Contents

1 General Introduction **5**

1.1 Genome stability and mutagenesis 5

1.2 Background to DNA replication 6

1.3 DNA replication in *E. coli* 8

1.3.1 Initiation 9

1.3.2 Elongation 10

1.3.3 Termination 12

1.3.4 Lagging strand synthesis 13

1.3.5 Stoichiometry in the replisome 14

1.3.6 Dynamics in replisome 16

1.4 DNA damage tolerance in *E. coli* 17

1.4.1 Translesion synthesis 17

1.4.2 Translesion synthesis by DNA polymerase IV 19

1.4.3 Translesion synthesis by DNA polymerase V 20

1.4.4 Translesion synthesis by DNA polymerase II 21

1.4.5 Polymerase exchange during translesion synthesis 21

1.4.6 Error-free DNA damage tolerance pathways 23

1.5 Thesis objectives 24

Bibliography **26**

2 Development of a single-molecule method to study DNA replication **42**

2.1 Introduction 42

2.1.1 Single-molecule methods for studying protein-DNA interactions 43

2.1.2 Aims 49

2.2 Materials and Methods 49

2.3 Results 50

2.3.1 Flow cell . . . 50
2.3.2 Surface chemistry . . . 51
2.3.3 Fluorophores . . . 54
2.3.4 Labeling strategies . . . 56
2.3.5 DNA substrates . . . 58
2.3.6 Setting up CoSMoS experiments to study protein-DNA interactions . . 60
2.3.7 Data acquisition . . . 61
2.3.8 Data analysis . . . 62
2.4 Discussion . . . 64

Bibliography **67**

3 Studying *E. coli* DNA replication by CoSMoS **71**
3.1 Introduction . . . 71
3.2 Materials and Methods . . . 72
3.2.1 Cloning of protein expression vectors . . . 72
3.2.2 Protein expression and purification . . . 73
3.2.3 Protein labeling . . . 73
3.2.4 ϕX174 primer extension assays . . . 75
3.2.5 DNA substrate . . . 75
3.2.6 Preparation of slides . . . 76
3.2.7 TIRF microscopy . . . 76
3.2.8 Data analysis . . . 77
3.3 Results . . . 77
3.3.1 Protein labeling and complex re-constitution . . . 77
3.3.2 Clamp loading . . . 80
3.3.3 Clamp unloading . . . 81
3.3.4 Polymerase loading . . . 82
3.3.5 Polymerase idling and extension . . . 86
3.4 Discussion . . . 86

Bibliography **89**

4 Studying *E.coli* translesion synthesis by CoSMoS **93**
4.1 Introduction . . . 93

4.2 Materials and Methods . . . 94
4.2.1 Cloning of protein expression vectors . . . 94
4.2.2 Protein expression and purification . . . 95
4.2.3 Protein labeling . . . 95
4.2.4 ϕX174 primer extension assays . . . 96
4.2.5 DNA substrate . . . 96
4.2.6 TIRF microscopy . . . 97
4.3 Results . . . 97
4.3.1 Protein labeling . . . 97
4.3.2 Pol IV and Pol IIIcore exchange . . . 98
4.3.3 Pol II and Pol IIIcore exchange . . . 101
4.3.4 Polymerases compete for binding to the hydrophobic groove on the β clamp . . . 102
4.3.5 Pol IV recruitment . . . 104
4.4 Discussion . . . 105

Bibliography **108**

5 A heterodimeric clamp for structural analysis of *E. coli* translesion synthesis **111**
5.1 Introduction . . . 111
5.1.1 Monomeric β clamp . . . 112
5.1.2 Aims . . . 114
5.2 Materials and Methods . . . 115
5.2.1 Protein expression and purification . . . 115
5.2.2 Analytical ultracentrifugation . . . 115
5.2.3 Cryo-electron microscopy . . . 115
5.2.4 Differential scanning fluorimetry . . . 116
5.2.5 Primer extension assays . . . 116
5.3 Results . . . 116
5.3.1 The β clamp monomer-dimer equilibrium . . . 116
5.3.2 Structural analysis of the heterodimeric β clamp . . . 118
5.3.3 Biochemical analysis of the heterodimeric β clamp . . . 120
5.3.4 Abolishing polymerase binding to the β clamp . . . 121
5.4 Discussion . . . 122

Bibliography **126**

6 Concluding remarks **128**

Bibliography **132**

Chapter 1

General Introduction

1.1 Genome stability and mutagenesis

In all species, genetic information is stored in the molecule deoxyribonucleic acid (DNA) inside the cell. DNA contains all the information that is required by a cell to perform its functions and this information is read out by first transcribing the DNA into messenger RNA, and then translating the messenger RNA into proteins that perform diverse functions in the cell. This is known as the central dogma in biology. DNA in a cell is copied and divided to daughter cells during cell division, which is required for the maintenance and growth of the organism. Given that all the information responsible for the survival of an organism is contained within DNA, and this information needs to be passed on accurately, it is vital that DNA is protected from errors or mutations that could otherwise have deleterious effects on the cell. In fact, a number of human diseases are caused by the accumulation of errors in DNA that are manifested in the altered behaviour of cells, for example cancer and neurodegenerative diseases.

However, changes in DNA are not always damaging. In fact, mutations underlie the process of evolution that give rise to selective advantages within a population. For example, bacterial strains can acquire mutations to allow them to remove or metabolize antibiotics more efficiently, and therefore adapt to the hostile living conditions. In eukaryotes, mutations in DNA are important for the generation of diversity during antibody production and also give cancer cells the ability to overcome cell division regulation, resulting in their increased growth rates. Therefore, there needs to be a delicate balance between preserving DNA in order for it to be accurately maintained and passed on, and the mutations that can give rise to adaptation and new functions in the cell.

Mutations in DNA arise from various endogenous and exogenous sources[74]. Cells are ex-

posed to a large number of exogenous insults, including mutagenic chemicals and UV irradiation. Inside the cell, DNA can undergo spontaneous depurination and are exposed to reactive oxygen species that are generated as a by-product of cellular metabolism. Mutations can also arise from mistakes made during DNA replication, when the information in DNA is copied in order to be passed to daughter cells. While DNA replication is a highly accurate process, it has a low intrinsic error rate that leads to the introduction of errors during the copying of DNA. To overcome the burden imposed on DNA by the various mutagenic sources, cells have evolved a number of repair processes that resolve the various different types of mutations, for example base excision repair that removes damaged bases, nucleotide excision repair that excises damaged nucleotides, and homologous recombination and non-homologous end-joining that deals with DNA double-strand breaks. However, when the DNA damage is located in front of the progressing replication fork, it can trigger replication fork stalling, which can be lethal for the cell. Under these circumstances, the cell needs to rapidly overcome the stalling and continue fork progression in order to prevent the processing of the single-stranded DNA at a stalled replication fork into a double-strand break, which can lead to chromosomal rearrangements. The mechanisms that exist in the cell to temporarily avoid DNA damages are known as the DNA damage tolerance pathways, and interestingly, some of these pathways are inherently mutagenic.

Therefore, in order to understand how the balance between genome stability and mutagenesis is achieved, it is essential to study the mechanisms that govern replicative fidelity.

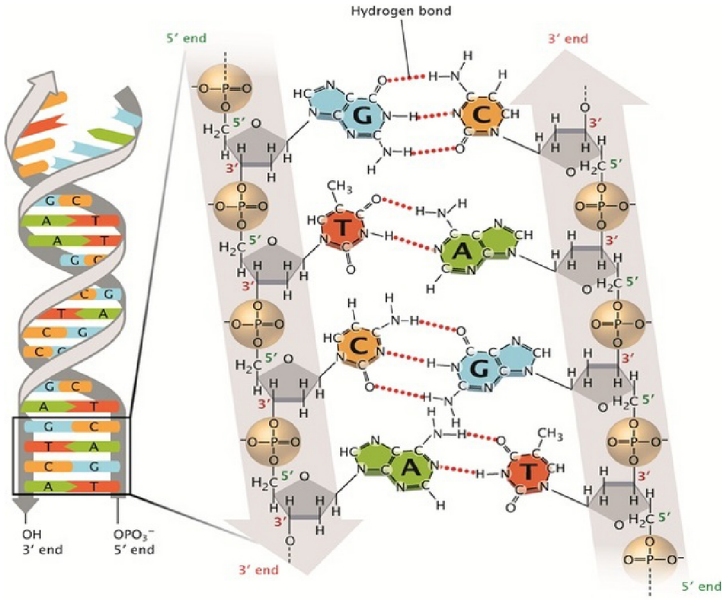
1.2 Background to DNA replication

DNA replication is the process by which the information in DNA is duplicated in order for it to be partitioned between the mother and daughter cell during cell division. The 3D structure of DNA determined by Watson and Crick in 1953 [134] shed light on the mechanism by which the process of DNA replication is achieved.

In the structure, DNA was shown to be a right-handed helix consisting of two anti-parallel strands of deoxyribonucleotides. Each deoxyribonucleotide consists of a deoxyribose sugar, a base and a phosphate group, and the individual nucleotides are covalently bonded together through the linking of the 3' hydroxyl group of the sugar on one nucleotide to the 5' phosphate group of another through the formation of phosphodiester bonds. These 3'-5' linkages form the sugar-phosphate backbone, which is located on the outside of the DNA molecule. Within the DNA molecule, the genetic information is stored in the bases that group together to form

sequences of instructions for the cell. DNA is made of four purine or pyrimidine bases, adenosine (A), guanosine (G), thymidine (T) and cytosine (C), that are arranged perpendicularly to the phosphate backbone.

The bases form pairs in a specific manner, where a purine (A and G) hydrogen bonds to a pyrimidine (T and C) and follow the rule of A with T and G with C pairing. The structure of DNA is further stabilized through π-π interactions between the adjacent vertically stacked base-pairs. The specific base-pairing ensures that if the sequence on one strand is known, the sequence on the other strand can be automatically determined. As Watson and Crick noted in their seminal paper, this base-pairing rule "immediately suggests a possible copying mechanism for the genetic material". This is shown in Fig. 1.1.



.

Figure 1.1. The structure of DNA reveals copying mechanism. Schematic representation of the anti-parallel double helical structure of DNA showing the sugar-phosphate backbone and complementary base pairing between the bases. Figure taken from Pray, 2008[96]

When DNA is replicated, each strand of the double helix acts as the template for the synthesis of a new strand with Watson and Crick base-pairing. Enzymes called DNA polymerases catalyze the incorporation of deoxyribosenucleotide triphosphates (dNTPs) during the synthesis of new DNA strands. When a dNTP is added, the triphosphate is hydrolyzed and provides the energy required for the synthesis reaction[114]. The semi-conservative copying mechanism as we know it today was first demonstrated by Meselson and Stahl in 1958, who used radioisotope labeling to show that each daughter DNA molecule received on parental DNA strand[84].

In order to ensure the accurate synthesis of DNA, DNA polymerases have a high fidelity in

the range of 10^{-4} to 10^{-6}, that is, they introduce one error every 10,000 to 1,000,000 nucleotides incorporated[68]. Mistakes arise from the incorrect selection of nucleotides to be incorporated, resulting in a mismatched base-pair. In addition, the polymerase may skip a base on the template strand, leading to base insertions or deletions that may become manifested at the level of protein synthesis in the form of frameshift mutations. The fidelity of DNA synthesis is further improved by the ability of DNA polymerases to proofread and remove any mis-incorporated nucleotides through their associated exonucleases. The proofreading activity of the exonucleases increases the accuracy of DNA synthesis by 1000-fold[74]. In the event that a mis-incorporated nucleotide is not removed by the exonuclease, a further fail-safe mechanism is in place, called the mismatch repair system, which scans the newly synthesized DNA and removes the mismatches by excising the wrongly inserted nucleotide. Together, these mechanisms increase the fidelity of DNA synthesis to 10^{-10}, which ensures that DNA is copied in a highly accurate manner[74].

Another challenge for the copying of DNA lies in the anti-parallel nature of the DNA double helix. DNA synthesis proceeds in the 5'-3' direction and due to the anti-parallel nature of the two strands with the same overall direction of replication fork progression, one strand is synthesized continuously while the other is synthesized discontinuously in shorter 1kb fragments called Okazaki fragments[93]. These are named the leading and lagging strands, respectively. Therefore, not only does DNA replication need to be highly accurate, the synthesis of the two anti-parallel DNA strands need to be coupled in order to ensure that two copies of the DNA are made simultaneously and divided between the mother and daughter cells. The simultaneous synthesis of the two DNA strands also ensures that long stretches of single-stranded DNA are not generated in one strand while the other is being duplicated, which would be highly unstable and prone to degradation by nucleases in the cell.

Therefore, to complete the challenging tasks of DNA replication by the replicative polymerase, a number of events need to be orchestrated. Each step is catalyzed by dedicated proteins that together form the multi-protein replisome that catalyzes high fidelity DNA replication.

1.3 DNA replication in *E. coli*

The mechanisms of DNA replication have been studied in a number of bacterial species, including *Escherichia coli* (*E. coli*), *Bacillus subtilis*, *Thermus thermophilus*, *Pseudomonas aeruginosa* and *Staphylococcus aureus*. Out of these, the best understood is the *E. coli* system, which has been used as the paradigm for understanding the processes involved in DNA replication.

The key player in DNA replication in *E. coli*, the replicative DNA polymerase, was first

identified and characterized in the 1970s. In fact, the first DNA polymerase discovered in *E. coli* was DNA polymerase I[70], which was later found to be involved in lagging strand maturation rather than serving as the main replicative DNA polymerase. It was thereafter confirmed that the replicative DNA polymerase in *E. coli* is the α subunit of the DNA polymerase III (Pol III) holoenzyme[65], which consists of more than ten different proteins that collectively orchestrate the simultaneous synthesis of the leading and lagging DNA strands[82]. The holoenzyme and the other proteins involved in DNA replication together form the replisome which travels with the replication fork, a structure that is formed between the parental and nascent DNA strands. An overview of some of the key enzymes involved in *E. coli* DNA replication is shown in Fig. 1.2 and are discussed in detail in the following section.

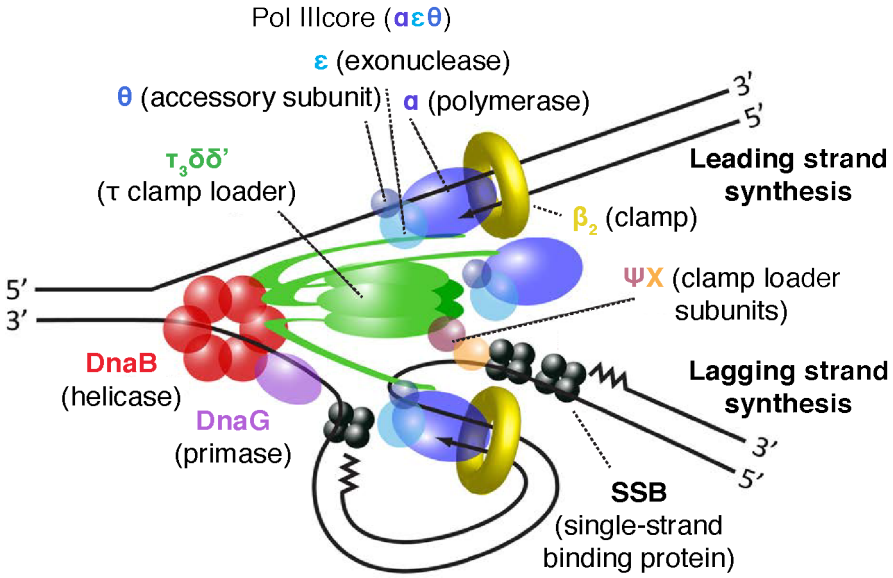


Figure 1.2. DNA replication in *E. coli*. Schematic representation of the key components of the *E. coli* replisome. Figure adapted from Beattie and Reyes-Lamothe, 2015[4].

1.3.1 Initiation

E. coli has a circular genome of 4.6 million base pairs that is replicated bi-directionally from a sequence in the genomic DNA called the origin of replication (ori). The *E. coli* genome has only one ori[125] and it is recognized and bound by replication initiation proteins, which marks the start of the replication cycle. The ori is an AT-rich sequence which allows it to be more easily melted and unwound, as the A-T base-pair consists of two hydrogen bonds instead of the three in the G-C base-pair. The ori is first recognized by DnaA, followed by the recruitment other replication initiation proteins that result in the unwinding of about 350 base pairs single-

stranded DNA (ssDNA)[38,108,12].

After the initial melting of the ori sequence, further unwinding of the DNA duplex is mediated by the hexameric DNA helicase (*DnaB*), which is loaded onto each single-stranded DNA by the helicase loader protein (*DnaC*)[2,39,11]. Several mechanisms exist to tightly control the loading the DnaB helicase to the ori during initiation to ensure that a single initiation event is achieved per cell cycle[16]. The helicase uses ATP hydrolysis to unwind the DNA duplex birectionally like a zipper[1] and creates a DNA replication bubble where the two replication forks proceed in the opposite directions. The ssDNA that protrudes from the unwinding by the helicase is coated with the tetrameric single-stranded DNA binding protein (SSB) that prevents DNA secondary structure formation and cleavage by nucleases[85].

The helicase is physically associated with the primase (*DnaG*)[77,127] that synthesizes short stretches of ribonucleic acid (RNA), from which the replicative DNA polymerase can extend[10]. The replicative DNA polymerase requires the primase activity to initiate DNA synthesis as it cannot synthesize polynucleotides *de novo*, that is, the polymerase is unable to link two nucleotides in the absence of an existing stretch of polynucleotides[54]. By coupling the activities of the helicase and primase, the replisome ensures that any DNA unwinding is coupled to the synthesis of primers, from which nascent DNA can be synthesized.

1.3.2 Elongation

Once the RNA primer has been synthesized, the replicative DNA polymerase Pol IIIα extends the primer and incorporates dNTPs in the 5' to 3' direction to form a nascent strand through complementary base-pairing to the the parental strand. The crystal structure of Pol IIIα shows that it has a narrow active site, which prevents the incorporation of the wrong nucleotides[69]. Pol IIIα is associated with the exonuclease ϵ subunit, which unlike in most other bacteria, is located on a separate protein instead of a domain in the polymerase[104]. The reason for the separation of the two proteins during evolution is not well understood. The exonuclease ϵ proofreads and removes the mis-incorporated nucleotides inserted by Pol IIIα through its 3'-5' exonuclease activity[68]. It has also been shown that the isolated Pol IIIα is a poor enzyme that has a low affinity for DNA[80,27] and that the interaction between ϵ and Pol IIIα stimulates the activity of Pol IIIα[59,115] and vice versa[78]. Together, Pol IIIα, ϵ and another subunit, θ, form the Pol IIIcore[81]. The θ subunit is referred to as an accessory subunit as it has no other known function apart from its role in stimulating the activity of ϵ. In fact, the θ subunit has been shown to be dispensable *in vivo*[110].

The isolated Pol IIIcore has low affinity for DNA and dissociates from DNA after synthesizing around ten nucleotides[26]. To improve its processivity on DNA, Pol IIIcore binds to a ring-shaped protein[63 42], the β clamp, which encircles the DNA and increases the processivity of Pol IIIα on DNA[116]. The *E. coli* holoenzyme has a processivity of more than 80,000 nucleotides per binding event[21]. The β clamp is a homodimeric protein with two conserved hydrophobic binding pockets (one on each monomer) for binding to its interaction partners. During DNA synthesis, the two binding pockets are occupied by Pol IIIα and ϵ[52 126 27]. Due to the ring shape of the β clamp, it needs to be loaded onto DNA by the clamp loader complex, which is pentameric and consists of three copies of the *dnaX* gene product τ or γ, and one copy each of the δ and δ' subunits[97]. τ is the full-length *dnaX* gene product whereas γ is a truncated version of τ arising from ribosomal frameshifting during protein synthesis[9 29 128]. The crystal structure of the γ clamp loader complex[53] shows that the subunits are arranged such that the γ subunits are are adjacent to one another and are flanked by δ and δ' on each side. Only the γ subunits are able to bind ATP, thus there are three ATPase sites in the clamp loader complex. The ATPase sites are located at the interface between two adjacent subunits and are formed by the insertion of a critical arginine finger from one subunit to the ATP binding site of the neighboring γ subunit[46 57].

The clamp loader complex is an ATPase that uses ATP hydrolysis to disrupt one dimer interface in the β clamp to break the ring structure prior to enclosing it on DNA. This is shown in the structure of the T4 bacteriophage clamp loader in complex with an open clamp in Fig. 1.3[56]. Given that the overall architecture of the clamp loader is widely conserved between species[46 57], it is assumed that the *E. coli* clamp loader operates by a similar mechanism.

In addition to its role in clamp loading, the clamp loader plays an important role in coordinating the different subunits present in the holoenzyme. This is achieved through the τ protein, which binds to Pol IIIα and the helicase via separate domains[41 40]. Two additional subunits, χ and ψ, bind to the clamp loader through the interaction between ψ and τ. These two subunits are not required for clamp loading but have been shown to stimulate the activity of the clamp loader complex and the χ subunit directly interacts with SSB[142 58]. The interaction between the χ and SSB is thought to be important for the recruitment of the clamp loader complex to the unwound DNA behind the helicase[142 58]. Therefore, the clamp loader acts as a bridge between the different components within the replisome, ensuring that the individual activities (helicase, primase, clamp loading and polymerase activities) are coupled to achieve greater efficiency in DNA synthesis.

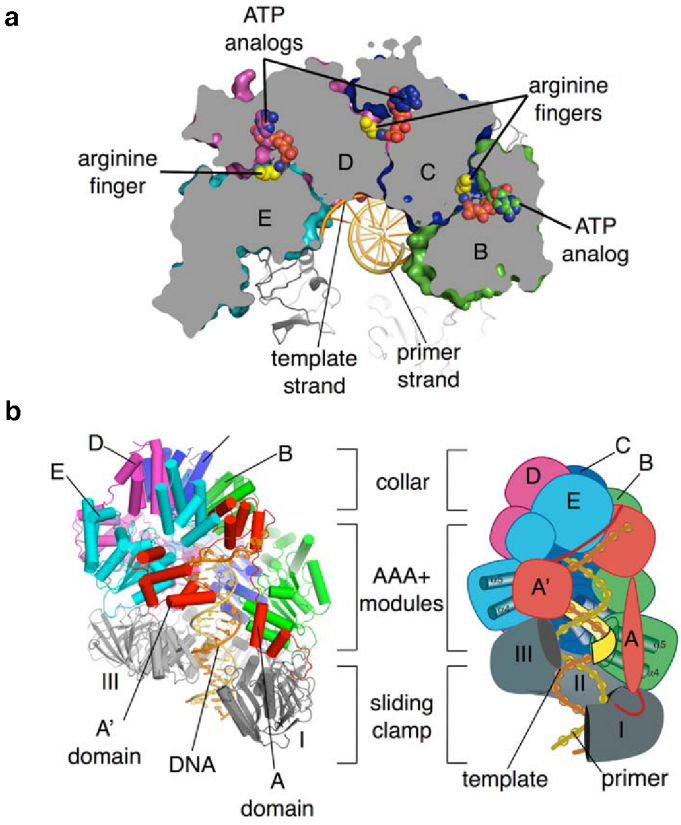


Figure 1.3. Structure of the T4 bacteriophage clamp loader bound to an open clamp. **a**, Top-view of the clamp loader-open clamp complex showing the location of the three ATP binding sites and the arginine fingers contributed to the ATPase site by the neighboring subunits. **b**, Structure and cartoon representation of the clamp loader-open clamp complex. The subunits are labeled from A-E, similar to the arrangement found in the *E. coli* clamp loader. An exception is the additional A' subunit, which is absent in the *E. coli* structure. Figure taken from Kelch *et al*, 2011[56].

1.3.3 Termination

After traveling bi-directionally from the origin, the two replication forks converge in the termination zone, which consists of 10 23-base-pair *ter* sites that are located in two oppositely oriented groups[48] Fig. 1.4. Each of the *ter* sites is bound by a monomeric DNA replication terminus site-binding proteins (Tus) and together, they form replication fork barriers to prevent further DNA synthesis[15 86].

The *ter* sites are oriented in such a way that the leftward and rightward replication forks can each enter the first five sites in the direction that they enter the termination zone but cannot pass through the next five *ter* sites that have an opposite orientation. This way, the *ter* sites act as traps for replication forks and ensure that the forks converge in the termination zone, independent of

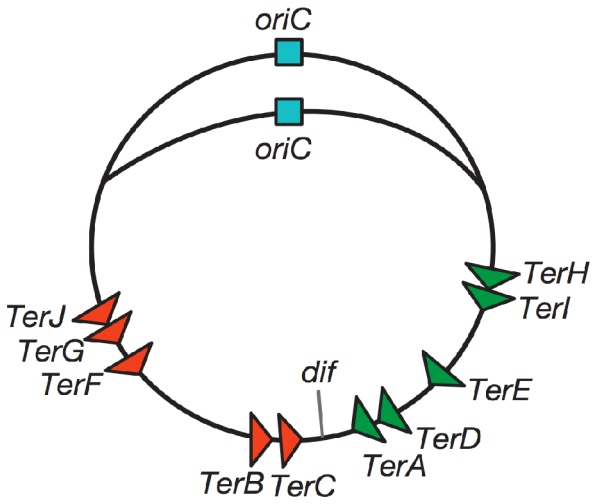


Figure 1.4. Sites of replication initiation and termination in *E. coli*. Bi-directional replication initiates at a single oriC site on the *E. coli* chromosome and terminates in the termination zone marked by 10 oppositely oriented *ter* sites. Each replication fork can pass through the first five *ter* sites in the direction of entry but is arrested by the opposite polarity of the next five *ter* sites. Figure taken from Elshenawy *et al*, 2015[25].

whether one fork enters the termination zone before the other[48 15 88 22]. Studies have shown that replication fork convergence leads to the formation of a 3' flap structure[47 135], which is normally removed, followed by gap filling and Okazaki maturation. The failure to remove the 3' flap can lead to replication re-start[105], and the extent of re-initiation events increases in the absence of the Tus protein[105 47]. While replication re-start is dangerous for the cell as it can lead to excess genetic material being synthesized in the cell, and therefore needs to be tightly controlled, it is curious that the *E. coli* termination seems to be highly susceptible to re-initiation[19].

1.3.4 Lagging strand synthesis

As mentioned in Section 1.2, the DNA synthesis proceeds in an overall 5' to 3' direction. However, due to the anti-parallel nature of the DNA duplex, the lagging strand is synthesized in the opposite direction compared to the unwinding of the DNA by the helicase. As a result, the DNA polymerase on the lagging strand can only synthesize shorter stretches of DNA of approximately 1 kb in size[93].

The synthesis of each Okazaki fragment requires the synthesis of a new RNA primer by the primase. When this is completed, the primase initially remains stably attached to its primer[142]. This interaction is thought to be stabilized through the binding of the primase to the SSB on the

ssDNA unwound by the helicase[142 118]. The primase is displaced from the primer by the clamp loader complex, which is directed to the site through the interaction between the χ subunit and SSB[142 58]. A new cycle of clamp loading and polymerase loading takes place, whereby the primer is extended to form an Okazaki fragment.

The mechanism by which the polymerase is released at the end of the Okazaki fragment remains poorly understood. Some studies suggest a collision model by which the polymerase physically bumps into the primer of the previous Okazaki fragment, causing it to release[72]. Another line of thought suggests that the active synthesis of primers by the primase may signal to the polymerase for its release, even if the Okazaki fragment has not been completed[137]. Regardless of the model, although the replicative DNA polymerase is released from the DNA at the end of each Okazaki fragment, it remains associated to the replisome through the tight interaction it has with τ protein of the clamp loader complex and may be recycled for the synthesis of another Okazaki fragment[117 72].

The Okazaki fragments that result from the discontinuous synthesis of the lagging strand need to be further processed to replace the RNA primers with DNA and to join the adjacent Okazaki fragments. This is carried out by the DNA polymerase I (Pol I), which has a 5'-3' exonuclease domain that catalyzes the removal of the RNA stretches. Pol I then fills the gap using its polymerase domain together with its 3'-5' proofreading exonuclease domain. Finally, the nicks in between the adjacent fragments are ligated by DNA ligase. The process by which the discontinous stretches of RNA and DNA on the lagging strand are converted to yield a complete, new daughter strand is called Okazaki fragment maturation[92].

1.3.5 Stoichiometry in the replisome

The coordinated synthesis of the leading and lagging strands requires two DNA polymerases to be tethered to the replisome. However, it has been suggested that there may be a third polymerase present at the replication fork. This stems from studies reporting differences in the composition of the clamp loader complex, which still remains a topic of debate.

As mentioned in section 1.3.2., the clamp loader complex consists of three copies of the *dnaX* gene product, and one copy each of the δ and δ' subunits[97]. The *dnaX* gene encodes for the τ subunit of the Pol III holoenzyme, which has a C-terminal domain that binds tightly to Pol IIIα[41]. The γ subunit is also expressed from the *dnaX* gene, but due to ribosomal frameshifting during protein synthesis, the γ subunit lacks the polymerase binding domain present in the full-length τ protein[9 29 128]. However, because the γ subunit retains the first three domains of the τ

protein that are responsible for carrying out clamp loading, a clamp loader with the composition $\gamma_3\delta_1\delta'_1$ is fully active in loading clamps onto DNA *in vitro*. Nonetheless, since only the τ protein is able to bind to the polymerase and two polymerases are required for the synthesis of the leading and lagging strands, the *in vivo* clamp loader is thought to be composed of at least two copies of the τ protein in the form of $\tau_2\gamma_1\delta_1\delta'_1$[20].

In contrast, other studies have suggested the presence of a third polymerase attached to the *in vivo* clamp loader complex in the replisome[43 83 102]. These studies show that replisomes consisting of the $\tau_3\delta_1\delta'_1$ clamp loader complex with three polymerases bound are more processive on DNA[43 83]. This is supported by studies showing that the τ protein, but not γ, is essential for viability in *E. coli*[8].

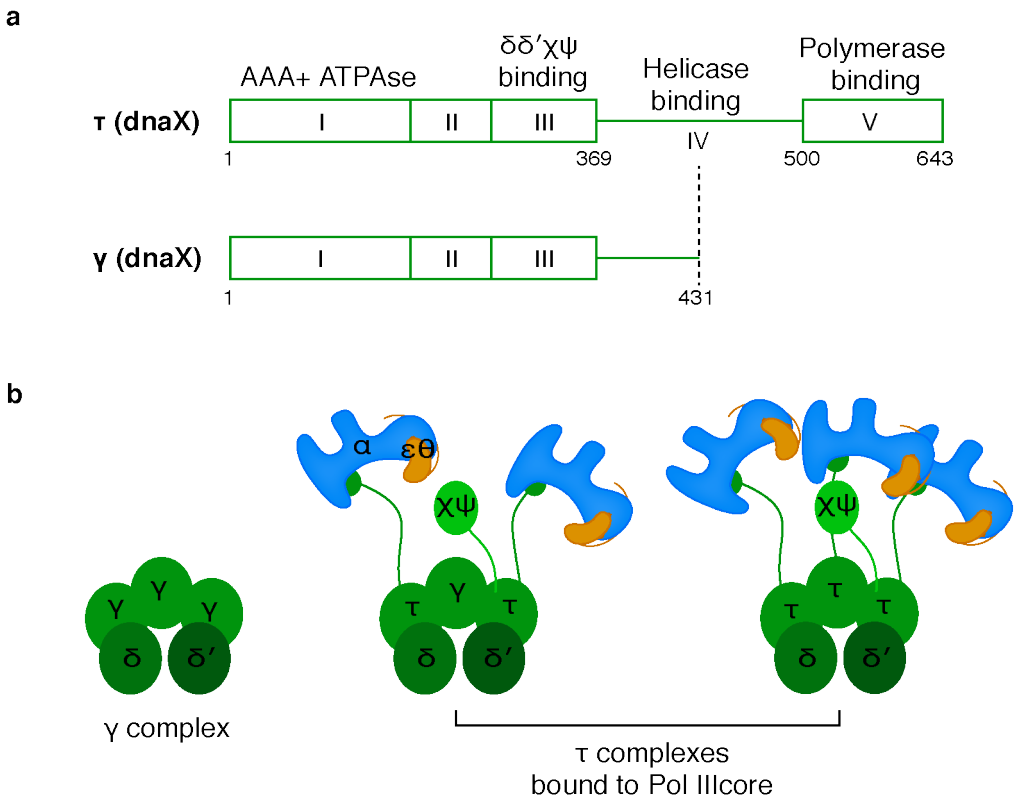


Figure 1.5. The *E. coli* clamp loaders. **a**, Cartoon representation of the domain composition of the two *dnaX* gene products τ and γ. **b**, Subunit arrangement in the γ and τ clamp loader complexes and their binding to Pol IIIcore.

The *in vivo* role of the γ subunit remains unclear, but it may be involved in clamp loader activity that is independent of polymerase binding. For example, the clamp loader complex is also known to be involved in the unloading of clamps once DNA synthesis has occurred, especially on the lagging strand where multiple clamps are loaded during Okazaki fragment synthesis. The unloading of clamps allows them to be recycled to allow for continuous DNA

synthesis[139]. Given that the unloading of clamps occurs post-replication, the $\gamma_3\delta_1\delta'_1$ clamp loader may be better suited for this role[13]. Furthermore, the β clamp is known to serve as a platform for other proteins, including those involved in DNA repair, and in these instances the loading of the clamp in the absence of Pol III may be desirable[43]. The domain composition of the γ and τ clamp loaders and binding to Pol IIIcore are shown in Fig. 1.5.

1.3.6 Dynamics in replisome

It has been long thought that the synthesis of the leading DNA strand is continuous, where a single replisome is capable of synthesizing large fragments of DNA. This has been supported by *in vitro* studies that have shown that single replisomes are capable of synthesizing 100 kbp of DNA[124 140].

However, recent studies suggest that the *E. coli* replisome may be much more dynamic than previously anticipated. Work done by Beattie *et al.* and Lewis *et al.* show that the Pol III*, which is the holoenzyme without the β clamp, can frequently exchange within the replisome[3 73], demonstrating that the synthesis of both the leading and lagging strands is discontinuous. The dynamics in DNA replication may allow the replicative DNA polymerase to overcome any roadblocks during DNA synthesis, for example in the form of lesions on the template DNA strand (see section 1.4) and protein-bound complexes on DNA[45], while maintaining the high processivity required for DNA replication.

Furthermore, it has been assumed that the physical association between the helicase and the clamp loader τ protein is important for coordinating the unwinding of the DNA duplex with DNA synthesis in order to prevent long stretches of ssDNA from being generated. However, a recent study shows that the activity of the helicase and the leading-strand polymerase can become temporarily uncoupled, with unwinding proceeding in the absence of DNA synthesis, albeit at a reduced rate[44]. Graham *et al.* also suggest that leading and lagging strand DNA synthesis can occur independently with no coordination between the leading and lagging polymerases. This stems from the observation that leading strand synthesis is independent of primase concentration and therefore lagging strand synthesis. This challenges the long-held view that the synthesis of the two anti-parallel DNA strands is coordinated.

Taken together, these studies show that while the composition of the *E. coli* replisome has been studied for a long time, the dynamics of the interactions between the individual components is only beginning to be understood.

1.4 DNA damage tolerance in *E. coli*

When the DNA polymerase makes an error in the incorporation of nucleotides, the mismatched nucleotide is removed by the proofreading exonuclease. However, when the error is located in the template DNA strand, the high fidelity polymerase is stalled. The replisome is stalled at sites of DNA damage on the template strand because the replicative DNA polymerases are specialized for high fidelity replication of DNA templates that contain A, T, G or C. When the bases are chemically altered, they may act as roadblocks for the progression of the replicative DNA polymerase.

Interestingly, not all chemically modified DNA bases lead to the stalling of the replication fork and these are called miscoding lesions[35]. Miscoding lesions normally correspond to only small changes to the bases, including modifications caused by reactive oxygen species (8-oxo-G). While the replicative DNA polymerase is able to incorporate dNTPs opposite miscoding lesions, these sites are very prone to point mutations.

Other modifications to DNA bases lead to the stalling of the replicative DNA polymerase and these are called replication-blocking lesions[35]. Stalled replication forks are dangerous for the cell because they are highly unstable in structure and can collapse, rearrange or break. Multiple mechanisms exist to relieve the stalling of the replication fork and allow the DNA damage to be repaired at a later point using the various DNA repair pathways in the cell.

Some of these mechanisms are error-prone whereas others are error-free. During translesion synthesis the high fidelity DNA polymerase is transiently replaced by specialized translesion DNA polymerases that use error-prone synthesis to bypass the DNA damage. Other DNA damage tolerance pathways such as post-replicative gap filling or replication fork regression are error-free as they use recombination with undamaged segments of the genome to replicate DNA[32]. It remains to be understood when and how the cell makes a choice between the various pathways for avoiding replication opposite a site of DNA damage.

1.4.1 Translesion synthesis

Translesion synthesis relies on the activity of alternative DNA polymerases to relieve the stalling of the replication fork. The first DNA polymerase isolated in *E. coli* was DNA polymerase I[70], which was originally thought to be the replicative polymerase until the isolation of a viable Pol I-deficient strain, which showed that DNA replication can resume in the absence of Pol I activity[18]. Two other DNA polymerases were then purified from *E. coli*, Pol II and Pol III[64,65]. Further characterization of these polymerases by biochemistry and genetics confirmed that Pol

III was in fact the main replicative DNA polymerase while Pol I was involved in Okazaki fragment maturation. The role of Pol II in the cell remained unclear. For a period of about 30 years after the isolation of Pol I, II and III, it was thought that there were only DNA polymerases present in *E. coli*.

Screening for mutations that prevented mutagenesis induced by UV-irradiation in *E. coli* led to the identification of the *UmuC* and *UmuD* genes[55] that were initially thought to alter the stringency of Pol III[23,24], leading to the bypass of arrest-inducing DNA lesions. In an early *in vitro* TLS assay, it was found that purified Pol III was unable to perform TLS, whereas the addition of the UmuC and UmuD proteins led to efficient synthesis past the lesion[100]. It was found later that TLS also occurred using purified UmuC and UmuD proteins[121] and that the *UmuDC* locus encoded a bona fide DNA polymerase and was named DNA polymerase V[101,123]. At around the same time, the *dinB* gene was found to encode for another DNA polymerase that is structural homologous to DNA polymerase V which was named DNA polymerase IV[131].

DNA polymerases are divided into different families based on the similarities between their domain structures[35]. In *E. coli*, the replicative Pol III belongs to the C family of DNA polymerases, while the specialized lesion bypass polymerases IV (dinB) and V (UmuC and UmuD) belong to the Y-family DNA polymerases[91]. The Y-family of DNA polymerases have unique structural properties that allow them to insert dNTPs across a site of lesion at which the replicative DNA polymerase is stalled[31,138,28]. For example, the DNA polymerases have low fidelity as a result of their enlarged active sites that allow bulky chemical groups on the modified bases to be accommodated. In addition, they lack the 3'-5' exonuclease domain that is present in the replicative polymerase and is responsible for removing any incorrectly incorporated nucleotides by the polymerase. Furthermore, the Y-family polymerases have low processivity on DNA and an extra domain not present in polymerases of the other families, called the little-fingers domain, which provides additional contacts to the DNA. This family of DNA polymerases are the least accurate and incorporate nucleotides with an error rate of 10^{-1} to 10^{-3}[68]. A comparison between the three-dimensional structures of the replicative and translesion polymerases is shown in 1.6.

In addition, another *E. coli* DNA polymerase, Pol II, has been shown to be involved in translesion synthesis, although it belongs to the B-family of high fidelity polymerases that possess proofreading activity[35]. All three polymerases Pol IV, Pol V and Pol II have been shown to bind to the circular β clamp to catalyze the bypass of lesions on the template DNA strand[7].

Given that the translesion DNA polymerases are more error-prone, their access to the replication fork needs to be tightly regulated, as otherwise they would result in increased mutation

rates. This is achieved by placing the expression of the translesion polymerase genes under a process called the SOS response[99]. Under normal cellular growth conditions, the expression of the SOS genes are repressed by a repressor called LexA[75]. However, as the replicative polymerase is stalled by DNA lesions, stretches of single-stranded DNA begin to accumulate in the cell due to the continued DNA unwinding by the helicase. These single-stranded regions become coated by a protein called RecA that recognizes and forms a filament on the single-stranded DNA. The RecA filament catalyzes the auto-proteolysis of the LexA repressor, which in turn allows the LexA-repressed genes to become transcribed[75].

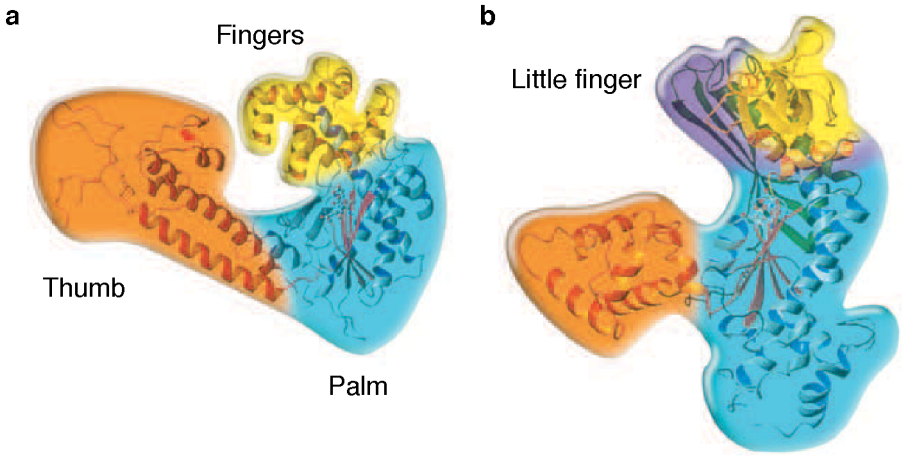


Figure 1.6. Comparing the structures of the replicative and the Y-family DNA polymerases. a, Schematic representation of the crystal structure of a high-fidelity replicative DNA polymerase and **b** an error-prone Y-family translesion DNA polymerase. The Y-family polymerase has an enlarged active site located between the thumb and fingers domains and an additional little finger domain shown in purple. Figure adapted from Friedberg, 2015[32].

While the roles of Pol II (*polB*), Pol IV (*dinB*) and Pol V (*umuDC*) in translesion synthesis have been established using a number of biochemical and genetics studies, it is interesting to note that these polymerases are non-essential inside the cell. In fact, strains in which all three polymerases have been mutated away do not show strong UV sensitivity[17], therefore suggesting that TLS plays a minor role in terms of survival to genotoxic agents.

1.4.2 Translesion synthesis by DNA polymerase IV

The expression levels of the Y-family polymerase Pol IV increases ten fold following the induction of the SOS response[61]. Studies have reported that when Pol IV is expressed at levels found under normal growth conditions (250 molecules/cell), it does not result in increased mu-

tation rates[67,136]. Over-expression of Pol IV at over 10,000 molecules/cell results in decreased replication fidelity and the cells show low level growth defects[60,132,66]. Finally, further increase in Pol IV expression to over 100,000 molecules/cell is lethal[129].

Pol IV has been shown to bypass a large number of lesions by biochemical experiments, including TT photo-adducts, acetylaminofluororene-derived DNA adducts, and AP sites[122,120,79]. However, genetic evidence for the role of Pol IV as a translesion polymerase has only been found for a few N2-guanine adducts[71,87,61]. The common characteristic between these lesions is their location in the minor groove of DNA, suggesting that Pol IV is efficient at bypassing minor groove DNA lesions.

Pol IV activity on DNA is distributive as it is only able to synthesize one nucleotide per binding event. However, this is increased upon binding to the ring-shaped β clamp, which increases the processivity of Pol IV on DNA to 30-400 nucleotides per binding event[130].

1.4.3 Translesion synthesis by DNA polymerase V

Unlike Pol IV that acts as a single protein, Pol V is a heterotrimer consisting of one copy of the *umuC* and two copies of the *umuD* gene products in the form of UmuD'$_2$C, where UmuD' is a proteolytic product of the UmuD protein[101,123]. Polymerase activity is associated with UmuC, while the UmuD'$_2$ homodimer acts as accessory proteins. *umuDC* strains of *E. coli* show much reduced UV-induced mutation rates[55,113], confirming its role in error-prone translesion synthesis. Pol V has been shown to bypass UV-induced lesions efficiently[122,36].

While Pol IV is normally present in the cell, Pol V cannot be detected biochemically in the cell under normal growth conditions. It has also been shown that a TT(6-4) lesion is bypassed at similar efficiency in UmuDC deficient and wild-type cells, suggesting that Pol V is not involved in UV-induced mutagenesis in non-SOS-induced cells[5]. The expression of the *umuDC* locus is SOS controlled, however interestingly, it is only induced during persistent SOS response with the level of Pol V peaking after 45 minutes of UV irradiation[112]. This shows that Pol V is not recruited as a first tier polymerase and instead forms the "late SOS response", which may be explained by the strong repression of the *umuDC* promoter by LexA. In addition to its late expression, Pol V activation proceeds via a complicated sequence of events, where the RecA filament induces the autoproteolysis of UmuD, forming the cleaved UmuD' product which dimerizes[14,89,109]. The UmuC protein is sequestered to the cell membrane, and is released from the membrane through binding to the UmuD'$_2$ dimer[103]. The sequestration of the UmuC to the cell membrane is a further mechanism to control the intracellular level of this error-

prone polymerase. Lastly, the UmuC-UmuD'$_2$ heterotrimer binds to RecA to become fully activated[107]. Given the complicated activation mechanism and its late appearance in the SOS response, it is likely that Pol V is recruited only when the other DNA damage tolerance and repair pathways have been exhausted.

Pol V has a processivity on DNA of approximately 25 nucleotides per binding event in the presence of the β clamp. However, it is significantly slower in DNA synthesis as compared to Pol III ($\sim$ 1000 nt/s) or Pol IV ($\sim$ 2 nt/s), all measured in the presence of the β clamp. The rate of synthesis by Pol V is only $\sim$ 0.3 nt/s[35].

1.4.4 Translesion synthesis by DNA polymerase II

As mentioned above, in contrast to Pol IV and Pol V that act as specialized DNA polymerases in translesion synthesis, Pol II is a high fidelity DNA polymerase with 3'-5' exonuclease activity. However, the expression of Pol II has been shown to increase approximately seven-fold during the SOS response[98] unlike the replicative Pol III, suggesting that it is involved in the DNA damage response.

The role of Pol II as a translesion DNA polymerase was confirmed by both *in vivo* and *in vitro* studies that show that it catalyzes the bypass of a G-AAF adduct located within a specific -2 frameshift sequence[87 6]. In addition, Pol II has shown to bypass AP-sites with high efficiency *in vitro*[94]. A crystal structure of Pol II in complex with an AP lesion containing DNA shows that the polymerase is able to accommodate for the mutation, despite its high fidelity active site[133]. This is achieved through an insertion that creates an additional cavity upstream of the active site to allow for the DNA lesion in the template to be accommodated. Interestingly, *E. coli* cells that lack Pol II do not display any phenotype under normal growth conditions.

1.4.5 Polymerase exchange during translesion synthesis

Given that *E. coli* has one replicative DNA polymerase and three polymerases involved in translesion synthesis, the interplay and switching between these polymerases have been a subject of interest.

Studies have shown that instead of delegating lesion bypass to the translesion DNA polymerases, the replicative Pol III may be actively involved in determining the outcome of the translesion synthesis event. When Pol V synthesizes four nucleotides or less across the site of DNA damage, Pol III will degrade the newly synthesized stretch of DNA using its 3'-5' exonuclease activity[36]. When the nascent stretch of DNA synthesized by Pol V is five nucleotides

or longer, Pol III extends it to complete the lesion bypass event. The same rule has also been demonstrated for Pol IV, where extension of at least four nucleotides after the insertion opposite the lesion is required to protect the nascent strand from degradation by Pol III's proofreading activity[51,30]. The importance of the length of DNA synthesized past the site of DNA damage underlines the requirement for the TLS polymerases to bind to the β clamp in order to synthesize a long enough stretch of DNA that is beyond the critical length required for Pol III extension. Consequently, the polymerases have been shown to bind to the β clamp[7].

In addition, Pol III has been shown to modulate the product of Pol II translesion synthesis. Studies have shown that Pol II produces a -2 frameshift mutation at a G-AAF adduct in a specific sequence context *in vivo*[34,62]. However, Pol II also produces a significant level of a -1 frameshift mutation *in vitro* that are not detected inside the cell[34,62]. This balance between the two outcomes of extending past the lesion is tilted towards the production of the -2 frameshift mutation in the presence of Pol III in the *in vitro* experiments[37].

Given that both the replicative and translesion DNA polymerases require binding to the β clamp to gain processivity on DNA[7], the mechanism by which they switch on the β clamp remains to be understood. A model for the exchange between a replicative and specialized polymerase, called the toolbelt model, predicts that the two can bind simultaneously to the clamp, one on each binding site on the clamp, and switch places across a site of DNA lesion 1.7[49]. This way, the specialized DNA polymerase acts in a manner analogous to the proofreading exonuclease: when the replicative polymerase inserts an incorrect nucleotide during synthesis, it is removed by the exonuclease, whereas when it is stalled due to the presence of a lesion on the template DNA strand, the DNA is channeled to the translesion polymerase for the incorporation of nucleotides immediately opposite the site of lesion. The exonuclease and translesion polymerases therefore act as as the tools to allow the replicative DNA polymerase to overcome potential roadblocks during DNA replication. The toolbelt model is discussed in more detail in Chapter 4.

An alternative to the toolbelt model is the competition for binding to the clamp on DNA by the replicative and translesion polymerases by 'mass action', where the outcome of the competition is simply determined by the relative concentrations of the different polymerases. This model is supported by studies showing that the bypass of a N2-acetylaminofluorene guanine adduct by Pol V or Pol II depends on the relative concentrations of the two polymerases[6,36]. The concentration-dependence of the polymerase competition may explain why the expression of the translesion polymerases is increased during the bacterial SOS DNA damage response[119].

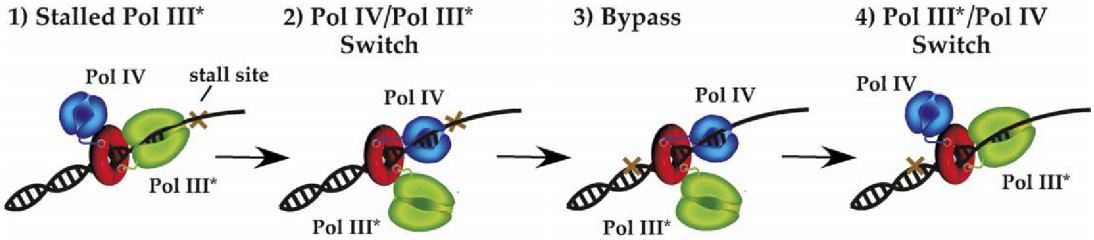


Figure 1.7. The toolbelt model. Model predicting the simultaneous binding of a high fidelity replicative DNA polymerase and a specialized translesion DNA polymerase to the β clamp. The polymerases alternate binding to DNA across a site of lesion on the template DNA. Figure taken from Indiani *et al*, 2015[49].

1.4.6 Error-free DNA damage tolerance pathways

As was mentioned previously, there are also multiple DNA damage tolerance pathways in *E. coli* that are independent of the translesion DNA polymerases. These pathways avoid replication directly opposite the damaged DNA base and use the information contained in the undamaged sister chromatid to continue with replication. Thus, these pathways are not mutagenic.

In a historically important paper published in 1968, Rupp and Howard-Flanders first identified what is known today as DNA damage tolerance[106]. They reported that after exposing *E. coli* cells to UV irradiation, DNA becomes synthesized in short fragments that are later converted into the larger molecular weight DNA present in the non-irradiated cells. This was the first indication that once the replication fork is arrested by the photo-adducts caused by exposing cells to UV, the replication fork re-initiates downstream of the site of lesion. The re-initiation of the replication fork therefore leaves gaps in the daughter DNA strand corresponding to the positions of UV-induced DNA damage. It was supported by the observation that the gaps are of approximately 1000 nucleotides in size, which corresponds to the average distance between UV-induced photo-adducts on DNA[50].

Re-priming downstream of the site of damage therefore relieves replication fork stalling as it removes the requirement for synthesis directly opposite the lesion. The re-priming events depend on the primosome, a protein complex that catalyzes the synthesis of a RNA primer. A central component of the primosome is the PriA helicase, and studies have shown that PriA deficient cells are poorly viable after exposure to UV irradiation[90]. However, PriA-independent replication fork re-priming events have also been identified[76]. More recently, it has been shown that re-priming downstream of a site of damage does not require the dissociation of the replisome or replication restart proteins, but instead, the replisome remains attached to the DNA

and re-initiates synthesis by a DnaG primase-dependent mechanism[141]. The gaps generated as a result of re-priming are filled through recombination events between the two newly synthesized daughter strands, as they have been shown to be dependent on the products of multiple recombination *rec* genes[111].

Another error-free pathway for avoiding DNA synthesis across a damaged DNA base is achieved through a structural rearrangement of the replication fork, whereby the nascent daughter strands are extruded through the re-pairing of the two parental strands and form a duplex[95]. This way, instead of using the damaged parental DNA strand as template for copying DNA, the daughter strand now uses the other undamaged daughter strand as template for synthesis. This pathway is known as replication fork regression or copy-choice replication after the fact that a choice is made between the strands that can act as the template for synthesis. When the replication fork has been resolved from the temporarily extruded daughter strands structure back to parental-daughter strand pairing, the damaged base is bypassed. The mechanisms involved in the resolution of such structures remain to be understood.

1.5 Thesis objectives

While the *E. coli* system has been the paradigm for studying DNA replication for many years, and the structure and function of the individual factors are well-characterized, it remains insufficiently understood how the factors with distinct functions are organized to carry out high fidelity DNA replication. Several individual interactions have been identified, for example the association between the helicase and the primase, the binding of Pol IIIα to the exonuclease ϵ to form the Pol IIIcore, and the interaction between the clamp loader protein τ and Pol IIIα. However, the overall structure of the replisome is highly dynamic, as evidenced by the challenge of solving structures of the replisome containing more than four factors (β, α, ϵ, and a C-terminal fragment of τ)[27] and the growing number of single-molecule studies showing the exchange of factors within the replisome[3 73]. Given the dynamics of the *E. coli* replisome, the temporal transition between the various sub-states of the replication cycle is crucial for understanding its function. One aim of this PhD thesis is therefore to study the temporal organization of the events involved in DNA replication.

Another area of study for this thesis are the events at the interface of DNA replication and translesion synthesis, and more specifically, how the exchange between the high-fidelity and specialized polymerases occurs. Given that the β clamp plays a pivotal role in controlling the access of the translesion polymerases to DNA, a major challenge for the access of the poly-

merases to the clamp is the repeated loading of clamps on the lagging strand due to its discontinuous synthesis. How the loading and unloading of the clamp and replicative polymerase Pol IIIα on the lagging strand is coordinated, while preventing the untimely association of the translesion polymerases, has not been studied so far.

In order to understand temporal organization of the processes in DNA replication and translesion synthesis and resolve whether the replicative and specialized DNA polymerases can simultaneously bind to the clamp, this PhD thesis aims to set up multi-color co-localization single molecule spectroscopy (CoSMoS)[33] as a means to directly visualize the binding and release of the various factors in real-time at the single-molecule level.

As a whole, this PhD thesis is a step towards better understanding how the cell achieves a balance between high fidelity DNA replication and error-prone DNA synthesis, and how a choice between replication fidelity and mutagenesis is made.

Bibliography

[1] T A Baker, B E Funnell, and A Kornberg. Helicase action of dnaB protein during replication from the Escherichia coli chromosomal origin in vitro. *The Journal of Biological Chemistry*, 262(14):6877–6885, May 1987.

[2] T A Baker, K Sekimizu, B E Funnell, and A Kornberg. Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the Escherichia coli chromosome. *Cell*, 45(1):53–64, April 1986.

[3] Thomas R Beattie, Nitin Kapadia, Emilien Nicolas, Stephan Uphoff, Adam Jm Wollman, Mark C Leake, and Rodrigo Reyes-Lamothe. Frequent exchange of the DNA polymerase during bacterial chromosome replication. *eLife*, 6:e21763, March 2017.

[4] Thomas R Beattie and Rodrigo Reyes-Lamothe. A Replisome's journey through the bacterial chromosome. *Frontiers in Microbiology*, 6:562, 2015.

[5] O J Becherel and R P Fuchs. SOS mutagenesis results from up-regulation of translesion synthesis. *Journal of Molecular Biology*, 294(2):299–306, November 1999.

[6] O J Becherel and R P Fuchs. Mechanism of DNA polymerase II-mediated frameshift mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15):8566–8571, July 2001.

[7] Olivier J Becherel, Robert P P Fuchs, and Jérôme Wagner. Pivotal role of the beta-clamp in translesion DNA synthesis and mutagenesis in E. coli cells. *DNA Repair*, 1(9):703–708, September 2002.

[8] A Blinkova, C Hervas, P T Stukenberg, R Onrust, M E O'Donnell, and J R Walker. The Escherichia coli DNA polymerase III holoenzyme contains both products of the dnaX gene, tau and gamma, but only tau is essential. *Journal of Bacteriology*, 175(18):6018–6027, September 1993.

[9] A L Blinkowa and J R Walker. Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Research*, 18(7):1725–1729, April 1990.

[10] J P Bouché, K Zechel, and A Kornberg. dnaG gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form. *The Journal of Biological Chemistry*, 250(15):5995–6001, August 1975.

[11] D Bramhill and A Kornberg. A model for initiation at origins of DNA replication. *Cell*, 54(7):915–918, September 1988.

[12] D Bramhill and A Kornberg. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. *Cell*, 52(5):743–755, March 1988.

[13] I Bruck and M O'Donnell. The DNA replication machine of a gram-positive organism. *The Journal of Biological Chemistry*, 275(37):28971–28983, September 2000.

[14] S E Burckhardt, R Woodgate, R H Scheuermann, and H Echols. UmuD mutagenesis protein of Escherichia coli: overproduction, purification, and cleavage by RecA. *Proceedings of the National Academy of Sciences of the United States of America*, 85(6):1811–1815, March 1988.

[15] F F Coskun-Ari and T M Hill. Sequence-specific interactions in the Tus-Ter complex and the effect of base pair substitutions on arrest of DNA replication in Escherichia coli. *The Journal of Biological Chemistry*, 272(42):26448–26456, October 1997.

[16] Alessandro Costa, Iris V Hood, and James M Berger. Mechanisms for initiating cellular DNA replication. *Annual Review of Biochemistry*, 82(1):25–54, 2013.

[17] Charmain T Courcelle, Jerilyn J Belle, and Justin Courcelle. Nucleotide excision repair or polymerase V-mediated lesion bypass can act to restore UV-arrested replication forks in Escherichia coli. *Journal of Bacteriology*, 187(20):6953–6961, October 2005.

[18] P De Lucia and J Cairns. Isolation of an E. coli strain with a mutation affecting DNA polymerase. *Nature*, 224(5225):1164–1166, December 1969.

[19] James M Dewar and Johannes C Walter. Mechanisms of DNA replication termination. *Nature Reviews Molecular Cell Biology*, 18(8):507–516, August 2017.

[20] Paul R Dohrmann, Raul Correa, Ryan L Frisch, Susan M Rosenberg, and Charles S McHenry. The DNA polymerase III holoenzyme contains γ and is not a trimeric polymerase. *Nucleic Acids Research*, 44(3):1285–1297, February 2016.

[21] D Allan Drummond and Claus O Wilke. The evolutionary consequences of erroneous protein synthesis. *Nature Reviews Genetics*, 10(10):715–724, October 2009.

[22] Iain G Duggin, R Gerry Wake, Stephen D Bell, and Thomas M Hill. The replication fork trap and termination of chromosome replication. *Molecular Microbiology*, 70(6):1323–1333, December 2008.

[23] H Echols and M F Goodman. Mutation induced by DNA damage: a many protein affair. *Mutation Research*, 236(2-3):301–311, September 1990.

[24] H Echols and M F Goodman. Fidelity mechanisms in DNA replication. *Annual Review of Biochemistry*, 60(1):477–511, 1991.

[25] Mohamed M Elshenawy, Slobodan Jergic, Zhi-Qiang Xu, Mohamed A Sobhy, Masateru Takahashi, Aaron J Oakley, Nicholas E Dixon, and Samir M Hamdan. Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. *Nature*, 525(7569):394–398, September 2015.

[26] P J Fay, K O Johanson, C S McHenry, and R A Bambara. Size classes of products synthesized processively by two subassemblies of Escherichia coli DNA polymerase III holoenzyme. *The Journal of Biological Chemistry*, 257(10):5692–5699, May 1982.

[27] Rafael Fernandez-Leiro, Julian Conrad, Sjors Hw Scheres, and Meindert H Lamers. cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *eLife*, 4:459, October 2015.

[28] Oliver Fleck and Primo Schär. Translesion DNA synthesis: little fingers teach tolerance. *Current Biology*, 14(10):R389–91, May 2004.

[29] A M Flower and C S McHenry. The gamma subunit of DNA polymerase III holoenzyme of Escherichia coli is produced by ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America*, 87(10):3713–3717, May 1990.

[30] James J Foti and Graham C Walker. Efficient extension of slipped DNA intermediates by DinB is required to escape primer template realignment by DnaQ. *Journal of Bacteriology*, 193(10):2637–2641, May 2011.

[31] E C Friedberg, P L Fischhaber, and C Kisker. Error-prone DNA polymerases: novel structures and the benefits of infidelity. *Cell*, 107(1):9–12, October 2001.

[32] Errol C Friedberg. Suffering in silence: the tolerance of DNA damage. *Nature Reviews Molecular Cell Biology*, 6(12):943–953, December 2005.

[33] Larry J Friedman, Johnson Chung, and Jeff Gelles. Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. *Biophysical Journal*, 91(3):1023–1031, August 2006.

[34] R P Fuchs, N Schwartz, and M P Daune. Hot spots of frameshift mutations induced by the ultimate carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature*, 294(5842):657–659, December 1981.

[35] Robert P Fuchs and Shingo Fujii. Translesion DNA synthesis and mutagenesis in prokaryotes. *Cold Spring Harbor Perspectives in Biology*, 5(12):a012682–a012682, December 2013.

[36] Shingo Fujii and Robert P Fuchs. Defining the position of the switches between replicative and bypass DNA polymerases. *The EMBO Journal*, 23(21):4342–4352, October 2004.

[37] Shingo Fujii and Robert P Fuchs. Interplay among replicative and specialized DNA polymerases determines failure or success of translesion synthesis pathways. *Journal of Molecular Biology*, 372(4):883–893, September 2007.

[38] R S Fuller, B E Funnell, and A Kornberg. The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. *Cell*, 38(3):889–900, October 1984.

[39] B E Funnell, T A Baker, and A Kornberg. In vitro assembly of a prepriming complex at the origin of the Escherichia coli chromosome. *The Journal of Biological Chemistry*, 262(21):10327–10334, July 1987.

[40] D Gao and C S McHenry. 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. *The Journal of Biological Chemistry*, 276(6):4441–4446, February 2001.

[41] D Gao and C S McHenry. tau binds and organizes Escherichia coli replication through distinct domains. Partial proteolysis of terminally tagged tau to determine candidate domains and to assign domain V as the alpha binding domain. *The Journal of Biological Chemistry*, 276(6):4433–4440, February 2001.

[42] Roxana E Georgescu, Seung-Sup Kim, Olga Yurieva, John Kuriyan, Xiang-Peng Kong, and Mike O'Donnell. Structure of a sliding clamp on DNA. *Cell*, 132(1):43–54, January 2008.

[43] Roxana E Georgescu, Isabel Kurth, and Mike E O'Donnell. Single-molecule studies reveal the function of a third polymerase in the replisome. *Nature Structural & Molecular Biology*, 19(1):113–116, December 2011.

[44] James E Graham, Kenneth J Marians, and Stephen C Kowalczykowski. Independent and Stochastic Action of DNA Polymerases in the Replisome. *Cell*, 169(7):1201–1213.e17, June 2017.

[45] Milind K Gupta, Colin P Guy, Joseph T P Yeeles, John Atkinson, Hazel Bell, Robert G Lloyd, Kenneth J Marians, and Peter McGlynn. Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli. *Proceedings of the National Academy of Sciences*, 110(18):7252–7257, April 2013.

[46] Mark Hedglin, Ravindra Kumar, and Stephen J Benkovic. Replication clamps and clamp loaders. *Cold Spring Harbor Perspectives in Biology*, 5(4):a010165–a010165, April 2013.

[47] H Hiasa and K J Marians. Tus prevents overreplication of oriC plasmid DNA. *The Journal of Biological Chemistry*, 269(43):26959–26968, October 1994.

[48] T M Hill, J M Henson, and P L Kuempel. The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication. *Proceedings of the National Academy of Sciences of the United States of America*, 84(7):1754–1758, April 1987.

[49] C Indiani, P McInerney, R Georgescu, and M F Goodman. A sliding-clamp toolbelt binds high-and low-fidelity DNA polymerases simultaneously. *Molecular Cell*, 2005.

[50] V N Iyer and W D Rupp. Usefulness of benzoylated naphthoylated DEAE-cellulose

to distinguish and fractionate double-stranded DNA bearing different extents of single-stranded regions. *Biochimica et Biophysica Acta*, 228(1):117–126, January 1971.

[51] Daniel F Jarosz, Susan E Cohen, James C Delaney, John M Essigmann, and Graham C Walker. A DinB variant reveals diverse physiological consequences of incomplete TLS extension by a Y-family DNA polymerase. *Proceedings of the National Academy of Sciences*, 106(50):21137–21142, December 2009.

[52] Slobodan Jergic, Nicholas P Horan, Mohamed M Elshenawy, Claire E Mason, Thitima Urathamakul, Kiyoshi Ozawa, Andrew Robinson, Joris M H Goudsmits, Yao Wang, Xuefeng Pan, Jennifer L Beck, Antoine M van Oijen, Thomas Huber, Samir M Hamdan, and Nicholas E Dixon. A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *The EMBO Journal*, 32(9):1322–1333, May 2013.

[53] D Jeruzalmi, M O'Donnell, and J Kuriyan. Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. *Cell*, 106(4):429–441, August 2001.

[54] Erik Johansson and Nicholas Dixon. Replicative DNA polymerases. *Cold Spring Harbor Perspectives in Biology*, 5(6):a012799–a012799, June 2013.

[55] T Kato and Y Shinoura. Isolation and characterization of mutants of Escherichia coli deficient in induction of mutations by ultraviolet light. *Molecular & General Genetics*, 156(2):121–131, November 1977.

[56] Brian A Kelch, Debora L Makino, Mike O'Donnell, and John Kuriyan. How a DNA polymerase clamp loader opens a sliding clamp. *Science*, 334(6063):1675–1680, December 2011.

[57] Brian A Kelch, Debora L Makino, Mike O'Donnell, and John Kuriyan. Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biology*, 10(1):34, April 2012.

[58] Z Kelman, A Yuzhakov, J Andjelkovic, and M O'Donnell. Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *The EMBO Journal*, 17(8):2436–2449, April 1998.

[59] D R Kim and C S McHenry. In vivo assembly of overproduced DNA polymerase III. Overproduction, purification, and characterization of the alpha, alpha-epsilon, and alpha-epsilon-theta subunits. *The Journal of Biological Chemistry*, 271(34):20681–20689, August 1996.

[60] S R Kim, G Maenhaut-Michel, M Yamada, Y Yamamoto, K Matsui, T Sofuni, T Nohmi, and H Ohmori. Multiple pathways for SOS-induced mutagenesis in Escherichia coli: an overexpression of dinB/dinP results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 94(25):13792–13797, December 1997.

[61] S R Kim, K Matsui, M Yamada, P Gruz, and T Nohmi. Roles of chromosomal and episomal dinB genes encoding DNA pol IV in targeted and untargeted mutagenesis in Escherichia coli. *Molecular Genetics and Genomics*, 266(2):207–215, October 2001.

[62] N Koffel-Schwartz, J M Verdier, M Bichara, A M Freund, M P Daune, and R P Fuchs. Carcinogen-induced mutation spectrum in wild-type, uvrA and umuC strains of Escherichia coli. Strain specificity and mutation-prone sequences. *Journal of Molecular Biology*, 177(1):33–51, July 1984.

[63] X P Kong, R Onrust, M O'Donnell, and J Kuriyan. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, 69(3):425–437, May 1992.

[64] T Kornberg and M L Gefter. Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 68(4):761–764, April 1971.

[65] T Kornberg and M L Gefter. Deoxyribonucleic acid synthesis in cell-free extracts. IV. Purification and catalytic properties of deoxyribonucleic acid polymerase III. *The Journal of Biological Chemistry*, 247(17):5369–5375, September 1972.

[66] Wojciech Kuban, Magdalena Banach-Orlowska, Malgorzata Bialoskorska, Aleksandra Lipowska, Roel M Schaaper, Piotr Jonczyk, and Iwona J Fijalkowska. Mutator phenotype resulting from DNA polymerase IV overproduction in Escherichia coli: preferential mutagenesis on the lagging strand. *Journal of Bacteriology*, 187(19):6862–6866, October 2005.

[67] Wojciech Kuban, Piotr Jonczyk, Damian Gawel, Karolina Malanowska, Roel M Schaaper, and Iwona J Fijalkowska. Role of Escherichia coli DNA polymerase IV in in vivo replication fidelity. *Journal of Bacteriology*, 186(14):4802–4807, July 2004.

[68] Thomas A Kunkel. DNA replication fidelity. *The Journal of Biological Chemistry*, 279(17):16895–16898, April 2004.

[69] Meindert H Lamers, Roxana E Georgescu, Sang-Gyu Lee, Mike O'Donnell, and John Kuriyan. Crystal structure of the catalytic alpha subunit of E. coli replicative DNA polymerase III. *Cell*, 126(5):881–892, September 2006.

[70] I R Lehman, M J Bessman, E S Simms, and A Kornberg. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. *The Journal of Biological Chemistry*, 233(1):163–170, July 1958.

[71] N Lenne-Samuel, R Janel-Bintz, A Kolbanovskiy, N E Geacintov, and R P Fuchs. The processing of a Benzo(a)pyrene adduct into a frameshift or a base substitution mutation requires a different set of genes in Escherichia coli. *Molecular Microbiology*, 38(2):299–307, October 2000.

[72] Frank P Leu, Roxana Georgescu, and Mike O'Donnell. Mechanism of the E. coli tau processivity switch during lagging-strand synthesis. *Molecular Cell*, 11(2):315–327, February 2003.

[73] Jacob S Lewis, Lisanne M Spenkelink, Slobodan Jergic, Elizabeth A Wood, Enrico Monachino, Nicholas P Horan, Karl E Duderstadt, Michael M Cox, Andrew Robinson, Nicholas E Dixon, and Antoine M van Oijen. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *eLife*, 6:4846, April 2017.

[74] T Lindahl. The Croonian Lecture, 1996: endogenous damage to DNA. In *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, pages 1529–1538. Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts, U.K., The Royal Society, November 1996.

[75] J W Little, D W Mount, and C R Yanisch-Perron. Purified lexA protein is a repressor of the recA and lexA genes. *Proceedings of the National Academy of Sciences of the United States of America*, 78(7):4199–4203, July 1981.

[76] Susan T Lovett. Filling the gaps in replication restart pathways. *Molecular Cell*, 17(6):751–752, March 2005.

[77] Y B Lu, P V Ratnakar, B K Mohanty, and D Bastia. Direct physical interaction between DnaG primase and DnaB helicase of Escherichia coli is necessary for optimal synthesis of primer RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23):12902–12907, November 1996.

[78] H Maki and A Kornberg. Proofreading by DNA polymerase III of Escherichia coli depends on cooperative interaction of the polymerase and exonuclease subunits. *Proceedings of the National Academy of Sciences of the United States of America*, 84(13):4389–4392, July 1987.

[79] Ayelet Maor-Shoshani, Ken Hayashi, Haruo Ohmori, and Zvi Livneh. Analysis of translesion replication across an abasic site by DNA polymerase IV of Escherichia coli. *DNA Repair*, 2(11):1227–1238, November 2003.

[80] Micah J McCauley, Leila Shokri, Jana Sefcikova, Ceslovas Venclovas, Penny J Beuning, and Mark C Williams. Distinct double- and single-stranded DNA binding of E. coli replicative DNA polymerase III alpha subunit. *ACS Chemical Biology*, 3(9):577–587, September 2008.

[81] C S McHenry and W Crow. DNA polymerase III of Escherichia coli. Purification and identification of subunits. *The Journal of Biological Chemistry*, 254(5):1748–1753, March 1979.

[82] Charles S McHenry. DNA replicases from a bacterial perspective. *Annual Review of Biochemistry*, 80(1):403–436, 2011.

[83] Peter McInerney, Aaron Johnson, Francine Katz, and Mike O'Donnell. Characterization of a triple DNA polymerase replisome. *Molecular Cell*, 27(4):527–538, August 2007.

[84] M Meselson and F W Stahl. THE REPLICATION OF DNA IN ESCHERICHIA COLI. *Proceedings of the National Academy of Sciences of the United States of America*, 44(7):671–682, July 1958.

[85] R R Meyer and P S Laine. The single-stranded DNA-binding protein of Escherichia coli. *Microbiological Reviews*, 54(4):342–380, December 1990.

[86] Mark D Mulcair, Patrick M Schaeffer, Aaron J Oakley, Hannah F Cross, Cameron Neylon, Thomas M Hill, and Nicholas E Dixon. A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell*, 125(7):1309–1319, June 2006.

[87] R Napolitano, R Janel-Bintz, J Wagner, and R P Fuchs. All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *The EMBO Journal*, 19(22):6259–6265, November 2000.

[88] Cameron Neylon, Andrew V Kralicek, Thomas M Hill, and Nicholas E Dixon. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiology and Molecular Biology Reviews*, 69(3):501–526, September 2005.

[89] T Nohmi, J R Battista, L A Dodson, and G C Walker. RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation. *Proceedings of the National Academy of Sciences of the United States of America*, 85(6):1816–1820, March 1988.

[90] P Nurse, K H Zavitz, and K J Marians. Inactivation of the Escherichia coli priA DNA replication protein induces the SOS response. *Journal of Bacteriology*, 173(21):6686–6693, November 1991.

[91] H Ohmori, E C Friedberg, R P Fuchs, M F Goodman, F Hanaoka, D Hinkle, T A Kunkel, C W Lawrence, Z Livneh, T Nohmi, L Prakash, S Prakash, T Todo, G C Walker, Z Wang, and R Woodgate. The Y-family of DNA polymerases. *Molecular Cell*, 8(1):7–8, July 2001.

[92] R Okazaki, M Arisawa, and A Sugino. Slow joining of newly replicated DNA chains in DNA polymerase I-deficient Escherichia coli mutants. *Proceedings of the National Academy of Sciences of the United States of America*, 68(12):2954–2957, December 1971.

[93] R Okazaki, T Okazaki, K Sakabe, K Sugimoto, and A Sugino. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, 59(2):598–605, February 1968.

[94] T Paz-Elizur, M Takeshita, M Goodman, M O'Donnell, and Z Livneh. Mechanism of translesion DNA synthesis by DNA polymerase II. Comparison to DNA polymerases I and III core. *The Journal of Biological Chemistry*, 271(40):24662–24669, October 1996.

[95] L Postow, C Ullsperger, R W Keller, C Bustamante, A V Vologodskii, and N R Cozzarelli. Positive torsional strain causes the formation of a four-way junction at replication forks. *The Journal of Biological Chemistry*, 276(4):2790–2796, January 2001.

[96] L Pray. Discovery of DNA structure and function: Watson and Crick. *Nature Education*, 1(1):100, 2008.

[97] A E Pritchard, H G Dallmann, B P Glover, and C S McHenry. A novel assembly mechanism for the DNA polymerase III holoenzyme DnaX complex: association of deltadelta' with DnaX(4) forms DnaX(3)deltadelta'. *The EMBO Journal*, 19(23):6536–6545, December 2000.

[98] Z Qiu and M F Goodman. The Escherichia coli polB locus is identical to dinA, the structural gene for DNA polymerase II. Characterization of Pol II purified from a polB mutant. *The Journal of Biological Chemistry*, 272(13):8611–8617, March 1997.

[99] M Radman. SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sciences*, 5A:355–367, 1975.

[100] M Rajagopalan, C Lu, R Woodgate, M O'Donnell, M F Goodman, and H Echols. Activity of the purified mutagenesis proteins UmuC, UmuD', and RecA in replicative bypass of an abasic DNA lesion by DNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22):10777–10781, November 1992.

[101] N B Reuven, G Arad, A Maor-Shoshani, and Z Livneh. The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication. *The Journal of Biological Chemistry*, 274(45):31763–31766, November 1999.

[102] Rodrigo Reyes-Lamothe, David J Sherratt, and Mark C Leake. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science*, 328(5977):498–501, April 2010.

[103] Andrew Robinson, John P McDonald, Victor E A Caldas, Meghna Patel, Elizabeth A Wood, Christiaan M Punter, Harshad Ghodke, Michael M Cox, Roger Woodgate, My-

ron F Goodman, and Antoine M van Oijen. Regulation of Mutagenic DNA Polymerase V Activation in Space and Time. *PLoS Genetics*, 11(8):e1005482, August 2015.

[104] Jeremy M Rock, Ulla F Lang, Michael R Chase, Christopher B Ford, Elias R Gerrick, Richa Gawande, Mireia Coscolla, Sebastien Gagneux, Sarah M Fortune, and Meindert H Lamers. DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. *Nature Genetics*, 47(6):677–681, June 2015.

[105] Christian J Rudolph, Amy L Upton, Anna Stockum, Conrad A Nieduszynski, and Robert G Lloyd. Avoiding chromosome pathology when replication forks collide. *Nature*, 500(7464):608–611, August 2013.

[106] W D Rupp and P Howard-Flanders. Discontinuities in the DNA synthesized in an excision-defective strain of Escherichia coli following ultraviolet irradiation. *Journal of Molecular Biology*, 31(2):291–304, January 1968.

[107] Katharina Schlacher, Kris Leslie, Claire Wyman, Roger Woodgate, Michael M Cox, and Myron F Goodman. DNA polymerase V and RecA protein, a minimal mutasome. *Molecular Cell*, 17(4):561–572, February 2005.

[108] K Sekimizu, D Bramhill, and A Kornberg. ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. *Cell*, 50(2):259–265, July 1987.

[109] H Shinagawa, H Iwasaki, T Kato, and A Nakata. RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 85(6):1806–1810, March 1988.

[110] S C Slater, M R Lifsics, M O'Donnell, and R Maurer. holE, the gene coding for the theta subunit of DNA polymerase III of Escherichia coli: characterization of a holE mutant and comparison with a dnaQ (epsilon-subunit) mutant. *Journal of Bacteriology*, 176(3):815–821, February 1994.

[111] K C Smith and D H Meun. Repair of radiation-induced damage in Escherichia coli. I. Effect of rec mutations on post-replication repair of damage due to ultraviolet radiation. *Journal of Molecular Biology*, 51(3):459–472, August 1970.

[112] S Sommer, F Boudsocq, R Devoret, and A Bailone. Specific RecA amino acid changes affect RecA-UmuD'C interaction. *Molecular Microbiology*, 28(2):281–291, April 1998.

[113] G Steinborn. Uvm mutants of Escherichia coli K12 deficient in UV mutagenesis. I. Isolation of uvm mutants and their phenotypical characterization in DNA repair and mutagenesis. *Molecular & General Genetics*, 165(1):87–93, September 1978.

[114] T A Steitz. A mechanism for all polymerases. *Nature*, 391(6664):231–232, January 1998.

[115] P S Studwell and M O'Donnell. Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme. *The Journal of Biological Chemistry*, 265(2):1171–1178, January 1990.

[116] P T Stukenberg, P S Studwell-Vaughan, and M O'Donnell. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *The Journal of Biological Chemistry*, 266(17):11328–11334, June 1991.

[117] P T Stukenberg, J Turner, and M O'Donnell. An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. *Cell*, 78(5):877–887, September 1994.

[118] W Sun and G N Godson. Binding and phasing of Escherichia coli single-stranded DNA-binding protein by the secondary structure of phage G4 origin of complementary DNA strand synthesis (G4oric). *The Journal of Biological Chemistry*, 268(11):8026–8039, April 1993.

[119] Mark D Sutton. Coordinating DNA polymerase traffic during high and low fidelity synthesis. *Biochimica et Biophysica Acta*, 1804(5):1167–1179, May 2010.

[120] N Suzuki, E Ohashi, K Hayashi, H Ohmori, A P Grollman, and S Shibutani. Translesional synthesis past acetylaminofluorene-derived DNA adducts catalyzed by human DNA polymerase kappa and Escherichia coli DNA polymerase IV. *Biochemistry*, 40(50):15176–15183, December 2001.

[121] M Tang, I Bruck, R Eritja, J Turner, E G Frank, R Woodgate, M O'Donnell, and M F Goodman. Biochemical basis of SOS-induced mutagenesis in Escherichia coli: reconstitution of in vitro lesion bypass dependent on the UmuD'2C mutagenic complex and RecA protein. *Proceedings of the National Academy of Sciences of the United States of America*, 95(17):9755–9760, August 1998.

[122] M Tang, P Pham, X Shen, J S Taylor, M O'Donnell, R Woodgate, and M F Goodman. Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature*, 404(6781):1014–1018, April 2000.

[123] M Tang, X Shen, E G Frank, M O'Donnell, R Woodgate, and M F Goodman. UmuD'(2)C is an error-prone DNA polymerase, Escherichia coli pol V. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16):8919–8924, August 1999.

[124] Nathan A Tanner, Gökhan Tolun, Joseph J Loparo, Slobodan Jergic, Jack D Griffith, Nicholas E Dixon, and Antoine M van Oijen. E. coli DNA replication in the absence of free β clamps. *The EMBO Journal*, 30(9):1830–1840, May 2011.

[125] J Tomizawa and G Selzer. Initiation of DNA synthesis in Escherichia coli. *Annual Review of Biochemistry*, 48(1):999–1034, 1979.

[126] Ana Toste Rêgo, Andrew N Holding, Helen Kent, and Meindert H Lamers. Architecture of the Pol III-clamp-exonuclease complex reveals key roles of the exonuclease subunit in processive DNA synthesis and repair. *The EMBO Journal*, 32(9):1334–1343, May 2013.

[127] K Tougu, H Peng, and K J Marians. Identification of a domain of Escherichia coli primase required for functional interaction with the DnaB helicase at the replication fork. *The Journal of Biological Chemistry*, 269(6):4675–4682, February 1994.

[128] Z Tsuchihashi and A Kornberg. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 87(7):2516–2520, April 1990.

[129] Kaori Uchida, Asako Furukohri, Yutaka Shinozaki, Tetsuya Mori, Daichi Ogawara, Shigehiko Kanaya, Takehiko Nohmi, Hisaji Maki, and Masahiro Akiyama. Overproduction of Escherichia coli DNA polymerase DinB (Pol IV) inhibits replication fork progression and is lethal. *Molecular Microbiology*, 70(3):608–622, November 2008.

[130] J Wagner, S Fujii, P Gruz, T Nohmi, and R P Fuchs. The beta clamp targets DNA polymerase IV to DNA and strongly increases its processivity. *EMBO Reports*, 1(6):484–488, December 2000.

[131] J Wagner, P Gruz, S R Kim, M Yamada, K Matsui, R P Fuchs, and T Nohmi. The dinB gene encodes a novel E. coli DNA polymerase, DNA pol IV, involved in mutagenesis. *Molecular Cell*, 4(2):281–286, August 1999.

[132] J Wagner and T Nohmi. Escherichia coli DNA polymerase IV mutator activity: genetic requirements and mutational specificity. *Journal of Bacteriology*, 182(16):4587–4595, August 2000.

[133] Feng Wang and Wei Yang. Structural insight into translesion synthesis by DNA Pol II. *Cell*, 139(7):1279–1289, December 2009.

[134] J D WATSON and F H CRICK. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, April 1953.

[135] Brian M Wendel, Charmain T Courcelle, and Justin Courcelle. Completion of DNA replication in Escherichia coli. *Proceedings of the National Academy of Sciences*, 111(46):16454–16459, November 2014.

[136] Erika Wolff, Mandy Kim, Kaibin Hu, Hanjing Yang, and Jeffrey H Miller. Polymerases leave fingerprints: analysis of the mutational spectrum in Escherichia coli rpoB to assess the role of polymerase IV in spontaneous mutation. *Journal of Bacteriology*, 186(9):2900–2905, May 2004.

[137] C A Wu, E L Zechner, J A Reems, C S McHenry, and K J Marians. Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. V. Primase action regulates the cycle of Okazaki fragment synthesis. *The Journal of Biological Chemistry*, 267(6):4074–4083, February 1992.

[138] Wei Yang. Damage repair DNA polymerases Y. *Current Opinion in Structural Biology*, 13(1):23–30, February 2003.

[139] N Yao, J Turner, Z Kelman, P T Stukenberg, F Dean, D Shechter, Z Q Pan, J Hurwitz, and M O'Donnell. Clamp loading, unloading and intrinsic stability of the PCNA, beta and gp45 sliding clamps of human, E. coli and T4 replicases. *Genes to Cells*, 1(1):101–113, January 1996.

[140] Nina Y Yao, Roxana E Georgescu, Jeff Finkelstein, and Michael E O'Donnell. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proceedings of the National Academy of Sciences*, 106(32):13236–13241, August 2009.

[141] Joseph T P Yeeles and Kenneth J Marians. The Escherichia coli replisome is inherently DNA damage tolerant. *Science*, 334(6053):235–238, October 2011.

[142] A Yuzhakov, Z Kelman, and M O'Donnell. Trading places on DNA–a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell*, 96(1):153–163, January 1999.

Chapter 2

Development of a single-molecule method to study DNA replication

2.1 Introduction

The first single-molecule experiments were performed on ion channels using patch-clamping in 1976[18]. Since then, a range of single-molecule techniques have been developed that have contributed to our understanding of protein folding[14 19], molecular motors[27 28], gene expression[12 25] and other important biological processes. The existing single-molecule methods can be broadly divided into two categories: force-based and fluorescence-based, and these will be discussed in more detail below.

Single-molecule techniques have generated much excitement in the molecular biology field because they have shifted the way we understand biological processes: instead of thinking in terms of the average of moles of molecules, we can now understand processes at the level of single molecules. This is important for understanding multi-step enzymatic reactions for example, because not all molecules will react at the same time and due to the stochastic nature of molecular interactions, individual events will rapidly become asynchronous. In an ensemble experiment, these asynchronous events will be averaged, leading to the blurring of the sequential steps in the enzymatic process and preventing the discrimination between the steps.

In addition, single-molecule methods reveal the heterogeneity in the behaviour of molecules and show a distribution of behavioural states rather than a single mean state that results from ensemble experiments. As described in Kapanidis and Strick, 2009[13], a given enzyme in two conformational states will result in a single value that represents the average of the two conformations in an ensemble measurement, whereas single-molecule measurements will reveal the

individual states and the transition between the states (Fig. 2.1).

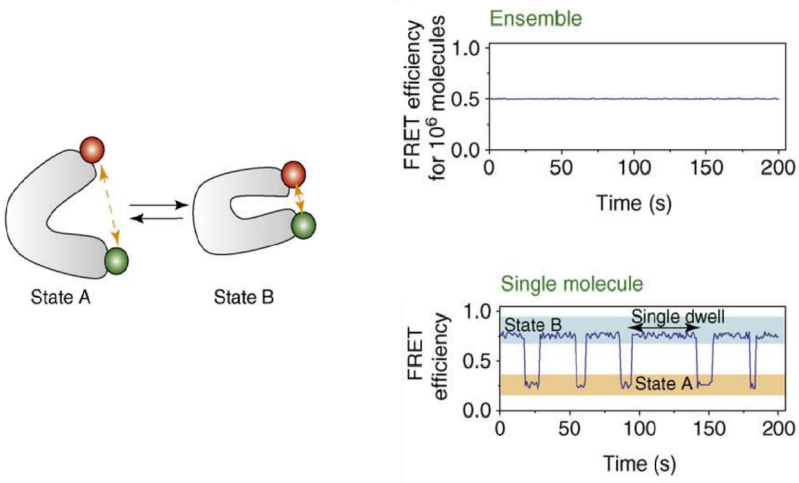


Figure 2.1. Single-molecule methods reveal underlying protein dynamics. The two conformational states of an enzyme is averaged out in an ensemble experiment, whereas single-molecule methods reveal the transition between the two states. Figure taken from Kapanidis and Strick, 2009[13].

Measuring events at the single-molecule level, however, means that careful statistical analysis of many individual events should be carried out in order to obtain a complete understanding of the underlying process. This may seem counter-intuitive as the measurement of a large number of molecules seems to approximate an ensemble experiment. However, the difference is that an ensemble measurement is a snapshot of all the states that a large number of molecules are in at any given time, while single-molecule studies can capture each individual state. However, obtaining a sufficient number of events for the interpretation of single-molecule statistics can be time-consuming.

The *E. coli* replisome is composed of over 10 different proteins that act in concert to catalyze multi-step enzymatic reactions. We want to understand the molecular details of how the different steps are temporally organized and how the transition between the different steps occurs, which are questions that can be uniquely addressed using single-molecule methods. However, to set up a method that can answer the specific questions we are interested in, we start by exploring the main single-molecule methods used to study protein-DNA interactions.

2.1.1 Single-molecule methods for studying protein-DNA interactions

Protein dynamics on DNA have been studied using both force-based and fluorescence-based techniques, each with their advantages and shortcomings. The main advantage of the force-

based methods, which include optical and magnetic tweezers, is the ability to detect physical changes in the DNA, for example when ssDNA is converted into dsDNA. However, the activity of proteins on DNA must be inferred indirectly from changes in the DNA tension, which makes it difficult to discriminate between multiple proteins as it is not possible to directly visualize which protein is bound. In addition, if the protein activity does not result in a significant change in the DNA tension, for example if the protein merely binds to the DNA, this may not be possible to characterize using the force-based methods.

The fluorescence-based methods such as single-molecule FRET (smFRET) and multi-color co-localization single-molecule spectroscopy (CoSMoS) have the advantage of directly visualizing the DNA and protein molecules and their physical locations. Fluorescence can also be used to probe structural dynamics in a single protein or the physical proximity of two molecules. The main limitation of the fluorescence methods is that a binding event does not always correlate with enzymatic activity. Since changes in DNA cannot be monitored, it is not always possible to distinguish random from productive binding events, for example whether or not a DNA nicking enzyme has introduced nicks in its substrate.

The attributes of each of these techniques are summarized in Table 1 and discussed below. A new family of methods that combine the force and fluorescence-based techniques is emerging to combine the advantages and overcome the limitations in the force-based and fluorescence-based single-molecule methods.

Optical and magnetic tweezers

In an optical tweezers experiment, the DNA is attached via a single or both ends to microspheres that are held in place in the focus of a laser beam (Fig. 2.2a). The force exerted on the DNA can be varied by adjusting the distance of the microspheres and vice versa, a change in the distance of the microspheres reflects the force experienced by the DNA. Optical tweezers can achieve extremely high spatial resolution at the angstrom scale[9], which is not yet possible using the other force-based methods.

One application of the optical tweezers is to study protein translocation on DNA, where the movement of a protein exerts a force on the DNA that is detected through a change in the distance between the microspheres[1]. Another application is to study how protein result in physical changes in the DNA, for example when a polymerase converts ssDNA to dsDNA. Given that ssDNA and dsDNA have different elastic properties, conversion between the two changes the force-distance relationship of a laser tweezers experiment[26].

Table 1. Summary of the single-molecule methods.

Attributes	Force-based		Fluorescence-based			
	Optical tweezers	**Magnetic tweezers**	**Single-molecule FRET (smFRET)**	**Protein-induced fluorescence enhancement (PIFE)**	**Co-localization single-molecule spectroscopy (CoSMoS)**	**DNA curtains**
DNA attachment	Microspheres in the focus of a laser beam	Magnetic beads	Surface-immobilization	Surface-immobilization	Surface-immobilization	Surface-immobilization
Detection	Bead position	Bead position	FRET signal	PIFE signal	Fluorescence	Fluorescence
Main advantages	High spatial resolution of 1 Å; detection of physical changes in DNA	High force sensitivity of 0.01 pN; detection of physical changes in DNA; application of torque on DNA; multiple DNA molecules studied at a time	Intra-molecular dynamics; physical proximity of binding partners; multiple DNA molecules studied at a time	Protein binding without the need to label protein; multiple DNA molecules studied at a time	Multiple laser channels for the direct visualization of molecules; multiple DNA molecules studied at a time; freedom in labeling position choice	Direct visualization of molecules; movement of proteins on DNA; discrimination between ssDNA versus dsDNA; multiple DNA molecules studied at a time
Main disadvantages	Cannot be used to study interactions that do not result in changes in DNA tension; manipulation of single DNA molecule at a time	Cannot be used to study interactions that do not result in changes in DNA tension	Detection of interactions only in the physical distance of the FRET signal; photo-bleaching	Detection of interactions only in the immediately vicinity of fluorophore on DNA; photo-bleaching; cannot be used to study physical changes in DNA	Lack of distance information and intra-molecular dynamics; cannot be used to study physical changes in DNA; photo-bleaching	Cannot be used to study intra-molecular dynamics; photo-bleaching
Examples of applications	Molecular motors on DNA (Abbondanzieri *et al*, 2005); DNA synthesis (Wuite *et al*, 2000)	DNA uncoiling (Strick *et al*, 2000)	Conformational changes in DNA upon protein binding (Taylor *et al*, 2015); protein movement on DNA (Senavirathne *et al*, 2015)	Protein binding (Markiewicz *et al*, 2012)	Step-wise assembly of factors in multi-enzymatic processes (Hoskins *et al*, 2011)	Diffusion of proteins on DNA during target search (Wang *et al*, 2013)

In the magnetic tweezers setup, the DNA is surface-immobilized on one end and the other end is vertically attached to a magnetic bead that is controlled using a pair of magnets (Fig. 2.2b). The magnets exert a constant force on the bead which can be adjusted by varying the distance of the bead from the surface[2]. Similar experiments previously described for the optical tweezers can also be performed using magnetic tweezers, although this method has not yet achieved the same spatial sensitivity as the optical tweezers. However, the advantage of magnetic tweezers is that they have greater force sensitivity and can be used to measure forces with 0.01 pN accuracy[13]. It is also easier to manipulate the DNA using magnetic tweezers, for example by applying torque and studying the effect of super-coiling on protein-DNA interactions. Finally, multiple DNA molecules can be measured simultaneously, which increases the

experimental throughput of the magnetic tweezers method compared to the optical tweezers that typically handle a single DNA molecule at a time. As an example, magnetic tweezers have been used to study the uncoiling of DNA[21].

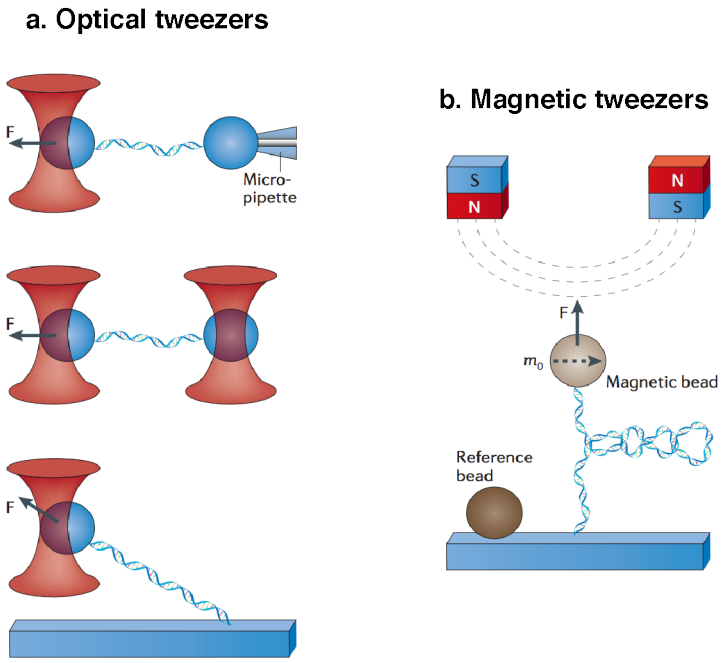


Figure 2.2. Optical and magnetic tweezers setup. **a**, Optical tweezers showing a DNA molecule attached via single or both ends to microspheres in the focus of laser beams. **b**, Magnetic tweezers showing a vertically stretched DNA that is attached to the surface via one end and to a magnetic bead via the other end. Figure taken from Dulin *et al*, 2013[2].

Single-molecule FRET and PIFE

In a single-molecule fluorescence resonance energy transfer (FRET) experiment, the DNA substrate and protein of interest are fluorescently labeled using a FRET fluorophore pair. The DNA molecules are immobilized via one end on a glass surface and the protein-DNA interactions are visualized by total internal reflection fluorescence (TIRF) microscopy. TIRF creates an evanescent wave that excites a thin layer of approximately 100 nm above the glass surface and allows the protein-DNA interactions to be followed over hundreds of DNA molecules.

Due to the distance-sensitivity of FRET at nanometer resolution, the FRET signal increases when the donor and acceptor are in close proximity and decreases when they are far apart. Therefore, the binding of a protein to DNA can only be detected by smFRET if the binding event brings the donor and acceptor fluorophores into close proximity (Fig. 2.3a). Intra-molecular

dynamics can be probed by placing both the donor and acceptor fluorophores on the DNA or protein. For example, this has been used to monitor conformational changes in the DNA when proteins bind to it[22]. Furthermore, the position of a protein on DNA can be measured when the donor and acceptor fluorophores are placed separately on the DNA and protein. As the protein moves along the DNA, the FRET readout can be used to give distance information[20].

While single-molecule FRET allows protein binding events and intramolecular dynamics to be studied, the caveat with this technique is that it does not allow the characterization of events beyond the FRET distance. Therefore, there is a strict requirement for the positioning of the fluorophores when studying processes by smFRET. The FRET signal can also be affected by factors besides distance, for example the physical orientation and rotational freedom of the fluorophore, which means that a FRET signal needs to be carefully interpreted.

The binding of a protein to DNA can also be measured using protein-induced fluorescence enhancement (PIFE) (Fig. 2.3b), a phenomenon where the close physical proximity of a protein to the fluorophore on the DNA enhances its fluorescence due to an induced cis-trans isomerization of a double bond present in the fluorophore[17]. The advantage of PIFE over FRET is that there is no need to fluorescently label the protein. This therefore avoids any effect caused by photo-bleaching of the protein signal on data analysis. However, the drawback is the loss of FRET signal that gives more sensitive distance information.

Co-localization single-molecule spectroscopy and DNA curtains

Another class of fluorescence-based single-molecule methods relies on the direct visualization of the interacting DNA and proteins through fluorescent labeling. Here, multiple proteins on DNA can be followed and the number of fluorescent factors that can be imaged in an experiment is only limited by the spectrally separate fluorophores and laser channels available. Therefore, an advantage of this type of fluorescence-based techniques over smFRET is an increase in the number of usable fluorophores beyond the FRET fluorophore pair. The lack of requirement for FRET in these methods also enables the studying of protein binding and processes happening at a distance beyond those required for energy transfer. Hence, there is a greater choice in the position of fluorophores when labeling molecules for such an experiment. However, the loss of the FRET signal also removes useful distance information, for example the conformational dynamics of a protein and its position on DNA.

In co-localization single-molecule spectroscopy (CoSMoS), fluorescently labeled DNA substrates are immobilized on a surface and the co-localization between fluorescently labeled pro-

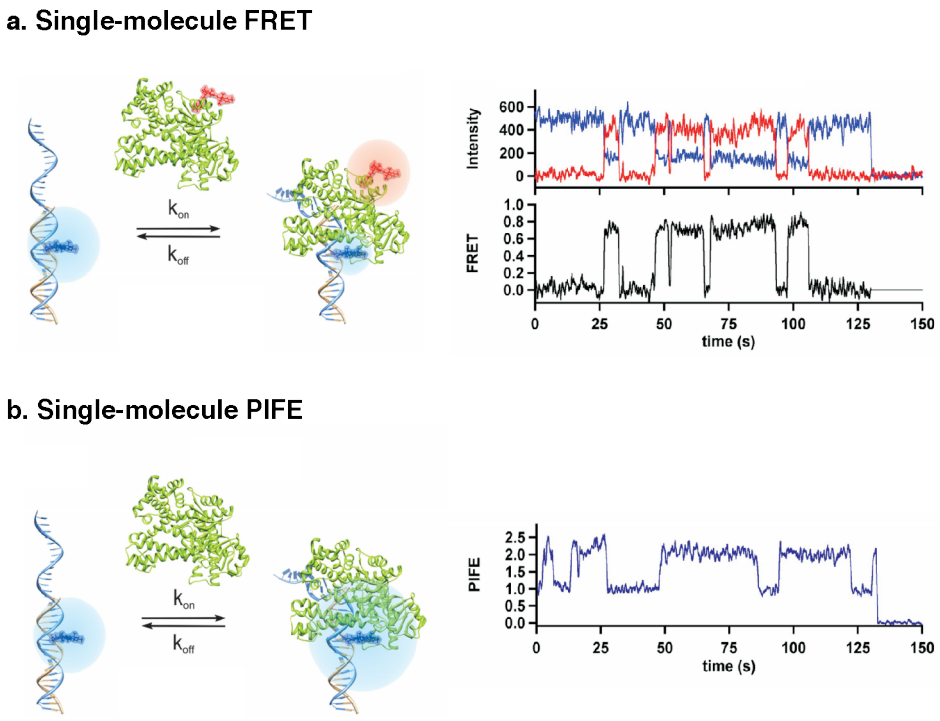


Figure 2.3. Single-molecule FRET and PIFE. a, The transfer of energy between a donor and an acceptor fluorophore can be used to detect binding events where the fluorophores are brought into close physical proximity. **b**, The binding of a protein to the vicinity of a fluorophore can lead to an enhancement of its fluoresence. Figure taken from Markiewicz *et al*, 2012[17].

teins and the DNA is followed by TIRF microscopy (Fig. 2.4a)[3]. The setup is similar to that for smFRET, the difference being an increase in the number of laser channels used. When the DNA is vertically immobilized on the surface, the specific location of the proteins on DNA cannot be resolved as there is no spatial information in the vertical plane of the TIRF field. Instead, information is obtained for the physical co-localization between proteins and DNA, including the order in which the proteins assemble and their lifetimes on the DNA[10 4 23].

The DNA curtains method also allows direct visualization of the proteins and DNA in real-time. However in contrast to CoSMoS, the DNA is surface-immobilized and then laterally stretched under hydrodynamic flow to create parallel arrays consisting of hundreds of DNA molecules (Fig. 2.4b)[7]. The advantage of this setup is that it allows the position of proteins on DNA to be followed in a high-throughput fashion, for example the three-dimensional diffusion of proteins during target search[24]. In addition, the incorporation of fluorophores that bind preferentially to ssDNA or dsDNA can be used to distinguish the physical state of the DNA molecule[6].

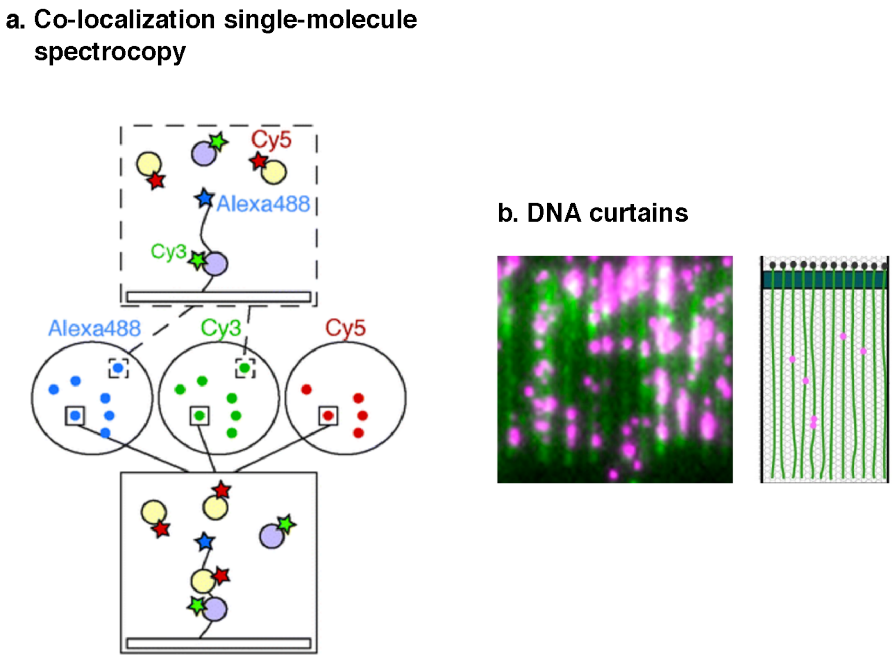


Figure 2.4. Colocalization single-molecule spectroscopy and DNA curtains. a, The interaction of multiple proteins on DNA is observed as the co-localization between the fluorescence signals. Figure taken from Larson *et al*, 2014[16]. **b**, Proteins (magenta) binding and movement on DNA (green), as well as changes to DNA length and structure, can be visualized on laterally stretched DNA curtains. Image by Dr Eric Greene, Columbia University.

2.1.2 Aims

Given that we want to study multiple protein factors in the *E. coli* replisome, we need a way to unambiguously identify the proteins and follow their binding to DNA. For these purposes, we have chosen to employ a fluorescence-based single-molecule technique, and specifically CoSMoS, to allow multiple proteins to be followed at the same time.

2.2 Materials and Methods

Since this chapter is on method development, all materials are specified as they are mentioned. More detailed experimental protocols are described in the Materials and Methods sections of Chapters 3 and 4.

2.3 Results

2.3.1 Flow cell

In order to image single molecules in the $\sim$ 100 nm TIRF evanescence field for CoSMoS, we required a small volume sample cell.

We started by creating a sandwich between a 22 mm $\times$ 22 mm and a 18 mm $\times$ 18 mm thickness 1 glass coverslip (VWR) using thin strips of parafilm (Fig. 2.5a). The sandwich was sealed by heating it to 150 °C, at which the melted parafilm glued the two coverslips together. Individual flow cells were separated from each other by the parafilm strips. While parafilm was effective at sealing the individual channels, imaging the channel showed substantial fluorescence in the parafilm, which was leaked into the flow cells during washing of the flow cell (data not shown). In addition, the dimensions of the coverslips were incompatible with the TIRF microscope stage, which meant that it required a special holder to be made in order to secure it on the stage.

Next, we created a flow cell between a 76 mm $\times$ 26 mm $\times$ 1.0-1.2 mm clear glass microscope slide (Thermo Scientific Menzel-Glaser) and a 22 mm $\times$ 22 mm, thickness 1 glass coverslip (VWR) using strips of transparent double-sided sticky tape (Tesa) (Fig. 2.5b). The volume of the flow cell was variable as this was dependent on the length of the cut tape strips and the width of the channel created manually between the strips but was typically around 10-15 μl. While the double-sided sticky tape was a clear improvement in reducing the background fluorescence, this flow cell was open-ended, which meant that the reaction could evaporate from the two ends. In addition, since the TIRF microscope is an objective-based microscope, the flow cell needs to be mounted onto the stage with the coverslip facing down. Therefore, any new additions into the flow cell requires the whole flow cell to be removed from the stage and placed back after the injections. This is a problem because co-localization experiments rely on the precise location of molecules and therefore any movement of the flow cell should be avoided.

To improve the reproducibility of the flow cells, we next sought to create a mould for cutting the double-sided tape so that the volume of the flow cell would stay constant. In order to do this, we purchased a Silhouette Portrait electronic cutting tool and 0.12 mm thick SecureSeal Adhesive Sheet from Grace Bio-labs, which enabled us to create a mould in a highly reproducible and fast manner. To further improve the reproducibility between experiments, we designed the mould to contain three channels so that the same surface-treated coverslip could be used for three experiments (Fig. 2.5c). This reduces any variation between the individual

coverslips when a direct comparison needs to be made between different experimental conditions. Furthermore, to allow easier manipulation of the experiment, we drilled holes in the glass slide at each end of the flow cell that were just big enough to use a pipette tip for injecting new reactions into the flow cell directly on the microscope stage. This type of flow cell was used for the experiments described in this Thesis.

In addition, we attempted to make microfluidic flow cells by inserting tubing into the drilled holes in the triple channels and securing the tubing into place using epoxy glue (5 minute Epoxy, Devcon Home) (Fig. 2.5d). We then connected the free end of one tubing to a syringe on a Pump11 Elite micro-pump and the free end of the other tubing to an Eppendorf tube containing the solution to be introduced into the flow cell. We then used the syringe to withdraw the reaction from the Eppendorf tube into the flow cell at speeds of 50-100 μl/min. Solution was withdrawn into the flow cell rather than injected into it to avoid loss of material to the syringe walls. The addition of tubing proved more difficult than anticipated: if the tubing was inserted immediately after the assembly of the flow cells, i.e. when the flow cell was empty, the final steps of surface passivation (see section 2.3.2 Surface chemistry) were time-consuming due to the slow flow rate. If the tubing was connected after all the surface passivation was done, i.e. the flow cell contained solution, it was very easy to introduce air bubbles into the system, which frequently rendered the flow cell unusable. On the infrequent occasion that a successful microfluidic flow cell was constructed, we also observed inconsistency in the amount of labeled molecules injected into the flow cell, despite using the same concentrations and flow rates. However, this may be simply due to nonspecific binding of labeled DNA and proteins to the tubing and could be easily resolvable by passivating the tubing with BSA.

Since direct manual injection of reaction into the flow cell suffices for our experiments, we did not pursue the construction of the microfluidic flow cell further. However, this would be of interest for the future as having an automated system, especially with a switchable valve to control the different inlets, can help improve the reproducibility and reduce the labour intensity in a single-molecule experiment.

2.3.2 Surface chemistry

Next we sought a way to immobilize DNA so that the protein-DNA interactions could be visualized at the surface using TIRF microscopy. Since biotinylation is a common DNA modification and biotin binds to streptavidin with high affinity, most surface chemistries utilize ways to capture the biotin on DNA. Apart from attaching the DNA, the glass surface must also reduce any

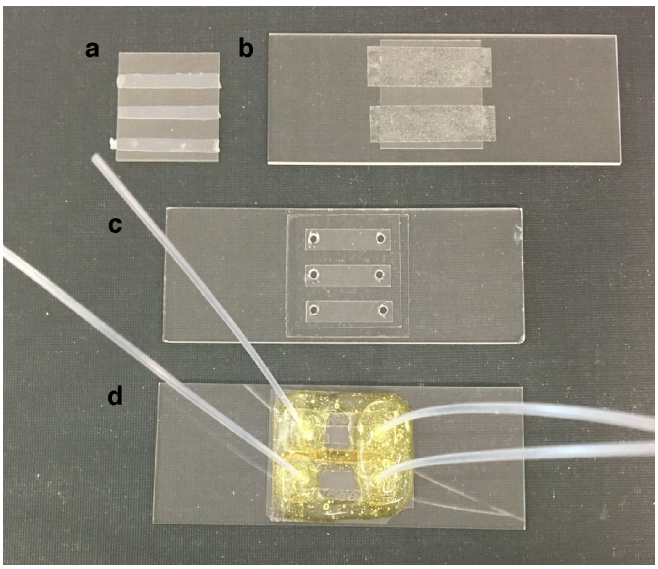


Figure 2.5. Various designs of flow cells. **a**, Dual flow cells created between two glass coverslips using molten parafilm. **b**, Single flow cell created between a glass microscope slide and a glass coverslip using double-sided tape. **c**, Triple flow cells created between a glass microscope slide and a glass coverslip using mould cut out of double-sided adhesive sheet. **d**, Tubing attached to the flow cell type described in c.

non-specific binding of proteins.

We first tested different dilutions of anti-biotin antibodies (Abcam) that non-specifically attach to the glass surface and mediate direct capture of biotin-DNA. We further passivated the surface using a layer of nonionic surfactant polyol called Pluronic F127 (Sigma). The advantage of this type of surface chemistry is its simplicity as a single antibody layer is sufficient to attach the bionylated DNA. However, doing some preliminary tests with this type of surface chemistry showed a significant level of non-specific protein binding to the surface (data not shown).

A more widely used surface chemistry for the attachment of DNA is using a polyether compound called polyethylene glycols (PEG)[8], which forms a dense layer of polymer brush at the surface. The PEG can be functionalized and a common modification is PEG-biotin. When a mixture of PEG and PEG-biotin is used to functionalize the glass surface, DNA-biotin can be attached via streptavidin to the PEG-biotin. We purchased HO-PEG-NH_2 and biotin-CONH-PEG-NH_2 from Rapp Polymere and mixed the two compounds in a 5:1 ratio of PEG to PEG-biotin. To attach the PEG onto glass, the glass needs to have the correct surface chemistry. We did this through silanizing the glass using (3-Glycidyloxypropyl)trimethoxysilane (GOPTS) (Sigma). GOPTS contains an epoxy group that reacts with the -NH_2 group of the PEG molecule and thereby forms a covalent linkage. GOPTS attaches to glass cleaned and oxidized using 3M NaOH and Piranha solution, respectively. Piranha solution a mixture of hydrogen peroxide and

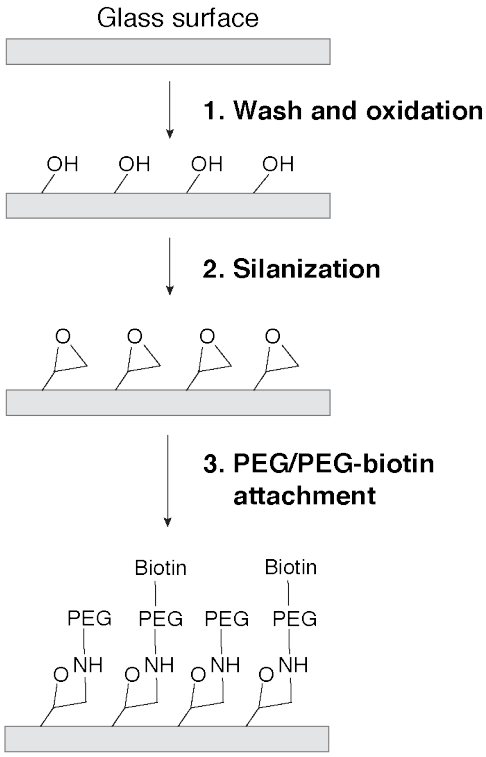


Figure 2.6. Surface chemistry. The glass surface was treated using a multi-step protocol: the surface was first washed and oxidized, followed by silanization and attachment of a layer of PEG and PEG-biotin.

concentrated sulfuric acid (Fig. 2.6).

Imaging using the PEG surface still resulted in some non-specific binding of proteins (data not shown) and therefore once the flow cell was assembled, the cleaned counter-surface and the reaction surface were further passivated using a layer of Poly(l-lysine)-g-poly(ethylene glycol) (SuSoS) which is PEG attached to a poly(L-lysine) backbone. The backbone interacts directly with the oxidized glass surface through electrostatic interactions. Furthermore, a layer of Pluronic F127 (Sigma) and bovine serum albumin (BSA, New England Biolabs) were added for further passivation of the glass surface. Lastly, streptavidin was introduced to bridge the biotin on the PEG layer and the biotin on the DNA substrate.

While this multi-step surface treatment protocol is labor-intensive, it resulted in a well-passivated surface with very little non-specific protein binding to the surface.

2.3.3 Fluorophores

Fluorophores with high photostability and long fluorescence lifetime are required in order to image protein-DNA interactions for long enough to capture the events of interest. Fluorescence lifetime refers to how long the fluorophore can be imaged before it is bleached, while photostability measures the blinking of a fluorophore when it is excited with a laser. Fluorophores with high photostability eliminates ambiguity in distinguishing between fluorophore blinking and the dissociation and rapid rebinding of a protein since both can be seen as a fast disappearance and re-appearance of the fluorescence signal. Our choice of using organic fluorophores instead of quantum dots for example, took into consideration the much larger size of quantum dots which may interfere with protein dynamics and protein-DNA interactions.

We began by testing the frequently used cyanine dyes (Cy3 and Cy5). To measure the fluorescence lifetime of the dyes and visualize two truly co-localizing molecules, we placed Cy3 on the template (Cy3 NHS ester from GE Healthcare; see section 2.3.4 Labeling strategies) and Cy5 on the primer (directly ordered from IDT) and then annealed the two DNA strands to form dsDNA (Fig. 2.7a). We then imaged the DNA substrate using the 561 nm and 638 nm lasers but to our surprise, there was little signal in the Cy3 channel and no co-localization with Cy5 signals was observed. This could be explained by the long wavelength of the laser used to excite the Cy3 dye, which is normally excited using a 532 nm laser. The laser wavelength of 561 nm is greater than the maximum excitation wavelength for Cy3 (554 nm, ThermoFisher Scientific), at which the fluorophore absorbs the most efficiently and after which the absorption efficiency decreases sharply. Therefore, the lack of signal may be accounted for by the low excitation efficiency (Fig. 2.7a). Furthermore, we observed that the Cy5 dye bleached away rapidly (Fig. 2.8).

Given that the cyanine dyes are unsuitable for the available lasers and microscope setup, we switched to other fluorophores. We replaced the Cy5 dye on the primer DNA to Atto 647N and imaged the Cy3/Atto 647N DNA duplex (Fig. 2.7b). The Atto 647N fluorophore gave bright, well-defined spots that had a significantly longer fluorescence lifetime than Cy5 under the same imaging conditions (Fig. 2.8). Next, we searched for fluorophores that are suitable for imaging using the 561 nm and 488 nm lasers. This was done by labeling the template DNA with Atto 647N and varying the fluorophore on the primer strand to Atto 565 (Fig. 2.7c), or Atto 550, Alexa 488, Atto 488 (Fig. 2.7d, data not shown for Atto 550 and Alexa 488). Given the higher stability and brighter signal observed for the Atto dyes, we decided to use Atto 488, Atto 565 and Atto 647N in our future CoSMoS experiments.

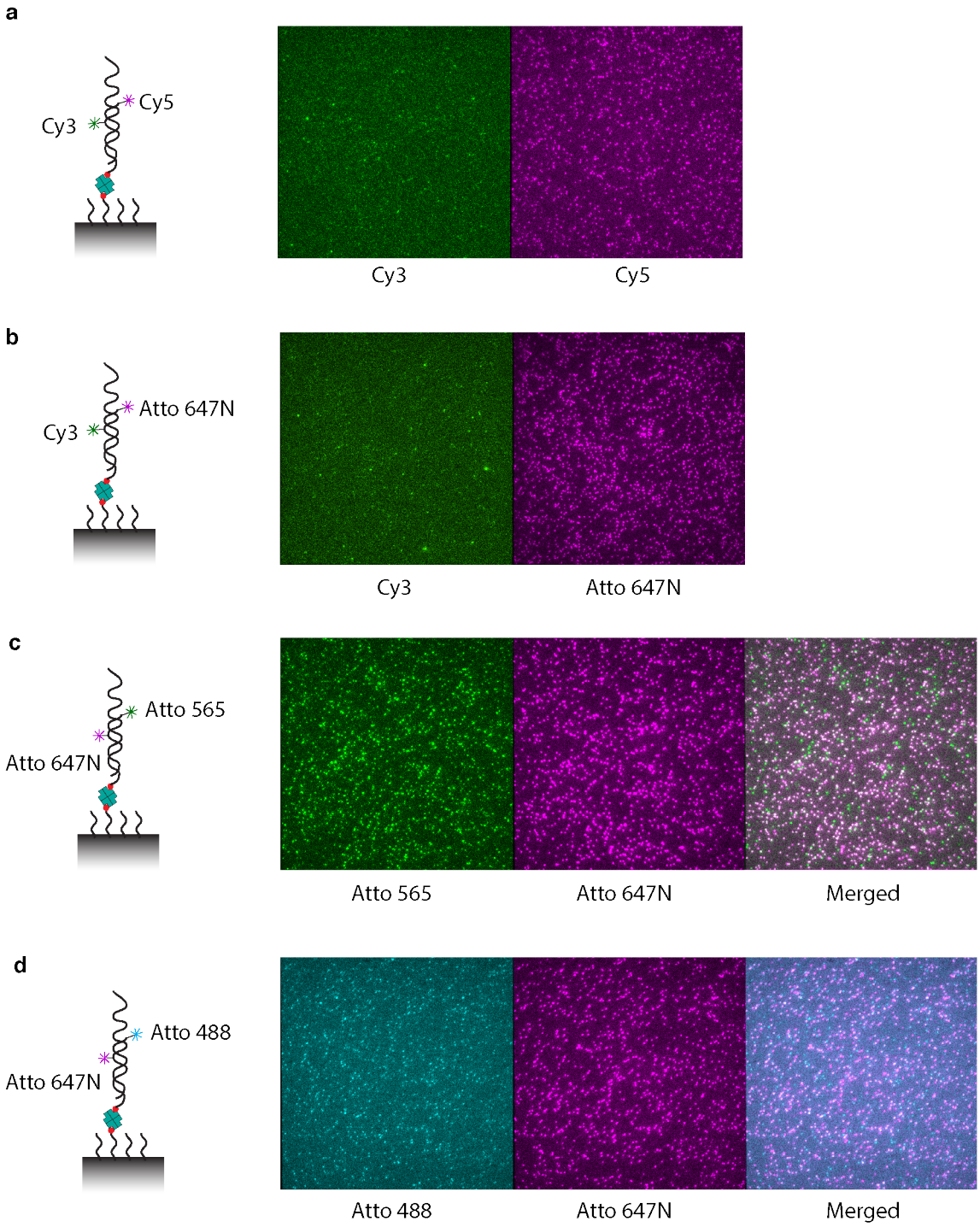


Figure 2.7. Atto fluorophores give bright signals. **a**, Low signal seen for the Cy3-Cy5 fluorophore pair. **b**, Replacing Cy5 with Atto 647N resulted in bright spots. **c**, The Atto 565-Atto 647N fluorophore pair gave bright signals and co-localization could be seen as white spots in the merged image. **d**, The Atto 488-Atto 647N fluorophore pair gave bright signals and co-localization could be seen as white spots in the merged image.

The search for a fluorophore that could be excited using the fourth laser of 405 nm was more problematic. We labeled the template DNA using the Atto 425 fluorophore and excited

it using the 405 nm laser. However, the signal on the DNA could also be observed in the 488 nm channel, suggesting that the two channels were not spectrally well-separated. In addition, considering that the 405 nm laser is an UV laser, excitation with such a short wavelength may result in mutations in the DNA substrates that could affect the protein-DNA interactions. For these reasons, this laser channel was not considered further for the co-localization experiments.

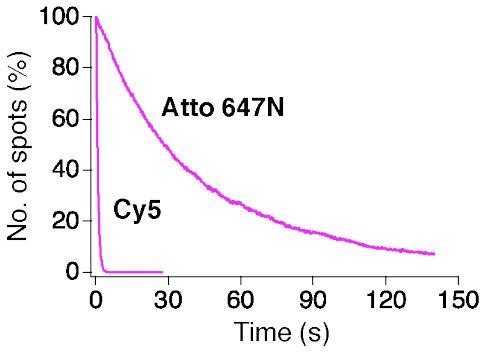


Figure 2.8. Atto 647N has a significantly longer fluorescence lifetime compared to Cy5. The Atto 647N and Cy5 fluorophores were imaged under the same conditions. The number of fluorescent spots were counted over the duration of the experiment.

Since it has been observed by us and others (Jeff Gelles, Brandeis University) that the lifetime of a fluorophore differs greatly when placed on DNA or protein, we next determined the fluorescence lifetime of the Atto dyes on proteins. This was done by labeling the β clamp protein of the *E. coli* Pol III holoenzyme, loading the protein onto DNA (see Chapter 3), and measuring the fluorescence lifetime of the fluorophores under the same imaging conditions as those used for the experiments presented in this Thesis. The β clamp was chosen for labeling because it is known to be highly stable once loaded onto DNA (see Chapter 3), and also due to the fact that most of the protein binding events studied in this Thesis occur on the β clamp. The determination of the fluorescence lifetimes of the Atto fluorophores ensures that the lifetime of the protein-DNA interactions we study is not limited by the photo-bleaching of the fluorophores. The fluorescence lifetimes of Atto 488, Atto 565 and Atto 647N on the β clamp on DNA are shown in Fig. 2.9.

2.3.4 Labeling strategies

We wanted to achieve site-specific labeling of the DNA and proteins so that the number of interacting molecules at any given time can be quantified. For protein labeling, we used two different labeling strategies, depending on the number of cysteine residues present in the protein

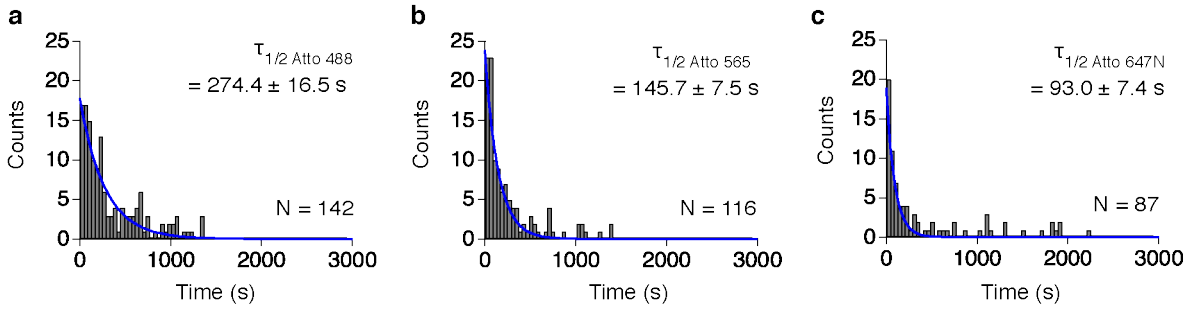


Figure 2.9. Fluorescence lifetimes of Atto 488, Atto 565 and Atto 647N on β clamp loaded onto DNA. The fluorescence lifetimes of **a**, Atto 488 **b**, Atto 565 and **c**, Atto 647N were determined under the same imaging conditions as those used for the experiments presented in this Thesis. Histograms show the distribution of lifetimes of the fluorophores on the β clamp and the values represent the mean lifetime $\pm$ s.e.m.

of interest. The labeling of the proteins presented in this Thesis is described in more detail in Chapters 3 and 4.

For proteins with one or two surface-exposed cysteine residues, we created a single labeling site using mutagenesis to remove the other cysteine residue. Maleimide-linked dyes react almost exclusively with the thiol groups in cysteine residues in the pH range 7.0-7.5 (Sigma). At higher pH, maleimide can also interact with amines in the protein. Therefore to specifically label the cysteine residue in the proteins, we conducted our labeling reaction at pH 7.5. Since reducing agents such as dithiothreitol (DTT) contain thiol groups, any DTT must be buffer exchanged out of the reaction prior to the introduction of the maleimide-dye. In fact, excess DTT is used to quench the labeling at the end of the reaction.

In the case where the protein of interest contains multiple cysteine residues, replacing all but one cysteine residue would involve multiple rounds of mutagenesis that could affect the protein function. Therefore in these cases, we directly added a Ybbr sequence DSLEFIASKLA[30] to the N-terminus of the protein. This sequence is specifically recognized by the *Bacillus subtilis* Sfp phosphopantetheinyl transferase enzyme, which conjugates any coenzyme A (CoA)-linked moiety to the protein. Therefore, when an Atto dye is covalently linked to CoA by reacting the maleimide functionalized Atto dye to CoA, the resulting CoA-linked dye can be enzymatically conjugated to the protein of interest[29].

Comparing these two protein labeling strategies, we consistently achieved 60-70 % labeling efficiency using the Sfp enzymatic method in contrast to the higher efficiencies of 80-100 % obtained by the direct labeling of cysteines using the maleimide dyes (see Chapters 3 and 4).

The lower dye incorporation may be due to the incomplete separation of the CoA-dye conjugate from free CoA, which would lead to the enzymatic incorporation of CoA to the protein of interest, thus making the site unavailable for reaction with CoA-dye. In contrast, labeling proteins with maleimide-dyes is a direct method, without any need to purify the reacting dye specifies and is therefore only dependent on the accessibility of the cysteine residue to the maleimide dye.

After labeling the proteins, we performed control experiments to show that the labeling did not affect the function of the proteins. We used a ϕX174 phage DNA (5.4 kb) primer extension assay to compare the activity of the unlabeled and labeled proteins. The individual proteins will be discussed in more detail in Chapters 3 and 4, but as shown in Fig. 2.10, the labeled proteins retained wild-type activity.

The primer-template DNA substrate was labeled internally on the template DNA strand using NHS ester-activated dyes that react with primary amines to form amide bonds. Therefore, the template DNA strand was internally amino-modified at a thymine base, and this was then reacted with a NHS ester-linked Atto dye. We typically achieved $\sim 70\ \%$ labeling efficiency on the DNA.

2.3.5 DNA substrates

A central component in the *E. coli* replisome is the β sliding clamp. Initial experiments of adding the isolated clamp to a primer-template DNA substrate with a free 5' end on the template DNA showed that the clamp is capable of threading and unthreading itself at the free end (Fig. 2.11b). Therefore, we biotinylated the free 5' end and bound it to streptavidin. Streptavidin is a tetrameric protein where each monomer is capable of binding to biotin. To avoid forming large streptavidin-DNA complexes, we purified a monovalent streptavidin that has only one monomer active for biotin binding (Fig. 2.11a). The DNA substrate end-blocked using monovalent streptavidin prevented the self-threading of clamps (Fig. 2.11c).

The monovalent streptavidin was made by purifying two different forms of streptavidin: streptavidin "alive" that binds biotin and streptavidin "dead" that no longer binds biotin due to three point mutations that abolish its biotin-binding activity[11]. In addition, the streptavidin "alive" contains a polyglutamate tail consisting of six glutamate residues at the C-terminus of the protein, which can be used to separate the different ratios of streptavidin "alive" and streptavidin "dead" subunits by ion exchange chromatography (Fig. 2.11a).

The fluorescently labeled template DNA was bound to the monovalent streptavidin and then

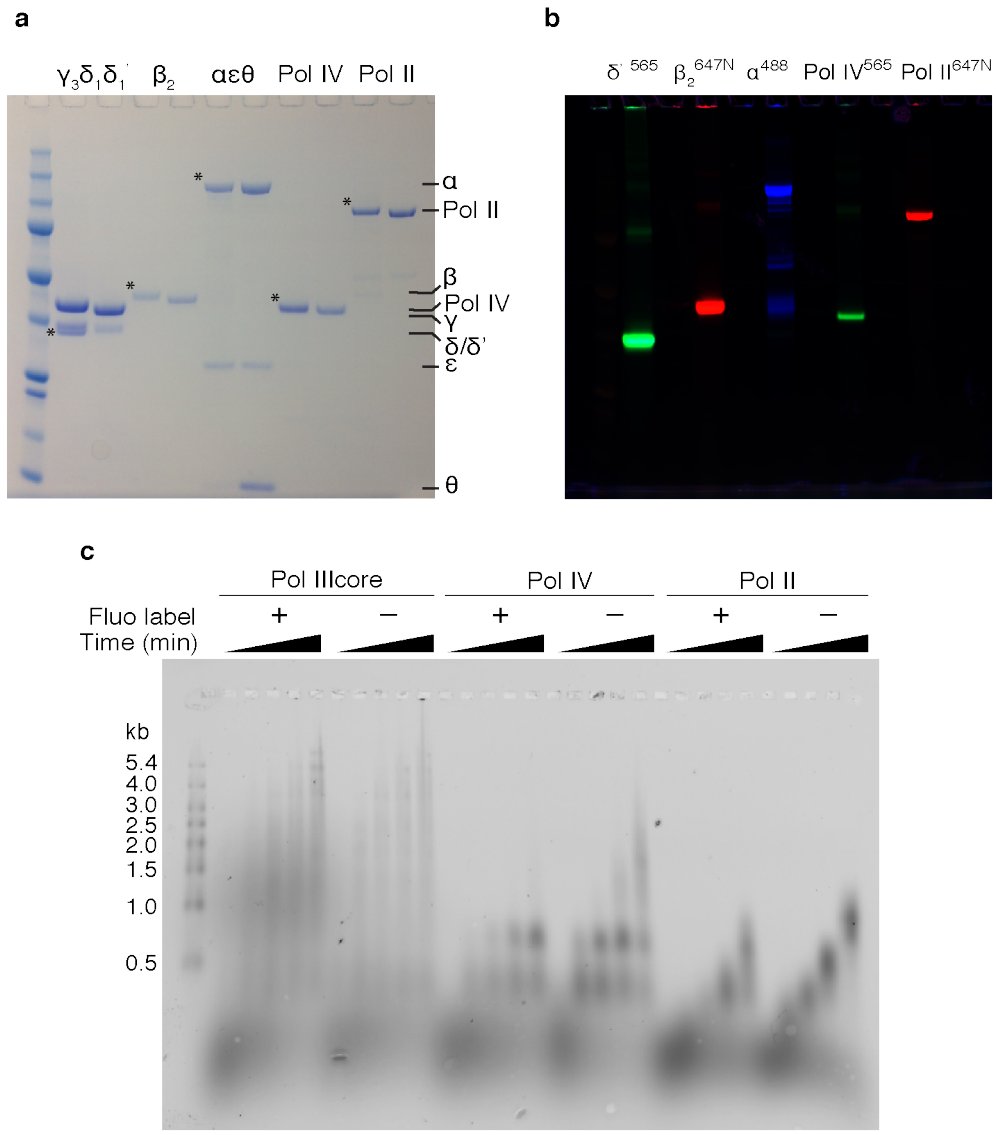


Figure 2.10. The fluorescently labeled proteins retained wild-type activity. **a**, Coomassie stained SDS-PAGE showing the unlabeled and labeled (*) proteins. **b**, Three-wavelength fluorescence scan of the gel shown in **b** before Coomassie staining, showing the labeled proteins. **c**, ϕX174 phage DNA (5.4 kb) primer extension assay showing the activity of the unlabeled and labeled proteins. In the labeled reactions, fluorescently labeled clamp, clamp loader and polymerase (Pol IIIcore, Pol II, or Pol IV) were used. The reactions were quenched at 0, 0.5, 1, 2, and 5 min. The increased background in the 'Pol IIIcore Fluo label +' lanes is caused by α-Atto 488, whose fluorescence signal overlaps with that of fluorescein on the DNA primer.

annealed to the primer DNA. This order is important because the primer DNA also contains a 5' biotin modification that is used for the surface-immobilization of the primer-template and therefore we do not want the primer biotin to be sequestered by binding to the monovalent streptavidin. In addition, the primer DNA contains a phosphothioate bond between the final

two nucleotides that is resistant to 3'-5' exonuclease activity of the *E. coli* ϵ subunit, in order to protect the primer strand from being degraded.

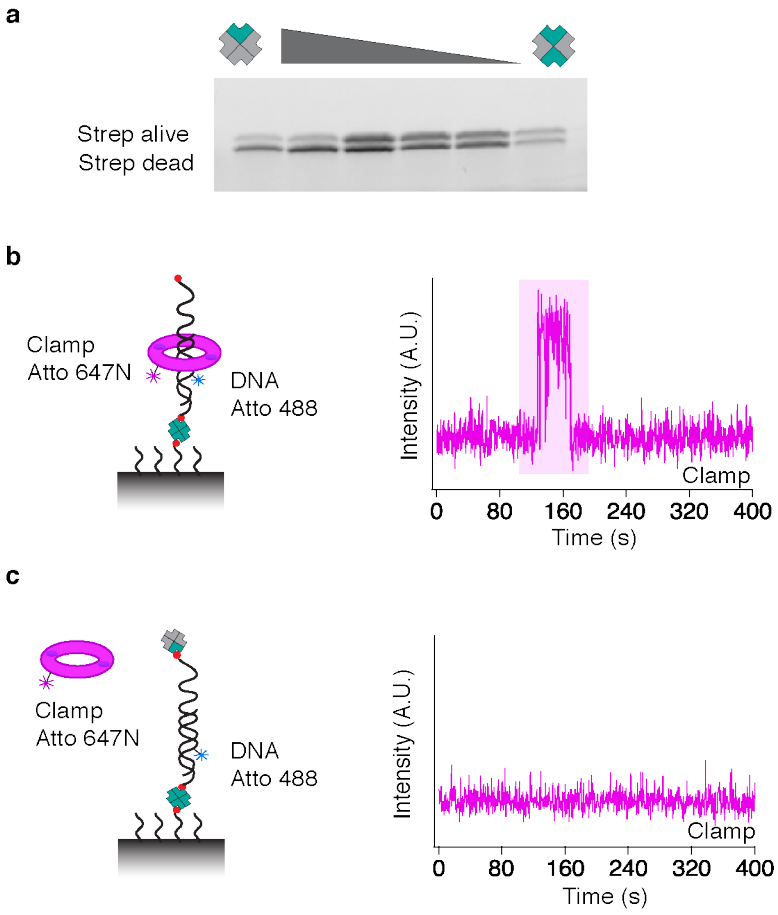


Figure 2.11. DNA is end-blocked with monovalent streptavidin. **a**, Purification of monovalent streptavidin by ion exchange chromatography. **b**, Clamp can thread and unthread itself on DNA with a free end. **c**, Clamp threading on DNA is blocked with monovalent streptavidin.

2.3.6 Setting up CoSMoS experiments to study protein-DNA interactions

To test whether the CoSMoS setup can be used to study protein-DNA interactions, we fluorescently labeled the Klenow fragment (KF) of the *E. coli* DNA polymerase I that results from proteolytic digestion of the full-length protein, which removes the 5' to 3' exonuclease domain. Previous bulk experiments have measured nanomolar affinity for KF binding to DNA[15] and KF-DNA interactions have also been shown using single-molecule FRET[17]. Therefore, KF is a good positive control for studying protein-DNA binding events by CoSMoS.

The labeling of KF using maleimide-Atto 647N is described in detail in Chapter 3. When KF-Atto 647N was incubated with Atto 565-labeled DNA, we observed frequent co-localization

events between the protein and the DNA signals (Fig. 2.12). To visualize the protein binding events on a single DNA molecule, a kymograph was created that shows the signal intensity over the course of the experiment. The kymograph shows multiple association and dissociation events of the KF on the DNA and therefore demonstrates that the CoSMoS setup can be used to study the co-localization between proteins and DNA.

In addition, we tested the binding of the single-strand binding (SSB) protein with the DNA substrate and here too we observed co-localization between SSB-Atto 647N and DNA-Atto 565 (Fig. 2.12). The co-localization between SSB and DNA can be explained by SSB binding to the single-stranded overhang of the template DNA strand. The SSB protein was labeled by Dr Emma Gleave in the lab.

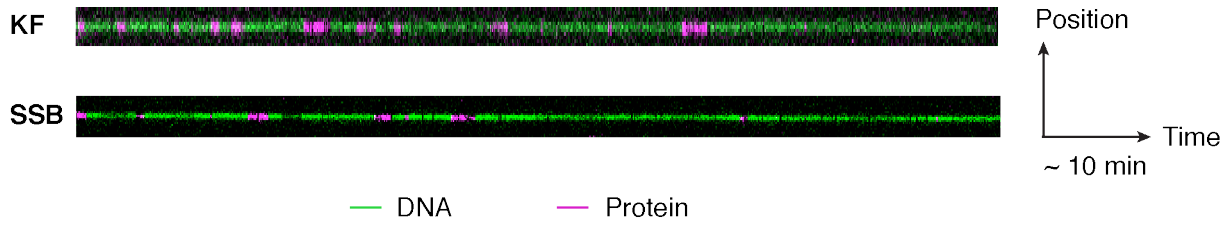


Figure 2.12. Kymographs showing protein-DNA binding. Initial experiments showing binding of Pol I Klenow fragment and single-strand binding protein to DNA.

2.3.7 Data acquisition

Since there are three laser channels in the current setup: 488 nm, 561 nm and 638 nm, and one laser channel is used for imaging the DNA, the number of labeled proteins that can be imaged is two. In order to increase this number to three, we imaged the DNA and proteins in series so that the same laser channel could be used twice. To do this, a short image stack was acquired of the DNA channel at high laser power to record the position of the DNA spots and then rapidly bleach away the signal (Atto 488, Fig. 2.14a). In order to control that the bleaching of the DNA did not affect protein binding on DNA, we measured the binding lifetime of Pol IIIcore on β clamp-DNA (see Chapter 3) with and without bleaching the DNA. As shown in Fig. 2.13, the bleaching of DNA did not affect the lifetime of Pol IIIcore binding. The same Atto 488 fluorophore that was placed on DNA can then be used to label a protein. Around 5-10 DNA spots were left unbleached (Fig. 2.14a) in order to map the position of the DNA spots to the protein channels during data analysis.

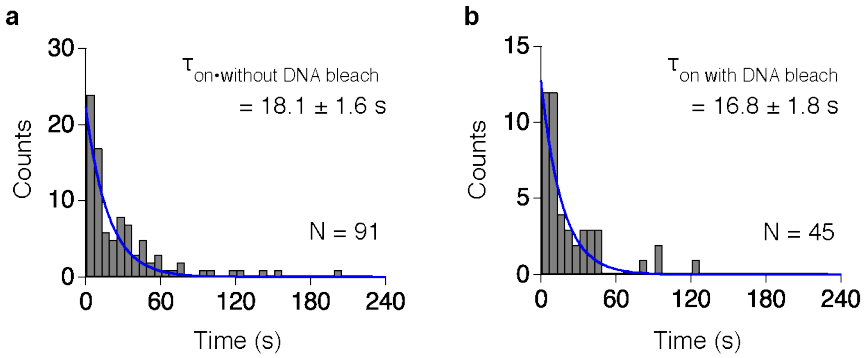


Figure 2.13. Bleaching of the Atto 488 fluorophore used to label DNA shows no effect on the binding of Pol IIIcore to clamp-DNA. a, The lifetime of Pol IIIcore on clamp-DNA without and **b**, with bleaching of the DNA signal.

The labeled proteins were manually injected into the flow cell and the acquisition of the protein channels was started immediately. The three laser channels were acquired sequentially, that is, one image of channel 1, one image of channel 2 followed by one image of channel 3 before the cycle starts again (Fig. 2.14b). We imaged each laser channel for 50 ms per frame for 1000 frames. There was a delay of 170 ms per frame due to the switching between the laser channels. Hence overall, the the time between two consecutive frames of the same laser channel was 660 ms (Fig. 2.14b).

2.3.8 Data analysis

Next, we needed an efficient way to analyze the data. This is not trivial as it requires i) the detection of binding events, ii) the determination of co-localization and iii) the quantitative analysis of hundreds of binding events. For this, we used an approach that was pioneered by Jeff Gelles and Larry Friedman at Brandeis University[5]. The analysis of multi-color image stacks involves multiple steps, and the analysis method specific to our data sets was developed in discussion with Larry and Jeff and are briefly described here. Firstly, the different laser channels need to be aligned to make sure that a spot in the 488 nm channel corresponds to the same spot in the 561 nm and the 638 nm channels. This was done by constructing a triple colored DNA substrate consisting of three annealed oligonucleotides each labeled with a different fluorophore (Atto 488, Atto 565 and Atto 647N). Then the spots containing all three colors are used to map the x and y coordinates in one channel to the others. The analysis revealed that the misalignment between the three laser channels is negligible (data not shown).

The images acquired of the DNA channel were used for selecting DNA spots to probe for

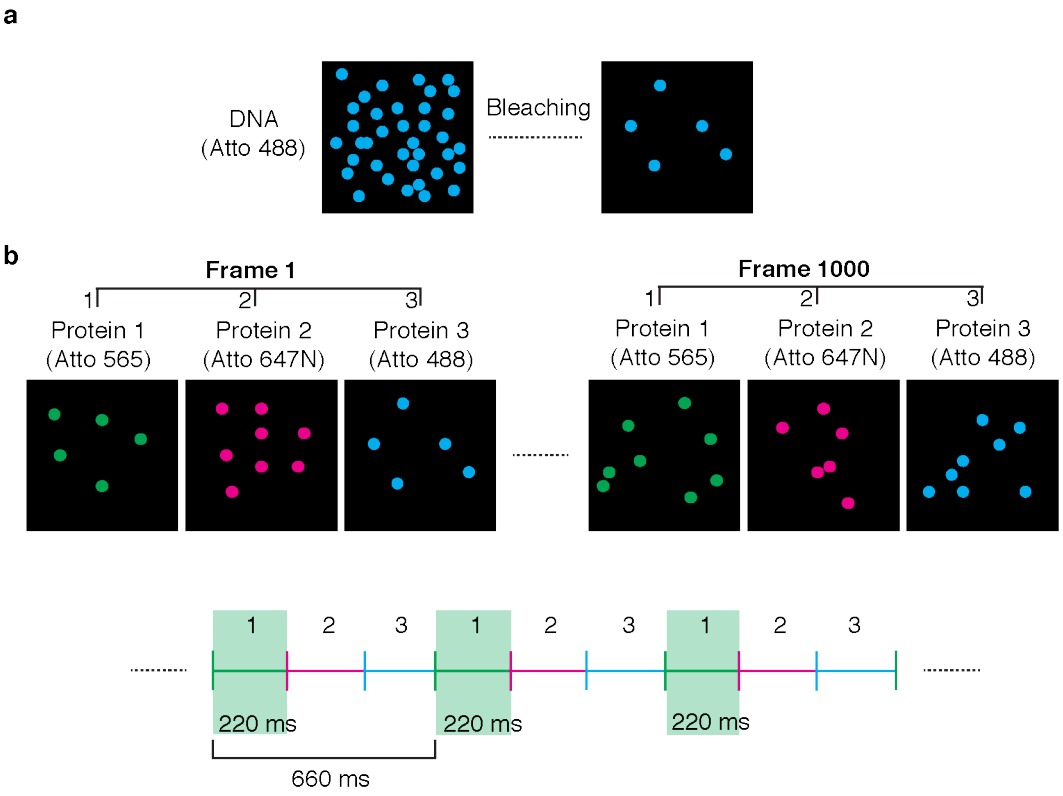


Figure 2.14. Data acquisition scheme. **a**, Image stack is acquired of the DNA channel, first to record the positions of the DNA molecules and then to bleach away the signal to recycle the fluorophore for protein imaging. **b**, Image stack is acquired of the protein channels where the channels are acquired sequentially within each image frame. The time between two consecutive frames of the same laser channel is 660 ms.

binding events in the protein channels. This selection was done in a semi-automated manner, where an intensity threshold input into the analysis package allowed all the spots above a certain intensity level to be picked. However, the package also picked closely spaced DNA spots that we then manually excluded from the analysis to avoid the protein binding events on a neighboring DNA molecule to be mistaken for a binding event on the DNA spot of interest. Once the DNA spots have been selected, mapping was done between the DNA channel and the protein channels in order to correct for any movement in the microscope stage in between acquiring the DNA images and injecting the labeled proteins into the flow cell. The mapping was done using the unbleached DNA molecules at the end of the DNA acquisition which were present in the first few frames of the protein channel imaged with the same laser (Fig. 2.14a and b).

Since the image stacks of the protein channels are acquired over ten minutes, the drift in the microscope stage also need to be corrected. This is because if the stage drift is significantly large over the course of the experiment, the integration of the signal intensity centered at single

xy coordinate will not be sufficient to follow the same spot over the stack of images. The correction for the stage drift was done by selecting for spots that remain for the whole duration of the image stack (i.e. proteins that are non-specifically bound to the surface) and using the changes in the coordinates of these spots to correct for the drift across the stack.

For each xy position of a DNA spot, the fluorescence intensity is integrated over the entire image stack in the protein channels. The output of this integration is a graph of signal intensity over time. However, an increase in the fluorescence intensity at a given position can be affected by fluorescent molecules in close proximity but not truly overlapping, and therefore looking at fluorescence intensity can be an error-prone way of studying co-localization events[5]. Instead, the true co-localization between two molecules is discriminated from the false positive events using a physical proximity threshold for the spots. Using this method, a positive co-localization event is scored if the distance between two molecules are less than 1.0 pixel (105 nm) apart. This threshold was selected by trial and error using different threshold values and picking a value that removes most false positive events but keeping the real co-localization events. Finally, each positive co-localization event was verified by visual inspection.

2.4 Discussion

Single-molecule methods offer previously unattained resolution when studying biological processes by revealing the dynamic behavior of individual molecules and the variation in a population of molecules. However, the development of a single-molecule method is not straightforward given the large number of variables that need to be tested and optimized for the specific experimental setup. There remains somewhat of a barrier between the developers of the single-molecule methods who commonly have backgrounds in physics and optics, and the biologists who want to apply these methods to study biological processes[13].

The initial stages of setting up the CoSMoS system was time-consuming as it involved significant trial and error. For example, a good amount of time was spent in testing various tapes for their contribution to the background fluorescence of the experiment; the first flow cells had high level of surface contamination that prompted more meticulous cleaning of the surfaces; the first CoSMoS experiments showed significant unspecific protein binding to the surface that encouraged further passivation of the surfaces. Building of this CoSMoS setup was a continuous improvement process that underwent various changes over time to improve the quality and reproducibility of the experiments. The final CoSMoS setup used for the experiments presented in this thesis is shown Fig. 2.15.

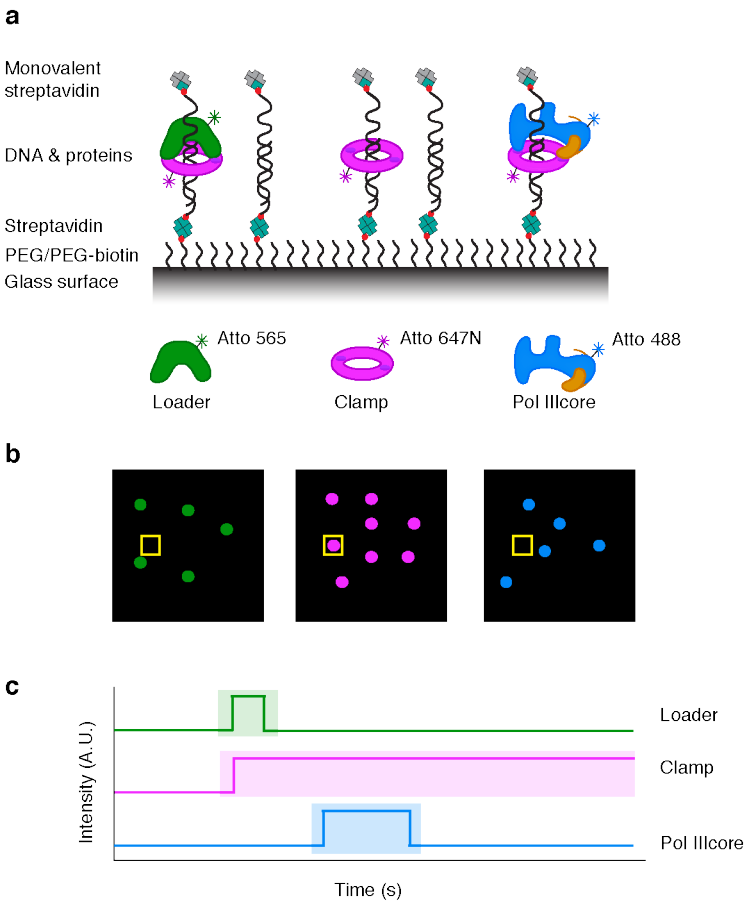


Figure 2.15. Experimental setup. a, Schematic representation of the experimental setup. DNA molecules are attached to a PEGylated glass slide via a biotin-streptavidin layer and are end-blocked with monovalent streptavidin. Fluorescently labeled proteins will be detected when bound to the DNA molecules. **b**, Schematic representation of the three fluorescent channels from a single image of a 1000-image series showing the presence of the fluorescently labeled molecules. **c**, Schematic representation of kymographs from a single position in the image series, revealing the binding and release of different proteins on the DNA at that position.

The biggest hurdle to increasing the throughput of single-molecule experiments is data analysis. While semi-automated, the current analysis still requires a significant amount of manual inspection. For example, the correction for drift in an image stack requires the identification and tracking of spots that remain for the whole duration of the experiment. Currently, these spots are identified manually by verifying their presence through the entire image stack, and this process could be simplified by automating the tracking of the spots in the different image frames. This would require a script that identifies the same spots over a number of images. In addition, given that the overall protocol for analyzing the CoSMoS images is the same for all experiments, it should be possible to automate the entire process from picking the DNA spots

to getting an output with the co-localization events.

In addition to the data analysis, the introduction of different fluids into the flow cell should also be possible to automate. Ideally, we would have multiple inlets into a flow cell and a micro-valve that governs the opening and closing of the individual inlets. The micro-pump would be under the control of a written program that controls the volume and speed that a solution is injected into the flow cell. This would be especially useful for the surface passivation steps that occur after the assembly of the flow cell, where the same volume of the same solutions are injected into the flow cell for every experiment.

While there are multiple variables to consider in a single-molecule experiment and currently it still takes a full day to analyze the results from a single experiment, with the automation of many of the processes, single-molecule experiments should be able to become a routine method in the laboratory in the near future.

Bibliography

[1] Elio A Abbondanzieri, William J Greenleaf, Joshua W Shaevitz, Robert Landick, and Steven M Block. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(7067):460–465, November 2005.

[2] David Dulin, Jan Lipfert, M Charl Moolman, and Nynke H Dekker. Studying genomic processes at the single-molecule level: introducing the tools and applications. *Nature Reviews Genetics*, 14(1):9–22, January 2013.

[3] Larry J Friedman, Johnson Chung, and Jeff Gelles. Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. *Biophysical Journal*, 91(3):1023–1031, August 2006.

[4] Larry J Friedman and Jeff Gelles. Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell*, 148(4):679–689, February 2012.

[5] Larry J Friedman and Jeff Gelles. Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods*, 86:27–36, September 2015.

[6] Bryan Gibb, Tim D Silverstein, Ilya J Finkelstein, and Eric C Greene. Single-stranded DNA curtains for real-time single-molecule visualization of protein-nucleic acid interactions. *Analytical Chemistry*, 84(18):7607–7612, September 2012.

[7] Jason Gorman, Teresa Fazio, Feng Wang, Shalom Wind, and Eric C Greene. Nanofabricated racks of aligned and anchored DNA substrates for single-molecule imaging. *Langmuir*, 26(2):1372–1379, January 2010.

[8] Taekjip Ha, Ivan Rasnik, Wei Cheng, Hazen P Babcock, George H Gauss, Timothy M Lohman, and Steven Chu. Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, 419(6907):638–641, October 2002.

[9] Iddo Heller, Tjalle P Hoekstra, Graeme A King, Erwin J G Peterman, and Gijs J L Wuite. Optical tweezers analysis of DNA-protein complexes. *Chemical Reviews*, 114(6):3087–3119, March 2014.

[10] Aaron A Hoskins, Larry J Friedman, Sarah S Gallagher, Daniel J Crawford, Eric G Anderson, Richard Wombacher, Nicholas Ramirez, Virginia W Cornish, Jeff Gelles, and Melissa J Moore. Ordered and dynamic assembly of single spliceosomes. *Science*, 331(6022):1289–1295, March 2011.

[11] Mark Howarth, Daniel J-F Chinnapen, Kimberly Gerrow, Pieter C Dorrestein, Melanie R Grandy, Neil L Kelleher, Alaa El-Husseini, and Alice Y Ting. A monovalent streptavidin with a single femtomolar biotin binding site. *Nature Methods*, 3(4):267–273, April 2006.

[12] Achillefs N Kapanidis, Emmanuel Margeat, Sam On Ho, Ekaterine Kortkhonjia, Shimon Weiss, and Richard H Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802):1144–1147, November 2006.

[13] Achillefs N Kapanidis and Terence Strick. Biology, one molecule at a time. *Trends in Biochemical Sciences*, 34(5):234–243, May 2009.

[14] M S Kellermayer, S B Smith, H L Granzier, and C Bustamante. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science*, 276(5315):1112–1116, May 1997.

[15] R D Kuchta, V Mizrahi, P A Benkovic, K A Johnson, and S J Benkovic. Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry*, 26(25):8410–8417, December 1987.

[16] Joshua D Larson, Margaret L Rodgers, and Aaron A Hoskins. Visualizing cellular machines with colocalization single molecule microscopy. *Chemical Society Reviews*, 43(4):1189–1200, February 2014.

[17] R P Markiewicz, K B Vrtis, D Rueda, and L J Romano. Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. *Nucleic Acids Research*, 40(16):7975–7984, September 2012.

[18] E Neher and B Sakmann. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature*, 260(5554):799–802, April 1976.

[19] Elizabeth Rhoades, Eugene Gussakovsky, and Gilad Haran. Watching proteins fold one molecule at a time. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6):3197–3202, March 2003.

[20] Gayan Senavirathne, Jeffrey G Bertram, Malgorzata Jaszczur, Kathy R Chaurasiya, Phuong Pham, Chi H Mak, Myron F Goodman, and David Rueda. Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution. *Nature Communications*, 6:10209, December 2015.

[21] T R Strick, V Croquette, and D Bensimon. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. *Nature*, 404(6780):901–904, April 2000.

[22] Martin R G Taylor, Mário Špírek, Kathy R Chaurasiya, Jordan D Ward, Raffaella Carzaniga, Xiong Yu, Edward H Egelman, Lucy M Collinson, David Rueda, Lumir Krejci, and Simon J Boulton. Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell*, 162(2):271–286, July 2015.

[23] Simina Ticau, Larry J Friedman, Nikola A Ivica, Jeff Gelles, and Stephen P Bell. Single-Molecule Studies of Origin Licensing Reveal Mechanisms Ensuring Bidirectional Helicase Loading. *Cell*, 161(3):513–525, April 2015.

[24] Feng Wang, Sy Redding, Ilya J Finkelstein, Jason Gorman, David R Reichman, and Eric C Greene. The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nature Structural & Molecular Biology*, 20(2):174–181, February 2013.

[25] Jin-Der Wen, Laura Lancaster, Courtney Hodges, Ana-Carolina Zeri, Shige H Yoshimura, Harry F Noller, Carlos Bustamante, and Ignacio Tinoco. Following translation by single ribosomes one codon at a time. *Nature*, 452(7187):598–603, April 2008.

[26] G J Wuite, S B Smith, M Young, D Keller, and C Bustamante. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*, 404(6773):103–106, March 2000.

[27] Ahmet Yildiz, Joseph N Forkey, Sean A McKinney, Taekjip Ha, Yale E Goldman, and Paul R Selvin. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science*, 300(5628):2061–2065, June 2003.

[28] Ahmet Yildiz, Michio Tomishige, Ronald D Vale, and Paul R Selvin. Kinesin walks hand-over-hand. *Science*, 303(5658):676–678, January 2004.

[29] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nature Protocols*, 1(1):280–285, 2006.

[30] Jun Yin, Paul D Straight, Shaun M McLoughlin, Zhe Zhou, Alison J Lin, David E Golan, Neil L Kelleher, Roberto Kolter, and Christopher T Walsh. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15815–15820, November 2005.

Chapter 3

Studying *E. coli* DNA replication by CoSMoS

3.1 Introduction

The *E. coli* β clamp is a circular protein that is loaded onto DNA by the clamp loader complex (see Chapter 1)[12,17]. Studies have shown that the loader binds and opens the clamp in an ATP-bound state and subsequent DNA binding triggers ATP hydrolysis and the release of the loader from clamp and DNA[12,17].

Much of the information on the general mechanism of clamp loading have been inferred from existing structures of the loader and clamp in the different stages of the clamp loading process. However, the existing structures are of clamp loaders from different species that are very divergent in sequence, and only single structures are available for some key intermediates. The existing structures are of the *E. coli* clamp loader in the absence[15] and presence[24] of DNA, the yeast clamp loader bound to a closed clamp in the absence of DNA[1], and the T4 bacteriophage clamp loader bound to an open clamp in the presence of DNA (Fig. 1.3)[16]. Therefore to better understand the clamp loading process, it is important to study the clamp loading cycle using the clamp loader and clamp from a single organism.

It has been estimated that 2000-4000 Okazaki fragments are needed for one round of replication of the *E. coli* 4.6 megabase genome[20]. However, each *E. coli* cell is thought to contain only 300 clamps[3] and therefore the clamps that are loaded onto the lagging strand need to be unloaded and recycled to allow continuous DNA synthesis to occur. The unloading of clamps is carried out by the same clamp loader complex that loads clamps onto DNA, but the mechanism of clamp unloading is much less studied.

The composition of the *E. coli* clamp loader complex is presented in Section 1.3.5 of Chapter 1. Briefly, the *in vivo* clamp loader is the τ complex, consisting of either two or three copies of the τ protein ($\tau_2\gamma_1\delta_1\delta'_1$ or $\tau_3\delta_1\delta'_1$) depending on the number of the replicative polymerase Pol IIIα tethered to it (Fig. 1.5, Chapter 1). However, the γ complex $\gamma_3\delta_1\delta'_1$, in which the full-length τ protein is replaced by the truncated protein γ, is fully functional in loading clamps onto DNA *in vitro* although it does not bind to Pol IIIα. In order to separate the processes of clamp loading and unloading from polymerase binding, we use the minimal clamp loader in the form $\gamma_3\delta_1\delta'_1$ in our experiments. We also study clamp loading using a fully saturated $\tau_3\delta_1\delta'_1$ complex with three polymerases bound as a comparison.

The replicative polymerase in *E. coli* Pol IIIα forms a stable trimeric core complex with the exonuclease ϵ subunit and the accessory θ subunit, termed the Pol IIIcore[23]. Pol IIIcore associates with the β clamp on DNA to initiate processive synthesis. While the transition from the loader bound clamp to the polymerase bound clamp needs to occur to allow for DNA synthesis, the molecular details of how this transition is organized and how the polymerase achieves processive DNA synthesis on the clamp remain unclear.

Therefore, the aim of this chapter is to study the loading of the replicative DNA polymerase onto the β clamp on DNA as well as the detailed kinetics and dynamics of the clamp loading and polymerase loading events. These questions can be uniquely addressed using CoSMoS given that the order of assembly and the lifetimes of single protein molecules on DNA can be measured.

3.2 Materials and Methods

3.2.1 Cloning of protein expression vectors

Genes for *E. coli* β (*dnaN*), ϵ (*dnaQ*), θ (*holE*), γ (*dnaX*) and *Bacillus subtilis* Sfp phosphopantetheinyl transferase were cloned into pET28a vectors, and genes for δ (*holA*), single cysteine δ' (*holB*) K83C/C217S/C294S[11] and τ (*dnaX*) were cloned into pET3d vectors. The sequence for Pol I (*polA*) Klenow fragment (residues 324-928) was cloned into a pETNKI-His-3C-LIC vector[21]. For labeling purposes the Ybbr sequence DSLEFIASKLA[32] was added N-terminally to Pol IIIα (*dnaE*) in a pETNKI-His-3C-LIC vector. The clamp binding mutants of ϵ were generated through site-directed mutagenesis: residues 182-187 were mutated from QTSMAF to QTSLPL for $\epsilon(\beta+)$[28] and to QTSAAA for $\epsilon(\beta-)$[7]. The plasmids for streptavidin 'alive' (biotin-binding) and streptavidin 'dead' (not biotin-binding)[13] were generous gifts from M. Howarth

(Univ. of Oxford).

3.2.2 Protein expression and purification

All columns were purchased from GE Healthcare. Unless otherwise specified, protein purifications were performed with the following gradients using the columns stated: nickel affinity using Histrap HP (25-500 mM Imidazole gradient in the presence of 500 mM NaCl), ion exchange using Hitrap Q HP for anion exchange and Hitrap SP HP for cation exchange (0-1 M NaCl gradient), Hitrap Phenyl HP for hydrophobic interaction chromatography (2-0 M ammonium sulfate gradient) and a 120 ml Superdex 200 column for gel filtration (150 mM NaCl).

Pol IIIα-Nybbr, β, θ and Sfp were expressed in *E. coli* BL21 (DE3) Gold cells for 2 hours at 30 °C and purified in 20 mM Hepes pH 7.5 and 2 mM DTT by nickel affinity chromatography, anion exchange and gel filtration.

The wild-type ϵ and the β binding mutants of ϵ were expressed in *E. coli* BL21 (DE3) Gold cells for 2.5 hours at 30 °C, and purified in 25 mM Hepes pH 8.2 and 2 mM DTT by nickel affinity under denaturing conditions (in the presence of 6 M Urea), followed by refolding overnight at 4 °C in 25 mM Hepes pH 8.2 and 10 mM DTT and anion exchange.

γ, δ and δ' K83C/C217S/C294S were expressed in *E. coli* BL21 (DE3) Gold cells for 2 hours at 25 °C followed by 2 hours at 18 °C. τ was expressed in *E. coli* BL21 (DE3) pLysS cells for 1 hour at 30 °C. δ and δ' K83C/C217S/C294S proteins were purified in 50 mM Hepes pH 7.5, 0.1 mM EDTA and 10 mM DTT by phenyl hydrophobic chromatography, anion and cation exchange. γ was purified in the same buffer by nickel affinity chromatography. τ was purified by 23 % to 17 % (w/v) ammonium sulfate cuts, followed by cation and anion exchange in 50 mM Hepes pH 7.5, 0.5 mM EDTA and 2 mM DTT. The $\gamma_3\delta_1\delta'_1$ and $\tau_3\delta_1\delta'_1$ complexes were assembled after the labeling of δ' K83C/C217S/C294S (see below).

Pol I Klenow fragment was expressed *E. coli* BL21 (DE3) Gold cells for two hours at 30 °C and purified in 20 mM Tris pH 8.0 and 2 mM DTT by nickel affinity chromatography, anion exchange and gel filtration.

Streptavidin alive and dead were purified and prepared as described[13].

3.2.3 Protein labeling

All gel filtration experiments in this section were carried out using a 2.4 ml Superdex 200 Increase column (GE Healthcare), with the exception for the β clamp, which is smaller size (80 kDa) compared to the other proteins and protein complexes and therefore run on a 2.4 ml

Superdex 75 column (GE Healthcare). The columns were equilibrated in 20 mM Tris-HCl pH 7.5, 50 mM potassium glutamate and 2 mM DTT.

Pol IIIα was labeled at a single site in the N-terminus using the Sfp method[31]. Atto 488 fluorophore was conjugated to Coenzyme A as described[32] and Pol IIIα was labeled in a reaction containing 10 mM $MgCl_2$, 50 mM Hepes 7.5, 10 mM DTT, 2 μM Sfp, 100 μM CoA-Atto 488, and 50 μM protein for 2 hours at room temperature. The labeled protein was purified away from the free dye by gel filtration and the labeling efficiency was determined spectrophotometrically using the protein and fluorophore absorption ratios, with the free fluorophore absorption at 280 nm subtracted from the protein absorption at 280 nm. The labeling efficiency of Pol IIIα was 60 $\%$. The labeling efficiency for Pol IIIα was verified at the single-molecule level by measuring the co-localization frequencies between labeled Pol IIIα and labeled ϵ (see Chapter 4) on clamp-DNA. Each co-localization event was scored as having labeled Pol IIIα only, labeled ϵ only, or both labeled proteins. Measuring the labeled efficiency this way showed 67 $\%$ labeling for Pol IIIα, which is similar to the efficiency measured by absorption.

The labeled Pol IIIα protein was then re-constituted into Pol IIIcore by incubating the protein together with wild-type ϵ and θ proteins at 1:1:1 ratio and Pol IIIcore was purified by gel filtration. Peak fractions from the gel filtration were analyzed by SDS-PAGE.

The β clamp is a homodimer and each monomeric subunit contains 4 cysteines. However each monomer was specifically labeled on a single cysteine residue as determined by mass spectrometry (data not shown). DTT was removed from the buffer by gel filtration prior to labeling as DTT contains thiol groups that would otherwise react with the maleimide dye. β clamp was then incubated with a 5 molar excess of maleimide-Atto 647N dye for 5 hours at 4 °C. After labeling, the reaction was quenched and the protein was purified away from the free dye by gel filtration in the presence of DTT. The labeling efficiency of the β monomer, 68 $\%$, was determined spectrophotometrically .

The δ' K83C/C217S/C294S protein was buffer exchanged by gel filtration to remove DTT. The protein was then labeled using a 5 molar excess of maleimide-Atto 565 dye for 1 hour at room temperature. The labeled protein was purified away from the free dye by gel filtration in buffer containing DTT to quench the labeling reaction. The labeling efficiency, 85 $\%$, was measured spectrophotometrically. The labeled δ' protein was then mixed together with the purified γ/τ and δ proteins at 1:2:1 ratio of γ/τ:δ:δ' and the re-constituted complexes ($\gamma_3\delta_1\delta'_1$ and $\tau_3\delta_1\delta'_1$) were purified by anion exchange. Peak fractions from the ion exchange were analyzed by SDS-PAGE. To fully occupy all Pol IIIcore binding sites on $\tau_3\delta_1\delta'_1$, a 5-fold excess of Pol IIIcore was added to the τ complex, and the complex was purified by gel filtration.

Pol I Klenow fragment was buffer exchanged on gel filtration to remove DTT. The single cysteine residue in Pol I Klenow fragment (C907) was labeled for 1 hour at room temperature using a 5 molar excess of maleimide-Atto 647N dye. The labeled protein was purified away from the free dye by gel filtration in the presence of DTT and the labeling efficiency was determined spectrophotometrically to be 100 %.

3.2.4 ϕX174 primer extension assays

Activity of the labeled proteins was tested using a single-stranded ϕX174 phage DNA (New England Biolabs) annealed to a primer with a 5' label (5'FAM-ACCAACATAAACATTATT-GCCCGGCGTACpG, where lowercase 'p' indicates a non-cleavable phosphothioate bond). Reactions were performed in 20 mM Tris pH 7.5, 2 mM DTT, 50 mM potassium glutamate, 8 mM magnesium acetate, and 0.05 mg/ml BSA. Each reaction contained 5 nM primed ϕX174 phage DNA, 50 nM β clamp, 10 nM γ clamp loader complex ($\gamma_3\delta_1\delta'_1$), and 30 nM Pol IIIcore. Reactions were stopped at 0, 0.5, 1, 2 and 5 minutes using a stop buffer containing 75 mM EDTA and 0.6 % (w/v) SDS, and were separated on an alkaline agarose gel (0.8 % agarose, 30 mM NaOH, 2 mM EDTA) for 15 hours at 14 V. Gels were scanned at 488 nM using a Amersham Typhoon (GE LifeSciences).

Results of the ϕX174 primer extension assays are presented in Section 2.3.4 of Chapter 2.

3.2.5 DNA substrate

The DNA oligos were purchased from IDT. The template DNA is 33 nt in length, with the sequence 5'Bio-CATAATATCCATGCTTCACC[amino-dT]TCATCCAAATCC, where 'amino-dT' represents an internal amino modification on a thymine base for the labeling of the DNA. The template DNA was labeled with excess NHS-ester activated Atto 488 fluorophore for 1 hour at room temperature in 0.1 M sodium phosphate buffer (0.1 M NaH_2PO_4 and 0.1 M Na_2HPO_4, pH 7.0) and the free dye was separated by gel filtration on a 2.4 ml Superdex 75 column (GE Healthcare). The labeled template DNA was then bound to monovalent streptavidin and purified by gel filtration before annealing to a 27-nt primer DNA with the sequence 5'Bio-GGATTTGGATGAAGGTGAAGCATGGApT, where the lowercase 'p' indicates the non-cleavable phosphothioate bond).

3.2.6 Preparation of slides

Glass slides and coverslips were washed in 3 M NaOH and Piranha solution (3:2 concentrated sulfuric acid: 30 % hydrogen peroxide) and then silanized (using 3-Glycidyloxypropyl)trimethoxysilane from Sigma) and pegylated (using 5:1 HO-PEG-NH_2 and biotin-CONH-PEG-NH_2 from Rapp Polymere). The imaging chamber (15 μL) was assembled by creating a sandwich between the coverslip and glass slide using double adhesive tape, and it was further passivated using 4 mg/ml PLL-PEG (SuSoS), 1 % (w/v) Pluronic F127 (Sigma) and 10 mg/ml BSA (New England Biolabs). Streptavidin (1 mg/ml) (New England Biolabs) was added last to bind the biotin-DNA.

3.2.7 TIRF microscopy

The labeled proteins were added together at 15 nM β clamp, 15 nM γ complex, 30 nM Pol IIIcore and 0.5 mM ATP in imaging buffer containing 20 mM Tris-HCl pH 7.5, 50 mM potassium glutamate, 8 mM $MgCl_2$, 4 % glycerol, 2 mM DTT, 0.1 % Tween20 and 1 mM Trolox. 10 nM of the τ complex was added in the reactions as it was saturated with Pol IIIcore on each of the three polymerase binding sites, and therefore the final concentration for Pol IIIcore in these experiments was 30 nM. In the Pol IIIcore idling experiments, 0.5 mM of each of dATP and dTTP was added to the reaction.

The reaction was manually injected into the imaging chamber and imaged on a Nikon Eclipse Ti-E microscope with ApoTirf 100X/1.49 Oil, 0.13-0.20 WD 0.12 objective controlled by the Micromanager software. The lasers used were 150 mW 488 nm, 150 mW 561 nm (both Coherent Sapphire) and 100 mW 638 nm (Coherent Cube) controlled by an acousto-optic tunable filter (Gooch & Housego). Movies were acquired on an Andor iXon (EM)+ CCD camera at 20-40 mW laser power (output from the laser) and an exposure of 50 ms per frame with rapid alternation between the laser channels (170 ms for the exchange). The frame rate was therefore 660 ms for a 3-color experiment, 440 ms for a 2-color experiment, and 86 ms for a single color experiment (as there was no need to switch between the laser channels). Prior to imaging the protein channels, a movie was acquired of the DNA channel at 50-60 mW laser power (output from the laser) and 1 s per frame exposure rate to record the positions of the DNA molecules and to bleach the DNA-Atto 488 signal in order for the channel to be re-used for Pol IIIα-Atto 488.

3.2.8 Data analysis

The acquired movies were analyzed using the Imscroll analysis GUI developed by Jeff Gelles and Larry Friedman[8] to find the individual landing events and their dwell times. Each field of view (54 μm $\times$ 54 μm, pixel size 105 $\times$ 105 nm) had an average density of 5900 DNA molecules of which 700-900 molecules were well-spaced and picked for analysis. Histograms were plotted in Igor Pro and the data was fitted with a single exponential. The decay constant τ represents the mean lifetime/lag time and the error in τ represents the error in the mean lifetime/lag time.

3.3 Results

3.3.1 Protein labeling and complex re-constitution

In order to able to quantify the number of interacting molecules in an experiment, we aimed to label the proteins at a single site.

Given that the γ complex consists of five polypeptides, each with multiple cysteine residues, we wondered whether we could develop a labeling method that avoids perturbations to the assembly of the γ complex. A previous study showed that the γ complex binds with high affinity to a short peptide of the ψ subunit of the holoenzyme[24]. Therefore, we cloned the sequence of the ψ peptide (residues 2-28) into a vector containing a SUMO tag to enable to expression of small peptides in *E. coli* and subsequently allow precise cleavage of the tag. Unfortunately the ψ peptide was sensitive to protease cleavage during purification and therefore could not be incorporated into the γ complex at any observable levels (data not shown).

Given the challenges of the indirect labeling method, we switched to directly labeling the γ complex at the δ' subunit by making a triple mutant (K83C/C217S/C294S) to achieve single site labeling[11]. The labeled δ' protein was purified by gel filtration which allowed the separation of the labeled protein from excess dye (Fig. 3.1a). The labeled δ' protein eluted from gel filtration at the same retention volume as the unlabeled protein prior to the labeling reaction, which showed that the labeling did not interfere with protein folding.

The labeled δ' protein was incubated with the δ and γ proteins at limiting γ concentrations for complex re-constitution. This is because the isolated γ protein elutes at a similar salt concentration as the γ complex during ion exchange and therefore we wanted to ensure that any γ protein will be incorporated into the complex. Excess labeled δ' and unlabeled δ proteins were separated from the re-constituted γ complex by ion exchange (Fig. 3.1b-c).

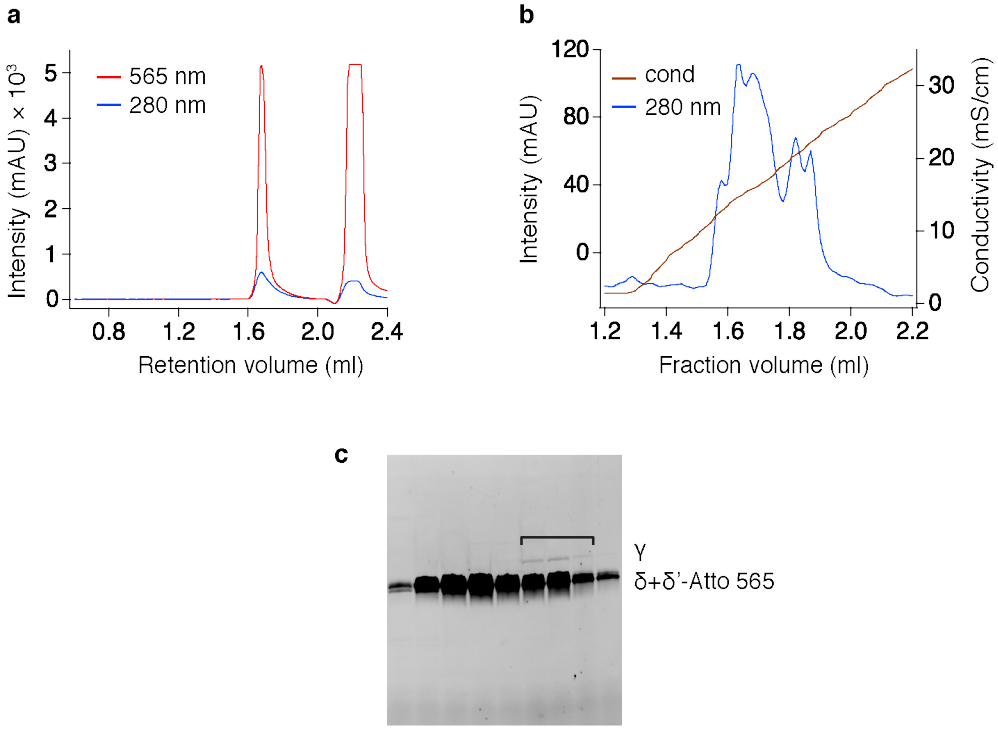


Figure 3.1. Labeling of the γ complex. **a**, Gel filtration run of the δ' labeling reaction. **b**, Ion exchange run of the re-constituted labeled γ complex. **c**, SDS-PAGE gel analysis of the peak fractions of the ion exchange run. The high intensity from the δ'-Atto 565 signal is due to fluorophore contribution to the gel visualized by UV.

Given that Pol IIIα contains 12 cysteine residues, it would be non-trivial to mutate away all cysteines apart from one for labeling with maleimide dyes at a single site. Therefore, we decided to label it at the N-terminus using the previously published Sfp method[32]. We tried a number of labeling conditions for the labeling of Pol IIIα including the duration of incubation (1, 2, 3 and 5 hours), temperature (room temperature or at 4 °C) and fluorophore to protein ratios (1:1 and 5:1). However, we were unable to increase the labeling efficiency beyond 60 %. The labeled Pol IIIα protein was separated from free dye by gel filtration (Fig. 3.2a). Incubation of the labeled Pol IIIα protein with exonuclease ϵ and the accessory protein θ at equimolar concentrations allowed the re-constitution of labeled Pol IIIcore, seen as a slight shift in the peak on gel filtration as compared to the labeled Pol IIIα alone (1.26 ml for Pol IIIcore and 1.29 ml for Pol IIIα) and the co-migration of Pol IIIα, ϵ and θ by gel analysis (Fig. 3.2b-c).

The labeling of the β clamp and the Klenow fragment of Pol I was optimized by testing a range of labeling conditions as done above for Pol IIIα, and the labeled protein was separated from free dye by gel filtration (Fig. 3.3a-b). In the case of labeling the β clamp, the free dye stuck to the gel filtration column and eluted after the run which explains why a peak

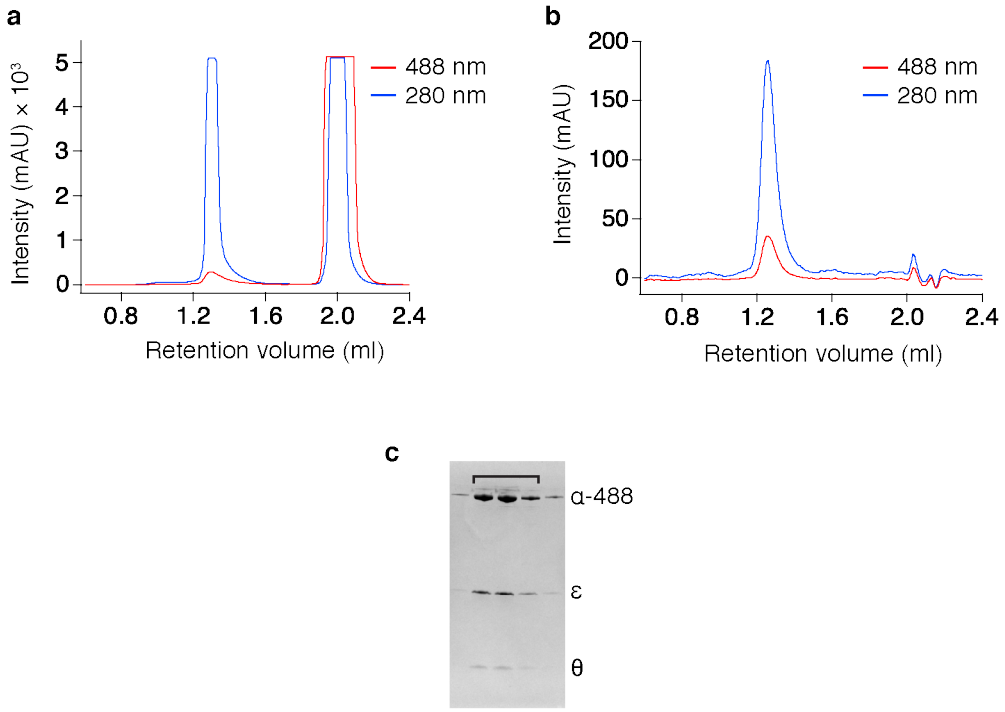


Figure 3.2. Labeling of Pol IIIcore. **a**, Gel filtration run of the Pol IIIα labeling reaction. **b**, Gel filtration run of the re-constituted labeled Pol IIIcore. **c**, SDS-PAGE gel analysis of the re-constituted Pol IIIcore. Gel stained using Coomassie.

corresponding to the free dye was not observed (Fig. 3.3a).

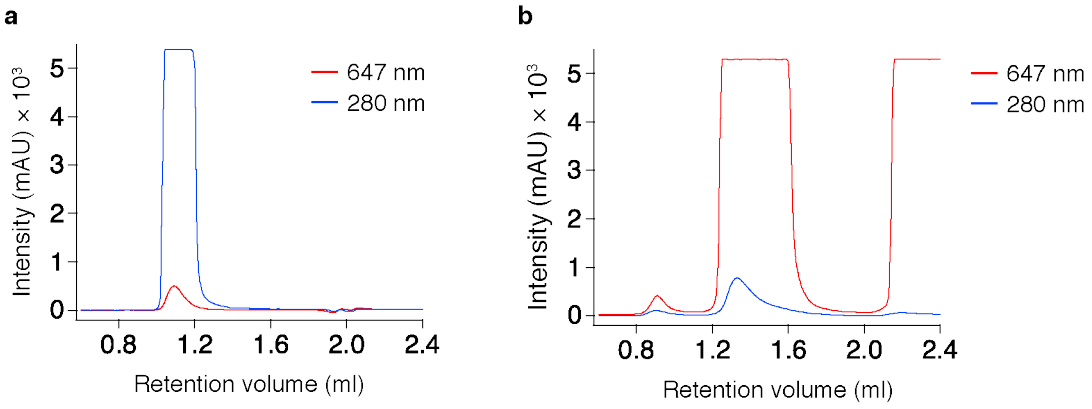


Figure 3.3. Labeling of β clamp and Klenow fragment. **a**, Gel filtration run of the β clamp labeling reaction. **b**, Gel filtration run of the Pol I Klenow fragment labeling reaction.

In addition to the assessment of correct protein folding by gel filtration, a further quality control of the labeled proteins was performed using a primer extension assay. Here, the ability of the labeled proteins to extend a primed single-stranded (ss) DNA substrate was compared to the unlabeled proteins and as shown in Fig. 2.10, Section 2.3.4 of Chapter 2, the labeled protein

retained wild-type activities.

3.3.2 Clamp loading

As shown in Fig. 2.11 in Section 2.3.5 of Chapter 2, isolated β clamps do not bind to the DNA substrate that is end-blocked with streptavidin. However, when the clamp is combined with the γ clamp loader complex, frequent clamp loading events are observed where the loader and clamp bind to the DNA simultaneously (Fig. 3.4a). The difference between the loader and clamp arrival times is zero in 67 $\%$ of the events and a single time frame in 33 $\%$ of the events (Fig. 3.4b). The difference in the arrival times can be explained by the sequential acquisition of the three laser channels within a single time frame (Fig. 3.4b-c). This shows that the *E. coli* loader and clamp form a pre-complex in solution, which was previously inferred from the yeast structure of the loader bound to a closed clamp in the absence of DNA[1].

Upon arrival at the DNA, the loader binds to the DNA very briefly, while the clamp remains bound for the remaining time of the data acquisition (Fig. 3.4a). Given that the lifetimes of the loader and the clamp are very different, it was not possible to accurately measure both lifetimes in a single experiment. To capture the long lifetime of the clamp on DNA, we first loaded the clamp onto DNA and then washed away the excess proteins. We imaged the clamp for one hour with a 10 s interval between the frames to avoid photo-bleaching of the signal. As shown in Fig. 3.4d, the clamp remained bound to DNA for 1430 $\pm$ 180 s or $\sim$ 23 minutes, excluding the time that was taken for the clamps to be loaded and the washing step prior to the start of the acquisition ($\sim$ 3 minutes).

In contrast, the lifetime of the γ clamp loader complex on DNA is very short-lived and in order to determine it accurately, we used only one laser channel to decrease the frame rate to 86 ms compared to the 660 ms for all three channels as is discussed in Section 2.3.8 of Chapter 2. The lifetime of the isolated loader on DNA is 1.20 $\pm$ 0.05 s, which is shortened to 0.41 $\pm$ 0.01 s in the presence of the clamp during clamp loading (Fig. 3.4e).

ATP hydrolysis is required for the release of the clamp loader from clamp-DNA as evidenced by the fact that in the presence of ATPγS, the slowly hydrolyzable ATP analogue, the loader and clamp still bind to the DNA simultaneously but also release together (Fig. 3.5a). The loader and clamp bind to the DNA briefly for 2.7 $\pm$ 0.2 s (Fig. 3.5b), which contrasts with the long lifetimes for the loaded clamps on DNA. Interestingly, the lifetime of the loader and clamp complex on DNA is longer than the lifetime of loader alone on DNA, which may be explained by the additional DNA-binding surfaces contributed by the clamp[9].

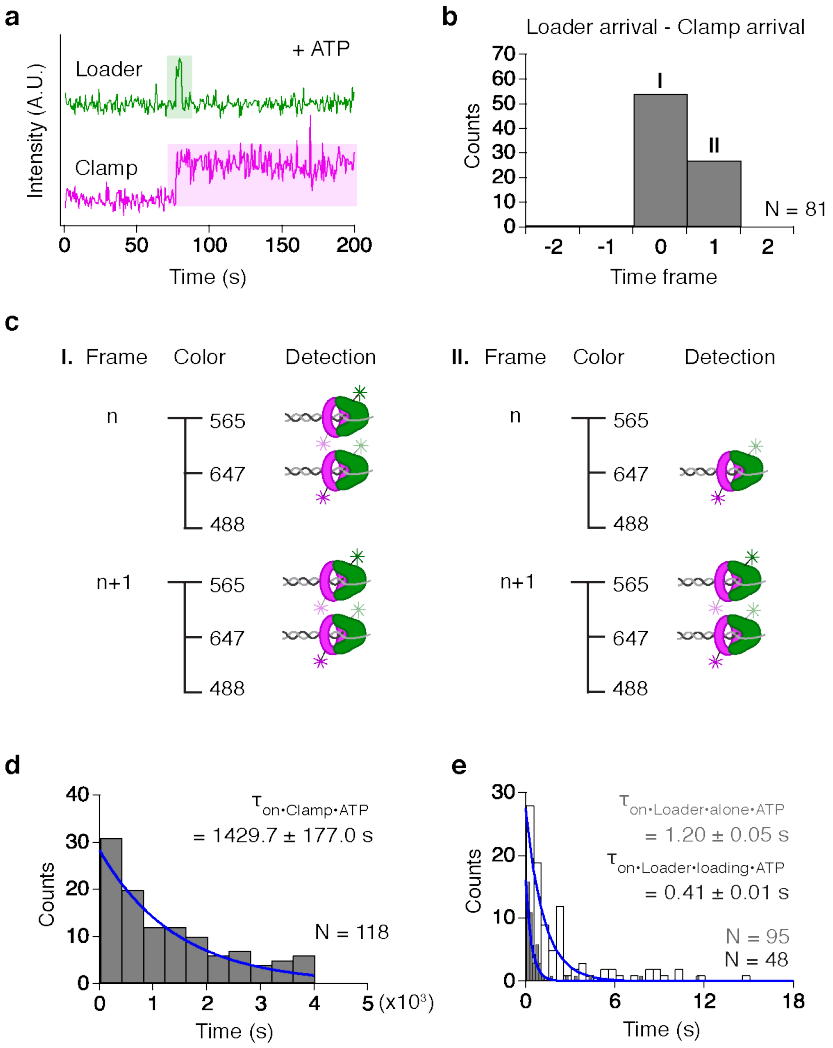


Figure 3.4. Clamp loading in the presence of ATP. a, Representative trace showing clamp loading on DNA in the presence of ATP. **b**, Binary distribution of the difference in the loader and clamp arrival times (I and II). **c**, Schematic representation of how the simultaneous arrival of loader and clamp can be recorded in a single (I) or two (II) sequential image frames. **d**, The distribution of lifetimes for the clamp and **e**, the loader on DNA. All lifetimes are represented as mean ± s.e.m.

Taken together, our data is consistent with previous studies showing that the ATPase activity of the clamp loader increases with clamp and DNA binding and that this triggers the release of the clamp loader from clamp-DNA [12 17].

3.3.3 Clamp unloading

Our data reveal a different temporal organization of the loader and the clamp during clamp unloading as compared to clamp loading. During unloading, the loader binds to the DNA for an

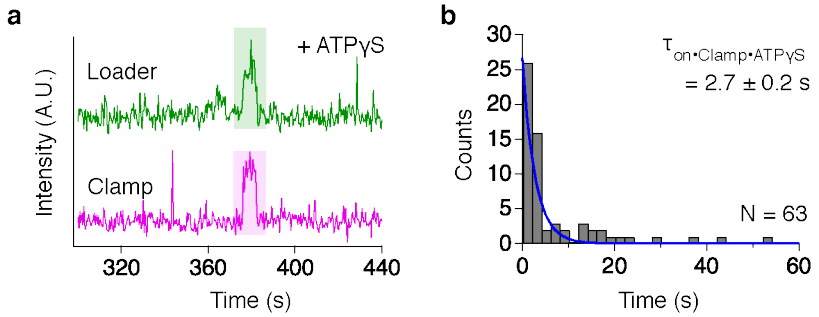


Figure 3.5. Clamp loading in the presence of ATPγS. a, Representative trace showing the simultaneous arrival and release of loader and clamp on DNA. **b**, The distribution of lifetimes for the clamp represented as mean ± s.e.m.

average of 10.8 ± 1.2 s but the clamp is released within 4.1 ± 0.4 s of its arrival (Fig. 3.6a-c). Therefore the loader remains bound to the DNA by itself once the clamp has been unloaded. It is curious that the isolated loader binds to the DNA substrate for 1.2 s (Fig. 3.4e) but once the clamp has been unloaded, the loader remains bound to the DNA for 6-7 s, which may suggest that the loader adopts an entirely different conformation during unloading compared to what can be achieved by the isolated loader.

In addition to the differences in the temporal organization of the unloading process, we also find that contrary to clamp loading, clamp unloading does not require ATP. To test the nucleotide requirement of the unloading process, we first loaded clamps onto DNA and then washed away the ATP and free proteins (Fig. 3.6d). The loaded clamps remained bound to the DNA and were unloaded by the introduction of new loader proteins in the absence of ATP. We controlled for the success of washing away the ATP in a separate experiment where new loader and clamp proteins were introduced after the wash. No loading events were observed which confirmed that there was no residual ATP in the chamber (data not shown). The loader lifetime and the lag time between loader arrival and clamp release are identical in the clamp unloading events in the presence and absence of ATP, showing that the unloading events are ATP-independent (Fig. 3.6e-f).

3.3.4 Polymerase loading

Next we studied the binding of the replicative DNA polymerase to DNA. Pol IIIcore associates with DNA very transiently, for less than the temporal resolution of a single laser channel (86 ms) (Fig. 3.7a-b). Multiple short-lived associations were observed on a single DNA molecule. This suggests that the polymerase makes short-lived contacts with the DNA but is unable to bind

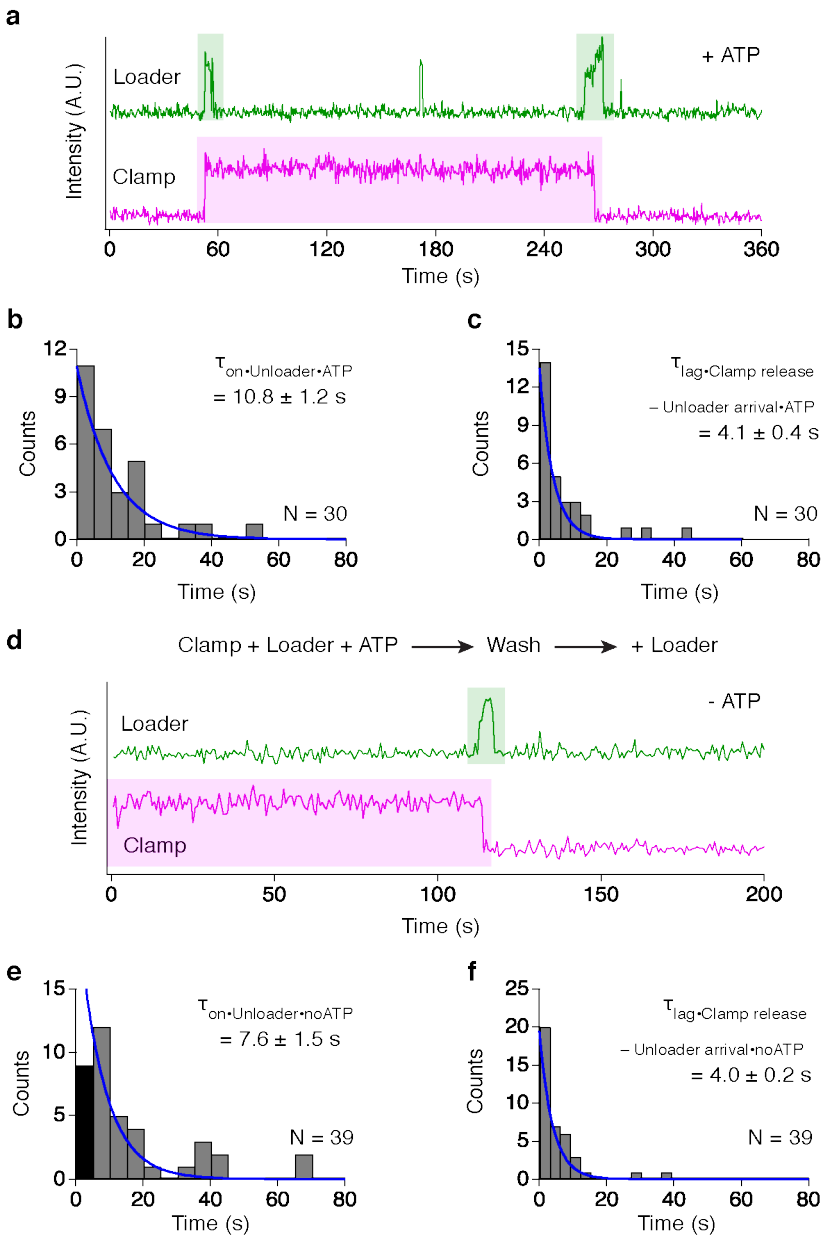


Figure 3.6. Clamp unloading in the presence and absence of ATP. a, Representative trace showing clamp loading and unloading by the loader in the presence of ATP. **b**, The distribution of lifetimes of the loader on DNA during clamp unloading. **c**, The distribution of lag times between the arrival of the loader and the release of the clamp. **d**, Representative trace showing the unloading of a pre-loaded clamp in the absence of ATP. **e**, The distribution of lifetimes of the loader on DNA. **f**, The distribution of lag times between the arrival of the loader and the release of the clamp. All lifetimes and lag times are represented as mean ± s.e.m.

stably, which agrees with previous studies showing that Pol IIIα is a poor enzyme in isolation[25 6] and reporting a low affinity for Pol IIIcore DNA binding[7].

The collisions nature of Pol IIIcore DNA binding is in contrast to another *E. coli* DNA

polymerase, Pol I, which is involved in Okazaki fragment maturation (Fig. 3.7a, c). The Klenow fragment results from proteolysis of Pol I and consists of the polymerase and 3'-5' exonuclease domains while the full Klenow fragment contains an additional 5'-3' exonuclease domain[2 18]. We measured a lifetime of 42.2 ± 1.8 s for the Pol I Klenow fragment on DNA (Fig. 3.7d) which is in agreement with previous studies reporting a high affinity for its binding to DNA[19 22 4].

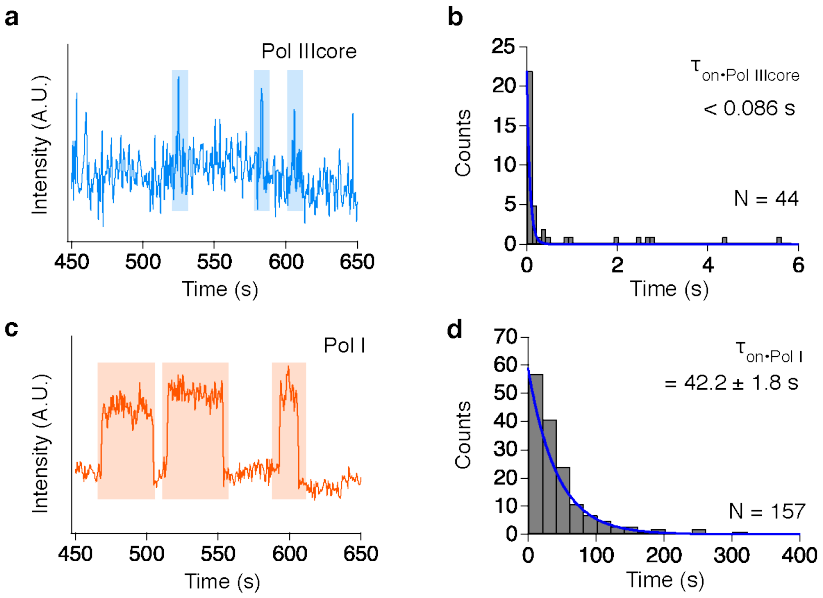


Figure 3.7. Pol IIIcore and Pol I binding to DNA. a, Representative trace showing short-lived Pol IIIcore binding events on DNA. **b**, The distribution of lifetimes for Pol IIIcore on DNA. **c**, Representative trace showing Pol I binding to DNA. **d**, The distribution of lifetimes for Pol I on DNA. All lifetimes are represented as mean ± s.e.m.

The behavior of Pol IIIcore on DNA changes dramatically when it binds to a clamp loaded onto DNA. Pol IIIcore associates with the clamp shortly after the release of the loader (Fig. 3.8a) and binds to the clamp-DNA for 15.7 ± 1.1 s (Fig. 3.8b). Pol IIIcore contains two β binding sequences: one located in the polymerase subunit α (QADMF, residues 920-924) and the other in the exonuclease subunit ϵ (QTSMAF, residues 182–187). The β binding sequence in the α subunit is absolutely required for clamp binding[5], while the sequence in the ϵ subunit stabilizes the Pol IIIcore complex on the clamp and has been shown to stimulate processive DNA synthesis and exonuclease activity[28 14]. In order to show that the Pol IIIcore binding events we see are occurring on the clamp, and that the β binding motif in the ϵ subunit indeed contributes to the lifetime of Pol IIIcore on the clamp, we measured the lifetime of Pol IIIcore on clamp-DNA using two variants of the ϵ subunit. As shown in Fig. 3.8c and d, the lifetime of Pol IIIcore on clamp-DNA increases when the β binding sequence in the ϵ subunit is enhanced (QTSMAF to QTSLPL)[28] and decreases when the sequence is weakened (QTSMAF to QTSAAA)[7].

To show how the hand-over from the clamp loader to Pol IIIcore occurs in the presence of the τ clamp loader complex, we pre-assembled the τ clamp loader complex ($\tau_3\delta_1\delta'_1$) with Pol IIIcore. Here, we see that the clamp, loader and Pol IIIcore bind to the DNA together (Fig. 3.8e). Different to the γ clamp loader complex, the τ clamp loader complex does not dissociate immediately once clamp loading has occurred. Instead, it remains bound for 14.8 $\pm$ 0.9 s and leaves together with Pol IIIcore (Fig. 3.8f). The presence of the τ clamp loader complex does not affect the lifetime of of the Pol IIIcore on the DNA substrate (3.8b and f).

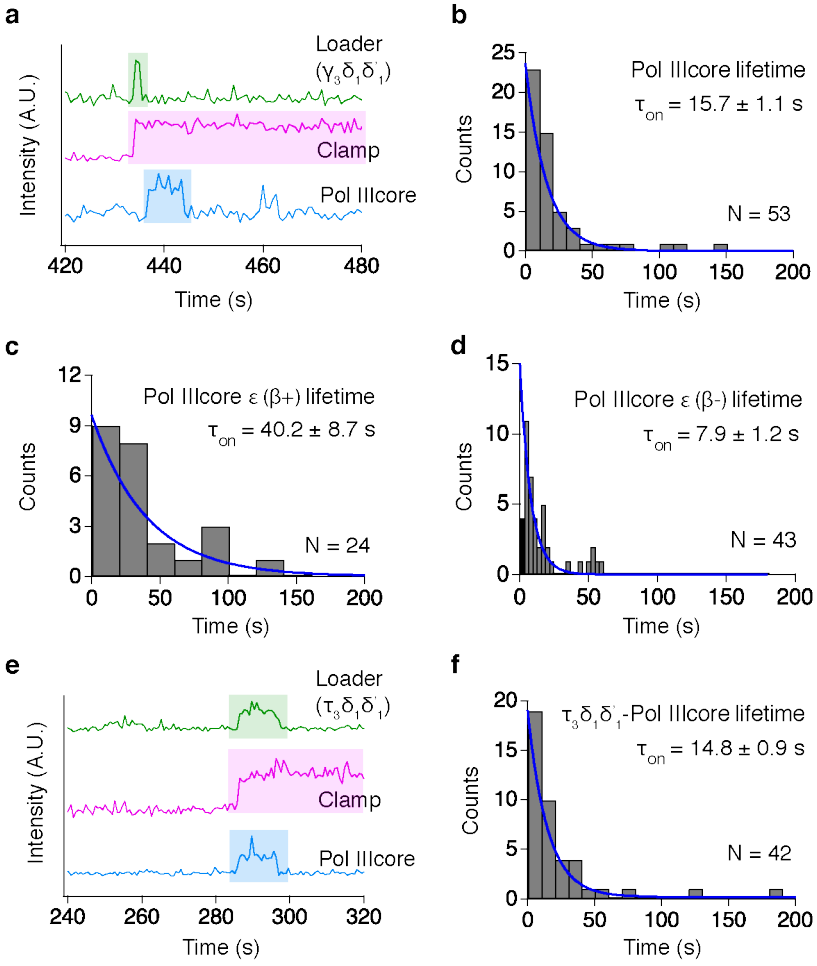


Figure 3.8. Polymerase loading onto clamp-DNA. a, Representative trace showing the binding of Pol IIIcore to clamp-DNA shortly after the release of the γ clamp loader. **b**, The distribution of lifetimes for Pol IIIcore on clamp-DNA. **c**, The lifetime of Pol IIIcore with an improved and **d**, weakened β binding motif in the ϵ subunit. **e**, Representative trace showing the simultaneous arrival of the clamp, loader and Pol IIIcore when using the τ clamp loader complex that was pre-assembled with Pol IIIcore. **f**, The distribution of lifetimes for the τ clamp loader-Pol IIIcore complex on clamp-DNA. All lifetimes and lag times are represented as mean $\pm$ s.e.m.

3.3.5 Polymerase idling and extension

The 15.7 $\pm$ 1.1 s lifetime of the stationary Pol IIIcore on clamp (Fig. 3.8b) seems short compared to the high processivity of the replisome during DNA synthesis (100 Kbp per binding event)[30 27]. Therefore, we tested the effect of nucleotides on the lifetime of Pol IIIcore on DNA. Given that the DNA substrate we used for the CoSMoS experiments has a short 6-nucleotide single-stranded overhang, we used only two of the four nucleotides (dATP and dTTP, 0.5 mM each) in the experiments. This is to prevent Pol IIIcore from dissociating at the end of the DNA substrate upon completing extension of the primer strand. The omission of two nucleotides is often used in DNA replication assays to induce 'polymerase idling', where the polymerase is held at the primer terminus due to the opposing polymerase and exonuclease activities[10]. Surprisingly, our results show that addition of dATP and dTTP has little effect on the lifetime of Pol IIIcore on clamp-DNA (Fig. 3.8b, Fig. 3.9a).

Given the limited length of our DNA substrates, experiments following the extension of the Pol IIIcore on clamp-DNA were carried out in collaboration with Dr. Iddo Heller at the Free University (VU), Amsterdam using a laser tweezers setup[29]. Our preliminary data using a 8.4 kb *pkyb1* DNA show that Pol IIIcore synthesizes DNA in short bursts (Fig. 3.9b) which is in agreement with previous observations that an actively synthesizing Pol IIIcore on a 48.5 kb λ DNA has a lifetime of $\sim$ 10 s[14]. Therefore, the presence of nucleotides do not significantly affect the lifetime of Pol IIIcore on clamp-DNA.

3.4 Discussion

Our CoSMoS setup has allowed us to label multiple components in the replisome and visualize them directly at a single-molecule level. Our results show that DNA replication entails a highly ordered sequence of events from clamp loading to polymerase loading to clamp unloading with sequential handover of DNA from one factor to the next.

We show that the clamp loading and unloading are not forward and backward reactions of the same mechanism but rather, they are different processes with distinct organization. Uncoupling clamp loading from unloading may be important to prevent futile cycles of clamp loading and unloading which would be energetically costly for the cell. The uncoupling between clamp loading and unloading may also reflect the fact that loading needs to occur at the same time as the synthesis of DNA, whereas unloading is a post-replication process that cleans up the DNA by removing the remaining clamps.

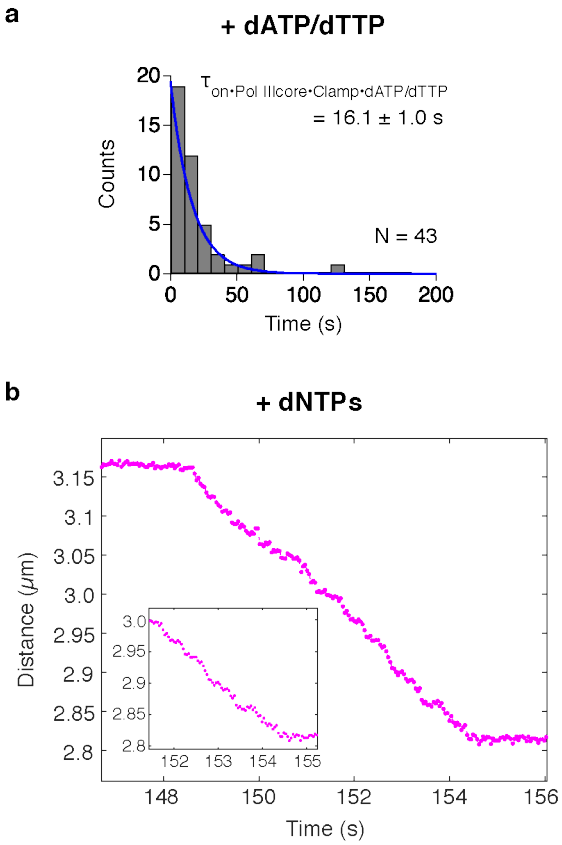


Figure 3.9. The effect of nucleotides on polymerase lifetime on clamp-DNA. a, The distribution of lifetimes for Pol IIIcore on clamp-DNA in the presence of dATP and dTTP represented as mean ± s.e.m. **b**, Representative trace showing segmented Pol IIIcore activity during active DNA synthesis on a 8.4 kb *pkyb1* DNA.

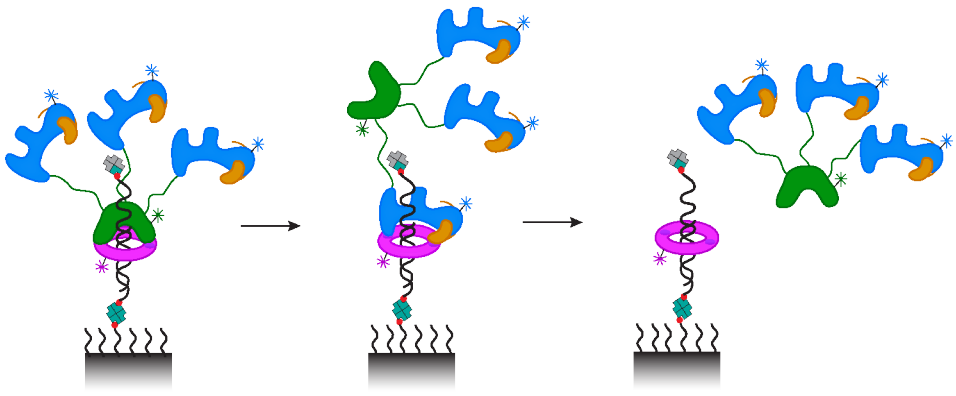


Figure 3.10. The τ clamp loader complex bridges the β clamp and Pol IIIcore. Cartoon representation of the switching of places between the τ clamp loader complex and Pol IIIcore on clamp-DNA.

Furthermore, our results show that the $\sim$ 10 s lifetime of Pol IIIcore on DNA is an intrinsic

property of the enzyme that is independent of whether it is stationary, idling or actively synthesizing DNA. However, during cellular replication Pol IIIcore is tethered to the replisome through the clamp loader protein τ. Our results show that the τ clamp loader complex mediates the immediate binding of the polymerase upon clamp loading by bridging the clamp and the polymerase. The τ clamp loader complex remains tethered to Pol IIIcore after the clamp has been loaded, which may represent the switching of places from the τ clamp loader complex to Pol IIIcore on the clamp-DNA (Fig. 3.10).

During the simultaneous synthesis of the leading and lagging strands, the polymerases on the two strands are linked together by the τ clamp loader complex. This prevents the polymerase from diffusing away and enables it to resume DNA synthesis upon dissociation from the DNA, which may explain the higher processivity measured for the replisome compared to the activity bursts of a single Pol IIIcore[30,26]. Having Pol IIIcore tethered to the replisome but not stably bound to DNA allows it to synthesize DNA with great processivity while maintaining the ability to exchange with other factors that have diverse roles on DNA (see Chapter 4).

Bibliography

[1] Gregory D Bowman, Mike O'Donnell, and John Kuriyan. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature*, 429(6993):724–730, June 2004.

[2] D Brutlag, M R Atkinson, P Setlow, and A Kornberg. An active fragment of DNA polymerase produced by proteolytic cleavage. *Biochemical and Biophysical Research Communications*, 37(6):982–989, December 1969.

[3] P M Burgers, A Kornberg, and Y Sakakibara. The dnaN gene codes for the beta subunit of DNA polymerase III holoenzyme of escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 78(9):5391–5395, September 1981.

[4] Thomas D Christian, Louis J Romano, and David Rueda. Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proceedings of the National Academy of Sciences*, 106(50):21109–21114, December 2009.

[5] Paul R Dohrmann and Charles S McHenry. A bipartite polymerase-processivity factor interaction: only the internal beta binding site of the alpha subunit is required for processive replication by the DNA polymerase III holoenzyme. *Journal of Molecular Biology*, 350(2):228–239, July 2005.

[6] P J Fay, K O Johanson, C S McHenry, and R A Bambara. Size classes of products synthesized processively by two subassemblies of Escherichia coli DNA polymerase III holoenzyme. *The Journal of Biological Chemistry*, 257(10):5692–5699, May 1982.

[7] Rafael Fernandez-Leiro, Julian Conrad, Sjors Hw Scheres, and Meindert H Lamers. cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *eLife*, 4:459, October 2015.

[8] Larry J Friedman and Jeff Gelles. Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods*, 86:27–36, September 2015.

[9] Roxana E Georgescu, Seung-Sup Kim, Olga Yurieva, John Kuriyan, Xiang-Peng Kong, and Mike O'Donnell. Structure of a sliding clamp on DNA. *Cell*, 132(1):43–54, January 2008.

[10] Roxana E Georgescu, Grant D Schauer, Nina Y Yao, Lance D Langston, Olga Yurieva, Dan Zhang, Jeff Finkelstein, and Mike E O'Donnell. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *eLife*, 4:e04988, April 2015.

[11] Eric R Goedken, Marcia Levitus, Aaron Johnson, Carlos Bustamante, Mike O'Donnell, and John Kuriyan. Fluorescence Measurements on the E.coli DNA Polymerase Clamp Loader: Implications for Conformational Changes During ATP and Clamp Binding. *Journal of Molecular Biology*, 336(5):1047–1059, March 2004.

[12] Mark Hedglin, Ravindra Kumar, and Stephen J Benkovic. Replication clamps and clamp loaders. *Cold Spring Harbor Perspectives in Biology*, 5(4):a010165–a010165, April 2013.

[13] Mark Howarth, Daniel J-F Chinnapen, Kimberly Gerrow, Pieter C Dorrestein, Melanie R Grandy, Neil L Kelleher, Alaa El-Husseini, and Alice Y Ting. A monovalent streptavidin with a single femtomolar biotin binding site. *Nature Methods*, 3(4):267–273, April 2006.

[14] Slobodan Jergic, Nicholas P Horan, Mohamed M Elshenawy, Claire E Mason, Thitima Urathamakul, Kiyoshi Ozawa, Andrew Robinson, Joris M H Goudsmits, Yao Wang, Xuefeng Pan, Jennifer L Beck, Antoine M van Oijen, Thomas Huber, Samir M Hamdan, and Nicholas E Dixon. A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *The EMBO Journal*, 32(9):1322–1333, May 2013.

[15] D Jeruzalmi, M O'Donnell, and J Kuriyan. Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. *Cell*, 106(4):429–441, August 2001.

[16] Brian A Kelch, Debora L Makino, Mike O'Donnell, and John Kuriyan. How a DNA polymerase clamp loader opens a sliding clamp. *Science*, 334(6063):1675–1680, December 2011.

[17] Brian A Kelch, Debora L Makino, Mike O'Donnell, and John Kuriyan. Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biology*, 10(1):34, April 2012.

[18] H Klenow and I Henningsen. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. *Proceedings of the National Academy of Sciences of the United States of America*, 65(1):168–175, January 1970.

[19] R D Kuchta, V Mizrahi, P A Benkovic, K A Johnson, and S J Benkovic. Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry*, 26(25):8410–8417, December 1987.

[20] F P Leu, M M Hingorani, J Turner, and M O'Donnell. The delta subunit of DNA polymerase III holoenzyme serves as a sliding clamp unloader in Escherichia coli. *The Journal of Biological Chemistry*, 275(44):34609–34618, November 2000.

[21] Mark P A Luna-Vargas, Evangelos Christodoulou, Andrea Alfieri, Willem J van Dijk, Magda Stadnik, Richard G Hibbert, Danny D Sahtoe, Marcello Clerici, Valeria De Marco, Dene Littler, Patrick H N Celie, Titia K Sixma, and Anastassis Perrakis. Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. *Journal of Structural Biology*, 175(2):113–119, August 2011.

[22] R P Markiewicz, K B Vrtis, D Rueda, and L J Romano. Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. *Nucleic Acids Research*, 40(16):7975–7984, September 2012.

[23] C S McHenry and W Crow. DNA polymerase III of Escherichia coli. Purification and identification of subunits. *The Journal of Biological Chemistry*, 254(5):1748–1753, March 1979.

[24] Kyle R Simonetta, Steven L Kazmirski, Eric R Goedken, Aaron J Cantor, Brian A Kelch, Randall McNally, Steven N Seyedin, Debora L Makino, Mike O'Donnell, and John Kuriyan. The Mechanism of ATP-Dependent Primer-Template Recognition by a Clamp Loader Complex. *Cell*, 137(4):659–671, May 2009.

[25] A Spanos, S G Sedgwick, G T Yarranton, U Hübscher, and G R Banks. Detection of the catalytic activities of DNA polymerases and their associated exonucleases following SDS-polyacrylamide gel electrophoresis. *Nucleic Acids Research*, 9(8):1825–1839, April 1981.

[26] Nathan A Tanner, Samir M Hamdan, Slobodan Jergic, Karin V Loscha, Patrick M Schaeffer, Nicholas E Dixon, and Antoine M van Oijen. Single-molecule studies of fork dy-

namics in Escherichia coli DNA replication. *Nature Structural & Molecular Biology*, 15(2):170–176, February 2008.

[27] Nathan A Tanner, Joseph J Loparo, Samir M Hamdan, Slobodan Jergic, Nicholas E Dixon, and Antoine M van Oijen. Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic Acids Research*, 37(4):e27, March 2009.

[28] Ana Toste Rêgo, Andrew N Holding, Helen Kent, and Meindert H Lamers. Architecture of the Pol III-clamp-exonuclease complex reveals key roles of the exonuclease subunit in processive DNA synthesis and repair. *The EMBO Journal*, 32(9):1334–1343, May 2013.

[29] G J Wuite, R J Davenport, A Rappaport, and C Bustamante. An integrated laser trap/flow control video microscope for the study of single biomolecules. *Biophysical Journal*, 79(2):1155–1167, August 2000.

[30] Nina Y Yao, Roxana E Georgescu, Jeff Finkelstein, and Michael E O'Donnell. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proceedings of the National Academy of Sciences*, 106(32):13236–13241, August 2009.

[31] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nature Protocols*, 1(1):280–285, 2006.

[32] Jun Yin, Paul D Straight, Shaun M McLoughlin, Zhe Zhou, Alison J Lin, David E Golan, Neil L Kelleher, Roberto Kolter, and Christopher T Walsh. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15815–15820, November 2005.

Chapter 4

Studying *E.coli* translesion synthesis by CoSMoS

4.1 Introduction

The mechanism of exchange between the replicative and translesion DNA polymerases has been dominated by the toolbelt model since its proposal in 2005[10]. In this study, bulk primer extension assays were used to show that Pol IV can replace a stalled Pol III on the β clamp, and Pol III can regain the primer-template from Pol IV. In addition, a FRET donor and acceptor was placed on Pol IV and Pol IIIα respectively and a FRET signal was only observed in the presence of clamp. Taken together, it was proposed that during replication, a replicative and a translesion DNA polymerase can each bind to one binding pocket on the clamp (the clamp is a dimer, with one hydrophobic binding pockets on each monomer, see Chapter 1). When the replicative polymerase is stalled at a lesion it temporarily switches place with the translesion polymerase, and once the lesion has been bypassed the replicative polymerase regains control of replication. Since this study, other studies have also shown that Pol IV[7] and Pol II[9] are able to replace an actively synthesizing Pol III replisome in bulk primer extension assays. Although the toolbelt model is very attractive, thus far there has not been any definitive proof for it. A caveat in the studies above supporting the toolbelt model is that they all use bulk methods to study the exchange between Pol II/IV and Pol IIIα. The averaging over many asynchronous events in a bulk experiment can only show that the translesion and replicative polymerases are capable of switching, but not the exact molecular mechanism of how the switching is achieved.

More recently, the exchange between Pol IIIα and Pol IV/Pol II has also been studied at the single-molecule level[15,14]. These studies show that Pol III, Pol IV and Pol II synthesize DNA at

different rates and use the change in the rate of DNA synthesis to infer the switching between the polymerases. However, given that the exchange between the polymerases is measured indirectly through the rate of DNA synthesis, it cannot be determined whether the polymerases bind simultaneously.

Recent studies have also revealed that during replication, the binding pockets on the clamp are occupied by Pol IIIα and ϵ[17 13]. The cryo-EM structure of the trimeric Pol IIIα-exonuclease-clamp complex shows that Pol IIIcore occupies the entire surface of the clamp and would therefore occlude the binding of a translesion DNA polymerase to the clamp during normal DNA synthesis[5]. However, a crystal structure of the Pol IV little finger domain on the clamp suggests that Pol IV can make secondary interactions with the rim of the clamp in addition to binding to the canonical binding site on the clamp[1]. It was proposed that this secondary interface keeps Pol IV in an inactive conformation by tethering it to the rim of the clamp and thereby regulates the switching between the replicative and translesion polymerases.

In summary, it remains controversial whether a replicative and translesion DNA polymerase can bind to the clamp simultaneously. This question can be uniquely addressed by CoSMoS, which allows the direct visualization of the replicative and translesion DNA polymerases. Therefore, this chapter aims to study the exchange between the different DNA polymerases in *E.coli*.

4.2 Materials and Methods

The cloning, expression, purification and labeling of many of the proteins used in this chapter are presented in Section 3.2 of Chapter 3. The preparation of proteins specifically used in this chapter are described below.

4.2.1 Cloning of protein expression vectors

Genes for Pol II (*polB*) and Pol IV (*dinB*) were cloned into pET11 vectors with the Ybbr sequence DSLEFIASKLA[18] added to the protein N-terminally. The β clamp binding mutants of Pol IV and Pol II were generated though site directed mutagenesis: residues 346-351 were mutated from QLVLGL to QLVAGA for Pol IV β groove mutant, residues 303-305 were mutated from VWP to AGA for Pol IV β rim mutant, and residues residues 779-783 were mutated from QLGLF to QLGAA for Pol II β groove mutant.

The gene for ϵ (*dnaQ*) was cloned into a pET28a vector with the Ybbr sequence DSLEFI-

ASKLA[18] added to the protein N-terminally.

4.2.2 Protein expression and purification

All columns were purchased from GE Healthcare. Unless otherwise specified, protein purifications were performed with the following gradients using the columns stated: nickel affinity using Histrap HP (25-500 mM Imidazole gradient in the presence of 500 mM NaCl), ion exchange using Hitrap Q HP for anion exchange and Hitrap SP HP for cation exchange (0-1 M NaCl gradient), Hitrap Phenyl HP for hydrophobic interaction chromatography (2-0 M ammonium sulfate gradient) and a 120 ml Superdex 200 column for gel filtration (150 mM NaCl).

Pol II-Nybbr and Pol II (β groove)-Nybbr were expressed in *E. coli* BL21 (DE3) Gold cells for 3 hours at 30 °C. The proteins were purified in 20 mM Tris pH 8.0, 0.5 mM EDTA and 2 mM DTT by anion exchange, phenyl hydrophobic chromatography, cation exchange and gel filtration.

Pol IV-Nybbr, Pol IV (β groove)-Nybbr and Pol IV (β rim)-Nybbr were expressed in *E. coli* BL21 (DE3) pLysS cells for 3 hours at 30 °C. The proteins were purified in 20 mM Hepes pH 7.5 and 2 mM DTT by cation exchange, phenyl hydrophobic chromatography and gel filtration.

ϵ-Nybbr was expressed in *E. coli* BL21 (DE3) Gold cells for 2.5 hours at 30 °C, and purified in 25 mM Hepes pH 8.2 and 2 mM DTT by nickel affinity under denaturing conditions (in the presence of 6 M Urea), followed by refolding overnight at 4 °C in 25 mM Hepes pH 8.2 and 10 mM DTT, and anion exchange.

4.2.3 Protein labeling

All gel filtration experiments in this section were carried out using a 2.4 ml Superdex 75 column (GE Healthcare), with the exception for Pol II, which is larger in size (90 kDa) and therefore was run a 2.4 ml Superdex 200 Increase column (GE Healthcare). The columns were equilibrated in 20 mM Tris-HCl pH 7.5, 50 mM potassium glutamate and 2 mM DTT.

Pol II was labeled at a single site in the N-terminus using the Sfp method[18]. Atto 647N fluorophore was conjugated to Coenzyme A as described[18] and Pol II was labeled in a reaction containing 10 mM $MgCl_2$, 50 mM Hepes 7.5, 10 mM DTT, 2 μM Sfp, 100 μM CoA-Atto 647N and 50 μM protein and the reaction was incubated overnight at 4 °C. The labeled protein was purified away from the free dye by gel filtration and the labeling efficiency was determined spectrophotometrically using the protein and fluorophore absorption ratios, with the free fluorophore absorption at 280 nm subtracted from the protein absorption at 280 nm. The labeling

efficiency of Pol II was 66 %.

Pol IV was labeled at the N-terminus using the Sfp method[18]. The labeling reaction was identical to that used to label Pol II except using CoA-Atto 565. The protein was labeled for 2 hours at 4 °C and the labeled protein was purified by gel filtration. The labeling efficiency for Pol IV was determined spectrophotometrically to be 62 %.

ϵ was labeled at the N-terminus using the Sfp method[18]. The protein was labeled for 2 hours at 4 °C using CoA-Atto 565 and the labeled protein was purified by gel filtration. The labeling efficiency was determined spectrophotometrically to be 78 %. The labeling efficiency for ϵ was verified at the single-molecule level by measuring the co-localization frequencies between labeled Pol IIIα (see Chapter 3) and labeled ϵ on clamp-DNA. Since the two proteins are known to form a tight complex, each co-localization event was scored as having labeled Pol IIIα only, labeled ϵ only, or both labeled proteins. Measuring the labeled efficiency this way showed 71 % labeling for ϵ, which is similar to the efficiency measured by absorption.

4.2.4 ϕX174 primer extension assays

Activity of the labeled proteins was tested using a single-stranded ϕX174 phage DNA (New England Biolabs) annealed to a primer with a 5' label (5'FAM-ACCAACATAAACATTATT-GCCCGGCGTACpG, where lowercase 'p' indicates a non-cleavable phosphothioate bond). Reactions were performed in 20 mM Tris pH 7.5, 2 mM DTT, 50 mM potassium glutamate, 8 mM magnesium acetate, and 0.05 mg/ml BSA. Each reaction contained 5 nM primed ϕX174 phage DNA, 50 nM β clamp, 10 nM γ clamp loader complex ($\gamma_3\delta_1\delta'_1$), and 30 nM Pol II or Pol IV. Reactions were stopped at 0, 0.5, 1, 2 and 5 minutes using a stop buffer containing 75 mM EDTA and 0.6 % (w/v) SDS, and were separated on an alkaline agarose gel (0.8 % agarose, 30 mM NaOH, 2 mM EDTA) for 15 hours at 14 V. Gels were scanned at 488 nM using a Amersham Typhoon (GE LifeSciences).

Results of the ϕX174 primer extension assays are presented in Section 2.3.4 of Chapter 2.

4.2.5 DNA substrate

The matched DNA substrate was prepared as described in section 3.2 of Chapter 3. The mismatched (C:T) substrate was constructed using the same streptavidin bound, Atto 488 labeled template DNA annealed to a 25-nt primer with the sequence 5' Biotin-GGATTTGGATGAAGG-TGAAGCATGpT 3' (IDT), where lowercase 'p' indicates the non-cleavable phosphothioate bond. The lesion substrate was constructed using a furfuryl-modified template with the se-

quence 5' Biotin-CATAA[N2-furfuryl-dG]ATCCATGCTTCACC[amino-dT]TCATCCAAAT-CC 3' (Eurogentec), which was bound to monovalent streptavidin through the 5' biotin moiety and labeled at the internal amino-dT modification using NHS-ester activated Atto 488 fluorophore as described for the non-lesion-containing template in Chapter 3. The primer used for the lesion substrate is the same as that used for the matched DNA substrate. The matched, mismatched and lesion DNA substrates are illustrated in (Fig. 4.1).

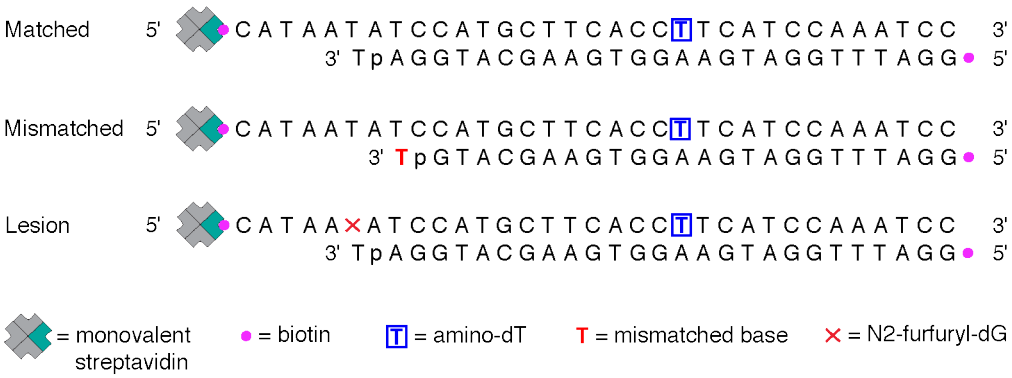


Figure 4.1. DNA substrates used in this Thesis. Nucleotide sequences of the matched, mismatched and lesion DNA substrates.

4.2.6 TIRF microscopy

Details of the TIRF microscope setup can be found in section 3.2 of Chapter 3. For the Pol III and Pol II/Pol IV co-localization experiments, the reaction contained 15 nM β clamp; 15 nM γ complex; 30 nM or 150 nM Pol IIIcore; 6 nM, 30 nM or 150 nM Pol IV; 30 nM Pol II and 0.5 mM ATP. Experiments in the presence of nucleotides were conducted with 0.5 mM of each of dATP and dTTP. The γ complex was unlabeled since we are limited to three laser channels per experiment.

4.3 Results

4.3.1 Protein labeling

In order to able to quantify the number of interacting molecules in an experiment, we aimed to label the proteins at a single site.

Pol II contains 7 cysteines and thus it would require multiple rounds of mutagenesis to obtain a protein construct for single site labeling using maleimide activated fluorophores. Therefore,

we chose the Sfp method for its labeling. Three different labeling conditions were tested, with 1 hour at room temperature yielding 46 %, and 2 hours at room temperature yielding 51 % labeling efficiency. Since the efficiency increased with the duration of labeling, the final labeling condition was chosen to be overnight at 4 °C, which yielded 66 % labeling efficiency. The labeling reaction was run on gel filtration to separate the labeled protein (at retention volume 1.2 ml) from free dye (at retention volume 2.3 ml) (Fig. 4.2a), and to assess for correct protein folding, shown by the same retention volume of the labeled protein compared to the protein prior to labeling.

Pol IV contains 6 cysteines and therefore the Ybbr method was the most straightforward way to achieve single site labeling. Reaction with CoA-Atto 565 at 4 °C for 2 hours yielded 62 % labeling while further overnight incubation at 4 °C resulted in the loss of the protein peak, suggesting that the protein is unstable when labeled for prolonged time. The 2-hour labeling reaction was analyzed by gel filtration, which separated the labeled protein (at retention volume 1.3 ml) from free dye (at retention volume 1.8 ml) (Fig. 4.2b). The gel filtration analysis was also used to assess for correct folding of the labeled protein, as shown by the same retention volume of the protein before and after labeling.

The exonuclease ϵ contains 3 cysteines and was labeled on the Ybbr tag as this construct was readily available in the lab. Labeling with CoA-Atto 565 at 4 °C for 1 hour yielded 42 % labeling and 2 hours yielded 78 % labeling. The 2-hour labeling reaction was injected onto a gel filtration column to separate the labeled protein (at retention volume 1.25 ml) from the free dye (at retention volume 1.8 ml) and to assess for correct protein folding (Fig. 4.2c).

4.3.2 Pol IV and Pol IIIcore exchange

To study whether Pol IIIcore and Pol IV can co-localize on the clamp, we followed the binding of Pol IIIcore and Pol IV to clamp-DNA. To do this, we loaded the β clamp onto DNA using the γ clamp loader complex and imaged the polymerases and the clamp over time.

At equimolar concentrations of the two polymerases (30 nM each), we observed a majority of events (92 % of 77 events) in which Pol IIIcore and Pol IV alternate binding to the clamp-DNA. 70 % of these events represents the switching from Pol IIIcore to Pol IV and 22 % represents the switching from Pol IV to Pol IIIcore (Fig. 4.3a). In some events, multiple switches between Pol IIIcore and Pol IV were seen (Fig. 4.3b). Likewise to the lifetime of Pol IIIcore on clamp-DNA shown in Chapter 3, the exchange between the polymerases was not affected by the addition of nucleotides. In the presence of 0.5 mM each of dATP and dTTP, the

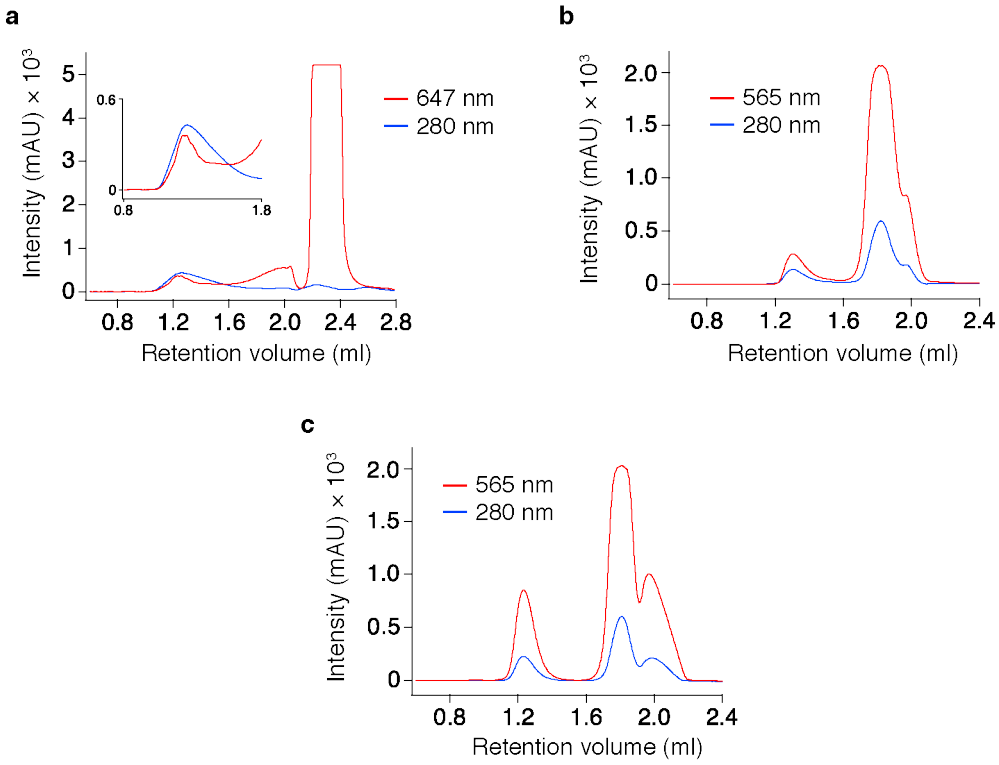


Figure 4.2. Labeling of Pol II, Pol IV and exonuclease ϵ. a, Gel filtration run of the Pol II **b**, Pol IV and **c**, exonuclease ϵ labeling reactions.

switch from Pol IIIcore to Pol IV accounts for 72 % of events and the switch from Pol IV to Pol IIIcore accounts for 21 % of the events.

In the Pol IIIcore to Pol IV polymerase switching events, there is a lag time between the release of Pol IIIcore and the arrival of Pol IV (Fig. 4.3b). This lag time decreases when the concentration of Pol IV in the experiments is increased (Fig. 4.4a-d), showing that the binding of the proteins to clamp-DNA is concentration-dependent. However, at all protein concentrations tested, the lifetime of Pol IIIcore on clamp-DNA remains unaltered at $\sim$ 16 seconds (Fig. 4.4e-h). This shows that Pol IV binding does not cause Pol IIIcore release, and therefore the two polymerases bind to the clamp-DNA independently.

In addition to the alternating binding between the polymerases that represents the majority of the events, we observed a small number of co-localization events between Pol IIIcore and Pol IV. When Pol IIIcore and Pol IV are present at equimolar concentrations (30 nM each), the co-localization between the two polymerases accounts for 9 % of the events (Fig. 4.3a). The co-localization events become more frequent at higher protein concentrations (Table 2).

Closer inspection of the co-localization events between Pol IIIcore and Pol IV shows that the two polymerases arrive and leave independent of one another and only co-localize on passing

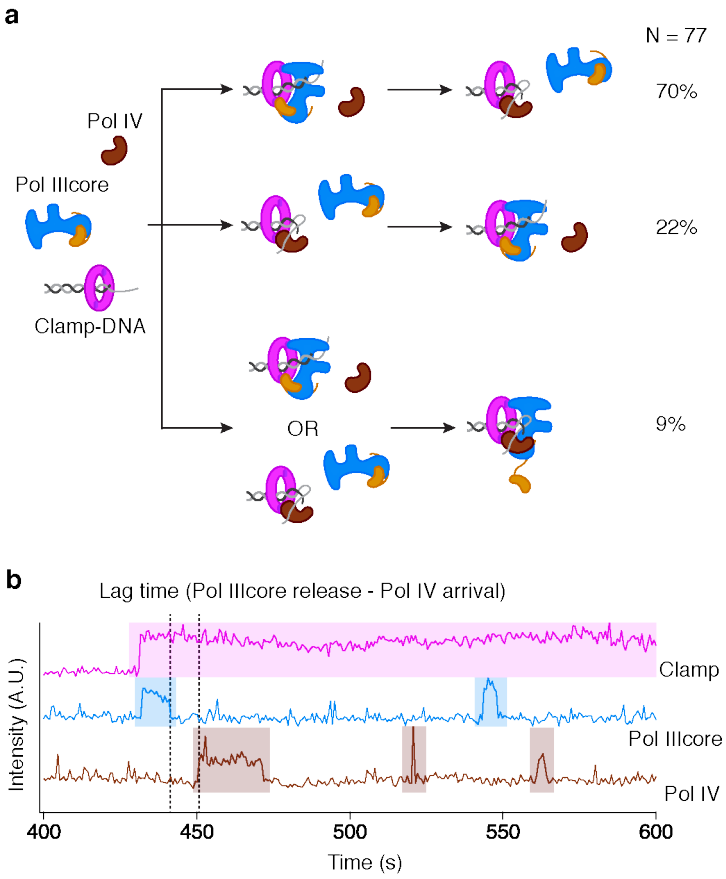


Figure 4.3. Exchange between Pol IV and Pol IIIcore on clamp-DNA. a, Cartoon representation of the frequency of each event type: Pol IIIcore to Pol IV switch (top), Pol IV to Pol IIIcore switch (middle) and Pol IIIcore-Pol IV co-localization (bottom). **b**, Representative trace showing alternating binding of Pol IIIcore and Pol IV on clamp-DNA.

(Fig. 4.5a). This is clearly different to the co-localization between Pol IIIα and ϵ, which are known to form a stable complex on the β clamp[17 13]. In an experiment using labeled Pol IIIα and labeled ϵ, we show that they arrive and depart from the clamp-DNA simultaneously (Fig. 4.5b). In addition, the co-localization time between Pol IIIcore and Pol IV is shorter than that between Pol IIIα and ϵ: 8.2 $\pm$ 0.6 s for Pol IIIcore and Pol IV co-localization compared to 15.8 $\pm$ 0.9 s for Pol IIIα and ϵ co-localization (Fig. 4.5c-d).

Taken together, our data shows that Pol IIIcore and Pol IV do not form a stable complex on clamp-DNA but that the two polymerases compete for binding to the β clamp in a concentration-dependent manner. This competition is strongly favoured towards Pol IIIcore in the presence of the τ clamp loader complex (Table 2). As shown in Chapter 3, the τ clamp loader is tethered to Pol IIIcore and immediately places it on the clamp upon clamp loading. Consequently, the exchange between Pol IIIcore and Pol IV at equimolar concentrations is dominated by Pol IIIcore to Pol IV switches (Table 2).

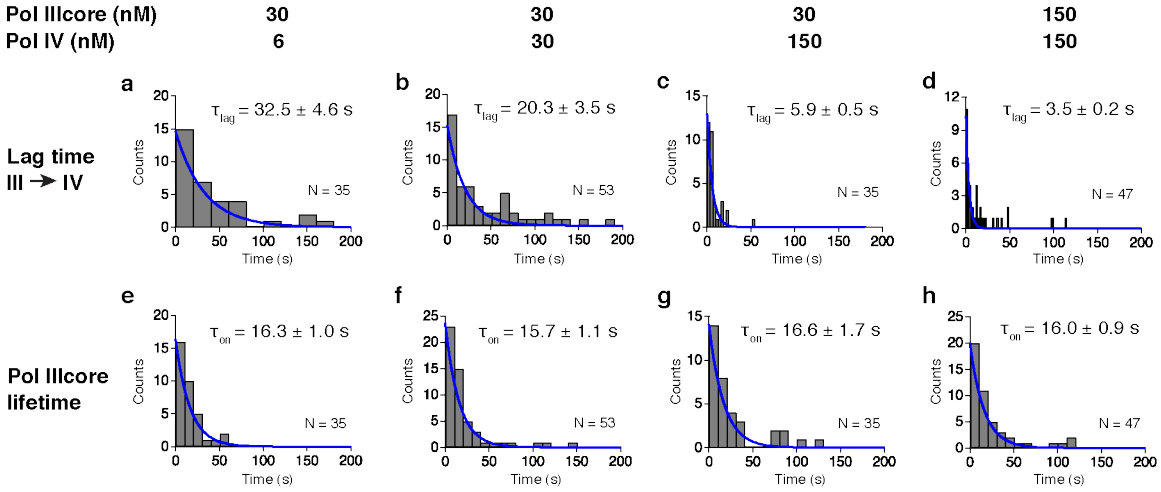


Figure 4.4. The lag time and the lifetime of Pol IIIcore during Pol IIIcore and Pol IV polymerase exchange at different protein concentrations. **a-d**, The lag time between Pol IIIcore release and Pol IV binding on clamp-DNA at increasing protein concentrations. **e-h**, The lifetime of Pol IIIcore on clamp-DNA at all four protein concentrations. All lifetimes are represented as mean ± s.e.m.

Table 2. Competition of Pol IIIcore and Pol IV

	Concentration (nM)		Polymerase exchange (%)			Lag time (s) [1]	Lifetime (s) [2]
Competition	Pol IIIcore	Pol IV/II	III→IV/II	IV/II→III	III+IV/II	III→IV/II	Pol IIIcore
Pol IIIcore - Pol IV	30	6	81	12	7	32.5 ± 4.6	16.3 ± 1.0
	30	30	70	22	9	20.3 ± 3.5	15.7 ± 1.1
	30	150	63	20	15	5.9 ± 0.5	16.6 ± 1.7
	150	150	51	24	26	3.5 ± 0.2	16.0 ± 0.9
τ-complex [3] - Pol IV	30 [4]	30	95	0	5	11.3 ± 1.3	14.8 ± 0.9

[1] Time between Pol IIIcore release and Pol IV arrival. [2] Lifetime on clamp-DNA. [3] τ-complex consists of τ clamp loader ($\tau_3\delta_1\delta'_1$) and three Pol IIIcore complexes (α, ε, θ). [4] Concentration of Pol IIIcore.

4.3.3 Pol II and Pol IIIcore exchange

Next, we studied the exchange between Pol II and Pol IIIcore on clamp-DNA using the same experimental setup as for Pol IV and Pol IIIcore. At equimolar concentrations of the two polymerases (30 nM), Pol II competes with Pol IIIcore for binding to clamp-DNA, with some events showing multiple polymerase exchanges (Fig. 4.6). However, no co-localization events were observed between Pol II and Pol IIIcore. This may be explained by the larger size of Pol II compared to Pol IV (90 kDa versus 40 kDa), which may prevent it from simultaneously binding to the clamp with Pol IIIcore.

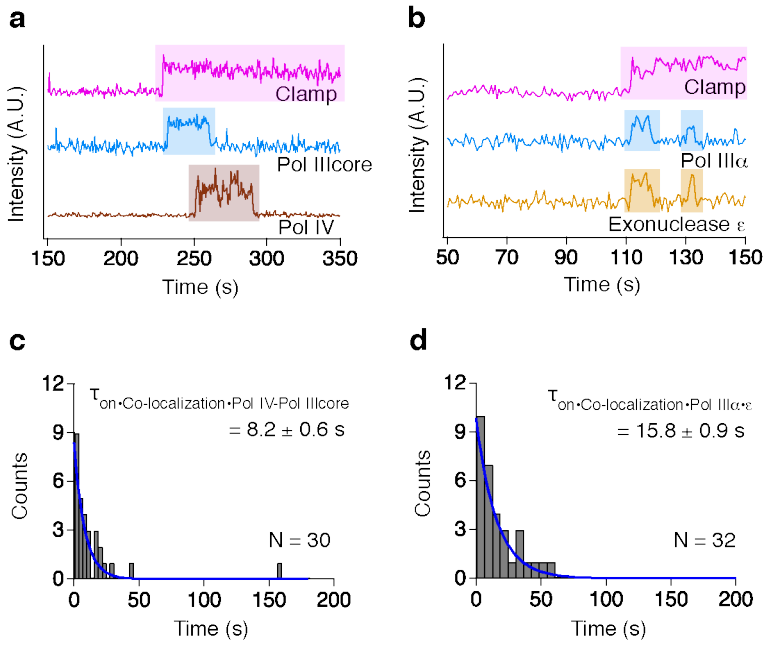


Figure 4.5. Co-localization between Pol IIIcore and Pol IV, and between Pol IIIα and ϵ. a, Representative trace showing the independent arrival and release of Pol IIIcore and Pol IV in the co-localization events. **b**, Representative trace showing the simultaneous arrival and release of Pol IIIα and ϵ during binding to clamp-DNA. **c**, Lifetime of the co-localization between Pol IIIcore and Pol IV. **c**, Lifetime of the co-localization between Pol IIIα and ϵ. All lifetimes are represented as mean $\pm$ s.e.m.

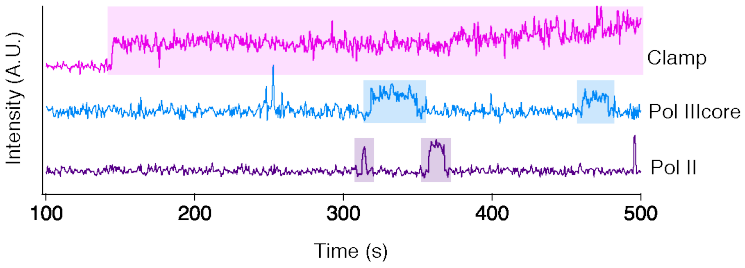


Figure 4.6. Exchange between Pol II and Pol IIIcore on clamp-DNA. Representative trace showing the alternating binding of Pol II and Pol IIIcore on clamp-DNA.

4.3.4 Polymerases compete for binding to the hydrophobic groove on the β clamp

To further investigate the competition between the polymerases on the β clamp, we created a number of polymerase mutants. Most of the clamp interacting proteins, including Pol IIIcore, Pol IV and Pol II, bind to a hydrophobic groove on the surface of the β clamp using the canonical sequence Qxx(L/M)xF[2]. The β clamp has two binding grooves due to its dimeric nature, and

as explained in Chapter 3, Pol IIIcore binds to both binding grooves, one via the β binding sequence in Pol IIIα and another via the β binding sequence in ϵ. Given that the sequence present in Pol IIIα has been shown to be crucial for clamp binding[4], the sequence in ϵ was mutated to enhance or weaken binding to the clamp in order to show that it also contributes to Pol IIIcore binding to the clamp. As presented in Section 3.3.4 of Chapter 3, the lifetime of Pol IIIcore is increased for the $\epsilon(\beta+)$ mutant and decreased for the $\epsilon(\beta-)$, showing that Pol IIIcore binds to both hydrophobic grooves on the clamp.

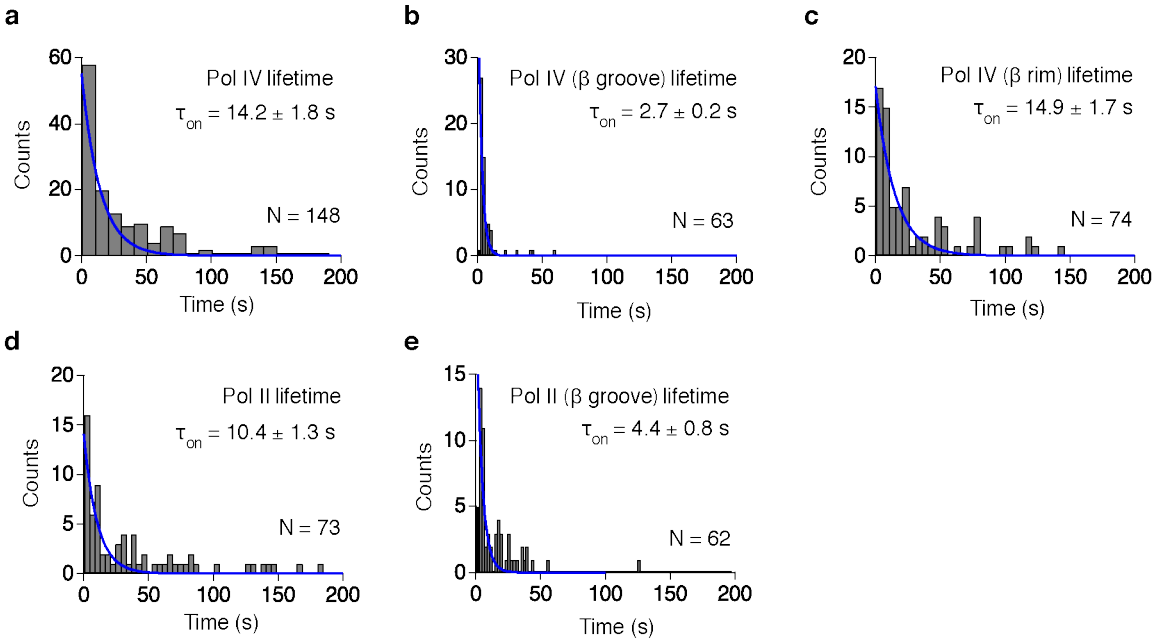


Figure 4.7. Lifetime of Pol IV and Pol II β binding mutants on clamp-DNA. a Lifetime of Pol IV wild-type, **b** β groove mutant, and **c** β rim mutant on clamp-DNA. **d** Lifetime of Pol II wild-type and **b** β groove mutant on clamp-DNA. All values represent mean lifetime $\pm$ s.e.m.

Two β clamp interacting motifs have been described for Pol IV: one is the canonical groove binding QLVLGL motif (residues 346-351), and the other is a rim binding sequence (VWP, residues 301-304) that interacts with the side of the clamp[18]. Previous studies have shown that mutating the groove binding motif in Pol IV inhibits clamp-dependent DNA synthesis, while mutating the rim contacts results in the loss of polymerase switching[8]. Our experiments measuring the lifetime of the Pol IV mutants on the clamp-DNA showed that mutating the groove binding sequence (QLVLGL to QLVAGA, residues 346-351) resulted in a significant reduction in the lifetime of binding (Fig. 4.7a-b). In contrast, mutation of the rim binding sequence (VWP to AGA, residues 303-305) had little effect on the lifetime of Pol IIIcore on clamp-DNA (Fig. 4.7a, c). The importance of Pol IV binding to the hydrophobic groove is further demonstrated by the loss of co-localization between Pol IIIcore and the Pol IV on clamp-

DNA for the β groove mutant, even at elevated concentrations of Pol IV, while the Pol IV β rim mutant shows little effect (Table 3).

Table 3. Lifetime of β-clamp binding mutants of Pol IIIcore, Pol IV and Pol II

			Polymerase exchange (%)		
Polymerase	Mutation	Lifetime (s) [1]	III→IV/II	IV/II→III	III+IV/II
Pol IIIcore	WT	15.7 ± 1.1	70	22	9
	ε (β-)	7.9 ± 1.2	70	26	3
	ε (β+)	40.2 ± 8.7	71	24	6
Pol IV	WT	14.2 ± 1.8	70	22	9
	β groove [2]	2.7 ± 0.2	40	60	0
	β rim	14.9 ± 1.7	66	29	5
Pol II	WT	10.4 ± 1.3	71	29	0
	β groove	4.4 ± 0.8	63	37	0

[1] Lifetime on clamp-DNA. [2] The Pol IV β cleft mutant was measured at high concentrations (90 nM) in an attempt to catch co-localization events.

Furthermore, mutating the β groove binding sequence in Pol II (QLGLF to QLGAA, residues 779-783) resulted in a decreased binding lifetime on the clamp (Fig. 4.7d-e).

The results therefore show that all three polymerases compete for binding to the same hydrophobic groove on the β clamp, and that the isolated polymerases compete with similar lifetimes ($\sim$ 10 seconds) on the clamp-DNA.

4.3.5 Pol IV recruitment

The lack of organization between Pol IIIcore and Pol IV raises the question of how Pol IV is recruited to a site of lesion. We wondered whether Pol IIIcore is released faster from DNA containing a lesion and therefore measured the lifetime of Pol IIIcore on the clamp on a DNA substrate containing N2-furfuryl-dG, a chemically modified guanine (Fig. 4.8a, b). This lesion was picked because it has previously been used to study translesion synthesis by Pol IV[12], and we confirmed that this substrate forms an effective block to Pol IIIcore but is bypassed by Pol IV using primer extension assays (Fig. 4.8c).

To our surprise, the lifetime of Pol IIIcore on clamp-DNA is not affected by the presence of the lesion: we measured 15.7 $\pm$ 1.1 s on the matched DNA substrate (Fig. 4.9a, also shown in Chapter 3) and 17.6 $\pm$ 2.1 s on the furfuryl DNA substrate (Fig. 4.9b). Furthermore, we tested the lifetime of Pol IIIcore on a mismatched DNA substrate and also did not observe a significant change in Pol IIIcore lifetime (19.0 $\pm$ 1.4 s, Fig. 4.9c) which confirms that Pol IIIcore does not

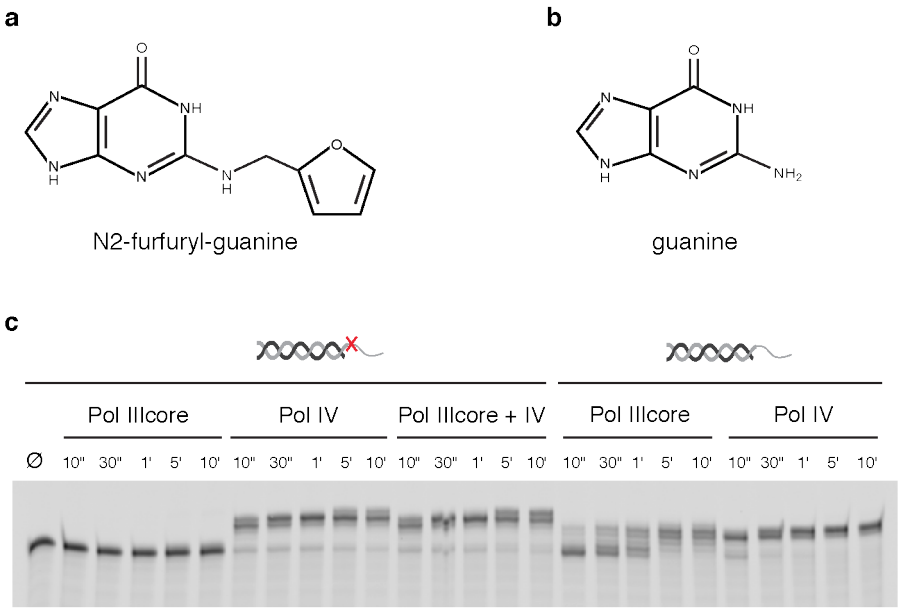


Figure 4.8. N2-furfuryl-dG lesion DNA substrate. **a**, Chemical structure of N2-furfuryl-guanine and **b**, guanine. **c**, Extension assays with the furfuryl lesion (red cross) and matched DNA substrates.

detect the differences between the DNA substrates.

In Chapter 3, we showed that addition of the two nucleotides dATP and dTTP (0.5 mM each) did not affect the lifetime of Pol IIIcore on the clamp on matched DNA, which is also shown in Fig. 4.9d. Similarly, the addition of dATP and dTTP does not significantly change the lifetime of Pol IIIcore on the clamp on the lesion and mismatched DNA substrates (Fig. 4.9e, f). The results therefore show that nucleotides do not facilitate the recognition of lesions or mismatches by Pol IIIcore.

Finally, also the exchange between Pol IIIcore and Pol IV is similar on the matched, mismatched and lesion DNA substrates in the absence of nucleotides, showing 13-18 % co-localization (Table 4).

Hence, Pol IIIcore dissociation is unaffected by the presence of the DNA lesion or mismatch, and Pol IV is not specifically recruited as a result of lesion recognition. Instead, Pol IV competes for binding when Pol IIIcore is released from DNA in a concentration-dependent manner.

4.4 Discussion

We have for the first time directly visualized the factors involved in *E.coli* translesion DNA synthesis and shown that in contrast to the highly organized sequence of events seen for DNA repli-

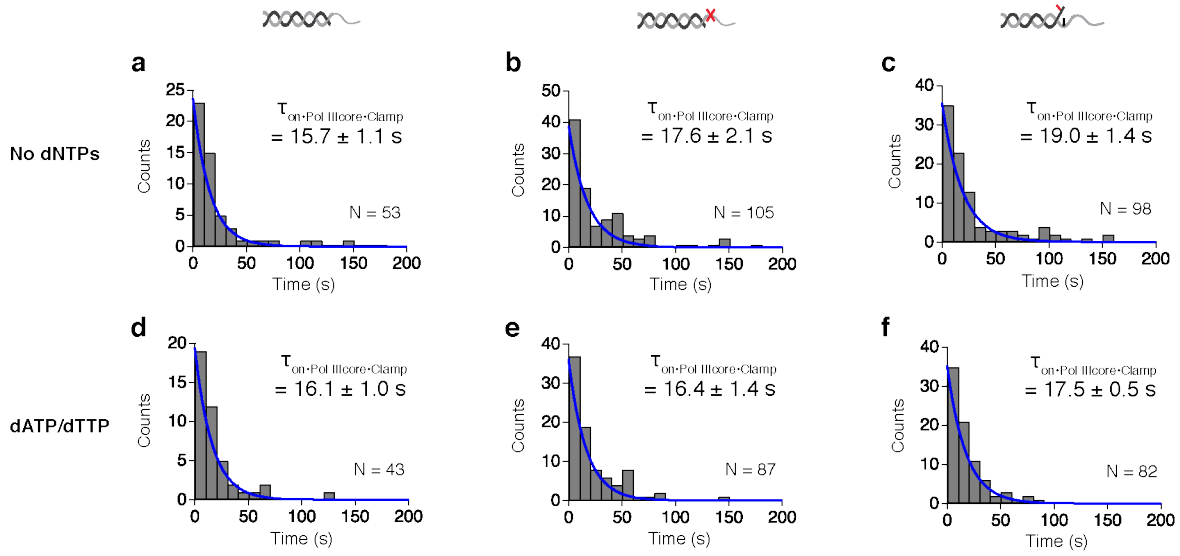


Figure 4.9. Pol IIIcore lifetimes on the clamp on the matched, lesion and mismatched DNA. a, Pol IIIcore lifetime on the matched, **b**, lesion and **c**, mismatched DNA in the absence of nucleotides. **d** Pol IIIcore lifetime on the clamp on the matched **e**, lesion and **f**, mismatched DNA in the presence of dATP and dTTP (0.5 mM each).

Table 4. DNA lesion and mismatch do not affect the lifetime of Pol IIIcore on clamp-DNA or its competition with Pol IV

	Lifetime (s)		Polymerase exchange [2] (%)		
	no dNTP	dATP/dTTP	III → IV	IV → III	III + IV
Matched	15.7 ± 1.1	16.1 ± 1.0	73	14	13
Lesion [1]	17.6 ± 2.1	16.4 ± 1.4	58	24	18
Mismatched [1]	19.0 ± 1.4	17.5 ± 0.5	64	23	13

[1] Lesion DNA: N^2-furfuryl-dG, mismatched DNA: G-T. [2] Polymerase exchange or co-localization in the absence of nucleotides.

cation (clamp loading, unloading and polymerase loading in Chapter 03), the switching between the replicative and the translesion polymerases occurs by concentration-dependent competition and shows little organization.

The ability of the translesion polymerases to compete with the replicative polymerase whenever it dissociates (even on a matched substrate) poses a need for the access of the error-prone translesion polymerases to be regulated. This may be achieved by multiple mechanisms: firstly, during normal DNA synthesis, once Pol IIIcore dissociates it remains tethered to the replisome through binding to the clamp loader protein τ. By simultaneously binding to two or three copies of Pol IIIcore[3,16], the clamp loader increases the effective concentration of Pol IIIcore at the replication fork, making it unlikely for the translesion polymerases to interchange with Pol IIIcore. Only when Pol IIIcore is unable to bypass the lesion, and therefore stalls or dis-

sociates, do the translesion polymerases have a chance to gain access to the clamp. Secondly, the levels of translesion polymerases in the cell under normal circumstances is low and is up-regulated during the SOS response[6]. Our observation that the translesion polymerases compete with the Pol IIIcore in a concentration-dependent manner is consistent with this: when lesions are abundant and the cell is under stress, more translesion polymerases are expressed in order to shift the equilibrium of the polymerase competition in favor of the translesion polymerases. The observation of increased co-localization between Pol IIIcore and Pol IV at higher protein concentrations suggests that Pol IV may also be able to frequently access the clamp-DNA during the SOS response. Furthermore, it has been proposed that the RecA protein may favor the translesion polymerases by inhibiting the Pol IIIcore replisome and activating the translesion replisomes, albeit by an unknown mechanism[11].

The lack of coordination between Pol IIIcore, Pol IV and Pol II on the matched, mismatched and lesion DNA substrates suggests that there is no specific selection for polymerases for lesion bypass. Instead, the process is entirely stochastic: whichever polymerase is recruited will attempt to synthesize past the lesion and when it is unable to do so other polymerases will compete for access to the DNA. Hence, the *E. coli* translesion polymerases are recruited for lesion bypass on a trial and error basis.

Bibliography

[1] Karen A Bunting, S Mark Roe, and Laurence H Pearl. Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. *The EMBO Journal*, 22(21):5883–5892, November 2003.

[2] B P Dalrymple, K Kongsuwan, G Wijffels, N E Dixon, and P A Jennings. A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20):11627–11632, September 2001.

[3] Paul R Dohrmann, Raul Correa, Ryan L Frisch, Susan M Rosenberg, and Charles S McHenry. The DNA polymerase III holoenzyme contains γ and is not a trimeric polymerase. *Nucleic Acids Research*, 44(3):1285–1297, February 2016.

[4] Paul R Dohrmann and Charles S McHenry. A bipartite polymerase-processivity factor interaction: only the internal beta binding site of the alpha subunit is required for processive replication by the DNA polymerase III holoenzyme. *Journal of Molecular Biology*, 350(2):228–239, July 2005.

[5] Rafael Fernandez-Leiro, Julian Conrad, Sjors Hw Scheres, and Meindert H Lamers. cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *eLife*, 4:459, October 2015.

[6] Robert P Fuchs and Shingo Fujii. Translesion DNA synthesis and mutagenesis in prokaryotes. *Cold Spring Harbor Perspectives in Biology*, 5(12):a012682–a012682, December 2013.

[7] Asako Furukohri, Myron F Goodman, and Hisaji Maki. A dynamic polymerase exchange with Escherichia coli DNA polymerase IV replacing DNA polymerase III on the sliding clamp. *The Journal of Biological Chemistry*, 283(17):11260–11269, April 2008.

[8] Justin M H Heltzel, Robert W Maul, Sarah K Scouten Ponticelli, and Mark D Sutton. A model for DNA polymerase switching involving a single cleft and the rim of the sliding clamp. *Proceedings of the National Academy of Sciences*, 106(31):12664–12669, August 2009.

[9] Chiara Indiani, Lance D Langston, Olga Yurieva, Myron F Goodman, and Mike O'Donnell. Translesion DNA polymerases remodel the replisome and alter the speed of the replicative helicase. *Proceedings of the National Academy of Sciences*, 106(15):6031–6038, April 2009.

[10] Chiara Indiani, Peter McInerney, Roxana Georgescu, Myron F Goodman, and Mike O'Donnell. A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. *Molecular Cell*, 19(6):805–815, September 2005.

[11] Chiara Indiani, Meghna Patel, Myron F Goodman, and Mike E O'Donnell. RecA acts as a switch to regulate polymerase occupancy in a moving replication fork. *Proceedings of the National Academy of Sciences*, 110(14):5410–5415, April 2013.

[12] Daniel F Jarosz, Veronica G Godoy, James C Delaney, John M Essigmann, and Graham C Walker. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature*, 439(7073):225–228, January 2006.

[13] Slobodan Jergic, Nicholas P Horan, Mohamed M Elshenawy, Claire E Mason, Thitima Urathamakul, Kiyoshi Ozawa, Andrew Robinson, Joris M H Goudsmits, Yao Wang, Xuefeng Pan, Jennifer L Beck, Antoine M van Oijen, Thomas Huber, Samir M Hamdan, and Nicholas E Dixon. A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *The EMBO Journal*, 32(9):1322–1333, May 2013.

[14] James E Kath, Seungwoo Chang, Michelle K Scotland, Johannes H Wilbertz, Slobodan Jergic, Nicholas E Dixon, Mark D Sutton, and Joseph J Loparo. Exchange between Escherichia coli polymerases II and III on a processivity clamp. *Nucleic Acids Research*, 44(4):1681–1690, February 2016.

[15] James E Kath, Slobodan Jergic, Justin M H Heltzel, Deena T Jacob, Nicholas E Dixon, Mark D Sutton, Graham C Walker, and Joseph J Loparo. Polymerase exchange on single DNA molecules reveals processivity clamp control of translesion synthesis. *Proceedings of the National Academy of Sciences*, 111(21):7647–7652, May 2014.

[16] Rodrigo Reyes-Lamothe, David J Sherratt, and Mark C Leake. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science*, 328(5977):498–501, April 2010.

[17] Ana Toste Rêgo, Andrew N Holding, Helen Kent, and Meindert H Lamers. Architecture of the Pol III-clamp-exonuclease complex reveals key roles of the exonuclease subunit in processive DNA synthesis and repair. *The EMBO Journal*, 32(9):1334–1343, May 2013.

[18] Jun Yin, Paul D Straight, Shaun M McLoughlin, Zhe Zhou, Alison J Lin, David E Golan, Neil L Kelleher, Roberto Kolter, and Christopher T Walsh. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15815–15820, November 2005.

Chapter 5

A heterodimeric clamp for structural analysis of *E. coli* translesion synthesis

5.1 Introduction

The background to *E. coli* translesion synthesis is discussed in Chapters 1 and 4. To better understand the process of polymerase switching, it is necessary to study how the replicative DNA polymerase Pol IIIcore and the translesion DNA polymerases Pol IV and Pol II bind to the β clamp. Indeed, if Pol IIIα and Pol IV formed a stable toolbelt on DNA, it should be possible to isolate the Pol IIIα-Pol IV-β clamp-DNA complex and study it structurally.

Two hydrophobic grooves on the β clamp serve as common binding sites for its interaction partners (Fig. 5.1a) through their well-conserved β clamp binding motif (CBM) with the amino acid sequence Qxx(L/M)xF[2]. Yet, there is no structure of a full translesion polymerase bound to the β clamp to inform understanding of how the translesion polymerases interact with the β clamp. Owing to the weak affinity of the Pol IV-β clamp interaction, previous efforts in the lab to crystallize the complex consistently yielded crystals of the isolated β clamp (unpublished data).

To study Pol II binding to the β clamp, the CMB on Pol II was mutated from the native QLGLF to the consensus sequence QLDLF to improve the binding affinity to the β clamp, and this complex was studied by cryo-EM (Dr Ana Toste Rego, unpublished data). A low resolution 3D map was obtained and clearly shows two polymerases bound to one β clamp (Fig. 5.1b) as a result of the presence of two binding sites on the β clamp. As it can be seen from the 3D map, the two Pol IIs seem to clash and therefore having two polymerases bound to one β clamp may prevent the polymerase from accessing all the possible conformations on the β clamp.

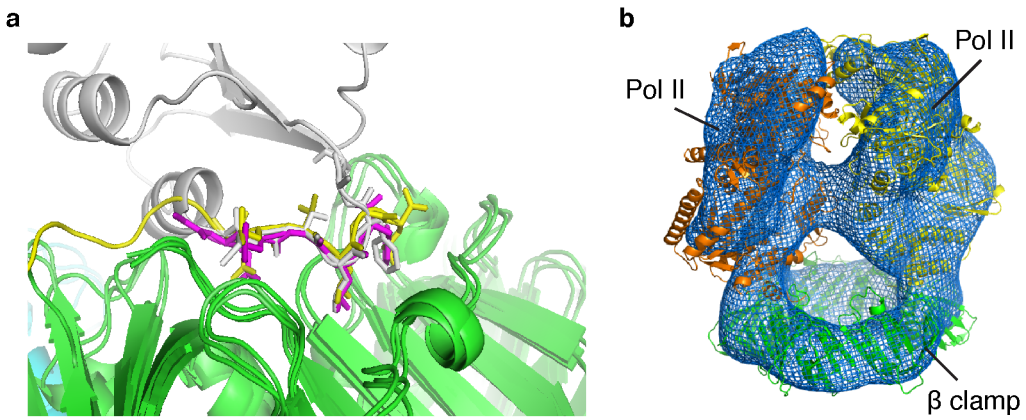


Figure 5.1. Interaction partners bind to a common hydrophobic cleft on the β clamp. a, The overlaid structures of the β clamp binding motifs of Pol IV (yellow), Pol II (magenta) and δ (gray) bound to the β clamp. **b**, Low resolution 3D map of Pol II improved β clamp binder bound to the β clamp obtained by cryo-EM (Dr. Ana Toste Rego, unpublished data).

Therefore, in order to study the binding of a single translesion polymerase structurally, the two polymerase binding sites need to be uncoupled. To do this, we aimed to create a heterodimeric β clamp. A β clamp heterodimer was previous purified by Scouten Ponticelli et al, who tagged the wild-type β clamp with His and Myc tags and separated the His/Myc heterodimers by affinity purification[10]. However, when this experiment was repeated, it was seen that the heterodimers re-equilibrated back into the respective homodimers over time (Dr Ana Toste Rego, unpublished data). Therefore, to obtain a more stable β clamp heterodimer, we mutated the β clamp dimer interface and looked for individual β clamp mutants that form monomers on their own but can be complemented by another mutant to form stable heterodimers. Given that the monomer mutants have low propensity to form homodimers, any re-equilibration over time should be insignificant.

I screened for β clamp monomer mutants during Part III of my undergraduate study at the University of Cambridge, and this Chapter is a continuation of the work. Therefore, in order to provide background for this Chapter, the findings in my Part III thesis are briefly summarized in Section 5.1.1.

5.1.1 Monomeric β clamp

Given that any mutation that breaks the β clamp dimer must be able to be complemented to form a heterodimer, point mutations were created to minimize perturbations to the β clamp interface. The β clamp mutants were individually purified and their oligomerization state was analyzed by gel filtration. If a mutation prevented the dimerization of the β clamp, the monomers are ex-

pected to have an increased retention volume on gel filtration compared to the wild-type dimeric β clamp. Out of the 15 tested β clamp mutants, seven formed monomers. However, only two monomer mutants, S104Q and E304N, complemented each other and formed a heterodimer (Fig. 5.2a).

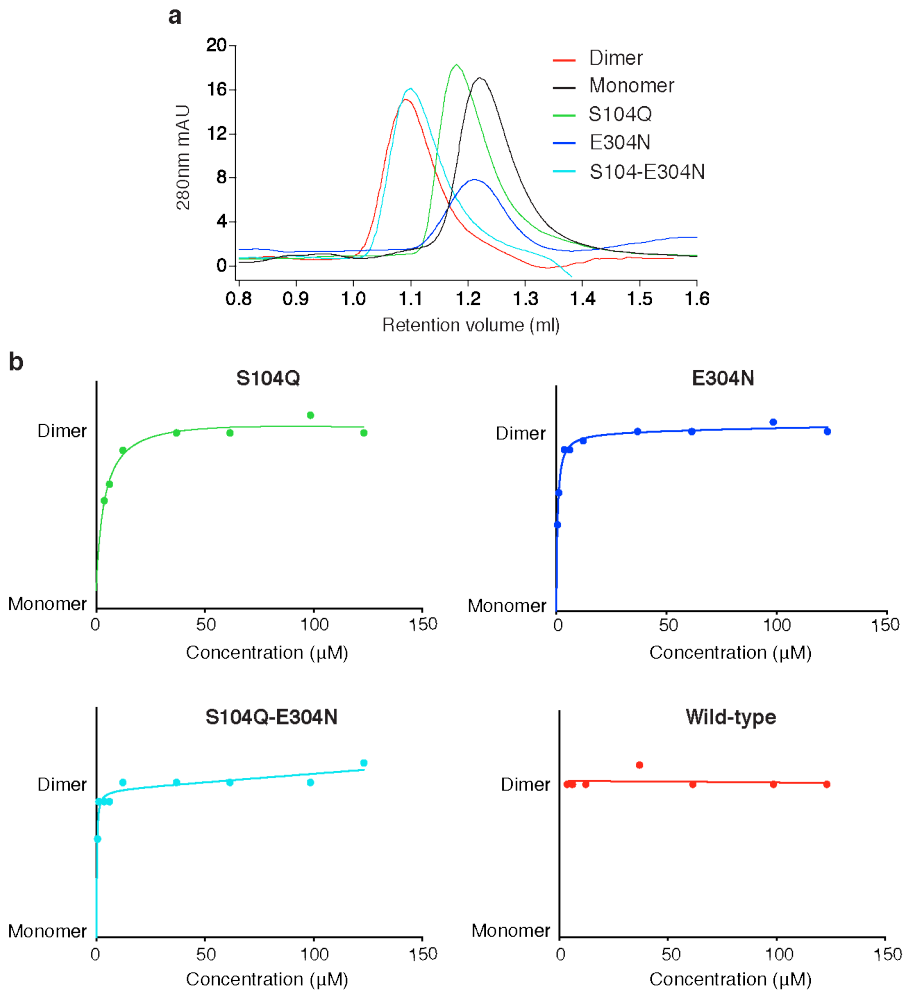


Figure 5.2. The two monomeric β clamp mutants S104Q and E304N form stable heterodimers. a, Gel filtration analysis of the β clamp mutants S104Q and E304N shows that they form monomers in isolation, and a heterodimer when both are present. The wild-type β clamp is included as a control for the dimeric species and the β I272A, L273A mutant[4] is included as a control for the monomeric species. **b**, Gel filtration analysis of S104Q, E304N, S104Q-E304N and wild-type β clamp at various protein concentrations shows that both S104Q and E304N form homodimers at high protein concentrations, but the S104Q-E304N is more stable. The wild-type β clamp is dimeric throughout the concentration range.

Next, we tested the stability of the β clamp monomers by analyzing their retention volumes at different protein concentrations by gel filtration chromatography (Fig. 5.2b). When the monomer-dimer state (1/retention volume) is plotted against protein concentration, it is seen

that the wild-type β clamp remains a stable dimer throughout the concentration range tested, while the S104Q and E304N mutants only dimerize at higher protein concentrations. This suggests that when the S104Q and E304N mutants are added together, there may be a mixture of the mutant homodimers as well as the S104Q-E304N heterodimer. However, the gel filtration analysis of the heterodimer suggests that it is more stable than the respective homodimers, as it stays as a dimer at the lower protein concentrations where the homodimers are dissociating (Fig. 5.2b). The loss of the 280 nm UV signal when the proteins are further diluted prevents analysis at even lower protein concentrations, and therefore the K_d values for the monomer-dimer equilibrium need to be determined using other methods.

5.1.2 Aims

There is some structural information to address whether the Pol IIIα-Pol IV-β clamp trimeric complex can exist. It is known that in solution, Pol IIIα forms a tight complex with the ϵ and θ subunits known as the Pol IIIcore[6]. Dr Rafael Fernandez-Leiro in our lab solved the cryo-EM structure of Pol IIIcore on the β clamp and DNA[3], which shows that the surface of the β clamp is fully occupied by Pol IIIcore and therefore there is no space to accommodate Pol IV. In addition, the characterization of Pol III, Pol IV and ϵ binding to the β clamp by size exclusion chromatography shows that in the presence of the ϵ subunit, Pol IV is excluded from the complex[11].

Another line of thought stems from the crystal structure of the 'little finger' (LF) domain of Pol IV on the β clamp[1] (see Chapter 1). This structure shows a substantial secondary interaction interface between Pol IV LF domain and the side of the β clamp, in addition to binding to the canonical hydrophobic pocket, which led to the proposal that these secondary interactions may be important in maintaining Pol IV in an inactive orientation during polymerase exchange with the replicative polymerase.

Given the disparity between the existing structural data, the use the β clamp heterodimer to solve the structure of the full-length Pol IV or Pol II on the β clamp and DNA would be helpful for understanding polymerase exchange at a structural level. Towards this goal, the S104Q-E304N β clamp heterodimer is further characterized, with mutagenesis of the hydrophobic binding pocket for abolishing polymerase binding on one β clamp monomer. This project was done with the help of summer student Timothy Venkatesan.

5.2 Materials and Methods

5.2.1 Protein expression and purification

The gene for the *E. coli* β *dnaN* was cloned into pET28a vectors with the heterodimer mutations S104Q and E304N and the hydrophobic binding pocket mutations M362D, M362R and M362S and M362del. The β proteins were expressed for two hours at 30 °C and purified in 20 mM Hepes pH 7.5, 500 mM NaCl, 2 mM DTT and 25-500 mM Imidazole by nickel affinity chromatography using a Histrap HP column (GE Healthcare). The N-terminal His-tag was cleaved overnight at 4 °C using the human rhinovirus 3C protease, and then the cleaved protein was purified by anion exchange in 20 mM Hepes pH 7.5, 2 mM DTT and 0-1 M NaCl using a Hitrap Q HP column. In the final step, the proteins were purified by size exclusion chromatography in 20 mM Hepes pH 7.5, 150 mM NaCl and 2 mM DTT using a 120 ml Superdex S75 column. All columns were purchased from GE Healthcare.

5.2.2 Analytical ultracentrifugation

For the analytical ultracentrifugation (AUC) experiments, β was labeled on a single surface-exposed cysteine residue using a five fold molar excess of maleimide-Alexa488 fluorophore. DTT was removed from the buffer (20 mM Hepes pH 7.5, 150 mM NaCl) by gel filtration prior to the labeling reaction and after 5 hours at 4 °C, the reaction was quenched by running it on a 2.4 ml Superdex 75 gel filtration column (GE Healthcare) in the presence of 2 mM DTT. A ProteomeLab XL-I analytical ultracentrifugation (Beckman Coulter) with fluorescence detection was used for velocity sedimentation experiments that were carried out at 45 000 rpm at 20 °C. The program SEDFIT[9] was used to fit the concentration gradients with sedimentation coefficient distributions (c(s)) by Dr Stephen McLaughlin. The c(s) distributions were then fitted to normal distributions and deconvoluted into monomeric and dimeric fractions. The area under the individual peaks were integrated as estimates for the concentration of the monomeric and dimeric species, which were used to determine the K_ds. The shifting peak positions were used to calculate the K_ds where the c(s) distributions were concentration-dependent.

5.2.3 Cryo-electron microscopy

The S104Q-E304N β heterodimer was assembled from the respective mutants at a final concentration of 48 μM and purified in 25 mM Hepes pH 7.5, 50 mM potassium glutamate and 2 mM DTT on a 2.4 ml Superdex 75 gel filtration column (GE Healthcare). The peak fraction

was retrieved and 0.1 volume of 0.05 % (V/V) Tween 20 was added. The sample was incubated for 5 min before pipetting onto a glow-discharged holey carbon cryo-EM grids (Quantifoil Cu R1.2/1.3) and frozen in liquid ethane using a Vitrobot (FEI). Images were taken on the Polara electron microscope (FEI) and the 2D classification was performed in the program Relion[8] by Dr Julian Conrad.

5.2.4 Differential scanning fluorimetry

The thermal stabilities of the β M362D, M362R and M362S and M362del proteins were measured using a Rotor-Gene 6000 real-time PCR machine (Corbett Research). Reactions contained 6 μM protein with 1X Sypro Orange dye in 25 mM HEPES-KOH pH 7.5, 150 mM NaCl and 2 mM DTT buffer. A reading was taken every 0.6 °C in the temperature range 30-85.2 °C.

5.2.5 Primer extension assays

The processivity of the β clamp mutants was tested in primer extension assays using a 5.4 kb single-stranded ϕX174 DNA primed with a FAM-labeled 30-nt primer. The extension reactions were performed at 25 °C at 10 nM final protein concentration (β clamp, γ complex, χ, ψ, Pol IIIcore) and the reaction was stopped after 15 s, 30 s, 1 min, 2 min and 5 min using a stop buffer containing 5 % Ficoll-400, 20 mM EDTA, 0.5 % SDS, and 0.5 mg/ml bromophenol blue. The products of the extension assay were separated on a 0.8 % alkaline agarose gel containing 3 M NaOH and imaged using a Typhoon imager.

5.3 Results

5.3.1 The β clamp monomer-dimer equilibrium

To study the monomer-dimer equilibrium and quantify the dimer affinities, the wild-type, S104Q, E304N and S104Q-E304N β clamp were subjected to analytical ultracentrifugation with fluorescence detection[7] with the help of Dr Stephen McLaughlin. To do this, we labeled the proteins using maleimide Alexa 488 dye. Curiously, we obtained close to 100 % labeling efficiency for the wild-type β clamp and the E304N mutant, but we could only achieve 35 % labeling of the S104Q mutant under the same reaction conditions. However, since only the fluorescently labeled molecules will be detected, this difference in labeling efficiency should not matter.

At a concentration of 1.8 μM wild-type β clamp, dimeric species were predominantly detected in the velocity sedimentation experiments (Fig. 5.3). Titrating the concentration down to 0.8 μM did not affect this monomer-dimer equilibrium, and the K_d for the homo-dimerization of the wild-type β clamp is approximated to be 5 nM.

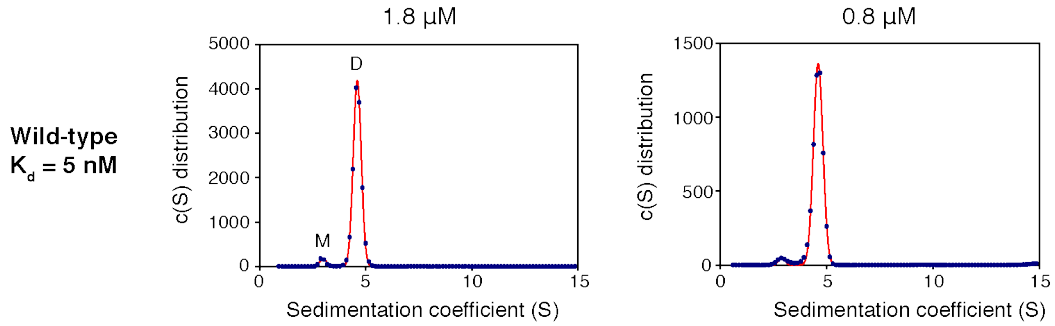


Figure 5.3. The wild-type β clamp forms a highly stable homodimer. The sedimentation coefficient distributions were obtained for the wild-type β clamp at 1.8 and 0.8 μM protein concentration. The β clamp is mostly dimeric and has a K_d of $\sim$ 5 nM. M = monomer, D = dimer.

In contrast, both mono- and dimeric species were detected for the β clamp mutants S104Q and E304N (Fig. 5.4). This sedimentation behaviour was concentration-dependent, as a larger proportion of dimers was observed when the protein concentration was increased from 1.8 μM to 3.5 μM. The K_ds for the homo-dimerization were approximately 3 μM for the S104Q mutant and 1.5 μM for the E304N mutant. This result suggests that the E304N mutant has a higher propensity to form dimeric β clamps compared to the S104Q mutant, which agrees with the previously described gel filtration experiments (Fig. 5.2b).

The S104Q-E304N heterodimer was made by mixing the S104Q-Alexa 488 and E304Q-Alex 488 proteins together at equimolar concentrations. When the S104Q-E304N heterodimer was subjected to analytical ultracentrifugation, a significantly larger proportion of dimers was observed at 3.5 μM protein concentration compared to the S104Q and E304N single mutants (Fig. 5.4). Calculating the K_d gave 43 nM, which is almost two orders of magnitude lower compared to the homodimerization of the single mutants. This shows that while the mutants are capable of forming weak S104Q and E304N homodimers, the S104Q-E304N heterodimeric species has a much higher affinity. Therefore, when the S104Q and E304N β clamp mutants are added together, a much greater proportion of the dimeric β clamps formed will be in the heterodimeric state. This result is in agreement with the results from the gel filtration titration experiments (Fig. 5.2b).

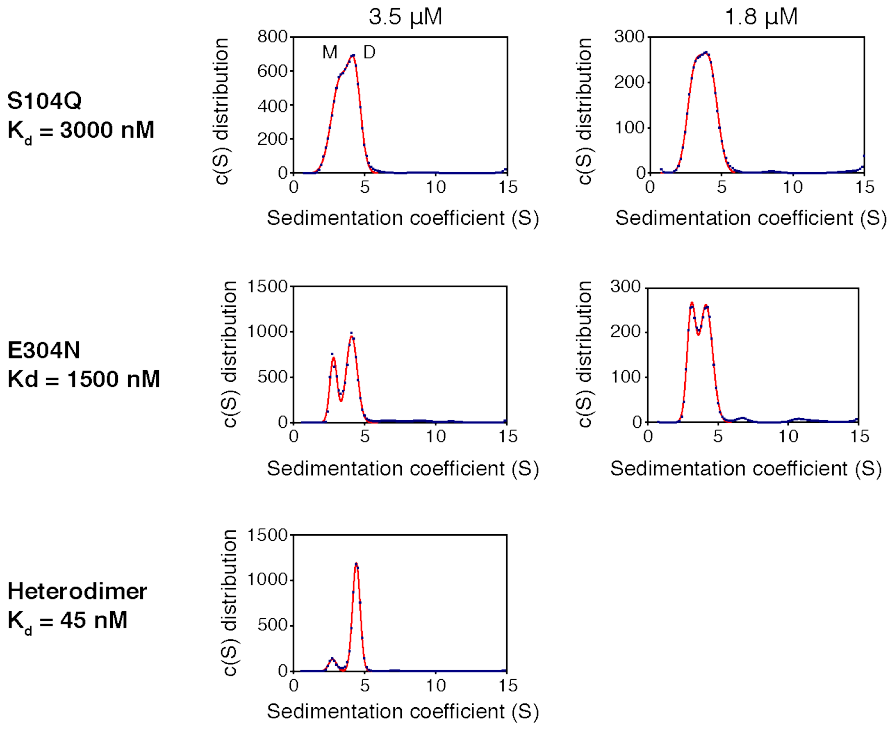


Figure 5.4. The heterodimeric β clamp is more stable compared to the respective homodimers. The sedimentation coefficient distributions were obtained for the β clamp mutants S104Q and E304N at 3.5 and 1.8 μM protein concentration. A larger proportion of the dimeric species is observed at the higher protein concentration for both mutants. The sedimentation coefficient distribution is obtained for the S104Q-E304N heterodimer at 3.5 μM protein concentration, at which a predominant dimeric species is detected. The K_d for the β clamp heterodimer is almost 100-fold lower than the respective homodimers.

5.3.2 Structural analysis of the heterodimeric β clamp

To ensure that the S104Q-E304N heterodimeric β clamp has retained the ring-shaped structure of the wild-type β clamp, the heterodimer was subjected to cryo-electron microscopy with the help of Dr Julian Conrad. This was to exclude the possibility that the monomers have associated together in an alternative conformation, for example by stacking on top of each other. Individual ring-shaped β clamp heterodimers can be seen in the micrographs (Fig. 5.5a) and the 2D class averages show a ring shaped structure (Fig. 5.5b). While the 2D class averages are not sufficient to give ultimate proof that the association between the monomeric β clamps is identical to the wild-type, they show that the overall ring shape of the β clamp is conserved.

The 2D class averages from the cryo-EM was confirmed by small angle X-ray scattering (SAXS) experiments performed by Dr Meindert Lamers that show that at 120 μM concentration, the S104Q-E304N heterodimer has the same overall shape as the wild-type β clamp (Fig. 5.5c).

At this concentration, the S104Q and E304N homodimers also have the same ring shape. These results are in agreement with the gel filtration analysis that shows that the β clamp mutants dimerize at 120 μM protein concentration (Fig. 5.2b). Taken together, the cryoEM 2D averages and the SAXS data show that the β clamp interface mutants have not interfered with the overall structure of the β clamp dimer ring.

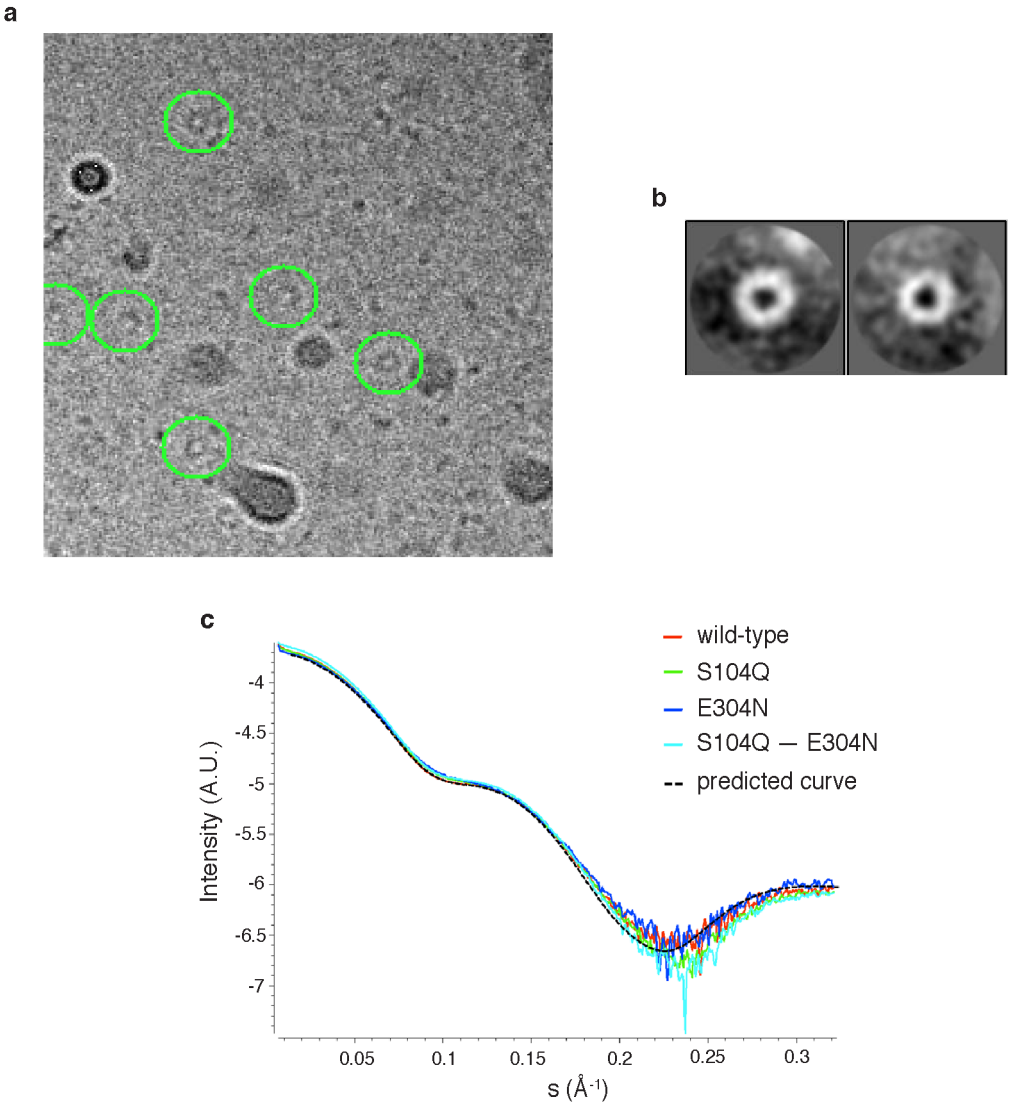


Figure 5.5. The S104Q-E304N heterodimeric β clamp has the same overall ring-shaped structure as the wild-type β clamp. a, A cryoEM micrograph showing individual heterodimer particles. **b**, The 2D classifications from manually picked 125 particles. **c**, Small-angle X-ray scattering data comparing the wild-type, S104Q, E304N and S104Q-E304N β clamps at 120 μM protein concentration to the predicted curve calculated from the wild-type β clamp crystal structure[5].

5.3.3 Biochemical analysis of the heterodimeric β clamp

The S104Q-E304N β heterodimeric β clamp was tested in a primer extension assay for its processivity, with the wild-type β clamp and the S104Q and E304N mutants included as controls. The results show that the S104Q-E304N heterodimer is processive; however, it does not produce as abundant extended DNA products as the wild-type β clamp (Fig. 5.6). The S104Q mutant almost completely abolishes extension but interestingly, the E304N mutant shows only slightly reduced activity compared to the S104Q-E304N heterodimer. This can be explained by the increased propensity of the E304N mutant to form homodimers compared to the S104Q mutant as shown by gel filtration and AUC (Fig. 5.2b and Fig. 5.4).

When the primer extension assay was repeated using a serial dilution of the S104Q-E304N heterodimer and the E304N mutant, we observed that the heterodimer still produces full-length products at the lower protein concentrations but less than the wild-type β clamp. On the other hand, the E304N homodimer appears to be not fully processive as shown by the lack of full-length products (Fig. 5.7).

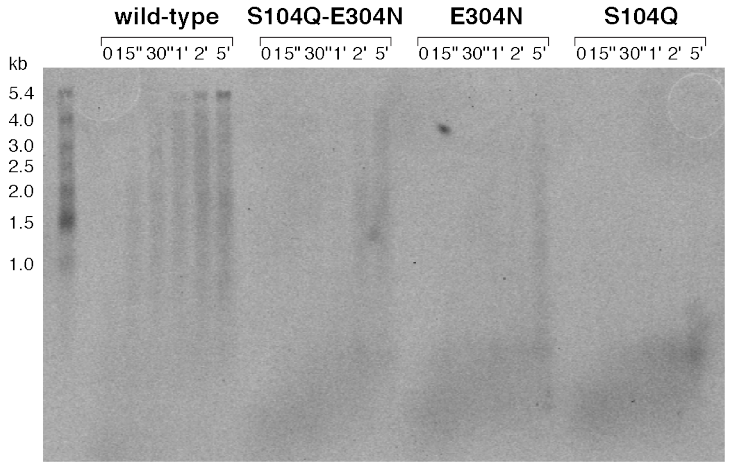


Figure 5.6. The S104Q-E304N heterodimeric β clamp is processive. Comparing the processivity of the wild-type β clamp to the S104Q-E304N heterodimer and the respective mutants in a primer extension assay using a 5.4 kb ssDNA. The reactions were quenched after 15 s, 30 s, 1 min, 2 min, and 5 min using EDTA.

The decrease in processivity of the S104Q-E304N heterodimeric β clamp compared to the wild-type β clamp suggests that disruption of the dimer interface has an effect on β clamp function. This could either be due to decreased β clamp loading caused by a disruption in the β clamp-loader interaction, or a disrupted β clamp-DNA interaction. Binding of Pol IIIα and ϵ to the heterodimeric β clamp is not affected (data not shown). In addition, the presence of

the mutant homodimeric β clamps, especially the S104Q homodimer which completely abolishes β clamp processivity, may reduce the apparent processivity of the S104Q-E304N β clamp heterodimer.

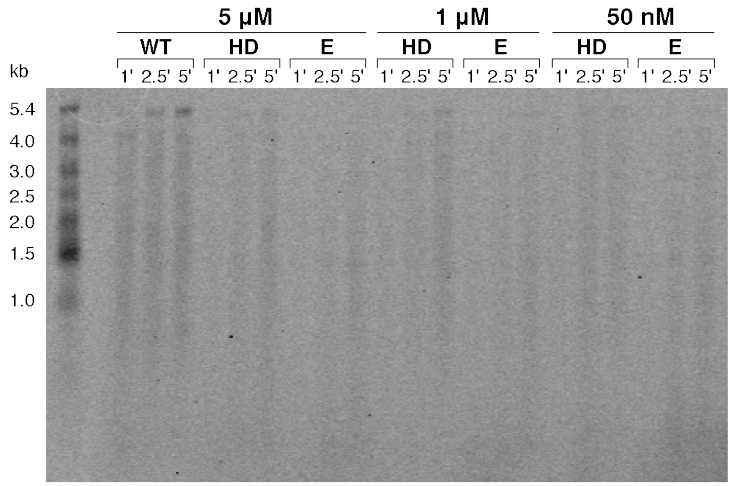


Figure 5.7. The E304N β clamp homodimer is not fully processive. Titration of the S104Q-E304N heterodimeric and the E304N homodimeric β clamp in a primer extension assay shows that the E304N homodimer does not form full-length products. WT = wild-type, HD = heterodimer, E = E304N mutant. The reactions were quenched after 1 min, 2.5 min and 5 min.

5.3.4 Abolishing polymerase binding to the β clamp

In order to create a single polymerase binding site on the heterodimeric clamp, we needed to disrupt one of the canonical hydrophobic clefts. To do this, we mutated the methionine 362 residue in this canonical binding pocket to M362D, M362R and M362S. In addition, a deletion of the β clamp C-terminal five residues starting from M362, termed M362del[10], was included as a control.

We incubated the β clamp M362 mutants with Pol IIIα and subjected the reaction to analysis by gel filtration on a 2.4 ml Superdex 200 Increase column (GE Healthcare). We see that the wild-type β clamp formed a complex with Pol IIIα (Fig. 5.8a and c) but this interaction was disrupted in the M362D and M362R mutants (Fig. 5.8b-c). The M362S mutant did not significantly disrupt Pol IIIα binding, which can be explained by the more subtle mutation from methionine to a polar rather than a charged residue. The M362del mutant also abolished complex formation (Fig. 5.8b-c).

Given that the exonuclease ϵ subunit enhances Pol IIIα binding to the β clamp, we wondered whether ϵ could restore the reduced binding of Pol IIIα to the mutant β clamps. To do this, we

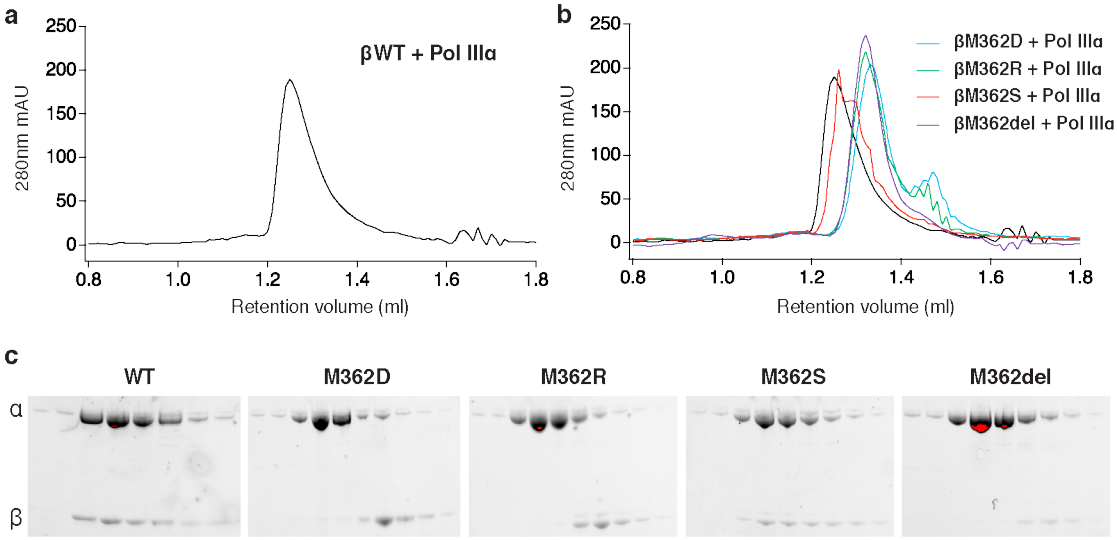


Figure 5.8. The M362D, M362R and M362del β clamp binding pocket mutants do not bind to Pol IIIα. a, Gel filtration analysis of the wild-type β clamp-Pol IIIα complex. **b**, Gel filtration runs of the mutant β clamp-Pol IIIα complexes. **c**, Gel analysis of the complexes.

subjected the reactions containing the β clamp M362 mutants, Pol IIIα and ϵ to gel filtration analysis. Consistent with the previous results with only Pol IIIα, the M362D, M362R and M362del mutants prevented complex formation as seen by a shift in the retention volumes, whereas the M362S did not have an effect (Fig. 5.9a-c). Closer inspection at the gel analyses revealed that the M362D mutant was more disruptive to complex formation compared to the M362R mutant (Fig. 5.9c).

The effect of the β clamp binding pocket mutations on processive DNA synthesis was tested in primer extension assays. In agreement to the gel filtration analysis, the M362D and M362R mutants were no longer processive, whereas the M362S mutant had no effect (Fig. 5.10).

Furthermore, we measured the thermal stability of these β clamp polymerase binding mutants by differential scanning fluorimetry. The thermal stability of the M362D, M362R and M362S mutants were comparable to the wild-type β clamp, within 2.6 °C difference in the melting temperature (Fig. 5.11). However, we observed a significantly lower melting temperature for the M362del mutant, which had a melting temperature 8.7 °C lower than the wild-type. Hence, we chose not to use M362del for future experiments.

5.4 Discussion

The structural and biochemical analyses of the S104Q-E304N heterodimeric β clamp consistently show that the heterodimer forms a ring structure and is functional in processive DNA

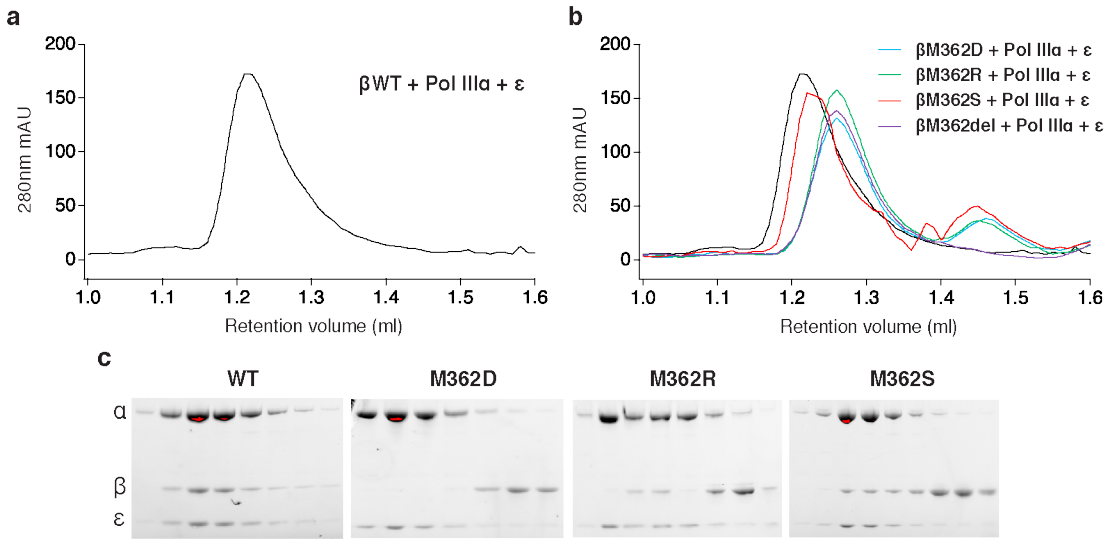


Figure 5.9. The M362D, M362R and M362del β clamp binding pocket mutants have reduced binding to Pol IIIα and ϵ. a, Gel filtration analysis of the wild-type β clamp-Pol IIIα-ϵ complex. **b**, Gel filtration runs of the mutant β clamp-Pol IIIα-ϵ complexes. **c**, Gel analysis of the complexes: gel for the M362del mutant is lacking.

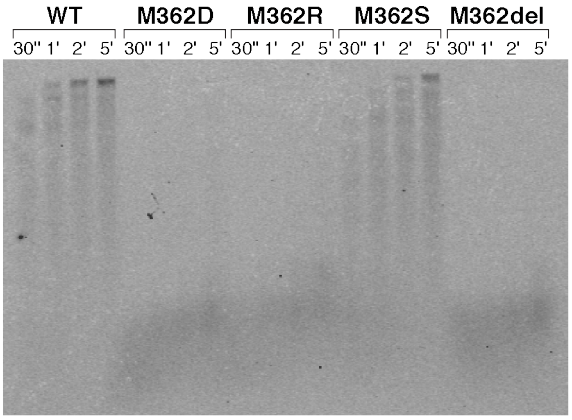


Figure 5.10. The M362D, M362R and M362del β clamp binding pocket mutants are no longer processive. Primer extension assay on a 5.4 kb ssDNA shows that the three mutations made in the β clamp binding pocket abolished processivity. The reactions were quenched after 30 s, 1 min, 2.5 min and 5 min using EDTA.

synthesis, that is, it is able to be loaded onto DNA by the β clamp loader and binds to the polymerase. The S104Q-E304N heterodimer is preferentially formed over the respective S104Q and E304N homodimers, but it is not as stable as the wild-type β clamp. Despite the slightly reduced stability, the heterodimeric β clamp allows the two binding sites of the β clamp to be independently manipulated, which is not possible using the wild-type β clamp.

One route of future work would be to improve the current S104Q-E304N heterodimer by

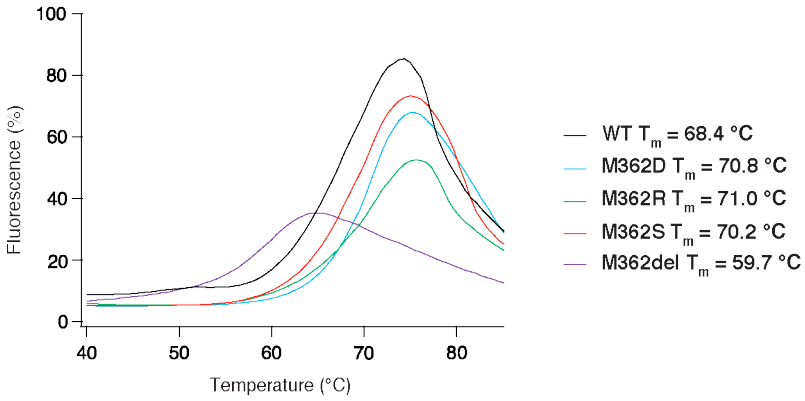


Figure 5.11. The M362del β clamp binding pocket mutants is not thermally stable. Differential scanning fluorimetry experiments show that the melting temperature T_m of the M362del mutant is significantly lower than the wild-type β clamp and the M362D, M362R and M362S mutants.

screening for additional mutations at the heterodimer interface. For example, if there were another pair of interface mutants that form monomers in isolation and can be complemented in the heterodimer, combining one mutant with S104Q and the other with E304N may give a more stable β clamp heterodimer that have reduced propensity to re-equilibrate into the respective monomers.

Another route is to utilize the S104Q-E304N β clamp heterodimer for studying translesion polymerase binding structurally. The low resolution cryo-EM map of the increased β binder of Pol II on the wild-type β clamp shows two polymerases bound (Fig. 5.1b). The problem of dual Pol II binding to the β clamp can be solved by using the β clamp heterodimer where the binding site on one β clamp monomer is mutated away. Given that the E304N mutant has a larger propensity to form homodimers, it is desirable to place the polymerase binding mutation M362D together with the E304N mutation. This way, any E304N homodimers formed would not bind to the polymerase and therefore any complex formed between a polymerase and a β clamp should have a single Pol II bound. The same can be done for Pol IV, where the CBM could be mutated to improve binding to the β clamp. However, Pol IV is much smaller with a molecular weight of 39.5 kDa and therefore this complex is expected to be more difficult to study by cryo-EM compared to Pol II, which has a molecular weight of 89.8 kDa. An additional advantage of using cryo-EM for the study of the single polymerase-β clamp complexes is that at the low (μM) protein concentrations used, any β clamps formed between the S104Q and E304N mutants should be predominantly heterodimeric given that the mutant homodimers are more prone to falling apart at low concentrations.

Hence, our heterodimeric clamp will be a valuable tool for the determination of structures of translesion polymerases bound to the β clamp.

Bibliography

[1] Karen A Bunting, S Mark Roe, and Laurence H Pearl. Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. *The EMBO Journal*, 22(21):5883–5892, November 2003.

[2] B P Dalrymple, K Kongsuwan, G Wijffels, N E Dixon, and P A Jennings. A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20):11627–11632, September 2001.

[3] Rafael Fernandez-Leiro, Julian Conrad, Sjors Hw Scheres, and Meindert H Lamers. cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *eLife*, 4:459, October 2015.

[4] D Jeruzalmi, O Yurieva, Y Zhao, M Young, J Stewart, M Hingorani, M O'Donnell, and J Kuriyan. Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of E. coli DNA polymerase III. *Cell*, 106(4):417–428, August 2001.

[5] X P Kong, R Onrust, M O'Donnell, and J Kuriyan. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, 69(3):425–437, May 1992.

[6] C S McHenry and W Crow. DNA polymerase III of Escherichia coli. Purification and identification of subunits. *The Journal of Biological Chemistry*, 254(5):1748–1753, March 1979.

[7] Sridharan Rajagopalan, Agnes M Jaulent, Mark Wells, Dmitry B Veprintsev, and Alan R Fersht. 14-3-3 activation of DNA binding of p53 by enhancing its association into tetramers. *Nucleic Acids Research*, 36(18):5983–5991, October 2008.

[8] Sjors H W Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, December 2012.

[9] P Schuck. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical Journal*, 78(3):1606–1619, March 2000.

[10] Sarah K Scouten Ponticelli, Jill M Duzen, and Mark D Sutton. Contributions of the individual hydrophobic clefts of the Escherichia coli beta sliding clamp to clamp loading, DNA replication and clamp recycling. *Nucleic Acids Research*, 37(9):2796–2809, May 2009.

[11] Ana Toste Rêgo, Andrew N Holding, Helen Kent, and Meindert H Lamers. Architecture of the Pol III-clamp-exonuclease complex reveals key roles of the exonuclease subunit in processive DNA synthesis and repair. *The EMBO Journal*, 32(9):1334–1343, May 2013.

Chapter 6

Concluding remarks

Mechanisms in the cell that ensure the accurate replication of DNA and maintenance of its correct information are crucial for the viability of the cell. However, mutations in DNA may also give rise to variability and adaptibility that confer the selective advantages required by a cell to survive. Therefore, cells rely on both the highly accurate and mutagenic events during DNA replication.

This PhD thesis studied the machinery responsible for the high fidelity DNA copying mechanism in *E. coli*. Consistent with the requirement for high accuracy, the findings presented in the previous chapters show that the factors in the replisome follow a highly controlled sequence of events from the loading of the clamp, to the switching of binding from the clamp loader to the replicative polymerase, to the unloading of the clamp by the clamp loader. The high fidelity of the replicative DNA polymerase, however, does not allow lesions on the template DNA strand to be bypassed, and the rescue of a stalled replication fork is dependent on the inherently mutagenic DNA polymerases known as the translesion polymerases.

Given the diversity of DNA lesions and only three translesion DNA polymerases in *E. coli*, it is difficult to envision that the polymerases specifically recognize certain lesions and therefore form lesion-polymerase pairs. Instead, the bypass of a lesion has been shown to be dependent on the amount of polymerases available in the cell[1]. The findings presented in this thesis support this stochastic mechanism for lesion bypass. We show that the translesion polymerases compete with the replicative polymerase for binding to the clamp on DNA in a concentration-dependent manner. Furthermore, in contrary to the toolbelt model, we see that the translesion and replicative DNA polymerases do not form a stable complex on the clamp-DNA.

Therefore, this work illustrates two distinct branches of copying DNA, where one branch is ordered and achieves high accuracy, while the other branch operates in a stochastic manner,

which allows it to overcome the burden imposed by the large diversity of DNA lesions. An overview of these pathways is shown schematically in Fig. 6.1.

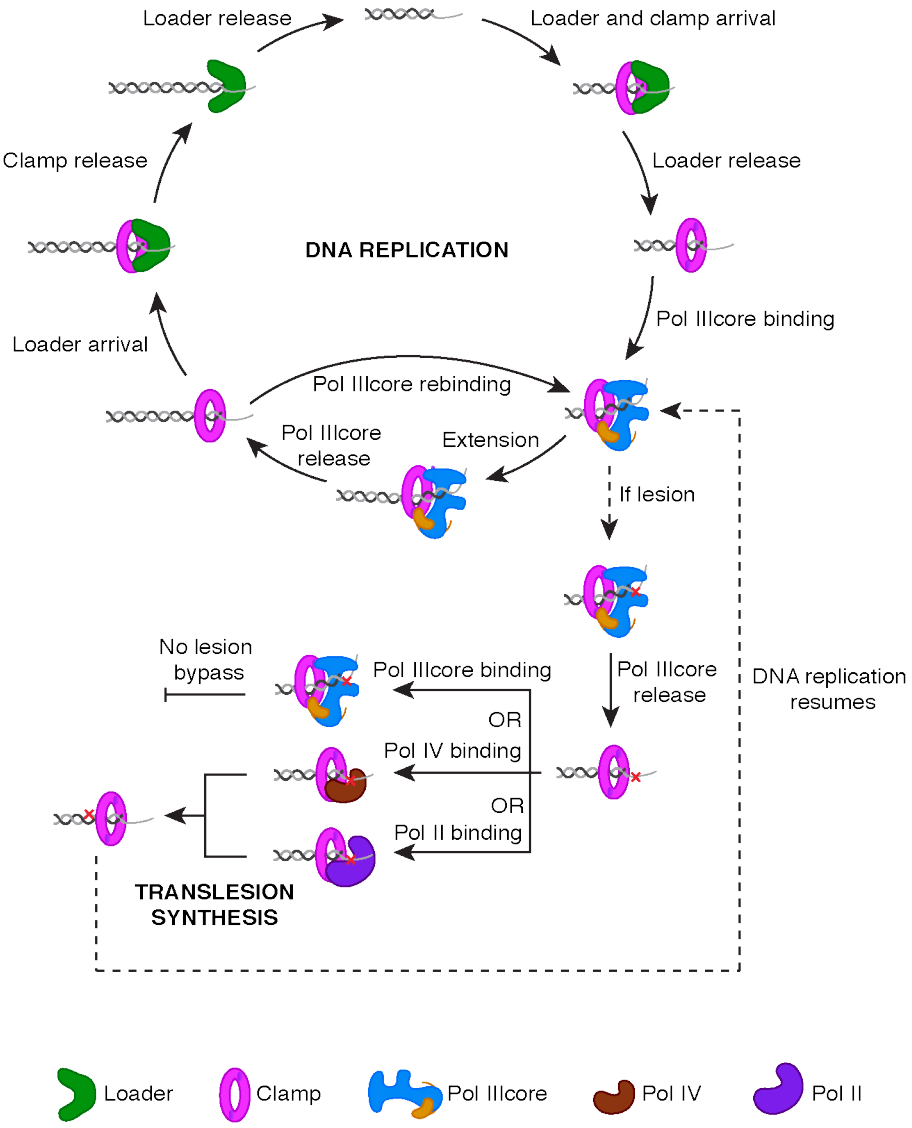


Figure 6.1. A model for DNA replication and translesion synthesis. The DNA replication cycle consists of an organized sequence of events involving clamp loading, polymerase loading, DNA synthesis, polymerase release and clamp unloading. In contrast, the events during translesion DNA synthesis are not coordinated, but are instead governed by a direct competition between the replicative and the translesion DNA polymerases. DNA replication resumes once the lesion has been bypassed by one of the translesion DNA polymerases.

The replisome consists of multiple other proteins that were not studied as part of this thesis, but would be interesting areas for future work. As presented in Chapter 1, DNA replication initiates at an origin of replication. Soon after, the DnaB helicase is loaded to promote the

unwinding of DNA. Due to the physical interaction between the helicase and the DnaG primase, the unwinding of DNA is coupled to primer synthesis. Only after these events will the clamp loader complex catalyze the loading the clamp onto DNA, followed by the association of the replicative DNA polymerase. Thus far, it has been suggested that the χ subunit of the clamp loader binds to SSB and displaces the primase, which is also shown to bind to SSB, from the newly synthesized primer. It would be of interest to study how the exchange between the primase and clamp loader occurs.

In addition, a standing question involves the release of the polymerase at the end of the Okazaki fragment during lagging strand synthesis. It would be of interest for future studies to visualize whether the collision of the polymerase to the primer of the previous Okazaki fragment causes its early release from DNA.

When the replicative DNA polymerase is stalled at a DNA lesion, it remains a topic of interest how the *E. coli* cell makes a choice between the various pathways in DNA damage tolerance. Does the re-priming of the replication fork downstream of the site DNA damage occur after the recruitment of a translesion DNA polymerase that fails to relieve the replication fork stalling, or are the re-priming and translesion synthesis pathways functional on distinct lesions? Given that the choice between translesion synthesis and re-priming or fork regression determines whether the lesion will be bypassed in an error-prone or error-free manner, studying how the choice between the two pathways is made would allow us to better understand the balance between fidelity and mutagenesis during the copying of DNA.

The multiple layers of control in the complex activation mechanism of Pol V in *E. coli* makes it both a difficult and interesting target for study. Thus far, Pol V has been purified in only a few laboratories and studies of Pol V have mainly come from the work done by Goodman, Woodgate, Fuchs and Fujii[4]. Given the challenge of its purification and the late appearance of this polymerase during the cellular SOS response[9], Pol V was not investigated in this thesis. However, it would be of interest to test whether Pol V can co-localize with Pol IIIcore on the clamp on DNA. This would further expand our model of the polymerase competition and test whether it can be applied to all three translesion DNA polymerases in *E. coli*.

The contrast between the highly ordered replication events and the stochastic competition between the replicative and translesion DNA polymerases suggests that the translesion polymerases are not an integral part of the replication machinery. This is supported by the observation that when the gene encoding for Pol II (*PolB*), Pol IV (*DinB*) or Pol V (*UmuC* and *UmuD*) are deleted in *E. coli*, the cells do not show growth defects under normal conditions. In fact, even the triple mutant does not show strong UV-sensitivity[11 3].

In the eukaryotes on the other hand, the deletion of a translesion DNA polymerase can result in severe consequences. For example, the loss of Pol η has been associated with Xeroderma Pigmentosum V, a disease characterized by extreme sensitivity to UV-light and cancer predisposition[5]; and deletion of REV1 has been shown to cause the loss of epigenetic information in chicken DT40 cells[8]. This suggests that the translesion DNA polymerases have more essential roles in the eukaryotes and may therefore have evolved into a more integral part of the replisome, which is supported by studies that report a higher degree of control for the access of the eukaryotic translesion polymerases to DNA. The control is achieved through the ubiquitination of the polymerases and of PCNA, the eukaryotic equivalent of the β clamp[2]. In addition, the translesion polymerase REV1 has shown to be capable of forming bridges between the other polymerases, thus facilitating their recruitment[7]. Despite the higher level of control present in the eukaryotic system, the number of translesion DNA polymerases is also higher and is reported to be approximately 15[6]. The large number of translesion polymerases present implies that in eukaryotes too, the exchange between the polymerases may operate at least in part by concentration-dependent competition. This is supported by studies showing that the intracellular levels of the human translesion polymerases Pol η, Pol κ and Pol ι are increased upon DNA damage[10 13 12]. A fruitful area for future study therefore lies in the investigation of the exchange between the eukaryotic replicative and translesion polymerases using CoSMoS.

The work presented in this PhD thesis has provided crucial insights into the interaction between the various components of the *E. coli* replisome and the different DNA polymerases, and is an important step towards a better understanding of when and how the cell achieves the important balance between fidelity and mutagenesis during DNA replication.

Bibliography

[1] O J Becherel and R P Fuchs. Mechanism of DNA polymerase II-mediated frameshift mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15):8566–8571, July 2001.

[2] Abel C S Chun and Dong-Yan Jin. Ubiquitin-dependent regulation of translesion polymerases. *Biochemical Society Transactions*, 38(Pt 1):110–115, February 2010.

[3] Charmain T Courcelle, Jerilyn J Belle, and Justin Courcelle. Nucleotide excision repair or polymerase V-mediated lesion bypass can act to restore UV-arrested replication forks in Escherichia coli. *Journal of Bacteriology*, 187(20):6953–6961, October 2005.

[4] Robert P Fuchs and Shingo Fujii. Translesion DNA synthesis and mutagenesis in prokaryotes. *Cold Spring Harbor Perspectives in Biology*, 5(12):a012682–a012682, December 2013.

[5] C Masutani, R Kusumoto, A Yamada, N Dohmae, M Yokoi, M Yuasa, M Araki, S Iwai, K Takio, and F Hanaoka. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature*, 399(6737):700–704, June 1999.

[6] Brian S Plosky and Roger Woodgate. Switching from high-fidelity replicases to low-fidelity lesion-bypass polymerases. *Current Opinion in Genetics & Development*, 14(2):113–119, April 2004.

[7] Julian E Sale. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 5(3):a012708–a012708, March 2013.

[8] Peter Sarkies, Charlie Reams, Laura J Simpson, and Julian E Sale. Epigenetic instability due to defective replication of structured DNA. *Molecular Cell*, 40(5):703–713, December 2010.

[9] S Sommer, F Boudsocq, R Devoret, and A Bailone. Specific RecA amino acid changes affect RecA-UmuD'C interaction. *Molecular Microbiology*, 28(2):281–291, April 1998.

[10] Maja T Tomicic, Dorthe Aasland, Steffen C Naumann, Ruth Meise, Christina Barckhausen, Bernd Kaina, and Markus Christmann. Translesion polymerase η is upregulated by cancer therapeutics and confers anticancer drug resistance. *Cancer Research*, 74(19):5585–5596, October 2014.

[11] Bethany Yeiser, Evan D Pepper, Myron F Goodman, and Steven E Finkel. SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proceedings of the National Academy of Sciences of the United States of America*, 99(13):8737–8741, June 2002.

[12] Huifang Zhu, Yanfeng Fan, Hongjuan Jiang, Jing Shen, Hongyan Qi, Ruhuan Mei, and Jimin Shao. Response of human DNA polymerase ι promoter to N-methyl-N'-nitro-N-nitrosoguanidine. *Environmental Toxicology and Pharmacology*, 29(1):79–86, January 2010.

[13] Huifang Zhu, Yanfeng Fan, Jing Shen, Hongyan Qi, and Jimin Shao. Characterization of human DNA polymerase κ promoter in response to benzo[a]pyrene diol epoxide. *Environmental Toxicology and Pharmacology*, 33(2):205–211, March 2012.