THE UNIVERSITY
of EDINBURGH

DNA Double-Strand Break Repair and the Termination of Replication in *Escherichia coli*

Ielyzaveta Iurchenko

Thesis presented for the degree of Doctor of Philosophy

Institute of Cell Biology

The University of Edinburgh

September 2016

Declaration

I hereby declare that this thesis was composed by me, and the research presented is my own, except where otherwise stated. This work has not been submitted for any other degree or professional qualification.

Ielyzaveta Iurchenko

2016

Мама, это тебе...

Спасибо за весь

вложенный в

меня труд.

So Long, and Thanks for All the Fish
D. Adams 'The Hitchhiker's Guide to the Galaxy'

Acknowledgements

First of all, I would like to thank my supervisor David Leach for the incredible opportunity to do my PhD in one of the best Universities in the world. Thanks for bearing with me during the difficult times for me and my country in 2014. Thanks for all the support and guidance through this challenging journey. Thanks for trying to dig out what I wanted to say exactly in that five-page-long sentences.

I would like to thank our beloved senior postdoc Elise Darmon. A huge thank you for the critical reading of everything I ever wrote (even my CV and cover letters). Thanks for being my bench neighbour. Also, thanks for the 'Doctor Who' DVDs!

I would like to thank our junior-postdocs-former-PhD-students Benura Azeroglu and Mahedi Hasan. Mahedi, thank you for the help with WGS data analyses. Especially, for the graphs that needed to be done right here right now. Benura, thank you for everything. If I start writing here about everything good you have done for me, it will be another PhD thesis on friendship. Thanks for meeting me that very first day in Edinburgh and introducing me to hangers, our scientific and casual discussions and the fun life together at Orrok lane! And, of course, thanks for letting me use your unpublished and published data.

I would like to thank our former-postdocs-former-PhD-students Martin White, Charlie Cockram and Julia Mawer. Thanks for allowing me to use your data and introducing me to 2D gels, WGS and ChIP-seq.

I would like to thank my second supervisor Sveta Makovets for the teabreaks, the Belarusian sweets, but most of all for being my mentor. Also, another thanks goes to all past and present members of the Leach lab, the Makovets lab, the Pilizota lab, the El Karoui lab and the Bayne lab. Special thanks goes to incredible Francesca Taglini from 'close to Bologna'. Thanks for your company on all the concerts (especially, the trip to Aberdeen), teaching me how to eat Italian food properly and all the cat videos you shared with my usual suspects.

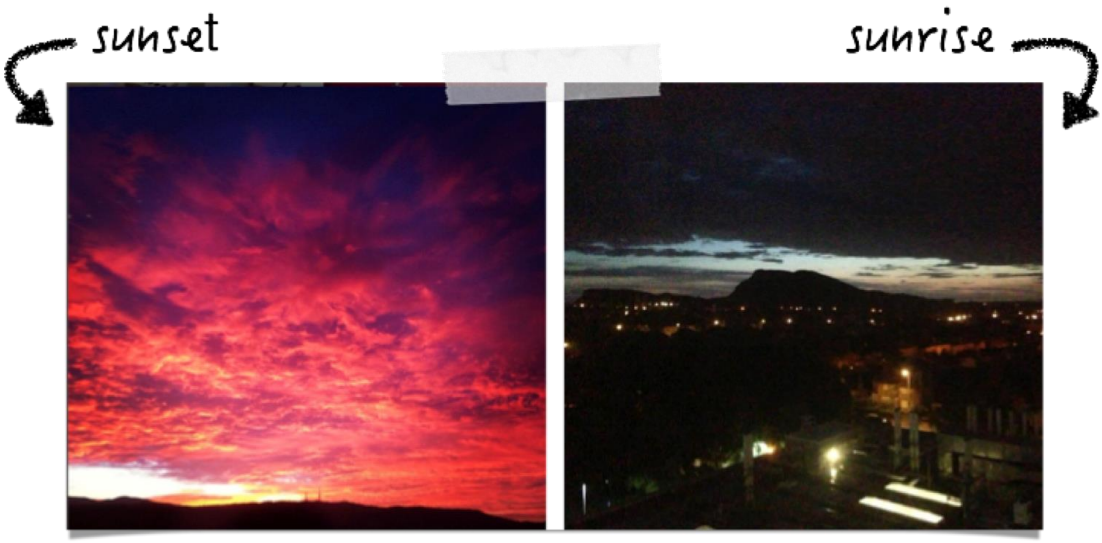
Another huge thanks goes to Delyan Mutavchiev. Thanks for dealing with me and cooking for me (occasionally) during the thesis writing. You'll be there soon as well, remember, I believe in you!

I would like to thank the SBS Scholarship from The University of Edinburgh for funding my degree.

Спасибо маме и дедушке, что, несмотря ни на все, поставили меня на ноги, дали мне образование и отправили меня за границу. Мне очень грустно, что дедушка так и не сможет увидеть этот закономерный конец всех моих многочисленных учеб. Спасибо мой собачке Ляльке, что вытянула меня из пневмонии. Спасибо всем, кто помогал и кто иногда вспоминает добрым словом.

Thanks to everyone who is reading this thesis and if you are doing a PhD, remember: 'Happiness can be found, even in the darkest of times, if one only remembers to turn on the light.' *Harry Potter, J.K. Rowling*.

Finally, thanks to Edinburgh, the best wee city in the world, and to the iconic Darwin building for amazing sunsets and sunrises and occasional Guy Fawkes and Fringe fireworks.



Thanks to everyone who supported me throughout this journey! Спасибо всем!

Щиро дякую! Merci à tous!

Lay Summary

DNA is called the molecule of life. Almost each cell of every organism contains this molecule, which stores the genetic code. In order to pass this information from the old cell to a new cell, DNA has to replicate. Each strand of a DNA double helix can serve as a template for its replication. The successful completion of the replication is critical, as each cell needs to have an exact copy of the DNA. DNA replication is a complex process that consists of three main steps. This work focuses on the last step of DNA replication, called termination. To study the termination of DNA replication I used bacteria called *Escherichia coli* as a model organism. These bacteria have one circular chromosome, which facilitates the study of the DNA replication. DNA replication in *E. coli* is facilitated by two replication machineries which originate from a single origin and move all the way around the chromosome replicating each DNA molecule. Termination starts when two replication machineries arrive to a specific region of the chromosome, called the terminus. The terminus region is also the place where chromosome segregation and dimer resolution occur. Dimer resolution is a very specific phenomenon to circular chromosomes. Dimer formation is one type of chromosome entanglement, which spontaneously occurs during the replication and which is resolved by a protein machinery, called XerCD/*dif*. A second type of entanglement is called catenation and is resolved by specific enzymes called TopoIII/TopoIV.

Here I investigate how all these events are linked to each other and if the interplay between XerCD/*dif* and TopoIII/TopoIV protein machineries could be the cause of additional DNA amplification and DNA damage in the terminus previously observed in Prof. Leach laboratory. Overall, this work sheds light on a very basic yet not fully understood mechanism of such a fundamental process as DNA replication termination.

Abstract

Faithful DNA replication is essential for the maintenance of genetic information. This complex process consists of 3 steps: initiation, elongation and termination. Although the first two steps are quite well understood in both eukaryotes and prokaryotes, many aspects of the termination of replication remain unclear.

Escherichia coli is an ideal organism to study termination of DNA replication. In *E. coli*, DNA replication starts by bidirectional firing of two replication forks from a unique origin and terminates when those forks collide in the terminus region of the circular chromosome. The terminus region is flanked by specific *ter* sequences, which ensure that termination of replication occurs within specific boundaries. Due to the circularity of the *E. coli* chromosome, once the replication is finished the dimers can be formed. To resolve the dimers, the *dif* sequences are aligned together and two chromosomes are then separated into two daughter cells.

Previous members of Prof. Leach laboratory have observed a stimulation of both double-strand break repair (DSBR) and DNA over-replication in the terminus region when DSBR was induced in the *lacZ* locus, half way between the origin and the terminus. In this work, I propose that these two phenomena, elevated levels of DSBR and DNA over-replication, are linked to each other. I confirm that the DSBs arise from the *dif* site and that the *dif* site is the source of DNA over-replication in the terminus. My results suggest that an attempted DSBR at *dif* leads to over-replication between *terA* and *terB*. Here, using next generation sequencing methods, I show that TopoIV and TopoIII topoisomerases introduce breaks in chromosome dimers that were not resolved by the XerCD/*dif* system, leading to DSBR and DNA over-replication.

Abbreviations

^{32}P	Phosphorus-32
ara	Arabinose
ascB::*246bp*	Allele of *ascB* with a 246 bp palindrome
ATP	Adenosine triphosphate
bp	Base pair
ChIP	Chromatin immunoprecipitation
ChIP-qPCR	ChIP-quantitative Polymerase Chain Reaction
ChIP-seq	ChIP-sequencing
Cm	Chloramphenicol
Cm^R	Chloramphenicol resistance cassette
Ct	Cycle threshold
dif	Deletion-induced filamentation
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSB	Double-strand break
DSBR	Double-strand break repair
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
glu	Glucose
HJ	Holliday junction
HR	Homologous recombination
kb	Kilo base pair

Km	Kanamycin
Km^R	Kanamycin resistance cassette
KOPS	FtsK orienting polar sequences
l	Litre
lacZ::*246bp*	Allele of *lacZ* with a 246 bp palindrome
LB	Luria broth
M	Molar
m	Mili
Mb	Mega base pair
MCS	Multiple cloning site
n	Nano
NGS	Next generation sequencing
OD_{600nm}	Optical density at 600 nanometers
ori	Origin
pal	Palindrome
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH	Power of hydrogen
PMGR	Plasmid mediated gene replacement
Pol	Polymerase
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RT-qPCR	Real-Time quantitative polymerase chain reaction
SDS	Sodium dodecyl sulphate

SSB	Single-stranded binding protein
ssDNA	Single-stranded DNA
Suc	Sucrose
Suc^S	Sucrose sensitivity cassette
TAE	Tris acetate-EDTA
Tc	Tetracycline
Tc^R	Tetracycline resistance cassette
ter	Termination site
Tm	Melting temperature
Ts	Temperature-sensitive
tus	Termination utilisation substance
UV	Ultraviolet light
v/v	Volume per unit volume
w/v	Weight per unit volume
WGS	Whole genome sequencing
µ	Micro

Content

Table of Figures

Table of Tables

Chapter I

1. Introduction

1.1 DNA replication

> "It has not escaped our notice that the specific pairing we have postulated immediately suggests a copying mechanism for the genetic material."
> *J. Watson, F. Crick*

DNA, or deoxyribonucleic acid, is a molecule that carries the hereditary information for life. The information in DNA is stored in the sequence of four chemical bases: adenine, guanine, cytosine and thymine. The pairing of adenine with thymine and of cytosine with guanine forms base pairs. A base bound to a sugar and a phosphate becomes a nucleotide. When nucleotides are arranged in two strands, because of the pairing between base pairs, they form a double helix, where the phosphates make a backbone (Watson and Crick 1953). To pass the information from one cell to its progeny, DNA needs to be replicated. DNA replication is a process where each strand of the double helix serves as a template for producing two identical copies of the DNA. It consists of 3 steps: initiation, elongation and termination. DNA replication has to be absolutely perfect to avoid mutations and be completed before a cell segregates its sister chromatids into two daughter cells. The misfunction of this system may cause serious problems, like deseases or cell death. Moreover, DNA is regularly damaged by various agents. Therefore, maintaining of DNA and regular repair of it are absolutely crucial for the cell to survive and pass its genetic information. Despite the high maintenance of DNA by very precise molecular mechanisms such as DNA replication and DNA repair, changes in DNA may occur. Some of these changes occur via DNA recombination. The interplay between DNA replication, repair and recombination with a focus on DNA replication termination using *Escherichia coli* as a model system will be the scope of this thesis.

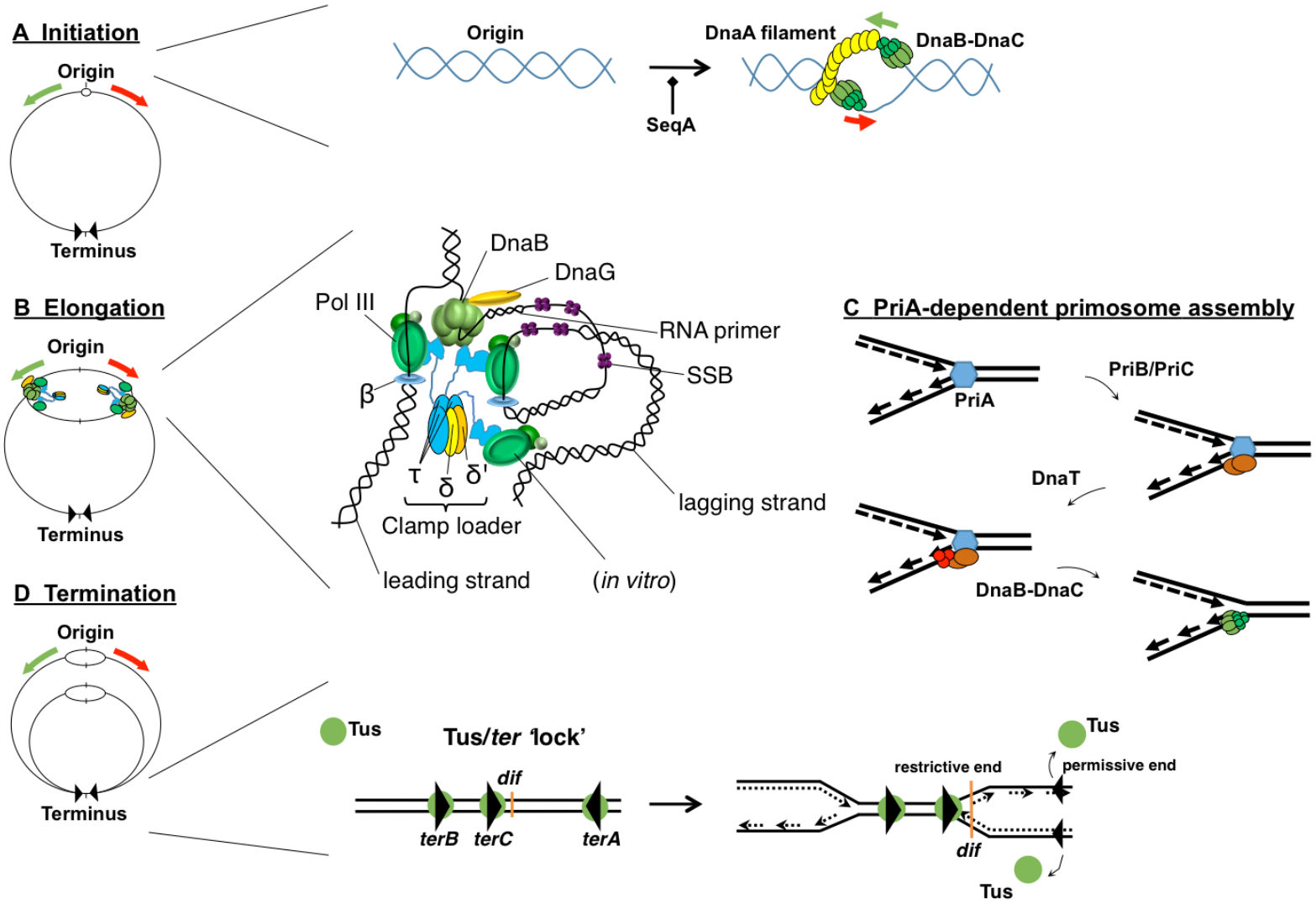

A Initiation
Origin
Terminus
Origin
DnaA filament
DnaB-DnaC
SeqA
B Elongation
Origin
Terminus
DnaB
DnaG
Pol III
RNA primer
SSB
β
τ
δ
δ'
Clamp loader
lagging strand
leading strand
(in vitro)
C PriA-dependent primosome assembly
PriA
PriB/PriC
DnaT
DnaB-DnaC
D Termination
Origin
Terminus
Tus
Tus/ter 'lock'
dif
terB
terC
terA
Tus
restrictive end
permissive end
dif
Tus

Figure 1.1. DNA replication in *E. coli*.

(A) Initiation of replication; (B) Elongation of replication, modified from (McInerney *et al*. 2007) (C) PriA-dependent primosome assembly and (D) Termination of replication. The direction of replication is shown using red and green arrows. When the DNA replication is initiated, origin is opened up by DnaA. DnaA forms a filament and interacts with one of the strands. Then, DnaA interacts with the DnaB helicase, and with the helicase loader DnaC. DnaA and DnaC together load DnaB on the lagging strand template. DnaA binding to the origin is regulated by the SeqA protein that inhibits DnaA filament formation. DnaG primase transiently binds the DnaB helicase and helps to release DnaC. When DnaG synthesises RNA primer, clamp loader loads a β-clamp on it. Finally, two (three *in vitro*) DNA polymerase III holoenzymes bind to this complex. One of the Pol III synthesises the continuous leading strand, another Pol III has to disassemble each time after it synthesised Okazaki fragment of the lagging strand and assemble to synthesise a new Okazaki fragment. This is where, *in vitro*, the presence of a third Pol III was proposed. Then, RNA primers are removed, a low processivity DNA polymerase I fills the gaps and DNA ends are ligated by DNA ligase. If the progression of the replication fork is compromised, DnaB disassembles and has to be loaded again by PriA-dependent primosome assembly. In this pathway the PriA protein loads to a 3-way junction, then either PriB or PriC is loaded. DnaT binds to this complex and loads DnaB together with DnaC. To ensure that replication termination occurs in a specific place on the chromosome, the replication forks are paused in a polar manner by the Tus/*ter* 'lock'. When the fork arrives from the permissive end, the DnaB helicase diassembles Tus from the *ter* sequence and the fork can pass. If the fork arrives from the non-permissive end, it will encounter a Tus/*ter* block that will pause the fork. It is not known if DnaB disassembles from the fork when reaches a non-permissive end of Tus/*ter*.

1.1.1 The Beginning: replication initiation

Replication starts at a specific place on the chromosome, called the origin. *E. coli*'s 4.6 Mb chromosome is circular and has a single unique origin, *oriC*, at which two replisomes assemble to duplicate the genome bidirectionally (Reyes-Lamothe *et al*. 2008). Replisome assembly is highly regulated within the cell cycle. Unlike eukaryotic organisms, *E. coli*'s origin is a well-defined sequence (Leonard and Méchali 2013). The initiation of DNA replication starts when multiple copies of initiator protein DnaA, bound to ATP, interact with 9-bp repeats upstream of *oriC*, DnaA boxes. It allows the origin's 13-mer AT-rich motifs to be opened up (Kaguni 2011; McGarry *et al*. 2004). DnaA forms a right-handed helical filament and directly interacts with one of the strands of the unwound region (Abe *et al*. 2007). After the unwinding of *oriC*, DnaA loads

the replisome components required for replication initiation and fork progression (Kaguni 2011; Reyes-Lamothe, Nicolas, and Sherratt 2012; Messer and Messer 1987). DnaA interacts with DnaB, the replicative helicase that forms a complex with DnaC, a helicase loader. DnaC also interacts with DnaA, facilitating the loading of the DnaB helicase (Kaguni 2011). DnaA loads two DnaB-DnaC complexes on ssDNA: one on the lower strand, second on the top strand of the unwinded 'bubble' (Figure 1.1 A). After DnaB is loaded onto ssDNA, DnaC is dissociated from the DNA to enable movement of the helicase. The DnaB helicase is a homohexamer with a ring-like structure that moves in the 5'-3' direction along the DNA onto which it is loaded. Due to this directionality of the helicase, it is placed on the lagging strand template of the replication fork, which is synthesised as Okazaki fragments (Balakrishnan and Bambara 2013). Okazaki fragments are 1-2 kb long newly formed DNA fragments on the lagging strand template only. When DnaB moves in one direction or another from the origin, it interacts with the primase, DnaG, and forms a ~12 nucleotides primer, which is needed to begin DNA replication (Frick and Richardson 2001). DnaG is a monomeric protein that transiently binds the DnaB helicase and helps to release DnaC. When the RNA primer is synthesised for the leading strand, the clamp loader loads a ring-shaped β-sliding clamp, which increases the polymerase's processivity (Johnson and O'Donnell 2005). Finally, DNA polymerase III holoenzyme binds to this complex and synthesis begins.

Normally, in rich medium, *E. coli* reinitiates DNA synthesis before the previous round of replication has been completed. To ensure that a second initiation does not occur immediately after the first one, DnaA binding at the origin is regulated by the SeqA protein. This protein can only bind to abundant hemimethylated GATC sites and by doing so inhibits DnaA filament

formation and delays remethylation by the Dam methylase (Dame, Kalmykowa, and Grainger 2011).

1.1.2 The Middle: replication elongation

1.1.2.1 Replication elongation

Replication is performed by a complex machinery, called the replisome, that contains DNA polymerases, sliding clamps, a clamp loader, a helicase, a primase and SSB (single-strand binding) proteins (Johnson and O'Donnell 2005) (Figure 1.1 B). *E.coli*'s genome is replicated in approximately 40 minutes at a rate of 1 Kb/sec. The pentameric clamp loader contains 3 γ subunits, 1 δ subunit and 1 δ' subunit, and connects DNA polymerases, the DnaB helicase and β-sliding clamps. DnaG, bound to DnaB, synthesises the RNA primer, which serves as a loading platform for the β-sliding clamp. Finally, the replisome is assembled by recruiting two replicative DNA polymerase III holoenzymes (Pol III), where each DNA polymerase consists of the α, ε and θ subunits. Pol III interacts with DnaB helicase via two τ subunits of the replisome (Johnson and O'Donnell 2005). It was determined that two (or three *in vitro*) polymerases are assembled in most replisomes and are performing simultaneous activity (McInerney *et al*. 2007). One polymerase is continuously replicating the leading strand template with high processivity; the other polymerase is involved in the more complicated replication of the lagging strand template. The polymerase on the lagging strand template is associated with a β clamp that is loaded at every RNA primer, which are separated by ~1.5 kb. When during replication the polymerase reaches the downstream primer, it must dissociate from the DNA in order to start the extension of a new Okazaki fragment. This dissociation takes about 1 sec, which is long enough for a polymerase on the leading strand template to advance further (Leu, Georgescu, and O'Donnell 2003). Also, the lagging strand template is

coated with single-strand binding proteins (SSB), which might pause the lagging strand DNA polymerase. SSB bound to ssDNA facilitates the dislodging of the primase from the primed site. Constant loading and dislodging of a DNA polymerase on the lagging strand template would slow down the DNA polymerase on the leading strand template. Therefore, it was proposed (and determined *in vitro*) that the second polymerase is involved in synthesising the lagging strand template. When the first polymerase dissociates from the β clamp, the second DNA polymerase could begin the extension of the next RNA primer (O'Donnell, Langston, and Stillman 2013; McInerney *et al*. 2007).

Finally, after Okazaki fragments synthesis, primers are removed by RNase H. The gaps between Okazaki fragments are filled by DNA polymerase I and DNA ends are ligated by DNA ligase.

1.1.2.2 PriA-dependent replication fork restart

When the progression of a replication fork is compromised, the replication fork is stalled, which becomes a hotspot for recombination events (Bierne 1991; Horiuchi *et al.* 1994). Replication forks could be stalled because of various factors, such as when encountering Lac or Tet repressors bound to their operators (Payne *et al.* 2006), or when encountering the ribosomal RNA transcriptional machinery (French 1992; Boubakri *et al.* 2010), or when the transcriptional machinery is stalled at GC rich sequences (Krasilnikova *et al*. 1998), or when encountering a Tus/*ter* block (described below). Stalling of replication forks leads to the disassembling of the replicative helicase and fork collapse. If replication forks are not restarted, eventually, it will lead to cell death.

To restart stalled replication forks and load DnaB in an *oriC* sequence-independent manner, the cell has evolved a mechanism called the PriA-dependent primosome assembly (Sandler and Marians 2000). PriA is a 3′ to 5′ DNA helicase that can recognise branched DNA structures. PriA, as studied *in vitro*, displays several roles in the cell: facilitation of DnaB loading onto the lagging strand template and unwinding of the lagging strand template for DnaB to generate single-stranded DNA (not confirmed *in vivo*). In this reaction, the loading of DnaB involves a series of complex protein-protein interactions through PriA, PriB/PriC, DnaT and DnaC (Sandler 2000). The roles of DnaT and PriC remain poorly understood. The growth of *priB* and *priC* single mutants is less affected than the *priA* mutant, which is hardly viable. In a PriA-PriB-DnaT reaction, first, PriA binds to ssDNA in a three-way junction (stalled fork). Then, PriB is bound to this protein-DNA complex, followed by DnaT. Afterwards, DnaT attracts DnaC that loads DnaB helicase (Ng and Marians 1996) (Figure 1.1 C).

1.1.3 The End: replication termination

1.1.3.1 Tus/*ter* 'lock' formation

As the two replication forks move on a circular chromosome in opposite direction from the origin, they meet and collide on the opposite side of the chromosome, at the terminus (Hill 1992). No essential genes are present in this part of the chromosome, so if removed, cells will still be viable (Henson and Kuempel 1985). While initiation and elongation in bacteria have been fairly well understood, the termination of replication remains mysterious in both prokaryotes and eukaryotes.

When DNA replication is complete, either one or both replisomes have to disassemble to prevent over-replication (Hiasa and Marians 1994). In

principal, the colliding of replication forks could occur anywhere in the chromosome but, in *E. coli*, the replication is directed towards a designated region. This ensures that the replichores are of the same size and that termination occurs in the region where other crucial events happen, like chromosome dimer resolution, chromosome decatenation and segregation of chromosomes. The convergence of replication forks in the terminus occurs within a ~200 kb region flanked by *ter* sites. *ter* sites are ~22-23 bp DNA sequences with a highly conserved 11 bp core sequence that are specifically recognised and bound by the DNA-binding Tus protein (Hill *et al*. 1988). *ter* sites, when bound by Tus, are able to arrest the progression of a replication fork in an orientation-specific manner by preventing the action of the DnaB helicase (Hill *et al.* 1987; Mulcair *et al.* 2006; Hill & Marians 1990; Bastia *et al.* 2008). When DnaB unwinds the duplex DNA from the permissive end of the Tus/*ter* complex, it displaces Tus protein from the *ter* site. If DNA replication of one of the replichores is delayed (e.g. due to ongoing repair on that replichore), the fork on the other replichore will encounter the non-permissive side of one of the *ter* sites bound by the Tus protein and pause there (Figure 1.1 D). It is not known whether the replisome disassembles when it arrives at Tus/*ter.* The *ter* sites can be divided into two groups: the ones that pause anti-clockwise moving replication forks and the ones that pause clockwise moving replication forks (Figure 1.2). Fourteen *ter* sites were identified in the *E. coli* chromosome by marker frequency assays or by sequence similarity (*terA – terL, terY* and *terZ*) (Duggin and Bell 2009a). Four of these *ter* sites are hypothetical and were found by sequence similarity (*terK*, *terL*, *terY* and *terZ*). Ten other *ter* sites, *terA – terJ,* are primary sites and are distributed over a 2,800 kb region. It was determined that paused replication forks were detected at each site except the *ter* sites identified by sequence similarity (Duggin and Bell 2009a). The biggest proportions of paused replication forks were identified at

terA, *terB* and *terC* sites. *terC* is located next to *dif*, ~180° from *oriC*, and is the first *ter* site to encounter replication forks arriving from the right replichore. *terA* and *terB* sites are also significantly used, although the *terB* site is oriented in a such way that it will pause forks arriving after they encountered *terC*. When the distance from *oriC* to *terB* was doubled compared to the distance from *oriC* to *terA*, it was observed that both *terA* and *terB* were still significantly used (Pelletier, Hill, and Kuempel 1988). Therefore, it was suggested that adjustments of the replication rates of individual replichores compensated for the difference and allowed both replication forks to arrive at the same time in the terminus region.

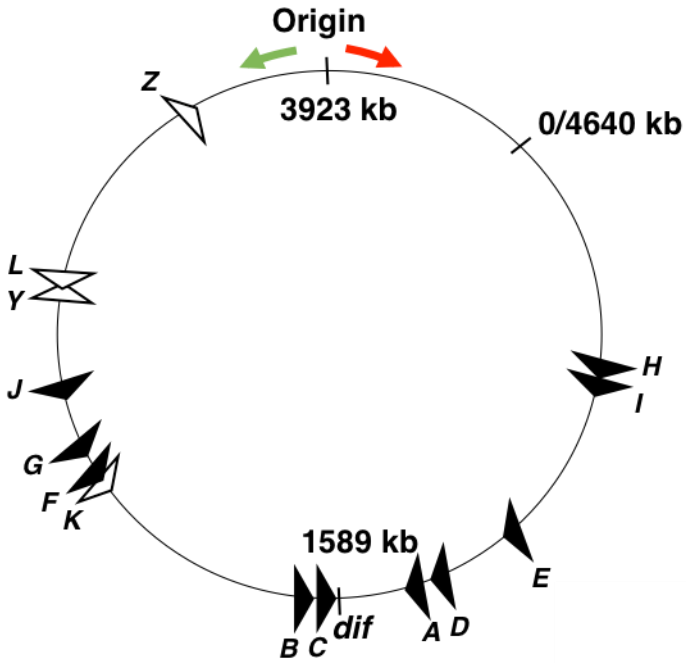


Figure 1.2. Location and orientation of all 14 *ter* sites in *E. coli*.

terH, *terI*, *terE*, *terD*, *terA*, *terY* and *terZ* are oriented in such a way that they would block anti-clockwise moving forks; *terC*, *terB*, *terK*, *terF*, *terG*, *terJ* and *terL* are oriented in such a way that they would block clockwise moving forks. *oriC*, *dif* and *tus* positions are indicated. The direction of replication is shown using red and green arrows. Origin and *dif* are indicated. *ter* sites marked as black triangles were shown to efficiently pause replication forks, *ter* sites marked as white triangles were shown to be very weak. Adapted from (Duggin and Bell 2009b).

Also, it has been shown that replication forks can pass through the non-permissive side of Tus/*ter* block with the help of the UvrD helicase (Bidnenko, Lestini, and Michel 2006). Another reason for the 'break-through' could be that the *ter* sites differ by affinity to the Tus protein. *terK*, *terL*, *terY* and *terZ* have the lowest affinity for Tus and the *terB* site has the highest affinity for this protein (Coscun-Ari 1997; Moreau and Schaeffer 2012). The significant fork pausing at *terB* is a consequence of a failure of *terC* to form a Tus/*ter* 'lock' in 15% of the time (Duggin and Bell 2009a; Moreau and Schaeffer 2012). Interestingly, the *terB* site is located immediately upsteam of the *tus* gene and when the Tus protein binds to *terB*, it occludes the RNA polymerase and represses the transcription of the *tus* gene, which might be another reason for the 'break-through' at other *ter* sites.

Additionally, it has been shown that the *ter* sites stimulate homologous recombination, therefore, replication fork arrest could be a source of genome instability (Rothstein, Michel, and Gangloff 2000). Arrest of a replication fork leads to the formation of a Y-structure with a single strand gap and Bidnenko and collaborators proposed that the replication forks from the next round of replication would arrive to the stalled fork and would create free double-strand ends (Bidnenko, Ehrlich, and Michel 2002). These double-strand ends then would be substrates for RecBCD-mediated DSB repair via homologous recombination. If cells are defective for homologous recombination, it would cause cell death.

1.1.3.2 Recent models of replication termination in *E. coli*

How termination of replication in *E. coli* occurs is still not clear. Recently, some models were proposed that involved advanced technology like whole genome sequencing of the *E. coli* genome that uncovered new features of replication termination (Rudolph *et al*. 2013; Wendel, Courcelle, and Courcelle 2014).

These groups studied the role of the RecG protein in the termination of replication. RecG is a DNA helicase and translocase that works on a D-loop and is involved in PriA-dependent replication restart. The *recG* mutant causes an over-replication in the terminus region. This phenotype is suppressed by the PriA helicase activity mutation (*priA300*).

Rudolph and collaborators proposed a model where collision of replication forks in the terminus leads to the formation of new divergent forks through PriA-mediated replisome assembly and RecBCD recombination (Figure 1.3 A). According to this model, when replisomes meet, the DnaB helicase displaces the nascent leading strand of the opposing fork and generates a 3′ flap. Then, this flap would be degraded by 3′ ssDNA exonucleases or unwound by RecG and converted into a 5′ flap that would be immediately degraded by a 5′ exonuclease. If RecG is absent, the 3′ flap will be a substrate for PriA and will result in a PriA-dependent replisome assembly that will set up a fork moving to a *ter* site. The double-strand end of this fork will recombine and will set up a fork moving the opposite direction to another *ter* site. The *ter* sites will pause the forks which will create an over-replication of the terminus region (Rudolph *et al*. 2013).

Wendel and collaborators proposed that converging replication forks pass each other without replisome disassemble and create an over-replication of the terminus (Figure 1.3 B). This leads to a transient intermediate that is unwound by RecG. Then, RecBCD in a RecA-independent manner with the help of SbcCD and ExoI exonucleases promotes the degradation of the over-replicated region. Finally, the gaps are ligated.

The latter model proposal was influenced by earlier foundings that claimed the presence of additional origins, *oriM*, *oriX* and *oriK*, in the terminus (Asai *et al*. 1993; de Massy *et al.* 1984; Kogoma 1997; Magee *et al.* 1992; Maduike,

Tehranchi, Wang 2014). These cryptic origins were proposed to be DNA-damage or RNaseHI-deficiency inducible. The presence of these origins was proposed because under various stress conditions, it was noticed that the DNA amount was increased, particularly in the terminus region between *terA* and *terB* sites. Therefore, Wendel and collaborators proposed that the observed increase in DNA copies would be created not by these origins but by over-replication of the terminus (Wendel, Courcelle, and Courcelle 2014).

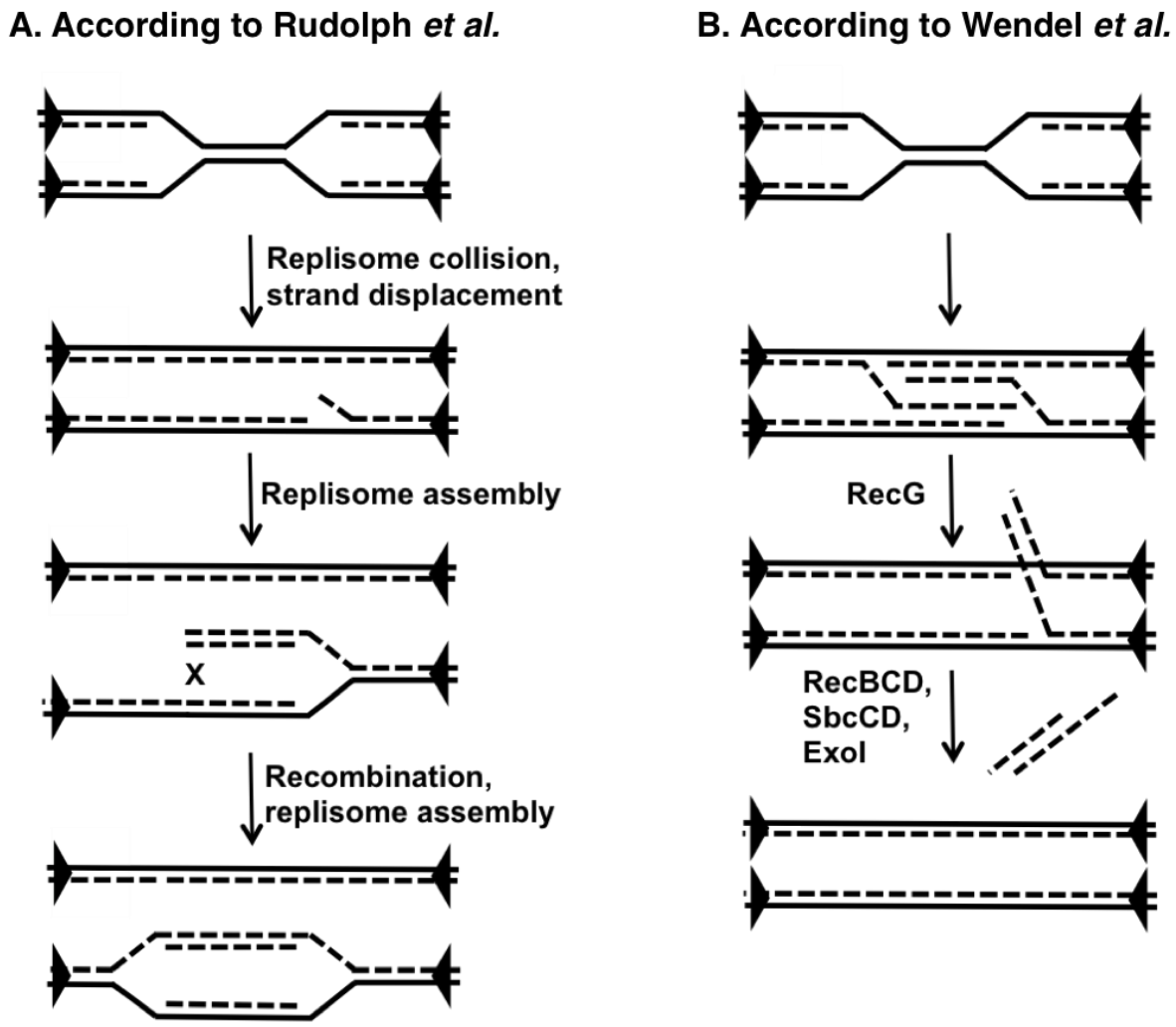


Figure 1.3. Proposed models for termination of replication in *E. coli*.

(A) According to Rudolph *et al., w*hen two replisomes meet, they disassemble leaving a 3' overhang. This overhang becomes a substrate for PriA-mediated replisome assembly and RecBCD-mediated recombination. (B) According to Wendel *et al.,* converging replication forks pass each other leaving over-replicated region. RecG unwinds the over-replicated intermediates. RecBCD and other exo- and endonucleases degrade and resolve the over-replicated region. Adapted from (Rudolph *et al.* 2013)*,* (Wendel, Courcelle, and Courcelle 2014)*.*

1.1.3.2 Termination in other organisms

1.1.3.2.1 Why do Tus/*ter* 'locks' exist?

So far, what is the advantage of Tus/*ter* systems and why do *ter* sites exist in bacteria remain a mystery. Cells without functional Tus/*ter* traps were reported to be perfectly viable, which suggests that the termination of replication can occur at any position in the chromosome. Replication fork arrest mechanisms similar to Tus/*ter* traps exist across diverse bacteria, including *Bacillus subtilis* (RTP/*ter*) and *Salmonella typhimurium* (Hill 1992). The Tus/*ter* block seems to be advantageous for circular chromosomes, like in most bacteria, due to the specific processes that have to be resolved in circular chromosomes, like chromosome dimers. Here, Tus/*ter* blocks ensure that the chromosome dimer is trapped near to the *dif* site needed for chromosome dimer resolution. Especially, it is required in rapidly growing bacteria, where termination of replication, chromosome dimer resolution, decatenation and cell division have to be co-ordinated (Duggin *et al*. 2008).

1.1.3.2.2 Termination in archaea

The archaeon *Sulfolobus solfataricus* has a single circular chromosome that has three bidirectional origins of replication that all fire at the same time in every cell cycle (Duggin, McCallum, and Bell 2008; Lundgren *et al*. 2004). However, only one *dif* site exists in this archaeon (Figure 1.4). This means that, unlike in *E. coli*, the termination of replication does not occur at *dif* but more like in eukaryotes stochastically between origins of replication. Indeed, Lundgren and collaborators used MFA (marker frequency analysis) to reveal that replication forks meet mid-way between the origins (Lundgren *et al.* 2004). Moreover, chromosome dimer resolution and termination of replication are spatially and temporally separated in archaea. *S. solfataricus* maintains its

chromosomes in a cohesed state for an extended period of time after replication and does not segregate chromosomes until very late in the cell cycle. This means that dimer resolution does not occur until the segregation machinery starts acting. These observations and the absence of *ter* sites in archaea (Duggin, Dubarry, and Bell 2011) suggest that dimer resolution and termination of replication do not overlap. This suggests that most probably the *ter* sites exist in bacteria to efficiently coordinate all complex processes that occur in the terminus region of a bacterial chromosome.

1.1.3.2.3 Termination in eukaryotes

Overall, eukaryotic cells undergo relatively uncontrolled termination of replication. Multiple origins that are located on a linear chromosomes fire bidirectionally throughout S phase. It was shown that termination occurs randomly when two forks fuse within a 4-5 kb zone (Greenfeder and Newlon 1992). It makes eukaryotic termination difficult to study. However, similar systems to Tus/*ter* were identified in yeast and peas (Brewer and Fangman 1988; Hernandez, Lamm 1988). 71 chromosomal termination regions (*TERs*) with an average length of 5 kb were identified in yeast. This region contain *TER* fork pausing elements that bind Top2 DNA topoisomerase (TopoIV in *E. coli*) (Fachinetti *et al*. 2010; Alver and Bielinsky 2010). According to this study, the topoisomerase facilitates fork collision at *TERs* together with the Rrm3 helicase that can overcome these fork progression barriers. This finding highlights the importance of replication fork barriers. Additionally, in rDNA locus, replication fork barriers (RFBs) exist that can stall replication forks in an orientation-dependent manner by interacting with Pol III, similarly to Tus/*ter*. RFBs pause the fork that moves opposite to the direction of transcription. However, these barriers destabilise the fork and could lead to the breakage and genomic rearrangements.

In eukaryotes, loading of the replicative helicase, hexamer Mcm2-7 (DnaB in *E. coli*), activated by recruited Cdc45 (RecJ in *E. coli*) and GINS, occurs during the G1 phase of the cell cycle (Yardimci and Walter 2014). The active Cdc45-Mcm2-7-GINS (CMG) complex forms the foundation for other components of the replisome to assemble (Tanaka 2013). Then, the assembled replisome replicates the DNA until it encounters another replisome and terminates replication. Until 2014 it was unknown if the replisome disassembles. In 2014, Moreno and collaborators and Maric and collaborators proposed a model for the disassembling of a eukaryotic replicative helicase during termination using budding yeast as a model organism (Maric *et al*. 2014; Moreno *et al.* 2014). These groups were studying the ubiquitylation of the replisome. They discovered that the Mcm7 subunit is polyubiquitylated with lysine 48 linking ubiquitins. Inhibiting replication prevented Mcm7 ubiquitylation. Blocking Mcm7 ubiquitylation caused an inhibition of the release of CMG complexes from the chromatin (Moreno *et al*. 2014). Therefore, polyubiquitylation of Mcm7 is required to disassemble the replisome following the release of the replicative helicase from the DNA during termination. The disassembly of the replisome occurs by Cdc48 ATPase. It specifically associates with ubiquitylated CMG complexes and acts by protein unfolding. Mcm7 gets removed from the CMG complex, disrupting the links in Mcm2-7 and opening up the hexameric ring, which, presumably, releases Cdc45 and GINS (Figure 1.5).

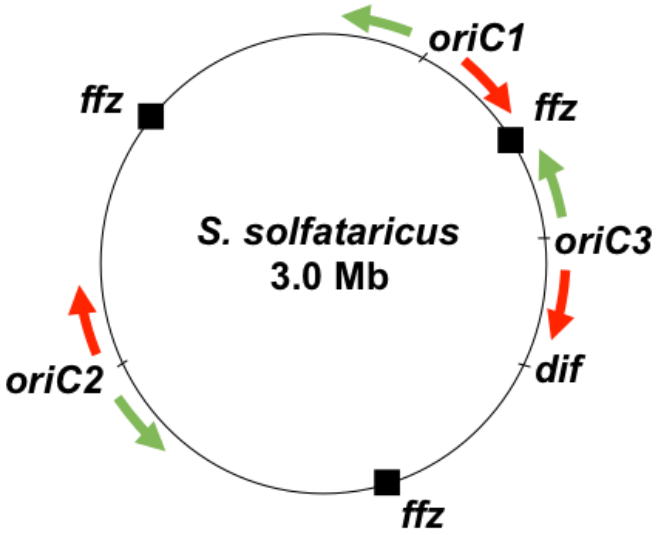


Figure 1.4. Replication organisation of the *Sulfolobus sulfataricus* chromosome.

Origins and the *dif* site are indicated. *ffz* stands for fork fusion zone. The direction of replication is shown using red and green arrows. Modified from (Lundgren *et al.* 2004)

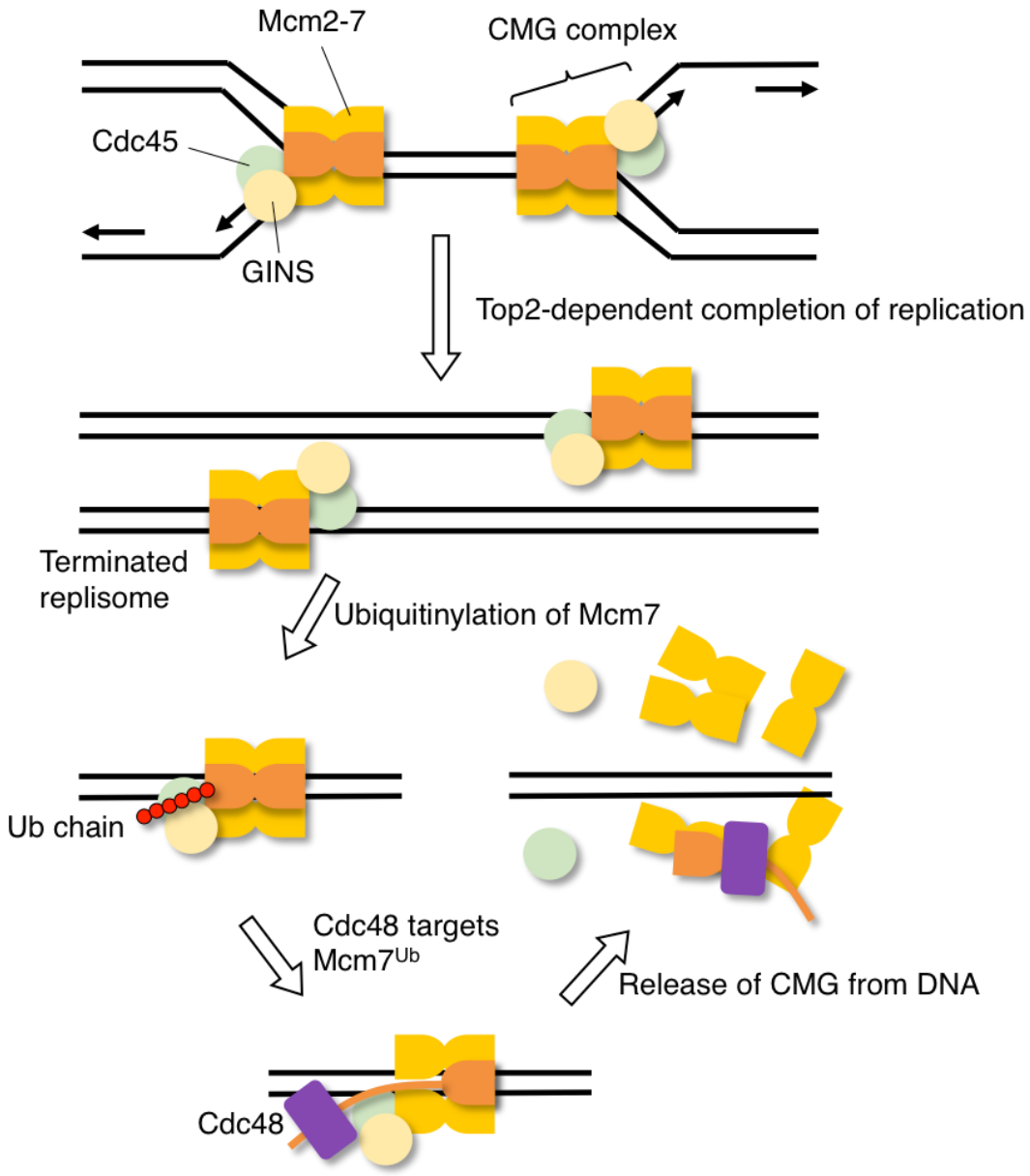


Figure 1.5. Termination of replication in eukaryotes.

When replication is completed, replisomes are targeted by Cdc48, resulting in CMG disassembly. Ub stands for ubiquitin. Adapted from (Maric *et al.* 2014; Moreno *et al.* 2014).

1.2 Chromosome dimer resolution by XerCD/*dif*

In earlier work, Lawrence and collaborators suggested, using a bioinformatic approach, that DNA replication termination would occur at the *dif* site (Lawrence 2007). The proposed model suggested that a nick would occur at *dif*, which would result in a double-strand break upon the arrival of forks. Duggin and collaborators (Duggin and Bell 2009a) demonstrated that a significant proportion of termination events occur at the *terC* site and argued against a possible termination at *dif* where no paused forks and less termination structures were detected.

The conserved *dif* site is a specific 28 bp DNA sequence located between *terA* and *terC* sites 18 kb from *terC.* It is required for the completion of crucial cellular processes, like chromosome decatenation and dimer resolution (Blakely, Davidson, and Sherratt 1997; El Sayyed *et al.* 2016). In fast growing bacteria, termination of replication, chromosome dimer resolution, chromosome decatenation, chromosome segregation and cell division all overlap spatially and temporally. An odd number of crossover events between sister chromosomes in a circular *E. coli* chromosome will generate a chromosome dimer. This dimer must be resolved accurately to generate two monomers, so that each daughter inherits a complete copy of genes. The *dif* locus is where site-specific XerCD recombination takes place to resolve a chromosome dimer (Blakely, Davidson, and Sherratt 1997) (Figure 1.6 A). A ring-shaped DNA translocase called FtsK, composed of 3 domains and anchored to the septum, translocates the DNA in the *oriC-dif* direction, so that two *dif* sites can be aligned together (Bigot *et al.* 2006). The directionality of the FtsK motor is controlled by KOPS (FtsK-orienting polar sequences, GGGNAGGG) that are located along the whole chromosome (Bigot *et al.* 2005). When FtsK aligns two *dif* sites together, it activates XerCD recombination

through a direct interaction with the pair of XerD recombinases that in their turn activate the pair of XerC recombinases (Aussel *et al.* 2002) (Figure 1.6 B). This process through cleavage and religation allows an exchange of strands between the two *dif* sites which yields two identical monomeric chromosomes. This recombination reaction is highly regulated to ensure proper chromosome dimer resolution through FtsK at the location of *dif*. FtsK also plays a crucial role in coordinating the late stages of cell division and chromosome segregation through the activities of its N-terminal transmembrane domain ($FtsK_N$) and its cytoplasmic C-terminal domain ($FtsK_C$), respectively (Sherratt, Lau 2001). $FtsK_N$ is attracted to the septum by FtsZ, and helps to recruit further division proteins to form a divisome. $FtsK_N$ is an essential domain, its absence results in an impared septum and is lethal for the cell (Barre *et al*. 2001; Steiner and Kuempel 1998). $FtsK_C$ is from the "ATPase associated with various activities" superfamily of proteins (AAA), therefore, the action of this protein requires ATP hydrolysis. The activity of $FtsK_C$ is independent of the DNA substrate topology and is required to switch the state of activity of XerC and XerD, so that the recombination reaction is initiated by XerD.

Chromosome dimer resolution coincides with the formation of a division septum by a tubulin-like protein, FtsZ (Löwe and Amos 1998). When the septal region is free of most chromosomal DNA, the formation of a septal ring begins. Precise chromosome positioning and segregation determines where the septum will form. A delayed termination of replication results in a septal ring formation and 'guillotining' the DNA at the terminus region leading to incompletely segregated chromosome (Hendricks *et al*. 2000).

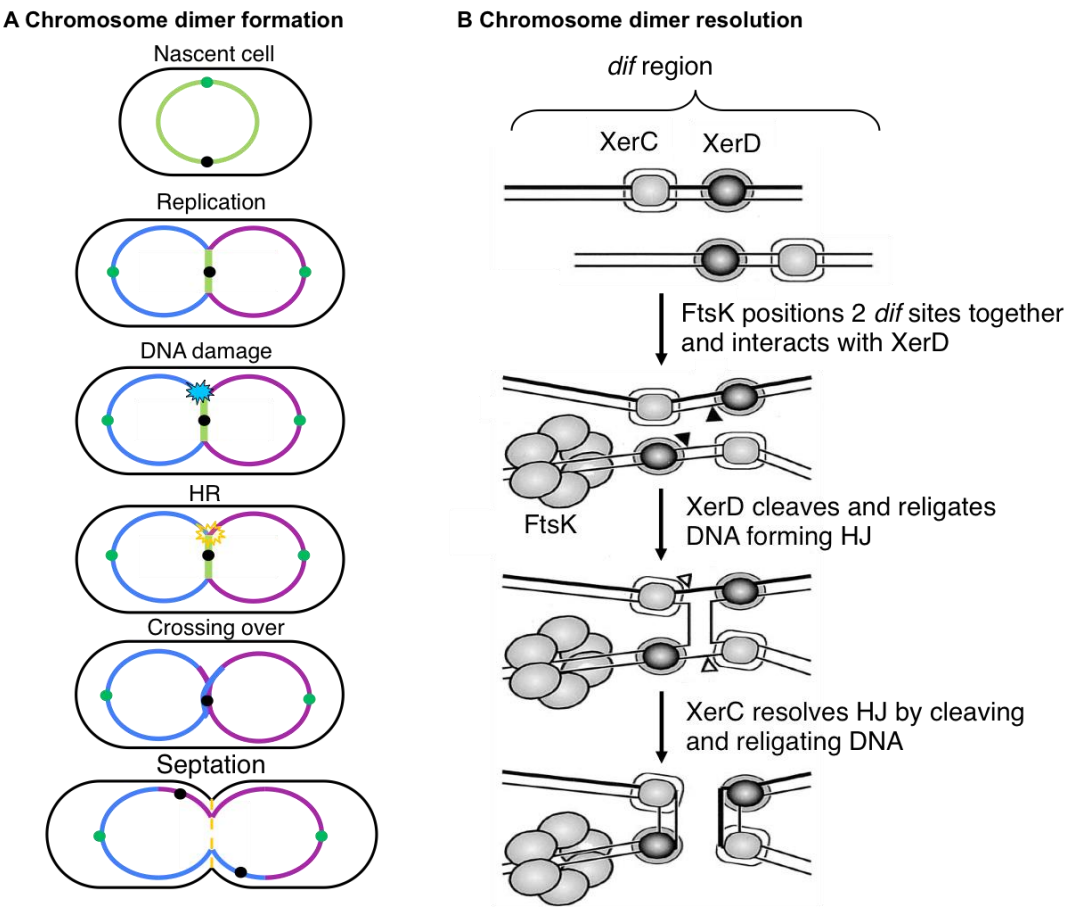


Figure 1.6. Chromosome dimer formation and resolution.

(A) Chromosome dimer forms due to the odd number of crossovers in the chromosome. (B) XerCD recombinases operate at the *dif* region to resolve chromosome dimers. First, FtsK brings *dif* sites together, second, FtsK attracts XerD to *dif* sites. XerD recombination creates Holliday junctions. Finally, XerD interacts with XerC that resolves Holliday junctions leaving two monomers. Adapted from (Aussel *et al*. 2002).

1.3 Chromosome decatenation by TopoIII and TopoIV

In addition to chromosome dimer resolution, complete segregation of sister chromosomes requires the removal of all catenation links (Figure 1.7). Catenation is the process when two fully replicated double-stranded chromosomes are still linked together. To resolve this, a double-strand break is required. Although, it is not the main mechanism of chromosomes decatenation, the XerCD/*dif* system is partially responsible for this (Grainge *et al*. 2007).

During replication, positive supercoiling accumulates in front of the forks, while precatenation (interwinding of replicated strands) occurs behind the replication forks (Wang, Reyes-lamothe, and Sherratt 2008; Espeli *et al*. 2003). Positive supercoiling ahead of the replication forks is removed by the type II topoisomerase II, the DNA gyrase. However, some of the positive tension ahead of the replication forks diffuses behind the forks, resulting in the appearance of precatenanes. This type of DNA interwrapping is removed by the type II topoisomerase IV (TopoIV). TopoIV is the main decatenating enzyme and is essential for cell survival. It is composed of four proteins, a ParC dimer, the catalytic subunit, and a ParE dimer, the ATPase subunit. Interestingly, ParC is associated with the replication factory whereas ParE is distributed in DNA-free spaces in the cell. Temperature-sensitive mutations in either *parE* or *parC* are conditionally lethal, leading to a large DNA mass accumulation in the centre of an elongated cell. TopoIV uses a double-strand passage mode of action, where it breaks, passages and rejoins double-stranded DNA. This mode of action is similar to gyrase. The main difference that distinguishes the functions of these enzymes is that the gyrase wraps the DNA around itself while TopoIV does not (Peng and Marians 1995). This gives gyrase a poor decatenating activity and great supercoiling and relaxing

activities. TopoIV can recognise crossovers, which explains why the decatenating activity of TopoIV is greater than its relaxing activity (Zechiedrich and Cozzarelli 1995). It was shown that the ParC subunit of topoisomerase IV interacts with FtsK through C-terminal domain 3 to focus unlinking of chromosomes at *dif* (El Sayyed *et al*. 2016; Espeli, Lee, and Marians 2003).

Accumulated catenanes can oppose DNA unwinding during the final steps of the termination of replication. To partially overcome this problem, the RecQ helicase specifically unwinds the remaining DNA and stimulates the type I topoisomerase III (TopoIII) to unlink dsDNA (Harmon, DiGate, and Kowalczykowski 1999; Suski and Marians 2008). TopoIII is encoded by the *topB* gene and is not essential for the cell to survive (DiGate and Marians 1989). Interestingly, Top3 (homologue of TopoIII) in eukaryotes is essential for cell survival and plays a significant role in the maintenance of genomic integrity (Wallis *et al.* 1989). TopoIII was originally identified to have a superhelical DNA relaxing activity and, then, it was shown to be able to segregate replicating daughter DNA molecules *in vitro* (DiGate and Marians 1988). It was also shown that TopoIII can act as a main decatenase *in vivo* but only when overexpressed (Nurse *et al.* 2003). Normally, TopoIII has a decatenation activity that requires a presence of ssDNA substrate, because type I topoisomerases can only break ssDNA. Therefore, TopoIII alone does not effectively decatenate chromosomes and TopoIV is the main decatenation machinery.

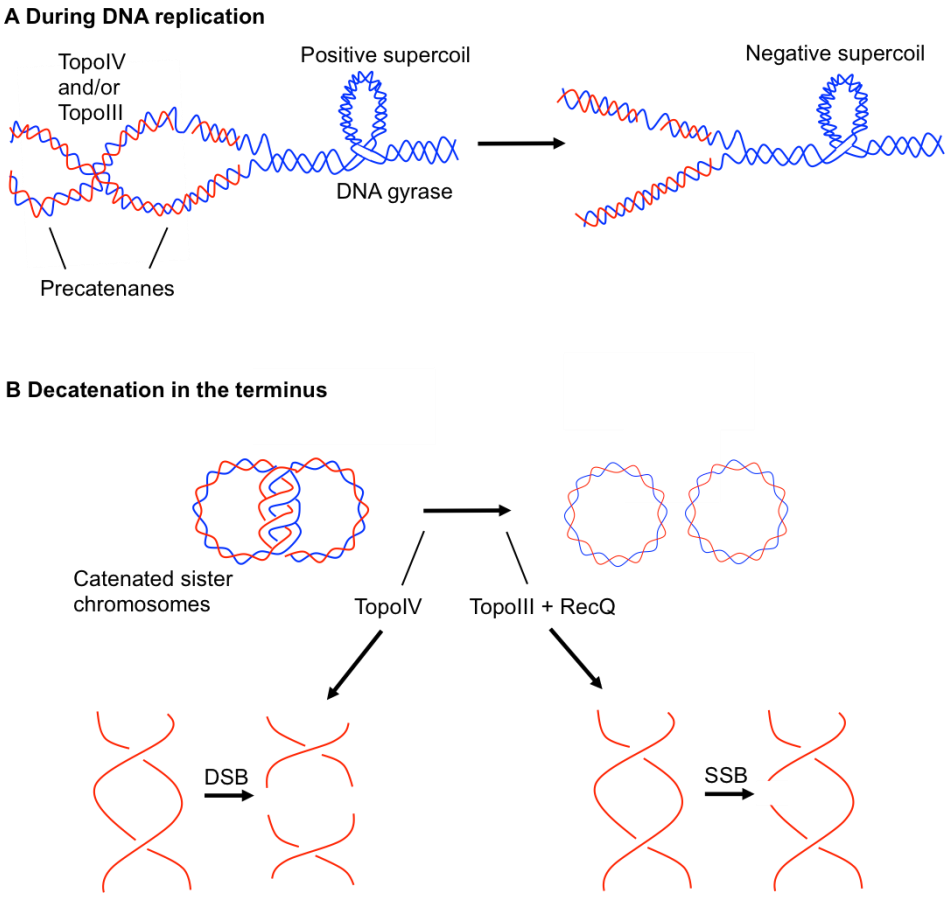


Figure 1.7. Chromosome decatenation.

(A) The DNA gyrase relaxes positive supercoils in front of the replication fork. Some of the supercoils in front of the fork are diffused behind the fork creating precatenanes, where TopoIV or TopoIII, through breaking and ligating, resolve them. (B) When two chromosomes are fully replicated, the tension is created in the terminus due to the catenation. TopoIV as a main decatenase resolves this tension through creating double-strand breaks and religating the DNA ends. TopoIII with the help of RecQ helicase can decatenate chromosomes as well through creating single-strand breaks and religating the DNA ends. Adapted from (Perez-Cheeks *et al.* 2012)

Two other topoisomerases, TopoI and the DNA gyrase (TopoII), have well defined roles (Champoux 2001). DNA gyrase is encoded by *gyrA* and *gyrB* and is the major facilitator of DNA replication. Mutations in both genes are lethal. TopoI is a type I topoisomerase that acts as a balancing force to DNA gyrase and is encoded by *topA*. Inactivation of TopoI leads to accumulation of suppressor mutations in DNA gyrase.

1.4 DNA damage and double-strand break repair

"Chekalinsky shuffled the cards and prepared to deal again."
A.S. Pushkin "The Queen of Spades"

1.4.1 DNA damage

DNA damage constitutes a serious threat to all living cells. Left unrepaired, DNA damage can cause genetic instability, mutations or cell death. In eukaryotes, it can lead to developmental disorders, cancer and various diseases. There are various types of DNA damage: DNA-protein crosslinks, insertions, deletions, substitutions, mismatches, intra- and interstrand crosslinks, single- and double-strand breaks (SSBs and DSBs) (Lindahl 1993). DSBs are one of the main sources of genetic rearrangements.

A DSB occurs when both strands of the DNA are broken simultaneously. Sources of DSBs can be divided into two groups: from exogenous agents and as a result of cellular metabolism. Exogenous agents can be anticancer DNA-damaging drugs such as mitomycin C, ionising radiation, ultraviolet light or mechanical stress (Lieber 1998). Endogenous agents can be oxidative free radicals, topoisomerase failures or the processing of different types of DNA lesions formed during DNA replication or transcription. Nevertheless, DSBs are essential in various processes, such as mating type switching in *S. cerevisiae*, meiotic recombination and faithful chromosome segregation. To preserve genomic integrity and ensure cell survival, it is essential to correctly repair DSBs.

1.4.2 DSB repair via homologous recombination

There are two main DSB repair pathways: NHEJ (non-homologous end joining) and HR (homologous recombination) (Iliakis *et al.* 2004). In *E. coli,* HR is the only pathway to repair DSBs (Figure 1.8). HR is important for the generation of genetic diversity and maintenance of genomic integrity. The

biochemical details of this complex and important process remain uncertain due to the difficulty of performing the biochemical assays with little or no DNA synthesis. Also, the isolation of intermediates remains complex due to their low abundance and, therefore, impractical. Additionally, more than 25 genes are involved in this process, which again makes it difficult to study (Clark and Margulies 1965). A sister chromosome or a homologous sequence of DNA is essential to repair a DSB by HR (Haber 2000). HR has 3 consecutive stages: pre-synapsis, synapsis and post-synapsis.

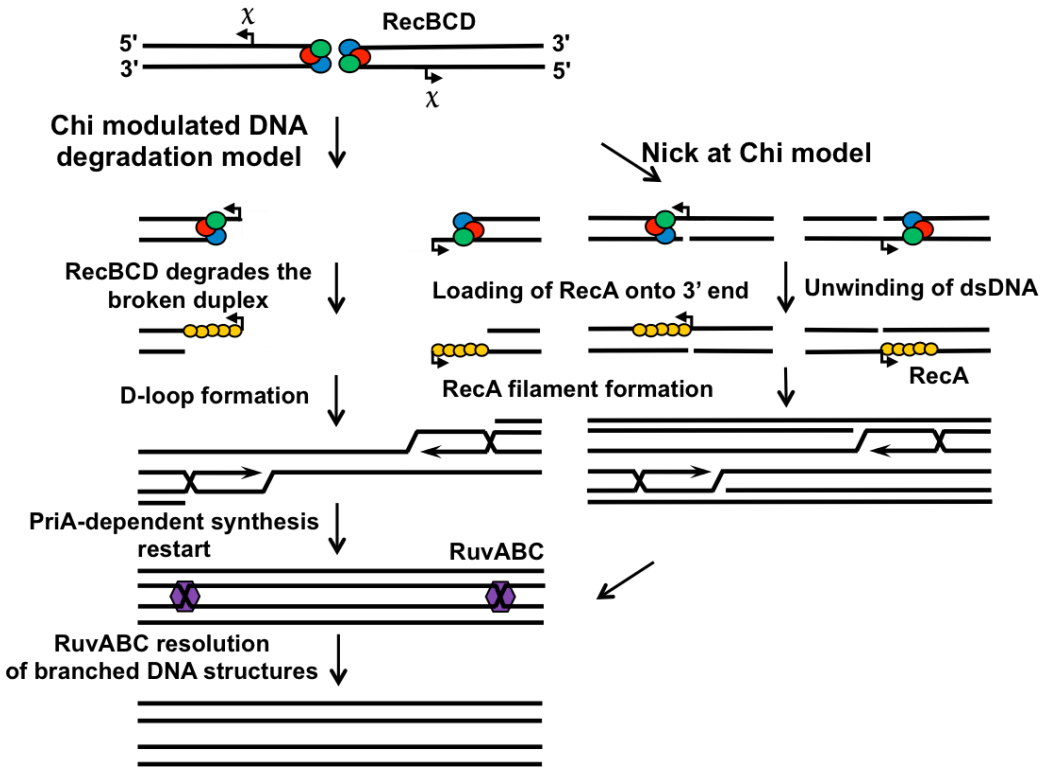


Figure 1.8. Two models of homologous recombination in *E. coli.*

Chi modulated DNA degradation model (Dixon and Kowalczykowski 1993). After the formation of a DSB, RecBCD processes the DNA ends. When RecBCD reaches a Chi (χ) site, it changes its conformation and forms 3' overhang. Meanwhile, it starts loading RecA on this 3' end, creating a RecA filament. RecA filament catalyses the homology search and forms a D-loop. Then, DnaB is loaded through RecG-PriA-PriB/PriC-DnaT. Finally, Holliday junctions are formed which are resolved by RuvABC and DNA ends are ligated by the DNA ligase.

Nick at Chi model (Taylor *et al*. 1985). After the formation of a DSB, RecBCD recognises double-strand ends and binds them. Then, it starts unwinding a DNA duplex. When RecBCD reaches a Chi (χ) site, it creates a nick at Chi site and starts loading RecA on 3' end, creating a RecA filament. RecA filament catalyses the homology search and forms a D-loop. Then, DnaB is loaded through RecG-PriA-PriB/PriC-DnaT. Finally, Holliday junctions are formed which are resolved by RuvABC and DNA ends are ligated by the DNA ligase.

1.4.2.1 Pre-synapsis: DNA end resection by RecBCD

Pre-synapsis starts with extensive processing of DNA double-strand ends by the RecBCD complex. RecBCD is a heterotrimeric multifunctional protein with activities such as DNA binding, RecA binding and loading, DNA-dependent ATPase, DNA helicase, ssDNA endonuclease, ssDNA exonuclease, dsDNA exonuclease and Chi-regulated nuclease (Dillingham and Kowalczykowski 2008). RecB and RecD proteins have independent helicase activities. They unwind DNA in presence of ATP with different speed. RecC is responsible for the recognition of a Chi site (χ) in an orientation-specific manner. A Chi site is the DNA sequence: 5′-GCTGGTGG-3′. This sequence is over-represented in the *E. coli* genome and its orientation is consistent with its function in the stimulation of DSB repair (DSBR) following replication fork collapse (Blattner *et al*. 1997).

When a DSB occurs, RecBCD recognises the DSB ends and binds them. Then, RecB starts processing the DNA and travelling 3′ to 5′ on one strand while RecD travels on the other strand in the direction 5′ to 3′. Actions of RecBCD separate the complementary strands of the DNA. There are two hypotheses that have been proposed to explain how the ends are processed (Taylor *et al*. 1985; Dixon and Kowalczykowski 1993). According to Dixon and collaborators, both strands are extensively degraded until RecBCD reach a correctly oriented Chi site. When the RecBCD complex reaches a Chi site, it changes its conformation and starts processing DNA in a different way. RecBCD switches its motor helicase from RecD to RecB and the processivity is reduced. 5′ end is continued to be degraded, therefore, producing a long 3′ overhang (Dixon and Kowalczykowski 1993). At the same time, RecB starts loading RecA protein on this 3′ end. According to Taylor and collaborators, RecBCD does not degrade DNA and just separates 3′ and 5′ strands. When

RecBCD encounters a correctly oriented Chi site, it nicks the 3′ strand and unwinds the duplex DNA to generate a 3′ overhang. Then, RecBCD loads RecA onto this overhang (Taylor *et al*. 1985).

1.4.2.2 Synapsis: RecA filament formation

Synapsis starts when the 3′ overhang generated by RecBCD is coated by RecA, to form a filament. Purified RecA protein is a DNA-dependent ATPase that can promote the formation of homologously paired joint molecules between suitable DNA substrate pairs. An absolute requirement for RecA work is the need of ssDNA (Cassuto *et al.* 1980; Ogawa *et al.* 1979). The nucleo-protein filament initiates a homology search and invades a homologous DNA duplex, creating a D-loop intermediate. The 3′ end of the invading strand primes DNA synthesis using the unbroken chromosome as a template. The RecA-ssDNA filament has two DNA-binding activities (Carra and Cucinotta 2011). Its first activity is to bind dsDNA to promote homology search. This binding is weak, sequence independent and quick. It is thought that the DNA duplex of the sister chromosome is temporally unwinded and tested for homology. If 15 bases are paired, the more stable second activity of the RecA filament is promoted (Cox 2007). Then, the RecA filament triggers strand exchange and forms a D-loop, therefore, leading to a structure formation that is called Holliday junctions (HJs) (Kowalczykowski *et al*. 1994).

1.4.2.3 Post-synapsis: RuvABC resolves Holliday junctions

During post-synapsis, DNA synthesis is initiated using the free 3′ end. Resolution of paired DNA molecules is essential. After the formation of HJs, branch migration is catalysed by the specialised DNA helicases, the RuvABC complex and, probably, by the RecG protein. RuvA is a DNA binding protein that is capable of recognising HJ structures. RuvB is an ATPase that is similar

to the DnaB helicase. RuvC is a nuclease and a DNA-binding protein. RuvC can specifically recognise and cleave HJ. It also binds to HJ regardless of the sequence (Sharples *et al.* 1991). RuvA binds to HJ, recruits RuvB and initiates branch migration in the 5′-3′ direction. Then, RuvC resolves the HJ by two symmetric cleavages and completes the recombination pathway. Then, DNA ends are directly joined by DNA ligase. In the absence of RuvAB another 3′-5′ helicase enzyme, called RecG, can promote HJ branch migration (Lloyd 1991; Mawer and Leach 2014).

1.5 DNA double-strand break formed by the cleavage of a 246 bp interrupted palindrome in *lacZ* by the SbcCD endonuclease

In Prof. David Leach laboratory, we have used a palindromic DNA sequence to form a DSB on only one of a pair of replicating sister chromosomes (Eykelenboom, Blackwood, Okely 2008) (Figure 1.9). This unique tool allows the introduction of a site-specific DSB once per replication cycle. Palindromes are two identical inverted repeat units that play a role in genomic instability due to their potential to form intra-strand structures (Leach 1994) (Figure 1.9 A). During DNA replication, a 246 bp interrupted palindrome (Figure 1.9 B), introduced in either the *lacZ* or the *ascB* gene, forms a hairpin-like structure on the lagging strand template of the chromosome, which is single-stranded between the Okazaki fragments. The structure-specific endonuclease SbcCD (Mre11-Rad50 in eukaryotes) cleaves this hairpin, leaving a two-ended DSB at the site of the palindrome (Figure 1.9 C). Then, the intact replicated leading strand serves as a template for HR-dependent DSB repair (Figure 1.8).

In this work, I used strains where the *sbcCD* operon was placed under the control of arabinose-inducible promoter $Para_{BAD}$. This promoter is controlled by the AraC protein. When arabinose is present in the medium, it binds to AraC and recruitment of RNA polymerase to the promoter occurs more rapidly. If arabinose is absent, AraC still binds to the RNA polymerase and recruits it to the operon. However, this occurs with very little binding affinity. To repress the promoter completely, glucose is added to the medium. Thus, this tool allows the control of the time of DSB induction in the *E. coli* (Eykelenboom, Blackwood, Okely 2008).

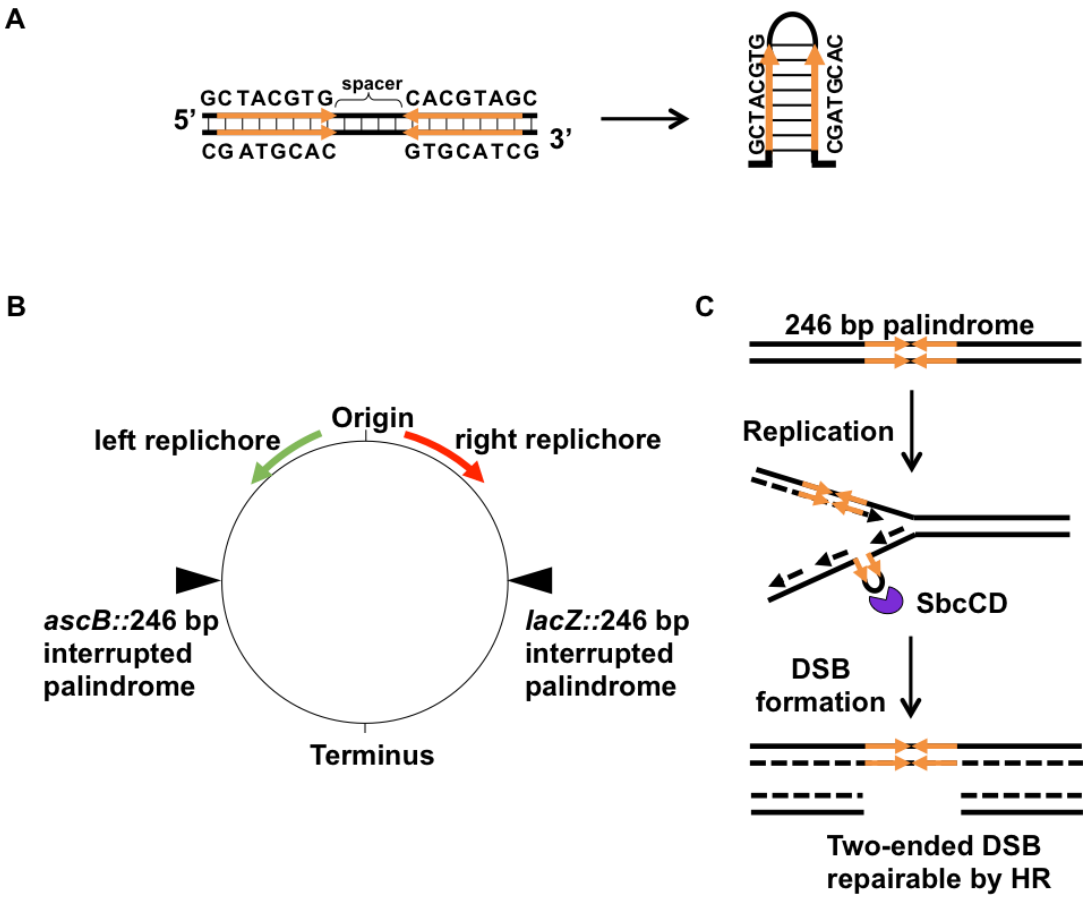


Figure 1.9. Replication-dependent cleavage of a 246 bp interrupted palindrome by SbcCD.

(A) Schematic representation of a hypothetical palindrome that forms a hairpin. (B) The location of the 246 bp interrupted palindromes on the *E. coli* chromosome. (C) The palindrome forms a hairpin only on the lagging strand template of the chromosome during DNA replication. The hairpin is cleaved by the structure-specific endonuclease SbcCD, forming a two-ended DSB that can be repaired by HR.

1.6 Repair of palindrome-mediated DSB at *lacZ* induces DSB and DNA over-replication in the terminus region

1.6.1 An introduction of a DSB in *lacZ* leads to an over-replication in the terminus region

To determine the genomic DNA abundance profile of *E. coli* strains subjected or not to a DSB, Dr. Martin White from Prof. Leach laboratory have used an approach based on whole genome sequencing (WGS). At the time of the beginning of this project, WGS had never been used to visualise a replication profile of the *E. coli* chromosome. Now, several papers have been published that describe the replication profiles of wild-type and mutant *E. coli* cells (Rudolph *et al*. 2013; Wendel, Courcelle, and Courcelle 2014; Azeroglu *et al*. 2016). To determine the replication profile of *E. coli* strains subjected or not to a DSB, Dr. Martin White used *pal*$^{-}$ *sbcCD*$^{+}$ and *pal*$^{+}$ *sbcCD*$^{+}$ strains. The presence of SbcCD and a 246 bp palindrome in the *lacZ* gene results in the formation of a repairable DSB on one sister chromosome at every replication cycle. The analyses of raw sequencing data were performed by Dr. Martin White. Briefly, the reads of obtained raw data were mapped to K12 MG1655 background that was taken from the NCBI database (NC000913.3). To correct for differences in sequence-based recovery across the genome, all the replication profiles of exponentially growing cultures were normalised to the profile of the non-replicating stationary phase wild-type culture (DL1777). The WGS technique counts the number of DNA copies across the whole genome. An exponentially growing strain replicates its DNA and therefore has more copies of the DNA at the origin than at the terminus (Figure 1.10). Surprisingly, the strain subjected to a DSB in *lacZ* had an excess of DNA in the terminus region between *terA* and *terB*. This difference was more striking when studying the ratio of the replication profile of the strain subjected to a DSB (DSB^{+}) to the

strain without a DSB (DSB^-). Figure 1.11 reveals an excess of DNA between *terA* and *terB*/*terC* loci with a peak in the region of the *dif* site. Additionally, there was less DNA detected on the right side of the replichore between the palindrome and the terminus region in the strain subjected to a DSB. These events suggest that the replication fork is delayed at the DSB site in *lacZ*, which probably shifts the termination of replication to the *terA* site.

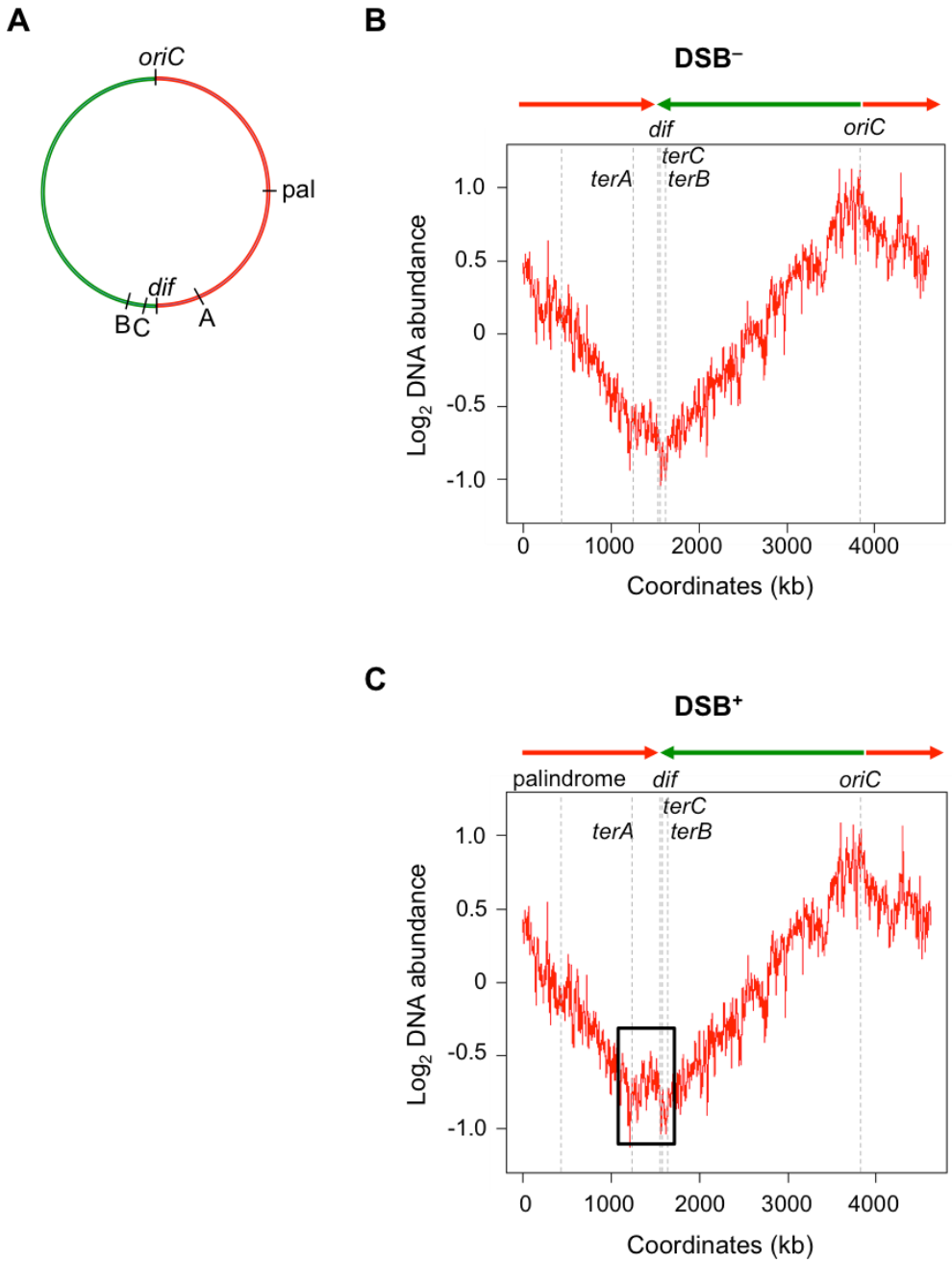


Figure 1.10. Whole genome sequencing profiles of DSB$^+$ and DSB$^-$ strains.

Replication profiles of exponentially growing cultures of strains with or without a 246 bp palindrome and with *sbcCD* under its native promoter are shown here. Log_2 DNA abundance represents log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of *terA*, *terB*, *terC*, *dif*, the palindrome and *oriC* are indicated. (A) Schematic of *E. coli* chromosome; (B) Strain DL1777 (*pal*$^-$ *sbcCD*$^+$); (C) Strain DL2859 (*pal*$^+$ *sbcCD*$^+$). Modified from White M., unpublished data.

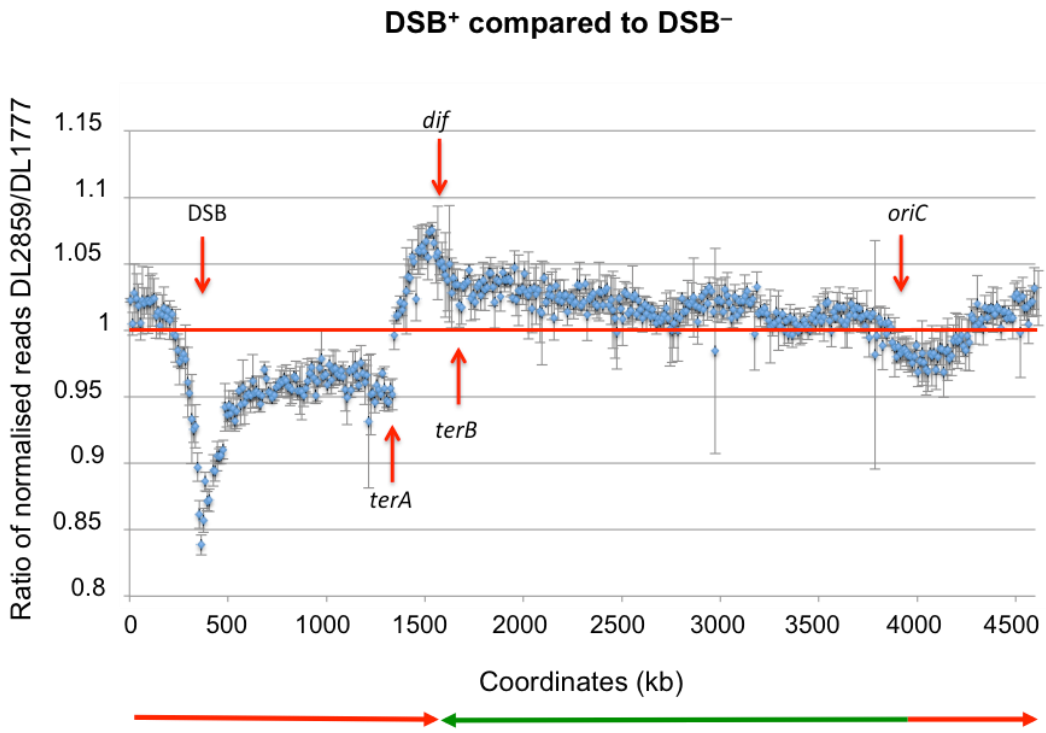


Figure **1.11**. **Ratio of replication profiles of DSB⁺ to DSB⁻ strains.**

The ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of *pal*⁺ *sbcCD*⁺ (DL2859) and *pal*⁻ *sbcCD*⁺ (DL1777) strains. The red line represents DL1777. The blue dots represent reads of the DL2859 strain. An average between two independent biological repeats is shown. The direction of replication is shown using green and red arrows. The positions of *terA*, *terB*, *dif*, the palindrome and *oriC* are indicated. Modified from White M., unpublished data.

1.6.2 Binding of RecA to the terminus region of the chromosome is dependent on repair of chronic double-strand break at *lacZ* under slow growth conditions

In Prof. David Leach laboratory, Dr. Charlotte Cockram and collaborators studied RecA-DNA interactions *in vivo* (Cockram *et al*. 2015). They used genome-wide analysis to investigate RecA-DNA binding during DNA double-strand break repair (Figure 1.12). MG1655 derivatives that were used in this study contained or not a 246 bp interrupted palindrome and SbcCD was constitutively expressed. The strains for ChIP-seq experiment were grown in the M9 minimal medium supplemented with 0.5% glucose and 0.2% casamino acids. Together with the expected RecA binding around the sites of the DSB in *lacZ*, surprisingly, ChIP-seq analyses revealed DSB-dependent binding of RecA in the terminus region around the *dif* site (Figure 1.13). The profile of this RecA binding was similar to the profile of RecA binding at the palindrome-induced DSB locus at *lacZ*, suggesting that the RecA enrichment at the terminus is due to RecBCD-mediated DSB processing. During the homologous recombination, RecBCD starts processing of DSB ends until it reaches properly oriented Chi sites (green and red circles in Figure 1.13). Here, RecBCD changes its conformation and starts loading RecA protein on the 3′ overhang, meanwhile continuing processing the 5′ end. Therefore, the profile of RecA enrichment in the terminus indicates the presence of a novel additional event that generates double-strand ends in the region of the chromosome where the termination of replication, chromosome dimer resolution and chromosomes decatenation occur.

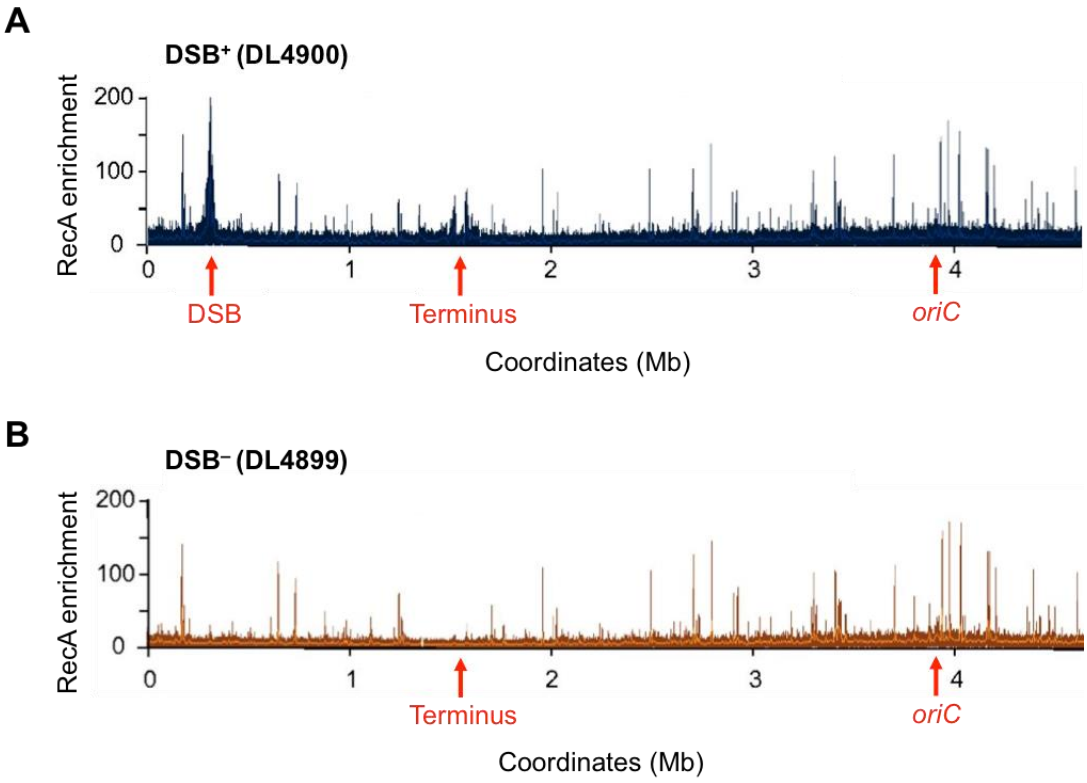


Figure 1.12. Genomic distribution of RecA binding in *E. coli.*

ChIP-seq analysis showing the genome-wide binding of the RecA protein. DSB-independent RecA binding was observed to highly expressed genes. In the presence of DSB, RecA binding was observed at the DSB locus in the *lacZ* gene and at the terminus region. *oriC* indicates the origin of replication; terminus is where DNA replication terminates. (A) Strain DL4900 (*pal*$^{+}$ *sbcCD*$^{+}$); (B) Strain DL4899 (*pal*$^{-}$ *sbcCD*$^{+}$). Modified from (Cockram *et al.* 2015).

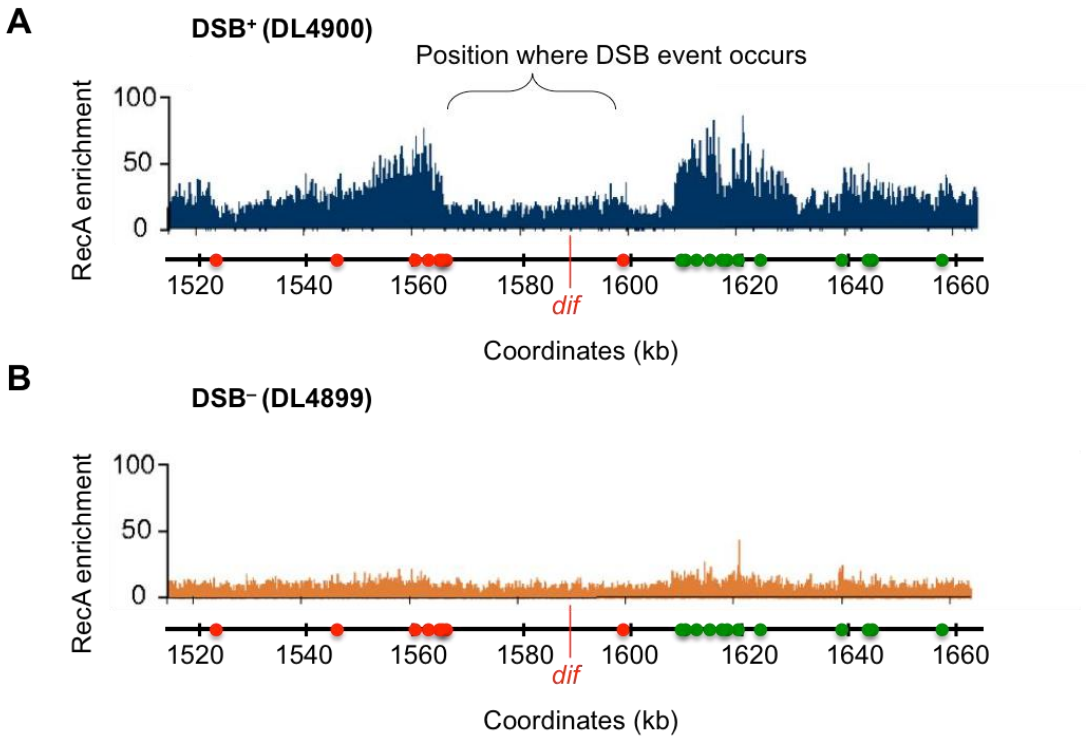


Figure 1.13. **DSB-dependent binding of RecA in the terminus region of *E. coli.***

ChIP-seq analysis revealed that, following a DSB in *lacZ,* RecA significantly binds DNA in the terminus region. The distribution of RecA binding correlates with the position of Chi sites, represented here by green and red circles. Green Chi sites are oriented in a such way that RecBCD will recognise them if it moves left to right on the chromosome, red Chi sites will be recognised by RecBCD that moves in the opposite direction – right to left. Therefore, the pattern of RecA enrichment indicates that a DSB-like event occurs between coordinates 1565 and 1598 kb and is repaired via RecBCD-dependent HR pathway. The *dif* site is located at 1588 kb. (A) Strain DL4900 (*pal*$^{+}$ *sbcCD*$^{+}$); (B) Strain DL4899 (*pal*$^{-}$ *sbcCD*$^{+}$). Modified from (Cockram *et al.* 2015).

1.6.3 Binding of RecA to the terminus region of the chromosome is independent of repair of acute double-strand break at *lacZ* in fast growth conditions

In parallel, in Prof. David Leach laboratory Dr. Benura Azeroglu was investigating the role of the RecG protein using the ChIP-seq assay developed by Dr. Charlotte Cockram (Azeroglu *et al.* 2016). Surprisingly, in this study the additional DSB event in the terminus region was independent from a DSB in *lacZ* (Figure 1.14). The discrepancy between the results of these two experiments can be explained by different growth conditions and different strain backgrounds. For all her experiments, Dr. Charlotte Cockram used MG1655 derivative strains with constantly expressed SbcCD, which allowed the palindrome to be cleaved all the time. The strains for this experiment were grown in M9 minimal medium supplemented with 0.5% glucose and 0.2% casamino acids. Dr. Benura Azeroglu, on the other hand, used Luria broth (LB) rich medium supplemented with 0.5% glucose to grow cells derived from BW27784 background, where SbcCD was placed under arabinose-induced promoter P_{BAD}. This allowed the control of palindrome cleavage. In this experiment, the DSB at *lacZ* was induced for 1 hour. The generation time of cells growing in LB medium is almost twice shorter than the generation time of cells growing in M9 medium (23 min vs 40 min). Therefore, there are more rounds of DNA replication due to more frequent origin firing in LB medium, and this can lead to replication forks being stalled. So, it means that due to the possibility of replication fork stalling, the spontaneous damage in the chromosome may be higher in LB medium. In M9 medium cells may have more time to divide and separate their chromosomes properly without any visible additional damage. All this could explain the difference between these ChIP-seq datasets.

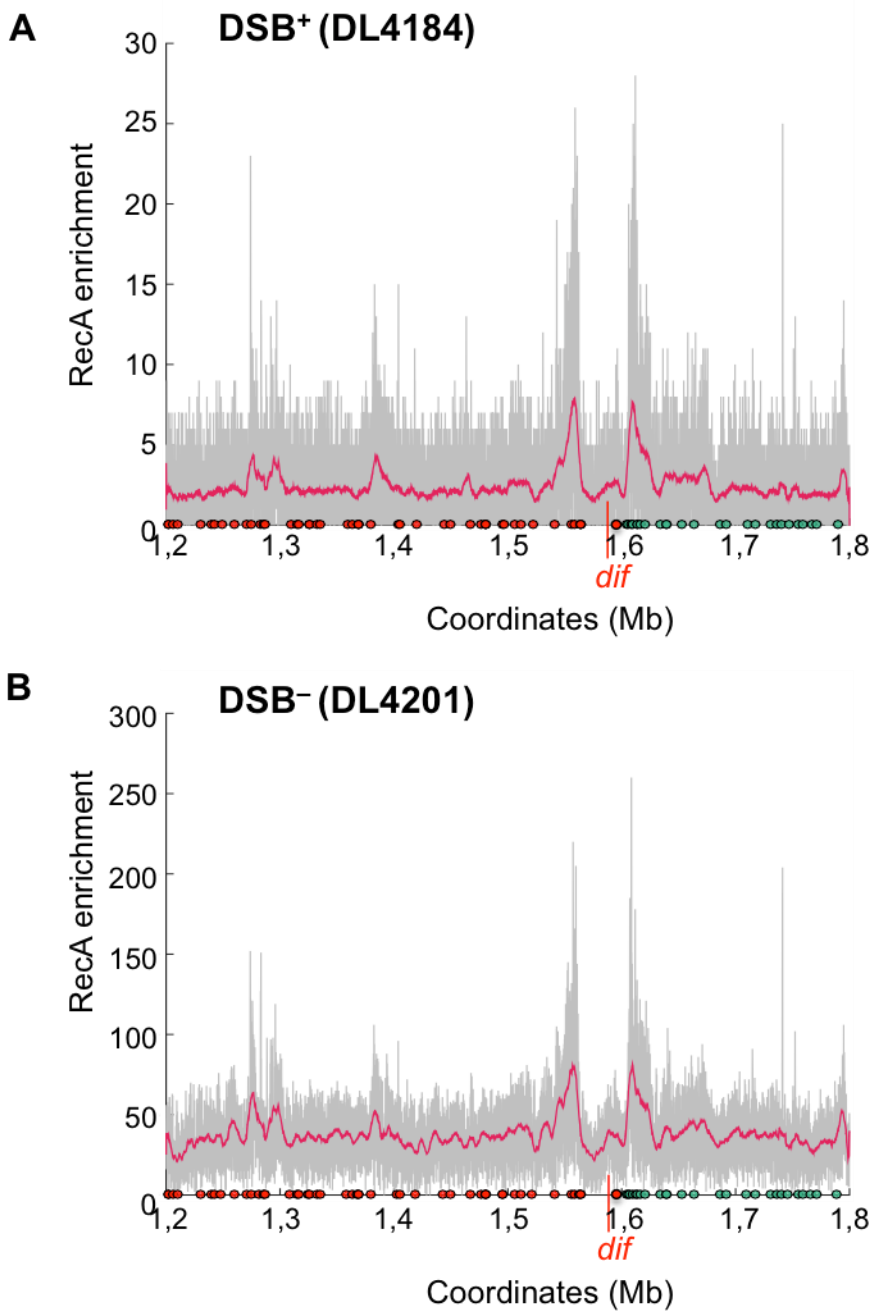


Figure 1.14. DSB-independent binding of RecA in the terminus region of *E. coli.*

ChIP-seq analysis revealed that the binding of RecA observed in the terminus is not dependent on DSB, induced in *lacZ*. The strain background and growth conditions were different compared to the observation made before. Green and red circles represent positions of Chi sites. Green Chi sites are oriented in a such way that RecBCD will recognise them if it moves left to right on the chromosome, red Chi sites will be recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site is located at 1588 kb. (A) Strain DL4184 (*pal*$^{+}$ P_{BAD}-*sbcCD*); (B) Strain DL4201 (*pal*$^{-}$ P_{BAD}-*sbcCD*). Modified from (Azeroglu *et al.* 2016).

1.7 Scope of this thesis

The study presented in this thesis investigates the termination of DNA replication in *E. coli*. More specifically, what is the nature of the over-replication detected in the terminus, why DSBs are observed around the *dif* site in cells that grow rapidly and whether the over-replication and DSB events in the terminus are linked to each other. Additionally, a potential effect of palindrome-mediated DSBs in *lacZ* on these events is investigated. To study these processes next generation sequencing methods combined with chromosome immunoprecipitation and two-dimensional native-native agarose gel electrophoresis were used. In this work, I used mutants defective in various stages of DNA replication termination: chromosome dimer resolution and chromosome decatenation. This study deepens our understanding of the nature of events in the terminus region and the interplay between such complicated processes as replication termination, chromosome dimer resolution and chromosome decatenation that overlap spatially and temporally.

Chapter 3 follows an earlier observation from Dr. Martin White in our laboratory, in which a DNA excess was shown in the terminus between *terA* and *terB* sites in a strain subjected to DSBs at *lacZ* that was not present in a strain without DSBs at *lacZ*. Using two-dimensional agarose gel electrophoresis, I showed that the DNA excess observed in the terminus is in fact DNA over-replication as more replication forks paused at *terA* and *terB* sites were detected in a strain subjected to DSBs at *lacZ*. Moreover, this chapter reports that the observed peak of over-replication in the terminus is at the *dif* site, which is a source of additional bi-directional replication that is activated by DSBs in *lacZ* or *ascB*.

Chapter 4 discusses the discrepancies between various datasets obtained in our laboratory about DSBs in the terminus observed using RecA-ChIP-seq. Also, this chapter justifies the choice of the genetic backgrounds for all experiments in this work and describes the choice of growth conditions for these strains.

Chapter 5 describes the choice of the XerCD/*dif* system as a candidate to test the cause of DSBs in the terminus and shows the role of the chromosome dimer resolution system in the appearance of DSBs around the *dif* site. Additionally, this chapter shows that the amount of DSBs at *ter* sites increases following a longer induction of DSBs at *lacZ* and presents a potential link of these *ter* DSBs with the observed over-replication in the terminus.

Chapter 6 focuses on the possibility that topoisomerases operating in the terminus (TopoIII and TopoIV) could be responsible for DSBs in the terminus. Also, this chapter describes the interplay between chromosome decatenation and chromosome dimer resolution by using TopoIV *dif* and TopoIII *dif* double mutants, where the amount of DSBs close to *dif* is reduced in TopoIV *dif* double mutant. This result suggests that TopoIV is involved in 'guillotining' DSBs in the terminus.

Finally, chapter 7 describes a tool that was recently developed by Dr. Benura Azeroglu in. Prof. David Leach laboratory. This tool utilises the pattern of RecA binding in a *recD* mutant to map DSB positions on the *E. coli* chromosome. In this work, this tool was used to map DSBs in the terminus in *tus* and in *dif* mutant. When *ter* DSBs were eliminated in *tus* mutant, the RecA enrichment pattern showed that the majority of DSBs occur at the *dif* site. At the same time, in a *dif* mutant, a wider distribution of RecA enrichment was detected suggesting that the presence of the *dif* site concentrates DSBs in the terminus.

Overall, this work presents new insight on the termination of DNA replication. It shows that topoisomerases (TopoIII and TopoIV) might be responsible for DSBs in the terminus. Moreover, these DSBs are concentrated in the terminus by XerCD/*dif* system. Also, it shows that the observed over-replication in the terminus is induced by DSBs at the *dif* site and that this over-replication leads to the DSBs at the *ter* sites. Finally, this study also demonstrates that the initial DSBs at the *dif* site are induced by palindrome-mediated DSBs at *lacZ* or *ascB*.

Chapter II

2. Materials and methods

2.1 Materials

2.1.1 Growth media

Media used in this study are indicated in Table 2.1. They were made in distilled water and autoclaved before use.

Table 2.1. Growth media.

Medium	Composition
LB (Luria Broth)	1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl; pH was adjusted to 7.5 with NaOH
LB Agar	1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% Bacto-agar; pH was adjusted to 7.5 with NaOH
LC Bottom Agar	1% tryptone, 0.5% yeast extract, 0.5% NaCl, 1% Difco-agar; pH was adjusted to 7.2 with NaOH
LC Top Agar	1% tryptone, 0.5% yeast extract, 0.5% NaCl, 0.7% Difco-agar; pH was adjusted to 7.2 with NaOH
Minimal Medium (M9)	33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM NH_4Cl, 0.2% casamino acids, 0.5% glucose, 1 mM $MgSO_4$, 5 μM $CaCl_2$; made up to 1 L with sterile ddH_2O

Minimal Agar (M9)	33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM NH_4Cl, 0.2% casamino acids, 0.5% glucose, 1 mM $MgSO_4$, 5 μM $CaCl_2$, 1.5% Bacto-agar; made up to 1 L with sterile ddH_2O

2.1.2 Medium complements and antibiotics

Medium complements and antibiotics are indicated in Table 2.2 and Table 2.3, respectively. They were added to the medium prior to use. All antibiotics were stored at -20°C and all complements were stored at room temperature.

Table 2.2. Sugars.

Sugar	Abbreviation Used	Solvent	Stock Concentration	Concentration Used
Arabinose	Ara	dH_2O	20% (w/v)	0.2% (w/v)
Glucose	Glu	dH_2O	20% (w/v)	0.5% (w/v)
Glycerol	Gly	dH_2O	80% (v/v)	40% (v/v)
Sucrose	Suc	dH_2O	20% (w/v)	5% (w/v)

Table 2.3. Antibiotics.

Antibiotic	Abbreviation Used	Solvent	Stock Concentration	Concentration Used
Chloramphenicol	Cm	100% Ethanol	50 mg/ml	50 μg/ml

Gentamycin	Gm	dH_2O	10 mg/ml	10 μg/ml
Kanamycin	Km	dH_2O	50 mg/ml	50 μg/ml
Tetracycline	Tc	50% Ethanol	15 mg/ml	15 μg/ml

2.1.3 Enzymes and enzyme buffers

All restriction enzymes and their buffers, as well as Quick Ligation™ Kit, were purchased from NEB (New England Biolabs). GoTaq® G2 and GoTaq® Flexi Polymerases were purchased from Promega. Herculase II Fusion DNA Polymerase was purchased from Agilent Technologies. All dNTPs (100 mM) were purchased from Roche, separately. They were mixed equally in order to obtain the final concentration of 25 mM. BigDye® terminator v3.1 Cycle Sequencing Kit and 2X Brilliant II SYBR® Green QPCR Master Mix were purchased from Applied Biosystems and Agilent Technologies, respectively. All enzymes were used according to the manufacturer's guidelines.

2.1.4 Buffers and solutions

2.1.4.1 Buffers for manipulations with phage

Phage buffer: 22 mM KH_2PO_4, 49 mM Na_2HPO_4, 85 mM NaCl, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 1% gelatine were dissolved in dH2O and autoclaved. Stored at room temperature.

TE buffer: 10 mM Tris and 1 mM EDTA were dissolved in dH2O; pH was adjusted to 7.4 with HCl. Stored at room temperature.

2.1.4.2 Buffers for agarose gel electrophoresis

50X Tris-acetate buffer (TAE): 2 M Tris base, 0.05 M EDTA (pH 8.0), 0.95 M glacial acetic acid; dissolved in 1 L of dH_2O and filter sterilised. The working concentration of TAE was 1X.

Molecular weight markers: 50 bp, 100 bp and 1 Kb DNA Ladders were used (New England Biolabs).

Agarose: agarose, molecular biology grade from MELFORD.

SafeView: purchased from NBS Biologicals Ltd, SafeView.

DNA Loading Buffer (6X): purchased from New England Biolabs.

2.1.4.3 Buffers for 2D-agarose gel electrophoresis

NDS: 0.5 M Na_2.EDTA, 10 mM Tris base, 0.6 mM NaOH and 34 mM N-lauryl sarcosine. The Na_2.EDTA, Tris base and NaOH were dissolved in 350 ml of dH_2O. N-lauryl sarcosine was dissolved in 50 ml of dH_2O and then added to the main solution. pH was adjusted to 8.0 with NaOH and the total volume was brought up to 500 ml.

TEN: 50 mM Tris, 50 mM EDTA and 100 mM NaCl were dissolved in dH_2O. pH was adjusted to 8.0 with HCl and the total volume was brought up to 1 L. Stored at +4°C.

5X Tris-borate buffer (TBE): 0.89 M Tris base, 0.89 M Boric acid and 23 mM EDTA were dissolved in dH_2O; pH was adjusted to 8.0 with HCl and the total volume was brought up to 500 ml. Stored at room temperature.

Proteinase K solution: 1 mg/ml Proteinase K and 1% (w/v) SDS were dissolved in NDS.

Depurination solution: 0.25 M HCl was completed in dH_2O. Stored at room temperature.

Transfer buffer: 10X SSC and 0.5 M NaOH were dissolved in dH_2O. Stored at room temperature.

Church-Gilbert buffer (hybridisation solution): 7% SDS, 0.5 M NaH_2PO_4 and 1 mM EDTA (pH 8.0) were dissolved in dH_2O.

The NaH_2PO_4, EDTA and 20 ml of distilled water were mixed until the NaH_2PO_4 was completely dissolved. The solution was heated for easy mixing. Then, SDS was added. pH was adjusted to 7.2 with NaOH. This hybridisation solution was filter sterilised using a 0.4 μm pore size filter.

20X SSC: 3 M NaCl and 300 mM Tri-sodium citrate were dissolved in dH_2O; pH was adjusted to 7.0 with HCl.

Wash buffer 1 (low stringency): 2X SSC and 0.1% SDS were completed in dH_2O.

Wash buffer 2 (high stringency): 0.5X SSC and 0.1% SDS were completed in dH_2O.

20X SSPE: 3 M NaCl, 200 mM NaH_2PO_4 and 20 mM EDTA were dissolved in dH_2O; pH was adjusted to 8.0.

Stripping buffer: 50% formamide and 5X SSPE.

2.1.4.4 Buffers and solutions for Chromatin Immunoprecipitation (ChIP) and ChIP-seq library preparation

Crosslinking agent: a solution of 37% formaldehyde containing 10-15% methanol, to prevent polymerisation, was purchased from Sigma. Stored at room temperature.

Quenching agent: 2.5 M Glycine was dissolved in dH_2O. Stored at room temperature.

ChIP (Lysis) buffer: 200 mM Tris-HCl (pH 8.0), 600 mM NaCl, 4% Triton™ X-100 (purchased from Sigma), Complete protease inhibitor cocktail (EDTA-free; Roche) were dissolved in dH_2O. Stored at room temperature before addition of the Complete protease inhibitor cocktail, after addition – kept at +4°C, used immediately.

Anti-RecA antibody: rabbit polyclonal antibody to recombinant full length RecA in working dilution 1:100 was purchased from abcam, Cat. No. ab63797. Stored at -20°C.

1X PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4 were dissolved in 800 ml of dH_2O. pH was adjusted to 7.4 with HCl, and dH_2O was added to 1 L. Solution was sterilised by autoclaving. Stored at room temperature.

Dynabeads Protein G: beads (30 mg/ml) are supplied in PBS (pH 7.4) supplemented with 0.1% Tween® 20 and 0.02% NaN_3 (purchased from invitrogen). Stored at +4°C.

Wash buffer: 1X PBS supplemented with 0.02% Tween® 20 (purchased from Fisher). Stored at room temperature.

ChIPed DNA purification: QIAquick® PCR Purification Kit purchased from QIAGEN.

Elution buffer: 10 mM Tris-HCl; pH was adjusted to 7.4. Stored at room temperature.

Library preparation kit: NEBNext ChIP-seq Library Preparation Master Mix Set for Illumina. Stored at -20°C.

Indexes: NEBNext Multiplex Oligos for Illumina, Index Primers Set 1 and Set 2 (Cat. No. E7335S and E7500S, respectively). Stored at -20°C.

Library DNA purification: MinElute® PCR Purification Kit and QIAquick® Gel Extraction Kit purchased from QIAGEN. MinElute columns stored at +4°C.

ChIP-Seq loading buffer: 50 mM Tris-HCl (pH 8.0), 40 mM EDTA and 40% (w/v) sucrose. Stored at +4°C.

2.1.4.5 Buffers and solutions for whole genome sequencing

Total genomic DNA extraction: Wizard® Genomic DNA Purification Kit was used (purchased from Promega).

Total genomic DNA purification: Genomic DNA Clean & Concentrator™-10 (purchased from Zymo Research) was used for purification of all genomic DNA shown in this work, unless otherwise stated.

Measurement of DNA concentration: Qubit® dsDNA BR kit was used (purchased from invitrogen). DNA and RNA buffers stored at room temperature, DNA and RNA dyes stored in a toxic cabinet, covered from light, RNA and DNA standards 1 and 2 stored at +4°C.

2.1.5 Oligonucleotides, plasmids and *E. coli* strains

2.1.5.1 Oligonucleotides

All oligonucleotides used in this work were designed using the Primer3 software (http://primer3.ut.ee). They were synthesised by MWG Biotech and are listed in Table 2.4. 100 mM stocks were completed in dH_2O and stored at -20°C. 10 mM working solutions were used.

Table 2.4. List of oligonucleotides used in this work.

Name	Sequence (5'-3')	Used for
dif.200bp.F	acggcagttgaaagcaatgt	PCR and sequencing to check the *dif* region deletion
dif.200bp.R	cagaaaagcacttcgcatca	
topB-PstI-F1	aaaaactgcagctccccggtgttgaagagta	
topB-R1	**agcactgccaccccaggtcaccac**ctgaccatt	Crossover PCR to make a fragment for the construction of pDL6028 permitting the deletion of the *topB* gene
topB-F2	gtggtgacctggggtggcagtgctgataagaaa	

topB-SalI-R2	aaaaagtcgacttatcgctgcgaacacaaac	
xerC-KO-F1[a]	aaaaactgcagtctcgacgatatgctgatgc	PCR and sequencing to check the *xerC* region deletion
xerC-KO-R2[a]	aaaaagtcgaccttcgccagctgtttaagg	
xerD-KO-F1[a]	aaaaactgcaggtttgtcatgcagcataccg	PCR and sequencing to check the *xerD* region deletion
xerD-KO-R2[a]	aaaaagtcgactgcagtttgtggcagtaacc	
Ex-test.F[a]	ttatgcttccggctcgtatg	PCR and sequencing to check for the presence of a palindrome in *lacZ*
Ex-test.R[a]	ggcgattaagttgggtaacg	
pKO.F[b]	agggcagggtcgttaaatagc	PCR and sequencing to check fragments inserted into pTOF24
pKO.R2[b]	agggaagaaagcgaaaggag	
List of the oligonucleotides used for Southern blot		
TerA-probe-F[c]	gaaggggttcacctaatcacg	To generate a 4 kb fragment Southern blot probe within *terA*
TerA-probe-R[c]	gctgtcggaggtgttattgg	
TerB-probe-F[c]	gcataaagcggtaggtgtcg	To generate a 4,5 kb fragment Southern blot probe within *terB*
TerB-probe-R[c]	atggtcggcagtatgaaagc	
TerC A F[b]	ccggtacccattgttattgc	To generate a 4 kb fragment Southern blot probe within *terC*
TerC B R[b]	ggtgcacagtatccagaacg	

List of the oligonucleotides used for qPCR

Q.bef.chi1.F	gatgtgcgctcattctcgta	qPCR the region 21,7 kb from *dif* before Chi sites on the left replichore
Q.bef.chi1.R	acgcaccaaaaggtcacttc	
Q.bef.chi3.F	aaccactggaagacctggaa	qPCR the region 9,4 kb from *dif* before Chi sites on the right replichore
Q.bef.chi3.R	gagcgaacggttttacttcg	
Q.Dif2.F	cgacattctaccgcctctga	qPCR the region 30,3 kb from *dif* after Chi sites on the right replichore
Q.Dif2.R	ggcagtgcgtttcgttatgt	
Q.Dif1.F	cacgttaaatgaaaacccgcg	qPCR the region 27,3 kb from *dif* after Chi sites on the left replichore
Q.Dif1.R	gaaggcgtcgaaaccaaaga	
Q.terA1.F	gccagagctgcttcgtaatc	qPCR the the region adjacent to *terA* site, includes the first Chi-site in proper orientation
Q.terA1.R	gccagagctgcttcgtaatc	
Q.terC2.F	cctgtcccactgctatttga	qPCR the the region adjacent to *terC* site, 2 kb from the first Chi-site in proper orientation
Q.terC2.R	cgggaatgatggagaacgta	
Q.terB2.F	tggaaaagaaactgggactg	

Q.terB2.R	tgaggatcagcatggcaaag	qPCR the the region adjacent to *terC* site, 2 kb from the first Chi-site in proper orientation
Q.cynT.F[d] Q.cynT.R[d]	cgtcattgctcagttggcta gtagcggtatggcacaaacg	qPCR the region 6,7 kb upstream from *pal* after Chi sites
hycGnew.F[d] hycGnew.R[d]	tggatgaatccatcgccagcatga tgacgcggtgaaggaacgacttta	Primers for qPCR of a control region used a a reference for normalisation

[a] – oligonucleotides designed by Ewa Okely, Leach lab, [b] – oligonucleotides designed by Julia Mawer, Leach lab, [c] – oligonucleotides designed by Martin White, Leach lab, [d] – oligonucleotides designed by Charlie Cockram, Leach lab. Cutting sites for PstI and SalI enzymes are underlined; crossover PCR homology arms are in **bold**.

2.1.5.2 Plasmids

All plasmids used in this work are listed in Table 2.5.

Table 2.5. List of plasmids used in this work.

Name	Description	Source
pDL1605	pTOF24 Cm^R Km^R Ts Suc^S	(Merlin, Mcateer, and Masters 2002)
pDL2712	pTOF24 + Δ*xerD* Cm^RTs Suc^S	Ewa Okely, Leach lab
pDL2732	pTOF24 + Δ*dif* Cm^RTs Suc^S	Ewa Okely, Leach lab
pDL2763	pTOF24 + Δ*xerC* Cm^RTs Suc^S	Ewa Okely, Leach lab
pDL6028	pTOF24 + Δ*topB* Cm^RTs Suc^S	This work

Cm^R - resistant to chloramphenicol; Km^R - resistant to kanamycin; Ts – temperature-sensitive ; Suc^S - sensitive to products of sucrose degradation.

2.1.5.3 *E. coli* strains

All bacterial strains used in this work are listed in Table 2.6.

Table 2.6. List of *E. coli* strains used in this work.

Name	Description	Source
BW27784	*Δ(araD-araB)567 Δ(araH-araF)570(::FRT) ΔaraEp-532::FRT ϕP_{cp18}araE533 Δ(rhaD-rhaB)568 hsdR514 ΔlacZ478(::rrnB-3)*	(Khlebnikov *et al*. 2001)
EJ812 (C600parC1215)	F- *thr-1 leu-6 thi-1 lacY1 supE44 tonA21 parC1215 met::Tn10*	(Kato *et al*. 1988; Zechiedrich and Cozzarelli 1995)
MG1655	F- *λ- ilvG- rfb-50 rph-1*	(Blattner *et al*. 1997)
MG1655 (DL)	F- *λ- ilvG- rfb-50 rph-1 fnr-267 (ΔynaJ ΔydaA Δfrn Δogt ΔabgT ΔabgB ΔabgA ΔabgR ΔydaL ΔydaM ΔydaN ΔdbpA ΔydaO)*	(Soupene *et al*. 2003)
W3110parE10	F- *λ- rph-1 INV(rrnD, rrnE) parE10*	(Kato, Nishimura, Imamura, Niki, Hiraga 1990;

		Zechiedrich and Cozzarelli 1995)
CAG18472	MG1655 F$^-$ λ^- *nupG511::Tn10 rph-1*	(Singer *et al*. 1989)
CAG18475	MG1655 F$^-$ λ^- *metC162::Tn10 rph-1*	(Singer *et al*. 1989)
CAG18527	MG1655 F$^-$ λ^- *metC3158*::*Tn10kan rph-1*	(Singer *et al*. 1989)
CAG18559	MG1655 F$^-$ λ^- *nupG3157*::*Tn10kan rph-1*	(Singer *et al*. 1989)
JJC256	MG1655 Δ*tus::kan*	(Bierne 1991)
JJC6762	MG1655 FtsK^{K997A} Cm^R	Benedicte Michel lab
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^{q}ZΔM15 Tn10*]	Stratagene
DL1777	*lacI^q lacZ$_{X-}$ fnr-267 (ΔynaJ ΔydaA Δfrn Δogt ΔabgT ΔabgB ΔabgA*	(Eykelenboom, Blackwood, Okely 2008)

	ΔabgR ΔydaL ΔydaM ΔydaN ΔdbpA ΔydaO)	
DL2151	*ΔsbcDC lacI^{q} lacZ$_{\chi-}$ fnr-267 (ΔynaJ ΔydaA Δfrn Δogt ΔabgT ΔabgB ΔabgA ΔabgR ΔydaL ΔydaM ΔydaN ΔdbpA ΔydaO)*	(Eykelenboom, Blackwood, Okely 2008)
DL2006	BW27784 *P$_{BAD}$-sbcDC lacZ::pal246 cynX::Gm^{R} lacI^{q} lacZχ^{-}*	(Eykelenboom, Blackwood, Okely 2008)
DL2573	BW27784 *P$_{BAD}$-sbcDC lacZ^{+} cynX::Gm^{R} lacI^{q} lacZχ^{-}*	(Eykelenboom, Blackwood, Okely 2008)
DL2859	*lacZ::246pal cynX::Gm^{R} lacI^{q} lacZ$_{\chi-}$ lacZ(L8) fnr-267 (ΔynaJ ΔydaA Δfrn Δogt ΔabgT ΔabgB ΔabgA ΔabgR ΔydaL ΔydaM ΔydaN ΔdbpA ΔydaO)*	(Eykelenboom, Blackwood, Okely 2008)
DL2874	*ΔsbcDC lacZ::246pal cynX::Gm^{R} lacI^{q} lacZ$_{\chi-}$ lacZ(L8) fnr-267 (ΔynaJ ΔydaA Δfrn Δogt ΔabgT ΔabgB ΔabgA ΔabgR ΔydaL ΔydaM ΔydaN ΔdbpA ΔydaO)*	(Eykelenboom, Blackwood, Okely 2008)

DL3683	DL4201 *mhpA::χχχ lacZY::χχχ*	Martin White, Leach lab
DL3684	DL4184 *mhpA::χχχ lacZY::χχχ*	Martin White, Leach lab
DL4184	DL2006 *mhpR::χχχ lacZ::χχχ proA::ISceIcs tsx::ISceIcs*	(Mawer and Leach 2014)
DL4201	DL2573 *mhpR::χχχ lacZ::χχχ proA::ISceIcs tsx::ISceIcs*	(Mawer and Leach 2014)
DL4899	MG1655 (DL) *lacI^{q} lacZχ^{-} lacZ^{+} lacZY::χχχ mhpA::χχχ*	(Cockram *et al*. 2015)
DL4900	MG1655 (DL) *lacZ::246pal cynX::Gm^{R} lacI^{q} lacZχ^{-} mhpA::χχχ lacyZY::χχχ*	(Cockram *et al*. 2015)
DL5107	*ascB::pal246 lacI^{q} lacZ$_{χ-}$ fnr-267 (ΔynaJ ΔydaA Δfrn Δogt ΔabgT ΔabgB ΔabgA ΔabgR ΔydaL ΔydaM ΔydaN ΔdbpA ΔydaO)*	Charlotte Cockram, Leach lab

DL5398	DL1777 Δ*tus::kan*	P1 using JJC256, this work
DL5399	DL2859 Δ*tus::kan*	P1 using JJC256, this work
DL5499	DL5107 Δ*tus::kan*	P1 using JJC256, this work
DL5699	DL4184 Δ*recD*	Benura Azeroglu, Leach lab
DL5780	DL4201 Δ*xerD*	PMGR[1] with pDL2712, this work
DL5782	DL4201 Δ*xerC*	PMGR with pDL2763, this work
DL5784	DL4184 Δ*xerC*	PMGR with pDL2763, this work
DL5786	DL4184 Δ*dif*	PMGR with pDL2732, this work
DL5788	DL4184 Δ*xerD*	PMGR with pDL2712, this work

[1] - Plasmid Mediated Gene Replacement

DL5796	DL5784 Δ*xerD*	PMGR with pDL2712, this work
DL5798	DL5782 Δ*xerD*	PMGR with pDL2712, this work
DL5812	DL4201 Δ*dif*	PMGR with pDL2732, this work
DL5894	DL4184 Δ*tus*	Benura Azeroglu, Leach lab
DL5895	DL4201 Δ*tus*	Benura Azeroglu, Leach lab
DL5966	W3110parE10 *nupG511::Tn10*	P1 using CAG18472, this work
DL5967	W3110parE10 *metC162::Tn10*	P1 using CAG18475, this work
DL5970	EJ812 *nupG511::Tn10*	P1 using CAG18472, this work

DL5973	EJ812 *metC162::Tn10*	P1 using CAG18475, this work
DL5975	DL4184 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL5976	DL4201 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL5996	DL4184 *metC162::Tn10 parC1215*	P1 using DL5973, this work
DL5998	DL4201 *metC162::Tn10 parC1215*	P1 using DL5973, this work
DL6022	DL5786 *metC162*::*Tn10 parE10*	P1 using DL5967, clone 1, this work
DL6023	DL5786 *metC162*::*Tn10 parE10*	P1 using DL5967, clone 2, this work
DL6024	DL5812 *metC162::Tn10 parE10*	P1 using DL5967, this work

DL6025	DL5812 *metC162::Tn10 parC1215*	P1 using DL5973, this work
DL6026	DL5786 *metC162::Tn10 parC1215*	P1 using DL5973, this work
DL6051	DL4184 Δ*topB*	PMGR with pDL6028, this work
DL6054	DL5812 Δ*topB*	PMGR with pDL6028, this work
DL6059	DL4201 Δ*topB*	PMGR with pDL6028, this work
DL6071	DL6058 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL6134	JJC6651 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL6141	DL4184 FtsK^{K997A} Cm^{R}	P1 using JJC6762, this work
DL6143	DL5812 FtsK^{K997A} Cm^{R}	P1 using JJC6762, this work

DL6146	MG1655 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL6148	DL5786 FtsK^{K997A} Cm^{R}	P1 using JJC6762, this work
DL6150	JJC6762 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL6153	DL4201 FtsK^{K997A} Cm^{R}	P1 using JJC6762, this work
DL6189	JJC6728 *metC162::Tn10 parE10*	P1 using DL5967, this work
DL6194	DL5812 Δ*recD*	Benura Azeroglu, Leach lab
DL6196	DL5895 Δ*recD*	Benura Azeroglu, Leach lab
DL6204	DL4201 Δ*recD*	Benura Azeroglu, Leach lab

2.2 Methods

2.2.1 Molecular biology methods

2.2.1.1 Genomic DNA extraction for PCR

The Promega Wizard® Genomic DNA purification kit was used to extract genomic DNA following the manufacturer's instructions. DNA was re-hydrated in milli-Q (MQ) water at 65°C for 1 hour and stored at -20°C.

2.2.1.2 Plasmid DNA extraction

An overnight of the strain that contains a desired plasmid was set up in 5 ml of LB medium (10 ml for a low copy number plasmid) with appropriate selection at the desired temperature (30°C for pTOF24 derivative). The QIAGEN QIAprep® Spin Miniprep Kit was used to extract plasmid DNA from the overnight culture following the manufacturer's instructions. DNA was eluted in 30 µl of MQ water and stored at -20°C.

2.2.1.3 Polymerase Chain Reaction (PCR)

To check if a strain contains a desired mutation, a polymerase chain reaction was done using the Go-Taq Polymerase and a peqlab Biotechnologie GmbH peqSTAR 96 Universal PCR machine. The reaction was made in MQ-H_2O, in a total volume of 25 µl per sample and contained 25 nM of dNTP, 5 µM of reverse and forward primers, 1X Mg^{2+} free reaction buffer, 2.5 mM of $MgCl_2$ solution, 1 unit of Go-Taq Polymerase and 1 µl of template (total-cell DNA, extracted as described above).

A cycle program was as follows:

Denaturation	96°C	10 min	
Denaturation 2	96°C	30 sec	
Annealing	(T_m -5) °C	30 sec	40X cycles
Elongation	72°C	1 min/kb	

Elongation 2	72°C	1 min
Storage	8°C	indefinite

2.2.1.4 PCR product purification for cloning

The QIAGEN PCR purification kit or QIAGEN Gel Extraction Kit were used for cleaning DNA fragments for cloning following manufacturer's instructions. DNA was eluted in 30 µl of MQ water and stored at -20°C.

2.2.1.5 Restriction digestion of purified DNA

30 µl of PCR purified DNA or plasmid DNA were digested following manufacturer's instructions (NEB, Roche). Samples were incubated at the optimum digestion temperature for a minimum of 1 hour to overnight.

2.2.1.6 Ligation of DNA fragments

A final volume of 20 µl per ligation reaction was used to ligate fragments of DNA following manufacturer's instructions. For this purpose, 10 µl of Quick Ligase Buffer (2X), 1 µl of Quick Ligase, 2 µl of purified plasmid and 7 µl of purified PCR product were used. The reaction mix was left at room temperature for 5 minutes.

2.2.1.7 Agarose gel electrophoresis of PCR products or plasmid DNA

A 1% (w/v) agarose gel was used in order to separate and visualise DNA fragments acquired from PCR reactions and digested or isolated plasmids. The agarose was melted in 1XTAE and allowed to cool to 55°C. Safeview (5 µl per 100 ml) was added to visualise the DNA under UV light following manufacturer's instructions. Gels were run from 80-120 V for 20-50 min. The DNA was visualised and its size was checked using DNA ladders on the GelDoc XR+ system (BioRad) using XcitaBlue conversion screen Cat. No. 1708182 (for DNA to be gel-extracted).

2.2.2 Microbiology methods

2.2.2.1 Overnight cultures

Overnight cultures were made in 5 ml of LB medium or M9 medium plus the relevant additives from a single colony after streaking from a -80°C stock the desired strain on a LB-agar plate with the appropriate complements. Cultures were incubated overnight under agitation at the required temperature.

2.2.2.2 Stocks at -80°C

Stocks were made from 700 µl of overnight culture mixed with 700 µl of 80% glycerol in an Eppendorf tube. The stock was well-mixed, sealed with Parafilm and stored at -80°C.

2.2.2.3 Transformation of *E. coli* by $CaCl_2$ treatment followed by heat shock

0.5 ml of an overnight culture to be transformed was diluted in 25 ml of LB medium and incubated 2 hours at 37°C under agitation (2.5 hours for *recA*$^-$ strain like XL1-Blue or slow-growing strains like dimer resolution mutants). During that time 10 ml of 0.1 M of $CaCl_2$ were prepared in dH_2O and this solution was kept on ice. The culture was divided in 1 ml aliquots and centrifuged for 1 minute at top speed. Supernatants were removed and pellets were re-suspended in 0.5 ml of fresh cold 0.1 M $CaCl_2$. The samples were left on ice for 30 minutes. Then, cells were centrifuged for 1 minute, the supernatant was removed and 0.1 ml of cold 0.1 M $CaCl_2$ was added along with 10 µl of newly ligated DNA (or 4 µl of previously isolated plasmid DNA). Samples were left on ice for 30 minutes. The cells were then heat shocked for 5 minutes at 37°C in a water bath (or heat-block) before being placed back on ice for 2 minutes. Next, 0.4 ml of LB medium was added to the cells, which were incubated at the desired temperature (30°C for pTOF24 derivatives) for an hour under agitation. Finally, the cells were harvested by centrifugation,

0.4 ml of supernatant was discarded and 100 μl of re-suspended cells were plated on LB agar plates with the selective marker and allowed to grow one (for the strains transformed with a plasmid DNA) or two nights (for the strains transformed with a ligation) at the desired temperature (30°C for pTOF24 derivatives). A control without DNA was always prepared in parallel.

2.2.2.4 Plasmid mediated gene replacement

Plasmid mediated gene replacement (PMGR) is one of the methods of strain construction in *E. coli* based on homologous recombination. It was designed by (Link and Phillips, Dereth 1997) and integrates/removes very precisely from the chromosome pieces of DNA without the introduction of a selection marker. The method uses a multicopy plasmid pTOF24 (Merlin, Mcateer, and Masters 2002) that contains a temperature sensitive replication protein (*repA*^{ts}), that replicates at 30°C but not at 42°C, two selection markers: *cat*, encoding the chloramphenicol resistance gene, and *aph*, encoding the kanamycin resistance gene, and a negative selection marker (*sacB*) that encodes the levansucrase gene from *B. subtilis*, toxic in the presence of sucrose (Suc). The kanamycin gene is flanked by a multiple cloning site (MCS) containing the PstI and SalI restriction sites (Figure 2.1).

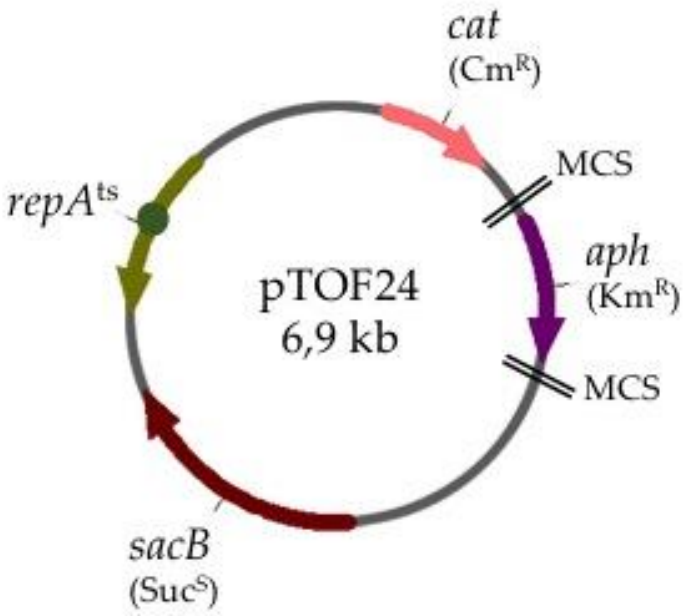


Figure 2.1. Map of pTOF24 plasmid.

pTOF24 plasmid encodes a temperature-sensitive replication initiation protein *repA*^{ts}, two positive selection markers *cat* (Cm^R) and *aph* (Km^R) and a negative selection marker *sacB* (Suc^S). The *aph* gene is flanked by a multiple cloning site (MCS). Dark green circle represents the origin of replication.

During the plasmid construction, the kanamycin gene in pTOF24 is replaced by a ~800 bp fragment of interest. This fragment consists of two ~400 bp DNA sequences that are homologous to the regions flanking the sequence-to-be-modified in *E. coli* chromosome. A crossover PCR was used to join these two ~400 bp fragments of DNA instead of restriction and ligation. Separate PCR reactions were done using 4 different primers (F1-R1 and F2-R2). Primers R1 and F2 were designed to have a ~20 bp homology to each other. To obtain a crossover PCR fragment, first, two ~400 bp fragments of DNA were separately amplified from DNA template (MG1655 or BW27784 extracted as described above) using primers F1-R1 and F2-R2 and the proof-reading polymerase Herculase. Second, the two DNA fragments were used as templates in a crossover PCR using primers F1-R2. During the second PCR reaction the homology fragments overlapped and created a ~800 bp product (Figure 2.2 A).

The cycle program was as follows:

Denaturation	98°C	30 sec	
Denaturation 2	98°C	5-10 sec	
Annealing	(T_m -5) °C	30 sec	30X
Elongation	72°C	30 sec/kb	
Elongation 2	72°C	10 min	
Storage	8°C	indefinite	

F1 and R2 primers for this crossover reaction were designed in order to flank the ~800 bp sequence with PstI and SalI cutting sites, allowing the cloning of the fragment into pTOF24 plasmid (Figure 2.2 B).

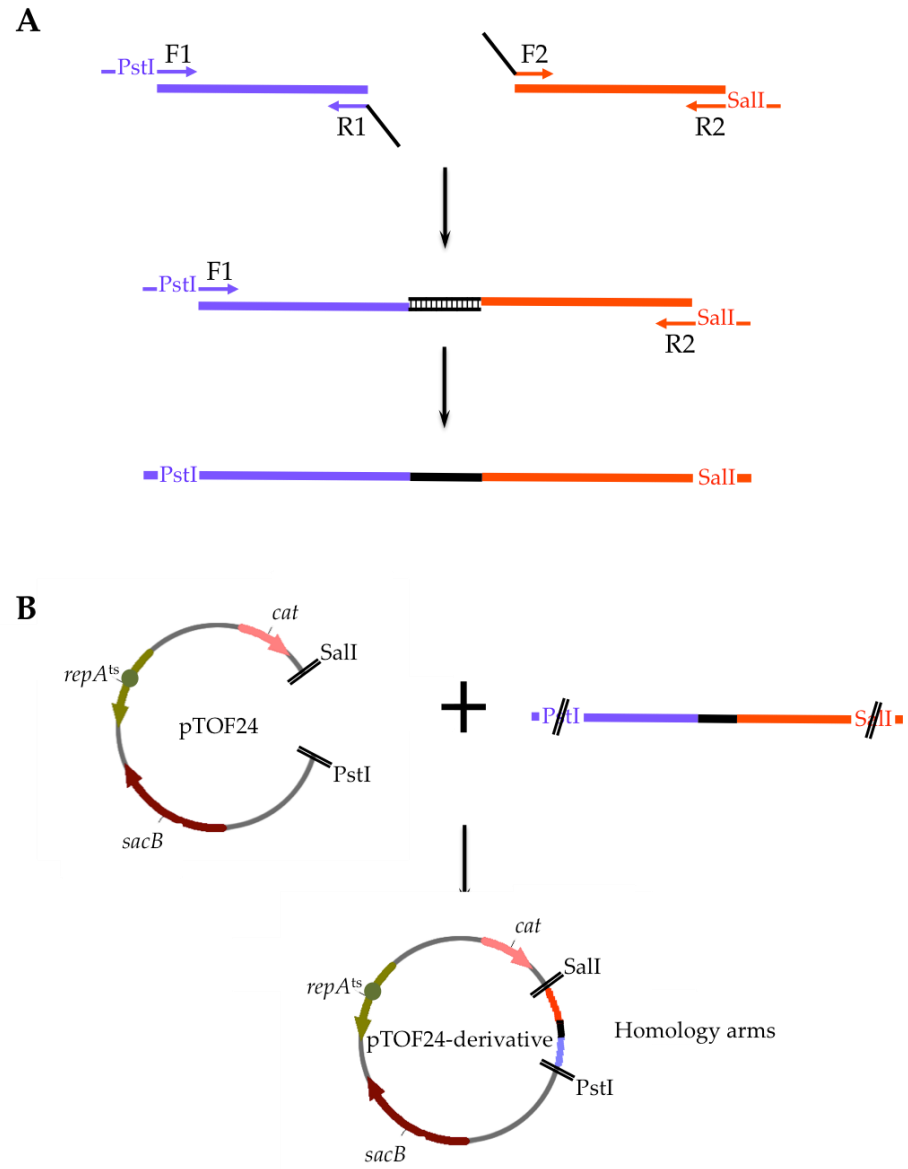


Figure 2.2. Construction of a pTOF24 derivative vector using crossover PCR.

A. Crossover PCR. First, two DNA fragments (purple and orange), that are homologous to the regions flanking the sequence-to-be-modified in *E. coli* chromosome, were amplified separately using two pairs of primers, F1 and R1, F2 and R2. Primer F1 adds a PstI cutting site to the first fragment whereas primer R2 adds a SalI cutting site to the second fragment. Primers R1 and F2 have ~20 bp of homology to each other (black lines). Then, the products from the first PCR were used as templates for the second PCR with primers F1 and R2. The region of homology permits the formation of a new, single crossover PCR product. B. pTOF24 plasmid and the crossover PCR product were digested with SalI and PstI enzymes and ligated together. This resulted in the *aph* gene being replaced with the crossover PCR product.

To introduce/remove a DNA fragment of interest using a pTOF24 derivative plasmid into/from the chromosome of *E. coli* (Figure 2.3), $CaCl_2$ transformation (described above) was performed allowing the introduction of this plasmid into the desired strain. Transformants were incubated overnight at 30°C on chloramphenicol plates. Successful transformants were then streaked onto fresh LB plates supplemented with chloramphenicol and incubated overnight at 42°C to allow the selection of cells that have successfully integrated the plasmid into their chromosome by homologous recombination. Integrants will be identified as large colonies and will be purified on LB plates supplemented with chloramphenicol and incubated at 42°C overnight. Then, a selection for the cells that have excised the plasmid from their chromosome was performed by selecting colonies from 42°C, inoculating them in liquid LB medium and allowing the culture to grow at 30°C overnight under agitation. During the incubation, the cells are expressing levansucrase from the *sacB* gene, which in presence of sucrose is lethal for *E. coli*. Then, 10^{-4} and 10^{-5} dilutions of the overnight cultures were plated onto LB agar plates supplemented with 5% sucrose. The next day colonies were checked for chloramphenicol sensitivity by patch test to ensure that the plasmid has been excised. Sucrose resistant chloramphenicol sensitive colonies were tested for the integration of the fragment of interest by PCR with the F1-R2 primers from the relevant crossover PCR. The chromosomal modification for positive colonies was confirmed by Sanger sequencing.

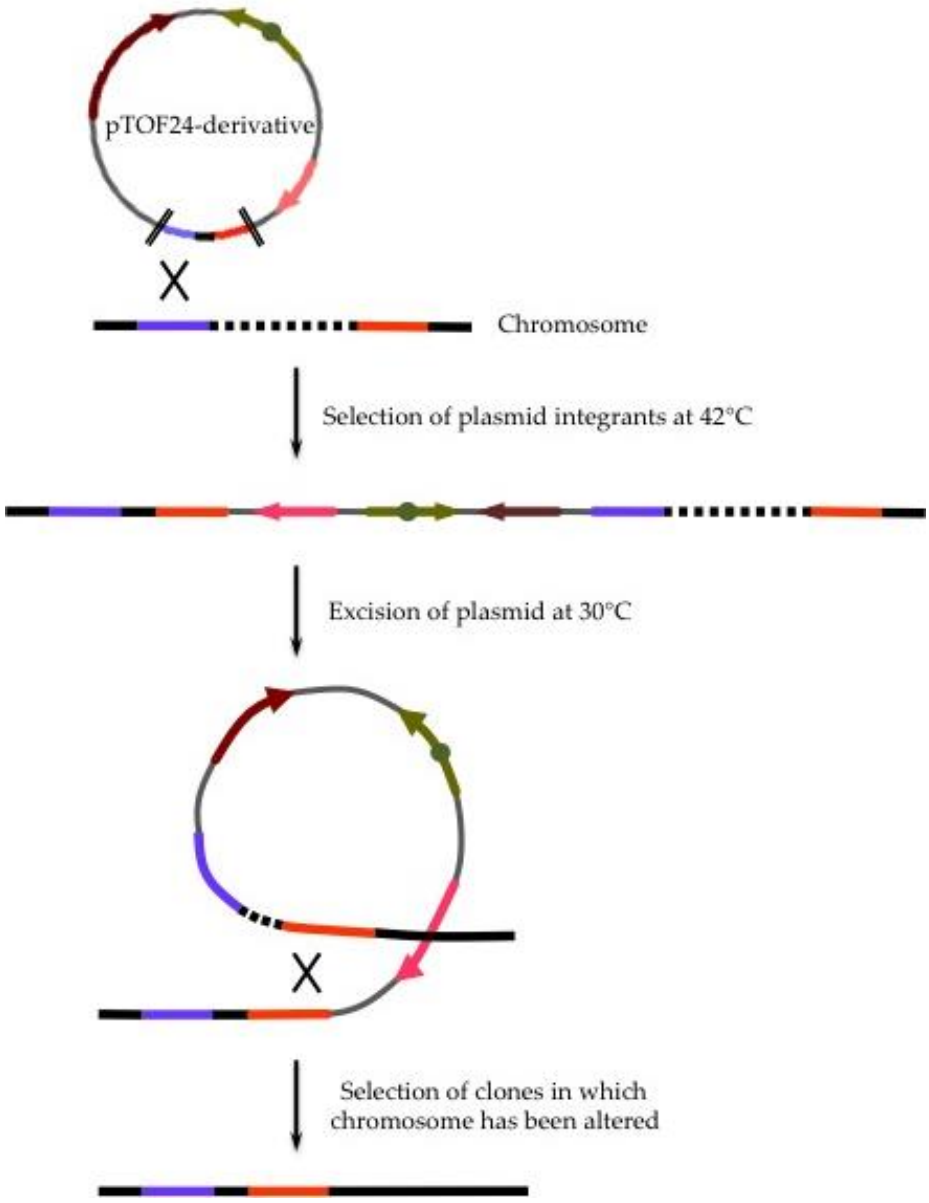


Figure 2.3. Plasmid mediated gene replacement (PMGR).

pTOF24-derivative plasmid has two regions of homology to the chromosome (purple and orange). At 42°C the plasmid can only be replicated if it integrates into the chromosome due to encoding a temperature-sensitive replication initiation protein. At this stage the plasmid integrants are selected by streaking them on chloramphenicol plates at 42°C. Growing cells in a liquid culture at 30°C allows the selection of the cells that have excised the plasmid from their chromosome. During the incubation, the cells are expressing levansucrase from the *sacB* gene, which in presence of sucrose is lethal for *E. coli*. Then, 10^{-4} and 10^{-5} dilutions of the overnight cultures are plated onto LB agar plates supplemented with 5% sucrose. The next day colonies are checked for chloramphenicol sensitivity by patch test to ensure that the plasmid has been excised. When integration and excision occur using different homology arms, the wild type chromosome region is replaced by the modified DNA insert from the plasmid. The chromosomal modification is tested by the PCR with the F1 and R2 primers and further confirmed by Sanger sequencing.

2.2.2.5 Sanger sequencing of DNA

Sanger sequencing was used to check genotypes of strains and constructions of plasmids. The BigDye® Terminator v3.1 Cycle-Sequencing Kit was used for sequencing DNA following manufacturer's instructions. Purified prior to sequencing PCR product (or plasmid DNA) was used as a template DNA. In a total volume of 10 μl per reaction, 2.5 μl of template DNA, 0.5 μl of primer pKO (F or R), 2 μl of BigDye and 5 μl of MQ water were used.

The sequencing cycle program was as follows:

Denaturation	96°C	1 min	
Denaturation 2	96°C	10 sec	
Annealing	(T_m -5) °C	5 sec	30X
Elongation	60°C	4 min	
Storage	10°C	indefinite	

Sequencing reactions were run by the Edinburgh Genomics, University of Edinburgh, using an ABI 3730XL capillary sequencing instrument (https://genomics.ed.ac.uk/work-with-us/sequencing).

2.2.2.6 Growth rate and spot tests

An overnight culture was diluted in 10 ml of LB or M9 medium supplemented with the relevant additives to an OD_{600nm} of 0.01. Cells were grown under agitation at the required temperature until an OD_{600nm} of 0.2-0.3 and re-diluted in new medium in order to have an OD_{600nm} of 0.01. This initial growth was done in order to give the time to all the cells to enter exponential phase after the stationary phase. Then, cells were grown until an OD_{600nm} of 0.2-0.3 within exponential phase. At this point, 1 ml of T0 sample was taken. Serial dilutions of T0 sample were performed, where cells were diluted from 10^0 to 10^{-6}. 5 μl of each dilution was taken and spotted on two LB agar plates supplemented

either with 0.5% glucose or 0.2% arabinose. The cultures were divided into two parts. In order to induce a DSB, 0.2% of arabinose was added to one subculture. As a control, 0.5% of glucose was added to the second subculture, to keep *sbcCD* supressed. The OD_{600nm} of these cultures was measured every hour for the duration of 5 hours in order to detect any changes in the growth behaviour of cells. Cultures were kept in logarithmic phase by dilution. Samples for spot-tests (T1-T5) were taken every hour and spotted on LB agar plates supplemented either with 0.5% glucose or 0.2% arabinose as described above. Plates were incubated at the required temperature overnight and analysed the next day. These experiments were repeated three times for each condition and each strain.

2.2.2.7 Preparation of P1 lysate

An overnight culture carrying the gene or mutation of interest was diluted 1/10 in 10 ml of LB containing 2.5 mM of $CaCl_2$ and this new culture was allowed to grow for 2 hours at the desired temperature under agitation. 0.2 ml of these cells were mixed with 0.1 ml of a previously made P1 lysate (10^0 to 10^{-5} dilutions in phage buffer were used). The mix was incubated at the required temperature for 30 minutes under gentle agitation. After phage adsorption, 2.5 ml of molten LC top agar plus 5 mM of $CaCl_2$ were added to the mix that was poured onto LC bottom agar plates containing 5 mM of $CaCl_2$. The plates were incubated non-inverted overnight at the required temperature. The next day, 5 ml of phage buffer was added on the plate that gave confluent lysis and the top agar and phage buffer were scrapped off into a 5 ml bottle using a pipette. 100 µl of chloroform were added to kill any bacterial cells present, the solution was well mixed and left at 4°C for 30 minutes. Then, the lysate was centrifuged for 5 minutes at 5,000 rpm at room temperature and the supernatant was

stored at 4°C in the dark in a sterile detergent-free 2 ml bottle with 100 µl of chloroform added to it.

2.2.2.8 P1 transduction

An overnight culture of a strain to be transduced was grown in LB supplemented with 2.5 mM of $CaCl_2$ at the desired remperature. The next day, cells were divided in 1 ml aliquots in 1.5 ml Eppendorf tubes that were centrifuged for 1 minute at top speed in a table-top centrifuge. Then, pellets were re-suspended in 0.1 ml of LB supplemented with 2.5 mM $CaCl_2$ and 0 µl, 1 µl, 10 µl and 100 µl of the desired P1 lysate were added. Controls containing 1 µl, 10 µl and 100 µl of phage without overnight culture were also prepared. Samples were incubated for 20 minutes at the desired temperature. 0.8 ml of LB supplemented with 2 mM of Sodium Citrate were added to prevent further infection by the phage. Cells were incubated for 1 h at the desired temperature under agitation to allow the selective marker to be expressed. Cells were centrifuged for 1 minute, 0.8 ml of supernatant were discarded, the rest of the supernatant was used to re-suspend cells and 0.1 ml of culture was plated on LB agar plates with the appropriate selective marker and incubated overnight at 37°C (two overnights at 30°C). In order to purify transductants from residual phages, individual colonies were streaked using selective LB agar plates, twice. To ensure the successful co-transduction, transductants were tested phenotypically or genotypically by PCR.

2.2.3 Two-dimensional agarose gel electrophoresis and Southern blot

2.2.3.1 Preparation of plugs for two-dimensional (2D) agarose gel electrophoresis and digestion of agarose embedded DNA

An overnight culture was diluted to an OD_{600nm} of 0.01 in 25 ml of LB supplemented with 0.5% of glucose. The cells were allowed to grow until an OD_{600nm} of 0.2-0.3, re-diluted in 25 ml of LB supplemented with 0.5% of glucose,

and 0.2% of arabinose was added to the culture in order to induce a DSB. The culture was left to grow for an additional hour. Then, the sample was spinned down at top speed for 7 minutes at 4°C and the pellet was re-suspended in 1 ml of cold TEN buffer and then centrifuged again for 7 minutes at 4°C. This wash was repeated one more time and the final amount of TEN added was calculated in order to have a final OD_{600nm} of 80. The cells were kept on ice at all times. The sample was mixed (50:50) with melted 0.8% (w/v) low melting point (LMP) agarose in sterile MQ water and poured into plug moulds in order to make plugs (the amount of cells used gave in average 3-5 plugs per strain). Plugs were incubated overnight at 37°C in 5 ml of proteinase K solution dissolved in NDS buffer. The next day, the solution was replaced with fresh proteinase K solution for another overnight incubation. Then, each plug was washed in 1.5 ml of 1X TE buffer for 5 hours, changing the buffer every hour. After washing, the DNA in the plugs was digested in 0.5 ml of 1X the appropriate restriction buffer with 500 units of the appropriate enzyme. The digestion occured overnight at 37°C. Unused plugs were stored at 4°C in fresh NDS buffer without proteinase K.

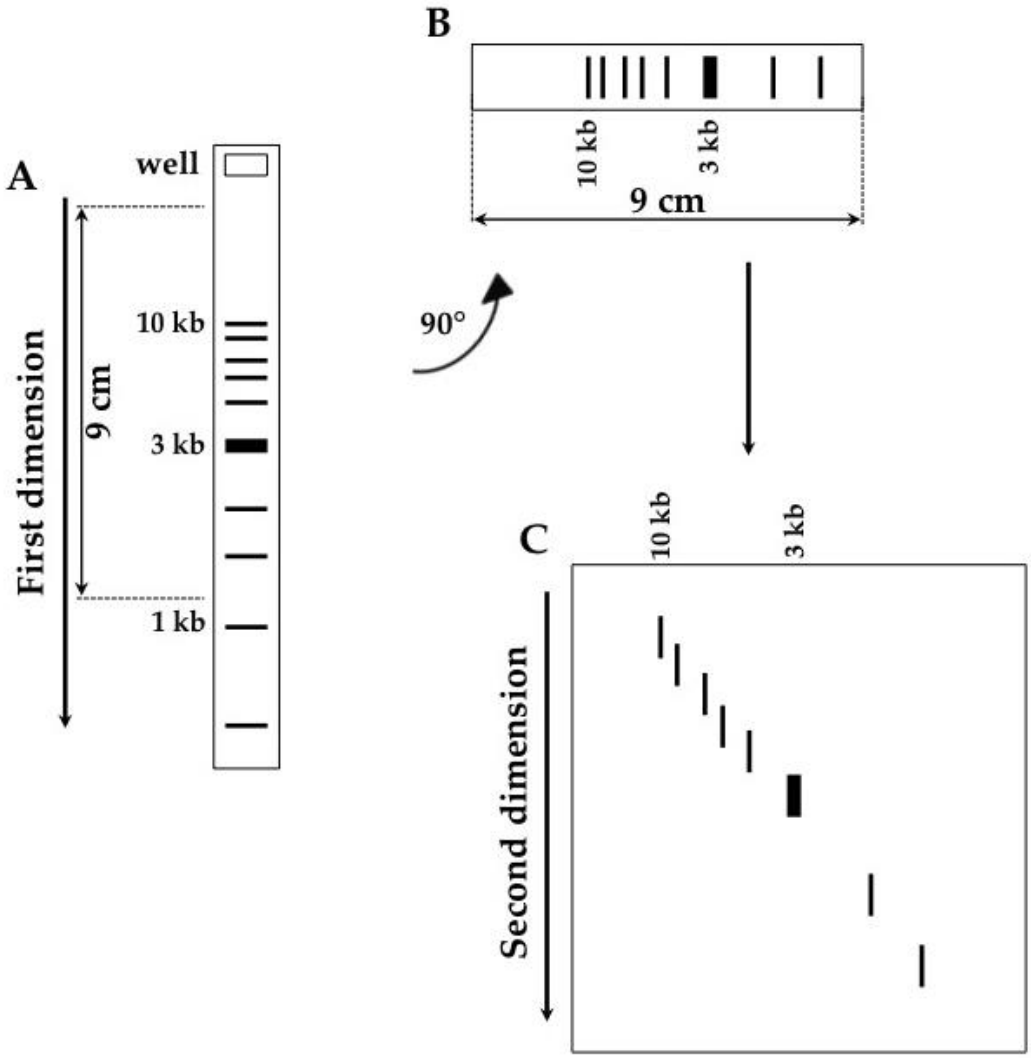


Figure 2.4. Two-dimensional agarose gel electrophoresis.

2D agarose gel electrophoresis is used to separate DNA molecules by size and by shape. An example of 2D agarose gel electrophoresis of 1 kb ladder is shown here. A. During the first dimension, DNA was separated by size. B. Then, a 9 cm lane containing DNA of interest was sliced out, turned 90° and run in the second dimension in order to separate DNA molecules by shape. C. The second dimension was run in the presence of 0.3 µg/ml of ethidium bromide to maximise the difference in shape between different branched structures.

2.2.3.2 Native-native two-dimensional (2D) agarose gel electrophoresis

2D agarose gel electrophoresis was used to separate different DNA structures and to detect a pause in DNA replication (Figure 2.4). In the first dimension, DNA was separated in function of size to minimise the structural differences between fragments. 5 µl of cooled-down 0.4% (w/v) MELFORD agarose in 1X TBE were used to attach plugs to a comb. Plugs were left for 30 min at 4°C. Then, 200 ml of the same agarose were carefully poured around the plugs in the gel tank (10x15 cm) and allowed to set for 30 more minutes at 4°C. 15 µl of 1 kb ladder were added to identify the size of the fragments after migration. The gel was run for 36 hours in 1X TBE at 27 V and 4°C. The lane with the ladder was cut out after the run and stained with 0.5 µg/ml of EtBr for 20 minutes to check the position of the 7 kb (*terC*) and 4 kb (*terA* and *terB*) fragments. 9 cm lines which contained the fragments of interest were then cut out of the gel. Thereafter, slices were turned 90° with the bigger-size product placed to the left and run in a second dimension to see the difference in DNA structure. 500 ml of cooled-down 1% (w/v) MELFORD agarose in 1X TBE were poured around the slices in 20x25 cm tank. The agarose and the 1X TBE running buffer contained 0.3 µg/ml of EtBr. This gel was run for 14 hours at 107 V and 4°C. A pump allowed the circulation of the EtBr across the tank. Thereafter, UV light was used to visualise the linear arc structures of the DNA.

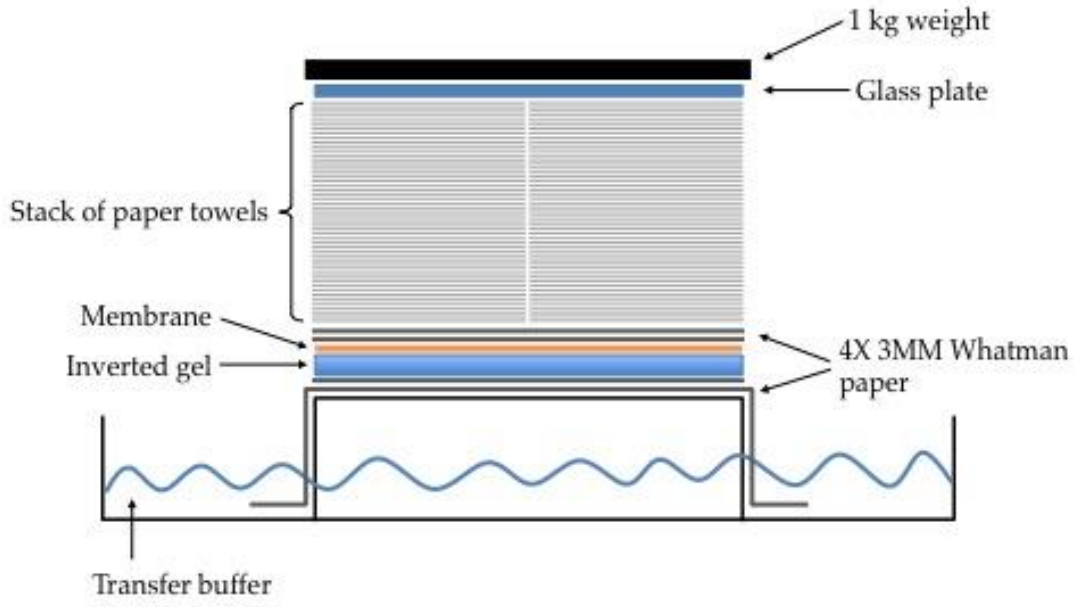


Figure 2.5. Transfer from an agarose gel to a positively charged membrane (Southern blot).

One sheet of Whatman paper soaked in transfer buffer was placed on top of a solid support in order to make a “bridge” with each edges dipping into a tray containing some transfer buffer. One small size Whatman paper was soaked in the transfer buffer and placed on top. The gel was inverted and placed on the top of the small size Whatman paper. A positively charged nylon membrane was placed on top of the gel. Two sheets of small size Whatman paper soaked in the transfer buffer, two blocks of tissues, a glass plate and 1 kg weight were placed on top of the membrane, in the written order.

2.2.3.3 Southern blot

A southern blot is used to transfer DNA fragments from a gel to a nylon membrane (Figure 2.5). As a result, DNA fragments are immobilised to the membrane and can be used for hybridisation analysis. A Southern blot was done immediately after running the second dimension gel. After visualisation of the DNA under UV light, the gel was washed with the depurination solution for 20 minutes and, afterwards, in the transfer buffer for 1 hour. Thereafter, the transfer was set up and the DNA was transferred overnight. To set up the transfer, the gel was placed facing down between four pieces of Whatmann paper (2 under the gel and 2 on top of the gel) soaked in transfer buffer. One of the Whatmann paper pieces was placed in a way to be always in contact with the transfer buffer. The membrane was marked in a corner in order to orientate it and placed on the top of the gel between the gel and the Whatmann paper. A block of tissue paper, a glass plate and a 1 kg weight were placed on the top of the gel, the membrane and Whatmann paper. The next day, the membrane was allowed to dry for 1 hour at room temperature and was UV cross-linked with a Stratagene UV Stratalinker™ 1800 using 1000 J/m^2. After cross-linking, the membrane was placed in a sealed plastic bag between two Whatmann papers to avoid any dust or humidity and stored at 4°C.

2.2.3.4 Preparation of ^{32}P labelled probe stock

Probes for radioactive probing of membranes were prepared using Prime-It II Random Primer Labeling kit (Agilent). DNA template for a probe was amplified using PCR. The reaction was made in MQ water, in a total volume of 50 µl per sample and contained 25 nM of dNTP, 10 µM of F and R primers, 1X Herculase reaction buffer, 1 unit of Herculase polymerase and 1 µl of MG1655 template (total-cell DNA, extracted as described above). To prepare a probe, a reaction mix of a total volume of 25 µl was made. First, 5 µl of 2.5

ng/μl of DNA template purified by gel extraction, 5 μl of random 9-mer oligonucleotides (27 OD units/ml) and 4 μl of sterile MQ water were mixed together and boiled in a thermocycler for 5 minutes to denature the DNA and primers. This mix was cooled down at room temperature for 5 minutes. Then, 5 μl of 5X dATP buffer (mixture containing dCTP, dTTP and dGTP, 0.1 mM of each) was added and 5 units of Exo(-) Klenow DNA polymerase (1 μl) were placed on the side of the tube. Following, 5 μl of 20 mCi/ml [α–^{32}P]dATP was added and well mixed by pippeting up and down. Then, the reaction mix was incubated in a thermocycler for 30 minutes at 37°C. The mix was completed up to 100 μl with distilled water and run through a GE Healthcare Illustra Microspin™ G-25 Column, according to manufacturer's instructions, in order to obtain a clean probe.

2.2.3.5 Membrane hybridisation and washing

Firstly, in order to pre-hybridise the membrane, it was placed in 25 ml of pre-hybridisation buffer (hybridisation buffer without the radioactive label) for 2-6 hours at 65°C in a hybridisation oven under rotation. During pre-hybridisation, the radioactive probe was boiled at 100°C for 5 minutes and then immediately transferred on ice for another 5 minutes. The pre-hybridisation buffer was then replaced with 20 ml of hybridisation buffer containing 25 μl of probe per membrane. The membrane was left at 65°C overnight. The next day, the membrane was rinsed at room temperature with 2X SSC and washed in 200 ml of Wash Buffer 1 in a 60°C oven for 15 minutes. A second wash was carried out in 200 ml of Wash Buffer 2 in the 60°C oven for 30 minutes. After these washes, the membrane was rinsed in 2X SSC at room temperature and placed in a GE Healthcare storage phosphor screen (Cat. No. 63-0034-86 and 63-0034-79) and left for 2-5 days to expose. The

images from the phosphor screen were taken using a Typhoon FLA 7000 phosphor imager scanner and analysed using the *ImageQuant™ TL* software.

2.2.3.6 Stripping the membrane

A probe was stripped from a membrane that needed to be probed by a different probe. The membrane was washed with 50 ml of stripping buffer for 1 hour at 65°C and then with 200 ml of Wash Buffer 2 for 30 minutes at 65°C. Afterwards, the membrane was rinsed in 2X SSC at room temperature. Membranes were placed in a GE Healthcare storage phosphor screen and exposed for 1 hour to check for the absence of the old probe before being probed again.

2.2.4 Next generation sequencing methods

2.2.4.1 Growth conditions of strains for next generation sequencing

Approximately 5 x 10^8 cells were used for each ChIP experiment. Cultures were grown in LB supplemented with 0.5% of glucose at 37°C. Overnight cultures were diluted to an OD_{600nm} of 0.01 and then left to grow until they reach an OD_{600nm} of 0.2, then re-diluted to an OD_{600nm} of 0.01 and allowed to grow to an OD_{600nm} of 0.2 – 0.25 and then harvested. The re-dilution step was included to increase the homogeneity of the culture and reproducibility. When cells containing the arabinose inducible *sbcCD* system were re-diluted the second time to an OD_{600nm} of 0.01, 0.2% of arabinose was added to the culture in order to induce the expression of the SbcCD endonuclease and, in the presence of the palindrome, the production of a DSB. Cultures were then left to grow for 3.5 hours at an OD_{600nm} maintained between of 0.1 and 0.3 by regular dilutions. Optimisation of the ChIP protocol showed a stronger signal of RecA binding to DNA in the terminus region after 3.5 hours of DSB induction.

Cultures of temperature-sensitive mutants, like TopoIVts, were initially grown in LB supplemented with 0.5% of glucose at permissive temperatue (30°C). After the second re-dilution step to an OD_{600nm} of 0.01, 0.2% of arabinose was added in order to induce a DSB and the cultures were grown for 3 hours at 30°C at an OD_{600nm} maintained between 0.1 and 0.3 by regular dilutions. Then, these cultures were shifted to non-permissive temperature (42°C) and allowed to grow for 30 minutes. At this point, 200 ml of cell culture were collected for the ChIP experiment and 25 ml of the culture were collected for DNA isolation for whole genome sequencing.

2.2.4.2 Chromatin Immunoprecipitation (ChIP)

Protein and DNA were crosslinked by the addition of 1% of formaldehyde to the cultures and incubation for 10 min at 22.5°C under agitation in a temperature-controlled water bath. Then, the reaction was quenched by the addition of 0.5 M of glycine. Cells were harvested by centrifugation at top speed for 5 min at 4°C and washed three times in 1 ml of ice-cold 1X PBS. The pellet was then re-suspended in 250 µl of ChIP buffer. Following, sonication of the crossinked samples was performed using a Diagenode Bioruptor at 30 s intervals for 10 min at high amplitude, which allowed to break the DNA molecule into ~300 bp size fragments. Samples were kept at 4°C at all times. After sonication, 350 µl of ChIP buffer were added to each sample and the samples were mixed by gentle pipetting. Afterwards, 100 µl of these lysates were removed and stored at -20°C as "Input". The remaining samples, marked as "IP", were rotated overnight at 4°C with 1/100 anti-RecA antibody (5 µl per sample). The next day, the IP samples were rotated with Protein G Dynabeads for 2 h at room temperature. All samples were then washed 3 times with 1X PBS supplemented with 0.02% Tween-20. Then the samples were re-suspended in 200 µl of 1X TE buffer + 1% SDS + 2 µl of 10 ng/µl of RNase A.

100 µl of the same buffer was added to the input samples and all samples were incubated in a thermocycler at 65°C for 10 h to reverse formaldehyde cross-links. The next day, DNA from all samples was purified using the Qiagen PCR purification kit. DNA was eluted in 50 µl of MQ water. Samples were stored at -20°C (Figure 2.6).

2.2.4.3 RT-qPCR

All real-time qPCR reactions were carried out in a final volume of 15 µl in the MX3000P qPCR machine (Agilent) or Roche LightCycler® 96 using the Brilliant II SYBR Green qPCR master mix. Input and ChIP DNA were diluted 10 times and 5 µl per reaction was used. 0.5 µl of each F and R 10 mM qPCR primer was used per reaction.

The cycle program was as follows:

Preincubation	95°C	10 min	
2 step amplification	95°C	30 sec	40X
	60°C	1 min	
Melting	95°C	1 min	
	55°C	30 sec	
Data aquisition	95°C	continuous	

All reactions were repeated as technical duplicates in the same run. The assay performance was checked by standard curve for all assays. Data were exported from the MxPro software (for Agilent qPCR machine) and from the LightCycler® 96 SW 1.1 (for Roche qPCR machine) to Microsoft Excel for analysis. Each sample was done in biological triplicates and the mean of all technical and biological repeats was taken.

2.2.4.4 Library preparation for next generation sequencing

ChIP samples were processed following NEB's protocol from the NEBNext® ChIP-seq library preparation kit for Illumina® (Figure 2.7). Briefly, 200 ng of ChIP enriched DNA eluted in 40 µl sterile water were mixed in a sterile microfuge tube with 5 µl of NEBNext End Repair Reaction Buffer (10X) and NEBNext End Repair Enzyme Mix (1 µl). Also, 4 µl of sterile water were added to obtain a total volume of 50 µl. The mixture was incubated in a thermal cycler for 30 minutes at 20°C in order to repair DNA ends by filling in ssDNA overhangs, removing 3'phosphates and 5'phosphorylating the sheared DNA. Here and after every step of library preparation, the DNA was cleaned up using the Qiagen MinElute PCR purification kit according to the manufacturer's instructions. Then, 1 µl of Klenow exo- per 44 µl of end-repaired, blunt DNA was used to adenylate the 3' ends of the DNA. 5 µl of NEBNext dA-tailing reaction buffer was also added to the tube to a total volume of 50 µl. The mixture was incubated in a thermal cycler for 30 minutes at 37°C. Next, 19 µl of dA-tailed DNA was used for ligation of 1 µl of NEXTflex DNA adaptors compatible with Illumina machines using 4 µl of T4 DNA ligase and 6 µl of quick ligation reaction buffer (5X). The mixture was incubated in a thermal cycler for 15 minutes at 20°C. Then, 3 µl of USER enzyme mix were added to excise uracil from the adaptors and create double-stranded ends. The mixture was left in a thermal cycler at 37°C for further 15 minutes. After adaptor ligation, 20 µl of the adaptor-modified DNA fragments together with addition of 2.5 µl of universal primer and 2.5 µl of individual indexes from NEBNext Multiplex Oligos for Illumina kit were PCR amplified using 25 µl of NEBNext Q5 Hot Start HiFi PCR Master Mix. PCR cycling conditions are below:

Initial denaturation	98°C	30 sec	1 cycle
Denaturation	98°C	10 sec	
Annealing/Extension	65°C	75 sec	15 cycles
Final extension	65°C	5 min	1 cycle
Hold	4°C	∞	

Individual indexes were added to each DNA sample in order to separate the libraries from each other during the data analysis.

2% agarose gel electrophoresis was used to run adaptor-ligated DNA fragments for 90 min at 100 V. Then, DNA fragments an average size of ~300 bp were gel-purified using QIAquick® Gel Extraction Kit from QIAGEN. The concentration of all samples was quantified by Qubit® using the dsDNA BR kit according to the manufacturer's instructions before being send to Edinburgh Genomics for sequencing using an Illumina HiSeq 2500/4000 (https://genomics.ed.ac.uk/).

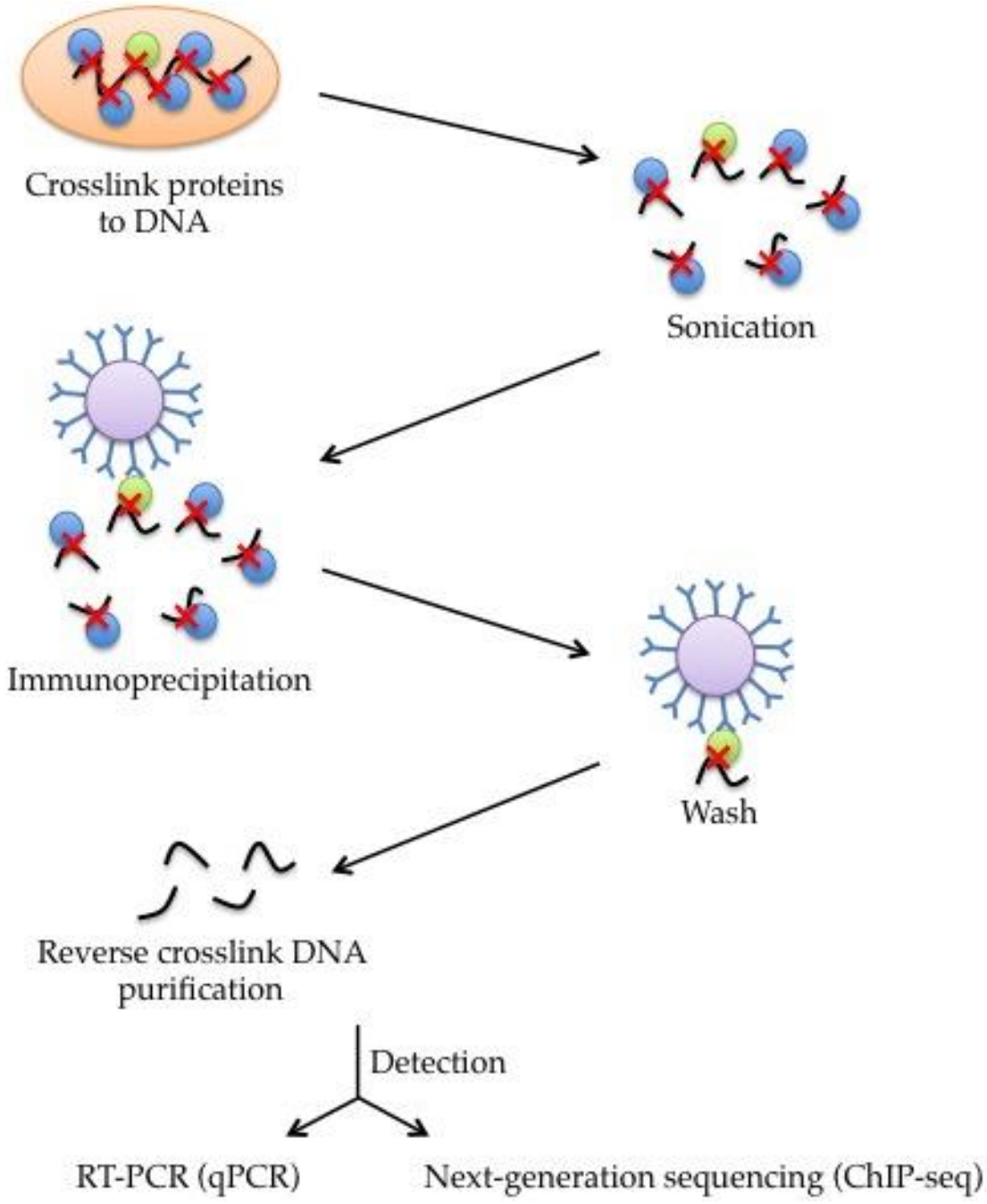


Figure 2.6. Workflow of Chromatin Immunoprecipitation (ChIP).

First, protein-DNA interactions were crosslinked with formaldehyde. Then, cells were sonicated in order to break the cells and sheer the DNA into small uniform fragments. Protein-DNA complexes were immunoprecipitated using an antibody against a protein of interest. Afterwards, antibody-protein-DNA complexes were purified, the crosslinking was reversed and DNA was purified and analysed.

2.2.4.5 Analysis of ChIP-seq data

A "read" is a sequence of nucleotides obtained from a DNA sequencer. The reads are used later to reconstruct the original sequence. 50 bp single-end reads (75 bp paired-end reads for Illumina HiSeq 4000) were mapped to the *Escherichia coli* K12 MG1655 reference genome using Novoalign version 2.07 (www.novocraft.com). Novoalign uses the Needleman-Wunsch algorithm to determine the optimal alignment of reads. Sequences were mapped using default settings, allowing for a maximum of one mismatch per read, except the strategy for reporting repeats (-r). When aligning reads across the whole genome, the –r 'Random' setting for reporting repeats was used. This setting instructs Novoalign to randomly allocate sequences that have multiple alignment locations, thereby equally distributing these hits to all locations. PyReadCounters was used to calculate the overlap between aligned reads and *E. coli* genomic features (Webb *et al*. 2014). The distribution of reads along the *E. coli* genome was visualised using the Integrated Genome Browser (Nicol, Helt, Blanchard, Raja 2009).

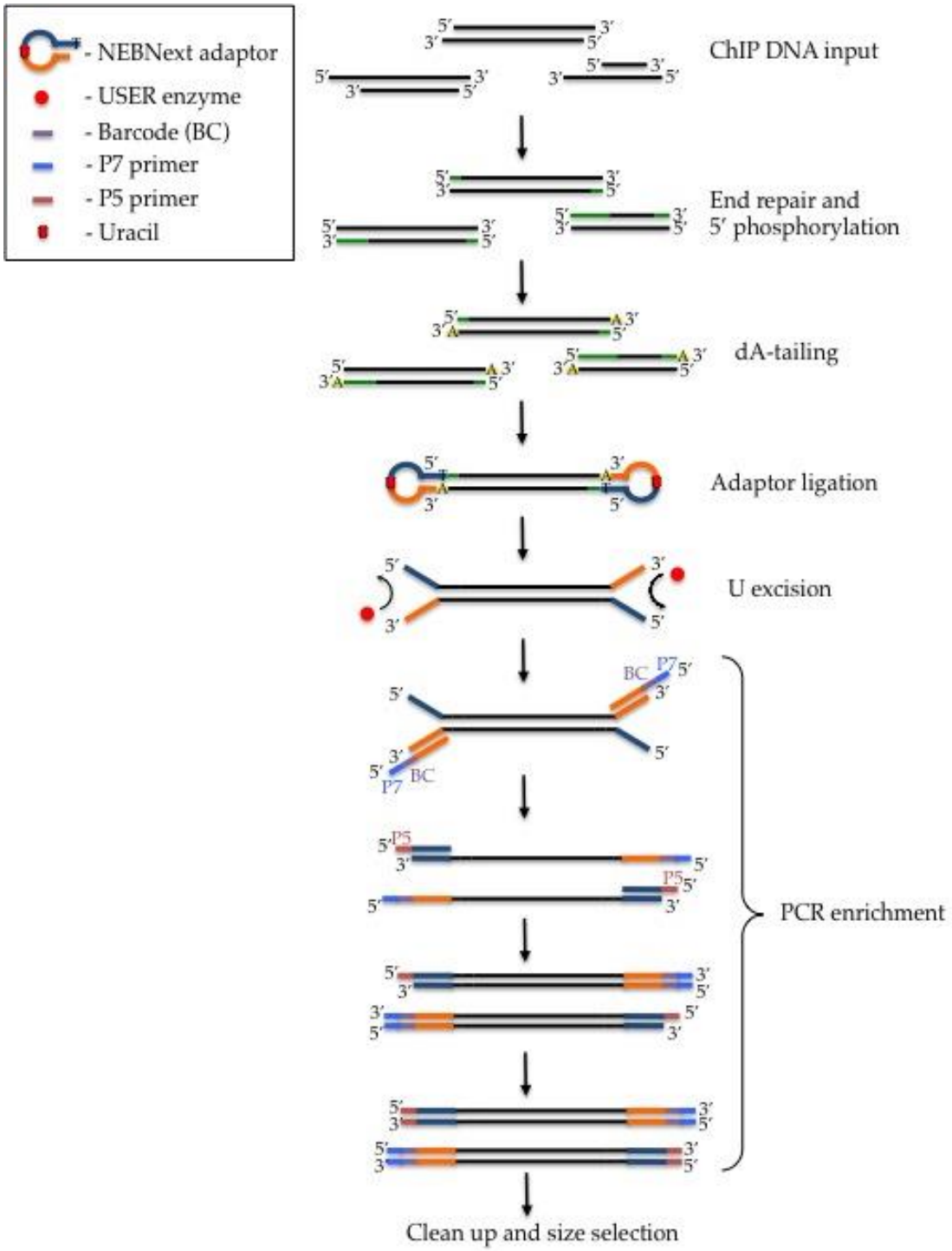


Figure 2.7. ChIP library construction.

Purified ChIP DNA was subjected to end repair to fill in ssDNA overhangs, remove the 3' phosphates and phosphorylate the 5' ends of the sheared DNA. Klenow exo- was used to adenylate the 3' ends of the DNA. DNA adaptors were ligated onto the ChIP DNA using T4 DNA ligase. Afterwards, DNA fragments were amplified using primers P5 and P7. Each P7 primer was barcoded in order to separate the libraries after the sequencing. Agarose gel electrophoresis was used to purify and size select enriched DNA so that the fragment size was ~350 bp.

2.2.4.6 Sample preparation for whole genome sequencing

The DNA used for whole genome sequencing was isolated from cultures using the Promega Wizard® Genomic DNA purification kit by following the manufacturer's instructions. RNase A (supplied with the kit) treatment was carried out for 50 minutes and the DNA was re-hydrated overnight in TE (10 mM Tris (pH 7.4), 1 mM EDTA) at 4°C. To further eliminate potential contamination by RNA, 3 units of Riboshredder (RNase Blend, purchased from epicentre) were added per sample according to the manufacturer's instructions. Samples were purified using the Zymo Genomic DNA Clean & Concentrator kit according to the manufacturer's instructions. The integrity of the DNA was verified by running the samples on a 0.8% agarose gel. The quality of the DNA was determined by Nanodrop analysis (Thermo Scientific) and the concentration of all samples was quantified by Qubit® using the dsDNA BR kit. Finally, the construction of libraries using Nextera DNA Library Prep kit by Illumina and DNA sequencing was carried out on an Illumina HiSeq 2500 platform by Edinburgh Genomics. All samples were sent for sequencing as biological duplicates.

2.2.4.7 Purification of DNA with phenol/chloroform

Extracted DNA samples were purified from Riboshredder using phenol/chloroform method. An equal volume of phenol:chloroform (1:1) was added to the samples in a 1.5 ml tubes. The content was vigorously mixed until an emulsion form was reached. White gel phase was added to the tube for better separation of the protein-free supernatant and the mixture was centrifuged at the top-speed in table-top centrifuge for 1 minute. The upper phase above the white gel phase was transferred to a fresh tube. Then, an equal volume of chloroform was added to the tube and the step was repeated. Recovery of the DNA was done by ethanol precipitation.

2.2.4.8 Ethanol precipitation

2.5-3 volumes of ice-cold 100% ethanol were added to the purified DNA and mixed well. The samples were kept on ice for 15 minutes to allow the precipitation to form. Mixture was centrifuged for 10 minutes at 4°C. The supernatant was removed and the pellet was washed in 1 ml 70% ethanol. The mixture was centrifuged at top-speed for 10 minutes at 4°C. The supernatant was removed and the pellet was dissolved in 50 µl of dH_2O.

2.2.4.9 Whole genome sequencing data analysis

The paired-end raw datasets from an Illumina HiSeq 2500 sequencing platform (obtained from Edinburgh Genomics) were mapped against the genomic sequence of the BW27784 reference strain using the BWA sequence aligner v.0.7.11 and were analysed using SAMtools v.1.2. BW27784 is a modified version of *E. coli* K12 MG1655. For modifications look at the *E. coli* strain Table 2.6. Replication profiles of exponentially growing cultures were calculated by normalising them to the reads of non-replicating stationary-phase wild-type culture. This was done in order to correct for differences in sequence-based recovery across the genome. Dr. Mahedi Hasan provided help with the visualisation of the replication profiles using R language (https://www.r-project.org/). Whole genome profiles were visualised by 1 kb non-overlapping windows binning.

Chapter III

3. DSBR in *lacZ* leads to DNA over-replication in the terminus region

3.1 Introduction

This chapter describes the consequences for the termination of DNA replication of the repair of a double-strand break in the *lacZ* gene in *E. coli.* Although the initiation and elongation of DNA replication are very well studied, the termination of replication remains relatively unknown in both prokaryotes and eukaryotes. In *E. coli,* bidirectional replication starts from a single origin, *oriC*, and completes at a specific region, known as the terminus. This region is flanked by *ter* sites, that when bound by Tus proteins, can stop the progression of replication forks in a polar manner. To understand the mechanism of replication termination in *E. coli*, I have used an approach based on whole genome sequencing (WGS) that was optimised by Dr. Martin White (described in detail in Section 1.6).

The experiments presented in this chapter were based on the data obtained by Dr. Martin White where, using WGS approach, he showed that the strain subjected to a DSB in *lacZ* had an excess of DNA in the terminus region between *terA* and *terB* with a peak in the region of the *dif* site (Section 1.6). Following this observation, I would expect the termination of replication to occur at both *terA* and *terB*/*C*.

The data presented in this chapter show that following a DSB in *lacZ* more replication forks stop at *terA* and *terB* sites confirming that observed DNA over-replication occurs between *terA* and *terB.* Moreover, the additional WGS data suggest that this DNA over-replication originates in the region around *dif* and is not dependent on the predicted place of termination of replication in the absence of over-replication. 3.2 The amount of paused replication forks detected at *terA* and *terB* sites is increased following DSBR at *lacZ.*

To quantify the amount of replication forks paused at *terA, terB* and *terC,* native/native two-dimensional (2D) agarose gel electrophoreses and Southern blots were performed using DNA isolated from strains containing or not *sbcCD* under its native promoter and containing or not a 246 bp interrupted palindrome in *lacZ*. Native/native 2D agarose gel electrophoresis is a technique that allows the separation and identification of various branched DNA molecules (Figure 3.1). In the first dimension, the DNA is separated primarily by size, then the gel is turned 90° and the second dimension separates DNA primarily by shape. Southern blotting allows the visualisation of the DNA fragment of interest on a nylon membrane using ^{32}P α-dATP labelling. To quantify the amount of replication forks paused at *ter* sites, 3 different radioactive probes for *terA*, *terB* and *terC* were used. These 4 kb probes, previously used by (Duggin and Bell 2009a), were amplified by PCR from the MG1655 strain. In the present study, the strains and the growth conditions were the same than that used by Martin White for the WGS described in Section 1.6. Additional control strains in which *sbcCD* was deleted were used in this experiment (DL2151 and DL2874).

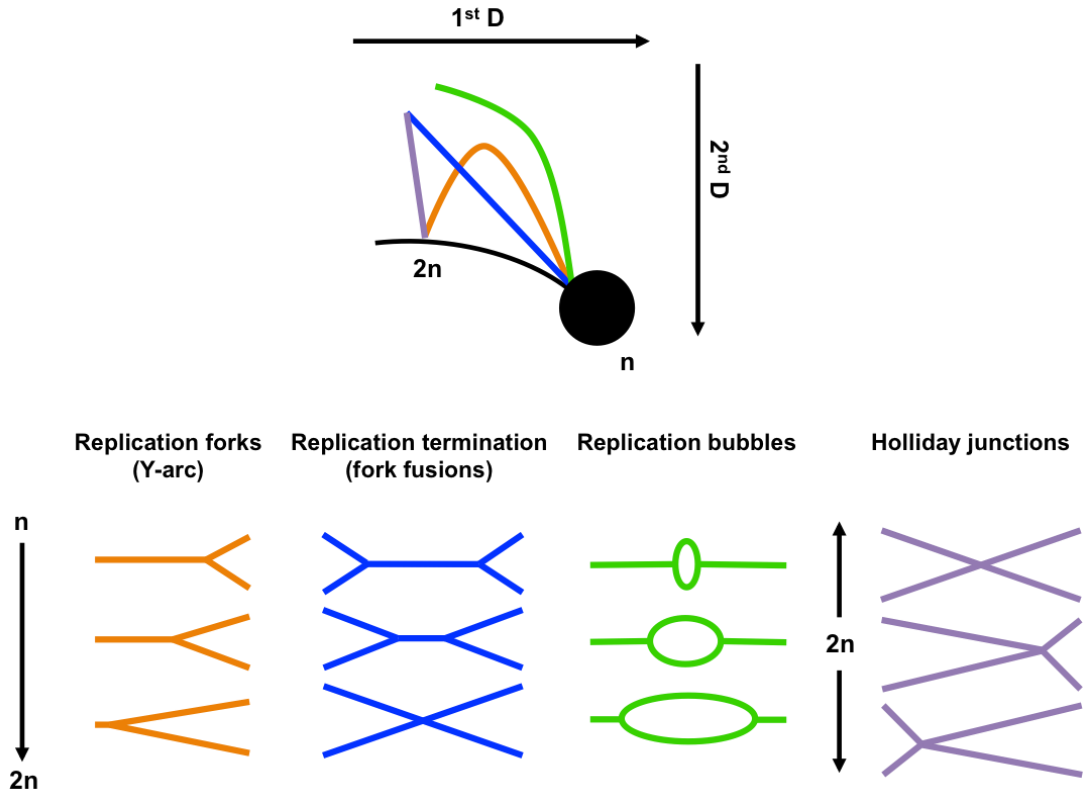


Figure 3.1. 2D agarose gel electrophoresis most common migration patterns.

Big black spots indicate the positions of linear DNA (n). 2n is the location of linear DNA prior to completion of replication. Different molecular shapes of DNA and their positions on the arc of linears are indicated as well. Modified from (Friedman KL 1995).

In the morning, overnight cultures were diluted in LB medium to an OD_{600nm} of 0.01 and grown to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and grown until they reached an OD_{600nm} of 0.2-0.25. Cells were harvested by centrifugation and embedded in 0.8% agarose plugs. Then, the cells were lysed during two overnights and the DNA in the plugs was digested by either NcoI or XmnI enzyme. The NcoI enzyme was used to generate a 6.7 kb fragment which is recognised by the *terC* probe (Figure 3.2 A). The *terC* site is located closer to the middle of the *terC* probe. This permits the visualisation of replication forks that pause at *terC* almost on the top of the Y-arc as showed on the schematic of Figure 3.2 B. The XmnI enzyme was used to generate 4.1 kb and 4.9 kb fragments to visualise *terA* and *terB* sites, respectively (Figure 3.3 A and Figure 3.4 A). The *terA* site is located 1.6 kb away from the beginning of the fragment. The schematic on Figure 3.3 B represents the position on the Y-arc of replication forks that pause at *terA*. The *terB* site is located 3.6 kb away from the end of the fragment. It will stop replication forks that arrive from this side of the fragment and the schematic on Figure 3.4 B represents the position on the Y-arc of replication forks that pause at *terB*. Replication forks (Y-arcs) as well as termination structures (fork fusions) are expected to appear in all fragments due to the nature of the replication termination in *E. coli* when replication forks collide at *ter* sites. Southern blots of 2D agarose gels of the *terC* fragment are shown in Figure 3.2 C. Linear DNA, Y-arcs and termination structures were detected in all strains. Paused replication forks at the *terC* site were detected in all strains at the expected place on the Y-arc. The amount of paused replication forks was larger in the pal^{+} $sbcCD^{+}$ strain compared to the three control strains. Southern blots of 2D agarose gels of the *terA* and *terB* fragments are shown in Figure 3.3 C and 3.6 C, respectively. Linear DNA was detected in all strains. Y-arcs and termination structures were fainter in the *terA* and *terB* fragments than in the

terC fragment. Paused replication forks at the *terA* and *terB* sites were detected in all strains at the expected places on the Y-arcs. Again, the amount of paused replication forks was larger in the *pal*$^{+}$ *sbcCD*$^{+}$ strain compared to the three control strains.

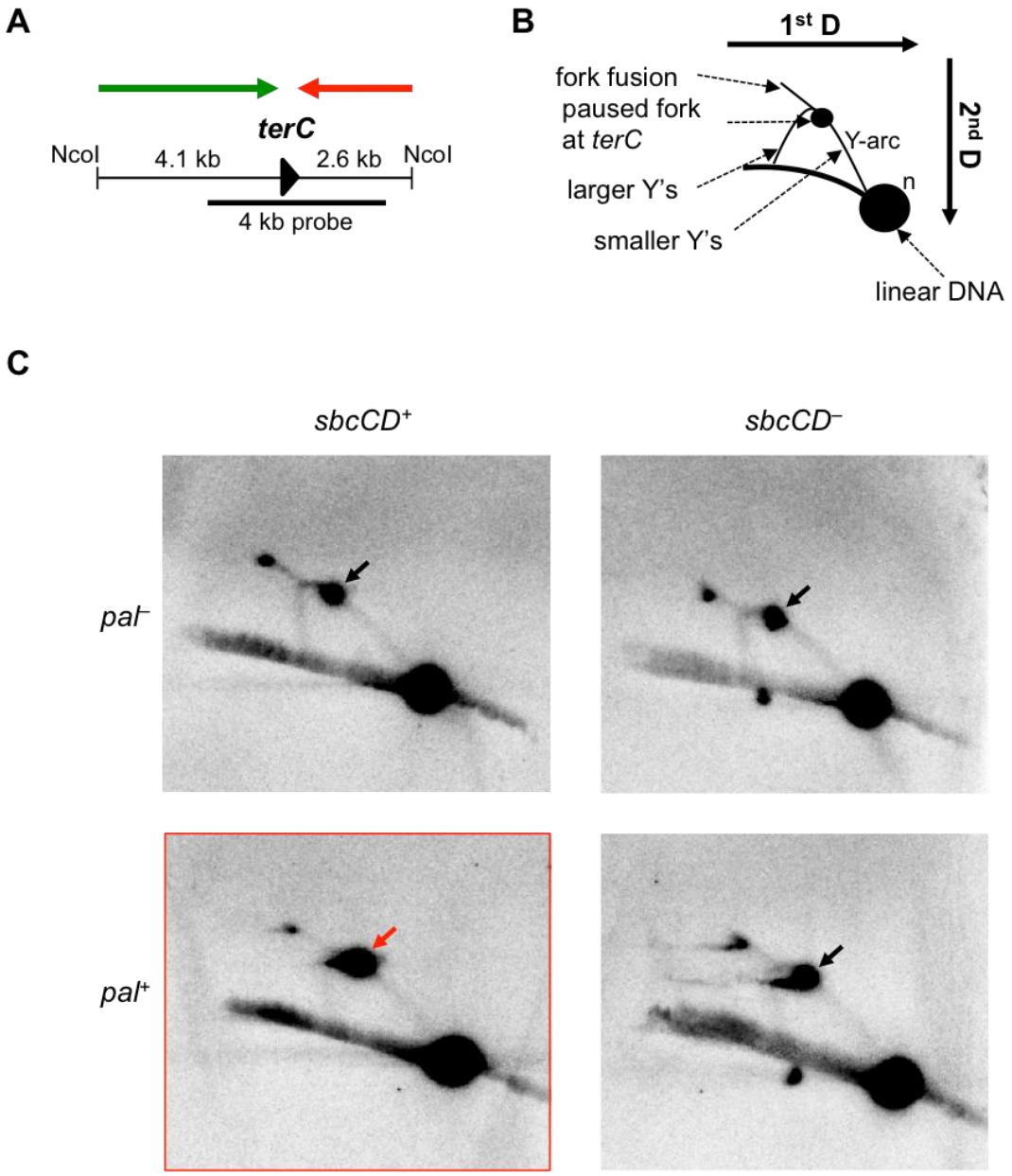


Figure 3.2..Two-dimensional agarose gel electrophoresis of the *terC* fragment.

(A) NcoI digestion map of the region surrounding the *terC* site. The orientation of the *terC* site is indicated by a black triangle. The black thick line is the position of the *terC* probe. The direction of the replication is indicated using red and green arrows. (B) Schematic representation of the migration patterns of branched DNA molecules when separated on a 2D agarose gel electrophoresis and hybridised with the *terC* probe. (C) 2D agarose gel electrophoresis of the *terC* fragment. Paused replication forks are marked by arrows. Strains used were DL1777 (*pal*$^{-}$ *sbcCD*$^{+}$), DL2151 (*pal*$^{-}$ *sbcCD*$^{-}$), DL2859 (*pal*$^{+}$ *sbcCD*$^{+}$) and DL2874 (*pal*$^{+}$ *sbcCD*$^{-}$).

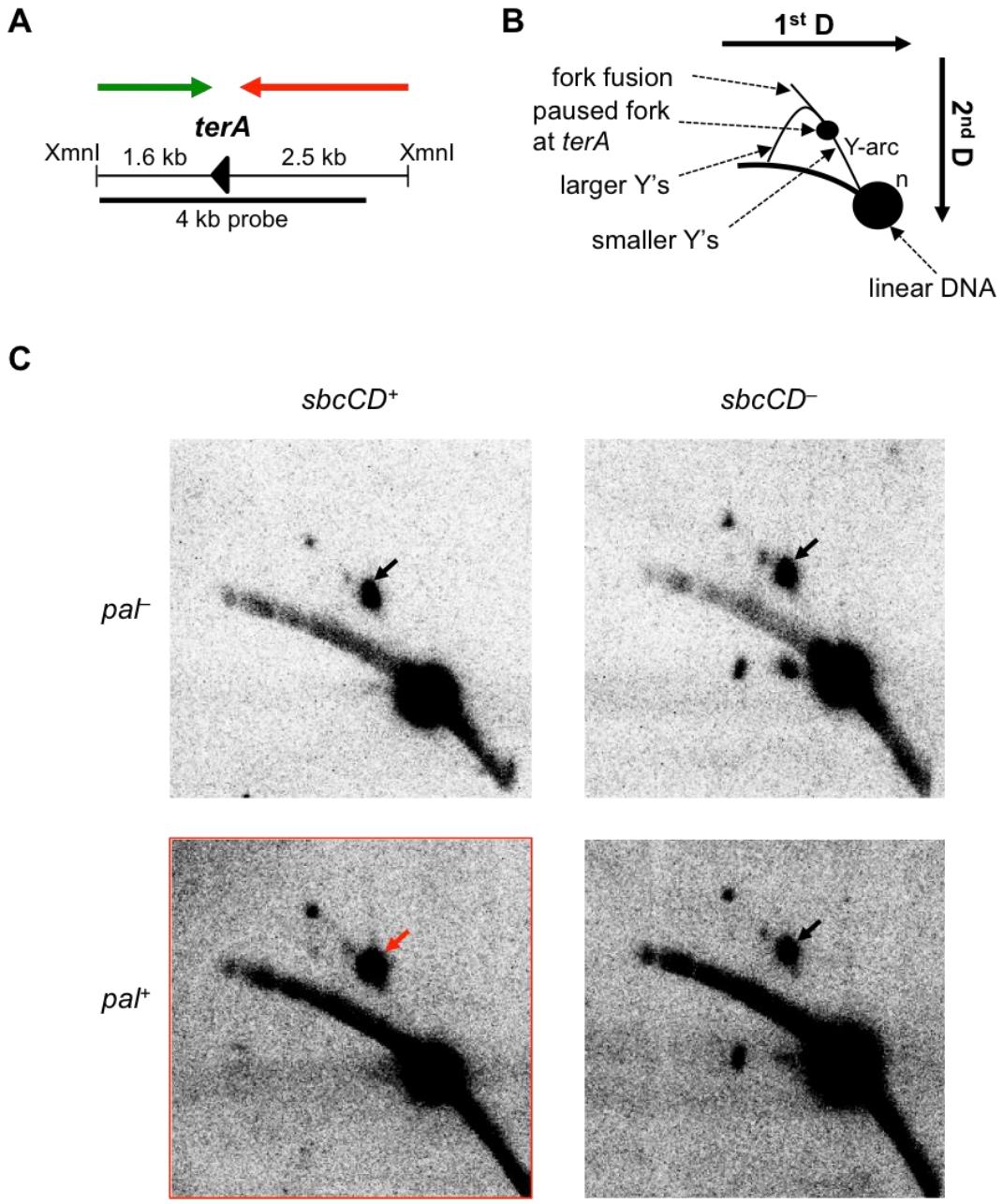


Figure 3.3. Two-dimensional agarose gel electrophoresis of the *terA* fragment.

(A) XmnI digestion map of the region surrounding the *terA* site. The orientation of the *terA* site is indicated by a black triangle. The black thick line is the position of the *terA* probe. The direction of the replication is indicated using red and green arrows. (B) Schematic representation of the migration patterns of branched DNA molecules when separated on a 2D agarose gel electrophoresis and hybridised with the *terA* probe. (C) 2D agarose gel electrophoresis of the *terA* fragment. Paused replication forks are marked by arrows. Strains used were DL1777 (*pal*$^{-}$ *sbcCD*$^{+}$), DL2151 (*pal*$^{-}$ *sbcCD*$^{-}$), DL2859 (*pal*$^{+}$ *sbcCD*$^{+}$) and DL2874 (*pal*$^{+}$ *sbcCD*$^{-}$).

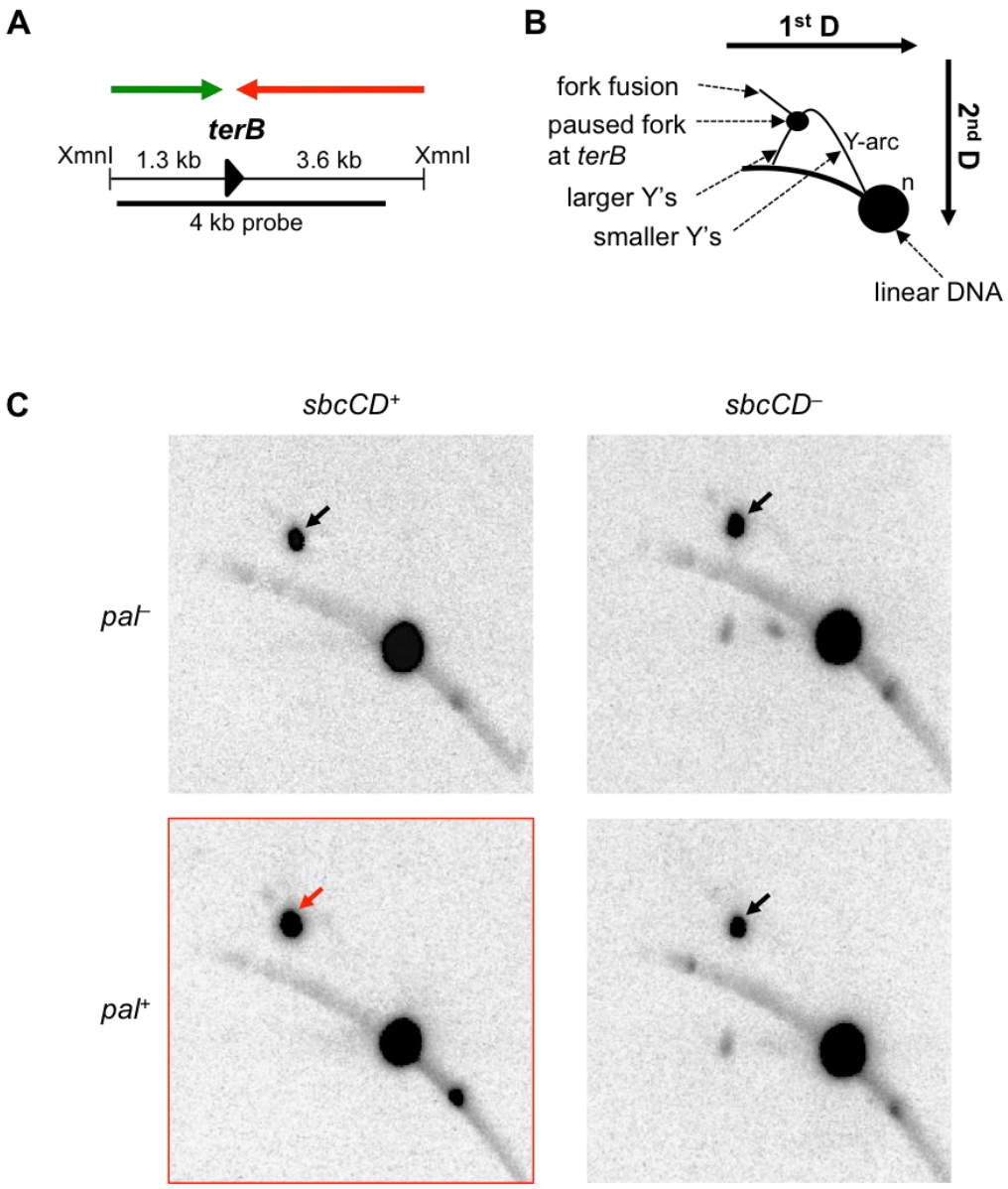


Figure 3.4. Two-dimensional agarose gel electrophoresis of the *terB* fragment.

(A) XmnI digestion map of the region surrounding the *terB* site. The orientation of the *terB* site is indicated by a black triangle. The black thick line is the position of the *terB* probe. The direction of the replication is indicated using red and green arrows. (B) Schematic representation of the migration patterns of branched DNA molecules when separated on a 2D agarose gel electrophoresis and hybridised with the *terB* probe. (C) 2D agarose gel electrophoresis of the *terB* fragment. Paused replication forks are marked by arrows. Strains used were DL1777 (*pal*$^{-}$ *sbcCD*$^{+}$), DL2151 (*pal*$^{-}$ *sbcCD*$^{-}$), DL2859 (*pal*$^{+}$ *sbcCD*$^{+}$) and DL2874 (*pal*$^{+}$ *sbcCD*$^{-}$).

To determine the amount of replication forks paused at each *ter* site, the detected signals were quantified using *ImageQuant* software. For this purpose the average intensity of the background signal was substracted from the average intensity of paused DNA and the average intensity of linear DNA. Then, the intensity of the signal from paused DNA was divided by the intensity of the signal from linear DNA plus the intensity of the signal from paused DNA. Individual quantifications for all *ter* fragments and all strains are showed in Figure 3.5.

Finally, the difference in the amount of paused replication forks after a DSB was calculated on the ratio of intensity of signal in DSB^+/DSB^- strains (DL2859/DL1777). The average of three independent biological experiments was plotted (Figure 3.6). This quantification showed that more replication forks stop at each *ter* site when the strain is subjected to a DSB in *lacZ*. Strikingly, 4 times and 8 times more paused replication forks were detected at *terB* and *terA* sites, respectively. This result confirms that the over-replication detected earlier by WGS occurs principally between *terA* and *terB* sites.

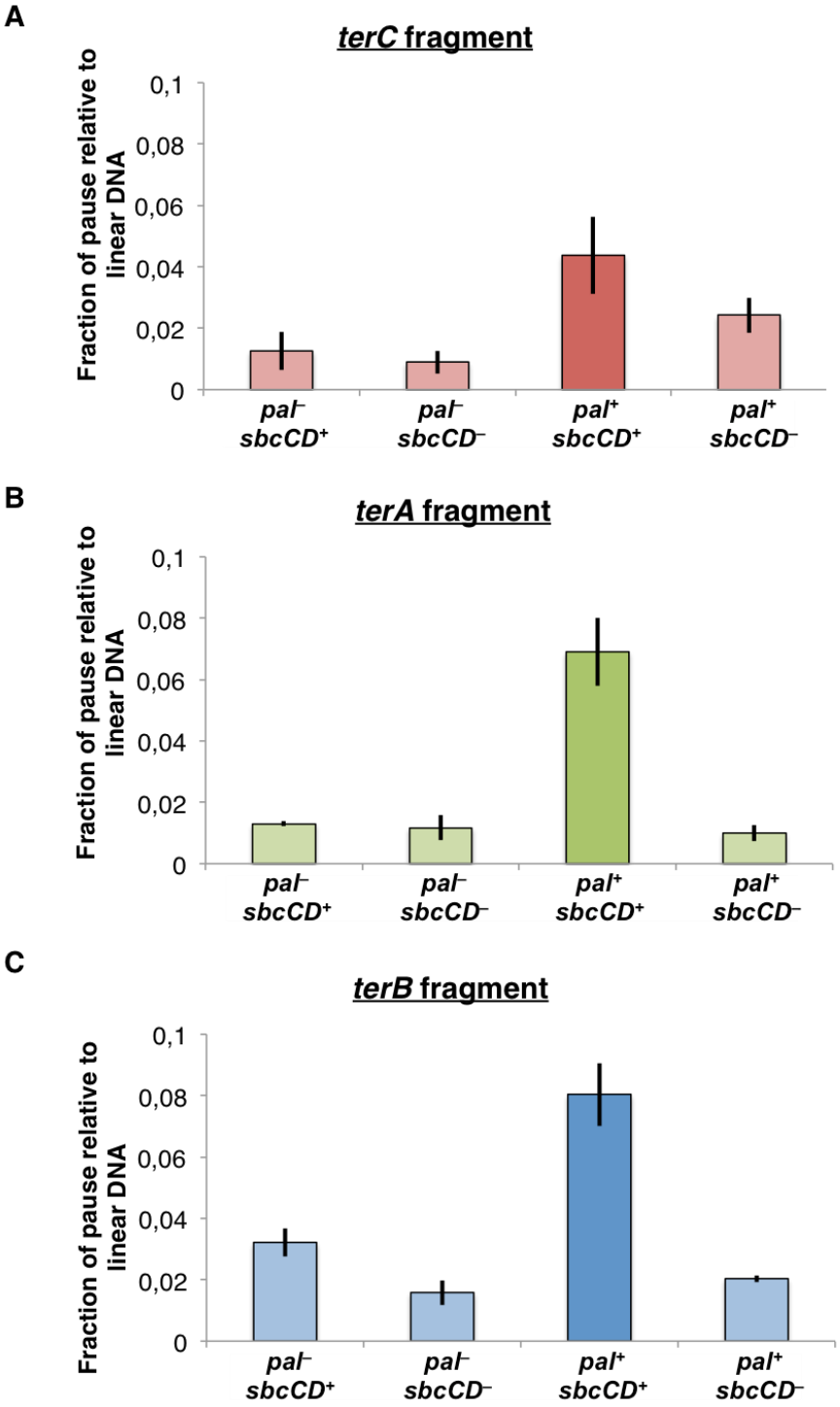


Figure 3.5. Larger amount of paused replication forks was observed in *pal*$^+$ *sbcCD*$^+$ strain at all *ter* sites studied.

Quantification of the amount of paused replication forks at each *ter* site studied. (A) For *terC* fragment. (B) For *terA* fragment. (C) For *terB* fragment. Error bars represent the standard error of the mean where n = 3. Strains used were DL1777 (*pal*$^-$ *sbcCD*$^+$), DL2151 (*pal*$^-$ *sbcCD*$^-$), DL2859 (*pal*$^+$ *sbcCD*$^+$) and DL2874 (*pal*$^+$ *sbcCD*$^-$).

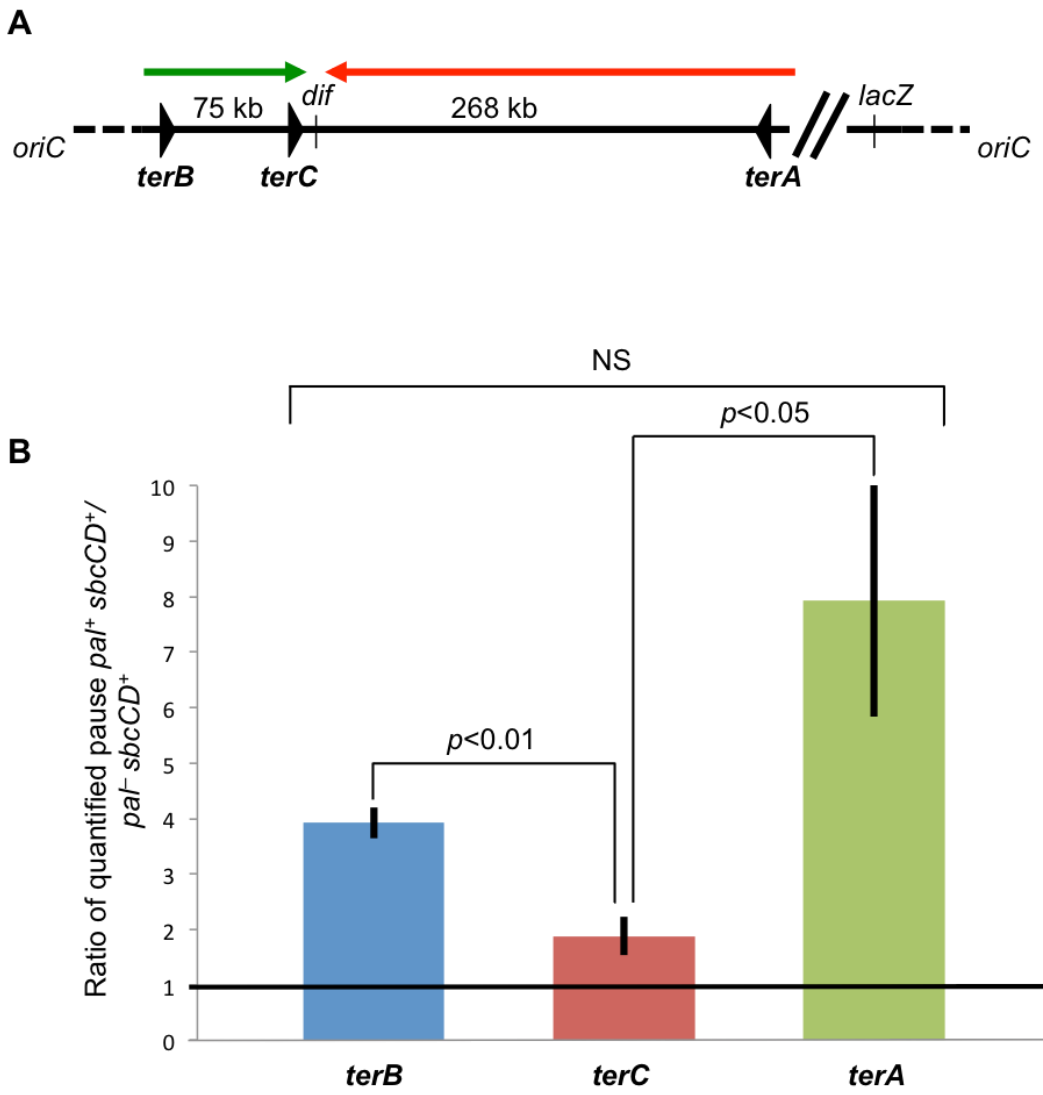


Figure 3.6. **The ratio of quantified amount of paused replication forks at *terA*, *terB* and *terC* sites in DSB$^+$ over DSB$^-$ strains.**

(A) Schematic representation of the positions of the *ter* sites on the *E. coli* chromosome. The orientation of the *ter* sites is indicated by black triangles. The position of the *dif* site is indicated. The direction of the replication is indicated by the red and green arrows. (B) The DSB$^+$/DSB$^-$ ratio of quantified amount of paused replication forks on the Y-arcs is normalised against the total linear DNA. Error bars represent the standard error of the mean where n = 3. *p* values represent unpaired *t* test where NS = not significant. Strains used were DL1777 (*pal*$^-$ *sbcCD*$^+$) and DL2859 (*pal*$^+$ *sbcCD*$^+$).

3.3 Removal of Tus/*ter* blocks did not affect the over-replication in the terminus

To test whether the presence of a DSB shifts the termination of replication from the *dif* site to the *ter* site closer to a DSB, I used strains containing a 246 bp palindrome introduced in different loci: *lacZ* and *ascB*. The *ascB* gene is located on the left replichore at the same distance from the origin as the *lacZ* gene is on the right replichore (Figure 3.7 A). This configuration should allow these genes to be replicated at the approximately the same time, resulting in a similar delay of the replication fork at the DSB site. So, if the presence of a DSB on one of the replichores shifted the replication termination either towards *terA*, when the DSB is at *lacZ*, or towards *terB*, when the DSB is at *ascB*, this change would be comparable in both strains. To test the presence of this potential shift, the strains were made *tus*$^-$ by P1 transduction. Binding of the Tus protein to a *ter* sequence is essential for pausing replication forks at the *ter* site. So, when the Tus/*ter* blocks are removed, replication forks can pass through *ter* sequences. If one of the replication forks is delayed and the Tus protein is removed, it allows the termination of replication to occur in other places than the terminus region. So, this shift becomes a good tool to test if the observed peak at *dif* depends on the termination of replication. If it does, then the peak would either disappear or be shifted towards the DSB site. The strains that were used for this WGS experiment were DL5398 (*pal*$^-$ *tus*$^-$), DL5399 (*lacZ::pal tus*$^-$) and DL5499 (*ascB::pal tus*$^-$), which were derived from MG1655 with *sbcCD* under its native promoter. As a control DL1777 (*pal*$^-$) was used (initially sequenced by Martin White).

For the WGS experiment, overnight cultures were grown at 37°C in LB media. In the morning, these overnight cultures were diluted in order to give an OD_{600nm} of 0.01 and left to grow until they reach an OD_{600nm} of 0.2. Then,

cultures were re-diluted to an OD_{600nm} of 0.01 and left to re-grow until they reach an OD_{600nm} of 0.2-0.3. The second dilution step increases the homogeneity of cultures by providing more time for cells to exit stationary phase and enter exponential phase. Cultures were harvested by centrifugation at maximum speed and DNA was isolated using a Wizard® kit as described in Chapter 2. The phenol-chloroform method was used to purify the isolated DNA and quality controls were performed. Edinburgh Genomics performed library preparations and whole genome sequencing using Illumina HiSeq 2500 platform. The obtained raw paired-end sequencing data were analysed using methods described in Chapter 2. Normally, the log_2 replication profile of a wild-type (DL1777) strain that grows in LB is V-shaped with increasingly more DNA at the origin site than at the terminus site (Figure 3.7 B). As expected, if the sites of termination are distributed over a wide region when the *ter* sites are inactive, *tus*$^-$ strains log_2 replication profiles were U-shaped (Figure 3.7 B, DL5398, DL5399 and DL5499). Furthermore, the *tus*$^-$ strains had lost more DNA in the regions of the 7 *rrn* operons than in the *tus*$^+$ strain. This is likely to be an artifact arising from DNA extraction as proteins bound to *rrn* operons might have been discarded together with the DNA following protein precipitation (Figure 3.7 B). The DNA loss at the *ascB* gene for DL5499 and at the *lacZ* gene for DL5399 indicates that DSBs were formed at the palindrome sites. As mentioned above, the log_2 replication profiles of the terminus regions were flattened due to the removal of the Tus/*ter* block. If this flattening were simply caused by a wider distribution of termination sites, a shift of the termination of replication cased by the palindrome was expected to be observed. Instead, the strains that were subjected to a DSB in different arms of the chromosome showed similar patterns and an excess of DNA in the terminus region compared to *pal*$^-$ *tus*$^-$ (Figure 3.7 B, DL5399 and DL5499).

To test with greater sensitivity, if the replication termination was shifted towards the DSB site, the ratios of *pal*$^{+}$ *tus*$^{-}$ to *pal*$^{-}$ *tus*$^{-}$ were determined (Figure 3.8). No shift of the replication termination was observed in either of the strains and the over-replication was maximal in the region of the *dif* site (Figure 3.8).

These results suggest that a DSB in either *ascB* or *lacZ* creates an excess of DNA in the terminus region, as was observed previoiusly for *lacZ tus*$^{+}$, and that this over-replication is independent of where the two replication forks are expected to collide given the observed delay to replication caused by the palindrome. Moreover, this over-replication is centred on the *dif* site.

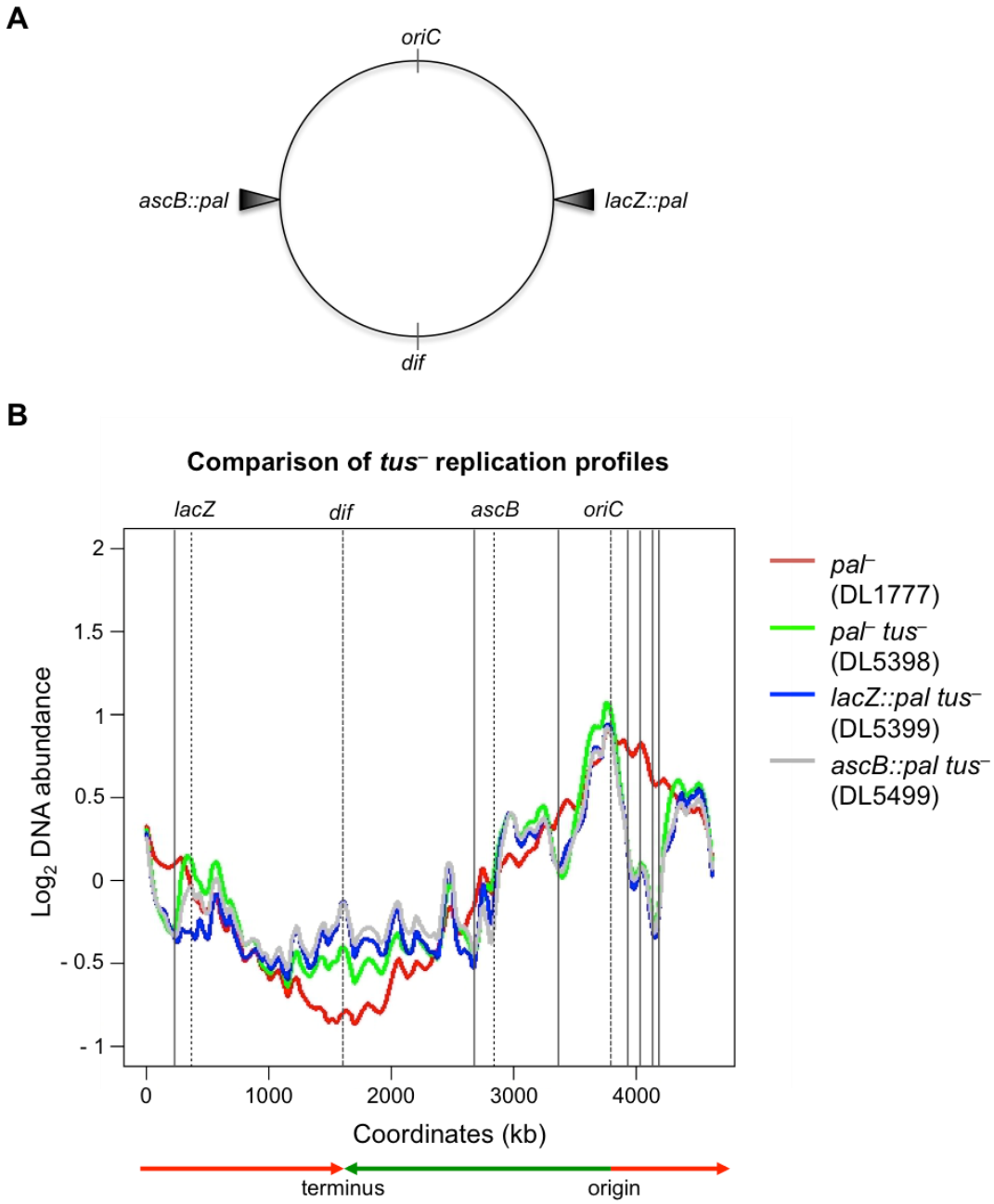


Figure 3.7. Whole genome sequencing profiles of various *tus*⁻ mutants.

(A) Map of the *E. coli* chromosome. *oriC* and *dif* are indicated. The locations of palindromes are shown by black triangles. (B) Replication profiles of exponentially growing cultures of *tus*⁻ strains with or without a 246 bp palindrome and with *sbcCD* under its native promoter. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of palindromes, *dif* site and *oriC* are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Strains used were DL1777 (*pal*⁻) analysed by Martin White in red, DL5398 (*pal*⁻ *tus*⁻) in green, DL5399 (*lacZ::pal tus*⁻) in blue and DL5499 (*ascB::pal tus*⁻) in grey.

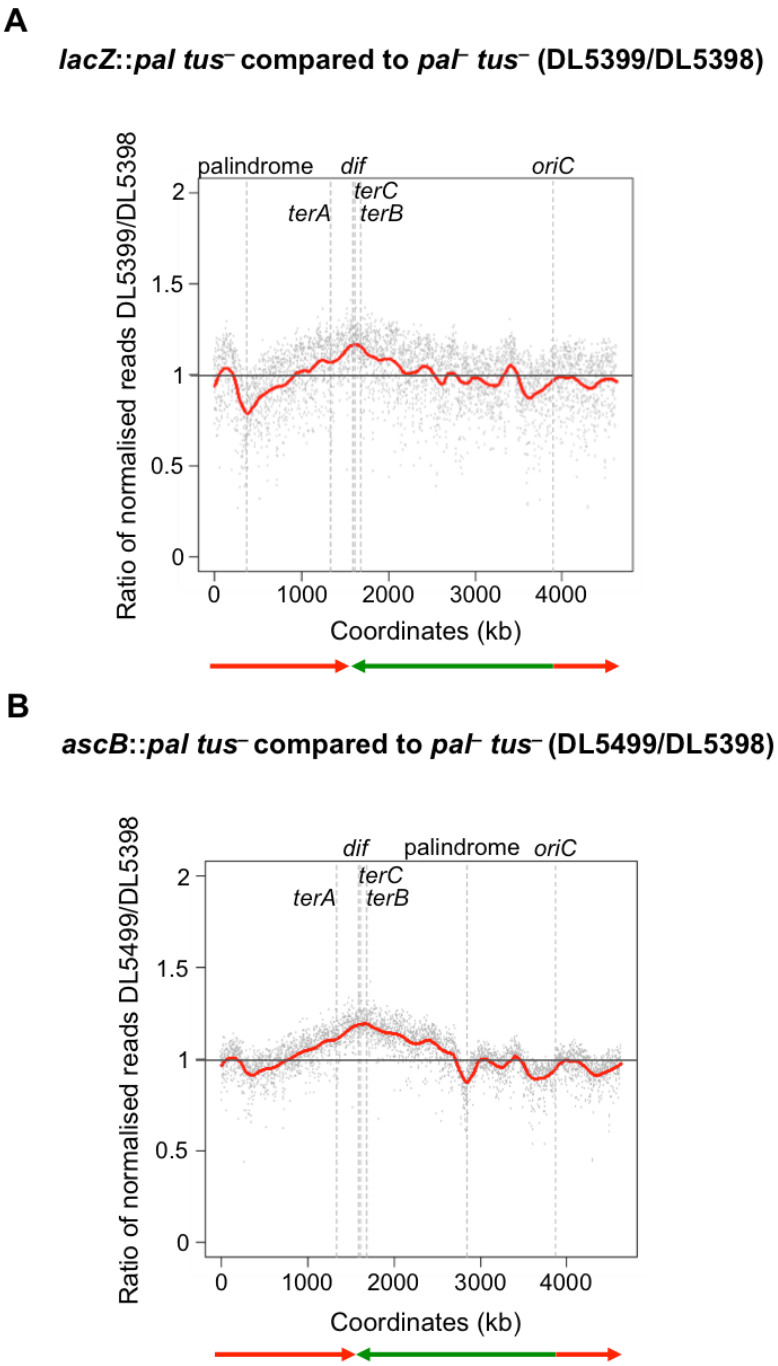


Figure 3.8. Ratios of replication profiles of *pal*⁺ *tus*⁻ to *pal*⁻ *tus*⁻ strains.

The ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of *lacZ::pal tus*⁻ to *pal*⁻ *tus*⁻ (A) and *ascB::pal tus*⁻ to *pal*⁻ *tus*⁻ (B) strains. Grey dots represent reads, red lines are fitted lines of the reads from DL5399 (A) or DL5499 (B). The direction of replication is shown using green and red arrows. The positions of *terA*, *terB*, *terC*, *dif*, palindromes and *oriC* are indicated. Strains used were DL5398 (*pal*⁻ *tus*⁻) showed as horizontal grey line, DL5399 (*lacZ::pal tus*⁻) and DL5499 (*ascB::pal tus*⁻).

3.4 Discussion

This study followed the work of Dr. Martin White in our laboratory who introduced and optimised the *E. coli* whole genome sequencing technique to study DSBR in *lacZ*. WGS data analysis showed an excess of DNA in the terminus region between *terA* and *terB* sites in the strain subjected to DSB in *lacZ*. Additionally, DSBR on the right replichore of the chromosome caused a delay of the replication fork on this replichore that, in the absence of any replication initiated in the terminus region, predicted a move of replication termination from the *terC*/*dif* region to the *terA* site. To quantify the amount of replication forks paused at *ter* sites and check if this excess of DNA in the terminus occured between *terA* and *terB*, 2D agarose gel electrophoresis combined with Southern blotting was performed. The quantified amount of paused replication forks at *terA, terB* and *terC* sites in the strain subjected to a DSB in *lacZ* was compared to the strain without a DSB. Larger amounts of replication forks stopped at *terA, terB* and *terC* were detected, inconsistent with a simple shift in position of termination caused by the palindrome. Instead the data were consistent with the presence of an excess of DNA between these termination sites in the strain subjected to a DSB in *lacZ*, due to more replication forks originated in this region. These forks were primarily blocked at the *terA* and *terB* sites, which showed 5.3- and 2.5-fold increase in the amount of replication forks respectively, as detected by 2D gel electrophoresis.

To determine whether replication termination in the absence of replication termination blocks was shifted towards a DSB site, I constructed *tus*$^-$ mutants with palindromes in different loci on the chromosome. The absence of Tus/*ter* blocks allowed the replication forks to pass through *ter* sequences and to detect the amended place of un-blocked replication termination in strains

subjected to DSBs. No shift in replication termination was detected. Instead, an excess of DNA centered on the *dif* region was observed. This *dif*-centered DNA excess was not affected by different loci of palindromes. Also, the $\log_2$ replication profile patterns in the terminus in the strains that were undergoing a DSBR in different loci in the chromosome were similar. This suggests that there is a source of additional bi-directional replication at *dif* that is activated by a DSB on either of the replichores.

Overall, the results suggest that when a DSB is induced elsewhere in the chromosome, over-replication is observed in the chromosomal terminus that originates in the *dif* region. Moreover, this over-replication is not dependent on the position where the primary replication forks are expected to collide in the terminus region. The origin of this terminus-replication is unknown and has to be investigated further.

Chapter IV

4. Choice of genetic background and appropriate growth conditions for chromatin immunoprecipitation to detect RecA enrichment in the terminus region

4.1 Introduction

Chromatin immunoprecipitation (ChIP) is a technique that is used to study DNA-protein interactions *in vivo*. In Prof. David Leach laboratory, we have been using ChIP in combination with next generation sequencing (NGS) or real-time quantitative PCR (qPCR) to visualise and/or quantify interactions between DNA and RecA protein. RecA is an essential protein for DNA double-strand break repair (DSBR) and single-strand gap repair by homologous recombination (HR).

Initially, ChIP was mainly used to study eukaryotic rather than prokaryotic organisms. In Prof. David Leach laboratory, Dr. Charlotte Cockram optimised the technique in order to determine the regions of DNA that are bound by RecA protein in the presence and absence of a DSB in *lacZ* gene in *E. coli* (Cockram *et al*. 2015). Briefly, to perform ChIP, cultures in exponential phase of the desired strains were chemically fixed using formaldehyde, which crosslinks all protein-DNA interactions. Then, the DNA was sheared into uniform fragments and a specific antibody directed against RecA was used to immunoprecipitate RecA-DNA complexes. Finally, the crosslinking was reversed and the DNA was purified. Afterwards, the regions of RecA-bound DNA were detected by next generation sequencing (ChIP-seq) and the levels of RecA enrichment were quantified by RT-qPCR (ChIP-qPCR). Both, Cockram and Azeroglu, using various genetic backgrounds and growth conditions have observed an appearance of additional DSB in the terminus region of the chromosome in the strain that was subjected to a DSB in *lacZ*.

However, in contradiction to Dr. Charlotte Cockram's results, Dr. Benura Azeroglu have observed an appearance of additional DSB in the terminus in the strain that was not subjected to a DSB in *lacZ*.

In this chapter, following an observation by Dr. Charlotte Cockram and Dr. Benura Azeroglu described in detail in Section 1.6, I provide an explanation for the discrepancies between Dr. Cockram's and Dr. Azeroglu's data and present the justification for the choice of the appropriate background strains and growth conditions to study RecA-DNA interactions in the terminus region of the *E. coli* chromosome.

4.2 Quantification of RecA enrichment levels in the terminus region by quantitative real-time PCR (qPCR)

4.2.1 Workflow of quantitative real-time PCR

Unfortunately, ChIP-seq assays cannot be used to quantitatively compare the amount of RecA enrichment between different strains. However, quantitative real-time PCR (qPCR) can be used to amplify and simultaneously quantify a site-specific target DNA. Therefore, ChIP-qPCR was used for the quantification of RecA enrichment in order to optimise the growth conditions for the strains for further ChIP-seq experiments. As a result, this method provides accurate and quantitative comparisons between samples.

During a qPCR reaction, SYBR green fluorescent dye is used to detect newly synthesised dsDNA (Pfaffl 2001). The dye binds to the minor groove of dsDNA, while it is generated by PCR, and provides a real-time signal of the dsDNA formation. The Ct values of the reaction are used to quantify the obtained product. Ct value is the cycle threshold value and it is the readout of the dsDNA product formation. qPCR consists of 40 cycles, which allows the fluorescent signal to increase above the background fluorescence. Therefore,

the Ct value is the first cycle at which the signal of fluorescence is higher than the background signal. The Ct values for quantification are taken where the threshold line crosses the amplification curve. The lower is the Ct value, the larger is the amount of target DNA in a given sample. Ct values above 35 represent non-specific binding, e.g. primer dimer formation.

qPCR amplifies a defined target DNA. Therefore, ChIP-qPCR requires prior knowledge of where a protein of interest binds (Hoffman and Jones 2009). As a consequence, the choice of the oligonucleotides is critical for successful qPCR reactions. Oligonucleotides used in this study were 20 bp in size and were designed to generate a product of 180 bp. The position of oligonucleotides was chosen in order to correspond to the ChIP-seq profile obtained by Charlotte Cockram (Figure 1.13). A melting curve is used to assess whether the reaction produced a single, specific product. It was run for each pair of oligonucleotides in each qPCR reaction. Each primer was designed to have a melting temperature of 55-60°C. If a single, specific product is formed during the reaction, a melting curve profile for a single pair of primers will result in a distinct single peak.

The efficiency of each qPCR reaction was checked by a standard curve. Serial dilutions of known DNA concentration were assayed with each pair of primers. The efficiency for each qPCR reaction was between 90-100% for the Agilent qPCR machine or 1.8 - 2 (out of 2) for the Roche LightCycler.

Relative quantification of ChIP-enriched DNA was performed. Relative quantification measures the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions (Cockram *et al*. 2015).

4.2.2 Quantification of the RecA enrichment level in the terminus in strains with constitutive DSBR in the *lacZ* gene

In order to determine the difference between the levels of RecA binding in the terminus of strains DL4899 and DL4900, DSB^- and DSB^+, respectively, the relative levels of RecA enrichment were quantified using qPCR (Figure 4.1). DL4899 and DL4900 are recombination-proficient strains derived from MG1655. DL4900 contains a 246 bp interrupted palindrome in the *lacZ* gene and the *sbcCD* operon under the control of its native promoter. This allows SbcCD to be constitutively expressed and to cleave hairpin-like structures, formed by the palindrome, once by replication cycle. As a control DL4899 (pal^- $sbcCD^+$), not subjected to periodic DSB, was used. Both strains contain artificially introduced three Chi sites (Chi array) positioned 3 kb either side of the palindrome locus.

Overnight cultures were grown at 37°C in M9 minimal medium supplemented with 0.5% glucose, 0.2% casamino acids, 5 μM $CaCl_2$ and 1 mM $MgSO_4$. In the morning, overnight cultures were diluted in order to give an OD_{600nm} of 0.01 and left to grow until they reached an OD_{600nm} of 0.2. Then, cultures were re-diluted to an OD_{600nm} of 0.01 and left to re-grow. The second dilution step increases the homogeneity of cultures by providing more time for cells to exit stationary phase and enter exponential phase. Cultures were harvested when they reached an OD_{600nm} of 0.2-0.3. Protein-DNA interactions in these cells were immediately crosslinked by formaldehyde and DNA was isolated as described in Chapter 2. The isolated DNA was analysed by qPCR.

For qPCR reactions, regions in *cynT* and *lacZ* genes were chosen as controls. The cynT primers amplify a 180 bp region after the Chi sites, 6.7 kb away from the DSB locus in *lacZ* on the origin-proximal side. The lacZ2 primers amplify a 180 bp region before the Chi sites, 2.1 kb away from the DSB locus in *lacZ* on

the origin-proximal side. As expected from previous studies in our laboratory, in the strain that does not contain the palindrome, we did not observe significant RecA binding, neither before nor after the Chi sites. Also, as expected, RecA enrichment in the strain subjected to DSBs was ~14-fold in the cynT region and ~1.5-fold in the lacZ2 region (Figure 4.1). This was consistent with previously observed data by Charlotte Cockram. To quantify the RecA enrichment in the terminus region, four pairs of primers were chosen based on the ChIP-seq data from (Cockram *et al*. 2015). Dif2, amplifies a 180 bp region after 4 endogenous Chi sites, 30.3 kb away from *dif* on the right replichore of the chromosome; bef.chi3, amplifies a 180 bp region before the Chi sites, 9.4 kb away from *dif* on the right replichore of the chromosome; bef.chi1, amplifies a 180 bp region before the Chi sites, 21.7 kb away from *dif* on the left replichore of the chromosome; and Dif1, amplifies a 180 bp region after 3 endogenous Chi 'array-like' sites, 27.3 kb away from *dif* on the left replichore of the chromosome. In the strain that was not subjected to DSBs (DL4899), as expected, we did not observe a significant RecA enrichment in the terminus. In the strain subjected to DSBs (DL4900), the relative RecA enrichment in Dif1 and Dif2 was ~4-fold compared to bef.chi1 and bef.chi3, and ~2-fold compared to the RecA enrichment in DL4899 (Figure 4.1). These data showed that ChIP-qPCR is a good and fast method to quantify and compare the RecA enrichment in the terminus between different strains.

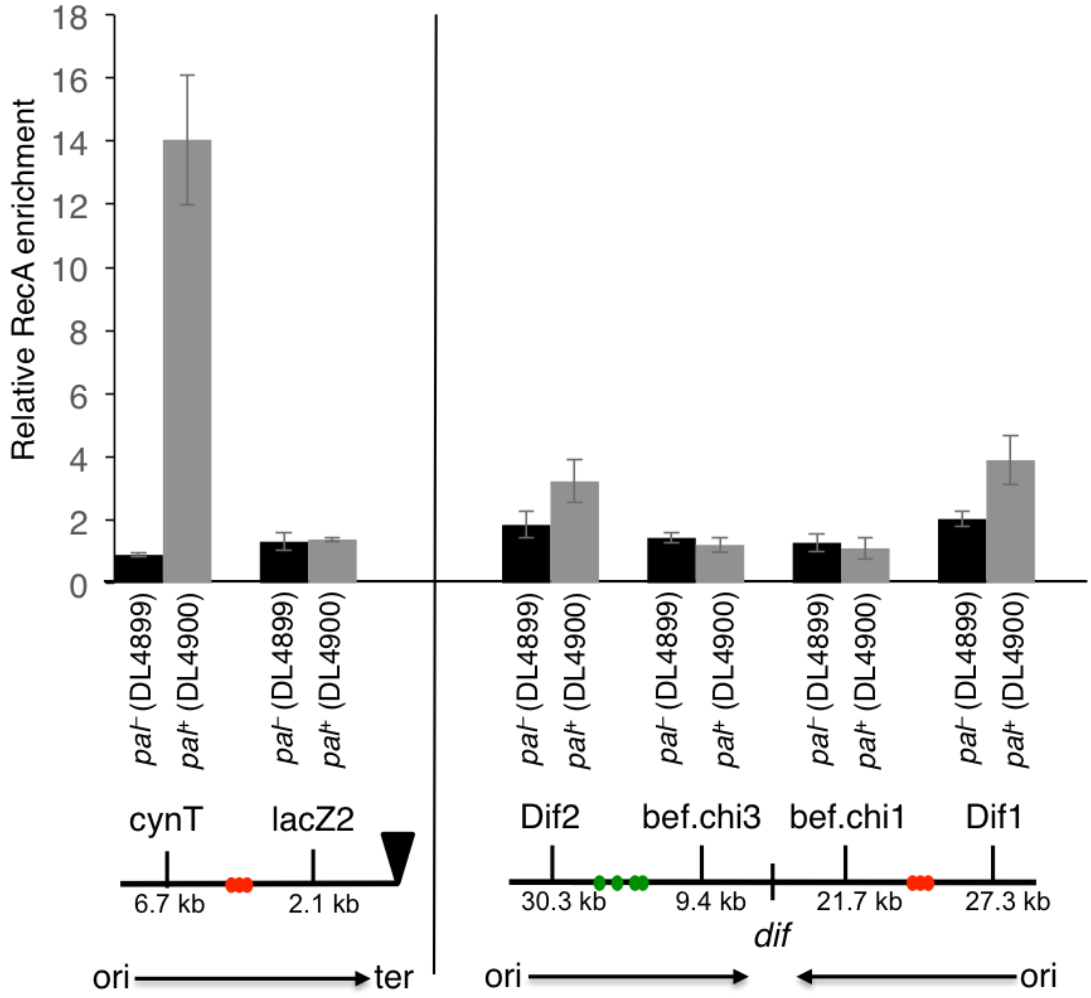


Figure 4.1. ChIP in combination with real-time quantitative PCR confirms the presence of a DSB event in the terminus.

ChIP-qPCR was used to quantify the levels of RecA enrichment in the terminus region. RecA enrichment levels in the *lacZ* locus were used as controls. Green and red circles represent differently oriented Chi sites (not to scale). Green Chi sites are oriented in a such way that RecBCD will recognise them if it moves left to right on the chromosome, red Chi sites will be recognised by RecBCD that moves in the opposite direction – right to left. A black triangle represents a 246 bp interrupted palindrome. The position of *dif* is marked here. Black arrows show the direction of the replication. Strains used were DL4900 (*pal*$^{+}$ *sbcCD*$^{+}$) and DL4899 (*pal*$^{-}$ *sbcCD*$^{+}$).

4.2.3 Quantification of RecA enrichment level in the terminus in strains subjected to an inducible DSB in the *lacZ* gene

To investigate if the termination of replication, chromosome dimer resolution, and/or chromosome decatenation systems were involved in the formation of a DSB in the terminus, mutants had to be constructed in a system that allowed the formation of a DSB in *lacZ* without palindrome excision. When the dimer resolution mutation (Δ*dif*) was introduced in MG1655 background strain, expressing SbcCD constitutively, the 246 bp interrupted palindrome was lost during exponential growth (Charlotte Cockram, unpublished data). In order to avoid this loss, mutants had to be constructed in a background, permitting SbcCD production over a desired period of time. For this purpose, strains derived from BW27784 that had the *sbcCD* operon under an arabinose-inducible promoter (P_{BAD}) were used. This allowed us to control the time of the DSB formation. The strains, used in this study, were DL2006 (*pal*$^{+}$ P_{BAD}-*sbcCD*) and DL2573 (*pal*$^{-}$ P_{BAD}-*sbcCD*) with the endogenous Chi sites (~5 kb from either side of the palindrome); DL4184 (*pal*$^{+}$ P_{BAD}-*sbcCD*) and DL4201 (*pal*$^{-}$ P_{BAD}-*sbcCD*) with Chi arrays introduced at 1.5 kb from either side of the palindrome; DL3683 (*pal*$^{-}$ P_{BAD}-*sbcCD*) and DL3684 (*pal*$^{+}$ P_{BAD}-*sbcCD*) with Chi arrays introduced at 3 kb from either side of the palindrome.

Initially, the growth conditions used for ChIP-qPCR experiments were the same as for the strains that express SbcCD constitutively: M9 minimal medium supplemented with 0.5% glucose, 0.2% casamino acids, 5 μM $CaCl_2$ and 1 mM $MgSO_4$. Time course was performed for the strains with endogenous Chi sites in *lacZ* (DL2006 and DL2573) to determine and optimise the exact time where the difference between strains in levels of RecA enrichment in the terminus region was significant. The overnight cultures were diluted to an OD_{600nm} of 0.01 and left to grow to an OD_{600nm} of 0.2. Then, the cultures were diluted to an

OD_{600nm} of 0.01, grown for 20 minutes and 0.2% arabinose was added (SbcCD expression induced). The cultures were maintained between an OD_{600nm} of 0.1-0.3. Time points were taken every 30 minutes for 150 minutes. 100 ml of culture were taken and immediately crosslinked with formaldehyde. DNA was isolated and analysed by qPCR. cynT fragment was used as a control. As expected, the level of RecA enrichment increased steadily and ~22-fold RecA enrichment was detected at 120 minutes after arabinose-induction (Figure 4.2 A). The cycling pattern was observed with low RecA enrichment level at 150 minutes. Surprisingly, the quantified levels of RecA enrichment in Dif1 and Dif2 fragments in the terminus region were at the lower level than expected with 1.5-2-fold RecA enrichment, suggesting that it is difficult to observe RecA enrichment in the terminus in M9 minimal medium (Figure 4.2 B). No difference in RecA enrichment levels between strains was detected. Experiment was done once for a control DL2573 (*pal*$^{-}$ P_{BAD}-*sbcCD*) strain.

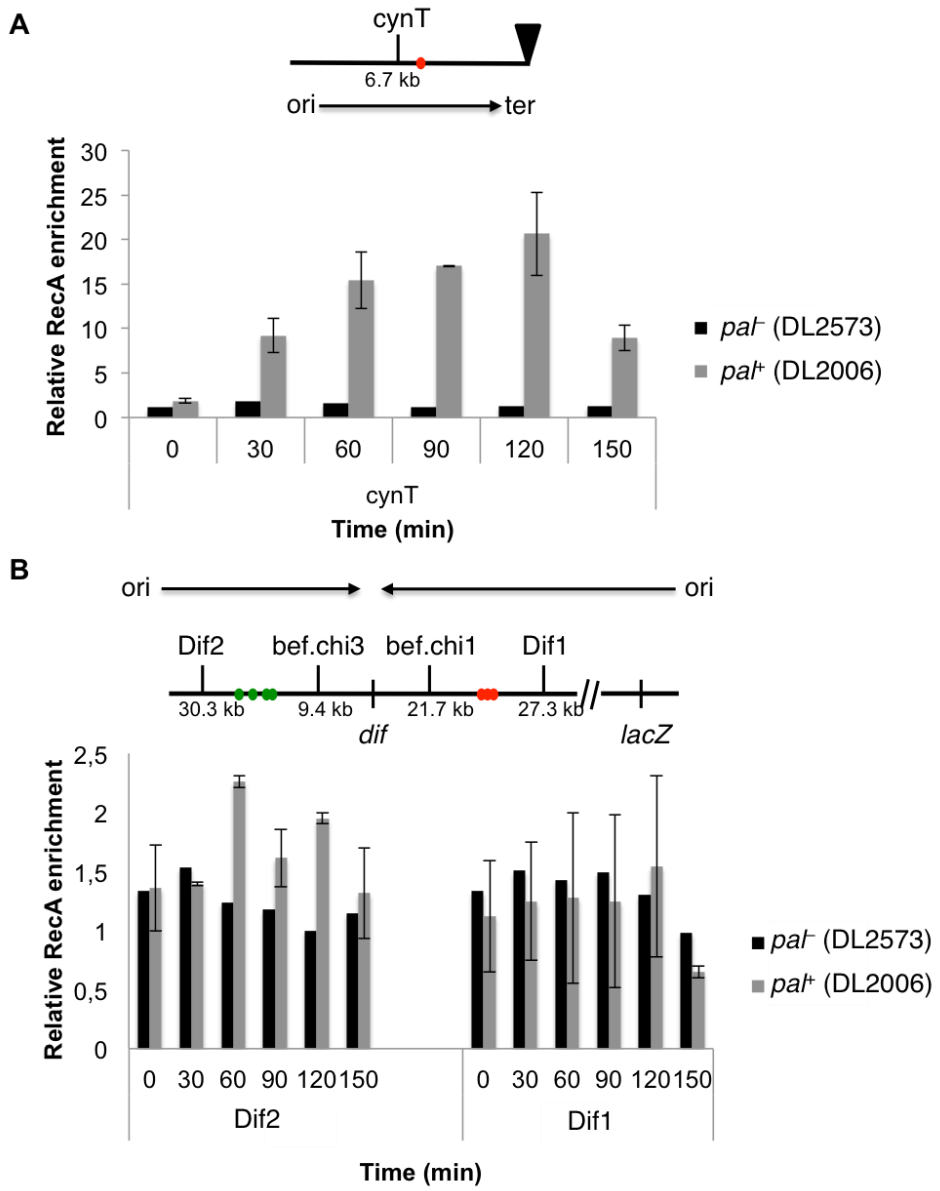


Figure 4.2. Time course for *pal*⁻ P$_{BAD}$-*sbcCD* and *pal*⁺ P$_{BAD}$-*sbcCD* strains with endogenous Chi sites in *lacZ* following arabinose-induced DSBs in M9 minimal medium.

Time course was performed to determine the exact time when the difference between strains in RecA enrichment levels in the terminus region was significant. ChIP-qPCR was used to detect RecA enrichment levels in the terminus region of inducible strains with endogenous Chi sites in *lacZ*. Green and red circles represent properly oriented Chi sites (not to scale). Green Chi sites are oriented in a such way that RecBCD will recognise them if it moves left to right on the chromosome, red Chi sites will be recognised by RecBCD that moves in the opposite direction – right to left. A black triangle represents a 246 bp interrupted palindrome. The position of *dif* is indicated here. Black arrows show the direction of the replication. (A) RecA enrichment levels in the *lacZ* locus were used as controls. (B) RecA enrichment levels in the terminus region. Strains used were DL2006 (*pal*⁺ P$_{BAD}$-*sbcCD*) and DL2573 (*pal*⁻ P$_{BAD}$-*sbcCD*) containing endogenous Chi sites in *lacZ*.

At the same time, the analysis of ChIP-seq data from the strains containing P_{BAD}-*sbcCD* and Chi arrays at 1.5 kb from either side of the palindrome (DL4184 and DL4201) revealed that after 1 hour of DSB induction in *lacZ* both strains showed RecA enrichment in the terminus (Figure 1.14). Therefore, all the strains that previously were grown in M9 minimal medium were grown as described in (Azeroglu *et al*. 2016) to check the levels of RecA enrichment in the terminus by qPCR (Figure 4.3).

Overnight cultures were grown at 37°C in LB medium supplemented with 0.5% glucose. In the morning, these overnight cultures were diluted in prewarmed LB medium supplemented with 0.5% glucose in order to give an OD_{600nm} of 0.01 and left to grow until they reached an OD_{600nm} of 0.2. Then, the cultures were re-diluted to an OD_{600nm} of 0.01 and left to grow for 20 minutes. The second re-dilution step increases the homogeneity of the cultures. After 20 minutes of growth, DSBs were induced by adding 0.2% arabinose and the cultures were left to grow for an hour until they reached an OD_{600nm} of 0.2-0.3. Protein-DNA interactions were immediately crosslinked and DNA was isolated as described in Chapter 2. The isolated DNA was analysed by qPCR.

For the qPCR reaction, the region in the *cynT* gene was chosen as a control. As expected, all strains that were not subjected to DSBs did not show significant RecA binding in *lacZ* (Figure 4.3). RecA enrichment in this region in DL2006 and in DL4184 was ~20-fold and ~14-fold in DL3684. To quantify and compare the RecA enrichment levels in the terminus region between these strains, two pairs of primers were chosen: Dif2 and Dif1. DL2006 and DL2573 showed 3-4-fold RecA enrichment in the terminus and no statistical difference was observed between these strains (Figure 4.3). DL3683 and DL3684 showed a 2-4-fold RecA enrichment and no statistical difference was observed between these strains (Figure 4.3). Surprisingly, only DL4201 and DL4184, strains

containing the 1.5 kb Chi array in *lacZ*, showed a palindrome-dependent difference in RecA binding in the terminus. The difference in RecA enrichment between the two strains was ~1.5-2-fold. Therefore, DL4184 and DL4201 were used as background strains to study the effect of termination of replication, chromosome dimer resolution, and chromosome decatenation on RecA enrichment levels in the terminus.

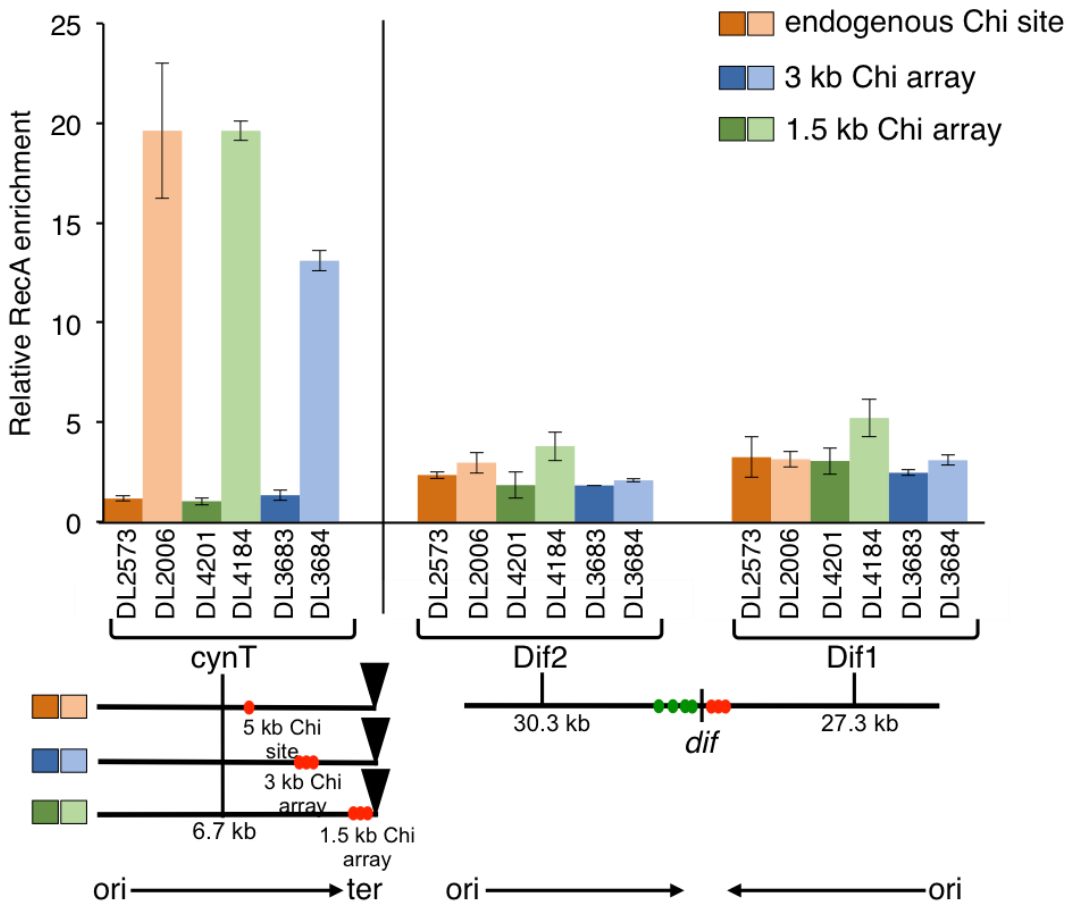


Figure 4.3. Choosing the optimal strain background to detect RecA enrichment in the terminus.

ChIP in combination with qPCR was used to detect and compare the difference in RecA enrichment levels between various inducible strain backgrounds. RecA enrichment levels in the *lacZ* locus were used as controls. Green and red circles represent properly oriented Chi sites (not to scale). Green Chi sites are oriented in a such way that RecBCD will recognise them if it moves left to right on the chromosome, red Chi sites will be recognised by RecBCD that moves in the opposite direction – right to left. A black triangle represents a 246 bp-interrupted palindrome. The position of *dif* is indicated here. Black arrows show the direction of the replication. Strains used were DL2006 (*pal*$^{+}$ P_{BAD}-*sbcCD*) and DL2573 (*pal*$^{-}$ P_{BAD}-*sbcCD*) with endogenous Chi sites in *lacZ*, DL4184 (*pal*$^{+}$ P_{BAD}-*sbcCD*) and DL4201 (*pal*$^{-}$ P_{BAD}-*sbcCD*) with Chi-arrays introduced at 1.5 kb from either side of the palindrome, DL3683 (*pal*$^{-}$ P_{BAD}-*sbcCD*) and DL3684 (*pal*$^{+}$ P_{BAD}-*sbcCD*) with Chi-arrays introduced at 3 kb from either side of the palindrome.

4.3 Discussion

The aim of this thesis is to investigate an additional DSB event in the terminus region of *E. coli.* Depending on the growth conditions and the strains used, this novel event in the terminus region is very difficult to detect even by such a sensitive approach as ChIP-seq. When growing cells in M9 minimal medium, RecA enrichment level for a chronic DSB in *lacZ* was ~10-fold, compared to background level, and RecA enrichment level in the terminus was ~4-fold. In the absence of the 246 bp palindrome in *lacZ*, an additional DSB event in the terminus was not detected in M9 minimal medium. Although, when the strains, where DSBs in *lacZ* are induced for a desired period of time, are grown in LB medium, the DSB event in the terminus is detected in both DSB^- and DSB^+ strains. Therefore, it is very important to choose the right background strain and growth conditions in order to detect this phenomenon in the terminus.

This chapter describes the optimisation of growth conditions for RecA-ChIP to detect the RecA enrichment in the terminus region. As datasets from ChIP-seq cannot be quantified, ChIP-qPCR was used for determining the amount of DNA where RecA was bound in a specific locus on the chromosome. ChIP-qPCR is less sensitive method than ChIP-seq and, therefore, RecA enrichment levels in the terminus were close to the background levels. Although, it was possible to detect 1.5-2-fold difference between the strain that was subjected to DSBs in *lacZ* and between the strain without DSBs in *lacZ*, which suggests that the DSB event in the terminus is partially DSB-dependent.

Also, the chapter describes the discrepancies between our ChIP-seq datasets and the growth medium choice for all experiments in this thesis. Due to the nature of chosen mutants for this thesis, the strains where DSBs in *lacZ* were induced for a desired period of time were more suitable for this study than

strains with chronic DSBs. The choice of medium was also important, because strains with acute DSBs grown using the same conditions as the strains with chronic DSBs (M9 minimal medium) did not show any difference between each other and the RecA enrichment levels were close to the background level. Although, these strains were used by Azeroglu *et al.* 2016 for the ChIP-seq experiments in LB medium, where both strains showed RecA enrichment in the terminus region. This might suggest that the phenomenon in the terminus is DSB-independent. However, qPCR quantifications that show 1.5-2-fold difference between DL4184 (*pal*$^{+}$ P$_{BAD}$-*sbcCD*) and DL4201 (*pal*$^{-}$ P$_{BAD}$-*sbcCD*) with Chi arrays introduced at 1.5 kb from either side of the palindrome grown in LB medium suggest that the phenomenon in the terminus is DSB-dependent. Finally, ChIP-qPCR quantifications led to the choice of DL4184 and DL4201, grown in LB medium, for the subsequent experiments, as only these strains showed a difference between each other in RecA enrichment levels in the terminus region.

Chapter V

5. The frequency of DSBs in the terminus region is increased in the absence of chromosome dimer resolution

5.1 Introduction

The *E. coli* chromosome is circular, hence recombination events happening during the cell cycle result in the formation of chromosome dimers. If these dimers are not resolved, sister chromosomes cannot segregate properly before cell division, leading to cell death (Pérals *et al.* 2000; Steiner and Kuempel 1998). During chromosome dimer resolution, the DNA translocase FtsK is loaded onto the chromosome on specific 8 bp sequences named KOPS (FtsK Orientating Polar Sequences) (Bigot *et al*. 2006). These sequences direct the FtsK motor towards the terminus allowing it to translocate DNA to ensure that specific 28 bp sequences required for chromosome dimer resolution, *dif* sites, are aligned together (Bigot *et al*. 2005). Then, FtsK attracts, by direct interaction, the site-specific recombinase XerD to the *dif* site and stimulates XerCD-recombination (Grainge, Lesterlin, and Sherratt 2011; Blakely, Davidson, and Sherratt 1997). Following the resolution of chromosome dimers, the sister chromosomes segregate into two daughter cells.

The aim of this study was to investigate the cause of DSBs in the terminus region of the *E. coli* chromosome. I hypothesised that failure of the XerCD/*dif* system might be responsible for this phenomenon. In this chapter I investigate whether the appearance of DSBs in the terminus region could be the result of incomplete chromosome dimer resolution: either misfunctioning of the XerCD/*dif* system that leads to incomplete dimer resolution or XerCD/*dif* failure to recognise dimers and perform the recombination reaction.

5.2 Chromosome dimer resolution mutants display a growth defect

To test whether the chromosome dimer resolution system is involved in the appearance of DSBs in the terminus region, I constructed *dif*, *xerC*, *xerD* and double *xerCD* mutants using the PMGR method. These mutations were introduced in cells that express *sbcCD* under an arabinose inducible promoter and contain Chi arrays at 1.5 kb on either side of the locus in the *lacZ* gene containing or not a 246 bp interrupted palindrome. Previously, it has been observed that the viability of *xerCD*/*dif* mutants is reduced by 10-15% (Pérals *et al*. 2000; Cornet *et al*. 1996). The viability of these mutants was tested in the presence and absence of a DSB in *lacZ* by growth curves and spot tests. The presence of DSBs in *lacZ* increases the frequency of HR, hence the amount of dimers formed in the terminus, which might lead to a larger growth defect in the strains that are subjected to a DSB in *lacZ*.

In the morning, overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 37°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and 0.2% arabinose was added to induce DSBs in *lacZ* (time 0). The OD_{600nm} of these exponentially growing cultures was measured every 30 min for 4.5 hours. The OD_{600nm} was regularly monitored and the cultures were diluted in fresh prewarmed medium in order to keep the cells in exponential phase (OD_{600nm} of 0.2-0.3). The growth profiles of *xerCD*/*dif* strains containing or not a 246 bp palindrome are shown in Figures 5.1 and 5.2. Mutants that did not contain a palindrome showed similar growth profiles with a slower growth than wild type strains. Mutants carrying the palindrome in *lacZ* showed an even slower growth rate. This suggests that the presence of DSBs in *lacZ* slows down the

growth of *xerCD*/*dif* mutants, possibly by introducing more chromosome dimers that could not be resolved.

To check the viability of *xerCD*/*dif* mutants, I performed spot tests. At the end of the growth curve, samples were diluted to an OD_{600nm} of 0.1. Then, 10-fold serial dilutions were spotted on LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Plates were incubated at 37°C overnight (Figure 5.3). These spot tests showed no reduction in viability on glucose-supplemented plates compared to wild type strains. However, mutants that were subjected to DSBs overnight (on the arabinose plates) grew as smaller colonies. These experiments suggest that DSBs in *lacZ* decrease cell growth and viability in absence of the XerCD/*dif* system.

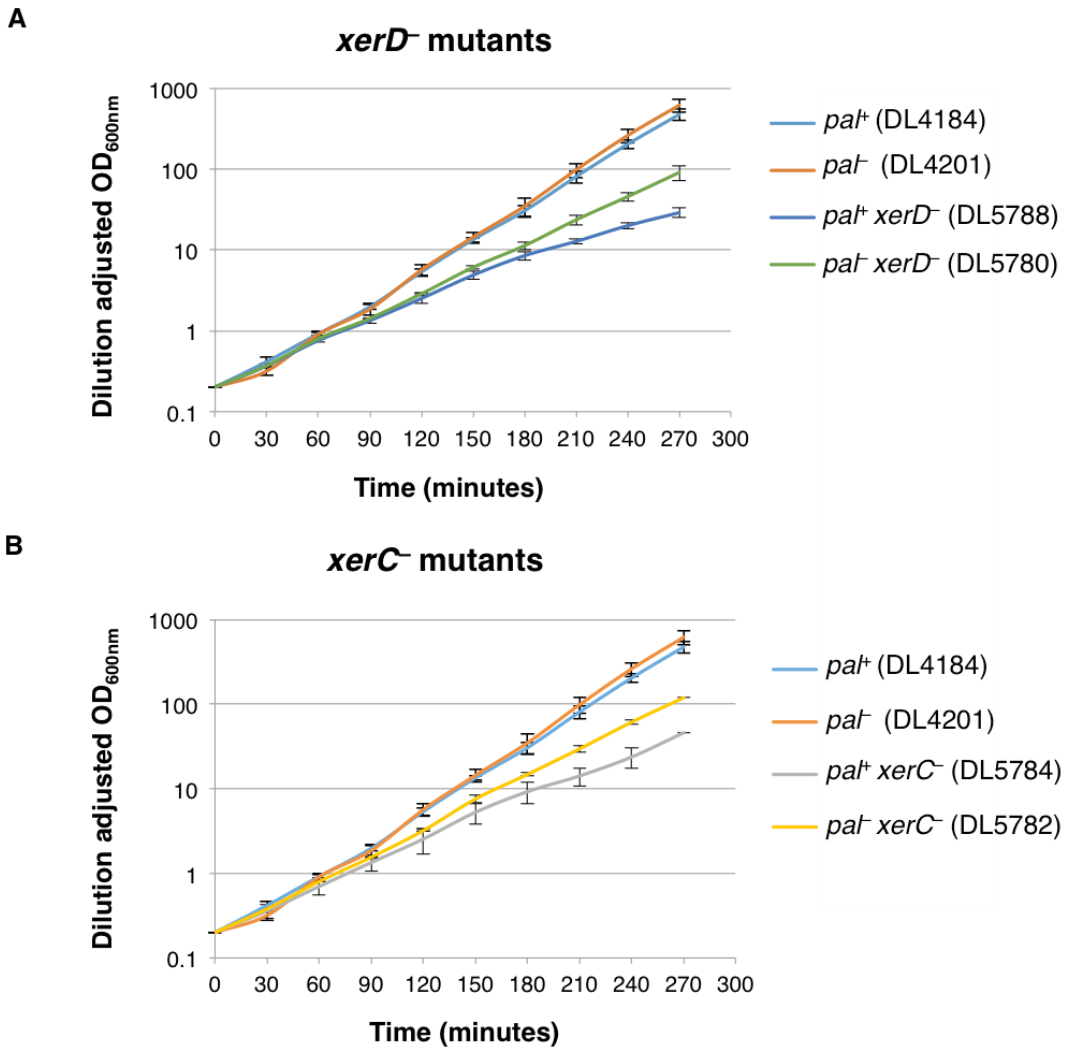
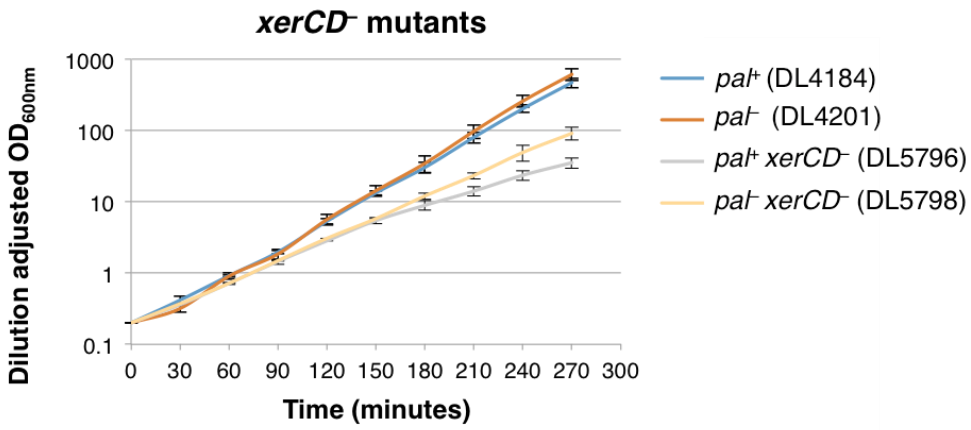


Figure 5.1. Absence of XerC and XerD affects cell growth.

Growth profiles of *xerC* and *xerD* mutants subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with 0.2% arabinose to induce SbcCD expression. Strains used were DL4184 (*pal*$^+$), DL4201 (*pal*$^-$), DL5788 (*pal*$^+$ *xerD*$^-$), DL5780 (*pal*$^-$ *xerD*$^-$), DL5784 (*pal*$^+$ *xerC*$^-$) and DL5782 (*pal*$^-$ *xerC*$^-$). (A) Growth profiles of *xerD*$^-$ mutants; (B) Growth profiles of *xerC*$^-$ mutants.

A



B

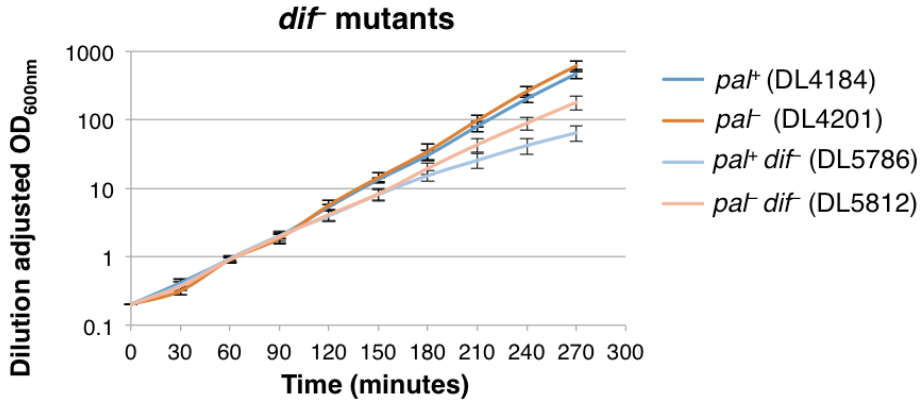


Figure 5.2. Absence of *xerCD/dif* affects cell growth.

Growth profiles of *xerCD* and *dif* mutants subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with 0.2% arabinose to induce SbcCD expression. Strains used were DL4184 (*pal*$^+$), DL4201 (*pal*$^-$), DL5796 (*pal*$^+$ *xerCD*$^-$), DL5798 (*pal*$^-$ *xerCD*$^-$), DL5786 (*pal*$^+$ *dif*$^-$) and DL5812 (*pal*$^-$ *dif*$^-$). (A) Growth profiles of *xerCD*$^-$ mutants; (B) Growth profiles of *dif*$^-$ mutants.

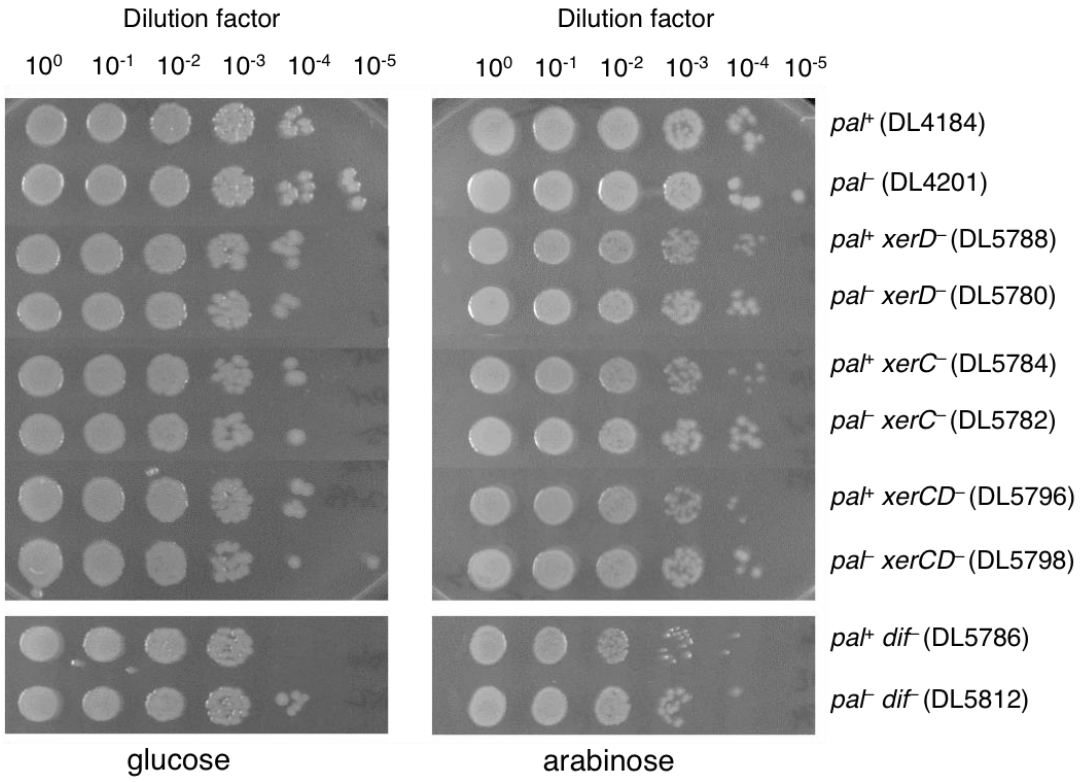


Figure 5.3. Dimer resolution mutants have small colony phenotype in presence of a DSB in *lacZ*.

Spot tests of 10-fold serially diluted samples collected after 4.5 hours of exponential growth in presence of arabinose and spotted onto LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Strains used were DL4184 (*pal*$^{+}$), DL4201 (*pal*$^{–}$), DL5788 (*pal*$^{+}$ *xerD*$^{–}$), DL5780 (*pal*$^{–}$ *xerD*$^{–}$), DL5784 (*pal*$^{+}$ *xerC*$^{–}$), DL5782 (*pal*$^{–}$ *xerC*$^{–}$), DL5796 (*pal*$^{+}$ *xerCD*$^{–}$), DL5798 (*pal*$^{–}$ *xerCD*$^{–}$), DL5786 (*pal*$^{+}$ *dif*$^{–}$) and DL5812 (*pal*$^{–}$ *dif*$^{–}$).

5.3 Chromosome dimer resolution mutants accumulate paused replication forks at *terA* and *terB* sites

If a formation of a DSB around the *dif* site leads to an over-replication between *terA* and *terB* sites and the XerCD/*dif* system is responsible for these DSBs in the terminus, I expect that in the absence of XerCD/*dif* system DSBs as well as over-replication in the terminus will disappear and, therefore, less replication forks will pause at *terA* and *terB* sites. To quantify the amount of replication forks that stop at *ter* sites in chromosome dimer resolution mutants or wild type cells, 2D gels and Southern blots were performed using DNA isolated from the *xerCD*/*dif* mutants described in section 5.2. Three different radioactive probes for *terA*, *terB* and *terC* were used, as described in Chapter 3.2. In the morning, overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 37°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and 0.2% arabinose was added to induce a DSB in *lacZ*. These cultures were grown for 1 hour until they reached an OD_{600nm} of 0.2-0.3. Then, cells were harvested by centrifugation and embedded in 0.8% agarose plugs. The cells were lysed for about 40 hours and the DNA in these plugs was digested by either NcoI or XmnI enzyme. NcoI enzyme was used to generate a 6.7 kb fragment which is recognised by the *terC* probe (Figure 5.4 A) and XmnI enzyme was used to generate either 4.9 kb or 4.1 kb fragments to visualise *terB* or *terA* sites, respectively (Figure 5.5 A and Figure 5.6 A). Southern blots of 2D gels carried out for visualising the *terC*, *terB* and *terA* fragments are shown in Figure 5.4 C, 5.5 C and 5.6 C. Linear DNA, Y-arcs and termination structures (fork fusions) were detected in all strains. Some of the strains in *terB* and *terA* fragments were partially digested by XmnI enzyme, therefore, the amount of paused replication forks and linear DNA (n) looks lower on the 2D gels. Additional experiments were

performed to confirm that partial digestion by XmnI enzyme did not affect the quantification of pause/n ratio (data not shown). Paused replication forks at *ter* sites were detected in all strains at the expected place on the Y-arc (Figures 5.4 B, 5.5 B and 5.6 B). The amount of replication forks paused at the *terC* fragment in chromosome dimer resolution mutants was similar (*pal*$^-$ strains) or lower (*pal*$^+$ strains) than the amount of paused replication forks in wild type strains (Figure 5.4 C). The amount of paused replication forks at the *terB* site was larger in the chromosome dimer resolution mutants than in the wild type strains (Figure 5.5 C). *terB* fragments of *dif* mutants were partially digested by XmnI enzyme, therefore, the amount of paused replication forks and linear DNA (n) looked lower on the 2D gels. The amount of paused replication forks at the *terA* fragment in chromosome dimer resolution mutants was similar (DL5788 strain) or larger (rest of the strains) than the amount of paused replication forks in wild type strains. *terA* fragments in *xerD*$^-$ and *dif*$^-$ strains were partially digested by XmnI enzyme which did not influence the quantification of pause/n ratio for these strains (Figure 5.6 C). To determine the amount of paused replication forks, I quantified all 2D gels as described in Chapter 3.2. The quantification of the fraction of paused replication forks for each strain is shown in Figures 5.4 D, 5.5 D and 5.6 D. More replication forks were paused at *terB* and *terA* sites in *xerCD*/*dif* mutants containing or not a palindrome in *lacZ* compared to wild type strains (Figures 5.5 D and 5.6 D). Similar amount of replication forks were paused at *terC* in *pal*$^-$ strains and more replication forks were observed in the wild type strain that contains a palindrome than in *xerCD*/*dif* mutants that contain a palindrome (Figure 5.4 D). This experiment suggests that *xerC, xerD, xerCD* and *dif* mutations lead to similar results where more replication forks are stopped at *terA* and *terB* sites which might lead to an additional over-replication in that region.

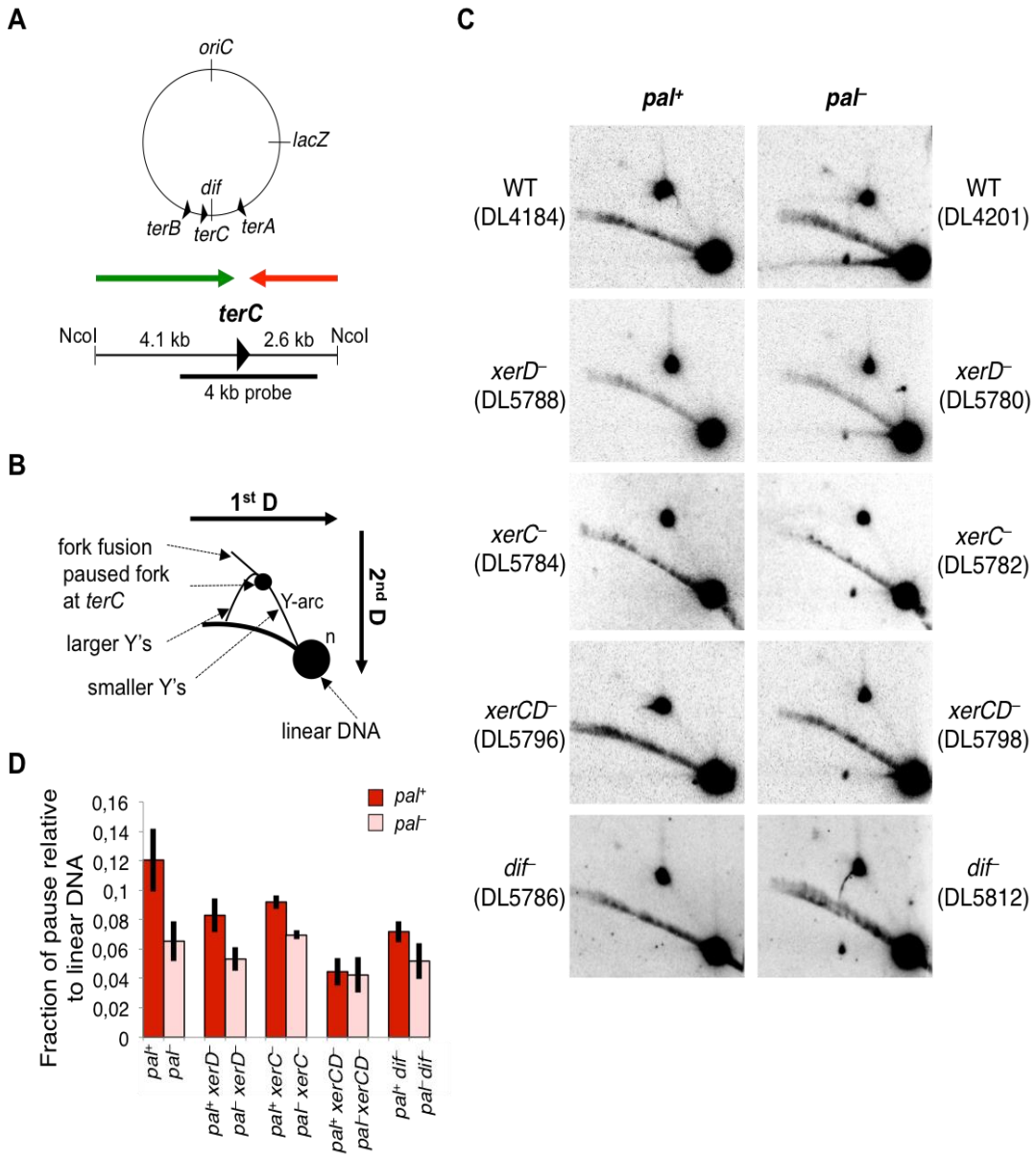


Figure 5.4. A lower amount of paused replication forks is detected at *terC* site in *xerCD/dif* mutants.

(A) *E. coli* map with the position of *ter* sites and NcoI digestion map of the region surrounding the *terC* site. The orientation of the *terC* site is indicated by a black triangle. The black thick line is the position of the *terC* probe. The direction of the replication is indicated by the red and green arrows. (B) Schematic representation of the migration patterns of branched DNA molecules when separated on a 2D agarose gel electrophoresis and hybridised with the *terC* probe. (C) 2D gel showing the *terC* fragment. (D) Individual quantifications of fraction of paused replication forks relative to linear DNA for *terC* fragment. Strains used were DL4184 (*pal*$^+$), DL4201 (*pal*$^-$), DL5788 (*pal*$^+$ *xerD*$^-$), DL5780 (*pal*$^-$ *xerD*$^-$), DL5784 (*pal*$^+$ *xerC*$^-$), DL5782 (*pal*$^-$ *xerC*$^-$), DL5796 (*pal*$^+$ *xerCD*$^-$), DL5798 (*pal*$^-$ *xerCD*$^-$), DL5786 (*pal*$^+$ *dif*$^-$) and DL5812 (*pal*$^-$ *dif*$^-$).

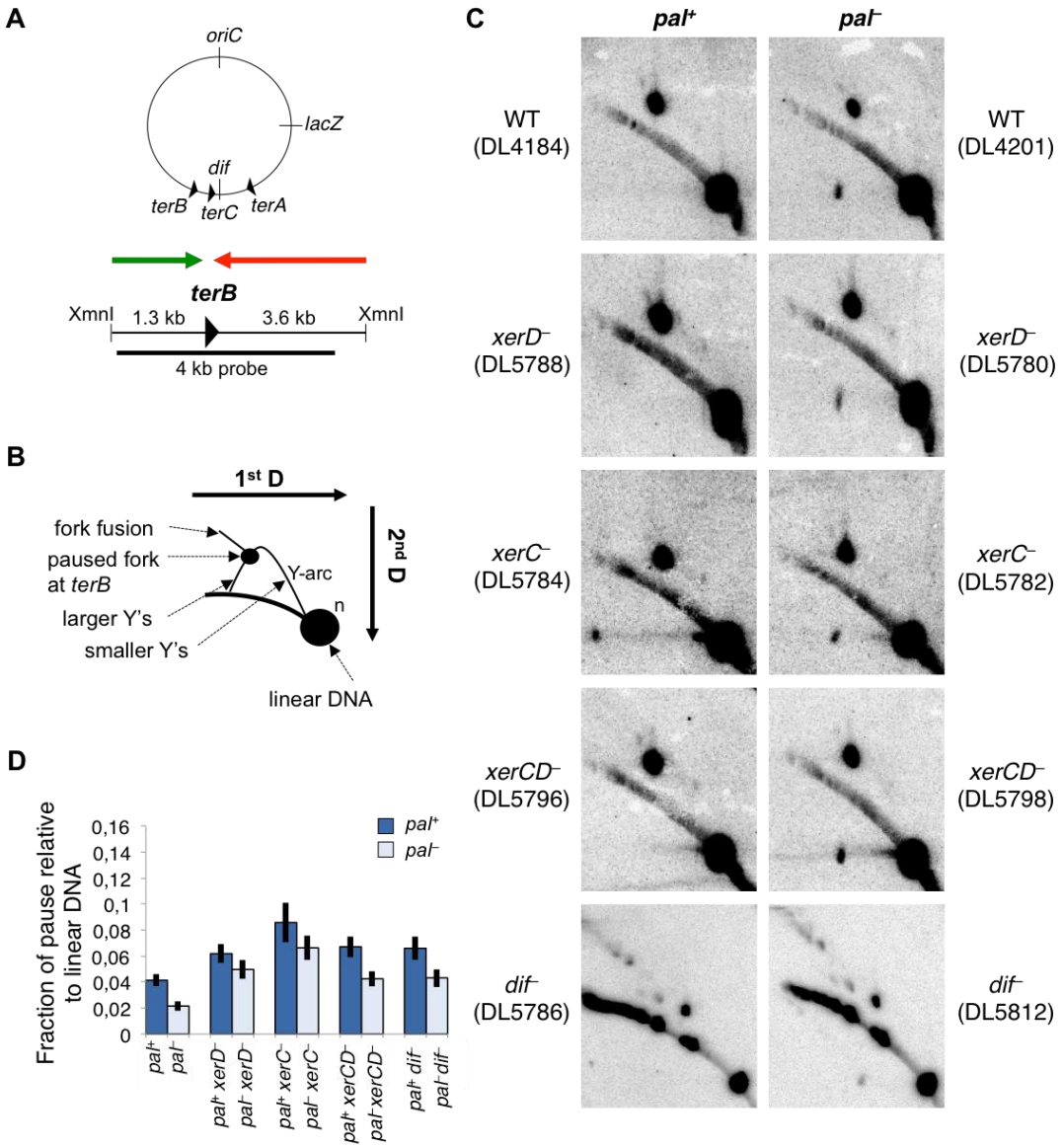


Figure 5.5. Larger amount of paused replication forks is detected at *terB* site in *xerCD/dif* mutants.

(A) *E. coli* map with the position of *ter* sites and XmnI digestion map of the region surrounding the *terB* site. The orientation of the *terB* site is indicated by a black triangle. The black thick line is the position of the *terB* probe. The direction of the replication is indicated by the red and green arrows. (B) Schematic representation of the migration patterns of branched DNA molecules when separated on a 2D agarose gel electrophoresis and hybridised with the *terB* probe. (C) 2D gel showing the *terB* fragment. (D) Individual quantifications of fraction of paused replication forks relative to linear DNA for *terB* fragment. Strains used were DL4184 (*pal*$^{+}$), DL4201 (*pal*$^{-}$), DL5788 (*pal*$^{+}$ *xerD*$^{-}$), DL5780 (*pal*$^{-}$ *xerD*$^{-}$), DL5784 (*pal*$^{+}$ *xerC*$^{-}$), DL5782 (*pal*$^{-}$ *xerC*$^{-}$), DL5796 (*pal*$^{+}$ *xerCD*$^{-}$), DL5798 (*pal*$^{-}$ *xerCD*$^{-}$), DL5786 (*pal*$^{+}$ *dif*$^{-}$) and DL5812 (*pal*$^{-}$ *dif*$^{-}$).

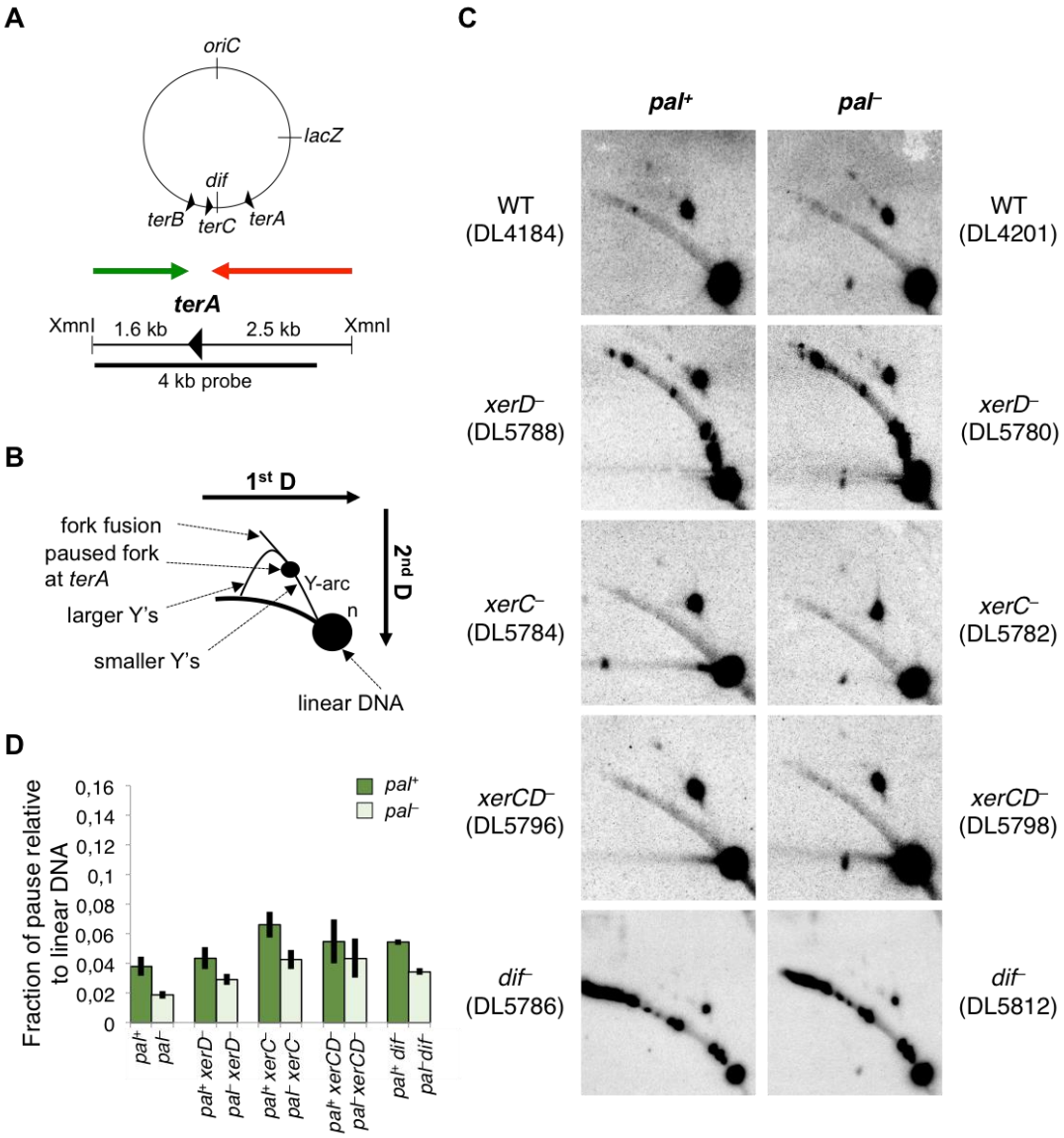


Figure 5.6. Larger amount of paused replication forks is detected at *terA* site in *xerCD/dif* mutants.

(A) *E. coli* map with the position of *ter* sites and XmnI digestion map of the region surrounding the *terA* site. The orientation of the *terA* site is indicated by a black triangle. The black thick line is the position of the *terA* probe. The direction of the replication is indicated by the red and green arrows. (B) Schematic representation of the migration patterns of branched DNA molecules when separated on a 2D agarose gel electrophoresis and hybridised with the *terA* probe. (C) 2D gel showing the *terA* fragment. (D) Individual quantifications of fraction of paused replication forks relative to linear DNA for *terA* fragment. Strains used were DL4184 (*pal*$^{+}$), DL4201 (*pal*$^{–}$), DL5788 (*pal*$^{+}$ *xerD*$^{–}$), DL5780 (*pal*$^{–}$ *xerD*$^{–}$), DL5784 (*pal*$^{+}$ *xerC*$^{–}$), DL5782 (*pal*$^{–}$ *xerC*$^{–}$), DL5796 (*pal*$^{+}$ *xerCD*$^{–}$), DL5798 (*pal*$^{–}$ *xerCD*$^{–}$), DL5786 (*pal*$^{+}$ *dif*$^{–}$) and DL5812 (*pal*$^{–}$ *dif*$^{–}$).

In all these strains the amount of paused replication forks seems higher in cells subjected to DSBs in *lacZ* than in cells without a DSB in *lacZ*. To compare the effect of a DSB in these different mutants to the wild type background, the amount of paused replication forks in *pal*$^-$ strain was substracted from the amount of paused replication forks in *pal*$^-$ strain. This was calculated for each strain and each *ter* fragment (Figure 5.7). Similar amount of replication forks paused at *terB* and *terA* for all strains suggests that despite higher amount of replication forks that stopped at *terA* and *terB* in *xerCD*/*dif* mutants than in wild type, the DSB at the palindrome site adds approximately the same amount of paused replication forks to *terA* and *terB* sites in both wild type and mutants (Figure 5.7). This observation suggests that the amount of paused replication forks observed in the absence of the XerCD/*dif* system and the amount of paused replication forks observed in the presence of DSBs in *lacZ* are additive.

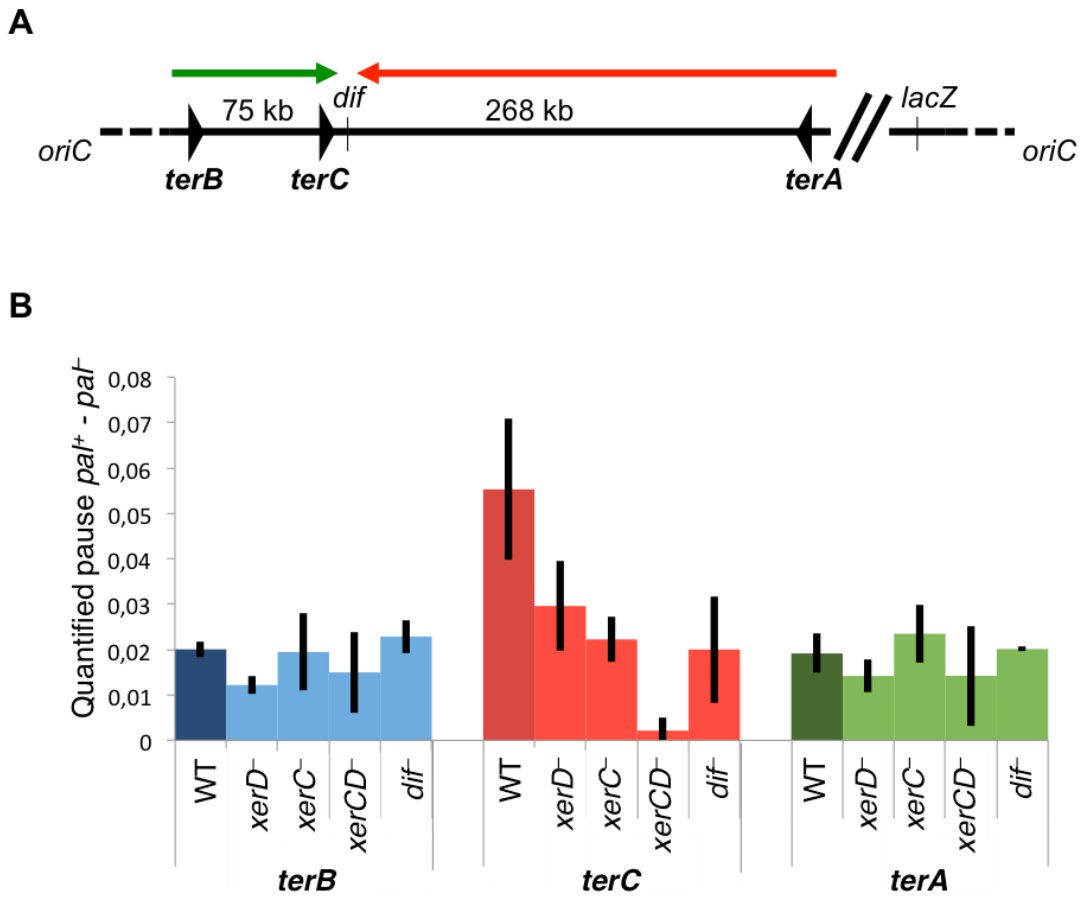


Figure 5 7. Difference in quantified amount of paused replication forks at *terB*, *terC* and *terA* sites in DSB⁻ substracted from DSB⁺ strains.

(A) Schematic representation of the position of *ter* sites on the *E. coli* chromosome. The orientation of *ter* sites is indicated by black triangles. The position of the *dif* site is showed. The direction of the replication is indicated by the red and green arrows. (B) DSB⁺ - DSB⁻ difference in quantified amount of paused replication forks on the Y-arcs normalised against the total linear DNA. Error bars represent the standard error of the mean where n = 3 (except for DL4201, DL4184 where n = 8). Strains used were DL4184 (*pal*⁺), DL4201 (*pal*⁻), DL5788 (*pal*⁺ *xerD*⁻), DL5780 (*pal*⁻ *xerD*⁻), DL5784 (*pal*⁺ *xerC*⁻), DL5782 (*pal*⁻ *xerC*⁻), DL5796 (*pal*⁺ *xerCD*⁻), DL5798 (*pal*⁻ *xerCD*⁻), DL5786 (*pal*⁺ *dif*⁻) and DL5812 (*pal*⁻ *dif*⁻).

5.4 WGS showed DNA loss around *dif* and an increase of DNA sequence in the terminus in *dif* mutants

Whole genome sequencing was performed to visualise an additional over-replication in the terminus of *xerCD*/*dif* mutants. Because so far chromosome dimer resolution mutants displayed similar behaviour, *dif* mutants were chosen for this experiment. The same growth conditions were used as described in section 5.3. The DNA for WGS was isolated by a Wizard kit and purified using a Zymo kit as described in Chapter 2. Edinburgh Genomics performed library preparations and WGS using an Illumina HiSeq 2500 platform. Obtained paired-end raw data were analysed as described in Chapter 2. Data from exponentially growing strains were normalised to data from stationary phase *pal*$^{-}$ (DL4201) strain and individual graphs for each biological repeat were carried out (Figure 5.8). Replication profiles of all these strains were V-shaped as expected. The DNA of all 7 *rrn* operons was poorly recovered (Figure 5.9). In a combined graph of the replication profiles of all strains in Figure 5.9, more DNA was observed in the terminus region of *dif* mutants compared to wild type strains. Interestingly, the peak of over-replication in the terminus in *dif* mutants was slightly shifted towards *terA*. This shift might be caused by a loss of DNA around the *dif*/*terC* region in *dif* mutants.

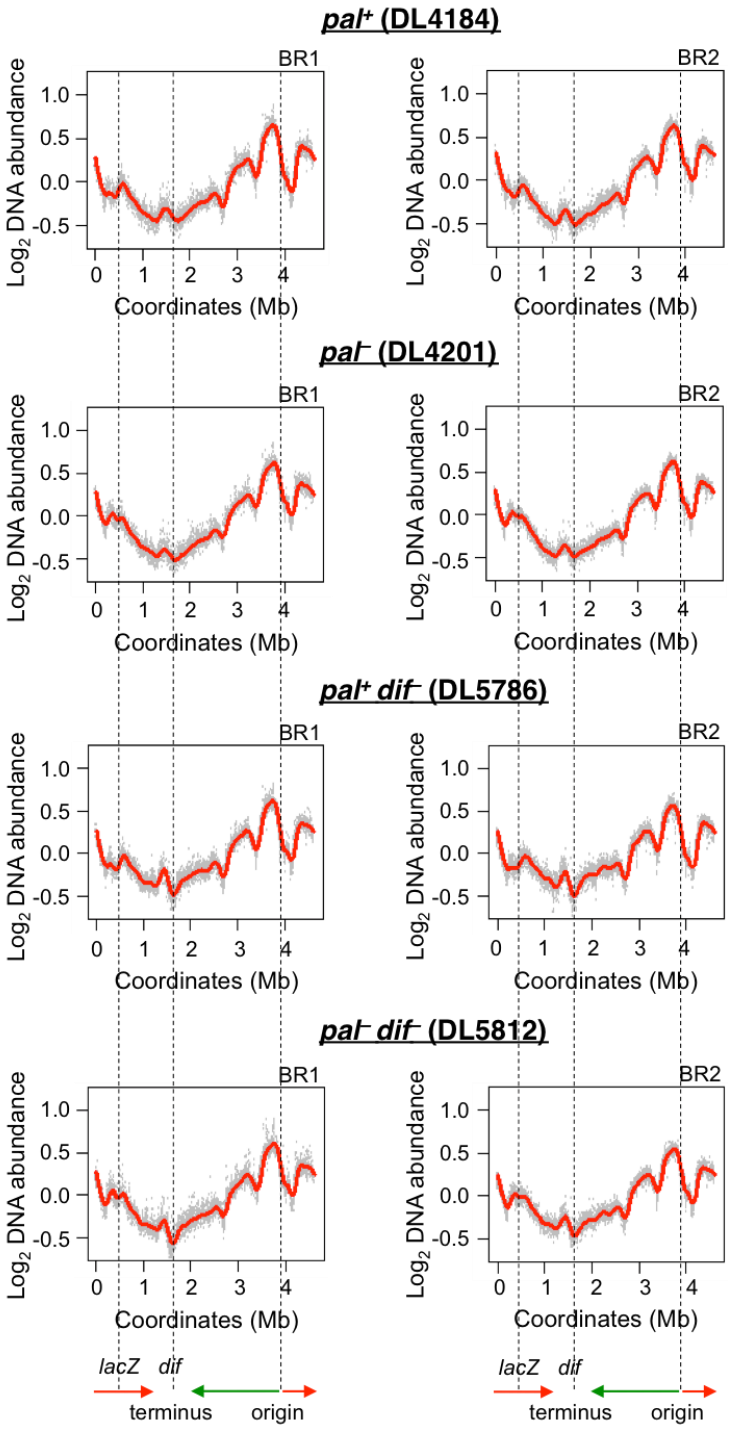


Figure 5.8. WGS profiles of biological repeats of *dif* mutants.

Replication profiles of exponentially growing cultures were normalised to data from stationary phase *pal*⁻ (DL4201) strain. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of the origin, *lacZ* and *dif* (terminus) are indicated using dash lines. BR1 is biological repeat 1; BR2 is biological repeat 2. Strains used were DL4184 (*pal*⁺), DL4201 (*pal*⁻), DL5786 (*pal*⁺ *dif*⁻) and DL5812 (*pal*⁻ *dif*⁻).

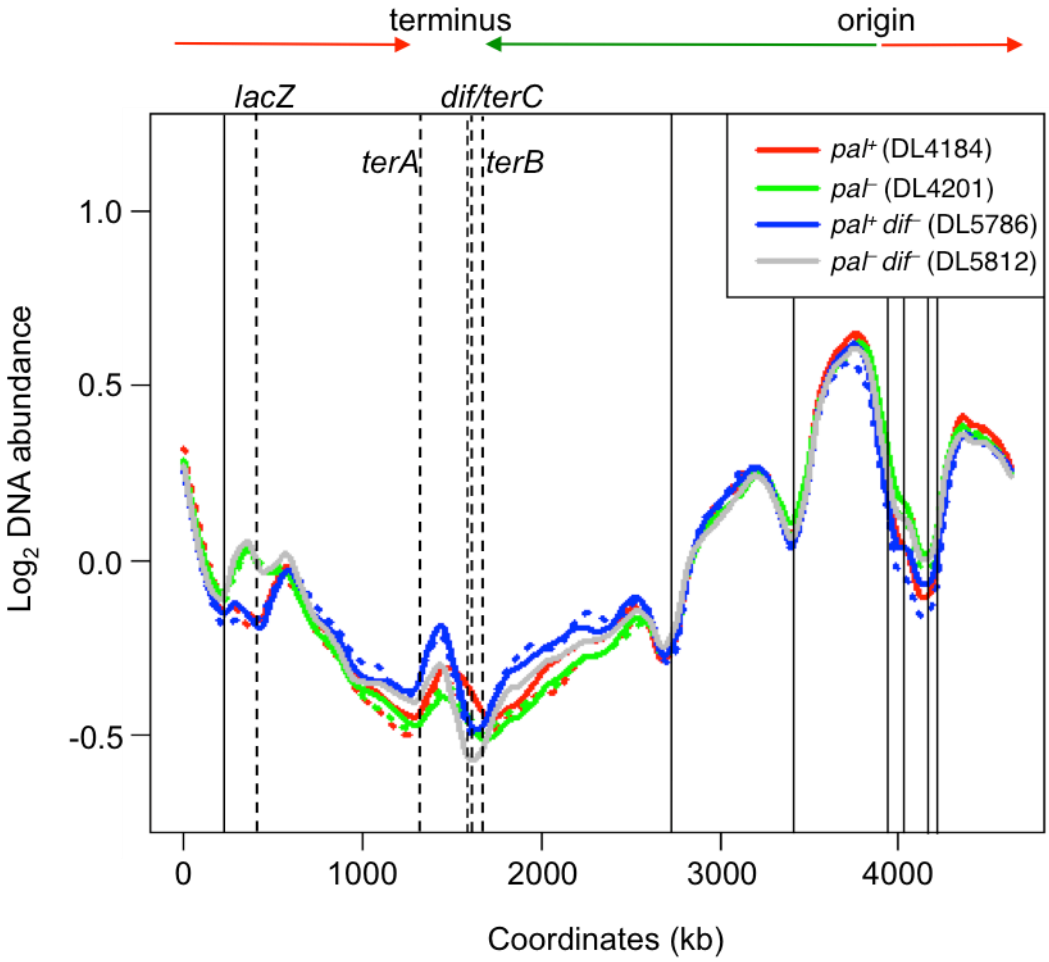


Figure 5.9. Combined replication profiles of all sequenced strains.

Replication profiles of exponentially growing cultures of *dif*⁻ and wild type strains containing or not a 246 bp palindrome and expressing SbcCD for 1 hour. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of the palindrome (*lacZ*), *dif, terA, terB* and *terC* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Biological repeats (BR1 and BR2) of the same strain are shown using dashed/solid lines of the same colour. Strains used were DL4184 (*pal*⁺) in red, DL4201 (*pal*⁻) in green, DL5786 (*pal*⁺ *dif*⁻) in blue and DL5812 (*pal*⁻ *dif*⁻) in grey.

To visualise the loss of DNA around the *dif* region, ratios of reads of *pal*$^{+}$ *dif*$^{-}$ (DL5786) over *pal*$^{+}$ (DL4184) and *pal*$^{-}$ *dif*$^{-}$ (DL5812) over *pal*$^{-}$ were carried out (Figure 5.10). Both ratios had similar patterns with a loss of DNA at the *dif* site. Also, in both graphs more DNA seems to be detected in the terminus and less DNA was detected around the origin region. The change in slope could be explained by different growth rates of *dif*$^{-}$ and wild type strains or by the death of some cells described in section 5.2 that stop initiating replication and, therefore, not contributing to the DNA replication. Because *dif*$^{-}$ strains are dying or growing slower compared to the wild type strains, the ratio looks 'inverted' with more DNA at the terminus and less DNA at the origin.

To visualise the impact of a DSB in *lacZ* in *dif* mutants in the terminus and abolish the differences in growth rate, the ratio of reads of *pal*$^{+}$ *dif*$^{-}$ (DL5786) over *pal*$^{-}$ *dif*$^{-}$ (DL5812) was carried out (Figure 5.11). The profile of this ratio looked similar to the *pal*$^{+}$/*pal*$^{-}$ (DL4184/DL4201) ratio with the same level of DNA excess in the terminus (Figure 5.11 A). This observation suggests that the amount of palindrome-mediated DNA excess in the terminus is the same in *dif* mutants and in wild type strains.

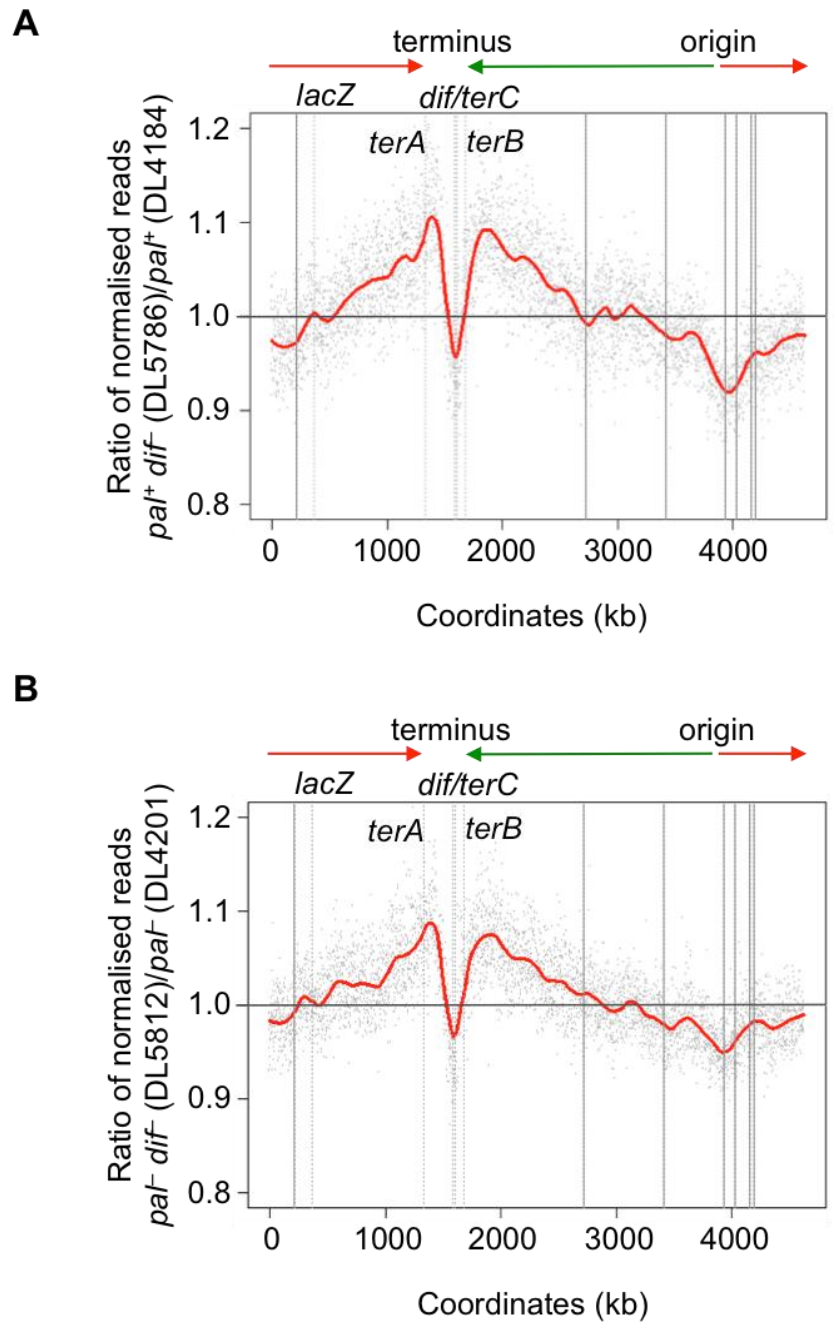


Figure 5 10. Ratio of replication profiles of *dif*⁻ strains over wild type strains.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of *pal*⁺ *dif*⁻ (DL5786) strain over *pal*⁺ (DL4184) strain (A) and of *pal*⁻ *dif*⁻ (DL5812) strain over *pal*⁻ (DL4201) strain (B). Grey dots represent reads, the red line is the fitted line of the reads from DL5786 (A) or DL5812 (B). The horizontal grey line represents DL4184 (A) or DL4201 (B). The direction of replication is shown using green and red arrows. The positions of the palindrome (*lacZ*), *terA*, *terB*, *terC* and *dif* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Strains used were DL4184 (*pal*⁺), DL4201 (*pal*⁻), DL5786 (*pal*⁺ *dif*⁻) and DL5812 (*pal*⁻ *dif*⁻).

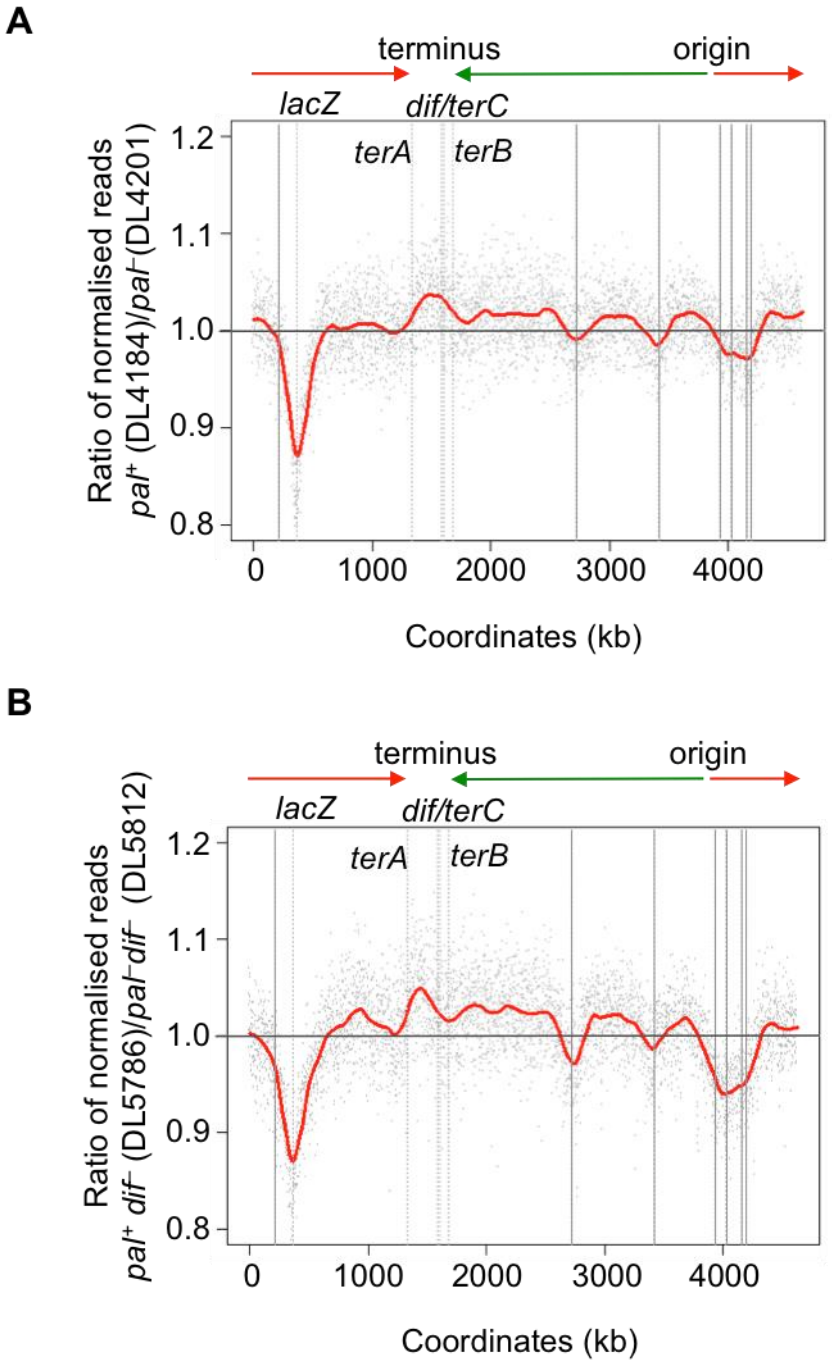


Figure 5.11. Ratio of replication profiles of *pal*$^{+}$ over *pal*$^{-}$ strains.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of *pal*$^{+}$ (DL4184) strain to *pal*$^{-}$ (DL4201) strain (A) and of *pal*$^{+}$*dif*$^{-}$ (DL5786) strain to *pal*$^{-}$ *dif*$^{-}$ (DL5812) strain (B). Grey dots represent reads, the red line is the fitted line of the reads from DL4184 (A) or DL5786 (B). The horizontal grey line represents DL4201 (A) or DL5812 (B). The direction of replication is shown using green and red arrows. The positions of the palindrome (*lacZ*), *terA*, *terB*, *terC* and *dif* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Strains used were DL4184 (*pal*$^{+}$), DL4201 (*pal*$^{-}$), DL5786 (*pal*$^{+}$ *dif*$^{-}$) and DL5812 (*pal*$^{-}$ *dif*$^{-}$).

5.5 DSBs at *ter* sites lead to over-replication in the terminus

5.5.1 *xerCD/dif* mutants reach steady state after 3.5 hours of SbcCD induction

Based on the growth curve data obtained for *xerCD*/*dif* mutants subjected or not to a DSB in *lacZ*, I identified three groups of strains that reached their steady state at the 3.5 h (210 min) time point (Figure 5.12). Steady state here means the time point where I could distinguish three groups of strains that were growing with different speed. Following this observation, I decided to check if there is a difference in the amount of DNA observed by WGS or the amount of RecA enrichment observed by ChIP-seq in the terminus in the strains where SbcCD was induced for 1 h or 3.5 h. Due to the observed differences between strains with either chronic DSB or short (1 h) acute DSB in *lacZ*, I expect a strain with longer time of acute DSB induction (3.5 h) to show similar result to a strain with a chronic DSB.

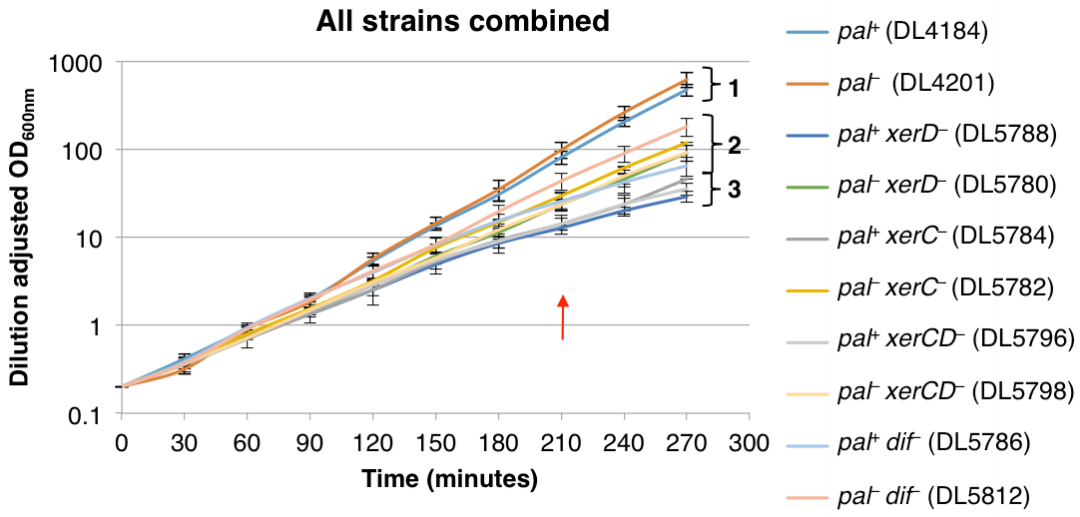


Figure 5.12. Growth curve revealed three different groups of strains after 3.5 hours (210 min) of SbcCD induction.

Growth profiles of *xerCD* and *dif* mutants subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with 0.2% arabinose to induce SbcCD expression. Strains used were DL4184 (*pal*$^{+}$), DL4201 (*pal*$^{-}$), DL5788 (*pal*$^{+}$ *xerD*$^{-}$), DL5780 (*pal*$^{-}$ *xerD*$^{-}$), DL5784 (*pal*$^{+}$ *xerC*$^{-}$), DL5782 (*pal*$^{-}$ *xerC*$^{-}$), DL5796 (*pal*$^{+}$ *xerCD*$^{-}$), DL5798 (*pal*$^{-}$ *xerCD*$^{-}$), DL5786 (*pal*$^{+}$ *dif*$^{-}$) and DL5812 (*pal*$^{-}$ *dif*$^{-}$).

5.5.2 Higher level of over-replication in the terminus is observed in the strain in which SbcCD was induced for 3.5 h

Following the observation described in section 5.5.1, I decided to compare *pal*$^+$ strains where SbcCD was induced for different length of time (1 h and 3.5 h) using whole genome sequencing to check if there is a difference in the amount of DNA in the terminus in these strains. Overnight culture of *pal*$^+$ (DL4184) was diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown to an OD_{600nm} of 0.2-0.25. Then, the culture was diluted again to an OD_{600nm} of 0.01, divided into 2 flasks and 0.2% arabinose was added to both flasks to induce SbcCD expression. Cultures in the first flasks were grown for 1 hour and cultures in the second flasks were grown for 3.5 hours. The OD_{600nm} was regularly monitored in order to keep these cultures at an OD_{600nm} of 0.2-0.3. The DNA for WGS was isolated using a Wizard kit and purified using a Zymo kit as described in Chapter 2. Edinburgh Genomics performed library preparations and WGS using an Illumina HiSeq 2500 platform. Obtained paired-end raw data were analysed as described in Chapter 2. Data from the exponentially growing *pal*$^+$ (DL4184) strain were normalised to data from the stationary phase *pal*$^-$ (DL4201) strain and individual graphs for each biological repeat were carried out (Figure 5.13). Replication profiles of all these strains were V-shaped as expected. The DNA of all 7 *rrn* operons was poorly recovered (Figure 5.14). The amount of DNA in the terminus region was higher in the *pal*$^+$ strain when SbcCD expression was induced for 3.5 hours than when SbcCD expression was induced for 1 hour. The depth of coverage (or average number of reads per nucleotide) for the strain in which SbcCD expression was induced for 3.5 hours was more variable than that for the strain in which SbcCD expression was induced for 1 hour.

To better visualise the difference between the replication profiles of the *pal*$^{+}$ strain subjected to 3.5 h or 1 h of SbcCD expression, the ratio of the number of reads of DL4184 (3.5 h) over DL4184 (1 h) was studied (Figure 5.14). This analysis revealed that the *pal*$^{+}$ strain in which SbcCD expression was induced for 3.5 hours had more DNA in the terminus region than the *pal*$^{+}$ strain in which SbcCD expression was induced for 1 hour. To confirm that this excess of DNA is correlated with the amount of paused replication forks at *terA* and *terB* sites and, therefore, if inducing an acute DSB for a longer period of time makes it similar to a chronic DSB (Chapter 3), additional 2D gels on this strain grown in the same conditions need to be performed. The loss of DNA at *rrn* genes was greater when the strain was subjected to a DSB for 3.5 hours. Additional peaks on the graph could be explained by the variability of depth of coverage between strains described above (Figure 5.13).

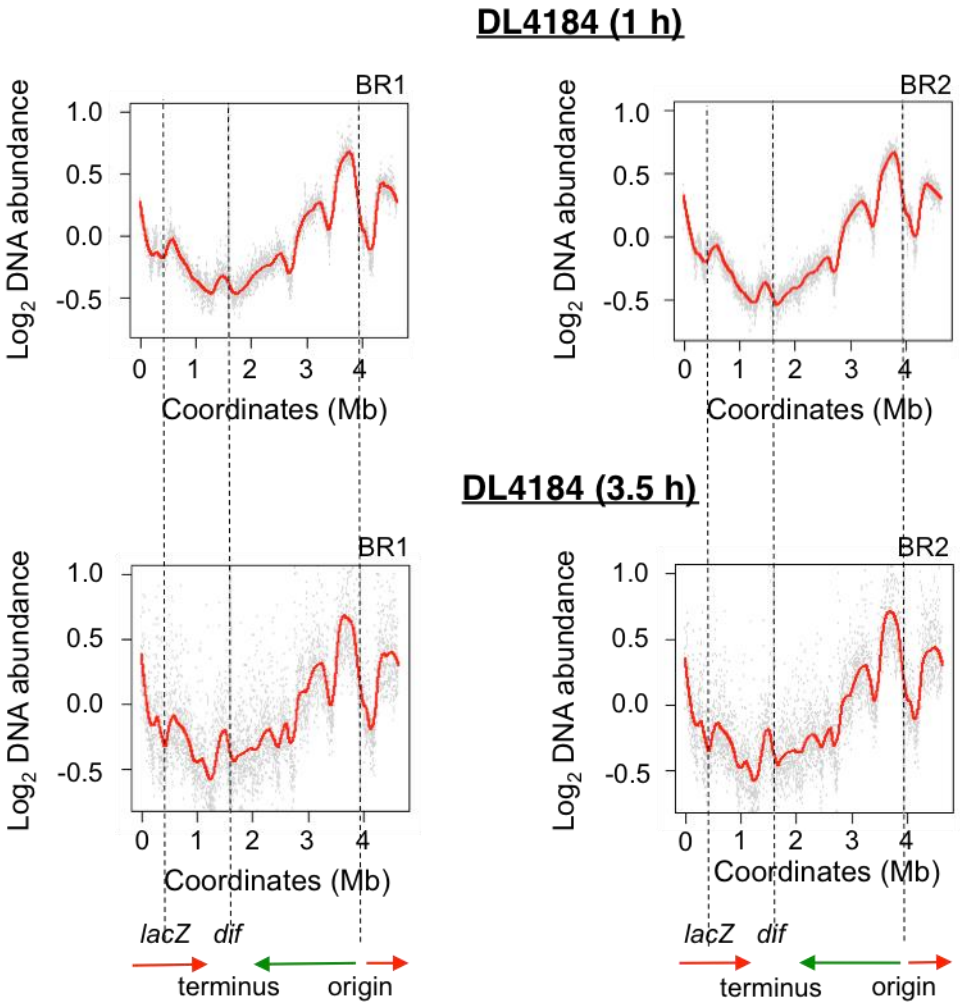


Figure 5.13. Biological repeats of WGS profiles of the *pal*⁺ (DL4184) strain in which SbcCD was induced for 1 h or 3.5 h.

Replication profiles of exponentially growing cultures of the *pal*⁺ (DL4184) strain in which SbcCD expression was induced for 1 hour or 3.5 hours. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. Grey dots represent reads, the red line is the fitted line of the reads from DL4184. The direction of replication is shown using green and red arrows. The positions of the origin, *lacZ* and *dif* (terminus) are indicated using dash lines. BR1 is biological repeat 1; BR2 is biological repeat 2.

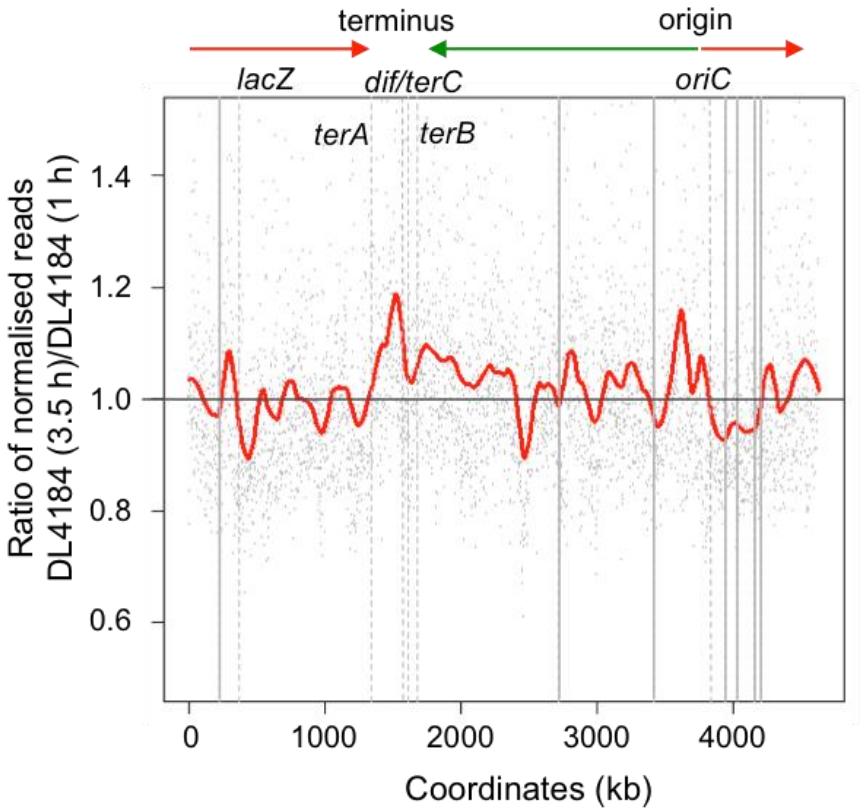


Figure 5.14. Ratio of replication profiles of the *pal*⁺ (DL4184) strain subjected to 3.5 h over 1 h of SbcCD expression.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of the *pal*⁺ (DL4184) strain in which SbcCD expression was induced for 3.5 hours over *pal*⁺ (DL4184) strain in which SbcCD expression was induced for 1 hour. Grey dots represent reads, the red line is the fitted line of the reads from DL4184 (3.5 h). The direction of replication is shown using green and red arrows. The positions of the palindrome (*lacZ*), *terA*, *terB*, *terC*, *dif* and *oriC* are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines.

5.5.3 The level of RecA enrichment at *ter* sites is increased when DSB in *lacZ* is induced for a longer period of time

To check if there is a difference in the amount of RecA enrichment and if it correlates with the observed excess of DNA in the terminus in *pal*$^{+}$ and *pal*$^{-}$ strains in which SbcCD was induced for different length of time (1 h and 3.5 h), I performed a RecA-ChIP-seq experiment. Overnight cultures of *pal*$^{-}$ (DL4201) and *pal*$^{+}$ (DL4184) strains were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and divided into 2 flasks, 0.2% arabinose was added to both flasks to induce SbcCD expression. Cultures in the first flasks were grown for 1 hour and cultures in the second flasks were grown for 3.5 hours. The OD_{600nm} was regularly monitored and the cells diluted in order to keep these cultures at an OD_{600nm} of 0.2-0.3. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 2500 platform (HiSeq 4000 for DL4201 3.5 h) by Edinburgh Genomics.

After the raw ChIP-seq data were analysed as described in Chapter 2, I noticed that RecA enrichment adjacent to *ter* sites was higher when SbcCD expression was induced for 3.5 hours than after 1 hour of SbcCD expression (Figure 5.15). As described in Chapter 4, it is not possible to quantitatively compare strains between each other using ChIP-seq data, therefore, all RecA peaks were compared within the strain, e.g. the level of RecA enrichment adjacent to *ter* sites was compared to the level of RecA enrichment around the *dif* site. Also, although DSBs at *ter* sites were detected in the *pal*$^{-}$ strain (DL4201), more RecA was bound to DNA around *ter* sites in a strain subjected to a DSB in *lacZ*

(DL4184). This suggests that when replication forks are less coordinated due to a delay at the DSBR site in *lacZ* at the right replichore, presumably, replication forks from the left replichore pause and break more often at the *terA* site, as it is the first *ter* site where the replication fork from the left replichore will pause. This phenomenon becomes more pronounced in the strain subjected to a DSB in *lacZ* for 3.5 hours due to a repetitive repair at *lacZ* locus.

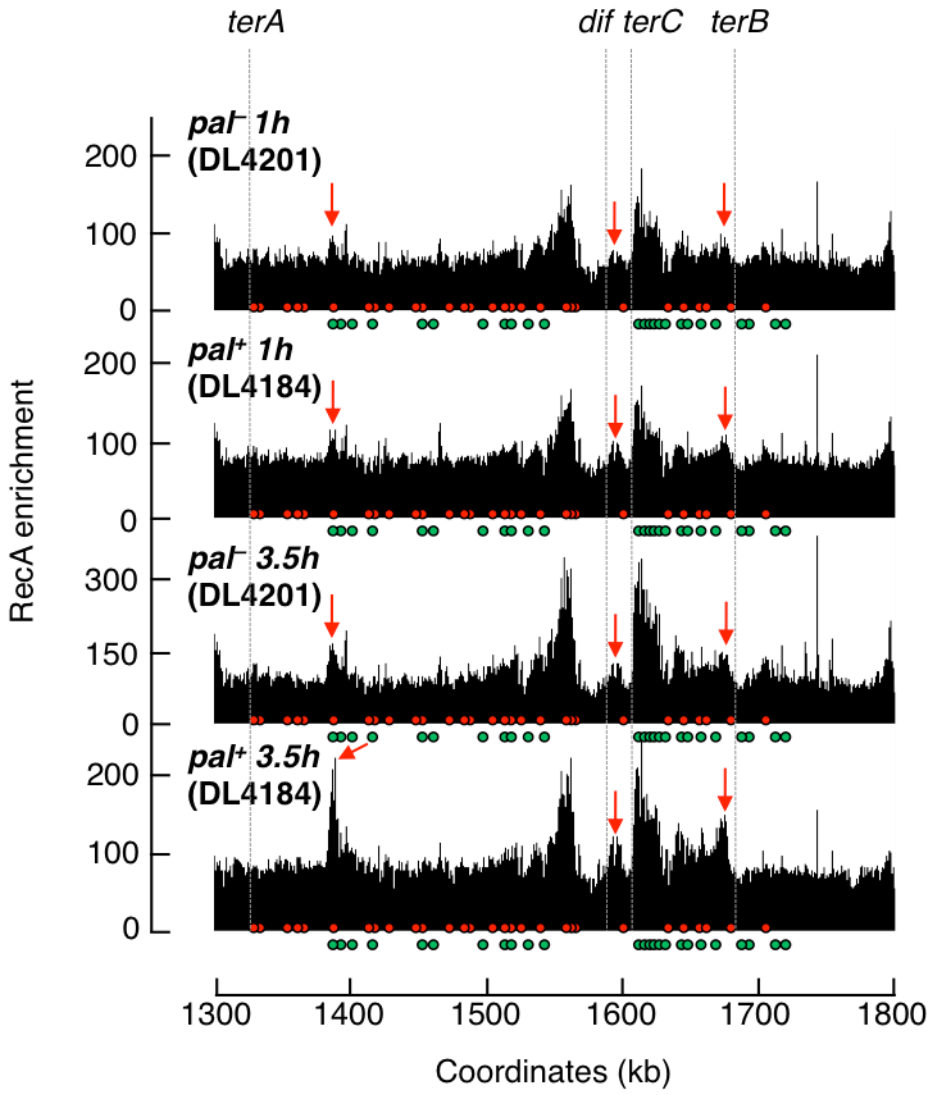


Figure 5 15. Comparison of RecA enrichment in the terminus region of strains in which SbcCD expression was induced for 1 h or 3.5 h.

ChIP-seq analyses revealed that more RecA binds adjacent to *ter* sites when a DSB in *lacZ* is induced for a longer period of time (3.5 h). The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Red arrows indicate RecA enrichment that corresponds to one-ended DSBs at *ter* sites. Strains used were DL4201 (*pal*$^-$) and DL4184 (*pal*$^+$).

To quantitatively determine the difference between the levels of RecA binding around *ter* sites in the *pal*$^{+}$ strain subjected to a DSB in *lacZ* for 1 h and 3.5 h, the relative levels of RecA enrichment were measured using qPCR (Figure 5.16). For this purpose, I designed 3 pairs of qPCR primers that were positioned after the appropriate Chi site in the correct orientation adjacent to each *ter* site (Figure 5.16 A). Relative quantification of ChIP-enriched DNA from the same samples described in Chapter 5.5.3 was performed. Relative quantification measures the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and shows the low level of RecA binding (Cockram *et al*. 2015). In the strain subjected to DSBs for 1 h, the relative RecA enrichment in terA1, terC2 and terB2 fragments was ~2-fold. In the strain subjected to DSBs for 3.5 h, the level of RecA enrichment was increased to ~2.5-3-fold (Figure 5.16 B). This quantification experiment supports the hypothesis that the level of RecA enrichment is increased following a longer DSB induction in *lacZ*.

Because RecA enrichment adjacent to *ter* sites was compared to the level of RecA enrichment around the *dif* site, I performed an additional control qPCR experiment using the same DNA, to confirm that the level around the *dif* site has not changed (Figure 5.17). From the ChIP-seq data in Figure 5.15, I expected the level of RecA enrichment around the *dif* site to stay the same. Indeed, the quantified amount of RecA bound to DNA stayed similar in *pal*$^{+}$ strain in both conditions, either subjected to a DSB in *lacZ* for 1 h or 3.5 h.

These observations correlate with the larger amount of DNA observed between *terA* and *terB* sites in this strain described in section 5.5.2. Hence, these results might suggest that DSBs at *ter* sites, rather than DSBs around the *dif* site, cause an increase of DNA over-replication in the terminus region.

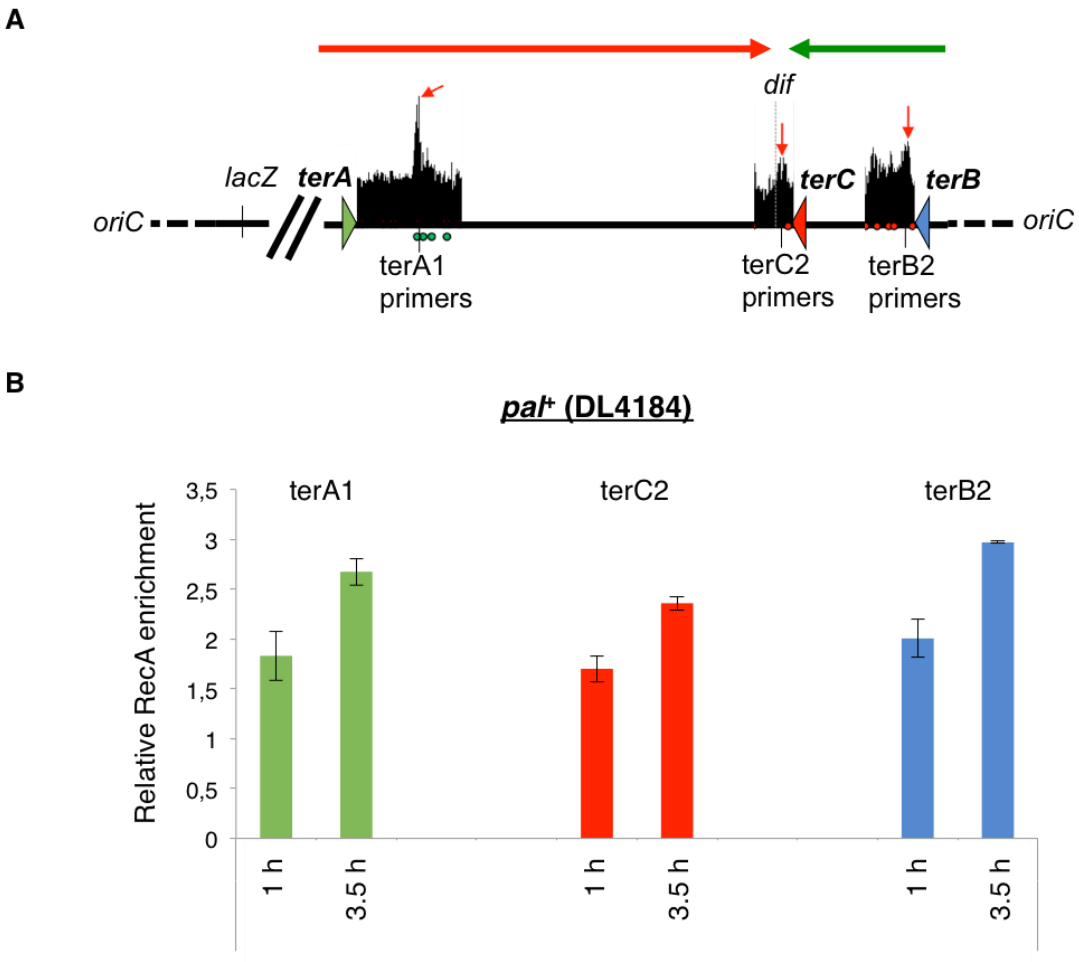


Figure 5.16. Quantification of RecA enrichment at *ter* sites in the *pal*$^+$ (DL4184) strain in which SbcCD expression was induced for 1 h or 3.5 h.

ChIP-qPCR revealed that more RecA binds adjacent to *ter* sites when a DSB in *lacZ* is induced for a longer period of time (3.5 h). (A) Schematic of qPCR primer positions. Coloured trianges show the position and orientation of *ter* sites. The data showed on top are taken from the ChIP-seq experiment of DL4184 strain subjected to a DSB for 3.5 h. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The palindrome position (*lacZ*), *dif*, *terA*, *terB, terC* sites and positions of qPCR primers are indicated. (B) Relative RecA enrichment around *ter* sites in DL4184 strain in which SbcCD expression was induced for 1 h or 3.5 h. RecA enrichment at *ter* sites is normalised to the RecA enrichment at the *hycG* site.

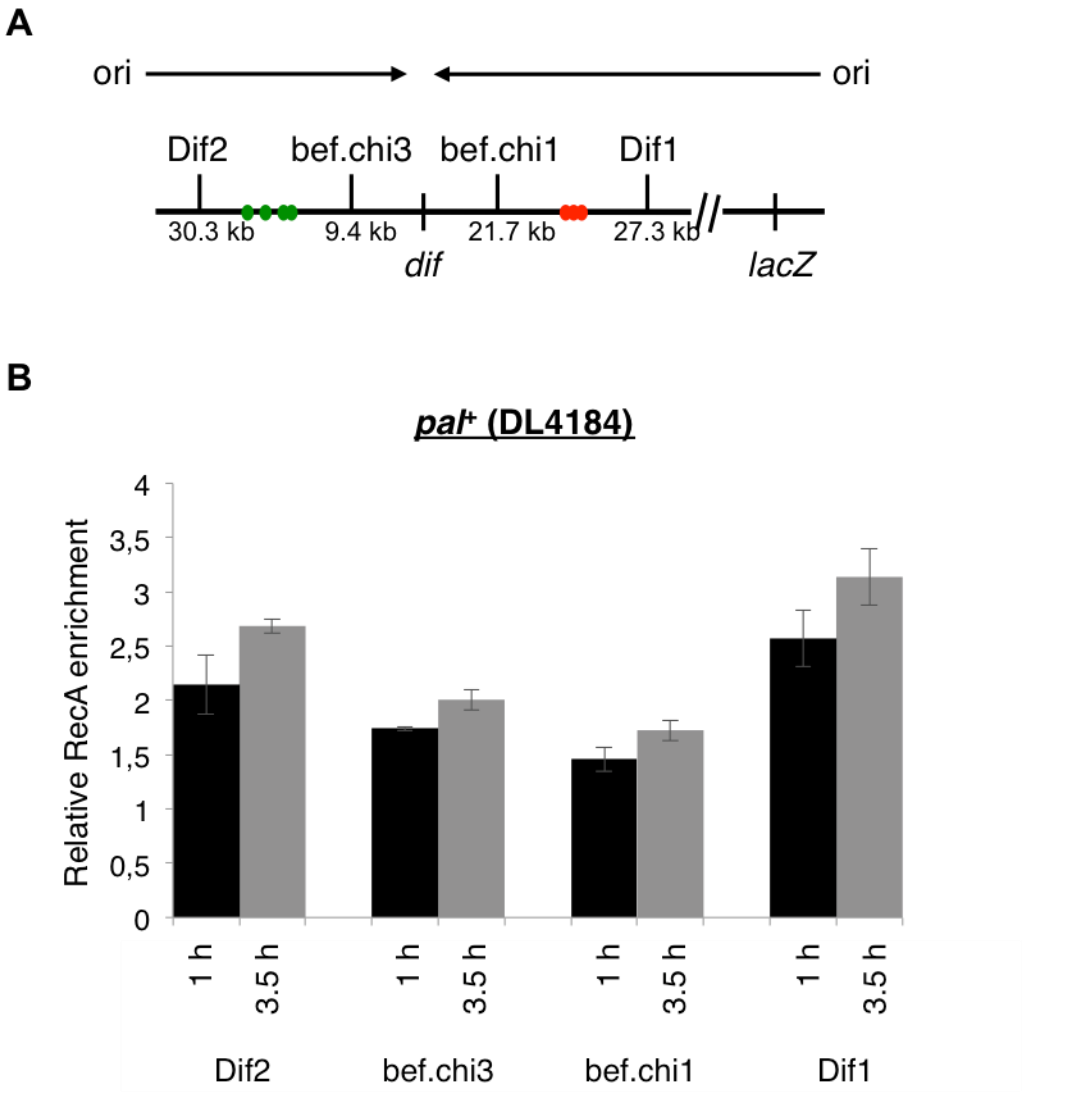


Figure 5.17. Quantification of RecA enrichment around the *dif* site in DL4184 strain in which SbcCD expression was induced for 1 h or 3.5 h.

ChIP-qPCR revealed that around similar amount of RecA binds around the *dif* site when a DSB in *lacZ* is induced for either 1 h or 3.5 h. (A) Schematic of qPCR primer positions. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The position of *dif* is indicated. (B) Relative RecA enrichment around the *dif* site in DL4184 strain with SbcCD expression induced for 1 h or 3.5 h. RecA enrichment at the *dif* site is normalised to RecA enrichment at the *hycG* site.

5.6 The level of RecA enrichment in the terminus is increased in chromosome dimer resolution mutants

5.6.1 *xerCD*/*dif* mutants show wider distribution of RecA enrichment in the terminus compared to wild type

Following the experiments from section 5.5, I decided to induce a DSB in *lacZ* for 3.5 hours in all future experiments. In order to determine if the XerCD/*dif* system is responsible for the appearance of DSBs in the terminus region, I performed a RecA-ChIP-seq experiment to visualise these DSBs. For this experiment, I used DL4201 (*pal*$^{-}$), DL4184 (*pal*$^{+}$), DL5796 (*pal*$^{+}$ *xerCD*$^{-}$), DL5798 (*pal*$^{-}$ *xerCD*$^{-}$), DL5786 (*pal*$^{+}$ *dif*$^{-}$) and DL5812 (*pal*$^{-}$ *dif*$^{-}$) strains. Overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and 0.2% arabinose was added to induce SbcCD expression. Cultures were grown for 3.5 hours. The OD_{600nm} was regularly monitored and the cells diluted in fresh prewarmed medium in order to keep these cultures at an OD_{600nm} of 0.2-0.3. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform (HiSeq 2500 for DL4184) by Edinburgh Genomics. Finally, raw ChIP-seq data were analysed as described in Chapter 2.

If the XerCD/*dif* system was responsible for DSBs in the terminus, I would expect the observed DSBs around the *dif* site to disappear. Surprisingly, it was not possible to determine if the XerCD/*dif* system was responsible for the initial DSBs around the *dif* site due to the appearance of even higher level of RecA enrichment in chromosome dimer resolution mutants (Figure 5.18). This

suggests that the presence of XerCD/*dif* prevents the appearance of more DSBs in the terminus. The RecA enrichment in these mutants was more distributed across the terminus compared to wild type strains but still centered around the *dif* region. This RecA binding was correlated with the presence of correctly oriented Chi sites, which suggests that it is RecBCD-mediated. There was no difference in RecA distribution between *pal*$^{+}$ and *pal*$^{-}$ strains, except for the RecA binding at *terA* and *terB* sites, suggesting that the DSB in *lacZ* did not change the pattern of RecA enrichment around the *dif* site. The level and distribution of RecA suggest that the 'guillotining' hypothesis described previously might be correct (Hendricks *et al*. 2000; Niki *et al*. 1991). When the XerCD/*dif* system is not present, *E. coli* is unable to resolve dimer of chromosomes and the DNA is sheared in the terminus, this process is called 'guillotining'. Then, the SOS response is induced, cells filament, viability is reduced and dead anucleate cells are produced. It was suggested that the chromosome dimer would be broken by the closing septum. The RecA-ChIP-seq results presented here are in accordance with the 'guillotining' of chromosome dimers. No significantly increased RecA enrichment between the *dif* site and the *terC* site suggests that the DSBs are still concentrated around the *dif* region, presumably, by the FtsK translocation that aligns two *dif* sites together and positions them at the septum.

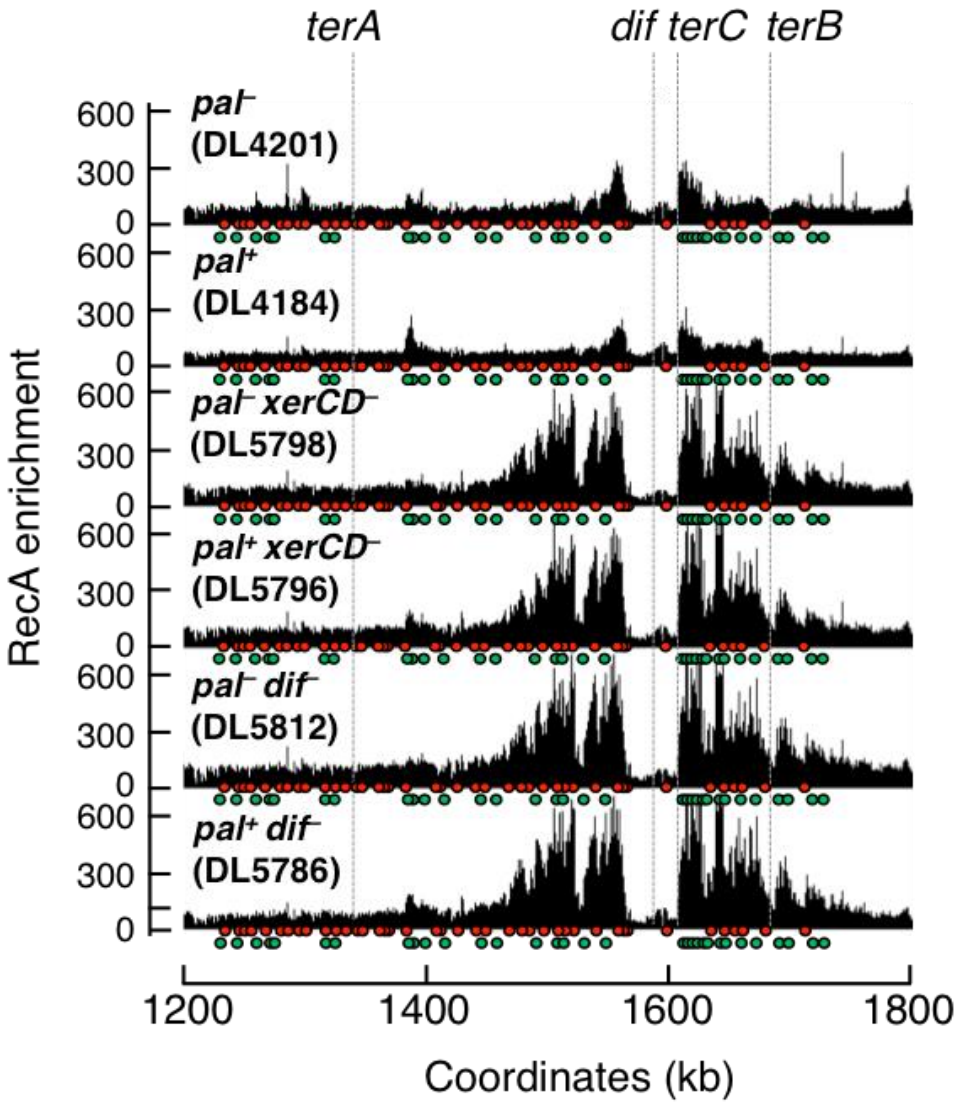


Figure 5.18. Comparison of RecA enrichment in the terminus region of *xerCD/dif* mutants.

ChIP-seq analyses revealed that more RecA binds in the terminus in *xerCD/dif* mutants. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal*$^{-}$), DL4184 (*pal*$^{+}$), DL5798 (*pal*$^{-}$ *xerCD*$^{-}$), DL5796 (*pal*$^{+}$ *xerCD*$^{-}$), DL5812 (*pal*$^{-}$ *dif*$^{-}$) and DL5786 (*pal*$^{+}$ *dif*$^{-}$).

5.6.2 RecA enrichment adjacent to *terA* site and not around the *dif* region of *xerCD*/*dif* mutants is stimulated by DSBR at *lacZ*

To confirm that additional DSB in *lacZ* do not increase the relative level of RecA binding in the terminus of *xerCD*/*dif* mutants, I performed ChIP-qPCR on the same DNA from the strains described in section 5.6.1 (Figure 5.19). For this experiment, I used 4 pairs of qPCR primers: 2 pairs of primers after Chi sites – Dif1 and Dif2 and 2 pairs of primers before Chi sites – bef.chi1 and bef.chi3 on each side of *dif* (Figure 5.19 A). I measured the relative quantification of RecA enrichment levels at each fragment, which was calculated as the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and shows background level of RecA binding (Cockram *et al*. 2015). As expected from the ChIP-seq experiments in section 5.6.1, this quantification experiment showed an increased level of RecA enrichment in the *xerCD*/*dif* mutants compared to the wild type. Although, at the same time, ChIP-qPCR revealed that the levels of RecA binding in the mutants were not increased when DSBs in *lacZ* were induced. In the *xerCD*/*dif* mutants that were not subjected to DSBs, the relative RecA enrichment in Dif1 and Dif2 was ~6- to 8-fold (Figure 5.19 B C). In the *xerCD* mutant subjected to DSBs, the level of RecA enrichment decreased to ~5-fold compared to *xerCD* mutant that was not subjected to DSBs in *lacZ* (Figure 5.19 B). On the other hand, the level of RecA enrichment stayed similar in the *dif* mutants subjected or not to DSBs (Figure 5.19 C).

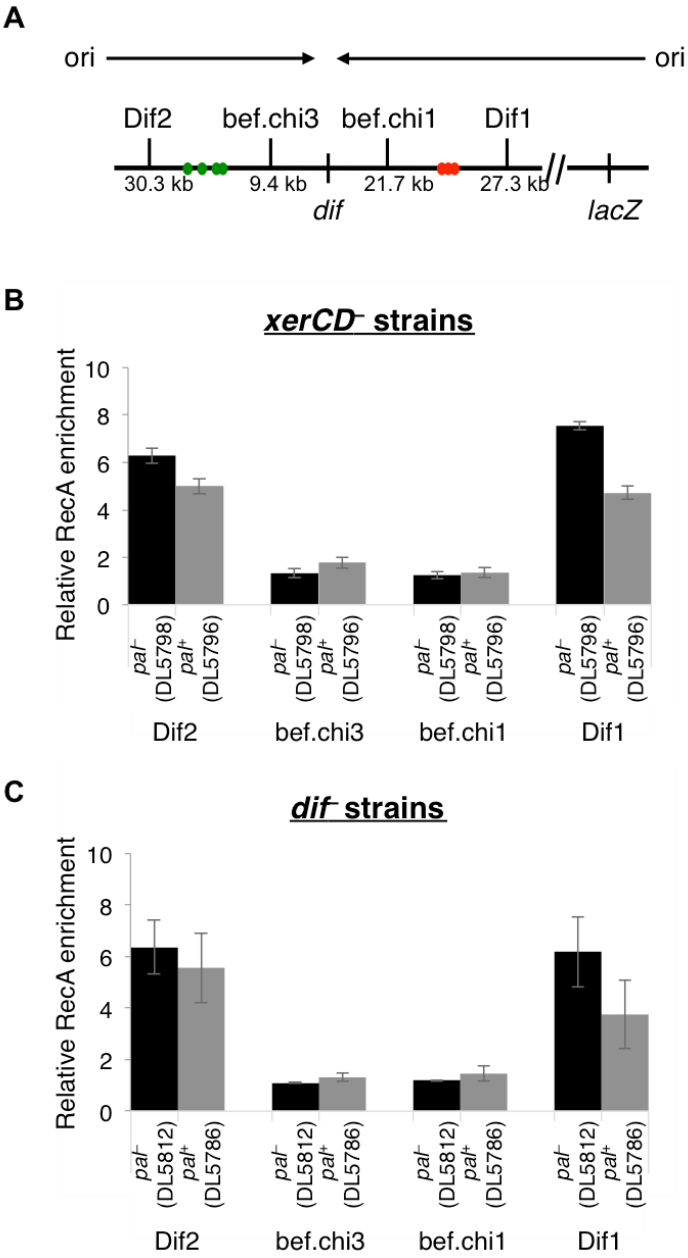


Figure 5.19. Quantification of RecA enrichment in the terminus region of *xerCD/dif* mutants.

ChIP-qPCR revealed that the levels of RecA binding were not increased when DSB in *lacZ* was induced. (A) Schematic of qPCR primer positions. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site and positions of qPCR primers are indicated. (B) Relative RecA enrichment in the terminus of *xerCD*⁻ strains. (C) Relative RecA enrichment in the terminus of *dif*⁻ strains. RecA enrichment in the terminus is normalised to the level of RecA enrichment at the *hycG* site. Strains used were DL5798 (*pal*⁻ *xerCD*⁻), DL5796 (*pal*⁺ *xerCD*⁻), DL5812 (*pal*⁻ *dif*⁻) and DL5786 (*pal*⁺ *dif*⁻).

To confirm that the level of RecA enrichment at *ter* sites was increased, following a DSB in *lacZ*, the relative levels of RecA enrichment adjacent to *ter* sites were measured using qPCR (Figure 5.20). For this purpose, the same DNA from *dif*⁻ strains, described in the ChIP-seq experiment earlier, was used. I used 3 pairs of qPCR primers that were positioned after the appropriate Chi site in the correct orientation adjacent to each *ter* site (Figure 5.20 A). Relative quantification measures the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and shows a background level of RecA enrichment (Cockram *et al*. 2015). In the strain subjected to DSBs in *lacZ*, the relative RecA enrichment in terA1 fragment was increased compared to the strain without a DSB in *lacZ* (~2.3- and ~1.4-fold, respectively). terC2 and terB2 fragments showed the same level of RecA enrichment for *dif* mutant subjected or not to a DSB (~1.6- and ~3-fold, respectively) (Figure 5.20 B). This quantification experiment supports the hypothesis that the level of RecA enrichment is increased at *ter* sites following a DSB induction in *lacZ*. Also, this result suggests that the over-replication in the terminus observed in *pal*⁺ *dif*⁻ strain in Figure 5.11 B might be linked with DSBs at *ter* sites and not with the DSBs around *dif* region. This leads to a conclusion that the DSBs at *ter* sites might induce an over-replication in the terminus.

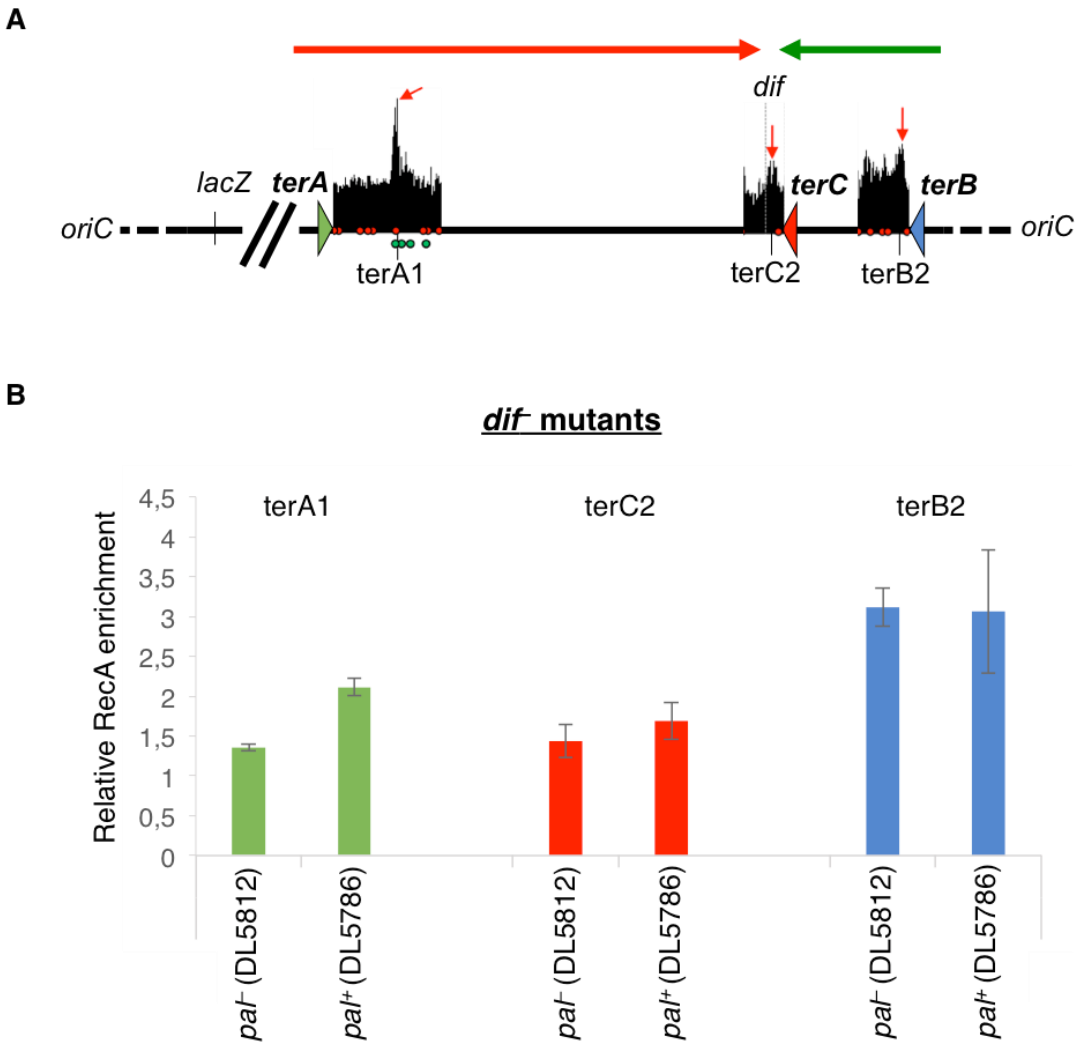


Figure 5.20. Quantification of RecA enrichment at *ter* sites in the *dif*⁻ mutants.

ChIP-qPCR revealed that more RecA binds adjacent to *terA* site when a DSB in *lacZ* is induced. (A) Schematic of qPCR primer positions. Coloured trianges show the position and orientation of *ter* sites. The data showed on top are taken from the ChIP-seq experiment of DL4184 strain subjected to a DSB for 3.5 h. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The palindrome position (*lacZ*), *dif*, *terA*, *terB terC* sites and the position of qPCR primers are indicated. (B) Relative RecA enrichment around *ter* sites in *dif*⁻ strains. RecA enrichment at *ter* sites is normalised to the RecA enrichment at the *hycG* site. Strains used were *pal*⁻ *dif*⁻ (DL5812) and *pal*⁺ *dif*⁻ (DL5786).

5.6.3 An ATPase mutation of FtsK leads to a larger distribution of DSBs in the terminus compared to *xerCD/dif* mutants

FtsK is a DNA translocase that activates the chromosome dimer resolution pathway. FtsK is an essential protein that is localised at the septum. It moves dsDNA in an ATP-dependent manner until it aligns two *dif* sites together and then interacts with XerD recombinase to stimulate XerCD recombination at *dif*. To check if an *ftsK* mutant shows the same RecA enrichment pattern as *xerCD*/*dif* mutants, I constructed an FtsK^{K997A} mutant. It encodes a single point mutation in the Walker A motif of the FtsK C-terminal domain that results in the loss of ATPase activity and results in inability to translocate DNA (Bigot *et al*. 2004). I used P1 transduction to co-transduce this mutation with a chloramphenicol resistance gene into the *ftsK* gene of *pal*$^{+}$ (DL4184) and *pal*$^{-}$ (DL4201) strains. I, also, introduced this mutation in *dif* mutants (DL5786 and DL5812). ChIP-seq experiments were performed as described in section 5.6.1.

These ChIP-seq experiments revealed that RecA binding was more distributed across the genome in FtsK^{K997A} mutants (~700 kb) than in single *dif* mutants (~100 kb) (Figure 5.21). Moreover, the pattern of RecA enrichment in a double FtsK^{K997A} *dif*$^{-}$ mutant was the same as in a single FtsK^{K997A} mutant (Figure 5.21). Also, the peak of RecA enrichment between the *dif* site and *terC* was more pronounced, which suggests that DSBs were distributed more randomly in *ftsK* mutants and are not primarily located at the *dif* site. This suggests that DNA translocation by FtsK is needed to concentrate the DSBs around *dif*. Importantly, there was no difference in RecA enrichment pattern between *pal*$^{+}$ and *pal*$^{-}$ strains, suggesting that DSBs in *lacZ* did not contribute to the RecA distribution in the terminus in these mutants.

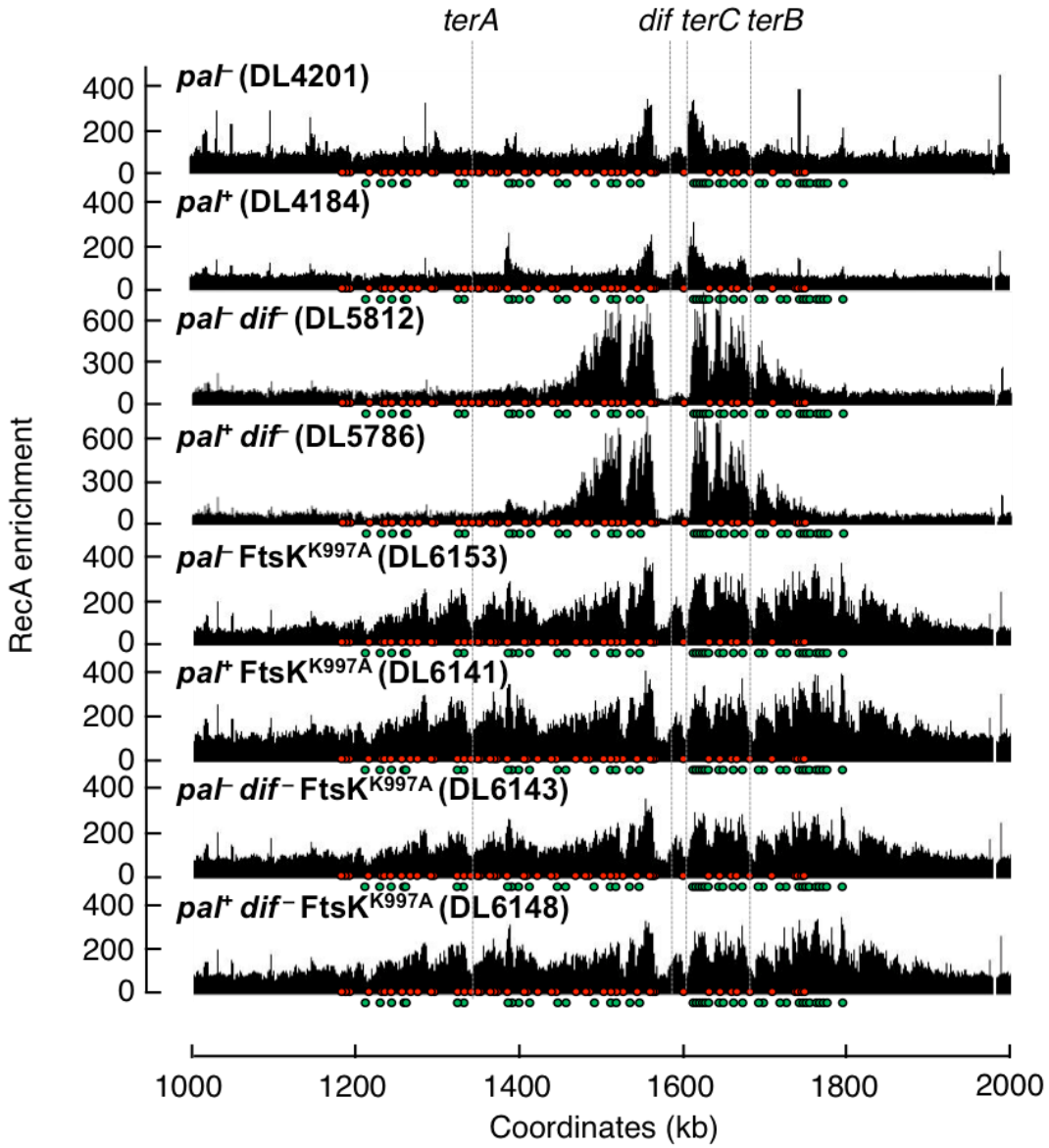


Figure 5.21. Comparison of RecA enrichment in the terminus region of *ftsK* mutants.

ChIP-seq analyses revealed that RecA binding in the terminus is more distributed in FtsK-ATPase mutants compared to *dif* mutants. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal*$^-$), DL4184 (*pal*$^+$), DL5812 (*pal*$^-$ *dif*$^-$), DL5786 (*pal*$^+$ *dif*$^-$), DL6153 (*pal*$^-$ FtsK^{K997A}), DL6141 (*pal*$^+$ FtsK^{K997A}), DL6143 (*pal*$^-$ *dif*$^-$ FtsK^{K997A}) and DL6148 (*pal*$^+$ *dif*$^-$ FtsK^{K997A}).

5.7 Discussion

This chapter describes the role of the chromosome dimer resolution system in the appearance of a DSB around the *dif* site. Initially, the XerCD/*dif* recombination system was chosen as a potential candidate to test the cause of DSBs in the terminus region because of higher amount of DSBs detected around *dif* in the strains that were subjected to a DSB in *lacZ*, especially when grown in M9 minimal medium (Figure 1.13) (Cockram *et al*. 2015). It is known that the amount of chromosome dimers increases following homologous recombination elsewhere in the chromosome (Barre *et al*. 2001). In this chapter, I have investigated whether incomplete chromosome dimer resolution could be implicated in the appearance of DSBs in the terminus region.

To answer this question, I constructed *xerCD*/*dif*/*ftsK* mutants and performed a set of experiments that included growth studies and viability tests, 2D gels, WGS, RecA-ChIP-seq and RecA-ChIP-qPCR.

Initially, for the 2D gel and WGS experiments I chose 1 hour of DSB induction in *lacZ* because it is an arbitrary length of induction time in our laboratory. For the mutants, 2D gels showed ~2 times more DNA detected at *terB* and *terA* sites than in wild type strains. Accordingly, WGS experiment showed more DNA between *terA* and *terB* sites in *dif* mutants compared to wild type strains. Also, WGS detected a loss of DNA at *terC* and, accordingly, 2D gels showed a decreased amount of paused replication forks at *terC* in *xerCD*/*dif* mutants compared to wild type strain. Moreover, from the same 2D gels and WGS data, I can conclude that the over-replication that is induced by the presence of a palindrome in *lacZ* is the same in wild type and in the dimer resolution mutant strains.

At the same time, the pattern of RecA binding around *dif* in *xerCD*/*dif* mutants was more distributed than in wild type strains. However, the quantified

amount of RecA bound to the region around *dif* was not increased in *pal*$^{+}$ strains compared to *pal*$^{-}$ strains in these mutants. This result suggests that the higher amount of DSBs in the terminus of *xerCD*/*dif* mutants does not correlate with the amount of excess of DNA detected in this region. This means that *xerCD*/*dif* DSBs do not induce over-replication in the terminus but that the presence of the DSB in *lacZ* does.

Overall, these observations suggest that there might be multiple causes of DSBs in the terminus. These experiments could not answer whether the failure of the XerCD/*dif* system causes the initial DSBs observed in the terminus. However, these experiments showed the importance of the XerCD/*dif* system in preventing of the formation of more DSBs in the terminus region of the *E. coli* chromosome. The initial break might be caused by incomplete chromosome dimer resolution by XerCD, which would lead to the DSBs around the *dif* site. However, in absence of the *dif* site, XerCD cannot bind to the chromosome and dimers cannot be resolved, therefore, the DNA is broken by the closing septum and the DSBs are distributed around the terminus. This hypothesis is confirmed by FtsK^{K997A} ChIP-seq experiments where RecA distribution in the terminus is much wider due to the inability of FtsK to translocate DNA and to place the *dif* region at the septum.

Additionally, this chapter shows that a higher RecA enrichment is observed at *ter* sites when DSBs in *lacZ* are induced for a longer time (3.5 h). This observation correlates with the detected over-replication in the terminus, where more DNA is observed in the terminus region of a strain that is subjected to DSBs in *lacZ* for a longer period of time (3.5 h). This might suggest that the observed over-replication in the terminus in both wild type and *xerCD*/*dif* mutants is caused by breaks at *ter* sites rather than by DSBs at the *dif*

site. The further explanation of this phenomenon is discussed in detail in Chapter 8.

Chapter VI

6. Is TopoIII or TopoIV involved in the formation of DSBs in the terminus region?

6.1 Introduction

Following the results from the previous chapters, I continued to investigate the causes of DSBs at *dif*. My next question was whether topoisomerases that act in the terminus region could be involved in the appearance of these DSBs. Two out of four topoisomerases in *E. coli* play a role in the terminus – TopoIII and TopoIV (Champoux 2001). TopoIII is a type I topoisomerase that relaxes supercoiled DNA produced during DNA replication, by cleaving, passaging and rejoining one strand of the DNA helix (Champoux 2001). TopoIV is a type II topoisomerase, which releases the supercoiling by cleaving both strands of the DNA duplex (Kato, Nishimura, Imamura, Niki, Hiraga 1990). Both of these topoisomerases are involved in chromosome catenation and decatenation (Harmon, Brockman, and Kowalczykowski 2003; Harmon, DiGate, and Kowalczykowski 1999; Seol *et al.* 2013; Suski and Marians 2008; Perez-Cheeks *et al.* 2012; Deibler, Rahmati 2001; Adams *et al.* 1992). TopoIV is the main decatenase, which is directly recruited by the FtsK protein to the *dif* region (El Sayyed *et al*. 2016; Espeli, Lee, and Marians 2003; Bigot and Marians 2010; Hojgaard *et al*. 1999). It is also known that the XerCD/*dif* system can decatenate chromosomes in the absence of TopoIV (Grainge *et al*. 2007; Shimokawa *et al*. 2013). Therefore, it would be interesting to test if TopoIII and/or TopoIV are involved together with the XerCD/*dif* system in DSB formation in the terminus. Importantly, TopoIV is an essential protein which consists of two subunits: a catalytic subunit, encoded by *parC*, and an ATP-binding subunit, encoded by *parE* (Kato, Nishimura, Imamura, Niki, Hiraga 1990). On the contrary, TopoIII is not essential and is encoded by the single *topB* gene.

In this chapter, I will determine if these topoisomerases could be a cause of DSBs at the *dif* site. I will also test if a DSB in *lacZ* decreases the viability of TopoIII and TopoIV mutants.

6.2 DSBs in *lacZ* had no effect on the viability of TopoIV mutants

Because TopoIV is essential in *E. coli*, I used temperature-sensitive mutations of each of the subunits, ParE and ParC. The permissive temperature for TopoIVts mutants is 30°C and the restrictive temperature is 42°C. To introduce the mutations in the DL4201 (*pal*$^{-}$) and DL4184 (*pal*$^{+}$) strains, I could not use the PMGR method because in order to integrate the plasmid DNA into the chromosome, I would have to grow the transformants at 42°C. Instead, I used the P1 transduction method, which is effective at 30°C. For this purpose, I requested strains containing the *parEts* (W3110parE10) or the *parCts* (C600parC1215) mutations from the laboratory of Professor Lynn Zechiedrich (Zechiedrich and Cozzarelli 1995) and strains containing a tetracycline or a kanamycine antibiotic resistance gene in close proximity of the *parE* (CAG18472 and CAG18559) or the *parC* (CAG18475 and CAG18527) gene from the *Coli* Genetic Stock Center (CGSC) (http://cgsc.biology.yale.edu). First, I introduced either the kanamycin or the tetracycline resistance gene in the strains containing a TopoIV mutation to link the mutation with an antibiotic resistance gene. Second, I used the antibiotic resistance gene to introduce this mutation into cells that express *sbcCD* under the arabinose inducible promoter and contain Chi arrays 1.5 kb either side from the locus in the *lacZ* gene containing or not the 246 bp interrupted palindrome. Then, I tested the viability of these mutants in the presence and absence of DSBs in *lacZ* by growth curves at 42°C and spot tests at 30°C. Strains used for these experiments were DL4201 (*pal*$^{-}$), DL4184 (*pal*$^{+}$), DL5976 (*pal*$^{-}$ *parEts*), DL5975 (*pal*$^{+}$ *parEts*), DL5998 (*pal*$^{-}$ *parCts*) and DL5996 (*pal*$^{+}$ *parCts*).

In the morning, overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 30°C to an OD_{600nm} of 0.2-

0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01, shifted to 42°C to inactivate either ParC or ParE of TopoIV and 0.2% arabinose was added to induce DSBs in *lacZ* (time 0). The OD_{600nm} of these cultures was measured every 30 min for 4.5 hours and the cultures were regularly diluted in fresh prewarmed medium in order to be kept in exponential phase (OD_{600nm} of 0.2-0.3). The growth profiles of the TopoIVts strains containing or not the 246 bp palindrome are shown in Figure 6.1. *parEts* mutants subjected or not to DSBs in *lacZ* showed similar growth profiles with a slower growth than wild type strains after 80 minutes at 42°C and reaching a plateau after 150 minutes. *parCts* mutants carrying or not the palindrome in *lacZ* showed an even slower growth rate than *parEts* mutants and reached a plateau at ~100 minutes. This experiment suggests that the presence of DSBs in *lacZ* does not further affect the growth of TopoIVts mutants and, also, that the mutation in the ParC catalytic subunit of TopoIV affects more cell growth than the mutation in ParE.

To check if the presence of DSBs in *lacZ* affects growth of a double XerCD/*dif* TopoIVts mutant, I constructed *dif⁻ parEts* and *dif⁻ parCts* mutants that contained or not the palindrome in *lacZ* and performed growth curves. The growth conditions were identical to those described above. Graphs shown in Figure 6.2 displayed a further decrease in growth compared to single TopoIVts mutants. In general, double mutants were showing the decrease in the growth almost immediately after the shift to 42°C. The *dif⁻ parCts* mutants were sicker than the *dif⁻ parEts* mutants. The presence of DSBs in *lacZ* did not further affect the growth of double mutants.

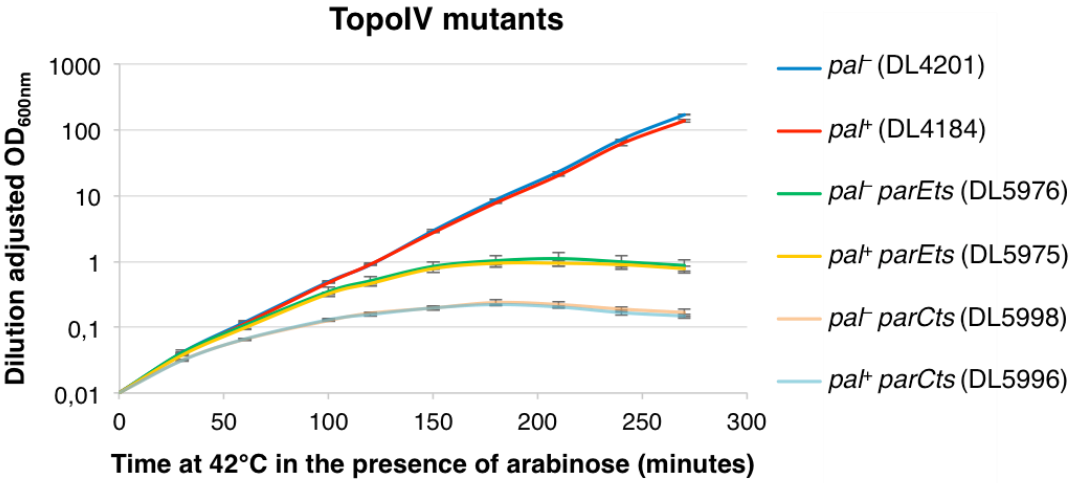


Figure 6.1. DSBs in *lacZ* had no effect on the growth of TopoIV mutants.

ParC inactivation is more detrimental to cell growth than ParE inactivation. Growth profiles of *parEts* and *parCts* mutants of TopoIV subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with 0.2% arabinose to induce SbcCD expression. Strains used were DL4201 (*pal*$^{-}$), DL4184 (*pal*$^{+}$), DL5976 (*pal*$^{-}$ *parEts*), DL5975 (*pal*$^{+}$ *parEts*), DL5998 (*pal*$^{-}$ *parCts*) and DL5996 (*pal*$^{+}$ *parCts*).

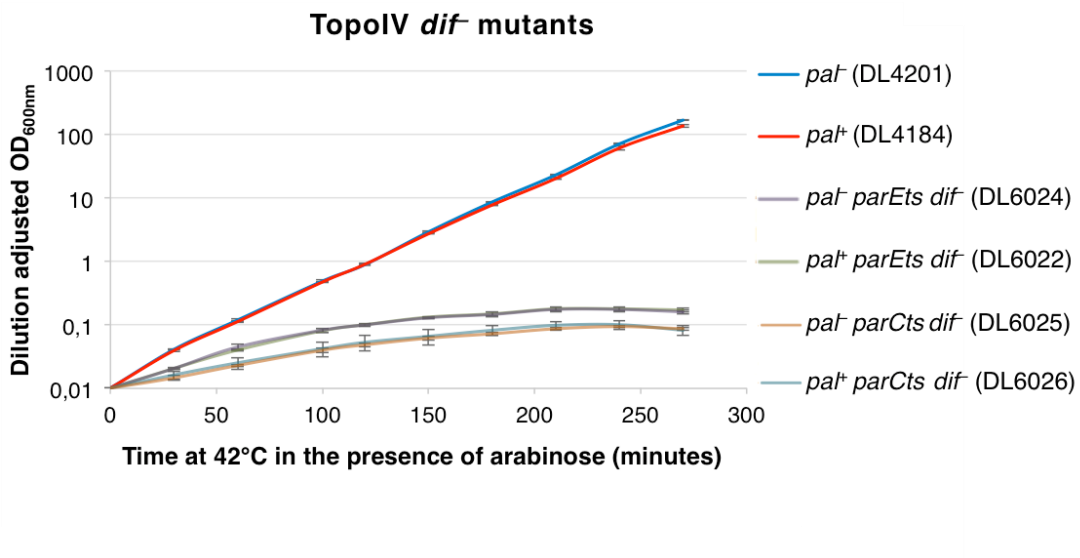


Figure 6.2. Double *dif* TopoIVts mutants display very slow cell growth and no difference between the strains that were subjected to DSBs in *lacZ* or not.

Growth profiles of *dif⁻ parEts* and *dif⁻ parCts* mutants subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with 0.2% arabinose to induce SbcCD expression. Strains used were DL4201 (*pal⁻*), DL4184 (*pal⁺*), DL6024 (*pal⁻ dif⁻ parEts*), DL6022 (*pal⁺ dif⁻ parEts*), DL6025 (*pal⁻ dif⁻ parCts*) and DL6026 (*pal⁺ dif⁻ parCts*).

To further check the viability of TopoIVts mutants subjected or not to DSBs, I performed spot tests at the permissive temperature. Overnight cultures grown at 30°C were diluted to an OD_{600nm} of 0.1. Then, 10-fold serial dilutions were spotted on LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Plates were incubated at 30°C overnight (Figure 6.3). These spot tests showed no reduction of viability of *parEts* mutants compared to wild type strains in presence or absence of DSBs. Double *dif⁻ parEts* mutants were similar to single *dif⁻* mutants and no difference in colony count between the strains subjected or not to a DSB was observed. This mutant grew as smaller colonies compared to single *dif* mutants, which suggests decreased viability of double mutants. However, in the *parCts* mutant subjected to DSBs overnight (on the arabinose plate), a 10-fold colony count reduction was observed. Also, this mutant grew as smaller colonies when subjected to DSBs. Surprisingly, the double *dif⁻ parCts* mutant when subjected to DSBs in *lacZ* showed significant reduction in viability compared to the double *dif⁻ parCts* mutant that was not subjected to a DSB. Overall, these experiments suggest that DSBs in *lacZ* decrease the viability of a strain containing a temperature-sensitive mutation in the catalytic subunit of TopoIV (ParC) even at the permissive temperature (30°C).

These observations indicate that the ParC subunit is partially inactive at the permissive temperature. Therefore, I decided to choose the *parEts* mutants that were not subjected to DSBs in *lacZ*, as they showed no difference in viability compared to the strains that were subjected to DSBs, for all experiments in this chapter (Figure 6.4).

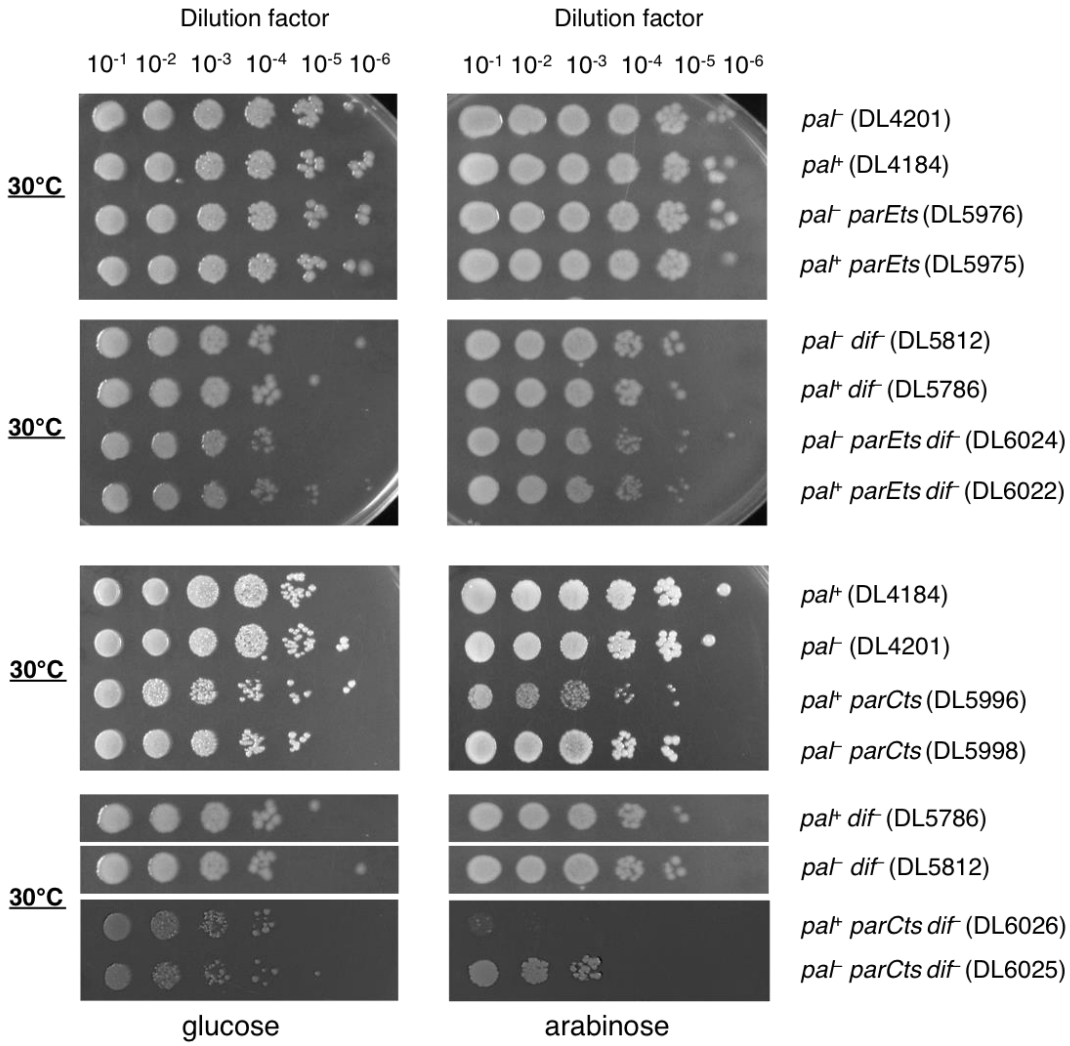


Figure 6.3. The viability of *parCts* mutants is affected when subjected to DSBs at 30°C.

Spot tests of 10-fold serially diluted overnight cultures were spotted onto LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Strains used were DL4201 (*pal*$^{-}$), DL4184 (*pal*$^{+}$), DL5976 (*pal*$^{-}$ *parEts*), DL5975 (*pal*$^{+}$ *parEts*), DL6024 (*pal*$^{-}$ *dif*$^{-}$ *parEts*), DL6022 (*pal*$^{+}$ *dif*$^{-}$ *parEts*), DL5996 (*pal*$^{+}$ *parCts*), DL5998 (*pal*$^{-}$ *parCts*), DL6026 (*pal*$^{+}$ *dif*$^{-}$ *parCts*) and DL6025 (*pal*$^{-}$ *dif*$^{-}$ *parCts*).

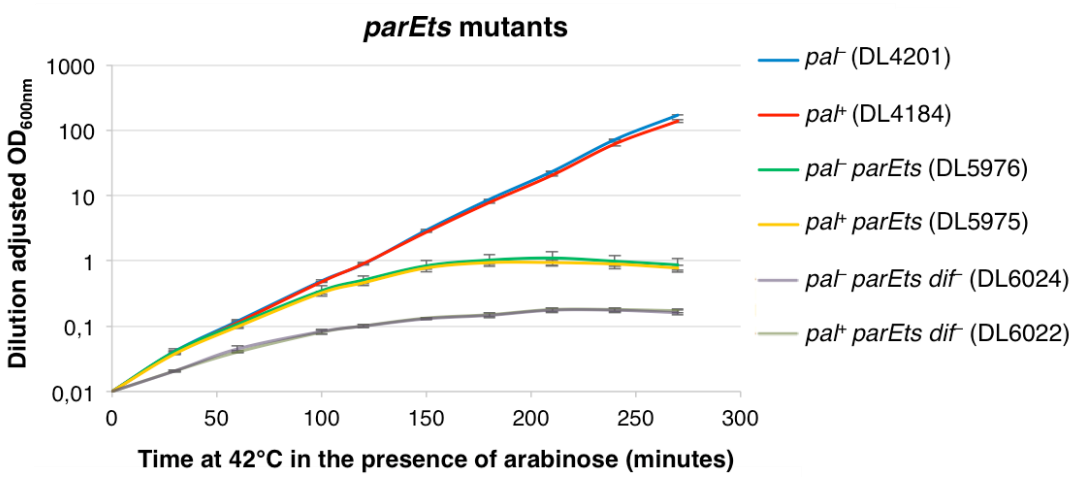


Figure 6.4. Combined growth profiles of *parEts* mutants.

Growth profiles of *parEts* and *dif⁻ parEts* mutants subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with 0.2% arabinose to induce SbcCD expression. Strains used were DL4201 (*pal⁻*), DL4184 (*pal⁺*), DL5976 (*pal⁻ parEts*), DL5975 (*pal⁺ parEts*), DL6024 (*pal⁻ dif⁻ parEts*) and DL6022 (*pal⁺ dif⁻ parEts*).

6.3 TopoIV contributes to the formation of DSBs in the terminus

6.3.1 TopoIVts mutant shows similar replication profile compared to the wild type strain

Whole genome sequencing was performed to visualise if a *parEts* mutant displayed additional DNA over-replication in the terminus. Strains used for this experiment were DL4201 (*pal*$^-$) and DL5976 (*pal*$^-$ *parEts*). In the morning, overnight cultures grown at 30°C were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 30°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and kept at 30°C until they reached an OD_{600nm} of 0.2. Then, the cultures were divided in two flasks. One flask stayed at 30°C when the other flask was incubated at 42°C for 30 minutes to inactivate ParE. I chose 30 minutes at 42°C because it was enough for *parEts* mutant to show accumulation of catenated chromosomes (Grainge *et al*. 2007). The DNA for WGS was isolated using the Wizard kit and purified using the Zymo kit as described in Chapter 2. Edinburgh Genomics performed library preparations and WGS using an Illumina HiSeq 2500 platform. Paired-end raw data obtained were analysed as described in Chapter 2. Data from exponentially growing cells were normalised to data from the stationary phase *pal*$^-$ (DL4201) strain and individual graphs for each biological repeat were carried out (Figure 6.5). Replication profiles of all these strains were V-shaped as expected. The DNA of all 7 *rrn* operons was partially recovered with an even lower recovery level in the *parEts* mutant grown at 42°C (Figure 6.6). Importantly, *parEts* cells grown at 30°C or 42°C showed similar replication profiles in the terminus to wild type profiles. In the combined of all strains shown in Figure 6.6, it is noticeable that biological replicates of *parEts* mutant grown at 30°C are

different from each other. This is due to a more variable range of reads in BR1 (Figure 6.5).

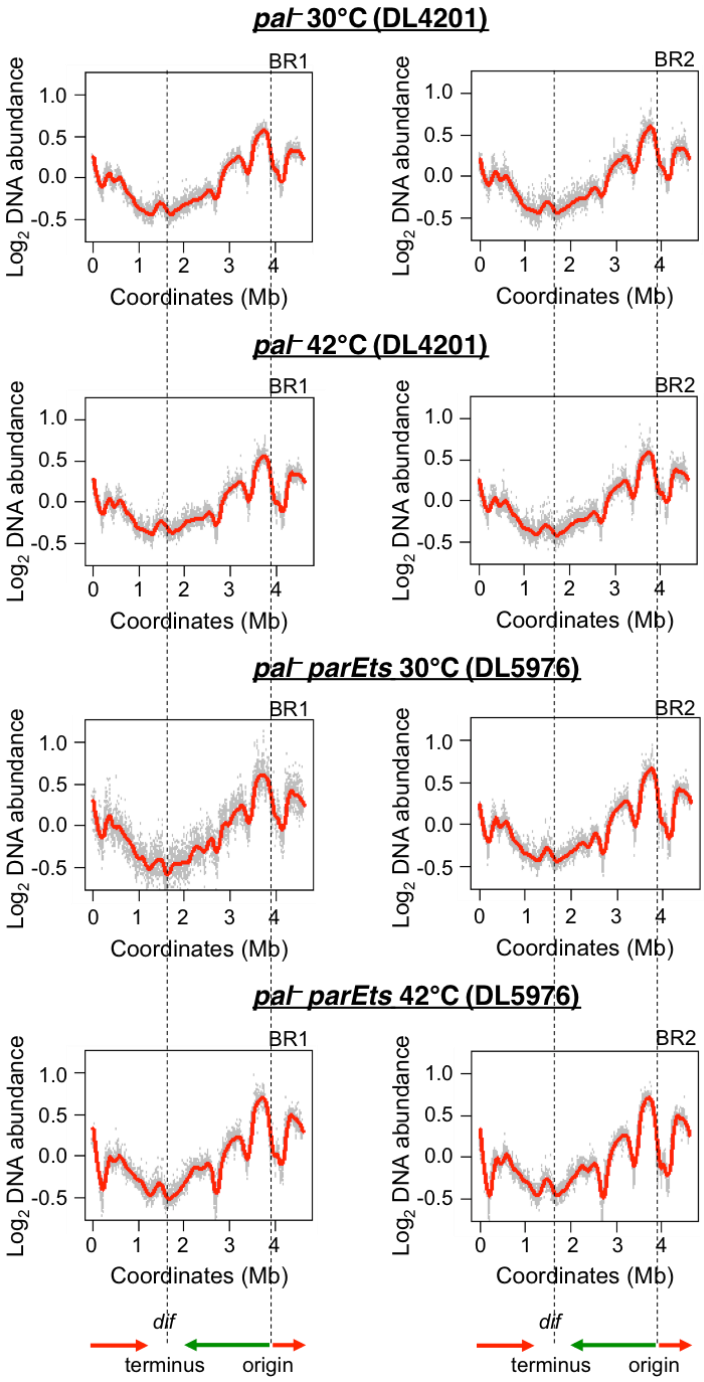


Figure 6.5. TopoIVts mutant shows similar replication profile compared to the wild type strain.

Replication profiles of exponentially growing cultures were normalised to data from the stationary phase *pal*⁻ (DL4201) strain. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The position of the origin and *dif* (terminus) is indicated using dash lines. BR1 is biological repeat 1; BR2 is biological repeat 2. Strains used were DL4201 (*pal*⁻) and DL5976 (*pal*⁻ *parEts*) grown at 30°C and 42°C.

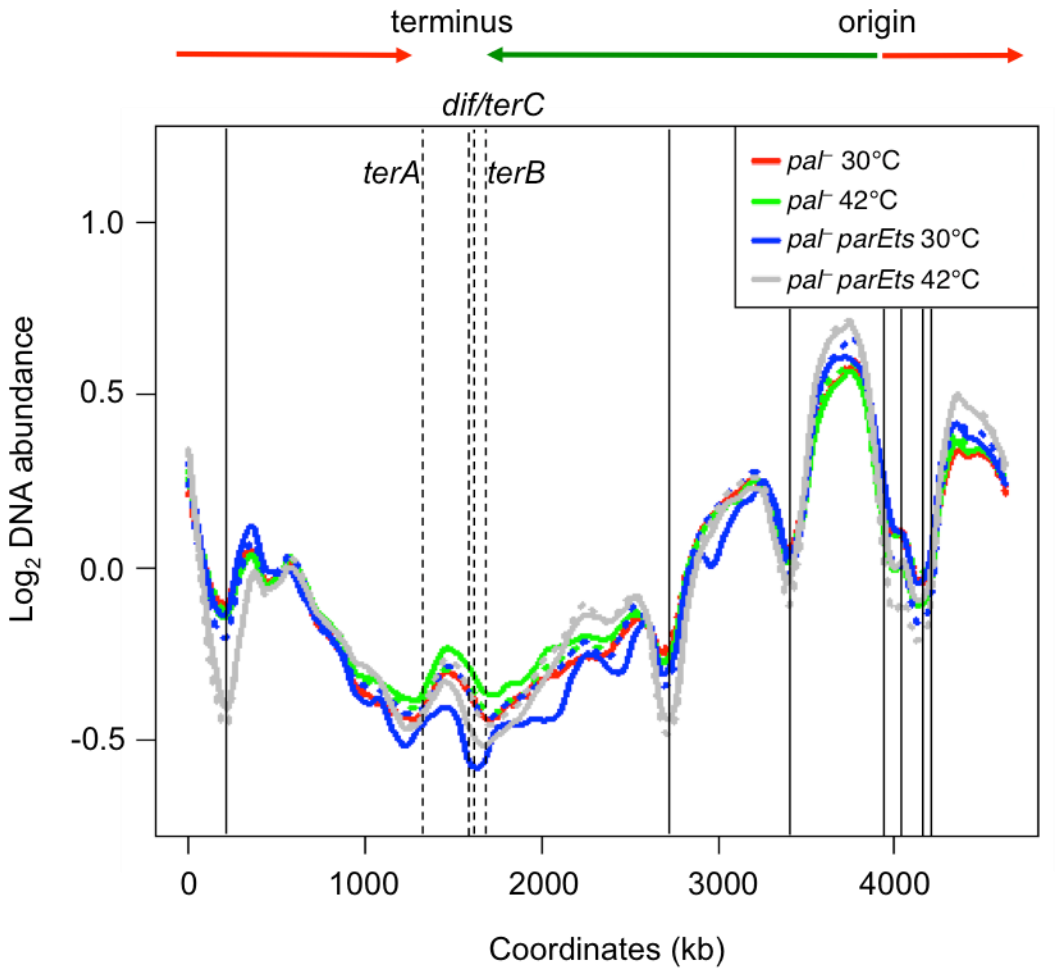


Figure 6.6. Combined replication profiles of all sequenced strains.

Replication profiles of exponentially growing cultures of *parEts* and wild type strains not containing a palindrome. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The position of *dif, terA, terB* and *terC* sites is indicated using dash lines. The position of *rrn* operons is indicated using solid lines. Biological repeats (BR1 and BR2) of the same strain are shown using dashed/solid lines of the same colour. Strains used were DL4201 (*pal*⁻) grown at 30°C in red, DL4201 (*pal*⁻) grown at 42°C in green, DL5976 (*pal*⁻ *parEts*) grown at 30°C in blue and DL5976 (*pal*⁻ *parEts*) grown at 42°C in grey.

To better visualise a potential difference in the amount of DNA sequence around the *dif* region in a strain mutated for TopoIV, ratios of reads of the *pal⁻ parEts* (DL5976) strain over the *pal⁻* (DL4201) strain were carried out (Figure 6.7). To calculate these ratios, I used the average of the two biological repeats presented in Figure 6.6. Figure 6.7 A represents the ratio of reads of the *pal⁻ parEts* (DL5976) strain grown at 42°C over the *pal⁻ parEts* (DL5976) strain grown at 30°C. This ratio shows that more DNA in *rrn* genes was lost in the strain that grew at 42°C. Also, more DNA was observed around the terminus. These increases in one location and decreases of DNA sequences in the other might be due to the normalisation. At the same time, the terminus regions of this strain grown at different temperatures looked similar. Figures 6.7 B and 6.7 C show the ratio of reads of the *pal⁻ parEts* (DL5976) strain grown at 30°C over the *pal⁻* (DL4201) strain grown at 30°C and the *pal⁻ parEts* (DL5976) strain grown at 42°C over the *pal⁻* (DL4201) strain grown at 42°C, respectively. The ratio in Figure 6.7 B looks stable showing no difference in the DNA levels of the two strains at permissive temperature. Finally, at non-permissive temperature (Figure 6.7 C), the TopoIVts mutant contains less DNA at the *rrn* gene loci and about the same or less amount of DNA in the terminus. These experiments demonstrate that the TopoIVts mutation does not introduce additional over-replication between *terA* and *terB*.

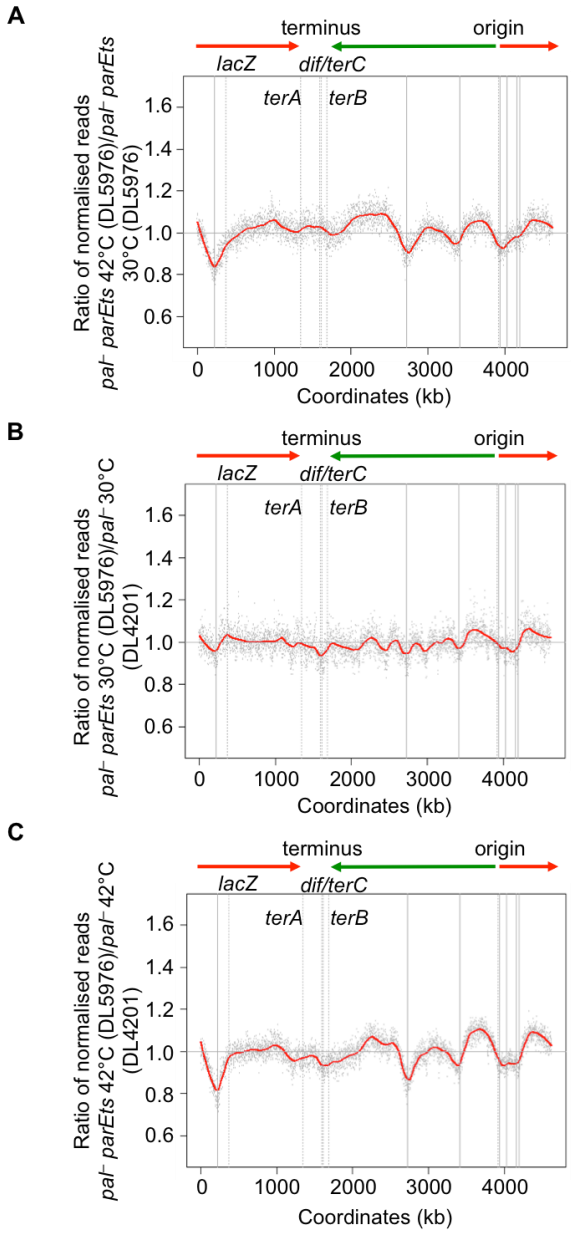


Figure 6.7. Ratio of replication profiles of *parEts* strains over wild type strains.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of (A) *pal⁻ parEts* (DL5976) strain at 42°C over *pal⁻ parEts* (DL5976) strain at 30°C; (B) *pal⁻ parEts* (DL5976) strain at 30°C over *pal⁻* (DL4201) strain at 30°C and (C) *pal⁻ parEts* (DL5976) strain at 42°C over *pal⁻* (DL4201) strain at 42°C. Grey dots represent reads, the red line is the ratio of the reads from DL5976 at 42°C (A); DL5976 at 30°C (B) or DL5976 at 42°C (C). The horizontal grey line represents DL5976 at 30°C (A); DL4201 at 30°C (B) or DL4201 at 42°C (C). The direction of replication is shown using green and red arrows. The position of *lacZ*, *terA*, *terB*, *terC* and *dif* sites is indicated using dash lines. The position of *rrn* operons is indicated using solid lines. Strains used were DL4201 (*pal⁻*) and DL5976 (*pal⁻ parEts*).

6.3.2 TopoIVts mutants show wild type level of RecA enrichment in the terminus

To determine the RecA enrichment patterns in the terminus of *parEts* mutants and to confirm that there is no effect of the presence of the palindrome in *lacZ* of these strains, I performed RecA-ChIP-seq experiments to visualise RecA enrichment in the *E. coli* chromosome. For this experiment, I used DL4201 (*pal*$^{-}$), DL5976 (*pal*$^{-}$ *parEts*), DL4184 (*pal*$^{+}$) and DL5975 (*pal*$^{+}$ *parEts*) strains. As in Chapter 5, I induced SbcCD expression in these strains for 3.5 hours. Overnight cultures grown at 30°C were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 30°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01, 0.2% arabinose was added to induce SbcCD expression and the cultures were kept at 30°C for 3 hours of exponential growth. The OD_{600nm} of the cultures was regularly monitored and the cells were diluted in fresh prewarmed medium to keep them at an OD_{600nm} of 0.2-0.25. Then, the cultures were divided in two flasks. One flask was incubated at 30°C when the other flask was incubated at 42°C for 30 minutes to inactivate ParE. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform by Edinburgh Genomics. Finally, raw ChIP-seq data were analysed as described in Chapter 2.

If TopoIV alone was responsible for the formation of DSBs in the terminus, I would have expected the disappearance of the observed DSBs around the *dif* site. Instead, the pattern of RecA enrichment in TopoIVts mutants was the same as in wild type strains (Figure 6.8). Moreover, strains that contain the palindrome in *lacZ* showed the same RecA binding pattern as strains without

the palindrome, except at *ter* sites, suggesting that DSBs in *lacZ* did not change the pattern of RecA enrichment around *dif*. This suggests that TopoIV alone is not the cause of DSBs around *dif*.

To confirm that TopoIVts mutation does not change the relative level of RecA binding in the terminus, I performed ChIP-qPCR on the same DNA as used for ChIP-seq from the DL4201 (*pal*$^{-}$) and DL5976 (*pal*$^{-}$ *parEts*) strains described above (Figure 6.9). For this experiment, I used 4 pairs of qPCR primers: 2 pairs of primers after Chi sites – Dif1 and Dif2 and 2 pairs of primers before Chi sites – bef.chi1 and bef.chi3, one on each side of *dif* (Figure 6.9 A). I measured the relative quantification of RecA enrichment levels of each fragment, which was calculated as the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and always shows background level of RecA binding (Cockram *et al*. 2015). As expected from the ChIP-seq experiments results above, these ChIP-qPCRs showed a similar level of RecA enrichment in the *parEts* mutants and the wild type strains. Additionally, these ChIP-qPCRs revealed that the levels of RecA binding in the terminus were slightly decreased in all tested strains grown at 42°C (from ~3- to ~2.5-fold) (Figure 6.9 B). This observation shows that the level of RecA enrichment in the terminus is affected by the temperature.

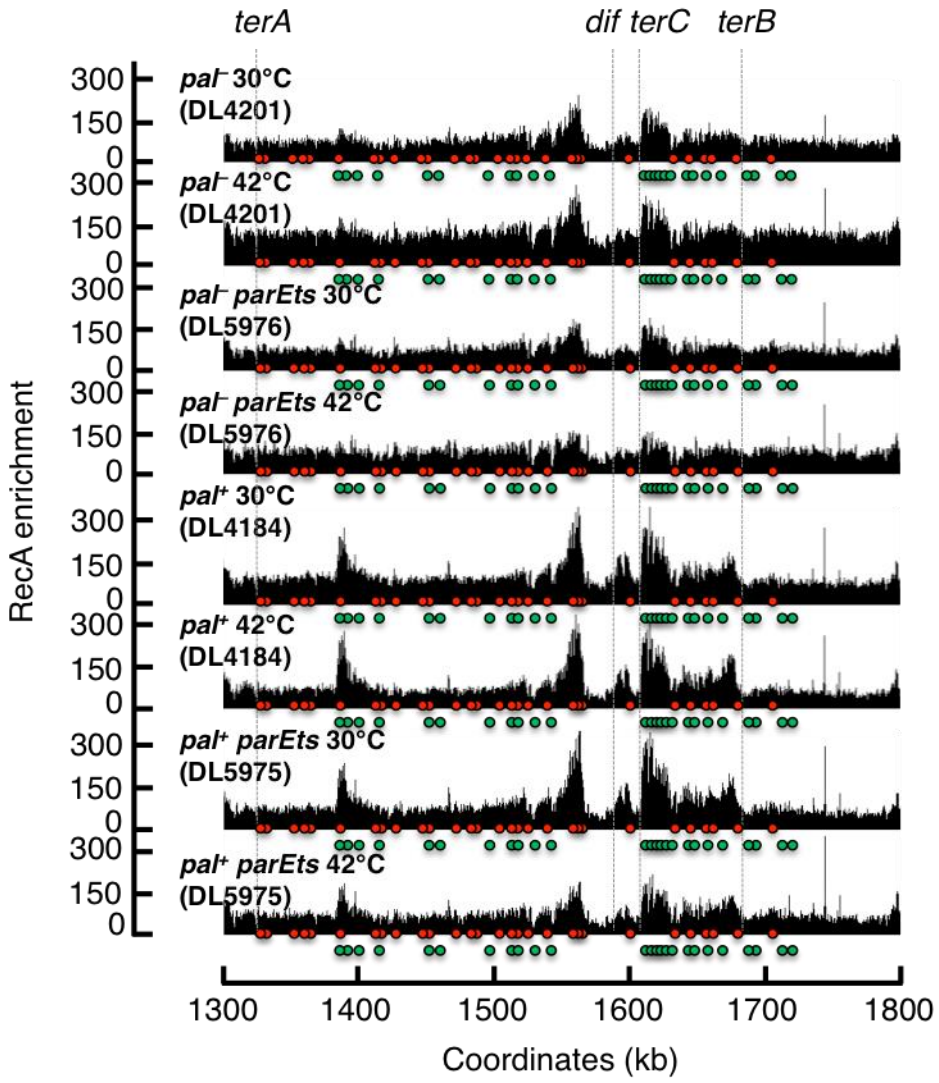


Figure 6.8. Comparison of RecA enrichment in the terminus region of *parEts* mutants.

ChIP-seq analyses revealed similar patterns of RecA enrichment in the terminus in *parEts* mutants and in wild type strains. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal*$^-$) grown at 30°C and 42°C, DL5976 (*pal*$^-$ *parEts*) grown at 30°C and 42°C, DL4184 (*pal*$^+$) grown at 30°C and 42°C and DL5975 (*pal*$^+$ *parEts*) grown at 30°C and 42°C.

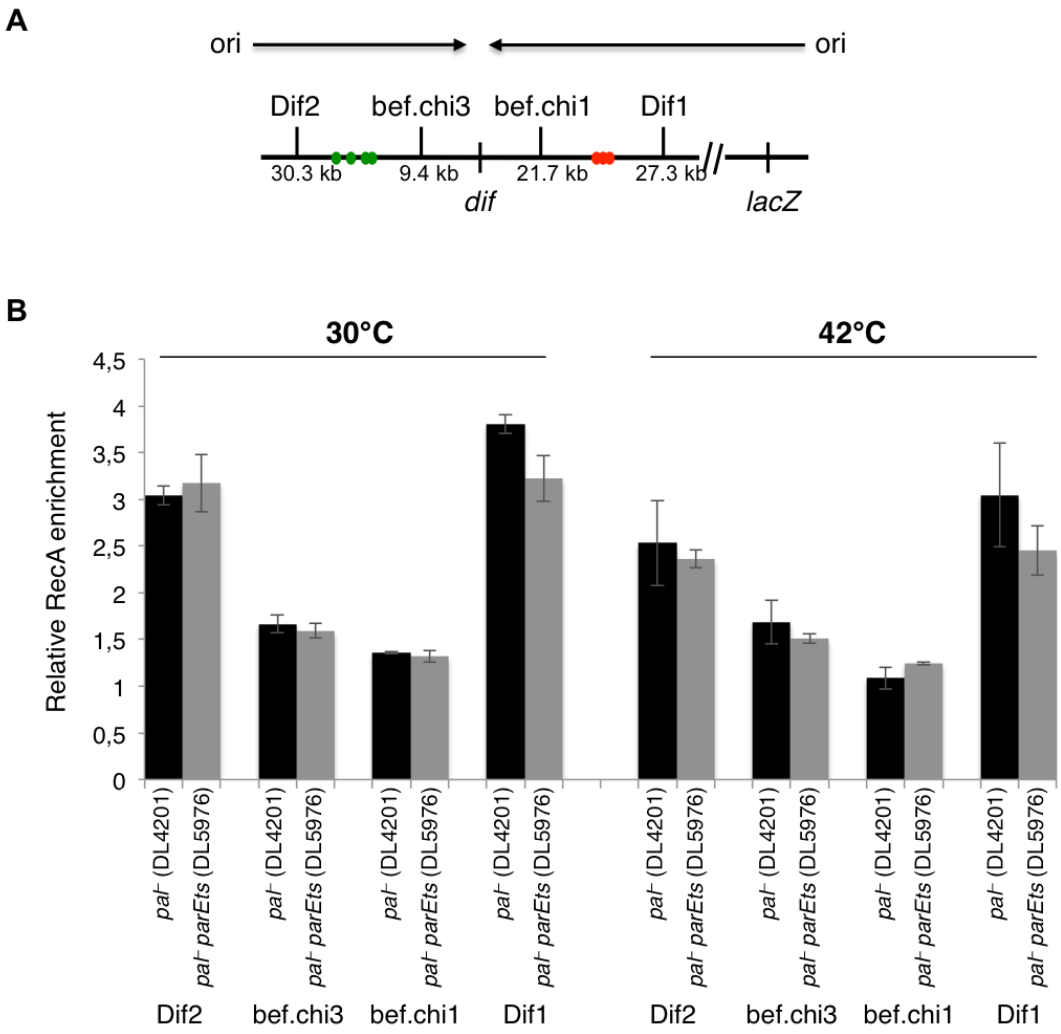


Figure 6.9. Quantification of RecA enrichment in the *dif* region of *pal*⁻ *parEts* mutant.

ChIP-qPCR revealed that the levels of RecA binding were the same in *pal*⁻ *parEts* mutant and in wild type strains. (A) Schematic of qPCR primer positions. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *lacZ* and positions of qPCR primers are indicated. Replication direction is shown using black arrows. (B) Relative RecA enrichment in the terminus of *pal*⁻ and *pal*⁻ *parEts* strains. RecA enrichment in the terminus was normalised to the level of RecA enrichment at the *hycG* site. Strains used were DL4201 (*pal*⁻) grown at 30°C and 42°C and DL5976 (*pal*⁻ *parEts*) grown at 30°C and 42°C.

6.3.3 The loss of DNA at *dif* is partially recovered in *dif⁻ parEts* double mutant

Whole genome sequencing was performed to visualise if double *dif⁻ parEts* mutants display the same replication profile as the single *dif⁻* mutants discussed in Chapter 5. These *dif⁻* mutants showed additional DNA in the terminus with a peak that was shifted to *terA* and a loss of DNA at *dif*. Strains used for this experiment were DL4201 (*pal⁻*) and DL6024 (*pal⁻ dif⁻ parEts*). In the morning, overnight cultures grown at 30°C were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 30°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and kept at 30°C until they reached an OD_{600nm} of 0.2. Afterwards, the cultures were shifted to 42°C for 30 minutes to inactivate ParE. The DNA for WGS was isolated using a Wizard kit and purified using a Zymo kit as described in Chapter 2. Edinburgh Genomics performed library preparations and WGS using an Illumina HiSeq 2500 platform. Obtained paired-end raw data were analysed as described in Chapter 2. Data from exponentially growing cells were normalised to data from the stationary phase *pal⁻* (DL4201) strain and individual graphs for each biological repeat were carried out (Figure 6.10). As expected, the overall replication profiles of all these strains were V-shaped with more DNA at the origin and less DNA at the terminus. The DNA of all 7 *rrn* operons was partially recovered (Figure 6.11). The *dif⁻ parEts* mutant grown at 30°C or 42°C showed similar replication profiles in the terminus region with a higher peak of over-replication that was shifted to *terA* and a loss of DNA at *dif*. These replication profiles were similar to the *dif⁻* profiles observed previously (Figure 6.11). Higher amount of DNA closer to *terA* and a larger loss at *dif* were observed in a TopoIVts *dif⁻* double mutant at 30°C and 42°C compared to a *dif⁻* single mutant at 37°C (Figure 6.11). The accumulation of DNA close to *terA* could be caused by an over-replication that

is elevated in a double mutant irrespective of the temperature. Unfortunately, I do not have *dif*⁻ mutant profiles grown at 30°C and 42°C, therefore, I cannot directly compare the double TopoIVts *dif*⁻ mutant to a single *dif*⁻ mutant.

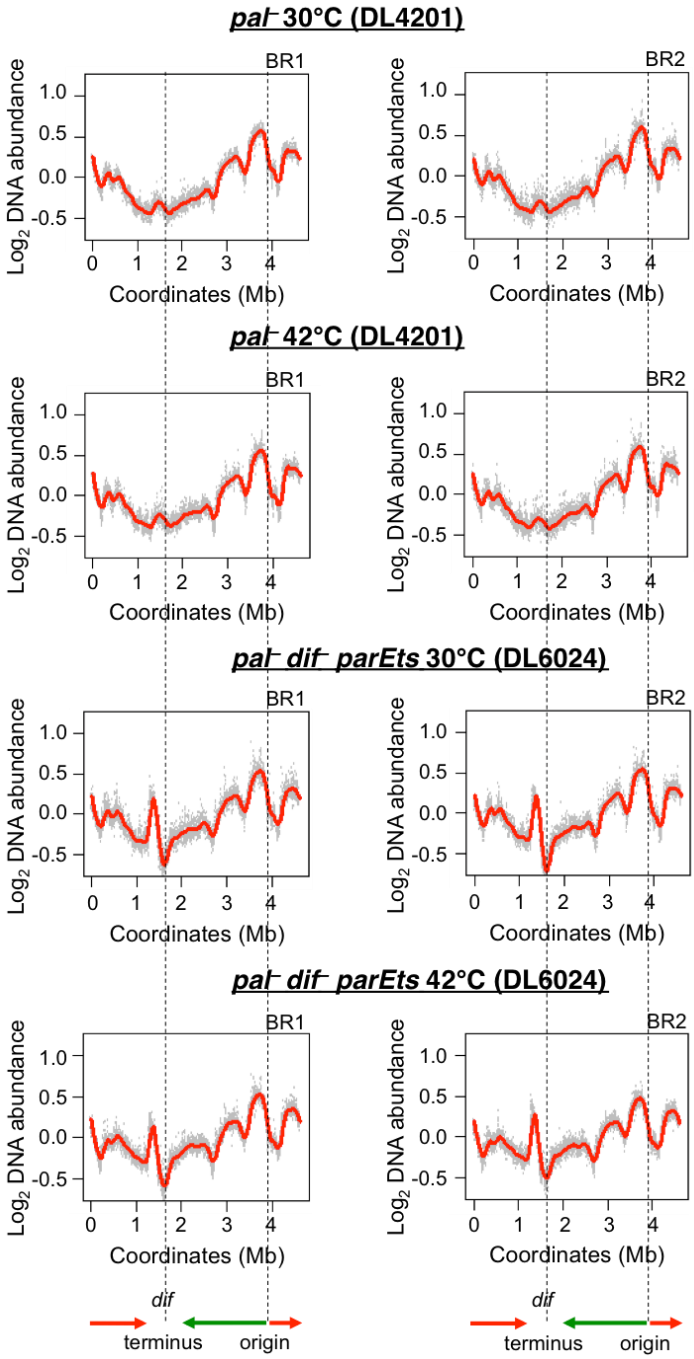


Figure 6.10. WGS profiles of *dif⁻ parEts* mutants showed an increase of the DNA excess in the terminus compared to wild type strains.

Replication profiles of exponentially growing cultures were normalised to data from stationary phase *pal⁻* (DL4201) strain. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The position of the origin and *dif* (terminus) is indicated using dash lines. BR1 is biological repeat 1; BR2 is biological repeat 2. Strains used were DL4201 (*pal⁻*) and DL6024 (*pal⁻ dif⁻ parEts*) grown at 30°C and 42°C.

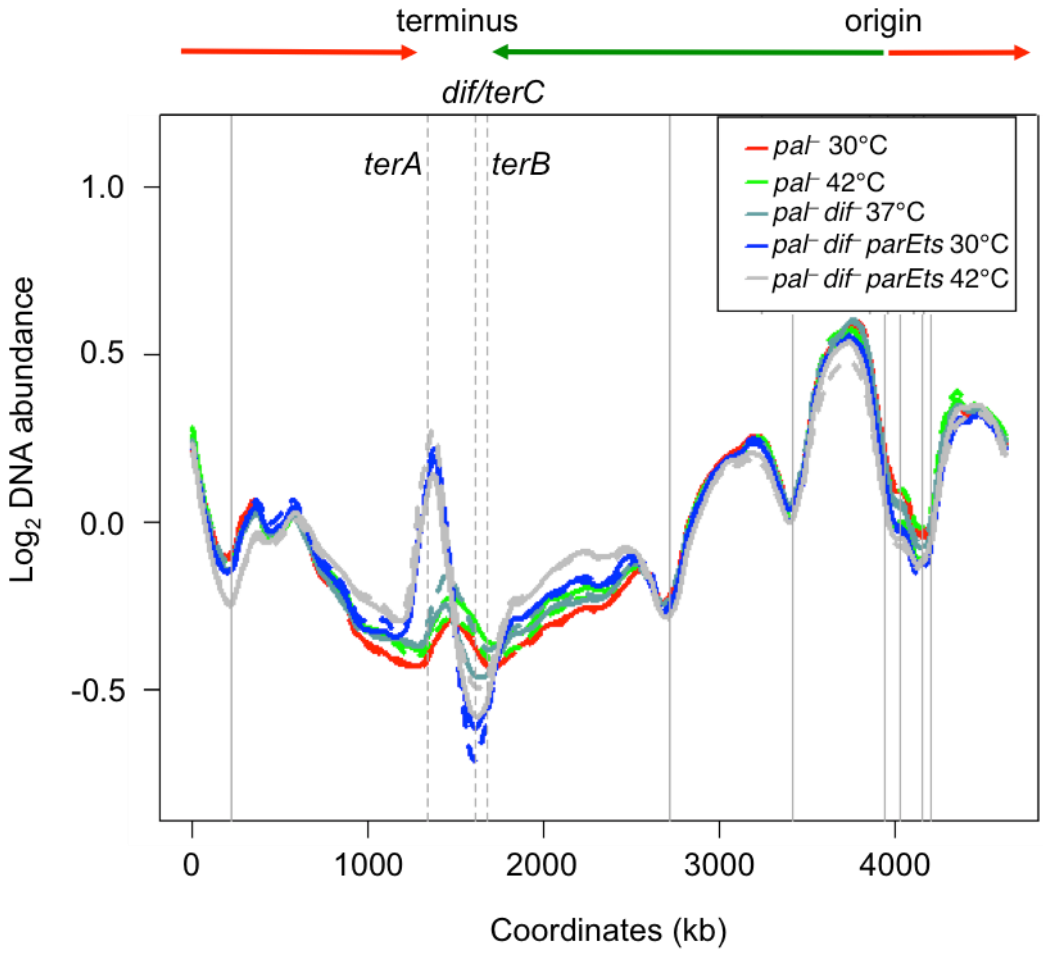


Figure 6.11. The loss of DNA at *dif* is partially recovered in *dif⁻ parEts* double mutant in combined replication profiles of sequenced strains.

Replication profiles of exponentially growing cultures of *dif⁻ parEts, dif⁻* and wild type strains without a palindrome. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of *dif, terA, terB* and *terC* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Biological repeats (BR1 and BR2) of the same strain are shown using dashed/solid lines of the same colour. Strains used were DL4201 (*pal⁻*) grown at 30°C in red, DL4201 (*pal⁻*) grown at 42°C in green, DL5812 (*pal⁻ dif⁻*) grown at 37°C in dark grey, DL6024 (*pal⁻ dif⁻ parEts*) grown at 30°C in blue and DL6024 (*pal⁻ dif⁻ parEts*) grown at 42°C in grey.

To confirm that in the absence of functional TopoIV, cells recover some DNA at *dif*, ratios of reads of the *pal⁻ dif⁻ parEts* (DL6024) strain over the *pal⁻* (DL4201) strain grown at 30°C and 42°C were carried out (Figure 6.12). Figure 6.12 A represents the ratio of reads of *pal⁻ dif⁻ parEts* (DL6024) strain grown at 42°C over *pal⁻ dif⁻ parEts* (DL6024) strain grown at 30°C. This ratio shows that some DNA loss at *dif* was recovered in the strain where TopoIV was inactive (42°C). Also, less DNA was detected in a peak near the *terA* site in this strain at 42°C.

To eliminate possible differences in growth rate, the ratio of reads of the *pal⁻ dif⁻ parEts* (DL6024) strain over the *pal⁻* (DL4201) strain grown at 30°C and the *pal⁻ dif⁻ parEts* (DL6024) strain over the *pal⁻* (DL4201) strain grown at 42°C were carried out (Figures 6.12 B and C, respectively). Overall, these ratios looked similar to each other irrespectively of the temperature. This observation suggests that TopoIV function might be affected even at 30°C.

TopoIV functions primarily at the *dif* site. So, it is possible that in the absence of the *dif* site, *parEts* mutant might be partially inactive even at 30°C, which was visible during the viability tests carried out in *dif parEts* double mutant (Figure 6.3).

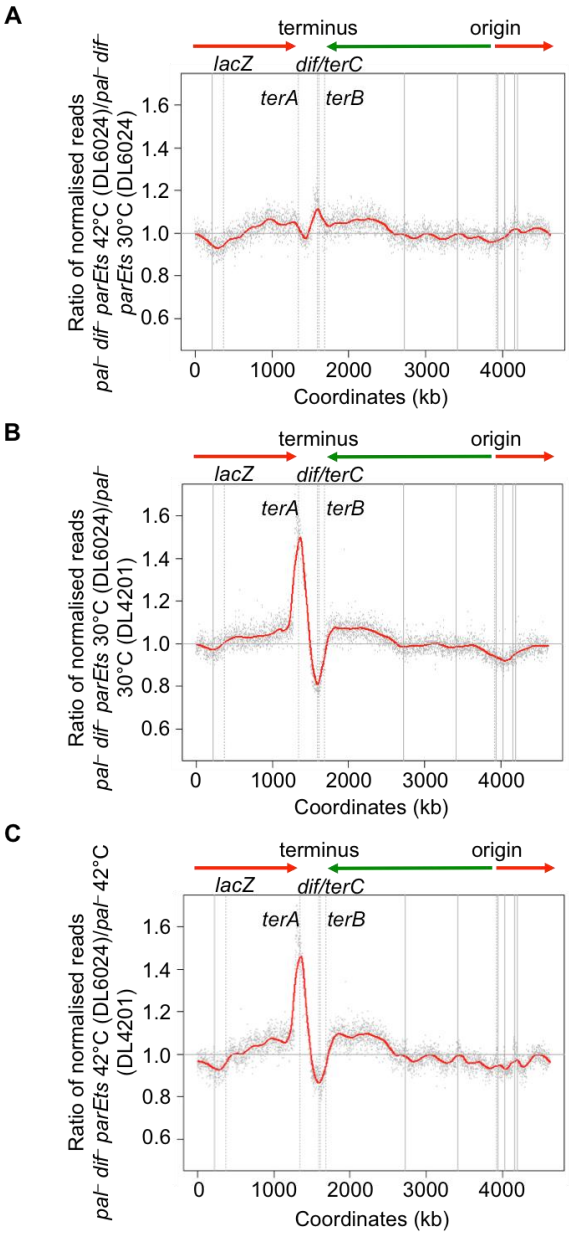


Figure 6.12. Ratio of replication profiles of the *dif⁻ parEts* strain over the wild type strain.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of (A) *pal⁻ dif⁻ parEts* (DL6024) strain at 42°C over *pal⁻ dif⁻ parEts* (DL6024) strain at 30°C; (B) *pal⁻ dif⁻ parEts* (DL6024) strain at 30°C over *pal⁻* (DL4201) strain at 30°C and (C) *pal⁻ dif⁻ parEts* (DL6024) strain at 42°C over *pal⁻* (DL4201) strain at 42°C. Grey dots represent DNA reads, the red line is the ratio of the reads from DL6024 at 42°C (A); DL6024 at 30°C (B) or DL6024 at 42°C (C). The horizontal grey line represents DL6024 at 30°C (A); DL4201 at 30°C (B) or DL4201 at 42°C (C). The direction of replication is shown using green and red arrows. The positions of *lacZ*, *terA*, *terB*, *terC* and *dif* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Strains used were DL4201 (*pal⁻*) and DL6024 (*pal⁻ dif⁻ parEts*).

6.3.4 TopoIV is partially responsible for DSBs around the *dif* site

To determine if the partial recovery of DNA loss at *dif* described in Chapter 6.3.3 correlates with the RecA enrichment patterns in the terminus of *dif⁻ parEts* mutants, I performed a RecA-ChIP-seq experiment to visualise RecA enrichment in the *E. coli* chromosome. For this experiment, I used the DL4201 (*pal⁻*), DL5976 (*pal⁻ parEts*), DL5812 (*pal⁻ dif⁻*) and DL6024 (*pal⁻ dif⁻ parEts*) strains. Overnight cultures grown at 30°C were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 30°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and were kept at 30°C until they reached an OD_{600nm} of 0.2. Afterwards, the cultures were divided in two flasks. One flask was incubated at 30°C when the other flask was incubated at 42°C for 30 minutes to inactivate ParE. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform by Edinburgh Genomics. Finally, raw ChIP-seq data were analysed as described in Chapter 2.

Due to the recovery of DNA loss at *dif* observed in WGS profiles, I expected that the amount of DSBs in the terminus would be lower in the *dif⁻ parEts* double mutant compared to the *dif⁻* single mutant. Overall, the RecA enrichment profile in the double *dif⁻ parEts* mutant had a similar RecA distribution compared to the single *dif⁻* mutant. However, when the mutant was grown at the restrictive temperature, the peaks of RecA enrichment closer to the *dif* site had less RecA bound to DNA (Figure 6.13). Also, when the *dif⁻ parEts* mutant was grown at permissive temperature, there was a lower amount of RecA bound to DNA in the peak of RecA enrichment that was nearest of the *dif* site. This might be, as mentioned above, because the *parEts*

mutant is partially defective for TopoIV activity even at 30°C, which is better observed in the *dif⁻ parEts* double mutant. Additionally, I noticed an increase of the RecA enrichment level in the *dif⁻ parEts* mutant grown at both temperatures on the *terA* side of *dif* for ~150 kb (highlighted in red in Figure 6.13, zoomed out). It was more pronounced when the mutant was grown at 42°C. This RecA enrichment was not Chi-dependent which might suggest that this increase of RecA level might actually correspond to an increase of DNA level at that position. This observation is consistent with the increased amount of DNA on the left side of *dif* (around *terA*) observed by WGS (Figure 6.10). Because this additional RecA enrichment was present in the double *dif⁻ parEts* mutant grown in both conditions and not in the single *dif⁻* mutant, I suggest that the excess of DNA observed using WGS in Figure 6.10 is not due to the temperature differences (30°C and 42°C against 37°C), but due to the double *dif⁻ parEts* mutation, which shows some TopoIV defects even at 30°C.

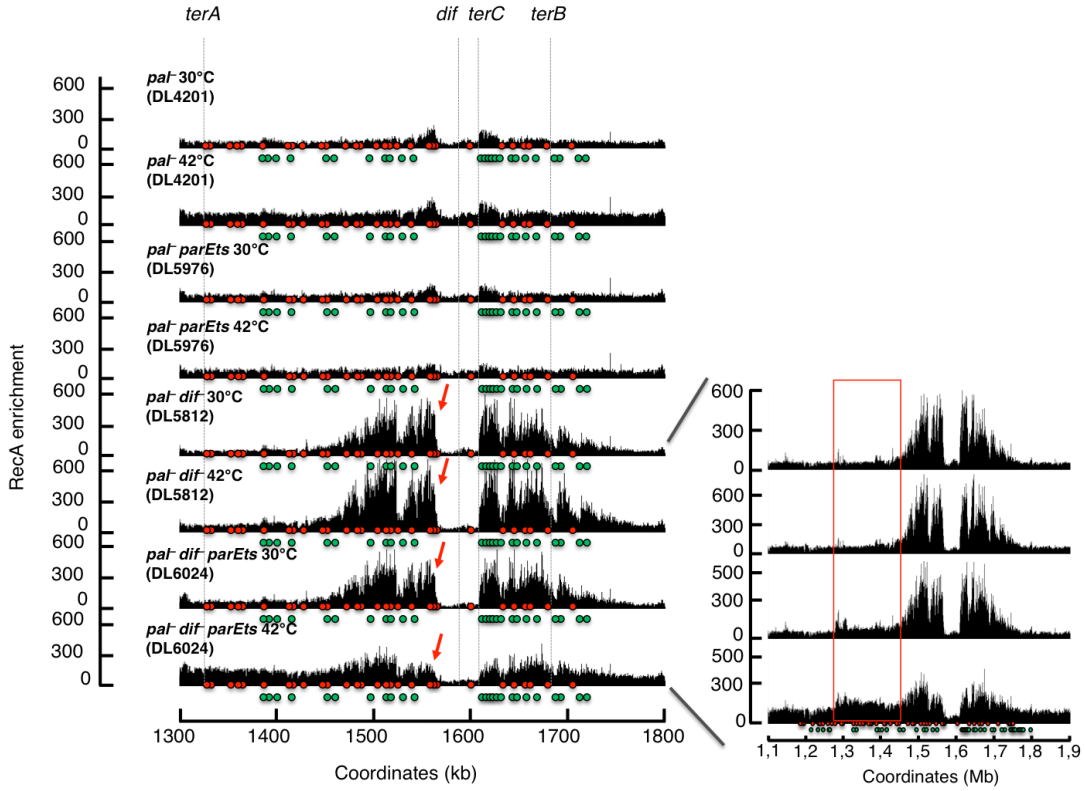


Figure 6.13. Comparison of RecA enrichment in the terminus region of *dif⁻ parEts* mutant.

ChIP-seq analyses revealed that less RecA binds closer to the *dif* site in *dif⁻ parEts* mutant. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Red arrows indicate a peak of RecA enrichment that decreases following the shift to 42°C. The zoomed out part of the graph shows the excess amount of RecA observed on the *terA* side of *dif* in *dif⁻ parEts* mutant (highlighted in red). Strains used were DL4201 (*pal⁻*) grown at 30°C and 42°C, DL5976 (*pal⁻ parEts*) grown at 30°C and 42°C, DL5812 (*pal⁻ dif⁻*) grown at 30°C and 42°C and DL6024 (*pal⁻ dif⁻ parEts*) grown at 30°C and 42°C.

To quantify the level of RecA enrichment close to *dif* in the presence and in the absence of functional TopoIV, I performed ChIP-qPCR experiments on the same DNA that I used for ChIP-seq (Figure 6.14). For these experiments, I used 4 pairs of qPCR primers: 2 pairs of primers after Chi sites – Dif1 and Dif2 and 2 pairs of primers before Chi sites – bef.chi1 and bef.chi3, one on each side of *dif* (Figure 6.14 A). I measured the relative quantification of RecA enrichment levels for each fragment, which was calculated as the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and always shows background level of RecA binding (Cockram *et al*. 2015). As expected from the ChIP-seq experiments above, this quantification experiment showed that the level of RecA enrichment after Chi sites in the *dif- parEts* double mutant grown at 42°C was lower (~4-5-fold) than this level in the *dif- parEts* double mutant grown at 30°C (~8-fold). Also, it was lower than in the single *dif-* mutants (~7-9-fold) grown at both temperatures (Figure 6.14 B). Overall, these observations suggest that the TopoIV is partially responsible for the 'guillotining' of chromosomes resulting in DSBs in the terminus described in Chapter 5.

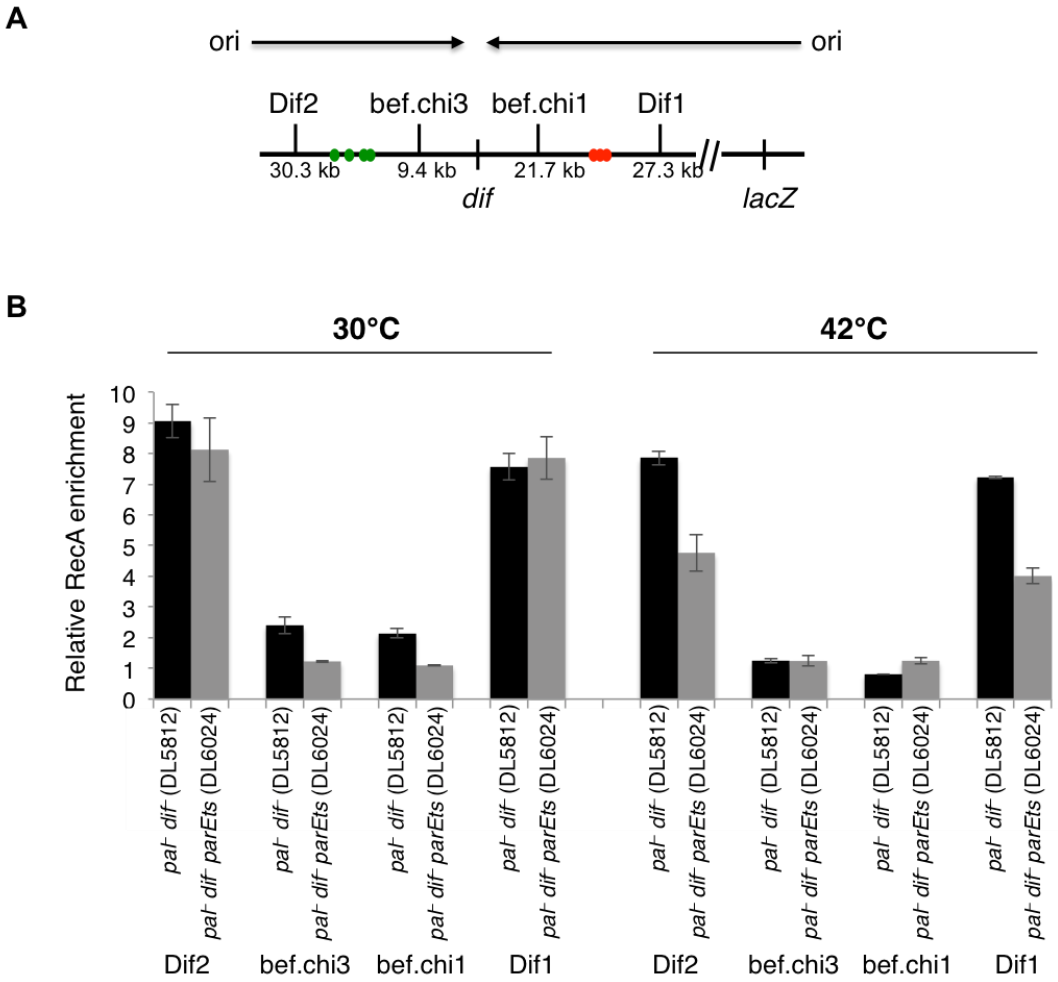


Figure 6.14. Quantification of RecA enrichment in the terminus region of *dif⁻ parEts* mutant revealed the decreased level of RecA binding at 42°C.

ChIP-qPCR revealed that the levels of RecA binding were lower in the *dif⁻ parEts* mutant at 42°C compared to the single *dif⁻* mutant. (A) Schematic of qPCR primer positions. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *lacZ* and positions of qPCR primers are indicated. Replication direction is shown using black arrows. (B) Relative RecA enrichment in the terminus of *pal⁻ dif⁻* and *pal⁻ dif⁻ parEts* strains. RecA enrichment in the terminus is normalised to the level of RecA enrichment at the *hycG* site. Strains used were DL5812 (*pal⁻ dif⁻*) grown at 30°C and 42°C and DL6024 (*pal⁻ dif⁻ parEts*) grown at 30°C and 42°C.

To investigate if the effects of *parEts* mutation are similar in other chromosome dimer resolution mutants, I constructed a FtsK^{K997A} *parEts* double mutant, where FtsK is deficient in ATPase activity and cannot translocate DNA to align two *dif* sites together. For ChIP-seq experiments, I used the DL4201 (*pal*$^{-}$), DL5976 (*pal*$^{-}$ *parEts*) and DL6150 (*pal*$^{-}$ FtsK^{K997A} *parEts*) strains. The cultures were grown using the same conditions as described for *dif* mutants above. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform by Edinburgh Genomics. Finally, raw ChIP-seq data were analysed as described in Chapter 2.

According to the previous observations (Figure 6.13), if the absence of chromosome dimer resolution system was affecting the *dif parEts* double mutant, I expected that the peaks of RecA enrichment close to *dif* would be lower in the FtsK^{K997A} *parEts* double mutant grown at 42°C than in this mutant grown at 30°C. Importantly, the distribution of RecA enrichment observed in the FtsK^{K997A} *parEts* double mutant grown at 30°C, shown in Figure 6.15 was similar to the previously observed amount of RecA enrichment in the FtsK^{K997A} single mutant shown in Figure 5.21. As expected, the FtsK^{K997A} *parEts* double mutant grown at 42°C showed a lower amount of RecA enrichment at the peaks near the *dif* site in the same mutant grown at 30°C (Figure 6.15).

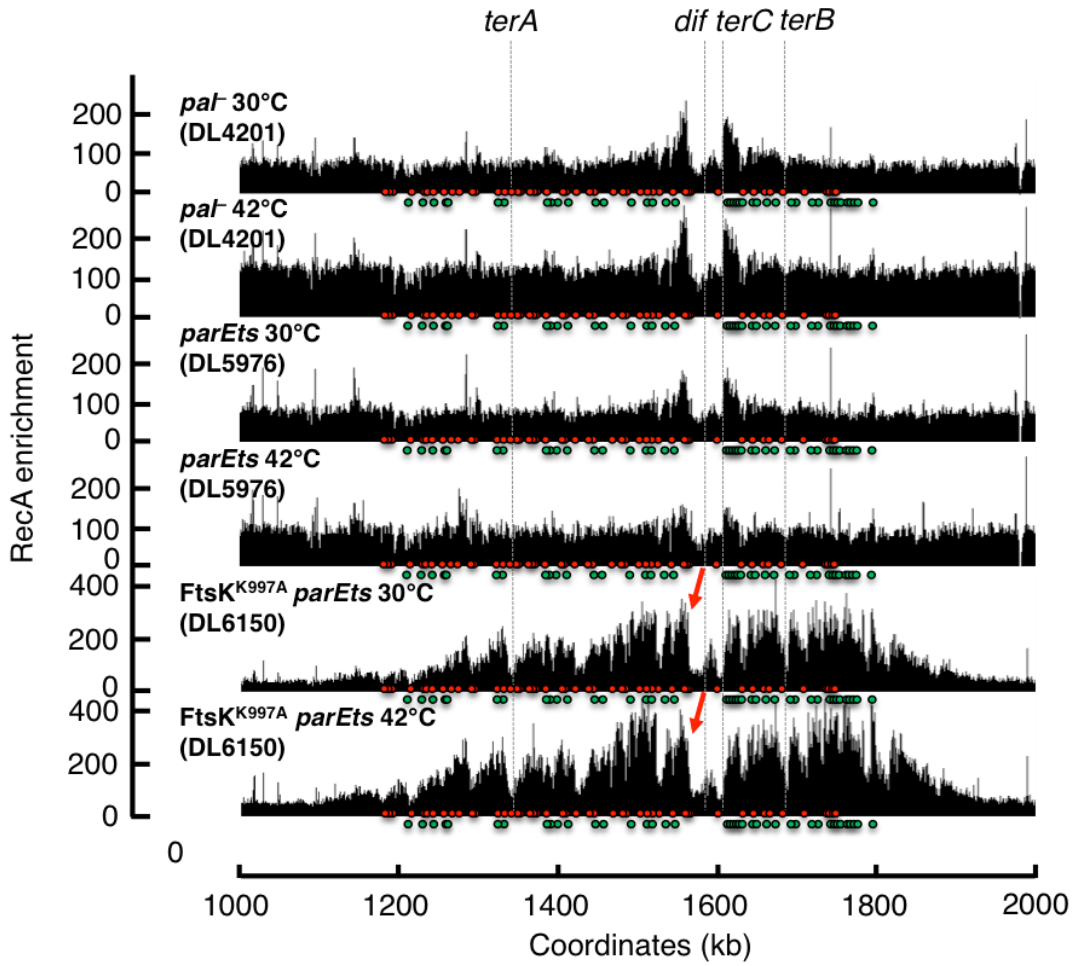


Figure 6.15. RecA enrichment profiles of a FtsK^{K997A} *parEts* mutant showed a decrease of RecA level in the peaks near *dif* at 42°C.

ChIP-seq analyses revealed that less RecA binds closer to the *dif* site in FtsK^{K997A} *parEts* mutant. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Red arrows indicate a peak of RecA enrichment that decreases following the shift to 42°C. Strains used were DL4201 (*pal*$^-$) grown at 30°C and 42°C, DL5976 (*pal*$^-$ *parEts*) grown at 30°C and 42°C and DL6150 (*pal*$^-$ FtsK^{K997A} *parEts*) grown at 30°C and 42°C.

To quantify the loss of RecA binding to DNA around *dif* in the absence of TopoIV in the FtsK^{K997A} mutant, ChIP-qPCR experiments on the FtsK^{K997A} *parEts* double mutant grown at 30°C and 42°C were performed (Figure 6.16). The DNA used in these experiments was the same as in the previous ChIP-seq experiments. The qPCR primers used were the same as described above (Figure 6.16 A). I measured the relative quantification of RecA enrichment levels for each fragment, which was calculated as the relative quantity of target DNA compared to an internal reference. As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and always shows background levels of RecA binding (Cockram *et al*. 2015). As expected from the ChIP-seq experiments above, this quantification experiment showed that the level of RecA enrichment after the Chi sites in the FtsK^{K997A} *parEts* double mutant grown at 42°C was lower (~6-fold) than in the FtsK^{K997A} *parEts* double mutant grown at 30°C (~10-14-fold) (Figure 6.16 B). Also, observed RecA enrichment before Chi sites (bef.chi1 and bef.chi3 fragments) enrichment in the FtsK^{K997A} *parEts* double mutant was higher (~3-5-fold) compared to the other strains studied (~1.5-2-fold). These observations are consistent with the hypothesis that the TopoIV mutant is partially responsible for the 'guillotining' of chromosomes and formation of DSBs in the terminus described in Chapter 5. Although, it was determined that TopoIV is responsible for some of the DSBs in the terminus, further investigations need to be done in order to determine what is the cause of the rest of the observed DSBs in the terminus.

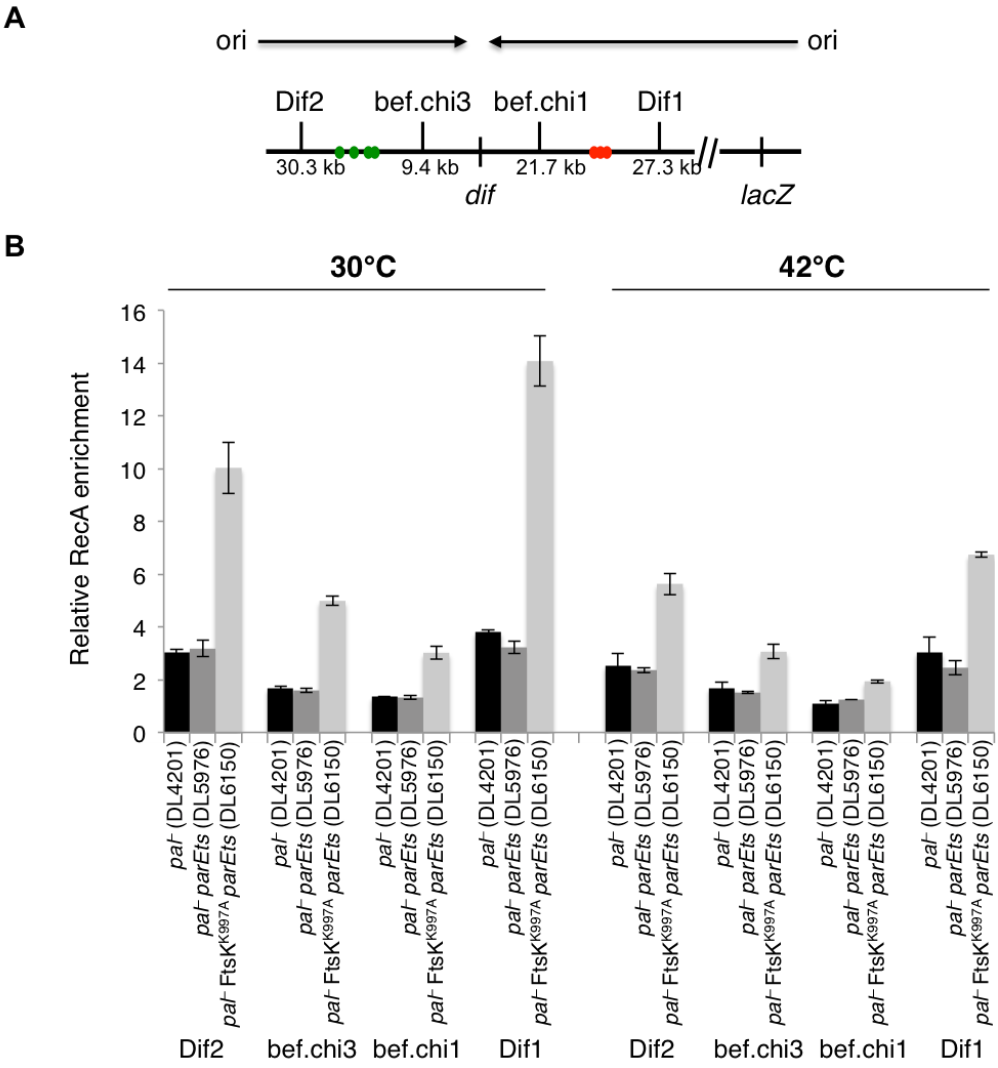


Figure 6.16. Quantification of RecA enrichment in the terminus region of FtsK^{K997A} *parEts* grown at 42°C mutant revealed a decrease of RecA level near *dif*.

ChIP-qPCR revealed that the levels of RecA binding were lower in FtsK^{K997A} *parEts* mutant at 42°C compared to 30°C. (A) Schematic of qPCR primer positions. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *lacZ* and positions of qPCR primers are indicated. Replication direction is shown using black arrows. (B) Relative RecA enrichment in the terminus of *pal*$^{-}$, *pal*$^{-}$ *parEts* and *pal*$^{-}$ FtsK^{K997A} *parEts* strains. RecA enrichment in the terminus is normalised to the level of RecA enrichment at the *hycG* site. Strains used were DL4201 (*pal*$^{-}$) grown at 30°C and 42°C, DL5976 (*pal*$^{-}$ *parEts*) grown at 30°C and 42°C and DL6150 (*pal*$^{-}$ FtsK^{K997A} *parEts*) grown at 30°C and 42°C.

6.4 Does TopoIII play a role in the DSB formation at the terminus?

6.4.1 DSBs in *lacZ* had no effect on the viability of TopoIII mutant

TopoIII is a type I topoisomerase that with the help of RecQ helicase is able to decatenate chromosomes instead of TopoIV by making single-strand breaks, passaging through a double-strand duplex of the sister chromosome and ligating the ends (Harmon, Brockman, and Kowalczykowski 2003; Seol *et al*. 2013; Nurse *et al*. 2003). The *topB* gene encodes the TopoIII topoisomerase, which is not essential for the cell survival. To study whether TopoIII is involved in DSB formation and/or appearance of over-replication in the terminus, I constructed *topB* deletion mutants using the PMGR method. TopoIII mutants do not display any growth defect (Perez-Cheeks *et al*. 2012). Also, *topB⁻ dif⁻* mutants were constructed to investigate if this double mutant shows similar replication profile to the *dif⁻ parEts* double mutant.

As it was reported that TopoIII mutants do not display any growth defect compared to wild type, I decided to investigate the viability of TopoIII mutants subjected to DSBs in *lacZ*. To check this, I performed spot tests for *topB⁻* and *topB⁻ dif⁻* subjected or not to DSBs. Overnight cultures grown at 37°C were diluted to an OD_{600nm} of 1. Then, 10-fold serial dilutions were spotted on LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Plates were incubated at 37°C overnight (Figure 6.17). These spot tests showed no reduction of viability of *topB* mutants compared to wild type strains in presence or absence of DSBs. Double *dif⁻ parEts* mutants were similar to single *dif⁻* mutants where the *dif⁻* strains that were subjected to DSBs showed smaller colony phenotype.

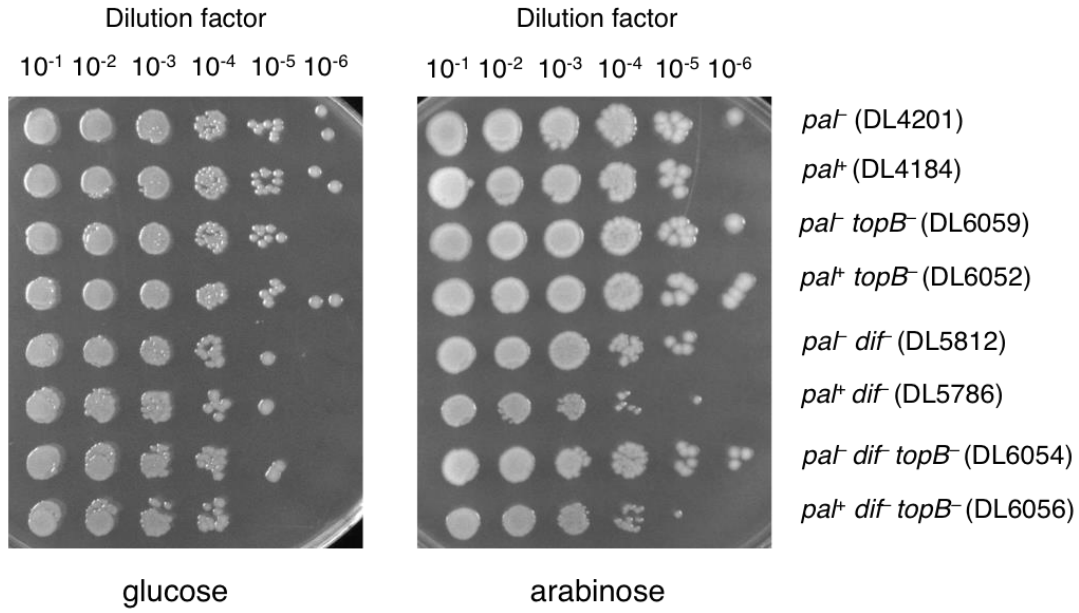


Figure 6.17. The viability of *topB*– mutants is not affected when subjected to DSBs in *lacZ*.

Spot tests of 10-fold serially diluted overnight cultures were spotted onto LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Strains used were DL4201 (*pal*–), DL4184 (*pal*+), DL6059 (*pal*– *topB*–), DL6052 (*pal*+ *topB*–), DL5812 (*pal*– *dif*–), DL5786 (*pal*+ *dif*–), DL6054 (*pal*– *dif*– *topB*–) and DL6056 (*pal*– *dif*– *topB*–).

These observations indicate that the TopoIII mutants display no growth defect when subjected to DSBs compared to the strains that are not subjected to DSBs in *lacZ*. Therefore, for simplicity of experiments and due to the cost of the next generation sequencing, I decided to use TopoIII mutants that were not subjected to DSBs in *lacZ* for all experiments in this chapter.

6.4.2 The replication profile of *topB*⁻ mutant looks similar to wild type

To investigate the amount of DNA in the terminus of a *topB* mutant and compare this to TopoIVts mutants described in a Section 6.1-6.3, I decided to do a WGS experiment. For this experiment, strains DL4201 (*pal*⁻), DL6059 (*pal*⁻ *topB*⁻), DL5812 (*pal*⁻ *dif*⁻) and DL6054 (*pal*⁻ *dif*⁻ *topB*⁻) were used. Overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 37°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and grown until they reach an OD_{600nm} of 0.2-0.3. The DNA for WGS was isolated using the Wizard kit and purified using the Zymo kit as described in Chapter 2. Edinburgh Genomics performed library preparations and WGS using an Illumina HiSeq 2500 platform. Paired-end raw data obtained were analysed as described in Chapter 2. Data from exponentially growing cells were normalised to data from stationary phase *pal*⁻ (DL4201) strain and individual graphs for each biological repeat were carried out (Figure 6.18). As expected, the overall replication profiles of all these strains were V-shaped with more DNA at the origin and less DNA at the terminus. The DNA of all 7 *rrn* operons was partially recovered (Figure 6.19).

If TopoIII was responsible for over-replication in the terminus in the same manner than TopoIV, I would have expected that the replication profile of the single TopoIII mutant would remain similar to the wild type replication profile. Also, according to the previous observation, I would have expected that over-replication would be higher in the *dif* region of the *dif*⁻ *topB*⁻ double mutant. The single *topB*⁻ mutant (DL6059) showed a replication profile in the terminus that was similar to the wild type strain profile (DL4201). No significant DNA excess was observed in the terminus (Figure 6.18). The *dif*⁻ *topB*⁻ double mutant showed a loss of DNA around the *dif* region and a peak

of over-replication shifted towards *terA* similar to the *dif*⁻ single mutant (Figure 6.19).

To visualise if there is a difference in the quantity of DNA in the terminus of these strains, the ratios of reads of *topB*⁻ (DL6059) over *pal*⁻ (DL4201) strains and *topB*⁻ *dif*⁻ (DL6054) over *dif*⁻ (DL5812) strains were carried out (Figure 6.20). The ratio of reads of *topB*⁻ (DL6059) over *pal*⁻ (DL4201) strains showed that some of the DNA was recovered in *rrn* genes (Figure 6.20 A). Also, this ratio showed that there is approximately the same or slightly less amount of DNA in the terminus region of the *topB*⁻ mutant compared to the wild type strain. The ratio of *topB*⁻ *dif*⁻ (DL6054) over *dif*⁻ (DL5812) strains revealed that there is a loss of DNA at *dif* in the *topB*⁻ *dif*⁻ mutant. This might be because TopoIV is involved in the decatenation of the chromosomes around *dif* and introduces some mistakes when TopoIII is absent. So far, *dif*⁻ *topB*⁻ mutant did not display the same over-replication phenotype as *dif*⁻ *parEts* mutants.

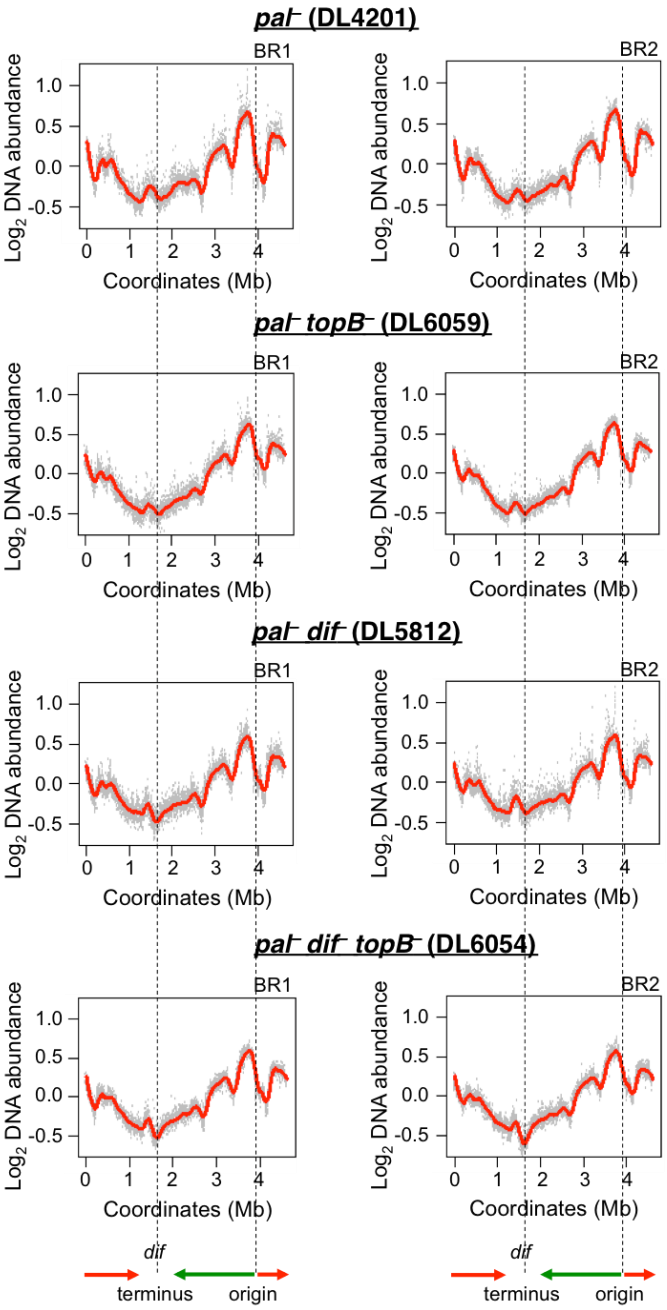


Figure 6.18. WGS profiles of *topB*⁻ single mutant show no difference compared to wild type strains; *dif*⁻ *topB*⁻ double mutant is the same as *dif*⁻ single mutant.

Replication profiles of exponentially growing cultures were normalised to data from the stationary phase *pal*⁻ (DL4201) strain. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of the origin and *dif* (terminus) are indicated using dash lines. BR1 is biological repeat 1; BR2 is biological repeat 2. Strains used were DL4201 (*pal*⁻), DL6059 (*pal*⁻ *topB*⁻), DL5812 (*pal*⁻ *dif*⁻) and DL6054 (*pal*⁻ *dif*⁻ *topB*⁻).

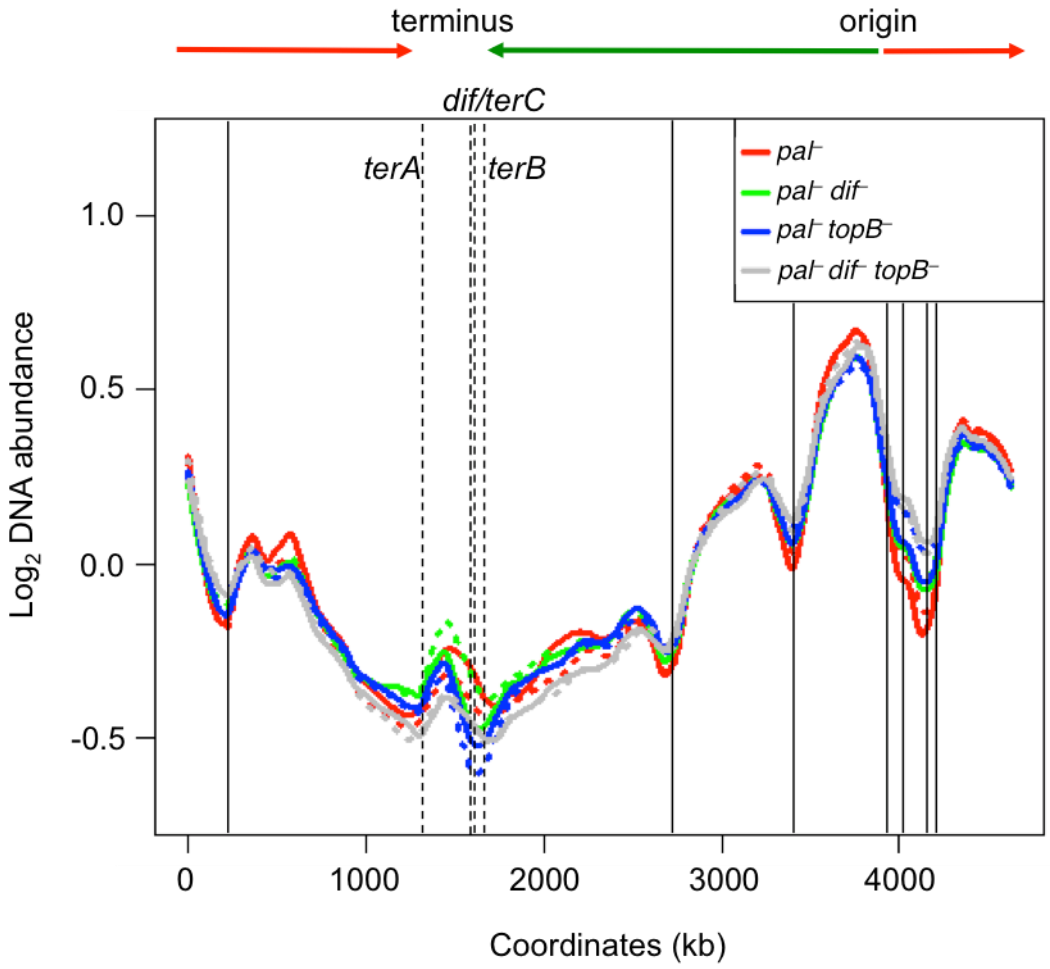


Figure 6.19. Combined replication profiles revealed no difference between the sequenced strains.

Replication profiles of exponentially growing cultures of wild type, *topB⁻*, *dif⁻* and *dif⁻ topB⁻* strains. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of *dif, terA, terB* and *terC* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Biological repeats (BR1 and BR2) of the same strain are shown using dashed/solid lines of the same colour. Strains used were DL4201 (*pal⁻*) in red, DL5812 (*pal⁻ dif⁻*) in green, DL6059 (*pal⁻ topB⁻*) in blue and DL6054 (*pal⁻ dif⁻ topB⁻*) in grey.

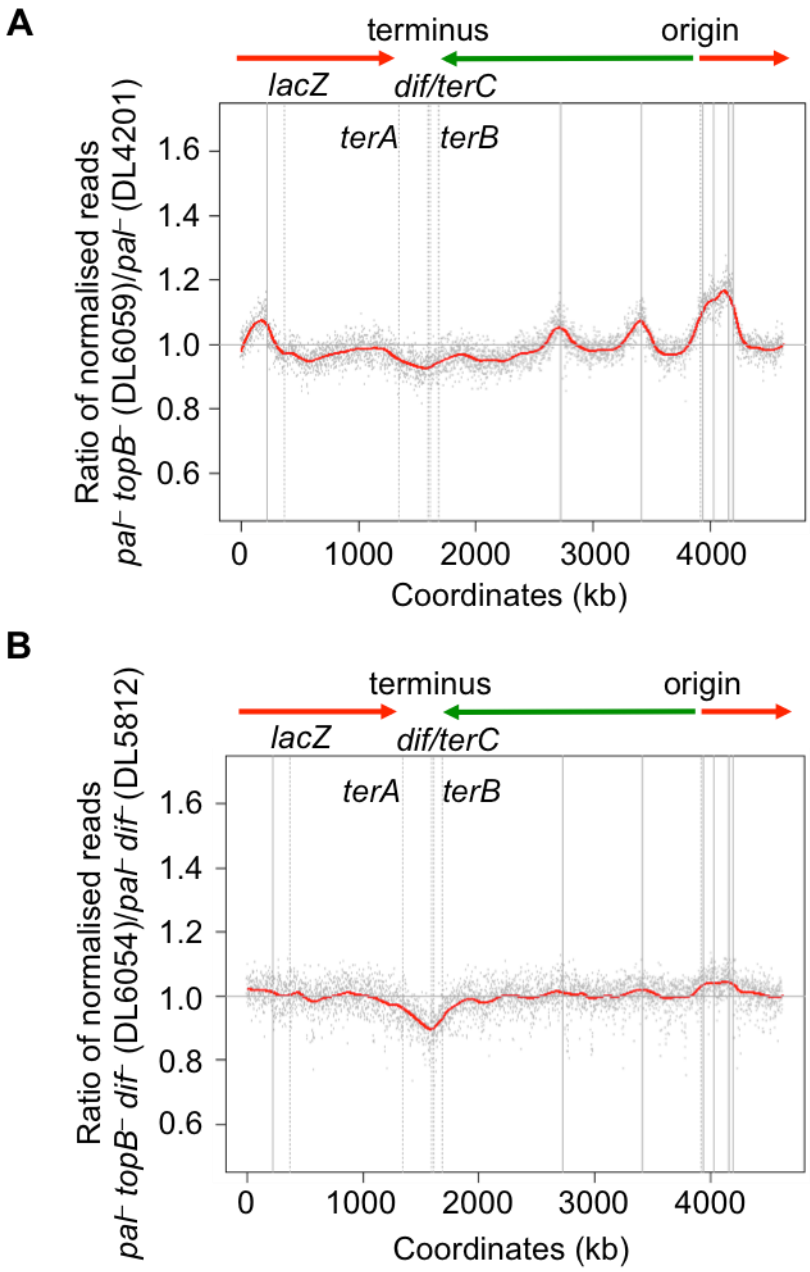


Figure 6.20. Ratio of replication profiles of *topB*⁻ over wild type strain and *topB*⁻ *dif*⁻ strain over *dif*⁻ strain.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of (A) *pal*⁻ *topB*⁻ (DL6059) over *pal*⁻ (DL4201) strains and (B) *pal*⁻ *topB*⁻ *dif*⁻ (DL6054) over *pal*⁻ *dif*⁻ (DL5812) strains. Grey dots represent reads, the red line is the ratio of the reads from DL6059 (A) and DL6054 (B). The horizontal grey line represents DL4201 (A) and DL5812 (B). The direction of replication is shown using green and red arrows. The positions of *lacZ*, *terA*, *terB*, *terC* and *dif* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Strains used were DL4201 (*pal*⁻), DL6059 (*pal*⁻ *topB*⁻), DL5812 (*pal*⁻ *dif*⁻) and DL6054 (*pal*⁻ *dif*⁻ *topB*⁻).

6.4.3 *topB*⁻ single and *topB*⁻ *dif*⁻ double mutants have the same RecA DNA binding levels as wild type and *dif*⁻ single mutant, respectively

To investigate whether TopoIII plays a role in the formation of DSBs in the terminus region, I performed RecA ChIP-seq experiments using the strains described in Chapter 6.4.1. Overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 37°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and grown until they reached an OD_{600nm} of 0.2-0.3. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 2500 platform by Edinburgh Genomics. Finally, raw ChIP-seq data were analysed as described in Chapter 2.

According to the previous observations, showing a loss of DNA in *topB*⁻ *dif*⁻ mutant, I expected that the RecA enrichment levels would remain the same in the single *topB*⁻ mutant as in the wild type strain and would be reduced in the double *topB*⁻ *dif*⁻ mutant. Indeed, the distribution of RecA enrichment observed in the single *topB*⁻ mutant in Figure 6.21 was similar to the distribution of RecA enrichment in the wild type strain (DL4201). However, the *topB*⁻ *dif*⁻ double mutant showed an identical RecA enrichment profile to the *dif*⁻ mutant where more RecA enrichment was observed in both strains compared to the single *topB*⁻ or the wild type strains (Figure 6.21).

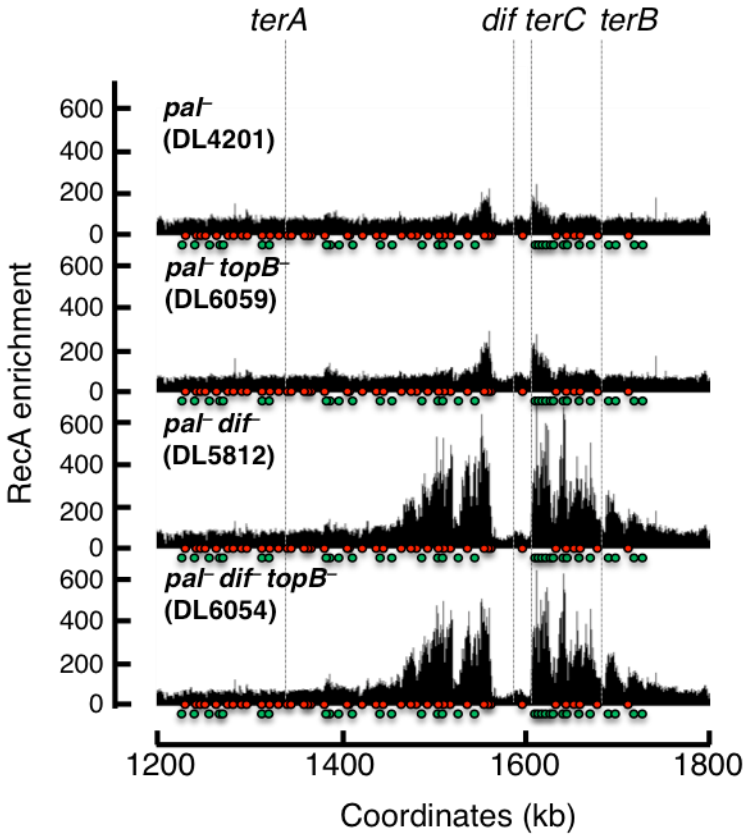


Figure 6.21. Comparison of RecA enrichment in the terminus region of *dif*⁻ *topB*⁻ double mutants showed no difference compared to single *dif*⁻ mutant.

ChIP-seq analyses revealed similar RecA enrichment profiles in the *topB*⁻ mutant compared to the wild type and in the *dif*⁻ *topB*⁻ mutant compared to the *dif*⁻ mutant. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal*⁻), DL6059 (*pal*⁻ *topB*⁻), DL5812 (*pal*⁻ *dif*⁻) and DL6054 (*pal*⁻ *dif*⁻ *topB*⁻).

To confirm that the levels of RecA enrichment were similar between the *topB*⁻ single mutant and the wild type and between the *topB*⁻ *dif*⁻ double mutant and the *dif*⁻ mutant, I performed ChIP-qPCR experiment, using the same DNA than for the ChIP-seq experiments. For those qPCR experiments, I used 4 pairs of primers as described in Chapter 6.3 (Figure 6.22 A). As a reference, a site in the *hycG* gene was used as this gene is constitutively expressed under all growth conditions and always shows background level of RecA binding (Cockram *et al*. 2015). As expected from the ChIP-seq experiments above, this quantification experiment showed that the level of RecA enrichment in the *topB*⁻ single mutant was the same or slightly lower than the level of RecA enrichment in the wild type strain (DL4201) (Figure 6.22 B). However, the RecA enrichment in *topB*⁻ *dif*⁻ double mutant was slightly higher than in the *dif*⁻ mutant (Figure 6.22 C). This result might suggest that there are more DSBs observed in *topB*⁻ *dif*⁻ double mutant, which is consistent with higher DNA loss in this mutant.

To summarise, these experiments suggest that the TopoIII topoisomerase alone is not exclusively involved in DSBs formation in the terminus. However, WGS and ChIP-seq result on the *dif*⁻ TopoIII double mutant suggest that TopoIII might be involved in the DSBs formation when XerCD/*dif* system does not operate.

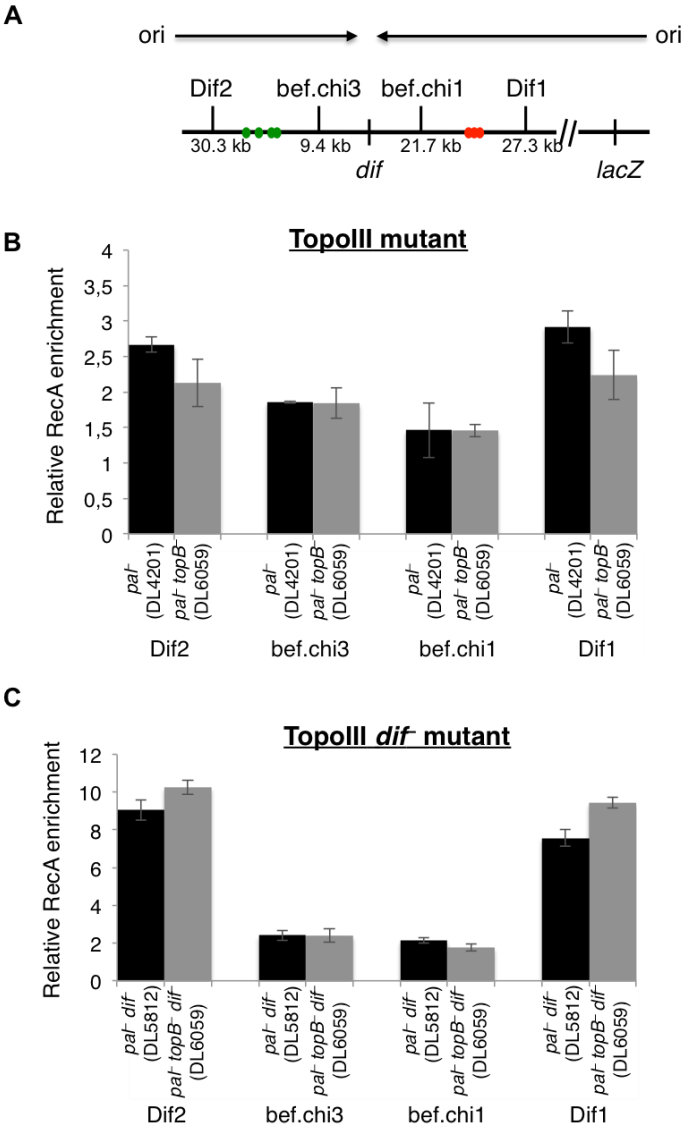


Figure 6.22. Quantification of RecA enrichment in the terminus region of *topB*⁻ mutants showed no difference compared to wild type (B) and of *dif*⁻ *topB*⁻ double mutant compared *dif*⁻ mutant.

ChIP-qPCR revealed the same level of RecA binding in the *topB*⁻ mutant than in the wild type (B) and in the *topB*⁻ *dif*⁻ mutant than in the *dif*⁻ mutant (C). (A) Schematic of qPCR primer positions. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *lacZ* and positions of qPCR primers are indicated. Replication direction is shown using black arrows. (B) Relative RecA enrichment in the terminus of *pal*⁻ and *pal*⁻ *topB*⁻ strains. (C) Relative RecA enrichment in the terminus of *pal*⁻ *dif*⁻ and *pal*⁻ *topB*⁻ *dif*⁻ strains. RecA enrichment in the terminus is normalised to the level of RecA enrichment at the *hycG* site. Strains used were DL4201 (*pal*⁻), DL6059 (*pal*⁻ *topB*⁻), DL5812 (*pal*⁻ *dif*⁻) and DL6054 (*pal*⁻ *dif*⁻ *topB*⁻).

6.5 TopoIII TopoIV together are involved in the formation of DSBs around the *dif* site

6.5.1 DSBs in *lacZ* had no effect on the viability of TopoIII mutant

As TopoIII can decatenate chromosomes in the absence of TopoIV, I decided to construct a TopoIII TopoIV double mutant to check if these topoisomerases could together be a cause of initially observed DSBs around the *dif* site. The *topB⁻ parCts* mutants were reported to be synthetically lethal and show an accumulation of unresolved catenanes even when grown at a permissive temperature (30°C) (Perez-Cheeks *et al*. 2012). Therefore, to construct a double TopoIII TopoIV mutant, I used the P1 transduction method to introduce the *parEts* mutation in the *topB⁻* strain. The *topB⁻* strain was kept at 30°C at all times to facilitate this strain construction.

To verify if a DSB elsewhere in the chromosome influences the growth of the TopoIV TopoIII double mutant, I decided to investigate the viability of these mutants subjected or not to DSBs in *lacZ*. To check this, I performed spot tests for *topB⁻ parEts* double mutants subjected or not to DSBs. Overnight cultures grown at 30°C were diluted to an OD_{600nm} of 1. Then, 10-fold serial dilutions were spotted on LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Plates were incubated at 30°C overnight (Figure 6.23). These spot tests showed 10-fold reduction of viability for *topB⁻ parEts* double mutants compared to single mutants or wild type strains in the absence of palindrome (Figure 6.23, upper panel). Also, *topB⁻ parEts* double mutants that contained a palindrome grew as smaller colonies even when SbcCD expression was suppressed (Figure 6.23, lower panel).

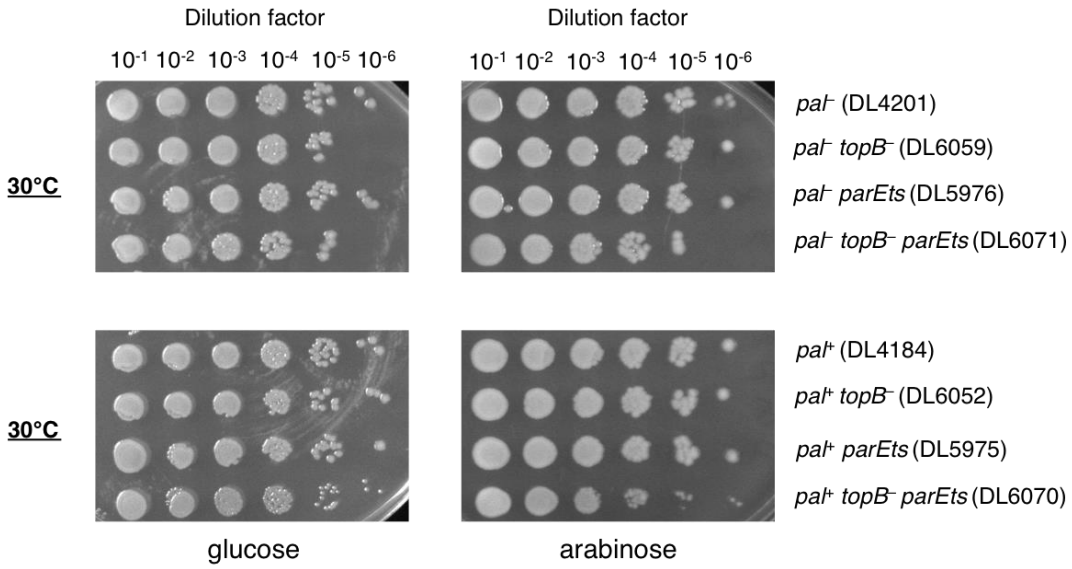


Figure 6.23. 10-fold reduction in the viability of *topB*– *parEts* mutants is observed.

Spot tests of 10-fold serially diluted overnight cultures were spotted onto LB plates supplemented with 0.5% glucose to repress SbcCD expression or 0.2% arabinose to induce SbcCD expression. Strains used were DL4201 (*pal*–), DL6059 (*pal*– *topB*–), DL5976 (*pal*– *parEts*), DL6071 (*pal*– *topB*– *parEts*) DL4184 (*pal*+), DL6052 (*pal*+ *topB*–), DL5975 (*pal*+ *parEts*) and DL6070 (*pal*+ *topB*– *parEts*).

This suggests that a presence of a palindrome reduces the viability of *topB*$^{-}$ *parEts* double mutant or that the *topB*$^{-}$ *parEts* double mutant without a palindrome accumulated suppressor mutations. Therefore, these spot tests have to be repeated. The chosen *topB*$^{-}$ *parEts* double mutant without a palindrome was growing slower than single mutants, as was reported before (Perez-Cheeks *et al*. 2012), so, it is unlikely that the mutant accumulated the suppressor mutations and, therefore, it was used for further experiments.

6.5.2 More DNA around the *dif* site is observed in the *parEts topB*⁻ double mutant when *parEts* is inactive

To check if the double mutant showed differences in its replication profile compared to the single mutants, I performed WGS on these strains. The strains used for this experiment were DL4201 (*pal*⁻) and DL6071 (*pal*⁻ *topB*⁻ *parEts*). The growth conditions used for this experiment were the same as described in Chapter 6.3.3. The DNA was isolated and purified as described in Chapter 2. Edinburgh Genomics performed library preparations and WGS using an Illumina HiSeq 2500 platform. Obtained paired-end raw data were analysed as described in Chapter 2. Data from exponentially growing strains were normalised to data from the stationary phase *pal*⁻ (DL4201) strain and individual graphs for each biological repeat were carried out (Figure 6.24). As expected, the overall replication profiles of all these strains were V-shaped with more DNA at the origin and less DNA at the terminus. The DNA of all 7 *rrn* operons was only partially recovered (Figure 6.25). Surprisingly, biological replicates of the *topB*⁻ *parEts* double mutant grown at either 30°C or 42°C were different from each other. BR2 of the double mutant grown at 42°C showed a similar profile to BR2 of the double mutant grown at 30°C, except that there was more DNA observed in the terminus at 42°C (Figure 6.25). Therefore, this experiment needs to be repeated.

To better visualise this difference, ratios of the amount of DNA in the double mutant over the wild type were calculated (Figure 6.26). For these ratios only BR2 of the *topB*⁻ *parEts* double mutant was taken to avoid extra peaks and gaps due to the problems in BR1. In Figure 6.26 A, the ratio of reads of the *topB*⁻ *parEts* double mutant grown at 42°C was compared to the *topB*⁻ *parEts* double mutant grown at 30°C. This ratio showed that more DNA was accumulated on both sides of the terminus region. The observed difference could be due to

the normalisation issues. Interestingly, this ratio showed a peak of over-replication at *dif* suggesting that there is less DNA in the double *topB⁻ parEts* mutant when TopoIV is active (30°C). The ratio of reads of the *topB⁻ parEts* double mutant grown at 30°C was compared to the wild type strain grown at 30°C (Figure 6.26 B). This ratio showed no difference between the compared strains. Then, the ratio of reads of the *topB⁻ parEts* double mutant grown at 42°C was compared to the wild type strain grown at 42°C (Figure 6.26 C). This ratio showed an excess of DNA in the terminus region of the double mutant.

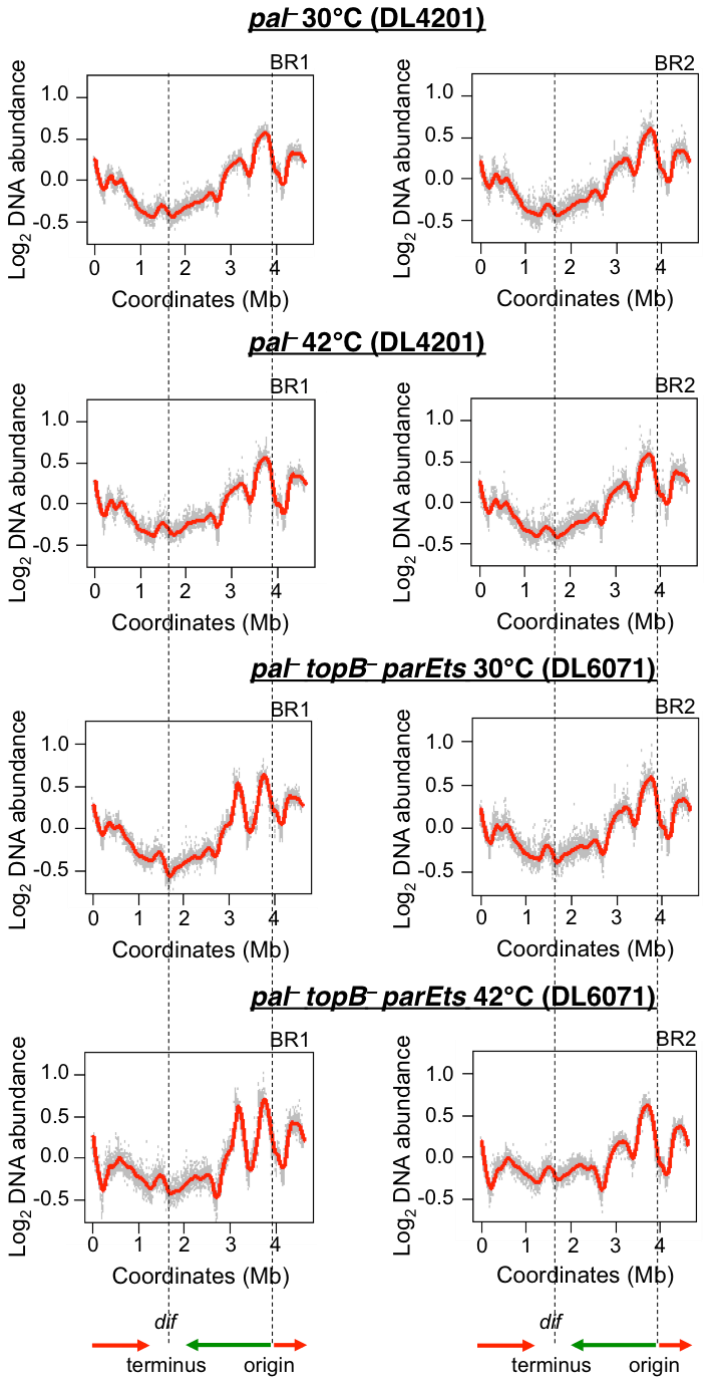


Figure 6.24. WGS profiles of biological repeats of *topB⁻ parEts* mutants.

Replication profiles of exponentially growing cultures were normalised to data from the stationary phase *pal⁻* (DL4201) strain. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of the origin and *dif* (terminus) are indicated using dash lines. BR1 is biological repeat 1; BR2 is biological repeat 2. Strains used were DL4201 (*pal⁻*) and DL6071 (*pal⁻ topB⁻ parEts*) grown at 30°C and 42°C.

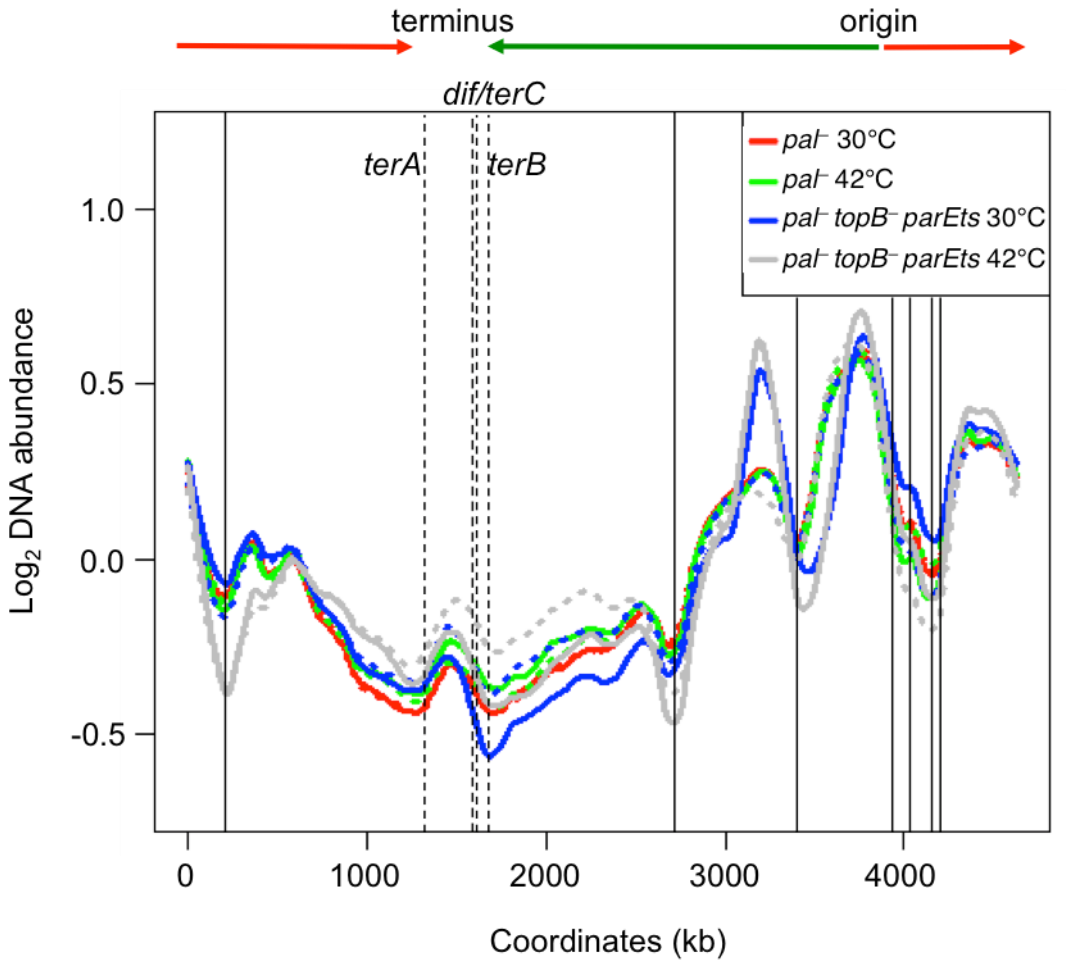


Figure 6.25. Combined replication profiles of all sequenced strains showed significant difference between different biological repeats of the *topB⁻ parEts* strain.

Replication profiles of exponentially growing cultures of *topB⁻ parEts* and wild type strains. Log_2 DNA abundance represents the log_2 of the normalised copy number of uniquely mapped sequence reads. The direction of replication is shown using green and red arrows. The positions of *dif, terA, terB* and *terC* sites are indicated using vertical dash lines. The positions of *rrn* operons are indicated using vertical solid lines. Biological repeats (BR1 and BR2) of the same strain are shown using dashed/solid lines of the same colour. Strains used were DL4201 (*pal⁻*) grown at 30°C in red, DL4201 (*pal⁻*) grown at 42°C in green, DL6071 (*pal⁻ topB⁻ parEts*) grown at 30°C in blue and DL6071 (*pal⁻ topB⁻ parEts*) grown at 42°C in grey.

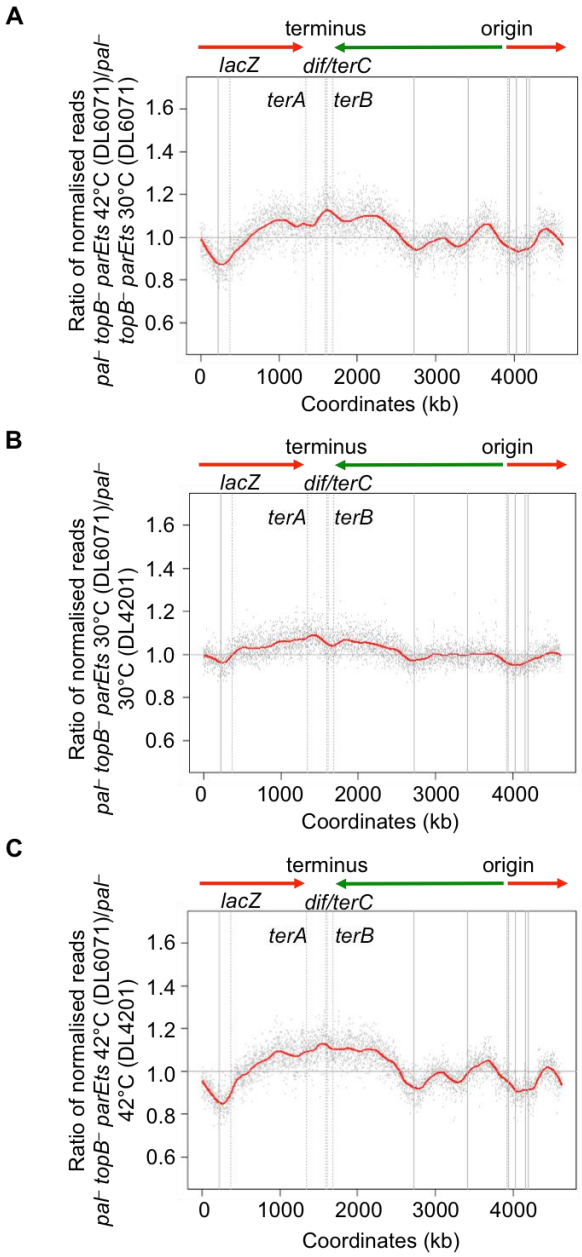


Figure 6.26. Ratio of replication profiles of the *topB⁻ parEts* strain over the wild type strain.

Ratio of normalised uniquely mapped sequence reads from exponentially growing cultures of (A) the *pal⁻ topB⁻ parEts* (DL6071) strain at 42°C over the *pal⁻ topB⁻ parEts* (DL6071) strain at 30°C; (B) the *pal⁻ topB⁻ parEts* (DL6071) strain at 30°C over the *pal⁻* (DL4201) strain at 30°C and (C) the *pal⁻ topB⁻ parEts* (DL6071) strain at 42°C over the *pal⁻* (DL4201) strain at 42°C. Grey dots represent reads, the red line is the ratio of the reads from DL6071 at 42°C (A); DL6071 at 30°C (B) or DL6071 at 42°C (C). The horizontal grey line represents DL6071 at 30°C (A); DL4201 at 30°C (B) or DL4201 at 42°C (C). The direction of replication is shown using green and red arrows. The positions of *lacZ*, *terA*, *terB*, *terC* and *dif* sites are indicated using dash lines. The positions of *rrn* operons are indicated using solid lines. Strains used were DL4201 (*pal⁻*) and DL6071 (*pal⁻ topB⁻ parEts*).

6.5.3 DSBs in the terminus are reduced in the *topB⁻ parEts* double mutant

To investigate whether the TopoIII TopoIV are together responsible for DSBs around the *dif* region, I performed a ChIP-seq experiment on the *topB⁻ parEts* double mutant grown at 30°C and 42°C using the strains described in Chapter 6.5.2. The growth conditions used for this experiment were as described in Chapter 6.3.4. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform by Edinburgh Genomics. Finally, raw ChIP-seq data were analysed as described in Chapter 2.

The *topB⁻ parEts* mutant was compared to the single *parEts* mutant and to the wild type strain (Figure 6.27). More DSBs at *ter* sites were observed in the *topB⁻ parEts* double mutant grown at permissive temperature. Interestingly, the *topB⁻ parEts* double mutant grown at non-permissive temperature showed an almost complete absence of DSBs in the terminus. These results need to be confirmed by qPCR.

All these experiments using the TopoIII TopoIV double mutant suggest that TopoIII and TopoIV might be a cause of initially observed DSBs around the *dif* site.

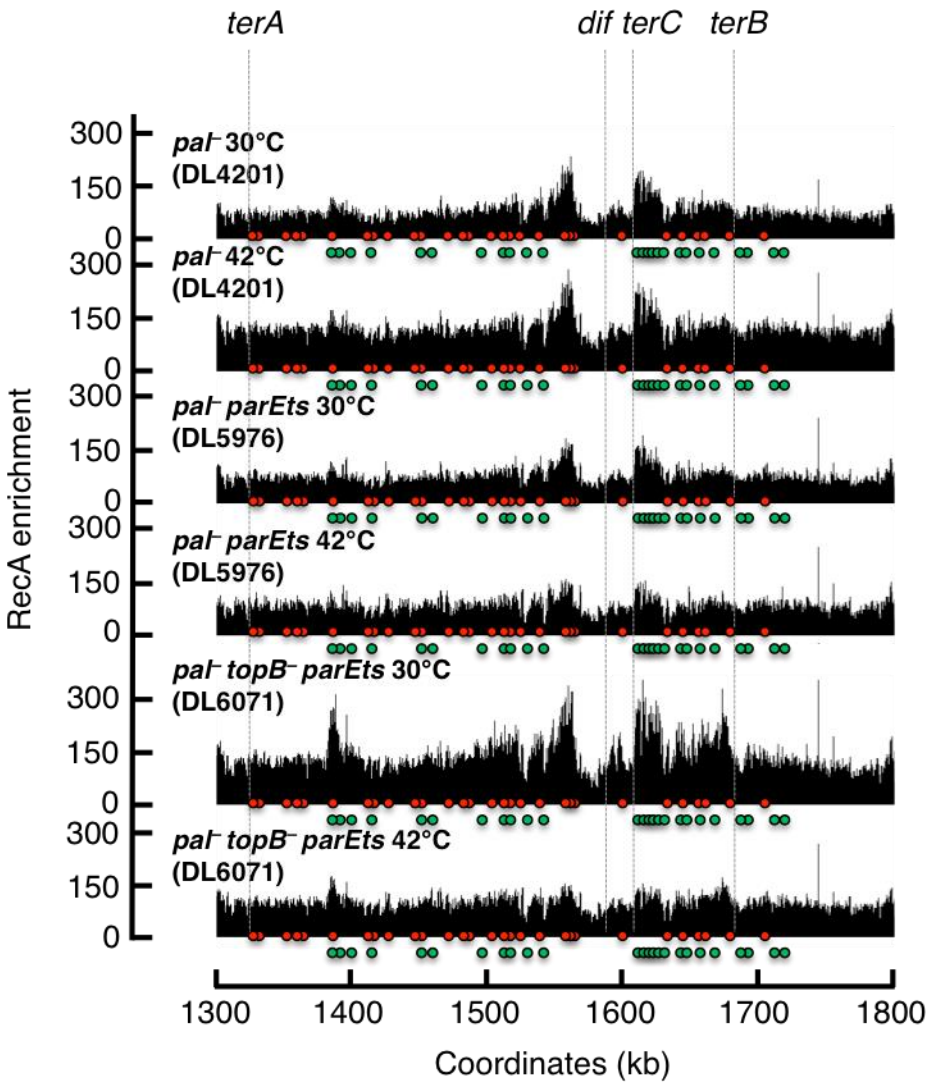


Figure 6.27. Comparison of RecA enrichment in the terminus region of *topB⁻ parEts* mutant revealed less RecA enrichment around the *dif* site when TopoIV is inactive.

ChIP-seq analyses revealed a lower level of RecA enrichment in the *topB⁻ parEts* double mutant. The positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal⁻*) grown at 30°C and 42°C, DL5976 (*pal⁻ parEts*) grown at 30°C and 42°C and DL6071 (*pal⁻ topB⁻ parEts*) grown at 30°C and 42°C.

6.6 Discussion

This chapter described the role of the topoisomerase IV and topoisomerase III in the formation of DSBs around the *dif* site. These topoisomerases were chosen because they function primarily in the terminus. TopoIV is an essential protein, therefore, I constructed temperature-sensitive mutants of each subunit of TopoIV, *parE* and *parC*. These mutants could grow at 30°C but were dead at 42°C. Surprisingly, a *parCts* mutant subjected to chronic DSBs showed a decreased viability even at 30°C. Hence, the *parEts* mutant of TopoIV was chosen for all further experiments. As *parEts* mutants subjected or not to palindrome-induced DSBs showed the same growth rates, I decided to only use the strains that did not contain the palindrome.

TopoIII, on the contrary, is not an essential protein and a full deletion of the *topB* gene that encodes TopoIII was performed. Viability tests showed no difference between TopoIII mutants subjected or not to DSBs in *lacZ*, therefore for simplicity of the experiments, the TopoIII mutant without the palindrome was chosen. Additionally, a double mutant of TopoIII TopoIV was constructed. On the spot tests this mutant showed 10-fold reduction in viability. Also, this mutant that contained a palindrome grew as smaller colonies even when SbcCD was not induced, which might suggest that even a presence of a palindrome has an effect on the strain viability or that the strain without a palindrome accumulated suppressor mutations and, therefore, grows as normal-sized colonies.

Finally, double mutants of TopoIV *dif* and TopoIII *dif* were constructed to avoid chromosome decatenation by the XerCD/*dif* system. Unfortunately, I could not construct the triple TopoIII TopoIV *dif* mutant, probably, because the viability of TopoIV *dif* double mutant is reduced even at permissive temperature.

A mutation in *parEts* or *topB* did not show any DSB or over-replication phenotype indicating that either TopoIV or TopoIII is not exclusively responsible for DSBs around the *dif* site. This is probably because these proteins work synergistically so that when one is absent the other one is enough to generate the DSBs. This is confirmed by the behaviour of the double TopoIV TopoIII mutant (see below).

Also, the XerCD/*dif* system might be involved in the decatenation if TopoIV is absent. The TopoIII *dif* double mutant did not show any DSB or over-replication phenotype either. This might be because TopoIV is primarily operating in the terminus and decatenating all chromosomes. TopoIV is targeted to *dif* by the FtsK protein, so both systems – decatenation and chromosome dimer resolution – operate together.

Surprisingly, the TopoIVts *dif*⁻ double mutant showed a significant increase of DNA amount on the left side of *dif* (next to *terA*). This increase could be observed in the WGS profiles when cells were grown at both restrictive and permissive temperatures. A loss of sequence around *dif* was also observed in a double TopoIVts *dif*⁻ mutant. This loss was greater at a permissive temperature and was partially recovered at a restrictive temperature.

Accordingly, RecA-ChIP-seq showed that the DSBs observed in the TopoIV *dif*⁻ mutant at restrictive temperature were less abundant at a peak that was closer to the *dif* site, which was confirmed by qPCR on the same DNA. Beside this peak, the ChIP-seq profile of the TopoIV *dif*⁻ mutant looked similar to the profile of the *dif*⁻ mutant.

The decrease of RecA binding and over-replication close to *dif* in TopoIV *dif*⁻ mutant suggests that TopoIV is partially responsible for 'guillotining' DSBs as described in Chapter 5. Also, this suggests that the over-replication and DSBs around *dif* are linked to each other. This provides some evidence that when

XerCD/*dif* is absent, TopoIV can break the DNA to resolve the accumulated chromosome dimers but, probably, does not operate properly there. Another double mutant that disrupts chromosome dimer resolution, $FtsK^{K997A}$ TopoIVts, confirmed that TopoIV is involved in the appearance of DSBs around *dif* when the chromosome dimer resolution system is not functional.

When the TopoIVts TopoIII mutant was tested at restrictive temperature for the levels of RecA enrichment in the terminus by ChIP-seq, it was shown that both proteins are involved in DSBs formation around *dif*. This result confirms that both topoisomerases function synergistically and are needed to decatenate chromosomes. Because both topoisomerases function through a DNA break (DSB for TopoIV and SSB for TopoIII that can be converted into a DSB), it is possible that, when the cells are growing in rich medium and the DNA replication is fast, the topoisomerases together with the XerCD/*dif* system are overloaded with chromosome dimers and chromosome catenanes, so they misfunction and leave DSBs that are not immediately ligated and could be recognised and repaired by RecBCD-mediated DSBR.

Another explanation could be that the observed DSBs around the *dif* region lead to DNA over-replication originated from *dif*. If this hypothesis is correct, then TopoIV, TopoIII and the XerCD/*dif* system would be required to break one of the three chromosomes trapped in the terminus to successfully resolve them into two separated sister chromosomes.

Chapter VII

7. Mapping the location of Tus-independent DSBs in the terminus

7.1 Introduction

To map the positions of DSBs in the terminus region, I used a tool that was developed by Benura Azeroglu (unpublished). This tool uses the pattern of RecA binding in a *recD* mutant. RecD protein is one of the subunits of the RecBCD complex that initiates HR-dependent DSBR. The RecD subunit is a 5'-3' helicase that starts unwinding double-stranded DNA at the break site. In a *recD* mutant, the initial unwinding of the DNA is slower and RecB starts loading RecA on the DNA immediately (Churchill, Anderson, and Kowalczykowski 1999). Hence, a RecA-ChIP-seq analysis of a *recD* mutant (strain DL5699) containing a 246 bp palindrome in the *lacZ* region, did not show loading of RecA in relation to Chi sites, but to the place of the break, at the palindrome (Figure 7.1 A). Therefore, I used this tool to more precisely localise the positions of DSBs in the terminus region of the *E. coli* chromosome.

Unfortunately, at first, the pattern of positions of DSBs in the *dif* region was more complex than expected due to other RecA-dependent events also happening in this region in a *recD* mutant (Figure 7.1 B). During termination of DNA replication, some replication forks are expected to pause at *ter* sites (*terA*, *terB* and *terC*). These events occur more frequently in a strain subjected to a palindrome-dependent DSB where one replication fork is delayed and, presumably, replication forks progression is less coordinated. Some paused replication forks break at *ter* sites, leaving one-ended double-strand breaks (Figure 7.1 B, upper panel). The pattern of RecA loading at Chi sites near these DSBs in the terminus suggests that repair of these breaks is RecBCD-mediated. When the level of RecA enrichment was analysed in a *recD* mutant, we noticed

that RecA loading at one-ended DSBs at *ter* sites was elevated and that binding in relation to the *terC* site overlapped with the region containing the DSBs surrounding the *dif* site (Figure 7.1 B).

Therefore, in this chapter, I have studied the consequences of the one-ended DSBs at *ter* sites and eliminated them in order to more precisely localise the region around *dif* subjected to DSBs.

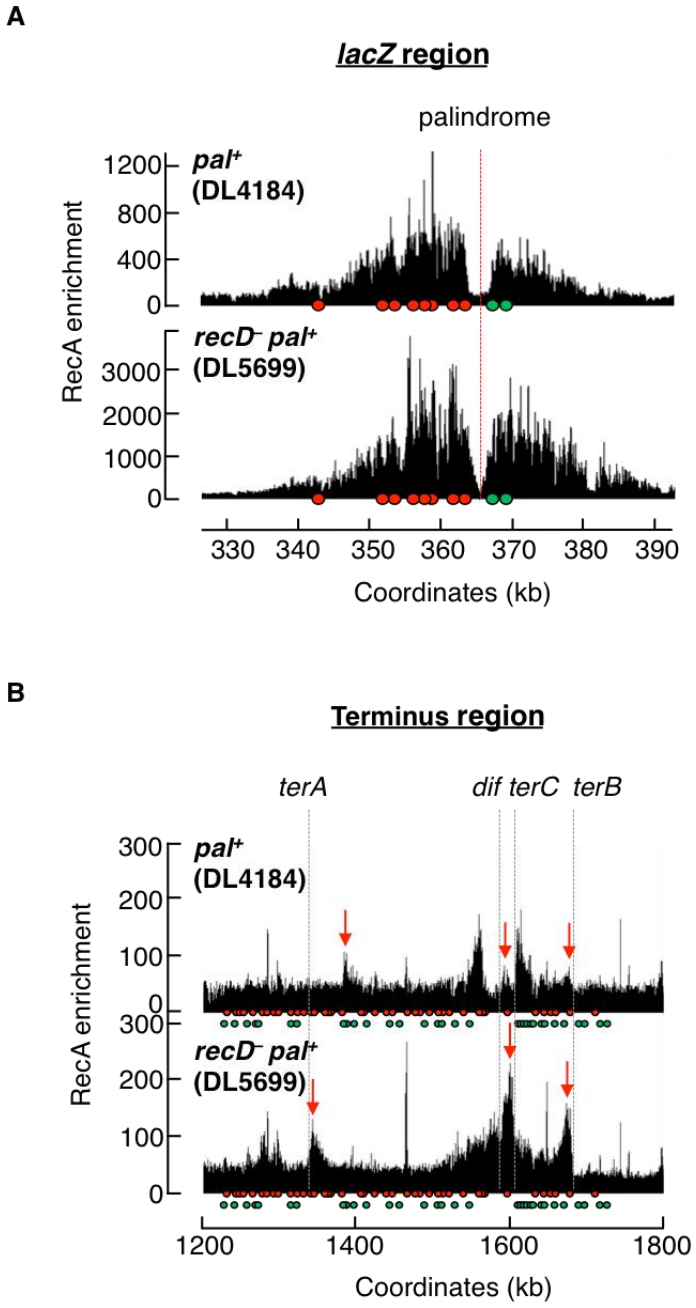


Figure 7.1. *recD* mutant as a tool to map the precise position of a DSB.

RecA-ChIP-seq analysis of *recD* mutant reveals the position of a DSB. The palindrome, *terA*, *terB*, *terC* and *dif* are indicated. Positions of Chi sites are shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. Red arrows are indicating the position of one-ended DSBs at *ter* sites. Strains used were DL4184 (*pal*$^+$) and DL5699 (*recD*$^-$ *pal*$^+$). (A) Mapping of RecA enrichment at the palindrome site in the *lacZ* region. (B) RecA binding at the terminus region. Dr. Benura Azeroglu prepared the samples and performed the analysis.

7.2 No growth defect in *tus* mutants is detected in presence or absence of a DSB at the *lacZ* locus

As RecA enrichment adjacent to the *terC* site is expected to be the result of paused replication forks and the Tus protein is essential for pausing replication forks at *ter* sites, *tus* mutants were constructed and studied. Due to the sensitivity of ChIP-seq and the terminus location of the *tus* gene, I could not use the P1 transduction method to construct *tus* mutants as previously (Chapter 3). During P1 transduction, 100 kb of DNA is packed in the P1 phage. The *tus* mutation might also bring other unwanted mutations in the terminus region that would affect the results of the experiment. Therefore, the PMGR method was used to construct these mutants (Chapter 2).

The *tus* mutation was introduced in cells that express *sbcCD* under an arabinose inducible promoter and contain Chi arrays 1.5 kb away either side of the locus in the *lacZ* gene containing or not a 246 bp interrupted palindrome. Because replication forks are not coordinated in a strain subjected to a DSB in *lacZ,* a decrease in viability might be observed in *tus* mutants where the replication terminates further from the *dif*/*terC* locus. Therefore, the viability of these mutants was tested by growth curve and spot tests.

In the morning, overnight cultures were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown at 37°C to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01, divided into 2 flasks and 0.2% arabinose was added to one flask to induce a DSB in *lacZ* (time 0). OD_{600nm} of these exponentially growing cultures was measured every hour for 5 hours (Figure 7.2). The OD_{600nm} was regularly monitored and the cultures were diluted in order to keep cultures in the exponential phase (OD_{600nm} of 0.2-0.3). The growth profile of strains containing or not *tus* with or without a 246 bp palindrome in *lacZ* remained the same during the whole

period of time while grown in arabinose or glucose. This result suggests that the presence of a DSB in *lacZ* of a *tus* mutant did not have a detectable effect on the growth of this strain.

To check the viability of *tus* mutants, I performed a spot test, where samples at the end of the growth curve were diluted to an OD_{600nm} of 0.1. Then, 10-fold serial dilutions were spotted on LB plates supplemented with 0.5% glucose to repress SbcCD expression and incubated at 37°C overnight (Figure 7.3). Although, *tus* mutation affected the shape of colonies, spot tests showed full viability of *tus* mutants subjected or not to a DSB in *lacZ*. I noticed that the colonies with *tus* mutation are less uniform and less shiny than the colonies of wild type cells. These experiments suggested that there is no effect on growth or viability when the *tus* gene is deleted even when DSBR at the site of a palindrome in *lacZ* had unbalanced the replication of the two replichores.

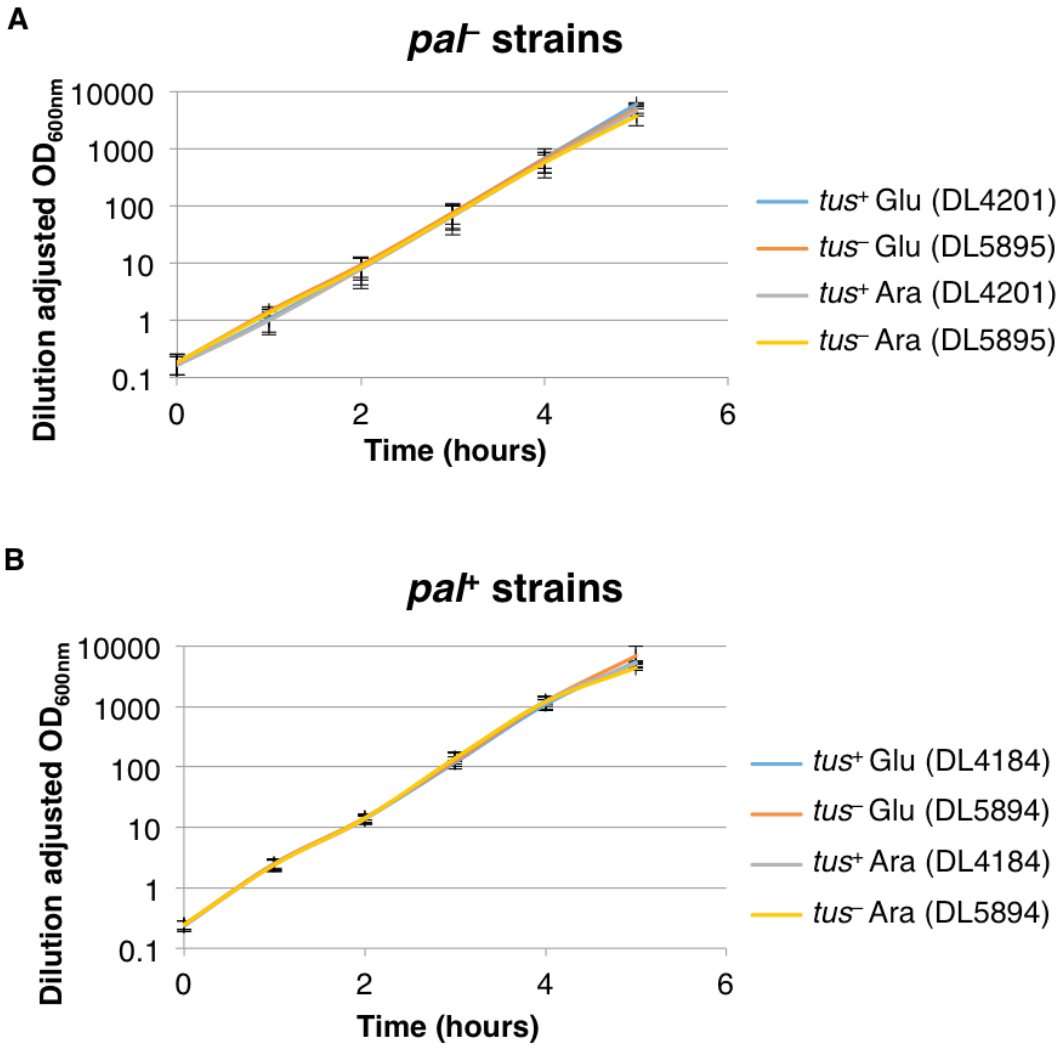


Figure 7.2. Absence of *tus* gene does not affect cell growth.

Growth profile of *tus* mutants subjected or not to DSBR in *lacZ*. Cultures were maintained in exponential phase by diluting them in fresh LB supplemented with either arabinose (Ara) or glucose (Glu) to induce or supress SbcCD expression. Strains used were DL4201 (*pal*⁻), DL4184 (*pal*⁺), DL5894 (*pal*⁺ *tus*⁻) and DL5895 (*pal*⁻ *tus*⁻). (A) Strains that do not contain the palindrome in *lacZ*; (B) Strains that contain the palindrome in *lacZ*.

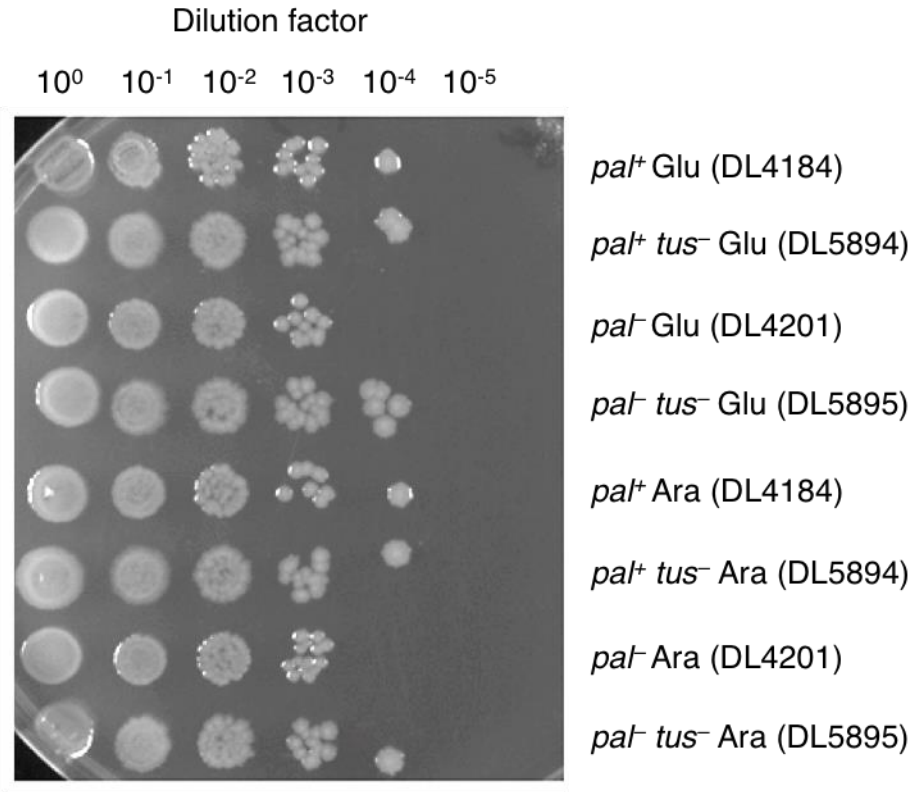


Figure 7.3. Viability of strains is not affected when a *tus* gene is deleted.

Spot tests of 10-fold serially diluted samples collected after 5 hours of exponential growth in presence of glucose (Glu) or arabinose (Ara) and spotted onto LB plates supplemented with 0.5% glucose to repress SbcCD expression. Strains used were DL4201 (*pal*$^{-}$), DL4184 (*pal*$^{+}$), DL5894 (*pal*$^{+}$ *tus*$^{-}$) and DL5895 (*pal*$^{-}$ *tus*$^{-}$).

7.3 Removal of Tus/*ter* blocks eliminates one-ended DSBs at *ter* sites

To demonstrate that RecBCD-mediated one-ended RecA enrichment near *ter* sites is Tus-dependent, I performed RecA-ChIP-seq on strains that lack the *tus* gene (Figure 7.4). Overnight cultures of *pal*$^{-}$, *pal*$^{+}$, *pal*$^{-}$ *tus*$^{-}$ and *pal*$^{+}$ *tus*$^{-}$ strains were diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown to an OD_{600nm} of 0.2-0.25. Then, the cultures were diluted again to an OD_{600nm} of 0.01 and 0.2% arabinose was added to induce SbcCD expression. Afterwards, these cultures were grown for 3.5 h. The OD_{600nm} was regularly monitored and the cultures were diluted in order to keep them at an OD_{600nm} of 0.2-0.3. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform by Edinburgh Genomics. The raw ChIP-seq data obtained were analysed as described in Chapter 2. As expected, RecA enrichment was not observed at *ter* sites in *tus* mutants with or without a DSB in *lacZ.* This means that observed one-ended DSBs adjacent to *ter* sites are due to paused replication forks at these sites. However, the previously observed RecA enrichment around *dif* remained at the same position, demonstating that the DSBs around *dif* are independent of Tus-*ter* blocks.

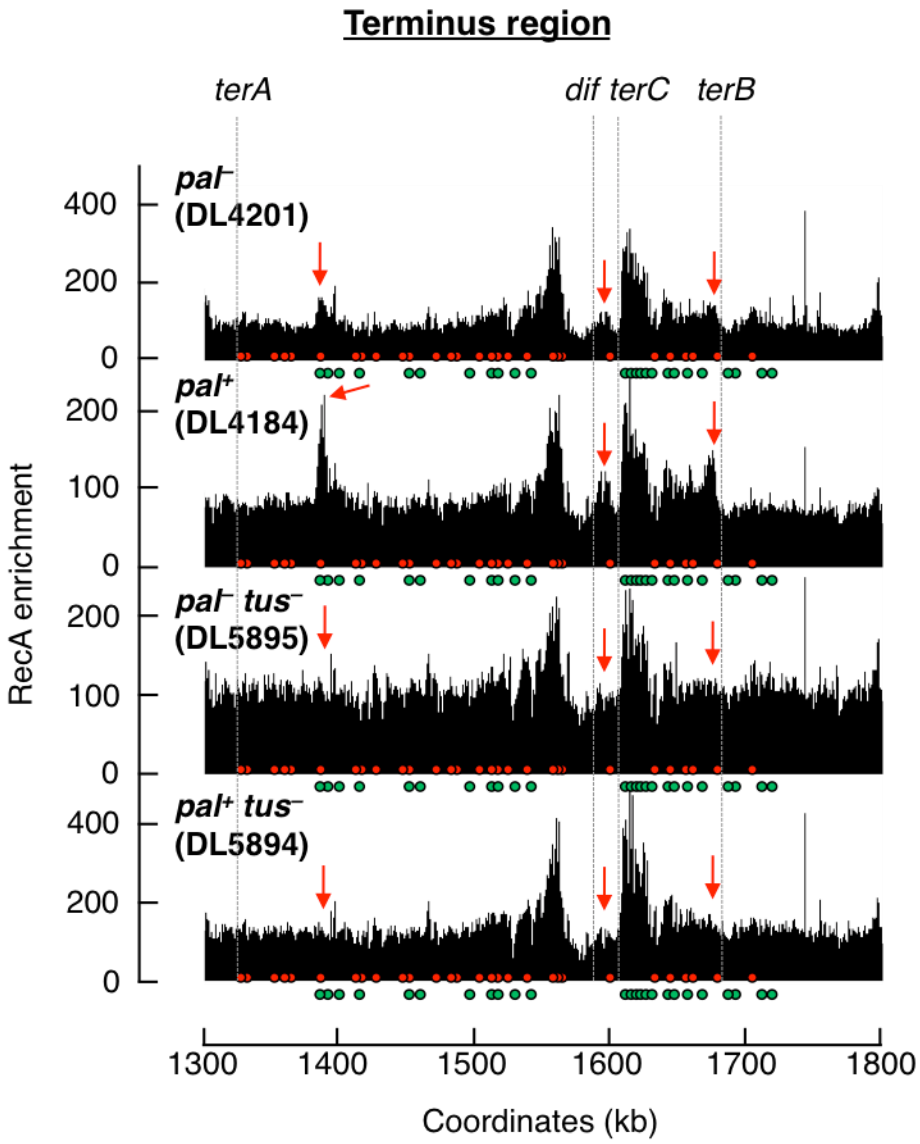


Figure 7.4. Removal of Tus/*ter* blocks eliminates one-ended DSBs at *ter* sites.

ChIP-seq analyses revealed that when *tus* is deleted RecA enrichment that corresponds to *ter* sites disappears. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4184 (*pal*$^{+}$); DL4201 (*pal*$^{-}$); DL5895 (*pal*$^{-}$ *tus*$^{-}$) and DL5894 (*pal*$^{+}$ *tus*$^{-}$). Scale is different.

7.4 DSBs in the terminus region occur close to the *dif* site

7.4.1 A recD mutant permits the localisation of the DSB in the dif region of the terminus

To map the location of the DSBs in the *dif* region of the terminus, I carried out RecA ChIP-seq of a *tus*⁻ *recD*⁻ double mutant. In this double mutant, the *tus* mutation allows the elimination of one-ended DSBs at *ter* sites and the *recD* mutation permits RecA loading directly at break sites. Therefore, RecA ChIP-seq analysis in this strain should reveal RecA binding at the loci of the DSBs in the *dif* region of the terminus. An overnight culture of the *pal*⁻ *tus*⁻ *recD*⁻ strain was diluted in LB medium supplemented with 0.5% glucose to an OD_{600nm} of 0.01 and grown to an OD_{600nm} of 0.2-0.25. Then, the culture was diluted again to an OD_{600nm} of 0.01 and left to grow until an OD_{600nm} of 0.2-0.3. Afterwards, all proteins were crosslinked to DNA with formaldehyde, cells were harvested by centrifugation and ChIP was performed as described in Chapter 2. Then, the DNA was purified, libraries were made and the DNA was sequenced on an Illumina HiSeq 4000 platform by Edinburgh Genomics.

This RecA-ChIP-seq experiment revealed that in a *tus*⁻ *recD*⁻ double mutant DSBs in the terminus occurs at or very close to the *dif* site (zoomed region in Figure 7.5). This DSB looks different from a palindrome-mediated DSB in *lacZ* (Figure 7.1 A), where the gap between two sides of the DSB was V-shaped that pointed directly at the palindrome with the same amount of RecA on each side of the palindrome. The RecA distribution at *dif* looks V-shaped but, on the contrary, the loaded amount of RecA on each side of *dif* was different: higher on the right side from *dif* than on the left side. This might be due to more RecA on the right side of *dif* overall (Figure 7.5, *pal*⁻ *tus*⁻ panel). The backround signal on the right side looks higher than the backround signal in the left side of *dif*.

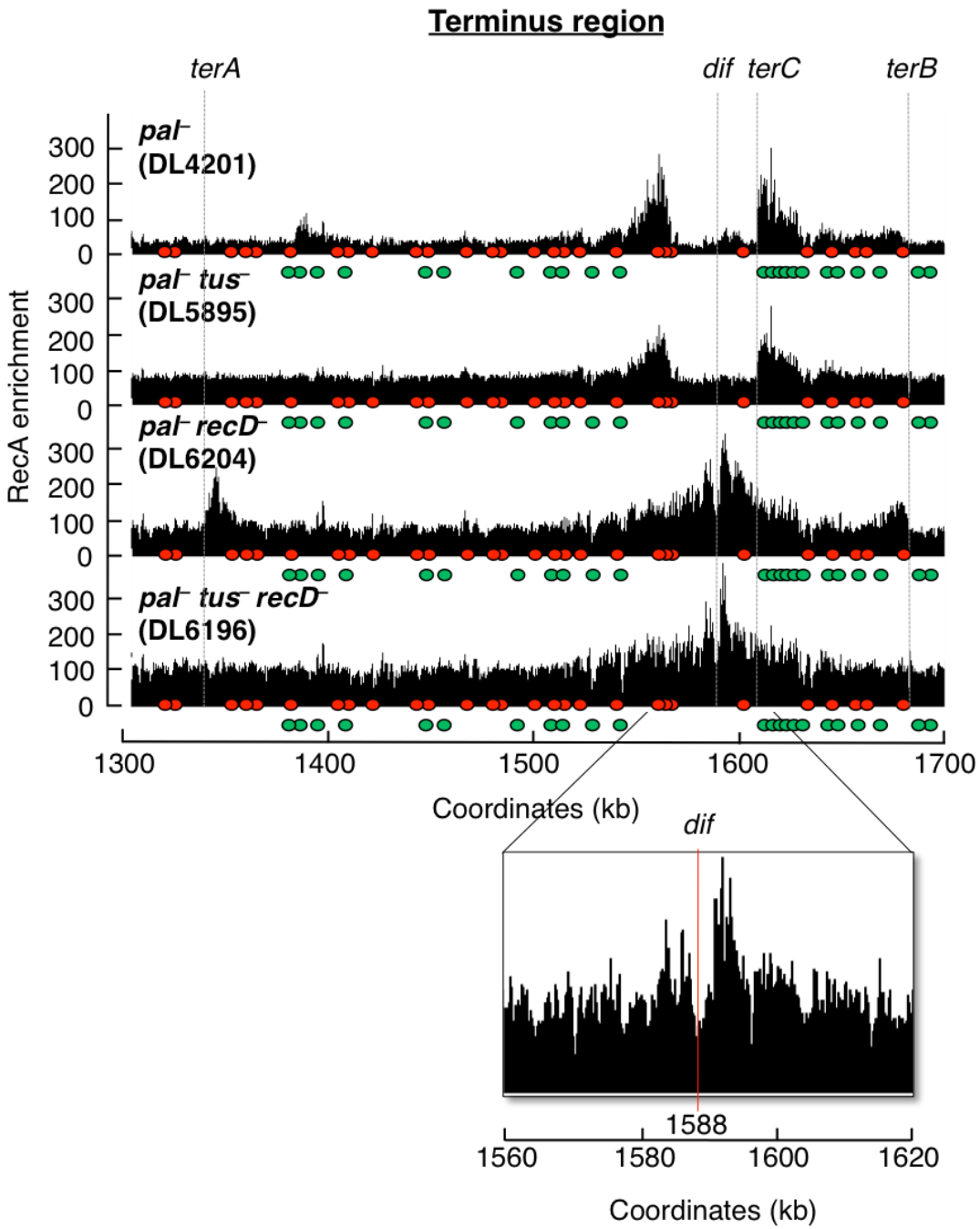


Figure 7.5. DSBs in the terminus occurs at or close to the *dif* site.

A ChIP-seq analysis of the *tus*⁻ *recD*⁻ double mutant revealed that DSBs in the terminus region occurs at or close to the *dif* site. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal*⁻), DL5895 (*pal*⁻ *tus*⁻), DL6204 (*pal*⁻ *recD*⁻), which was prepared and analysed by Benura Azeroglu and DL6196 (*pal*⁻ *tus*⁻ *recD*⁻).

7.4.2 Removal of the *dif* site leads to the 'guillotining' of chromosomes

To confirm that the DSB occurs at the *dif* site and to test the 'quillotining' hypothesis described in Chapter 5, I performed a ChIP-seq experiment in *dif*⁻ *recD*⁻ double mutant. During 'guillotining', when *E. coli* is unable to resolve dimer of chromosomes, the DNA is sheared in the terminus. If the 'guillotining' hypothesis is correct, then in *dif* mutant I expect to observe DSBs occurring in multiple positions around the place where the location of the *dif* site used to be.

All growth and experimental conditions were the same as described in section 7.4.1. As expected, in the *dif*⁻ *recD*⁻ double mutant, DSBs were distributed for ~100 kb and centered around the region where the *dif* site used to be (Figure 7.6). The distribution of RecA showed that DSBs might be occurring in two positions around the *dif* region. This was concluded from two peaks of RecA enrichment around this site. This result suggests that the 'guillotining' of chromosome dimer that was described in (Hendricks *et al.* 2000) might be correct.

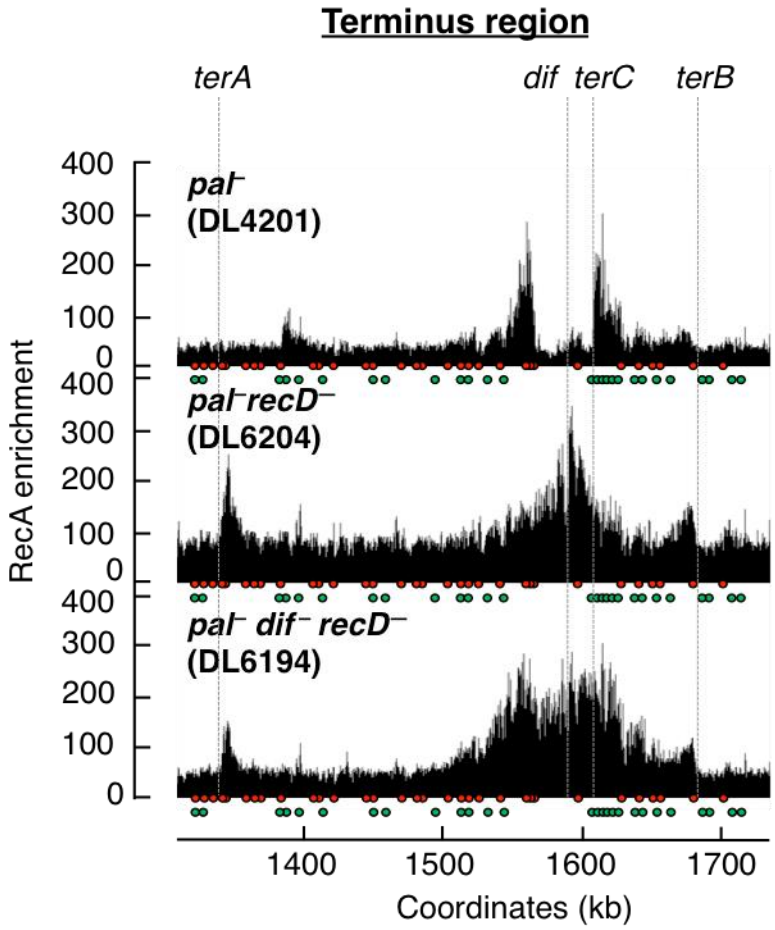


Figure 7.6. Removal of the *dif* site leads to 'guillotining' of chromosome dimers.

A ChIP-seq analysis of the *dif*⁻ *recD*⁻ double mutant revealed that there are more DSBs in the terminus region and that they are distributed across ~100 kb around the position of *dif* locus. The position of Chi sites is shown using red (5'-gctggtgg-3') and green (3'-ccaccagc-5') circles. Green Chi sites are oriented in a such way that RecBCD recognises them if it moves left to right on the chromosome. Red Chi sites are recognised by RecBCD that moves in the opposite direction – right to left. The *dif* site, *terA*, *terB* and *terC* are indicated. Strains used were DL4201 (*pal*⁻), DL6204 (*pal*⁻ *recD*⁻), which was prepared and analysed by Benura Azeroglu and DL6194 (*pal*⁻ *dif*⁻ *recD*⁻).

7.5 Discussion

In this chapter I show the location of two-ended DSBs in the *dif* region of the terminus using a novel tool (*recD* mutant) developed in our laboratory. During HR-dependent DSBR, the RecBCD complex binds to double-strand ends and starts unwinding and degrading DNA until it reaches a Chi site. Because in a *recD*$^{-}$ strain the RecD protein is missing from the RecBCD complex, the unwinding of the DNA is slow and, therefore, RecB starts loading RecA immediately. This tool was used in a combination with a *tus* mutation. These mutants do not form Tus/*ter* blocks and cannot stop the progression of replication forks. Cells that lack Tus proteins were viable and the growth was not distinguishable from that of wild-type cells. ChIP-seq analysis revealed that using the *tus* mutation eliminated one-ended DSBs that occur at *ter* sites. Therefore, the source of observed one-ended DSBs at *ter* sites in *tus*$^{+}$ cells is a blocked replication fork. RecA enrichment from one-ended DSBs at the *terC* site overlaps with the DSBs in *dif* region, so it was crucial to remove this type of RecA enrichment in order to visualise the location of DSBs around the *dif* site. Using a *recD*$^{-}$ *tus*$^{-}$ double mutant, I confirmed that the DSBs in the *dif* region of the terminus occur at or very close to the *dif* site.

The distribution of RecA in the absence of the *dif* site showed that DSBs might be occurring in two positions around the *dif* region. This was concluded from two peaks of RecA enrichment around this site. The peak on the right side of *dif* was difficult to observe due to Tus/*terC* DSBs in that region. Unfortunately, I could not construct *dif tus recD* mutant to observe distinguished peaks better. Interestingly, although *tus*$^{-}$ mutants do not have any observed phenotype, it was reported that RTP (Tus in *E. coli*) facilitates chromosome dimer resolution by RipX (XerD in *E. coli*) in *B. subtilis* (Lemon, Kurtser, and Grossman 2001). Similarly, Tus could be involved in XerCD/*dif* dimer resolution in *E. coli*.

Absence of Tus/*ter* block could lead to more severe defects in the dimer resolution, therefore, I could not construct this mutant.

Because the RecA distribution was different in the *dif*$^{-}$ *recD*$^{-}$ mutant compared to the *dif*$^{+}$ *recD*$^{-}$ strain, the mechanism of the DSB formation might be different as well. In the *dif*$^{+}$ strain, the DSBs might be caused by either XerCD or TopoIV/TopoIII that misfunction. When the *dif* site is absent, the DSBs could be caused by the 'guillotining' of chromosome dimer described in Chapter 5. When the *dif* site is not present in the terminus region, XerCD complex is not able to bind to the terminus and resolve a chromosome dimer. That is why the chromosome dimer is broken by a septum that forms there. The 'guillotining' of dimer chromosomes was reported previously (Hendricks *et al*. 2000), although it has not been shown that this broken chromosome dimer could be repaired.

Although, in Chapter 6, I described that TopoIV is partially responsible for 'guillotining', therefore, both mechanisms of DSBs formation in the terminus in the presence or in the absence of *dif* might be TopoIV/TopoIII dependent.

Chapter VIII

8. Discussion

The aim of this thesis was to investigate a cause of additional DSBs formation around the *dif* site and of DNA over-replication between *terA* and *terB* sites in the terminus region of *Esherichia coli.* The terminus region of *E. coli* is the location of multiple processes that are crucial for cell survival. These processes include DNA replication termination, chromosome dimer resolution, chromosome decatenation and chromosome segregation. The first three of these processes were tested as potential candidates for the formation of DSBs and over-replication in the *E. coli* terminus region. Also, it was tested whether the over-replication and the formation of DSBs in the terminus were linked to each other.

Lawrence and collaborators suggested that DNA replication termination does not occur at *ter* sites but at the *dif* site. Their proposed model suggested that a nick would occur at *dif* which would create a DSB when encountered by replication forks (Lawrence 2007). Duggin and collaborators demonstrated that a significant proportion of termination events occurs at the *terC* site in wild type cells and argued against a possible termination at *dif* where no paused forks and less termination were detected (Duggin and Bell 2009b).

In accordance with Lawrence and collaborators and opposed to Duggin and collaborators, a DSB event was observed in the terminus around the position where replication forks are expected to terminate (*dif*) in wild type cells. On the other hand, opposed to Lawrence and collaborators, where they proposed that this DSB might be replication-dependent, in this study, I showed that this DSB event was not associated with fork collision. This was achieved by delaying one of the replication forks and, thus, changing the position of the termination of replication, which did not change the position of DSBs in the terminus. More experiments could be done to confirm this observation, for

example introducing an additional *ter* site between *dif* and *terA* in order to shift the termination from *dif*.

Interestingly, the *terC* site, which is positioned on the chromosome to catch forks travelling clockwise earlier than *terB*, was not functioning properly allowing a majority of forks to pass through in a strain subjected to DSBs in *lacZ*. According to previous studies a significant amount of paused replication forks was observed at *terB* and *terA* sites even when the distance between them was increased (Pelletier, Hill, and Kuempel 1988). The proposed hypothesis explaining this phenomenon was that individual replichores compensate somehow for the distance difference. One possible explanation for this observation could be that extra origins in the terminus would send additional replication forks towards *terA* and *terB* sites (Magee *et al*. 1992). Here, I propose an alternative explanation for the appearance of DNA over-replication in the terminus that does not depend on new origins but on the formation of DSBs. Either a DSB is formed at *dif*, which creates an over-replication, or the termination of replication is less coordinated due to the ongoing DSBR on one of the replichores and this leads to more forks pausing at *ter* sites and creating an over-replication (Figure 8.1).

Additionally, Lawrence and collaborators indicated that *ter* sites primarily act to pause replication arising from DNA repair events not associated with chromosome replication (Lawrence 2007). Duggin and collaborators argued that 'there is no reason to suggest that repair-associated replication would only take place in the terminus region' (Duggin and Bell 2009b). Here, we complete the hypothesis of Lawrence and collaborators by showing that the *ter* sites act to halt replication arising from DSB events in the terminus region as well as replication forks that are produced by *oriC*.

Therefore, in this work, we are proposing an advantage of the Tus/*ter* block system for the bacterial chromosomes. Since observed DSBs in the terminus region might generate additional replication forks moving from the terminus to the origin, it would be required to stop this progression in order to avoid the collision of these forks with the transcriptional machinery.

In this study, I observed that the longer the cells were subjected to DSBs at *lacZ* the higher was the level of DSBs in the terminus. DNA over-replication observed between *terA* and *terB* sites was also increased with the time of DSB induction in *lacZ*. This suggests that the DSBs in the terminus are linked to DNA over-replication in the terminus. In this work, I propose a model that shows that DSBs at the *dif* site cause an over-replication between *terA* and *terB* sites (Figure 8.1). This model describes two types of observed DSBs in the terminus: DSBs at *dif* and DSBs at *ter* sites. DSBs at *dif* are formed by TopoIV and TopoIII and could be divided into two types: focused by XerCD/*dif* system and unfocused (in the absence of XerCD/*dif* system). One-ended DSBs at *ter* sites are a consequence of DNA over-replication in the terminus and are formed when a new round of replication forks runs into already stalled replication forks at *ter* sites. Bidnenko and collaborators described in detail how this type of DSBs could be formed (Bidnenko, Ehrlich, and Michel 2002).

Overall, I showed, that initially observed DSBs at the *dif* site by Dr. Charlotte Cockram could be a cause of over-replication in the terminus (Cockram *et al*. 2015) (Chapter 4). Indeed, following DSBs in *lacZ,* the amount of DSBs at the *dif* site was increased as well as DNA amount in the terminus (Chapter 3), suggesting that these DSBs contribute to the over-replication in the terminus. In Chapter 7, I showed that DSBs are centered at *dif* and, in Chapter 3, I showed that DNA over-replication originates from the *dif* site, when DSBs in *lacZ* are induced. This result confirms that over-replication in the terminus is linked to

DSBs formation at *dif*. Also, I observed that the level of one-ended DSBs at *ter* sites is increased following DSBs induction in *lacZ*. Because both kinds of DSBs, one-ended at *ter* sites and, presumably, two-ended at the *dif* site, are palindrome-stimulated, I suggest that these DSBs are linked.

During the investigation of the cause of initial DSBs at the *dif* site observed by Dr. Charlotte Cockram (Cockram *et al*. 2015), I decided to eliminate the FtsK/XerCD/*dif* dimer resolution system. Homologous recombination at the palindrome site generates more dimers, therefore, it was essential to test dimer resolution system as a candidate for DSBs formation at *dif* (Chapter 5). Removal of the XerCD/*dif* system led to the formation of DSBs, previously described as 'guillotining', distributed around the *dif* region. This type of DSBs also leads to DNA over-replication in the terminus, presumably, through uncoordinated re-invasion of double-strand ends into the intact duplex and setting up new replication forks. This suggests that, in fact, the presence of the FtsK/XerCD/*dif* system prevents the formation of a large number of DSBs in the terminus and that removing this system leads to severe defects in chromosome segregation, which force chromosomes to separate using other mechanisms.

Here, I showed that one of these mechanisms that improperly resolves chromosome dimers in the absence of the FtsK/XerCD/*dif* system is topoisomerase IV (Chapter 6). It was reported that TopoIV primarily acts at *dif* and is attracted there by FtsK (El Sayyed *et al*. 2016; Espeli, Lee, and Marians 2003). When TopoIV and XerCD/*dif* are inactive, the amount of DSBs closer to *dif* is reduced, suggesting that TopoIV is partially involved in 'guillotining' of chromosomes. This type of DSBs also led to an elevated amount of DNA between *terA* and *terB*.

Interestingly, TopoIII also showed that it might be involved in DSBs formation at *dif*. However, the effect of TopoIII elimination was not as easily observed as the effect of TopoIV inactivation, presumably, because TopoIV is the primary decatenation machinery. That TopoIV and TopoIII act synergistically was confirmed by the inactivation of both enzymes. When both TopoIV and TopoIII were eliminated, the level of DSBs at *dif* was significantly decreased, leading to a conclusion that both TopoIV and TopoIII are responsible for the formation of DSBs at *dif* and could be involved in the removal of a potentially unresolvable and problematic situation of over-replication in the terminus of the *E. coli* chromosome (Figure 8.1).

Finally, with the proposed model, I suggest that the primary cause of DSBs at *dif* is TopoIV and TopoIII topoisomerases. During fast growing conditions or when DSBs are induced elsewhere in the chromosome (e.g. *lacZ* or *ascB*), more chromosome dimers could be formed. The XerCD/*dif* dimer resolution system might be over-loaded with dimers and TopoIII and TopoIV are required to resolve the potential problem. After the termination of replication, the XerCD/*dif* system focuses both topoisomerases at *dif* and they form DSBs there. This creates two-ended DSBs, which, with the help of RecBCD-mediated DSBR, could be invaded back to the intact chromosome and could be repaired by the homologous recombination. However, the re-invasion of these double-stranded ends is not coordinated, which allows the replication forks to be set up that move in the opposite directions. This leads to DNA over-replication in the terminus (between *terA* and *terB*). This over-replication creates 3 chromosomes in the terminus instead of two and has to be resolved. Because three chromosomes are not a substrate for the XerCD/*dif* system, topoisomerases are involved in this cleavage leaving double-stranded ends. Again, RecBCD-mediated machinery is involved to repair these DSBs setting

up new round of replication forks in the opposite directions. This leads to another round of over-replication, which, upon reaching the *ter* sites, will create double-stranded ends. Indeed, one-ended DSBs were observed at *ter* sites. This leads to the appearance of a linear chromosome that could be degraded leaving two forks that pause at *ter* sites. These two forks can be later resolved by the new round of replication that arrives from the origin (Figure 8.1).

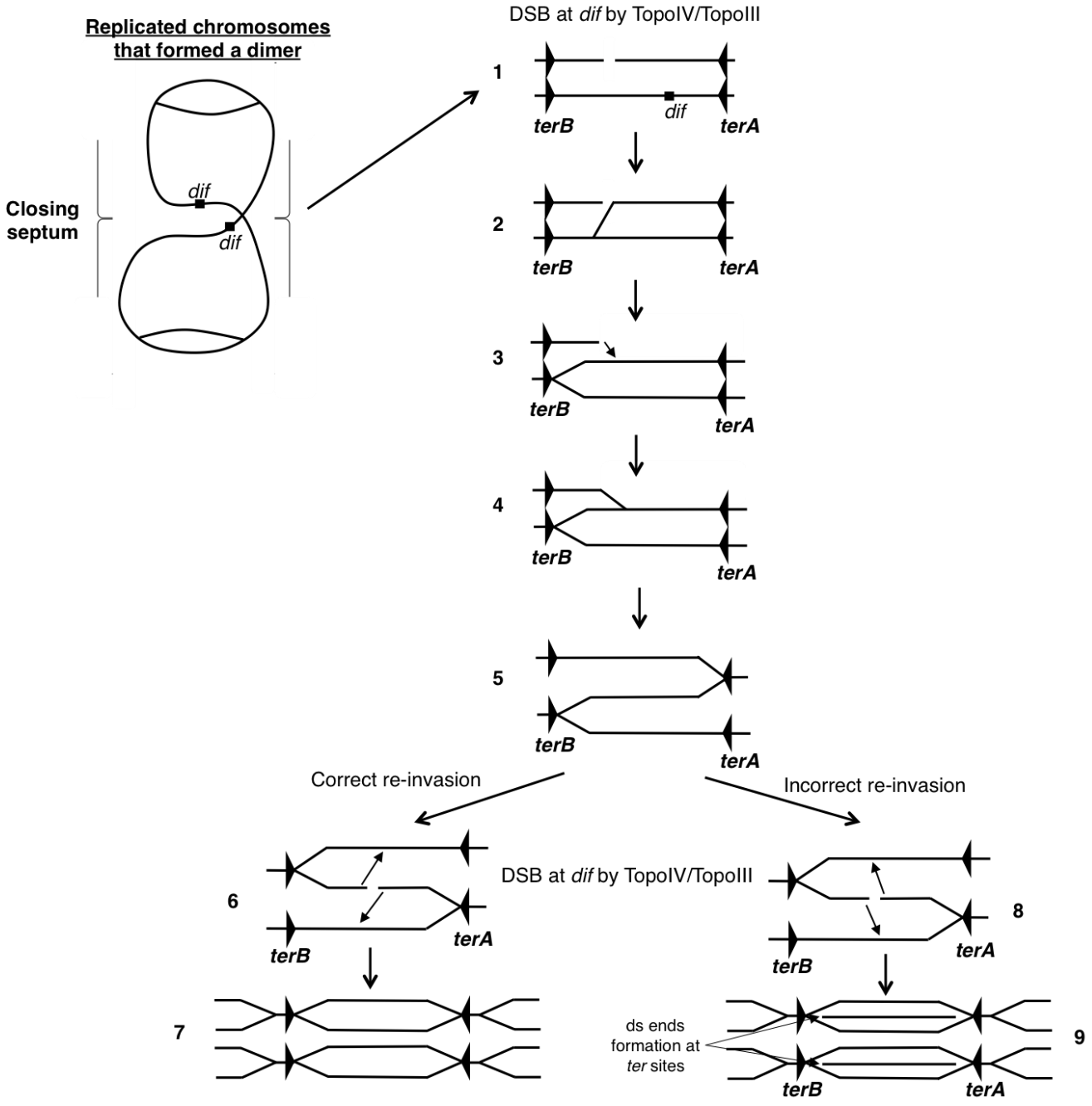

Replicated chromosomes that formed a dimer
Closing septum
dif
dif
DSB at dif by TopoIV/TopoIII
1
dif
terB
terA
2
terB
terA
3
terB
terA
4
terB
terA
5
terB
terA
Correct re-invasion
Incorrect re-invasion
DSB at dif by TopoIV/TopoIII
6
terB
terA
7
8
terB
terA
ds ends formation at ter sites
9
terB
terA

Figure 8.1. Proposed model for DSBs formation and DNA over-replication in the terminus region of the *E.coli* chromosome.

When the chromosome dimer is formed and the *dif* sites are not yet aligned, topoisomerases (TopoIV and TopoIII) are involved at the DSB formation at one of the *dif* sites (1). This DSB could be repaired by homologous recombination. Due to the uncoordinated invasion of replication forks mediated by RecA, one of the forks is set up, which starts replicating towards *terB* (2). Meanwhile, the second replication fork is re-invaded in the intact sister chromosome, which started the DNA replication towards *terA* (3). This creates an over-replication between *terA* and *terB* (4), which leaves two chromosomes interlinked with each other (5). Therefore, another DSB is required to resolve this problem. TopoIII and TopoIV cleave the middle chromosome (middle chromosome needs to be cleaved otherwise it will lead to the same interlinked structure) at *dif* leaving a DSB that can be repaired via HR. Then, two scenarios could occur. If double-stranded ends are correctly re-invaded, each to their own chromosome, it will lead to the formation of two replication forks that will pause at *terA* and *terB*. These forks are then resolved by the next round of replication forks that travel from the origin (6-7). If double-stranded ends are invaded in the incorrect way, using the opposite chromosome, it will lead to the formation of the replication forks that will run into the forks that are already paused at *terA* and *terB* sites. This situation will lead to the formation of a linear chromosome with free double-stranded ends at *ter* sites (8). This chromosome then could be degraded by RecBCD leaving two forks paused at *ter* sites that could be resolved by the next round of replication from the origin (9).

References

Abe, Yoshito, Takaaki Jo, Yusaku Matsuda, Chika Matsunaga, Tsutomu Katayama, and Tadashi Ueda. 2007. “Structure and Function of DnaA N-Terminal Domains: Specific Sites and Mechanisms in Inter-DnaA Interaction and in DnaB Helicase Loading on *oriC*.” *Journal of Biological Chemistry* 282 (24): 17816–27.

Adams, E. David, Eugene M. Shekhtman, E. Lynn Zechiedrich, Molly B. Schmid, and Nicholas R. Cozzarelli. 1992. “The Role of Topoisomerase IV in Partitioning Bacterial Replicons and the Structure of Catenated Intermediates in DNA Replication.” *Cell* 71 (2): 277–88.

Alver, Robert C., and Anja-Katrin Bielinsky. 2010. “Termination at sTop2.” *Molecular Cell* 39 (4). Elsevier: 487–89.

Asai, Tsuneaki, Suzanne Sommer, Adriana Bailone, and Tokio Kogoma. 1993. “Homologous Recombination-Dependent Initiation of DNA Replication from DNA Damage-Inducible Origins in *Escherichia Coli*.” *The EMBO Journal* 12 (8): 3287–95.

Aussel, Laurent, François-Xavier Barre, Mira Aroyo, Andrzej Stasiak, Alicja Z. Stasiak, and David Sherratt. 2002. “FtsK Is a DNA Motor Protein That Activates Chromosome Dimer Resolution by Switching the Catalytic State of the XerC and XerD Recombinases.” *Cell* 108 (2): 195–205.

Azeroglu, Benura, Julia S. P. Mawer, Charlotte A. Cockram, Martin A. White, A. M. Mahedi Hasan, Milana Filatenkova, and David R. F. Leach. 2016. “RecG Directs DNA Synthesis during Double-Strand Break Repair.” *PLoS Genetics* 12 (2): 1–23.

Balakrishnan, Lata, and Robert A. Bambara. 2013. “Okazaki Fragment Metabolism.” *Cold Spring Harbor Perspectives in Biology* 5 (2): 1–12.

Barre, François-Xavier, Britta Søballe, Bénédicte Michel, Mira Aroyo, Malcolm Robertson, and David Sherratt. 2001. “Circles: The Replication-Recombination-Chromosome Segregation Connection.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (15): 8189–95.

Bidnenko, Vladimir, S. Dusko Ehrlich, and Bénédicte Michel. 2002. “Replication Fork Collapse at Replication Terminator Sequences.” *EMBO Journal* 21 (14): 3898–3907.

Bidnenko, Vladimir, Roxane Lestini, and Bénédicte Michel. 2006. “The *Escherichia Coli* UvrD Helicase Is Essential for Tus Removal during Recombination-Dependent Replication Restart from *Ter* Sites.” *Molecular Microbiology* 62 (2): 382–96.

Bigot, Sarah, Jacqueline Corre, Jean Michel Louarn, François Cornet, and François -Xavier Barre. 2004. “FtsK Activities in Xer Recombination, DNA Mobilization and Cell Division Involve Overlapping and Separate Domains of the Protein.” *Molecular Microbiology* 54 (4): 876–86.

Bigot, Sarah, and Kenneth J. Marians. 2010. “DNA Chirality-Dependent Stimulation of Topoisomerase IV Activity by the C-Terminal AAA+ Domain of FtsK.” *Nucleic Acids*

Research 38 (9): 3031–40.

Bigot, Sarah, Omar A. Saleh, François Cornet, Jean-François Allemand, and François-Xavier Barre. 2006. “Oriented Loading of FtsK on KOPS.” *Nature Structural & Molecular Biology* 13 (11): 1026–28.

Bigot, Sarah, Omar A. Saleh, Christian Lesterlin, Carine Pages, Meriem El Karoui, Cynthia Dennis, Mikhail Grigoriev, Jean-François Allemand, François-Xavier Barre, and François Cornet. 2005. “KOPS: DNA Motifs That Control *E. Coli* Chromosome Segregation by Orienting the FtsK Translocase.” *The EMBO Journal* 24 (21): 3770–80.

Blakely, Garry W., Anne O. Davidson, and David Sherratt. 1997. “Binding and Cleavage of Nicked Substrates by Site-Specific Recombinases XerC and XerD.” *Journal of Molecular Biology* 265 (1): 30–39.

Blattner, Frederick R., Guy Plunkett III, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides. Jeremy D. Galsner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, Ying Shao. 1997. “The Complete Genome Sequence of *Escherichia Coli* K-12 .” *Science* 277 (September): 1453–62.

Boubakri, Hasna, Anne Langlois, De Septenville, and Enrique Viguera. 2010. “The Helicases DinG, Rep and UvrD Cooperate to Promote Replication across Transcription Units *in Vivo*”. *EMBO Journal* 29 (1): 145–57.

Brewer, Bonita J., and Walton L. Fangman. 1988. “A Replication Fork Barrier at the 3’ End of Yeast Ribosomal RNA Genes”. *Cell* 55: 637–43.

Carra, Claudio, and Francis A. Cucinotta. 2011. “Binding Selectivity of RecA to a Single Stranded DNA, a Computational Approach.” *Journal of Molecular Modeling* 17 (1): 133–50.

Cassuto, Era, Stephen C. West, Jane Mursalim, Stephen Conlon, and Paul Howard-Flanders. 1980. “Initiation of Genetic Recombination: Homologous Pairing between Duplex DNA Molecules Promoted by recA Protein.” *Proceedings of the National Academy of Sciences* 77 (7): 3962–66.

Champoux, James J. 2001. “DNA Topoisomerases: Structure, Function, and Mechanism.” *Annu. Rev. Biochem.* 70: 369–413.

Churchill, Jason J., Daniel G. Anderson, and Stephen C. Kowalczykowski. 1999. “The RecBC Enzyme Loads recA Protein onto ssDNA Asymmetrically and Independently of Chi, Resulting in Constitutive Recombination Activation.” *Genes and Development* 13 (7): 901–11.

Clark, Alvin J., and Ann Dee Margulies. 1965. “Isolation and Characterization of Recombination-Deficient Mutants of *Escherichia Coli* K12.” *Proceedings of the National Academy of Sciences* 53 (1): 451–59.

Cockram, Charlotte A, Milana Filatenkova, Vincent Danos, Meriem El Karoui, and David R. F. Leach. 2015. “Quantitative Genomic Analysis of RecA Protein Binding during DNA

Double-Strand Break Repair Reveals RecBCD Action *in Vivo*." *Proceedings of the National Academy of Sciences of the United States of America* 112 (34): E4735-42.

Cornet, Francois, Jacqueline Louarn, Josette Patte, and Jean-Michel Louarn. 1996. "Restriction of the Activity of the Recombination Site *Dif* to a Small Zone of the *Escherichia Coli* Chromosome." *Genes & Development* 10: 1152–61.

Cox, Michael M. 2007. "Regulation of Bacterial RecA Protein Function." *Critical Reviews in Biochemistry and Molecular Biology* 42 (1): 41–63.

Dame, Remus T., Olga J. Kalmykowa, and David C. Grainger. 2011. "Chromosomal Macrodomains and Associated Proteins: Implications for DNA Organization and Replication in Gram Negative Bacteria." *PLoS Genetics* 7 (6).

David J. Sherratt, Ivy F. Lau, and Francois-Xavier Barre. 2001. "Chromosome Segregation." *Current Opinion in Microbiology* 4: 653–59.

de Massy, Bernard, Josette Patte, Jean-Michel Louarn, and Jean-Pierre Bouche 1984. "*oriX*: A New Replication Origin in *E. Coli*." *Cell* 36 (1): 221–27.

Deepak Bastia, Shamsu Zzaman, Gregor Krings, Mukesh Saxena, Xiaohua Peng, and Marc M. Greenberg. 2008. "Replication Termination Mechanism as Revealed by Tus-Mediated Polar Arrest of a Sliding Helicase." *Proceedings of the National Academy of Sciences of the United States of America* 105 (35): 12831–36.

DiGate, Russell J., and Kenneth J. Marians. 1988. "Identification of a Potent Decatenating Enzyme from *Escherichia Coli*." *Journal of Biological Chemistry* 263 (26): 13366–73.

DiGate, Russell J., and Kenneth J. Marians. 1989. "Molecular Cloning and DNA Sequence Analysis of *Escherichia Coli topB*, the Gene Encoding Topoisomerase III ." *The Journal of Biological Chemistry* 264 (30): 17924–30.

Dillingham, Mark S., and Stephen C. Kowalczykowski. 2008. "RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks." *Microbiology and Molecular Biology Reviews : MMBR*.

Dixon, Dan A., and Stephen C. Kowalczykowski. 1993. "The Recombination Hotspot Chi Is a Regulatory Sequence That Acts by Attenuating the Nuclease Activity of the *E. Coli* RecBCD Enzyme." *Cell* 73 (1): 87–96.

Duggin, Iain G., and Stephen D. Bell. 2009. "Termination Structures in the *Escherichia Coli* Chromosome Replication Fork Trap." *Journal of Molecular Biology* 387 (3): 532–39.

Duggin, Iain G., R. Gerry Wake, Stephen D. Bell, and Thomas M. Hill. 2008. "The Replication Fork Trap and Termination of Chromosome Replication." *Molecular Microbiology* 70 (6): 1323–33.

Duggin, Iain G, Nelly Dubarry, and Stephen D. Bell. 2011. "Replication Termination and Chromosome Dimer Resolution in the Archaeon *Sulfolobus Solfataricus*." *The EMBO Journal* 30 (1). Nature Publishing Group: 145–53.

Duggin, Iain G, Simon A. McCallum, and Stephen D. Bell. 2008. "Chromosome Replication Dynamics in the Archaeon *Sulfolobus Acidocaldarius*." *Proceedings of the National*

Academy of Sciences of the United States of America 105 (43): 16737–42.

El Sayyed, Hafez, Ludovic Le Chat, Elise Lebailly, Elise Vickridge, Carine Pages, Francois Cornet, Marco Cosentino Lagomarsino, and Olivier Espéli. 2016. “Mapping Topoisomerase IV Binding and Activity Sites on the *E. Coli* Genome.” *PLOS Genetics* 12 (5): e1006025.

Espeli, Olivier, Chong Lee, and Kenneth J. Marians. 2003. “A Physical and Functional Interaction between *Escherichia Coli* FtsK and Topoisomerase IV.” *Journal of Biological Chemistry* 278 (45): 44639–44.

Espeli, Olivier, Cindy Levine, Heide Hassing, and Kenneth J. Marians. 2003. “Temporal Regulation of Topoisomerase IV Activity in *E. Coli.*” *Molecular Cell* 11 (1): 189–201.

Fachinetti, Daniele, Rodrigo Bermejo, Andrea Cocito, Simone Minardi, Yuki Katou, Yutaka Kanoh, Katsuhiko Shirahige, Anna Azvolinsky, Virginia A. Zakian, and Marco Foiani. 2010. “Replication Termination at Eukaryotic Chromosomes Is Mediated by Top2 and Occurs at Genomic Loci Containing Pausing Elements.” *Molecular Cell* 39 (4). Elsevier: 595–605.

Fatma Filiz Coscun-Ari, and Thomas M. Hill. 1997. “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): 26448–56.

French, Sarah. 1992. “Consequences of Replication Fork Movement through Transcription Units *in Vivo*.” *Science* 258 (November): 1362–65.

Frick, David N., and Charles C. Richardson. 2001. “DNA Primases.” *Annu. Rev. Biochem.* 70: 39–80.

Friedman Katherine L., Bonita J. Brewer. 1995. “Analysis of Replication Intermediates by Two-Dimensional Agarose Gel Electrophoresis.” *Methods Enzymol.* 262 (1983): 613–27.

Grainge, Ian, Migena Bregu, Mariel Vazquez, Viknesh Sivanathan, Stephen CY Ip, and David J. Sherratt. 2007. “Unlinking Chromosome Catenanes *in Vivo* by Site-Specific Recombination.” *The EMBO Journal* 26 (19): 4228–38.

Grainge, Ian, Christian Lesterlin, and David J. Sherratt. 2011. “Activation of XerCD-*Dif* Recombination by the FtsK DNA Translocation." *Nucleic Acid Res.* 39(12):5140-5148.

Greenfeder, Scott A., and Carol S. Newlon. 1992. “A Replication Map of a 61-Kb Circular Derivative of *Saccharomyces Cerevisiae* Chromosome *III.*” *Molecular Biology of the Cell* 3 (9): 999–1013.

Haber, James E. 2000. “Partners and Pathways - Repairing a Double-Strand Break.” *Trends in Genetics* 16 (6): 259–64.

Harmon, Frank G., Joel P. Brockman, and Stephen C. Kowalczykowski. 2003. “RecQ Helicase Stimulates Both DNA Catenation and Changes in DNA Topology by Topoisomerase III.” *Journal of Biological Chemistry* 278 (43): 42668–78.

Harmon, Frank G., Russell J. DiGate, and Stephen C. Kowalczykowski. 1999. “RecQ Helicase and Topoisomerase III Comprise a Novel DNA Strand Passage Function: A Conserved

Mechanism for Control of DNA Recombination." *Molecular Cell* 3 (5): 611–20.

Helene Bierne, S. Dusko Ehrlich, and Benedicte Michel. 1991. "The Replication Termination Signal *terB* of the *Escherichia Coli* Chromosome Is a Deletion Hot Spot." *The EMBO Journal* 10 (9): 2699–2705.

Hendricks, E. Cale, Heather Szerlong, Thomas Hill, and Peter Kuempel. 2000. "Cell Division, Guillotining of Dimer Chromosomes and SOS Induction in Resolution Mutants (*Dif*, *xerC* and *xerD*) of *Escherichia Coli.*" *Molecular Microbiology* 36 (4): 973–81.

Henson, Joan M., and Peter L. Kuempel. 1985. "Deletion of the Terminus Region (340 Kilobase Pairs of DNA) from the Chromosome of *Escherichia Coli*." *Proceedings of the National Academy of Sciences of the United States of America* 82 (11): 3766–70.

Hiasa, Hiroshi, and Kenneth J. Marians. 1994. "Tus Prevents Overreplication of *oriC* Plasmid DNA." *Journal of Biological Chemistry* 269 (43): 26959–68.

Hill, Thomas M. 1992. "Arrest of Bacterial DNA Replication." *Annu Rev Microbiol* 46: 603–33.

Hill, Thomas M., Joan M. Henson, and Peter L. Kuempel. 1987. "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): 1754–58.

Hill, Thomas M., and Kenneth J. Marians. 1990. "*Escherichia Coli* Tus Protein Acts to Arrest the Progression of DNA Replication Forks *in Vitro*." *Proceedings of the National Academy of Sciences of the United States of America* 87 (April): 2481–85.

Hill, Thomas M., Anthony J. Pelletier, Marianne L. Tecklenburg, and Peter L. Kuempel. 1988. "Identification of the DNA Sequence From the *Escherichia Coli* Terminus Region That Halts Replication Forks." *Cell* 55 (3): 459–66.

Hoffman, Brad G., and Steven J. M. Jones. 2009. "Genome-Wide Identification of DNA-Protein Interactions Using Chromatin Immunoprecipitation Coupled with Flow Cell Sequencing." *Journal of Endocrinology* 201 (1): 1–13.

Hojgaard, Andrias, Heather Szerlong, Camille Tabor, and Peter Kuempel. 1999. "Norfloxacin-Induced DNA Cleavage Occurs at the *Dif* Resolvase Locus in *Escherichia Coli* and Is the Result of Interaction with Topoisomerase IV." *Molecular Microbiology* 33 (5): 1027–36.

Horiuchi, Takashi, Yohko Fujimura, Hideo Nishitani, Takehiko Kobayashi, and Masumi Hidaka. 1994. "The DNA Replication Fork Blocked at the *Ter* Site May Be an Entrance for the RecBCD Enzyme into Duplex DNA." *Journal of Bacteriology* 176 (15): 4656–4663.

Iliakis, G., H. Wang, A. R. Perrault, W. Boecker, B. Rosidi, F. Windhofer, W. Wu, J. Guan, G. Terzoudi, and G. Pantelias. 2004. "Mechanisms of DNA Double Strand Break Repair and Chromosome Aberration Formation." *Cytogenetic and Genome Research* 104 (1–4): 14–20.

John K. Eykelenboom, John K. Blackwood, Ewa Okely, and David R.F. Leach. 2008. "SbcCD Causes a Double-Strand Break at a DNA Palindrome in the *Escherichia Coli* Chromosome." *Molecular Cell* 29: 644–51.

John W. Nicol, Gregg A. Helt, Steven G. Blanchard, Jr., Archana Raja, and Ann E. Loraine. 2009. “The Integrated Genome Browser: Free Software for Distribution and Exploration of Genome-Scal Datasets.” *BMC Bioinformatics* 25 (20): 2730–31.

Johnson, Aaron, and Mike O’Donnell. 2005. “Cellular DNA Replicases: Components and Dynamics at the Replication Fork.” *Annual Review of Biochemistry* 74 (1): 283–315.

Jun-ichi Kato, Yukinobu Nishimura, Ryu Imamura, Hironori Niki, Sota Hiraga, and Hideho Suzuki. 1990. “New Topoisomerase Essential for Chromosome Segregation in *E . Coli.*” *Cell* 63: 393–404.

Kaguni, Jon M. 2011. “Replication Initiation at the *Escherichia Coli* Chromosomal Origin.” *Curr. Opin. Chem. Biol.* 15 (5): 606–13.

Kato, Jun-ichi, Yukinobu Nishimura, Masao Yamada, Hideho Suzuki, and Yukinori Hirota. 1988. “Gene Organization in the Region Containing a New Gene Involved in Chromosome Partition in *Escherichia Coli*.” *Journal of Bacteriology* 170 (9): 3967–77.

Khlebnikov, Artem, Kirill A. Datsenko, Tove Skaug, Barry L. Wanner, and Jay D. Keasling. 2001. “Homogeneous Expression of the PBAD Promoter in *Escherichia Coli* by Constitutive Expression of the Low-Affinity High-Capacity *araE* Transporter.” *Microbiology* 147 (12): 3241–47.

Kogoma, Tokio. 1997. “Stable DNA Replication: Interplay between DNA Replication, Homologous Recombination, and Transcription.” *Microbiology and Molecular Biology Reviews : MMBR* 61 (2): 212–38.

Kowalczykowski, Stephen C., Dan A. Dixon, Angela K. Eggleston, Scott D. Lauder, and William M. Rehrauer. 1994. “Biochemistry of Homologous Recombination in *Escherichia Coli*.” *Microbiological Reviews* 58 (3): 401–465.

Krasilnikova, Maria M., George M. Samadashwily, Andrey S. Krasilnikov, and Sergei M. Mirkin. 1998. “Transcription through a Simple DNA Repeat Blocks Replication Elongation.” *EMBO Journal* 17 (17): 5095–5102.

Lawrence, Heather Hendrickson and Jeffrey G. 2007. “Mutational Bias Suggests That Replication Termination Occurs near the *Dif* Site, Not at *Ter* Sites.” *Molecular Microbiology* 64 (1): 42–56.

Leach, David R. F. 1994. “Long DNA Palindromes, Cruciform Structures, Genetic Instability and Secondary Structure Repair.” *BioEssays* 16 (12): 893–810.

Lemon, Katherine P., Iren Kurtser, and Alan D. Grossman. 2001. “Effects of Replication Termination Mutants on Chromosome Partitioning in *Bacillus Subtilis*.” *Proceedings of the National Academy of Sciences of the United States of America* 98: 212–17.

Leonard, Alan C., and Marcel Méchali. 2013. “DNA Replication Origins.” *Cold Spring Harbor Perspectives in Biology* 5 (10): a010116–a010116.

Leu, Frank P., Roxana Georgescu, and Mike O’Donnell. 2003. “Mechanism of the *E. Coli Tau* Processivity Switch during Lagging-Strand Synthesis.” *Molecular Cell* 11 (2): 315–27.

Lieber, Michael R. 1998. “Warner-Lambert/Parke-Davis Award Lecture. Pathological and

Physiological Double-Strand Breaks: Roles in Cancer, Aging, and the Immune System." *The American Journal of Pathology* 153 (5): 1323–32.

Lindahl, Tomas. 1993. "Instability and Decay of the Primary Structure of DNA." *Nature* 362: 709–15.

Link, Andrew J., Dereth Phillips, and George M. Church. 1997. "Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type *Escherichia Coli:* Application to Open Reading Frame Characterization." *Journal of Bacteriology* 179 (20): 6228–37.

Lloyd, Robert G. 1991. "Conjugational Recombination in Resolvase-Deficient *ruvC* Mutants of *Escherichia Coli* K-12 Depends on *recG*." *Journal of Bacteriology* 173 (17): 5414–18.

Löwe, Jan, and Linda A. Amos. 1998. "Crystal Structure of the Bacterial Cell-Division Protein FtsZ ." *Nature* 391 (6663): 203–6.

Lundgren, Magnus, Anders Andersson, Lanming Chen, Peter Nilsson, and Rolf Bernander. 2004. "Three Replication Origins in *Sulfolobus* Species: Synchronous Initiation of Chromosome Replication and Asynchronous Termination." *Proceedings of the National Academy of Sciences of the United States of America* 101 (18): 7046–51.

Magee, Thomas R., Tsuneaki Asai, Deborah Malka, and Tokio Kogoma. 1992. "DNA Damage-Inducible Origins of DNA Replication in *Escherichia Coli*." *The EMBO Journal* 11 (11): 4219–25.

Maric, Marija, Timurs Maculins, Giacomo De Piccoli, and Karim Labib. 2014. "Cdc48 and a Ubiquitin Ligase Drive Disassembly of the CMG Helicase at the End of DNA Replication." *Science* 346 (6208): 1253596.

Mawer, Julia S. P., and David R. F. Leach. 2014. "Branch Migration Prevents DNA Loss during Double-Strand Break Repair." *PLoS Genetics* 10 (8): e1004485.

McGarry, Kevin C., Valorie T. Ryan, Julia E. Grimwade, and Alan C. Leonard. 2004. "Two Discriminatory Binding Sites in the *Escherichia Coli* Replication Origin Are Required for DNA Strand Opening by Initiator DnaA-ATP." *Proceedings of the National Academy of Sciences* 101 (9): 2811–16.

McInerney, Peter, Aaron Johnson, Francine Katz, and Mike O'Donnell. 2007. "Characterization of a Triple DNA Polymerase Replisome." *Molecular Cell* 27 (4): 527–38.

Merlin, Christophe, Sean Mcateer, and Millicent Masters. 2002. "Tools for Characterization of *Escherichia Coli* Genes of Unknown Function." *Journal of Bacteriology* 184 (16): 4573–81.

Messer, W. 1987. "Initiation of DNA Replication in *Escherichia Coli.*" *Proceedings of the National Academy of Sciences* 169: 3395–99.

Moreau, Morgane J. J., and Patrick M. Schaeffer. 2012. "Differential Tus-*Ter* Binding and Lock Formation: Implications for DNA Replication Termination in *Escherichia Coli*." *Molecular bioSystems* 8: 2783–2791.

Moreno, Sara Priego, Rachael Bailey, Nicholas Campion, Suzanne Herron, and Agnieszka Gambus. 2014. “Polyubiquitylation Drives Replisome Disassembly at the Termination of DNA Replication.” *Science* 346 (6208): 477–81.

Mulcair, Mark D., Patrick M. Schaeffer, Aaron J. Oakley, Hannah F. Cross, Cameron Neylon, Thomas M. Hill, and Nicholas E. Dixon. 2006. “A Molecular Mousetrap Determines Polarity of Termination of DNA Replication in *E. Coli.*” *Cell* 125 (7): 1309–19.

Ng, Jenny Y., and Kenneth J. Marians. 1996. “The Ordered Assembly of the Phi X174-Type Primosome.” *The Journal of Biological Chemistry* 271 (26): 15642–48.

Niki, Hironori, Aline Jaffe, Ryu Imamura, Teru Ogura, and Sota Hiraga. 1991. “The New Gene *mukB* Codes for a 177 Kd Protein with Coiled-Coil Domains Involved in Chromosome Partitioning of *E. Coli.*” *The EMBO Journal* 10 (1): 183–93.

Nkabuije Z. Maduike, Ashley K. Tehranchi, Jue D. Wang, and Kenneth N. Kreuzer. 2014. “Replication of the *Escherichia Coli* Chromosome in RNase HI- Deficient Cells: Multiple Initiation Regions and Fork Dynamics.” *Molecular Microbiology* 91 (1): 39–56.

Nurse, Pearl, Cindy Levine, Heide Hassing, and Kenneth J. Marians. 2003. “Topoisomerase III Can Serve as the Cellular Decatenase in *Escherichia Coli.*” *Journal of Biological Chemistry* 278 (10): 8653–60.

O’Donnell, Michael, Lance Langston, and Bruce Stillman. 2013. “Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya.” *Cold Spring Harbor Perspectives in Biology* 5 (7): 1–13.

Ogawa, T., H. Wabiko, T. Tsurimoto, and T. Horii. 1979. “Characteristics of Purified *recA* Protein and the Regulation of Its Synthesis *In Vivo*.” *Cold Spring Harb. Symp. Quant. Biol.* 43 (2): 909–15.

P. Hernandez, S. S. Lamm, C. A. Bjerknes and J. Van’t Hof. 1988. “Replication Termini in the rDNA of Synchronized Pea Root Cells (*Pisum Sativum*).” *The EMBO Journal* 7 (2): 303–8.

Payne, Bryony T. I., Ingeborg C. van Knippenberg, Hazel Bell, Sergio R. Filipe, David J. Sherratt, and Peter McGlynn. 2006. “Replication Fork Blockage by Transcription Factor-DNA Complexes in *Escherichia Coli.*” *Nucleic Acids Research* 34 (18): 5194–5202.

Pelletier, Anthony J., Thomas M. Hill, and Peter L. Kuempel. 1988. “Location of Sites That Inhibit Progression of Replication Forks in the Terminus Region of *Escherichia Coli.*” *Journal of Bacteriology* 170 (9): 4293–98.

Peng, Hong, and Kenneth J. Marians. 1995. “The Interaction of *Escherichia Coli* Topoisomerase IV with DNA.” *Journal of Biological Chemistry* 270 (42): 25286–90.

Pérals, Koryn, François Cornet, Yann Merlet, Isabelle Delon, and Jean-Michel Louarn. 2000. “Functional Polarization of the *Escherichia Coli* Chromosome Terminus: The *Dif* Site Acts in Chromosome Dimer Resolution Only When Located between Long Stretches of Opposite Polarity.” *Molecular Microbiology* 36 (1): 33–43.

Perez-Cheeks, Brenda A., Chong Lee, Ryo Hayama, and Kenneth J. Marians. 2012. “A Role

for Topoisomerase III in *Escherichia Coli* Chromosome Segregation." *Molecular Microbiology* 86 (4): 1007–22.

Pfaffl, Michael W. 2001. "A New Mathematical Model for Relative Quantification in Real-Time RT-PCR." *Nucleic Acids Research* 29 (9): e45.

Reyes-Lamothe, Rodrigo, Emilien Nicolas, and David J. Sherratt. 2012. "Chromosome Replication and Segregation in Bacteria." *Annual Review of Genetics*. 46: 121–43.

Reyes-Lamothe, Rodrigo, Christophe Possoz, Olessia Danilova, and David J. Sherratt. 2008. "Independent Positioning and Action of *Escherichia Coli* Replisomes in Live Cells." *Cell* 133 (1). Elsevier Inc.: 90–102.

Richard W. Deibler, Sonia Rahmati, E. Lynn Zechiedrich. 2001. "Topoisomerase IV, Alone, Unknots DNA in *E. Coli*." *Genes & Development* 15: 748–61.

Rothstein, Rodney, Benedicte Michel, and Serge Gangloff. 2000. "Replication Fork Pausing and Recombination or "gimme a break"" *Genes Dev.* 14 (1): 1–10.

Rudolph, Christian J., Amy L. Upton, Anna Stockum, Conrad A. Nieduszynski, and Robert G. Lloyd. 2013. "Avoiding Chromosome Pathology When Replication Forks Collide." *Nature* 500 (7464): 608–11.

Sandler, Steven J. 2000. "Multiple Genetic Pathways for Restarting DNA Replication Forks in *Escherichia Coli* K-12." *Genetics* 155 (2): 487–97.

Sandler, Steven J., and Kenneth J. Marians. 2000. "Role of PriA in Replication Fork Reactivation in *Escherichia Coli*." *Journal of Bacteriology* 182 (1): 9–13.

Seiji Tanaka, and Hiroyuki Araki. 2013. "Helicase Activation and Establishment of Replication Forks at Chromosomal Origins of Replication." *Cold Spring Harbor Perspectives in Biology* 5:a010371.

Seol, Yeonee, Ashley H. Hardin, Marie Paule Strub, Gilles Charvin, and Keir C. Neuman. 2013. "Comparison of DNA Decatenation by *Escherichia Coli* Topoisomerase IV and Topoisomerase III: Implications for Non-Equilibrium Topology Simplification." *Nucleic Acids Research* 41 (8): 4640–49.

Sharples, Gary J., Robert G. Lloyd, Steven C. West. 1991. "Resolution of Holliday Junctions in *Escherichia Coli:* Identification of the *ruvC* Gene Product as a 19-Kilodalton Protein." *Journal of Bacteriology* 173 (23): 7711–15.

Shimokawa, Koya, Kai Ishihara, Ian Grainge, David J. Sherratt, and Mariel Vazquez. 2013. "FtsK-Dependent XerCD-*Dif* Recombination Unlinks Replication Catenanes in a Stepwise Manner." *Proceedings of the National Academy of Sciences of the United States of America* 110 (52): 20906–11.

Singer, Mitchell, Tania A. Baker, Gavin Schnitzler, Shawn M. Deischel, Manju Goel, William Dove, Kathryn J. Jaacks, Alan D. Grossman, James W. Erickson, and Carol A. Gross. 1989. "A Collection of Strains Containing Genetically Linked Alternating Antibiotic Resistance Elements for Genetic Mapping of *Escherichia Coli*." *Microbiological Reviews* 53 (1): 1–24.

Soupene, Eric, Wally C. Van Heeswijk, Valley Stewart, Daniel Bertenthal, Haidy Lee, Gyaneshwar Prasad, Oleg Paliy, Parinya Charernnoppakul, Sydney Kustu, and Jacqueline Plumbridge. 2003. “Physiological Studies of *Escherichia Coli* Strain MG1655 : Growth Defects and Apparent Cross-Regulation of Gene Expression.” *Journal of Bacteriology* 185 (18): 5611–26.

Steiner, Walter W., and Peter L. Kuempel. 1998. “Cell Division Is Required for Resolution of Dimer Chromosomes at the *Dif* Locus of *Escherichia Coli*.” *Molecular Microbiology* 27 (2): 257–68.

Suski, Catherine, and Kenneth J. Marians. 2008. “Resolution of Converging Replication Forks by RecQ and Topoisomerase III.” *Molecular Cell* 30 (6): 779–89.

Taylor, Andrew F., Dennis W. Schultz, Alfred S. Ponticelli, and Gerald R. Smith. 1985. “RecBC Enzyme Nicking at Chi Sites during DNA Unwinding: Location and Orientation-Dependence of the Cutting.” *Cell* 41 (1): 153–63.

Wallis, John W., Gary Chrebet, Gary Brodsky, Mark Rolfe, and Rodney Rothstein. 1989. “A Hyper-Recombination Mutation in *S. Cerevisiae* Identifies a Novel Eukaryotic Topoisomerase.” *Cell* 58 (2): 409–19.

Wang, Xindan, Rodrigo Reyes-Lamothe, and David J. Sherratt. 2008. “Modulation of *Escherichia Coli* Sister Chromosome Cohesion by Topoisomerase IV.” *Genes & Development* 22: 2426–33.

Watson, James D., and Francis H. C. Crick. 1953. “Molecular Structure of Nucleic Acids.” *Nature* 4356(171):737-738.

Webb, Shaun, Ralph D Hector, Grzegorz Kudla, and Sander Granneman. 2014. “PAR-CLIP Data Indicate That Nrd1-Nab3-Dependent Transcription Termination Regulates Expression of Hundreds of Protein Coding Genes in Yeast.” *Genome Biology* 15 (1): R8.

Wendel, Brian M., Charmain T. Courcelle, and Justin Courcelle. 2014. “Completion of DNA Replication in *Escherichia Coli.*” *Proceedings of the National Academy of Sciences of the United States of America* 111 (46): 16454–59.

Yardimci, Hasan, and Johannes C. Walter. 2014. “Prereplication-Complex Formation: A Molecular Double Take?” *Nature Structural & Molecular Biology* 21 (1). Nature Publishing Group: 20–25.

Zechiedrich, E. Lynn, and Nicholas R. Cozzarelli. 1995. “Roles of Topoisomerase IV and DNA Gyrase in DNA Unlinking during Replication in *Escherichia Coli*.” *Genes & Development* 9: 2859–69.