteins required for synthesis to buffer flow, replication begins. The DNA synthesized by the leading strand polymerase becomes the template for the lagging strand polymerase. Taken with permission from [27].

DnaB unwinds DNA through a steric exclusion model, whereby the leading strand is excluded from the central chamber of the helicase. Acting alone, DnaB has a slow unwinding rate (~30nt/s), but in association with the PolIII holoenzyme, its unwinding rate greatly increases about 10-fold, suggesting a safety mechanism to prevent excess unwinding in the absence of polymerases [53]. Although unclear as to how the presence of the holoenzyme stimulates helicase activity, it has been suggested that the activity of the polymerases helps propel/push the helicase forward. Also, the bound polymerases are thought to provide a tethering support to help stabilize the helicase [52].

In addition to its helicase activity, DnaB is a recruiting platform for the RNA primase, DnaG, which synthesizes a RNA primer of around 12 nucleotides – an important step considering replicative polymerases cannot synthesize DNA *de novo*. Structural studies indicate that DnaB contains three binding sites for DnaG [54, 55]. This does not necessarily mean that all three sites are occupied as DnaG is distributive with DnaB; changing concentrations of DnaG results in changes to Okazaki fragment length, which are in the range of 0.5-2kb [27, 56]. DnaG alone has a very slow synthesis rate, but its activity is greatly stimulated to *in vivo* levels through its interaction with DnaB, by increasing its affinity for template and its catalytic activity [57, 58]. Priming is partly sequence dependent with DnaG being activated upon recognizing a trinucleotide sequence, although priming does not occur at every site [55, 59]. Although DnaG is monomeric in its nonactive form, it has been suggested that it forms a dimer in its active form where priming is done in trans; an intermolecular mechanism where one monomer will recognize the signal sequence while the other will synthesize the primer [55].

The actual workhorse unit responsible for DNA synthesis is the PolIII holoenzyme. Pol III (there are between 2-3 with one on leading strand and the other two on the lagging strand) is responsible for incorporating dNTPs into the elongating DNA strand as well as removing incorrect bases through their proofreading activity, performed by the α subunit and ε subunit, respectively [60]. There is also a θ subunit, although the exact function is unknown. The clamp loader serves many functions including depositing β clamps onto the lagging strand: binding to polymerases and helicase, thereby promoting coordinated synthesis between polymerases and to helicase unwinding, and with its χ subunit interacting with SSB, it is thought it provides a stabilizing force during synthesis [61].

It is thought that PolIII* is highly processive; *in vitro* studies have indicated an average processivity of ~ 80kb, which is still less than the 2.3Mbp required for replicating the *E.coli* chromosome, indicating that excess copies are required for complete genome duplication [27, 51]. Nonetheless, these results indicated that polymerases are stably bound to the replisome [27, 51]. However, these experiments were done without excess PolIII* so there were no components for the engaged PolIII* to exchange with, so it remains an open question regarding the dynamics of replisomal subunits under conditions of excess copies.

In contrast to the leading-strand, multiple activities occur to synthesize the lagging-strand: RNA primer synthesis by DnaG, loading of clamp onto primer template by the clamp loader, loading of PolIII onto template, and release of PolIII from both the clamp and DNA upon competition of an Okazaki fragment. DnaG dissociates after synthesizing a RNA primer (accelerated by the χ subunit of the clamp loader), leaving an exposed primer template. The δ subunit of the clamp loader, breaks open the β clamp ring through ATP hydrolysis, allowing for the clamp to be loaded onto the primer template [60]. The τ subunit of the clamp loader, which interacts with the α subunit of PolIII, helps recruit PolIII to the β clamp at the primer template, whereupon synthesize will begin [25, 60]. Given that the lagging-strand polymerase must be recycled to the next primer template upon completion of synthesis, it raises the question as to how the polymerase is removed from both DNA and β clamp. There are two common models on PolIII release: 1) the collision-release model and 2) signal-release model. The first posits that PolIII and the τ subunit can sense the loss of ssDNA during synthesis which triggers the release of PolIII and the second purposes that a signal triggers release of PolIII when a new primer is synthesized [62, 63]. It is inconclusive as to which model is correct, although PolIII release may be the result of other factors, discussed below. Finally, for Okazaki fragment maturation, the RNA left from the primer is removed by Pol I, which replaces the RNA with DNA, and the left over nick is ligated by DNA ligase A[25]. This process is purposed to be facilitated by β clamp, which has been reported to bind both Pol I and ligase, and is left behind on DNA after being used by PolIII.

Previous models have suggested that the lagging strand polymerase had a higher rate of synthesis to compensate for lagging-strand synthesis events to permit coordinated synthesis [52]. However, to explain near-synchronous synthesis on both strands, a mechanism of stochastic switching to faster or slower rates by the leading and lagging strand polymerases has been proposed [64]. Interestingly, it was found that the polymerases do not synthesize DNA in a kinetically continuous manner, but rather they pause for several seconds, possibly due to unbinding from the β clamp. Upon rebinding to the clamp, they sample a rate of synthesis from the distribution, independent of the previous rate. Therefore, in this model, near-synchronous synthesis is the result of an averaged effect of stochastic sampling of rates. A safety mechanism where DnaB unwinding is slowed down in the absence of synthesis ensures that DNA is not significantly unwound during these pauses. The authors also found that lagging strand synthesis does not perturb the speed of leading strand synthesis, suggesting that the polymerases act independently from one another. [64]. These results highlight that the *E.coli* replisome may operate in a dynamic and noisy manner [65]. Consistent with this idea, it has also been proposed that dynamic behaviour of polymerases may help relieve torsional stress build-up, as a result of the coupled polymerases having to follow a helical path along DNA [66]. Chapter 3 will discuss observing kinetics of PolIII* subunits directly in their natural cellular context.

Saccharomyces cerevisiae

In the eukaryotic replisome, the core components responsible for synthesizing DNA are: CMG helicase (unwinds DNA), Pol ε (leading strand polymerase), Pol δ (lagging strand polymerase), Pol α primase (synthesizes RNA-DNA primer), proliferating nuclear cell antigen (PCNA, processivity clamp), replication factor C (RFC, clamp loader) and replication protein A (RPA, single-stranded binding protein) [5] (**Figure 2.7**). It has been shown that the budding yeast replisome travels at a uniform speed of 25bp/s (@ room temperature), consistent across the genome [67], to replicate on average 15kb [16].

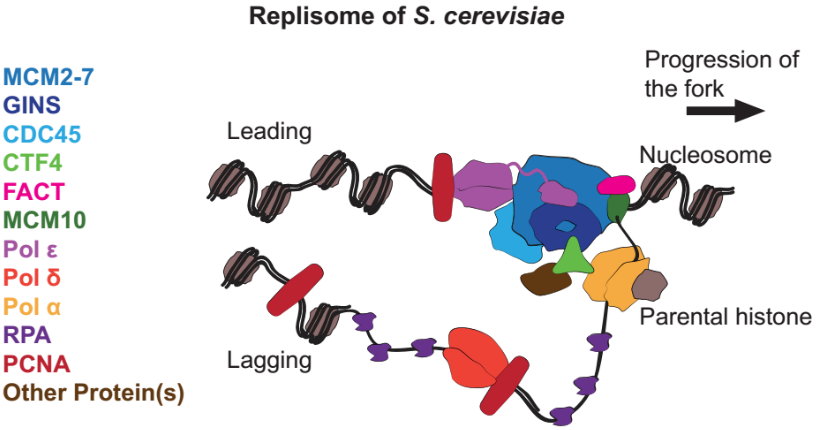


Figure 2.7: Model of budding yeast replisome, showing core components. Taken with permission from [4].

The CMG helicase is thought to be a recruiting platform for other proteins, thus being for an integral part for replisome stability. Although the helicase activity resides in the MCM2-7 subunits through the ATPase activities located at their C-termini, MCM2-7 requires GINS and Cdc45 to be an effective helicase [5]. The mechanism is unclear, but Cdc45 and GINS may help close a gate formed during translocation by MCM2 and MCM5 [5]. It was initially thought that CMG only translocates on the leading strand, but structural studies and *in vitro* reconstitutions with streptavidin roadblocks on DNA fork structures, suggests that dsDNA enters the N-termini of the ring, with an internal unwinding point to separate the two strands [68, 69]. However, a recent study using both *Drosophila melanogaster* CMG and budding yeast CMG, has shown that streptavidin roadblocks on the lagging-strand do not stall CMG unwinding and that the previous results may have been due to impaired productive binding of CMG onto the DNA fork [70].

Pol ε and Pol δ are the main replicative polymerases while Pol α is responsible for synthesizing RNA/DNA primers. Pol ε has been shown to have a flexible linker between the catalytic NTD and non-catalytic CTD, allowing it to engage and disengage with the leading-strand [71]. While Pol ε has a stable interaction with GINS, Pol δ does not appear to be stably connected to the replisome, aside from binding to PCNA [41, 72-74]. Pol α binds to Ctf4, which binds to

GINS as well, implying that there is some connection between the leading and lagging strand components [75]. Interestingly, RFC has been suggested to only act transiently at the replisome, unlike the bacterial replisome, where the clamp loader forms a relatively stable interaction with DnaB and PolIII.

As discussed in [4], a model of Pol δ being physically disconnected from the replisome raises questions as to how identical rates of synthesis on leading and lagging strands, observed *in vitro* and *in vivo,* is achieved [76-78]. Although the replisome does move slowly at a rate of 25bp/s, the *in vitro* rate of synthesis under physiological conditions of Pol δ is only ~ 50bp/s, which likely would not be fast enough to compensate for the delay caused by recruitment of RFC, followed by Pol δ, from the diffusing pool [79]. Also, the current estimates of the copy number of Pol δ subunits, would not support efficient lagging-strand synthesis especially at peak S-phase [30, 80-82]. It is possible that clustering of origins may help increase the local concentration of Pol δ and facilitate recycling of Pol δ to neighboring replisomes. Chapter 5 will explore the use of single-molecule fluorescence microscopy to study budding yeast replisomal factors and proposes a model based on Pol δ being physically connected to the replisome.

Pol α initially synthesizes an RNA primer (7-12nt) but then begins to synthesize DNA for a short stretch, creating a RNA-DNA hybrid primer [83]. While the replicative polymerases have 3'-5' exonuclease activity for proofreading, Pol α does not, which makes it highly error-prone and unsuited for extensive DNA synthesis. Fortunately, Pol δ has strand displacement activity which allows it to remove the DNA and RNA synthesized by Pol α, by creating a flap that is subsequently cleaved by flap endonuclease – 1 (FEN-1) [79, 84]. Pol ε does not possess strand-displacement activity, but it does have an ability to wrap around dsDNA, making it inherently processive unlike Pol δ, which requires PCNA to be processive; however, it has been shown that Pol ε still requires PCNA for maximal rate of synthesis [85-87].

The division of labour in polymerase usage – which is quite different than in *E.coli*, where the replicative polymerase on both strands is PolIII – has been observed *in vivo* both genome-wide and at specific loci, as well as with *in vitro* reconstitutions [74, 78, 88-92]. The *in vivo* approach consisted of using variants of the polymerases that created signature mutations. Essentially, it was found that signatures of Pol δ were located on the lagging strand, while Pol ε signatures were found on the leading strand. Pol α signatures were primarily found on the lagging strand, which is consistent with its role as laying down primers for Okazaki fragment synthesis [84]. *In vitro* experiments have provided a more mechanistic understanding of how this division of labour is achieved (**Figure 2.8**). It is thought that the polymerase strand usage is due to suppression mechanisms; the different polymerases can be incorporated into both strands but are suppressed from their non-canonical strands [74, 78]. In the case of Pol ε, it can be incorporated into the lagging strand, but its affinity for PCNA is quite poor and it is competitively inhibited by RFC, leading to its ejection from the lagging strand. On the leading strand, a strong affinity for the CMG helicase through its interaction with GINS, helps retain Pol ε on the leading strand [74]. In contrast, Pol δ has a strong interaction with PCNA so it is retained on the lagging strand, but on the leading strand it is thought it collides with the CMG – which it does not have a strong affinity for – leading to its release, known as the collision-release model [74].

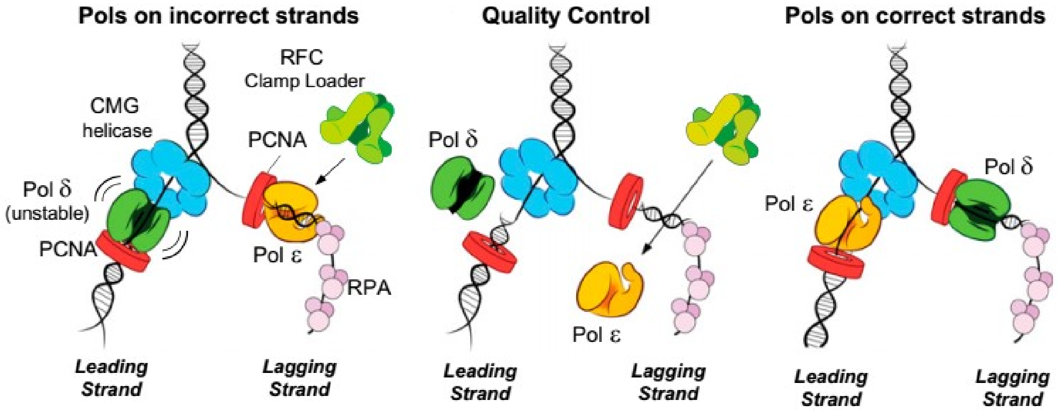


Figure 2.8: Model for how the division of labour of the eukaryotic polymerases is achieved. Left) Pol δ does not have a stable interaction with CMG on the leading strand, and colliding with it may lead to its release from the strand. Pol ε is not stable with PCNA and is inhibited by RFC. Middle) Ejection of polymerases from their non-canonical strands. Right) Polymerases shown on their correct strands. Taken with permission from [74].

A recent study has indicated that Pol δ synthesizes both strands, while Pol ε is merely used for removing errors created by Pol δ [93]. This would be consistent with genetic evidence indicating that the Pol ε catalytic domain is dispensable for viability, possibly because Pol δ and/or Pol α synthesizes the leading strand [94]. However, the cells are not very healthy and point mutations in the catalytic domain are lethal, possibly because Pol ε simply sits at the leading strand preventing access to Pol δ or Pol α[5, 94]. Therefore, although Pol δ or Pol α can potentially synthesize the leading strand in the absence of Pol epsilon, they may not be efficient.

A more appropriate question might not be *which* strand each polymerase synthesizes, but *when* they synthesize either strand. It is possible that all three polymerases may act on both strands with varying degrees of synthesis at different times during replication [78]. This would be consistent with recent evidence both *in vitro* and *in vivo* suggesting that Pol δ may help establish leading strand synthesis, but given its inefficiency on the strand, Pol ε will eventually take over for bulk leading strand synthesis [87, 95, 96]. It has been suggested that the flexible linker found in Pol ε, allows it to engage in leading-strand synthesis once Pol δ has initiated it [71]. Pol α can also potentially synthesize both strands, but its DNA synthesis – and not priming activity – is thought to be inhibited by RFC, through competition for the primer terminus [74]. Therefore once, priming has occurred, Pol α will only synthesize a short fragment of DNA before RFC binds to the primer terminus and loads PCNA, leading to the recruitment of the replicative polymerases [74]. This would help ensure large fragments of DNA synthesized by the error-prone Pol α, is not incorporated into nascent DNA.

In vitro reconstitutions have indicated that having a eukaryotic replisome with simply the core components - known as the minimal replisome - is not sufficient to replicate at speeds determined *in vivo* (mean ~ 25bp/s) [67, 87, 97]. For this to be achieved, it requires additional proteins to form what is known as the replisome progression complex (RPC). Some of the proteins include: Ctf4, Csm3, FACT, Mrc1, Tof1, MCM10, and Top1 [72, 73]. This in contrast to the bacterial replisome which requires only a few components to replicate at a very efficient rate (600bp/s @ 37 degrees) [26].

One of the key complexes governing replisome speed is the Mrc1-Tof1-Csm3 (MTC) complex. This complex has been thought to form a stable complex with the rest of the replisome during replication stress [98]. In addition, it is thought to be responsible for increasing fork speed during normal replication [99, 100]. *In vitro* reconstitutions have demonstrated that addition of the MTC to the minimal replisome helps achieve fork speeds similar to *in vivo* measurements [87]. Interestingly, single-molecule *in vitro* experiments have indicated that MTC acts only transiently with the replisome, but upon binding it increases fork speed, suggesting that fork speed may fluctuate during replication [101]. The mechanism behind the increase in fork speed is inclear, although it has been suggested that Mrc1 may stimulate CMG unwinding through its interaction with MCM and Pol epsilon [101]. Tof1 and Csm3 may simply help stimulate Mrc1 activity.

MCM10 is also thought to play a diverse role in the replisome. It has been suggested that it helps the transition of loaded CMGs on dsDNA into loaded CMGs on ssDNA, through melting of origin DNA [5]. In addition, it has been shown to not only stabilize the CMG, but also increase elongation rates both *in vivo* and *in vitro* [101, 102]. *In vitro* experiments have also indicated that it may help with lesion bypass on the lagging strand by inducing conformational changes in the MCM2-7 helicase to exclude the lagging strand from the MCM ring [103].

It is important to keep in mind that eukaryotic replisomes encounter obstacles similar to the bacterial replisome (e.g. lesions, transcription machinery, DNA-bound proteins, etc.) but unlike their bacterial counterpart, they also frequently encounter nucleosomes approximately every 160bp [5]. FACT specifically has been shown to be required *in vitro* for chromatin replication to achieve *in vivo* rates, by being involved in disassembling nucleosomes ahead of the fork and possibly in redepositing histones behind the fork as well [97]. In addition, multiple replisomal proteins are involved in histone chaperone activities including: MCM2, Pol α, Ctf4, and Pol ε [5, 104-108].

Both *in vivo* and *in vitro* studies indicate that chromatin also plays a role in Okazaki fragment length. In *E.coli,* Okazaki fragments have a relatively broad range from 0.5-2kb [27, 56]. In contrast, budding yeast Okazaki fragments have a narrow range of ~160bp, roughly the distance between nucleosomal midpoints [109]. This is consistent with more recent *in vitro* data with chromatinized templates that show a similar distribution of Okazaki fragments [77, 97]. Non-chromatinized templates resulted in a broader distribution of fragment lengths [77]. The strict Okazaki fragment length is likely a consequence of two factors: 1) Non-distributive behaviour of Pol α to prime in regular intervals and 2) Pol δ synthesis being inhibited by nucleosomes. Pol α is thought to be recruited to the replisome through its interaction with Ctf4. However, it has been shown that Pol α still can prime in the absence of Ctf4 *in vitro,* with fragment lengths similar to *in vivo* measurements, and a strict Okazaki fragment distribution with chromatinized templates, suggesting that chromatin itself, helps retain Pol α [78, 97]. One question would be how chromatin specifically recruits Pol α to the replisome and not to other sites on DNA, and the possibility that the interaction of Pol α to Ctf4 is more important for its histone chaperon activities [105, 106]. Chapter 5 will offer more evidence on the role (or lack thereof) of Ctf4 in retaining Pol α at the replisome.

Regarding the role of Pol δ in the periodicity of Okazaki fragments, Pol δ with PCNA could presumably synthesize and strand displace DNA for long stretches generating long Okazaki fragments, as well as displacing DNA that was already synthesized by Pol δ [77, 79]. The fact that Okazaki fragments are much shorter is likely a consequence of it encountering an obstacle, and the best candidate for this would be nucleosomes. Evidence for this was observed *in vivo* using a method to map Okazaki fragments and found termination of Okazaki fragments within nucleosomes [109]. The inhibition of elongation by nucleosomes was also confirmed *in vitro* [77]. These findings suggest an important role of chromatin for lagging strand dynamics. Leading-strand synthesis is likely unperturbed given that histones are deposited in the opposite direction of synthesis.

Among the RPC proteins, for many of which we do not have a clear functional understanding, Ctf4 is one where we do have some clear understanding in its structure and function. The EM and crystal structure of Ctf4 indicates that it exists as a disk-shaped trimer [75]. The structure indicated that it simply linked lagging strand and leading strand components, through its interaction with both Pol α and GINS. However, recent evidence indicates that Ctf4 also acts as a hub at the replisome, recruiting factors with diverse roles, some not even related directly to DNA replication [110]. Some of these include: Dna2 (processing Okazaki fragments and DNA end resection), Tof2 (role in rDNA), Dpb2 (subunit of Pol epsilon, interacts with GINS as well), Ch1 helicase (involved in sister chromatid cohesion) [110, 111]. These proteins have a Ctf4-interacting peptide (CIP) box that allow them to bind to Ctf4. It is still unclear how these various proteins, including Pol α, compete for the binding sites on Ctf4. It has been suggested that some of these interactions are weaker than others leading to a hierarchy of recruitment [110]. Ctf4 also plays a role in preventing double strand breaks and end resection that could potentially lead to genomic rearrangements, possibly by stabilizing replisome components at stalled forks [112].

2.4 Termination of Replication

The finale to the replisome's journey is termination, leading to its disassembly. This process however is not well understood in bacteria, and even less so in eukaryotes.

Escherichia coli

E.coli has multiple termination sites – known as *Ter* sites - located approximately at the midpoint on the chromosome. These sites have a polarity in their sequence giving them a permissive (P) and non-permissive face (NP) [113] (**Figure 2.9**). Once bound by the monomeric protein Tus they can form a barrier for the replisome depending on whether the replisome approaches the block from the permissive or non-permissive face. This would effectively stall one replisome, while the other sister replisome approaches from the opposite direction, eventually resulting in fork fusion and disassembly; it is thought that there is no active mechanism to cause disassembly, which is a possible explanation for why there is a reloading mechanism in bacteria [114] (**Figure 2.9**). The mechanism for this polarity is thought to do with the structural changes induced at the Tus-*Ter* interface by the unwinding activity of DnaB: At the NP interface, the unwinding of DNA can result in a locked conformation of Tus-Ter, while in contrast, unwinding can rapidly displace Tus at the P interface [115]. However, *in vivo*, Tus can be displaced by ~50% of replication forks at the NP interface [116, 117]. Elshenawy, M.M., and colleagues have shown that one possibility is that the efficiency of the Tus-Ter bypass is a consequence of the replisome speed: faster replisomes can pass through the Tus-*Ter* block before it adopts the "locked" conformation, while slower ones get trapped [118].

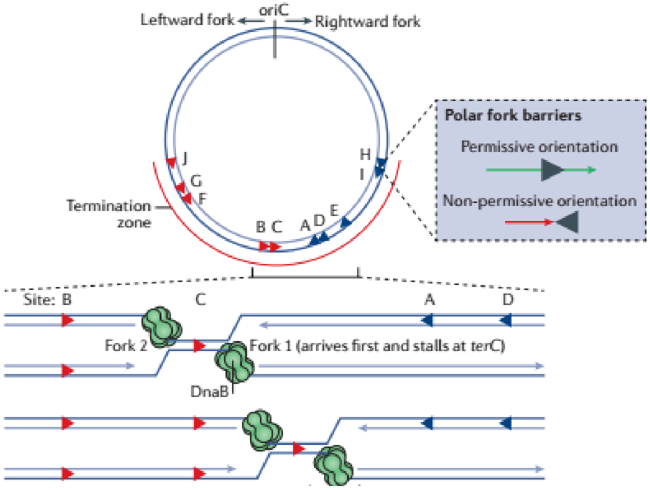


Figure 2.9: Diagram of termination in *E.coli*. Taken with permission from [114].

Interestingly, Δ*tus* does not seem to have abnormal growth phenotypes, and recent evidence suggests that the termination area evolved to limit over-replication, rather than simply specifying termination sites [114, 119].

Saccharomyces cerevisiae

In contrast to *E.coli*, budding yeast appears to not have strict termination sites. Termination sites seem to be dictated by when replisomes from neighbouring origins encounter one another, which would be dictated by origin firing time and individual replisome speed [5]. However, it does appear that there is an active mechanism to cause disassembly of the CMG helicase [114]. The process that drives the disassembly is ubiquitylation, specifically of the MCM7 subunit of the helicase [120] (**Figure 2.10**). The ubiquitylation is done by SCF^{Dia2} (SCF [Skp1/cullin/Fbox protein]), which is tethered to the replisome [121]. Subsequently, Cdc48 segregase is recruited and disassembles the CMG into its individual components, which also leads to removal of the other replisomal proteins bound to CMG [120].

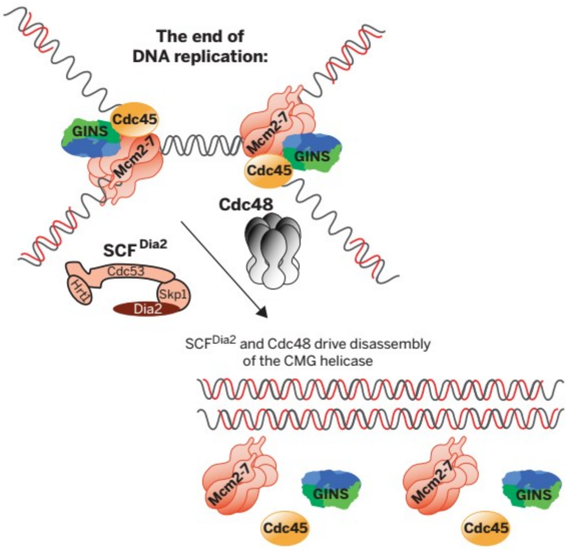


Figure 2.10: Diagram of termination in budding yeast, showing the role of ubiquitination in driving disassembly. Taken with permission from [120].

One unresolved question is how this ubiquitylation-dependent disassembly is targeted to terminating forks and not active ones, although some have suggested the distinct CMG conformations during termination and elongation, may help target the disassembly process at the right stage [114].

2.5 Fluorescence Microscopy

Fluorescence microscopy has been an invaluable method to study the localization, dynamics, and stoichiometries of proteins, especially in living cells. In particular, single molecule approaches have allowed us to obtain distributions of variables rather than simply ensemble averages, giving us a more stochastic and dynamic view of biology, with a stronger emphasis on the probabilistic nature of biological processes.

In vitro studies of *E.coli* DNA replication have provided important information on the factors affecting the processivity and rates of individual replisomes, but an alternative is to study them in live cells. Fortunately, with recent advances in microscopy and genetics, this is now a feasible option. Specifically, single-particle tracking photoactivated localization microscopy (sptPALM), Fluorescence Recovery After Photobleaching (FRAP) and Stepwise Photobleaching (SPB) have been useful methods to probe the dynamics and architecture of proteins in their native cellular context [122-125]. SptPALM particularly has the advantage of being able to directly observe the dynamics of single copies of proteins [123] (**Figure 2.11**).

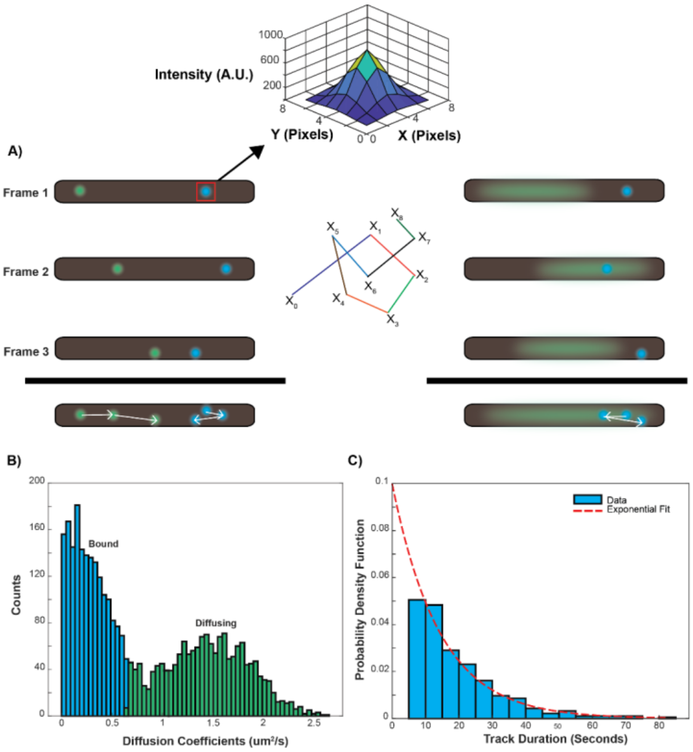


Figure 2.11: A) Illustration of single-molecule tracking experiments. Fluorescent spots are localized in each frame typically through fitting the intensity profile with a Gaussian function. Localizations are linked to together to form tracks. Left) With short camera exposure times (few milliseconds), one can detect both diffusing and bound molecules. Right) With longer camera exposures (hundreds of milliseconds), the diffusing molecules are blurred out, but the bound molecules are still detectable. B) Estimation of diffusion coefficients from data collected with short camera exposures, allows one to calculate the proportion of different diffusive states. C) Estimation of the mean track duration from data collected with long camera exposures, which can be used to estimate the mean bound time, after correcting for photobleaching. Adapted with permission from [4].

SptPALM requires tagging the protein of interest (POI) with either a photoactivatable (PA) or photoconvertible (PC) fluorescent protein (FP). These are proteins that are normally in one fluorophore state, but upon excitation with low 405nm wavelength light, they stochastically convert a small subpopulation of tagged proteins, to another state (fluorescent state for PA proteins and different fluorescent state for PC proteins). PA/PC-FPs have the advantage that they allow single molecules of high copy-number proteins to be tracked [14]. Tracking is done by linking the same molecule frame to frame, and various algorithms are available for this non-trivial process [123, 126-128]. After tracking, one can extract a variety of parameters for the POI such as its diffusion coefficient, proportions of different diffusive states, residence time, search time, cellular localization, etc. [123] (**Figure 2.11**). These are tremendously useful parameters to understand how replisomes behave - in terms of their stability and dynamics – in live cells. Chapters 3 and 5, will illustrate this point in more detail.

Unfortunately, while there has been significant single-molecule work done in bacteria, there has been much less in budding yeast, especially in live cells [129-135]. The challenges for doing single-molecule work in budding yeast are: the size of the yeast cell (4um diameter for haploid cells, compared to 0.7 um for *E.coli*) , autofluorescence, light refraction due to the thick cell wall, and the higher number of replisomes active [30, 132, 136, 137]. To detect single-molecules, one requires high sensitivity and good signal-to-noise ratio (SNR); as such, a large cell volume means more out-of-focus fluorescence coming from tagged proteins and cellular autofluorescence, thereby hindering detection of single-molecules. Fortunately, new advances in microscope configurations have allowed for selective illumination to minimize background fluorescence while retaining high sensitivity. These include total internal reflection fluorescence microscopy (TIRFM), light sheet microscopy (LSM), and highly-inclined and laminated optical sheet (HILO) [138, 139] (**Figure 2.12**).

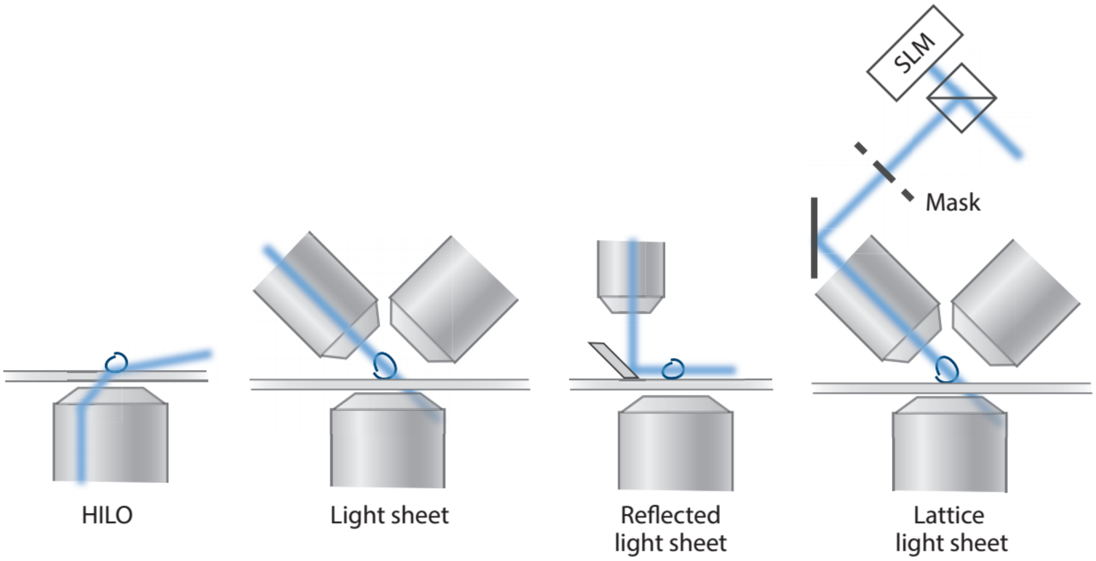


Figure 2.12: Diagram showing different microscopy techniques to achieve selective illumination**.** Taken with permission from [139].

In addition, the development of brighter and more photostable fluorescent dyes/proteins (e.g. HaloTag and SNAPTag) have improved detection of single-molecules, while also allowing for them to be observed for longer periods of time [14, 140, 141]. While incorporating the dye into cells can be difficult – and some approaches may perturb cell physiology – some have found introducing deletions of transporter proteins, thereby preventing export of dye, to be an effective way to increase dye incorporation, without significantly altering cell physiology [134].

In Chapter 5, I will describe the use of HaloTag, HILO, and sptPALM to characterize the binding kinetics of budding yeast replisomal factors.

Once images have been acquired, the next step is the analysis to extract quantitative information. Fortunately, recent advances in machine learning and deep learning has the potential to accelerate this step [142]. The most commonly used approach is supervised learning where a labeled data set (training data) is fed into an algorithm, which then learns a model that can be used to classify new data [142]. This approach has the potential to be a more robust method for classifying data or objects in images than traditional classification methods, and in a more high-throughput manner. Of course, tests must be made to check for overfitting (fitting a too complex model than required) and biases [142]. In Chapter 4, I will describe the use of machine learning for sptPALM experiments and tracking to identify DNA-bound molecules.

Recent advances in genetics have also paved the way to perturb conditions commonly done *in vitro*, directly in the cell. One example is using degron tags to rapidly deplete the POI, in a controlled manner, thereby lowering concentrations of competitor molecules[143]. Likewise, the use of temperature-sensitive mutants of POIs allows one to inactivate the protein rapidly. The advantages of these approaches compared to simply using deletion mutants, is that we can perform them on essential proteins and they do not accumulate suppressor mutations [143].

There are many unresolved questions on how replisomes function *in vivo* but these recent advances in microscopy, computational methods, and genetic methods, could pave the way to answering some of them.

Chapter 3

Frequent Exchange of the DNA Polymerase During Bacterial Chromosome Replication

This chapter is based on a published manuscript: Beattie, T.R.,* Kapadia, N.,* Nicolas, E., Uphoff, S., Wollman, A.J.M., Leake, M.C., and Reyes-Lamothe, R. Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife* (2017). e21763.

* Co-first authors

As highlighted in Chapter 2.3, *in vitro* studies of *E.coli* DNA replication have shown high processivity of the PolIII* complex, indicating that the polymerases act stably in the replisome. However, these studies were typically done in the absence of excess proteins, so we sought to use advanced microscopy methods (FRAP, single-molecule tracking) to study the binding kinetics of *E.coli* replisomal subunits directly in live cells. My main contribution was that I developed methods to analyze single-molecule data to obtain accurate, quantitative information on the kinetics. Surprisingly, we found that the PolIII* subunits exchanged quite rapidly, while DnaB helicase showed stable binding. These results suggest that the dynamics of polymerases may be governed by the surrounding pool of excess copies, and the *E.coli* replisome can undergo frequent changes to its composition over the course of replication.

3.1 Abstract:

The replisome is a multiprotein machine that carries out DNA replication. In *Escherichia coli*, a single pair of replisomes is responsible for duplicating the entire 4.6 Mbp circular chromosome. *In vitro* studies of reconstituted *E. coli* replisomes have attributed this remarkable processivity to the high stability of the replisome once assembled on DNA. By examining replisomes in live *E. coli* with fluorescence microscopy, we found that the Pol III* subassembly frequently disengages from the replisome during DNA synthesis and exchanges with free copies from solution. In contrast, the DnaB helicase associates stably with the replication fork, providing the molecular basis for how the *E. coli* replisome can maintain high processivity and yet possess the flexibility to bypass obstructions in template DNA. Our data challenges the widely-accepted semi-discontinuous model of chromosomal replication, instead supporting a fully discontinuous mechanism in which synthesis of both leading and lagging strands is frequently interrupted.

3.2 Introduction:

DNA replication is carried out by a multifunctional machine, the replisome. The *E. coli* replisome has been characterized *in vitro* and *in vivo* and is composed of more than 12 different proteins [29, 66]. DNA synthesis is performed by the Pol III polymerase (αεθ). Three copies of Pol III are incorporated into the replisome through an interaction with the τ subunit of the pentameric clamp loader complex ($\tau_3\delta\delta'$). Together, these constitute the Pol III* subassembly ($((\alpha\varepsilon\theta)_3\text{-}\tau_3\delta\delta')$). The clamp loader is also responsible for loading the β clamp dimer onto DNA, which is required for processive synthesis by Pol III. Addition of β clamp to Pol III* forms the Pol III holoenzyme. At the core of the *E. coli* replisome is the replicative helicase, DnaB, which encircles the lagging strand template and unwinds parental DNA. The Pol III holoenzyme associates with DnaB through the τ subunit of the clamp loader (**Figure 3.1A**). In addition, the DnaB helicase recruits the primase, DnaG, which synthesizes RNA primers. Due to the antiparallel nature of DNA, synthesis of one the strands – the leading strand – occurs co-directionally with progression of the replication fork, while the second strand – the lagging strand – is synthesized by repeated cycles of primer synthesis and DNA extension.

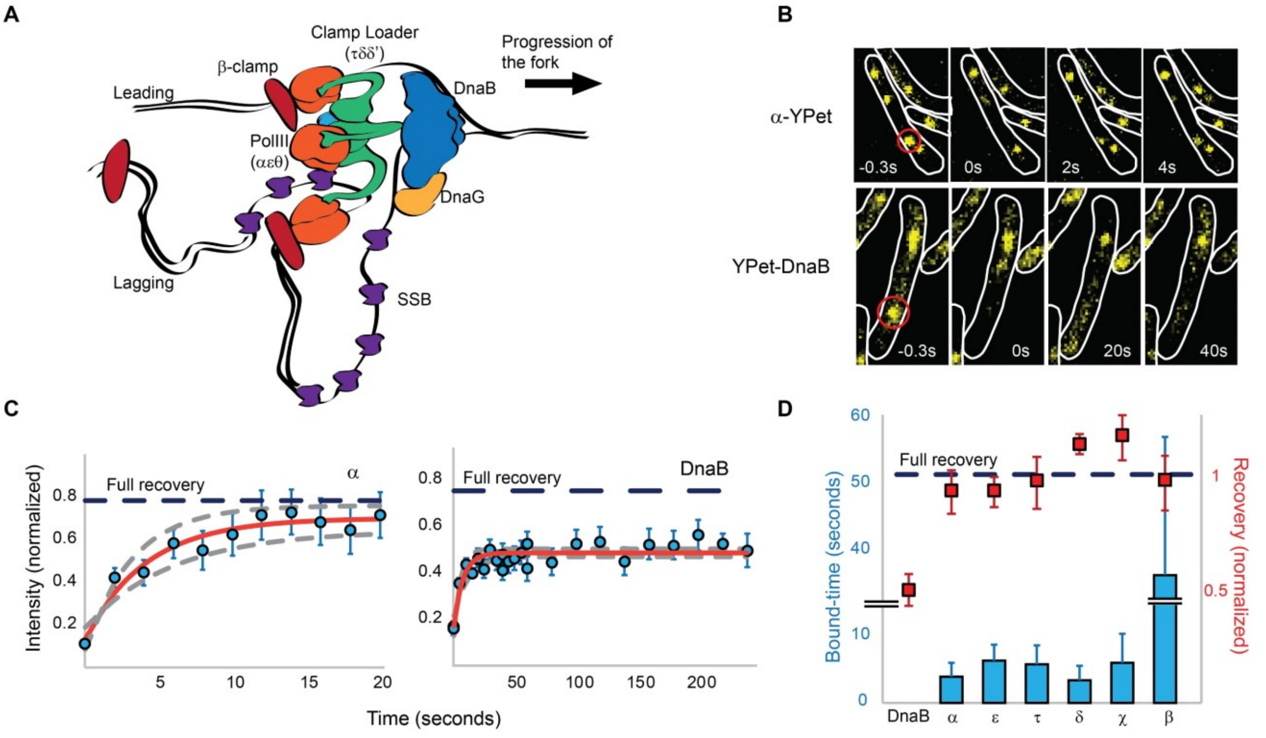


Figure 3.1: Most replisome subunits exchange frequently with the diffusing pool

(A) Model illustrating the architecture of a replisome at the *E. coli* replication fork.

(B) Representative fluorescence images of FRAP experiments for the Pol III α subunit and the DnaB helicase. Cell boundaries shown as white lines, red circle shows the location of the bleached focus.

(C) Representative examples of the FRAP curves for Pol III α subunit (N=48) and DnaB (N=96). Red line shows a reaction-diffusion model fit to the data, dashed grey lines show SE for the model. Dashed blue line represents the estimated maximum possible fluorescence recovery after correcting for photobleaching.

(D) Analysis summary of the replisome by FRAP. Bars represent average bound-times. Red squares represent level of recovery normalised to the intensity before bleaching. Dashed blue line represents maximum possible fluorescence recovery. It was not possible to estimate the bound-time for DnaB. Error bars represent SE.

Replication of the circular chromosome of *E. coli* proceeds bidirectionally from a single, defined locus: *oriC*. Multiple mechanisms tightly restrict DnaB loading, and therefore replisome assembly, to the *oriC* locus during initiation, with a single initiation event per cell cycle [20]. Two sister replisomes are assembled at *oriC* during initiation, and each is responsible for replicating half of the ~4.6Mbp chromosome. At 37°C, it takes 40-60 minutes of continuous DNA synthesis to complete chromosomal replication, at rates of 600-1000bp sec^{-1}.

Given the extent of DNA synthesis required, it has been assumed that the replisome is a stable protein complex capable of replicating large fragments of the chromosome without disassembling. This is supported by *in vitro* data showing that a single purified replisome, once assembled on DNA, is capable of synthesizing DNA with an average length of 70kbp without requiring replacement of the Pol III* subassembly or DnaB [27, 51]. Even greater stability has been inferred from *in vivo* experiments that suggest infrequent replication fork collapse during chromosome replication in *E. coli* [144].

Chromosomal DNA presents multiple potential obstacles to replisome progression. DNA lesions can result in the stalling of the replisome due to Pol III's inability to use damaged DNA as a template [145]. In addition, the replisome frequently encounters DNA-bound proteins, potentially resulting in disassembly or pausing [41, 146]. Multiple mechanisms have been proposed that allow replisome integrity to be maintained during bypass of such obstacles [147-149] and that remove bound proteins from DNA [41]. In cases where these strategies are insufficient, the cell also has mechanisms to mediate the reassembly of the replisome at specific DNA structures that arise following replisome collapse [150]. The frequency at which replisomes encounter these obstacles and the efficiency of the bypass mechanisms are still unclear. It is also uncertain in which way the architecture and stability of replisome play a role during these events.

The replisome is likely to be affected by multiple factors present inside cells which have not been accounted for in reconstituted systems. However, a direct measurement of the stability of the replisome has not been undertaken *in vivo*. Here we measure the binding kinetics of replisome subunits during DNA replication using two independent fluorescence-based methods in living cells. Our results show that the entire Pol III* subassembly is replaced within the replisome at a frequency equivalent to a few cycles of Okazaki fragment synthesis. This leads us to conclude that DNA replication is a discontinuous process on both strands. We also find that the DnaB helicase remains bound to DNA for tens of minutes, preventing disassembly of the replisome likely by serving as a dock during Pol III* subassembly turnover. We propose that this dynamic stability provides the replisome with flexibility to bypass frequent obstacles on DNA while maintaining the necessary processivity for chromosomal replication.

3.3 Results:

3.3.1 Fluorescence Recovery After Photobleaching reveals frequent exchange of subunits in active replisomes

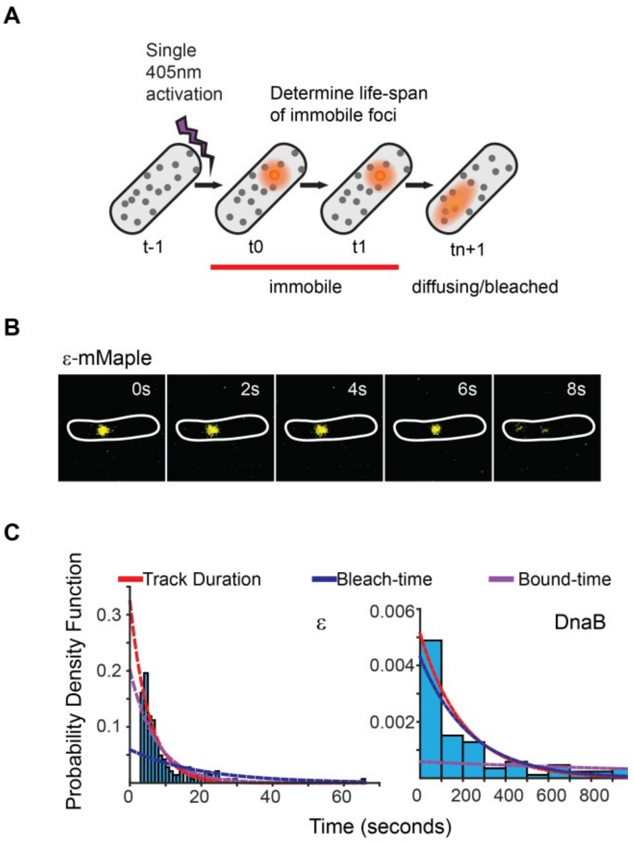
To assess the stability of the *E. coli* replisome when replicating chromosomal DNA *in vivo*, we first measured the binding kinetics of replisome subunits using fluorescence recovery after photobleaching (FRAP) in strains possessing fluorescent YPet derivatives of key replisome components [29, 122]. Using actively replicating cells in growth conditions that permit a single replication event per cell cycle, we bleached individual foci of fluorescent replisome subunits using a focused laser pulse and measured their recovery over time (**Figure 3.1B**). The dimensions of *E. coli* – in our conditions typically ~0.7μm diameter and few microns in length – and the low number of replisome subunit molecules per cell – a few hundred for most subunits [122] – increased the difficulty of selectively bleaching replisome foci without affecting the remaining fluorescent pool. To minimize photobleaching of the diffusing pool of fluorescent proteins we increased cell volume by treatment with cephalexin; this did not affect DNA replication (**Figure A1.1**).

To our surprise, we found that the initial focus fluorescence recovered in a few seconds for Pol III and clamp loader components (**Figure 3.1B**). Fluorescence recovery is not explained by the photophysical properties of YPet, like photoblinking, and instead it represents protein exchange (**Figure A1.2**). We used a reaction-diffusion model in a reaction-limited regime to fit the average fluorescence recovery curve of individual subunits, and calculated a time constant for binding (bound-time) which represents the average time that a molecule is bound to the replisome before exchanging. The bound-time was 4 ± 2 and 6 ± 2 seconds (mean ± SE) for the α and ε subunits of Pol III, respectively. Similarly, the τ, δ and χ subunits of the clamp loader had bound times of 6 ± 3, 3 ± 2 and 6 ± 4 seconds, respectively (**Figure 3.1C-D**). Molecules of the β clamp exchanged at a slower rate, remaining associated for an average of 36 ± 21 seconds, consistent with its binding to newly-synthesized DNA behind the replisome [151, 152]. The timescale of Pol III holoenzyme exchange is in striking contrast to the ~150 minutes required for two replisomes to complete chromosomal replication under our microscopy conditions (**Figure A1.3)**

3.3.2 sptPALM demonstrates fast turnover of the Pol III* subassembly

To confirm these results, we used single-particle tracking Photoactivated Localization Microscopy (sptPALM) [153] to determine the bound-times of replisome subunits [154]. We constructed *E. coli* strains with functional fusions of replisome subunits and the photoconvertible fluorescent protein, mMaple [155] (**Figure A1.4**). We used a single low intensity pulse of 405nm-laser activation per experiment to switch, on average, a single molecule per cell into a red fluorescence state. Long (500ms) camera exposure times – to motion-blur fast diffusing molecules – were used, spaced by 1s or 5s intervals, to track non-diffusing replisome-associated molecules as foci (**Figure 3.2A**). This illumination protocol did not perturb cell growth (**Figure A1.5**). Track duration distributions for labelled replisome subunits were calculated from the number of frames individual molecules appeared as foci (**Figure 3.2B-C**). Bound times were calculated by correcting for the disappearance of foci due to photobleaching, which was characterized using a strain carrying the transcriptional repressor LacI fused to mMaple and a chromosomal array of *lacO* binding sites. We also assessed the effect photoblinking using this same strain (**Figure A1.6 A-C**). We expect that in the timescale of our experiments, the lifetime of LacI-mMaple foci will be dictated by photobleaching, with dissociation from DNA being negligible [156].

The single-molecule results are consistent with our FRAP data. Pol III subunit and clamp loader components indeed exchanged rapidly, with ε,τ, and δ remaining replisome-associated for only 10 ± 0.7, 10 ± 0.7 and 12 ± 0.9 seconds (mean ± SE), respectively (**Figure 3.2B and 3.2D**). We found no strong evidence for multiple binding behaviors of individual subunits (**Figure A1.6D and Table A1.4),** suggesting that both leading and lagging polymerases behave similarly. As with FRAP, we observed similar bound-times for all subunits of both the DNA polymerase III and clamp loader complexes despite a difference in stoichiometry – δ, τ and ε are present in 1, 3 and 3 copies per replisome, respectively (**Figure 3.1A**). As such, exchange of individual subunits independently from one another, although still possible, does not easily explain our results. We therefore propose that the unit of exchange of Pol III and clamp loader subunits is the Pol III* subassembly ($(\alpha\varepsilon\theta)_3$-$\tau_3\delta\delta'$). This idea is supported by *in vitro* data that shows that exchange of Pol III at the replication fork requires it to be part of the Pol III* subassembly [157]. Cells have an excess of free Pol III that does not interact with the clamp loader [158]. Consistent with the notion that only Pol III subunits found within a Pol III* subassembly are competent for exchange, we observed re-binding of single molecules of ε, often to a different replisome, at a much higher frequency than would be predicted if all of the ~270 molecules of ε in the cell were in direct competition [122] (**Figure 3.3A** and **Figure A1.8**).



	Bound time (seconds)
ε	10.4 ± 0.8
τ	10.22 ± 0.7
δ	12.1 ± 0.9
β	47.2 ± 2.9
DnaB	913 ± 508

Figure 3.2: Exchange of Pol III* subassembly and DnaB occur on different timescales

(A) Diagram illustrating the sptPALM experimental design used to measure bound-times.

(B) Representative example of the focus life span for the Pol III ε subunit.

(C) Representative examples of the distribution of fluorescent foci life-spans (blue bars) for Pol III ε subunit and DnaB, showing fitting of a single-exponential decay model (red line), the estimated bleaching rate in the same conditions (blue line) and the corrected estimated bound-time (purple line). Note that to improve accuracy in single-molecule detection tracks shorter than four localizations were removed in the case of ε but corrected during curve fitting, hence the lower bar near 0 seconds time point. ε data was collected using 500ms exposure time and 1 second intervals (N=143), DnaB data was collected using 2 second exposure time and 10 second intervals (N=86). The plot for DnaB shows binned data for presentation purposes. (D) Summary of estimated average bound-times. Errors in the table represent SE.

The β clamp again showed a longer bound-time of 47 ± 3 seconds in sptPALM. Our estimate is broadly consistent with a previous estimate from *E. coli*, although ~4 times shorter [151]. Assuming similar numbers of DNA-bound β as in that earlier report, we estimate from our results that a new β dimer is loaded at each replication fork every ~2 seconds. Considering an average rate of fork progression of ~260bp s^{-1} – calculated from the duration of replication in the conditions used (**Figure A1.2**) – we find an average Okazaki fragment length of 520bp, in close agreement with an *in vitro* measurement of 650bp at room temperature [27]. In contrast, we estimate from FRAP and sptPALM an average replication fork progression of 1-3kbp prior to Pol III exchange, which would allow completion of multiple Okazaki cycles. We therefore think it unlikely that subunit exchange within the replisome is exclusively linked to the dynamics on the lagging strand.

3.3.3 DnaB may act as a stable platform upon which the Pol III* subassembly exchanges

The DnaB helicase displayed very different dynamics to other replisome subunits when assessed by FRAP. Crucially, we never observed full fluorescence recovery of DnaB over 5 minutes of measurement (**Figure 3.1B-D**), indicating that it is stably associated with the replisome on this timescale. Our analysis showed an initial recovery of fluorescence with a 7 ± 4 second time constant, which we attribute to the signal from diffusing molecules moving into the bleached area. However, we did not observe a significant increase in the intensity after this initial time point, preventing us from accurately estimating a bound-time for DnaB (**Figure 3.1C**). We conclude that in contrast to Pol III holoenzyme subunits, replisome-associated DnaB does not exchange frequently, and is instead a stable component of the replisome.

Analysis of DnaB by sptPALM confirmed that it is the most stable subunit in the replisome. To eliminate incorrect assignment of DnaB fluorescence by our software we used even longer (2s) exposure times, thus further blurring slow-diffusing molecules (**Figure A1.7**). Control experiments with ε under the same conditions had no significant effect on calculated bound-times (**Figure A1.6E**). Crucially, we estimate that single molecules of DnaB remained bound to the replisome for 913 ± 508 seconds, significantly longer than any other component (**Figure 3.2B-D**). The width on the distribution for this estimate is inherently large due to a close similarity between DnaB foci lifetimes and the bleaching time of mMaple (**Figure 3.2C**). However, our long bound-time estimate is supported by frequent examples of DnaB fluorescent foci that last for tens of minutes (**Figure 3.3B**). Currently we cannot assess the extent at which the turnover detected represents PriABC mediated re-loading of helicase [150]. Altogether, our data supports a role for DnaB as the primary determinant of replisome integrity.

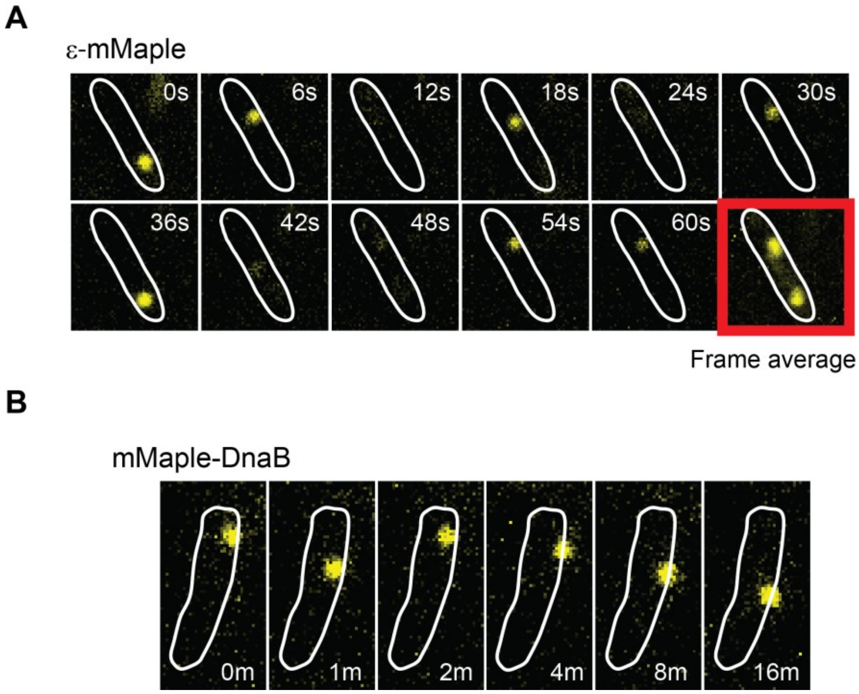


Figure 3.3: DnaB may serve as a platform for frequent Pol III* subassembly re-binding

(A) Representative example of a cell where a single activate copy of ε-mMaple shows multiple cycles of binding and unbinding (time in seconds). Obtained from an experiment using 2 second intervals between consecutive images. Frame average shows the cellular location of two replisomes.

(B) Example of a cell where a single activated mMaple-DnaB molecule remains localized as a focus for several minutes (time in minutes). Obtained from an experiment using 10 second intervals between consecutive images. Boundaries of the cells at the beginning of the experiment are shown as white outlines.

3.3.4 Active synthesis is only partially responsible for turnover

To determine if subunit exchange occurs as a consequence of active DNA synthesis, we measured bound-times by FRAP and sptPALM in cells treated with the DNA polymerase inhibitor hydroxyurea (HU) [159]. Components of the Pol III* subassembly and the β-clamp showed bound-times two- to six fold higher than in untreated cells (**Figure 3.4A-B**). HU had no apparent effect on DnaB FRAP estimates (**Figure 3.4A**). We conclude that the exchange of replisome components is at least partially dependent on active DNA synthesis. Remaining turnover may reflect residual DNA synthesis after HU treatment. In addition, it may indicate that the replisome is intrinsically dynamic as a multiprotein complex, with DNA synthesis further increasing subunit exchange.

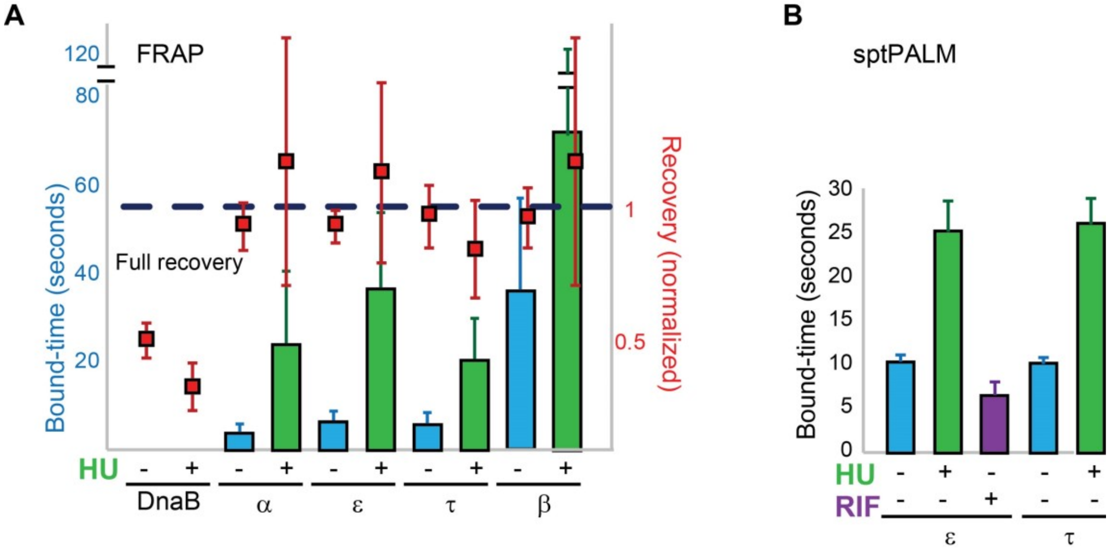


Figure 3.4: Replisome dynamics are partly dependent on active DNA synthesis

(A) Summary of the average bound-times in cells treated with HU, estimated by FRAP. The results for untreated (blue bars) and treated cells (green bars) is shown. Data for the untreated condition is presented to facilitate comparison and is identical to that in Figure 1D. Red squares represent the normalised level of fluorescence recovery. Dashed blue line shows estimated maximum possible recovery. It was not possible to estimate the bound-time for DnaB.

(B) Summary of bound-times estimated by sptPALM (weighted average). Results for the untreated (blue bars) and treated cells (green bars) is shown. Data obtained using cells treated with the RNA Polymerase inhibitor Rifampicin is also shown for cells carrying ε-mMaple (purple bars). Data for the untreated condition is presented to facilitate comparison and is identical to that in Figure 2D. Error bars represent SE.

3.4 Discussion:

3.4.1 A dynamic replisome may help to minimize delays in replication fork progression in the presence of roadblocks

Obstructions in template DNA, particularly protein-DNA complexes, have been shown to cause frequent pausing of the *E. coli* replisome *in vivo* [41]. *In vitro* studies have demonstrated that the *E. coli* replisome is capable of bypassing such obstacles by interrupting leading strand synthesis and resuming extension downstream of the obstruction from a leading strand primer deposited by DnaG or an mRNA synthesized by RNA polymerase [147-149]. Our data is entirely consistent with this mechanism, whereby bypass could be achieved through detachment of the stalled Pol III* subassembly from DnaB and its replacement downstream of the obstacle with another Pol III* from solution. Because DnaB translocates on the lagging strand, small lesions and large protein blockages on the leading strand can both be bypassed by Pol III* exchange. Note however that we did not observe any apparent effect on the dynamics of Pol III after inhibiting transcription, suggesting that this process is not the main cause of exchange (**Figure 3.4B**). On the lagging strand, small template lesions capable of passing through the central pore of DnaB may also be bypassed through Pol III* exchange. In contrast, obstacles on the lagging strand that destabilize DnaB, such as proteins stably bound to DNA or strand discontinuities, would likely result in the disassembly of the replisome. We propose that obstacle bypass along template DNA may be the primary selection pressure that has driven the evolution of a dynamic replisome.

In addition to the model above, we acknowledge that other processes may also exert selective pressure for the generation of the observed replisome binding kinetics. First, unbinding of Pol III* subassembly may result from build-up of helical torsion in the template DNA generated by the coupled synthesis of both DNA strands [66]. This is consistent with longer binding times when synthesis was inhibited by HU. Unbinding of a single polymerase from DNA or release of the whole Pol III* subassembly would have the same effect on stress relief. Second, the dynamics observed may be a byproduct of the highly regulated interaction between Pol III and β-clamp. Even though Pol III tightly binds the β-clamp to ensure highly processive synthesis, these two proteins rapidly unbind from each other upon completion of the duplex DNA. The strength of this protein-protein interaction is modulated by the OB domain in the α subunit of Pol III, which binds to ssDNA ahead of the catalytic domain, and the C-terminus of the τ subunit of the clamp loader [62, 160]. Premature activation of such a switch in both leading and lagging strand polymerases would weaken the grip of the Pol III* subassembly on the replication fork and potentially result in its displacement. This idea is consistent with the presence of ssDNA gaps in the lagging strand [161], which may be explained by a premature loss of Pol III processivity.

Presumably, exchange of Pol III* subassembly within the replisome occurs rapidly enough to minimize potentially deleterious ssDNA gaps between fragments of nascent DNA on the leading strand. However, the rate of DNA unwinding by DnaB decreases by more than 10-fold when DnaB is detached from the τ subunit of the clamp loader [162], providing a potential safety mechanism to limit DNA unwinding and exposure of ssDNA during Pol III* subassembly exchange. It remains to be determined if a newly associated copy of Pol III uses the existing 3' end at the leading strand or require the activity of primase to resume synthesis.

3.4.2 Protein excess in cells is a key factor in the regulation of replisome subunit turnover

Our data apparently contradict *in vitro* studies which have demonstrated that a single reconstituted *E. coli* replisome can operate without subunit exchange in synthesizing an average of ~80kbp [27, 51]. Measurements in those reports were performed by removing all diffusing Pol III* subassembly and DnaB subunits from the reaction. In contrast, in the cell there is a permanent excess of diffusing replisome subunits. We believe this explains the differences observed with our *in vivo* data. Competition for binding sites between DNA-bound and diffusing molecules has been shown to change the DNA-binding kinetics of proteins such as Fis, HU and NHP6A [163], EcoRI [164], RPA [165] and the transcription factor CueR [166]. Furthermore, mathematical modelling has shown that it is theoretically possible for a replisome to be stable under conditions in which no extra subunits are present, as *in vitro*, and yet undergo frequent subunit exchange in the presence of extra subunits, as *in vivo*, due to subunit competition [167]. We think that active synthesis may enhance exchange with the diffusing pool, consistent with our results using HU.

Frequent exchange of DNA polymerases in the presence of extra subunits has been observed in the replisomes of bacteriophages T4 and T7 *in vitro* [168, 169]. In T7, this occurs through a mechanism in which extra DNA polymerases associate with the bacteriophage DNA helicase and exchange with the active DNA polymerase through competition for DNA binding [170, 171]. It will be interesting in the future to determine the mechanisms that exist in *E. coli* to ensure efficient capture and exchange of the low-abundance Pol III* subassembly.

3.4.3 DNA synthesis of both leading and lagging strands is discontinuous in *E. coli*

One predicted consequence of the Pol III* subassembly exchanging as a single unit is that synthesis of both leading and lagging strands will be frequently interrupted, resulting in discontinuities on both strands. This contrasts the widely-accepted semi-discontinuous model of DNA replication. However, while this model is strongly supported by *in vitro* experiments [63], the mechanism that operates *in vivo* has long been unclear [172, 173]. Okazaki and colleagues' original characterization of replication intermediates demonstrated that all DNA is initially synthesized as short fragments, supporting fully discontinuous DNA replication [174]. More recent *in vivo* experiments performed in the absence of DNA ligase support the idea that discontinuities are produced on both leading and lagging strands during DNA replication in *E. coli* [175, 176]. Our data provide a mechanistic explanation for these observations, and supports a discontinuous model of DNA synthesis in *E. coli*.

3.5 Materials and Methods:

3.5.1 Strains and growth conditions

All strains used are derivatives of AB1157. Cells were routinely grown in LB or in M9 minimal media. M9 was supplemented with glycerol (final concentration 0.2%); 100µg/ml of amino acids threonine, leucine, proline, histidine and arginine; and thiamine (0.5µg/ml). When required, antibiotics were added at the following concentrations: ampicillin (100µg/ml), kanamycin (30µg/ml), chloramphenicol (25µg/ml), cephalexin (40µg/ml), rifampicin (Rif) (300µg/ml) and hydroxyurea (HU) (60-100 mM). For microscopy, cells were spotted on a 1% agarose pad in M9-Glycerol. DAPI was used at a working concentration of 300nM as recommended by manufacturer. Ethanol fixation was done using 70% ethanol in water, followed by two washes with PBS. For TetR-YPet strain, fixation was done using 4% formaldehyde and incubating 15 minutes at room temperature, 15 minutes on ice, followed by 2 washes with PBS.

Chromosomal replacement of replisome genes by fluorescent derivatives was done by lambda red [29, 177]. In short, we used plasmids carrying a copy of *ypet*[29, 178] or *mMaple*[155] followed or preceded by a kanamycin resistance cassette flanked by *frt* sites as PCR templates. Flexible peptides with sequences SAGSAAGSGEF (YPet C-ter fusions), SAGSAAGSGAV (mMaple C-ter fusions) or SAGSAAGSGSA (YPet and mMaple N-ter fusions) were used as a linker between the fluorescent protein (FP) and the protein targeted. Primers carrying 40-50nt tails with identical sequence to the chromosomal locus for insertion were used to amplify the *linker-FP-kan*^{R} (or *kan*^{R}*-FP-linker* in the case of N-terminal fusions) from template plasmids. The resulting PCR product was transformed by electroporation into a strain carrying the lambda red-expressing plasmid pKD46. Colonies were selected by kanamycin resistance and ampicillin sensitivity, screened by PCR using primers annealing to regions flanking the insertion, and sequenced. In the case of N-terminal fusions, in order to minimize the effect of the insertion on the expression levels of the gene we removed the kanamycin cassette by expressing the Flp recombinase from plasmid pCP20[177]. Gene fusions did not have any apparent detrimental effect on cell growth (**Figure A1.4**).

LacI-mMaple was generated through lambda red using the strain MG1655. The gene fusion was then transduced, using P1 phage, into an AB1157 derivative carrying a 256-*lacO* array replacing the *pheA* gene (chromosomal position 2735kb)[179]. Similarly, a TetR-YPet fusion expressed from a lac promoter [151] was transduced into a strain carrying a 256-*tetO* operator array at R3 (chromosomal position 852kb) [179].

3.5.2 Fluorescence Recovery after Photobleaching (FRAP)

Cells were grown in M9-Glycerol at 30°C, treated with cephalexin for 2h, harvested at early log-phase (OD_{600} 0.1-0.2), concentrated and spotted onto a pad of 1% agarose in M9-Glycerol, contained in a gene frame (Thermo Scientific™). Treatment with hydroxyurea was done on the agarose pad by mixing HU with media and agarose. Cells were incubated on the slides for 10 minutes before imaging.

Most FRAP experiments, except for the TetR-YPet control, were performed using a spinning disk imaging system (PerkinElmer) with a 100x NA1.35 oil objective and an ImagEM EMCCD camera (Hamamatsu Photonics). Images were acquired using Volocity imaging software. An image was acquired in the brightfield channel at the beginning of the experiment to serve as a reference. FRAP was performed by pulse-bleaching using a 488nm laser for 10-15 ms and 30-50% laser intensity (radius of the spot was diffraction limited at ~300nm). Two pre-bleach images were captured, the bleach spot was centered on one replisome focus and recovery of the bleached region was recorded at different intervals after bleaching. Image capture was done at a 300ms frame rate (4-6% 515nm laser) for most replisome components except for DnaB helicase, for which 500ms capture rate was used (2% 515nm laser). For α, ε, τ, δ and χ, we used intervals between pictures of 2s, 5s and 10s. For DnaB and β, we used intervals between pictures of 5s, 10s and 20s. Experiments were done at room temperature.

FRAP to control for photoblinking was done using an epifluorescence system, Leica DMi8, with a 100x oil objective (Leica 100x/NA 1.47 HL PL APO) and an iXon Ultra 897 EMCCD camera (Andor). FRAP was performed using an iLas2 unit (Roper Scientific) using an ILE laser combiner (Andor) and a 150mW 488nm laser. Both bleaching and excitation of YPet were done using the 488nm laser. Acquisition was done using 100ms exposure at 5 second intervals.

3.5.3 FRAP Analysis

Initial position of spots was manually selected using the coordinates for localized bleaching in the image recorded by the acquisition software. Tracking was then done automatically using a previously developed custom program in MATLAB (Mathworks), ADEMS code[180] (freely available at https://sourceforge.net/projects/york-biophysics/). Most experiments analyzed had a pixel size of 100nm, for which we used a search window with a radius of 5 pixels and an initial guess for the PSF of 3 pixels when fitting candidate spots. For a minority of the experiments, with pixel size of 140nm, analysis was done using 4-pixel search window and a 2-pixel radius for initial fitting.

Intensity traces were filtered to retain only those where clear bleaching was observed. We removed any trace where the intensity at any of the pre-bleach time points was below the value of the ROI immediately after bleaching (0-second time point). In addition, the intensity at the 0-second time point had to be below 40% of the mean pre-bleach intensity. FRAP data were then normalized by the average intensity of the pre-bleached data points.

To estimate the maximum possible fluorescence recovery (Max recovery), we used the corresponding brightfield image to draw a polygon in ImageJ[181] around cells containing a bleached spot. We used these ROIs to obtain the intensities across the experiment in the fluorescence channel. Max recovery was calculated by dividing the intensity of the cell at 0-second time point by the average intensity before bleaching. An average Max recovery value was obtained from all the bleached cells in the experiment.

To correct for photobleaching during the experiment, a different set of spots was manually selected in cells not exposed to localized bleaching, so they could serve as a baseline control. An average bleaching curve was produced using the intensity traces from these fluorescent foci. All data used to generate the bleaching curve were obtained in the same day using the same strain, excitation settings and interval between pictures as for the FRAP experiment. The average curve was fitted to an exponential decay function. FRAP intensity traces were corrected by dividing each time point by the corresponding normalized value in the fitted bleaching curve.

Data from the same set of experiments were averaged. Data from experiments performed the same day, but having different intervals between pictures, were collated into a single recovery curve. Data were then fitted by an exponential solution of the reaction-diffusion equation in a reaction-limited regime using MATLAB:

1)

$$y = c - ae^{-bt}$$

where *c* is the asymptote for recovery, *a* the amplitude of recovery, and *b* the rate of unbinding (i.e. k_{off}). Bound-times are the reciprocal of k_{off}.

Upper boundary for *c* during fitting was set to the Max recovery (see above), plus ten percent of this value to account for measurement error. In addition to R squared, which is not recommended for non-linear models, goodness of fit was assessed using the Kolmogorov-Smirnov test by measuring the normality in the distribution of residuals[182]. Standard errors and 95% confidence intervals (CI) on the parameter estimates were calculated using the variable values previously obtained, as initial estimates, and bootstrap sampling was performed over 10,000 samples (**Table A1.2**). The values reported in the figures are weighted averages of all the experiments done for the same subunit.

We expect that co-localization of sister replisome will have no effect on the rates calculated since the intensity of every spot is normalized against itself in FRAP, and the average rate of recovery is the same at every replisome. Similarly, in sptPALM binding time of individual molecules should not be influenced by a nearby replisome, resulting only a minimal increase in the probability of re-binding to the same place.

3.5.4 sptPALM

Cells were harvested from early log-phase cultures in M9-Glycerol (OD_{600} 0.1-0.2), concentrated and spotted onto a pad of 1% agarose in M9-Glycerol, contained in a gene frame. Coverslips cleaned with versa-clean, acetone and methanol were used to minimize fluorescent background. Treatment with hydroxyurea was done on the agarose pad, by mixing HU with media and agarose.

Imaging was performed at room temperature on an inverted Olympus IX83 microscope using a 60x oil objective lens (Olympus Plan Apo 60X NA 1.42 oil) or 100x oil objective lens (Olympus Plan Apo 100X NA 1.40 oil). Images were captured using a Hamamatsu Orca-Flash 4.0 sCMOS camera. Excitation was done from an iChrome Multi-Laser Engine from Toptica Photonics. Laser triggering was done through a real-time controller U-RTCE (Olympus). Experiments were done using HiLo illumination setup[183] from a single-line cell^TIRF illuminator (Olympus). Olympus CellSens 2.1 imaging software was used to control the microscope and lasers.

For experiments with replisome subunits fused to mMaple, a single 405nm wavelength activation event, typically lasting less than 20ms, was followed by multiple 561nm wavelength excitation events with camera captures of 500ms spaced by 1s or 5s intervals, or camera captures of 2s with continuous excitation (2s rates) or 10s intervals. Low levels of exposure to violet-blue light were used to minimize photoxicity and allow cells to continue growing during the experiments (**Figure A1.3**). To image LacI, we used continuous illumination of 561nm wavelength after a single 405nm wavelength activation event at capture rates and intervals of 500ms or 2s. We also used 2s capture with 10s intervals to characterize LacI bleaching in long experiments. Rifampicin experiments were done in a similar manner except Rifampicin was added to the M9-Gly agarose pad, and imaging was done after a 20-minute incubation on the agarose pad. We noticed that fewer spots were detected, consistent with inhibition of replication initiation through Rifampicin.

3.5.5 sptPALM Analysis

Images were first segmented in order to remove out-of-cell noise coming from contaminants on the coverslip. Binary masks were created using ImageJ, either from the differential interference contrast (DIC) channel or the green fluorescent channel of mMaple. For DIC, alignment was done by first obtaining a maximum-intensity projection of the PALM timelapse, and subsequently aligning it to the reference DIC, using ImageJ. Each slice of the PALM timelapse was then multiplied by the binary mask, to retain intensities within cells only. An average value of the out-of-cell background was added to regions outside of the ROIs to minimize incorrect assignment by the detection program due to sharp intensity increases.

PALM tracking was performed using previously developed software[154], based on the DAOSTORM[184] localization algorithm. An intensity threshold was used to find candidate molecules. The positions of the candidate molecules were then used as initial guesses for a 2D-elliptical Gaussian fit. The fitted parameters were: x-position, y-position, x-standard deviation, y-standard deviation, intensity, brightness, elliptical rotation angle, and background. Tracking was done based on a widely used algorithm[126]. Localizations were linked if they appeared within a 300nm radius between consecutive frames, using a memory parameter of one frame to account for blinking or missed localizations (i.e. the molecule can go missing for one frame and still be linked).

Further refinement of the recorded tracks was done to analyze only those that represented immobile single-molecules. To remove slow-diffusing molecules, we plotted a histogram of the PSFs in x and y for all localizations, and performed a two-component Gaussian mixture fit using Maximum Likelihood Estimation. The component with the smaller mean PSF likely represents bound molecules, whereas the other component represents unbound molecules. The two-component Gaussian mixture model has the following form:

2)

$$p\left(\frac{1}{\sigma_1\sqrt{2\pi}}\right)e^{\frac{-(x-\mu_1)^2}{2(\sigma_1)^2}} + (1-p)\left(\frac{1}{\sigma_2\sqrt{2\pi}}\right)e^{\frac{-(x-\mu_2)^2}{2(\sigma_2)^2}}$$

Where p is the mixture probability, σ_1 and μ_1 are the standard deviation and mean of normal distribution 1, respectively. Likewise, σ_2 and μ_2 are the standard deviation and mean of normal distribution 2, respectively. From the fit, we identified the mean and standard deviation of the component representing bound molecules. We then took 2 standard deviations above the mean to obtain an initial estimation of the threshold. We assessed the accuracy of tracking by manually comparing the tracking results for a subset of fluorescent spots with their lifetime in the original images. Using this method we determined that a threshold of x $\leq$170 and y $\leq$215, placed on the mean PSF over the track, helped to eliminate most of the unbound molecules from subsequent analysis.

In addition, we varied the threshold on the number of localizations for track acceptance across different time-intervals of capture. Our reasoning was that the probability that a track represents a genuinely bound molecule becomes higher as the time interval used increases[124]. Therefore, the thresholds for removing tracks were <4, <3, <2, and <2 localizations for interval times of 1s, 2s, 5s, and 10s, respectively. The thresholds were selected by comparing the raw image by eye to the tracks found by the tracking software. Technically no tracks were removed for 5s and 10s since tracks with 1 localization cannot be used to calculate track durations.

To quantify only single-molecule tracks, we plotted a histogram of the mean intensity of a track and fit using a Gaussian Mixture Model (GMM), utilizing the Expectation-Maximization (EM) algorithm. The intensity values were clustered based on membership probabilities (i.e. the probability of belonging to a particular Gaussian component). We used a 2 component GMM fit for most cases and isolated the cluster having the lowest mean, as the intensity values from this cluster likely represent single molecules. We used a 3 component GMM fit in some cases where a significant portion of the molecules seemed out of focus, resulting in a sharp spike of low intensity values in the histogram. In such cases, we isolated the cluster with the second lowest mean. This was especially important when studying proteins with long bound-times, where track fragmentation has a greater relative effect in underestimating the real track duration. We also performed a Bayesian Information Criterion (BIC) test to confirm that the 3 component GMM fit better than the 2 component model. We used only track durations with single molecule intensity values for subsequent analysis.

To avoid track fragmentation in the analysis of proteins with long bound-times, as in LacI and DnaB, caused by fluctuations in intensity or the molecule moving transiently out-of-focus, we determined the typical length of time that the localization software misses spot detection during long tracks (gap time). We did this by manually comparing the outcome of the analysis to the lifetime of a subset of spots in the original images. We found that on average, the gap was ~4 frames. Therefore, we linked tracks based on the criteria that their mean positions were ≤300nm apart and gap time between them was <=4 frames. For these data sets, we performed the GMM fit for isolation of single-molecule tracks after track linkage.

The track durations of multiple samples taken on the same day and time-interval were amalgamated into one data set. In order to get the average track duration, we fitted the track durations using MLE. The reason for our choice of MLE over the more commonly used Least Squares-Estimation (LSE) method is that it is invariant to the bin size (i.e. the parameter estimate is the same regardless of how we bin the data) and it allows us to infer what the population parameter is. Essentially, we use information from our sample data (track duration times and track acceptance threshold) as input into MLE, in order to find the population probability density function (PDF), that makes our data the most likely. The fitted lines represent this PDF [185, 186].

Histograms were binned based on the square-root rule, where the number of bins is equal to the square root of the sample size. We binned our data for presentation purposes only, in order to reduce noise associated with a finite sample size and reveal our sample distribution more clearly.

The PDF of the track durations is related to bleaching and unbinding as follows:

3)

$$k_{track}e^{-k_{track}t} = (k_{off} + k_{bleach})e^{-(k_{off}+k_{bleach})t}$$

Where k_{track} is the rate of track durations ending, k_{off} is the rate of unbinding, and k_{bleach} is the rate of bleaching. The model PDF we used for fitting was a left-truncated exponential distribution. This was used to compensate for the fact that we removed short duration tracks from analysis. The general form of this PDF is:

4)

$$\left(\frac{1}{\tau}\right)e^{\frac{-(x-L)}{\tau}}$$

where τ is the mean time, and L is the truncation point/origin of exponential distribution[187]. Note that the equation has the same form as expected for a translated exponential distribution and so we used this form for all data sets.

To correct for photobleaching, we used LacI tagged with mMaple. LacI is expected to have a binding time significantly longer than the bleaching time of mMaple in our experimental conditions[156]. Therefore, since the k_{off} term is much smaller than the k_{bleach} term, the average track duration is equivalent to the average bleach time for mMaple. Note that previous estimates of LacI bound-time at the *lacO* operator were determined using a single copy of the operator, while we used an array composed of 256 copies of *lacO*. This should result in even longer apparent binding of the repressor protein and increase the likelihood that focus disappearance is solely due to bleaching. We obtained a constant exposure bleaching curve, which we used for the 1s, and 5s intervals (500ms exposure data). We scaled the average bleach time from the constant exposure bleaching data, in order to use it for different time intervals. The constant exposure bleaching time is related to the average bleaching time as follows:

5)

$$T_{bleach} = \left(\frac{t_{interval}}{t_{exp}}\right) T_{constant}$$

Where $t_{interval}$ is the interval time, t_{exp} is the exposure time, and $T_{constant}$ is the constant exposure bleaching time.

We then calculated the average bound time using the following equation:

6)

$$T_{bound} = \frac{T_{track} T_{bleach}}{T_{bleach} - T_{track}}$$

In order to calculate the SE and 95% CI on the parameter estimates, we used the right-hand side of Equation 3. We used the bleaching times and bound times calculated previously, as initial estimates, and then performed bootstrap sampling over 10,000 samples, in order to calculate the standard errors and confidence intervals on the bound time estimates. We used the 'bias and accelerated percentile method' (BCA) algorithm when calculating CI, in order to compensate for any bias or skewness in the bootstrap distribution.

Previous characterization of photoblinking of mMaple found 49% probability of blinking and an average of 3.4 blinking events per molecule [188]. This same study set a cutoff time of 2.6 seconds to account for over 99% of the blinking events. We expected to detect fewer blinking events since the shorter ones will be recorded only as intensity fluctuations, and not discontinuities in the track, due to the use of longer capture rates, 500ms instead of 100ms. In addition, lower exposure intensity would likely contribute to a decrease rate of blinking [189]. To estimate the effect of blinking in our analysis, we used the data of LacI using 500ms capture-times. We analysed the data as previously except that we did not apply the one-frame memory parameter during tracking. We then determined the number of frames between two consecutive tracks at the same position of the field of view. We used a 2.6 second cutoff in our data since longer gap times likely represent new binding events instead of blinking.

For DnaB, since 500ms capture-times were not efficient at preventing diffusing molecules from being detected, even after the PSF threshold was applied (Figure 2-figure supplement 4G), we used exposure times of 2s and spaced capture by intervals of 2 seconds and 10 seconds. We used Pol III ε subunit as a control to ensure that increasing the exposure time does not significantly alter the bound time estimates (Figure 2-figure supplement 3).

Since the track durations of DnaB are similar to those of LacI, we determined a weighted average of the track duration times obtained and for each data set of DnaB performed a constrained fit (i.e. fitting with bounds placed on the estimates). We calculated a bound time from the weighted average in order to generate an initial estimate of the bound time, which was then used for the constrained fit. We allowed for 20% variation in the bleaching time in order to determine physically reasonable estimates. The lower and upper bounds for the DnaB bound time were 1 second and 90 minutes, respectively.

For the fitting procedure, we calculated the negative log-likelihood function of the two parameter (bleach time, bound time) left-truncated exponential distribution, as well as the gradients. We then used the MATLAB minimization function, *fmincon*, in order to find the parameters that minimize the negative log-likelihood function. This was done to improve the convergence to the correct solution, especially if the initial estimates were far from the actual solution, and to simplify the estimation procedure. We subsequently performed bootstrap sampling as discussed before to calculate standard errors and confidence intervals (**Table A1.3**).

The final estimates for bound times were calculated by doing a weighted average of data taken on multiple days and with different time intervals.

We performed a chi-square goodness of fit test under the null hypothesis that our data is sampled from a single-exponential distribution, and the alternative hypothesis that it does not. It is possible however that even if the fit is good, that a different model fits the data better (e.g. two exponential model). We wanted to determine the best model for the data and we performed two tests in this regard: 1) Log-Likelihood Ratio(LLR) test and 2) Bayesian Information Criterion (BIC) test.

Log-Likelihood Ratio Test- The LLR test tries to test if an unconstrained model statistically significantly fits the data better than the constrained model, by comparing the likelihood values obtained from the unconstrained versus the constrained. In our case, the unconstrained model is the two-exponential model while the one-exponential model is the constrained model, as shown below:

7)

$$p\left(\frac{1}{\tau_1}\right)e^{\frac{-(x-L)}{\tau_1}}+(1-p)\left(\frac{1}{\tau_2}\right)e^{\frac{-(x-L)}{\tau_2}}$$

where $\tau_1=(T_{bleach}+T_{bound\alpha})/T_{bleach}*T_{bound\alpha}$, $\tau_2=(T_{bleach}+T_{bound\beta})/T_{bleach}*T_{bound\beta}$, p is the mixture probability, and L is the truncation point.

Note that if we constrain p=1, we recover the single-exponential model.

Bayesian Information Criterion (BIC) test- The BIC test determines which model fits the data better, but penalizes for greater complexity (i.e. more parameters), to prevent over-fitting the data. The lowest number obtained through the test indicates the model that fits the data the best with the least complexity. When calculating the MLE estimates and log-likelihood values for the two-exponential model, the lower and upper bounds for the two timescales were 0.1 and 5400 seconds, respectively. The bounds for the bleaching constant were placed such that it allowed for 20% variation in the estimate.

The criterion we used to judge if the two-exponential model was the better model, was if the BIC test gave the lowest value for the two-exponential model and the LLR test gave a $p<0.01$. Also, the estimates obtained from the two-exponential should be sensible, and especially, they should not give us simply the values of the bounds, as that indicates that no estimates were found.

We found a few cases where the dataset passed the criterion. The timescales estimated were not consistent however, and upon further examination we realized that it was due to a few noticeable outliers in the dataset, possibly from noise due to dirt still on the coverslip. When we removed the outliers, it resulted in these datasets not passing the criterion, but without significantly changing the bound times previously obtained in the single-exponential fits.

3.5.6 Blinking Analysis of mMaple for sptPALM

We used LacI data collected with 500ms exposure as fast as possible (~500ms interval time), to characterize mMaple under our acquisition settings. The mean positions of single-molecule tracks initiating at the first frame were used as ROIs around which a 7x7 pixel window was drawn to extract intensity-time traces. Fluorescence from a bound molecule was identified as being 2 standard deviations above the mean cellular background, and the signal had to be above this threshold for >3 localizations (similar to track acceptance threshold previously described). Gap durations were calculated as the number of frames between bound fluorescence signals. Fits to the gap durations were done through MLE using a truncated exponential model. The resulting fit was used to calculate the probability of a gap duration lasting greater than a specified value, through integration.

3.5.7 Estimation of β-clamp loading rate

To estimate the effective loading rate we followed the following equation described elsewhere [151]:

8)

$$\frac{1}{T_{eff\_load}} = \frac{\beta_{2bound}}{T_{unload}}$$

where T_{eff_load} represents the loading rate, β_{2bound} represents the number of copies of β clamp at the fork and T_{unload} represents the bound-time of a β clamp. In our estimations we assume that there are 23 β dimers per fork as previously estimated [151].

3.6 Acknowledgements:

We thank Ann McEvoy for kindly providing a plasmid carrying mMaple. We thank David Sherratt, Adam Hendricks, James Graham, Jackie Vogel, Daniel Jarosz, Charl Moolman and members of the Reyes-Lamothe lab for discussion and helpful comments on this work. S.U. was supported by a Sir Henry Wellcome Fellowship by the Wellcome Trust and a Junior Research Fellowship at St John’s College Oxford. A.J.M.W. and M.C.L were funded by the Medical Research Council U.K.

(MRC# MR/K01580X/1), the Biology and Biotechnology Research Council U.K. (BBSRC# BB/N006453/1) and the Biological Physical Sciences Institute (BPSI) at the University of York, U.K. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC# 435521-2013), the Canadian Institutes for Health Research (CIHR MOP# 142473), the Canada Foundation for Innovation (CFI# 228994), and the Canada Research Chairs program.

3.7 Author contributions

TRB, NK, Formal analysis, Investigation, Writing—original draft, Writing—review and editing; EN, Investigation, Writing—original draft; SU, Software, Investigation, Writing—original draft, Writing—review and editing; AJMW, Software, Writing—original draft, Writing—review and editing; MCL, Formal analysis, Writing—original draft, Writing—review and editing; RR-L, Conceptualization, Supervision, Funding acquisition, Investigation, Writing—original draft, Writing—review and editing

Chapter 4

A Machine Learning Approach for Classification of DNA-Bound Proteins from Single-Molecule Tracking Experiments

This chapter is based on a manuscript in preparation: Kapadia, N., El-Hajj, Z.W., and Reyes-Lamothe, R. A machine learning approach for classification of DNA-bound proteins from single-molecule tracking experiments.

Building off the work I did in Chapter 3, I decided to create a method that would be more robust in estimating residence times of proteins bound to DNA, from single-molecule tracking experiments, and in a more high-throughput manner. In order to have confidence in the estimates, one needs to ensure that they are primarily analyzing tracks representing DNA-bound molecules, and not diffusing molecules or false positives due to errors in the tracking algorithm. Therefore, I developed a machine-learning approach to classify tracks, that is robust, user-friendly, and generalizable to a range of conditions. I show that this approach is accurate *in silico*, and in *E.coli* and budding yeast.

4.1 Abstract

Many proteins act on DNA for a wide range of processes, including DNA replication, DNA repair, and transcription. Their time spent on DNA can provide insight into these processes and their stability within complexes to which they may belong. Single-particle tracking allows for direct visualization of protein-DNA kinetics, however, identifying whether a molecule is bound to DNA can be non-trivial. Further complications arise with tracking molecules for extended durations in cases of processes with slow kinetics. We developed a machine learning approach, using output from a widely used tracking software, to robustly classify tracks in order to accurately estimate residence times. We validated our approach *in silico*, and in live-cell data from *Escherichia coli* and *Saccharomyces cerevisiae*. Our method has the potential for broad utility and is applicable to other organisms.

4.2 Introduction

Quantitative information regarding the kinetics of a protein provides valuable insight in the behaviour of the respective protein, as well as its relationship with other proteins in cases where it belongs to a protein complex. This in turn may inform on the activity of the protein. The residence times of DNA-bound proteins (DBP) can reveal important details on basic cellular processes such as transcription, DNA repair, and DNA replication, at the timescales at which they operate [15, 124, 135, 154, 190-192]. This is true of not only proteins that bind directly at sites on DNA such as initiator proteins, repair proteins, and chromatin remodellers, but also proteins that are part of complexes that translocate on DNA such as the DNA replication complex (replisome) and RNA polymerase (RNAP) [15, 193].

Recent advances in fluorescence microscopy have allowed us to study protein kinetics directly in living cells, with the most common techniques being single-particle tracking (SPT), fluorescence recovery after photobleaching (FRAP), and fluorescence correlation spectroscopy (FCS) [124, 138, 153, 190, 193-196]. SPT has the particular advantage of being able to directly observe protein behaviour, allowing for a wealth of information to be extracted from the images, both qualitatively and quantitatively [14, 139]. Typically, SPT has been used to determine binding kinetics of DBP with very fast kinetics (hundreds of milliseconds to a few seconds), by using capture rates of few to tens of milliseconds. However, certain processes operate on much longer timescales posing issues for SPT, including photobleaching and tracking single-molecules for long durations. One way to bypass these issues is by using long-exposure times to blur out diffusing molecules, in combination with low-laser power and stroboscopic illumination to minimize photobleaching. Nonetheless, tracking single-molecules over long durations can be difficult due to intensity fluctuations caused by the molecule moving out of focus and photophysics, resulting in fragmentation of tracks [15, 197] (**Figure 4.1A**). Reducing the intensity threshold required for localizing molecules can compensate for this, although with the disadvantage of introducing false positives. In addition, even with motion blurring and tracking parameters to select for only bound molecules, it is possible to detect diffusing molecules or have false positives due to localization and tracking errors; therefore filtering steps on tracks need to be included to only isolate tracks representing DNA-bound proteins [15, 134, 154, 198]. All these complications arise from the tracking algorithms and automated analysis, since the user can typically distinguish immobile DBP when looking at the raw images when using long-exposure and stroboscopic illumination as they appear to wiggle around.

Here, we provide a user-friendly and robust method to determine residence times of DBP using machine learning and SPT in live *Escherichia coli* (*E.coli*) and *Saccharomyces cerevisiae* (budding yeast), that can be easily extended to other organisms.

4.3 Results

4.3.1 Random forest for single-molecule tracking classification

Machine learning (ML), and its branch of deep-learning, is a powerful tool for image analysis and classification, with the most common implementation being supervised learning, whereby a labelled training data set is given to the ML algorithm, which then builds a model for subsequent classification [142, 199]. Our motivation arose from recognizing the limitations associated with automated detection and tracking, and we wanted to develop a classification method to compensate for these limitations. We recognized that a known DBP (e.g. histone H3) could be used to construct a training data set manually, for how DBP in general should move and how single-molecules should look like in images, from which we can build a ML model. We can also use this control as a photobleaching control when we estimate the residence time for our DNA-bound protein of interest (POI) [15].

We used Trackmate, a freely available plugin in Fiji, to track molecules and used variables from the tracking output that we believed to be good predictors of single-molecule DBP and help to reduce the cross-validation error (CVE) (**Figure 4.1B**) [128]. To predict whether a molecule moves like a DBP we used the following variables from the tracking output: mean speed, maximum speed, minimum speed, and median speed. For predicting whether the track represented a genuine molecule, we used the maximum quality variable – a parameter corresponding to the intensity of the molecule as well as its shape [128]. Classification of the track was done manually by looking at the raw image data, with binary classification: 0 for a diffusing molecule or noise, and 1 for a track representing a genuine immobile, DNA-bound molecule. We chose to use the random forest algorithm to construct a model from the training data, as it is accurate, tolerant of noise, and less prone to overfitting [200] (**Figure 4.1C**). The various parameters used to construct the different random forest models are listed in **Table A2.1.** We used this model to classify tracks for our POI (**Figure 4.1C**), and subsequently determine their residence time after correcting for photobleaching.

We first tested our approach using computer simulations of single-molecule timelapses in *E.coli* and nuclei of budding yeast (See Materials and Methods). The main difference between the two simulations aside from cell shape, was that the budding yeast simulation had a distortion in the shape of the molecule based on its position in z, to model the point spread function (PSF) and emulate the molecule going out of the focal depth of the objective. Under our experimental conditions, *E.coli* cells have diameters of 0.7μm, which is not far from the estimated focal depth of high numerical aperture objectives commonly used in single-molecule studies (~0.4 μm), so we assumed molecules to be in focus, regardless of their z position[15]. We first constructed training data sets from simulated data with 500ms exposure, no time interval, and a mean bleach time of 10s (**Table A2.2)**. We had a stable, bound fraction (Diffusion coefficient (D) of D_{bound}= 0.005$\mu m^2/s$, bound time >>> bleach time), along with a mobile fraction (D_{mobile} = 0.5$\mu m^2/s$) **(Figure 4.1D).**

We only considered tracks with ≥ 4 localizations, as it is difficult to discern the state of the molecule for shorter tracks. In order to make use of a single training data set, instead of constructing multiple training data sets for classifying data sets collected under different conditions, we constructed two models: ML model 1 only has the speed variables and ML model 2 has the speed variables along with maximum quality (**Figure 4.1D)**. We calculate the mean of mean speed as well as the mean of maximum quality from the tracks classified as being bound in the training data and use these values to scale the speed and quality variables, respectively, for data collected with different time intervals and/or illumination intensities **(Figure A2.1).**

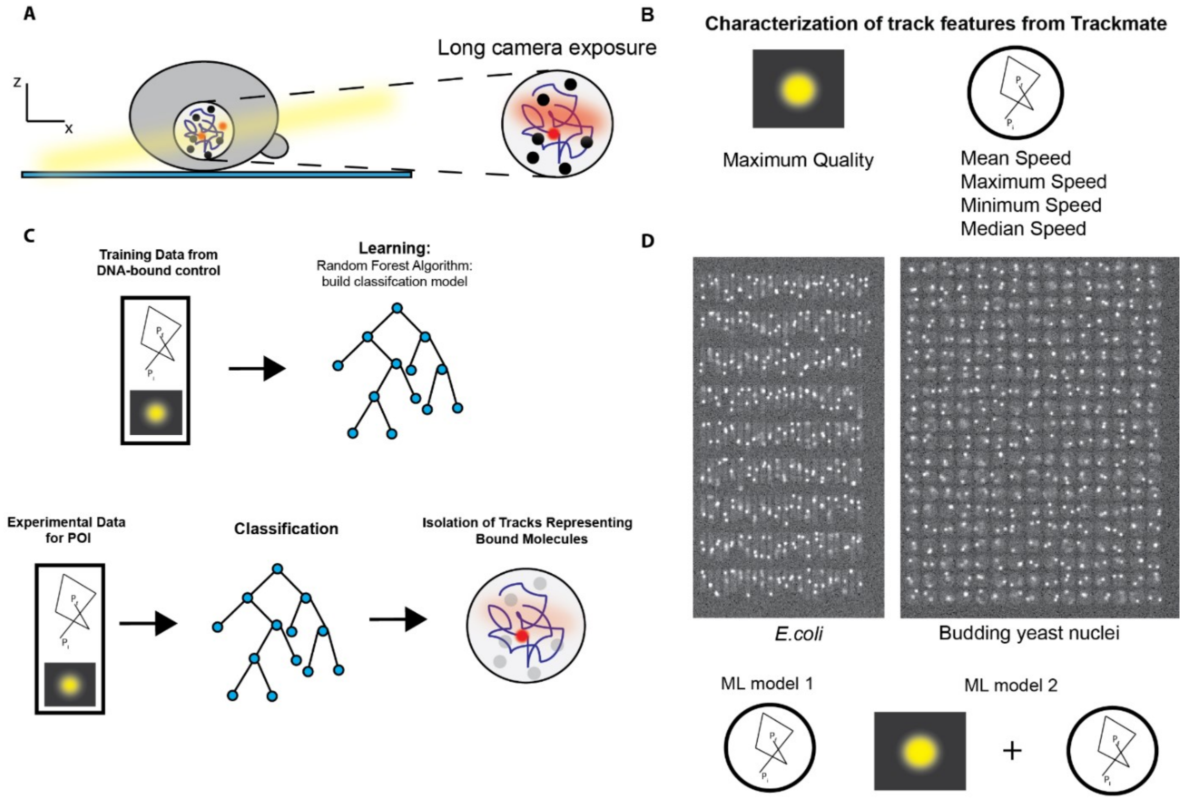


Figure 4.1: Approach to isolate DNA-bound molecules. A) Diagram of experimental setup to do single-molecule tracking with photoactivatable/photoconvertible fluorophores to calculate residence times on DNA. B) Variables from Trackmate used to predict if a track represents a genuine bound molecule. C) Illustration of the general procedure for ML and how it can be used to predict tracks from experimental data for POI. D) Top - Example images from simulations of single-molecule movies representing *E.coli* cells and budding yeast nuclei. Bottom – Illustration of the variables used in ML models 1 and 2.

The entire procedure for the classification of tracks for a POI is as follows:

1) We performed a two-component Gaussian mixture model (GMM) fit on the log of mean speed, of all the tracks, and the component with the lowest mean is selected as representing the immobile molecules. This step does not need to be robust as it will be used to do some initial filtering for subsequent steps.

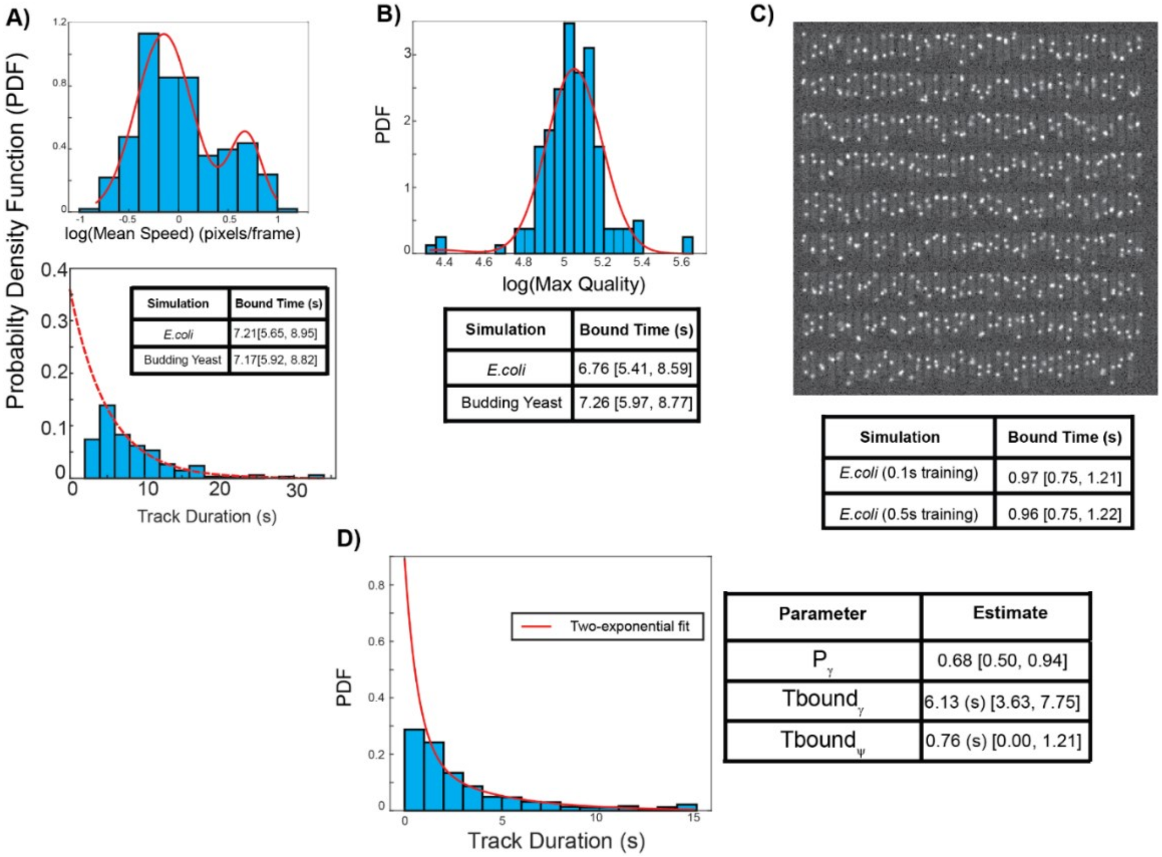
2) Using this value, we scale the speed variables accordingly, using the mean of mean speed calculated from the training data, as follows: scale factor$_{speed}$ = mean (mean speed$_{training}$)/mean(mean speed$_{boundPOI}$). This helps to ensure the ML models can classify tracks obtained from different time intervals than the training data. We then run the tracks through ML model 1 for initial filtering.
3) From the resulting tracks, we calculate the mean of maximum quality using a two-component GMM, similar to step 1 except selecting the component with the higher mean and use the same variable from the training data to scale the quality variable, similar to the previous step. We then run the tracks through ML model 2 for final classification.

To assess how well the procedure worked, we quantified both the accuracy (proportion of tracks accurately predicted to be bound), and the recovery error (RE) (the fraction of tracks known to be representing bound molecules recovered by the classification procedure). While high classification accuracy is important to ensure accurate estimation of residence time, we also wanted to make sure that the recovery error is small, as single-molecule studies are often plagued by low sample sizes thereby reducing confidence in the estimate [14, 15]. In addition, the out-of-bag (OOB) errors, equivalent to the CVE for random forests [200], were estimated for each ML model **(Table A2.1)**. On test data, we obtained an accuracy of 0.96 and RE of 0.10 for the *E.coli* simulation, and an accuracy of 0.97 and RE of 0.06, for the budding yeast simulation (**Table A2.3)**.

4.3.2 Accurate estimation *in silico* of residence times under different conditions

Next, we tested how well the ML models would work on simulated data with different time intervals and spot intensities (**Table A2.2)**. First, we tested on data with a 1s time interval and spot intensities the same as the training data set. The residence time was set to 8s, while the bound fraction was set to 0.5. As **Figure 4.2A** (top) illustrates, we first calculate the mean of mean speed of the bound population, in order to scale the speed variables. After the final classification step, for both the *E.coli* simulation and the budding yeast, we were able to obtain an estimate for the residence time that was within error of 8s (**Figure 4.2A** (bottom)). In addition, the accuracy values were 0.93 and 0.96, with recovery errors of 0.11 and 0, respectively, for the *Ecoli* and budding yeast simulations (**Table A2.4)**. We then tested whether we can obtain an accurate estimate in a situation representing poorer image quality (e.g. lower spot intensity). As **Figure 4.2B** (top) illustrates, we use the GMM fit on the max quality values to find an appropriate scaling factor.

After classification, we find that we can still recover an estimate of the residence that is within error of the known bound time, for both simulations, with accuracy values and REs similar to the previous condition (**Figure 4.2B, Table A2.4**).



Simulation	Bound Time (s)
E.coli	7.21[5.65, 8.95]
Budding Yeast	7.17[5.92, 8.82]

Simulation	Bound Time (s)
E.coli	6.76 [5.41, 8.59]
Budding Yeast	7.26 [5.97, 8.77]

Simulation	Bound Time (s)
E.coli (0.1s training)	0.97 [0.75, 1.21]
E.coli (0.5s training)	0.96 [0.75, 1.22]

Parameter	Estimate
P_γ	0.68 [0.50, 0.94]
$Tbound_\gamma$	6.13 (s) [3.63, 7.75]
$Tbound_\psi$	0.76 (s) [0.00, 1.21]

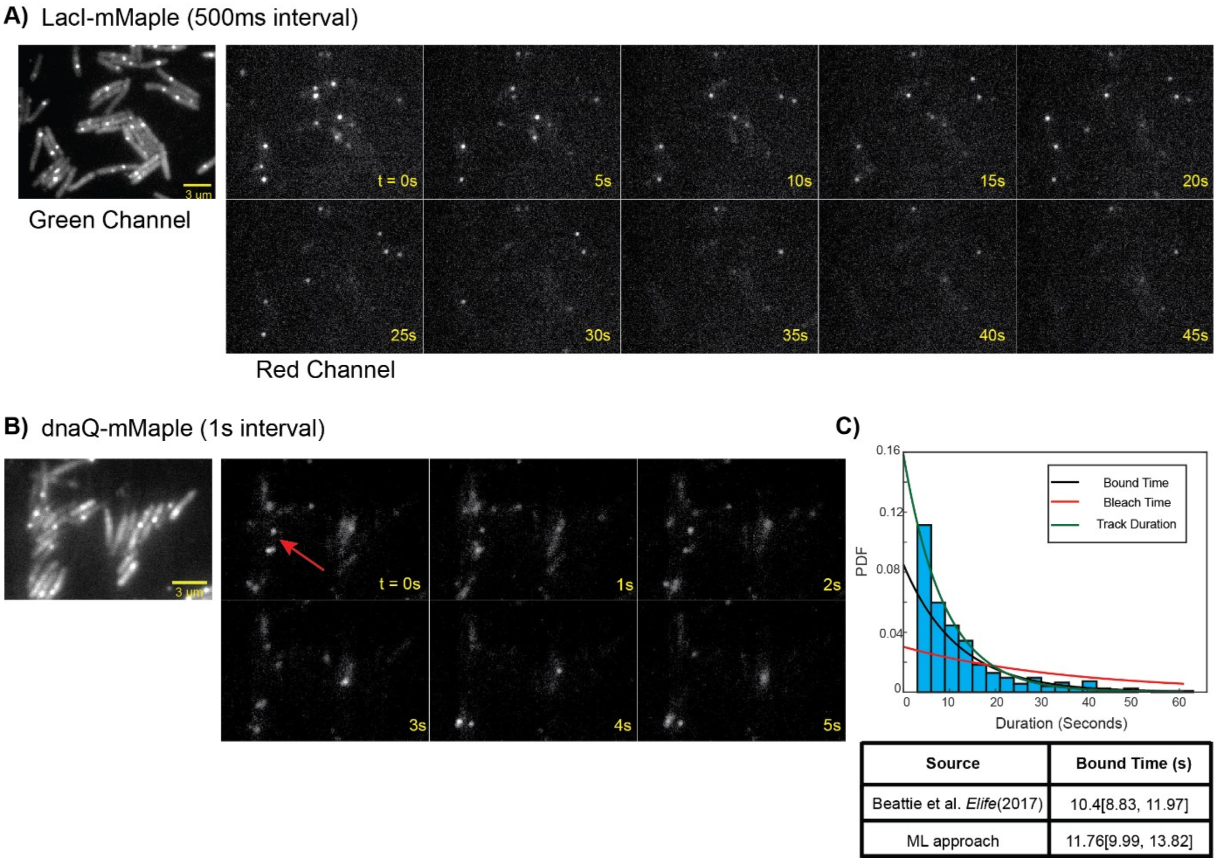
Figure 4.2: Estimating residence times in a range of conditions. A) Top – Diagram illustrating the GMM fit to determine the mean of mean speed for 1s interval data, which can be used to rescale the speed variables. Bottom – Estimates of residence times in both simulations. B) Top – Diagram illustrating the GMM fit to maximum quality values from a data set with lower fluorescent intensities. Bottom – Estimates of residence times from simulated data representing poorer image quality. C) Top – Example of image from simulated 100ms timelapse in *E.coli*. D) Two-exponential fitting to a data set with heterogenous population of bound molecules. For all estimates, 95% confidence intervals are shown.

While a 500ms exposure helps to blur out diffusing molecules, it might be too long for faster processes. Therefore, we constructed a training data set with 100ms exposure and tested on data with a mean residence time set to 1s while the bleach time for this condition was set to 2s (20 frames) (**Figure 4.2C)**. Once again, we were able to obtain an estimate within error of 1s, although with lower accuracy and higher recovery errors than with 500ms (**Table A2.4**), as one would expect under conditions where the separation between two diffusive states is more difficult to discern. (**Figure 4.2C)**. We tested this by changing D_{mobile} to 5 $\mu m^2/s$, and found the errors were drastically reduced **(Table A2.4)**. We then asked whether the 500ms training data can be used to classify the 100ms data, and surprisingly, we found that it still performed well (**Figure 4.2C)**, with comparable errors to the 100ms training data (**Table A2.4)**. We also found that we could get accurate estimates of the residence times of a heterogenous population of bound molecules, where two distinct binding regimes are present (**Figure 4.2D).** We note that for the simulation representing a heterogenous population of bound molecules, we had to change the bleach time to 10s, in order to recover the two binding times. As others have alluded to, detecting multiple populations is highly dependent on acquisition settings which is why we had to change the bleach time to recover binding times of 1s and 7s [201]. These results suggest that one can use ML models constructed from a single training data set and use them for classifying tracks obtained from data collected under widely different conditions, along with obtaining accurate residence times of homogenous and heterogenous bound populations.

4.3.3 Experimental validation in *E.coli*

We then wanted to determine if our approach could work with estimating residence times in *E.coli*. For our DNA-bound control we used LacI, a transcriptional repressor, fused to the photoconvertible mMaple, which upon illumination with 405nm light, converts to the red form [155]. LacI has been reported to bind stably to the *lacO* array site on DNA (~5 minutes), thus making it suitable as a photobleaching control, under our acquisition times [156] **(Figure 4.3A)**. Although we had initially used this strain for photobleaching correction in [15], we decided to use it for constructing a training data set as well. The LacI data used for the training data was collected with 500ms exposure, as fast as possible.

Once the ML models were constructed, we asked whether we could get an accurate estimate of the residence time for *E.coli* DNA polymerase, PolIII, for which we had previously reported to have a residence time of around ~10s, [15]. We reanalyzed a dataset from [15] taken of the PolIII subunit, ε, also tagged with mMaple (dnaQ-mMaple), to determine if the ML approach could give identical results (**Figure 4.3B)**. We found that we were able to recover an estimate of the residence time that was very similar to the previously reported estimate **(Figure 4.3C),** and with small confidence intervals, confirming the robustness of our approach in estimating residence times of DBP in *E.coli*.



Source	Bound Time (s)
Beattie et al. *Elife*(2017)	10.4[8.83, 11.97]
ML approach	11.76[9.99, 13.82]

Figure 4.3: Estimating residence times for PolIII subunit, ε - A) Example of LacI-mMaple data collected with 500ms exposure. Red channel represents image after 405 induced photoconversion of mMaple. B) Example of dnaQ-mMaple timelapse. Red arrow indicates molecule that we observed dissociating. C) Estimation of residence times and comparison to estimate from [15], with 95% confidence intervals.

4.3.4 Estimating residence time of Topoisomerase II in budding yeast

We asked if our approach could be used in live haploid budding yeast. For our photobleaching control and training data construction, we used histone H3 fused to HaloTag (H3-HaloTag), due to its expected long residence time and high bound fraction [202] (**Figure 4.4A)**. In order to detect the protein, we incubated the cells with the cell-permeable photoactivatable (PA) dye, JF-PA549, and the experimental protocol was very similar to that described in *E.coli* [15], including highly-inclined and laminated optical (HILO) sheet stroboscopic illumination with 500ms exposure, but with cycles of low-dose 405nm activation every 40 frames **(Materials and Methods)**. To improve image quality with budding yeast, we also used the refractive-index matching media, Optiprep, to minimize light refraction due to the cell wall [203]. The mean photobleaching duration estimated, with 1s interval acquisition, was ~23 seconds **(Materials and Methods)**.

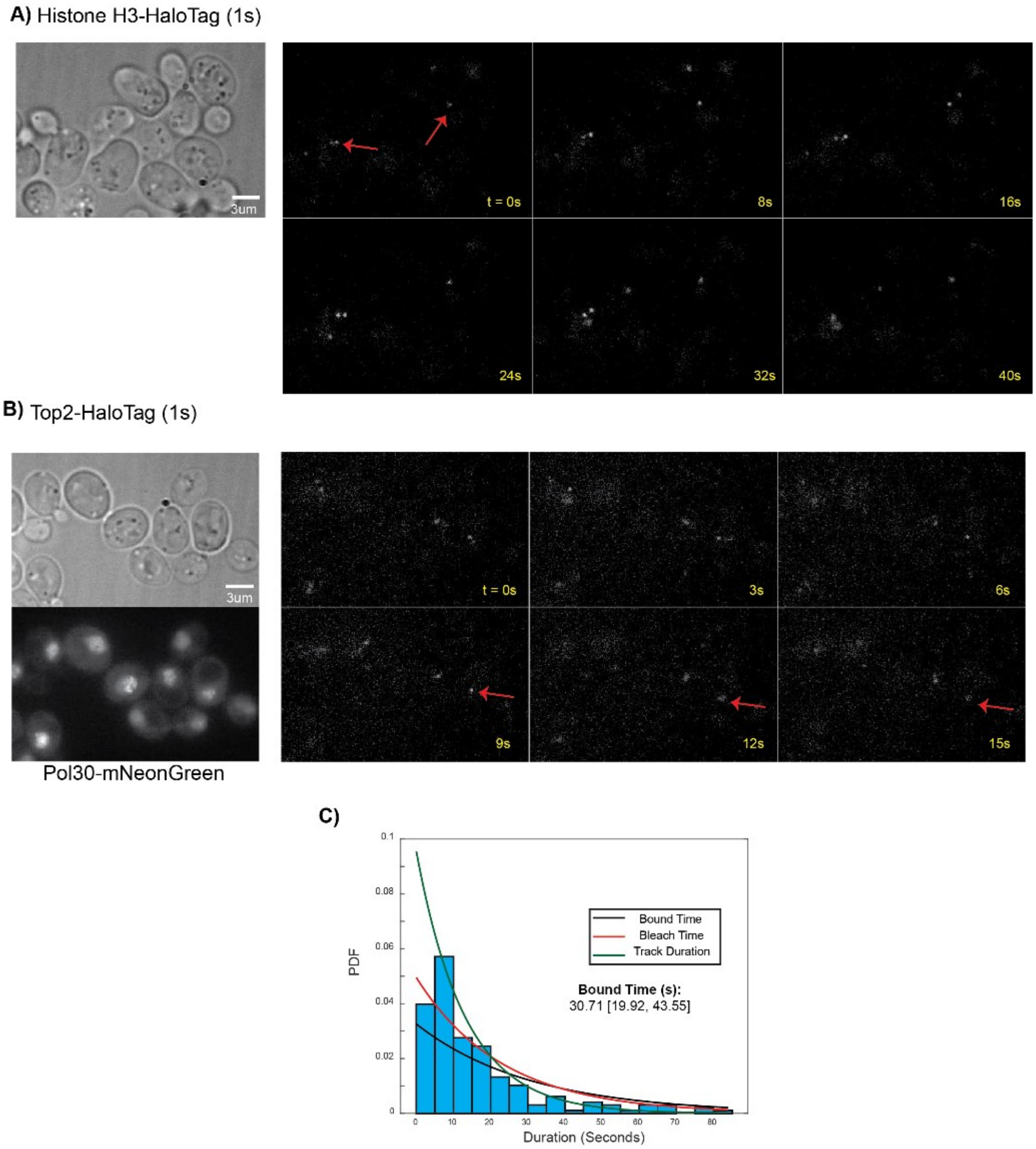


Figure 4.4: Estimating residence time of Top2. A) Example of H3-HaloTag images, with timelapse shown after photoactivation with 405. B) Example of Top2-HaloTag timelapse after photoactivation. Red arrows indicate molecule we observed dissociating. Also shown is the SME of a Pol30-mNeonGreen z-stack. C) Estimation of residence time with 95% confidence intervals.

For our experimental test, we tagged Topoisomerase II (Top2) with HaloTag (Top2-HaloTag) and to segment nuclei in order to isolate tracks found only in nuclei, we tagged proliferating nuclear cell antigen (PCNA) with mNeonGreen and acquired a z-stack in the green channel prior to acquisition **(Figure 4.4B)**. The resulting z-stack went through a smooth manifold extraction (SME) process for better image quality of the nuclei [204]. We chose Top2 as mammalian Top2 enzymes have been reported to be dynamic and can complement Top2 function in budding yeast *Δtop2* strains, suggesting they function similarly, albeit with the possibility of slightly different dynamics [196]. We collected the data under the same acquisition settings as H3-HaloTag. As indicated in **Figure 4.4B** with the red arrows, we could visualize dissociation of the protein under our acquisition settings, suggesting it is dynamic. Following tracking analysis and ML classification, we obtained an estimate for the residence time of Top2 of ~30s, consistent with dynamic behaviour reported previously (**Figure 4.4C)**. As one can notice, the spot quality in live budding yeast, is of slightly poorer quality compared to *E.coli* (compare **Figure 4.3B vs Figure 4.4B)**, but we surprisingly found our ML approach still worked quite well in this system, suggesting our approach can be used in a range of conditions and organisms.

4.4 Discussion

Here we have provided a robust, easy-to-use classification approach that can be used on long-capture single-molecule data, to isolate tracks of DBP in order to accurately estimate their residence times. While there are alternatives to classify bound molecules such as estimating their diffusion coefficient, step-size per frame of DBP, PSF, etc., and applying a threshold for acceptance, they often suffer from not being precise estimators or require additional analysis steps and/or experiments, and setting hard thresholds may discard too many tracks [15, 134, 154, 205].

We do acknowledge the limitations with our approach, namely that in cases where the diffusing fraction has similar mobility to the bound fraction, the error in classification will increase, although this would likely be the case for most classifiers. Also, constructing training data sets manually can take significant time and choosing the right parameter values for building random forest models does require some expertise; however, our results suggest that once the initial models have been created, they can be used across a range of conditions whether that be different exposure times, time intervals, and data quality, not to mention the potential to estimate residence times in a high-throughput manner for cases where one is interested in an entire protein complex. Finally, we did have to adjust our simulation parameters (akin to adjusting experimental conditions) to detect two binding regimes in the case where there was a heterogenous population of bound molecules, similar to what a recent study had reported [201]. Nonetheless, this is not an issue with our classification approach but rather the acquisition settings used (or in our case, simulation parameters), and our results suggest simulations can help identifying the optimal experimental conditions to use for detecting multiple binding states.

Finally, although we have shown our ML approach, along with the experimental protocol, works well in small organisms such as *E.coli* and budding yeast, it should be easily extendable to other systems such as mammalian cells.

4.5 Materials and Methods

4.5.1 Computer simulations of single-molecule timelapses

Computer simulations of images were written in Python 3.6.

First, a 5000x5000 array was constructed as an image. Each array element was 10nm, therefore the total size was meant to represent 50µm x 50µm. Cells were placed in a grid-like pattern to prevent overlap and given a low intensity value of around 0.2 per array element to represent cell autofluorescence. *E.coli* cells were modelled as 3D rectangles, with average width of 0.7µm, and an average length of 3µm., while budding yeast nuclei were modelled as spheres of average diameter 2µm. For each cell, unless stated otherwise, two fluorescent molecules were assigned such that their initial locations were confined to the interior of their respective cell. A weighted sampling method was used to assign their initial diffusive state.

To model fluorescent spots, a spot intensity was assigned at the center of the molecules and using a standard deviation of 130nm, the intensity was spread across the region using a Gaussian filter. In the case of the spots in the budding yeast simulation, the standard deviation used was increased by 5% for every 150nm increase/decrease in z position of the molecule's position. For example, from +/- 150nm from the z-origin of 0, the molecule was assumed to be in focus. However, if the molecule moved to a |z position| of >150nm but less than 300nm, its standard deviation was increased by 5%. After the Gaussian filter, poisson shot noise at each element was added by sampling from a poisson distribution with mean equal to the initial intensity value at that element.

Kinetics of the molecule were determined using transition matrices: one for the photophysics (e.g. photobleaching), and the other for transitions to different diffusive states. We determined the probability of transition using a time step (τ) of 5ms. We omitted photoblinking as the lower laser powers used in long-exposure experiments typically do not cause it. This simplified the photophysics transition matrix to:

	$F_{state}(t+\tau)$	$B_{state}(t+\tau)$
$F_{state}(t)$	$1-CDF_{bleach}$	CDF_{bleach}
$B_{state}(t)$	0	1

Where the F_{state} is the fluorescent state, B_{state} is the irreversible photobleached state, and CDF_{bleach} is the probability of photobleaching within τ, calculated from the cumulative distribution function (CDF) of an exponential distribution with a specified mean bleach time. If the molecule went to the B_{state}, it was removed from subsequent iterations.

The transition matrix for transitions to different diffusive states was:

	$Mobile_{state}$ $(t+\tau)$	$Bound_{state}(t+\tau)$
$Mobile_{state}(t)$	$1-CDF_{search}$	CDF_{search}
$Bound_{state}(t)$	CDF_{bound}	$1- CDF_{bound}$

Where $Mobile_{state}$ represents the state with $D = D_{mobile}$, and $Bound_{state}$ represents the state with $D = D_{bound}$, CDF_{search} is probability that the molecule switches to the $Bound_{state}$, based on an exponential distribution with a specified mean search time, and CDF_{bound}, is the probability that the molecule switches to the $Mobile_{state}$ based on an exponential distribution with a specified mean bound time. In the case of the heterogenous population of bound molecules, a second $Bound_{state}$ was added with a different bound time.

From the transition matrices and the respective probabilities, we would select the next state using sampling from a multinomial distribution with the weights given by the probabilities.

To simulate molecule movement, the step size in each x,y,z direction was picked from a Gaussian distribution with variance $2D\tau^{\alpha}$, and $\alpha = 1$ for the mobile fraction, and 0.4 for the bound fraction to represent subdiffusive behaviour of genomic loci, due to the Rouse model of DNA polymer motion, although we do acknowledge that loci in *E.coli* can undergo more ballistic motions [206-208].

After each time step, the sum of the values of 10x10 array elements was taken to simulate a 100nm pixel size and get a 500x500 pixel image. Subsequently, camera noise was added by sampling from gaussian distribution with mean = 150 and standard deviation 20. To obtain images of different exposure times, time steps were integrated (e.g. 100ms = 5ms * 20 steps of integration). In the case of time intervals, we allowed molecule movement but no image formation until the next image is taken. E.g. for 1s interval, a 500ms image was followed by 500ms of only molecule movement but no image formation, followed by another 500ms image.

4.5.2 Construction of Top2-Halo

Histone H3-Halo (YTK1434) and the *pdr5*Δ::KanMX deletion (YTK1414) were a gift from Tatiana Karpova [134]. The PCNA-mNeonGreen (YTB31) and Top2-Halo (ZEY078) used in this study are from a BY4741 background. Plasmids used in this study are derivatives of pUC18 (ColE1 origin, Ampicillin resistance); pTB16 carries mNeonGreen and a downstream NatMX marker, while pSJW01 carries the HaloTag gene with a HygB marker. Both mNeonGreen and HaloTag genes encode an 8 amino acid linker at the 5' end (sequence: GGTGACGGTGCTGGTTTAATTAAC). Plasmids were maintained in *E. coli* DH5α and were extracted by growing in LB with 100 μg/ml ampicillin then using the Presto Mini Plasmid Kit (Geneaid).

All PCRs were made using either Phusion or Q5 (NEB). PCR reaction were in a volume of 50 μl and included water, 3% DMSO, the reaction buffer, 2.5 mM of each dNTP, 0.2 μM of each primer, either 1ng of plasmid DNA (for insertions) or 1 μl of genomic DNA (for screening insertions), and 0.5 μl of polymerase.

Fluorescent fusions were made by PCR amplification from pTB16 (for PCNA-mNeonGreen) or pSJW01 (for Top2-Halo) using the primers listed in **Table A2.5**. PCR products were transformed into wild-type diploid BY4743. A single colony was grown at 30°C in 5 ml yeast peptone dextrose (YPD) overnight, then diluted to a final OD600 of 0.1 in 10ml of YPD. Cells were taken at OD600=0.5-0.6 and centrifuged at 4000 rpm for 5 min. The pellet was washed twice with 25 ml of sterile deionized water, then once with 1ml of 100mM lithium acetate. Cells were then resuspended in 240 μl of 50% PEG, then 50 μl of salmon sperm DNA (thawed at 95°C for 5min, then incubated on ice for at least 10min), 50 μl of the PCR product and 36 μl of 1M lithium acetate were added in this order. The mixture was thoroughly mixed by pipeting and incubated on a rotator at 25°C for 45min, followed by heat shock at 42°C for 30min. Cells were pelleted in a microcentrifuge, washed in 500 µl of sterile water, then resuspended in 200 µl of YPD and plated on YPD agar. After growing at 30°C overnight, the cell lawn was replica-plated onto selective YPD agar, either with 100 µg/ml cloNAT (Werner) for PCNA-mNeonGreen, or with 200 µg/ml Hygromycin B (Life Technologies) for Top2-HaloTag. Transformants were screened for the presence of an insert by PCR using the indicated primers in **Table A2.5**.

Confirmed clones were then sporulated by taking 750 µl of a YPD overnight cultures, washing 4 times with 1 ml sterile deionized water, then washing once with 1 ml of potassium acetate sporulation medium (KAc), and finally resuspending in 2 ml of KAc and incubating at 25°C with shaking. After 5 days the sporulating cultures were checked by microscopy for the appearance of numerous tetrads, then 750 µl was taken and washed 3 times in sterile water before final resuspension in 1ml water and storage at 4°C. For dissection, 45 µl of spores were treated with 5 µl of zymolase for 10 min, then tetrads were dissected on YPD plates to isolate haploids with the tagged fusion. Genomic DNA was isolated from the haploid by vortexing the cells in the presence of zirconia/silica beads, followed by phenol extraction and ethanol precipitation. The insertion site was once again amplified using the same screening primers as above, and the PCR product was sequenced to confirm that the tag and linker were both mutation-free.

The Top2-HaloTag haploid was then combined with PCNA-mNeonGreen (from YTB31) and the *pdr5*Δ::KanMX deletion (from a haploid sporulated from YTK1414) by mating. To do this, 10 µl of water was spotted on a YPD plate, and colonies from the Mat a and Mat α haploids to be combined were mixed together into the water drop and incubated at 30°C overnight. Cells were then restreaked on an auxotrophic -Met-Lys plate on which only the mated diploid could grow. Diploids resulting from a mating were dissected as above, and eventually haploids with all three markers were isolated and used for imaging.

4.5.3 Single molecule imaging in budding yeast

A single colony from a YPD plate was placed in 5mL synthetic complete (SC) medium and grown with shaking at 30 degrees Celsius for ~5-6 hours. This culture was diluted by transferring ~50uL into 5mL of fresh S.C and grown overnight at 30 degrees. The overnight culture was diluted to

0.15 the next day and grown until the optical density (OD) reached 0.30. 1mL of this culture was spun down for 1 min @ 13000RPM, and the pellet was resuspended in 500uL of fresh S.C. Janelia Farms photoactivatable 549 (JF-PA549) was added to the 500uL culture for a final dye concentration of 50nM, except for Histone H3-Halo (YTK1434), where a concentration of 10nM was used to compensate for the higher copy number. This culture was placed in a thermomixer at 30 degrees and 500RPM for 40 minutes. After incubation, 3 wash cycles using fresh S.C were done to wash away unbound dye. After the final wash step, the pellet was resuspended in 50uL of S.C, and 3uL of the culture was placed on an agarose pad consisting of SC and Optiprep (Sigma), within a Gene Frame (Thermo Scientific). The pad was made by taking a 2% agarose Optiprep mixture (0.02g in 1mL Optiprep) - that was heated to 90 degrees - and mixing 500uL with 500uL 2xSC, resulting in a 1% agarose 30% Optiprep SC mixture. Approximately 140uL of this mixture was placed within the Gene Frame, with excess being removed with a KimWipe. Prior to imaging, we waited ~15 minutes to let any unbound dye be released.

Coverslips were cleaned with the following steps: 1) Place in 2% VersaClean detergent solution overnight. 2) Wash with MilliQ water 3x. 3) Sonicate in acetone for 30 minutes. 4) Wash with MilliQ water 3x. 5) Place in methanol and flame coverslips using Bunsen burner. 6) Place in Plasma Etch plasma oven for 10 minutes.

Microscopy was done at 23 degrees, on a Leica DMi8 inverted microscope with a Roper Scientific iLasV2 (capable of ring total internal reflection fluorescence (TIRF)), and an Andor iXON Ultra EMCCD camera. An Andor ILE combiner was used, and the maximum power from the optical fiber was 100mW for the 405nm wavelength, and 150mW for the 488nm and 561nm wavelengths. The iLasV2 was configured to do ring highly inclined and laminated optical sheet (HILO), for selective illumination and single-molecule sensitivity. Metamorph was used to control acquisition. A Leica HCX PL APO 100x/1.47 oil immersion objective was used, with 100nm pixel size. Any z-stacks were doing using a PInano piezo Z controller**.**

Single-particle photoactivated localization microscopy (sptPALM) experiments were performed by activating molecules with low power (0.5% in software) 405nm light to photoactivate ~1 molecule/cell, followed by stroboscopic, long-exposure (500ms) illumination with 561nm light (5% in software) to image primarily bound molecules. A brightfield image and a z-stack of 6um (0.3um step size) in the 488nm channel, was taken before and after each timelapse, to ensure normal cell health and to find nuclei.

4.5.4 Tracking analysis

Tracking was done with Trackmate [128]. Spots were localized using the Laplacian of Gaussian (LoG) method, with an estimated spot size of 2.5 pixels, with the exception of the *E.coli* experimental data where it was set to 5 pixels. The intensity threshold was set a bit lower to prevent track fragmentation due to intensity fluctuations. The linear assignment problem (LAP) algorithm was used to form tracks with costs on quality ranging from 0.1-0.5. We set a gap frame of 1 to allow temporary disappearance of the molecule, and track merging and splitting was allowed in cases where multiple molecules crossed paths with one another.

To isolate tracks found only in cells/nuclei, we used the binary images to locate tracks whose mean positions coincided with values of 1 in the binary image.

4.5.5 Machine learning and track analysis

All machine learning and subsequent analysis for estimation of residence times was done using Matlab.

To construct training data sets, we had binary classification, with a value of 0 assigned to false positive/diffusing molecule, and a value of 1 to a track representing a bound molecule. We manually looked at the raw image data to determine if the molecule appeared immobile.

For the learning procedure, the "TreeBagger" function in Matlab was used, representing the random forest algorithm. The hyperparameters that were adjusted were: InBagFraction (representing the fraction of the training data given to each tree), MinLeafSize (minimum leaf size), NumPredictorstoSample (number of predictors to sample at each node), and NumTrees (the number of trees to construct). The hyperparameters were adjusted until the best OOB error was achieved.

For GMM fitting, the expectation-maximization (EM) algorithm was used.

After the final classification, we analyzed the tracks to extract residence times.

We then fit the track durations of the remaining tracks with a truncated exponential model, to compensate for discarding short duration tracks, using Maximum Likelihood Estimation (MLE) through Matlab's "mle" function, to calculate the mean track duration.

2)

$$PDF = \left(\frac{1}{\tau}\right) e^{\frac{-(x-L)}{\tau}}$$

where τ is the mean track duration, and L is the truncation point. For photobleaching controls, this was equivalent to estimating the mean bleach time.

The 95% confidence intervals were calculated by bootstrapping 1000 samples. Bound times were calculated using the following equation:

3)

$$Tbound = Ttrack * Tbleach/(Tbleach - Ttrack)$$

To calculate the errors on the estimate, we performed bootstrap sampling on the track durations to the following equation:

4)

$$(\frac{1}{T_{track}})e^{-t/T_{track}} = \left((\frac{1}{T_{bound}}) + (\frac{1}{T_{bleach}})\right) e^{-((\frac{1}{T_{bound}})+(\frac{1}{T_{bleach}}))t}$$

With 10% variation allowed for the T_{bleach} estimate, in order to obtain biologically sensible results.

To check for two-exponential mixtures, the track durations were fit with the following two-exponential model:

5)

$$p\left(\frac{1}{\tau_1}\right) e^{\frac{-(x-L)}{\tau_1}} + (1-p)\left(\frac{1}{\tau_2}\right) e^{\frac{-(x-L)}{\tau_2}}$$

where τ_1 = ($Tbleach$ + $Tbound_\gamma$)/($Tbleach$ *$Tbound_\gamma$), τ_2= ($Tbleach$ + $Tbound_\psi$)/($Tbleach$ * $Tbound_\psi$), p is the mixture proportion, and L is the truncation point.

The lower and upper bounds on the two binding timescales were 0.0001s and 6000s, respectively, while allowing for a 10% variation in the bleaching estimate.

To check for overfitting and to identify whether the two-exponential model significantly fit the data better, we used the BIC test and the Loglikelihood ratio (LLR) test, as described in [15]. We looked for cases when the two-exponential model estimates did not simply return the lower and/or upper bounds as this would indicate that no physically sensible solution was found.

4.6 Acknowledgements

We thank Tatiana Karpova (NCI) for providing the yeast strains carrying H3-Halo and pdr5Δ;Luke Lavis (Janelia Farms) for providing the Halo substrate coupled to PA-JF549 dye. **Funding:** The experiments were done using equipment from the Integrated Quantitative Bioscience Initiative (IQBI), funded by CFI 9. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC# 435521-2013), the Canadian Institutes for Health Research (CIHR MOP# 142473), the Canada Foundation for Innovation (CFI# 228994), and the Canada Research Chairs program.

4.7 Author contributions

NK, Conceptualization, Formal analysis, Investigation, and Writing. ZWEH, Investigation, RR-L, Supervision, Funding acquisition.

Chapter 5

Processive Activity of the Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells

This chapter is based on a submitted manuscript: Kapadia, N., El-Hajj, Z.W., Beattie, T.R., Yu, A., and Reyes-Lamothe, R. Processive activity of the replicative DNA polymerases in the replisome of live eukaryotic cells.

This chapter is the culmination of the work I did in Chapters 3 and 4, where I use the techniques from those chapters to ask whether the eukaryotic replisome behaves in a similar manner to bacterial replisome, with respect to the dynamics of individual replisomal subunits. As discussed in Chapter 2.5, there has not been much progress in single-molecule microscopy in budding yeast. Therefore, a significant amount of work went into establishing budding yeast as system for these types of experiments. Surprisingly, we found that the subunits of the budding yeast replisome showed stable binding which was quite unexpected for Pol δ, given previous models suggested it had no physical link to the replisome. Our results suggest that while the eukaryotic replisome may not have a dynamic composition as the bacterial replisome, it might share a similar architecture, with both the leading and lagging strands physically connected to the replisome. This would also help explain synchronized synthesis (equivalent rates of synthesis on leading and lagging strand), observed in budding yeast.

5.1 Abstract

DNA replication is carried out by a multi-protein machine called the replisome. In *Saccharomyces cerevisiae*, the replisome is composed of over 30 different proteins arranged into multiple subassemblies, each performing distinct activities. Synchrony of these activities is required for efficient replication and preservation of genomic integrity. How this is achieved is particularly puzzling at the lagging strand, where current models of the replisome architecture propose turnover of the canonical lagging strand polymerase, Pol δ, at every cycle of Okazaki fragment synthesis. Here we established single-molecule fluorescence microscopy protocols to study the binding kinetics of individual replisome subunits in live *S. cerevisiae*. Our results show long residence times for most subunits at the active replisome, supporting a model where all subassemblies bind tightly and work in a coordinated manner for extended periods, including Pol δ, hence redefining the architecture of the active eukaryotic replisome.

5.2 Introduction

A crucial step during cell proliferation is the duplication of the genome, composed of long DNA molecules. DNA replication is carried out by a specialized multiprotein machine called the replisome. In all organisms, the replisome accomplishes a few basic functions: progressive unwinding of the double-stranded DNA, the synthesis of an RNA molecule that serves as primer, and the polymerization of DNA. In eukaryotic cells, unwinding is catalysed by the CDC45/MCM2-7/GINS (CMG) helicase. Pol α carries out the RNA primer synthesis at the lagging strand – whose synthesis proceeds in an opposite direction to DNA unwinding – and extends primers by its DNA polymerase activity. Pol ε and Pol δ serve as the replicative DNA polymerases, synthesizing most of the DNA on the leading and lagging strands, respectively, and while Pol ε is thought to interact stably with CMG, Pol δ does not appear to have a stable interaction with the replisome [41, 74]. In addition to these basic functions, the replisome contains multiple other proteins serving structural or other specialized roles, such as Ctf4, which is thought to act as a recruiting hub for other proteins, including Pol α (**Fig. 5.2A**) (reviewed in [5]).

It is thought that a stable architecture of the replisome helps ensure that the basic activities occur in an orderly and synchronized manner, thus limiting the accumulation of single-strand DNA (ssDNA) near the sites of replication, which can be a source of genome instability [209, 210].

However, plasticity of the architecture over time – through dissociation of subunits – has been observed in bacteria [211-213], leading to questions on how this balance of synchronized work and dynamic behaviour of individual subunits is achieved. *In vitro* studies of the binding kinetics of individual subunits of the replisome have been limited in the past as they could not take into account the great number of components and the presence of post-translational modifications, both of which can influence the strength of interaction between proteins and replisome function. Here we characterized the binding kinetics of eukaryotic replisome subunits in live cells, using single-molecule fluorescence microscopy, to gain insight into replisome function and architecture.

5.3 Results

5.3.1 Single-molecule tracking in budding yeast with machine learning classification

We first established a set of experimental and analysis protocols for the use of single-molecule fluorescence microscopy in budding yeast. We used a strain carrying a fusion of the histone H3 to HaloTag (H3-Halo) [134, 214] as a control, due to the expected long-residence binding of H3 on chromatin [202]. Using H3-Halo we characterized the rate of bleaching of the fluorophore, the localization linking distance for tracking chromatin-bound proteins in our timelapses, and the traits of a single-fluorescent molecule in our system, which we exploited for robust classification using a machine-learning (ML) approach (**Figure 5.1B-E**). For detection, we used a HaloTag ligand coupled to the cell permeable photoactivable dye PA-JF549 [141]. The imaging protocol used to track single-molecules is similar to that previously described for *Escherichia coli* [211], where a single exposure to a low dose of 405nm activating light was followed by multiple events of 561nm-excitation at fixed intervals. However, here we did multiple rounds of activation-imaging to maximize the sample size (**Figure 5.1A**). Camera integration times of 500 milliseconds resulted in the motion-blurring of diffusive molecules, facilitating the tracking of chromatin-bound molecules represented as foci. All our imaging was done using HILO illumination to increase the signal-to-noise ratio [183]. In addition, to minimize light refraction caused by the cell wall [132], we did our experiments in presence of 30% of the refractive media-matching Optiprep [203] (**Figure A3.4**). Unbiased identification of chromatin-bound single-molecules was done using a random forest-based machine learning algorithm, which used multiple traits of molecular movement and intensity for classification (**Figure 5.1C-D**) (**Materials and Methods**).

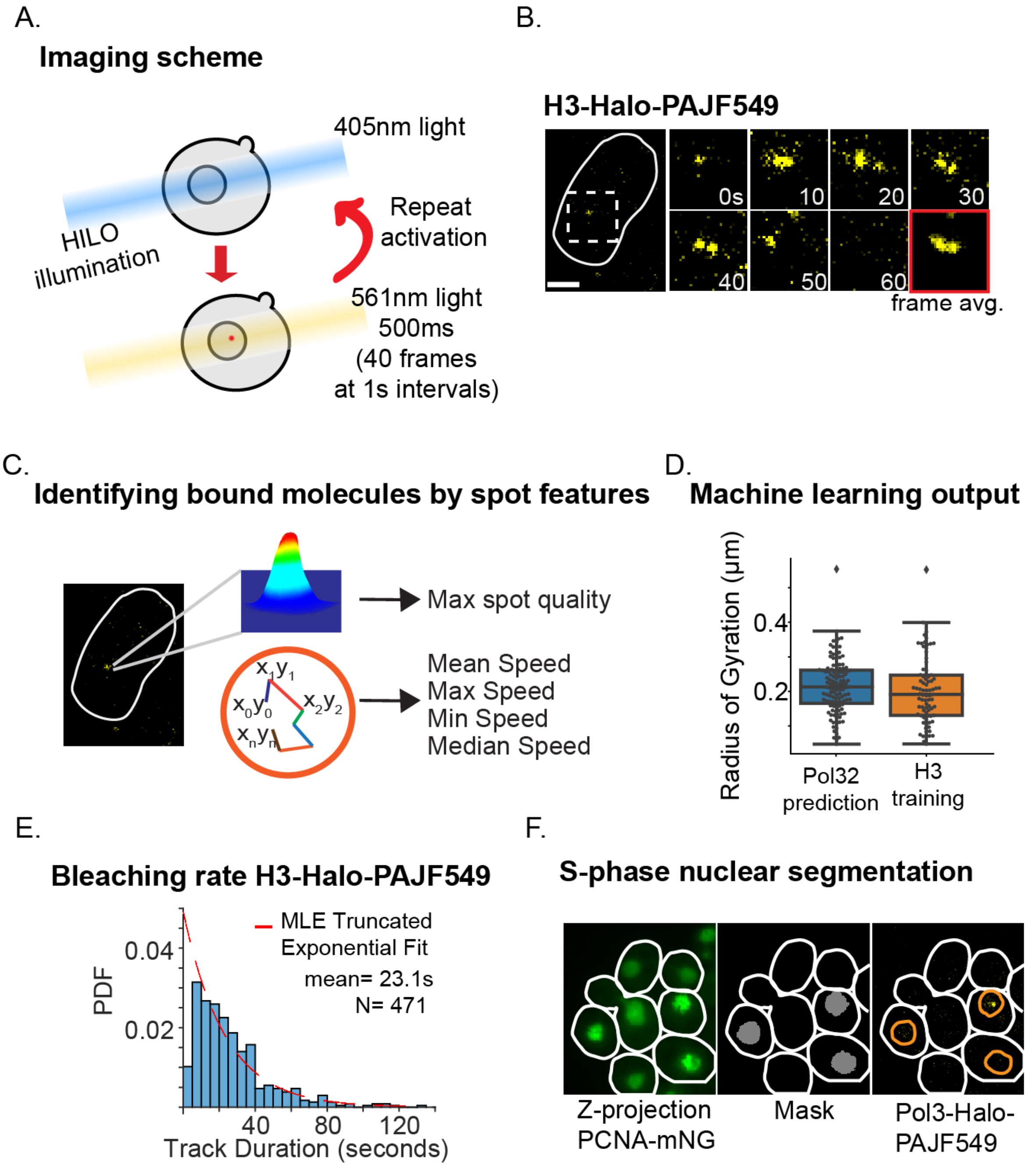


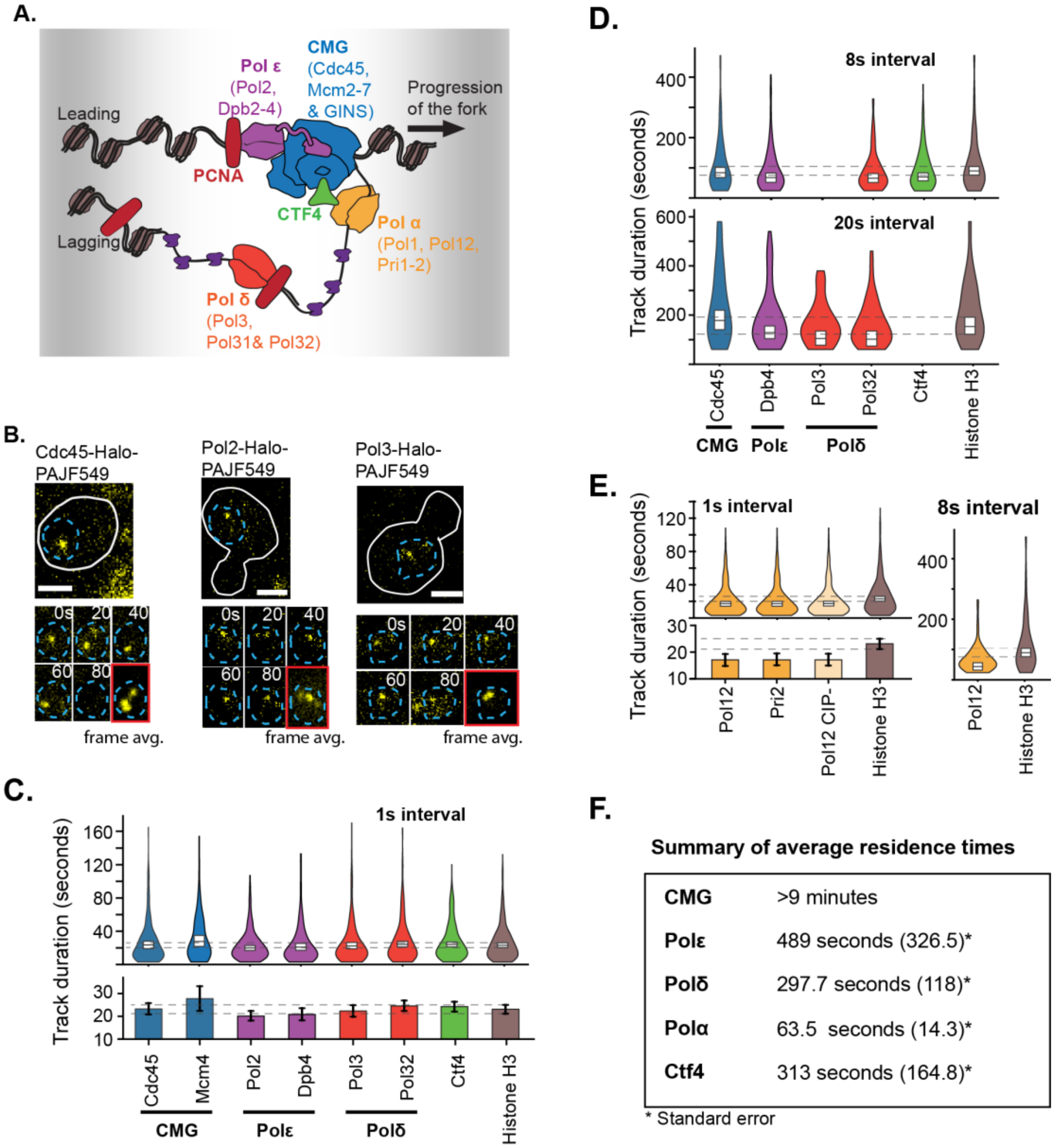
Figure 5.1: Single-molecule tracking in the yeast nucleus. (**A**) To track single-molecules, we exposed Halo-tagged yeast cells to a single exposure of low-intensity 405nm light, resulting in the stochastic activation of on average one molecule per nucleus. This was followed by time-lapse imaging using Highly Inclined and Laminated Optical sheet (HILO) illumination and 561nm light for fluorescence detection. Multiple rounds of activation and imaging were performed to increase sample size. (**B**) Representative images of single-molecules of H3-Halo-PAJF549 in live cells. The image on the left shows the entire cell. Images on the right show a magnified view of the

square drawn on the left image, and show individual timepoints or a frame average of the experiment over 60 seconds. Scale bar 2 μm. (**C**) Information on the intensity and speed of single molecules of H3-Halo-PAJF549 was extracted and used to identify chromatin-bound molecules of replisome subunits. (**D**) Example of the classification outcome of the machine-learning algorithm used to identify chromatin-bound molecules. The radius of gyration of individual molecules is shown for the training data of H3, and for molecules of Pol32 and H3 classified as bound. (**E**) Characteristic bleaching rate for the fluorophore PAJF549 using H3-Halo as control, whose molecules should bleach before unbinding. (**F**) Representative image of the procedure followed to segment nuclei of cells undergoing S-phase. Smooth Manifold Extraction (SME) was used to project the intensity of a Z-stack of PCNA-mNeonGreen covering the whole cell while preserving spot-like features. S-phase nuclei were selected based on high heterogeneity in the distribution of intensities.

5.3.2 Stable binding of eukaryotic replisome subunits

We applied these newly developed strategies to evaluate the residence times of multiple replisome subunits. We expected long residence times for the CMG helicase and the leading-strand Pol ε. In contrast, we expected much shorter binding for Pol δ, consistent with the use of a different copy of the polymerase for the synthesis of each Okazaki fragment. We constructed and characterized haploid strains carrying HaloTag fusions of replisome subunits as single gene copies at their original chromosomal location (**Figure A3.5-3.7**). Analysis was done in cells undergoing S-phase, as judged by the heterogeneity of intensity within the nucleus of PCNA-mNeonGreen (**Figure 5.1F**). Using 1-second intervals in a single-imaging plane, we observed lifetimes of fluorescent foci that were indistinguishable from H3-Halo (our bleaching control) in cells carrying Cdc45-Halo and Mcm4-Halo (**Figure 5.2B-C**) (**Table A3.5-3.7**), consistent with a reported tight binding of CMG at chromatin [215]. Similarly, Pol ε (Pol2-Halo and Dpb4-Halo) and the Ctf4 subunit – both of which have a direct interaction with the CMG helicase [41, 75] – exhibited lifetimes similar to those of the CMG complex, suggesting they interact tightly with it. Unexpectedly, Pol3-Halo and Pol32-Halo, two subunits of Pol δ, also showed long track durations indicative of stable binding (**Figure 5.2C**). This was surprising since there is no reported connection between Pol δ and the CMG helicase.

We corroborated and extended the results above by doing similar experiments using longer time intervals between acquisitions. To reduce the likelihood of chromatin-bound molecules moving out of the focal plane, we imaged three and four different planes separated by 500nm, for 8s and 20s intervals, respectively, and performed maximum intensity projections (MIP) on the resulting z-stacks. These experiments suggest that the CMG helicase, Ctf4, Pol δ, and Pol ε, remain bound to chromatin for an average time that exceeds 5 minutes (**Figure 5.2D**) (**Materials and Methods**).



Summary of average residence times

CMG	>9 minutes
Polε	489 seconds (326.5)*
Polδ	297.7 seconds (118)*
Polα	63.5 seconds (14.3)*
Ctf4	313 seconds (164.8)*

* Standard error

Figure 5.2: Binding kinetics of replisome subunits. (**A**) Diagram of the architecture of the replisome in *S. cerevisiae*. (**B**) Images for long-lived single-molecule fluorescent foci in strains carrying replisome fusions to Halo-PAJF549. Images on the top show the entire cells. Images on the bottom show a magnified view of the nuclei and show individual timepoints or a frame average of the experiment over 80 seconds. Scale bar 2 μm. (**C**) Distribution of track durations (spot lifetimes) for different replisome subunits imaged at intervals of 1 second (top). White boxes

represent the boundaries of confidence intervals of the mean (represented as the line in the box). Dashed lines, representing the confidence intervals for Histone H3, are drawn for comparison. Colours correspond to the diagram shown in A. Histogram showing only the average and confidence intervals to facilitate comparison (bottom). (**D**) Distribution of track durations for the same subunits as in C using intervals between pictures of 8 or 20 seconds. (**E**) Distribution of track durations for subunits for Pol α and comparison with histone H3. Plot on the top left shows the distribution obtained using 1-second intervals. Histogram is shown for these same distributions so that the difference with H3 can be appreciated (bottom left). The distribution of track durations using 8-second intervals is shown on the right. (**F**) Summary of the residence times obtained. With exception for CMG and Ctf4, the weighted averages were used from different time intervals and different subunits in the case of Pol δ and Pol α. For all subcomplexes except Pol α, the track durations were very close to the bleaching duration, resulting in large errors. For CMG, a lower-bound estimate was calculated from data obtained with the 8-second intervals, since the track duration was nearly identical to the bleaching duration.

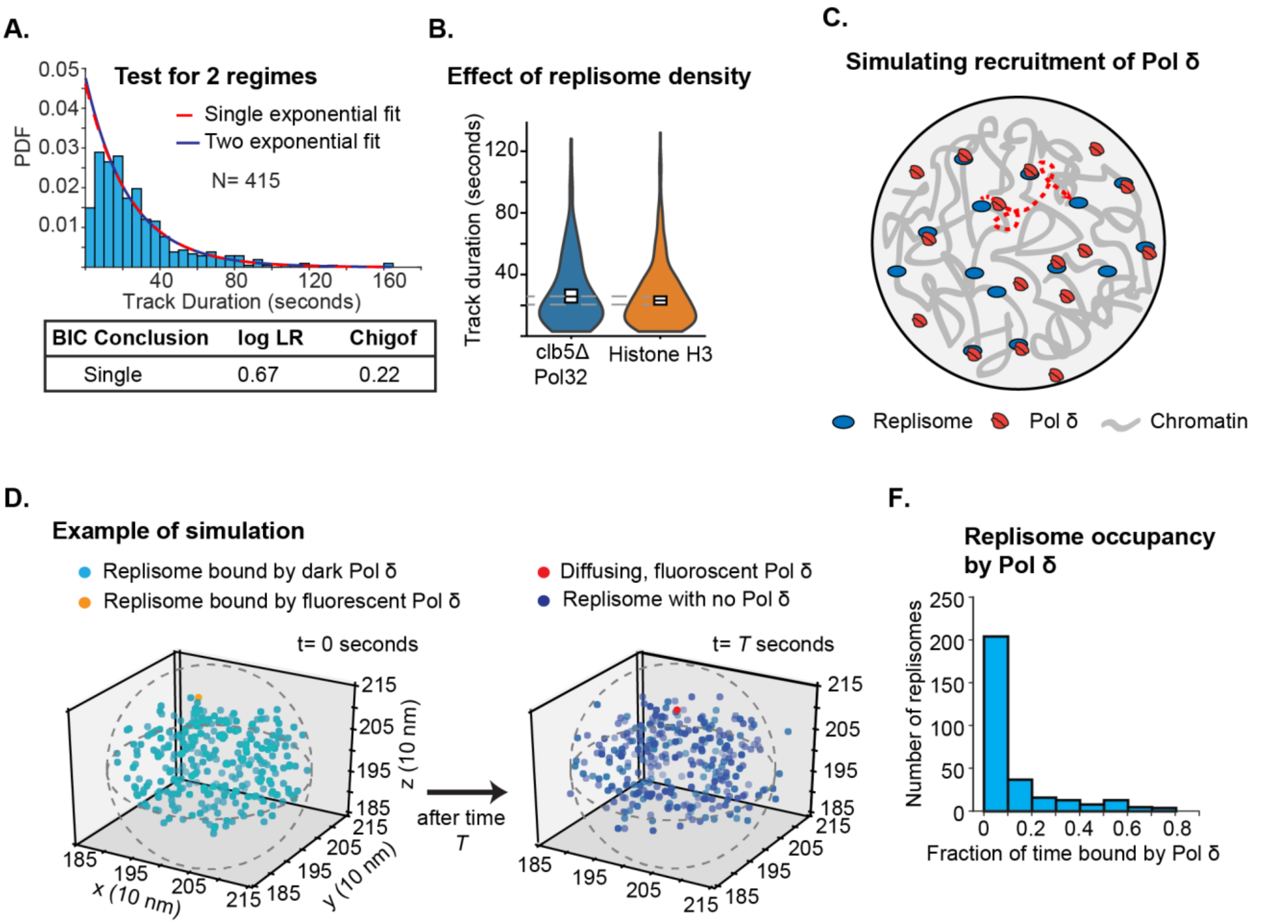
Note that the similarity to the bleaching control results in a high uncertainty in our estimates. Furthermore, given that our measurements originate from replisomes at varying stages between initiation and termination, the real residence times may be longer than our estimates. To put these numbers into perspective, the average segment replicated by an individual replisome, defined as half the length of a replicon, is 18.5kb [216]. At an average replication rate of 1.6 kbp/min [67], the time for synthesizing this segment is ~11 minutes. Our estimates approach the time required for completion of synthesis, making it likely that some copies of the DNA polymerases remain associated to the replisome from initiation to termination.

Pol α was the only subcomplex that exhibited a dynamic behaviour in our experiments. Using 1-second intervals between consecutive pictures, both Pol12-Halo and Pri2-Halo had track durations shorter than the bleach time (**Figure 5.2E**). This observation was corroborated by results using 8-second intervals for Pol12-Halo. A weighted average of the two subunits and intervals results in an estimated residence time of 63.5 seconds**,** indicating that Pol α can perform multiple cycles of priming before unbinding (**Figure 5.2F**). To test if the binding kinetics of Pol α reflect a dynamic binding to the stably-bound Ctf4, we repeated the experiments of Pol12-Halo in a mutant strain of Pol1 lacking the Ctf4-interacting peptide (CIP), termed Pol12-Halo CIP- (**Figure 5.2E**). Surprisingly, we observed no significant difference in the CIP- strain when compared to the wt strain, suggesting that the interaction with Ctf4 is not essential for retaining Pol α at the replisome.

This is consistent with *in vitro* results that suggest that Ctf4 does not retain Pol α at the replisome [97].

5.3.3 Rebinding of Pol δ does not explain long residence time

Four different models can explain the observed stable-binding of Pol δ: first, stable binding at the end of Okazaki fragments; second, usage at both leading and lagging strands; third, efficient recycling and quick rebinding in replisome-dense nuclear regions; fourth, tight binding to the replication fork. Slow recycling after completion of the Okazaki fragments, possibly through binding to PCNA, is incompatible with the estimated copy numbers of the Pol δ subunits. Considering that at the peak of S-phase there are ~300 replisomes per nucleus [217], and that the estimated copy number of Pol31 is ~2200 (reportedly the least abundant subunit)[80], cells would only be able to undergo up to 9 cycles at the lagging strand, spanning about 1 minute (165bp Okazaki fragments x 9 / 1.6kbp min^{-1} rate), before depleting the available Pol δ. The second model, where Pol δ works at both strands, would potentially result in two different binding regimes: long residence and short residence times at leading and lagging strands, respectively. However, we observe no evidence for the presence of two regimes in our data (**Figure 5.3A and Table A3.3**). The third model considers that Pol δ efficiently rebinds chromatin after completion of an Okazaki fragment. This idea is especially compelling due to the high density of replisomes at mid S phase. To test this model, we artificially reduced the replisome density by using a strain carrying a deletion of clb5, which should result in the inactivation of about half of the origins of replication [218] (**Figure A3.6**). However, we did not observe a significant difference in the lifetimes of Pol32-Halo when comparing wt and clb5Δ cells, suggesting that fast rebinding cannot explain the long residence times observed (**Figure 5.3B**). We also performed computer simulations of Pol δ binding in the context of high density of replisomes at mid S-phase and found no evidence of frequent rebinding under our experimental and biologically sensible parameters (**Figure 5.3C-E and Table A3.4**) (**Materials and methods**). Furthermore, these simulations indicated that a model of transient Pol δ, where each copy dissociates every 4 seconds after completion of an Okazaki fragment, would lead to many replisomes being left without a Pol δ for a significant fraction of time (**Figure 5.3F-G**). Hence, we propose that a single copy of Pol δ synthesizes multiple Okazaki fragments. This model would explain how the eukaryotic replisome achieves near identical rates of synthesis between the leading and lagging strand (discussed in [4]).



BIC Conclusion	log LR	Chigof
Single	0.67	0.22

E.

Expected spot lifetimes with a 4s residence time of Pol δ

$D_{Pol\delta}$ (μm²/s)	$D_{Replisome}$ (μm²/s)	T_{Track} [CI]
0.5	0.005	2.89 [2.00, 3.79]
0.5	0.05	3.17 [2.19, 4.16]
5	0.005	4.15 [2.86, 5.44]
5	0.05	2.94 [2.03, 3.85]

G.

Mean replisome occupancy by Pol δ

$D_{Pol\delta}$ (μm²/s)	$D_{Replisome}$ (μm²/s)	Mean Occupancy
0.5	0.005	0.1207
0.5	0.05	0.1275
5	0.005	0.1219
5	0.05	0.1269

Figure 5.3: Testing for recycling of Pol δ. (**A**) Statistical testing of track durations obtained from Pol32 with 500ms interval, to check if a more complex model fits better than a single-exponential, indicating that recycling may be frequent or if there are multiple binding behaviours at the replisome. The tests used were the loglikelihood ratio (LLR) test and the Bayesian information criterion (BIC) test. (**B**) Comparison of track durations obtained of Pol32 in a clb5Δ background with a 1s interval, and from the histone H3 data. (**C**) Diagram for recycling of Pol δ among multiple replisomes. (**D**) Diagram of general simulation setup zoomed into the sphere where all the replisomes were initially placed. We used a 4-second residence time for Pol δ to account for a model where a different copy is needed for synthesis of each Okazaki fragment (**Materials and**

Methods). Initially, 300 replisomes were placed inside a sphere that was 300nm in diameter to represent a diffraction-limited region, represented by the dotted line. This unnaturally high density was used to maximize the probability of recycling. The replisomes are subsequently allowed to move outside the initial sphere and throughout a nucleus of diameter 1.5μm. Unbound, non-fluorescent copies of Pol δ are not shown for clarity. **E)** Track durations obtained from simulations, over a range of diffusion coefficients. The initial fraction of replisomes bound by Pol δ was 0.75. The estimated track duration for a single unbinding event was estimated to be 3.41 seconds, based on a 4s residence time and a 23s bleach time. **F)** Representative histogram of the fraction of time replisomes had a bound Pol δ, over the course of 200s. **G)** Results for mean occupancy of replisomes (i.e. the average fraction of time replisomes had a bound Pol δ), over a range of diffusion coefficients.

5.4 Discussion

The structural, biophysical, and biochemical characterization of replisome architecture in bacteria shows a link between their replicative DNA polymerases and their helicase, and that there is some level of synchrony between helicase unwinding and polymerase activity [162, 219]. This is presumably important to limit the generation of ssDNA during elongation of the strands. Here we provide evidence that supports a similar architecture for the eukaryotic replisome. We hypothesize that a yet uncharacterized protein-protein interaction bridges Pol δ and the CMG helicase (**Figure 5.4**). It is possible that interaction with Pol α may help to retain Pol δ at the replisome [220]; however, Pol α does have a shorter residency time than Pol δ, indicating that there are likely other important interactions retaining Pol δ at the replisome. In humans an interaction between Pol δ and AND1, the orthologue of Ctf4, has been described [221], which if present in budding yeast may contribute to the longer residence times. Future work will define how Pol δ is kept at the replication fork.

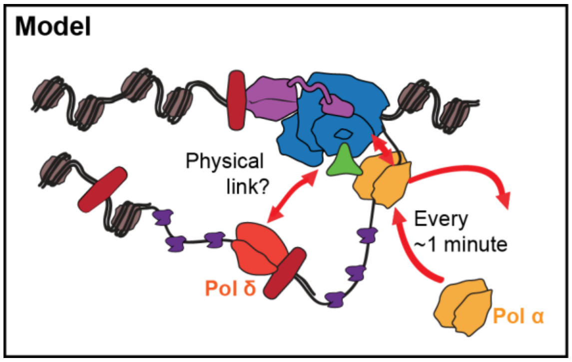


Figure 5.4: Model for the architecture of the eukaryotic replisome. The CMG helicase, Pol ε and CTF4 bind tightly to chromatin and act processively. In addition, our data suggest the existence of a direct or indirect interaction between Pol δ and the stably bound CMG helicase. Pol δ acts processively, with a single copy of this polymerase potentially synthesizing all the Okazaki fragments of an individual replisome from initiation to termination. Pol α interacts with the CMG independently of Ctf4 and turnovers at rates that permit it to synthesize multiple primers at every binding event. The function of the interaction between Pol α and Ctf4 at the replisome is unclear.

Long residence times of both the leading and lagging strand DNA polymerases in yeast contrast with the fast dynamics reported for the bacterial replisome [211-213], where DNA Pol III is exchanged every few seconds. However, the rate of replication, which is about 30 times slower in budding yeast compared to *E. coli*, would reduce the rate of accumulation of torsional stress near the fork – a possible factor triggering unbinding of the polymerases in bacteria [66] – and allow time for its dissipation by other means. Alternatively, strategies unique to eukaryotes may result in a similar transient binding of the DNA polymerases to DNA that do not require complete unbinding from the helicase. For example, despite the stable binding of Pol ε to CMG, its flexible linker potentially allows it to unbind from DNA while remaining bound to the rest of the replisome [222]. While our results suggest that the polymerases are stable at the replisome, they do not discard the possibility of them being dynamic on DNA, due to the resolution limit of our microscope.

S. cerevisiae continues to be a powerful model organism for the understanding of basic eukaryotic cellular mechanisms due to its genetic tractability. However, the development of single-molecule imaging techniques, which have advanced the understanding and are widely applied in bacterial systems, have had limited use in this system due to higher background fluorescence and light scattering. We expect that the single-molecule protocols described here will facilitate its future use to study a wide range of questions in this organism.

5.5 Materials and Methods

5.5.1 Strains constructions

Strains used in this study are all from a BY4741 background (with the exception of the CIP- mutant, see below) and are shown in **Table A3.1**. Plasmids used in this study are derivatives of pUC18 (ColE1 origin, Ampicillin resistance); pTB16 carries mNeonGreen and a downstream NatMX marker, while pSJW01 carries the HaloTag gene with a HygB marker. Both mNeonGreen and HaloTag genes encode an 8 amino acid linker at the 5' end (sequence: GGTGACGGTGCTGGTTTAATTAAC). Plasmids were maintained in *E. coli* DH5α and were extracted by growing in LB with 100 µg/ml ampicillin then using the Presto Mini Plasmid Kit (Geneaid).

All PCRs were made using either Phusion or Q5 (NEB). PCR reaction were in a volume of 50 µl and included water, 3% DMSO, the reaction buffer, 2.5 mM of each dNTP, 0.2 µM of each primer, either 1ng of plasmid DNA (for insertions) or 1 µl of genomic DNA (for screening insertions), and 0.5 µl of polymerase.

Fluorescent fusions were made by PCR amplification from pTB16 or pSJW01 using the primers listed in **Table A3.2**. PCR products were transformed into wild-type diploid BY4743. A single colony was grown at 30°C in 5 ml yeast peptone dextrose (YPD) overnight, then diluted to a final OD_{600} of 0.1 in 10ml of YPD. Cells were taken at OD_{600}=0.5-0.6 and centrifuged at 4000 rpm for 5 min. The pellet was washed twice with 25 ml of sterile deionized water, then once with 1ml of 100mM lithium acetate. Cells were then resuspended in 240 µl of 50% PEG, then 50 µl of salmon sperm DNA (thawed at 95°C for 5min, then incubated on ice for at least 10min), 50 µl of the PCR product and 36 µl of 1M lithium acetate were added in this order. The mixture was thoroughly mixed by pipeting and incubated on a rotator at 25°C for 45min, followed by heat shock at 42°C for 30min. Cells were pelleted in a microcentrifuge, washed in 500 µl of sterile water, then resuspended in 200 µl of YPD and plated on YPD agar. After growing at 30°C overnight, the cell lawn was replica-plated onto selective YPD agar, either with 100 µg/ml cloNAT (Werner) for mNeonGreen, or with 200 µg/ml Hygromycin B (Life Technologies) for HaloTag. Transformants were screened for the presence of an insert by PCR using the indicated primers in **Table A3.2**.

Confirmed clones were then sporulated by taking 750 µl of a YPD overnight cultures, washing 4 times with 1 ml sterile deionized water, then washing once with 1 ml of potassium acetate sporulation medium (KAc), and finally resuspending in 2 ml of KAc and incubating at 25°C wish shaking. After 5 days the sporulating cultures were checked by microscopy for the appearance of numerous tetrads, then 750 µl was taken and washed 3 times in sterile water before final resuspension in 1ml water and storage at 4°C. For dissection, 45 µl of spores were treated with 5 µl of zymolase for 10 min, then tetrads were dissected on YPD plates to isolate haploids with the tagged fusion. Genomic DNA was isolated from the haploid by vortexing the cells in the presence of zirconia/silica beads, followed by phenol extraction and ethanol precipitation. The insertion site was once again amplified using the same screening primers as above, and the PCR product was sequenced to confirm that the tag and linker were both mutation-free.

The HaloTag haploids were combined with PCNA-mNeonGreen (from YTB31) and the pdr5Δ::KanMX deletion (from a haploid sporulated from YTK1414) by mating. To do this, 10 µl of water was spotted on a YPD plate, and colonies from the Mat a and Mat α haploids to be combined were mixed together into the water drop and incubated at 30°C overnight. Cells were then restreaked on an auxotrophic -Met-Lys plate on which only the mated diploid could grow. Diploids resulting from a mating were dissected as above, and eventually haploids with all three markers were isolated and used for imaging. These haploids are all listed in **Table A3.1**.

The CIP- strain YCE449 is from a W303 background and carries a Pol1-4A allele, where four amino acids are substituted with alanine at D141, D142, L144 and F147 [75]. Pol12-Halo and CIP- were combined by mating YJL10 with YCE449, then sporulating and dissecting the resulting diploid to get YAJ05.

5.5.2 Western blot

To prepare crude cell lysates, YPD cultures were grown to exponential phase (OD600=0.5) and fixed in 10% TCA. The pellet was washed in cold acetone then in Beating Buffer (8 M urea, 50 mM ammonium bicarbonate, 5 mM EDTA), then resuspended in Beating Buffer with glass beads and vortexed at 4°C for 5min. The bottom of the tube was pierced with a needle, placed inside another tube and the cell debris and buffer was centrifuged into the clean tube. This was centrifuged again and the supernatant was collected. Protein concentration was determined using a Bradford assay.

SDS-PAGE was performed using 4-20% Mini-Protean TGX precast gels (Biorad). Lysates were prepared in Laemmli sample buffer with β-mercaptoethanol. Gels were run in Tris-Glycine-SDS at 100V for 2-3h. Proteins were transferred onto a nitrocellulose membrane using Biorad's Trans-Blot Turbo system and transfer packs running the MixedMW preset program. The membrane was incubated in Blocking Buffer (5% milk and 3% BSA in TBS-Tween) for 1h. The membrane was probed with α-Halotag mouse monoclonal antibody (Promega, diluted 1:1000 in Blocking Buffer), washed 3 times in TBS-Tween, then probed with goat α-mouse HRP-conjugated secondary antibody (Promega, diluted 1:10000 in Blocking Buffer) and finally washed 3 times in TBS-Tween. The membrane was treated with Clarity Western ECL substrate (Biorad) and exposed to autoradiography film (Diamed).

5.5.3 Flow cytometry

Cells were prepared for flow cytometry by growing cultures in YPD to exponential phase (OD600=0.5) and fixing them in 70% ethanol at 4°C. Cells were washed in Tris-Cl and incubated with RNAse A at 42°C for 3h, then with Proteinase K at 50°C for 30min. Cells were centrifuged and the pellet was resuspended in Tris-Cl and stained with propidium iodide (8 µg/ml). Samples were run on a FACSCalibur (Becton Dickinson) using the following filters and detector settings: FSC E01; SSC 396V and 4.61 gain; FL2 730V and 4.10 gain. Cytometer was calibrated using parental haploid (BY4741) and diploid (BY4743) asynchronous exponential cultures.

5.5.4 Microscopy

A single colony from a YPD plate was placed in 5mL synthetic complete (SC) medium and grown with shaking at 30 degrees Celsius for ~5-6 hours. This culture was diluted by transferring ~50uL into 5mL of fresh S.C and grown overnight at 30 degrees. The overnight culture was diluted to 0.15 the next day and grown until the optical density (OD) reached 0.30. 1mL of this culture was spun down for 1 min @ 4000RPM, and the pellet was resuspended in 500uL of fresh S.C. Janelia Farms photoactivatable 549 (JF-PA549) was added to the 500uL culture for a final dye concentration of 50nM, except for YTK1434-Halo (Histone H3), where a concentration of 10nM was used to compensate for the higher copy number. This culture was placed in a thermomixer at 30 degrees and 500RPM for 40 minutes. After incubation, 3 wash cycles using fresh S.C were done to wash away unbound dye. After the final wash step, the pellet was resuspended in 50uL of S.C, and 3uL of the culture was placed on an agarose pad consisting of SC and Optiprep (Sigma), within a Gene Frame (Thermo Scientific). The pad was made by taking a 2% agarose Optiprep mixture (0.02g in 1mL Optiprep) - that was heated to 90 degrees - and mixing 500uL with 500uL 2xSC, resulting in a 1% agarose 30% Optiprep SC mixture. Approximately 140uL of this mixture was placed within the Gene Frame, with excess being removed with a KimWipe. Prior to imaging, we waited ~15 minutes to let any unbound dye be released.

Coverslips were cleaned with the following steps: 1) Place in 2% VersaClean detergent solution overnight. 2) Wash with MilliQ water 3x. 3) Sonicate in acetone for 30 minutes. 4) Wash with MilliQ water 3x. 5) Place in methanol and flame coverslips using Bunsen burner. 6) Place in Plasma Etch plasma oven for 10 minutes.

Microscopy was done at 23 degrees, on a Leica DMi8 inverted microscope with a Roper Scientific iLasV2 (capable of ring total internal reflection fluorescence (TIRF)), and an Andor iXON Ultra EMCCD camera. An Andor ILE combiner was used, and the maximum power from the optical fiber was 100mW for the 405nm wavelength, and 150mW for the 488nm and 561nm wavelengths. The iLasV2 was configured to do ring highly inclined and laminated optical sheet (HILO), for selective illumination and single-molecule sensitivity. Metamorph was used to control acquisition. A Leica HCX PL APO 100x/1.47 oil immersion objective was used, with 100nm pixel size. Any z-stacks were doing using a PInano piezo Z controller**.**

Single-particle photoactivated localization microscopy (sptPALM) experiments were performed by activating molecules with low power (0.5-2% in software) 405nm light to photoactivate ~1 molecule/cell, followed by stroboscopic, long-exposure (500ms) illumination with 561nm light (5% in software) to image primarily bound molecules. The time intervals used were 1s, 8s, and 20s, with cycles of activation every 40 time points for the 0.5s and 1s interval data, and every 15 and 10 timepoints, for the 8s and 20s interval data, respectively. For the 8s and 20s time intervals, a mini 561nm z-stack of 1um and 1.5um was done, respectively, with 0.5um step sizes, in order retain molecules in focus. Maximum intensity projections (MIP) of these stacks were used for subsequent tracking analysis. A brightfield image and a z-stack of 6um (0.3um step size) in the 488nm channel, was taken before and after each timelapse, to ensure normal cell health and to find cells undergoing S-phase through the presence of POL30-mNG foci.

5.5.5 Tracking Analysis

Aside from tracking, all analysis was done using Matlab.

Tracking was done using Trackmate [128]. First, molecules were localized in each frame using a Laplacian of Gaussian (LoG) method, with an estimated diameter of 2.5 pixels. An intensity threshold was chosen that was slightly low, to still detect molecules that were moving out of the focal plane. Subsequent classification helped discard potential false-positive tracks from analysis that will be discussed later. After localization, tracks were formed using the Linear Assignment Problem algorithm by linking molecules in consecutive frames. The linking distance was chosen based on calculating the cumulative distribution function of the step sizes from the Histone H3 data, thus providing information on the step size of chromatin-bound proteins. Essentially, we varied the linking distance when analyzing the Histone H3 data until a noticeable plateau in the cumulative distribution function (CDF) was observed. We determined the linking distance as being the step size giving the 0.95 probability value in the CDF and did this for multiple time intervals: 3 pixels for 0.5s and 1s interval, 5 pixels, and 7 pixels for 8s and 20s time intervals, respectively. A gap frame – to allow for missed localization – of 1 was used the gap-linking distance was set to 2 pixels more than the linking distance for that time interval (e.g. 5 pixel gap-linking distance when using a 3 pixel linking distance). Linking also had cost of 0.3 for the "Quality" parameter to ensure that correct molecules were linked. We also allowed for track merging and splitting with a 5 pixel distance for all time intervals, in some cases where multiple molecules were activated within the region and they were near one another. Tracks with less than four localizations were discarded as they were unreliable. Furthermore, since we are using track merging and splitting, we can exploit this property to discard tracks that are found in noisy regions or too many molecules active. In these cases, one would expect more spots localized in a track than predicted given a certain track duration (e.g. for a track duration of 20 frames, one would expect 21 spots, but may get 40). Therefore, we calculated a ratio of number of spots in track vs. number of spots estimated based on track duration and used a cut-off of 1.5 to remove tracks with too many spots localized.

To isolate tracks found in the nuclei of cells undergoing S-phase, we used the Pol30-mNG z-stacks, as PCNA is active during S-phase, thus resulting in fluorescent foci[32]. First, we performed a Smooth Manifold Extraction (SME) [204] on the z-stack, as an alternative to MIP, to distinguish more clearly Pol30-mNG foci. We then generated a binary mask of this image, resulting in all the nuclei regions have intensity values of 1, and zero elsewhere. To isolate S-phase nuclei, we used a threshold on the standard deviation of the intensities within the regions, as an indicator of heterogenous intensity caused by fluorescent foci. From the resulting binary mask, we isolated tracks found only in S-phase nuclei by calculating the mean positions of tracks and determining if those positions land in the binary mask.

For robust classification of tracks representing bound molecules, we employed a machine learning (ML) approach, using the output from Trackmate. First, we generated a training data set of ~750 tracks from the Histone H3 data, taken with 0.5s interval (**Table A3.7, date: 20180511**). Given that we used long-exposure to blur out diffusing molecules, it is quite easy to detect bound molecules from the raw timelapses. Therefore, we manually classified tracks by assigning a value of 1 to tracks representing genuine bound molecules vs 0 for tracks representing diffusing molecules or noise. Once we had the classifications, we performed the learning procedure using the algorithm, Random Forest ([200]**)**, to develop two classification models, referred to as Model 1 and Model 2, with out-of-bag errors (equivalent to estimated cross-validation error) of 0.10 and 0.05, respectively. Model 1 had the predictor variables from Trackmate: mean speed, max speed, min speed, and median speed. Model 2 had all those variables as well, but in addition, included max quality. Both models were built with 6000 trees during the learning procedure, with following hyperparameters: InBagFraction (fraction of training data set given to each tree) = 0.50, MinLeafSize (minimum leaf size) = 50, NumPredictorstoSample (number of predictors to sample at each node) = 2. We then could output the predictor variables from tracks derived from timelapses of replisomal proteins, and input them into the models, and they would classify tracks as being bound.

The motive for building two classification models was to scale the predictor variables accordingly based on different time intervals and quality of data (e.g. slightly different laser intensities), compared to the training data set. First, we calculated the mean of the mean speed and the mean of the max quality from the Histone H3 training data set. Next, we performed a Gaussian Mixture Model (GMM) fit, with two components, on the log-transformed mean speed values of a given data set. We then took the mean of the first Gaussian component, likely representing bound molecules, and with the mean of mean speed values obtained from the training data set, we scaled all 4 speed values accordingly. For example, if the mean of mean speed from the training data set was 1.71 and the mean of the first Gaussian component was 2.5 (assuming a longer time interval), then the scaling factor would be 1.71/2.5 = 0.68. We then performed classification with Model 1.

From the tracks extracted with Model classification, we again performed a GMM fit, but this time on the log-transformed max quality values (**Figure A3.1**).

We then took the mean of the second Gaussian component (likely representing genuine molecules), and used the mean of the max quality value from the training data, to scale the quality values in a similar manner as described before. This would help ensure that despite slight differences in the quality of data compared to the training data set, tracks would still be robustly classified. After scaling, we input the tracks into Model 2, as the final classification step. To confirm that ML approach can extract bound molecules we calculated the radius of gyration (R_g) of tracks from the training data set as well as tracks from Pol32 data. The R_g was calculated as follows:

1)

$$R_g = \sqrt{\left(\frac{1}{n}\right)\sum_{i}^{n}((x_i - \bar{X})^2 + (y_i - \bar{Y})^2)}$$

where n is the number of localizations of the track, $\overline{X}$ is the mean x-position, $\overline{Y}$ is the mean y-postion, and x_i and y_i, are the x and y coordinates, respectively of the ith localization.

We performed a GMM fit, with the Bayesian Information Criterion (BIC) test incorporated to prevent overfitting, on the mean intensities of tracks followed by clustering, to isolate tracks representing single-molecules and not multiple ones (**Figure A3.2**). Also, given that we use very low dye concentrations, thus <<100% dye labelling, and the very low laser power for activation, it is unlikely that we are activating multiple molecules at once.

We then fit the track durations of the remaining tracks with a truncated exponential model, to compensate for discarding short duration tracks, using Maximum Likelihood Estimation (MLE) through Matlab's "mle" function, to calculate the mean track duration.

2)

$$PDF = \left(\frac{1}{\tau}\right) e^{\frac{-(x-L)}{\tau}}$$

where τ is the mean track duration, and L is the truncation point.

The 95% confidence intervals were calculated by bootstrapping 1000 samples. Bound times were calculated using the following equation:

3)

$$Tbound = Ttrack * Tbleach/(Tbleach - Ttrack)$$

To calculate the errors on the estimate, we performed bootstrap sampling on the track durations from the combined data set (when multiple data sets of the same time interval were taken) to the following equation:

4)

$$(\frac{1}{T_{track}})e^{-t/T_{track}} = \left((\frac{1}{T_{bound}}) + (\frac{1}{T_{bleach}})\right)e^{-((\frac{1}{T_{bound}})+(\frac{1}{T_{bleach}}))t}$$

With 10% variation allowed for the T_{bleach} estimate, in order to obtain biologically sensible results.

For Cdc45, since its track durations overlapped strongly with bleaching duration, we could not calculate a reliable estimate, so we estimated a minimum bound time based on the bleach duration with 8s time interval (**Figure A3.3**). Using equation 3, we calculated track durations by varying the bound time with a fixed bleaching time. The minimum bound time we chose was the one that gave us a track duration time equal to 1 second lower than the lower bound of the CI from the estimated bleaching time.

To check for two-exponential mixtures, the track durations were fit with the following two-exponential model:

5)

$$p\left(\frac{1}{\tau_1}\right)e^{\frac{-(x-L)}{\tau_1}} + (1-p)\left(\frac{1}{\tau_2}\right)e^{\frac{-(x-L)}{\tau_2}}$$

where τ_1 = (*Tbleach* + *Tboundalpha*)/(*Tbleach* **Tboundalpha*), τ_2 = (*Tbleach* + *Tboundbeta*)/(*Tbleach* * *Tboundbeta*), p is the mixture proportion, and L is the truncation point.

The lower and upper bounds on the two binding timescales were 0.1s and 900s, respectively, while allowing for a 20% variation in the bleaching estimate.

To check for overfitting and to identify whether the two-exponential model significantly fit the data better, we used the BIC test and the Loglikelihood ratio (LLR) test, as described in [211]. We looked for cases when the two-exponential model estimates did not simply return the lower and/or upper bounds as this would indicate that no physically sensible solution was found. We also performed a chi-square goodness of fit test, under the null hypothesis that the data comes from a single-exponential distribution.

5.5.6 Simulations of Pol δ

Simulations were performed using Python 3.7.

A 3D array of 400 x 400 x 400, was generated, with each element corresponding to a 10nm x 10nm x 10nm region. The nucleus was modelled as a sphere of diameter of 1.5um. Inside the nucleus was another sphere – termed the replisome sphere – that was 300nm in diameter, to resemble a diffraction-limited region. Both of these spheres had their centers placed at position (200, 200, 200) of the 3D array. Replisomes were placed randomly within the replisome sphere but are subsequently allowed to move throughout the nucleus. A fraction of Pol δ were bound to replisomes while the unbound fraction (referred to as excess copy numbers) were placed throughout the nucleus. One of the bound Pol δ molecules was assigned to be fluorescent, in order to test for any rebinding behavior of a fluorescent molecule. To simulate diffusive behavior of the unbound fraction, we calculated the step size in each direction, by sampling from a normal distribution of mean 0, and a standard deviation of $\sqrt{2 * D * dt^{\alpha}}$, where dt is the time lag (0.05 seconds), α is the diffusion exponent (equal to 1 for diffusing Pol δ, and 0.4 for replisomes, representing chromosomal loci movement), and D is the diffusion coefficient. If the step size caused the molecule to go out of the nucleus, the step size was resampled until it remained in the nucleus. Movement of replisomes based on genomic loci movement, was assumed to be subdiffusive as explained by the Rouse model [207]. To represent the 200 seconds of acquisition time for the 1 second time interval data, we performed 200s/(0.05s/time step) = 4000 time steps. The residence time of Pol δ – assuming one copy per Okazaki fragment – was estimated to be 4s, based on the size of Okazaki fragments measured *in vivo* (~160bp) [109], and rate of synthesis of

Pol δ under physiological conditions (~50bp/s) [79], while also allowing for delays due to strand-displacement activity and engagement with PCNA.

Two sets of transition probabilities were setup; one for the fluorescence state kinetics (e.g. photobleaching), with two states (fluorescent vs. bleaching) and the other for the molecule kinetics (e.g. unbinding), with two states (bound vs. unbound) since these were assumed to be independent from one another. Binding of Pol δ to a replisome was assumed to happen when the molecule entered a region occupied by a replisome and bound molecules would move along with their respective replisomes until dissociation. The next state of the molecule after a time step, was calculated by a sampling from a multinomial distribution with proportions given by the appropriate transition probabilities.

To determine rebinding of a fluorescent Pol δ, we calculated the overall time it spent in the fluorescent, bound state, regardless of the specific replisome it was bound to. This gives us a track duration time for one molecule within a nucleus. A fraction of 0.75 of the replisomes had a Pol δ bound at the start of the simulation. We also performed with the simulation with a fraction 0.25 with no significant effect (data not shown). We repeated this simulation 40 times (representing 40 nuclei sampled) under the same parameter set, in order to obtain a distribution of track duration times, which we can fit to using MLE, to extract the average track duration. Based on the bound time we input (4 seconds), and the measured photobleaching time with 1s interval (23 seconds), we can calculate the predicted track duration using equation 3, representing a model where there is no rebinding, and compare it to the estimated average track duration from the simulations.

For the simulations regarding the occupancy of replisomes by Pol δ, they were identical to the ones described above, except we only performed one simulation per parameter set. We then calculated for each replisome, the fraction of time it was bound by a Pol δ.

5.6 Acknowledgments

We thank Tatiana Karpova (NCI) for providing the yeast strains carrying H3-Halo and pdr5Δ;Luke Lavis (Janelia Farms) for providing the Halo substrate coupled to PA-JF549 dye; and Karim Labib for the strain carrying the CIP-binding mutant of Pol1. We thank Thomas Gligoris and Stephan Uphoff for initial help working with HaloTag fusions. **Funding:** The experiments were done using equipment from the Integrated Quantitative Bioscience Initiative (IQBI), funded by CFI 9. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC# 435521-2013), the Canadian Institutes for Health Research (CIHR MOP# 142473), the Canada Foundation for Innovation (CFI# 228994), and the Canada Research Chairs program.

5.7 Author Contributions

NK, Conceptualization, Formal analysis, Investigation, Writing; ZWEH, Formal analysis, Investigation, Writing; AY, TRB, Investigation; RR-L, Conceptualization, Supervision, Funding acquisition, Writing.

Chapter 6

Conclusion

The cell is highly vulnerable during DNA replication as many factors can pose as obstacles to replication, resulting in potentially life-threatening changes to its genome. And yet, the replisome efficiently and accurately is able to meet this demanding challenge [223]. The work presented here reveals how replisomes function at the single-molecule level, in their native cellular context, within two domains of life.

The remarkable finding that *E.coli* PolIII* dissociates from the replisome every few seconds (Chapter 3) stands in contrast to our intuition of how replication would proceed, as conceivably, a replisome with subunits falling off could lead to increased exposure to ssDNA gaps, replisome disassembly, and issues with genome stability. However, the observation that dissociation is frequent only when excess copies are available to be incorporated (i.e. a concentration dependent mechanism), provides a mechanism by which the *E.coli* replisome can balance flexibility in composition to deal with potential roadblocks with high fidelity replication [167, 211, 212]. In addition, as PolIII* is the unit of exchange, this could help ensure that both strands are synthesized simultaneously, in contrast to the case where the polymerases are recruiting to each strand individually, potentially leading to increased ssDNA exposure. Nonetheless, one important question is how does the *E.coli* replisome move at a fast rate of 300nt/s with PolIII* dissociating frequently. One model is that excess PolIII* are recruited to the replisome efficiently through protein-protein interactions, creating a high local concentration, which is part of an ongoing project in the lab.

The observation that the residence times of PolIII* subunits are exponentially-distributed implies that the binding kinetics are governed by a Poisson process, emphasizing the influence of stochasticity and noise in how the replisome functions. Much has been written on stochasticity from a systems biology perspective (e.g. gene expression noise, random partitioning noise), but less so at the level of a protein complex [65, 224, 225]. Recent evidence using single-molecule methods on replisomes suggests that protein complexes function in a noisy manner with dissociation of components and stochastic sampling of different molecular pathways [64, 65, 211, 212]. This would invoke a stronger role for compensatory mechanisms (e.g. checkpoint pathways) to ensure genome integrity is preserved. Furthermore, stochastic models might require a revision – away from deterministic thinking – of the way we describe how complexes function, and expressions personifying protein activities (e.g. Protein A recruits Protein B, Protein A coordinates the activities of Proteins B and C) might not be appropriate ways to describe how they work.

Chapter 5 addressed the question as to whether what is true with the bacterial replisome is also true of the eukaryotic replisome. Surprisingly, we found that the eukaryotic replisome forms a stable complex, with both polymerases remaining bound to the replisome for at least a few minutes. This was unexpected for Pol δ, as previous models suggested it was not physically coupled to the replisome and therefore would dissociate after completion of an Okazaki fragment. Interestingly, a recent *in vitro* single-molecule study with a reconstituted budding yeast replisome, obtained results consistent with ours, providing further support to our observations [226]. Ongoing work in the lab using the proximity-based TurboID assay will hopefully elucidate potential protein interactions retaining Pol δ to the replisome [227].

A model where both polymerases are physically coupled to the replisome can help explain the identical rates of synthesis, on the leading and lagging strands, observed *in vitro* and *in vivo* [76-78]. Moreover, the results suggest that the eukaryotic replisome may share architectural features similar to the *E.coli* replisome; namely, it may also have a trombone model structure, with ssDNA loop formation on the lagging strand (Chapter 2). Thus, while the dynamics are not conserved between the two replisomes, the architectures may resemble one another.

This leads to the question of why the eukaryotic and bacterial replisomes differ in their polymerase dynamics. As highlighted in Chapter 2, bacteria and eukaryotes differ quite significantly in their chromosome structure and it could be that those differences resulted in the distinct dynamics. This includes the fact that many eukaryotic replisome proteins perform additional roles aside from replicating DNA, including histone chaperone activities, which may require different dynamic behaviours to facilitate their roles (Chapter 2). Single-molecule tracking (SMT) might be limited to address some of these questions but advances in cryo-electron microscopy may help resolve the structure of the eukaryotic replisome in the context of chromatin, which would provide insight into how histone chaperone activities are orchestrated [228]. How stochasticity influences these processes is unclear, but it may have important consequences for how preservation of epigenetic marks is achieved.

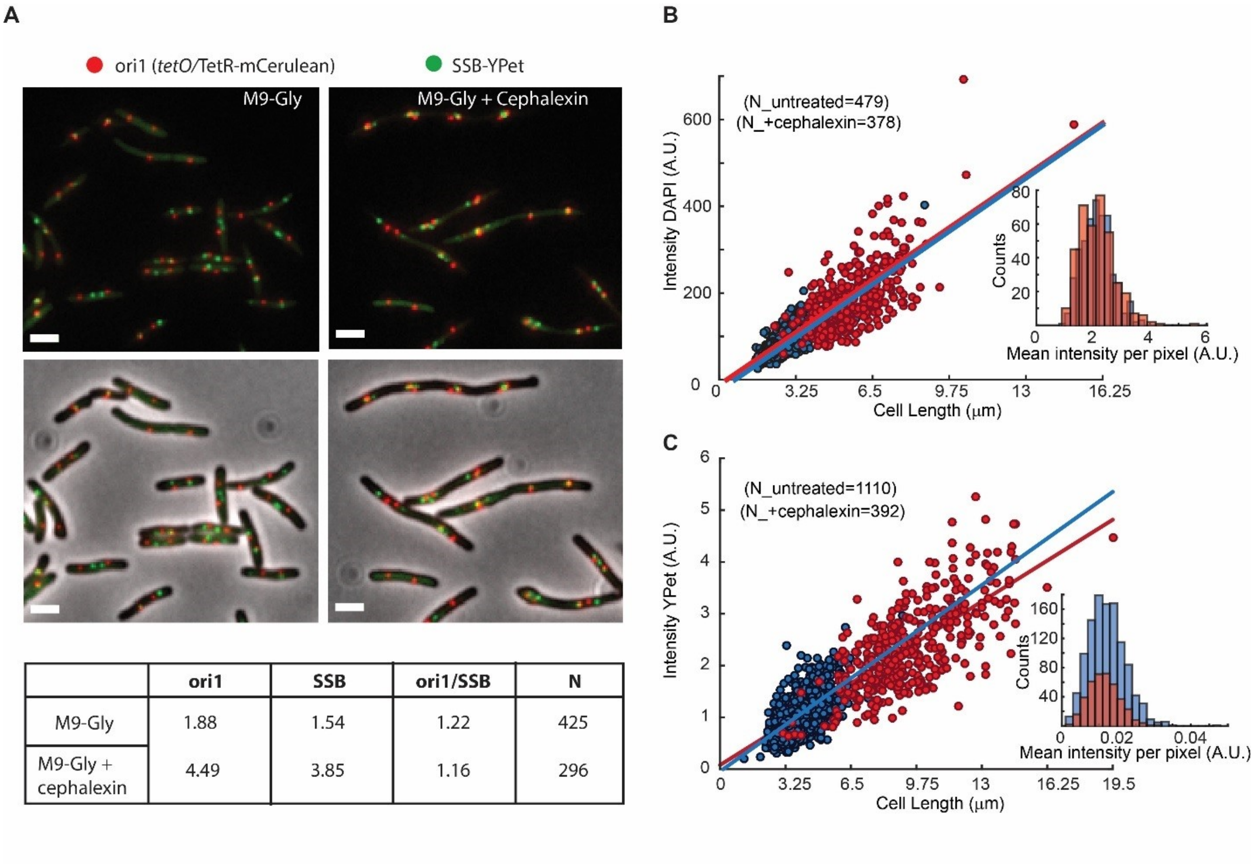
There are other questions regarding the eukaryotic replisome that we can now address with our system. For example, is MCM10 a stable component of the replisome and what are the dynamics of other replication progression complex proteins such as FACT, Top1, and MTC (Chapter 2)? We can also determine if the dynamics of the replisome change as a consequence of replication stress and DNA damage through treatment with hydroxyurea (results in the depletion of dNTPs) or UV light (formation of thymine dimers). These questions will be of great interest to characterize the entire eukaryotic replisome *in vivo*.

This work would not have been possible without SMT. However, SMT requires significant experimental expertise and extensive analysis. Chapter 4 provides a method that attempts to solve the latter issue by providing a user-friendly method to classify DNA-bound proteins that can be used in a high throughput manner for multiple data sets, which is beneficial for studying multiple proteins in a complex. We are only beginning to grasp the power of machine learning for image and data analysis, and future improvements will help with a lot of the manual, difficult, or time-consuming tasks in analysis.

In conclusion, the work presented in this thesis offers a complementary approach to traditional methods that allows us to work in parallel with biochemists, structural biologists, and geneticists to address some of the longstanding questions on replisomes in live cells. The system can potentially be adapted to other organisms, thus allowing us to illuminate how DNA replication is accomplished across the tree of life, one molecule at a time.

Appendix

A1-Supplementary Information: Frequent Exchange of the DNA Polymerase During Bacterial Chromosome Replication



	ori1	SSB	ori1/SSB	N
M9-Gly	1.88	1.54	1.22	425
M9-Gly + cephalexin	4.49	3.85	1.16	296

Figure A1.1: Artificial elongation of cells by cephalexin treatment does not interfere with DNA replication or protein expression

(A) Representative images of non-treated (M9-Gly) or cells treated with 40μg/ml of cephalexin (M9 Gly + Cephalexin) are shown. The position of the *ori1* locus (mapped at 3908kb in the chromosome), labelled by TetR-mCerulean bound to a *tetO* array, is shown in red, while the position of the replisome component SSB labelled with YPet is shown in green. Top panels do not show the phase contrast. Scale bar represents 2μm. A table summarizes analysis of this data, showing the number of *ori1* and SSB foci per cell, and the ratio between them.

(B) Distribution of the total intensity of DAPI signal against cell length in cells untreated (blue dots) and treated with cephalexin for one hour (red dots). Ethanol fixation was used to ensure homogenous permeability to the dye. Fitting to a linear model is shown in the respective colours. The inset shows the distribution of mean intensities per pixel for both conditions.

(C) Distribution of the total intensity of ε-YPet signal against cell length in cells untreated (blue dots) and treated with cephalexin for one hour (red dots). Fitting to a linear model is shown in the respective colours. The inset shows the distribution of mean intensities per pixel for both conditions.

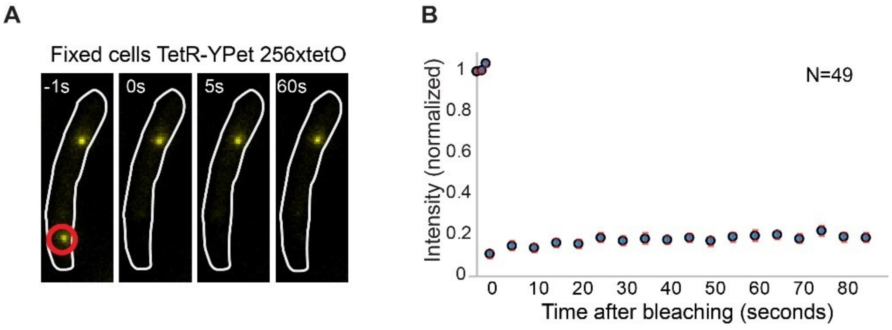


Figure A1.2: Minimal contribution of YPet photoblinking during FRAP

(A) Representative fluorescence images of a cell carrying a *tetO* operator array and expressing TetR-YPet. Cells were fixed with formaldehyde to avoid protein exchange. Cell boundaries are represented with a white line. The point of localized bleaching is shown with a red circle.

(B) Average distribution of fluorescence recovery after photobleaching of 49 cells. Note that intensity increase after bleaching is minimal (<5% of the total intensity), consistent with stochastic fluctuations and experimental measurement error. Error bars represent SE.

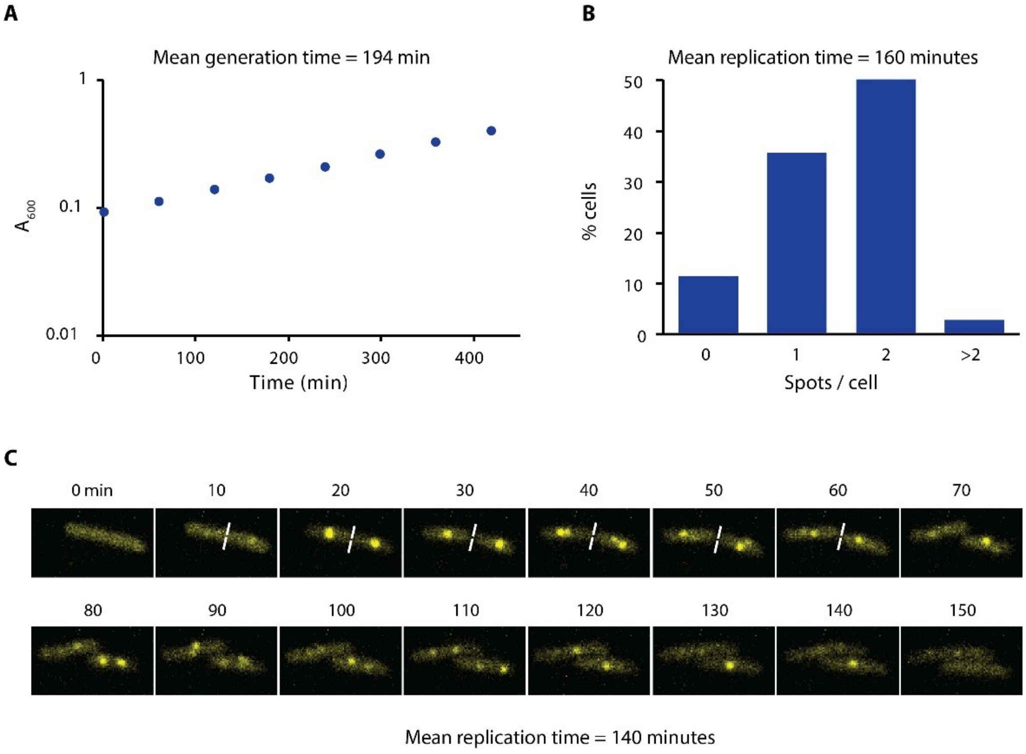
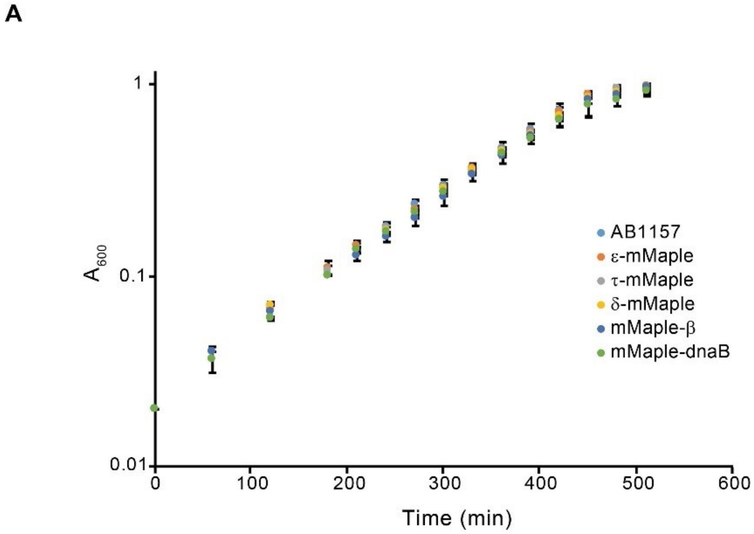


Figure A1.3: Growth rate and replication time of *E. coli* in our experimental conditions

(A) Growth curve of AB1157 in M9-Glycerol at 22ºC is shown. Samples were taken every hour for 7 hours.

(B) Distribution of the number of spots per cell of a strain carrying ε-YPet grown in M9-Glycerol at 22ºC (N=1403 cells). We estimated the replication time by taking into account the generation time and the number of cells with spots. We made the assumption that initiation of DNA replication occurs at cell birth to account for the uneven distribution of cell ages in the population.

(C) Representative images obtained from a time-lapse experiment of a strain carrying SSB-YPet grown on a 1% agarose pad in M9-Glycerol at 22ºC. Pictures were taken at 10 minute intervals. Replication time was determined from the time point of first appearance of SSB-YPet spot to its subsequent disappearance (N=56 cells). The average of the two methods (150 minutes) is reported in the main text.



B

Strain	Mean generation time	Standard deviation
AB1157	87.0	2.2
ε-mMaple	89.7	1.8
τ-mMaple	88.5	2.8
δ-mMaple	87.8	2.2
mMaple-β	85.3	2.2
mMaple-dnaB	90.4	1.4

Time in minutes

Figure A1.4: Characterisation of mMaple fusions

(A) Plot showing growth curves of the AB1157 and derivative strains carrying mMaple fusions to replisome components. Experiment was done in M9-Glycerol at 37ºC. Average and standard deviation of three experiments are shown.

(B) Table that summarizes the results from the growth curve experiments

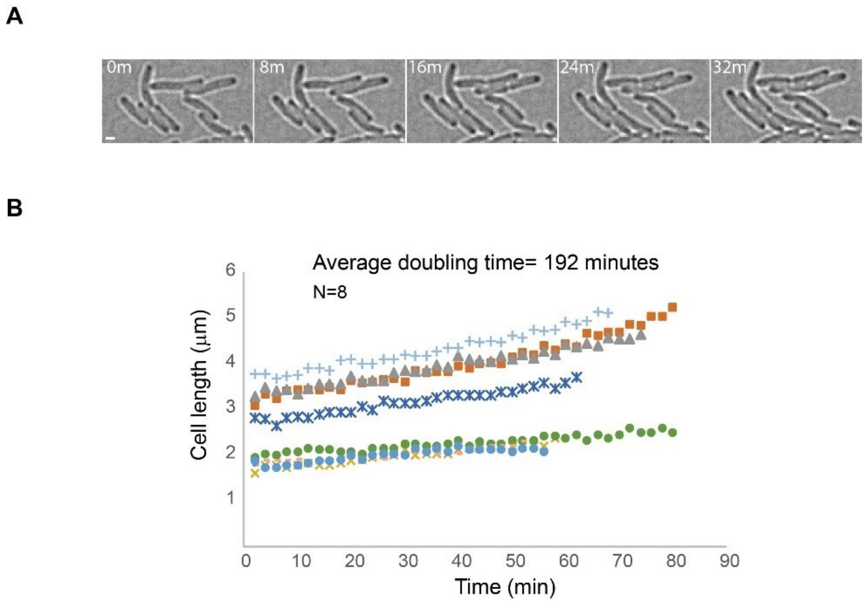


Figure A1.5: Minimal exposure to 405nm activation light allows continuation of cell growth

(A) Images obtained from an sptPALM experiment of a strain carrying mMaple-DnaB using 2s exposure times of the 561nm laser and 2 minute intervals (time in minutes). Note the growth of cells despite exposure to a single event of 405nm wavelength activation and multiple exposures to 561nm wavelength light. Scale bar = 2µm.

(B) Plot showing lengths of cells over time for 8 different cells. The average of doubling-time is similar to the generation time of AB1157 at 22°C

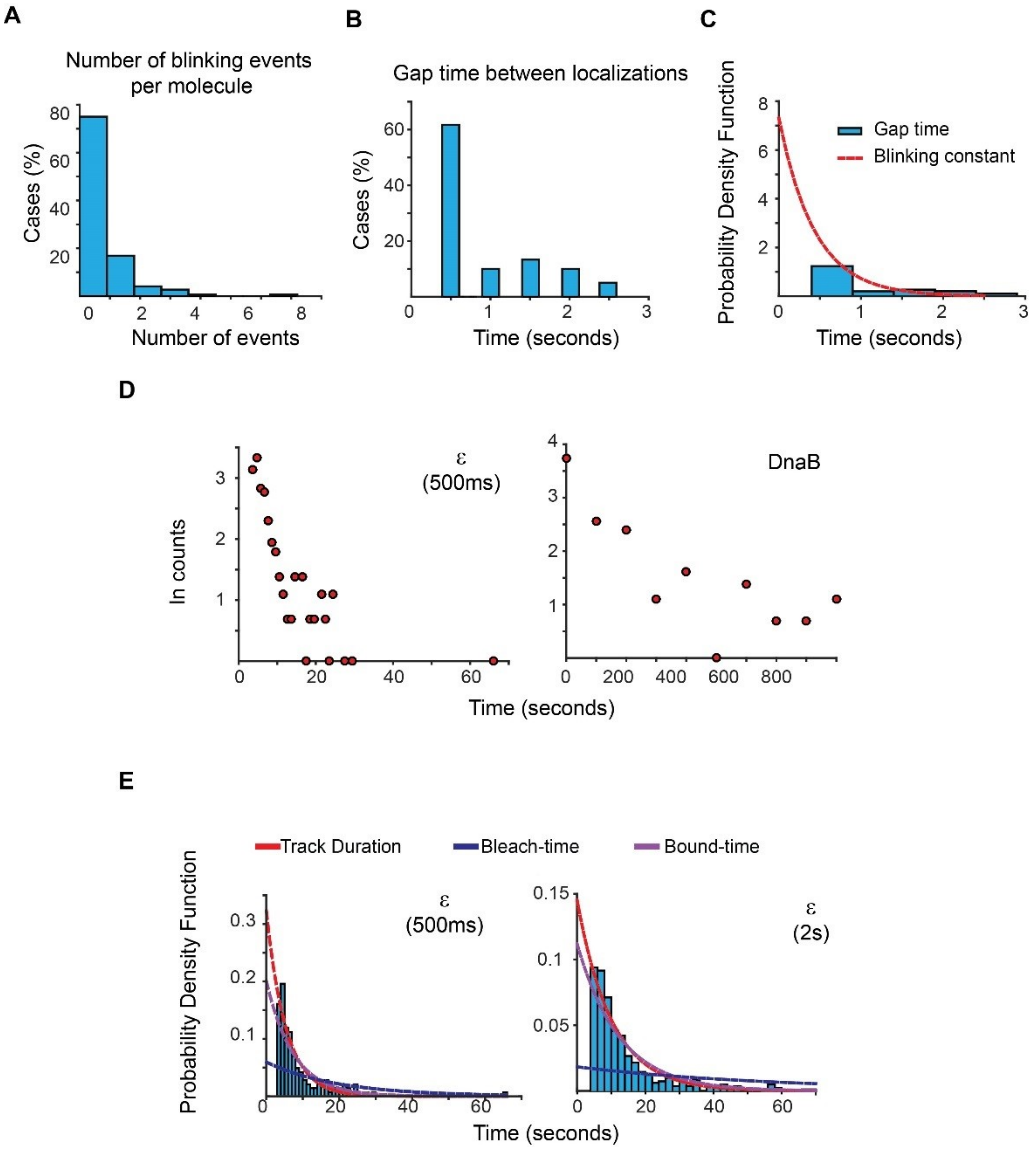


Figure A1.6: Estimation of photoblinking, test for two binding kinetic regimes and characterisation of the effect of longer 2s capture rates in our estimation of bound-times

(A) Frequency of detected short gaps, likely representing photoblinking, during the tracking of a population of LacI-mMaple molecules using 500ms capture rates. We applied a cut-off threshold at 2.6 seconds for the maximum duration of photoblinking based on previous characterisation of mMaple [188]. More than 75% of the molecules did not show photoblinking (N=148 molecules)

(B) Distribution of gap times between subsequent localizations at the same location of the field of view. Note that most events lasted for only one frame. N=60 events.

(C) We fitted the distribution of gap times to a single exponential function using a truncated form of MLE. This was done to account for the fraction of events shorter than 500ms, which would be missed in our experiments. Using a 1-frame memory parameter we estimate that our analysis will prematurely terminate less than 7.5%, 3% and 0.0001% of the tracks due to blinking when using a 1-second, 2-second, and 5-second intervals, respectively.

(D) Semi-log plots of the data presented in Figure 2C for ε and DnaB. The plots show a relatively linear relation between number of cases and time, which is indicative of a single regime of binding for both subunits. Further support of a single binding behaviour is presented in Supplementary file 1C.

(E) Plots showing the PDF curves of bound, bleaching, and tracking times for representative results from a single experiment of ε imaged with 500ms (left) (N=143) and 2s exposure (right) (N=415). The bound-time was 7.44s (SE ±1.07s) and 12.34s (SE ±1.36s) for 500ms and 2s, respectively. The plot for the 500ms example is presented to facilitate comparison and is identical to that in Figure 2C.

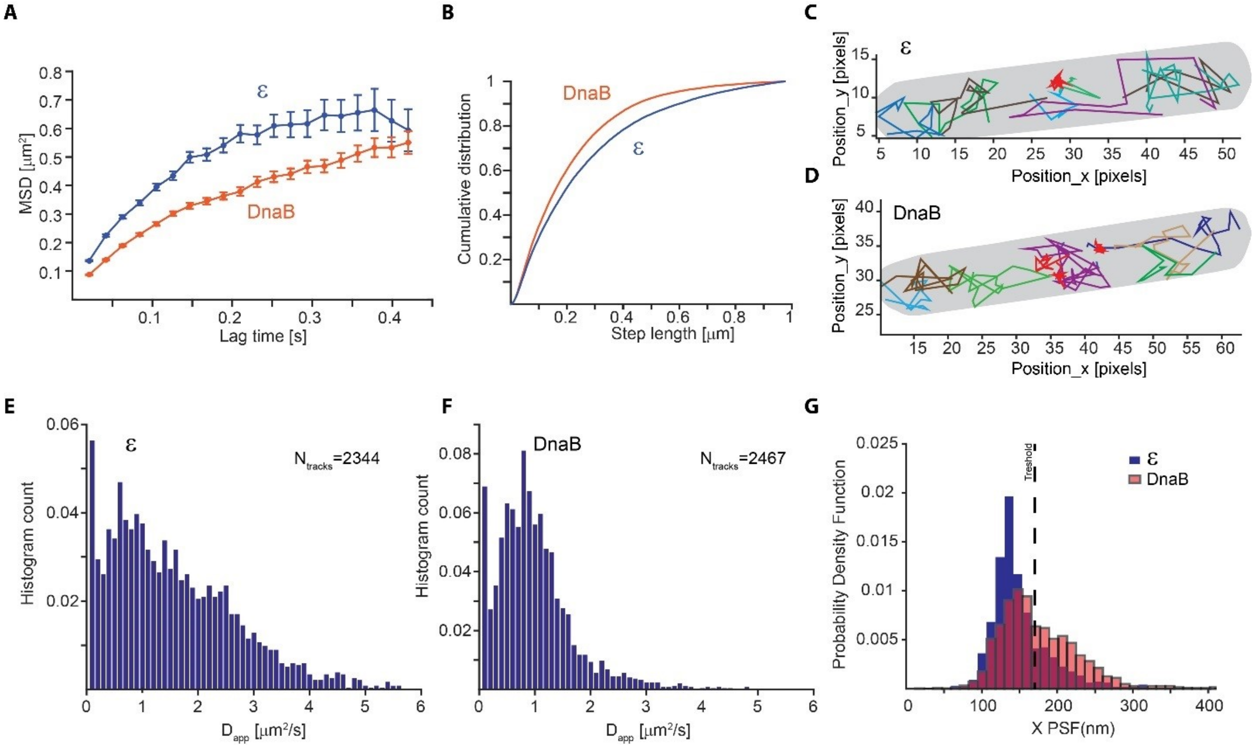


Figure A1.7: Slow diffusion of DnaB helicase complicates correct assignment of immobile molecules at sub-second capture rates

(A-F) Analysis from PALM experiments of ε-mMaple and mMaple-DnaB using a 21ms capture rate. (A) Calculated MSDs for the two proteins. In both cases, curves plateau at around 600nm in agreement with the dimension of the short axis of the cell. (B) Comparison of step lengths in the tracks of diffusing molecules analysed. (C-D) Examples of detected tracks for ε and DnaB. Lines of different colours are used to facilitate the observation of individual tracks. A red line was only used to show the position of tracks representing immobile tracks where the apparent diffusion coefficient is close to 0. The outline of the cell is shown in grey. (E-F) Distribution of the apparent diffusion coefficients calculated. 2344 tracks obtained from 77 cells and 2467 tracks obtained from 90 cells were used for ε and DnaB, respectively.

(G) Distribution of mean PSFs for the x-axis for tracks of ε-mMaple or mMaple-DnaB obtained from experiments done with 500ms capture rates. In the case of ε, a clear peak close to 100nm shows the PSF dimensions of immobile molecules, while a second peak close to 200nm represents diffusing molecules. In contrast, the dimensions of the PSFs from immobile and diffusive molecules is less clear for DnaB. The dashed line shows the threshold used to assign immobile and diffusing molecules in our analysis.

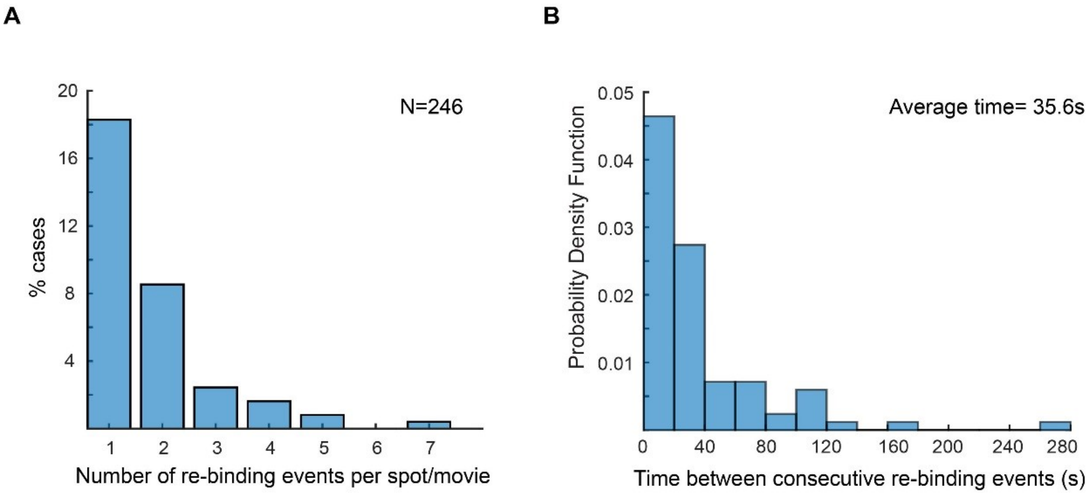


Figure A1.8: Re-binding of copies of ε at the same position are unexpectedly frequent

(A) Plot showing the number of rebinding events during a single experiment.

(B) Distribution of gap times, calculated from data obtained using 2 second exposure time. Rebinding analysis was done similarly to the track linkage analysis, except no frame threshold was applied. Time between re-binding events was obtained by determining the time interval between the end of a single-molecule track and the beginning of a new track at the same position in the cell. Note that in reality the frequency of binding will likely be higher, since molecules have a similar probability of binding to a different replisome between consecutive events. We manually analysed cells with two replisome spots and estimated that 52% of consecutive spot reappearance events occur at a different position in cells with two replisomes (N=30 cells; 199 re-binding events).

Table A1.1 Strains used for this study

Strain	**Relevant genotype**	**Source**
AB1157	*thr-1, araC14, leuB6(Am), DE(gpt-proA)62, lacY1, tsx-33, qsr'-0, glnV44(AS), galK2(Oc), LAM-, Rac-0, hisG4(Oc), rfbC1, mgl-51, rpoS396(Am), rpsL31(strR), kdgK51, xylA5, mtl-1, argE3(Oc), thi-1*	Dewitt, adelberg, 1962
RRL27	*holC-ypet kan*	Reyes-Lamothe 2008
RRL30	*holE-ypet kan*	Reyes-Lamothe 2008
RRL32	*ssb-ypet kan*	Reyes-Lamothe 2008
RRL33	*holA-ypet kan*	Reyes-Lamothe 2008
RRL34	*holD-ypet kan*	Reyes-Lamothe 2008
RRL35	*dnaE-ypet kan*	Reyes-Lamothe 2008
RRL36	*dnaQ-ypet kan*	Reyes-Lamothe 2008
RRL51	*dnaX-ypet kan*	Reyes-Lamothe 2008
RRL196	*frt ypet-dnaN*	Reyes-Lamothe 2010
RRL368	*frt-ypet-dnaB*	Reyes-Lamothe 2010
RRL537	*dnaQ-mmaple kan*	This study
RRL538	*holA-mmaple kan*	This study
RRL541	*tetR-ypet kan, [tetO240-gm]852*	This study
RRL553	*dnaX-mmaple kan*	This study
RRL557	*frt mmaple-dnaB*	This study
RRL558	*frt mmaple-dnaN*	This study
TB44	*dnaB-mMaple kan*	This study
TB54	*lacI-mMaple kan, [lacO240-hyg]2735::ΔpheA*	This study

Plasmid	**Features**	**Source**
pKD46	Expression of lambda red genes	Datsenko and Wanner
pCP20	Expression of Flp recombinase	Datsenko and Wanner
pROD61	mYPet Kan R6K gamma ori. For C-ter insertions	This study
pROD83	YPet Kan R6K gamma ori. For N-ter insertions	This study
pROD93	mMaple Kan R6K gamma ori. For C-ter insertions	This study
pROD160	mMaple Kan R6K gamma ori. For N-ter insertions	This study

Table A1.2 Analysis of FRAP data

Strain	N	MaxR	A	Koff	TimeC	Asym	Rsqrd	DoF	redChi	KS test	pKS	SE A	SE Koff	SE Asym	A 95% Conf	Koff 95% Conf	Asym 95% Conf
DnaB	96	0.749	0.312	0.135	7.388	0.483	0.835	19	0.015	Nor	0.876	0.059	0.051	0.016	0.131,0.343	0.018,0.213	0.455,0.514
DnaB	36	0.770	0.277	0.100	9.971	0.467	0.369	16	0.167	Nor	0.674	0.065	0.062	0.035	0.129,0.387	0.014,0.190	0.408,0.528
DnaB	25	0.823	0.225	3.540	0.282	0.484	0.204	22	0.092	Nor	0.440	0.018	0.000	0.018	0.189,0.262	3.54,3.54	0.450,0.521
DnaB +HU	35	0.793	0.189	0.122	8.165	0.388	0.775	10	0.011	Nor	0.985	0.056	0.058	0.024	0.082,0.358	0.022,0.232	0.355,0.417
α	44	0.924	0.795	0.272	3.677	0.907	0.665	24	0.111	Nor	0.458	0.072	0.065	0.039	0.698,0.890	0.155,0.399	0.834,0.984
α	48	0.783	0.564	0.240	4.172	0.699	0.942	8	0.014	Nor	0.920	0.112	0.072	0.059	0.415,1.07	0.134,0.455	0.622,0.753
α +HU	55	0.787	0.555	0.042	23.978	0.747	0.828	18	0.064	Nor	0.487	0.179	0.018	0.197	0.384,1.054	0.012,0.079	0.576,1.297
ß	54	0.762	0.454	0.027	37.341	0.670	0.749	31	0.070	Nor	0.117	0.044	0.008	0.060	0.387,0.536	0.007,0.038	0.629,0.866
ß	38	0.755	0.637	0.029	34.942	0.830	0.841	16	0.020	Nor	0.579	0.139	0.007	0.195	0.582,1.095	0.003,0.036	0.776,1.459
ß +HU	71	0.744	0.516	0.014	72.568	0.818	0.824	16	0.073	Nor	0.616	0.207	0.007	0.258	0.374,0.979	0.002,0.026	0.679,1.412
ε	17	0.734	0.435	0.068	14.666	0.808	0.682	14	0.200	Nor	0.997	0.159	0.039	0.183	0.267,0.827	0.011,0.129	0.666,1.288
ε	42	0.820	0.461	0.206	4.852	0.699	0.608	22	0.124	Nor	0.979	0.069	0.085	0.027	0.249,0.534	0.105,0.391	0.642,0.747
ε	71	0.729	0.523	0.167	5.999	0.771	0.804	22	0.032	Nor	0.335	0.058	0.036	0.027	0.365,0.598	0.108,0.273	0.725,0.828
ε	65	0.768	0.437	0.188	5.323	0.668	0.805	18	0.039	Nor	0.511	0.048	0.044	0.022	0.323,0.501	0.104,0.279	0.626,0.709
ε +HU	28	0.743	0.498	0.028	35.101	0.784	0.873	10	0.043	Nor	0.726	0.224	0.018	0.270	0.207,0.946	0.006,0.054	0.590,1.329
ε +HU	47	0.762	0.617	0.027	37.488	0.837	0.924	10	0.032	Nor	0.677	0.242	0.013	0.280	0.330,1.171	0.007,0.050	0.611,1.472
τ	43	0.721	0.487	0.235	4.247	0.662	0.617	22	0.068	Nor	0.997	0.075	0.059	0.029	0.410,0.576	0.158,0.390	0.607,0.719
τ	27	0.807	0.580	0.330	3.029	0.684	0.861	10	0.018	Nor	0.991	0.298	0.161	0.022	0.0580,1.102	0.033,0.627	0.640,0.724
τ	47	0.793	0.693	0.115	8.733	0.872	0.862	10	0.017	Nor	0.523	0.177	0.037	0.051	0.438,1.317	0.062,0.217	0.798,0.948
τ +HU	49	0.740	0.537	0.025	39.458	0.799	0.891	10	0.047	Nor	0.942	0.245	0.016	0.291	0.217,1.021	0.006,0.048	0.584,1.372
τ +HU	41	0.738	0.350	0.261	3.829	0.455	0.675	16	0.029	Nor	0.509	0.108	0.114	0.017	0.035,0.665	0.099,0.496	0.421,0.486
τ +HU	71	0.646	0.416	0.058	17.152	0.602	0.870	10	0.012	Nor	0.676	0.123	0.014	0.023	0.312,0.791	0.036,0.088	0.567,0.650
δ	48	0.847	0.757	0.292	3.428	0.932	0.816	21	0.056	Nor	0.718	0.093	0.120	0.028	0.314,0.823	0.130,0.554	0.887,0.993
χ	42	0.804	0.583	0.167	6.000	0.884	0.765	26	0.063	Nor	0.435	0.060	0.051	0.048	0.445,0.687	0.071,0.282	0.815,0.985

The list of FRAP experiments for each the subunits is shown. Each row represents a single set of experiments where the data from different time intervals were combined to obtain a single set of estimates. N represents the number of cells measured.

The model used for fitting the recovery curves was: $y = c - a\,\mathrm{e}^{-bt}$

where ‘a’ is the amplitude of recovery (**A**), ‘b’ is the rate of recovery (**Koff**) and ‘c’ is the asymptote for curve (**Asym**).

MaxR= Total cell intensity after bleaching step, defining the maximum possible recovery

Time Constant (**TimeC**) = 1/Koff

Degrees of Freedom (**DoF**) were defined as the number of time-points minus the number of parameters in the model
Kolmogorov-Smirnov (**KS**) test is used to determine if the distribution of the residuals is normal (Nor) as a measure of the goodness of fit
pKS is the p-value of the Kolmogorov-Smirnov test of the residuals
Standard Errors (**SE**) were calculated using Bootstrap resampling of the data with 10,000 iterations
95% Confidence values (**95% Conf**) were calculated from Bootstrap resampling

Table A1.3 Analysis of sptPALM data

Strain	Exposure Time, Interval Time	Number of Molecules	Bound Time (s)	Std. Error (s)	95% Confidence Interval
ε	500ms,1 second	40	4.0781	0.8499	2.6011 , 6.0413
ε	500ms,1 second	143	7.4411	1.0731	5.3831,9.6147
ε	500ms,1 second	57	5.3643	0.9316	3.7838,7.5324
ε	2seconds, 2seconds	163	10.4058	1.9365	8.1133,16.5602
ε	2seconds, 2seconds	415	12.3396	1.3616	10.2301,16.0835
ε	500ms,5 seconds	106	12.2612	2.4566	8.4361 , 18.3918
ε	2 seconds, 10 seconds	771	15.0272	1.2051	10.1861,31.5054
ε	2 seconds, 10 seconds	307	16.5094	4.2964	12.6881,17.1335
ε	500ms,1 second	145	10.8298	2.147	8.1978,16.9549
ε HU	500ms, 1 second	94	22.2076	5.0136	13.2196,33.3087
ε HU	500ms, 5 seconds	194	26.996	4.6829	19.5857 ,39.6222
ε RIF	500ms,1 second	69	6.1274	1.4109	4.1338,10.2727
ε RIF	500ms,1 second	60	7.081	2.8167	3.8225,19.7814
δ	500ms, 1 second	78	11.5957	2.9982	6.8584, 17.3936
δ	500ms, 1 second	92	8.697	1.5627	6.0750,12.4884
δ	500ms, 5 seconds	139	13.5089	1.9453	10.3240 , 18.0131
δ	500ms, 5 seconds	441	12.4475	1.2638	10.6088,16.0456
γ / τ	500ms, 1 second	167	10.2711	1.54	7.5117, 13.5108
γ / τ	500ms, 1 second	166	7.731	0.9497	6.7347,10.1444
γ / τ	500ms, 1 second	64	10.0856	3.0839	5.5159 ,15.1284
γ / τ	500ms, 5 second	276	11.0481	1.289	8.9082 ,14.0603
γ / τ	500ms, 5 second	109	11.9405	2.0243	8.4691 16.7055
γ / τ HU	500ms, 1 second	190	27.5391	4.1893	21.2238 , 40.0616
γ / τ HU	500ms, 5 second	168	24.9486	3.7809	18.5154,33.7872
β	500ms, 1 second	336	57.1851	15.5542	41.8224 ,82.1468
β	500ms, 1 second	118	31.8378	4.6764	24.8192,44.4616
β	500ms, 1 second	181	29.8311	7.1506	22.6254,59.9923
β	500ms, 5 seconds	464	59.7503	6.0364	53.4034, 68.3717
β	500ms, 5 seconds	273	41.2986	2.9074	35.7621, 49.6564
β	500ms, 5 seconds	428	47.226	2.5643	42.8505,52.3607
β	500ms, 10 seconds	178	38.8389	5.1799	30.1962,50.7123

β	500ms, 10 seconds	263	43.0297	5.0548	34.324,54.2780
dnaB	500ms, 1 second	81	58.6854	16.3232	31.8179, 88.0281
dnaB	500ms, 1 second	39	43.9003	15.2807	21.9501, 65.8504
dnaB	500ms, 5 seconds	60	37.7639	8.1653	24.1637, 56.6458
dnaB	2seconds,2seconds	36	1681.2	2178.1	63.3,5395.3
dnaB	2seconds,2seconds	57	1202.8	2103.6	99.8,5395.3
dnaB	2seconds,2seconds	160	5285	1.93E+03	5035.4,5.3955
dnaB	2seconds,10 seconds	124	407.5475	149.3787	217.7521,846.2488
dnaB	2seconds,10 seconds	86	1864.6	1845.3	76.2,5398.1
dnaB	2seconds,10 seconds	115	746.5645	365.5298	379.3,1720.8
dnaB HU	500ms, 1 second	75	32.8548	6.7004	21.1815 , 48.1302
dnaB HU	500ms, 5 seconds	95	69.0107	16.8591	41.1344,103.5161

The list of sptPALM experiments is shown. Each row represents an independent set of experiments using the same time interval between pictures.

Table A1.4 Results of goodness-of-fit tests for sptPALM

Strain	Exposure Time, Interval Time	Number of Molecules	Chi square ($p<0.01$)	Likelihood ratio Test ($p<0.01$)	BIC Test (Single vs Double)	Talpha (s)	Tbeta (s)
ε	500ms,1 second	40	0.7482	7.41E-02	184.6894 188.8769	7.7879	1.776
ε	500ms,1 second	143	0.497	1.29E-06	777.7910 764.2785	14.5014	1.5326
ε	500ms,1 second	57	0.3432	0.1911	285.9711 292.3479	6.9698	1.4735
ε	2seconds, 2seconds	163	0.1035	0	1.0531 0.8344*10^3	16.8438	0.1
ε	2seconds, 2seconds	415	0.0149	0	2.7687 2.6880*10^3	32.4254	4.3654
ε	500ms,5 seconds	106	0.5092	0.00E+00	727.6712 474.5390	22.8176	0.1
ε	2 seconds, 10 seconds	771	3.81E-27	0	5.6370 2.3824*10^3	34.5085	0.1
ε	2 seconds, 10 seconds	307	0.0245	0	2.2915 1.0299*10^3	37.4506	0.1
ε	500ms,1 second	145	0.4957	1.82E-10	797.3873 766.6519	11.3787	0.1
ε HU	500ms, 1 second	94	0.3673	1.64E-08	637.4230 614.6339	43.8949	0.1
ε HU	500ms, 5 seconds	194	0.0486	0.00E+00	1.5828 1.2822*10^3	43.6251	0.1
ε RIF	500ms, 1 second	69	0.9248	0.0484	336.9174 341.4590	84.3264	4.1604
ε RIF	500ms, 1 second	60	0.0265	0.0054	264.7774 265.1736	10.738	1.9023
δ	500ms, 1 second	78	0.007	6.96E-07	474.4378 458.5241	2.94E+03	4.2
δ	500ms, 1 second	92	0.4768	4.56E-14	523.3928 475.5273	13.6277	0.1
δ	500ms, 5 seconds	139	0.3431	0.00E+00	975.6567 686.1073	23.426	0.1
δ	500ms, 5 seconds	441	2.17E-08	0.00E+00	3.0126 2.2388*10^3	19.8989	0.1
γ/τ	500ms, 1 second	167	0.3493	1.99E-08	981.3559 960.0837	39.6638	2.8935
γ/τ	500ms, 1 second	166	0.1804	0.00E+00	910.7361 831.9867	11.4531	0.1
γ/τ	500ms, 1 second	64	0.0469	4.03E-10	378.9490 348.1694	5399.9	4.1
γ/τ	500ms, 5 second	276	0.0968	0.00E+00	1.8303 1.2212*10^3	19.554	0.1
γ/τ	500ms, 5 second	109	0.1869	0.00E+00	742.9151 495.1825	21.6465	0.1
γ/τ HU	500ms, 1 second	190	0.0104	2.45E-11	1.3167 1.2826 *10^3	2.73E+03	3.9

γ/τ HU	500ms, 5 second	168	9.29E-04	0.00E+00	1.3508	1.0929*10^3	39.8463	0.1
β	500ms, 1 second	336	0.0028	0	2.4845	2.4201 *10^3	5400	6.4
β	500ms, 1 second	118	0.7531	2.75E-11	874.4232	839.6167	65.93	0.1
β	500ms, 1 second	181	0.2842	5.44E-15	1.2309	1.1801*10^3	105.3916	0.1
β	500ms, 5 seconds	464	1.45E-08	0.00E+00	4.2937	3.6104 *10^3	116.3723	0.1
β	500ms, 5 seconds	273	0.0031	0	2.4390	2.0324*10^3	65.0855	0.1
β	500ms, 5 seconds	428	1.89E-04	0	3.7845	3.2143*10^3	86.8245	0.1
β	500ms, 10 seconds	178	0.0793	0	1.6199	1.1354*10^3	63.1876	0.1
β	500ms, 10 seconds	263	0.0013	0	2.3954	1.7464*10^3	70.2949	0.1
dnaB	500ms, 1 second	81	0.0233	5.54E-07	605.9887	589.7109	5400	3.5
dnaB	500ms, 1 second	39	0.5106	6.32E-06	288.5588	275.4979	1049	0.1
dnaB	500ms, 5 seconds	60	0.0235	0.00E+00	524.7951	427.8399	67.4397	0.1
dnaB	2seconds,2seconds	36	N/A	1.92E-05	358.2261	347.1240	5400	7
dnaB	2seconds,2seconds	57	0.0879	1.53E-09	562.4826	534.0724	1353.3	0.1
dnaB	2seconds,2seconds	160	0.0089	0	1.5839	1.5203*10^3	5400	4.1
dnaB	2seconds,10 seconds	124	2.28E-04	0	1.4453	1.1710*10^3	5.3992	0.1
dnaB	2seconds,10 seconds	86	0.1985	0	1.0963	0.8884 *10^3	5400	0.1
dnaB	2seconds,10 seconds	115	0.0258	0	1.3971	1.1542*10^3	5381.5	0.1
dnaB HU	500ms, 1 second	75	0.0796	1.17E-11	534.6678	497.2866	161.2	0.1
dnaB HU	500ms, 5 seconds	95	0.0013	0.00E+00	902.2316	764.1085	147.6445	0.1
ε	500ms,1 second	142	0.5658	2.89E-15	750.1909	697.7604	9.3415	0.1
ε	2seconds, 2seconds	413	0.0136	0.00E+00	2.7043	2.3847 *^3	16.3533	0.1
δ	500ms, 1 second	76	0.1063	6.18E-06	421.5906	409.8209	10.305	0.1
γ/τ	500ms, 1 second	161	0.4302	1.02E-12	877.2507	836.6067	10.1318	0.1
γ/τ HU	500ms, 1 second	186	0.0979	6.24E-13	1.2295	1.1882 *10^3	33.771	0.1

The summary of the fitting test results for sptPALM experiments is presented. Chi square test was used to assess the goodness-of-fit of a single exponential function. Likelihood ratio test and BIC tests were used to test for a second exponential behavior in the distributions.

***Data sets highlighted in yellow indicate ones that initially suggested a two -exponential model. Data highlighted in green represents the results after outliers were removed from these data sets. The difference in the number of molecules indicates the number of outliers removed.**

A2-Supplementary Information: A Machine Learning Approach for Classification of DNA-Bound Proteins from Single-Molecule Tracking Experiments

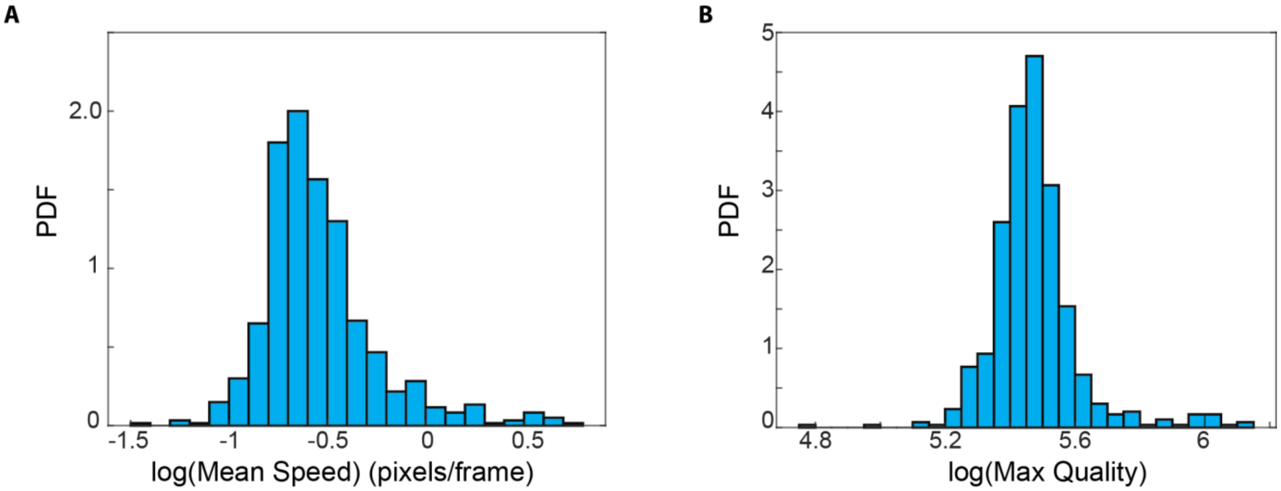


Figure A2.1: A) Representative distribution of the log(mean speed) values for tracks classified as being bound from training data set (500ms, *E.coli*). B) – Representative distribution of the log(maximum quality) values for tracks classified as being bound from training data set (500ms, *E.coli*).

Table A2.1: Parameter values used to construct random forests.

Training Data	Number of Trees	Minimum Leaf Size	Predictors to Sample at Node	Bag Fraction	OOB error	Number of Tracks
***E.coli* (simulation)**						
500ms (ML model 1)	6000	50	2	0.5	0.034	1263
500ms (ML model 2)	6000	50	2	0.5	0.0079	1263
100ms (ML model 1)	6000	50	2	0.5	0.0195	1075
100ms (ML model 2)	6000	50	2	0.5	0.014	1075
Budding Yeast (simulation)						
500ms (ML model 1)	10000	50	2	0.8	0.03	967
500ms (ML model 2)	10000	50	2	0.8	0.0269	967
LacI						
500ms (ML model 1)	10000	70	2	0.5	0.1634	1438
500ms (ML model 2)	10000	70	2	0.5	0.0709	1438
Histone H3						
500ms, 1s time interval (ML model 1)	6000	50	2	0.5	0.1151	1251
500ms, 1s time interval (ML model 2)	6000	50	2	0.5	0.036	1251

Table A2.2: Parameters used for simulations of training data and experimental data.

Training Data	**Tbleach (s)**	**Tbound(s)**	**Tsearch(s)**	**D_{mobile} (um^2/s)**	**D_{bound} (um^2/s)**	**Mobile fraction**	**Bound Fraction**	**Integrated Spot Intensity**
Ecoli								
500ms exposure (no time interval)	10	100	100	0.5	0.005	0.3	0.7	3000
100ms exposure (no time interval)	2	100	100	0.5	0.005	0.3	0.7	3000
Budding Yeast								
500ms exposure (no time interval)	10	100	100	0.5	0.005	0.3	0.7	3000
Experimental								
Ecoli								
500ms exposure (1s time interval)	20	8	10000000	0.5	0.005	0.5	0.5	3000
500ms exposure (1s time interval, lower spot intensity)	20	8	10000000	0.5	0.005	0.5	0.5	2000
100ms exposure (no time interval)	2	1	10000000	0.5	0.005	0.7	0.3	3000
100ms exposure (no time interval, higher D_{mobile})	2	1	10000000	5	0.005	0.7	0.3	3000
100ms exposure (no time interval, mixed bound population)	10	1s/7s	10000000	0.5	0.005	0.1	0.45/0.45	3000
Budding Yeast								
500ms exposure (1s time interval)	20	8	10000000	0.5	0.005	0.5	0.5	3000
500ms exposure (1s time interval, lower spot intensity)	20	8	10000000	0.5	0.005	0.5	0.5	2000

Table A2.3: Accuracy and RE on test data after final classification step.

E.coli	**Prediction Accuracy**	**Recovery Error (RE)**
500ms	0.96	0.1
100ms	0.97	0.04
Budding Yeast		
500ms	0.97	0.06

Table A2.4: Results including errors for simulated experimental data.

E.coli	**Spot Intensity = 3000**	**Spot Intensity = 2000**
1s Interval (500ms exposure)		
Bound Time [95% Confidence Interval]	6.76[5.41, 8.59]	7.21[5.65, 8.95]
Number of tracks predicted to be bound	169	159
Prediction Accuracy	0.93	0.99
Recovery Error	0.11	0.11
100ms Exposure (no time interval)		
Bound Time [95% Confidence Interval]	0.97[0.75, 1.21]	
Number of tracks predicted to be bound	156	
Prediction Accuracy	0.85	
Recovery Error	0.18	
100ms Exposure (no time interval). Predicted using 500ms Training Data		
Bound Time [95% Confidence Interval]	0.96[0.75, 1.22]	
Number of tracks predicted to be bound	158	
Prediction Accuracy	0.86	
Recovery Error	0.16	
100ms Exposure (no time interval). D_{mobile} = 5um^2/s		
Bound Time [95% Confidence Interval]	0.90[0.71, 1.13]	
Number of tracks predicted to be bound	142	
Prediction Accuracy	0.98	
Recovery Error	0.08	
Budding Yeast		
Bound Time [95% Confidence Interval]	7.26[5.97, 8.77]	7.17[5.92, 8.82]
Number of tracks predicted to be bound	232	227
Prediction Accuracy	0.97	0.97
Recovery Error	0	0

Table A2.5: Primers used for yeast strains.

Primer	**Description**	**Sequence**
TB81	C-terminal mNeonGreen tagging of PCNA (F)	cctacagtttttcttggctcctaaatttaatgacgaagaaGGTGACGGTGCTGGTTTA
TB82	C-terminal mNeonGreen tagging of PCNA (R)	tttattatttttagtatacaactatatagataatttacatCACAGGAAACAGCTATGACC
TB98	Screen C-terminal tag of PCNA (F)	AGAGTTGGTATCAGGCTCTC
TB99	Screen C-terminal tag of PCNA (R)	AAGCTGATATTTAACGCATCTTAG
TOP2insF	C-terminal Halo tagging of Top2 (F)	aggaaaaccaaggatcagatgtttcgttcaatgaagaggatGGTGACGGTGCTGGTTTA
TOP2insR	C-terminal Halo tagging of Top2 (R)	acatataaaaagaatggcgctttctctggataaatattatCACAGGAAACAGCTATGACC
TOP2seqF	Screen C-terminal tag of Top2 (F)	ACTATCTGGTGAAAGCGACC
TOP2seqR	Screen C-terminal tag of Top2 (R)	ACGATGTTTTTCGCCCAGGC

A3-Supplementary Information: Processive Activity of the Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells

A3.1 Optiprep viability

Iodixanol (Optiprep) was shown to be compatible with a variety of live specimens and to have no effect on their growth or viability [203]. Since none of the specimens tested had been yeast cells, we decided to verify that it did not affect growth and viability of *S. cerevisiae* under our specific experimental and imaging conditions.

In an initial attempt, we tried to perform growth curves on parental yeast strains with and without Optiprep. Due to the refractive index matching of Optiprep, the light absorption (**Figure A3.4A**) and subsequently the OD measurements of cultures with the compound are lowered by an order of magnitude, which makes comparisons more difficult as the measurements are close to the lower detection limit of a spectrophotometer, which increases measurement error (**Figure A3.4B**).

To alleviate this problem we approached growth measurement differently, by spotting YTB31 cells on Geneframes with SC-agarose pads, where the agarose is made in either water or in Optiprep (final concentration 30%). Cells were imaged in bright-field at 5min intervals over 15h at 22°C, with the Geneframe preventing evaporation and drying of the agarose pad. Each cell was then analyzed for cell cycle duration, by measuring the time between the appearance of a bud and the appearance of the next bud after division has occurred (**Figure A3.4C**). The average of this time for all cells form multiple fields of view was used as a measure of doubling time. The doubling times with and without Optiprep were almost identical (125min and 126min, N=53 and N=31, respectively), which confirms that the presence of Optiprep in the agarose pad does not affect growth, and therefore is unlikely to affect DNA replication.

A3.2 Growth and DNA replication in Halo-tagged strains

To ensure the presence of the Halotag at the C-terminus of replisome proteins, the mNeonGreen fusion to PCNA and the deletion of *pdr5* were not affecting DNA replication, we performed growth curves on the strains we used for imaging and compared this to the parental strain BY4741 as well as YTB31 (PCNA-mNeonGreen fusion alone). Cultures were grown in SC in the same way used for imaging (see **Microscopy**). OD was measured at 30min intervals over 9h and the slope of the resulting curve was used to calculate the doubling time of each strain. The average of two repeats is shown in the figure. The average doubling times of Halo-tagged Pol2 (YJL02), Pol3 (YJL24), Pol32 (YJL11), Mcm4 (YAY256), Cdc45 (ZEY158), Ctf4 (ZEY077), and the combined Pol12-Halo CIP- mutant (YAJ05) are all similar to those of BY4741 and YTB31 (**Figure A3.5**), suggesting the genetic modifications in these strains do not affect growth and viability. In contrast, a strain where Dpb2 (another subunit of Pol ε) was tagged shows a higher doubling time than any of the other strains, indicative of potential growth problems. This strain was therefore not used for imaging. The histone H3-Halo strain YTK1434 [134] also showed significantly slower growth rate, though this should not affect our results since it was only used as a bleaching control.

Although we were confident the Halo tags did not affect growth, we analyzed the imaging strains by flow cytometry (see **Materials and Methods**) so we could be certain it was not causing more subtle problems with DNA replication that could affect our estimates of residency times without affecting growth rates. Strains with Halo-tagged Pol12 (YJL10), Pol2 (YJL02), Pol3 (YJL24), Pol32 (YJL11), Mcm4 (YAY256) and Cdc45 (ZEY158) all showed similar profiles to the parental BY4741, with a similar proportion of cells in S-phase (between the left-side G1 peak and the right-side G2 peak), in contrast to the Halo-tagged Dpb2, which shows an increased proportion of cells in S-phase indicative of slower or problematic DNA replication and an extended S-phase (**Figure A3.6A**). This increased S-phase proportion can also be seen in a clb5 deletion mutant (YHZ09), which is known to have a lengthier S-phase (**Figure A3.6B**). A small but reproducible change in the pattern, indicative of a longer S-phase, was also observed in the Cdc45-Halo strain (ZEY158). Taken together, the growth curves and flow cytometry data show that the Halo tags and other modifications to our strains do not influence DNA replication and growth and that our measurements and imaging accurately reflect the state of wild-type untagged cells.

A3.3 Western blot

To ensure that the fluorescence tracks seen in imaging were all due to tagged proteins of interest and not to free or cleaved Halo protein, we performed Western blots as described in Supplementary Methods. A different amount was used for each lysate in order to enhance detection of low abundance proteins, prevent saturation by high copy number proteins, and allow easier comparison between both. Since the primary antibody targets an epitope in the Halo tag, we expect to find a band for each protein shifted to a higher molecular weight (due to the added presence of the tag), and to not see any lower molecular weight bands if the Halo tag is not being cleaved. The representative blot (**Figure A3.7**) shows bands of the expected size when including the Halo tag for each of the replisome proteins (Pol12 113kDa, Pri2 96 kDa, Pol2 290kDa, Dpb4 56kDa, Pol3 159kDa, Pol32 75kDa, Mcm4 140kDa, Cdc45 109kDa, Ctf4 139 kDa) and no visible band where free Halo is expected (34kDa), which confirms that the fusions are intact and the Halo tag is properly associated with and not being cleaved from the replisome proteins. The expected molecular weight of HHT1 (Histone H3) with Halo is 50kDa, but other bands can also seen near 30kDa. H3 is known to undergo cleaving at various sites in multiple cellular processes and we hypothesize the bands we see are not free Halo, but rather these cleaved forms of H3. The absence of any secondary bands in any of the replisome proteins confirms our belief that the Halo tag fusion is stable and the observed fluorescence is indicative of actual replisome dynamics.

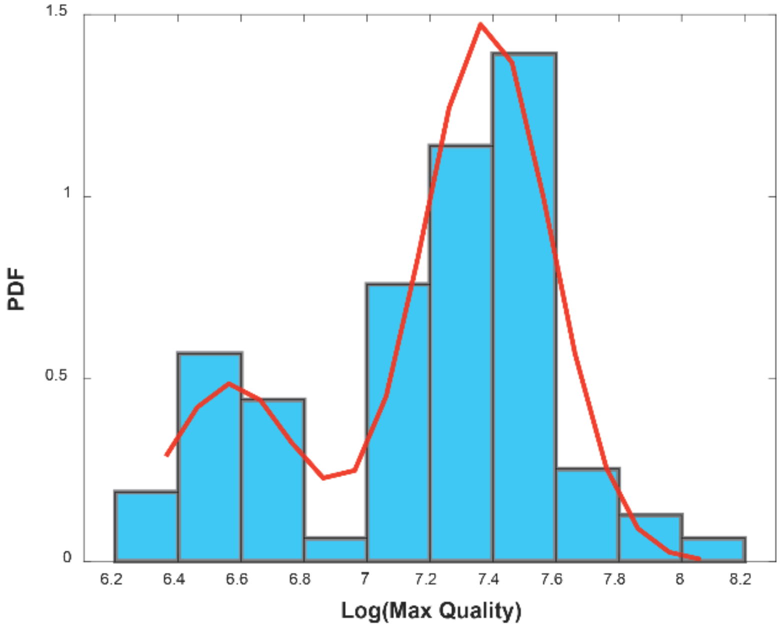


Figure A3.1: GMM fitting to quality values. After the first classification step using Model1, the log-transformed max quality values were fitted with a GMM, and the second largest peak (shown here as being ~ 7.4), was used to scale the quality parameters, prior to classification using Model 2.

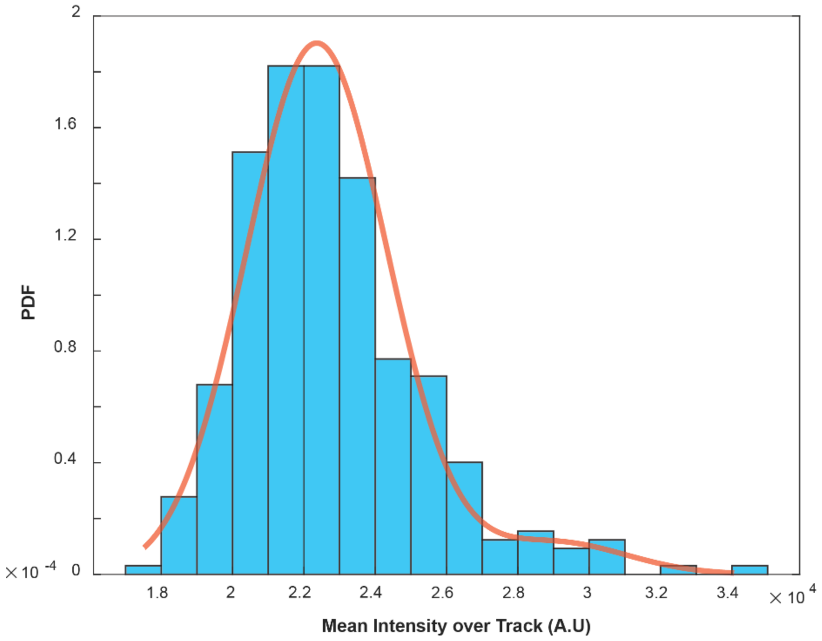


Figure A3.2: GMM fitting on mean intensities of tracks. A GMM was fit to the mean intensities of tracks, with a maximum of 3 components allowed to fit. With the GMM, we clustered the data to isolate tracks representing single molecules.

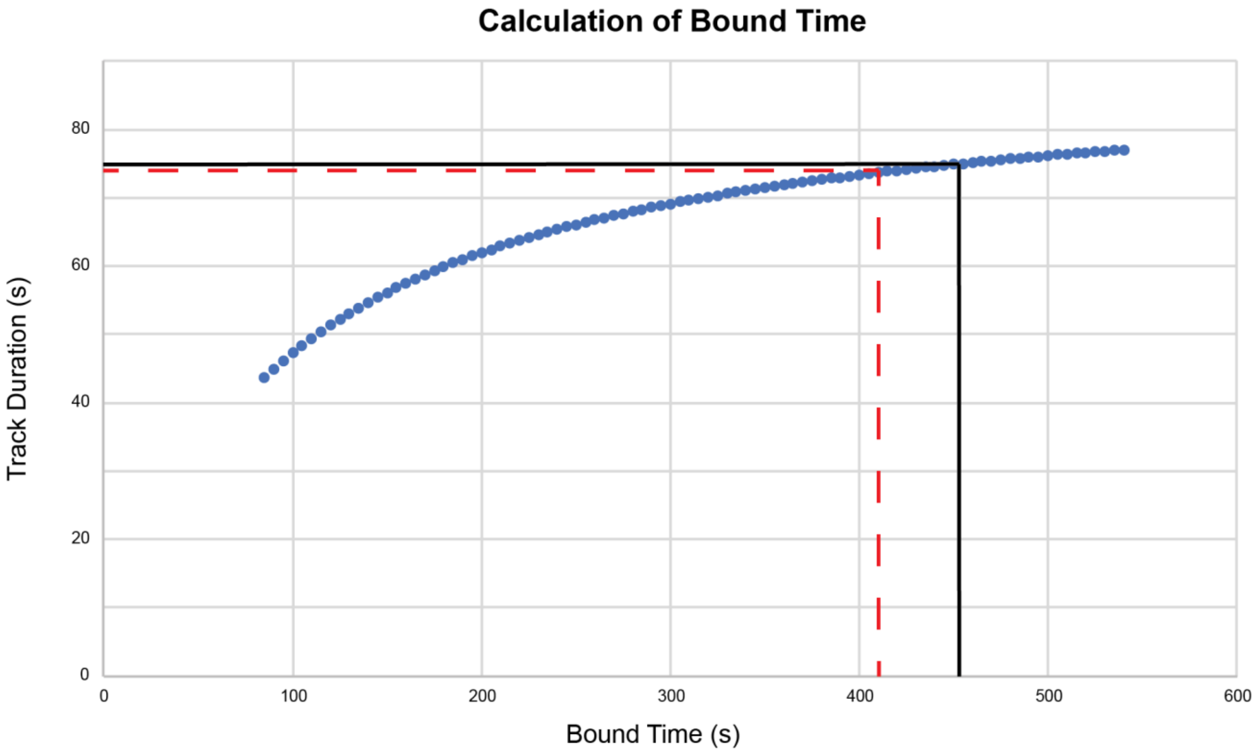
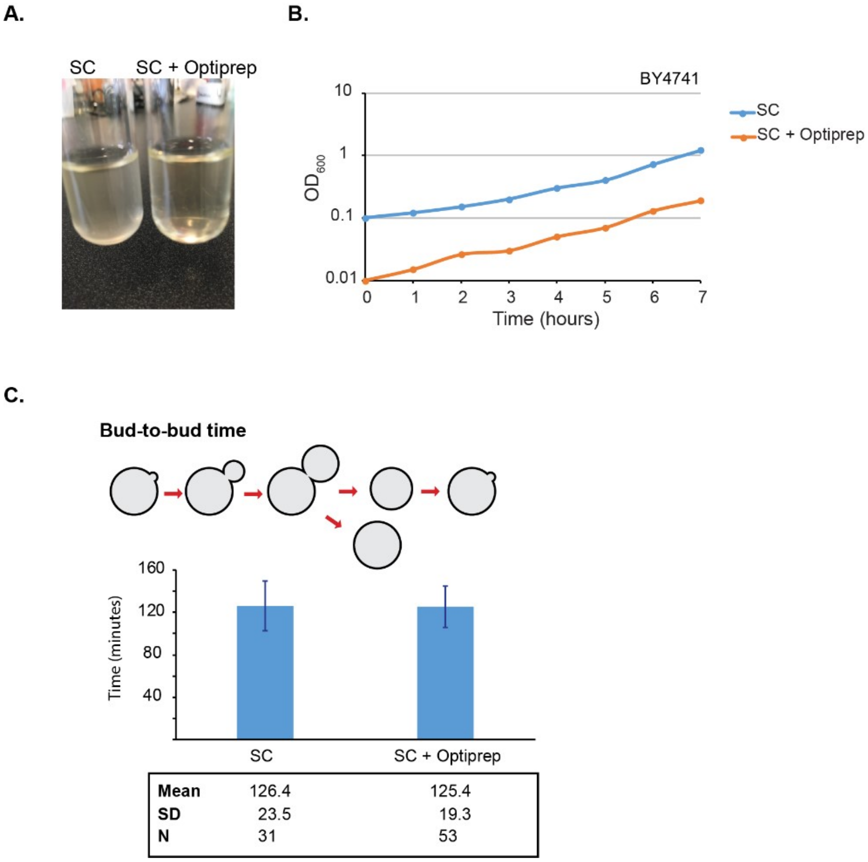
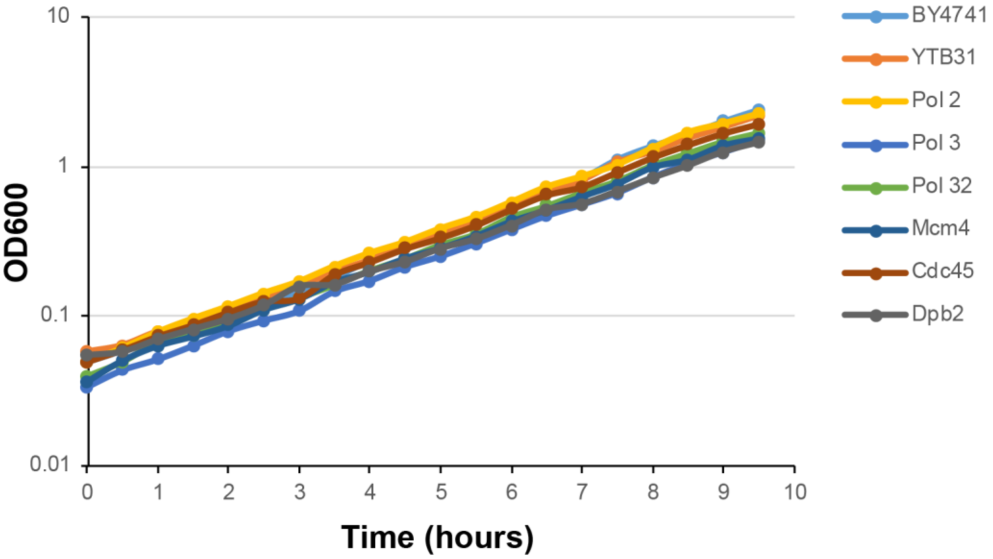


Figure A3.3: Calculation of minimum bound time using 8 interval data. Black solid indicated the bound time value that results in a track duration time equal to the lower bound of the CI for the bleaching data. The red dotted line represents the track duration that is 1s lower than the lower bound of the bleaching estimate CI.



	SC	SC + Optiprep
Mean	126.4	125.4
SD	23.5	19.3
N	31	53

Figure A3.4: Optiprep does not affect cell growth. A) Addition of Optiprep to an SC culture visibly reduces light absorption by the cells. Both tubes in the image contain two identical halves of the same cultures, with Optiprep added to one of them. The colony counts plated from both cultures are also identical. B) Growth curves of the parental BY4741 strain in SC and SC with 30% Optiprep. Both cultures were diluted from an exponentially growing culture at 30°C and OD measurements were taken every 1h. C) Bud-to-bud measurements on agarose pads. An exponentially growing culture of YTB31 was spotted on SC-agarose pads with the agarose made in either water or Optiprep. The slides were placed on an Olympus microscope at 22°C and bright-field timelapses were taken with 5min intervals for 15h. Single cells were individually timed from the formation of a bud until the formation of the next bud from the cell mother cell and this was used as a measure of doubling time. The mean doubling time for both conditions is shown, with error bars representing standard deviation of the mean.



Generation times in minutes

Strain	Repeat 1	Repeat 2	Average	StDev
BY4741	99	108	**104**	4.4
YTB31	104	110	**107**	2.9
Pol2	103	103	**103**	0.3
Pol3	105	111	**108**	3.2
Pol32	105	114	**109**	4.7
Mcm4	107	112	**109**	2.5
Cdc45	106	103	**105**	1.1
Ctf4	107	108	**108**	0.6
Dpb2	118	124	**121**	3.2
Pol12 CIP-	101	99	**100**	1.1
H3	127	126	**127**	0.6

Figure A3.5: Generation times of Halo-tagged strains used for imaging. Cultures were diluted from exponential parental cultures and grown in SC at 30°C. OD measurements were taken every 30 minutes and used to plot growth curves. The slopes of the curves were used to calculate the generation time of each strain. The representative graph shown here is from one of two independent experiments. The calculated generation times from both repeats are shown in the table along with their average.

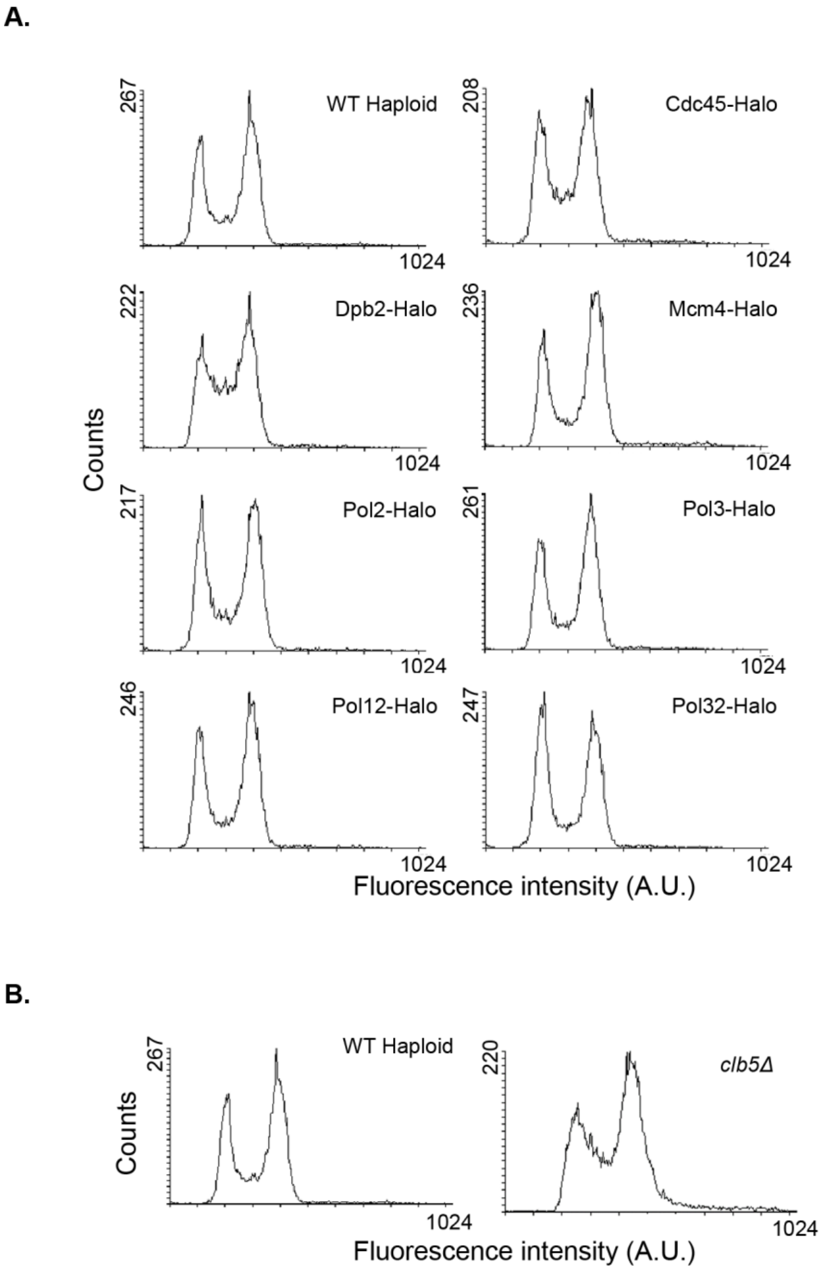
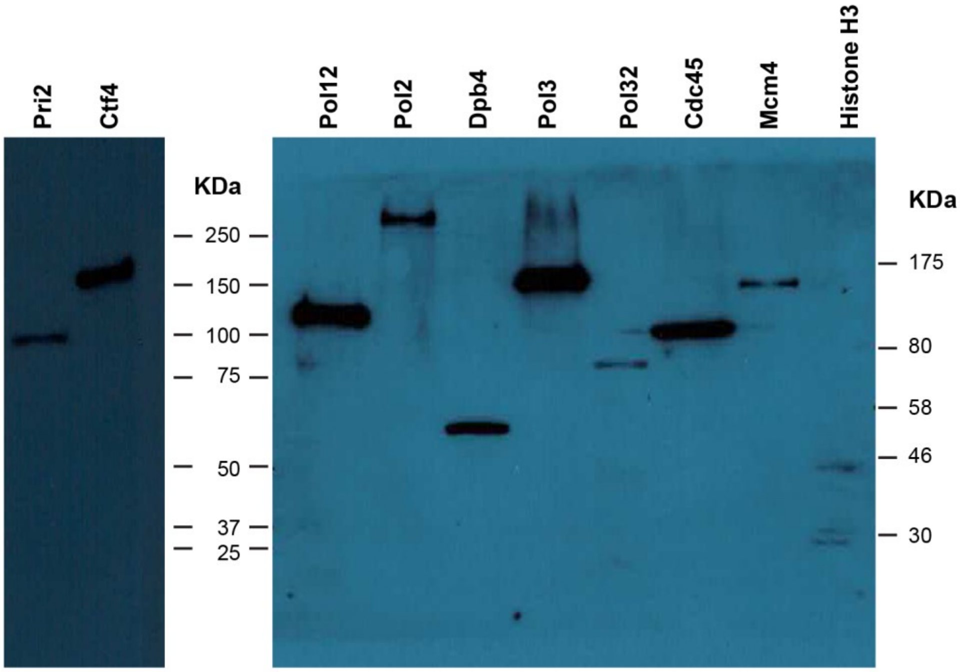


Figure A3.6: Flow cytometry analysis of Halo-tagged strains. Exponential YPD cultures were fixed in 70% ethanol then analyzed by flow cytometry. The cytometer was calibrated using exponentially-growing BY4741 (haploid) and BY4743 (diploid) parental strains. The first peak represents cells with a single chromosomal content (G1), the second peak at double the fluorescence intensity represents cells with fully replicated chromosomes (G2). The area between both peaks represents cell in S-phase. A) Halo-tagged strains compared to the parental wild-type. All HaloTag fusion have similar profiles to the parental wild-type, except for Dpb2-Halo, which has a higher proportion of cells in S-phase. B) The parental wild-type and a clb5 deletion strain. clb5Δ shows a higher proportion of S-phase cells.



Protein	Size (kDa)	Size with Halotag (kDa)
Pol12	78	113
Pri2	62	96
Pol2	255	290
Dpb4	22	56
Pol3	124	159
Pol32	40	75
Mcm4	105	140
Cdc45	74	109
Ctf4	104	139
HHT1	15	50
Halo alone		34

Figure A3.7: Western blot of HaloTag fusions. Cell lysates were made from cultures of the various imaging strains. The table shows the size of each protein as well as the expected size when accounting for the presence of the HaloTag. The band for each protein corresponds to the expected size and no smaller bands are seen for the replisome proteins.

Table A3.1: The *S. cerevisiae* strains used in this study are listed along with their genotypes.

Strain	Genotype
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
BY4742	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0
BY4743	MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0
YHZ09	MATa his3Δ1 leu2Δ0 LYS2 met15Δ0 ura3Δ0 clb5Δ0::KanMX POL30-mNeonGreen-Nat
YTB31	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 POL30-mNeonGreen-Nat
YTK1414	MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0 PDR5/pdr5Δ::KanMX
YJL10	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat POL12-Halo-HygB
ZEY136	MATa his3Δ1 leu2Δ0 LYS2 met15Δ0 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat PRI2-Halo-HygB
YJL02	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat POL2-Halo-HygB
YJL18	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat DPB4-Halo-HygB
YJL24	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat POL3-Halo-HygB
ZEY057	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat POL3-Halo-HygB clb5Δ0::KanMX
YJL11	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat POL32-Halo-HygB
YAY256	MATa his3D1 leu2D0 lys2D0 MET15 ura3D0 pdr5Δ::KanMX POL30-mNeonGreen-Nat MCM4-Halotag-HygB
ZEY158	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat CDC45-Halo-HygB
ZEY077	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat CTF4-Halo-HygB
YAJ05	MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 pdr5Δ0::KanMX POL30-mNeonGreen-Nat POL12-Halo-HygB Pol14A-URA3
YTK1434	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pdr5Δ0::KanMX HHT1-Halo-URA3

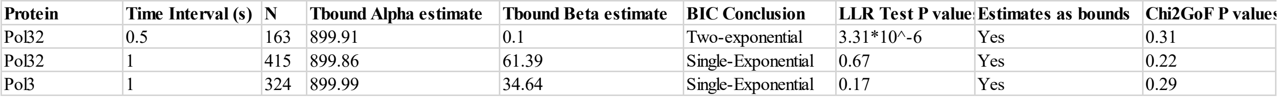
Table A3.2: Primers used in this study, with a short description and their sequence.

Primer	Description	Sequence
TB81	C-terminal mNeonGreen tagging of PCNA (F)	cctacagtttttcttggctcctaaatttaatgacgaagaaGGTGACGGTGCTGGTTTA
TB82	C-terminal mNeonGreen tagging of PCNA (R)	tttattatttttagtatacaactatatagataatttacatCACAGGAAACAGCTATGACC
TB98	Screen C-terminal tag of PCNA (F)	AGAGTTGGTATCAGGCTCTC
TB99	Screen C-terminal tag of PCNA (R)	AAGCTGATATTTAACGCATCTTAG
TB123	C-terminal Halo tagging of Pol12 (F)	caacgtgtggaagcgcgctagagttgacttgattgctagtGGTGACGGTGCTGGTTTA
TB124	C-terminal Halo tagging of Pol12 (R)	accttgagctattccattagttaagtttgaattaaatataCACAGGAAACAGCTATGACC
AY9	Screen C-terminal tag of Pol12 (F)	CTTGTTGAAGGTGAAGAGCC
AY10	Screen C-terminal tag of Pol12 (R)	GCCAGTTTCAAGGTCGATAG
TB125	C-terminal Halo tagging of Pri2 (F)	gaagctggaaaaggaaaaactattcaataatggtaatcatGGTGACGGTGCTGGTTTA
TB126	C-terminal Halo tagging of Pri2 (R)	tttagttatctcttcgctttttccttttccctttctgcaCACAGGAAACAGCTATGACC
AY11	Screen C-terminal tag of Pri2 (F)	CGAAAGATCAAGGCAACTGC
AY12	Screen C-terminal tag of Pri2 (R)	TTTTTGACCATACTTACAGTAGAC
TB61	C-terminal Halo tagging of Pol2 (F)	ttttgatatattattgagttgtattgctgatttgaccataGGTGACGGTGCTGGTTTA
TB62	C-terminal Halo tagging of Pol2 (R)	ggtaaagaggccattgaacctcgcgttatatactgcttacCACAGGAAACAGCTATGACC
TB44	Screen C-terminal tag of Pol2 (F)	TGCCCACTGTCCATGTGC
TB45	Screen C-terminal tag of Pol2 (R)	CAACTTCCGGAGTGGTCAC
AY23	C-terminal Halo tagging of Dbp4 (F)	ccaagatgtagaaactagagttcaaaaccttgagcaaacgGGTGACGGTGCTGGTTTA

AY24	C-terminal Halo tagging of Dbp4 (R)	gagtggtggcaagcactactagacagtttccatagcggggCACAGGAAACAGCTATGACC
AY43	Screen C-terminal tag of Dpb4 (F)	AAGGCGATGCATTACAGGAC
AY44	Screen C-terminal tag of Dpb4 (R)	TTCCCCGGCTTGCAAATAAC
TB63	C-terminal Halo tagging of Pol3 (F)	aaaagagctgcaggagaaagtagaacaattaagcaaatggGGTGACGGTGCTGGTTTA
TB64	C-terminal Halo tagging of Pol3 (R)	cctttcttaatcctaatatgatgtgccaccctatcgttttCACAGGAAACAGCTATGACC
TB71	Screen C-terminal tag of Pol3 (F)	GCGCTGGTAACTTACATAGTG
TB72	Screen C-terminal tag of Pol3 (R)	TGAATCTGGATTTTTCCAAGTATC
AY27	C-terminal Halo tagging of Pol32 (F)	gcaaggaacattggaaagctttttcaaaagaaaggcaaaaGGTGACGGTGCTGGTTTA
AY28	C-terminal Halo tagging of Pol32 (R)	tcacaattagtaatggaaagtgtttggaaaaaaaagaagaCACAGGAAACAGCTATGACC
AY47	Screen C-terminal tag of Pol32 (F)	AAGCAAGAAACGCCGTCATC
AY48	Screen C-terminal tag of Pol32 (R)	TTCTATCACGTAAGTTGACATTTG
AY29	C-terminal Halo tagging of Mcm4 (F)	cgagggtgtaaggagatcagttcgcctgaataaccgtgtcGGTGACGGTGCTGGTTTA
AY30	C-terminal Halo tagging of Mcm4 (R)	ttattaattgttacgcagggaatgattgtagtagacagcaCACAGGAAACAGCTATGACC
AY49	Screen C-terminal tag of Mcm4 (F)	GGAAGCCTTGTCAAGATTGC
AY50	Screen C-terminal tag of Mcm4 (R)	ATCGAGCCTACATACAGTATTG
TB87	C-terminal Halo tagging of Cdc45 (F)	ttcaccattcctggagaagctgaccttgagtggattgttaGGTGACGGTGCTGGTTTA
TB88	C-terminal Halo tagging of Cdc45 (R)	tatgctggtatatatgtacgactaaataatataaatttgaCACAGGAAACAGCTATGACC
TB104	Screen C-terminal tag of Cdc45 (F)	AAATAACTGCAGAAACGGATGC

TB105	Screen C-terminal tag of Cdc45 (R)	AGAGCCGCGCACAAAATATG
TB59	C-terminal Halo tagging of Ctf4 (F)	taataatataagggaagctagatatgaacagcaattgaaaGGTGACGGTGCTGGTTTA
TB60	C-terminal Halo tagging of Ctf4 (R)	tcaaataattgtctcttgcgtatatatattttacatttttCACAGGAAACAGCTATGACC
TB69	Screen C-terminal tag of Ctf4 (F)	CACTTACTGCAGCCGTTAAG
TB70	Screen C-terminal tag of Ctf4 (R)	TAATGTGGGAGCATTTTGAACG
NK57	Sequence mutations in POL1-4A CIP mutant (F)	ATATACGACGAAATCGACG
NK58	Sequence mutations in POL1-4A CIP mutant (R)	CCACATCATCCAATAAATCC

Table A3.3: Results from statistical tests to test for two-exponential behaviour. The lower and upper bounds on the estimates were 0.1s and 900s, respectively. The bleach time was 15s and 23s, for the 0.5s and 1s time interval, respectively. We used the BIC test and LLR test, under the null hypothesis of single-exponential distribution, to compare whether a two-exponential (more complex) model significantly fits better than the single-exponential. If the estimates were simply the bounds, it indicated that the algorithm could not find two different behaviours within the bleaching time. We also performed the chi square goodness of fit test, under the null hypothesis that the distribution comes from a single-exponential distribution.



Protein	Time Interval (s)	N	Tbound Alpha estimate	Tbound Beta estimate	BIC Conclusion	LLR Test P value	Estimates as bounds	Chi2GoF P value
Pol32	0.5	163	899.91	0.1	Two-exponential	3.31*10^-6	Yes	0.31
Pol32	1	415	899.86	61.39	Single-Exponential	0.67	Yes	0.22
Pol3	1	324	899.99	34.64	Single-Exponential	0.17	Yes	0.29

Table A3.4: Results of simulations for mean occupancy through a range of parameters.

Replisomes	Excess Copy Number	D_Pol Delta (um^2/s)	D_Replisome (um^2/s)	Mean Occupancy
300	3000	0.5	0.005	0.1249
300	3000	0.5	0.05	0.1261
300	3000	5	0.005	0.1221
300	3000	5	0.05	0.1243
300	6000	0.5	0.005	0.1773
300	6000	0.5	0.05	0.1822
300	6000	5	0.005	0.18
300	6000	5	0.05	0.1844

Table A3.5: Track durations from combined data sets**.** Standard errors and CI were calculated through bootstrapping 1000 samples.

	Estimate	Std.Error	CI	N
0.5s				
POL32	13.23	1.11	[11.3,15.59]	163
1.0s				
CDC45	23.24	1.2	[21.10,25.75]	406
DPB4	20.84	1.38	[18.28, 23.60]	269
POL3	22.32	1.33	[19.76, 24.94]	324
POL32	24.51	1.26	[22.25, 27.10]	415
POL12	17.09	1.12	[15.05, 19.30]	253
POL2	20.11	1.01	[18.31, 22.27]	305
CTF4	24.23	1.08	[22.36,26.63]	384
PRI2	14.75	1.33	[12.58, 18.00]	118
POL12_CIP	15.88	1.09	[14.02, 18.18]	240
CLB5	23.75	2	[20.25,28.22]	114
8s Interval				
CDC45	84.16	7.71	[70.29, 101.73]	129
DPB4	66.94	6.06	[56.64, 80.89]	128
POL32	65.54	5.61	[56.06, 78.09]	130
POL12	43.96	4.64	[36.49, 55.61]	95
CTF4	69.61	5.18	[60.09, 80.50]	147
YTK1434 (H3)	89.52	5.33	[80.01, 100.44]	263
20s Interval				
CDC45	178.44	18.24	[143.61, 217.19]	64
DPB4	128.91	11.95	[107.83, 155.94]	92
POL3	104.55	13.34	[81.36,136.01]	44
POL32	102.27	13.89	[77.73, 135.32]	44
YTK1434(H3)	153.67	15.29	[126.33, 185.26]	60

Table A3.6: Bound times for different subunits. Errors were calculated using bootstrap sampling.

1s	**Bound Time Estimate**	**Std error**	**CI**
Pol12	65.53	15.29	[45.971, 107.24]
Pri2	40.74	9.97	[27.38, 71.64]
Pol12_CIP	50.71	11.32	[34.34, 73.10]
8s			
Dpb4	265.39	141.02	[138.89, 625.63]
Pol32	244.67	88.30	[148.90, 486.98]
Pol12	86.38	17.01	[60.60,127.43]
Ctf4	312.98	164.77	[186.9, 1037.3]
20s			
Dpb4	800.06	584.62	[405.30, 2531.00]
Pol32	305.76	124.65	[155.99, 660.27]
Pol3	327.08	128.16	[170.05, 674.19]

Table A3.7: Results showing the track durations of all experiments. Histone H3 (20180511) was used for the ML training data set.

Date (year/month/day)	Protein	Sample Size	Mean Track Duration (s) [95% CI]
500ms Interval			
20181206	Pol32	119	12.91 [10.59, 15.23]
20181122	Pol32	44	14.11 [9.94, 18.28]
20180511	Histone H3	126	15.69[12.95, 18.44]
1s Interval			
20180417	Cdc45	160	23.55[19.90, 27.2]
20180427	Cdc45	173	22.06[18.77, 25.34]
20180830	Cdc45	73	25.37[19.55, 31.19]
20180810	Mcm4	95	27.76[22.18, 33.34]
20180412	Pol2	107	20.73 [16.8, 24.66]
20180802	Pol2	120	18.75[15.4, 22.1]
20180803	Pol2	78	21.37[16.63, 26.11]
20180808	Dpb4	88	20.16[15.95, 24.37]
20180424	Dpb4	181	21.17[18.08, 24.25]
20180621	Pol3	138	22.8 [19, 26.61]
20180622	Pol3	39	33.15 [22.75, 43.56]
20180724	Pol3	69	15.84 [12.1, 19.58]
20190429	Pol3	78	21.8 [16.95, 26.63]
20180406	Pol32	182	26.95[23.04, 30.87]
20180429	Pol32	127	23.17 [19.14, 27.2]
20180809	Pol32	106	21.94 [17.77, 26.12]
20190531	ΔClb5-Pol3-Halo	52	25.06 [18.25, 31.87]
20190602	ΔClb5-Pol3-Halo	62	22.65 [17.0, 28.28]
20180618	Pol12	54	19.81[14.53, 25.1]
20180626	Pol12	140	16.67 [13.91, 19.43]
20190108	Pol12	59	15.59[11.61, 19.57]
20190713	Pri2	45	17.31 [12.25, 22.37]
20190718	Pri2	73	13.18 [10.16, 16.2]
20190806	Pol12-CIP	56	18.96 [14, 23.92]
20190807	Pol12-CIP	77	15.84 [12.31, 19.39]
20190813	Pol12-CIP	49	15.26 [10.99, 19.54]
20190814	Pol12-CIP	58	13.48 [10.01. 16.95]
20190111	Ctf4	306	23.72 [21.06, 26.38]
20190112	Ctf4	78	26.23 [20.41, 32.05]
20180904	Histone H3	471	23.12[21.03, 25.21]
8s Interval			
20180906	Cdc45	16	76.5 [39.02, 113.98]
20180913	Cdc45	14	58.86 [28.03, 89.69]
20180927	Cdc45	13	53.54 [24.44, 82.64]
20181005	Cdc45	68	92.23 [70.32, 114.16]
20181125	Cdc45	18	102.22 [55, 149.45]
20180817	Dpb4	53	53.28 [38.94, 67.63]
20180921	Dpb4	44	58.36 [41.12, 75.61]
20181204	Dpb4	31	102.45 [66.39, 138.52]
20180827	Pol32	31	72.26 [46.82, 97.69]
20180919	Pol32	22	53.45 [31.12, 75.79]
20181130	Pol32	36	75.11 [50.57, 99.65]
20190216	Pol32	41	58.54 [40.62, 76.45]
20180914	Pol12	12	70 [30.39, 109.61]
20180925	Pol12	28	28.29 [17.81, 38.76]
20181129	Pol12	15	51.2 [25.29, 77.11]
20190109	Pol12	12	34.67 [15.05, 54.28]
20190129	Pol12	28	48.57 [30.58, 66.56]
20190115	Ctf4	73	68.6 [52.86, 84.34]
20190116	Ctf4	74	70.59 [54.51, 86.68]
20180822	Histone H3	155	79.33 [66.84, 91.82]
20180907	Histone H3	108	104.15 [84.51, 123.79]
20s Interval			
20181009	Cdc45	15	192 [94.84, 289.16]
20181030	Cdc45	18	147.78 [79.51, 216.05]
20181107	Cdc45	31	189.68 [122.91, 256.45]
20181017	Dpb4	21	153.33 [87.75, 218.91]
20181102	Dpb4	51	114.12 [82.8, 145.44]
20181113	Dpb4	20	141 [79.21, 202.79]
20181011	Pol32	16	87 [44.63, 130.37]
20181031	Pol32	15	128 [63.22, 192.78]
20181106	Pol32	13	90.77 [41.43, 140.11]
20190329	Pol3	21	131.43 [75.22, 187.65]
20190403	Pol3	23	80 [47.31, 112.69]
20181116	Histone H3	60	153.67 [114.78, 192.55]

Bibliography

1. Watson, J.D. and F.H.C. Crick, *Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid.* Nature, 1953. **171**(4356): p. 737-738.
2. Meselson, M. and F.W. Stahl, *The Replication of DNA in Escherichia-Coli.* Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(7): p. 671-682.
3. Okazaki, R., et al., *Mechanism of DNA Replication - Possible Discontinuity of DNA Chain Growth.* Japanese Journal of Medical Science & Biology, 1967. **20**(3): p. 255-&.
4. Kapadia, N. and R. Reyes-Lamothe, *A quest for coordination among activities at the replisome.* Biochemical Society Transactions, 2019. **47**: p. 1067-1075.
5. Bell, S.P. and K. Labib, *Chromosome Duplication in Saccharomyces cerevisiae.* Genetics, 2016. **203**(3): p. 1027-1067.
6. Beattie, T.R. and R. Reyes-Lamothe, *A Replisome's journey through the bacterial chromosome.* Frontiers in Microbiology, 2015. **6**.
7. Forterre, P., *Why Are There So Many Diverse Replication Machineries?* Journal of Molecular Biology, 2013. **425**(23): p. 4714-4726.
8. Zeman, M.K. and K.A. Cimprich, *Causes and consequences of replication stress.* Nature Cell Biology, 2014. **16**(1): p. 2-9.
9. Reyes-Lamothe, R. and D.J. Sherratt, *The bacterial cell cycle, chromosome inheritance and cell growth.* Nature Reviews Microbiology, 2019. **17**(8): p. 467-478.
10. Dame, R.T., F.M. Rashid, and D.C. Grainger, *Chromosome organization in bacteria: mechanistic insights into genome structure and function.* Nat Rev Genet, 2019.
11. Taddei, A. and S.M. Gasser, *Structure and Function in the Budding Yeast Nucleus.* Genetics, 2012. **192**(1): p. 107-129.
12. Grover, P., J.S. Asa, and E.I. Campos, *H3-H4 Histone Chaperone Pathways.* Annual Review of Genetics, Vol 52, 2018. **52**: p. 109-130.
13. Feynman, R.P., *There's Plenty of Room at the Bottom An Invitation to Enter a New Field of Physics.* Resonance-Journal of Science Education, 2011. **16**(9): p. 890-904.
14. Elf, J. and I. Barkefors, *Single-Molecule Kinetics in Living Cells.* Annu Rev Biochem, 2018.
15. Beattie, T.R., et al., *Frequent exchange of the DNA polymerase during bacterial chromosome replication.* Elife, 2017. **6**.
16. Duina, A.A., M.E. Miller, and J.B. Keeney, *Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system.* Genetics, 2014. **197**(1): p. 33-48.
17. Duan, Z., et al., *A three-dimensional model of the yeast genome.* Nature, 2010. **465**(7296): p. 363-367.
18. Kim, S., et al., *The dynamic three-dimensional organization of the diploid yeast genome.* Elife, 2017. **6**.
19. Garcia-Muse, T. and A. Aguilera, *Transcription-replication conflicts: how they occur and how they are resolved.* Nature Reviews Molecular Cell Biology, 2016. **17**(9): p. 553-563.
20. Costa, A., I.V. Hood, and J.M. Berger, *Mechanisms for initiating cellular DNA replication.* Annu Rev Biochem, 2013. **82**: p. 25-54.

21. Katayama, T., et al., *Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC.* Nature Reviews Microbiology, 2010. **8**(3): p. 163-170.
22. Arias-Palomo, E., et al., *The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker.* Cell, 2013. **153**(2): p. 438-448.
23. Kaguni, J.M., *Replication initiation at the Escherichia coli chromosomal origin.* Current Opinion in Chemical Biology, 2011. **15**(5): p. 606-613.
24. Chodavarapu, S. and J.M. Kaguni, *Replication Initiation in Bacteria.* DNA Replication across Taxa, 2016. **39**: p. 1-30.
25. Lewis, J.S., S. Jergic, and N.E. Dixon, *The E. coli DNA Replication Fork.* DNA Replication across Taxa, 2016. **39**: p. 31-88.
26. Pham, T.M., et al., *A single-molecule approach to DNA replication in Escherichia coli cells demonstrated that DNA polymerase III is a major determinant of fork speed.* Molecular Microbiology, 2013. **90**(3): p. 584-596.
27. Yao, N.Y., et al., *Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(32): p. 13236-13241.
28. Tanner, N.A., et al., *Single-molecule studies of fork dynamics in Escherichia coli DNA replication.* Nature Structural & Molecular Biology, 2008. **15**(2): p. 170-176.
29. Reyes-Lamothe, R., et al., *Independent positioning and action of Escherichia coli replisomes in live cells.* Cell, 2008. **133**(1): p. 90-102.
30. Saner, N., et al., *Stochastic association of neighboring replicons creates replication factories in budding yeast.* Journal of Cell Biology, 2013. **202**(7): p. 1001-1012.
31. Yuan, Z.N., et al., *Ctf4 organizes sister replisomes and Pol alpha into a replication factory.* Elife, 2019. **8**.
32. Kitamura, E., J.J. Blow, and T.U. Tanaka, *Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories.* Cell, 2006. **125**(7): p. 1297-1308.
33. Yardimci, H., et al., *Uncoupling of Sister Replisomes during Eukaryotic DNA Replication.* Molecular Cell, 2010. **40**(5): p. 834-840.
34. Bleichert, F., M.R. Botchan, and J.M. Berger, *Mechanisms for initiating cellular DNA replication.* Science, 2017. **355**(6327).
35. Bell, S.P. and B. Stillman, *Atp-Dependent Recognition of Eukaryotic Origins of DNA-Replication by a Multiprotein Complex.* Nature, 1992. **357**(6374): p. 128-134.
36. Remus, D., et al., *Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing.* Cell, 2009. **139**(4): p. 719-730.
37. Labib, K., J.F.X. Diffley, and S.E. Kearsey, *G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus.* Nature Cell Biology, 1999. **1**(7): p. 415-422.
38. Ilves, I., et al., *Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins.* Molecular Cell, 2010. **37**(2): p. 247-258.
39. Yeeles, J.T.P., et al., *Regulated eukaryotic DNA replication origin firing with purified proteins.* Nature, 2015. **519**(7544): p. 431-+.
40. Muramatsu, S., et al., *CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol epsilon, and GINS in budding yeast.* Genes & Development, 2010. **24**(6): p. 602-612.

41. Sengupta, S., et al., *Dpb2 Integrates the Leading-Strand DNA Polymerase into the Eukaryotic Replisome.* Current Biology, 2013. **23**(7): p. 543-552.
42. Aparicio, O.M., *Location, location, location: it's all in the timing for replication origins.* Genes & Development, 2013. **27**(2): p. 117-128.
43. Knott, S.R.V., et al., *Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in S. cerevisiae.* Cell, 2012. **148**(1-2): p. 99-111.
44. Mantiero, D., et al., *Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast.* Embo Journal, 2011. **30**(23): p. 4805-4814.
45. Tanaka, S., et al., *Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing.* Curr Biol, 2011. **21**(24): p. 2055-63.
46. Ostrow, A.Z., et al., *Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in S. cerevisiae.* Proc Natl Acad Sci U S A, 2017. **114**(12): p. E2411-E2419.
47. Ebrahimi, H., et al., *Early initiation of a replication origin tethered at the nuclear periphery.* Journal of Cell Science, 2010. **123**(7): p. 1015-1019.
48. Sinha, N.K., C.F. Morris, and B.M. Alberts, *Efficient in vitro replication of double-stranded DNA templates by a purified T4 bacteriophage replication system.* J Biol Chem, 1980. **255**(9): p. 4290-3.
49. Chastain, P.D., et al., *Architecture of the replication complex and DNA loops at the fork generated by the bacteriophage T4 proteins.* Journal of Biological Chemistry, 2003. **278**(23): p. 21276-21285.
50. Park, K., et al., *Formation of a DNA loop at the replication fork generated by bacteriophage T7 replication proteins.* Journal of Biological Chemistry, 1998. **273**(9): p. 5260-5270.
51. Tanner, N.A., et al., *E. coli DNA replication in the absence of free beta clamps.* Embo Journal, 2011. **30**(9): p. 1830-1840.
52. Georgescu, R.E., et al., *Replisome mechanics: lagging strand events that influence speed and processivity.* Nucleic Acids Research, 2014. **42**(10): p. 6497-6510.
53. Kim, S.S., et al., *Coupling of a replicative polymerase and helicase: A tau-DnaB interaction mediates rapid replication fork movement.* Cell, 1996. **84**(4): p. 643-650.
54. Mitkova, A.V., S.M. Khopde, and S.B. Biswas, *Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of Escherichia coli in RNA primer synthesis.* Journal of Biological Chemistry, 2003. **278**(52): p. 52253-52261.
55. Bailey, S., W.K. Eliason, and T.A. Steitz, *Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase.* Science, 2007. **318**(5849): p. 459-463.
56. Zechner, E.L., C.A. Wu, and K.J. Marians, *Coordinated Leading-Strand and Lagging-Strand Synthesis at the Escherichia-Coli DNA-Replication Fork .3. A Polymerase-Primase Interaction Governs Primer Size.* Journal of Biological Chemistry, 1992. **267**(6): p. 4054-4063.
57. Swart, J.R. and M.A. Griep, *Primer synthesis kinetics by Escherichia coli primase on single-stranded DNA templates.* Biochemistry, 1995. **34**(49): p. 16097-16106.
58. Johnson, S.K., S. Bhattacharyya, and M.A. Griep, *DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates.* Biochemistry, 2000. **39**(4): p. 736-744.
59. Yoda, K. and T. Okazaki, *Specificity of Recognition Sequence for Escherichia-Coli Primase.* Molecular & General Genetics, 1991. **227**(1): p. 1-8.

60. Johnson, A. and M. O'Donnell, *Cellular DNA replicases: Components and dynamics at the replication fork.* Annual Review of Biochemistry, 2005. **74**: p. 283-315.
61. Kurth, I. and M. O'Donnell, *Replisome Dynamics during Chromosome Duplication.* EcoSal Plus, 2009. **3**(2).
62. Georgescu, R.E., et al., *Mechanism of polymerase collision release from sliding clamps on the lagging strand.* Embo Journal, 2009. **28**(19): p. 2981-2991.
63. Wu, C.A., E.L. Zechner, and K.J. Marians, *Coordinated Leading-Strand and Lagging-Strand Synthesis at the Escherichia-Coli DNA-Replication Fork .1. Multiple Effectors Act to Modulate Okazaki Fragment Size.* Journal of Biological Chemistry, 1992. **267**(6): p. 4030-4044.
64. Graham, J.E., K.J. Marians, and S.C. Kowalczykowski, *Independent and Stochastic Action of DNA Polymerases in the Replisome.* Cell, 2017. **169**(7): p. 1201-+.
65. Scherr, M.J., B. Safaric, and K.E. Duderstadt, *Noise in the Machine: Alternative Pathway Sampling is the Rule During DNA Replication.* Bioessays, 2018. **40**(2).
66. Kurth, I., R.E. Georgescu, and M.E. O'Donnell, *A solution to release twisted DNA during chromosome replication by coupled DNA polymerases.* Nature, 2013. **496**(7443): p. 119-+.
67. Sekedat, M.D., et al., *GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome.* Molecular Systems Biology, 2010. **6**.
68. Georgescu, R., et al., *Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(5): p. E697-E706.
69. Langston, L. and M. O'Donnell, *Action of CMG with strand-specific DNA blocks supports an internal unwinding mode for the eukaryotic replicative helicase.* Elife, 2017. **6**.
70. Kose, H.B., et al., *Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model.* Cell Reports, 2019. **26**(8): p. 2113-+.
71. Zhou, J.C., et al., *CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(16): p. 4141-4146.
72. Gambus, A., et al., *GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks.* Nature Cell Biology, 2006. **8**(4): p. 358-U41.
73. Gambus, A., et al., *A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome.* Embo Journal, 2009. **28**(19): p. 2992-3004.
74. Schauer, G.D. and M.E. O'Donnell, *Quality control mechanisms exclude incorrect polymerases from the eukaryotic replication fork.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(4): p. 675-680.
75. Simon, A.C., et al., *A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome.* Nature, 2014. **510**(7504): p. 293-+.
76. Gan, H.Y., et al., *Checkpoint Kinase Rad53 Couples Leading- and Lagging-Strand DNA Synthesis under Replication Stress.* Molecular Cell, 2017. **68**(2): p. 446-+.
77. Devbhandari, S., et al., *Chromatin Constrains the Initiation and Elongation of DNA Replication.* Molecular Cell, 2017. **65**(1): p. 131-141.

78. Georgescu, R.E., et al., *Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation.* Elife, 2015. **4**.
79. Stodola, J.L. and P.M. Burgers, *Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale.* Nature Structural & Molecular Biology, 2016. **23**(5): p. 402-408.
80. Ho, B., A. Baryshnikova, and G.W. Brown, *Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome.* Cell Systems, 2018. **6**(2): p. 192-+.
81. Chong, Y.T., et al., *Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis.* Cell, 2015. **161**(6): p. 1413-1424.
82. Kulak, N.A., et al., *Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells.* Nature Methods, 2014. **11**(3): p. 319-U300.
83. Pellegrini, L., *The Pol alpha-primase complex.* Subcell Biochem, 2012. **62**: p. 157-69.
84. Pavlov, Y.I., et al., *Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta.* Current Biology, 2006. **16**(2): p. 202-207.
85. Swan, M.K., et al., *Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta.* Nature Structural & Molecular Biology, 2009. **16**(9): p. 979-U107.
86. Hogg, M., et al., *Structural basis for processive DNA synthesis by yeast DNA polymerase epsilon.* Nature Structural & Molecular Biology, 2014. **21**(1): p. 49-+.
87. Yeeles, J.T.P., et al., *How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication.* Molecular Cell, 2017. **65**(1): p. 105-116.
88. Pursell, Z.F., et al., *Yeast DNA polymerase epsilon participates in leading-strand DNA replication.* Science, 2007. **317**(5834): p. 127-130.
89. McElhinny, S.A.N., et al., *Division of labor at the eukaryotic replication fork.* Molecular Cell, 2008. **30**(2): p. 137-144.
90. Georgescu, R.E., et al., *Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork.* Nature Structural & Molecular Biology, 2014. **21**(8): p. 664-670.
91. Clausen, A.R., et al., *Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation.* Nature Structural & Molecular Biology, 2015. **22**(3): p. 185-191.
92. Koh, K.D., et al., *Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA.* Nature Methods, 2015. **12**(3): p. 251-+.
93. Johnson, R.E., et al., *A Major Role of DNA Polymerase delta in Replication of Both the Leading and Lagging DNA Strands.* Molecular Cell, 2015. **59**(2): p. 163-175.
94. Dua, R., D.L. Levy, and J.L. Campbell, *Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain.* Journal of Biological Chemistry, 1999. **274**(32): p. 22283-22288.
95. Garbacz, M.A., et al., *Evidence that DNA polymerase delta contributes to initiating leading strand DNA replication in Saccharomyces cerevisiae.* Nature Communications, 2018. **9**.
96. Aria, V. and J.T.P. Yeeles, *Mechanism of Bidirectional Leading-Strand Synthesis Establishment at Eukaryotic DNA Replication Origins.* Molecular Cell, 2019. **73**(2): p. 199-+.
97. Kurat, C.F., et al., *Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates.* Molecular Cell, 2017. **65**(1): p. 117-130.

98. Katou, Y., et al., *S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex.* Nature, 2003. **424**(6952): p. 1078-1083.
99. Szyjka, S.J., C.J. Viggiani, and O.M. Aparicio, *Mrc1 is required for normal progression of replication forks throughout chromatin in S-cerevisiae.* Molecular Cell, 2005. **19**(5): p. 691-697.
100. Hodgson, B., A. Calzada, and K. Labib, *Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S Phase.* Molecular Biology of the Cell, 2007. **18**(10): p. 3894-3902.
101. Lewis, J.S., et al., *Single-molecule visualization of Saccharomyces cerevisiae leading-strand synthesis reveals dynamic interaction between MTC and the replisome.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(40): p. 10630-10635.
102. Looke, M., M.F. Maloney, and S.P. Bell, *Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase.* Genes & Development, 2017. **31**(3): p. 291-305.
103. Langston, L.D., et al., *Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks.* Elife, 2017. **6**.
104. Foltman, M., et al., *Eukaryotic Replisome Components Cooperate to Process Histones During Chromosome Replication.* Cell Reports, 2013. **3**(3): p. 892-904.
105. Evrin, C., et al., *Histone H2A-H2B binding by Pol alpha in the eukaryotic replisome contributes to the maintenance of repressive chromatin.* Embo Journal, 2018. **37**(19).
106. Gan, H.Y., et al., *The Mcm2-Ctf4-Pol alpha Axis Facilitates Parental Histone H3-H4 Transfer to Lagging Strands.* Molecular Cell, 2018. **72**(1): p. 140-+.
107. Yu, C.H., et al., *A mechanism for preventing asymmetric histone segregation onto replicating DNA strands.* Science, 2018. **361**(6409): p. 1386-+.
108. Liu, S.F., et al., *RPA binds histone H3-H4 and functions in DNA replication-coupled nucleosome assembly.* Science, 2017. **355**(6323): p. 415-419.
109. Smith, D.J. and I. Whitehouse, *Intrinsic coupling of lagging-strand synthesis to chromatin assembly.* Nature, 2012. **483**(7390): p. 434-U80.
110. Villa, F., et al., *Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase.* Molecular Cell, 2016. **63**(3): p. 385-396.
111. Samora, C.P., et al., *Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome.* Molecular Cell, 2016. **63**(3): p. 371-384.
112. Sasaki, M. and T. Kobayashi, *Ctf4 Prevents Genome Rearrangements by Suppressing DNA Double-Strand Break Formation and Its End Resection at Arrested Replication Forks.* Molecular Cell, 2017. **66**(4): p. 533-+.
113. Neylon, C., et al., *Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex.* Microbiol Mol Biol Rev, 2005. **69**(3): p. 501-26.
114. Dewar, J.M. and J.C. Walter, *Mechanisms of DNA replication termination.* Nat Rev Mol Cell Biol, 2017. **18**(8): p. 507-516.
115. Mulcair, M.D., et al., *A molecular mousetrap determines polarity of termination of DNA replication in E. coli.* Cell, 2006. **125**(7): p. 1309-19.
116. Valjavec-Gratian, M., T.A. Henderson, and T.M. Hill, *Tus-mediated arrest of DNA replication in Escherichia coli is modulated by DNA supercoiling.* Molecular Microbiology, 2005. **58**(3): p. 758-773.

117. Duggin, I.G. and S.D. Bell, *Termination Structures in the Escherichia coli Chromosome Replication Fork Trap.* Journal of Molecular Biology, 2009. **387**(3): p. 532-539.
118. Elshenawy, M.M., et al., *Replisome speed determines the efficiency of the Tus-Ter replication termination barrier.* Nature, 2015. **525**(7569): p. 394-+.
119. Midgley-Smith, S.L., et al., *Chromosomal over-replication in Escherichia coli recG cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed.* Nucleic Acids Research, 2018. **46**(15): p. 7701-7715.
120. Maric, M., et al., *Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication.* Science, 2014. **346**(6208): p. 440-+.
121. Maculins, T., et al., *Tethering of SCFDia2 to the Replisome Promotes Efficient Ubiquitylation and Disassembly of the CMG Helicase.* Current Biology, 2015. **25**(17): p. 2254-2259.
122. Reyes-Lamothe, R., D.J. Sherratt, and M.C. Leake, *Stoichiometry and Architecture of Active DNA Replication Machinery in Escherichia coli.* Science, 2010. **328**(5977): p. 498-501.
123. Shen, H., et al., *Single Particle Tracking: From Theory to Biophysical Applications.* Chemical Reviews, 2017. **117**(11): p. 7331-7376.
124. Mazza, D., et al., *A benchmark for chromatin binding measurements in live cells.* Nucleic Acids Research, 2012. **40**(15).
125. Stracy, M., et al., *Single-Molecule Imaging of Transcription, Chromosome Organization, and DNA Repair in Live Bacteria.* Biophysical Journal, 2016. **110**(3): p. 20a-21a.
126. Crocker, J.C. and D.G. Grier, *Methods of digital video microscopy for colloidal studies.* Journal of Colloid and Interface Science, 1996. **179**(1): p. 298-310.
127. Jaqaman, K., et al., *Robust single-particle tracking in live-cell time-lapse sequences.* Nature Methods, 2008. **5**(8): p. 695-702.
128. Tinevez, J.Y., et al., *TrackMate: An open and extensible platform for single-particle tracking.* Methods, 2017. **115**: p. 80-90.
129. Ries, J., et al., *A simple, versatile method for GFP-based super-resolution microscopy via nanobodies.* Nat Methods, 2012. **9**(6): p. 582-4.
130. Puchner, E.M., et al., *Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory.* Proc Natl Acad Sci U S A, 2013. **110**(40): p. 16015-20.
131. Kaplan, C. and H. Ewers, *Optimized sample preparation for single-molecule localization-based superresolution microscopy in yeast.* Nature Protocols, 2015. **10**(7): p. 1007-1021.
132. Smith, C., et al., *In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p.* Journal of Cell Biology, 2015. **211**(6): p. 1121-1130.
133. Wollman, A.J.M. and M.C. Leake, *Millisecond single-molecule localization microscopy combined with convolution analysis and automated image segmentation to determine protein concentrations in complexly structured, functional cells, one cell at a time.* Faraday Discussions, 2015. **184**: p. 401-424.
134. Ball, D.A., et al., *Single molecule tracking of Ace1p in Saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin.* Nucleic Acids Res, 2016. **44**(21): p. e160.
135. Mehta, G.D., et al., *Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast.* Mol Cell, 2018.

136. Phillips, R., *Physical biology of the cell*. Second edition / ed. 2013, London New York, NY: Garland Science. xxx, 1057 pages.
137. Maslanka, R., M. Kwolek-Mirek, and R. Zadrag-Tecza, *Autofluorescence of yeast Saccharomyces cerevisiae cells caused by glucose metabolism products and its methodological implications.* Journal of Microbiological Methods, 2018. **146**: p. 55-60.
138. Gebhardt, J.C.M., et al., *Single-molecule imaging of transcription factor binding to DNA in live mammalian cells.* Nature Methods, 2013. **10**(5): p. 421-+.
139. Yu, J., *Single-Molecule Studies in Live Cells.* Annual Review of Physical Chemistry, Vol 67, 2016. **67**: p. 565-585.
140. England, C.G., H.M. Luo, and W.B. Cai, *HaloTag Technology: A Versatile Platform for Biomedical Applications.* Bioconjugate Chemistry, 2015. **26**(6): p. 975-986.
141. Grimm, J.B., et al., *Bright photoactivatable fluorophores for single-molecule imaging.* Nature Methods, 2016. **13**(12): p. 985-+.
142. Moen, E., et al., *Deep learning for cellular image analysis.* Nat Methods, 2019. **16**(12): p. 1233-1246.
143. Natsume, T. and M.T. Kanemaki, *Conditional Degrons for Controlling Protein Expression at the Protein Level.* Annual Review of Genetics, Vol 51, 2017. **51**: p. 83-102.
144. Maisnier-Patin, S., K. Nordstrom, and S. Dasgupta, *Replication arrests during a single round of replication of the Escherichia coli chromosome in the absence of DnaC activity.* Mol Microbiol, 2001. **42**(5): p. 1371-82.
145. Moore, P.D., et al., *Sites of termination of in vitro DNA synthesis on ultraviolet- and N-acetylaminofluorene-treated phi X174 templates by prokaryotic and eukaryotic DNA polymerases.* Proc Natl Acad Sci U S A, 1981. **78**(1): p. 110-4.
146. Mettrick, K.A. and I. Grainge, *Stability of blocked replication forks in vivo.* Nucleic Acids Res, 2016. **44**(2): p. 657-68.
147. Heller, R.C. and K.J. Marians, *Replication fork reactivation downstream of a blocked nascent leading strand.* Nature, 2006. **439**(7076): p. 557-62.
148. Yeeles, J.T.P. and K.J. Marians, *The Escherichia coli Replisome Is Inherently DNA Damage Tolerant.* Science, 2011. **334**(6053): p. 235-238.
149. Pomerantz, R.T. and M. O'Donnell, *The replisome uses mRNA as a primer after colliding with RNA polymerase.* Nature, 2008. **456**(7223): p. 762-6.
150. Heller, R.C. and K.J. Marians, *Replisome assembly and the direct restart of stalled replication forks.* Nat Rev Mol Cell Biol, 2006. **7**(12): p. 932-43.
151. Moolman, M.C., et al., *Slow unloading leads to DNA-bound beta2-sliding clamp accumulation in live Escherichia coli cells.* Nat Commun, 2014. **5**: p. 5820.
152. Su'etsugu, M. and J. Errington, *The replicase sliding clamp dynamically accumulates behind progressing replication forks in Bacillus subtilis cells.* Mol Cell, 2011. **41**(6): p. 720-32.
153. Manley, S., et al., *High-density mapping of single-molecule trajectories with photoactivated localization microscopy.* Nature Methods, 2008. **5**(2): p. 155-157.
154. Uphoff, S., et al., *Single-molecule DNA repair in live bacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(20): p. 8063-8068.
155. McEvoy, A.L., et al., *mMaple: a photoconvertible fluorescent protein for use in multiple imaging modalities.* PLoS One, 2012. **7**(12): p. e51314.

156. Hammar, P., et al., *Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation.* Nat Genet, 2014. **46**(4): p. 405-8.

157. Yuan, Q., et al., *DNA Polymerase III, but Not Polymerase IV, Must Be Bound to a tau-Containing DnaX Complex to Enable Exchange into Replication Forks.* J Biol Chem, 2016. **291**(22): p. 11727-35.

158. Maki, H. and A. Kornberg, *The polymerase subunit of DNA polymerase III of Escherichia coli. II. Purification of the alpha subunit, devoid of nuclease activities.* J Biol Chem, 1985. **260**(24): p. 12987-92.

159. Sinha, N.K. and D.P. Snustad, *Mechanism of inhibition of deoxyribonucleic acid synthesis in Escherichia coli by hydroxyurea.* J Bacteriol, 1972. **112**(3): p. 1321-4.

160. Leu, F.P., R. Georgescu, and M. O'Donnell, *Mechanism of the E. coli tau processivity switch during lagging-strand synthesis.* Mol Cell, 2003. **11**(2): p. 315-27.

161. Li, X. and K.J. Marians, *Two distinct triggers for cycling of the lagging strand polymerase at the replication fork.* J Biol Chem, 2000. **275**(44): p. 34757-65.

162. Kim, S., et al., *Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement.* Cell, 1996. **84**(4): p. 643-50.

163. Graham, J.S., R.C. Johnson, and J.F. Marko, *Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA.* Nucleic Acids Res, 2011. **39**(6): p. 2249-59.

164. Sidorova, N.Y., T. Scott, and D.C. Rau, *DNA concentration-dependent dissociation of EcoRI: direct transfer or reaction during hopping.* Biophys J, 2013. **104**(6): p. 1296-303.

165. Gibb, B., et al., *Concentration-dependent exchange of replication protein A on single-stranded DNA revealed by single-molecule imaging.* PLoS One, 2014. **9**(2): p. e87922.

166. Chen, T.Y., et al., *Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells.* Nat Commun, 2015. **6**: p. 7445.

167. Aberg, C., K.E. Duderstadt, and A.M. van Oijen, *Stability versus exchange: a paradox in DNA replication.* Nucleic Acids Research, 2016. **44**(10): p. 4846-4854.

168. Yang, J., et al., *The dynamic processivity of the T4 DNA polymerase during replication.* Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8289-94.

169. Johnson, D.E., et al., *Exchange of DNA polymerases at the replication fork of bacteriophage T7.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5312-7.

170. Geertsema, H.J., et al., *Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery.* Proc Natl Acad Sci U S A, 2014. **111**(11): p. 4073-8.

171. Loparo, J.J., et al., *Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3584-9.

172. Yeeles, J.T., *Discontinuous leading-strand synthesis: a stop-start story.* Biochem Soc Trans, 2014. **42**(1): p. 25-34.

173. Wang, T.C., *Discontinuous or semi-discontinuous DNA replication in Escherichia coli?* Bioessays, 2005. **27**(6): p. 633-6.

174. Okazaki, R., et al., *Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains.* Proc Natl Acad Sci U S A, 1968. **59**(2): p. 598-605.

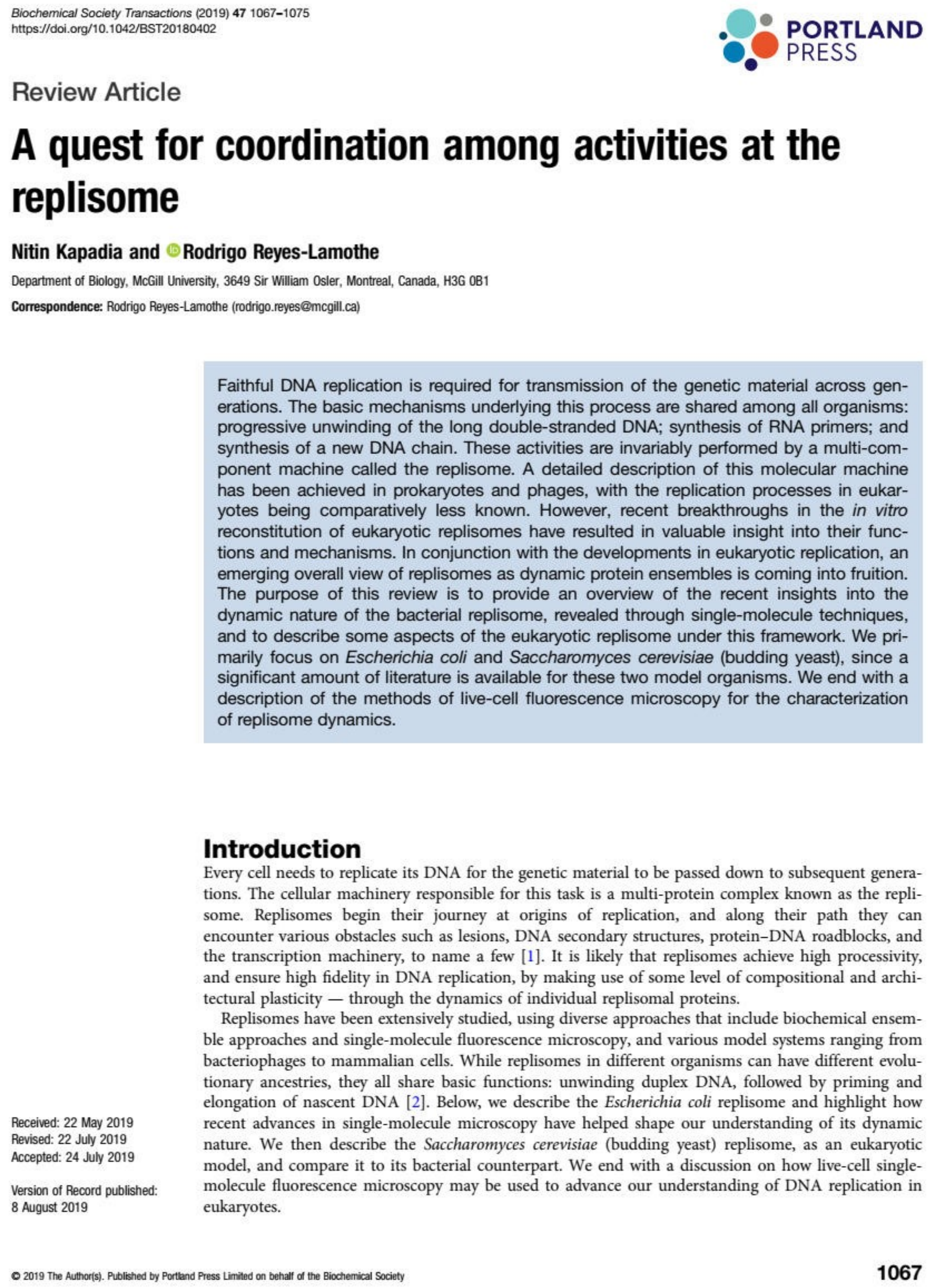
175. Amado, L. and A. Kuzminov, *Low-molecular-weight DNA replication intermediates in Escherichia coli: mechanism of formation and strand specificity.* J Mol Biol, 2013. **425**(22): p. 4177-91.
176. Amado, L. and A. Kuzminov, *The replication intermediates in Escherichia coli are not the product of DNA processing or uracil excision.* Journal of Biological Chemistry, 2006. **281**(32): p. 22635-22646.
177. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-6645.
178. Nguyen, A., Daugherty PS., *Evolutionary optimization of fluorescent proteins for intracellular FRET.* Nat Biotechnol, 2005. **23**(3): p. 355-360.
179. Wang, X., et al., *The two Escherichia coli chromosome arms locate to separate cell halves.* Genes Dev, 2006. **20**(13): p. 1727-1731.
180. Miller, H., et al., *Superresolution imaging of single DNA molecules using stochastic photoblinking of minor groove and intercalating dyes.* Methods, 2015. **88**: p. 81-8.
181. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
182. Andrae, R., T. Schulze-Hartung, and P. Melchior, *Dos and don'ts of reduced chi-squared.* arXiv preprint arXiv:1012.3754, 2010.
183. Tokunaga, M., N. Imamoto, and K. Sakata-Sogawa, *Highly inclined thin illumination enables clear single-molecule imaging in cells.* Nature Methods, 2008. **5**(2): p. 159-161.
184. Holden, S.J., S. Uphoff, and A.N. Kapanidis, *DAOSTORM: an algorithm for high-density super-resolution microscopy.* Nat Methods, 2011. **8**(4): p. 279-80.
185. Myung, I.J., *Tutorial on maximum likelihood estimation.* Journal of mathematical Psychology, 2003. **47**(1): p. 90-100.
186. Woody, M.S., et al., *MEMLET: An Easy-to-Use Tool for Data Fitting and Model Comparison Using Maximum-Likelihood Estimation.* Biophys J, 2016. **111**(2): p. 273-82.
187. Balakrishnan, N. and A.P. Basu, *The Exponential Distribution: Theory.* Methods and Applications, Gordon and Breach Publishers, 1995.
188. Durisic, N., et al., *Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate.* Nature Methods, 2014. **11**(2): p. 156-+.
189. Garcia-Parajo, M.F., et al., *Real-time light-driven dynamics of the fluorescence emission in single green fluorescent protein molecules.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7237-42.
190. Hager, G.L., J.G. McNally, and T. Misteli, *Transcription Dynamics.* Molecular Cell, 2009. **35**(6): p. 741-753.
191. Stracy, M., et al., *Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(32): p. E4390-E4399.
192. Xie, X.S., et al., *Single-molecule approach to molecular biology in living bacterial cells.* Annual Review of Biophysics, 2008. **37**: p. 417-444.
193. Sprouse, R.O., et al., *Regulation of TATA-binding protein dynamics in living yeast cells.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(36): p. 13304-13308.

194. Mueller, F., et al., *FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know?* Current Opinion in Cell Biology, 2010. **22**(3): p. 403-411.
195. Sprague, B.L., et al., *Analysis of binding reactions by fluorescence recovery after photobleaching.* Biophysical Journal, 2004. **86**(6): p. 3473-3495.
196. Christensen, M.O., et al., *Dynamics of human DNA topoisomerases II alpha and II beta in living cells.* Journal of Cell Biology, 2002. **157**(1): p. 31-44.
197. Etheridge, T.J., et al., *Quantification of DNA-associated proteins inside eukaryotic cells using single-molecule localization microscopy.* Nucleic Acids Research, 2014. **42**(19).
198. Callegari, A., et al., *Single-molecule dynamics and genome-wide transcriptomics reveal that NF-kB (p65)-DNA binding times can be decoupled from transcriptional activation.* Plos Genetics, 2019. **15**(1).
199. Angermueller, C., et al., *Deep learning for computational biology.* Mol Syst Biol, 2016. **12**(7): p. 878.
200. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.
201. Ho, H.N., et al., *Identification of Multiple Kinetic Populations of DNA-Binding Proteins in Live Cells.* Biophysical Journal, 2019. **117**(5): p. 950-961.
202. Kimura, H. and P.R. Cook, *Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B.* J Cell Biol, 2001. **153**(7): p. 1341-53.
203. Boothe, T., et al., *A tunable refractive index matching medium for live imaging cells, tissues and model organisms.* Elife, 2017. **6**.
204. Shihavuddin, A., et al., *Smooth 2D manifold extraction from 3D image stack.* Nat Commun, 2017. **8**: p. 15554.
205. Michalet, X. and A.J. Berglund, *Optimal diffusion coefficient estimation in single-particle tracking.* Phys Rev E Stat Nonlin Soft Matter Phys, 2012. **85**(6 Pt 1): p. 061916.
206. Javer, A., et al., *Persistent super-diffusive motion of Escherichia coli chromosomal loci.* Nature Communications, 2014. **5**.
207. Hajjoul, H., et al., *High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome.* Genome Res, 2013. **23**(11): p. 1829-38.
208. Weber, S.C., A.J. Spakowitz, and J.A. Theriot, *Bacterial Chromosomal Loci Move Subdiffusively through a Viscoelastic Cytoplasm.* Physical Review Letters, 2010. **104**(23).
209. Lopes, M., M. Foiani, and J.M. Sogo, *Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions.* Mol Cell, 2006. **21**(1): p. 15-27.
210. Yang, Y., et al., *Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast Saccharomyces cerevisiae.* PLoS Genet, 2008. **4**(11): p. e1000264.
211. Beattie, T.R., et al., *Frequent exchange of the DNA polymerase during bacterial chromosome replication.* Elife, 2017. **6**.
212. Lewis, J.S., et al., *Single-molecule visualization of fast polymerase turnover in the bacterial replisome.* Elife, 2017. **6**.
213. Liao, Y., et al., *Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells.* Biophys J, 2016. **111**(12): p. 2562-2569.

214. Los, G.V., et al., *HaloTag: a novel protein labeling technology for cell imaging and protein analysis.* ACS Chem Biol, 2008. **3**(6): p. 373-82.
215. Yeeles, J.T., et al., *Regulated eukaryotic DNA replication origin firing with purified proteins.* Nature, 2015. **519**(7544): p. 431-5.
216. Liachko, I., et al., *A comprehensive genome-wide map of autonomously replicating sequences in a naive genome.* PLoS Genet, 2010. **6**(5): p. e1000946.
217. Saner, N., et al., *Stochastic association of neighboring replicons creates replication factories in budding yeast.* J Cell Biol, 2013. **202**(7): p. 1001-12.
218. McCune, H.J., et al., *The temporal program of chromosome replication: genomewide replication in clb5{Delta} Saccharomyces cerevisiae.* Genetics, 2008. **180**(4): p. 1833-47.
219. Sanders, G.M., H.G. Dallmann, and C.S. McHenry, *Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases.* Mol Cell, 2010. **37**(2): p. 273-81.
220. Huang, M.E., et al., *The Saccharomyces cerevisiae protein YJR043C (Pol32) interacts with the catalytic subunit of DNA polymerase alpha and is required for cell cycle progression in G2/M.* Mol Gen Genet, 1999. **260**(6): p. 541-50.
221. Bermudez, V.P., et al., *Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication.* J Biol Chem, 2010. **285**(13): p. 9493-505.
222. Zhou, J.C., et al., *CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome.* Proc Natl Acad Sci U S A, 2017. **114**(16): p. 4141-4146.
223. Marians, K.J., *Lesion Bypass and the Reactivation of Stalled Replication Forks.* Annual Review of Biochemistry, Vol 87, 2018. **87**: p. 217-238.
224. Sanchez, A., S. Choubey, and J. Kondev, *Regulation of Noise in Gene Expression.* Annual Review of Biophysics, Vol 42, 2013. **42**: p. 469-491.
225. Huh, D. and J. Paulsson, *Non-genetic heterogeneity from stochastic partitioning at cell division.* Nature Genetics, 2011. **43**(2): p. 95-U32.
226. Lewis, J.S., et al., *Tunability of DNA Polymerase Stability during Eukaryotic DNA Replication.* Mol Cell, 2019.
227. Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nature Biotechnology, 2018. **36**(9): p. 880-+.
228. O'Donnell, M. and H.L. Li, *The Eukaryotic Replisome Goes Under the Microscope.* Current Biology, 2016. **26**(6): p. R247-R256.

Annex

A Quest for Coordination Among Activities at the Replisome



Biochemical Society Transactions (2019) **47** 1067–1075
https://doi.org/10.1042/BST20180402

PORTLAND PRESS

Review Article

A quest for coordination among activities at the replisome

Nitin Kapadia and Rodrigo Reyes-Lamothe

Department of Biology, McGill University, 3649 Sir William Osler, Montreal, Canada, H3G 0B1

Correspondence: Rodrigo Reyes-Lamothe (rodrigo.reyes@mcgill.ca)

Faithful DNA replication is required for transmission of the genetic material across generations. The basic mechanisms underlying this process are shared among all organisms: progressive unwinding of the long double-stranded DNA; synthesis of RNA primers; and synthesis of a new DNA chain. These activities are invariably performed by a multi-component machine called the replisome. A detailed description of this molecular machine has been achieved in prokaryotes and phages, with the replication processes in eukaryotes being comparatively less known. However, recent breakthroughs in the *in vitro* reconstitution of eukaryotic replisomes have resulted in valuable insight into their functions and mechanisms. In conjunction with the developments in eukaryotic replication, an emerging overall view of replisomes as dynamic protein ensembles is coming into fruition. The purpose of this review is to provide an overview of the recent insights into the dynamic nature of the bacterial replisome, revealed through single-molecule techniques, and to describe some aspects of the eukaryotic replisome under this framework. We primarily focus on *Escherichia coli* and *Saccharomyces cerevisiae* (budding yeast), since a significant amount of literature is available for these two model organisms. We end with a description of the methods of live-cell fluorescence microscopy for the characterization of replisome dynamics.

Introduction

Every cell needs to replicate its DNA for the genetic material to be passed down to subsequent generations. The cellular machinery responsible for this task is a multi-protein complex known as the replisome. Replisomes begin their journey at origins of replication, and along their path they can encounter various obstacles such as lesions, DNA secondary structures, protein–DNA roadblocks, and the transcription machinery, to name a few [1]. It is likely that replisomes achieve high processivity, and ensure high fidelity in DNA replication, by making use of some level of compositional and architectural plasticity — through the dynamics of individual replisomal proteins.

Replisomes have been extensively studied, using diverse approaches that include biochemical ensemble approaches and single-molecule fluorescence microscopy, and various model systems ranging from bacteriophages to mammalian cells. While replisomes in different organisms can have different evolutionary ancestries, they all share basic functions: unwinding duplex DNA, followed by priming and elongation of nascent DNA [2]. Below, we describe the *Escherichia coli* replisome and highlight how recent advances in single-molecule microscopy have helped shape our understanding of its dynamic nature. We then describe the *Saccharomyces cerevisiae* (budding yeast) replisome, as an eukaryotic model, and compare it to its bacterial counterpart. We end with a discussion on how live-cell single-molecule fluorescence microscopy may be used to advance our understanding of DNA replication in eukaryotes.

Received: 22 May 2019
Revised: 22 July 2019
Accepted: 24 July 2019

Version of Record published:
8 August 2019

E. coli replisome

The *E. coli* replisome is composed of at least 12 different proteins that arrange into the following stable subcomplexes: the helicase (DnaB), the primase (DnaG), the DNA polymerase III (Pol III), the processivity factor β clamp, the clamp loader, and the single-strand binding protein SSB (SSB_4) [3,4] (Figure 1A). Multiple protein–protein interactions among these subcomplexes result in the formation of a compact protein structure. DnaB and DnaG interact with each other [5]. Up to three copies of Pol III and the clamp loader are physically coupled by the interaction between α and τ subunits, forming the Pol III* subcomplex [6]. The interaction between Pol III* and β clamp form the Pol III holoenzyme (HE) [7]. Finally, the τ subunit mediates the interaction between DnaB helicase and Pol III HE [8]. This architecture is thought to facilitate the coordination of activities during DNA synthesis, specially at the lagging strand: primer synthesis, clamp loading, and recruitment of the Pol III — potentially the same copy that synthesized the previous Okazaki fragment — can rapidly succeed each other. Replisome architecture also increases the chances of simultaneous synthesis at both strands, due to the physical connection between multiple copies of Pol III. Coupling of these activities was suggested in the trombone model [9], which likens the growth of the single-strand DNA (ssDNA)-loop between the Pol III and the helicase at the lagging strand to the movement of a trombone's slide.

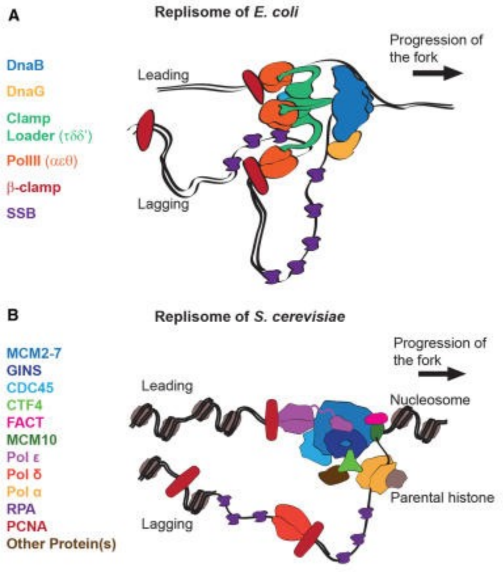


Figure 1. Model of the replisome.

(**A**) *E. coli* replisome. The diagram shows a model for the architecture of the bacterial replisome. The homohexameric DnaB helicase opens the double-stranded DNA, generating a forked DNA structure. The DnaG primase interacts with DnaB. The τ subunit, which is part of the clamp loader, interacts with DnaB and with the DNA Pol III, serving as a link between these two subassemblies. Multiple copies of τ in the clamp loader result in the oligomerization of Pol III, allowing for the coupled synthesis of the two strands. A copy of the β clamp binds to each engaged copy of Pol III. At the end of each Okazaki fragment, the Pol III at the lagging strand unbinds from β clamp, which remains bound on DNA. SSB covers the single-strand DNA that accumulates during Okazaki fragment synthesis. (**B**) Budding yeast replisome. The diagram shows a model for the architecture of the yeast replisome. The functional helicase (called CMG) is composed of the heterohexamer MCM2-7, the heterotetramer GINS complex, and CDC45. At the leading strand, Pol ε interacts with the CMG helicase through GINS. At the lagging strand, Pol α interacts with the homotetramer CTF4, which in turn interacts with the CMG helicase. In contrast, Pol δ which is expected to work at the lagging strand has no reported connection with the CMG. MCM10 is thought to have a diverse role at the replisome, not discussed here. Mrc1–Tof1–Csm3 (MTC) complex and RFC are omitted for clarity.

The physical link between enzymes at the replisome has led to the assumption that their activities are actively synchronized. However, an emerging perspective rejects this idea and instead proposes that near-synchronous synthesis can still be achieved with a stochastic sampling of activity rates by individual copies of enzymes (discussed in [10]). An example of these two views — deterministic and stochastic — is found in the models that match the DNA synthesis rates at the leading and lagging strand. A number of deterministic models have proposed mechanisms that increase the synthesis rate of the Pol III at the lagging strand, or that decrease the rate at the leading strand, to compensate for the time invested in priming, clamp loading, and Pol III recruitment [10,11]. In contrast, a stochastic model proposes that the rate of DNA synthesis at either strand is assigned stochastically and independently of the rate of the other strand [10,12]. The evidence for the stochastic model comes from visualizing synthesis of single DNA molecules, through single-molecule microscopy, where individual traces indicated that synthesis was interspersed with stochastic pauses, followed by a switch in the rate of synthesis, independent from the previous rate [12]. In this model, variance in the rates will frequently cause one of the two strands to trail behind, despite similar overall average synthesis rates. However, switches in the rates of synthesis can potentially reverse this discrepancy, leading to an averaged effect of near-synchronous synthesis.

The second shift in our understanding of the replisome has come from studying the binding stability among subunits, also through single-molecule microscopy. Until recently, it was assumed that once assembled, the helicase and Pol III* would act processively for hundreds of kilobase pairs. Studies of single replisomes using rolling-circle replication — which consists of a double-stranded circular template with a single-stranded tail, where the replisome can cycle potentially an unlimited number of times — demonstrated processivities of over 40 kbp, but with individual instances reaching lengths of few hundreds kbp [13–15]. Subsequent experiments using live-cell single-molecule fluorescence microscopy have shown that turnover is much more frequent than previously appreciated, with dissociation occurring every few seconds, indicating a much lower processivity of Pol III* [16–18]. Frequent turnover can only be observed *in vitro* when an excess of Pol III* is present in the buffer [17], which explains why it was not detected in previous studies. Turnover is more rapid during active replication, suggesting that DNA synthesis may exert physical strain on the replisome, and result in dissociation of some of its components. This model matches the proposed torsion buildup on DNA as a result of the physical connection between active polymerases at the two strands [19]. A dynamic composition of the replisome might also be advantageous to facilitate lesion skipping on damaged DNA — suggesting that the replisome has an inherent ability to bypass lesions — and to co-ordinate handover of the elongating strand to specialized polymerases for damage bypass [20].

Saccharomyces cerevisiae replisome

Many of the eukaryotic replisome subunits do not share common ancestry with bacteria [2]. However, since they accomplish similar functions, it is unknown the extent to which the strategies they used are shared. The core eukaryotic replisome is composed of over 30 different proteins that include the CMG helicase (composed of CDC45, MCM2–7, GINS), Pol ε, Pol δ, Pol α, the structural protein CTF4, the processivity factor proliferating nuclear cell antigen (PCNA), the clamp loader replication factor C (RFC) and the single-stranded binding protein, replication protein A (RPA) [21] (Figure 1B). Multiple other proteins interact with the replisome (described in [21]), known as the replication progression complex (RPC), but here we focus on the minimal composed replisome for conciseness. Pol ε and Pol δ are the main replicative polymerases, while Pol α is responsible for synthesizing RNA/DNA primers, primarily on the lagging-strand. Multiple sources of evidence suggest that Pol δ and Pol ε synthesize the lagging strand and leading strands, respectively, although alternative models exist which assign a role at the leading strand to Pol δ [22–33]. Reminiscent of the bacterial replisome, Pol ε forms a stable interaction with the GINS subunit of CMG helicase, [28,34] while Pol α binds to CTF4, which in turn binds to GINS [35], suggesting there is some connection between the leading-strand and lagging-strand components. However, there is no reported interaction between Pol δ and the CMG [28,34,36,37]. Furthermore, it is believed that the RFC clamp loader only acts transiently at the replisome, without forming stable interactions that could bridge it to the CMG helicase.

The current model for the architecture of the eukaryotic replisome with Pol δ being physically disconnected, suggests replisome dynamics that differ to the bacterial replisome, and leads to question how excess ssDNA exposure is minimized when there is no coupling between the helicase and the lagging-strand polymerase. It also leaves unexplained how synchronous synthesis between the two strands, reported using both *in vivo* and *in vitro* approaches, is achieved [26,38,39].

A possible explanation is that the slow progression of the replication fork, estimated at ~ 25 bp/s in budding yeast [40], may provide enough time for the efficient recruitment of RFC and subsequently Pol δ, and that a much faster rate of synthesis by Pol δ can compensate for the delay. In contrast with bacteria, where the same polymerase synthesizes the two strands, faster synthesis by Pol δ over Pol ε could be explained by their difference in structure. This model also requires that the time needed for the recruitment of a new copy of RFC and Pol δ at the replication fork, which is given by their inherent diffusion coefficients and the number of available copies, must be short. However, in physiological conditions, Pol δ synthesizes at a rate of ~50 bp/s, which is not sufficiently different from the speed of the replication fork [41]. Furthermore, the reported abundance of Pol δ in the cell suggests that it is not high enough for efficient lagging-strand synthesis, especially at the peak of S-phase when hundreds of replisomes are estimated to be active [42–45]. Thus, it seems unlikely that the rate of synthesis by Pol δ can compensate for the additional events at the lagging strand. It is nevertheless conceivable that some aspects of DNA replication in cells may minimize DNA synthesis interruptions even in scenarios dependent on the stochastic recruitment of Pol δ. One such possible mechanism is the proposed clustering of replication origins [46,47], which could help recycle these proteins from neighbouring replisomes and facilitate their recruitment at the fork by increasing their local concentration.

In eukaryotes, DNA is wrapped around histone cores (collectively referred to as nucleosomes) that are spaced apart ~160 bp in chromatin [48]. This implies that in addition to activities directly related to DNA replication, the eukaryotic replisome must also progress through and help preserve chromatin. Some activities enable the progression of the replication fork (the DNA structure at the replisome) as they encounter nucleosomes (e.g. FACT [49]), while others facilitate the transfer of nucleosomes — which may carry posttranslational modifications — from the front of the replication fork to the nascent DNA. Multiple replisome subunits have histone chaperone activity, including MCM2, Pol α, CTF4, Pol ε, and RPA [21,50–54], but the details of how they contribute to these processes is still emerging. Given the recent insight into the role of stochasticity from the bacterial replisome, we think it is unlikely that histones are being transferred down a fixed, deterministic path — one histone chaperone protein to the next — but rather, the multiple histone interaction sites on the replisome serve to retain them locally and increase the likelihood of them binding to nascent DNA. Chromatin also has important roles during the synthesis and maturation of Okazaki fragments. In budding yeast, Okazaki fragments have a narrow range of ~160 bp, nearly matching the inter-nucleosomal distance [48]. *In vitro*, chromatin is required to maintain this tight distribution of lengths [39,49], and can still be observed even in conditions where Pol α is not coupled to the CMG — by removing CTF4 from the reaction [49].

Chromatin may also play a role in the dynamics of the leading-strand polymerase, and influence whether it frequently unbinds from DNA like its bacterial counterpart, Pol III. While Pol ε may likely encounter obstacles and DNA topology issues like PolIII, it also has an unusual flexible linker that could allow it to bypass obstacles, while remaining stably connected to the replisome [33] (Figure 1B). It is possible that the flexible linker was selected for to increase the frequency of deposition of PCNA clamps on the leading strand, as seen *in vitro* [55]. PCNA plays a role in assembling new nucleosomes, though its interaction with chromatin assembly factor 1 (CAF-1), which is involved in depositing H3–H4 tetramers [21,56]. Given how frequently new nucleosomes need to be assembled, and the weak interaction of Pol ε to PCNA, it is possible that the flexible nature allows Pol ε to remain attached to CMG helicase when it unbinds from PCNA — giving access to the clamp loader to load new PCNA — and subsequently engage again with the leading-strand. This behaviour can also be exploited when bypassing obstacles on DNA. Another suggested use for the flexible linker is that it allows Pol ε to take over synthesis after Pol δ initiation in models where Pol δ contributes to the synthesis of the leading-strand [32,33].

Study of DNA replication using live-cell fluorescence microscopy

Fluorescence microscopy has been an invaluable method to study the localization, dynamics, and stoichiometries of proteins, especially in living cells. In particular, single-molecule approaches have allowed us to obtain distributions of variables rather than simply ensemble averages, giving us a more stochastic and dynamic view of biology, with a stronger emphasis on the probabilistic nature of biological processes. Indeed, the recent discoveries of the dynamic and stochastic nature of replisome in bacteria are all due to these approaches.

In vitro studies of *E. coli* DNA replication have provided important information on the factors affecting the processivity and rates of individual replisomes [11,13,57]. A complementary approach is to study this process in live cells, where enzymes are in their natural milieu. With recent advances in microscopy and genetics, we and others have been able to successfully pursue this avenue. Specifically, single-particle tracking photoactivated localization microscopy (sptPALM), has the advantage of being able to directly observe the dynamics of single copies of replisomal proteins [16,18,58]. It was through this technique, as highlighted above, that we were able to directly visualize the dissociation of PolIII*, as indicated by the disappearance of fluorescent foci — much faster than the measured photobleaching rate — using long-capture (Figure 2) [16].

SptPALM requires tagging the protein of interest (POI) with a photoactivatable (PA) fluorescent protein (FP) [59]. These are proteins that are normally in a dark state, but upon excitation with low 405 nm wavelength

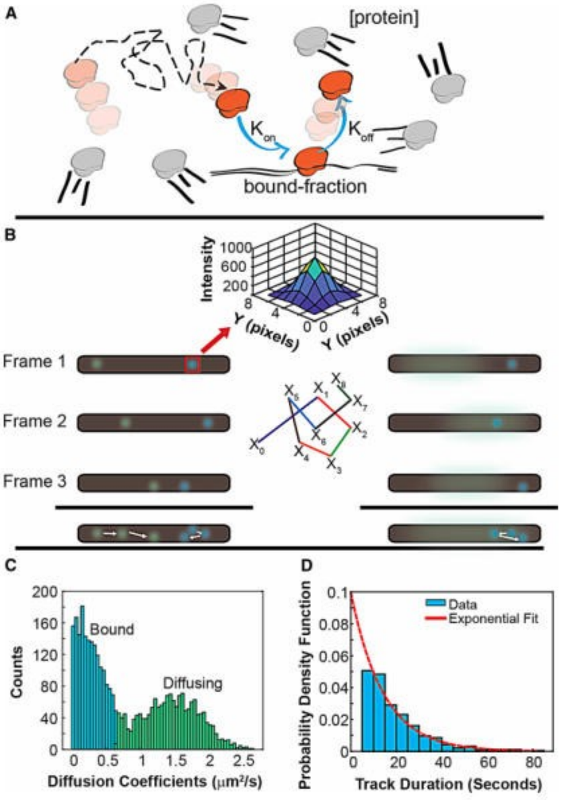


Figure 2. Single-molecule approaches.

(**A**) Study of DNA binding proteins using single-molecule approaches allow us to obtain information about their binding kinetics, diffusion coefficients and the proportion of DNA-bound copies. (**B**) Schematic representation of the data obtained in single-molecule fluorescence microscopy of live cells. Individual spots, representing single copies of the protein studied, are detected and localized at each frame (using a 2D-gaussian fitting of their intensities to determine their localization with high precision). Diffusive molecules can be accurately tracked through a time-lapse if the integration times at the camera are in the range of millisecond to low tens of milliseconds (left). Alternatively, longer camera integration times (ranging from hundreds of milliseconds to few seconds) results in average blurring of diffusive molecules, but still allows detection of immobile molecules (right). (**C**) Example of the analysis of fast-capture data. The distribution of diffusion coefficients and the fraction of bound and diffusive molecules can be obtained from them. (**D**) Example of the analysis of slow-capture data, where the distribution of fluorescent spot lifetimes can be used to determine the residence times of proteins on DNA.

light, a small subpopulation of tagged proteins will stochastically convert into a fluorescent state. A variation is photoconvertible (PC) FPs which convert into a different fluorescent state. PA/PC-FPs have enabled tracking of single molecules even for high copy-number proteins [60]. Tracking of single-molecules is done by linking their localizations frame to frame, and various algorithms are available for this non-trivial process [58,61–63]. After tracking, one can extract a variety of parameters for the POI such as its diffusion coefficient, proportions of different diffusive states, residence time, search time, and high precision mapping of cellular localization; all valuable for determining kinetics of replisome subunits in live cells (Figure 2) [16,18,58].

Recent advances in genetics have also paved the way to perturb conditions commonly done *in vitro*, directly in the cell. One example is using degron tags to rapidly deplete the POI, in a controlled manner, thereby lowering concentrations of competitor molecules [64]. Likewise, the use of temperature-sensitive mutants of POIs allows one to inactivate the protein rapidly. The advantages of these approaches compared with simply using deletion mutants, is that we can perform them on essential proteins and cells do not accumulate suppressor mutations [64].

Development of budding yeast as a system for single-molecule microscopy has been slower than in bacteria, having only few pioneering contributions [65–71]. The main challenges for doing single-molecule work on replisomes in budding yeast are: (1) the size of the yeast cell (4 μm diameter for haploid cells, compared with 0.7 μm for *E. coli*), resulting in higher background fluorescence, (2) cellular autofluorescence, (3) the higher number of replisomes active at a given time [45,72,73], and (4) the light diffraction from the thick cell wall [68]. Even though more work is needed to make of budding yeast an amenable model for single-molecule microscopy, new advances in microscope configurations have allowed for selective illumination to minimize background fluorescence while retaining high sensitivity. These include light sheet microscopy (LSM) and highly inclined and laminated optical sheet (HILO) [74,75]. In addition, the development of brighter and more photostable fluorescent dyes/proteins (e.g. Halo and SNAP tags) have improved detection of single-molecules, while also allowing for them to be observed for longer periods of time [60,69,76,77].

Perspectives

- Advances in fluorescence microscopy have allowed us to probe the dynamics and architecture of replisomes in living cells.
- The recent view of the bacterial replisome as being a dynamic cellular machine, with inherent stochastic processes, would not have been revealed without the advent of single-molecule fluorescence techniques.
- It will only be a matter of time before these techniques are applied to study the eukaryotic replisome, providing a complementary approach to traditional biochemical approaches, revealing a richer understanding on how it operates: is it inherently dynamic and stochastic like the bacterial replisome? How is chromatin assembly coupled to DNA synthesis? What are the consequences when it encounters obstacles on DNA?
- The capability to study replication complexes in a range of organisms, especially in live cells, will help to illuminate the factors governing their architectures and dynamics, and how they perform one of life's fundamental processes: DNA replication.

Abbreviations

CAF-1, chromatin assembly factor 1; FP, fluorescent protein; HE, holoenzyme; HILO, highly inclined and laminated optical sheet; LSM, light sheet microscopy; PA, photoactivatable; PC, photoconvertible; PCNA, proliferating nuclear cell antigen; POI, protein of interest; RFC, replication factor C; RPA, replication protein A; RPC, replication progression complex; sptPALM, single-particle tracking photoactivated localization microscopy; ssDNA, single-strand DNA.

Author Contribution

N.K. and R.R.-L. wrote the manuscript.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada [Discovery Grant,435521-2013], and the Canadian Institutes of Health Research [Operating Grant,142473].

Acknowledgements

We like to thank Ziad El-Hajj, Maxime Leroux, Nicolas Soubry for their feedback on the manuscript.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1 Zeman, M.K. and Cimprich, K.A. (2014) Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 https://doi.org/10.1038/ncb2897

2 Forterre, P. (2013) Why are there so many diverse replication machineries? *J. Mol. Biol.* **425**, 4714–4726 https://doi.org/10.1016/j.jmb.2013.09.032

3 Beattie, T.R. and Reyes-Lamothe, R. (2015) A Replisome's journey through the bacterial chromosome. *Front Microbiol.* **6**, 562 https://doi.org/10.3389/fmicb.2015.00562

4 Lewis, J.S., Jergic, S. and Dixon, N.E. (2016) The *E. coli* DNA replication fork. *Enzymes* **39**, 31–88 https://doi.org/10.1016/bs.enz.2016.04.001

5 Tougu, K. and Marians, K.J. (1996) The interaction between helicase and primase sets the replication fork clock. *J. Biol. Chem.* **271**, 21398–21405 https://doi.org/10.1074/jbc.271.35.21398

6 Gao, D.X. (2001) Mchenry CS. tau binds and organizes Escherichia coli replication proteins through distinct domains — Partial proteolysis of terminally tagged tau to determine candidate domains and to assign domain V as the alpha binding domain. *J. Biol. Chem.* **276**, 4433–4440 https://doi.org/10.1074/jbc.M009828200

7 Stukenberg, P.T., Studwellvaughan, P.S. and Odonnell, M. (1991) Mechanism of the sliding beta-clamp of DNA polymerase-III holoenzyme. *J. Biol. Chem.* **266**, 11328–11334 PMID:2040637

8 Gao, D.X. and McHenry, C.S. (2001) Tau binds and organizes *Escherichia coli* replication proteins through distinct domains — Domain IV, located within the unique C terminus of tau, binds the replication fork helicase, DnaB. *J. Biol. Chem.* **276**, 4441–4446 https://doi.org/10.1074/jbc.M009830200

9 Sinha, N.K., Morris, C.F. and Alberts, B.M. (1980) Efficient in vitro replication of double-stranded DNA templates by a purified T4 bacteriophage replication system. *J. Biol. Chem.* **255**, 4290–4293 PMID:6989836

10 Scherr, M.J., Safaric, B. and Duderstadt, K.E. (2018) Noise in the machine: alternative pathway sampling is the rule during DNA replication. *Bioessays* **40**, 1700159 https://doi.org/10.1002/bies.201700159

11 Georgescu, R.E., Yao, N.N., Indiani, C., Yurieva, O. and O'Donnell, M.E. (2014) Replisome mechanics: lagging strand events that influence speed and processivity. *Nucleic Acids Res.* **42**, 6497–6510 https://doi.org/10.1093/nar/gku257

12 Graham, J.E., Marians, K.J. and Kowalczykowski, S.C. (2017) Independent and stochastic action of DNA polymerases in the replisome. *Cell* **169**, 1201–1213.e17 https://doi.org/10.1016/j.cell.2017.05.041

13 Yao, N.Y., Georgescu, R.E., Finkelstein, J. and O'Donnell, M.E. (2009) Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proc. Natl Acad. Sci. U.S.A.* **106**, 13236–13241 https://doi.org/10.1073/pnas.0906157106

14 Tanner, N.A., Tolun, G., Loparo, J.J., Jergic, S., Griffith, J.D., Dixon, N.E. et al. (2011) *E. coli* DNA replication in the absence of free beta clamps. *EMBO J.* **30**, 1830–1840 https://doi.org/10.1038/emboj.2011.84

15 Wu, C.A., Zechner, E.L. and Marians, K.J. (1992) Coordinated leading-strand and lagging-strand synthesis at the *Escherichia coli* DNA-replication fork. 1. Multiple effectors act to modulate Okazaki fragment size. *J. Biol. Chem.* **267**, 4030–4044 PMID:1740451

16 Beattie, T.R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A.J.M., Leake, M.C. et al. (2017) Frequent exchange of the DNA polymerase during bacterial chromosome replication. *eLife* **6**, e21763 https://doi.org/10.7554/eLife.21763

17 Lewis, J.S., Spenkelink, L.M., Jergic, S., Wood, E.A., Monachino, E., Horan, N.P. et al. (2017) Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *eLife* **6**, e23932 https://doi.org/10.7554/eLife.23932

18 Liao, Y., Li, Y.L., Schroeder, J.W., Simmons, L.A. and Biteen, J.S. (2016) Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells. *Biophys. J.* **111**, 2562–2569 https://doi.org/10.1016/j.bpj.2016.11.006

19 Kurth, I., Georgescu, R.E. and O'Donnell, M.E. (2013) A solution to release twisted DNA during chromosome replication by coupled DNA polymerases. *Nature* **496**, 119–122 https://doi.org/10.1038/nature11988

20 Marians, K.J. (2018) Lesion bypass and the reactivation of stalled replication forks. *Annu. Rev. Biochem.* **87**, 217–238 https://doi.org/10.1146/annurev-biochem-062917-011921

21 Bell, S.P. and Labib, K. (2016) Chromosome duplication in *Saccharomyces cerevisiae*. *Genetics* **203**, 1027–1067 https://doi.org/10.1534/genetics.115.186452

22 Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. and Kunkel, T.A. (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**, 127–130 https://doi.org/10.1126/science.1144067

23 McElhinny, S.A.N., Gordenin, D.A., Stith, C.M., Burgers, P.M.J. and Kunkel, T.A. (2008) Division of labor at the eukaryotic replication fork. *Mol. Cell* **30**, 137–144 https://doi.org/10.1016/j.molcel.2008.02.022

24 Georgescu, R.E., Langston, L., Yao, N.Y., Yurieva, O., Zhang, D., Finkelstein, J. et al. (2014) Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 https://doi.org/10.1038/nsmb.2851

25 Clausen, A.R., Lujan, S.A., Burkholder, A.B., Orebaugh, C.D., Williams, J.S., Clausen, M.F. et al. (2015) Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat. Struct. Mol. Biol.* **22**, 185–191 https://doi.org/10.1038/nsmb.2957

26 Georgescu, R.E., Schauer, G.D., Yao, N.Y., Langston, L.D., Yurieva, O., Zhang, D. et al. (2015) Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *eLife* **4**, e04988 https://doi.org/10.7554/eLife.04988

27 Koh, K.D., Balachander, S., Hesselberth, J.R. and Storici, F. (2015) Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat. Methods* **12**, 251–257 https://doi.org/10.1038/nmeth.3259

28 Schauer, G.D. and O'Donnell, M.E. (2017) Quality control mechanisms exclude incorrect polymerases from the eukaryotic replication fork. *Proc. Natl Acad. Sci. U.S.A.* **114**, 675–680 https://doi.org/10.1073/pnas.1619748114

29 Johnson, R.E., Klassen, R., Prakash, L. and Prakash, S. (2015) A major role of DNA polymerase delta in replication of both the leading and lagging DNA strands. *Mol. Cell* **59**, 163–175 https://doi.org/10.1016/j.molcel.2015.05.038

30 Yeeles, J.T.P., Janska, A., Early, A. and Diffley, J.F.X. (2017) How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol. Cell* **65**, 105–116 https://doi.org/10.1016/j.molcel.2016.11.017

31 Garbacz, M.A., Lujan, S.A., Burkholder, A.B., Cox, P.B., Wu, Q.Q., Zhou, Z.X. et al. (2018) Evidence that DNA polymerase δ contributes to initiating leading strand DNA replication in *Saccharomyces cerevisiae*. *Nat. Commun.* **9**, 858 https://doi.org/10.1038/s41467-018-03270-4

32 Aria, V. and Yeeles, J.T.P. (2019) Mechanism of bidirectional leading-strand synthesis establishment at eukaryotic DNA replication origins. *Mol. Cell* **73**, 199 https://doi.org/10.1016/j.molcel.2018.10.019

33 Zhou, J.C., Janska, A., Goswami, P., Renault, L., Ali, F.A., Kotecha, A. et al. (2017) CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc. Natl Acad. Sci. U.S.A.* **114**, 4141–4146 https://doi.org/10.1073/pnas.1700530114

34 Sengupta, S., van Deursen, F., de Piccoli, G. and Labib, K. (2013) Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr. Biol.* **23**, 543–552 https://doi.org/10.1016/j.cub.2013.02.011

35 Simon, A.C., Zhou, J.C., Perera, R.L., van Deursen, F., Evrin, C., Ivanova, M.E. et al. (2014) A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature* **510**, 293–297 https://doi.org/10.1038/nature13234

36 Gambus, A., van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R.C., Edmondson, R.D. et al. (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J.* **28**, 2992–3004 https://doi.org/10.1038/emboj.2009.226

37 Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D. et al. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol.* **8**, 358–366 https://doi.org/10.1038/ncb1382

38 Gan, H.Y., Yu, C.H., Devbhandari, S., Sharma, S., Han, J.H., Chabes, A. et al. (2017) Checkpoint kinase Rad53 couples leading- and lagging-strand DNA synthesis under replication stress. *Mol. Cell* **68**, 446–455.e3 https://doi.org/10.1016/j.molcel.2017.09.018

39 Devbhandari, S., Jiang, J.Q., Kumar, C., Whitehouse, I. and Remus, D. (2017) Chromatin constrains the initiation and elongation of DNA replication. *Mol. Cell* **65**, 131–141 https://doi.org/10.1016/j.molcel.2016.10.035

40 Sekedat, M.D., Fenyo, D., Rogers, R.S., Tackett, A.J., Aitchison, J.D. and Chait, B.T. (2010) GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.* **6**, 353 https://doi.org/10.1038/msb.2010.8

41 Stodola, J.L. and Burgers, P.M. (2016) Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nat. Struct. Mol. Biol.* **23**, 402–408 https://doi.org/10.1038/nsmb.3207

42 Ho, B., Baryshnikova, A. and Brown, G.W. (2018) Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Syst.* **6**, 192–205.e3 https://doi.org/10.1016/j.cels.2017.12.004

43 Chong, Y.T., Koh, J.L.Y., Friesen, H., Duffy, K., Cox, M.J., Moses, A. et al. (2015) Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413–1424 https://doi.org/10.1016/j.cell.2015.04.051

44 Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 https://doi.org/10.1038/nmeth.2834

45 Saner, N., Karschau, J., Natsume, T., Gierlinski, M., Retkute, R., Hawkins, M. et al. (2013) Stochastic association of neighboring replicons creates replication factories in budding yeast. *J. Cell Biol.* **202**, 1001–1012 https://doi.org/10.1083/jcb.201306143

46 Knott, S.R.V., Peace, J.M., Ostrow, A.Z., Gan, Y., Rex, A.E., Viggiani, C.J. et al. (2012) Forkhead transcription factors establish origin timing and long-range clustering in *S. cerevisiae*. *Cell* **148**, 99–111 https://doi.org/10.1016/j.cell.2011.12.012

47 Mantiero, D., Mackenzie, A., Donaldson, A. and Zegerman, P. (2011) Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J.* **30**, 4805–4814 https://doi.org/10.1038/emboj.2011.404

48 Smith, D.J. and Whitehouse, I. (2012) Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434–480 https://doi.org/10.1038/nature10895

49 Kurat, C.F., Yeeles, J.T.P., Patel, H., Early, A. and Diffley, J.F.X. (2017) Chromatin controls DNA replication origin selection, lagging-strand synthesis, and replication fork rates. *Mol. Cell* **65**, 117–130 https://doi.org/10.1016/j.molcel.2016.11.016

50 Foltman, M., Evrin, C., De Piccoli, G., Jones, R.C., Edmondson, R.D., Katou, Y. et al. (2013) Eukaryotic replisome components cooperate to process histones during chromosome replication. *Cell Rep.* **3**, 892–904 https://doi.org/10.1016/j.celrep.2013.02.028

51 Evrin, C., Maman, J.D., Diamante, A., Pellegrini, L. and Labib, K. (2018) Histone H2A-H2B binding by Pol alpha in the eukaryotic replisome contributes to the maintenance of repressive chromatin. *EMBO J.* **37** https://doi.org/10.15252/embj.201899021

52 Gan, H.Y., Serra-Cardona, A., Hua, X., Zhou, H., Labib, K., Yu, C.H. et al. (2018) The Mcm2-Ctf4-Pol alpha axis facilitates parental histone H3-H4 transfer to lagging strands. *Mol. Cell* **72**, 140–151.e3 https://doi.org/10.1016/j.molcel.2018.09.001

53 Yu, C.H., Gan, H.Y., Serra-Cardona, A., Zhang, L., Gan, S.L., Sharma, S. et al. (2018) A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. *Science* **361**, 1386–1389 https://doi.org/10.1126/science.aat8849

54 Liu, S., Xu, Z., Leng, H., Zheng, P., Yang, J., Chen, K. et al. (2017) RPA binds histone H3-H4 and functions in DNA replication-coupled nucleosome assembly. *Science* **355**, 415–420 https://doi.org/10.1126/science.aah4712

55 Chilkova, O., Stenlund, P., Isoz, I., Stith, C.M., Grabowski, P., Lundstrom, E.B. et al. (2007) The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.* **35**, 6588–6597 https://doi.org/10.1093/nar/gkm741

56 Ben-Shahar T, R., Castillo, A.G., Osborne, M.J., Borden, K.L., Kornblatt, J. and Verreault, A. (2009) Two fundamentally distinct PCNA interaction peptides contribute to chromatin assembly factor 1 function. *Mol. Cell. Biol.* **29**, 6353–6365 https://doi.org/10.1128/MCB.01051-09

57 Tanner, N.A., Hamdan, S.M., Jergic, S., Schaeffer, P.M., Dixon, N.E. and van Oijen, A.M. (2008) Single-molecule studies of fork dynamics in *Escherichia coli* DNA replication. *Nat. Struct. Mol. Biol.* **15**, 170–176 https://doi.org/10.1038/nsmb.1381

58 Shen, H., Tauzin, L.J., Baiyasi, R., Wang, W.X., Moringo, N., Shuang, B. et al. (2017) Single particle tracking: from theory to biophysical applications. *Chem. Rev.* **117**, 7331–7376 https://doi.org/10.1021/acs.chemrev.6b00815

59 Manley, S., Gillette, J.M., Patterson, G.H., Shroff, H., Hess, H.F., Betzig, E. et al. (2008) High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155–157 https://doi.org/10.1038/nmeth.1176

60 Elf, J. and Barkefors, I. (2019) Single-molecule kinetics in living cells. *Annu. Rev. Biochem.* **88**, 635–659 https://doi.org/10.1146/annurev-biochem-013118-110801

61 Crocker, J.C. and Grier, D.G. (1996) Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.* **179**, 298–310 https://doi.org/10.1006/jcis.1996.0217

62 Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L. et al. (2008) Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 https://doi.org/10.1038/nmeth.1237

63 Tinevez, J.Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E. et al. (2017) Trackmate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 https://doi.org/10.1016/j.ymeth.2016.09.016

64 Natsume, T. and Kanemaki, M.T. (2017) Conditional degrons for controlling protein expression at the protein level. *Annu. Rev. Genet.* **51**, 83–102 https://doi.org/10.1146/annurev-genet-120116-024656

65 Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. and Ewers, H. (2012) A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods* **9**, 582–584 https://doi.org/10.1038/nmeth.1991

66 Puchner, E.M., Walter, J.M., Kasper, R., Huang, B. and Lim, W.A. (2013) Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory. *Proc. Natl Acad. Sci. U.S.A.* **110**, 16015–16020 https://doi.org/10.1073/pnas.1309676110

67 Kaplan, C. and Ewers, H. (2015) Optimized sample preparation for single-molecule localization-based superresolution microscopy in yeast. *Nat. Protoc.* **10**, 1007–1021 https://doi.org/10.1038/nprot.2015.060

68 Smith, C., Lari, A., Derrer, C.P., Ouwehand, A., Rossouw, A., Huisman, M. et al. (2015) In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. *J. Cell Biol.* **211**, 1121–1130 https://doi.org/10.1083/jcb.201503135

69 Ball, D.A., Mehta, G.D., Salomon-Kent, R., Mazza, D., Morisaki, T., Mueller, F. et al. (2016) Single molecule tracking of Ace1p in *Saccharomyces cerevisiae* defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Res.* **44**, e160 https://doi.org/10.1093/nar/gkw257

70 Mehta, G.D., Ball, D.A., Eriksson, P.R., Chereji, R.V., Clark, D.J., McNally, J.G. et al. (2018) Single-molecule analysis reveals linked cycles of RSC chromatin remodeling and Ace1p transcription factor binding in yeast. *Mol. Cell* **72**, 875–887.e9 https://doi.org/10.1016/j.molcel.2018.09.009

71 Wollman, A.J., Shashkova, S., Hedlund, E.G. Friemann, R., Hohmann, S. and Leake, M.C. (2017) Transcription factor clusters regulate genes in eukaryotic cells. *eLife* **6**, e27451 https://doi.org/10.7554/eLife.27451

72 Phillips, R. (2013) *Physical Biology of the Cell*, 2nd edn, pp. 1057, Garland Science, London, New York, NY

73 Maslanka, R., Kwolek-Mirek, M. and Zadrag-Tecza, R. (2018) Autofluorescence of yeast *Saccharomyces cerevisiae* cells caused by glucose metabolism products and its methodological implications. *J. Microbiol. Methods* **146**, 55–60 https://doi.org/10.1016/j.mimet.2018.01.017

74 Gebhardt, J.C.M., Suter, D.M., Roy, R., Zhao, Z.Q.W., Chapman, A.R., Basu, S. et al. (2013) Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421–426 https://doi.org/10.1038/nmeth.2411

75 Yu, J. (2016) Single-molecule studies in live cells. *Annu. Rev. Phys. Chem.* **67**, 565–585 https://doi.org/10.1146/annurev-physchem-040215-112451

76 England, C.G., Luo, H.M. and Cai, W.B. (2015) Halotag technology: a versatile platform for biomedical applications. *Bioconjugate Chem.* **26**, 975–986 https://doi.org/10.1021/acs.bioconjchem.5b00191

77 Grimm, J.B., English, B.P., Choi, H., Muthusamy, A.K., Mehl, B.P., Dong, P. et al. (2016) Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* **13**, 985–988 https://doi.org/10.1038/nmeth.4034