DNA replication and replication termination in *Escherichia coli*

A thesis submitted
for the degree of Doctor of Philosophy
by ***Sarah Midgley-Smith***

Department of Biosciences,
Brunel University London

October 2018

Abstract

A prerequisite for successful cell division is the generation of an accurate copy of the entire genome as well as faithful segregation into the daughter cells. In the bacterium *Escherichia coli,* replication of the circular chromosome is initiated at a single origin (*oriC*) where two replication forks are assembled and proceed bi-directionally until they converge within a defined termination region opposite *oriC* and fork fusion takes place. This region is flanked by *ter* sequences, which, when bound by Tus protein, form a replication fork trap that allows forks to enter but not to leave. While the events associated with initiation as well as the elongation of replication have been extensively studied, the molecular details associated with the fusion of two replisomes are far less well characterised.

The data presented here significantly extend our understanding of the molecular mechanics associated with the fusion of two replisomes in *E. coli.* My results strongly support the idea that RecG is a key player in processing intermediates that arise as two replication forks fuse. In the absence of RecG, over-replication of the chromosome is initiated at fork fusion intermediates, a process that can take place outside of the native termination area if forks are forced to fuse in an ectopic location. RecG has also been implicated in processing recombination intermediates. Over-replication in the absence of RecG is dependent on recombination and my data support the idea that RecG is important in limiting replication that initiates at recombination intermediates, some of which arise as a consequence of fork fusion events. In contrast, my data do not support the notion that the over-replication of chromosomal DNA in cells lacking RecG is in any way triggered by R-loops, as suggested previously in the literature.

The observation that over-replication is taking place in cells lacking exonuclease I strongly supports the idea that 3' single-stranded DNA structures are a key intermediate of fork fusion events. My data demonstrate that 3' flaps accumulating in the absence of ExoI can be converted into 5' flaps and degraded by 5' exonucleases such as ExoVII and RecJ. RecG helicase is likely to be involved in this conversion. Thus, the data presented in this thesis highlight the complexity of replication fork fusion events and demonstrate that multiple protein activities are required to process fork fusion intermediates in order to allow DNA replication to be completed with high accuracy and enable faithful segregation of complete chromosomes into daughter cells.

Declaration

The research presented in this thesis is my own work, unless otherwise specified, and has not been submitted for any other degree.

Sarah Midgley-Smith

Acknowledgements

First and foremost, I would like to thank Dr Christian Rudolph. It has been a privilege and a pleasure to work in your lab and I am immensely grateful to you for supporting me in achieving this feat. I will miss your enthusiasm, stimulating discussions about our work, and especially your disregard for the instruction manual. Thank you for teaching me so much about what it means to be a scientist.

A great team of people at Brunel have made my time here that much more enjoyable. Thanks to the technicians, especially Helen, for always helping out when needed. Special thanks must go to the other members of the Rudolph lab, in particular Juachi; I will miss working with you and I am so grateful for your kindness and willingness to share your knowledge in the lab. Thanks to some first-rate undergrad students, Toni, Monja and Ewa, for producing some great work and ensuring that I was never lonely in the lab! Thanks to my fellow PhD students past and present for being an excellent source of procrastination; there have been some interesting conversations. Ezgi, I could not have wished for a better desk neighbour. Being in the midst of a PhD often felt quite lonely and it was nice to know that you were right there with me, sharing successes and (more often) frustration! I am so grateful for all your encouragement.

I could not possibly fail to mention the best friends I can imagine, Louisa, Liz, Tone and Kate. You are outstanding. I wish I could see you all every day and I am so lucky to have you as my friends. Thanks also to Caroline, Naomi, Pippa, J-Lo and Steph for fab times away from work. Special mention to Tommy, for everything.

To my parents, Dave and Kate; your support has been incredible. Thank you for instilling in me as a child the importance of education and always encouraging me to do well, and for being endlessly supportive on the many occasions that I have not. I know that you are as pleased about this success as I am, and that is very special. I am thrilled to be able to share this with you. Thanks for the well-timed Vespers and food that kept me writing through the night!

My sisters Eleanor and Harriet are second to none. Eleanor, thank you for always being there on the end of the phone with all the right things to say to cheer me up; I think the highlight was firework cells! Harri, you have provided many welcome distractions from my work, like cooking chilli for 30 people, or carrying a washing machine up and down stairs! Thank you both for being so fab all the times you knew I was not quite on form; you girls are the best.

Finally, to my husband and my best friend, Greg. I cannot write anything that will do justice to what you have done in supporting me through the past 10 years. This success belongs to you as much as to me, as the Big Science Story would not have happened without you! You cheer me up and make me laugh when I need it most and are always on my side. I am unbelievably lucky to have your constant love, help and advice and I know I have not always made it easy. Thank you for being right there next to me through all the ups and downs; pub?

Publications

As a result of my work, several publications have already been published with data that will be presented in this thesis, and those publications are:

Midgley-Smith, S.L.; Dimude, J.U. and Rudolph, C.J. (2019) A role for 3' exonucleases at the final stages of chromosome duplication in *Escherichia coli* **Nucleic Acids Research** 47(4):1847-1860

Dimude, J.U.; Midgley-Smith, S.L. and Rudolph, C.J. (2018) Replication-transcription conflicts trigger extensive DNA degradation in *Escherichia coli* cells lacking RecBCD. **DNA Repair (Amst)** 70:37–48

Midgley-Smith, S.L.; Dimude, J.U.; Taylor, T.; Forrester, N.M.; Upton, A.L.; Lloyd, R.G. and Rudolph, C.J. (2018) Chromosomal over-replication in *Escherichia coli recG* cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed. **Nucleic Acids Research** 46(15):7701-7715

Dimude, J.U.; Midgley-Smith, S.L.; Stein, M. and Rudolph, C.J. (2016). Replication Termination: Containing Fork Fusion-Mediated Pathologies in *Escherichia coli*. **Genes** 7(8), 40

Dimude, J.U.; Stockum, A; Midgley-Smith, S.L.; Upton, A.L.; Foster, H.A.; Khan, A; Saunders, N.J.; Retkute, R. and Rudolph, C.J. (2015). The Consequences of Replicating in the Wrong Orientation: Bacterial Chromosome Duplication without an Active Replication Origin. **mBio**. 6(6) e01294-15

Darja Ivanova, Toni Taylor, Sarah L. Smith, Juachi U. Dimude, Amy L. Upton, Mana M. Mehrjouy, Ole Skovgaard, David J. Sherratt, Renata Retkute, Christian J. Rudolph. (2015) Shaping the landscape of the *Escherichia coli* chromosome: replication-transcription encounters in cells with an ectopic replication origin. **Nucleic Acids Research.** 43(16): 7865–7877

List of abbreviations

A_{600}	absorption at 600 nm
ADP	adenosine diphosphate
amp	gene for ampicillin resistance
apra	gene for apramycin resistance
ATP	adenosine triphosphate
bp	base pair
BrdU	bromodeoxyuridine
cat	chloramphenicol acetyl transferase, gene for chloramphenicol resistance
cfu	colony forming units
cSDR	constitutive stable DNA replication
Dam	DNA adenine methyltransferase
dhfr	dihydrofolate reductase, gene for trimethoprim resistance
dif	deletion-induced filamentation, dimer resolution site
D-loop	displacement-loop where the invading species is DNA
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DSBR	double-strand break repair
dsDNA	double-stranded DNA
DUE	DNA-unwinding element
EDTA	Ethylenediaminetetraacetic acid
EdU	5-ethynyl-2'-deoxyuridine
FP	fluorescent fusion protein
frt	flippase recognition target
Hfr	high frequency of recombination
IPTG	Isopropyl-β-D-thiogalactoside
iSDR	inducible stable DNA replication
kan	gene for kanamycin resistance
Kb	kilobases
LOESS	locally estimated scatterplot smoothing
Mbp	mega base pairs
MFA	marker frequency analysis
MMR	methyl-directed mismatch repair
PCR	polymerase chain reaction
PMSF	phenylmethane sulphonyl fluoride

Pol III	DNA polymerase III
RIDA	regulatory inactivation of DnaA
R-loop	displacement-loop where the invading species is RNA
RNA	ribonucleic acid
RNAP	RNA polymerase
rrn operon	ribosomal RNA operon
SDR	stable DNA replication
SDS	sodium dodecyl sulphate
SSB	single-stranded DNA binding protein
ssDNA	single-stranded DNA
ter	termination sequence
tos	telomerase occupancy site
tRNA	transfer RNA
ts	temperature sensitive
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Table of contents

Introduction

Understanding how a single cell can divide to become two cells is one of the most fundamental questions in biology. The genome must be duplicated for a cell to divide and as the chromosomes contain all the genetic information required to direct the growth and development of a cell, the resulting two copies must be identical. Maintenance of genomic stability is an essential task for all organisms. A number of different, interrelated processes combine to ensure accurate replication and division of the chromosomes, and precise distribution of the genomic material among the daughter cells. Mutations in the genome affect genomic stability, which can have serious consequences. In multicellular organisms, mechanisms exist in normal healthy tissues to prevent the survival or proliferation of cells that have accumulated genomic alterations (Zhivotovsky and Kroemer, 2004) highlighting the importance of managing genomic stability. In humans, genomic instability is a prominent feature of cancers (Negrini et al., 2010; Shen, 2011) and the cause of a number of diseases and genetic disorders (Aguilera and Gómez-González, 2008; Castel et al., 2010; Hou et al., 2017; Sasaki et al., 2010). Recent work by Tomasetti and colleagues (2017) suggests that replication errors are responsible for two thirds of the mutations that lead to cancers in humans (Tomasetti et al., 2017), which highlights how important it is to study the mechanistic details of DNA replication (Hsieh and Yamane, 2008). In addition, links between defects in a repair pathway (mismatch repair) that functions as part of DNA replication and the Lynch cancer syndrome or hereditary non-polyposis colorectal cancer have been reported (Wimmer and Etzler, 2008)

In single-celled organisms such as bacteria, genomic instability can cause pathological problems including delayed or unsuccessful chromosome replication and cell division, which can threaten the viability of the cells. It was recently reported that the fusion of two replication forks has the potential to cause pathology (Rudolph et al., 2013). This is somewhat surprising considering replication fork fusion is an integral part of DNA replication. My research has focussed on furthering our understanding of the events at replication termination using *Escherichia coli* as a model organism.

DNA replication in Escherichia coli

The *E. coli* genome consists of a single, circular chromosome of 4.6 Mb, containing over 4000 protein-coding genes that account for >80 % of the genome (Blattner et al., 1997). Bidirectional replication is initiated at a single, sequence-specific origin of replication, *oriC*, located at 84.6 min on the *E. coli* genetic map (Skarstad and Katayama, 2013). The replication machinery is assembled into two

replication forks that move in opposite directions around the chromosome until they meet in a defined region located approximately opposite *oriC* where replication terminates, inter-linked daughter chromosomes are resolved and segregation of the chromosomes to each cell half takes place before the cell divides.

Cell cycle

In all living cells, a number of different processes must be coordinated and managed for progression through the cell cycle, such as chromosome duplication, cell growth and cell division. The *E. coli* cell cycle in slow-growing cells appears relatively simple compared with that of eukaryotic cells, where there are numerous checkpoints that ensure progression through the cycle is not initiated until the previous step is completed (Cooper, 2000). It was thought that bacterial cells such as *E. coli* followed a relatively rigid cell cycle where achievement of a critical cell mass was the trigger for DNA replication initiation (Donachie, 1993). This concept has been increasingly challenged (Bates and Kleckner, 2005; Boye and Nordström, 2003; Nordstrom et al., 1991; Wold et al., 1994). More recent data suggest that the bacterial cell cycle is actually not such a simple step-by-step affair and consists instead of a series of cell cycle events that can be overlapping and are coordinated by multiple interactions between these events (Haeusser and Levin, 2008). These are divided in to three periods; B-, C-, and D-period (Figure 1).

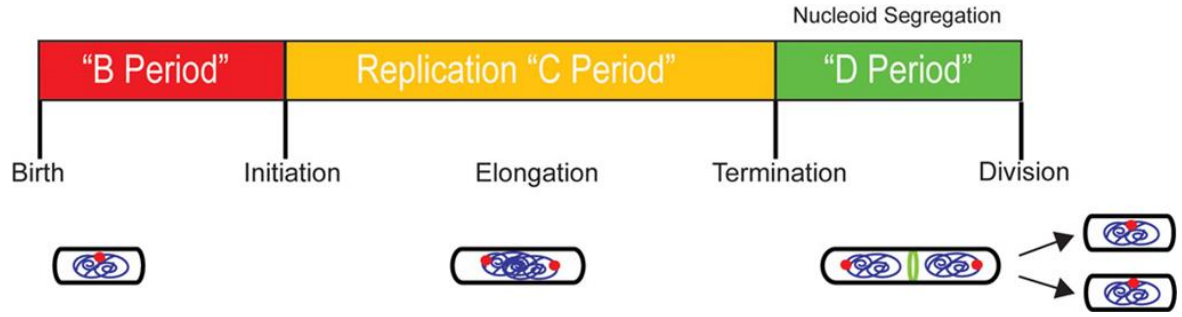

Figure 1: The bacterial cell cycle. The schematic diagram shows the three separate periods that are identifiable in slow-growing bacterial cells. The red dot indicates the origin of replication in the cell diagrams and the cytokinetic ring, on which the division machinery assembles, is shown in green. C-period covers DNA replication to termination and D-period covers cell division. Cell growth happens in B-period but continues throughout the cell cycle (figure reproduced from Haeusser and Levin, 2008 with permission).

E. coli cells can grow at very different growth rates. When environmental conditions support fast-growth, the time needed for DNA replication (C-period) can exceed the total doubling time (i.e. B + C + D periods). This situation is possible as *E. coli* can grow with overlapping DNA replication cycles, where two or four origins are initiated in one generation whilst a round of replication is already underway through initiation from the previous generation. Replication is initiated at all origins simultaneously once per generation (Skarstad et al., 1986). Cell growth is continuous throughout the whole cell cycle (Bates et al., 2005).

Despite a large body of research over many years, the intricacies of the cell cycle and regulatory mechanisms of the processes involved are not fully understood, although progress is being made (Adiciptaningrum et al., 2015; Bates and Kleckner, 2005; Bates et al., 2005; Grant et al., 2011). Recent developments in single-cell research are starting to allow investigation in to the behaviour of single cells as opposed to population-based data, which are already challenging our understanding of the cell cycle (Osella et al., 2017)

DNA replication initiation

DNA replication in *E. coli* is initiated at a single, defined DNA sequence called *oriC*. The 245 bp *oriC* region contains multiple asymmetric 9-bp sites called DnaA boxes that have high or low affinity for binding of the DnaA initiator protein (*dnaA*). DnaA is bound to the high affinity sites for the majority of the cell cycle and only binds the lower-affinity sites at initiation. An AT-rich region to one side of the origin, called the DNA-unwinding element (DUE) and consisting of three 13-bp repeats, contains a third type of binding site that is similar to the weak-affinity boxes from *oriC*. DnaA is complexed with ATP (adenosine triphosphate) or ADP (adenosine diphosphate) at the origin and both complexes are able to bind the high affinity boxes but only ATP-DnaA can bind the low-affinity sites. ATP-DnaA is the active form of DnaA and is key for initiating replication at *oriC* (reviewed in Hansen and Atlung, 2018; Kaguni, 2006; Messer, 2002; Mott and Berger, 2007).

As the DnaA:origin ratio increases prior to an initiation event, ATP-DnaA binds the lower affinity boxes. The binding of DnaA to the origin distorts the DNA and facilitates unwinding of the DNA within the DUE, resulting in single-stranded DNA (ssDNA) in this region (Figure 2) (Leonard and Grimwade, 2005).

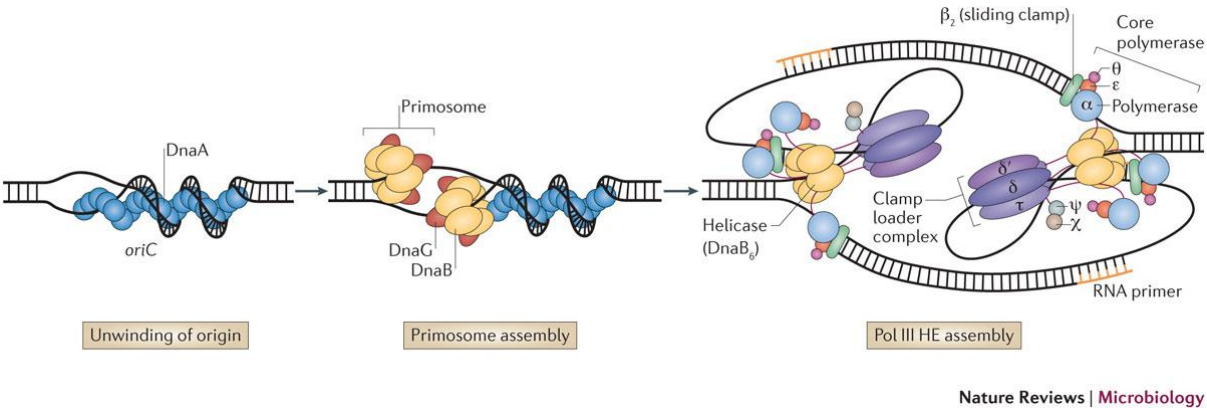

Figure 2: DNA replication initiation in *Escherichia coli* (figure reproduced from Robinson and van Oijen, 2013 with permission).

Once the DNA in the AT-rich DUE region has been unwound, the stage is set for the assembly of the replisome components. The replisome is a large, multiprotein complex (see Figure 21) that contains all the components required for chromosome duplication (Beattie and Reyes-Lamothe, 2015; Johnson and O'Donnell, 2005; Kurth and O'Donnell, 2009; O'Donnell, 2006). The first replisome component to

be recruited is the replicative DnaB helicase. The single-stranded DUE regions will quickly be bound by single-stranded binding protein (SSB; page 19), which hinder DnaB binding. DNA-bound DnaA, aided by the DnaB loading factor DnaC, facilitates the loading of two circular homo-hexamer DnaB molecules on to opposite strands at the unwound region, one for each replication fork. DnaC then leaves, activating the DnaB molecules (reviewed in Costa et al., 2013; Leonard and Grimwade, 2005). DnaB helicase encircles the lagging template strand and uses energy released from ATP hydrolysis to translocate along the DNA in a 5' to 3' direction, which physically forces the two strands apart (Kurth and O'Donnell, 2009). Upon activation, the two DnaB helicases proceed in opposite directions and open up the loop further (Fang et al., 1999). This allows DnaG primase access to synthesise two leading strand primers. The β sliding clamp and the DNA polymerase III holoenzymes (page 102) are then loaded at each primed template, forming the replication forks. At this point, initiation proceeds immediately to elongation as the replication forks are established and set off around the chromosome (Fang et al., 1999; Messer, 2002; Mott and Berger, 2007; Nielsen and Løbner-Olesen, 2008).

Regulation of replication initiation

Accurate duplication and segregation of genomic material is important in all organisms and to achieve this, it is essential that the entire genome is replicated exactly once per cell cycle. In order to limit DNA replication initiation to once per generation and maintain replication at an appropriate time in the cell cycle, a number of regulatory strategies are employed in *E. coli* cells following a successful initiation event that prevent untimely replication reinitiations. Both *oriC* and DnaA are targeted in the control of replication initiation. These mechanisms ensure that replication is initiated at all origins present in a cell only once per cell cycle and that the initiation events occur simultaneously (Hansen and Atlung, 2018; Katayama et al., 2017; Messer, 2002; Mott and Berger, 2007; Skarstad and Katayama, 2013). It has been suggested that initiation at multiple origins present in the cell due over-lapping cell cycles occurs as an initiation cascade, where DnaA released from the first origin to fire transiently increases the ratio of free DnaA to *oriC* sequences for the remaining old origins and promotes another initiation event by immediately binding to other origins present (Løbner-Olesen et al., 1994).

Methylation and sequestration of the origin

During the initiation cascade, a mechanism called sequestration is important in preventing immediate re-initiation of an origin that has just fired, which could occur while the initiation potential remains high until all origins have fired, and therefore directs successive initiations to the remaining 'old' origins.

In addition to the four normal bases, *E. coli* DNA contains two bases modified by methylation; 6-methyl-adenine and 5-methylcytosine. The Dam (DNA adenine methyltransferase) enzyme is responsible for adenine methylation. The target recognition sequence for Dam methyltransferase is 5'-GATC-3', which is palindromic in duplex DNA. The chromosome is normally fully methylated, which

means that this sequence is methylated throughout the chromosome on both strands of the DNA, but during DNA replication, there is a transient state of hemimethylation immediately behind the replication forks as the parent strand is methylated and the newly synthesised strand is not. This allows for discrimination between the two strands, which is of importance in the mismatch repair process (page 20) and in maintaining synchronous replication initiation, although it is not essential for viability of *E. coli* cells (Marinus and Løbner-Olesen, 2014). GATC sites on the nascent strand are very quickly targeted by Dam protein for remethyation, ≤1 min in most areas of the chromosome in exponentially growing cells, resulting once more in fully methylated DNA (Adhikari and Curtis, 2016; Marinus and Løbner-Olesen, 2014).

Whilst studies involving analysis of DNA sequences indicate that the GATC sequence appears in the chromosome at an average of once per 243 nucleotides, close to the once per 256 expected for a 4 bp sequence, the sequences are actually distributed unevenly around the chromosome and so there are a number of GATC-rich regions and conversely areas with fewer GATC sites than average (Barras and Marinus, 1988; Hénaut et al., 1996; Marinus and Løbner-Olesen, 2014). *oriC* and the promoter for *dnaA* contain a high number of GATC sites. The full methylation of the *oriC* region facilitates duplex opening at DNA replication initiation (Yamaki et al., 1988). Following replication of the origin, the resulting hemimethylated GATC sequences within *oriC* and the *dnaA* promoter remain hemimethylated for longer than the sequences in other parts of the chromosome due to the action of Sequestration A (SeqA) protein, which competes with Dam methyltransferase for access to GATC sites within these regions (Figure 3) (Lu et al., 1994; Marinus and Løbner-Olesen, 2014; Messer, 2002; Skarstad and Katayama, 2013).

SeqA protein discriminates between an origin that is about to initiate replication (old origin) and an origin that has just been replicated (new origin) by the fully methylated or hemimethylated state respectively of the GATC sequences present in each origin region. SeqA preferentially binds to hemimethylated GATC sequences and following DNA replication initiation, SeqA rapidly binds to the hemimethylated origins. This is referred to as origin sequestration and the action of SeqA prevents immediate re-initiation of an origin that has just fired by preventing re-methylation by Dam and denying the DnaA initiator protein access to a number of low-affinity DnaA binding sites within *oriC* that contain the GATC sequence recognised by SeqA (Figure 3) (Kaguni, 2006; Nievera et al., 2006; Skarstad and Katayama, 2013).

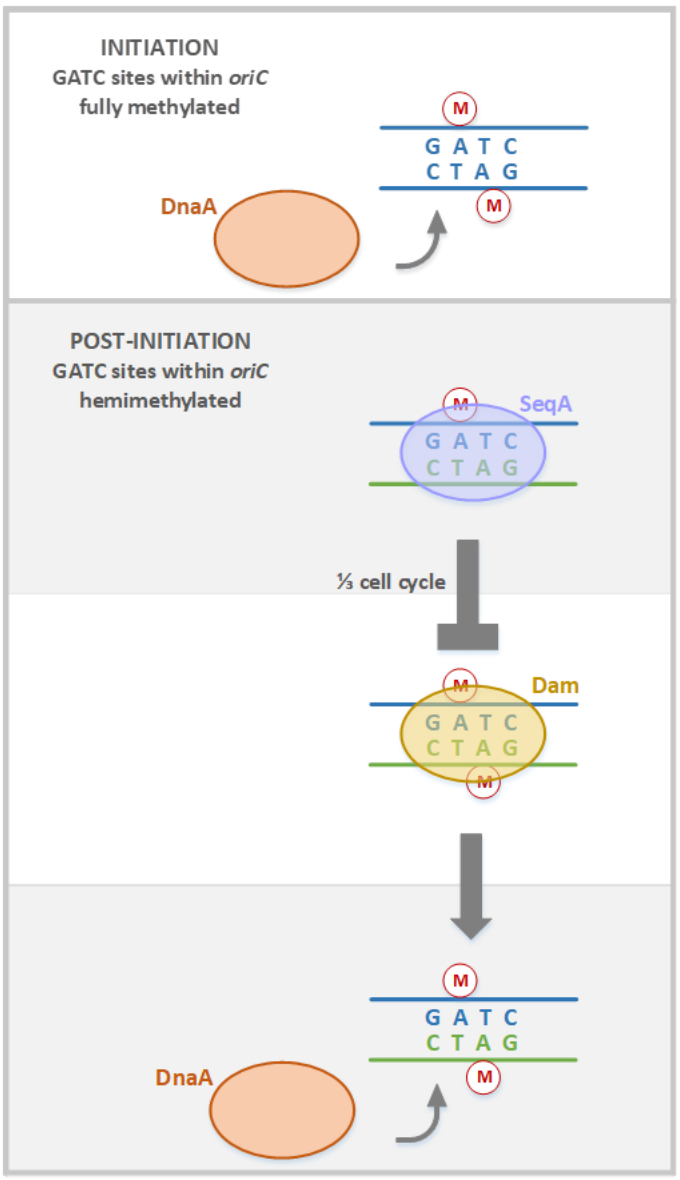

Figure 3: The GATC sites within the origin region and the *dnaA* promoter undergo changes to their methylation status that enable modulation of replication initiation and sequestration by SeqA.

These sites within a new origin remain inaccessible to DnaA for about one third of the cell cycle due to sequestration (Figure 3) (Slater et al., 1995), whereas DnaA binding at the three high affinity boxes R1, R2 and R4 is detected during the sequestration period, allowing resetting of the origin complexes (Nievera et al., 2006). Both *Δdam* and *ΔseqA* mutants are viable but show asynchronous replication initiation, consistent with the idea that methylation status and sequestration of origin regions are both important in promoting synchronous initiation of all replication origins within a cell and contribute to limiting initiation events at each origin to once per cell cycle (Løbner-Olesen et al., 1994; Lu et al., 1994; Marinus and Løbner-Olesen, 2014).

Reducing DnaA availability

By preventing re-initiation of newly replicated origins, origin sequestration is important in allowing time for the initiation potential to be reduced. A number of strategies are proposed to be involved in achieving this and they are centred around regulating the availability and activity of the DnaA initiator protein such that available active DnaA is reduced to levels that cannot sustain initiation until the cell is ready to begin the next round (Skarstad and Katayama, 2013).

Regulatory inactivation of DnaA – DnaA has affinity for both ATP and ADP, but ADP-DnaA is considered the inactive form of DnaA as only ATP-DnaA is able to effect unwinding of the dsDNA at the origin to initiation replication (Leonard and Grimwade, 2011). Following an initiation event, DnaA protein is inactivated by hydrolysis of the DnaA-bound ATP to produce the inactive ADP-DnaA complex instead. This process is termed regulatory inactivation of DnaA (RIDA) (Katayama et al., 1998) and involves an interaction between the β sliding clamp binding to DNA during replication initiation and the active ATP-DnaA, which promotes the conversion to ADP-DnaA. This interaction is mediated by Hda protein (Katayama, 2001; Katayama et al., 2001; Kato and Katayama, 2001).

Titration of DnaA to reservoir sites – There are many DnaA-binding sites present in locations around the chromosome outside of the origin region. As replication progresses, the increasing number of DnaA-binding sites formed by duplication of the chromosome contribute to the titration of DnaA away from the origin, reducing the availability of free DnaA molecules in the cell. One of these sites that has a particularly high affinity for DnaA is the *datA* region, which is located near to *oriC* and so is replicated soon after initiation of replication. *datA* is estimated to bind approximately eightfold more DnaA molecules than the *oriC* region (Hansen et al., 2007; Kitagawa et al., 1996; Skarstad and Katayama, 2013). It has been proposed that *datA* therefore is key in providing a reservoir that titrates DnaA molecules away from the origin and in line with this, it was reported that in the absence of the *datA* locus, growth is normal but cells display an asynchronous initiation phenotype, indicating that *datA* is essential for maintaining timely replication initiation (Kitagawa et al., 1998; Ogawa et al., 2002).

Regulation of *dnaA* expression – The *dnaA* gene is located near *oriC* and as a consequence, the *dnaA* promoter is subject to sequestration by SeqA for a similar length of time as the origin (Campbell and Kleckner, 1990), and it has been shown that this contributes to maintaining initiation control and preventing asynchronous initiation events (Riber and Løbner-Olesen, 2005).

In addition to sequestration, *dnaA* expression is regulated by the DnaA protein itself (Atlung et al., 1985). *dnaA* is transcribed from two promoters (Messer, 2002). The promoter region contains DnaA boxes and the binding of DnaA protein to its own promoter region results in the negative regulation of both promoters and repression of the transcription of *dnaA* (Hansen and Atlung, 2018). Whilst both ATP-DnaA and ADP-DnaA can bind to this region, ATP-DnaA is much more effective in repressing *dnaA* expression (Speck et al., 1999).

Elongation

Once the replisomes are established at each end of the initiation bubble, replication begins as the two replication forks move away from the origin in opposite directions, proceeding at high speed of 550 – 750 bp s^{-1} (Pham et al., 2013) with the left-hand, anticlockwise fork often slightly ahead of the right-hand fork (Breier et al., 2005). The two strands of the DNA template are separated by the replicative helicase DnaB (LeBowitz and McMacken, 1986). The presence of SSB at the replication fork protects the resulting single-stranded template DNA until it is duplicated. The replicative polymerase, DNA Pol III, is tethered to the DNA template by the β sliding clamp, which increases the speed and processivity of the polymerase (Burgers et al., 1981; Johnson and O'Donnell, 2005; LaDuca et al., 1986). The sliding clamp is actively assembled around the DNA by the clamp loader at primed sites generated by DnaG, a primase that is transiently recruited to the replication fork by DnaB and acts in a distributive fashion to synthesise short RNA primer sequences to provide a substrate for DNA polymerase to initiate DNA synthesis (Rowen and Kornberg, 1978; Wu et al., 1992). The replisome is a multi-subunit complex. The components of the replisome are described in more detail later (page 102).

Coupling the duplication of each strand of the template causes a challenge for cells that results from the antiparallel nature of a DNA molecule. DNA Pol III moves in a 3' to 5' direction along the template strand and begins synthesising the nascent strand at the 3'-hydroxyl end of a primer. The leading strand synthesis occurs in the same direction as the replication fork is moving whereas the lagging strand is orientated in the opposite direction. As a result, the lagging strand is synthesised in short, discontinuous sections called Okazaki fragments, each of which is primed by DnaG as replication progresses (Balakrishnan and Bambara, 2013; Okazaki et al., 1968). The generally accepted view is that leading strand synthesis is continuous following a single priming event at the origin, although there is evidence for both continuous and discontinuous replication of the leading strand (Amado and Kuzminov, 2006; Graham et al., 2017; Langston et al., 2009; Wang, 2005); indeed, discontinuous synthesis of both strands was suggested very early on in the effort to elucidate the details of DNA replication (reviewed in Ogawa and Okazaki, 1980). The RNA primers needed for the lagging strand synthesis are subsequently replaced with DNA nucleotides by DNA polymerase I and the gaps between two Okazaki fragments are joined by ligase (Pomerantz and O'Donnell, 2007).

Single-stranded DNA binding protein

Whilst the normal conformation of the DNA that comprises the genome of *E. coli* is as a double-stranded, helical molecule folded into a compact structure and organised into superhelical domains, the processes surrounding nucleic acid metabolism that are essential for life require access to the information contained within the dsDNA, which means that regions of unwound, ssDNA must be formed within the chromosome when necessary (Rocha, 2008; Toro and Shapiro, 2010). ssDNA is vulnerable to attack that might damage the DNA molecule. To protect the integrity of the genetic information contained within the DNA molecule and to prevent untimely reannealing of the single

strands, single-stranded DNA binding (SSB) proteins bind to ssDNA in a sequence-independent manner, and so SSB is targeted to sites of active replication, recombination and repair (Meyer and Laine, 1990; Shereda et al., 2008). In addition to providing protection to ssDNA, SSB has more recently been shown to interact with a number of proteins in order to target DNA metabolism proteins to appropriate regions of the DNA molecule, including Exonuclease I (ExoI) (Lu and Keck, 2008; Lu et al., 2011), RecJ exonuclease (Han et al., 2006), RecG helicase, PriA (Yu et al., 2016) and other DNA metabolism proteins involved in replication, recombination and repair (Shereda et al., 2008), and additionally the interaction can stimulate the catalytic activity of a number of these enzymes (Shereda et al., 2008).

Errors during Replication

There are numerous factors that might affect the complete and accurate duplication of the *E. coli* chromosome and the replication forks must overcome these obstacles in order to avoid the risk of genomic instability. Two main problems that elongating replisomes must deal with are instances of polymerase-associated replication errors and blocks to progression, such as small lesions or protein roadblocks.

Polymerase-associated replication errors

The replicative polymerase Pol III has a relatively low error rate in nucleotide incorporation due to its ability to discriminate between correct and incorrect base pairing (Kunkel, 2004). Errors that are made can be resolved by the 3' to 5' exonuclease proofreading capability of Pol III, which is provided by the DnaQ subunit (Scheuermann and Echols, 1984). Despite the relatively high fidelity resulting from these combined actions of DNA Pol III (Kunkel, 2004; Loeb and Kunkel, 1982), occasionally an error such as a base-base mismatch or small insertion-deletion mutation is made during DNA replication that is not resolved by Pol III and instead is corrected by the DNA methyl-directed mismatch repair (MMR) pathway. The combined effect of all three steps results in high fidelity of replication with an overall error rate of DNA replication reported to be between 10^{-9} and 10^{-10} (Fijalkowska et al., 2012). In addition, the MMR machinery has a role in regulating recombination through repairing mismatches and preventing recombination between homeologous sequences (Evans and Alani, 2000; Modrich and Lahue, 1996; Spies and Fishel, 2015). The combined effect of the activities of MMR is important in promoting genomic stability and the absence of a functional MMR system results in a mutator phenotype with increased spontaneous mutation rates (Schofield and Hsieh, 2003).

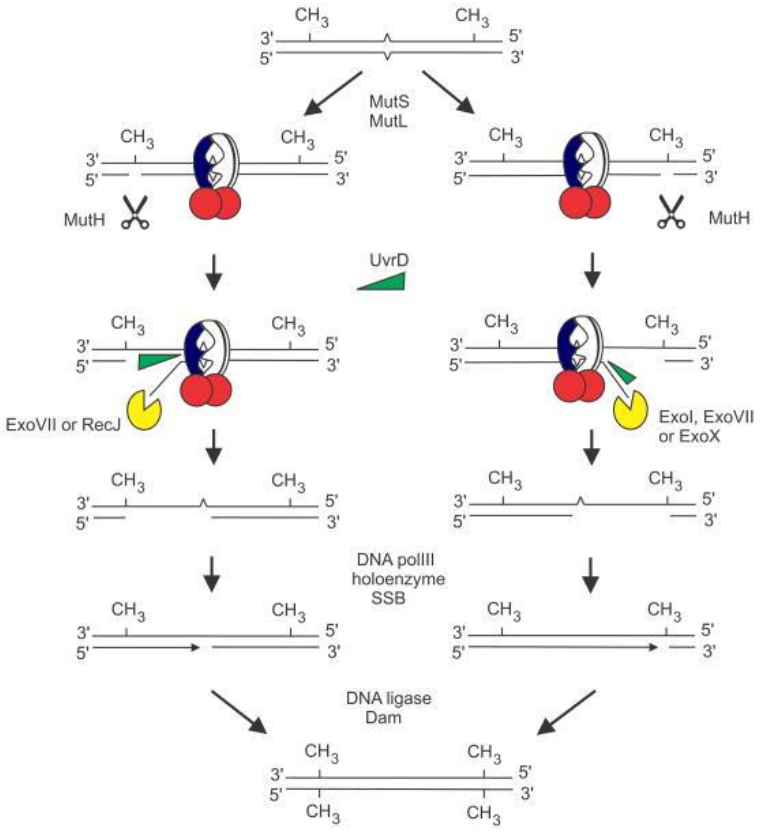

Figure 4: Schematic depicting the key steps of the MMR processing of a replication error. MutS is represented by the white and blue pierced circle and MutL is represented by the filled red circles. All other components are labelled within the diagram. The bidirectional nature of mismatch repair necessitates both exonucleases with 5' – 3' activity and exonucleases with 3' – 5' activity (reproduced from Marinus, 2012; ©2012 American Society for Microbiology. Used with permission. No further reproduction or distribution os permitted without the prior written permission of American Society for Microbiology).

MMR is initiated when MutS recognises and binds to a replication error such as a base pair mismatch (Su and Modrich, 1986; Su et al., 1988). MutS then recruits the MutL protein to the MutS-heteroduplex DNA complex. Once the mismatch has been identified, MutS-MutL complex interacts with MutH, a sequence-specific endonuclease that recognises GATC sites and introduces a ssDNA nick in the nascent strand of a hemimethylated GATC duplex (Welsh et al., 1987) such as that found transiently behind replication forks. The innate endonuclease activity of MutH is extremely weak (Welsh et al., 1987). Once activated by MutS-MutL, MutH activity was shown *in vitro* to be increased over 30-fold (Au et al., 1992) and produces a single stranded nick at a GATC sequence on the nascent, unmethylated strand. The GATC site targeted in MMR can be either 3' or 5' to the site of the mismatch (Grilley et al., 1993; Modrich, 2016). MutL facilitates UvrD (DNA helicase II) loading at the site of the endonucleolytic incision created by MutH (Matson and Robertson, 2006). UvrD then unwinds the dsDNA starting from the nick and proceeding towards and beyond the site of the mismatch error, generating a ssDNA flap. UvrD helicase has 3' – 5' polarity and so must be loaded on to the parent strand when the nick is 5' to the mismatch and on to the nascent strand when the nick is 3' to the mismatch, resulting in a 5' or 3' ssDNA flap respectively, which indicates that MMR requires the ability to remove single-strand flap structures of both polarities (Matson and Robertson, 2006). This functionality is provided by ssDNA

exonuclease proteins. ExoI and Exonuclease X (ExoX) are 3' ssDNA exonucleases that catalyse the degradation of the 3' ssDNA generated when MutH nicks the DNA 3' to a mismatch. RecJ and Exonuclease VII (ExoVII) are 5' ssDNA exonucleases that catalyse the hydrolysis of the ssDNA generated when MutH nicks the DNA 5' to a mismatch. ExoVII is also able to contribute to degradation of the 3' substrate as it has 3' – 5' activity in addition to its 5' – 3' activity (Burdett et al., 2001; Viswanathan et al., 2001), although it was found *in vitro* to be much less effective in the 3' – 5' direction (Grilley et al., 1993). The resulting ssDNA gap is stabilised by the binding of SSB. The normal replication polymerase, Pol III, is responsible for re-synthesising the section of excised nascent strand DNA, and DNA ligase seals the nick, completing the restoration of the duplex to the parental genotype. The basic steps of MMR described here are shown in Figure 4 (Reviews on MMR include: Harfe and Jinks-Robertson, 2000; Kunkel and Erie, 2005; Marinus, 2012; Modrich, 2016).

Cells lacking all four of the exonucleases involved in MMR were found to suffer from a loss of viability (Burdett et al., 2001), but the presence of any one of ExoI, ExoX, RecJ or ExoVII has been shown to be enough to support full mismatch correction and restore normal mutation rates (Viswanathan et al., 2001).

Barriers to replication elongation

In addition to managing misincorporations, replication forks will inevitably encounter blocks to progression, and it has been reported that in fact a significant portion of replication forks in normal chromosome replication stall or collapse due to obstacles such as lesions, protein-DNA complexes or DNA secondary structures (Cox et al., 2000; Lindahl, 1993; Maisnier-Patin et al., 2001; Mirkin and Mirkin, 2007; Rudolph et al., 2007a; Sandler and Marians, 2000). The consequences of such impediments to replication fork progression depend on the type of obstruction encountered.

Encounters between the replication machinery and transcription complexes have been shown to be a major cause of replication fork pausing but do not necessarily lead to replisome disassembly and fork inactivation (Gupta et al., 2013).

There are many repair processes in place to deal with the lesions resulting from DNA damage that are continuously arising (Kisker et al., 2013; Lindahl, 1993), however some lesions are not repaired in time and so are encountered by replication forks. The cellular response to replication forks encountering DNA lesions in the template during replication is complex and an extensive body of work surrounding the events that follow such occurrences has been generated, with controversy surrounding the exact mechanisms employed by cells (for reviews see Bichara et al., 2011; Fuchs, 2016; Lehmann and Fuchs, 2006; Lovett, 2017; McGlynn and Lloyd, 2002a and references therein). The response to a small lesion on the DNA differs depending on if it occurs on the lagging strand or the leading strand. There is plenty of evidence supporting the idea that the leading strand polymerase is blocked by lesions in its template strand and that replication cannot resume until the lesion has been repaired (Rudolph et al., 2007b). In contrast, it is generally accepted that a lesion occurring on

the lagging strand template does not pose such a problem to the progression of a replication fork and can be by-passed by the replisome skipping to the next Okazaki fragment. The resulting ssDNA gap containing the lesion is subsequently resolved through recombination with the sister chromosome (McGlynn and Lloyd, 2002a). Recently, the idea that the leading strand synthesis is able to re-initiate downstream of the site of a lesion and is only dependent on DnaG primase and not replication re-start proteins has garnered renewed interest (Lehmann and Fuchs, 2006; Yeeles and Marians, 2011); it remains to be determined exactly what occurs when a replication fork encounters a lesion in the leading strand template.

PriA-mediated replication restart

Regardless of the cause, if a stalled replication fork is unable to resume replication once the block to replication is resolved, chromosome duplication will be unsuccessful unless the DNA downstream of the stalled fork is replicated. The fork replicating the other replichore is blocked from progressing out of the termination region opposite *oriC* by the presence of the Tus/*ter* replication fork trap (page 24) and so is unable to rescue the stalled fork, and the DnaA-*oriC* initiation system is unable to catalyse the assembly of replication forks away from the origin of replication. It was recently reported that cells in which the replication origin has been moved to a location roughly 1 Mbp away from its original location (*ΔoriC oriZ*$^{+}$) have a slow growth phenotype (Ivanova et al., 2015). The replichore arrangement has been disrupted in these cells and instead of replicating roughly equal halves of the chromosome, the anti-clockwise replication fork must replicate ¾ of the chromosome whilst the clockwise fork replicates just ¼ before it is blocked at a Tus/*ter* complex. The introduction of a *tus* deletion was found to be a good suppressor of the slow growth, which highlights the importance of this aspect of replication (Ivanova et al., 2015).

Replication restart pathways enable the re-assembly of the replisome at the location of the stalled fork and the resumption of DNA synthesis. The stochastic nature of replication fork stalling means that unlike DnaA-mediated replication initiation, specific DNA sequences cannot form the basis of initiation and instead, restart proteins recognise specific DNA structures where replisome assembly is required (Heller and Marians, 2006). PriA is the main replication restart protein and *ΔpriA* mutants have a complex phenotype that includes slow growth and reduced viability (Lee and Kornberg, 1991) and rapidly acquire suppressor mutations, demonstrating the importance of the replication restart process (Sandler and Marians, 2000). PriA has been shown to recognise and load a replisome at multiple DNA structures, including various stalled fork structures and D-loops that might be formed via homologous recombination (Gabbai and Marians, 2010; McGlynn et al., 1997; Nurse et al., 1999; Tanaka and Masai, 2006). The major restart pathway proposed is referred to as the PriA-PriB-DnaT pathway (Michel and Sandler, 2017). PriA binding to a stalled fork structure is followed by the binding of PriB. DnaT is then required to facilitate DnaC in the loading of the replicative helicase DnaB (Michel and Sandler, 2017). DnaB encircles the lagging strand and so a single-stranded region on the lagging

strand template is required for DnaB loading. PriA has a 3' – 5' DNA helicase activity that is able to remodel stalled fork structures where necessary to accommodate this requirement, although this function is not essential for replication restart (Gabbai and Marians, 2010; Zavitz and Marians, 1992).

In addition to the PriA-PriB pathway, *E. coli* cells have a second PriA-mediated replication restart capability, the PriA-PriC pathway. There is redundancy between these two pathways, demonstrated by the almost wild type phenotypes of either a *priB* or *priC* deletion and the fact that *ΔpriB ΔpriC* double mutants have very poor viability and slow growth phenotype (Sandler et al., 1999), although the PriA-PriB pathway is thought to be the dominant pathway in replication restart (Windgassen et al., 2018). PriC has been shown *in vitro* to bind preferentially to ssDNA and be most active on stalled replication forks that have a short stretch of ssDNA between the nascent leading strand and the fork junction (Windgassen et al., 2018).

Termination

Whilst replication initiation and elongation have been extensively studied and are generally well understood, the events that take place during the final phase of DNA replication, called termination, have received much less attention. The two replication forks traversing the circular chromosome in opposite directions approach each other. The area between the converging replisomes must be replicated in its entirety and with high accuracy and the two duplex DNA molecules generated by one replisome must be fused with the corresponding duplexes of the opposing replisome. Duplication of the chromosome is then complete (Dewar and Walter, 2017).

The replication fork trap

Given the circular nature of the *E. coli* chromosome and the fact that there is a single origin where replication is initiated bidirectionally, it follows that there will be a single fork fusion event per cell cycle where the two replisomes will meet each other head on. In *E. coli*, this termination event takes place in a region approximately opposite *oriC*. By introducing ectopic origins at different locations in the chromosome and supressing initiation at *oriC*, it was found that the termination location is not affected by the position of the origin of replication. The realisation that replication forks were unable to progress out of the terminus area led to the discovery of the replication fork trap in the area of the chromosome opposite *oriC* (reviewed in Neylon et al., 2005). This region is flanked by ten primary *ter* (termination) sites (*terA–J*). The sequence of each of the 23 bp *ter* sites is asymmetric and has a strictly conserved GC6 base pair followed by a very highly conserved consensus sequence forming the core region (Coskun-Ari and Hill, 1997; Duggin and Bell, 2009; Neylon et al., 2005).

In addition to the *ter* sites, a DNA-binding protein encoded by the *tus* gene is involved in forming the replication fork trap (Hill et al., 1989). The terminus utilization substance, Tus protein, recognises and binds asymmetrically to *ter* sequences, and the resulting Tus/*ter* complex forms a barrier to the progress of replication forks in an orientation-specific manner. A replication fork is blocked when

approaching a *ter*-Tus complex from a non-permissive direction but is able to pass when approaching from the opposite direction. The *ter* sites are arranged into two opposing groups spanning approximately 45 % of the chromosome, as shown in Figure 5. All ten primary *ter* sites are orientated to allow replication forks to proceed when travelling in an origin to terminus direction but to block progression of forks in the terminus to origin direction, resulting in a region of over 260 kb between the innermost *ter* sites that replication forks can enter but cannot leave (Duggin et al., 2008; Neylon et al., 2005) (Figure 5). The presence of a block preventing replication fork progression out of the terminus region dictates that half of the chromosome is replicated by the fork moving clockwise from *oriC* and the other half is replicated by the fork moving anti-clockwise, dividing the chromosome into two approximately equal replichores (Reyes-Lamothe et al., 2012).

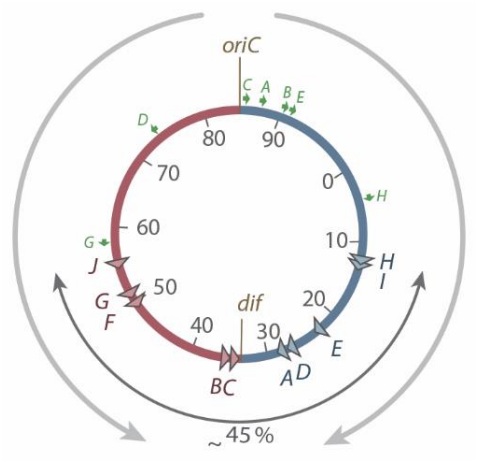

Figure 5: Map of the *E. coli* chromosome showing the location of the primary *ter* sites (*terA-J*, triangles), the highly transcribed *rrn* operons A-D, G and H (green arrows) and the location of the chromosome dimer resolution site, *dif*. Numbers represent the minutes of the standard genetic map. Tus bound to a *ter* site forms a semi-permissive block to replication. Forks traversing the left hand replichore (red) from *oriC* are able to pass *terJ*, *G*, *F*, *B* and *C* but are blocked by *terA*. Conversely, *terH*, *I*, *E*, *D* and *A* are permissive to forks traversing the right hand replichore (blue) from *oriC*, but these forks are blocked at *terC* (figure reproduced from Dimude et al., 2016 with permission).

The gene encoding Tus is located within the termination region just downstream of *terB*. The *tus* promoter and ribosome binding site overlap *terB*. As might be expected, it was found that the transcriptional start site is within *terB* and the expression of *tus* is autoregulated through the binding of Tus protein to *terB*, which inhibits transcription of *tus* (Neylon et al., 2005).

Tus is a monomeric protein and a single Tus molecule binds to each *ter* site (Coskun-Ari et al., 1994). The binding affinity of Tus to *ter* is very high; the *in vitro* equilibrium dissociation constant (K_d) of *terB*-Tus was reported as 3.4×10^{-13} mol L^{-1} (Gottlieb et al., 1992), but it does not bind to all the *Ter* sites equally. Moreau and Schaeffer (2012) reported that Tus binds strongly to *terA-E* and *terG* and less strongly to the remaining *ter* sites. Given that *terA-D* are the four innermost *ter* sites, it is not surprising that Tus forms a tight bond with these *ter* sites; they are the most likely to encounter replication forks from the blocking orientation (Figure 5). The data from Moreau and Schaeffer are in line with the work by Duggin and Bell (2009) on the fork pausing efficiencies of each *ter*-Tus complex,

which they found differed considerably between the different *ter* sites, and they observed that sites *terA-D* have the highest innate efficiency of fork pausing.

Mechanism of polar replication fork arrest

Following the identification of the *ter* sites and the *ter*-binding protein Tus as the components needed to form a polar block to replication fork progression, the next question was how a *ter*-Tus complex achieves this. Despite a large body of work, the mechanism of polar replication fork arrest has not yet been fully determined and there are conflicting reports on what the key interactions might be.

One suggestion is that the presence of Tus bound to the DNA acts as a simple clamp and presents a non-specific but polar block to proteins translocating along the DNA, including the replisome. Using the *oriC* plasmid with ectopic *terB* sequences inserted in various orientations, Lee and colleagues (1989) showed that *ter*-Tus complexes can form a block to helicase-mediated strand separation, which includes Rep and UvrD helicases in addition to the replisome helicase, DnaB, in an orientation-specific manner. Additionally, the strand displacement activity of PriA has also been shown *in vitro* to be inhibited by *ter*-Tus complexes (Hiasa and Marians, 1992; Lee and Kornberg, 1992). This means that the non-permissive end of the polar *ter*-Tus complex that arrests DnaB translocating in the 5'-3' direction also inhibits progress of other helicases translocating on the other strand in the 3'-5' direction (Neylon et al., 2005). It has also been shown that Tus-*ter* complexes blocks RNAPs in a polar fashion (Mohanty et al., 1998), and that Tus-*ter* complexes maintain functionality when engineered in the genome of *Saccharomyces cerevisiae* and cause a block to replication forks in a polar manner (Larsen et al., 2014). These findings are in line with the idea that a Tus-*ter* complex acts as a 'molecular clamp' and is non-specific in its inhibition of movement.

However, many have challenged the idea that DNA binding alone can explain the mechanism of a Tus-*ter* block, and there are numerous reports that support a model of specific protein-protein interactions between a replisome component, presumably the DnaB helicase, with the non-permissive face of the Tus-*ter* complex. Mutations in *tus* have been identified that result in a reduced ability to block replication fork progression but do not affect the DNA-binding affinity of Tus and the mutant proteins form stable complexes with *ter* sites (Henderson et al., 2001; Mulugu et al., 2001; Skokotas et al., 1995). It has been reported that a Tus-*ter* complex inhibits translocation of DnaB in a polar fashion even in the absence of strand separation (Bastia et al., 2008), which is in contrast to the findings of Mulcair *et al.* (Mulcair et al., 2006) that a structural change in the DNA that occurs only when a helicase elicits strand separation from the non-permissive direction is responsible for the polar arrest of replication forks. Additionally, it was shown that a *tus* mutant that retained normal DNA binding affinity but a reduced interaction of DnaB-Tus also resulted in defective arrest of the translocating DnaB (Bastia et al., 2008).

The structural change identified by Mulcair and colleagues (2006) was shown to occur when strand separation exposed the strictly conserved C(6) residue at the non-permissive end of *ter*. Their

data indicate that the strand separation caused by the approaching replisome results in the C(6) residue flipping out of the double helix and binding to a specific location of the Tus protein, which results in the formation of a very stable 'locked' Tus-*ter* complex (C(6) lock complex), which they suggest provides an explanation for the polarity of the blocking action of Tus-*ter* complexes. More recent studies reinforce the importance of the C(6) residue in the formation of a robust block to replication fork progression (reviewed in Berghuis et al., 2018).

Given the numerous studies in support of different models and presenting conflicting information, the interaction appears to be quite complex and it is likely that multiple effects contribute to the tight block in *E. coli*. Despite the work by Larsen and colleagues showing that *E. coli* Tus-*ter* complexes function in *Saccharomyces cerevisiae* (Larsen et al., 2014), they also revealed that the Tus-*ter* complex presents a far less effective block to replication in yeast compared to its activity in *E. coli* as three complexes in tandem were still overcome by the yeast replisome, whereas in *E. coli*, a single *ter* site is very effective (Bidnenko et al., 2002; Dimude et al., 2016; Ivanova et al., 2015). It has also been shown that while the functionally similar but unrelated RTP-*Ter* system of *B. subtilis* operates effectively in *E. coli* (Andersen et al., 2000; Kaul et al., 1994), the efficiency of replication fork arrest by Tus-*ter* complexes differed in the two organisms and was much higher in *E. coli* than in *B. subtilis* (Andersen et al., 2000). The crystal structure of the Tus-*ter* complex led Kamada and colleagues (1996) to suggest that the polar nature of the block to replication fork progression might be due to differing, orientation-specific interactions between Tus and an approaching helicase protein. The *ter*-binding regions of Tus are inaccessible to the helicase at the non-permissive end of the complex due to bulky domains in this region of Tus. Tight binding of Tus to *ter* sequences would be required to form this steric block to fork progression. This steric block is not present when approaching from the permissive orientation, and so helicase unwinding could disrupt the DNA-binding of Tus and therefore is able to displace Tus and continue moving past the *ter* site (Kamada et al., 1996). Together, these observations support the suggestion that both protein-protein and protein-DNA interactions are important in ensuring an efficient and polar block to replication fork progression (for reviews, see Berghuis et al., 2018; Duggin et al., 2008; Neylon et al., 2005).

The role of the replication fork trap

Despite the fact that much of the work on termination systems so far has focussed on understanding how they operate, arguably the more interesting question is: what is the purpose of a defined termination area that allows replication forks to enter but not to leave this region? If, as has been reported, a significant portion of replication forks in normal chromosome replication stall or collapse due to impediments such as endogenous DNA damage, protein-DNA complexes or DNA secondary structures even in the absence of exogenous damage (Cox et al., 2000; Lindahl, 1993; Maisnier-Patin et al., 2001; Mirkin and Mirkin, 2007; Rudolph et al., 2007a; Sandler and Marians, 2000), then it would make sense for whichever fork arrives first at the half way point opposite the replication origin to pass

through the terminus and meet the opposing fork that may have been delayed on its journey. Whilst this would reduce any delay to chromosome duplication and cell division, the replication fork trap ensures that this does not happen. Moreover, if a blocked fork fails to reach the termination region, replication of the chromosome will be incomplete and the cell might not survive (Maisnier-Patin et al., 2001; Sharma and Hill, 1995). This highlights the importance of the mechanisms that enable repair and restart of stalled replication forks in bacteria, such as PriA-mediated replication restart (Michel and Sandler, 2017) (page 23) in cellular DNA metabolism as part of normal bacterial growth (Cox et al., 2000). If the presence of a replication fork trap can pose a threat to the viability of the cell, there must be some advantage that outweighs the risks posed by preventing one replication fork from leaving the terminus to rescue a stalled fork.

It is not immediately obvious what the role of the fork trap might be and there are a number of observations that confuse the matter. With almost no exception, all bacterial species have circular chromosomes like that in *E. coli* and yet a fork trap system has only been identified in a limited number of species, demonstrating that it is not an essential feature for successful replication termination in a circular chromosome (Neylon et al., 2005). On the other hand, the components of the functionally-related RTP-*ter* replication fork trap of *B. subtilis* show no structural or sequence homology with the *E. coli* system, indicating that the two systems have evolved independently and that a fork trap fulfils an important physiological function (Neylon et al., 2005). Inactivation of the replication fork trap by the deletion of *tus* does not induce a particularly obvious phenotype (Roecklein et al., 1991); single mutant cells exhibit a slow growth phenotype, but the effect is very mild (Ivanova et al., 2015). In line with this, the location of replication fork fusion is unaffected by the presence or absence of the fork trap (Dimude et al., 2016; Ivanova et al., 2015; Rudolph et al., 2013). This was also shown to be the case in *B. subtilis* (Kono et al., 2014).

Despite a lot of focus on replication forks stalling at Tus-*ter* complexes, the number of forks that terminate replication at a Tus-*ter* complex *in vivo* is relatively low in wild type cells (Duggin and Bell, 2009). Recent data suggest that arrest of forks by Tus-*ter* can in fact lead to problems (Midgley-Smith et al., 2018). If one fork is significantly slower than the other in reaching the termination region, maybe following delay at a replication block, then the fork that arrives first will eventually encounter a Tus-*ter* complex from the non-permissive direction. Whilst a fork being stably arrested at a Tus-*ter* complex might provide an advantage at first and allows the second fork time to catch up, *in vitro* and *in vivo* measurements of fork stability at obstacles including nucleoprotein roadblocks suggest a limited half-life of 4–6 mins (Marians et al., 1998; McGlynn and Guy, 2008; Mettrick and Grainge, 2016). This suggests that after a relatively limited period of time, forks become inactivated and can be processed by recombination proteins, which can have pathological consequences that result in destabilisation of the genome (Lambert et al., 2005; Midgley-Smith et al., 2018). In line with this, it has been reported previously that the termination region is a hotspot for recombination (Horiuchi et al., 1994). Our work suggests that this might be an important reason why forks do not get held by Tus-*ter* as a general rule (Midgley-Smith et al., 2018). It seems therefore that the fork trap is not needed to

mediate fork fusion events in normal replication termination so there must be another reason for it to exist.

Encounters between replication and transcription

The polymerases that carry out replication and gene transcription share the same template and in the case of bacteria, there is no temporal separation of the two processes and so they occur simultaneously. To complicate matters further, the two complexes are often moving in opposing directions and they progress at very different rates, with transcription speed reported to be 10-20 times slower than replication (Dennis et al., 2009; Gotta et al., 1991; Pham et al., 2013). Collisions between replication and transcription complexes are inevitable and have been shown to be problematic; in particular, head on collisions between the replication and transcription machineries have been shown to result in serious inhibition of replication fork progression (Ivanova et al., 2015; Kim and Jinks-Robertson, 2012; McGlynn et al., 2012; Mirkin and Mirkin, 2005; Poveda et al., 2010; Rudolph et al., 2007a). Given that there are over 4000 genes within the genome sequence of *E. coli*, distributed throughout the chromosome (Blattner et al., 1997), conflicts between replication and transcription must be unavoidable, especially in rapidly growing cells that not only exploit overlapping rounds of DNA replication, resulting in multiple replisomes operating on the DNA at once, but that also will require increased transcription to keep up with demand for cellular components.

The strong indication that head-on conflicts are problematic provides a strong argument for the importance of maintaining co-directionality between replication and transcription. Maybe the function of the replication fork trap is to maintain co-directionality of replication and transcription by preventing movement of replication forks out of the terminus region and in to the opposite replichore (Brewer, 1988). This set up dictates that each replichore is always replicated in the same orientation. This creates a situation where it is possible to align transcription with this. Whilst encounters between replication and transcription machineries moving in the same direction can still cause problems (Ivanova et al., 2015; Merrikh et al., 2011), avoiding head-on collisions would undoubtably be an advantage to cells.

In line with this, the large majority of highly transcribed genes in *E. coli* are located on the leading strand template and so are transcribed co-directionally with the movement of the replisome during replication. A similar situation is found in *B. subtilis*, which displays a particularly high co-orientation of the *rrn* operons with replication, but also substantial co-directionality in general with an over-all co-directionality of almost 75% (McLean et al., 1998). The overall co-orientation in *E. coli* however is much lower, with approximately 55% of genes aligned with replication (Blattner et al., 1997; Brewer, 1988; McLean et al., 1998). An analysis of a number of different genes that are highly transcribed under fast growth conditions has recently been conducted. The findings suggest that while a high proportion of these genes have been found to display co-directionality with replication (Brewer, 1988; Ellwood and Nomura, 1982; Jin et al., 2012; McLean et al., 1998), the location of these genes should

also be taken in to consideration (Dimude et al., 2016). Both the *rrn* operons, which encode ribosomal RNA, and the ribosomal protein genes are highly skewed to the origin-proximal half of the chromosome. In order for the observed co-directionality to be relevant in explaining the presence of the fork trap, there must be a threat that replication forks escaping the termination area in the absence of Tus will reach these regions of the chromosome. It is unlikely that this situation would occur very often. Using strains carrying an ectopic replication origin, it was possible to investigate chromosome dynamics in the presence and absence of functioning replication fork blocks under conditions where the midpoint of replication (normally within the terminus) is skewed in one direction or the other (Ivanova et al., 2015). It was found that replication forks that are able to progress through the termination region and in to the opposite replichore were able to proceed with few problems (Dimude et al., 2018a; Ivanova et al., 2015). In line with this, tRNA genes, which are much more evenly distributed than *rrn* operons and the ribosomal protein genes, were found to have a high degree of co-directionality when located in the origin-proximal half of the chromosome but those located in the origin-distal half showed a slight bias for the opposite orientation, which would see replication-transcription encounters at these genes occurring head-on (Dimude et al., 2016). Whilst co-orientation of replication and transcription may provide an advantage that exists as a result of the presence of the replication fork trap, it does not seem to be the driving force behind its existence, possibly in line with the observation that not all bacterial species have a fork trap.

Over-replication at fork fusion locations

Despite the presence of the replication fork trap, the fraction of forks stalled at Tus-*ter* complexes is relatively low (Duggin and Bell, 2009) and instead the majority of replication forks fuse freely in the region between the innermost *ter* sites and not at a Tus-*ter* complex, as outlined above. There have been a number of reports using various experimental approaches that have shown that over-replication of the DNA can occur at fork fusion sites.

Hiasa and Marians studied replication termination using an *in vitro* system of minichromosomes containing *oriC* for replication initiation and two *ter* sites arranged opposite the origin with a slight asymmetry, simulating the situation found in the chromosome (Hiasa and Marians, 1994). They found that in the presence of Tus, replication terminated at one or the other Tus-*ter* complexes. In the absence of Tus, termination occurring through free-fusing fork events led to over-replication of the DNA, resulting in double-stranded nascent DNA. The suggested pathway that could result in over-replication at replication fork fusion was the displacement by DnaB helicase of the 3' end of the nascent strand, which is unable to occur when a fork is blocked by a Tus-*ter* complex. A similar situation was observed in an *in vivo* investigation of replication, supporting this suggestion. When investigating replication intermediates arising in the R1 plasmid, which initiates unidirectional replication from a single origin, Krabbe et al (1997) found that in the absence of Tus, which ordinarily binds to two naturally occurring, oppositely orientated *ter* sites that are situated between the origin

and the DNA synthesis start site of the leading strand, the plasmid becomes very unstable. This is coupled with an accumulation of multimeric plasmid forms and rolling circle replication. The authors suggested that this might arise as a result of the replisome progressing past the origin and encountering the already replicated region of the plasmid, where the replicative helicase then proceeds to unwind the nascent strand, and which by strand switching of the leading strand polymerase becomes the new leading strand template (Krabbe et al., 1997).

The absence of the terminator protein, RTP, from *B. subtilis* cells was shown to contribute to problems at replication termination when combined with mutations that hinder dimer resolution and chromosome partitioning following the fusion of replication forks (Lemon et al., 2001). Similarly, Markovitz (2005) combined *Δtus* with different *polA* alleles encoding mutant DNA polymerase I proteins and observed over-production of DNA.

Origin-independent DNA synthesis at replication termination

So far, all phenotypes of cells lacking Tus are very mild. Recent work has revealed a more conspicuous phenotype of *Δtus* cells than has been seen in the past. Rudolph and colleagues found that cells lacking RecG helicase took far longer than wild type cells to recover from DNA damage and formed long filaments with high incidences of initiation of DNA synthesis (Rudolph et al., 2009a, 2009b). They hypothesised that RecG might have a role in managing replication fork fusions, which normally take place within the defined termination region. When they investigated chromosome dynamics in exponential phase *ΔrecG* cells in the absence of DNA damage-induced synthesis, they found that pathological origin-independent over-replication occurs specifically within the termination region (Rudolph et al., 2013), which has been confirmed a number of times since (Azeroglu et al., 2016; Dimude et al., 2015; Wendel et al., 2014). This synthesis is robust enough to sustain cell growth in the absence of origin firing if fork traps are inactivated by the deletion of the *tus* gene, which allows the DNA synthesis initiated in the termination region to extend beyond this area and proceed back towards *oriC* (Dimude et al., 2015; Rudolph et al., 2013). Replication forks moving in a direction opposite to normal can cause problems for cells (Ivanova et al., 2015; Lang and Merrikh, 2018; Wang et al., 2011); it is therefore unsurprising that the origin-independent growth seen in the absence of RecG is improved further by the introduction of a point mutation in the RNA polymerase (Dimude et al., 2015; Rudolph et al., 2013) that reduces replication-transcription conflicts (Dutta et al., 2011). When these two mutations are combined in a *ΔrecG* background, the aberrant DNA synthesis seen in the termination area of *ΔrecG* cells is able to sustain robust cell growth in the absence of DnaA activity and the cells can tolerate the deletion of the entire oriC region (Rudolph et al., 2013).

The work by Rudolph and colleagues indicates that RecG is a key player in managing events at replication termination by processing intermediates generated at replication fork fusions (Rudolph et al., 2009a, 2009b, 2010a, 2013). Over-replication of the termination region is also observed in cells lacking 3' exonucleases (Rudolph et al., 2010a, 2013; Wendel et al., 2014). A model to explain how the

over-replication might arise via the formation and subsequent exploitation of a 3' ssDNA flap has been suggested (Dimude et al., 2016; Lloyd and Rudolph, 2016; Rudolph et al., 2009b, 2010b, 2013) (Figure 6). As two replication forks meet, the DnaB helicase of one replisome may displace the leading strand polymerase of the opposing replisome and will encounter dsDNA; the parent strand and the newly replicated daughter strand. If DnaB unwinds the dsDNA, this will result in a 3' ssDNA flap as the newly replicated daughter strand is displaced from the parent strand. 3' flaps can be degraded by 3' ssDNA exonucleases such as ExoI, ExoVII and SbcCD (Lovett, 2011) (page 36) but are also excellent substrates for RecG. *In vitro*, RecG is able to act on many branched DNA structures (Briggs et al., 2004; McGlynn and Lloyd, 2002b; Rudolph et al., 2010b) (page 34). RecG has a high affinity for 3' single-stranded flap structures and has been shown to remodel such a structure into a 5' flap as it can unwind the 5' end at the branch point of a 3' flap while simultaneously reannealing the 3' single-strand flap (Bianco, 2015; Briggs et al., 2004; McGlynn et al., 2001; Tanaka and Masai, 2006). In the absence of RecG or 3' exonucleases, 3' flaps can persist and can instead be targeted by PriA, which promotes *oriC*-independent loading of DnaB and subsequent replication fork assembly, to re-start replication (Windgassen et al., 2018). The resulting DNA duplex on the leading strand will have a dsDNA end, which is a substrate for RecBCD to process and load RecA (Kowalczykowski, 2000), thereby promoting homologous recombination with the sister chromosome and generating a substrate from which PriA can establish another replication fork that moves in the opposite direction.

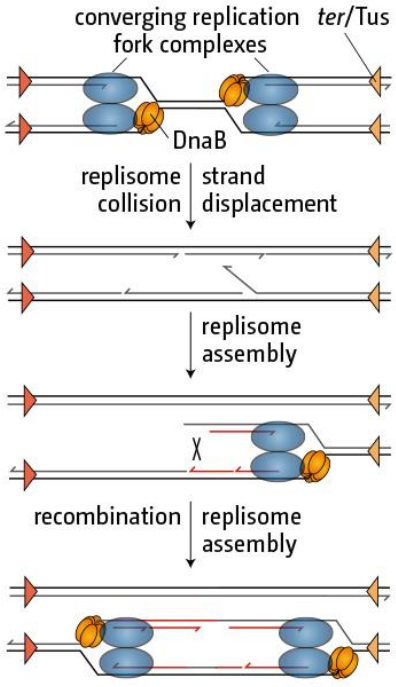

Figure 6: Schematic representation of how the fusion of two replication forks at replication termination might lead to the initiation of origin-independent over-replication (reproduced from Rudolph et al., 2013 with permission).

This over-replication is contained by the replication fork trap, which prevents forks established within the termination region from leaving. The threat to genomic stability in areas where replication forks normally meet would suggest that the termination area exists in order to limit the location of where these collisions are able to occur and to contain any problems that result from collisions, for example, over-replication of an already-replicated chromosome, thereby contributing to the maintenance of genomic stability. In line with this, the termination area has previously been described as a recombinational hotspot (Horiuchi et al., 1994) and the events outlined above provide additional explanation for why that might be. The molecular details of exactly what happens when two replication forks meet and fuse is not yet well understood in either prokaryotic or eukaryotic cells. The findings that replication fork fusion events have the potential to cause pathogenicity began to highlight some of the key players involved in protecting cells from the effects of fork collisions (Rudolph et al., 2009a, 2009b, 2013).

Post-replicative processing of daughter chromosomes

The converging of two replication forks at termination presents a number of potential challenges that must be overcome by the cell and as such, there are a number of processes in addition to the fusion of two replication forks that are localised to the termination region and that promote resolution of the daughter chromosomes.

As DNA unwinds during replication, twisting tension builds up in the rest of the coiled molecule. This results in the DNA molecule twisting around itself to form positive supercoils to accommodate the helical tension. If no action is taken, it will become impossible to separate the strands of DNA any further as it will be too tightly twisted. Topoisomerases type I and type II are enzymes that manage the topology of DNA by either increasing the forming supercoils or releasing supercoils (Champoux, 2001; Wang, 1996). As the replication forks converge, DNA gyrase, a type II topoisomerase that normally removes positive supercoils (Nöllmann et al., 2007), will at some point be unable to access the DNA. It has been shown that the over-winding ahead of a replication fork can force the fork to rotate, allowing the supercoils to diffuse behind the replication fork and resulting in an intertwining of the two daughter chromosomes (Cebrián et al., 2015). In addition to positive supercoiling, the two circular chromosomes are topologically intertwined (catenanes) due to the helical nature of DNA. All links must be resolved in order to allow separation of the chromosomes before cell division occurs, and this function is provided by topoisomerase IV, a type II topoisomerase that unwinds the linkages between the chromosomes by passing DNA though a transient dsDNA break (Zechiedrich and Cozzarelli, 1995; Zechiedrich et al., 1997).

Homologous recombination that occurs during a round of DNA replication can result in the formation of a chromosome dimer, which is a structure formed when two fully replicated chromosomes are joined together as a single, large circle (Barre et al., 2001; Lesterlin et al., 2004). Dimer resolution occurs at the *dif* (deletion-induced filamentation) site where two site-specific

recombinases, XerC and XerD, are recruited and function in a sequential manner to introduce a crossover at this location, which resolves the dimer to two monomers. FtsK is a multifunctional DNA translocase that is targeted to the division septum and coordinates segregation of the chromosome with cell division (Bigot et al., 2007; Reyes-Lamothe et al., 2012; Sherratt et al., 2010). As part of this function, FtsK mediates Xer-*dif* site-specific recombination in the resolution of chromosome dimers. FtsK utilises polar sequences called KOPS (FtsK orientating polar sequences) to sort the DNA into daughter chromosomes on either side of the septum (Bigot et al., 2005). This activity results in the two *dif* sites co-localising to the septum region and the XerCD-*dif* complex undergoes resolution to chromosome monomers (Barre et al., 2001; Reyes-Lamothe et al., 2012).

RecG

The *recG* gene encodes the monomeric RecG protein (McGlynn et al., 2000; Singleton et al., 2001) and is not highly expressed (Lloyd and Sharples, 1991); one report suggests that RecG is present at 7 molecules per cell (Taniguchi et al., 2010). RecG is classified as a Superfamily 2 (SF2) DNA helicase and in contrast to most other helicases, it unwinds DNA by translocating on double-stranded DNA (dsDNA) rather than on ssDNA (Singleton et al., 2007). RecG has been shown to target a variety of branched DNA structures *in vitro* (Briggs et al., 2004; McGlynn and Lloyd, 2002b; Rudolph et al., 2010b). Although the gene for RecG was first discovered during a screen for recombination-deficient mutants of *E. coli* almost 50 years ago (Storm et al., 1971), it has proved difficult to determine exactly what its role in cellular metabolism is. In line with this, besides the significant role of RecG at replication termination outlined above (page 31), a number of other roles for RecG have been proposed.

An investigation into recombination mutants 20 years after the initial characterisation of the gene resulted in further analysis of the *recG* locus, revealing that *ΔrecG* mutants have a mild sensitivity to UV light and mitomycin C and increased sensitivity to ionising radiation, indicating a reduction in DNA repair capability (Lloyd and Buckman, 1991). The extensive work carried out by the Lloyd laboratory also revealed that RecG can catalyse branch migration of Holliday junctions (Lloyd and Sharples, 1993). The observation that cells lacking both RecG and the Holliday junction resolvase RuvABC show a synergistic sensitivity to DNA damage (Lloyd, 1991) led to the suggestion that both RecG and RuvABC might have overlapping functionalities. The reduced recovery of conjugational recombinants in *Δruv ΔrecG* double mutants further contributed to this idea (Lloyd, 1991). However, whilst RecG can catalyse branch migration of Holliday junctions, it cannot function in Holliday junction resolution, and a resolvase that works with RecG in a situation analogous to the RuvABC system (West, 1997) has not been found.

Further *in vitro* studies showed that RecG is able to process R-loops by unwinding the DNA-RNA hybrid (Fukuoh et al., 1997; Vincent et al., 1996), and can unwind D-loops (McGlynn et al., 1997)

leading to suggestions that the function of RecG is to process these structures in order to avoid interference with DNA replication or transcription (McGlynn and Lloyd, 2002a). RecG has also been implicated as a key player in replication fork repair by promoting the regression of replication forks that have been stalled at a block of some kind, such as sites of DNA damage, protein-DNA complexes or DNA secondary structures, which allows repair or bypass of a lesion via a nonmutagenic process (Bianco, 2015; Cox et al., 2000; Gupta et al., 2014; Manosas et al., 2013; McGlynn and Lloyd, 2002a; McGlynn et al., 2001; Michel et al., 2001). More recently, work on the CRISPR-Cas adaptive immune system has shown that RecG is needed for primed adaptation to take place (Ivančić-Baće et al., 2015). In line with reports that show that the absence of RecG results in an increase in certain types of recombination (Lloyd and Rudolph, 2016; Lovett, 2006; Lovett et al., 1993), Azeroglu and colleagues (2016) recently proposed that RecG is important in managing replication initiated at branched intermediates of recombination associated with double-strand break repair (DSBR) to ensure that converging replication forks are established and over-replication via divergent replication is prevented (Azeroglu and Leach, 2017).

It was recently reported that RecG co-localises with sites of active DNA replication (Upton et al., 2014). RecG has previously been shown to interact with SSB *in vitro* (Buss et al., 2008; Shereda et al., 2008; Zhang et al., 2010). It was recently shown that RecG binds to SSB *in vivo* (Yu et al., 2016) and that this interaction is mediated through the C-terminal tail of SSB (Bianco and Lyubchenko, 2017). This interaction localises RecG to ssDNA areas of the chromosome where SSB is bound, which, given that SSB is concentrated around active replisomes (page 19), results in RecG being localised to sites of DNA synthesis and allows access of RecG to the replication fork, which might be responsible for enabling RecG monomers to be present in the cell at such low frequencies (Taniguchi et al., 2010).

The range of suggestions for what RecG is doing *in vivo* is reflected in the reported affinity of RecG for a variety of dsDNA substrates, which crucially contain a 3’ ssDNA branch (McGlynn et al., 2001; Tanaka and Masai, 2006), and the crystal structure reveals how RecG might process these substrates. The crystal structure of RecG protein bound to a synthetic DNA substrate is shown in Figure 7A. The interaction between the protein and the DNA revealed how it can unwind a forked DNA structure (Singleton et al., 2001), whereby the template strands run either side of a wedge-like structure that is part of Domain 1 in grooves that do not accommodate duplex DNA (Singleton et al., 2001). It has been proposed that the helicase activity provided by Domains 2 and 3 pulls the template strands through the grooves, resulting in the separation of the nascent strands from their corresponding template due of the channelling of the wedge domain, and the subsequent reannealing of the two parent strands (Singleton et al., 2001).

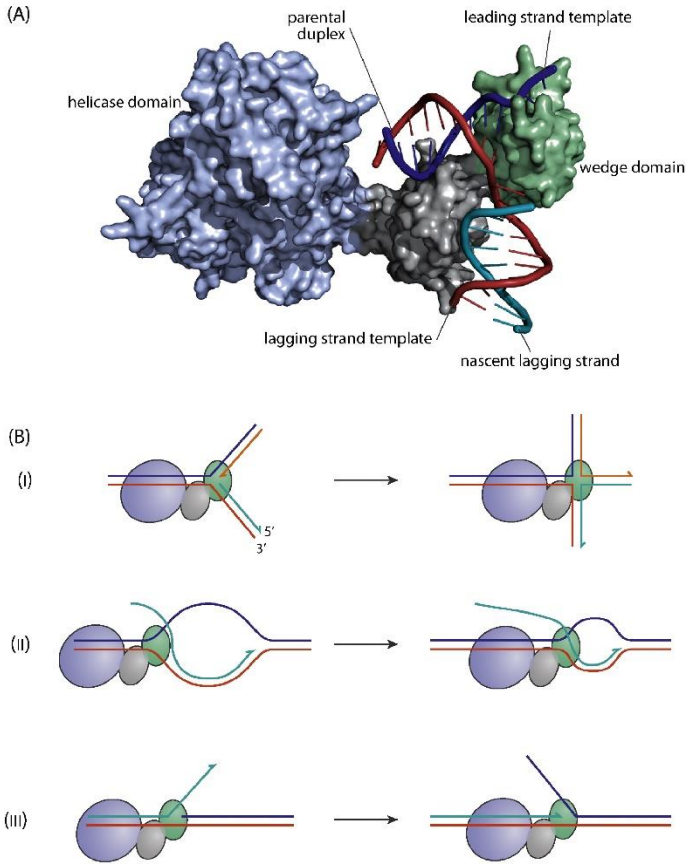

Figure 7: RecG at a lagging strand replication fork. A) *E. coli* RecG modelled onto the X-ray crystal structure of *Thermotoga maritima* RecG is shown bound to a lagging strand fork. The wedge of Domain 1 is highlighted in green. B) RecG has been shown to catalyse a number of reactions at branched DNA structures *in vitro*; (I) fork reversal leading to the formation of a Holliday junction, (II) unwinding a D- or R-loop and (III) the conversion of a 3' flap to a 5' flap (reproduced from Rudolph et al., 2010b with permission).

RecG homologues are present in most bacterial species (Rocha et al., 2005; Sharples et al., 1999) but have not been identified in other organisms. However, there are a number of studies that have identified potential functional homologues in the form of helicases in human and yeast cells that are able to remodel branched DNA structures similarly to RecG (Bétous et al., 2013; Killen et al., 2012; Ralf et al., 2006; Whitby, 2010).

3' exonucleases

Nuclease enzymes are responsible for cleaving the phosphodiester bonds between nucleotide subunits of nucleic acids. Exonuclease enzymes process DNA molecules specifically from an end rather than cleaving DNA internally and can be further classified according to their preference to degrade ss- or dsDNA, and if they proceed in a 3' to 5' or 5' to 3' direction. They can be processive, catalysing many hydrolysis reactions before dissociating from the substrate molecule, or distributive, releasing a single nucleotide each time they bind their substrate (Lovett, 2011).

Exonuclease I (ExoI), encoded by the *xonA* gene (also designated *sbcB*) (Phillips and Kushner, 1987) is a highly processive exonuclease specific to ssDNA. It digests ssDNA to nucleotide monomers (Lehman and Nussbaum, 1964; Thomas and Olivera, 1978) at a rate of up to 10,000 nucleotides per minute, and dissociates when it encounters dsDNA (Lovett, 2011). ExoI catalysis operates exclusively in a 3' – 5' direction (Lehman and Nussbaum, 1964; Lovett, 2011).

Exonuclease VII (ExoVII) is composed of two non-identical subunits; a large subunit with the catalytic activity encoded by *xseA* and a smaller subunit encoded by *xseB* that is likely to regulate the activity of ExoVII (Lovett, 2011). It was proposed that the subunits formed a pentamer protein in the ratio of 1:4 XseA:XseB (Chase et al., 1986; Vales et al., 1982, 1983), although more recently Poleszak et al. (Poleszak et al., 2012) have reported results that support a heptamer structure for ExoVII with a stoichiometry of 1:6. ExoVII is a processive enzyme that initiates degradation of specifically ssDNA in both 3' – 5' and 5' – 3' directions, releasing oligonucleotides that vary in length depending on the substrate (Chase and Richardson, 1974a, 1974b). Strains lacking either *xseA* or *xseB* were found to be defective in exonuclease activity (Vales et al., 1983) and Jung et al (Jung et al., 2015) recently found that both subunits were needed for ExoVII to have enzymatic activity, which both suggest that a single gene can be deleted in order to remove ExoVII activity from cells. They also reported that expression of *xseA* in the absence of XseB resulted in cell death. ExoVII-deficient strains in the studies in this thesis were generated by deleting the *xseA* gene.

The *sbcC* and *sbcD* genes are co- transcribed to form a heterodimer protein, SbcCD. The SbcD subunit provides the nuclease activity and the SbcC subunit is an ATPase (Lovett, 2011). It has been shown that SbcCD is a processive nuclease that can act on a variety of genetic substrates and digests DNA mainly in a 3' – 5' direction and has both ssDNA endonuclease and 3' – 5' dsDNA exonuclease activity, meaning it degrades the 3' strand of the duplex (Connelly and Leach, 1996; Connelly et al., 1997, 1999). SbcCD is able to cleave hairpin structures that can form at palindromic sequences or inverted repeat sequences during DNA replication, resulting in a double strand break that can be repaired by homologous recombination (Connelly et al., 1998; Darmon et al., 2010; Eykelenboom et al., 2008; Leach et al., 1997).

The cellular function of 3' exonucleases

ExoI, ExoVII and SbcCD are all involved in aspects of DNA repair but their precise cellular function remains poorly understood and can be complicated by the fact that there often appears to be functional redundancy between the exonucleases (Lovett, 2011). ExoI has been shown to interact with SSB in order to function in DNA replication and repair (Lovett, 2011). The interaction between ExoI and SSB stimulates the catalytic activity of ExoI, possibly through the action of SSB in removing secondary structures that might otherwise hinder the processivity of ExoI, and also through stabilising ExoI substrate binding (Lu and Keck, 2008; Lu et al., 2011; Sandigursky and Franklin, 1994). As discussed above, ExoI and ExoVII have both been shown to function in the MMR system, degrading

ssDNA excised from the daughter strand behind a replication fork when a mismatch replication error is detected (Burdett et al., 2001; Marinus, 2012; Viswanathan and Lovett, 1998; Viswanathan et al., 2001). SbcCD has been implicated in a role preventing the formation of inverted chromosome duplications by cleaving hairpin structures that form at palindromic sequences (Darmon et al., 2010; Eykelenboom et al., 2008) and, in conjunction with ExoI, in processing ssDNA ends to blunt dsDNA ends that RecBCD is then able to process, thereby promoting RecBCD-mediated recombination (Lovett, 2011; Zahradka et al., 2009).

Stable DNA replication

Despite the control mechanisms in place to ensure that replication initiates specifically at *oriC* and only once per cell cycle, replication can initiate via alternative mechanisms independently of both DnaA and *oriC*. This origin-independent replication was termed stable DNA replication (SDR) because, in contrast to DnaA-dependent initiation at *oriC*, SDR does not require concomitant protein synthesis to replenish an unstable factor (Kogoma and Lark, 1970, 1975). SDR found to be activated under the same conditions that induced the SOS response was named inducible stable DNA replication (iSDR) and a distinct mechanism of SDR initiation that appears transiently as cells enter stationary phase independent of SOS induction was nSDR (Kogoma, 1997).

A third classification of SDR is observed in cells with defects in nucleic acid metabolism and was first identified in cells lacking RNase HI (Horiuchi et al., 1984; Ogawa et al., 1984). RNase HI, encoded by the *rnhA* gene, is an endonuclease that specifically degrades the RNA in a DNA:RNA hybrid such as an RNA-loop (R-loop) (Tadokoro and Kanaya, 2009). An R-loop is a structure in which ssRNA is bound to a complementary sequence on one strand of duplex DNA, thereby displacing the second DNA strand. It was shown that in the absence of RNase HI, SDR is constitutively active and so this was named constitutive stable DNA replication (cSDR) (Kogoma, 1997). cSDR in the absence of RNase HI was found to be sufficiently robust to sustain growth independently of *oriC* firing and this DNA synthesis does not require the DnaA initiator protein; indeed, *ΔrnhA* cells can tolerate the deletion of the entire *oriC* region (Dimude et al., 2015; Kogoma, 1997; Maduike et al., 2014; Rudolph et al., 2013). It was suggested by Kogoma and co-workers that the origin-independent replication seen in *ΔrnhA* cells initiates via the formation of R-loops, which are able to persist in the absence of RNase HI (Kogoma et al., 1985; Meyenburg et al., 1987; Kogoma, 1997), which suggests that RNase HI has a role in conferring specificity of replication initiation at *oriC* by denying initiation at other chromosomal locations by removing R-loops that might otherwise provide initiation substrates. SDR was later discovered in *recG* mutants, and cells lacking both RNase HI and RecG were found to be inviable (Hong et al., 1995; Rudolph et al., 2009a). Coupled with the fact that RecG was shown *in vitro* to be able to unwind the RNA from R-loops, this led to the suggestion that the SDR seen in both *ΔrecG* and *ΔrnhA* mutants arises via a common underlying mechanism (Gowrishankar, 2015; Hong et al., 1995; Kogoma, 1997). This

hypothesis is investigated further as part of the Results section of this thesis, and I revisit the suitability of the term SDR to describe origin-independent over-replication in light of my findings.

Objectives

DNA replication termination contributes to the successful duplication of an organism's genome and, like initiation and elongation, is likely to be tightly regulated and managed to ensure that all areas of the genome are copied once and only once per cell cycle. Despite the fact that the occurrence of termination events is on par with initiations, as highlighted by Dewar and Walter (2017), the question of what happens when two replication forks meet has received relatively little attention in any organism. Rudolph and colleagues have identified that in *E. coli*, RecG is a key player at replication termination and in the absence of RecG, origin-independent over-replication of the termination region is seen (Lloyd and Rudolph, 2016; Rudolph et al., 2013). Their work supports the idea that this over-replication arises as a result of pathological events at replication fork fusions, which generate intermediates that can be exploited in PriA-mediated replisome assembly (Rudolph et al., 2013). The absence of RecG is known to result in an increase in certain types of recombination (Lloyd and Rudolph, 2016; Lovett, 2006; Lovett et al., 1993) and in line with this, the origin-independent synthesis in cells lacking RecG is dependent on the function of both RecA and RecBCD (Rudolph et al., 2013). Origin-independent over-replication within the termination area has also been observed in cells lacking 3' exonucleases (Rudolph et al., 2013). These findings suggest that both RecG and 3' exonucleases have a role in managing events at termination to ensure that a single copy of the chromosome is generated at each round of replication.

Therefore, the objectives of my thesis have been:

1. Further investigating the molecular mechanism of fork fusion events and how RecG is involved in processing the resulting intermediates
2. Investigating and clarifying the role of 3' exonucleases in replication termination
3. Determining whether the increased numbers of recombination events in cells lacking RecG lead to a localised increase of the recombination frequencies in areas where replication forks fuse

Materials and Methods

Materials

All liquid media and agar used for microbial growth were autoclaved at 121°C and 103 kPa for 15 min at the time of making.

Broth and plate media

LB medium

Broth: 1% bacto tryptone (BD Biosciences), 0.5% yeast extract (BD Biosciences), 0.05% sodium chloride (Fisher), 0.002 M sodium hydroxide (Fisher), pH~7.0. To make agar, the LB broth was distributed into 200 ml and 300 ml aliquots and 1.5% agar (Sigma Aldrich) was added (3 g and 4.5 g respectively).

Mu medium

Broth: 1% bacto tryptone (BD Biosciences), 0.5% yeast extract (BD Biosciences), 1% sodium chloride (Fisher), 0.002 M sodium hydroxide (Fisher), pH~7.0. To make agar, the Mu broth was distributed into 200 ml and 300 ml aliquots and 1% agar (Sigma Aldrich) was added (2 g and 3 g respectively).

56 salts medium

74 mM potassium phosphate monobasic (KH_2PO_4), 120 mM sodium phosphate dibasic (Na_2HPO_4), 1.7 mM magnesium sulphate ($MgSO_4$), 30.5 mM ammonium sulphate ($(NH_4)_2SO_4$), 0.085 mM calcium nitrate tetrahydrate ($Ca(NO_3)_2 \cdot 4\ H_2O$), 0.0036 mM iron(III) sulphate heptahydrate ($FeSO_4 \cdot 7\ H_2O$) (all chemicals from Sigma Aldrich). 100 × and 1000 × stocks were made of calcium nitrate tetrahydrate and iron(III) sulphate heptahydrate respectively and appropriate volumes of each were used in making the 56 salts medium.

56/2: 56 salts were diluted 2-fold with sterile deionized water. For 56/2 growth medium, an appropriate carbon source (arabinose or glucose) was added and, if needed, casamino acids solution was added (final concentration 0.1%).

56/2 agar: 200 ml of 56 salts was mixed with 200 ml of 3% water agar and the required supplements were added.

M9 minimal medium

Difco™ M9 Minimal Salts base (BD Biosciences) was used in the preparation of M9 minimal media. The powder makes a 5 × M9 minimal salts solution, which was distributed in to 50 ml aliquots and autoclaved. One 50 ml aliquot was diluted 5-fold, re-distributed in to 50 ml aliquots and autoclaved.
M9 agar: A 50 ml 5 × M9 minimal salts aliquot was added to 200 ml molten 1.8% water agar (3.6 g agar). Final concentrations of 2 mM magnesium sulphate ($MgSO_4$), 0.1 mM calcium chloride ($CaCl_2$) (Sigma Aldrich), 0.4% glucose and 0.05% casamino acids were aseptically added to the M9 agar, ensuring that the bottle remained at 50°C to prevent it setting prematurely.
M9 minimal medium: The 1 × M9 minimal salts was used for serial dilutions of bacterial cultures and bacteriophage P1 lysate cultures. 2 mM magnesium sulphate and 0.1 mM calcium chloride were aseptically added to a 50 ml aliquot of 1 × M9 salts.

Antibiotics and other supplements

Antibiotics and other supplements were made in deionized water*. Antibiotics (with the exception of trimethoprim) and other supplements were sterile filtered at the time of making using single use Ministart NML syringe filters (0.45 µm, Sartorius Stedim Biotech GmbH) and added to the media at the time of use.

Table 1: Antibiotics used

	Stock concentration	Storage of stock	Final concentration
Ampicillin (Melford)	5 mg/ml	4°C	50 µg/ml
Apramycin (Sigma Aldrich)	4 mg/ml	–20°C	40 µg/ml
Chloramphenicol (Sigma Aldrich)	1 mg/ml	4°C	10 µg/ml
Kanamycin (Melford)	4 mg/ml	4°C	10 - 40 µg/ml
Rifampicin (Sigma Aldrich)	10 mg/ml	–20°C	5 – 15 µg/ml
Tetracycline (Sigma Aldrich)	1 mg/ml	–20°C	10 µg/ml
Trimethoprim (Sigma Aldrich)	50 mg/ml	4°C	10 µg/ml
Arabinose (Sigma Aldrich)	20% w/v	Room temp	0.05 – 0.2%
Glucose (Sigma Aldrich)	20% w/v	Room temp	0.2 – 0.32%
IPTG (Melford)	0.1 M	– 20°C	0.15 mM
X-Gal (Melford)	20 mg/ml	– 20°C	66.67 µg/ml

*Rifampicin was made in ethanol. Trimethoprim and X-Gal were made in DMSO.

Buffers

Elution buffer

10 mM Tris base (Melford), 0.1 mM EDTA (VWR) pH 8.0

Extraction of chromosomal DNA for deep sequencing and subsequent marker frequency analysis (page 59) was done using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). The deep sequencing performed by Earlham Institute required a lower concentration of EDTA in the sample than was achieved using the elution buffer supplied with the GenElute kit. The following was used instead. 1 ml of Tris base and 20 µl of 0.5 M EDTA were mixed and the solution was made up to 100 ml with deionized water.

MC buffer

100 mM Magnesium sulphate ($MgSO_4$); 5 mM Calcium chloride ($CaCl_2$)

For 50 ml: A 500 mM $CaCl_2$ stock solution was made by dissolving 5.549 g $CaCl_2$ in 100 ml deionized water. 0.6 g $MgSO_4$ was weighed and transferred to a measuring cylinder, which was then topped up to ~ 40 ml with deionized water. 500 µl of 500 mM $CaCl_2$ stock was added and the contents of the measuring cylinder were mixed by covering the opening with parafilm and carefully inverting the measuring cylinder until the contents were dissolved. The volume was topped up to 50 ml with deionized water, mixed once again and then sterile filtered in to ~ 5 ml aliquots.

TAE buffer

2 M Tris base, 1 M acetic acid, 0.05 M EDTA pH 8

For 1 litre 50 × TAE stock: 242 g Tris base, 57.1 ml acetic acid, 100 ml 0.5 M EDTA pH 8

For use: 1 litre of 1× TAE was prepared by diluting 20 ml 50 × TAE in 980 ml water.

TBE buffer

0.89 M Tris base, 0.89 M boric acid (Sigma Aldrich), 0.02 M EDTA

For 1 litre 10 × TBE stock: 108 g Tris base and 55 g boric acid were dissolved in approximately 900 ml of deionized water. 40 ml of 0.5 M EDTA (pH 8.0) was added and the solution was adjusted to a final volume of 1000 ml.

For use: 1 litre of 1 × TBE was prepared by diluting 100 ml 10 × TBE in 900 ml water.

TEE buffer

10 mM Tris base • HCl, 10 mM EGTA, 100 mM EDTA, pH 8.0

For 1 litre: 1.21 g Tris base, 3.80 g EGTA and 37.22 g EDTA were dissolved in approximately 900 ml of deionized water. The solution was then adjusted to a final volume of 1000 ml.

Strains and plasmids

Table 2: *Escherichia coli* K-12 strains

Strain number	Relevant Genotype[a]	Source
General P1 donors		
DL729	*ΔsbcCD::kan recD1009 supE supF*	David Leach
RRL36	*dnaQ-ypet-kan-frt*	David Sherratt
RUC663	*tnaA::*Tn*10 dnaA46*	Tove Atlung
STL2694	*xonAΔ300::cat thr-1 leuB6 proA2 supE44 kdg51 rfbD1 araC14 lacY1 galK2 xyl-5 mtl-1tsx-33 rpsL31 rac*$^-$	Susan Lovett
WX296	*oriZ-<cat>*	(Wang et al., 2011)
MG1655 derivatives		
MG1655	F$^-$ *rph-1*	(Bachmann, 1996)
AM1666	*ΔrecA::apra*	(Mahdi et al., 2006)
AM1775	*Δtus::cat*	(Rudolph et al., 2013)
AM1874	*ΔxseA::dhfr*	(Rudolph et al., 2010a)
APS345	*att*Tn*7::lacO240-kan zdd/e::tetO240-gen*	(Rudolph et al., 2007b)
AS1059	*dnaQ-YPet-<kan>*	MG1655 × P1.RRL36 to Km^r
AS1103	*ΔlacIZYA ΔsbcCD::spc ΔxseA::dhfr ΔxonA::apra* pAM401 pAST116	N7684 × pAST116 to Km^r
AU1015	*ΔlacIZYA<> ΔrecG::apra*	Plasmid-free derivative of JJ1119
AU1054	*dnaA46 tnaA::*Tn*10*	(Rudolph et al., 2007b)
AU1066	*ΔlacIZYA tnaA::*Tn*10 dnaA46 rnhA::cat*	(Stockum et al., 2012)
AU1091	*tnaA::*Tn*10 dnaA46 ΔrecG263::kan*	(Stockum et al., 2012)
JD1104	*ΔlacIZYA srgA1 argE86::*Tn*10*	JJ1264 × P1.RCe300 to Tc^r
JD1107	*ΔlacIZYA srgA1 rpoB*35*	JD1104 × P1.RCe395 to Tc^s Arg$^+$
JD1152	*priA300 rpoB*35 dnaA46 tnaA::*Tn*10*	N5535 × P1.RUC663 to Tc^r
JD1153	*ΔlacIZYA srgA1 rpoB*35 dnaA46 tnaA::*Tn*10*	JD1107 × P1.RUC663 to Tc^r
JD1350	*ΔlacIZYA oriZ-<cat> ΔxonA::apra ΔoriC::kan* pAM488	SLM1210 × P1.RCe576 to Km^r
JD1351	*ΔlacIZYA oriZ-<cat> ΔxseA::dhfr ΔxonA::apra ΔoriC::kan* pAM488	SLM1215 × P1.RCe576 to Km^r
JJ1119	*ΔlacIZYA<> ΔrecG::apra* pJJ100	(Zhang et al., 2010)
JJ1261	*ΔlacIZYA metB1*	JJ1257 × P1.N4441 Arg$^+$ Met$^-$
JJ1264	*ΔlacIZYA srgA1*	JJ1261 × P1.N3695 to Met$^+$
JJ1359	*ΔlacIZYA dam1::kan ΔrecG::apra tus1::dhfr*	(Rudolph et al., 2013)
N3695	*ΔrecG263::kan srgA1*	(Al-Deib et al., 1996)
N4560	*ΔrecG265::cat*	(Meddows et al., 2004)

N4704	*rnhA::cat*	(Rudolph et al., 2013)
N4837	*argE::*Tn*10*	(Mahdi et al., 2006)
N4849	*rpoB*35*	(Mahdi et al., 2006)
N4934	*recJ284::*Tn*10*	(Rudolph et al., 2008)
N5286	*xonAΔ300::cat*	MG1655 × P1.STL2694 to Cm^r
N5296	*xonAΔ300::cat ΔsbcCD::kan*	N5286 × P1.DL729 to Km^r
N5535	*priA300 rpoB*35*	(Mahdi et al., 2006)
N6576	*ΔlacIZYA ΔrecG::apra*	(Zhang et al., 2010)
N6796	*tus1::dhfr*	(Stockum et al., 2012)
N6798	*ΔrecG265::cat tus1::dhfr*	N4560 × P1.JJ1359 to Tm^r
N6953	*ΔxonA::apra ΔxseA::dhfr ΔsbcCD::kan*	(Rudolph et al., 2010a)
N7684	*ΔlacIZYA ΔsbcCD::spc ΔxseA::dhfr ΔxonA::apra* pAM401	(Rudolph et al., 2010a)
RCe203	*tnaA::*Tn*10 dnaA46 Δtus::kan*	(Rudolph et al., 2013)
RCe262	*rpoB*35 tnaA::Tn10 dnaA46*	N4849 × P1.RUC663 to Tc^r
RCe267	*rpoB*35 Δtus::cat dnaA46 tnaA::*Tn*10*	(Rudolph et al., 2013)
RCe268	*rpoB*35 ΔrecG::apra Δtus::cat tnaA::*Tn*10 dnaA46*	(Rudolph et al., 2013)
RCe300	*att*Tn*7::lacO240-kan zdd/e::tetO240-gen argE::*Tn*10*	APS345 × P1.N4837 to Tc^r Arg$^-$
RCe303	*rpoB*35 tnaA::*Tn*10 dnaA46 rnhA::cat*	RCe262 × P1.N4704 to Cm^r
RCe309	*rpoB*35 tnaA::*Tn*10 dnaA46 rnhA::cat tus1::dhfr*	RCe303 × P1.N6796 to Tm^r
RCe326	*rpoB*35 ΔrecG::apra Δtus::cat dnaA46 tnaA::*Tn*10* pDIM104	RCe268 × pDIM104 to Ap^r
RCe395	*rpoB*35 dnaA46 tnaA::*Tn*10 ΔrnhA::cat tus1::dhfr ΔoriC::kan*	(Rudolph et al., 2013)
RCe427	*tos-kan*	(Rudolph et al., 2013)
RCe504	*oriZ-<cat>*	(Ivanova et al., 2015)
RCe508	*oriZ-<cat> ΔrecG::apra*	RCe504 × P1.AU1015 to Apra^r
RCe528	*rpoB*35 Δtus::cat dnaA46 tnaA::*Tn*10 ΔxonA::apra*	RCe267 × P1.AS1103 to Apra^r
RCe544	*ΔlacIZYA oriZ-<cat>*	TB28 × P1.WX296 to Cm^r
RCe552	*ΔlacIZYA tna::*Tn*10 dnaA46 rnhA::cat* pECR22	AU1066 × pECR22 to Ap^r
RCe553	*rpoB*35 Δtus::cat dnaA46 tnaA::*Tn*10 ΔsbcCD::kan*	RCe267 × P1.N5296 to Km^r
RCe554	*rpoB*35 Δtus::cat dnaA46 tnaA::*Tn*10 ΔsbcCD::kan ΔxonA::apra*	RCe553 × P1.AS1103 to Apra^r
RCe557	*ΔlacIZYA tna::*Tn*10 dnaA46 rnhA::cat* pLAU17	AU1066 × pLAU17 to Ap^r
RCe562	*ΔsbcCD::kan*	MG1655 × P1.N5296 to Km^r
RCe563	*ΔxonA::apra*	MG1655 × P1.AS1103 to Apra^r
RCe569	*ΔxonA::apra ΔsbcCD::kan*	RCe563 × P1.N5296 to Km^r

RCe576	*rpoB*35 oriZ-<cat> tus1::dhfr ΔoriC::kan*	(Ivanova et al., 2015)
RCe714	*ter4.44<> oriZ-<cat> ter4.57-<kan> ΔrecG::apra*	SLM1197 × P1.AU1015 to Apra^r
RCe745	*ter4.44<> oriZ-<cat> ter4.57-<kan> tus1::dhfr*	SLM1197 × P1.N6798 to Tm^r
RCe760	*ter4.44<> oriZ-<cat> ter4.57-<kan> tus1::dhfr ΔrecG::apra*	RCe745 × P1.AU1015 to Apra^r
SLM1008	*rpoB*35 ΔrecG::apra Δtus::cat tnaA::*Tn*10 dnaA46* pECR22	RCe268 × pECR22 to Ap^r
SLM1010	*rpoB*35 ΔrecG::apra Δtus::cat tnaA::*Tn*10 dnaA46* pLau17	RCe268 × pLAU17 to Ap^r
SLM1037	*kankanMX4-<cat>-narU*	This study
SLM1039	*kankanMX4-<cat>-yhjR*	This study
SLM1042	*kankanMX4-<cat>-narU*	MG1655 × P1.SLM1037 to Cm^r
SLM1043	*kankanMX4-<cat>-yhjR*	MG1655 × P1.SLM1039 to Cm^r
SLM1048	*kankanMX4-<cat>-narU ΔrecG::apra*	SLM1042 × P1.AU1015 to Apra^r
SLM1049	*kankanMX4-<cat>-yhjR ΔrecG::apra*	SLM1043 × P1.AU1015 to Apra^r
SLM1051	*oriZ-<>*	RCe504 × pCP20 to Cm^s Ap^s
SLM1052	*oriZ-<> ΔrecG::apra*	RCe508 × pCP20 to Cm^s Ap^s
SLM1058	*oriZ-<> kankanMX4-<cat>-narU*	SLM1051 × P1.SLM1037 to Cm^r
SLM1059	*oriZ-<> ΔrecG::apra kankanMX4-<cat>-narU*	SLM1052 × P1.SLM1037 to Cm^r
SLM1060	*oriZ-<> kankanMX4-<cat>-yhjR*	SLM1051 × P1.SLM1039 to Cm^r
SLM1061	*oriZ-<> ΔrecG::apra kankanMX4-<cat>-yhjR*	SLM1052 × P1.SLM1039 to Cm^r
SLM1104	*ΔlacIZYA tnaA::*Tn*10 dnaA46 rnhA::cat* pDIM104	AU1066 × pDIM104 to Ap^r
SLM1107	*priA300 rpoB*35 dnaA46 tnaA::*Tn*10 Δtus::cat*	JD1152 × P1.AM1775 to Cm^r
SLM1108	*ΔlacIZYA srgA1 rpoB*35 dnaA46 tnaA::*Tn*10 Δtus::cat*	JD1153 × P1.AM1775 to Cm^r
SLM1109	*priA300 rpoB*35 dnaA46 tnaA::*Tn*10 Δtus::cat ΔxonA::apra*	SLM1107 × P1.AS1103 to Apra^r
SLM1110	*ΔlacIZYA srgA1 rpoB*35 dnaA46 tnaA::*Tn*10 Δtus::cat ΔxonA::apra*	SLM1108 × P1.AS1103 to Apra^r
SLM1113	*Δtus::cat ter4.44-<kan>*	This study
SLM1115	*Δtus::cat ter4.57-<kan>*	This study
SLM1125	*ΔterC-kan*	This study
SLM1134	*ΔterC-kan*	MG1655 × P1.SLM1125 to Km^r
SLM1140	*ΔterC-kan ΔrecG::apra*	SLM1134 × P1.AU1015 to Apra^r
SLM1142	*dnaQ-YPet-<>*	AS1059 × pCP20 to Km^s Ap^s
SLM1144	*ΔlacIZYA::dnaQ-mTagRFP-<kan>*	This study
SLM1146	*dnaQ-YPet-<> ΔlacIZYA::dnaQ-mTagRFP-<kan>*	SLM1142 × P1.SLM1144 to Km^r
SLM1170	*tnaA::cat*	This study

SLM1171	*ΔlacIZYA srgA1* pAM488	JJ1264 × pAM488 to Ap^r
SLM1172	*ΔlacIZYA srgA1 ΔxonA::apra*	SLM1171 × P1.RCe563 to Ap^s $Apra^r$
SLM1174	*tos-kan ΔxonA::apra*	RCe427 × P1.RCe563 to $Apra^r$
SLM1178	*ΔxonA::apra recJ284::*Tn*10*	RCe563 × P1.N4934 to Tc^r
SLM1182	*ter4.44<kan>*	MG1655 × P1.SLM1113 to Km^r
SLM1184	*rpoB*35 ΔxseA::dhfr*	N4849 × P1.AM1874 to Tm^r
SLM1185	*ΔxseA::dhfr*	MG1655 × P1.AM1874 to Tm^r
SLM1186	*ΔlacIZYA srgA1 ΔxonA::apra ΔxseA::dhfr*	SLM1172 × P1.AM1874 to Tm^r
SLM1187	*tos-kan ΔxonA::apra ΔxseA::dhfr*	SLM1174 × P1.AM1874 to Tm^r
SLM1188	*ΔxonA::apra recJ284::*Tn*10 ΔxseA::dhfr*	SLM1178 × P1.AM1874 to Tm^r
SLM1189	*dnaA46 tnaA::cat*	AU1054 × P1.SLM1170 to Cm^r
SLM1190	*rpoB*35 ΔxseA::dhfr Δtus::kan*	SLM1184 × P1.RCe203 to Km^r
SLM1191	*rpoB*35 ΔxseA::dhfr Δtus::kan ΔxonA::apra*	SLM1190 × P1.RCe563 to $Apra^r$
SLM1192	*ter4.44<>*	SLM1182 × pCP20 to Km^s (Ap^s)
SLM1193	*ter4.44<> oriZ-<cat>*	SLM1192 × P1.RCe544 to Cm^r
SLM1194	*rpoB*35 ΔxseA::dhfr Δtus::kan ΔxonA::apra dnaA46 tnaA::*Tn*10*	SLM1191 × P1.RUC663 to Tc^r
SLM1195	*rpoB*35 Δtus::kan*	N4849 × P1.RCe203 to Km^r
SLM1196	*rpoB*35 ΔxseA::dhfr Δtus::cat*	SLM1184 × P1.AM1775 to Cm^r
SLM1197	*ter4.44-<> oriZ-<cat> ter4.57-<kan>*	SLM1193 × P1.SLM1115 to Km^r
SLM1198	*priA300 rpoB*35 dnaA46 tnaA::*Tn*10 Δtus::cat ΔxonA::apra ΔxseA::dhfr*	SLM1109 × P1.AM1874 to Tm^r
SLM1199	*ΔlacIZYA srgA1 rpoB*35 dnaA46 tnaA::*Tn*10 Δtus::cat ΔxonA::apra ΔxseA::dhfr*	SLM1110 × P1.AM1874 to Tm^r
SLM1201	*rpoB*35 Δtus::kan ΔxonA::apra*	SLM1195 × P1.RCe563 to $Apra^r$
SLM1202	*rpoB*35 ΔxseA::dhfr Δtus::cat ΔxonA::apra*	SLM1196 × P1.RCe563 to $Apra^r$
SLM1203	*ΔxseA::dhfr ΔxonA::apra*	SLM1185 × P1.RCe563 to $Apra^r$
SLM1204	*ΔxseA::dhfr recJ284::*Tn*10*	SLM1185 × P1.N4934 to Tc^r
SLM1205	*ter4.44-<> oriZ-<cat> ter4.57-<kan> ΔrecG::apra*	SLM1197 × P1.AU1015 to $Apra^r$
SLM1206	*ΔlacIZYA oriZ-<cat> ΔxonA::apra*	RCe544 × P1.RCe563 to $Apra^r$
SLM1208	*ΔlacIZYA oriZ-<cat> ΔxseA::dhfr*	RCe544 × P1.AM1874 to Tm^r
SLM1209	*ΔxseA::dhfr ΔsbcCD::kan*	SLM1185 × P1.RCe562 to Km^r
SLM1210	*ΔlacIZYA oriZ-<cat> ΔxonA::apra* pAM488	SLM1206 × pAM488 to Ap^r
SLM1211	*ΔlacIZYA oriZ-<cat> ΔxseA::dhfr* pAM488	SLM1208 × pAM488 to Ap^r
SLM1212	*tos-kan ΔxonA::apra ΔxseA::dhfr* N15 lysogen	SLM1187 × N15 to $N15^r$
SLM1213	*ΔxseA::dhfr ΔxonA::apra* N15 lysogen	SLM1203 × N15 to $N15^r$
SLM1215	*ΔlacIZYA oriZ-<cat> ΔxseA::dhfr ΔxonA::apra* pAM488	SLM1211 × P1.RCe563 to $Apra^r$ Ap^r
SLM1217	*ΔlacIZYA oriZ-<cat> ΔxseA::dhfr ΔxonA::apra*	Plasmid-free derivative of SLM1215

SLM1218	*rpoB*35 Δtus::kan ΔxonA::apra dnaA46 tnaA::cat*	SLM1201 × P1.SLM1189 to Cm^r
SLM1219	*rpoB*35 ΔxseA::dhfr Δtus::cat dnaA46 tnaA::*Tn*10*	SLM1196 × P1.RUC663 to Tc^r
SLM1220	*rpoB*35 ΔxseA::dhfr Δtus::cat ΔxonA::apra tos-kan*	SLM1202 × P1.RCe427 to Km^r
SLM1221	*rpoB*35 ΔxseA::dhfr Δtus::cat ΔxonA::apra dnaA46 tnaA::*Tn*10*	SLM1202 × P1.RUC663 to Tc^r
SLM1222	*rpoB*35 Δtus::kan dnaA46 tnaA::cat*	SLM1195 × P1.SLM1189 to Cm^r
SLM1223	*rpoB*35 ΔxseA::dhfr Δtus::cat dnaA46 tnaA::*Tn*10 ΔsbcCD::kan*	SLM1219 × P1.RCe562 to Km^r
SLM1224	*rpoB*35 Δtus::kan ΔxonA::apra dnaA46 tnaA::cat recJ284::*Tn*10*	SLM1218 × P1.N4934 to Tc^r
SLM1225	*rpoB*35 ΔxseA::dhfr Δtus::cat ΔxonA::apra tos-kan dnaA46 tnaA::*Tn*10*	SLM1220 × P1.RUC663 to Tc^r
SLM1226	*rpoB*35 ΔxseA::dhfr Δtus::cat dnaA46 tnaA::*Tn*10 ΔsbcCD::kan ΔxonA::apra*	SLM1223 × P1.RCe563 to Apra^r
SLM1230	*rpoB*35 ΔxseA::dhfr Δtus::cat ΔxonA::apra tos-kan dnaA46 tnaA::*Tn*10* N15 lysogen	SLM1225 × N15 to N15^r
SLM1232	*rpoB*35 ΔxseA::dhfr Δtus::cat ΔxonA::apra dnaA46 tnaA::*Tn*10* N15 lysogen	SLM1221 × N15 to N15^r
SLM1233	*rpoB*35 Δtus::kan dnaA46 tnaA::cat recJ284::*Tn*10*	SLM1222 × P1.N4934 to Tc^r
SLM1236	*rpoB*35 ΔxonA::apra*	N4849 × P1.RCe563 to Apra^r
SLM1238	*rpoB*35 ΔxonA::apra ΔxseA::dhfr*	SLM1236 × P1.AM1874 to Tm^r
SLM1242	*ΔxseA::dhfr ΔxonA::apra Δtus::cat*	SLM1203 × P1.AM1775 to Cm^r
SLM1244	*ΔxseA::dhfr ΔxonA::apra Δtus::cat dnaA46 tnaA::*Tn*10*	SLM1242 × P1.RUC663 to Tc^r
SLM1245	*rpoB*35 ΔxonA::apra ΔxseA::dhfr dnaA46 tnaA::*Tn*10*	SLM1238 × P1.RUC663 to Tc^r
SLM1246	*ΔxseA::dhfr ΔxonA::apra dnaA46 tnaA::*Tn*10*	SLM1203 × P1.RUC663 to Tc^r
TB28	*ΔlacIZYA*	(Bernhardt and de Boer, 2003)

a – Only the relevant additional genotype of the derivatives is shown. The abbreviations *apra*, *kan*, *cat* and *dhfr* refer to insertions conferring resistance to apramycin (Apra^r), kanamycin (Km^r), chloramphenicol (Cm^r) and trimethoprim (Tm^r).

Table 3: Plasmids used

Plasmid	Description	Reference
pAM401	pRC7 *recG*$^{+}$. The entire *sbcCD* gene, including the native promoter area, was amplified with primers introducing *Apa*I sites at either end of the gene and cloned in to the *Apa*I site of pRC7.	(Rudolph et al., 2010a)
pAM488	pRC7 *xonA*$^{+}$. Multiple cloning site of pGEM backbone was cloned into the *Apa*I site of pRC7. This allowed cloning of the entire *xonA* gene, including its promoter region, into the *Bam*HI and *Hin*dIII sites of the integrated MCS.	A.A. Mahdi and R.G. Lloyd, unpublished
pAST116	Human mitochondrial FEN1 gene amplified from cDNA with primers introducing *Nco*I and *Xba*I sites at the ends was cloned in to pLau17 via *Nco*I and *Xba*I, thereby releasing the eCFP gene. The amplification has a point mutation at position 2017 (A->C) and encodes a catalytically inactive D223A version of Fen1.	AS and CJR, unpublished
pCP20	*oriR101(ts)*, gene for FLP recombinase from *Saccharomyces cerevisiae* is under control of phage λ repressor. A temperature-sensitive variant allows FLP expression at high temperature.	(Cherepanov and Wackernagel, 1995)
pDIM104	*recG* expression plasmid. *Escherichia coli recG* expressed via p*ara* promoter. pBAD24 plasmid backbone.	(Rudolph et al., 2009a)
pDIM141	pBAD24 *mRFP1-<kan>* for generating protein fusions with RFP. The kanamycin resistance cassette is encoded within the *Hin*dIII fragment of this vector.	T. Moore and R.G. Lloyd, unpublished.
pECR22	pBAD24 Sc*RNH1* for expression of Saccharomyces cerevisiae RNase H1 in *E. coli*.	(Dimude et al., 2015)

pJJ100	pRC7 *recG*$^{+}$. The entire *recG* gene, including the native promoter area, was amplified with primers introducing *Apa*I sites at either end of the gene and cloned into the *Apa*I site of pRC7.	(Zhang et al., 2010)
pKD46	*oriR101*, Red recombinase plasmid	(Datsenko and Wanner, 2000)
pLau17	pBAD24 eCFP for generating protein fusions with eCFP.	(Lau et al., 2003)
pRS316-KanKanMX4	colE1, *ARS6*, shuttle plasmid for yeast and *E. coli*, containing the *KanKanMX4* fragment cloned via *Eco*RI and *Sal*I from pFA6a-KanKanMX4.	(Ede et al., 2011)
pSLM001	pRS316-KanKanMX4 with added <*cat*> fragment via PCR amplification of the <*cat*> cassette, adding *Bgl*II at both ends. Cloning into pRS316-KanKanMX4 via *Bgl*II.	This study

Methods

The optical density of cultures was measured using a Jenway 7300 spectrophotometer. Cultures were grown in glass tubes with an internal diameter of 1.5 cm and the spectrophotometer was set up to accommodate these tubes instead of a 1 cm cuvette. A culture of wild type *E. coli* cells grown to an optical density of A_{600} = 0.5 in a glass tube equates to A_{600} = 0.42 in a 1 cm cuvette, which equates to 6 × 10^7 cells/ml.

Genetic and genomic strain manipulation

Mutational analysis of genes is an important approach in gaining insight in to their function. Gene disruptions were introduced to the chromosome using λ Red-mediated recombination and transduction using P1*vir*, and further genomic manipulation was achieved via cloning and transformation of strains with plasmids by electroporation.

Recombineering

To introduce a gene disruption, Datsenko and Wanner's one step method of gene inactivation is used (Datsenko and Wanner, 2000) where a gene for an antibiotic resistance determinant is recombined directly in to the chromosome in the place of the target gene. Primers with homology to regions either side of the target gene are used in the PCR amplification of a template plasmid containing the antibiotic resistance gene. This marker gene allows for selection of cells that have successfully lost the gene of interest, and it can subsequently be lost itself if necessary. This method was first developed in *Saccharomyces cerevisiae* (Baudin et al., 1993). Datsenko and Wanner (Datsenko and Wanner, 2000) subsequently adapted this method for use in *E. coli*, which was necessary because the native exonuclease RecBCD hydrolyses linear DNA (Goldmark and Linn, 1972) and so would target the PCR-amplified DNA fragment. To overcome this, the lambda red recombination system from the lambda red bacteriophage is used. The Red system gene products, Gam (*γ*), Bet (*β*) and Exo (*exo*), block RecBCD activity and promote recombination between the PCR product and the chromosome (Murphy, 1991, 1998).

A fresh overnight of a wild type strain carrying pKD46, a λ Red recombinase expression plasmid, was used to inoculate 11 ml LB. The LB was supplemented with ampicillin to select for cells carrying the plasmid, which contains a gene for ampicillin resistance, and with arabinose (0.2% final concentration) in order to induce expression of the lambda red genes. The culture was incubated in a shaking water bath at 30°C as pKD46 has a temperature sensitive replicon that is maintained at 30°C and lost at 37°C. The culture was grown to an optical density of A_{600} = 0.6. The cells were pelleted (5000 rpm, 4°C, 5 mins, Eppendorf 5804 R) and then washed four times in ice cold 10 % glycerol solution (see page 56 for a detailed description of the electroporation procedure), ensuring that the cells and reagents were kept on ice throughout. 100 – 200 ng of the PCR-amplified DNA was added to

the resuspended cells and incubated on ice for ~ 10 mins before being transferred to a cuvette for transformation by electroporation. Immediately after the shock was delivered, 2 × 800 µl of enriched LB was added to the cuvette and the cells were then allowed to recover in a tube rotator (8 rpm, Stuart rotator SB3) at 37 °C for 60 mins. 800 µl of the culture was spun down (Eppendorf Mini Spin or Stuart Microfuge, 3 mins, 13400 rpm) and the cell pellet was resuspended in ~ 100 µl of the supernatant. This was spread on an agar plate containing the appropriate antibiotic to select for recombinant cells and incubated overnight at 37°C; the pKD46 plasmid should be lost at 37°C. The remaining half of the culture was left on the bench overnight. If no transformants grew on the first plate within 24 hrs, it was spun down and plated in the same manner. Recombinant colonies were purified once non-selectively and tested by a streak test (page 52) and a verification PCR. P1 liquid culture lysates of the correct recombinant strains were prepared and used to transfer the mutant construct to a clean wild type strain by P1*vir* transduction (see below) for further use. This final step is important because in the recombineering strains, it is possible that recombination events might have occurred in addition to those that were planned and so it is cleanest to transfer to a clean wild type background.

Removing the selectable marker

The primers used to PCR-amplify the gene for an antibiotic resistance determinant contain the 65-nt *FRT* site (FLP-recombination target) that is recognised by Flp recombinase from *Saccharomyces cerevisiae*. The gene for the recombinase, *FLP*, is present on pCP20 under thermal induction. The plasmid contains genes for ampicillin and chloramphenicol resistance and has temperature sensitive replication; it is maintained at 30°C and lost at 42°C (Cherepanov and Wackernagel, 1995).

To remove the selectable antibiotic resistance marker from a recombinant strain, the strain was transformed (page 56) with pCP20. Following electroporation, the cells were allowed to recover at 30°C for 60 mins. 50 µl were spread on an ampicillin plate and incubated at 30°C overnight. Routinely, four colonies were picked from the ampicillin plate and purified once non-selectively overnight at 42°C, which simultaneously induces expression of *FLP* and represses plasmid replication. A streak test was carried out to confirm loss of both the FRT-flanked resistance gene and pCP20.

Transduction

Bacteriophage P1 – liquid culture lysate

If the deletion or mutation of a particular gene has been done previously, the strain carrying this mutation or deletion was used as a donor.

The *E. coli* phage P1*vir* was used to introduce specific genetic deletions and point mutations via transduction. A stock of bacteriophage grown on wild type cells (P1.MG1655) is kept available in the lab and this is used to generate any new strains of bacteriophage needed.

To prepare a new liquid culture lysate, an overnight culture of the donor strain was used to inoculate 11 ml Mu broth to an optical density of A_{600} = 0.05 and the culture was incubated in a shaking water bath at either 30°C or 37 °C until it reached an optical density of A_{600} = 0.3. The phage require calcium ions to be available in order to infect cells, so 200 µl of calcium chloride ($CaCl_2$) solution (0.5 M) were added and the culture was returned to the water bath for 10 min. 30 µl of the stock wild type P1*vir* was used to inoculate the culture and it was incubated at 30 °C/37 °C until lysis was complete, normally 3 – 4 hours. It is important to remove any uninfected and non-lysed cells as they would contaminate downstream applications so 0.5 ml chloroform was added and the culture vortexed after lysis to kill any remaining cells. 10 ml of the lysate was spun in 2 ml aliquots (Eppendorf Mini Spin or Stuart Microfuge, 8 min, 13400 rpm) to remove the bacterial cell components as well as unlysed cells and the supernatant was transferred to a glass tube and a further 0.5 ml of chloroform was added. The P1*vir* lysates were stored at 4 °C and can be used for several years.

P1vir Transduction

A fresh overnight culture of the recipient strain was used to inoculate 11 ml Mu and the culture was grown to an optical density of A_{600} = 0.8 or higher in a shaking water bath at either 30°C or 37°C. The cells were pelleted and re-suspended in 1 ml MC buffer (page 42), vortexed briefly and inoculated at room temperature for 10 min. 50 µl and 200µl aliquots of the appropriate P1 lysate were mixed with 200 µl aliquots of the recipient strain and incubated in a water bath at 37 °C for 30 min. Control tubes containing the recipient strain only and P1 only were also set up and carried through for the rest of the protocol. Sodium citrate ($C_6H_7NaO_7$ 1 M, pH 6.8) was added to all tubes to prevent further infections of recipient cells by the phage after 30 min. The cultures were mixed with 3 ml molten 0.6% Mu top agar kept at 42°C, or 3 ml 0.6% molten water top agar when selecting for trimethoprim resistance, poured on to the relevant antibiotic plates and incubated at either 30 °C or 37 °C until there were visible transductant colonies (16 – 72 hrs). After primary selection, transductant colonies were purified on medium without an antibiotic. Streak tests (see below) were carried out to determine the antibiotic resistance profile and preliminary phenotype of the transductants compared with that of the original strain. One transductant culture showing the correct resistance profile was selected and a sample was frozen for further use.

Streak test

Four single colonies were picked from the non-selective colony purification plate for testing following a genetic manipulation procedure. The colonies were used to inoculate 5 ml LB broth and the cultures were grown overnight at 30°C or 37°C. 10 µl of each culture was plated on a selection of agar plates; routinely, two LB, ampicillin, apramycin, chloramphenicol, kanamycin, tetracycline and trimethoprim. One of the LB plates was exposed to 60 J/m^2 UV for 60 s. Additional plates were sometimes used, such as LB incubated at 42°C when a temperature sensitive allele was involved, or plates containing

Mitomycin C. The original parent strain was plated alongside the recombinants/transductants/transformants to allow comparison of the resistance pattern of each. The plates were incubated overnight at 30°C or 37°C.

When a *dnaA46*(ts) allele encoding a temperature sensitive DnaA initiator protein was introduced to strains that will grow in the absence of origin firing, the LB plate incubated at 42 °C would not show a definitive difference between the parent strain and the transductants. On these occasions, functional testing for the presence of the mutant *dnaA46* allele was carried out. A P1 liquid culture lysate of a transductant strain was prepared and used to transduce the *dnaA46* allele in to wild type cells to confirm that there were temperature sensitive transductants generated, which confirms the presence of the mutant allele in the strain used to generate the P1.

Freezing a strain

For long term storage of a bacterial strain, 900 µl of a fresh overnight culture was mixed with 900 µl of 80% glycerol solution in a cryotube (Nunc) and stored at –20°C. Strains were frozen in duplicate and stored in two separate freezers.

Cloning

Preparation of plasmid DNA from E. coli

DH5α is a general *E. coli* cloning strain which is designed to have high transformation efficiency, high plasmid yield from minipreps, increased DNA insert stability, and blue/white screening capability. Routinely, plasmids were stored in DH5α at –20°C instead of using wild type cells in order to minimise plasmid degradation by endogenous nucleases, which can be a problem even at –20°C. When a plasmid was needed, a fresh culture of the DH5α strain containing the plasmid was inoculated in 5 ml LB with ampicillin present to select for cells carrying the plasmid, which contains a gene for ampicillin resistance. The culture was grown overnight at 30 °C if the plasmid had a temperature sensitive replicon and at 37°C if not. The entire 5 ml culture was spun down to pellet the cells and the NucleoSpin® Plasmid miniprep kit (Macherey-Nagel) was used to extract plasmid DNA according to the manufacturer's instructions.

Briefly, the cells were lysed in an SDS/alkaline buffer. The high pH conditions denature the DNA content of the cells. Once neutralised, the small, supercoiled plasmid DNA strands were able to reanneal. Due to the size of the chromosomal DNA, the denatured strands were unable to accurately reanneal and instead formed very large, insoluble conglomerates along with proteins and other cell components, which were removed by centrifugation. The supernatant containing plasmid DNA was loaded on to a Nucleospin® Plasmid column where the DNA bound to the silica membrane. After washing with an ethanolic buffer, the plasmid DNA was eluted and measured using a Biodrop micro-volume spectrophotometer.

PCR

DNA fragments for cloning and gene disruption via recombineering were amplified from a template using the polymerase chain reaction (PCR). PCR was also used in a verification capacity to confirm changes to the genome following genetic or genomic manipulation. MyTaq™ DNA polymerase (Bioline) was used for routine amplifications. When high-fidelity amplification was needed, VELOCITY DNA polymerase (Bioline) was used. Velocity possesses a proofreading capability that results in a low error rate and high PCR fidelity. When a colony was used as the template DNA, a Hot Start PCR was carried out instead (see below).

All PCR reactions were set up on ice. Primers were manufactured by Eurofins Genomics and were supplied at 100 mM concentration. Primers were used at 10 mM concentration so working stock solutions were generated by a 10 dilution in sterile distilled water. MyTaq™ is supplied with an optimised buffer system that dNTPs and $MgCl_2$ and so it was not necessary to add these components to the PCR reaction separately. Both MyTaq™ and VELOCITY were used according to the manufacturer's recommendations.

Hot Start PCR

Hot-start PCRs were used when cells from a colony or a culture provided the template in the form of chromosomal DNA. Taq DNA Polymerase (NEB) was used for this type of PCR. To prepare the template DNA, a pipette tip was dipped in to a colony and the cells were then resuspended in 10 µl sterile distilled water. 1 µl of this dilution was used as the PCR template. When using cells in broth culture, 1 µl of a 1:10 dilution of an overnight culture was used as the PCR template.

A PCR tube containing appropriate volumes of buffer, colony suspension and sterile distilled water to a total volume of 30 µl was loaded in to the thermocycler. The other components were set up in 20 µl a separate tube using appropriate volumes for 50 µl total reaction volume. The PCR programme ran an initial denaturation of 98° for 5 min. This breaks open the cells and releases the DNA in to the reaction mixture. The reaction was then cooled to 85°C and the mixture containing the remaining components was added. The remainder of the programme was run according to the polymerase manufacturer's instructions.

Gel electrophoresis

In order to visualise the products following PCR amplification, DNA fragments were generally separated on a 1 % agarose gel (SeaKem® LE Agarose, Lonza) made with 1× TBE buffer. DNA fragments were loaded in to the wells after being mixed with 5 × loading dye. At least 100 ng of DNA is required for each well in a maximum volume of 20 µl. 10 µl of 2-log DNA ladder (NEB) was run on every gel for size confirmation. The gel was run at 100 V in a 15 cm chamber, equating to a gradient voltage of ~ 6 V/cm. To visualise the DNA, the gel was stained with 250 ml of 2 x GelRed® solution (Biotium) for 15 – 30 min (freshly made Gel Red solution stained the DNA more quickly) and visualised

using the Gel Doc Molecular Imager (Bio-Rad). To run a quick gel, 0.5 × TBE buffer was used both for gel and running buffer. These gels could be run at a higher voltage of 140V.

1 × TAE gels were also used. Gels made with TBE can be run at higher voltage for less time than TAE gels, as TAE has a lower buffer capacity than TBE, but the boric acid interacts with downstream applications of the DNA so if the DNA was needed for further work, a TAE gel was used instead. TAE gels were run at 80 V in a 15 cm chamber, equating to a gradient voltage of ~ 5 V/cm. Instead of Gel Red, SYBR® Gold (Life Technologies) stain was used with TAE gels as it is excitable with blue light transillumination, which does not cause DNA damage. 3.5 µl of 10 x SYBR® Gold was added to each 20 µl well. The DNA was then visualised using a blue-light transilluminator and the necessary bands were cut out using a new scalpel blade for each band and purified (see below).

DNA extraction from agarose gels

The bands containing the DNA required were purified using the NucleoSpin® Gel and PCR CleanUp kit (Macherey-Nagel) according to the manufacturer's instructions. Briefly, the plugs cut out of a TAE gel were mixed with a chaotrophic salt-containing binding buffer and incubated at 50°C until the agarose gel plugs dissolved. The sample was loaded on to a NucleoSpin® Gel and PCR Clean-up Column where, in the presence of chaotrophic salt in the binding buffer, the DNA binds to the silica membrane. A number of washes were carried out with ethanolic buffer to remove any impurities and finally the DNA was released from the silica by eluting with a low ionic salt solution. The sample was measured using a Biodrop micro-volume spectrophotometer.

The cloning process

To clone a DNA fragment in to a plasmid vector, the DNA fragment was amplified by PCR using primers that contain a restriction site that is also present in the plasmid. 5 µl of each PCR reaction were run on a 0.5 ×TBE 1% agarose gel to check the PCR had been successful. The rest of the PCR products were then run on a TAE gel and the appropriate bands were cut out and purified. The plasmid and DNA fragment were digested separately with the appropriate restriction enzyme in 20 µl reactions containing DNA (PCR-amplified insert or plasmid vector), buffer, restriction enzyme and sterile distilled water according to the restriction enzyme manufacturer's instructions. The digested DNA was purified using the NucleoSpin® Gel and PCR CleanUp kit (Macherey-Nagel; see above). This is important because it removes the restriction enzymes prior to ligation and prevents the DNA being recut each time a ligation event happens.

The ligation and ligation control (vector only, no insert DNA) reactions were set up with a molar ratio of 1:3 vector:DNA fragment, T4 DNA Ligase (NEB), 2 µl 10 × T4 DNA Ligase Buffer and sterile distilled water to bring the reaction volume to 20 µl, although the reaction volume can be larger if necessary. The ligation reactions were incubated at 4°C overnight. The following day, each ligation reaction underwent microdialysis using Millipore membrane filters (0.025 µm) floated on Molecular

Biology Grade water (0.03 μm filtered, Fisher Bioreagents) for 30 mins. This removed the ions from the reaction mixture.

The ligation and vector control were used to transform DH5α (see below). After the recovery phase, 100 μl of each culture was used to inoculate a spread plate containing ampicillin. The rest of the culture was spun down and plated on ampicillin.

Transformation of bacterial strains

Preparation of competent cells

A fresh overnight culture of the recipient strain was used to inoculate 11 ml LB to an optical density of A_{600} = 0.05 and the culture was grown to an optical density of A_{600} = 0.6 in a shaking water bath. Once grown to the correct density, the cells and all reagents were kept on ice for the rest of the procedure (methodology of transformation reviewed in Aune and Aachmann, 2010).

Due to the high voltage of electroporation, it is necessary for the culture of cells to have very low conductivity to prevent arcing of the electric current. To achieve this, the cells were washed four times in order to reduce the ionic strength of the culture; the cells were pelleted and then resuspended in 10 ml ice cold 10% glycerol solution. This was repeated a further three times with decreasing volumes of glycerol solution (5 ml, 1 ml and 0.5 ml). After the final wash step, the cell pellet was resuspended in the drop that remained after discarding the final 0.5 ml of glycerol solution.

Cell transformation via electroporation

After washing, the plasmid was added to the cells and the mixture was incubated on ice for 10 min before being transferred to a chilled electroporation cuvette. Immediately after electroporation at 1.75 kV for at least 4 ms (Eppendorf Eporator), the cells were mixed with 1.6 ml SOC and incubated at 30°C or 37°C for one hour in a tube rotator (Stuart Rotator SB3, 8 rpm). This is to allow cells to recover from the electroporation and to have time to synthesise the antibiotic resistance determinant conferred by the plasmid before being incubated with this antibiotic. Following the recovery stage, 50 μl of the transformation culture was plated on to the relevant antibiotic plates for selection of transformants and streaked to single colonies. The remaining transformation culture was left on the bench overnight. When cells were transformed with a DNA fragment, 100 μl of the culture was used to inoculate a spread plate containing the appropriate antibiotic. The remaining culture was spun down to pellet the cells, which were then re-suspended in ~ 150 μl of the supernatant and used to inoculate a second spread plate. The plates were incubated at 30°C or 37°C for at least 24 hours.

Chromosome Linearisation

Linearisation of the E. coli chromosome

The autonomy of the linearization system of the *E. coli* bacteriophage N15 was exploited by Cui and colleagues (2007) in the development of a simple and effective method to linearise the chromosome in *E. coli* cells (see page 161 for more detail).

P1*vir* transduction (page 51) was used to introduce the phage DNA sequence *tos* in to the chromosome of the target strain at a location in the termination area near to *dif*. The *tos* sequence is linked to a gene for kanamycin resistance as a selection marker but the gene displays weak resistance. Therefore, the transduction cultures were plated on Mu agar plates with kanamycin at 20 µg/ml instead of the normal concentration of 40 µg/ml. Using the high salt medium, Mu, instead of LB, enhances the effectiveness of the weak kanamycin resistance determinant for unknown reasons. The transductants were infected with N15 bacteriophage and lysogens were subsequently isolated and tested with re-infection to confirm lysogenic state; lysogens cannot be re-infected. The N15 phage expresses the *telN* gene, which encodes the TelN telomerase protein. TelN processes a region within *tos* to linearise the chromosome and generate two termini with hairpin structures. Linearisation of the chromosome was subsequently confirmed using PCR and pulsed-field gel electrophoresis.

Confirmation of linearisation by pulsed-field gel electrophoresis

Conventional gel electrophoresis is not suitable for the separation of very large DNA fragments. In order to resolve DNA fragments larger than the 23 kb upper limit of this technique, a change of direction of the electric field in the electrophoresis tank is introduced periodically. Fragments will reorientate to the new direction at different rates according to their molecular weight and this causes the separation of fragments of different sizes (Nassonova, 2008; Southern et al., 1987)

An overnight culture was used to inoculate 11 ml LB broth and the cells were grown in a shaking water bath to an optical density of A_{600} = 0.3 – 0.5. A 2 ml aliquot was transferred to a microtube and pelleted and the supernatant was discarded. The pellet was resuspended in 100 µl TEE buffer containing 0.05% lauroylsarcosine and 0.5% SDS. 1 % molten low melting point agarose (SeaPlaque™ GTG™ Agarose, Lonza) was prepared and 100 µl was added to each sample and mixed swiftly by pipetting up and down five times being careful not to create bubbles. Each sample was transferred to wells in a disposable plug former (Bio Rad) and put to 4°C to solidify. The plugs were removed from the wells and incubated for 2 hours at 42°C with 10 mg/ml lysozyme in 500 µl TEE containing 0.05% lauroylsarcosine and 0.5% SDS. The solution was removed after 2 hours and replaced with 500 µl TEE containing 1% SDS and 5 mg/ml proteinase K and put at 52°C overnight. The plugs were washed twice in 500 µl TEE at 37°C for 30 min then treated with 1 mM (5 µl) phenylmethane sulphonyl fluoride (PMSF) (freshly prepared as 100 mM stock solution in methanol) in 500 µl fresh TEE at 37°C for 30 min. A further 5 µl of PMSF was added to the plugs and incubated for 1 hour at 37°C. The plugs were

washed three times with 500 µl aliquots of fresh TEE at 37°C for 30 min followed by a final wash with 500 µl 0.1 x TEE at 37°C for 30 min. The plugs were transferred into 300 µl of 1 × 3.1 restriction buffer (NEB) and 3 µl of *Not*I restriction enzyme (NEB) were added to each tube. The plugs were digested overnight at 37°C. They were loaded in to a 1% agarose gel (Pulse field certified agarose, Bio-Rad) in 0.5 x TBE and run on a CHEF Mapper PFGE system (Bio-Rad) for 20 hours at 14°C:
Gradient voltage = 6.0 V/cm
Included angle = 120°
Initial switch = 1.7 s
Final switch = 32.5 s
The gel was stained with 2 x Gel Red solution (Biotium) and visualised using the Gel Doc Molecular Imager (Bio-Rad).

Origin-independent DNA synthesis

A temperature-sensitive allele, *dnaA46*, of the DNA replication initiator protein, DnaA, was utilised in investigating origin-independent DNA synthesis. This protein is active at 30°C but is inactive above 37°C (Frey et al., 1981). Strains carrying the wild type *dnaA* allele are able to maintain 'steady state balanced exponential growth' between 30°C and 42°C (Frey et al., 1981) whereas cells that carry the *dnaA46* allele and are able to grow at 42°C do so independently of *oriC* firing. This was an important tool for studying origin-independent DNA replication that occurs in some of the strain backgrounds where the genes for proteins of interest have been deleted.

Spot dilution assay

Fresh overnight cultures of the strains to be investigated were used to inoculate 11 ml LB to A_{600} = 0.05. The cultures were grown to an optical density of A_{600} = 0.48 in a shaking water bath at 30°C. The cultures would normally grow at different rates and so they were stored on ice until all had reached the correct density. 100 µl of a culture was diluted in 56/2 salts or M9 minimal salts to 10^{-5} using serial dilutions of 10^{-1}. 10 µl of the 10^{-5} dilution of the culture was plated as a single spot on two LB plates. This was repeated with the rest of the dilutions of the culture. One plate was incubated at 30°C (permissive temperature) overnight and the other at 42°C (restrictive temperature). The plates were incubated for 24 – 72 hours. Images were taken using a Gel Doc Molecular Imager and growth of a strain at 42°C was compared to relevant control strains to assess the effect of various genetic and genomic manipulations on origin-independent synthesis. Experiments were performed at least twice independently to confirm the reproducibility of the results. One set of images is presented.

EdU Incorporation

EdU Click-iT labelling of newly replicated DNA – Incorporation of EdU in to newly replicated DNA and preparation of samples for Click-labelling with fluor were carried out as described (Ferullo et al.,

2009). Briefly, fresh overnight cultures of *dnaA46* derivatives were used to inoculate 11 ml LB and the cultures were incubated in a shaking water bath at 30°C and grown to an optical density of A_{600} = 0.3. The cultures were then shifted to the restrictive temperature of 42°C and incubated for 90 min. This was to allow all current rounds of replication to finish whilst preventing new replication from initiating via DnaA. EdU was added to a final concentration of 30 µg/ml (132 µl of EdU in to 11 ml of culture) and the cultures were incubated at 42°C. After 15 min, a 2 ml sample was extracted and fixed in 13 ml 90% methanol. The Click-iT® Plus EdU Alexa Fluor 488 kit (Life Technologies) was used to label the EdU with Alexa Fluor 488. The components of the click labelling reaction were assembled in a specific order, according to the manufacturer's instructions. Click-labelling with fluor was performed as previously described (Ferullo et al., 2009).

Imaging flow cytometry – EdU incorporation in cells was visualised using the Amnis ImageStream Mark II. The settings used were 488 nm excitation with a green emission filter, 50 mW laser, 60 x magnification for bright-field microscopy, and it was set to capture at least 10,000 in focus cells, which was later gated to around 5000 in focus, single cells. The ImageStream data was analysed using the Amnis Ideas imaging flow cytometry software v6.1.

Marker frequency analysis by deep sequencing

A fresh overnight culture was used to inoculate 11 ml LB to A_{600} = 0.05. The culture was incubated in a shaking water bath at 37°C and grown to an optical density of A_{600} = 0.48. In order to allow for a robust number of divisions in exponential phase, the culture was diluted a second time by transferring 100 µl in to 11 ml of fresh LB, pre-warmed to the appropriate temperature. This was then incubated in the shaking water bath until the culture once again reached an optical density of A_{600} = 0.48. The culture was swiftly transferred to a 15 ml Falcon tube and flash frozen in liquid nitrogen. This ensured that DNA replication was arrested whilst the cells were in exponential growth phase. The frozen culture was stored for a minimal time at -20°C before subsequent DNA extraction. A wild type exponential sample and a wild type stationary phase sample were prepared with every batch of samples sent for marker frequency analysis. Marker frequency analysis was performed using Illumina HiSeq 2500 sequencing (fast run) to measure sequence copy number. FastQC was used for a basic metric of quality control in the raw data. Bowtie2 was used to align the sequence reads to the reference. Samtools was used to calculate the enrichment of uniquely mapped sequence tags in 1 kb windows. Bioinformatics analysis was carried out by Earlham Institute.

The data were presented as a marker frequency replication profile, as described previously (Ivanova et al., 2015; Müller et al., 2014; Skovgaard et al., 2011). To present the data as a marker frequency profile, the raw read counts of a sample were first divided by the average of all read counts across the entire genome. This compensates for variation in the absolute numbers of aligned reads in the various samples, which would otherwise affect the position of the data set on the y-axis when plotted as a marker frequency and instead enables all data from a single sequencing run to be plotted

with the same scale on the y-axis. The normalised read counts were then divided by the corresponding normalised read count from the wild type stationary sample. Data points that are outliers due to various technical reasons will occur similarly in both the exponential phase and the stationary phase sample and so in doing this division, these outliers are neutralised and the data set is 'cleaned' significantly. Graphs were then generated for each strain in Microsoft Excel and Adobe Illustrator. Marker frequency profiles for key constructs were generated independently twice.

Chromosome extraction

For replication profiling: The frozen culture was defrosted as swiftly as possible by rubbing in between hands. It was important to proceed quickly in order to minimise the potential for replication restarting. The 11 ml culture was centrifuged to pellet the cells and the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) was used for chromosome extraction. The cells were lysed using a chaotrophic salt-containing buffer followed by RNase A and Proteinase K treatments. The DNA was then bound to a silica membrane in a spin column, washed to remove contaminants and finally released from the silica with elution buffer. The kit was used according to manufacturer's instructions except that the second wash step was carried out twice and 200 µl of a low EDTA concentration elution buffer (page 42) was used instead of the kit elution buffer.

For other purposes: when genomic DNA was needed as a PCR template for example, 2 ml of a fresh overnight culture was spun down to pellet the cells and the kit was then used according to the manufacturer's instructions.

Determination of reversion rates

Lea and Coulson (Lea and Coulson, 1949) devised a system of statistical calculations to quantify the distribution of number of mutants in a culture of bacteria, taking in to account both the total number of mutants present in the culture and the number of spontaneous mutation events that took place in order to result in the total number of mutants. This approach was used to determine the reversion rate of a duplicated sequence to a single copy of the sequence in various strain backgrounds. The *KanKanMX4* duplication is within a gene for kanamycin resistance (Ede et al., 2011). The reversion rate of the *KanKanMX4* construct was estimated using a fluctuation assay based on the method of the median by Lea and Coulson (Foster, 2006; Lea and Coulson, 1949).

To carry out a fluctuation assay, a number of cultures inoculated with a single cell should be grown in parallel, but this is not possible in practice as *E. coli* will not readily grow from such a low cell density in liquid broth. Using full single colonies extracted from a plate culture to inoculate LB broth cultures avoids this problem as each colony will have resulted from a single cell and the cell number present in a plate colony is enough to enable the culture to grow. A fresh overnight culture of each strain was used to inoculate 11 ml LB. The cultures were incubated in a shaking water bath at 37°C and grown to

an optical density of A_{600} = 0.48, at which point the cells were in mid-log phase growth. 50 µl aliquots of the cultures were transferred to 450 µl M9 minimal medium and diluted in 10-fold steps from 10^{-1} to 10^{-5}. 100 µl of each 10^{-5} dilution was spread on an LB agar plate to single colonies and incubated at 37 °C overnight. Using a sterile scalpel, 9 single colonies were cut out of the agar of each plate and transferred to test tubes containing 5 ml LB. The tubes were vortexed to separate the cells and then incubated at 37°C for at least 24 hours. After 24 hours, a second serial dilution was carried out. Six cultures of each strain were diluted 100-fold to 10^{-2} and the remaining three cultures were diluted to 10^{-7}. 100 µl aliquots of the appropriate dilutions were spread on LB plates and Mu plates containing 20 µg/ml kanamycin and put at 37°C overnight. Colonies were scored on LB (viable titre) and kanamycin for reversion frequency of each culture. The data from the fluctuation assay were used to calculate the most likely number of reversion events per culture using the Lea and Coulson method of the median (Foster, 2006; Lea and Coulson, 1949).

Synthetic lethality assay

A plasmid-based lethality assay developed by Bernhardt and de Boer (Bernhardt and de Boer, 2004) was used to establish whether cells lacking 3' exonucleases can tolerate the deletion of the replication origin, *oriC*. A wild type copy of *xonA* under its native promoter was cloned into pRC7, a mini-F plasmid that is unstable and rapidly lost from cells (Bernhardt and de Boer, 2004). The plasmid was used to cover *ΔxonA::apra* in the chromosome of a *Δlac* background. Due to the presence of the *lac* genes in pRC7, loss of the plasmid can be detected using agar plates containing IPTG and X-gal. IPTG inhibits the action of the lac repressor LacI, thereby inducing expression of the *lac* operon (Marbach and Bettenbrock, 2012). This results in the production of β-galactosidase, which cleaves the synthetic substrate, X-gal, resulting in a dark blue precipitate. Cells lacking the plasmid will form white colonies whereas blue colonies show the presence of the plasmid. White sectors within blue colonies form when the plasmid was lost after plating. This system allows detection of synthetically lethal mutations by scoring numbers of blue and white colonies.

ΔoriC::kan was introduced to *ΔxonA* pAM488 and *ΔxonA ΔxseA* pAM488 strains. Cultures of the resulting strains were grown overnight in LB broth containing ampicillin to maintain plasmid selection. These overnight cultures were used to inoculate 11 ml LB to A_{600} = 0.05, which were then grown without ampicillin selection to an optical density of A_{600} = 0.48 in a shaking water bath at 37°C. 100 µl of each culture was diluted in M9 minimal salts to 10^{-5} using serial dilutions of 10^{-1}. 50, 100 and 200 µl aliquots of the 10^{-4} dilutions and 100 and 200 µl aliquots of the 10^{-5} dilutions were spread on LB agar or M9 glucose minimal salts agar supplemented with IPTG and X-gal. The plates were incubated at 37 °C for 48 hr (LB agar) or 72 hr (M9 agar), after which they were photographed and scored for blue and white colonies.

Fluorescence Microscopy

A fresh overnight culture was used to inoculate 11 ml LB to an optical density of A_{600} = 0.05. The culture was incubated in a shaking water bath at 37°C and grown to an optical density of A_{600} = 0.3. Agar platforms were prepared on microscope slides. An adhesive gene frame (1.5 × 1.6 cm, Thermo Scientific) was applied to a glass microscope slide. 110 µl of molten 1% agarose solution in M9 0.2 % glucose minimal medium was pipetted in to the well, taking care to ensure no air bubbles were generated, and was immediately covered with a second slide to ensure a flat and even surface was achieved as the agarose set. After 10 mins, the slides were separated and allowed to dry for a further ~20 mins. 1.5 µl of the culture was transferred to the agar platform on a microscope slide. The slide was rotated slowly a number of times to encourage the drop of culture to spread out, thereby ensuring good areas of a single cell layer of coverage. The agar platform is necessary because *E. coli* cells are motile and so need to be fixed to the microscope slide in order that they can be viewed in focus. Once the drop of culture had dried it was covered with a cover slip and examined using a Nikon T*i*-U inverted microscope equipped with a DS-Qi2 camera (Nikon).

Chromosomal recombineering

Ectopic replication fork trap

The ectopic termination area was created using the recombineering method described (page 50; Datsenko and Wanner, 2000). A kanamycin gene with the 34 bp FLP recognition target (FRT) sequence at either end was amplified from pDIM141 by PCR using two pairs of primers, with one primer in each pair containing the 23-bp sequence of *terA* (Duggin and Bell, 2009). The primer pairs also contained sequence homology for two different locations on the chromosome corresponding to the two integration locations of the *terA* PCR constructs; one at 4.44 Mbp with the *terA* sequence in an orientation that is permissive for replication forks coming from *oriC* (*terA4.44*), and the second at 4.57 Mbp with the *terA* sequence in an orientation that blocks forks arriving from *oriC* (*terA4.57*). The pDIM141 PCR template plasmid was first digested with *Sca*I restriction enzyme in order to minimise false positives at later stages in the process caused by the plasmid reannealing and being present without the PCR fragment inserted. To further prevent this, the PCR products were purified on a TAE gel (page 55) before being used to separately transform a *Δtus* strain carrying pKD46, a lambda Red recombinase expression plasmid, which facilitated the recombination event inserting the DNA fragments in to the chromosome (page 50; Datsenko and Wanner, 2000). A *Δtus* background was used because in the case of the *terA4.57* construct, the *ter* site is in the non-permissive orientation with respect to *oriC* and so it was essential to prevent Tus protein binding whilst the strain had only a single origin. Four colonies from each transformation plate were purified once non-selectively, and the integration of the *terA4.44* and *terA4.57* constructs was confirmed in all eight isolates by colony PCR (page 54).

Two isolates of each of the recombineering strains, *Δtus terA4.44* and *Δtus terA4.57*, were selected to take forward. P1 liquid culture lysates of each of these were prepared. *terA4.44* was introduced via P1*vir* transduction to the chromosome of a strain carrying two copies of *oriC*; the second copy, *oriZ*, is located approximately half way around the right hand replichore (Wang et al., 2011). pCP20, a helper plasmid expressing the FLP recombinase that processes FRT sites (Datsenko and Wanner, 2000), was then used to remove the kanamycin resistance marker, resulting in *oriC⁺ oriZ⁺ terA4.44* cells with a single FRT scar in the place of the kanamycin resistance gene. Both ectopic *ter* constructs have kanamycin resistance as the marker gene and so it was necessary to remove the marker once the *terA4.44* construct was inserted in to the chromosome in order that transductant colonies could then be selected when the *terA4.57* construct was subsequently introduced. The streak test following pCP20 treatment showed that all transformed strains were sensitive to kanamycin, and PCR was used to further confirm the successful removal of the kanamycin resistance marker from two purified colonies following pCP20 treatment.

The streak test also showed that the parent strain was resistant to chloramphenicol but that the transformants were sensitive, indicating that the FLP recombinase had removed the FRT-flanked gene for a chloramphenicol resistance that is linked to the ectopic origin, *oriZ*. The final step to complete the strain construction was to insert the *terA4.57* construct in to the *oriC⁺ oriZ⁺ terA4.44* strain, which was done via P1*vir* transduction and resulted in the final double origin strain with an ectopic termination area, *oriC⁺ oriZ⁺ terA4.44 terA4.57*.

Functional analysis of ter *sites*

To check the ectopic *ter* sites were functional, the *terA4.57* construct was crossed in to wild type and *oriC⁺ oriZ⁺* cells via P1*vir* transduction. Only the transduction in to the double origin strain background produced transductant colonies. Viability in this case is expected because replication forks from *oriC* blocked at the ectopic *ter* site will be met by replication arriving from *oriZ*, whereas in wild type cells, replication forks from *oriC* would reach *terA4.57* and then be unable to progress further, resulting in over a quarter of the chromosome un-replicated. Chromosomal DNA from the origin recombineering strains was amplified using the following primers:

5'terA4_4seq TCTGGTTAATGCAGGTTGCCA, 3'terA4_4seq TAAGGATGGTCGCGTGCAAT

5'terA4_6seq ATCTGCCGGGTACAGGACAT, 3'terA4_6seq AGTGTCAGGGTGCGTGAGAA.

Sanger sequencing was used to confirm that the sequences of both *terA4.44* and *terA4.57* were identical.

Strain reconstruction

The generation of the replication profiles from marker frequency data of the original *oriC⁺ oriZ⁺ ter4.44 ter4.57* strain created as outlined above revealed a chromosomal rearrangement (Figure 14). This is likely to be the result of the action of FLP recombinase (page 51), which was used to remove the FRT-

flanked gene for kanamycin resistance linked to the *terA4.44* construct. To reconstruct the strain and avoid this problem, the *terA4.44* construct was transferred to a wild type background by P1*vir* transduction and the resulting strain was treated with pCP20 to remove the kanamycin resistance gene before any other FRT sites were present. The ectopic replication origin, *oriZ*, was then introduced to the *terA4.44* strain via P1*vir* transduction and finally the *terA4.57* construct was inserted, resulting in the *oriC*$^{+}$ *oriZ*$^{+}$ *terA4.44 terA4.57* strain. *ΔrecG*, *Δtus* and *ΔrecG Δtus* derivatives of this strain were subsequently generated.

Deletion of terC

A gene for kanamycin resistance (*<kan>*) was amplified from pDIM141 by PCR using primers that introduced 40 bp homology to the DNA sequences on either side of *terC*. The resulting DNA fragment was recombineered in to the chromosome of wild type cells using the one-step gene inactivation method (page 50) (Datsenko and Wanner, 2000), resulting in the deletion of the *terC* sequence while the surrounding sequences remain intact.

Fluorescently-labelled dnaQ in an ectopic location

The *dnaQ-mtag<kan>* construct was amplified from strain RRL546 by PCR using primers that introduced homology to the chromosomal DNA sequences on either side of the *lac* operon. The resulting DNA fragment was recombineered in to the chromosome of wild type cells (MG1655) using the one-step gene inactivation method (page 50) (Datsenko and Wanner, 2000), resulting in the deletion of *lacIZYA*. Correct recombinants were those that were resistant to kanamycin and formed white colonies on IPTG X-gal containing agar plates. In cells in which the *lac* operon remains intact, (*lac*$^{+}$), IPTG induces expression of the *lac* operon, which results in the production of β-galactosidase. This hydrolises X-gal and the product of this reaction produces a blue pigment, resulting in blue colonies on IPTG X-gal plates. *Δlac* cells in which the *lac* operon has been recombined with the *dnaQ-mtag-<kan>* construct form white colonies on IPTG X-gal plates. P1 liquid culture lysates of the correct recombinant strains were prepared.

The *frt*-flanked gene for kanamycin resistance associated with the *dnaQ-ypet* construct in AS1059 cells was removed using the FLP/*frt* site-directed recombination system, resulting in strain SLM1142 *dnaQ-ypet*. The P1 liquid culture lysates containing the *ΔlacIZYA::dnaQ-mtag<kan>* construct were used to transfer this construct in to SLM1142 *dnaQ-ypet* cells via P1*vir* transduction (page 51). An IPTG X-gal plate was included in the streak test to confirm the absence of the *lac* operon in the transductant strains.

RecG and DNA replication termination

RecG is a multifunctional DNA translocase that has been shown to target a number of different branched structures *in vitro*, including D-loops and R-loops, Holliday junctions and a variety of replication fork structures (Bianco, 2015; Briggs et al., 2004; Fukuoh et al., 1997; Gupta et al., 2014; Lloyd and Sharples, 1993; Manosas et al., 2013; McGlynn and Lloyd, 2000; McGlynn et al., 1997; Rudolph et al., 2010b; Vincent et al., 1996). It has proven difficult to determine the role RecG has *in vivo* and seemed most likely that RecG might have multiple functions in nucleic acid metabolism within a cell. In line with this, the phenotype of cells lacking RecG is pleiotropic and introducing *ΔrecG* in to otherwise wild type cells results in multiple, but relatively modest, mutant phenotypes, such as mild sensitivity to UV light and ionising radiation, sensitivity to the DNA-damaging agent mitomycin C and a slight reduction in recovery of recombinants in Hfr crosses (reviewed in Lloyd and Rudolph, 2016).

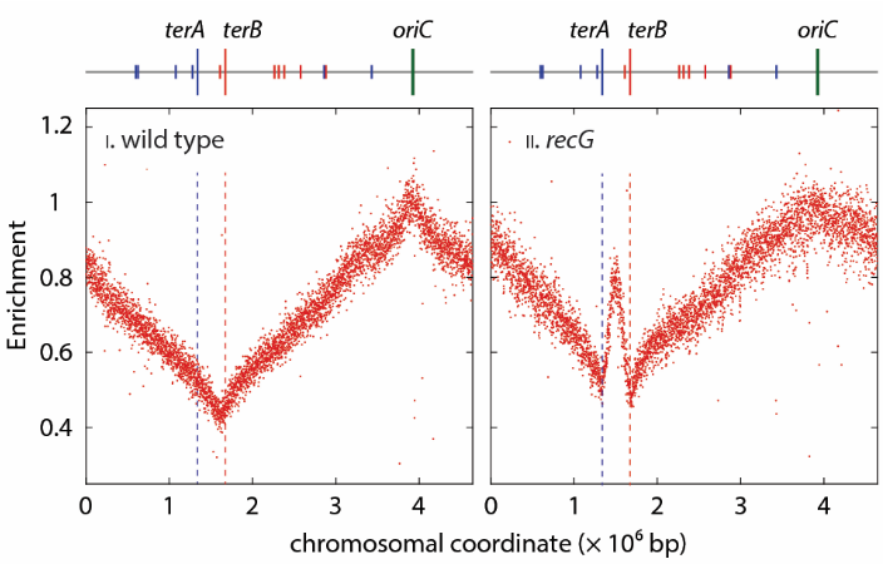

Figure 8: Replication profiles of *E. coli* cells in the presence and absence of RecG. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of oriC (green line) and ter sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore). Sequencing templates were isolated from MG1655 (wild type) and N6576 (*ΔrecG*). Figure reproduced from Rudolph et al., 2013 with permission.

Rudolph and colleagues recently revealed a more pronounced phenotype of *ΔrecG* cells. Marker frequency analysis (MFA) via deep sequencing can be used to generate replication profiles and analyse the replication dynamics in various genetic backgrounds. Replication profiles are generated by plotting copy number (determined as the ratio of each sequence location in a replicating sample to that in a non-replicating control) against each chromosome location (in 1 kb fragments) (page 59).

This assay was used by Rudolph and colleagues to generate replication profiles of wild type cells and cells lacking RecG (Rudolph et al., 2013).

The replication profile of wild type cells (Figure 8, panel I) shows that in an exponential phase culture, the chromosome location present in the highest frequency is the origin of replication, *oriC*, as previously reported (Skovgaard 2011). Conversely, the lowest point of the profile is near *terC* within the termination area of the chromosome. The replication profile of cells lacking RecG is different. The high point is still at *oriC* and the low point is within the termination area. However, there is a peak within the profile that is not seen in the wild type profile (Figure 8, panel II). This peak shows that the chromosomal locations represented within that section of the replication profile are present at a higher frequency in the *ΔrecG* culture than in the wild type culture; there is an amplification of the termination area (Rudolph et al., 2013). The peak shows that in a culture of *ΔrecG* cells, DNA synthesis initiates away from *oriC*. This over-replication of the termination area is not detected when RecG protein is present (Figure 8, panel I wild type profile), which suggests that RecG performs some function that inhibits this origin-independent replication. Rudolph and colleagues suggest that RecG may actually have a single important role in DNA metabolism that may account for the pleiotropic phenotypic traits seen in the absence of RecG, and that is in preventing re-replication of the chromosome by processing DNA substrates that arise as a result of replication fork fusion events (Lloyd and Rudolph, 2016; Rudolph et al., 2009a, 2009b, 2010b, 2013).

Cells lacking RNase HI also show over-replication in the termination area

Origin-independent over-replication was identified in cells lacking RNase HI (encoded by the *rnhA* gene) (Horiuchi et al., 1984; Ogawa et al., 1984), an endonuclease that specifically degrades the RNA strand in a DNA:RNA hybrid (an R-loop) (Tadokoro and Kanaya, 2009) and was termed cSDR (page 38). It was suggested by Kogoma and co-workers that the origin-independent replication seen in *ΔrnhA* cells initiates via the formation of R-loops, which are able to persist in the absence of RNase HI (Kogoma et al., 1985; Meyenburg et al., 1987; reviewed in Kogoma, 1997). Further work revealed that cells lacking RecG also show SDR activity and that loss of both RNase HI and RecG activity is lethal to cells (Hong et al., 1995). It has been shown *in vitro* that RecG can also process R-loops (Fukuoh et al., 1997; Vincent et al., 1996) and as a result, it has been suggested that the DnaA-independent replication seen in the absence of either protein might have a common underlying initiation mechanism (Hong et al., 1995).

Replication profiling of *ΔrnhA* cells (Maduike et al., 2014; Dimude et al., 2015) reveal that, as seen in the absence of RecG, exponentially growing *ΔrnhA* cells show an amplification of the termination area, demonstrating that origin-independent over-replication initiates in this region of the chromosome in both *ΔrecG* and *ΔrnhA* cells (Figure 8 and Figure 9 respectively).

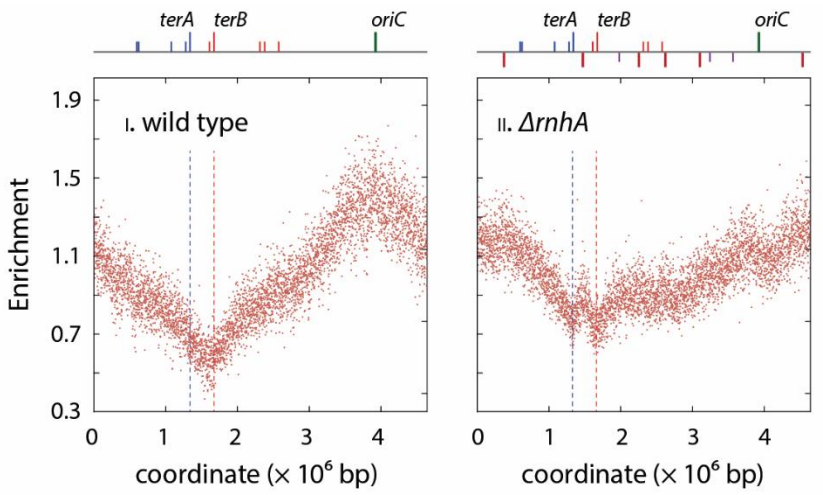

Figure 9: Replication profiles of *E. coli* cells in the presence and absence of RNase HI. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore). Sequencing templates were isolated from MG1655 (wild type) and N4704 (*ΔrnhA*). Figure reproduced from Dimude et al., 2015 with permission.

In the absence of RNase HI however, there were further deviations from the wild type profile. There are additional peaks along the profile outside of the termination area (Dimude et al., 2015; Maduike et al., 2014). Kogoma and colleagues first identified five sites of replication initiation in the absence of RNase HI and *oriC*, which were termed *oriK*s (Kogoma, 1997). The peaks seen in the replication profile of *ΔrnhA* cells suggest that origin-independent DNA synthesis is indeed able to initiate at multiple locations of the chromosome and they correspond with the original data as well as more recent experiments (Dimude et al., 2015; Maduike et al., 2014). Whilst it is clear that cells lacking RecG exhibit cSDR (Hong et al., 1995), the fact that the marker frequency profiling reveals that it is most prevalent in the termination area (Rudolph et al., 2013) and not at multiple defined chromosomal locations like that seen in *ΔrnhA* cells lead to questions about the similarities and differences of the over-replication in the two genetic backgrounds, with implications for the potentially different roles of RecG and RNase HI in DNA metabolism.

Over-replication sustaining growth

Despite the origin-independent over-replication seen in the replication profiles of both *ΔrecG* and *ΔrnhA* cells, it has been shown previously that while cells lacking RNase HI are able to grow in the absence of origin firing, cells lacking RecG are not. Normal DNA replication initiated at *oriC* is mediated by the main initiator protein, DnaA, encoded by *dnaA*. This initiation system can be manipulated through the introduction of a temperature sensitive allele of DnaA. *dnaA46* (referred to as *dnaA* from here onwards) encodes a DnaA(ts) protein that is active at 30 °C and inactive at 42 °C (Frey et al., 1981), meaning that origin firing can be prevented by incubating cells at the restrictive temperature.

A spot dilution assay was used to assess growth that arises as a result of origin-independent DNA synthesis in *dnaA ΔrecG* and *dnaA ΔrnhA* cells (Dimude et al., 2015; Rudolph et al., 2013).

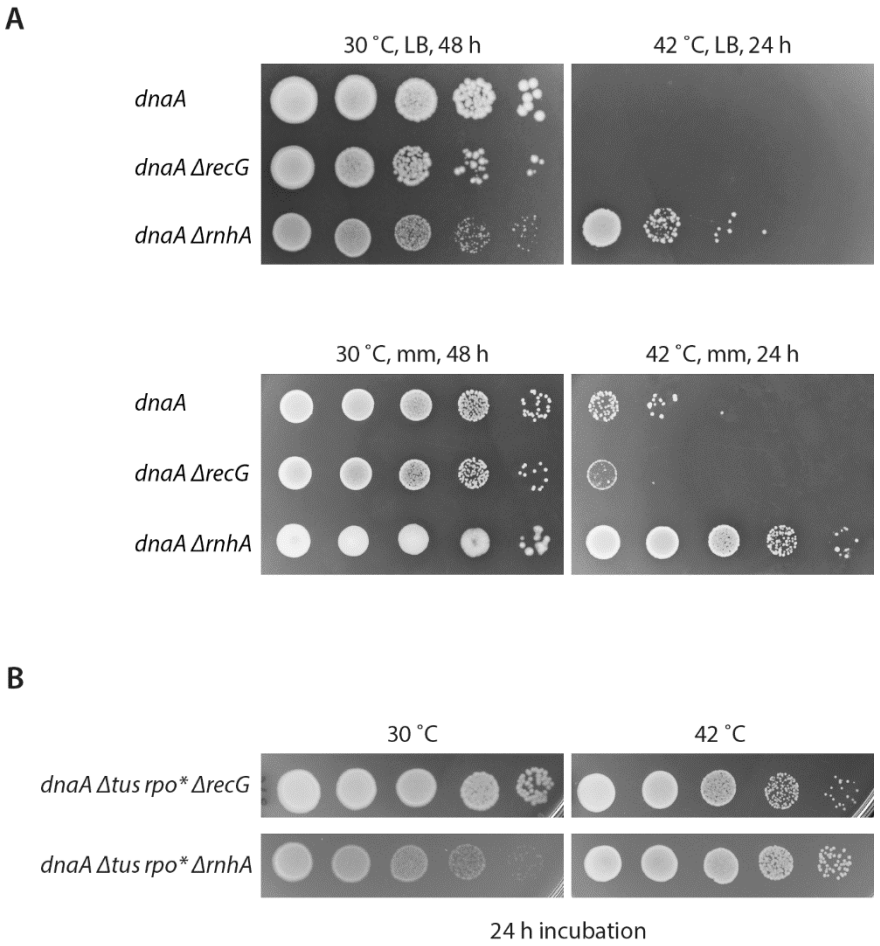

Figure 10: Cell growth in the absence of origin firing. **A**) A spot dilution assay was used to evaluate growth arising via origin-independent DNA replication in *dnaA(ts) ΔrecG* and *dnaA(ts) ΔrnhA* cells. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The strains used were AU1054 (*dnaA46*), AU1091 (*dnaA46 ΔrecG*) and AU1066 (*dnaA46 ΔrnhA*). Figure reproduced from Dimude et al., 2015 with permission. **B**) The replication fork trap in the termination area was inactivated in *dnaA ΔrecG* and *dnaA ΔrnhA* cells by deletion of the *tus* gene (*Δtus*) and an *rpoB*35* point mutation was introduced (*rpo**), which destabilises ternary RNA polymerase complexes. The experiment was carried out on LB agar plates. The strains used were RCe268 (*dnaA46 Δtus rpo* ΔrecG*) and RCe309 (*dnaA46 Δtus rpo* ΔrnhA*). Figure is modified from Dimude et al., 2015 and Rudolph et al., 2013 with permission.

This assay revealed that unlike cells lacking RecG, *dnaA ΔrnhA* cells can establish colonies at the restrictive temperature, confirming that the origin-independent over-replication seen in the absence of RNase HI (Figure 9) is robust enough to sustain cell growth, especially on minimal salts agar (Figure 10A; Dimude et al., 2015; Kogoma, 1997; Maduike et al., 2014; Rudolph et al., 2013). *dnaA ΔrecG* cells that grew robustly at the permissive temperature were only able to grow at the restrictive temperature when a combination of two mutations were introduced; *Δtus* and *rpoB*35* (Figure 10B; Rudolph et al., 2013). *tus* encodes the terminus utilization substance, Tus, which binds to *ter*

sequences that flank the termination area and forms a unidirectional block to replication fork progression (page 24; Duggin and Bell, 2009). *Δtus* inactivates the *ter*/Tus replication fork trap. *rpoB*35* (referred to as *rpo** from here onwards) is a point mutation in the β subunit of RNA polymerase that reduces conflicts between replication and transcription complexes by reducing the ability of RNAP to pause and backtrack (Dutta et al., 2011). That these two mutations allow *dnaA ΔrecG* cells to grow in the absence of origin firing is in line with the replication profile of *ΔrecG* cells (Figure 8; Rudolph et al., 2013), which shows a high level of over-replication confined to the termination area, prevented from progressing further by the Tus/*ter* complexes. Once the replication fork trap is inactivated, the over-replication would then be able to escape from the termination area and so has the potential to maintain chromosome replication in the absence of origin firing.

The work presented in this chapter builds on an extensive body of work (Lloyd and Rudolph, 2016; Rudolph et al., 2009a, 2009b, 2010b, 2010a, 2013) and forms part of a bigger study within the Rudolph lab to further characterise the origin-independent replication seen in *ΔrecG* cells with the aim of increasing our understanding of the role RecG plays in limiting genomic instability through DNA metabolism. With the exception of the *terC* deletion work (page 80), the data presented in this chapter have been published (Dimude et al., 2015; Midgley-Smith et al., 2018).

Origin-independent synthesis in ΔrecG and ΔrnhA cells in the absence of origin firing

To further characterise the origin-independent over-replication seen in cells lacking RecG or RNase HI, newly replicated DNA was visualised via pulse-labeling with 5-ethynyl-2'-deoxyuridine (EdU) in order to determine if origin-independent synthesis arises in all cells of a population of *ΔrecG* or *ΔrnhA* cells or if it is limited to a subset of the population.

The *dnaA46* allele was introduced to *ΔrecG* and *ΔrnhA* cells. Cultures of *dnaA*, *dnaA ΔrecG* and *dnaA ΔrnhA* strains were grown at the permissive temperature of 30 °C until the cultures reached exponential growth phase. The cultures were then shifted to the restrictive temperature for 90 minutes to inhibit origin firing and allow all current rounds of DnaA(ts)-initiated replication to finish. Cells were pulse labelled with the nucleotide analogue EdU for 15 minutes at 42 °C before being fixed with ethanol, and a click-labelling reaction was used to fluorescently label EdU incorporated in to chromosomal regions where nascent DNA synthesis had taken place (Ferullo et al., 2009) (page 58). EdU was used as the nucleotide analogue in this assay instead of bromodeoxyuridine (BrdU) as the immunostaining technique used for detection of BrdU incorporation includes a denaturing treatment. EdU allows for quantification of the intensity and location of fluorescence, of both a sample population and individual cells, under non-denaturing conditions, and so allows the use of microscopy to visualise the fluorescent signal in individual cells (Ferullo et al., 2009). The fluorescently labelled samples were

visualised via high-resolution microscopy in flow using an Amnis Imagestream^{x} Mark II and a minimum of 5000 in-focus cells were analysed for each strain (page 59).

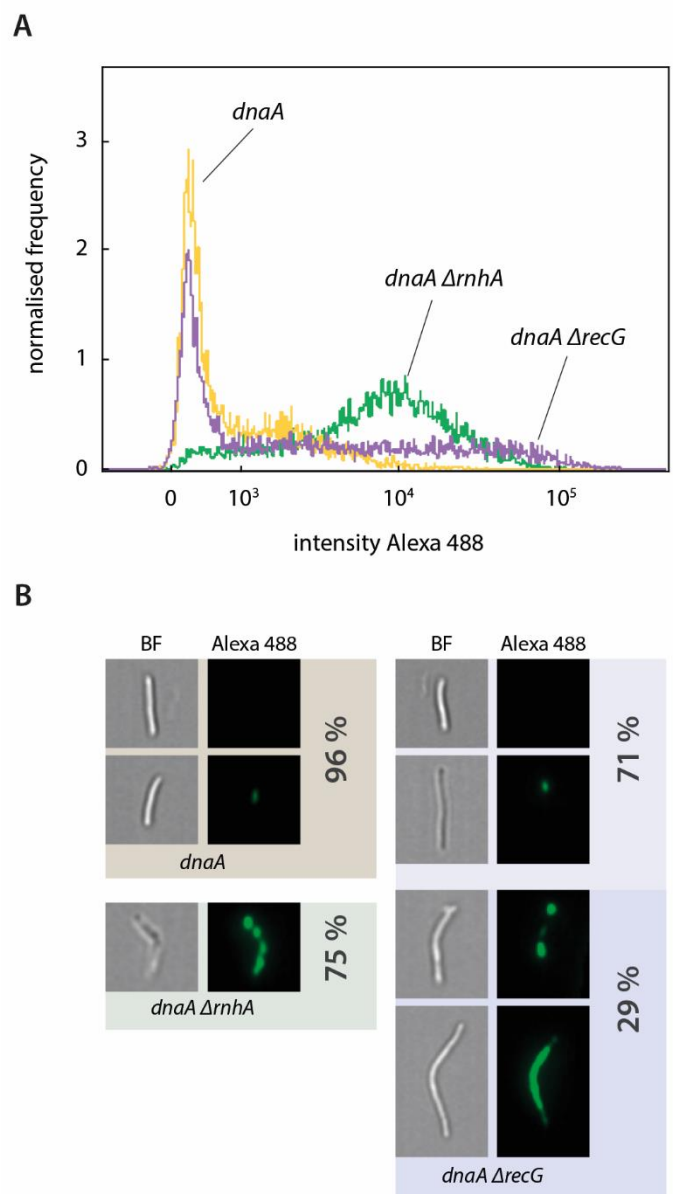

Figure 11: Origin-independent -DNA synthesis in *dnaA(ts)* cells lacking RecG or RNase HI. **A**) Imaging flow cytometry was used to evaluate DnaA-independent DNA synthesis arising in *dnaA*, *dnaA ΔrecG* and *dnaA ΔrnhA* cells. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Synthesis occurring at the restrictive temperature is achieved in the absence of *oriC* firing and was detected by pulse-labelling samples at 42 °C with the nucleotide EdU followed by labelling with Alexa Fluor 488. The fluorescent signal of each sample was analysed using the Amnis Imagestream^{x} Mark II. A minimum of 5000 in-focus cells were analysed for each strain. The data plotted were from a single experiment. The experiment was repeated with a similar number of cells and was reproducible. **B**) 60 × magnification was used to observe the distribution of the fluorescent signal within the population and within individual cells; representative images are shown. The strains used were AU1054 (*dnaA46*), AU1091 (*dnaA46 ΔrecG*) and AU1066 (*dnaA46 ΔrnhA*).

A *dnaA* single mutant was used as a control to establish a baseline result in the absence of origin firing. The graph of the total fluorescent intensity of the cells in each culture (Figure 11A) shows two distinct

data distributions; *dnaA* and *dnaA ΔrecG* cells show similar fluorescence intensity on a population level whilst *dnaA ΔrnhA* cells are distinct from those two strains. The vast majority of *dnaA* cells (96 %) exhibited very little or no fluorescent signal, which means that DnaA(ts)-initiated replication had essentially finished after the 90-minute incubation at 42 °C and no further replication was initiated. In *dnaA* cells lacking RecG, 71 % of cells show comparable levels of fluorescence to that seen in *dnaA* single mutant cells, indicating that in the majority of *dnaA ΔrecG* cells, DNA synthesis ceases once ongoing rounds of replication from *oriC* terminate. However, there is a second population of cells within the *dnaA ΔrecG* sample that show high levels of fluorescence distributed along the whole cell or spots of intense fluorescent signal, suggesting high levels of DNA synthesis occurring and meaning that DnaA-independent synthesis is able to initiate in 29 % of *dnaA ΔrecG* cells (Figure 11B).

In contrast to what is seen in cells lacking RecG, the majority of *dnaA ΔrnhA* cells showed a high frequency of strong fluorescent signal (Figure 11A). The image in Figure 11B is representative of 75% of the sample investigated and shows robust fluorescent signal distributed along the entire length of the cells, indicating high levels of EdU incorporation and therefore high levels of DNA synthesis taking place at multiple chromosome locations even in the absence of origin firing. The distinct spots of fluorescence are in line with the multiple initiation sites observed via the MFA of *ΔrnhA* cells, which shows a number of deviations from the wild type profile (Figure 9) (Dimude et al., 2015), as has been previously reported (Maduike et al., 2014). These peaks seen in other chromosomal locations in addition to that at *oriC* indicate amplification of the sequences in these locations, which implies that DNA synthesis initiates at these locations, and the number of peaks combined with our LOESS regression data suggest five main initiation sites in addition to the normal replication origin (Dimude et al., 2015). Only 4 % of *dnaA ΔrnhA* cells showed no fluorescent signal.

Despite the high level of origin-independent synthesis detected in a portion of *dnaA ΔrecG* cells via the EdU pulse-labelling assay (Figure 11), it has been shown previously that *dnaA ΔrecG* cells are not able to grow in the absence of origin firing (Figure 10; Rudolph et al., 2013) and so this DNA synthesis is unable to sustain growth, in contrast to the synthesis and origin-independent growth seen in *ΔrnhA* cells (Figure 10; Dimude et al., 2015; Rudolph et al., 2013).

Over-replication is likely to initiate from different substrates in cells lacking RecG or RNase HI

It has been shown that both RecG and RNase HI can process R-loops and so it was thought that over-replication seen in the absence of either protein initiates via the formation of R-loops that are able to persist in *ΔrecG* or *ΔrnhA* cells (Kogoma, 1997). If this is the case, RNase HI should be able to compensate for RecG and vice versa. To investigate if this is the case, the effect of an increased cellular concentration of RNase HI on the origin-independent growth phenotype of *dnaA ΔrnhA* and *dnaA*

*ΔrecG Δtus rpo** cells was investigated. The *Saccharomyces cerevisiae* (*S. cerevisiae*) RNase H1 protein (encoded by the *RNH1* gene) was used for this as it will degrade the RNA from an RNA:DNA hybrid and has been shown to reduce some of the phenotypes of *E. coli ΔrnhA* cells (Cerritelli and Crouch, 2009), but will not interact with *E. coli* proteins. Using a protein from a different organism separates the structural and protein-protein interactions from functionality, which allows more certainty that it is indeed the degradation of RNA:DNA hybrids that causes any effect on the phenotypes observed. The *RNH1* gene was cloned in to a high copy-number expression plasmid under the control of the P_{araBAD} promoter from the arabinose operon, which allows the expression of the gene to be controlled. When arabinose is present in the growth medium, expression is induced, and when glucose is present in the growth medium, expression is repressed (Guzman et al., 1995). *dnaA ΔrecG Δtus rpo** and *dnaA ΔrnhA* cells were transformed with the yeast *RNH1* expression plasmid and *dnaA ΔrecG Δtus rpo** and *dnaA ΔrnhA* strains carrying pLAU17 (Lau et al., 2003) were generated as vector control strains. Using a spot dilution assay, the resulting strains were used to investigate if increasing the capability to deal with R-loops has an effect on the origin-independent growth phenotype seen in the absence of either RecG or RNase HI. Briefly, samples of exponential phase cultures were spotted in duplicate on to two sets of minimal salts agar plates supplemented with ampicillin in order to maintain selection of plasmid-containing cells; one set contained arabinose as the carbon source to induce expression of *RNH1* and the second set contained glucose as the carbon source in order to repress expression of *RNH1*. One of each plate type was incubated at 30 °C and the second at 42 °C to prevent replication initiating via DnaA(ts). Single colonies formed as a result of origin-independent over-replication (42 °C) were scored as a fraction of the number of colonies on the corresponding control plate (30 °C), which were set to 1.

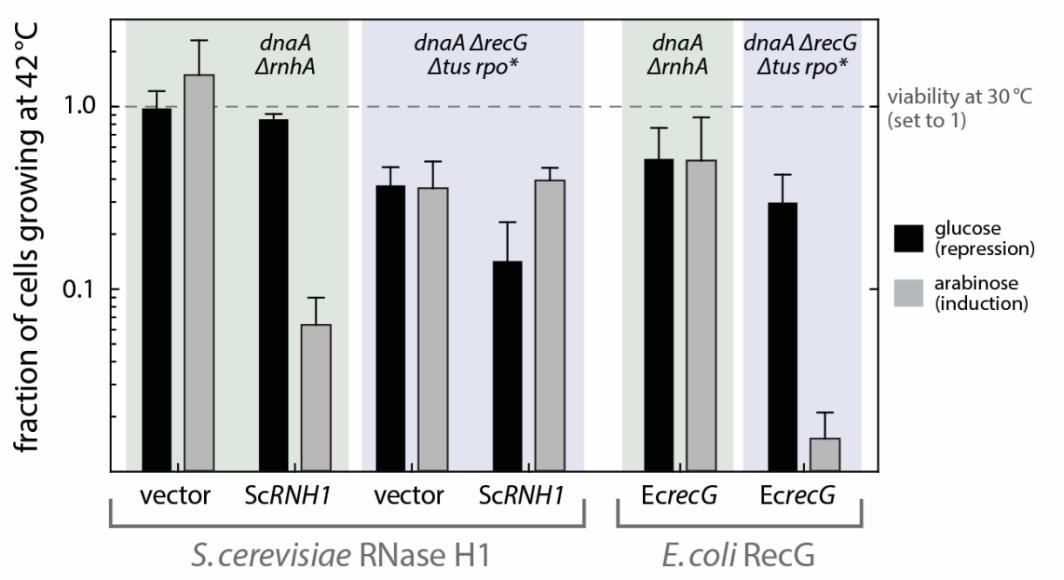

Figure 12: Complementation of *ΔrnhA* and *ΔrecG* with *S. cerevisiae* RNase H1 and *E. coli* RecG. A spot dilution assay was used to evaluate the effect of over-expression of yeast *RNH1* (pECR22) or *E. coli recG* (pDIM104) on origin-independent growth in *dnaA(ts) ΔrecG Δtus rpo** and *dnaA(ts) ΔrnhA* cells. pLAU17 was used as the vector control. Expression of the ectopic genes is repressed by glucose (0.2 %) and

induced by arabinose (0.05 %). The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA*46, which produces a protein that is functional at 30°C but not at 42°C. Viability was determined by growth at 30 °C and is set to 1 (dashed line) and the corresponding growth occurring at the restrictive temperature, in the absence of *oriC* firing, is shown as a fraction relative to this. The strains used were RCe557 (*dnaA*46 Δ*rnhA* pLAU17), RCe552 (*dnaA*46 Δ*rnhA* pECR22), SLM1010 (*dnaA*46 Δ*recG* Δ*tus rpo** pLAU17), SLM1008 (*dnaA*46 Δ*recG* Δ*tus rpo** pECR22), SLM1104 (*dnaA*46 Δ*rnhA* pDIM104) and RCe326 (*dnaA*46 Δ*recG* Δ*tus rpo** pDIM104). Data are means from at least 3 independent experiments (±standard deviation [SD]).

As seen before, *dnaA* cells lacking RNase HI were able to sustain cell growth in the absence of origin firing (*dnaA ΔrnhA* cells in the *S. cerevisiae* sector of the graph in Figure 12). The presence of the vector without the yeast gene had no effect and similarly when the yeast gene was repressed (*dnaA ΔrnhA* Sc*RNH1* glucose repression column, Figure 12) there was no change to the fraction of cells growing at 30 °C that were also able to grow at 42 °C. However, the expression of yeast *RNH1* resulted in more than a twenty-fold reduction in growth seen in the absence of origin firing compared to that seen in the same strain at the permissive temperature (*dnaA ΔrnhA* Sc*RNH1* arabinose induction column, Figure 12). The suppression of this phenotype of *ΔrnhA* cells by *RNH1* supports the idea that origin-independent over-replication in *ΔrnhA* cells initiates via R-loops and demonstrates that yeast RNase H1 can compensate for *E. coli* RNase HI by processing R-loops and preventing over-replication of the chromosome. In contrast, expression of yeast *RNH1* did not result in a reduction of origin-independent growth in *dnaA ΔrecG Δtus rpo** cells (*dnaA ΔrecG Δtus rpo** Sc*RNH1* arabinose induction column, Figure 12). It cannot be excluded that RecG is involved in R-loop processing. However, for the ability of *dnaA ΔrecG Δtus rpo** cells to grow in the absence of origin firing, R-loops either do not play a role or, alternatively, they are for some reason structurally so different that they are not accessible to RNase H1.

Can RecG compensate for RNase HI? The assay was carried out again, this time using a high copy-number expression plasmid containing the *recG* gene under the control of the P_{araBAD} promoter instead of the yeast *RNH1* gene. *dnaA ΔrnhA* and *dnaA ΔrecG Δtus rpo** cells were transformed with the plasmid, generated spot dilution plates and counted colonies formed after incubation at 30 °C and 42 °C. The over-expression of *E. coli recG* reduced the growth of *ΔrecG* cells occurring in the absence of origin firing over 20-fold (*dnaA ΔrecG Δtus rpo** Ec*recG* arabinose column, Figure 12), as would be expected, whereas the origin-independent growth in *dnaA ΔrnhA* cells was unaffected by the increase in RecG (*dnaA ΔrnhA* Ec*recG* glucose column, Figure 12). RecG is either unable to access R-loops in *ΔrnhA* cells for some reason and so cannot prevent replication from initiating at these structures, or R-loops do not contribute to origin-independent growth and so replication is able to initiate in the absence of RNase HI even when *recG* is expressed from a high copy-number plasmid.

Over-replication in ΔrecG cells occurs at fork fusion locations

Cells carrying a second copy of the origin of replication have been used to further investigate over-replication in the termination area of *ΔrecG* cells. The ectopic origin, called *oriZ*, is located approximately half way around the right hand replichore (Figure 13). Wang and colleagues (2011) showed that both origins in *oriC*$^{+}$ *oriZ*$^{+}$ cells functioned normally in initiating bidirectional replication and fired with equal efficiency. The introduction of *oriZ* will cause replication forks to meet and fuse outside of the normal termination region as replication forks moving clockwise from *oriC* and anticlockwise from *oriZ* will meet in the middle of the short interval between the two origins, which becomes an ectopic fork fusion region (Figure 13A). Replication profiles of *oriC*$^{+}$ *oriZ*$^{+}$ cells confirmed this, showing two distinct low points corresponding to the termination area and the ectopic fork fusion region (Ivanova et al., 2015; Rudolph et al., 2013). In the absence of RecG, there is a significant increase in the peak of over-replication seen in the native termination area compared to that seen in single origin *ΔrecG* strain. Crucially, there is also a change to the profile in the ectopic fork fusion region, which becomes noticeably shallower in the *ΔrecG* derivative. This is consistent with the notion that when replication forks meet, over-replication is able to initiate and will trigger forks moving back towards each origin, resulting in the broad amplification of the shorter region between the origins seen in *oriC*$^{+}$ *oriZ*$^{+}$ *ΔrecG* cells (Midgley-Smith et al., 2018; Rudolph et al., 2013).

The effect of *ΔrecG* on the ectopic fork fusion region, whilst present, is not particularly conspicuous. If *ter* sites are inserted into this region in such a way that they form an ectopic fork trap around the new fork fusion area, does a peak of over-replication appear?

To investigate this and to determine whether over-replication initiation in *ΔrecG* cells is co-localised with fork fusion events and not specific to defined chromosomal locations, I created an ectopic replication fork trap in a double origin strain. Two *ter* sites were inserted into the chromosome between *oriC* and *oriZ* flanking the region where most replication forks meet, which is the low point of the profile of the short interval between the two origins (Ivanova et al., 2015).

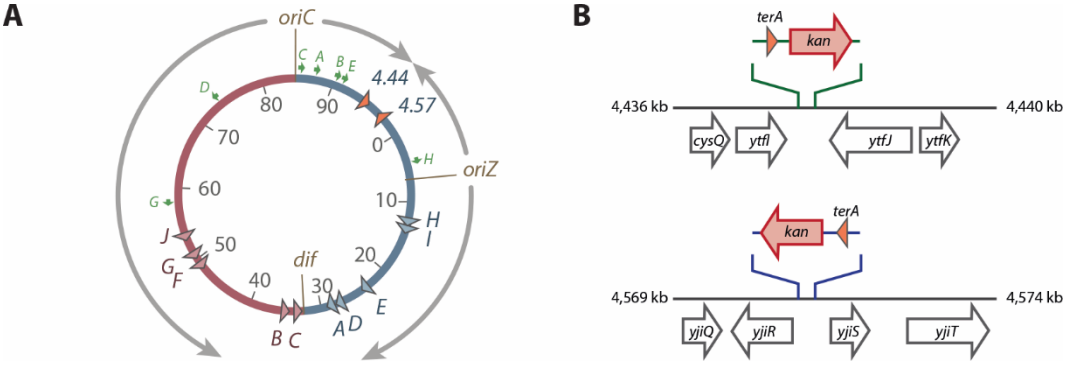

Figure 13: The ectopic *ter* sites. **A**) Schematic representation of the chromosome in *oriC*⁺ *oriZ*⁺ cells that have two ectopic *ter* sites, indicated by orange triangles, located on either side of the ectopic replication fork fusion region. **B**) Schematic representation of the integration locations of the *terA* sequence at 4.44 Mb and 4.57 Mb.

The ectopic termination area was created using the recombineering method described (page 50; Datsenko and Wanner, 2000) to insert the *terA* sequence at two chromosomal locations, one at 4.44 Mb with the *terA* sequence in an orientation that is permissive for replication forks coming from *oriC* (*terA4.44*), and the second at 4.57 Mb with the *terA* sequence in an orientation that blocks forks arriving from *oriC* (*terA4.57*) (Figure 13). The constructions were initially carried out in *Δtus* cells to prevent the interference of functioning Tus/*ter* complexes with chromosome replication in the single origin strain background used initially. The *ter* constructs were then transferred via P1 transduction in to *oriC*⁺ *oriZ*⁺ cells. As both *terA* constructs carried the same antibiotic resistance marker (*kan*), the two integrations were done separately, with the marker of the *terA4.44* removed from *oriC*⁺ *oriZ*⁺ *terA4.44* cells before the integration of *terA4.57* was carried out (see page 62 for more detail of strain construction).

Functional analysis of ter sites

To check the ectopic *ter* sites were functional, the *terA4.57* construct was crossed in to wild type and *oriC*⁺ *oriZ*⁺ cells via P1*vir* transduction. Only the transduction into the double origin strain background produced transductant colonies. Viability in this case is expected because replication forks from *oriC* blocked at the ectopic *ter* site will be met by replication arriving from *oriZ*, whereas in wild type cells, replication forks from *oriC* would reach *terA4.57* and then be unable to progress further, resulting in over a quarter of the chromosome un-replicated. Sanger sequencing was used to confirm that the sequences of both *terA4.44* and *terA4.57* were identical.

Over-replication in an ectopic termination area

Does a peak of over-replication form in the fork fusion area of the shorter interval between the two origins in *oriC*⁺ *oriZ*⁺ *ΔrecG* cells when ectopic *ter* sites are inserted in this region to create a replication fork trap? Replication profiles of exponential phase *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57* cells in the presence

and absence of RecG and Tus were established through marker frequency analysis by deep sequencing (Figure 14).

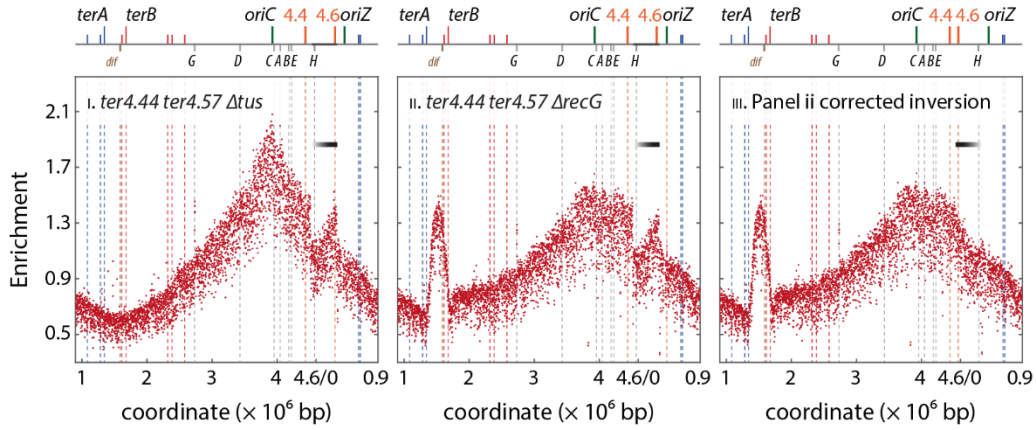

Figure 14: Replication profiles of *E. coli oriC*⁺ *oriZ*⁺ cells with an ectopic replication fork trap. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal coordinates, which are offset to start at 0.9 Mb to make it easier to see both replication origins. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* and *oriZ* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and *dif* below. Sequencing templates were isolated from SLM1161 (*oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57 Δtus*) and SLM1132 (*oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57 ΔrecG*). The section of the profile in panels I and II that deviates from the trend indicates an inversion. The data points for this section of the profile for *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57 ΔrecG* were subsequently inverted (panel III), which restored the continuity of the profile.

It very quickly became clear when the profiles were generated that there was a problem with the double origin strain construction. Previously established profiles of double origin strains show two peaks of equal height coinciding with the replication origin chromosomal locations, demonstrating that both the original and ectopic origin fire at equal rates (Ivanova et al., 2015; Rudolph et al., 2013). The profile generated for *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57* cells is so noisy as to render it unreadable (data not shown). The *Δtus* derivative in this case can substitute for that strain as it has been shown previously that the replication profile of *Δtus* cells is very similar to that of wild type cells (Ivanova et al., 2015; Rudolph et al., 2013). In a double origin strain, the profile of a *Δtus* derivative differs from the *tus*⁺ only in the region of the native termination area, where the low point of the profile is shifted to the midpoint between the two origins (Ivanova et al., 2015).

The *Δtus* derivative of the *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57* strain (Figure 14 panel I) shows that there is a peak of initiation at *oriC* but not at *oriZ*, which means that *oriZ* is not functional in this strain. There is a clear deviation from the downward trend of the replication profile from *oriC* in the right hand replichore, which indicates that this section of the chromosome has been inverted (Skovgaard et al., 2011). The same anomaly is seen in the *ΔrecG* derivative in panel II, which indicates that the error occurred in the construction of the ectopic termination area. The ectopic replication origin construct, *oriZ*, consists of the origin sequence and a selectable marker, a chloramphenicol resistance gene, which is flanked by *frt* sites. It is likely that the FLP recombinase, which was used to remove the *frt*-

flanked gene for kanamycin resistance linked to the *terA4.44* construct, had a role in enabling an inversion event between the *frt* sites of the ectopic origin construct and one of the ectopic *ter* sites. The replication profile of the *ΔrecG* derivative in which the data points that span this area have been inverted reveals that the continuity of the profile is then restored (Figure 14 panel III).

Strain reconstruction

To reconstruct the strain and avoid the problems seen above, the *terA4.44* construct was transferred to a wild type background by P1*vir* transduction and the resulting strain was treated with pCP20 to remove the kanamycin resistance gene before any other *frt* sites were present. The ectopic replication origin, *oriZ*, was then introduced to the *terA4.44* strain via P1*vir* transduction and finally the *terA4.57* construct was inserted, resulting in the *oriC*$^{+}$ *oriZ*$^{+}$ *terA4.44 terA4.57* strain. *ΔrecG*, *Δtus* and *ΔrecG Δtus* derivatives of this strain were subsequently generated.

Over-replication in an ectopic fork fusion area

Replication profiles of exponential phase *oriC*$^{+}$ *oriZ*$^{+}$ *terA4.44 terA4.57* cells in the presence and absence of RecG and Tus were established through marker frequency analysis by deep sequencing. The reconstruction of the double origin ectopic *ter* strain was successful. The profile of *oriC*$^{+}$ *oriZ*$^{+}$ *terA4.44 terA4.57* cells shows two clear peaks that correspond with *oriC* and *oriZ* and show that both the native and ectopic origins are functional (Figure 15, panel II). The peaks are a similar height, which means that the chromosomal locations represented by those data points are present in the library prep at similar frequencies and indicates that the origins are firing with equal efficiency on a population basis, in line with previous work on double origin strains (Ivanova et al., 2015; Rudolph et al., 2013; Wang et al., 2011). The low points in two regions of the profile correspond with fork fusion areas.

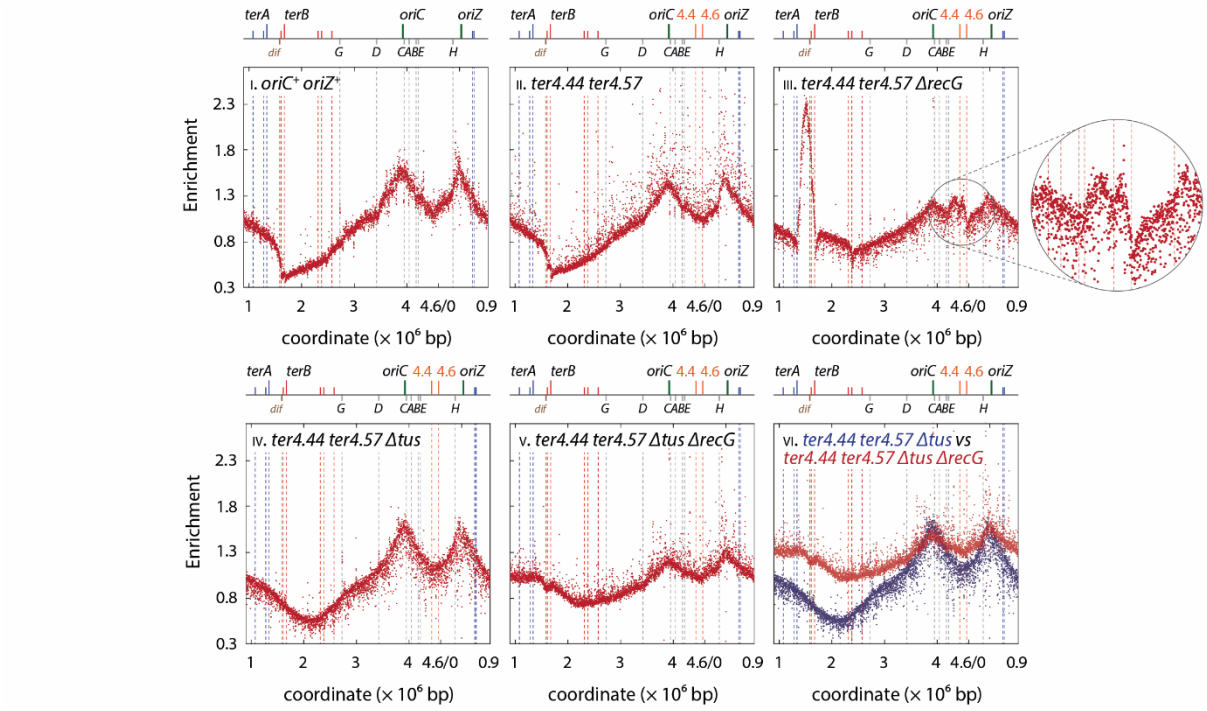

Figure 15: Marker frequency analysis of exponential phase *E. coli* cells with two replication origins and an ectopic termination area, in the presence and absence of RecG and Tus. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal coordinates, which are offset to start at 0.9 Mb to make it easier to see both replication origins. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* and *oriZ* (green lines) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and *dif* below. Orange lines represent the ectopic *ter* sites. The profiles from panel IV and V were overlaid in panel VI and were aligned using the origin peaks. See text for details. The strains used were RCe504 (*oriC*⁺ *oriZ*⁺), SLM 1197 (*oriC*⁺ *oriZ*⁺ *ter4.44 ter4.57*), RCe714 (*oriC*⁺ *oriZ*⁺ *ter4.44 ter4.57 ΔrecG*), RCe745 (*oriC*⁺ *oriZ*⁺ *ter4.44 ter4.57 Δtus*) and RCe760 (*oriC*⁺ *oriZ*⁺ *ter4.44 ter4.57 Δtus ΔrecG*). These data have been published (Midgley-Smith et al., 2018). Replication profiles of *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57* and *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57 ΔrecG* cells have been generated independently twice; see page 163 for the second data set.

The profile of *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57* is in essence the same as that for *oriC*⁺ *oriZ*⁺ cells (Figure 15, panels I and II). There is a distinct step in the profile of double origin strains at the native termination area between *terA* and *terC*, which is due to the asymmetric nature of the chromosome with an ectopic origin inserted. On replication initiation, one replication fork from each origin will proceed towards the native termination area, but the fork from *oriZ* will reach it first and be blocked by the Tus/*ter* block at *terC* until the fork from the opposite replichore arrives from *oriC*. On a population basis, the quarter of the chromosome between *oriZ* and *terC* (Figure 13, right hand replichore) will be present at a higher frequency than the corresponding quarter in the left hand replichore, resulting in the step seen in the replication profile between the left and right hand replichores.

This is not seen in the ectopic termination region in *oriC*⁺ *oriZ*⁺ *terA4.44 terA4.57* cells (Figure 15, panel II) even though the slightly asymmetric shape of the valley indicates that on a population basis forks progress further towards *oriZ* from *oriC* than they do towards *oriC* from *oriZ*. However, it is

unlikely that the ectopic termination area will be as defined as the native termination area. The distance between the ectopic *ter* sites is less than the distance between the innermost *ter* sites of the native termination area, and there are only single *ter* sites inserted in each orientation, which is in contrast to the situation in the native termination area. In addition, the *rrnH* operon will be encountered by the replication fork from *oriZ* in the opposite direction to normal, giving rise to potential head-on replication-transcription encounters, which have previously been shown to cause problems for replication progression (Dimude et al., 2015; French, 1992; Ivanova et al., 2015; Mirkin and Mirkin, 2007; Wang et al., 2007). On average forks fuse near to the ectopic *ter* site at 4.57 Mb, as indicated by the low point between the origins, although it is likely to vary from cell to cell. Forks might get arrested at the 4.57 Mbp location, but the lack of distortion to the shape of the valley between the two origins compared to that seen in the native termination area suggests that any blocks, if they occur, are only very short-lived.

There are in fact two low points in the native termination area of *oriC*$^{+}$ *oriZ*$^{+}$ and *oriC*$^{+}$ *oriZ*$^{+}$ *terA4.44 terA4.57* cells, one at *terC* and an even lower one at *terB* (Figure 15, panels I and II). The fact that there is a step in the profile at the *terC* location, combined with the results of 2-D gel analysis showing that *terC* is functional (JUD and CJR, unpublished results), confirm that a fork block is successfully established a *terC*. The observed shape of the profile between *terC* and *terB* can be explained by the understanding that a Tus/*ter* complex is not a perfect block. Recent work by Moolman *et al.* (Moolman et al., 2016) demonstrated that a replisome travelling in a clockwise direction is impeded by Tus/*terC* but it is capable of overcoming this block and so is not halted indefinitely. In the asymmetric double origin strain, the fork from *oriZ* will always arrive at the termination area before the fork from *oriC*, and so will proceed through the termination area until it is blocked by the Tus/*terC* barrier. If a fork blocked by *terC* is restarted, it will proceed to *terB*.

As seen previously, there is a peak of over-replication in the native termination area in the double origin ectopic *ter* strain in the absence of RecG. As has previously been reported, there is a significant increase to this amplification of the termination region in *ΔrecG* cells that have two copies of the origin compared with that seen in the single origin *ΔrecG* strain (cf. Figure 8, panel II and Figure 15, panel III; Rudolph et al., 2013).

As we suspected based on our data thus far, when *recG* is deleted from *oriC*$^{+}$ *oriZ*$^{+}$ cells and an ectopic replication fork trap is created in these cells (*terA4.44 terA4.57*), a peak of over-replication is seen in the ectopic fork fusion region (Figure 15, panel III). This peak of over-replication in the ectopic termination area is not seen when RecG is present (Figure 15, panel II). These data strongly support the idea that over-replication initiation in *ΔrecG* cells is co-localised with fork fusion events and not specific to a particular part of the chromosome.

Whilst *terA4.44* clearly forms a block to replication forks moving towards *oriC*, the peak of over-replication in the ectopic termination region is able to extend out beyond the *ter*/Tus block on that side until it encounters *rrn* operons, where RNAP complexes halt progress of the replication forks. Having confirmed via sequencing and functional analysis that the *terA* construct is reliable, it is

possible that this situation is similar to that in the native termination area of double origin strains described above. In single origin *ΔrecG* cells, *terC* is entirely overcome by replication forks initiating in the termination area (discussed below; Figure 17).

When the replication fork traps are inactivated by the deletion of *tus*, the peaks of origin-independent synthesis in both the native and ectopic termination areas disappear. The valleys between the peaks are noticeably shallower compared with *recG*$^+$ strains, as reported previously (Rudolph et al., 2013). Overlaying the *oriC*$^+$ *oriZ*$^+$ *Δtus* profile on to the *oriC*$^+$ *oriZ*$^+$ *Δtus ΔrecG* profile, aligned using origin peaks (Figure 15, panel VI), confirms this and is consistent with bidirectional origin-independent synthesis initiating in *ΔrecG* cells in the absence of forks being blocked at *ter*/Tus traps. Growth experiments were carried out to determine the doubling times of *oriC*$^+$ *oriZ*$^+$ *Δtus* cells in the presence and absence of RecG and confirmed that they are very similar (data not shown; Midgley-Smith et al., 2018). A similar doubling time strongly suggests that the frequency of origin firing is unlikely to be affected and if this is the case, the alignment via peak height is justified (Midgley-Smith et al., 2018).

Deletion of terC does not shift the low point of the peak of over-replication

In cells lacking RecG, marker frequency analysis reveals a defined peak of over-replication within the termination region of exponentially growing cells (Figure 17, panel IV). The low point on the right-hand side of the peak is at *terA*, the innermost *ter* site of the right hand replichore, but not at *terC*, the innermost *ter* site of the left hand replichore (Figure 16). Instead, the low point is at *terB*, the second innermost *ter* site of the left hand replichore, as was recently independently confirmed (Azeroglu et al., 2016). This was unexpected. Functional analysis of *terC* via 2-D gels confirmed that in principle, Tus protein is able to bind and form a block to replication forks (Duggin and Bell, 2009). It is possible that the *terC* construct within our bacterial strains might have a point mutation that inactivates the ability of *terC* complexed with Tus to form a block to replication forks. However, we have recently reported data demonstrating that *terC* does indeed represent a strong block to replication fork progression (Dimude et al., 2016; Ivanova et al., 2015) and so there is little doubt that *terC* is fully functional in the strains used in this study, which is also seen in recent work from Maduike and colleagues, who also use derivatives of MG1655 in their work (Maduike et al., 2014). The fact that *terC* is functional in wild type cells led to the question of what was happening to the Tus/*ter* trap at *terC* in the *ΔrecG* profile. As stated above, recent work by (Moolman et al., 2016) demonstrates that replisomes are not halted indefinitely at Tus/*terC* and can overcome the block. However, this is not the explanation in this case, as without even a distortion at the *terC* location in *ΔrecG* cells, which

would represent forks being stalled at a *ter*/Tus trap, there is no indication that forks get blocked at Tus/*terC* at all.

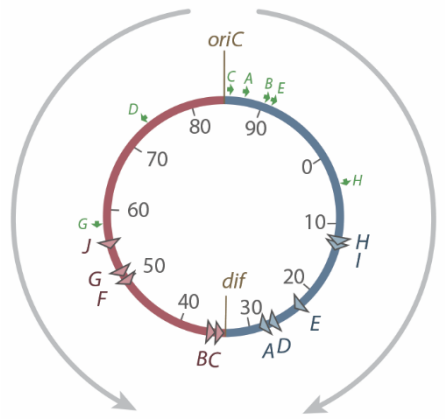

Figure 16: Schematic representation of the *E. coli* chromosome. The *E. coli* chromosome is replicated by two replication forks moving in opposite directions from the origin, *oriC*. The *ter* sites (triangles and labelled A-J) of the left hand (red) and right hand (blue) replichores are orientated to allow replication forks (grey arrows) to move from *oriC* towards the termination region between *terA* and *terC*. The chromosome dimer resolution site *dif* within this region is also labelled.

Tus protein is displaced from a *ter* site if a replisome arrives from the permissive orientation. In order for a replication fork to pass a *ter* site in the opposite direction to normal, Tus protein must not be bound. It is possible that a replication fork arrives at *terC* from the origin and displaces Tus protein. It meets a fork that has been initiated in the termination area and held at *terC* and new forks are established in each direction, one further in to the termination area and the other towards *terB*. Due to the number of proteins in the vicinity at the time, Tus protein would simply have no space to gain access and bind to *terC* and so there is no block until *terB*. This could potentially explain the perceived failure of Tus/*terC* to block replication forks from leaving the termination area in *ΔrecG* cells (Figure 17). Whilst this might provide a temptingly simple explanation for the situation at *terC*, it is not clear why the same would then not be seen at the other end of the replication fork trap; replication profiles of *ΔrecG* cells indicate that *terA* represents a strong block to replication forks and prevents progression out of the termination region (Figure 17).

If *terC* is deleted in a *ΔrecG* strain, does the low point of the peak of over-replication shift towards *terF*? According to the hypothesis above, in *ΔrecG ΔterC* cells, a fork established in the termination area would then arrive first at *terB*, which, at 75 kb away (Duggin and Bell, 2009), is relatively close to the location of *terC*. This fork would be blocked until a replication fork arrives from the origin, displaces Tus, and establishes two new replication forks again with one moving further in to the termination area and the other now free to move to *terF*.

To investigate this possibility, the one-step inactivation method developed by Datsenko and Wanner (page 50; Datsenko and Wanner, 2000) was used to delete *terC* from a wild type MG1655 strain background. *ΔrecG* was then moved into the resulting *ΔterC* strain via P1*vir* transduction. Marker frequency analysis by deep sequencing was performed and replication profiles were

determined for *ΔterC* cells in the presence and absence of RecG. The data for the replication profile of *ΔrecG* single mutant cells shown in Figure 17, panel IV was generated from a different sequencing run to the other strains. The information required from this control is only concerned with demonstrating that the low point of the peak of over-replication in *ΔrecG* cells is at *terB* and not *terC* and other characteristics such as relative peak height are unimportant and so in this instance, it is acceptable to use this profile alongside the others.

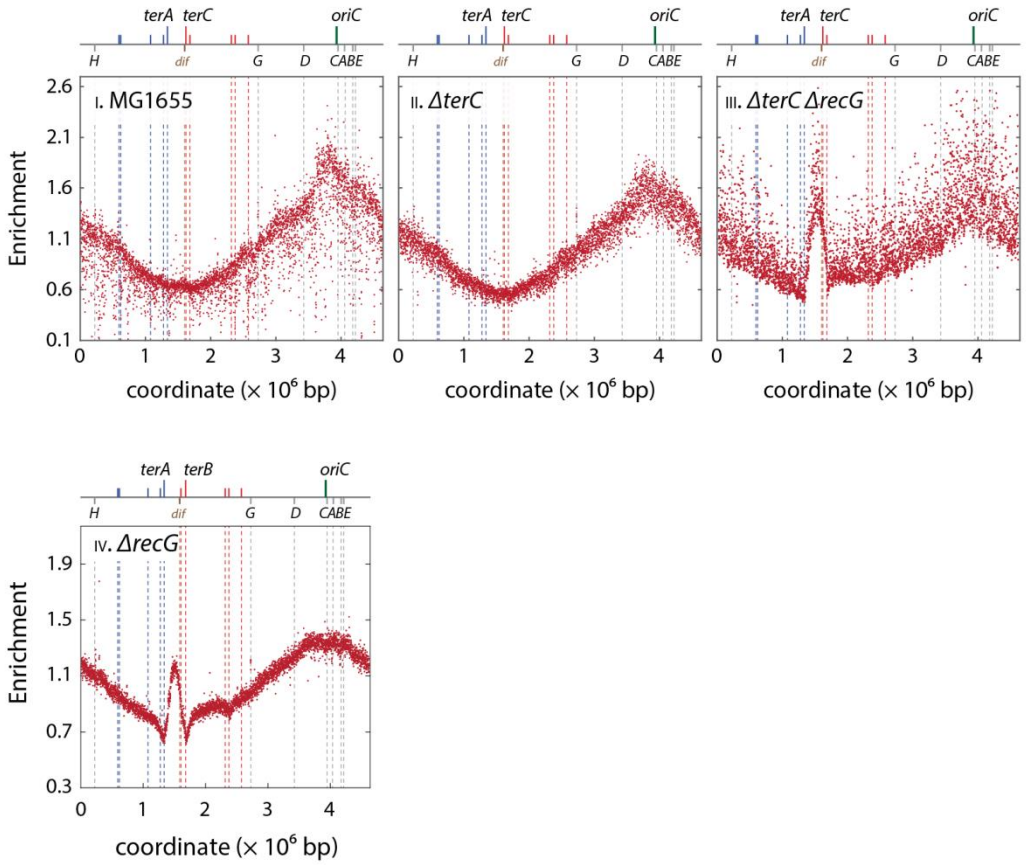

Figure 17: Replication profiles of *E. coli* cells showing the effect of Δ*ter*C on the peak of over-replication in cells lacking RecG. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *ori*C (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven rrn operons and the chromosome dimer resolution site *dif* below. The data for the Δ*rec*G profile (IV) were generated from a different sequencing run and are plotted on a different axis to the other profiles. Sequencing templates were isolated from MG1655 (wild type), SLM1134 (Δ*ter*C), SLM1140 (Δ*ter*C Δ*rec*G) and N4560 (Δ*rec*G). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

Whilst there is some noise in the replication profiles we have established for the *ΔterC* derivatives, the major features are clear (Figure 17).The profile of the *ΔterC* single mutant appears in essence to be the same as the wild type profile (Figure 17, panels I and II), indicating that the absence of a Tus/*ter* block at *terC* does not cause any major problems for the cells.

The replication profile of *ΔterC ΔrecG* cells reveals that the low point of the peak remains at *terB* even in the absence of *terC*. One possible explanation for this observation might be that a Tus/*terC* complex is overcome more easily than the Tus/*terB* complex. However, this explanation is rather unlikely, as in other contexts a Tus/*terC* complex is a very efficient block for a proceeding replisome (Dimude et al., 2018a; Duggin and Bell, 2009; Ivanova et al., 2015; Moreau and Schaeffer, 2012). Instead, it seems more likely that the absence of any indication of a block to replication fork progression at *terC* is not due to displacement of Tus from *terC* caused by initiation of replication forks at the Tus/*terC* complex via origin-independent DNA synthesis. If this was the case, there would be sufficient time in between rounds of chromosome replication in *ΔterC* cells for replication forks established via origin-independent initiation to proceed to Tus/*terB* instead, and the whole scenario would be shifted to the right in the replication profile of *ΔterC ΔrecG* cells. This is not the case.

The *ter* sites in *E. coli* are not evenly distributed either side of the midpoint of the chromosome (i.e. the location directly opposite *oriC* in the context of a circular chromosomal diagram); the midpoint is instead near to *terC* (Figure 16; Ivanova et al., 2015). This means that the two replichores are slightly asymmetrical in that forks approaching the termination area from *oriC* will on average arrive at the termination area at *terA* (right-hand fork) before arriving at *terC* (left-hand fork). This is reflected in the replication profiles of wild type cells where the low point of the profiles is near *terC* and not at the midpoint between the innermost *ter* sites (Figure 8, panel I; note that the right-hand replichore is shown on the left hand side of a replication profile with respect to the termination area and vice versa). In the profile of a *ΔrecG* single mutant, the low points either side of the peak of over-replication are approximately symmetrical whereas in the *ΔterC ΔrecG* strain, the profile is higher on the right of the peak compared to the left and so the chromosomal sequences represented there are present at a higher frequency on a population level than those to the left of the peak. It is unlikely that *ΔterC* would have any affect that allows forks to reach the termination region from the origin faster than in *terC*$^{+}$ cells, and so it is possible that this is an effect of *ΔterC* on forks that originate in the termination region via origin-independent synthesis. As seen in Figure 17 panel IV, this increase in marker frequency extending from *terB* to the next fork traps at *terF/G* is seen in replication profiles of *ΔrecG* single mutants however and so we are unable to attribute the differences seen between the profiles to the absence of *terC*. Taken together with the fact that the same situation is not seen at *terA*, which forms a strong block to replication fork progression out of the termination region, my data confirm that the activity resulting from origin-independent over-replication within the termination region does not provide an explanation for the observed lack of a replication fork block at *terC* in *ΔrecG* cells and indicate instead that the situation is more complex.

Discussion

In the absence of either RecG or RNase HI, DNA synthesis is able to initiate independently of both the initiator protein DnaA and the origin of replication, *oriC*. Both RecG and RNase HI can remove RNA:DNA hybrids *in vitro* (Fukuoh et al., 1997; Tadokoro and Kanaya, 2009; Vincent et al., 1996), which led to the idea that the over-replication in both *ΔrnhA* and *ΔrecG* cells might arise via a common underlying mechanism enabled by the persistence of R-loops (Hong et al., 1995; Kogoma, 1997). Our data show that despite this apparent similarity however, there are a number of significant differences between the over-replication initiating in *ΔrecG* cells and that seen in *ΔrnhA* cells. Whilst the data presented in this chapter support that origin-independent synthesis indeed arises via R-loops in the absence of RNase HI, they do not support R-loop involvement in the origin-independent synthesis seen in *ΔrecG* cells.

We recently reported that the genetic requirements for origin-independent over-replication in the termination area differ in the absence of RecG or RNase HI (Dimude et al., 2015). Different substrates are exploited by PriA to initiate origin-independent synthesis. We found that, whereas the helicase activity of PriA, and specifically the ability to unwind a forked structure equivalent to a 3' flap, was essential for origin-independent replication in the termination region of *ΔrecG* cells (Rudolph et al., 2013), this specific substrate activity of PriA is not essential for origin-independent synthesis in cells lacking RNase HI (Dimude et al., 2015). The effect of chromosome linearisation near to *dif* on the origin-independent synthesis seen in the termination region differed between *ΔrecG* and *ΔrnhA* cells. This manifested in a striking difference in the ability of *dnaA Δtus rpo** cells with a linearised chromosome to grow in the absence of origin firing; cells lacking RecG were unable to grow at all whereas there was no detectable effect on the growth of cells lacking RNase HI, which grew robustly at 42 °C (Dimude et al., 2015; Rudolph et al., 2013). These data suggest that different mechanisms are responsible for the origin-independent synthesis in *ΔrecG* or *ΔrnhA* cells.

This is supported by the fact that the RNase H1 protein from *S. cerevisiae* can compensate for *ΔrnhA* and reduces growth in the absence of origin firing over 20-fold but has no effect on the growth of *dnaA ΔrecG Δtus rpo** cells at restrictive temperature (Figure 12). Given that RNase H1 cleaves the RNA from RNA:DNA hybrids (Cerritelli and Crouch, 2009), this result strongly suggests that in contrast to the situation in *ΔrnhA* strains, R-loops are not important for the majority of origin-independent DNA synthesis in the absence of RecG. The data in Figure 12 suggest that the effects of *S. cerevisiae* RNase H1 and *E. coli* RecG are specific; there is no obvious complementation between the two proteins.

The imaging flow cytometry analysis of EdU incorporation in to newly synthesised DNA provides insight in to the frequency of origin-independent over-replication seen in cells lacking RecG or RNase HI on a population basis, but also how that synthesis is distributed among the individual cells within a culture (Figure 11). The data show very clear differences in the fluorescent signal detected in the absence of either protein. In contrast to *dnaA* single mutants, the majority of *dnaA ΔrnhA* cells show

either robust signal along the entire length of the cell or multiple intense spots, indicating high levels of EdU incorporation and therefore high levels of DNA synthesis taking place at multiple chromosome locations even in the absence of origin firing (Figure 11). In contrast, there is very little or no fluorescent signal in the majority (71 %) of *dnaA ΔrecG* cells and instead they appear like the *dnaA* single mutant, indicating that in these cells, DNA synthesis is not able to occur 90 minutes after *oriC* firing has been inhibited (Figure 11).

The imaging flow cytometry data for both *dnaA ΔrecG* and *dnaA ΔrnhA* cells are in line with the growth capability seen in the absence of each protein when origin firing is prevented through the use of DnaA(ts). *dnaA ΔrnhA* cells are able to form colonies at restrictive temperature, especially on minimal media, whereas *dnaA ΔrecG* cells cannot (Figure 10) (Dimude et al., 2015; Rudolph et al., 2013). Significant levels of origin-independent DNA synthesis must be occurring in *dnaA ΔrnhA* cells to support the growth at restrictive temperature and this is exactly what we see in the EdU incorporation assay. *dnaA ΔrecG* cells do not grow in the absence of origin firing, consistent with the lack of fluorescence seen in the majority of these cells when probed for origin-independent DNA synthesis. This also means that the significant level of DNA synthesis implied by the intense fluorescent signal in 29 % of *dnaA ΔrecG* cells is not able to sustain growth (Figure 10, Figure 11). It is possible that this synthesis arises as a result of spontaneous damage to DNA in these cells. It has been shown previously that UV-irradiated *ΔrecG* cells exhibit significantly increased levels of DnaA-independent synthesis, which arises as a result of UV-induced damage to the DNA (Rudolph et al., 2009a) and as spontaneous DNA damage is well known to occur under normal growth conditions even in the absence of exogenous insult (Lindahl, 1993), this provides a reasonable explanation for the high levels of synthesis seen in a subset of the population, as indicated by the intense fluorescent signal of these cells.

In contrast to the situation in the absence of RNase HI, origin-independent synthesis in *dnaA ΔrecG* cells is only able to sustain growth at the restrictive temperature when the replication fork trap is inactivated by the introduction of a *tus* deletion (Dimude et al., 2015; Rudolph et al., 2013). This allows the over-replication within the termination area to progress out of this region, and it is robust enough to duplicate the whole chromosome and support cell growth, especially when the impact of replication-transcription encounters is reduced by the introduction of the *rpoB*35* allele of RNA polymerase (Figure 10) (Dutta et al., 2011; Rudolph et al., 2013). These findings strongly suggest that whatever generates the over-replication takes place within the termination region in *ΔrecG* cells, which is different in cells lacking RNase HI (Dimude et al., 2015; Rudolph et al., 2013).

As a result of the data presented here, taken together with the rest of the study (Dimude et al., 2015), it is clear that there are significant differences in the occurrence of and genetic requirements for origin-independent synthesis in the absence of either RecG or RNase HI. Whilst we cannot exclude that RecG might have a role in removal of R-loops, our data suggest that R-loops contribute very little to the over-replication observed in *ΔrecG* cells.

Instead there is a growing body of work that implicates RecG in a role in managing events at the fusion of two replication forks, preventing re-replication of the chromosome by processing 3' flap structures that arise as a result of replication fork fusion events (Lloyd and Rudolph, 2016; Rudolph et al., 2009a, 2009b, 2010b, 2013). One aspect of this work was to generate replication profiles of strains carrying two copies of the replication origin in the presence and absence of RecG. Replication forks initiated bi-directionally at each origin will move towards each other resulting in two forks meeting within the fork trap in the native termination region whereas the other two forks will meet in an ectopic location between the two origins. The replication profiles revealed that the region between the two origins that becomes an ectopic fork fusion location is shallower and less defined in the *ΔrecG* derivative, consistent with an amplification of a broad region of the chromosome (Figure 15) (Midgley-Smith et al., 2018; Rudolph et al., 2013). If origin-independent over-replication is initiated at fork fusion events in the absence of RecG, this is exactly what would be expected. It was reasoned that the lack of a defined peak like that seen in the native termination region was due to the fact that there are no Tus-*ter* traps present to contain the over-replication and so it is able to spread out from the initiation location and amplify a larger section of the chromosome (Rudolph et al., 2013). I developed this work further by inserting *ter* sequences between the two origins to create a replication fork trap, thereby creating a proper ectopic termination area. As expected, the presence of functional Tus/*ter* fork traps in *oriC*$^{+}$ *oriZ*$^{+}$ cells resulted in a defined peak of over-replication in the ectopic fork fusion region in the *ΔrecG* derivative that is not seen when RecG is present (Figure 15) (Midgley-Smith et al., 2018). This data strongly supports the idea that the over-replication that arises in cells lacking RecG initiates as a result of events at fork fusions. The logical follow up control that would confirm this idea would be to generate a single origin derivative (*oriC*$^{+}$ *terA4.44 terA4.57 ΔrecG* or *oriZ*$^{+}$ *terA4.44 terA4.57 ΔrecG*) to show that in the absence of fork fusion events, the peak disappears. However, this is of course not possible, as without a functional replication origin on either side of the ectopic fork trap, the Tus/*ter* complexes would prevent replication of a large region of the chromosome.

We have suggested that as two replication forks converge at replication termination, the helicase of one replisome might displace the leading-strand polymerase of the opposing replisome and unwind the newly created duplex DNA. This would result in the formation of a 3' flap structure, which might normally be processed by RecG or degraded by 3' exonucleases, but in the absence of either, PriA is able to exploit these structures to re-initiate replication. D-loop recombination intermediates are formed via the recombinogenic substrates generated by the newly established replication forks, which are also substrates for PriA replication restart, and results in bidirectional replication.

Despite that fact that a number of observations can be explained by this model and a growing body of data that supports the idea of fork fusions having the potential to result in over-replication of the chromosome, the molecular details of what happens within the termination region are largely unknown. Our hypothesis that the low point of the peak of over-replication in *ΔrecG* cells would be shifted from *terB* towards *terF* in the absence of *terC* was based on our model of over-replication at

fork fusion events. If over-replication is initiated as a result of replication fork fusions, newly established replication forks within the termination area will progress until they are stopped by Tus/*ter* complexes at *terC* and *terA*. Blocked forks will be met by replication forks initiated at *oriC* and so secondary initiation events are expected to occur in this scenario. We hypothesised the sheer number of proteins involved might prevent access of Tus to the DNA at Tus/*terC*, therefore allowing replication forks that are established there to proceed to Tus/*terB*, where they are then blocked. The data presented here show that even in the absence of *terC*, the position of the low point of the peak of over-replication in *ΔrecG* cells was unaffected (Figure 17). Whilst the data do not invalidate any part of the model, they do suggest that the situation in the termination region is more complex than is currently understood. The fact that the right-hand low point of the peak of over-replication was not altered by the deletion of *terC* was to some degree unsurprising, as the reasoning used to develop the hypothesis could not explain why the same situation was not seen at *terA*. Further investigation is needed to unravel the molecular details of the over-replication in *ΔrecG* cells, and by extension, the details of fork fusion events during normal replication termination. It would be of interest to determine in the first instance what happens to the shape of the peak of over-replication in *ΔrecG ΔterB* and *ΔrecG ΔterB ΔterC* cells. In this instance, the *tus* gene would need to be expressed from a different location because *terB* overlaps with *tus*. Furthermore, a similar situation is seen in the ectopic fork fusion location of *oriC⁺ oriZ⁺ ΔrecG* cells in which an ectopic fork trap has been created (Figure 15). In contrast to the situation at *terC*, the distortion in the replication profile at the location of the Tus/*terA4.44* complex in *oriC⁺ oriZ⁺ terA4.44 terA4.57 ΔrecG* cells indicates that there is a block to fork progression but that the over-replication is on occasion able to proceed further towards *oriC*. This could be investigated further by introducing double *ter* sites to create the ectopic replication fork trap, which would also more closely resemble the situation found at the native termination region.

The observation that cells lacking both RecG and the Holliday junction resolvase RuvABC show a synergistic sensitivity to DNA damage (Lloyd, 1991) has led to the suggestion that both RecG and RuvABC might have overlapping functionalities. The reduced recovery of conjugational recombinants in *Δruv ΔrecG* double mutants further contributed to this idea (Lloyd, 1991). More recently, the focus has been on a role for RecG at replication termination. So does RecG have a function in the resolution of recombination intermediates? Such a role certainly cannot be ruled out categorically. However, the finding that over-replication in cells lacking RecG depends on the presence of both RecA and RecBCD (Midgley-Smith et al., 2018; Rudolph et al., 2013) strongly suggests that in the absence of RecG, fork fusion intermediates lead to increased levels of recombination. In contrast, in cells lacking both RecG and RuvABC the over-replication observed in the termination area of *ΔrecG* cells remains unchanged. However, while *dnaA Δtus rpo* ΔrecG* cells grow robustly at restrictive temperature (Figure 10), growth is abolished if *ruv* is deleted in addition (Midgley-Smith et al., 2018; Rudolph et al., 2013). Given that in *ΔrecG Δruv* cells over-replication is still detected in the termination area (Midgley-Smith et al., 2018) the most likely explanation for this observation is that the over-replication initiated at both fork fusion intermediates and subsequently at recombination intermediates results in an

increased number of recombination intermediates, which prevent the segregation of the chromosomes into daughter cells. Thus, recombination intermediates are still formed, but if they cannot be resolved because of a lack of Holliday junction resolution, the cells affected become inviable, which might explain the observed reduced recovery of conjugational crosses in *Δruv ΔrecG* cells (Lloyd, 1991). In line with this hypothesis, it was reported recently that recombination levels can be significantly increased in *Δruv ΔrecG* cells if the RecFOR recombination system is deactivated (Lloyd and Rudolph 2016). Indeed, in *Δruv ΔrecG ΔrecF priA300* cells the synergism between *recG* and *ruv* is virtually eliminated, suggesting that RecG has little to do with providing an alternative pathway for the resolution of recombination intermediates (Lloyd and Rudolph, 2016).

The data presented here show that in *ΔrecG* cells carrying a second copy of the origin, the peak of over-replication is dramatically increased (Figure 15), as reported previously (Midgley-Smith et al., 2018; Rudolph et al., 2013). The over-replication cannot be due to forks being held specifically at Tus/*ter* complexes as growth of *ΔrecG* cells in the absence of origin firing depends specifically on the absence of Tus, not its presence (Figure 10) (Rudolph et al., 2013) and in line with this, we recently reported data suggesting that there are increased numbers of active replication forks in *ΔrecG* and *oriC⁺ oriZ⁺ ΔrecG* cells if *tus* is deleted (Midgley-Smith et al., 2018). Instead, we suggest that the increased peak height is related to the period of time that a fork is held rather than the fact that they are blocked (Midgley-Smith et al., 2018). The presence of the ectopic replication origin will result in clockwise forks arriving at the termination area much sooner than anticlockwise forks coming from *oriC*, and so forks will be held at *terC/B* for a considerable period of time before a fork fusion event takes place. Reports on the stability of a replisome at obstacles including nucleoprotein roadblocks suggest a limited half-life of 4–6 min (Marians et al., 1998; McGlynn and Guy, 2008; Mettrick and Grainge, 2016), which suggests that after a relatively limited period of time, the replication fork complexes might start to disassemble. Forks permanently stalled at obstacles are processed by recombination proteins RecBCD (Dimude et al., 2018b; Michel and Leach, 2012) and so the opposing fork is likely to meet a complex recombination intermediate. If RecG has a role in managing recombination, as the result of the tandem repeat deletion assay suggests (page 89), and in doing so, affects the ability of PriA to initiate replication at these intermediates, then it would be expected that the over-replication seen in the absence of RecG would increase in an *oriC⁺ oriZ⁺* background where forks are stalled for an extended period of time.

Origin-independent over-replication and genomic instability

Recent work has shown that in the absence of RecG, origin-independent over-replication occurs specifically in the termination area of the chromosome, and that this replication is robust enough to sustain cell growth independent of replication initiated via both the normal origin, *oriC,* and the initiator protein DnaA (page 68) (Dimude et al., 2015; Rudolph et al., 2013). In the absence of RecA or RecBCD recombinase activities, the over-replication is abolished (Midgley-Smith et al., 2018; Rudolph et al., 2013), indicating that it is dependent on recombination. A model to explain how the over-replication might arise as a result of a replication fork fusion event via the formation and subsequent exploitation of a 3' ssDNA flap has been developed (Dimude et al., 2015; Lloyd and Rudolph, 2016; Rudolph et al., 2010a, 2013). PriA-mediated replication re-start at the 3' flap (Windgassen et al., 2018) converts it to a dsDNA flap, which is then a substrate for RecBCD to process and initiate homologous recombination (Kowalczykowski, 2000). This results in a further substrate for PriA to establish another replication fork that moves in the opposite direction (see page 31 for more detail).

Whilst recombination is an essential process for DNA repair and successful DNA replication (Cox et al., 2000; Cromie et al., 2001), it can come at a cost to genome integrity, and increased levels of recombination lead to destabilisation of the genome. This can have pathological consequences, and in humans, genomic instability is a prominent feature of cancers (Negrini et al., 2010; Shen, 2011) and the cause of a number of diseases and genetic disorders (Aguilera and Gómez-González, 2008; Castel et al., 2010; Hou et al., 2017; Sasaki et al., 2010). It has been shown that aberrant homologous recombination can be a cause of genomic instability (Shammas et al., 2009). If the over-replication seen in *E. coli* cells in the absence of RecG is triggered by an increased frequency of recombination, then it has the potential to destabilise the genome.

If replication termination triggers recombination events, we should be able to measure that the recombination frequency is increased. Marker frequency analyses of exponentially growing cells lacking RecG show a peak of over-replication specifically within the termination region (Dimude et al., 2015; Rudolph et al., 2013; Wendel et al., 2014), and so we investigated recombination frequency in this location and in a second, control location away from the termination region. It has been reported previously that the rate of recombination can be altered in the absence of RecG. Lovett and colleagues showed that *ΔrecG* cells exhibit increased genomic instability using an assay measuring deletion of a tandem repeated DNA sequence from both a plasmid- or chromosome-based repeat construct; *ΔrecG* cells show an increased deletion rate compared to wild type cells (Lovett, 2006; Lovett et al., 1993). Similar results in the reversion rate of a tandem repeat sequence using a plasmid-based assay were generated by Lloyd and Rudolph (2016). We decided to use a similar assay for our investigation. If

there is a correlation between replication termination and recombination frequency, we would expect to see a location-specific increase in recombination specifically in the absence of RecG, with the increase in the termination region being higher than the increase at the control location assayed. This would be in line with our model, which is based around the idea that the majority of over-replication seen in the absence of RecG results from events associated with replication fork fusions, and so supports the prediction of an increased frequency of recombination in the termination region.

A tandem repeat cassette termed *KanKanMX4* (Ede et al., 2011) located on a plasmid similarly showed increased reversion rates in the absence of RecG (Lloyd and Rudolph, 2016). The *KanKanMX4* cassette contains a gene for kanamycin resistance that has an internal duplication of a 266 bp fragment of the gene (Figure 18; Ede et al., 2011), which inactivates the resistance gene and cells carrying this cassette are susceptible to the antibiotic kanamycin.

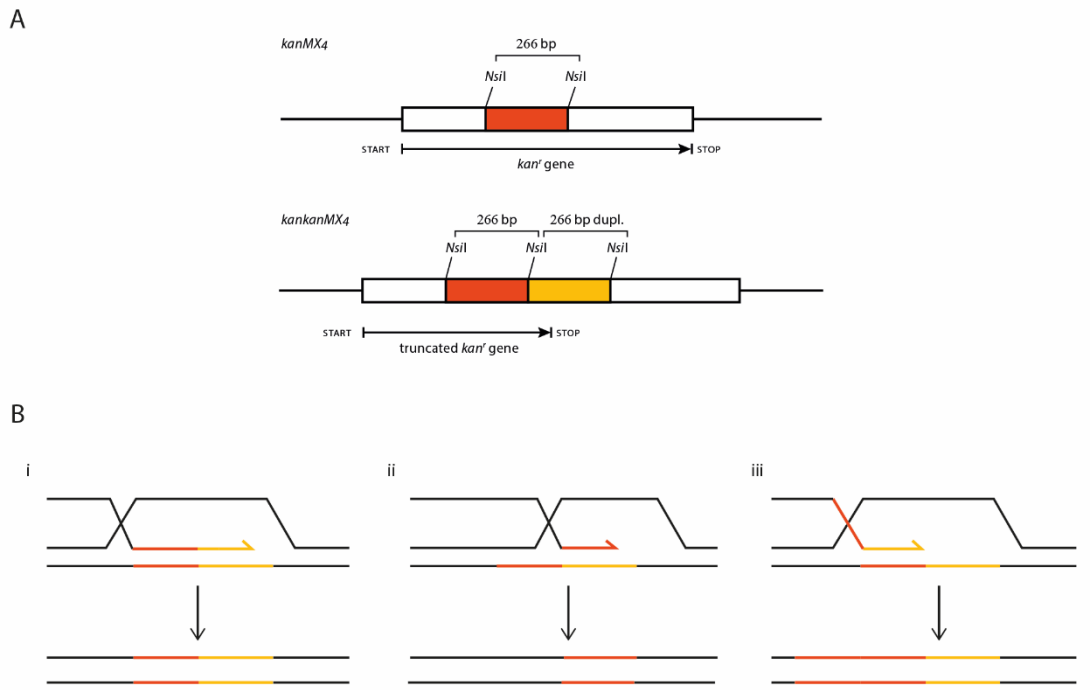

Figure 18: The KanKanMX4 tandem repeat cassette. **A)** Diagram of the KanMX4 and KanKanMX4 module showing the duplication within the kanamycin resistance gene, represented by the coloured boxes. The duplication causes a frameshift and introduces a stop codon in the second copy of the repeat, resulting in a gene expression and a non-functional kanamycin resistance protein. **B)** Schematic representation of gene conversion events occurring via homologous recombination. The duplication is represented by the coloured sections as in A). i. If replication or end processing is stopped within the second copy of the repeat and invasion of the second copy on the sister chromosome takes place, the duplication is maintained. ii. If it stops in the first copy of the repeat, invasion into the first copy on the sister chromosome will also result in the duplication being maintained but if the invasion occurs in the second copy on the sister chromosome, one of the repeats is lost and functional kanamycin resistance is restored. iii. If replication or end processing stops in the second repeat but the ssDNA invades the first repeat of the sister chromosome, a triplication is generated. Only the outcome of ii) can be detected by our assay.

Gene function is restored when one of the two repeats is deleted, and this event is quantified experimentally by scoring kanamycin-resistant colonies. This deletion event can be mediated via homologous recombination or by recombination-independent misalignment errors during replication

that are associated with repeat sequences such as replication slippage errors (Bzymek and Lovett, 2001; Lovett, 2004). I investigated whether it was possible to use the *KanKanMX4* construct to detect different reversion rate increases when it was inserted in to two different locations of the chromosome. The work by Lloyd and Rudolph (2016) on reversion rates of the plasmid-based *KanKanMX4* tandem repeat cassette revealed that the increase seen in the absence of RecG is abolished in *ΔrecG ΔrecA* and *ΔrecG ΔrecB* double mutants, indicating that the reversion events in *ΔrecG* cells are mediated by recombination (Lloyd and Rudolph, 2016), further confirming its suitability for our aim.

It has been shown that RecA and RecBCD are required for both the origin-independent over-replication and origin-independent growth seen in the absence of RecG (Midgley-Smith et al., 2018; Rudolph et al., 2013). It was recently reported that in contrast, the RuvABC Holliday junction resolvase is dispensable for the peak of over-replication in the replication profiles of *ΔrecG* cells but that growth of *dnaA Δtus rpo* ΔrecG* cells is abolished at restrictive temperature in the absence of RuvABC, indicating that RuvABC is important for segregation of the replicated chromosomes (Midgley-Smith et al., 2018). The tandem repeat deletion rates determined by Lloyd and Rudolph are in line with these data; the increase in the reversion rate in the absence of RecG depends on RecA and RecB but not Ruv (Lloyd and Rudolph, 2016).

Inserting a tandem repeat cassette at two chromosomal locations

The *KanKanMX4* module was originally cloned in to the shuttle vector pRS316 (Sikorski and Hieter, 1989), generating pRS316-*KanKanMX4* (Ede et al., 2011). Given that I wanted to investigate the effect that specific chromosome location has on the reversion rate of the construct, it was not possible to use a plasmid-based assay and instead the tandem repeat cassette had to be integrated in to the chromosome. To achieve this, it was necessary to first link it to a selectable marker; in this case, a gene that conferred resistance to chloramphenicol (*cat*) flanked by *frt* sites (<*cat*>) was used. The <*cat*> construct was amplified by PCR from the synthetic *oriZ*-<*cat*> construct present in the chromosome of TB28 and cloned in to pRS316-*KanKanMX4*, resulting in plasmid pSLM001. This was subsequently used as the template for the PCR-amplification of the *KanKanMX4*-<*cat*> cassette, using primers that carry 50 bp ends that contain homology to two different regions of the chromosome; one near *narU* in the termination area and a second, control region near *yjhR* at 4.53 Mb of the *E. coli* chromosome. The insert locations were selected to minimise the risk of interfering with expression of local genes. In the termination region, the insertion was located after the stop codon of one gene and well before the start of the next gene, and the control location insertion was between two stop codons (Figure 19 A and C respectively). The resulting PCR products were integrated separately on to the chromosome of wild type cells using the lambda Red recombination system (Datsenko and Wanner, 2000). Colony PCR was employed to verify the correct insert at each chromosomal location (Figure 19).

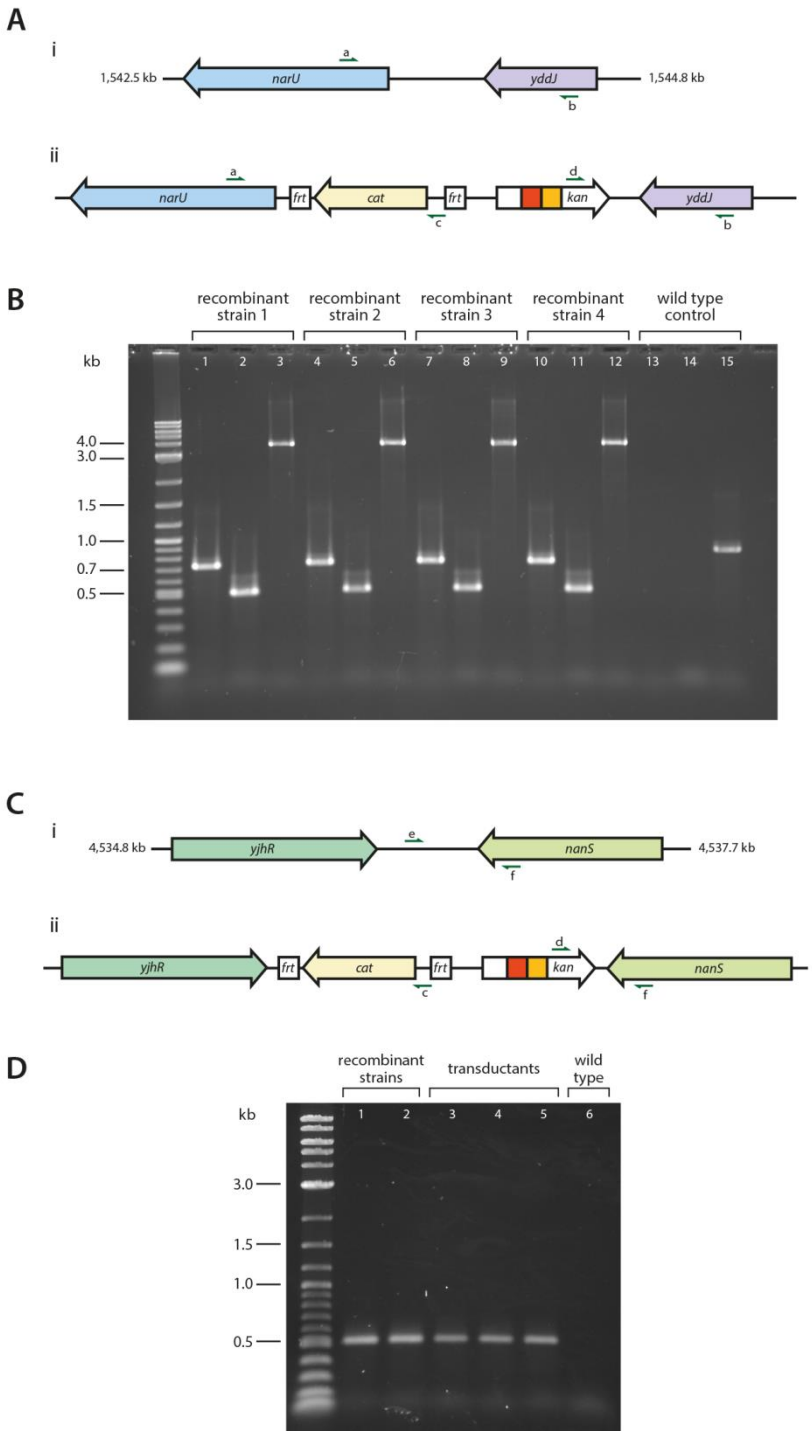

Figure 19: Confirmation of successful integration of the *KanKanMX4* cassette at two chromosomal locations. Schematic representation is shown of the area around *narU* within the termination area (**A**i) and around *yjhR* at the control location (**C**) with and without the integrated *KanKanMX4-<cat>* construct. The verification primers are shown as green arrows. PCR products generated with the primers for the termination area (**B**) are shown for four recombinant strains and a wild type control. Primer combinations used were a/c (lanes 1, 4, 7, 10 and 13), d/b (lanes 2, 5, 8, 11, and 14) and a/b (lanes 3, 6, 9, 12 and 15). Colony PCR products were run on 1 % agarose 1 × TBE gels with a gradient voltage of ~6 V/cm. 2-log ladder (NEB) was run for size reference. PCR products generated with the primers for the control region (**D**) are shown. The *KanKanMX4-<cat>* constructs in two recombineering clones were transduced in to wild type cells, producing three transductant strains. Chromosomal DNA from the original recombineering clones, three transductant strains and wild type control was used as the template DNA. Problems with the integration initially were resolved by altering the integration strategy slightly. The initial verification primer (e) was

lost and so we could not test the left-hand flank of the integration. We confirmed integration of the construct using primer combination d/f, in addition to observing the expected antibiotic resistance pattern. Colony PCR products were run on 1 % agarose 1 × TBE gels with a gradient voltage of ~6 V/cm. 2-log ladder (NEB) was run for size reference.

P1 liquid culture lysates of the correct recombinant strains were prepared and used to transfer the *KanKanMX4-<cat>* construct in to wild type cells at either the termination area or the control region in order to avoid taking forward any aberrant recombination events that may have occurred during the recombineering process. Subsequently *ΔrecG* was moved in to each of the strains via P1*vir* transduction, resulting in *KanKanMX4narU ΔrecG* and *KanKanMX4yjhR ΔrecG*.

Reversion rate at two distinct locations of the chromosome

To determine if there is an increase in the rate of recombination, it is not so simple as counting the number of reversion mutants and presenting that as a fraction of the total colony forming units (cfu) of the culture, as is highlighted by the diagram below (Figure 20). A recombination event, leading to the reversion of the *KanKanMX4* module to a functional kanamycin resistance gene, has happened after three divisions in Culture A in Figure 20. The red cell is kanamycin resistant. There is a reversion event much earlier in Culture B and again after three divisions (Figure 20). Although there are now eight reversion mutants in total in this culture, there have only been two reversion events.

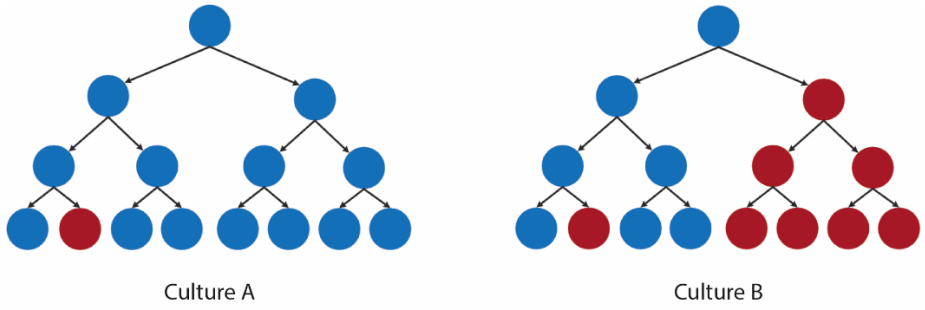

Figure 20: Schematic diagram explaining mutation rate versus number of mutants. Circles represent bacterial cells, which divide and result in two daughter cells. The blue circles are non-mutant cells and the red circles represent a mutant cell that differs genetically from the original parent cell. See the text for details.

Whilst it is easier to measure the frequency of reversion mutants, even across a number of independent cultures of the same strain, it does not take in to consideration that events will, by chance, occur either early or late in some cultures (Rosche and Foster, 2000). Instead, the reversion rate of the *KanKanMX4* construct in the presence and absence of RecG was estimated using a fluctuation assay based on the method of the median by Lea and Coulson (Foster, 2006; Lea and Coulson, 1949), which

utilises a system of statistical calculations that compensates for this eventuality (Lea and Coulson, 1949) (page 60).

It was predicted that the reversion rate of the *KanKanMX4* construct would increase in the absence of RecG, as observed before (Lloyd and Rudolph, 2016; Lovett, 2006; Lovett et al., 1993) and that reversion events would be particularly prevalent in the termination area (*KanKanMX4narU*) in *ΔrecG* cells compared to the control location (*KanKanMX4yjhR*).

In the tandem repeat assay, there will be a certain frequency of kanamycin-resistant cells across the cultures of each strain due to reversion events that occur whether RecG is present or not. Both the mutation and reversion events are extremely susceptible to media composition and other experimental conditions, and so control strains that contained the *KanKanMX4* construct but were otherwise wild type (*KanKanMX4narU* and *KanKanMX4yjhR*) were used to establish a baseline rate of reversion events at each insert location in order to be able to determine if there is an increase in one location or another when RecG was absent. The reversion rate of the control strains is only relevant for that particular batch of media and so within each experiment at least, the media must all come from the same batch. The reversion rate established for each control strain was set to 1 and the reversion rate of the corresponding strain lacking RecG was used to calculate the factor increase in reversion rate compared to the control strain reversion rate.

Table 4: Reversion rates of the *KanKanMX4* in the termination area and the control region in wild type and *ΔrecG* cells. "*Termination area*" and "*control location*" refer to the *KanKanMX4* insert location at *narU* and *yjhR* respectively. The strains used were SLM1042 (*KanKanMX4narU*), SLM1048 (*KanKanMX4narU ΔrecG*), SLM1043 (*KanKanMX4yjhR*) and SLM1049 (*KanKanMX4yjhR ΔrecG*). Experiments were performed at least twice independently, producing very similar results.

Strain	*Reversion rate*	*Factor increase*
Wild type *termination area*	$9.13 \times 10^{-8} \pm 1.18 \times 10^{-8}$	1
ΔrecG termination area	$9.74 \times 10^{-7} \pm 1.02 \times 10^{-7}$	10.67
Wild type *control location*	$1.18 \times 10^{-7} \pm 1.44 \times 10^{-8}$	1
ΔrecG control location	$1.62 \times 10^{-6} \pm 1.63 \times 10^{-7}$	13.76

The reversion rate of the *KanKanMX4* tandem repeat construct to a functional kanamycin resistance gene is very similar in each of the two wild type strains, which means that the reversion rate in wild type cells is not affected by the chromosome location of the tandem repeat construct (Table 4).

In the absence of RecG, the reversion rate of the *KanKanMX4* construct is higher in both *ΔrecG KanKanMX4narU* and *ΔrecG KanKanMX4yjhR* cells when compared to the reversion rate of the corresponding wild type strains (Table 4). This is in line with previous work using tandem repeat

systems, which report an increase in reversion rate in cells lacking RecG compared to the wild type reversion rate (Lloyd and Rudolph, 2016; Lovett, 2006; Lovett et al., 1993).

If there is a correlation between over-replication and recombination frequency, then we should observe a greater increase in reversion rate in the termination region in the absence of RecG compared to the increase seen in the control region in the absence of RecG. However, this is not what was observed. The factor increase in reversion rate at each location was instead very similar (Table 4). If anything, the increase in *ΔrecG KanKanMX4yjhR* cells is slightly higher than the increase seen in *ΔrecG KanKanMX4narU* cells.

Effect of an ectopic origin on reversion rate

Replication profiles determined by high resolution marker frequency analysis of *ΔrecG* cells revealed a defined peak of over-replication in the termination region that is not present in wild type cells (Dimude et al., 2015; Rudolph et al., 2013; Wendel et al., 2014). When a second copy of the replication origin was introduced at an ectopic location roughly a quarter of the way round the right-hand replichore (*oriZ*; Figure 13), there was a dramatic increase in the amplification of the termination region in cells lacking RecG compared to *ΔrecG* cells with a single replication origin (Rudolph et al., 2013). We introduced *oriZ* to our tandem repeat construct strains to see if we would then observe a stronger increase in the reversion rate of *KanKanMX4narU* in the absence of RecG in comparison to the increase in reversion rate of the *KanKanMX4yjhR* construct, in line with what we see in the replication profiles of *oriC⁺ oriZ⁺ ΔrecG* cells. If the data show that the biggest factor increase in reversion rate is seen in *oriC⁺ oriZ⁺ ΔrecG KanKanMX4narU* cells, it would indicate a strong correlation between the over-replication seen in the absence of RecG and recombination.

In double origin strains, replication initiates from both *oriC* and *oriZ* with equal efficiency (Dimude et al., 2018a; Ivanova et al., 2015; Wang et al., 2011) and this results in replication forks meeting in an ectopic region in between the two copies of the origin in addition to the native termination region (Ivanova et al., 2015; Midgley-Smith et al., 2018). The control insert location of the *KanKanMX4* cassette at *yjhR* is within this ectopic fork fusion region. Based on our model that proposes that fork fusion events lead to recombination-dependent over-replication in the absence of RecG, recombination frequency will be increased in this region in *oriC⁺ oriZ⁺ KanKanMX4yjhR* cells and so we would expect to see a greater increase in the reversion rate of *KanKanMX4yjhR ΔrecG* cells in a double origin background compared to the increase seen in the comparative strains in a single origin background.

oriZ⁺ derivatives of the *KanKanMX4* strains were created via P1*vir* transduction and the reversion rates for each *KanKanMX4* insert location were determined for the double origin strains in the presence and absence of RecG.

Table 5: *KanKanMX4* reversion rates in the presence and absence of RecG in strains carrying two copies of the replication origin. "*Termination area*" and "*control location*" refer to the *KanKanMX4* insert location at *narU* and *yjhR* respectively. The strains used were SLM1058 (*oriC*$^+$ *oriZ*$^+$ *KanKanMX4narU*), SLM1059 (*oriC*$^+$ *oriZ*$^+$ *ΔrecG KanKanMX4narU*), SLM1060 (*oriC*$^+$ *oriZ*$^+$ *KanKanMX4yjhR*) and SLM1061 (*oriC*$^+$ *oriZ*$^+$ *ΔrecG KanKanMX4yjhR*). Experiments were performed at least twice independently, producing very similar results.

Strain	*Reversion rate*	*Factor increase*
oriZ$^+$ *termination area*	$2.58 \times 10^{-7} \pm 2.92 \times 10^{-8}$	1
oriZ$^+$ *ΔrecG termination area*	$8.36 \times 10^{-7} \pm 8.88 \times 10^{-8}$	3.24
oriZ$^+$ *control location*	$1.26 \times 10^{-6} \pm 1.23 \times 10^{-7}$	1
oriZ$^+$ *ΔrecG control location*	$1.72 \times 10^{-6} \pm 1.81 \times 10^{-7}$	1.37

In a double origin background, the absence of RecG increases the reversion rate of the *KanKanMX4* module inserted in the termination area by a factor of 3.24 (Table 5). The factor increase in the reversion rate of the control location was, at 1.37, slightly less. Whilst this is in principle what was expected based on the hypothesis, the increase in over-replication affected by the addition of *oriZ* in *oriC*$^+$ *oriZ*$^+$ *ΔrecG* cells is so striking (Rudolph et al., 2013) that a more significant increase in reversion rate in the termination area of the double origin *ΔrecG* strain compared to the control region was expected.

Moreover, the actual reversion rate of the *KanKanMX4* construct at the control location is higher than that at the termination region in both *recG*$^+$ and *ΔrecG* cells, in contrast to what we would expect (Table 5). The replication profiles of *ΔrecG* cells show clearly that over-replication occurs in the termination region (Dimude et al., 2015; Rudolph et al., 2013; Wendel et al., 2014) and it has been found to be dependent on both RecB and RecA (Midgley-Smith et al., 2018; Rudolph et al., 2013), thereby indicating the necessity of recombination in enabling the over-replication to take place. Taking this in to account, it is somewhat surprising that the *KanKanMX4yjhR* construct demonstrated a greater recombination rate than the *KanKanMX4narU* construct in a double origin background.

Discussion

A tandem repeat reporter cassette inserted at two different locations of the chromosome was employed to measure the reversion rate of each location in *ΔrecG* cells with one or two replication origins in an attempt to provide a direct link between over-replication in *ΔrecG* cells and homologous recombination. The over-replication seen in the termination region of *ΔrecG* cells depends on both

RecA and RecB but not Ruv, and previous studies using the *KanKanMX4* plasmid-based assay reported that the same is seen in the increase in the reversion rate of the tandem repeat specifically in the absence of RecG (Lloyd and Rudolph, 2016; Midgley-Smith et al., 2018; Rudolph et al., 2013), which supports the suitability of this assay for our aims. In line with the model for how over-replication in the absence of RecG occurs as a result of replication fork fusions, a larger increase in the reversion rate of the tandem repeat cassette in the termination region was anticipated, where replication predominantly terminates, compared to a control region away from termination. In cells containing an additional, ectopic replication origin, *oriZ*, the control region becomes an ectopic fork fusion region, and we expected to see a bigger increase in reversion rate in *ΔrecG* derivatives of these cells compared with the single origin *ΔrecG* cells.

Had we seen the trends that had been expected, I would have carried out a variety of control experiments such as creating *ΔrecA* derivatives to confirm that the increase in the reversion rate in the absence of RecG was due to an increase in homologous recombination, and *ΔrecB* derivatives to support the suggestion that the initial substrate for a recombination event was a dsDNA end that is processed by RecBCD. There are a number of further experiments that would have been carried out to test for a correlation between increased recombination frequency and fork fusion events. Previous work has shown that initiation of over-replication that occurs in the absence of RecG is dependent on the helicase activity of PriA, specifically the 3' flap processing ability (Rudolph et al., 2013). The *srgA1* allele of *priA* encodes a mutant protein that no longer has this ability (Gregg et al., 2002) and the over-replication is abolished in *ΔrecG srgA1* cells (Dimude et al., 2015; Rudolph et al., 2013). By introducing the *srgA1* allele in to the *KanKanMX4* strains, we would expect that any increase in reversion rate in the *ΔrecG* derivatives would be abolished, which would provide further support to the model of recombination-dependent over-replication initiating at fork fusion locations.

An important control would be needed to support that any increase in reversion rate seen in the absence of RecG at both *narU* and *yjhR* following the introduction of the ectopic replication origin, *oriZ*, was due to replication fork fusions and not a peculiarity of a double origin strain. In both insert locations, an increase would be expected when *oriZ* was introduced. Therefore, there should be a third integration site of the *KanKanMX4* construct in the opposite replichore to *oriZ*. This would be needed in order to establish that there is no increase in this location even when *oriZ* is introduced so that the increases seen in the other two locations can be attributed to replication fork fusion events and are not just a feature of cells with two replication origins. As discussed below, this was not the case as the system did not work quite as expected and so this further work was not undertaken.

The data show an increase in the reversion rate in cells lacking RecG, in line with previous studies (Lloyd and Rudolph, 2016; Lovett, 2006; Lovett et al., 1993), and a smaller increase in *ΔrecG* cells with a double origin strain background. We hypothesised that there would be a larger increase in reversion in cells with the repeat cassette inserted in the termination region in the single origin strains and a significantly larger increase in the termination region in double origin strains, but the data do not show this (Table 4, Table 5). There was a significant increase in the reversion rate in both

KanKanMX4narU and *KanKanMX4yjhR* cells in the absence of RecG as was expected, but the increase was not higher in *KanKanMX4narU* cells and instead, the factor increase in reversion in the absence of RecG in single origin strains was very similar at both *KanKanMX4* construct locations. The increases in reversion rate in double origin *ΔrecG* cells were also significantly less than we had expected, particularly for the native termination region (*KanKanMX4narU*). Moreover, the actual reversion rates of *oriC⁺ oriZ⁺ KanKanMX4yjhR* cells were higher than in *oriC⁺ oriZ⁺ KanKanMX4narU* cells both in the presence and absence of RecG.

Whilst the reversion rate data for the single origin constructs were not generated in parallel with the double origin constructs and so are not directly comparable, it is worth noting that the reversion rates of both double origin *recG⁺* constructs are higher than their single origin counterparts, whilst the values for the *ΔrecG* strains are really quite similar (*KanKanMX4* construct location specific) in both single origin and double origin constructs (Table 4, Table 5). There is almost a 3-fold increase in the reversion rate of the *KanKanMX4narU* construct in the double origin background compared to the rate in the corresponding single origin strain. More notably, there is a 10-fold difference in reversion rate between the single and double origin *KanKanMX4yjhR* cells whilst the values for the two *ΔrecG* derivatives are relatively similar. Given that the increase occurred in the control strains where all processing enzymes are present, it is a surprising observation that we would not expect based on our model.

The results of the tandem repeat deletion assay were not fully in line with what we had expected, indicating that either our model is wrong and so the predictions we made based on that were wrong, or that the assay is not suitable for measuring what we aimed to measure.

Our model hypothesises that RecBCD is involved in remodelling fork fusion intermediates to promote recombination and generation of a substrate for PriA-mediated replication fork assembly, and also in processing the double strand end of a newly-established replication fork and promoting invasion of the re-replicated DNA behind the fork or the sister chromosome, leading to further replication initiation (Dimude et al., 2016; Rudolph et al., 2009b, 2010a, 2013). Both *in vitro* and *in vivo*, RecBCD degrades dsDNA at a very fast rate (Dillingham and Kowalczykowski, 2008; Wiktor et al., 2018), and with high processivity of over ~100 kb. The *KanKanMX4* construct employed in this reversion rate assay contains an internal duplication that is 266 bp long and was not specifically engineered to contain *chi* sites (χ), the specific DNA sequence that regulates RecBCD activity by switching degradation of the dsDNA to the 5'end only, resulting in a 3' ssDNA overhang that is a substrate for RecA-mediated strand invasion of homologous sequences (Anderson and Kowalczykowski, 1997; Arnold and Kowalczykowski, 2001; Bianco and Kowalczykowski, 1997; Smith, 2012). One explanation for the discrepancy between the hypothesis and what was observed could be that the action of RecBCD enzyme in dsDNA degradation is very likely to degrade far more than the entire *KanKanMX4* construct and so the subsequent recombination event would occur outside of the *KanKanMX4* construct. In this event, the tandem repeat within the kanamycin resistance gene would remain intact and so would not result in kanamycin resistance. The assay was only ever going to be

successful in detecting an effect in the absence of RecG purely because, on a population basis, there would be so many reversion events that on occasion, the degradation of dsDNA by RecBCD would end in the vicinity of the *KanKanMX4* construct and the subsequent RecA-mediated strand invasion would happen to result in the deletion of one of the duplicated sequences. Given the reported speed and processivity of RecBCD *in vivo*, it is possible that in the majority of cases where RecBCD activity is triggered, the enzyme is so fast and processive that the whole tandem repeat construct is simply degraded, and the RecA-mediated strand invasion event takes place upstream of the *KanKanMX4* construct, which will therefore leave the duplication intact more often than not, even after a homologous recombination event has taken place. This provides a potential explanation or part thereof for the difference between what we expected to see and the actual data.

Why might the reversion rates of *ΔrecG* strains with the *KanKanMX4yjhR* construct be similar to or even higher than the construct in the termination region in both single and double origin strain backgrounds? The *yjhR* insert location, which functions as a control region in the single origin strains, becomes an ectopic fork fusion region in the double origin strains, although there are no *ter*/Tus complexes to form a replication fork trap like that seen in the native termination region. This structural difference could go some way in explaining why the actual reversion rate of the *KanKanMX4yjhR* construct is higher than that of the *KanKanMX4narU* construct. Over-replication within the native termination region is contained within the ~250 kb between the innermost *ter* sites (Dimude et al., 2016; Duggin and Bell, 2009). The speed of dsDNA degradation by RecBCD according to a recent *in vivo* study (~1.6 kb s^{-1}) (Wiktor et al., 2018) is faster than DNA synthesis by PolIII (550 – 750 bp s^{-1}) (Pham et al., 2013). If RecBCD degrades so quickly and the progression of replication forks within the termination region is blocked by the *ter*/Tus replication fork pause sites, the over-replication will be digested away very quickly. In contrast, the *yjhR* location in the double origin strains becomes an ectopic termination region but without the constraints of the *ter*/Tus complexes. Recent work has shown that activity at the *rrnCABE* operon cluster impedes replication forks progressing in a direction opposite to normal (Ivanova et al., 2015; Midgley-Smith et al., 2018) but even taking this in to account, the area between *rrnCABE* and *oriZ* is much larger than between *terA* and *terC*. This might provide a higher chance that over-replication initiating within this region is able to persist before RecBCD can catch up.

Additionally, given that there is less confinement of replication forks within the ectopic termination region between *oriC* – *oriZ* and the *rrn* operons (*rrnH* and *rrnCABE*) will lead to replication fork pausing, there is likely to be more variability in the location of where forks actually fuse in the ectopic fork fusion region compared to the native termination area, which means a wider spread of initiating events. If the proximity of the initiating event to the tandem repeat cassette is the problem, then a wider spread would be in favour of seeing more reversion events because if by chance there is a fork fusion event that results in over-replication that happens far enough away from the *KanKanMX4yjhR* construct so that over the course of that experiment, RecBCD degradation happens to stop in the duplicated area, then that might still lead to a reversion event even though there are no

chi sites. So a higher variability of fork fusion locations in the ectopic termination region might lead to more reversion events being detected.

A prediction of this scenario would be that if the replication fork trap is inactivated, this should cause an increase in the reversion rate of the tandem repeat construct located at *narU* in the native termination region. To test this, I planned to inactivate the replication fork pause sites (*ter*/Tus) by deleting *tus* to see if that lead to an additional increase in the reversion rate in the native termination area. The introduction of *Δtus* to *oriC⁺ oriZ⁺ KanKanMX4narU* cells would create a similar situation in the native termination region as that in the ectopic termination region, and so the fork fusion locations would become more spread out. In *ΔrecG* derivatives, this would result in degradation as part of RecBCD-mediated recombination initiating in more diverse locations compared to the situation in *tus⁺* cells, and given the large number of cells assayed in an experiment, it should increase the chance of the assay detecting a number of extra reversion events whereby just by chance, the degradation ends near to the tandem repeat cassette and then causes a reversion event. Whilst it would have been interesting to investigate why we did not see the results we had expected, this was beyond the scope of the timeframe of my PhD studies.

In a growing population, there is a gradient in the copy number of chromosomal sequences. Homologous recombination can only take place if there is a second copy of the chromosomal sequence present. Given that within a population there will be more copies of *KanKanMX4yjhR* than *KanKanMX4narU* this might facilitate more recombination events in the *yjhR* location compared to the native termination region, which might also contribute to the unexpectedly high reversion rates detected in the *yjhR* location.

If there are factors affecting the accuracy of the data in reflecting the levels of recombination actually occurring in each chromosome location tested, as we have suggested above, there are a number of approaches that might mitigate this. Given the speed of RecBCD dsDNA degradation, using a construct with a much longer repeat sequence might help to increase the chance that RecBCD degradation ends in the vicinity and RecA-mediated strand invasion occurs within the tandem repeat, thereby increasing the detection frequency of recombination events. Lovett and colleagues have previously used a 750 bp tandem repeat to investigate reversion rates in a number of strain backgrounds, and reported a factor of 4 increase in reversion rate in *ΔrecG* cells with the tandem repeat integrated in to the chromosome at *lac* (Lovett, 2006; Lovett et al., 1993). This repeat sequence is almost three times the length of the 266 bp repeat within the *KanKanMX4* construct and so would be a good starting point, but it is likely that even this might not be long enough.

To increase the sensitivity of the assay, an artificial tandem repeat construct could be generated that is engineered to contain a *chi* sequence, which would increase the likelihood of strand invasion occurring within the tandem repeat sequence; recombination is increased 5- to 10-fold in the vicinity of a χ site (Arnold and Kowalczykowski, 2001). Wiktor et al. (Wiktor et al., 2018) recently used live cell imaging to investigate the efficiency of RecBCD in χ site recognition and report a calculated efficiency of 26 % indicating that at least two if not three χ sites would be needed to provide a good

chance of limiting the degradation activity of RecBCD. It was shown previously in an *in vivo* assay measuring the degradation of plasmid DNA linearized by the introduction of a DSB that the presence of a single χ site increases the fraction of undegraded DNA from <4 % to almost 14 %, and the addition of a second χ site increases the survival to >30 % (Kuzminov et al., 1994). Despite the fact that the 8 bp χ sequence is over-represented in the *E. coli* genome (Cardon et al., 1993), generating a large tandem repeat that contains two or three χ sites and yet provides a mechanism of detecting reversion events would require quite some genetic engineering and I did not have time to complete this work within the timescale of my PhD studies.

The data presented in this chapter raise questions about a potential correlation between increasing the number of origins and an increase in recombination frequency. We do not currently have further insight in to what might be happening where the introduction of an ectopic origin in *recG*$^+$ cells seems to result in such an increase in reversion rate, and in order to be certain of any significance in the differences of reversion rate, the experiments would need to be repeated with all strains assayed in parallel. If the observation of the increase in recombination frequency in cells with a second copy of the replication origin is shown to be real, and by inference that double origin strains have increased genomic instability, this would help to explain why *E. coli* employs the replication system that it does, with only one origin instead of multiple origins around the chromosome, and so it would be interesting to investigate this further.

Differentially labelling replisomes

Replication of chromosomal DNA is carried out prior to cell division in all cells by a complex molecular machine called the replisome (Yao and O'Donnell, 2010). Replication initiates at a replication origin, where two replisomes are assembled and proceed until termination, where two converging forks fuse. In the circular chromosome of *E. coli*, there is a single origin of replication, *oriC*, and the two replisomes move independently along the DNA (Reyes-Lamothe et al., 2008). Fork fusions normally occur within a defined location called the termination area, which is flanked by unidirectional fork traps (*ter*/Tus) that allow forks to enter the termination region but block them from leaving (page 24; Neylon et al., 2005).

The *E. coli* replisome is a large, multiprotein complex that contains all the components required for chromosome duplication (Figure 21)(Beattie and Reyes-Lamothe, 2015; Johnson and O'Donnell, 2005; Kurth and O'Donnell, 2009; O'Donnell, 2006). The replicative DNA polymerase is present in the replisome as part of a multi-subunit complex called DNA polymerase III holoenzyme (Pol III H.E.), which has three major functional units; the replicative Pol III core, the β sliding clamp and the clamp loader. The Pol III core is a heterotrimeric complex consisting of the α subunit polymerase that carries out DNA synthesis (encoded by *dnaE*), the ε subunit (encoded by *dnaQ*) that provides 3'-5' exonuclease proofreading activity (Scheuermann and Echols, 1984) and the θ subunit (encoded by *holE*), whose function is yet to be fully understood but there is some suggestion that it might have a role in stabilising the ε subunit, thereby enhancing ε subunit proofreading activity (Kurth and O'Donnell, 2009; Taft-Benz and Schaaper, 2004). It was historically understood that there were two copies of Pol III present in a replisome, one associated with each strand of the DNA template, but investigations by McInerney et al. (McInerney et al., 2007) demonstrated that three DNA Pol IIIs are present in the Pol III H.E *in vitro*. Using a technique that allowed millisecond-level live cell imaging (Plank et al., 2009), the Sherratt lab confirmed that there are in fact three molecules of Pol III in an active replisome *in vivo* in *E. coli* (Beattie and Reyes-Lamothe, 2015; Reyes-Lamothe et al., 2010).

The speed and processivity of Pol III is due to its interaction with the β sliding clamp (encoded by *dnaN*), a ring-shaped homodimer that encircles the DNA and tethers Pol III to the template via a specific interaction with the α subunit (Burgers et al., 1981; LaDuca et al., 1986; Marians et al., 1998). The final component of the Pol III H.E. is the clamp loader, which actively assembles the β sliding clamp around the DNA. The multiprotein complex is a heptamer ($\tau_3\delta\delta'\psi\chi$) in which the δ subunit (encoded by *holA*) binds to the β sliding clamp dimer and opens it (Jeruzalmi et al., 2001; Turner et al., 1999), and the τ subunits (encoded by *dnaX*) form connections between the Pol III core units and the DnaB helicase (Gao and McHenry, 2001a, 2001b). A combination of three molecules of Pol III in association with the clamp loader is termed Pol III* ($(\alpha\varepsilon\theta)_3$-$\tau_3\delta\delta'$). Less is known about the roles of the final two subunits of the clamp loader, χ (*holC*) and ψ (*holD*). They form a heterodimer and it was thought that

there was one copy in the replisome (Johnson and O'Donnell, 2005) but *in vivo* microscopy of fluorescently-labelled χ and ψ has shown instead that there are four copies of the heterodimer present in a replisome (Reyes-Lamothe et al., 2010).

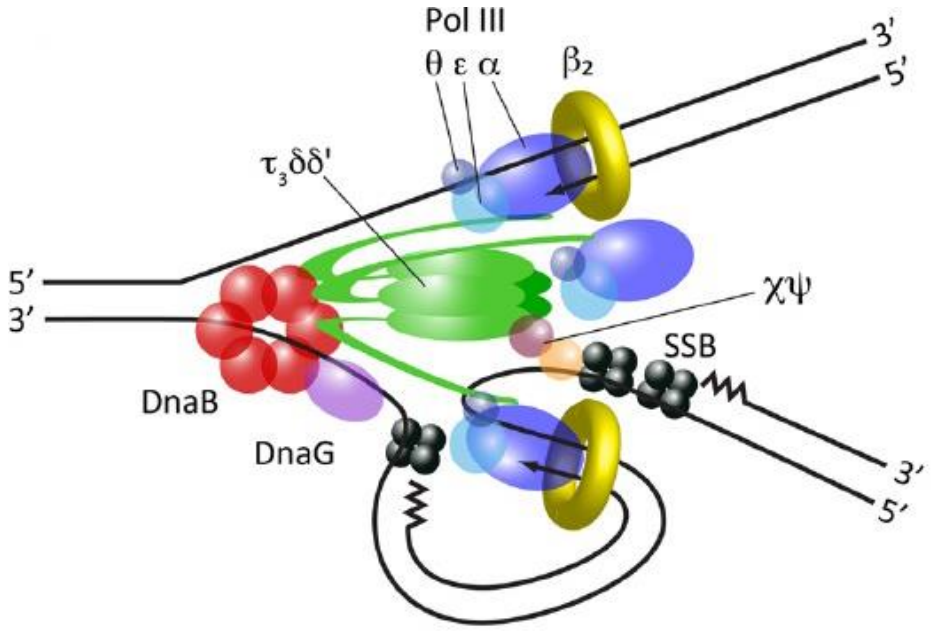

Figure 21: Components of the *E. coli* replisome. The clamp loader complex is shown in green and the β sliding clamps are shown in yellow. All other components are labelled within the diagram. Figure reproduced from Beattie and Reyes-Lamothe, 2015 with permission.

Exactly what happens when two replication forks meet and fuse is not well understood (Beattie and Reyes-Lamothe, 2015; Dewar and Walter, 2017; Dimude et al., 2016; Neylon et al., 2005). In *E. coli*, there is a single fork fusion event per cell cycle between two replication forks, and this takes place in a well-defined termination region. If we could label the replication forks differentially, we would be able to study in detail the events associated with replication fork fusion.

I explored the feasibility of differentially labelling replisomes in single cells using two different fluorescent proteins, with the aim of using this as a tool for investigating replisome dynamics in individual cells. The independent progression of the two replisomes in the bidirectional replication initiated at *oriC* in *E. coli* enables visualisation of the replication forks as two separate foci at certain points of DNA replication (Reyes-Lamothe et al., 2008), namely sometime after initiation and before termination, when replisome components are fluorescently labelled. Cells in which each replisome is labelled with a different colour could potentially be identified at that point and, using time lapse microscopy, the replisomes might then be tracked, even if the foci subsequently come together as a single focus once more, by using the different colour channels of the fluorescent microscope. These cells would enable us to monitor the progression of the replisomes as well as gain further insight in to replication termination and the fusion of two replication forks. Does one fork disassemble before the other at fork fusion or upon arriving at a *ter*/Tus trap? Is this different in a strain that has a second, ectopic copy of *oriC*, which ensures one fork is always stalled at a *ter*/Tus trap prior to fork fusion?

Cells with differentially labelled replisomes would help us to answer questions such as these. Derivatives of these strains with deletion mutations that have been shown to result in origin-independent DNA synthesis could then be created in order to investigate over-replication initiating at fork fusion events, for example *ΔrecG* or *ΔxonA ΔxseA ΔsbcCD*.

Labelling the replisomes

A number of replisome component fluorescent fusion proteins have been successfully created by other labs. The stoichiometry of the various replisome components was taken in to consideration when choosing which would be most appropriate for our aim. For example, β sliding clamp molecules are left behind the replisome on the lagging strand as part of Okazaki strand synthesis. This results in a trail of signal instead of a defined focus (Georgescu et al., 2010; Moolman et al., 2014; Reyes-Lamothe et al., 2010; Stukenberg et al., 1994). The DnaB replicative helicase is a hexamer and so DnaB proteins are present at a relatively high copy number in each replisome, which makes it very unlikely that there will be single colour replisomes by chance. The ε subunit of the Pol III core is well suited for differential labelling as it is located only at the replication fork and is present at a low enough copy number (three) that we were optimistic about the chances of achieving single-coloured replisomes.

To generate the appropriate strain, two copies of the gene for *dnaQ* fused to two different fluorescent protein genes had to be combined in to a single strain. To avoid fluorescent signal generated from plasmid replication, we chose to chromosomally express both fluorescent protein (FP) constructs instead of using an expression vector, which also avoided the problem of plasmids being present in cells in multiple copies. One of the *dnaQ*-FP fusions was located in the place of the native *dnaQ* gene. The second *dnaQ*-FP fusion was integrated in an ectopic location on the chromosome close to the site of the original *dnaQ* gene. Both were important considerations with respect to managing gene dosage during replication. Gene dosage refers to the number of copies of a gene present in the genome of a cell. Gene expression is affected by the position of a gene in relation to the origin of replication, with those located close to the origin being present at a higher copy number than those located further away from the origin (Sousa et al., 1997). To avoid bias in favour of either of the fusions, and to ensure the best chance of achieving single colour labelling of individual replisomes, it was important to ensure that the two *dnaQ* FP variants were present at a ratio as close to 1:1 as possible.

DnaQ-YPet and DnaQ-mTag fusions have previously been generated by the Sherratt lab (Reyes-Lamothe et al., 2008) and these were used for this work. Both the *dnaQ-ypet* and the *dnaQ-mtag* constructs were associated with an *frt*-flanked gene for kanamycin resistance (*kan*) (Reyes-Lamothe et al., 2008) to enable the constructs to be transferred between strains and selection for cells carrying the transferred DNA to be carried out. In order to allow the two constructs to be combined in a single strain, it was necessary to first remove the *kan* gene from *dnaQ-ypet* cells using the FLP/*frt* site-directed recombination system (page 51; Datsenko and Wanner, 2000), after which we were able to

move the *dnaQ-mtag* construct in to the resulting strain via P1*vir* transduction, generating *dnaQ-ypet dnaQ-mtag* cells (see page 64 for full strain construction).

The replisomes can be differentially labelled

The foci generated in *dnaQ-ypet dnaQ-mtag* cells were examined to see if it was possible to generate differentially labelled replisomes in single cells where the three molecules of DnaQ in a single replisome were all YPet fusions and the three molecules of DnaQ in the other replisome were all mTag fusions.

dnaQ-ypet dnaQ-mtag cells were cultured in M9 minimal medium supplemented with 0.2% glucose to ensure slow growth and limit replication initiation to a single event per cell cycle (page 13). A sample of the culture was taken at early exponential growth phase. This was to ensure that cells would be actively dividing and so replication would be taking place but also that the culture was not so grown that the cells would be too dense to allow for clear visualisation of single cells when observed via fluorescent microscopy. The sample of liquid culture was applied to a minimal medium agarose pad on a microscopy slide and allowed to dry (page 62). Once dry, the samples were covered with a cover slip, which gently squeezes the bacteria between the agarose surface and the cover slip and fixes the bacteria in place without the need to fix the cells chemically, allowing for live cell microscopy. Images were taken using conventional fluorescence microscopy.

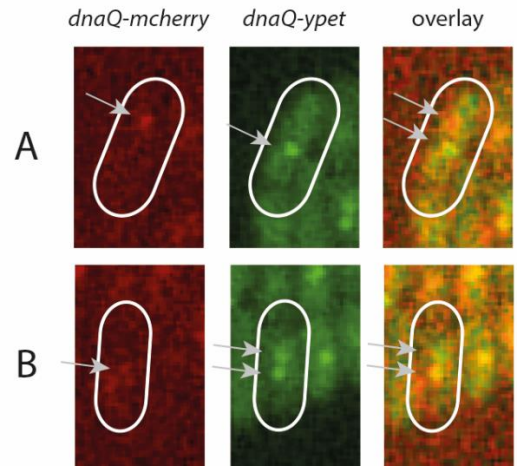

Figure 22: Visualisation of replisomes in actively replicating live *E. coli* cells via fluorescently-labelled DnaQ molecules. Snapshot images of cells containing chromosomally-expressed *dnaQ-ypet* and *dnaQ-mtag* constructs, in an otherwise wild type background. Cells were grown in minimal medium to early exponential phase before being transferred to a minimal medium pad on a microscopy slide and viewed using conventional fluorescence microscopy. Cells were visualised with a Nikon T*i*-U inverted microscope equipped with a DS-Qi2 camera (Nikon). Images were taken and the colour of the foci pairs in individual cells was assessed. White arrows highlight the location of the foci. Brightfield microscopy was used alongside fluorescent channels to provide cell boundaries, which are shown as white lines. The strain used was SLM1146 (*dnaQ-ypet*<> *dnaQ-mtag*<>).

With two *dnaQ* FP variants (YPet and mTag), the composition of a single replisome focus can theoretically be 3:0, 2:1, 1:2 or 0:3 DnaQ-YPet:DnaQ-mTag.

The panels in Figure 22A show one focus visible in the red channel (*dnaQ-mtag*) and a separate focus visible in the green channel (*dnaQ-ypet*), each representing a replisome within a single cell. The fact that they are in distinct locations in the cell, coupled with the foci appearance in the overlay image, shows that both foci are pure in colour. The panels in Figure 22B show an example of a cell that contains one pure colour focus (*dnaQ-ypet*, green) and one mixed colour focus, which contains both *dnaQ-ypet* and *dnaQ-mtag* molecules and appears as orange in the overlay image. Whilst a large proportion of the cells contained two foci that were both orange, meaning that the three DnaQ molecules in each replisome were a mixture of DnaQ-mTag and DnaQ-YPet, the two cells shown in the images in Figure 22 are both examples of cells that would be appropriate for use in investigating replisome dynamics as at least one replisome in each case contained three copies of the same fluorescently-labelled DnaQ construct. The images in Figure 22 show that it is indeed possible to generate differentially labelled replisomes in cells containing two different fluorescently-labelled *dnaQ* constructs.

Discussion

The data presented show that is it possible to differentially label the replisomes of a single cell by creating cells with two copies of the gene for DnaQ, one of the subunits of the replicative polymerase Pol III, and labelling each gene with a different coloured fluorescent protein. The benefit of labelling a replisome subunit that is present at a low copy number is that there is a greater chance of achieving pure coloured-replisomes and so a greater chance at generating cells with differentially labelled replisomes. The intensity of the fluorescent signal is unavoidably compromised as a consequence, and as can be seen from Figure 22, the capability of basic fluorescent microscopy was pushed to the limit in order to detect the three fluorescently labelled copies of ε in each replisome as a single focus.

Rolling circle replication assays have been used extensively to study replication *in vitro*. For example, Yao et al. (Yao et al., 2009) used this technique to investigate features of the replisome and reported that the DnaB and Pol III* subunits of the replisome were stably associated with the replication fork for anywhere between <30 kb and >300 kb, with an average of ~ 80 kb. It was somewhat surprising that not long after we had successfully achieved proof of concept in differential labelling of replisomes, it came to light from two independent studies that components of the replisome do not remain stably associated and instead are frequently swapped during DNA replication elongation. The publications were preceded by a report of a theoretical explanation for the high rates of polymerase exchange seen in the T7 phage (Åberg et al., 2016). The authors suggested that the presence and absence of competing replisome factors could be the explanation for what is seen *in vitro* (e.g. from rolling circle replication assays) and what has been seen *in vivo* in T7, which they then

showed to be possible using mathematical modelling. Beattie and co-workers (Beattie et al., 2017) investigated the stability of the replisome of *E. coli* cells *in vivo* by examining fluorescently-labelled replisome components using two independent methods; fluorescence recovery after photobleaching (FRAP) and single-particle tracking Photoactivated Localisation Microscopy (sptPALM). They found that Pol III and the clamp loader (Pol III*) components recover fluorescence very quickly after bleaching, which is consistent with protein exchange taking place. Pol III H.E. components were found to be bound to the replisome for no more than a matter of seconds at a time, and the fact that the timings for the Pol III* components were very similar despite a difference in stoichiometry led to the suggestion that the exchange was likely not individual protein components from the replisome but instead was the whole Pol III* subassembly ($(\alpha\varepsilon\theta)_3$-$\tau_3\delta\delta'$). An estimation of the fork progression between each exchange under the conditions of their assay was significantly longer than the average Okazaki fragment length and so they conclude that their observations are likely not linked to mechanistics of the lagging strand synthesis. In contrast, DnaB helicase displayed significantly more stability at the replisome than any other component and showed no recovery of fluorescent signal over a 5 min period of measurement.

At the same time, Lewis and colleagues published a combined *in vitro* and *in vivo* study showing that indeed as Åberg et al. (2016) had suggested, availability of the replisome components could be the factor that influences the stability of the replisome (Lewis et al., 2017).

Unfortunately, the finding that *in vivo*, Pol III molecules, and therefore the ε subunit (*dnaQ*), are rapidly exchanged with copies present in the environment around the replisome renders this technique of differential labelling of the replisomes futile for studying replisome dynamics. As already discussed above, components outside of the Pol III* subunit assembly that might be more stably associated with the replication fork do not present as good candidates for achieving differential labelling of the replisomes, rendering it unlikely that a workable solution will be found to make this aim possible.

3’ exonucleases and DNA replication termination

Recent work has shown that the DnaA-independent over-replication of the chromosome in cells lacking RecG is dependent on specifically the helicase activity of the replication restart protein PriA (Rudolph et al., 2013). PriA has multiple biochemical activities. In an otherwise wild type background, mutations that disable the ATPase, helicase or translocase activities of PriA were shown to have very little or no effect on normal cellular physiology, whereas retaining the ability to assemble protein complexes on to the chromosome was found to be instrumental in maintaining a near-wild type phenotype (Sandler, 2005; Sandler et al., 2001; Zavitz and Marians, 1992). *priA300* (PriAK230R) contains a mutation in the *priA* gene that inactivates the helicase, ATPase and translocase activities of PriA whilst leaving the ability to assemble active primosomes intact (McGlynn et al., 1997; Zavitz and Marians, 1992). A mutation in the *priA* gene called *srgA1* is even more specific in its effect on PriA activity than *priA300*. *srgA1* encodes a protein that has an L557P substitution that specifically affects its ability to unwind the lagging strand of a replication fork that is missing the leading strand, a branched DNA structure that is the equivalent of a 3’ flap. The ability of the mutant protein to unwind replication forks in situations with both the leading and lagging strands present is essentially the same as for the wild-type protein (Gregg et al., 2002). The introduction of the *srgA1* allele of *priA* in to *ΔrecG* cells abolishes the DnaA-independent over-replication. This suggests that the DNA structure exploited by PriA to initiate origin-independent DNA synthesis in the absence of RecG is a 3’ flap (Rudolph et al., 2010b, 2013).

3’ ssDNA can be degraded by 3’ exonucleases. This led to questions about what might happen in cells lacking 3’ exonucleases; does origin-independent over-replication occur in the termination region of these cells? Rudolph and colleagues (2013) generated a replication profile for a strain that lacked all three major 3’ exonucleases, ExoI, ExoVII and SbcCD (Δ*xonA ΔxseA ΔsbcCD*). Exonuclease I (ExoI), is a highly processive exonuclease specific to ssDNA, with 3’ to 5’ activity (Lehman and Nussbaum, 1964; Thomas and Olivera, 1978). Exonuclease VII (ExoVII) is a processive exonuclease specific to ssDNA and has both 3’ to 5’ and 5’ to 3’ activity (Chase and Richardson, 1974a, 1974b). ExoVII is composed of two non-identical subunits; a large subunit encoded by *xseA* and a smaller subunit encoded by *xseB* (Vales et al., 1983). ExoVII-deficient strains in these studies were generated by deleting the *xseA* gene. SbcCD (*sbcCD*) has both ssDNA endonuclease and 3’ – 5’ dsDNA exonuclease activity (Connelly and Leach, 1996; Connelly et al., 1999).

As predicted, there was a peak of over-replication in the replication profile of this strain (Figure 23) (Rudolph et al., 2013). As with the over-replication seen in the profile of *ΔrecG* cells, the peak occurs in the termination region, although it bleeds through both Tus/*terC* and Tus/*terB* in to the left

hand replichore. The frequency of the chromosomal locations represented at the tip of the peak is equal to or even slightly higher than the frequency of the region of *oriC*, which means that the number of copies of this area of the chromosome present in the sample is at least as many as the number of copies of the origin region present in the sample.

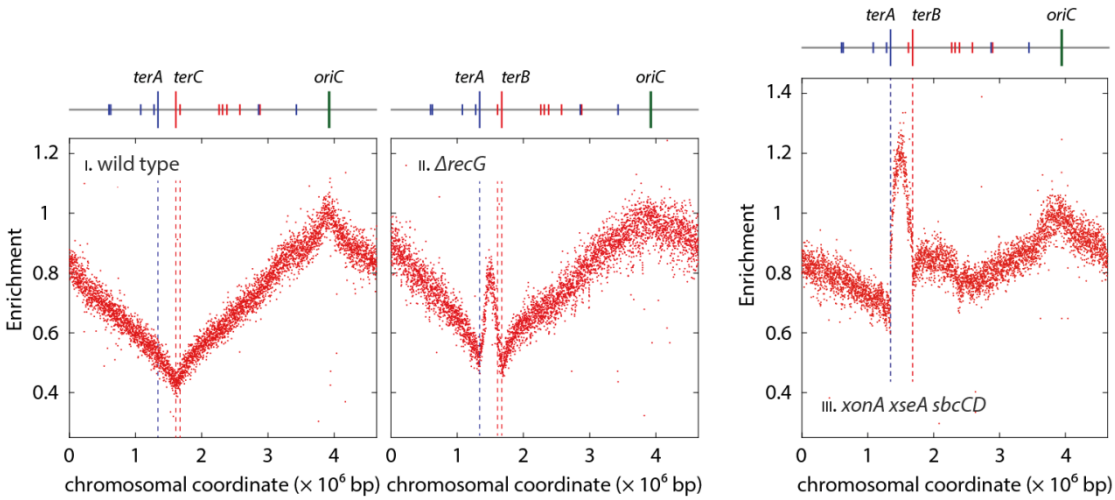

Figure 23: Replication profile of *E. coli* cells in the absence of RecG or the exonucleases ExoI, ExoVII and SbcCD. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary wild type control strain) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and the *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore). Sequencing templates were isolated from MG1655 (wild type), N6576 (*ΔrecG*) and N6953 (*ΔxonA ΔxseA ΔsbcCD*). Figure modified from Rudolph et al., 2013 with permission.

These findings triggered the following work to investigate the over-replication that occurs in the absence of the three major 3' ssDNA exonucleases and to provide further insight in to what might be occurring at fork fusion events. My data indicate that both RecG and 3' exonucleases have a role in managing events at replication termination. The data presented in this chapter have recently been published (Midgley-Smith et al., 2019).

Origin-independent growth

Cells lacking 3' exonucleases can grow in the absence of origin firing

The replication profile in Figure 23 demonstrates a clear similarity between cells lacking RecG and cells lacking 3' exonucleases; in the absence of either RecG or 3' exonucleases, there is a defined peak of over-replication in the termination area. In *ΔrecG* cells, this origin-independent over-replication is robust enough to sustain cell growth in the absence of origin firing (Rudolph et al., 2013).

Is the same true for the over-replication seen in the absence of 3' exonucleases? In order to assess if growth does arise as a result of this origin-independent over-replication, we created strain constructs lacking the three 3' exonucleases in which we were able to prevent *oriC* from firing using DnaA(ts). We conducted spot dilution assays of *dnaA Δtus rpo* ΔxonA ΔxseA ΔsbcCD* cells and compared growth in the presence (30 °C) and absence (42 °C) of origin firing. We also assayed the single and double mutant strains in a *dnaA Δtus rpo** background in order to assess if origin-independent growth can occur in the absence of only one or two of the proteins. Briefly, a sample of each exponentially growing culture at 30 °C was taken and serially diluted. Dilutions were spotted on to LB and minimal salts agar plates in duplicate and one set was then incubated at the permissive temperature for DnaA(ts) of 30 °C and the other at the restrictive temperature of 42 °C.

The spot dilution assay revealed that, as is seen in the absence of RecG, *dnaA Δtus rpo** cells lacking all three 3' exonucleases can indeed grow in the absence of origin firing (Figure 24, *dnaA Δtus rpo* ΔxonA ΔxseA ΔsbcCD* at 42 °C). These cells show robust levels of origin-independent growth, which is consistent with the clear peak of over-replication seen in Figure 23 and establishes further similarity between cells lacking RecG and cells lacking 3' exonucleases. The smaller colony sizes of *dnaA Δtus rpo** cells lacking all three 3' exonucleases at 30 °C after 24 h incubation (Figure 24, 30 °C, 24 h incubation) demonstrates that these cells have a mild slow-growth phenotype that is not seen in the absence of one or two 3' exonucleases, but subsequent growth following incubation for 48 h confirmed that the viability of the cells was not significantly affected (Figure 24, 30 °C, 48 h incubation).

The absence of each 3' exonuclease individually resulted in very little or no origin-independent growth, indicating that the presence of any two of the 3' exonuclease proteins is sufficient to prevent origin-independent growth occurring in these cells. However, the small amount of growth seen at 42 °C in the absence of ExoI only (Figure 24, *dnaA Δtus rpo* ΔxonA*) and not in the absence of either ExoVII or SbcCD (Figure 24, *dnaA Δtus rpo* ΔxseA* and *dnaA Δtus rpo* ΔsbcCD* respectively) suggests that ExoVII and SbcCD are not as effective at preventing origin-independent replication initiation as ExoI. The fact that growth is inhibited in *dnaA Δtus rpo* ΔxseA ΔsbcCD* cells at 42 °C even after 48 hours shows clearly that the presence of ExoI alone is enough to prevent origin-independent growth. Introduction of *ΔsbcCD* to *dnaA Δtus rpo* ΔxonA* cells improved growth at 42 °C significantly, and an even stronger effect was seen by the introduction of *ΔxseA* to *dnaA Δtus rpo* ΔxonA* cells. This assay demonstrates the very clear hierarchy of the 3' exonucleases in preventing origin-independent growth, with ExoI being the most important, followed by ExoVII and finally SbcCD.

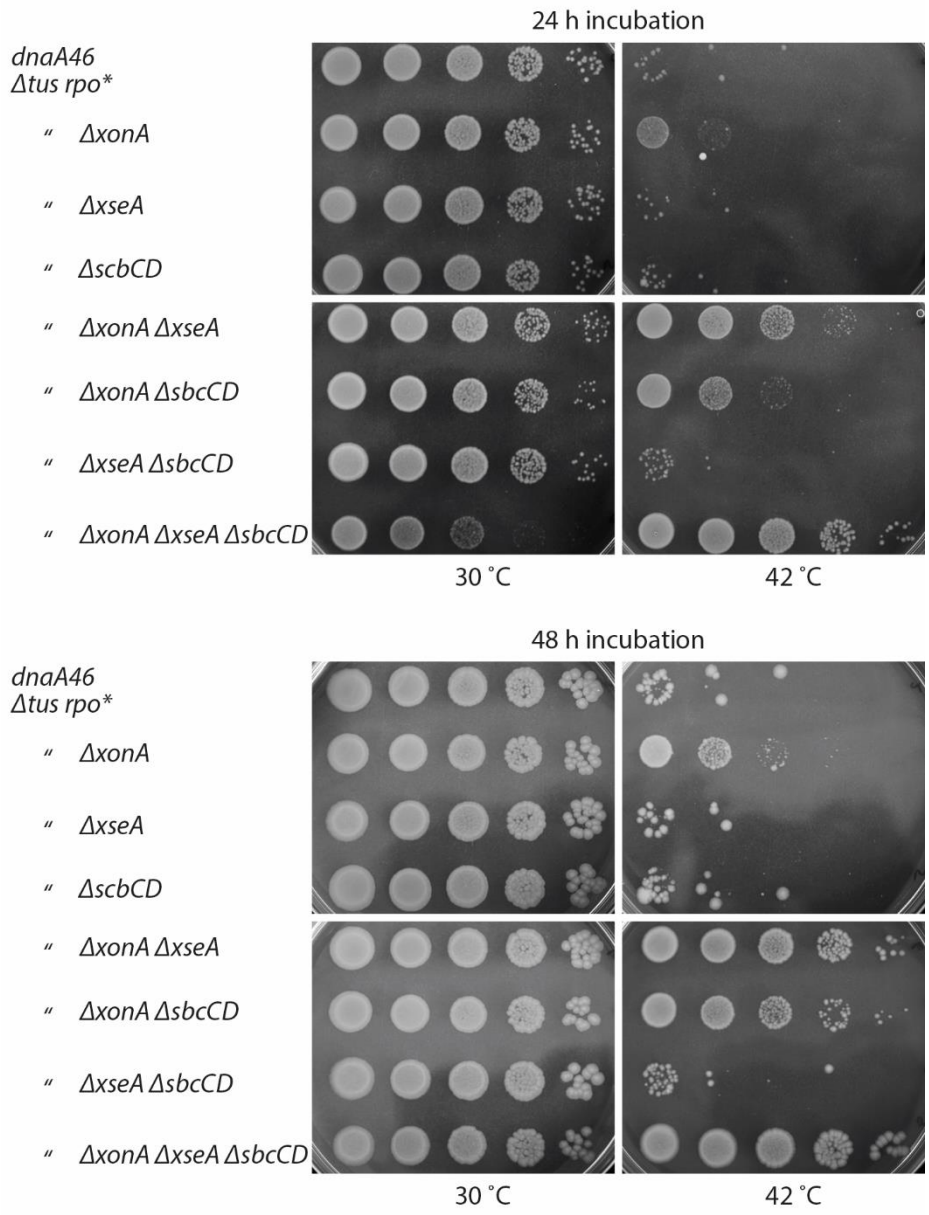

Figure 24: The effect of Δ*xonA*, Δ*xseA* and Δ*sbcCD* on origin-independent growth. A spot dilution assay was used to evaluate the relative contribution of Exo I, Exo VII and SbcCD in supressing growth arising via origin-independent DNA replication. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The replication fork trap in the termination area was inactivated by deletion of the *tus* gene and an *rpoB*35* point mutation was introduced, which destabilises ternary RNA polymerase complexes. The strains used were RCe267 (*dnaA46 Δtus rpo**), RCe528 (*dnaA46 Δtus rpo* ΔxonA*), SLM1219 (*dnaA46 Δtus rpo* ΔxseA*), RCe553 (*dnaA46 Δtus rpo* ΔsbcCD*), SLM1194 (*dnaA46 Δtus rpo* ΔxonA ΔxseA*), RCe554 (*dnaA46 Δtus rpo* ΔxonA ΔsbcCD*), SLM1223 (*dnaA46 Δtus rpo* ΔxseA ΔsbcCD*) and SLM1226 (*dnaA46 Δtus rpo* ΔxonA ΔxseA ΔsbcCD*).

To investigate chromosome replication in cells lacking 3' exonucleases in more detail, we conducted marker frequency analysis via deep sequencing to generate replication profiles of cells lacking either ExoI (*ΔxonA*), ExoVII (*ΔxseA*) or SbcCD (*ΔsbcCD*).

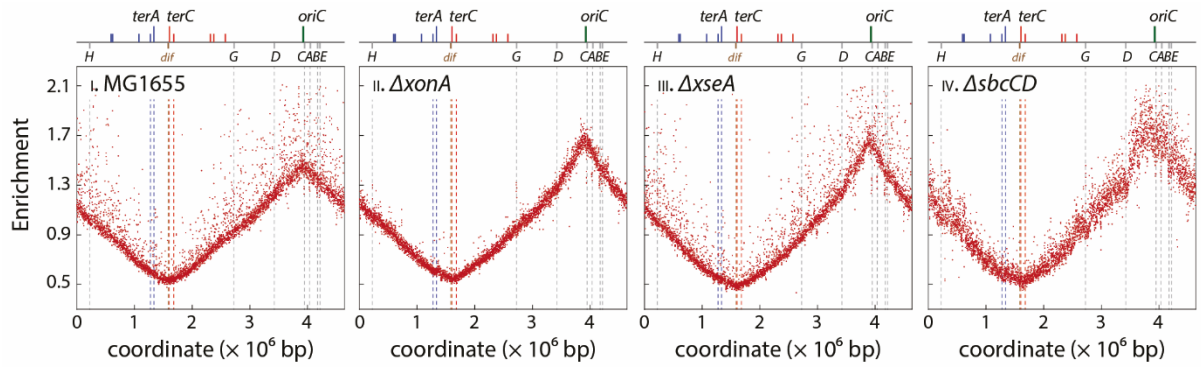

Figure 25: Replication profiles of *E. coli* cells in the presence and absence of the exonucleases ExoI, ExoVII or SbcCD. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Sequencing templates were isolated from MG1655 (wild type), RCe563 (*ΔxonA*), SLM1185 (*ΔxseA*) and RCe562 (*ΔsbcCD*). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

The replication profile of wild type cells (Figure 25, panel I) fits well with previously reported marker frequency analysis, with a clearly defined origin region present at the highest frequency on a population level and the lowest copy numbers in the termination region on a population level (Ivanova et al., 2015; Maduike et al., 2014; Rudolph et al., 2013; Skovgaard et al., 2011). The replication profiles of *ΔxonA*, *ΔxseA* and *ΔsbcCD* cells show that on a population level, in the absence of ExoI, ExoVII or SbcCD (Figure 25, panels II, III and IV respectively), there is no peak of over-replication in the termination area and the profiles look like that of wild-type cells (Figure 25, panel I), in stark contrast to the high frequency of initiation within the termination area seen in the profile of cells lacking all three exonucleases (Figure 23 panel III). As seen from the growth assay in Figure 24 however, all single mutants in a *dnaA Δtus rpo** background show essentially no growth in the absence of origin firing and so it is not particularly surprising that the single mutant replication profiles do not show any over-replication.

In line with the origin-independent growth seen in cells combining *ΔxonA* with *ΔxseA* or *ΔsbcCD* in a *dnaA Δtus rpo** background (Figure 24), replication profiles of *ΔxonA ΔxseA* cells and *ΔxonA ΔsbcCD* cells show peaks of over-replication within the termination area (Figure 26, panels I and II), as was shown previously for *ΔxonA ΔsbcCD* cells (Wendel et al., 2014, 2018). However, the peaks are small, especially compared to those seen in the absence of RecG or the triple 3' exonuclease mutant strain (Rudolph et al., 2013; Figure 23) which suggests that the presence of any one of the 3' exonuclease proteins will compensate for the absence of the other two to some extent, limiting the over-replication seen in the termination region in the replication profiles of the double mutant strains (Figure 26).

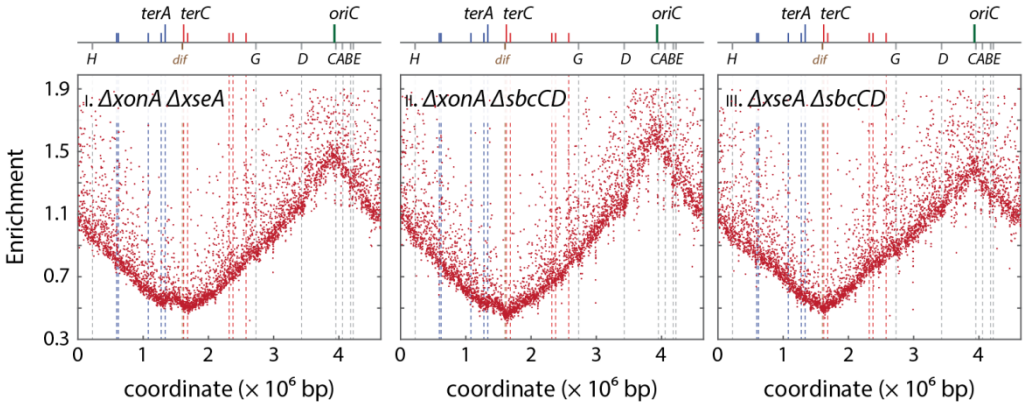

Figure 26: Replication profiles of *E. coli* cells in the absence of two exonucleases. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Sequencing templates were isolated from SLM1203 (*ΔxonA ΔxseA*), RCe569 (*ΔxonA ΔsbcCD*) and SLM1209 (*ΔxseA ΔsbcCD*). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

It is also possible that the peaks of over-replication seen in *ΔxonA ΔxseA* and *ΔxonA ΔsbcCD* cells (Figure 26) are less defined than that seen in the absence of RecG partly because the profiles are slightly shallower. In the profile of *ΔrecG* cells (Figure 23), the downwards slope of both the left- and right-hand replichores either side of the termination area is more acute than that in cells lacking 3' exonucleases, and this makes the peak more defined and visible.

The growth assay of *dnaA Δtus rpo** cells lacking 3' exonucleases (Figure 24) revealed that cells lacking all three 3' exonucleases have a mild slow growth phenotype (Figure 24; 30 °C, 24 h incubation). *dnaA Δtus rpo* ΔxonA ΔxseA* cells show no growth defects (30 °C) and grow robustly in the absence of origin firing even in the presence of SbcCD (42 °C) and so this strain was used for further experiments.

Origin-independent growth requires inactivation of the replication fork trap

The origin-independent over-replication in cells lacking RecG is only able to sustain cell growth in the absence of origin firing when a combination of two mutations are introduced to *dnaA ΔrecG* cells; *Δtus* inactivates the replication fork trap and an *rpoB*35* point mutation in the β subunit of the RNA polymerase alleviates conflicts between replication and transcription complexes (Rudolph et al., 2013). Without these two mutations, *dnaA ΔrecG* cells are unable to grow at the restrictive temperature (page 68).

The fact that over-replication seen in the replication profiles of cells lacking 3' exonucleases is confined to the termination area suggested that the same would apply to origin-independent growth in *dnaA ΔxonA ΔxseA* strains. In order to confirm if this was the case, we carried out a spot dilution assay of *dnaA ΔxonA ΔxseA* cells in the presence and absence of Tus protein and the *rpo** point mutation separately and in combination.

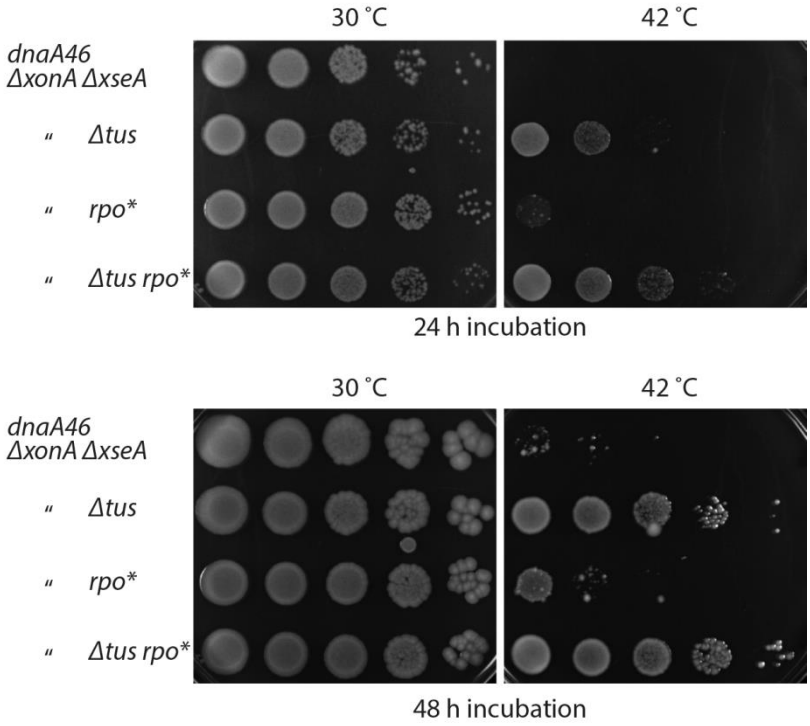

Figure 27: The effect of the replication fork trap and replication-transcription encounters on origin-independent growth in the absence of ExoI (Δ*xonA*) and ExoVII (Δ*xseA*). A spot dilution assay was used to evaluate the growth arising via origin-independent DNA replication in strains carrying a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The replication fork trap in the termination area was inactivated by deletion of the *tus* gene. The effects of replication-transcription encounters were reduced by introducing a point mutation, *rpo*B*35, which destabilises ternary RNA polymerase complexes. The strains used were SLM1246 (*dnaA46 ΔxonA ΔxseA*), SLM1244 (*dnaA46 ΔxonA ΔxseA Δtus*), SLM1245 (*dnaA46 ΔxonA ΔxseA rpo**) and SLM1194 (*dnaA46 ΔxonA ΔxseA Δtus rpo**).

The growth assay confirmed that, like *dnaA ΔrecG* cells, *dnaA ΔxonA ΔxseA* cells are unable to form colonies at the restrictive temperature of 42 °C (Figure 27). The introduction of an *rpo** point mutation resulted in a marginal increase in growth. Inactivation of the replication fork trap by deleting *tus* resulted in much more significant levels of growth, especially following 48 hrs incubation, showing that origin-independent over-replication is able to sustain growth once the replication forks are able to leave the termination area. This growth was improved even further in *dnaA ΔxonA ΔxseA* cells that carried a combination of both Δ*tus* and *rpo** mutations (Figure 27). This is in line with the idea that origin-independent replication initiates in the termination area in cells lacking the 3' exonucleases and is only able to sustain growth at the restrictive temperature in a *dnaA* background once the fork

trap is inactivated. Forks leaving the termination region will be moving in the opposite orientation to normal and so will encounter problems resulting from replication-transcription conflicts, and the *rpo** mutation relieves these conflicts to some extent (Dutta et al., 2011; Trautinger et al., 2005), leading to a stronger origin-independent growth phenotype (Figure 27).

Over-replication of the termination area is abolished if the chromosome is linearised

Although the chromosomes of most prokaryotes, including *E. coli*, are circular, some bacteriophages as well as some bacterial species have linear chromosomes like eukaryotes. Cui and colleagues (Cui et al., 2007) exploited the linearization mechanism of the *E. coli* bacteriophage N15 in order to investigate whether the *E. coli* chromosome could be linearised or not and to assess any effects on prokaryotic cells that result from having a linear chromosome. In this system, the *tos* sequence is introduced to the chromosome near to the dimer resolution site, *dif*, within the termination region. Lysogenic infection introduces the N15 telomerase TelN to the cells, which mediates linearization by processing the DNA at a site within *tos* to generate two termini with hairpin ends (see page 161 for more detail) (Cui et al., 2007). They reported that the *E. coli* cells remained viable and the linearised chromosome structure was stable. Linearisation did not affect the growth of wild type cells (Cui et al., 2007), a finding that was subsequently independently confirmed (Rudolph et al., 2013).

It was recently reported that chromosome linearization affects the origin-independent over-replication seen in cells lacking RecG (Dimude et al., 2015; Rudolph et al., 2013). The over-replication within the termination area is substantially reduced in *ΔrecG* cells with a linearised chromosome. In line with this, *dnaA Δtus rpo* ΔrecG* cells with a linearised chromosome are no longer able to grow in the absence of origin firing. This supports the idea that replication originating in the termination region is responsible for the origin-independent growth in *dnaA Δtus rpo* ΔrecG* cells and if replication forks are prevented from meeting by chromosome linearization, this growth is abolished. Given that over-replication is observed in the termination area of cells lacking 3' exonucleases, we wanted to investigate how the over-replication seen in these cells was affected when replication forks were prevented from meeting. Would it be a similar situation to that seen in *ΔrecG* cells, indicating a common underlying mechanism in initiation of origin-independent replication in the two strain backgrounds?

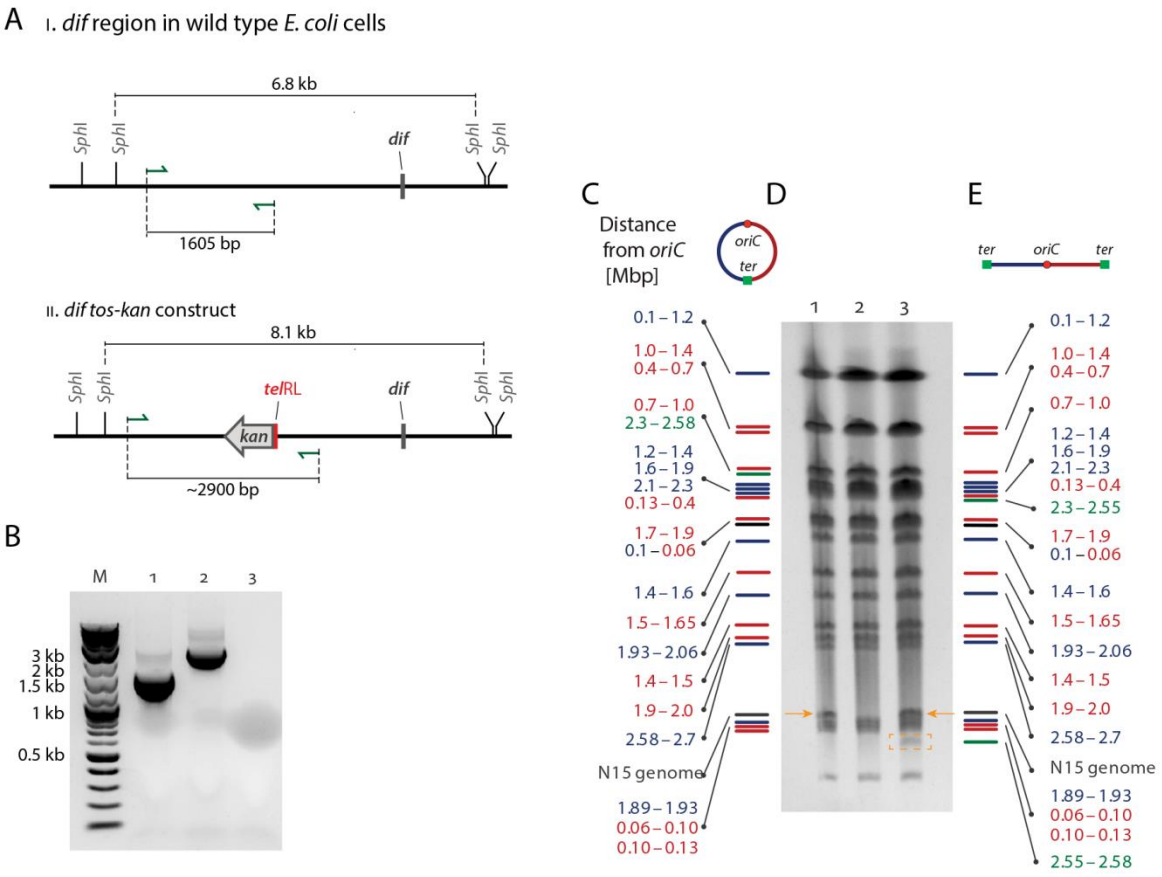

Figure 28: Confirmation of chromosome linearisation. **A**) Schematic representation of the region surrounding *dif* in the termination area in wild type cells and in cells with the *tos-kan* construct integrated. The linearisation verification primers are shown in green (Cui et al., 2007). **B**) PCR products generated with the linearisation verification primers for wild type (lane 1), *tos-kan* cells (lane 2) and N15 lysogen of *tos-kan* with a linearised chromosome (lane 3). The increase in product size between lane 1 and lane 2 indicates the presence of the *tos-kan* cassette. The N15 telomerase TelN is present in N15 lysogens. TelN cleaves its target sequence within *tos*, resulting in linearisation of the chromosome in *tos-kan* N15 lysogens. The linearisation site is located between the primer binding sites and so prevents the formation of a PCR product. **C** – **E**) Pulse field gel electrophoresis confirming linearisation of the chromosome. The *tos* sequence is inserted into a 273.6 kb *Not*I fragment shown in green in **C**, which is divided in to two fragments shown in green in **E** when *tos* is processed by TelN. One of these fragments forms an additional band near the base of the gel highlighted by an orange box that is absent from un-linearised chromosomal samples. The chromosomal DNA was prepared from SLM1213 (*ΔxonA ΔxseA* N15 lysogen) in lane 1, SLM1187 (*ΔxonA ΔxseA tos-kan*) in lane 2 and SLM1212 (*ΔxonA ΔxseA tos-kan* N15 lysogen) in lane 3.

We utilised the system described above to achieve chromosome linearization in strains lacking two of the three main 3' exonucleases, *ΔxonA ΔxseA*. Linearisation was confirmed (Figure 28) and marker frequency analysis by deep sequencing was conducted to establish replication profiles of *ΔxonA ΔxseA* cells with either a circular or linearised chromosome. The replication profiles show that the over-replication in the termination region of *ΔxonA ΔxseA* cells is unaffected by the insertion of the *tos-kan* construct or by the lysogenic infection with N15, as the peak in the termination region in these cells is unchanged (Figure 29, panels I and II). If the over-replication was initiated from an origin-like sequence within the termination region (Kogoma, 1997), it is likely that the insertion of the *tos-kan* construct would disrupt it and the over-replication would be abolished but this is not what we see and

instead the peak appears the same as that seen in *ΔxonA ΔxseA* cells (cf. Figure 26 and Figure 29). However, origin-independent over-replication in the termination area of cells lacking 3' exonucleases is completely abolished when the linearization sequence *tos* is combined with N15 lysogenic infection to linearise the chromosome (Figure 29, panel III).

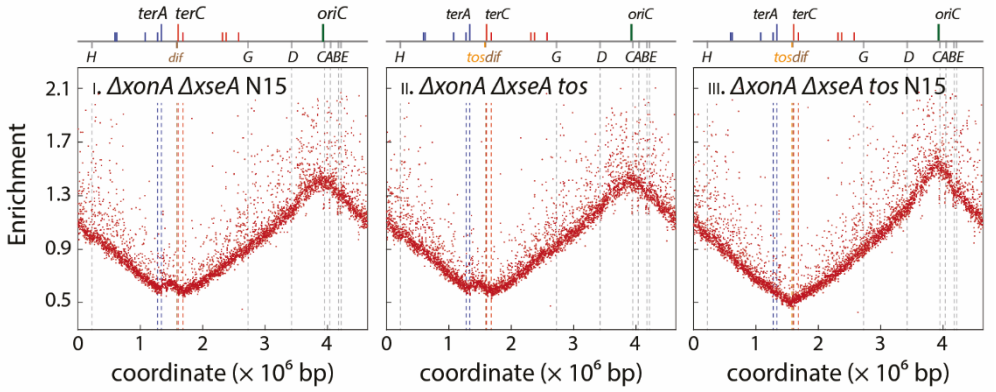

Figure 29: Replication profiles of *E. coli* cells showing the effect of chromosome linearization on replication in *ΔxonA ΔxseA* cells. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Introduction of the *tos* linearization sequence to *ΔxonA ΔxseA* cells was followed by lysogenic infection with N15, resulting in the expression of the telomerase TelN that processes *tos*. Sequencing templates were isolated from SLM1213 (*ΔxonA ΔxseA* N15), SLM1187 (*ΔxonA ΔxseA tos*) and SLM1212 (*ΔxonA ΔxseA tos* N15 lysogen). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

The effect of chromosome linearisation on over-replication in the termination area of *ΔxonA ΔxseA* cells differs slightly to what is seen in *ΔrecG* cells following chromosome linearisation. The profile of a linearised *ΔrecG* strain showed that the over-replication within the termination area is significantly reduced following chromosome linearisation but it is not abolished entirely. Instead, there is amplification of the sequences on either side of the *tos-kan* construct (Rudolph et al., 2013), which is different to the complete lack of amplification seen in *ΔxonA ΔxseA* cells with a linearised chromosome.

The growth of *dnaA Δtus rpo* ΔxonA ΔxseA* cells with either a circular or a linearised chromosome was assessed in the presence and absence of origin firing.

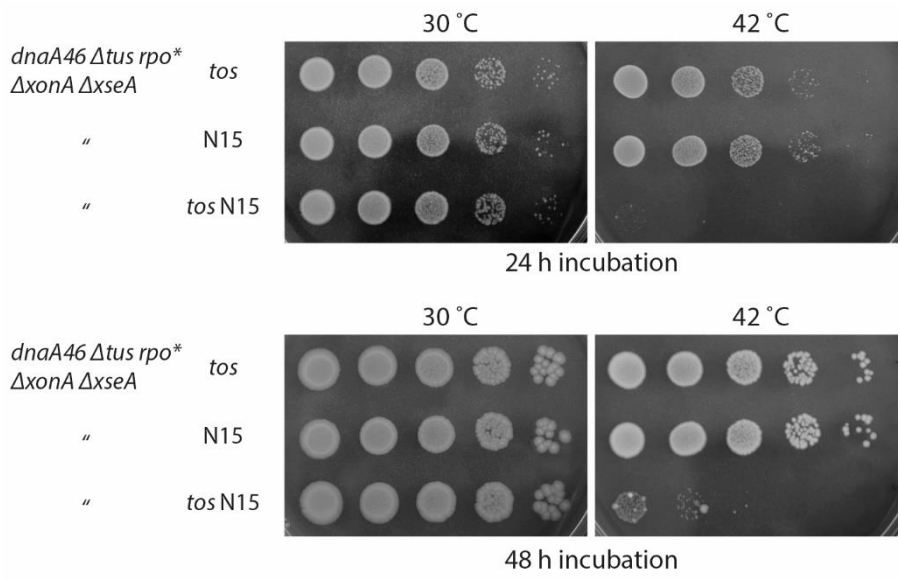

Figure 30: Spot dilution assays to evaluate DnaA-independent growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells with a linearised chromosome. Introduction of the *tos* linearization sequence to *dnaA Δtus rpo* ΔxonA ΔxseA* cells was followed by lysogenic infection with N15, resulting in the expression of the telomerase TelN that processes *tos*. The strains used were SLM1225 (*dnaA46 Δtus rpo* ΔxonA ΔxseA tos*), SLM1232 (*dnaA46 Δtus rpo* ΔxonA ΔxseA* N15) and SLM1230 (*dnaA46 Δtus rpo* ΔxonA ΔxseA tos* N15 lysogen).

In line with the replication profile data, the spot dilution assay in Figure 30 shows that the ability of *dnaA Δtus rpo* ΔxonA ΔxseA* cells to grow in the absence of origin firing is almost entirely abolished in the strain with a linearised chromosome, as is seen in cells lacking RecG. Linearisation does not affect the growth of the strain at the permissive temperature of 30 °C when the DnaA(ts) protein is functional and able to initiate DNA replication, which reiterates that *E. coli* cells with a linearised chromosome are viable and are able to grow as normal. The non-lysogenic *tos* strain and the N15 lysogen without the linearization sequence both demonstrate that there is no reduction in origin-independent growth in *dnaA Δtus rpo* ΔxonA ΔxseA* strains that retain a circular chromosome (Figure 30; see Figure 24 and Figure 27 for the *dnaA Δtus rpo* ΔxonA ΔxseA* control). Therefore, the almost total loss of ability to grow in the absence of origin firing seen in the *dnaA Δtus rpo* ΔxonA ΔxseA tos* N15 lysogen is a direct result of linearising the chromosome. The most obvious change effected by the linearization of the chromosome at *dif* is that replication forks will not meet in the termination region but will instead run in to the hairpin structures generated by phage N15 processing of the *tos* sequence (Figure 41) (Cui et al., 2007). Whilst it is possible that other explanations might apply, the clear absence of origin-independent DNA synthesis and origin-independent growth in *ΔxonA ΔxseA* cells with a linearised chromosome strongly supports the idea that replication fork fusions are responsible for generating intermediates that can trigger the observed over-replication.

Origin-independent over-replication requires PriA helicase activity in cells lacking 3' exonucleases

Growth experiments of cells lacking RecG in the presence and absence of PriA helicase activity show that the robust origin-independent growth seen in *ΔrecG* cells is abolished in *priA300* mutants. In line with this, marker frequency analysis revealed that there is no peak of origin-independent over-replication in the termination area of *ΔrecG* cells when the helicase activity of PriA is disabled by the introduction of the *priA300* allele (Rudolph et al., 2013).

To establish further the requirements for origin-independent replication and growth seen in cells lacking 3' exonucleases, we investigated how the presence and absence of the helicase activity of PriA affected *ΔxonA ΔxseA* cells.

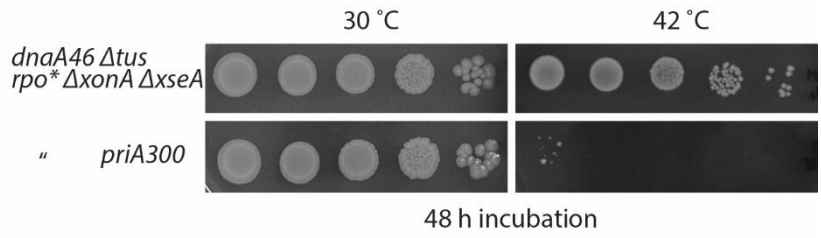

Figure 31: DnaA-independent growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells in the presence and absence of the helicase activity of PriA. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The strains used were SLM1194 (*dnaA46 Δtus rpo*ΔxonA ΔxseA*) and SLM1198 (*dnaA46 Δtus rpo* ΔxonA ΔxseA priA300*).

Introduction of a *priA300* allele in to a *dnaA Δtus rpo* ΔxonA ΔxseA* strain had no effect on growth at the permissive temperature of 30 °C when DnaA(ts) protein is functional and able to initiate DNA replication at *oriC*. As is seen in the *ΔrecG* background however, the ability of *dnaA Δtus rpo* ΔxonA ΔxseA priA300* cells to grow at 42 °C was abolished, which means that *oriC*- and DnaA-independent DNA replication is not able to sustain growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells when the helicase activity of PriA is disabled (Figure 31).

In cells lacking RecG, the effect of *priA300* on origin-independent growth is reflected in the disappearance of the peak of over-replication seen in the termination area of *ΔrecG* cells from the replication profile. This is explained by the model proposed for initiation of origin-independent DNA replication that a 3' flap generated as a result of a replication fork fusion event is able to persist in the absence of RecG and can be targeted by PriA to load DnaB helicase and initiate replication. Helicase activity is required by PriA to modulate the DNA fork structure in order to achieve this (Rudolph et al., 2009b, 2010a, 2013).

Given the results of the growth experiments and the *ΔrecG priA300* replication profile, it seemed likely that the peak of over-replication seen in the termination area of *ΔxonA ΔxseA* cells would be

eradicated in a *priA300* strain. We constructed a *ΔxonA ΔxseA priA300* strain and prepared the chromosomal DNA sample from exponentially growing cells in order to generate a replication profile through marker frequency analysis. The samples have been sent to Earlham Institute for marker frequency analysis via deep sequencing, but we have not received the data back yet and so are unable to generate a replication profile for this strain.

srgA1 was found to be a suppressor of *ΔrecG* mutant phenotypes (Al-Deib et al., 1996). The fork structure that *srgA1* is unable to process is effectively a 3' flap, which is the predicted substrate generated in replication fork fusion events in the absence of RecG according to the model for origin-independent DNA replication developed by Rudolph and colleagues (Dimude et al., 2015; Rudolph et al., 2009b, 2010a, 2013). The peak of over-replication in the termination region of *ΔrecG* cells is abolished in the replication profile of *ΔrecG srgA1* double mutants (Rudolph et al., 2013). Growth experiments in *dnaA Δtus rpo* ΔrecG srgA1* cells revealed that the origin-independent growth seen in *ΔrecG* cells is also almost entirely abolished (reduced over 1000-fold) in the presence of the mutant PriA[L557P] protein, supporting the idea that a 3' flap structure generated during replication termination, which would normally be eliminated by RecG, is able to persist in the absence of RecG and be targeted by PriA for replication initiation (Rudolph et al., 2013).

In order to establish if the 3' flap processing ability of PriA is essential for the origin-independent growth seen in cells lacking 3' exonucleases, we utilised the spot dilution assay to assess origin-independent growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells in the presence and absence of the *srgA1* allele.

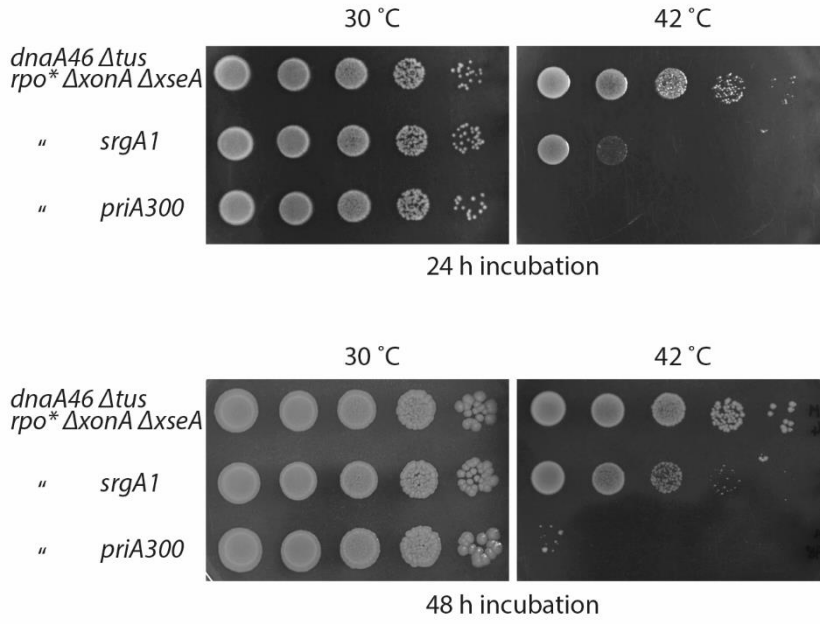

Figure 32: Spot dilution assays to evaluate the effect of PriA helicase mutations on DnaA-independent growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The strains used were SLM1194 (*dnaA46 Δtus rpo* ΔxonA ΔxseA*), SLM1199 (*dnaA46 Δtus rpo* ΔxonA ΔxseA srgA1*) and SLM1198 (*dnaA46 Δtus rpo* ΔxonA ΔxseA priA300*). SLM1194 and SLM1198 are reproduced from Figure 31.

As with the *priA300* mutation, the presence of the *srgA1* allele had no effect on growth of *dnaA Δtus rpo* ΔxonA ΔxseA* cells at the permissive temperature of 30 °C when DnaA(ts) protein is functional and able to initiate DNA replication. In contrast to what was observed in *ΔrecG* cells (Rudolph et al., 2013), we found that the *dnaA Δtus rpo* ΔxonA ΔxseA srgA1* strain was able to grow at 42 °C via origin-independent DNA replication (Figure 32). The colonies grow slowly and the inactivation of the 3' flap processing ability of PriA (*srgA1*) does reduce growth by at least one order of magnitude, but it is not abolished to the same extent as that seen in both the *dnaA Δtus rpo* ΔxonA ΔxseA* and *dnaA Δtus rpo* ΔrecG* strains lacking the full helicase activity of PriA or the *dnaA Δtus rpo* ΔrecG srgA1* strain (Figure 31; Rudolph et al., 2013). This suggests that in contrast to *ΔrecG* cells, in the absence of 3' exonucleases there are at least two independent pathways for initiation of origin-independent DNA replication, one which relies on PriA targeting of 3' flaps and a second one which does not.

This idea is supported by the replication profiles generated via marker frequency analysis. The peak of over-replication seen in the termination area of the replication profile of *ΔxonA ΔxseA* cells is not abolished by the presence of the *srgA1* helicase (Figure 33 A) in contrast to what is seen in *ΔrecG srgA1* cells (Rudolph et al., 2013). Therefore, origin-independent replication is able to initiate in Δ*xonA ΔxseA srgA1* cells without the 3' flap processing ability of PriA helicase. The profiles of the *ΔxonA* and *ΔxseA* single mutant strains are unaffected by the addition of the *srgA1* allele (compare Figure 33 B, panels I and II with Figure 25, panels II and III respectively. All profiles were produced from the same sequencing run). The profile of *srgA1* itself is equivalent to the wild-type profile (compare Figure 33 B, panel III with Figure 25, panel I; the wild-type profile was produced from the same sequencing run as the *srgA1* profile).

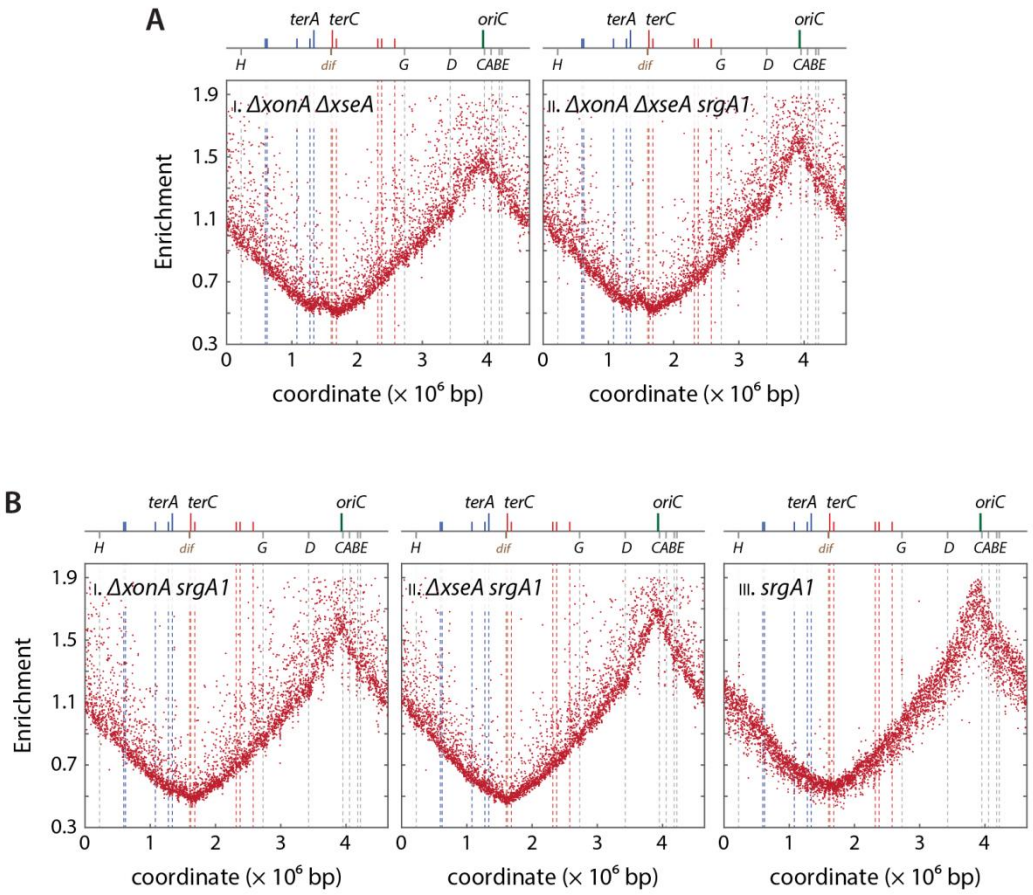

Figure 33: Replication profiles of *E. coli* cells with a *srgA1* point mutation in the gene for PriA. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Sequencing templates were isolated from SLM1203 (*ΔxonA ΔxseA*), SLM1186 (*ΔxonA ΔxseA srgA1*), SLM1172 (*ΔxonA srgA1*), SLM1214 (*ΔxseA srgA1*) and JJ1264 (*srgA1*). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

Over-replication initiating at 5' flap structures

The accumulation of 3' flap structures in the termination region is what would be expected in cells lacking 3' exonucleases, which could subsequently be targeted by PriA to initiate new replication forks that are then responsible for the over-replication seen in the absence of 3' exonucleases (Figure 23, Figure 26) (Rudolph et al., 2013). So why does the introduction of the *srgA1* allele of PriA not abolish all origin-independent replication and growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells (Figure 32, Figure 33) when it does have that affect in *ΔrecG* derivatives? RecG protein is present in cells lacking one or more 3' exonucleases, and in fact becomes essential for viability in the absence of all three 3' exonucleases (Rudolph et al., 2010b). Whilst 3' exonucleases have a relatively narrow activity and

substrate specificity (Lovett, 2011), RecG has been shown to unwind a variety of structures (Lloyd and Rudolph, 2016). If RecG is able to convert 3' flap structures to 5' flaps *in vivo*, then one explanation might be that because RecG is still present in these cells, it is able to convert 3' flaps generated at replication fork fusion events to 5' flaps, which are normally degraded by exonucleases with 5' to 3' polarity, such as RecJ or ExoVII (Lovett, 2011) but might occasionally provide a substrate for the initiation of origin-independent replication instead. If this is the case, cells lacking 5' flap processing should show an increase in origin-independent over-replication. This prompted the idea that the growth of *dnaA Δtus rpo* ΔxonA ΔxseA srgA1* cells at 42 °C might be due to the absence of ExoVII reducing the 5' flap processing capability of the cells and allowing PriA the opportunity to establish origin-independent synthesis at 5' flaps. To establish if this is the case, we utilised the spot dilution assay to assess the growth of *dnaA Δtus rpo* ΔxonA* and *dnaA Δtus rpo* ΔxonA srgA1* cells at permissive and restrictive temperatures.

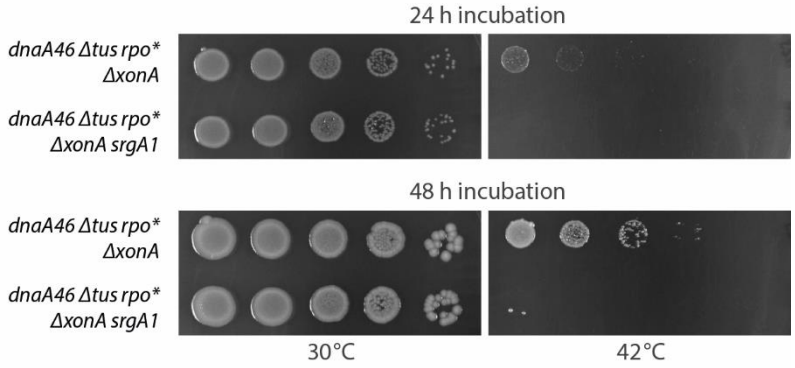

Figure 34: Spot dilution assay to evaluate the effect of *srgA1* allele of PriA on DnaA-independent growth in *dnaA Δtus rpo* ΔxonA* cells. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The strains used were RCe528 (*dnaA46 Δtus rpo*ΔxonA*) and SLM1110 (*dnaA46 Δtus rpo*ΔxonA srgA1*).

As shown in Figure 34, in the presence of ExoVII, the origin-independent growth seen in *dnaA Δtus rpo* ΔxonA* cells is abolished when the *srgA1* allele of PriA is introduced, demonstrating that the origin-independent growth in the absence of ExoI is in fact dependent on the 3' flap-processing ability of PriA. *dnaA Δtus rpo* ΔxonA srgA1* cells can only grow at the restrictive temperature when the 5' exonuclease capability of the cells has been diminished via the deletion of *ΔxseA*.

To investigate this further, we then generated replication profiles of *ΔxonA ΔxseA* strains in the presence and absence of RecJ, an exonuclease with 5'-3' activity that specifically degrades ssDNA (Lovett, 2011). If origin-independent over-replication can initiate from 5' flap substrates in some way, then removing RecJ should result in an increase in over-replication in cells lacking 3' exonucleases.

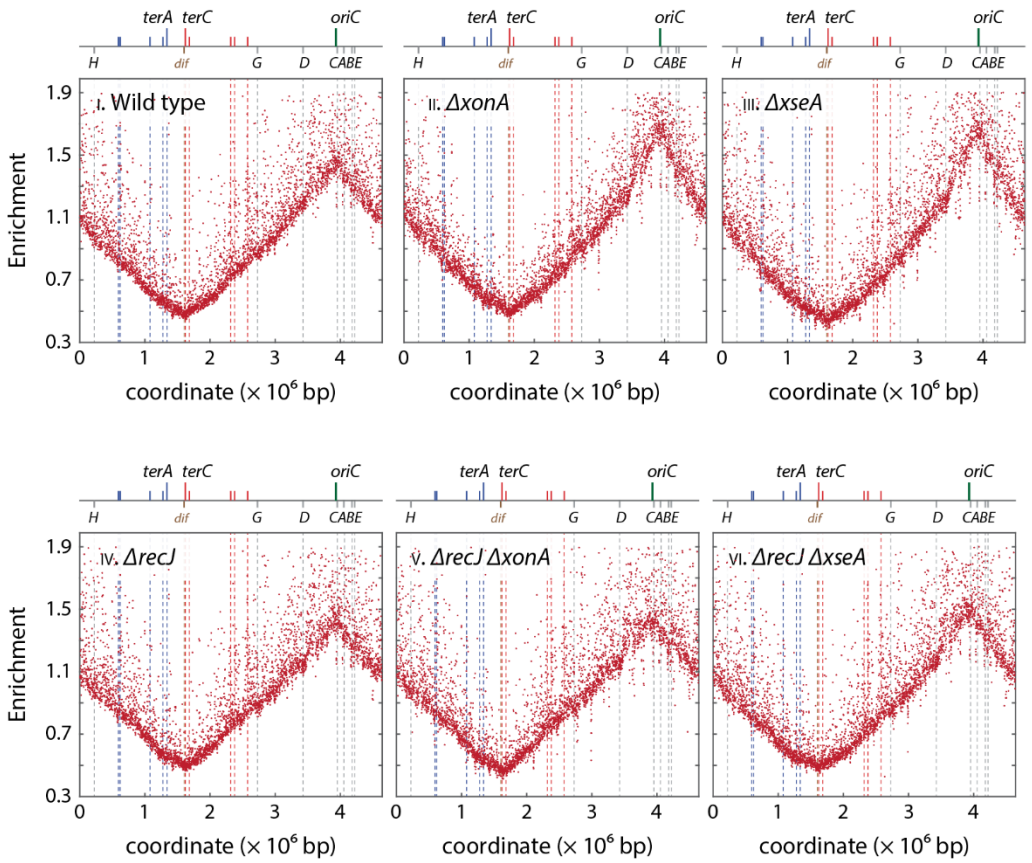

Figure 35: Replication profiles of *E. coli* cells showing the effect of *recJ* deletion on replication in Δ*xonA* and Δ*xseA* cells. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Sequencing templates were isolated from MG1655 (wild type), RCe563 (Δ*xonA*), SLM1185 (Δ*xseA*), N4934 (Δ*recJ*), SLM1178 (Δ*recJ* Δ*xonA*) and SLM1204 (Δ*recJ* Δ*xseA*). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

The absence of RecJ alone in an otherwise wild-type background does not alter the replication profile from that seen in wild-type cells. The replication profiles of *ΔxonA* and *ΔxseA* cells are also unaffected by the absence of RecJ (Figure 35). It is possible that origin-independent over-replication is not able to initiate from 5' flaps that persist in *ΔrecJ* cells, and so no effect is seen in the replication profiles of *ΔrecJ* derivatives. It is also possible that ExoVII (*xseA*) masks any effect of the absence of RecJ in *ΔrecJ ΔxonA* cells, as ExoVII has 5' – 3' exonuclease activity as well as being a 3' exonuclease, and so this enzyme is likely to be compensating for *ΔrecJ*. However, as no peak of over-replication is seen in the *ΔxonA* and *ΔxseA* single mutants (Figure 25) it is perhaps not surprising that the absence of RecJ does not influence the replication profiles of these strains.

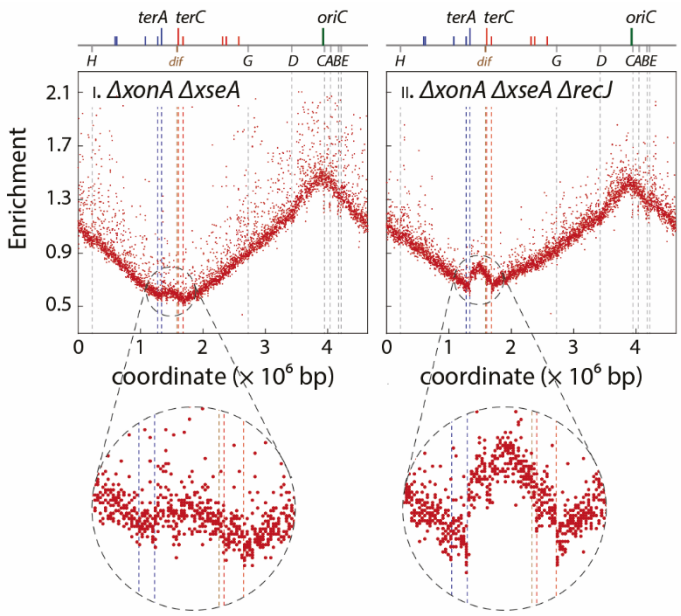

Figure 36: Replication profiles of *E. coli* cells showing the effect of *recJ* deletion on replication in Δ*xonA* Δ*xseA* cells. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Sequencing templates were isolated from SLM1203 (Δ*xonA* Δ*xseA*) and SLM1188 (Δ*xonA* Δ*xseA* Δ*recJ*). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

In contrast, the absence of RecJ in a *ΔxonA ΔxseA* background results in a perceptible difference to the replication profile. There is a significant increase in the size of the peak of over-replication in the termination area of these cells (Figure 36), which means that there is a higher incidence of origin-independent replication in the termination area in *ΔxonA ΔxseA* cells in the absence of RecJ compared to that seen when RecJ is present to degrade 5' flap structures. If in the absence of 3' exonucleases, 3' flaps are converted to 5' flaps and in the absence of 5' exonucleases ExoVII and RecJ, instead of being degraded can be targeted as a substrate for replication initiation, then we would expect an increase in the over-replication seen in the termination region of *ΔxonA ΔxseA* cells when RecJ is absent, and this is exactly what we see (Figure 36).

To establish if the increase in over-replication of the termination area of *ΔxonA ΔxseA* cells in the absence of RecJ translates in to an increase in origin-independent growth, a spot dilution assay was used to assess the growth of *dnaA Δtus rpo* ΔxonA* cells at 42 °C in the presence and absence of RecJ. Until this point, *ΔxonA ΔxseA* cells had been utilised instead of cells lacking all three 3' exonucleases. The origin-independent growth in *dnaA Δtus rpo* ΔxonA ΔxseA* cells (Figure 24) is robust enough that it would not be possible using the spot dilution assay to measure any improvement in growth that deleting *recJ* might allow. *dnaA Δtus rpo* ΔxonA* cells achieve single colony growth at the 1×10^{-3} dilution at 42 °C and therefore allow for an increase in origin-independent growth to be assessed and the effect of Δ*recJ* on origin-independent growth in *dnaA Δtus rpo* ΔxonA* cells to be determined.

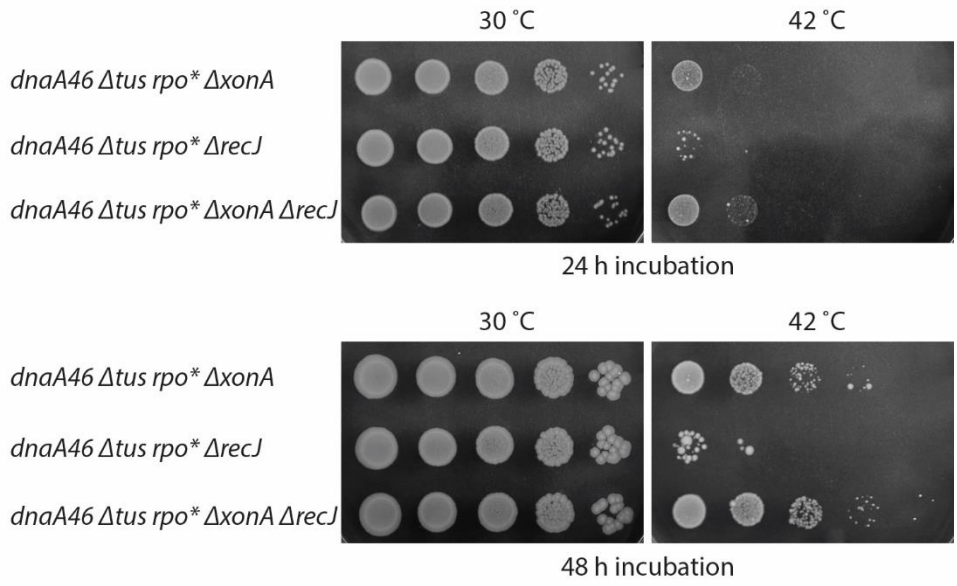

Figure 37: Spot dilution assays to evaluate the effect of *recJ* deletion on DnaA-independent growth in *dnaA tus rpo* xonA* cells. The strains carry a temperature sensitive allele for the DnaA initiator protein, *dnaA46*, which produces a protein that is functional at 30°C but not at 42°C. Growth occurring at the restrictive temperature is achieved in the absence of *oriC* firing. The strains used were SLM1218 (*dnaA46 Δtus rpo* ΔxonA*), SLM1233 (*dnaA46 Δtus rpo* ΔrecJ*) and SLM1224 (*dnaA46 Δtus rpo* ΔxonA ΔrecJ*).

There was no difference in growth of *dnaA Δtus rpo** cells at either the permissive (30 °C) or restrictive (42 °C) temperature in the absence of RecJ (Figure 37). The minimal growth and large colony variants seen at 42 °C in the absence of origin firing is equivalent to that seen previously in the *dnaA Δtus rpo** strain (Figure 24). In *dnaA Δtus rpo* ΔxonA* cells however, origin-independent growth is at least doubled in the absence of RecJ, suggesting that replication initiating at 5' flaps in the absence of RecJ can indeed contribute to the origin-independent growth seen in *dnaA Δtus rpo* ΔxonA* cells. This is in line with the observed increase of origin-independent DNA synthesis seen in the termination area of *ΔxonA ΔxseA* cells in the absence of RecJ (Figure 36).

DNA replication in strains carrying an ectopic replication origin

To gain further insights in to the events that trigger initiation in the termination area in the absence of 3' exonucleases, we generated strains with an additional copy of *oriC* inserted into the chromosome. This copy of *oriC*, termed *oriZ*, is located roughly a quarter of the way around the right-hand replichore, clockwise from *oriC* (Figure 13) (Ivanova et al., 2015; Wang et al., 2011). Previous work on double origin strains lacking RecG shows that there is a substantial increase in the peak of over-replication seen in the termination area of Δ*recG* cells when *oriZ* is present (Rudolph et al., 2013). A further significant result was that the shorter interval between *oriC* and *oriZ* that forms a new, ectopic

location of fork fusion events was significantly shallower in the absence of RecG. This is indicative of amplification of the region, which is consistent with over-replication initiated as a result of fork fusions taking place in this location due to the presence of the ectopic origin. In the absence of *ter*/Tus complexes creating a defined termination area, over-replication of the chromosome will not appear as a defined peak like that seen in the native termination area and instead will lead to a broader increase in marker frequency of the region (Midgley-Smith et al., 2018; Rudolph et al., 2013).

Is the peak of over-replication in *ΔxonA ΔxseA* cells affected by the presence of an ectopic replication origin? To investigate this, replication profiles of *ΔxonA*, *ΔxseA* and *ΔxonA ΔxseA* cells in an *oriC*⁺ *oriZ*⁺ background were established (Figure 38). The equal peak height of the two origin regions confirms synchronous initiation occurs at *oriC* and *oriZ* as reported previously (Rudolph et al., 2013; Wang et al., 2011). The double origin profiles are asymmetric as is expected, which is due to the fact that the clockwise replication fork initiating at *oriZ* will reach the termination area in advance of the anti-clockwise fork initiated at *oriC*, and so will proceed through the termination area and be blocked at *terC*/Tus or *terB*/Tus. On a population basis, this results in many more cells having replicated the lower half of the right-hand replichore including the *terA – terC/B* region compared with the area behind *terB*. This is what causes the step in the replication profile at the termination area.

In contrast to what is seen in *oriC*⁺ *oriZ*⁺ *ΔrecG* cells, the replication profile of *oriC*⁺ *oriZ*⁺ *ΔxonA ΔxseA* cells shows that the peak of over-replication in the termination area of *ΔxonA ΔxseA* cells is not substantially increased by the presence of an ectopic replication origin, and in fact is not visible in the profile (Figure 38, panels I and IV). This does not necessarily mean that the over-replication is abolished however; the replication profile is so skewed in the termination area due to the distorted physiology of the chromosome and it is possible that this is enough to mask the relatively minor peak that is present in *ΔxonA ΔxseA* cells.

Is the profile of the shorter chromosomal region between the two origins altered in the absence of ExoI and ExoVII, as is seen in cells lacking RecG (Midgley-Smith et al., 2018; Rudolph et al., 2013). In the absence of a replication profile of *oriC*⁺ *oriZ*⁺ cells from this sequencing run, the *oriC*⁺ *oriZ*⁺ *ΔxonA* profile was used as the comparison, as the profile of a *ΔxonA* single mutant shows no obvious differences to the profile of wild type cells (Figure 25). An overlay of the replication profiles for *oriC*⁺ *oriZ*⁺ *ΔxonA* and *oriC*⁺ *oriZ*⁺ *ΔxseA* cells shows the similarity between the two profiles (Figure 44, panel I), indicating that any differences between the profiles of *oriC*⁺ *oriZ*⁺ *ΔxonA* and *oriC*⁺ *oriZ*⁺ *ΔxonA ΔxseA* must arise as a result of the combined deletion of *xonA* and *xseA*. All strains involved were sequenced as part of the same sequencing run, and the same stationary phase sample was used for normalising the data sets of all three strains. If over-replication occurs in the ectopic fork fusion region in *oriC*⁺ *oriZ*⁺ *ΔxonA ΔxseA* cells, an increase in the marker frequency of this region in comparison to *xonA*⁺ *xseA*⁺ cells would be expected. Overlaying the *oriC*⁺ *oriZ*⁺ *ΔxonA ΔxseA* profile on to the *oriC*⁺ *oriZ*⁺ *ΔxonA* profile, aligned according to the *oriC* peak height, reveals that the valley between the origins is less defined in the absence of both ExoI and ExoVII (Figure 38, panel V). The effect is subtle, and there might be alternative explanations that explain this, but the result warrants a more detailed analysis in

the future. Growth experiments confirmed that the doubling times of *oriC*⁺ *oriZ*⁺ *ΔxonA* and *oriC*⁺ *oriZ*⁺ *ΔxonA ΔxseA* cells are very similar (data not shown; Midgley-Smith et al., 2019), suggesting that the frequency of origin firing is unlikely to be affected by the presence or absence of *xseA* and if this is the case, the alignment via the origin peal height is justified.

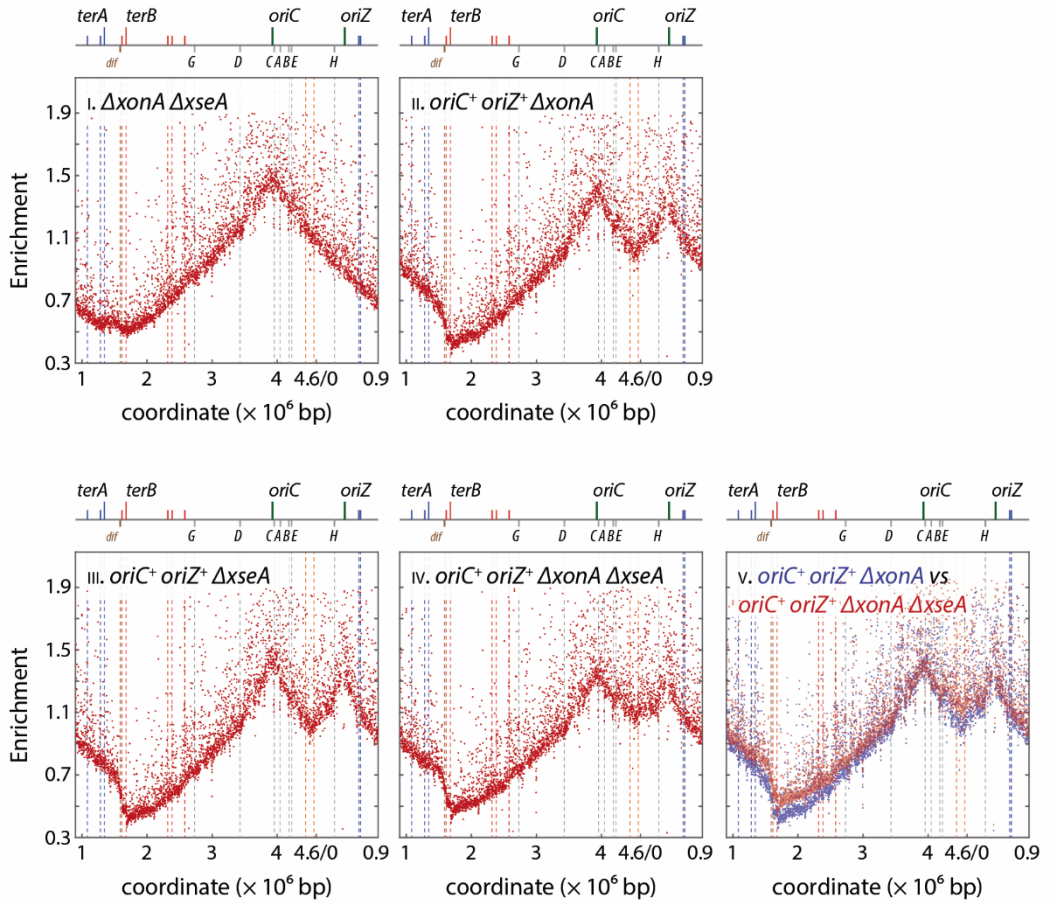

Figure 38: Replication profiles of *E. coli* cells with two replication origins lacking *ΔxonA*, *ΔxseA* or both. Shown is the marker frequency analysis of exponentially growing cultures generated via deep sequencing. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal location. The chromosome coordinates are offset to start 0.9 Mb to make it easier to see the shorter interval between the replication origins. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* (green line) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and the chromosome dimer resolution site *dif* below. Sequencing templates were isolated from SLM1203 (*ΔxonA ΔxseA*), SLM1206 (*oriC*⁺ *oriZ*⁺ *ΔxonA*), SLM1208 (*oriC*⁺ *oriZ*⁺ *ΔxseA*) and SLM1217 (*oriC*⁺ *oriZ*⁺ *ΔxonA ΔxseA*). Marker frequency profiles for key constructs have been generated independently twice. Only one representative replication profile is shown.

That the difference between the two profiles (Figure 38, panel V) is subtle is not surprising given the relatively small peak seen in the native termination area of *ΔxonA ΔxseA* cells despite the fact that in this region, *ter*/Tus traps are present to focus the over-replication. I have since created a double origin strain that has *ter* sites inserted in between the two origins, flanking the region where most forks meet, thereby creating an ectopic fork trap. When *ΔrecG* is introduced to this strain, the broad amplification of the ectopic fork fusion region seen in the replication profile of *oriC*⁺ *oriZ*⁺ *ΔrecG* cells

without the ectopic *ter* sites is replaced by a clear peak of over-replication, which strongly supports the idea that origin-independent over-replication in *ΔrecG* cells is co-localised with fork fusion events (Figure 15) (Midgley-Smith et al., 2018). The double origin ectopic *ter* strain could be used to determine if the same is true for the over-replication in the ectopic termination region of cells lacking 3' exonucleases. These experiments are currently being undertaken in the Rudolph lab.

ExoI and ExoVII are not essential for the survival of *oriZ*$^+$ cells lacking *oriC*

It has previously been shown that it is possible to delete the entire *oriC* region in a strain carrying the ectopic replication origin, *oriZ*, which can provide insights in to the established architecture of *E. coli* chromosomes and problems that arise when this is altered (Ivanova et al., 2015). We recently found that it was not possible to delete *oriC* in this strain background in the absence of RecG. By exploiting a synthetic lethality assay (page 61), we were able to confirm that *ΔoriC oriZ*$^+$ *ΔrecG* cells are not viable (Midgley-Smith et al., 2018). This assay uses a gene expressed from pRC7, an unstable plasmid that is rapidly lost and carries a functional *lac* operon, to cover chromosomal deletion of the gene in a *Δlac* strain background. Following introduction of further mutations, plasmid-free colonies can be distinguished from those where the covering plasmid is present by plating the cultures on to plates containing IPTG and X-gal (Bernhardt and de Boer, 2004). Mutations that abolish over-replication in *ΔrecG* cells (*priA300*) or that allow forks arising from origin-independent replication to leave the termination area (*Δtus*) were found to restore viability of *ΔoriC oriZ*$^+$ *ΔrecG* cells (Midgley-Smith et al., 2018).

We investigated *oriC*$^+$ *oriZ*$^+$ strains lacking 3' exonucleases further by attempting to delete *oriC* in this strain background. Given that this is not possible in *ΔrecG* cells without a covering plasmid, we investigated this initially using a synthetic lethality assay. We created *ΔxonA* and *ΔxonA ΔxseA* derivatives of *oriC*$^+$ *oriZ*$^+$ *Δlac* cells, and then transformed the resulting strains with the pAM488 *lac*$^+$ plasmid expressing *xonA* to cover the chromosomal *xonA* deletion. Using P1*vir* transduction, we then generated *ΔoriC::kan* derivatives of these strains. We confirmed the deletion of *oriC* from the resulting strains through PCR analysis of the *oriC* chromosomal region (Figure 45).

Briefly, samples of *ΔoriC oriZ*$^+$ *Δlac ΔxonA* p*xonA*$^+$ and *ΔoriC oriZ*$^+$ *Δlac ΔxonA ΔxseA* p*xonA*$^+$ cells were extracted from exponential phase cultures and dilutions were plated on LB agar or M9 glucose minimal salts agar supplemented with IPTG and X-gal. IPTG induces expression of the *lac* operon, which results in the production of β-galactosidase. This hydrolyses X-gal and the product of this reaction produces a blue pigment. Cells without the plasmid do not produce a blue colour. Blue and white colonies were counted after incubation at 37 °C. If the deletion of *oriC* is lethal in cells lacking 3' exonucleases, cells in which the covering plasmid is lost will fail to grow and the *lac*$^+$ cells that retain the plasmid will form blue colonies only. If viability is not affected, both blue and white colonies will

form, and if viability is reduced but not eliminated, plasmid-free cells will form white colonies that will be noticeably smaller than the blue colonies.

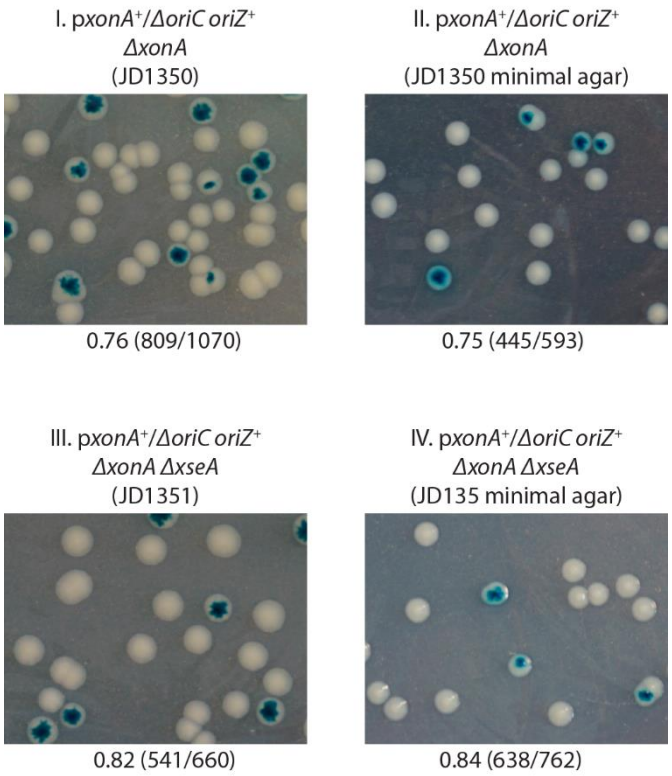

Figure 39: Viability of Δ*oriC oriZ*+ cells in the absence of 3' exonucleases. Photographs of synthetic lethality assays show the effect of *xonA* and *xseA* deletions on the viability of Δ*oriC oriZ*+ cells. The images are representative of the total plate in each case. The relevant genotype and strain number of each strain used is shown above each image, and the number of white colonies (which have lost the covering plasmid) is shown below as a fraction of the total number of colonies of that plate. The actual colony numbers for each plate is shown in parentheses as number of white/total colonies counted.

The presence of both white and blue colonies and segregation within the blue colonies in the synthetic lethality assay revealed that *ΔoriC oriZ*+ cells are viable in the absence of *ΔxonA* or *ΔxonA ΔxseA*, in contrast to what is seen in the absence of RecG (Figure 39) (Midgley-Smith et al., 2018). The fact that the white and blue colonies are the same size shows that plasmid-free cells are able to grow as robustly as those that retain the plasmid, which means that *ΔoriC oriZ*+ *ΔxonA* and *ΔoriC oriZ*+ *ΔxonA ΔxseA* cells are not only viable but are healthy.

When ampicillin-containing agar plates were initially inoculated with frozen stocks of the two strains used in this assay, large colony variants present on both plates indicated that suppressor mutations develop in both strain constructs, as reported previously for *ΔoriC oriZ*+ constructs (Ivanova et al., 2015). However, I noticed that on both these plates and the subsequent LB IPTG X-gal plates from the initial experiment that there were more suppressor colonies on the Δ*oriC oriZ*+ Δ*xonA* pAM488 (*xonA*+) (JD1350) plates than on the *ΔoriC oriZ*+ *ΔxonA ΔxseA* pAM488 (*xonA*+) plates. Given the fact that the presence of ampicillin selects for cells carrying the plasmid, which contains a gene for ampicillin resistance, the Δ*oriC oriZ*+ Δ*xonA* pAM488 (*xonA*+) strain is purely Δ*oriC oriZ*+, and JD1351

is the same but with *xseA* deleted. It is therefore possible that pathology seen normally in Δ*oriC oriZ*$^{+}$ strains is reduced slightly in the absence of ExoVII; further investigation would be required to confirm this.

Discussion

It was shown previously that in the absence of either RecG helicase or 3' exonucleases, replication can initiate independently of both DnaA and *oriC* and it was proposed that this might arise as a result of PriA-mediated replication initiated at 3' flap structures generated as two replication forks fuse (Rudolph et al., 2010a, 2013). Recent work investigating the origin-independent synthesis seen in *ΔrecG* cells provides compelling evidence in support of this and is in line with a role for RecG in managing events at replication termination, including data presented earlier in this thesis (page 65) (Dimude et al., 2015; Lloyd and Rudolph, 2016; Midgley-Smith et al., 2018; Rudolph et al., 2009a, 2009b, 2013). Based on the hypothesis that 3' flap structures might result from replication fork fusions and the fact that over-replication of the termination area was observed in *ΔxonA ΔxseA ΔsbcCD* cells, it was suggested that 3' exonucleases might also have a role at replication termination (Rudolph et al., 2013).

While the results presented previously are certainly in line with this hypothesis, other explanations could apply, as the involvement of 3' exonucleases in termination had not been rigorously tested. The data presented here confirm that in the absence of 3' exonucleases, origin-independent synthesis leads to over-replication of the termination region specifically (Figure 26), as reported previously (Rudolph et al., 2013; Wendel et al., 2014). I have shown that in cells lacking 3' exonucleases, this over-replication is able to sustain robust growth in the absence of origin firing (Figure 24), and that this is only possible when the replication fork trap is inactivated by the deletion of *tus* (Figure 27). Replication forks leaving the termination area will proceed in an orientation opposite to normal, resulting in conflicts with transcription complexes, and indeed, origin-independent growth is improved further when a point mutation that alleviates replication-transcription conflicts (*rpo**) (Dutta et al., 2011; Ivanova et al., 2015; McGlynn et al., 2012) is introduced (Figure 27). The fact that *dnaA ΔxonA ΔxseA* cells cannot grow in the absence of origin firing when *dnaA Δtus rpo* ΔxonA ΔxseA* cells can (Figure 27), despite the over-replication seen in the termination region of *ΔxonA ΔxseA* cells (Figure 26), strongly supports the idea that the origin-independent DNA synthesis in cells lacking 3' exonucleases is initiated as a result of events occurring within the termination region, in line with the idea that 3' exonucleases normally perform some function at termination that inhibits this over-replication. These data show similarities to the origin-independent synthesis in cells lacking RecG (Rudolph et al., 2013).

In line with this, similarities were found in the genetic requirements for origin-independent synthesis to initiate in cells lacking either RecG or 3' exonucleases. A circular chromosome has been

found to be a requirement for the over-replication in both *ΔrecG* and 3' exonucleases-deficient strains. Replication terminates within the termination region when two converging forks fuse. If in the absence of 3' exonucleases origin-independent over-replication initiates due to events at replication fork fusions, then it would be expected that by preventing forks from meeting, the over-replication would disappear, and this is exactly what is seen. Linearising the chromosome at a location near to *dif* within the termination region (Cui et al., 2007) prevents replication forks from meeting. The data reported here show that origin-independent growth is almost entirely abolished in *dnaA Δtus rpo* ΔxonA ΔxseA* cells when the chromosome is linearised (Figure 30) as is seen in cells lacking RecG (Dimude et al., 2015; Rudolph et al., 2013). The peak of over-replication in the replication profile of *ΔxonA ΔxseA* cells is also abolished on chromosome linearization (Figure 29), demonstrating that in these cells, the origin-independent over-replication cannot initiate. Thus, in the absence of either RecG or 3' exonucleases, origin-independent synthesis is only able to initiate in cells with a circular chromosome, and the necessity for fork fusion events in generating substrates that PriA can target for re-initiation provides an explanation for this observation.

In addition, previous work has shown that in *ΔrecG* cells, the initiation of origin-independent replication is dependent on the helicase activity of PriA, specifically its ability to process 3' flaps (Dimude et al., 2015; Rudolph et al., 2013). Similarly, my data show that the helicase activity of PriA is also essential for origin-independent growth in the absence of 3' exonucleases (Figure 31). Initially, it appeared that the specific 3' flap processing ability of PriA was not essential for origin-independent growth as in contrast to the situation in the absence of RecG, origin-independent growth in cells lacking ExoI and ExoVII was not abolished when a *srgA1* mutant allele was introduced (Figure 32). However, further investigation revealed that when ExoVII is present, the origin-independent growth of *dnaA Δtus rpo* ΔxonA* cells is in fact entirely abolished by the introduction of the *srgA1* allele (Figure 34). This confirmed that the growth of *dnaA Δtus rpo* ΔxonA ΔxseA srgA1* cells seen at 42°C is very likely to be dependent on the absence of ExoVII, which has 5' to 3' exonuclease activity in addition to 3' to 5' activity, and that the 3' flap processing ability is therefore essential for origin-independent growth in *dnaA Δtus rpo* ΔxonA* cells (Figure 34), strongly suggesting that 3' flap structures accumulate in the termination area in the absence of ExoI (Figure 40). Replication initiating at 3' flap structures does underpin the majority of the origin-independent growth seen in cells lacking 3' exonucleases, as in *dnaA Δtus rpo* ΔxonA ΔxseA* cells carrying the *srgA1* allele, growth at restrictive temperature was reduced substantially (over 10-fold) (Figure 32), but the significant growth that remains demonstrates that there is at least one other mechanism for initiating origin-independent over-replication in these cells. Our *ΔrecJ* data support the idea that initiation at 5' flaps might explain this to some degree. If RecG converts a 3' flap to a 5' flap, as suggested previously (Rudolph et al., 2010b), the resulting structure would be accessible for primosome assembly by PriA (Tanaka and Masai, 2006) even in cells carrying the mutant protein (PriA L557P) encoded by *srgA1* (Figure 40). In line with this, when 5' flap degradation is prevented in cells lacking 3' exonucleases by introduction of a *recJ* deletion, the peak of origin-independent over-replication seen in these cells increases (Figure

36) and the origin-independent growth of *dnaA Δtus rpo* ΔxonA* cells increases in the absence of RecJ (Figure 37) even though ExoVII, which has both 3' and 5' exonuclease activity (Lovett, 2011), is present in these cells. This result provides an explanation for the observation that *dnaA Δtus rpo* ΔxonA ΔxseA srgA1* cells are able to grow in the absence of origin firing.

There is a striking difference in the size of the peak of over-replication in cells lacking RecG compared to that seen in *ΔxonA ΔxseA* cells considering that both show similar viability of ~ 60% in a *dnaA Δtus rpo** background when grown at restrictive temperature (this study and Rudolph et al., 2013), revealing that there is not a direct correlation between the levels of over-replication and the ability of cells to grow in the absence of origin firing. Replication profiles of cells lacking ExoI and SbcCD have been generated previously by the Courcelle laboratory and show a peak of over-replication in the termination region of *ΔxonA ΔsbcCD* cells (Wendel et al., 2014, 2018). Whilst our data are in line with this work in principle, the peak in the replication profile of our *ΔxonA ΔsbcCD* strain was much less pronounced (Figure 26). Their replication profiles also show a small but perceptible peak in the termination region of *ΔxonA* and *ΔsbcCD* single mutant strains, which are not defined in our profiles. Wendel and colleagues have also profiled cells lacking RecG (Wendel et al., 2014) and again, the peak of over-replication seen in the replication profile of *ΔrecG* cells is significantly higher than that seen in previous work (Rudolph et al., 2013), which suggests that there is more origin-independent over-replication occurring in these strains than in our strains. The parental wild type strain, SR108, used by the Courcelle group is a derivative of W3110 (Donaldson et al., 2004). The genome sequences of W3110 and the parental wild type strain that we use, MG1655, are almost identical except that W3110 has been shown to have a large chromosomal inversion spanning *rrnD-rrnE* (Hayashi et al., 2006; Hill and Harnish, 1981; Skovgaard et al., 2011). The two operons are in opposite replichores but are both relatively close to *oriC*; *rrnD* is in the left hand replichore and *rrnE* is in the right hand replichore. The result of this inversion is that compared to MG1655, *oriC* is shifted slightly in to the left hand replichore because the distance between *rrnD* and *oriC* is greater than the distance between *oriC* and *rrnE*. In MG1655 cells, replication termination occurs very close to *terC* on average whereas in W3110, because of the inversion, it will be closer to *terA*. In W3110 derivatives, the anticlockwise fork will progress far enough to reach *terA* in the termination region, where it will then be held by the Tus/*terA* complex before fusing with an opposing fork. Previous work carried out in the Rudolph laboratory has shown that over-replication in *ΔrecG* cells is elevated substantially when one fork is trapped at a Tus/*ter* barrier before the converging fork arrives (Midgley-Smith et al., 2018; Rudolph et al., 2013), as discussed earlier (page 88) and so the asymmetric position of the origin in a W3110 background certainly explains the dramatic levels of over-replication seen in the replication profile of the Courcelle laboratory *ΔrecG* strain (Wendel et al., 2014) compared to our own. Could it be that an asymmetric replichore arrangement has a similar effect in cells lacking 3' exonucleases? The marker frequency replication profiles of double origin strains lacking 3' exonucleases presented here (Figure 38) demonstrate that an asymmetric replichore arrangement certainly does not have the same effect as in the absence of RecG. It could still

have a mild effect but the distortion in the termination area resulting from the presence of the ectopic replication origin is likely to obscure any mild increase to the relatively small levels of over-replication that occurs in *ΔxonA ΔsbcCD* cells (cf. Figure 38 and Figure 26) whereas it may be more evident in a strain with a single origin that is mildly shifted (Wendel et al., 2014). The data presented here suggests that in cells lacking 3' exonucleases, an asymmetric replichore arrangement does not trigger much higher levels of over-replication compared with single *oriC*$^{+}$ backgrounds, in contrast to what is seen in *ΔrecG* cells (Rudolph et al., 2013).

Could the differing levels of over-replication seen in the absence of either RecG or 3' exonucleases be explained by the idea that synthesis may occur in only a subset of cells? However, the robust viability at 42 °C, which is observed in both *ΔrecG* and 3' exonuclease-deficient strains, makes this very unlikely. So, despite the differences in over-replication, the viabilities of cells lacking RecG and 3' exonucleases are very similar. Instead, a key difference between RecG and 3' exonucleases that might help to explain the differing levels of over-replication observed is the involvement of RecG in recombination, as proposed before (Rudolph et al., 2013) and as discussed earlier (page 88) (Figure 40). In line with this, Azeroglu and colleagues (2016) recently reported that RecG is important in managing replication initiation at sites of DSBR and that in the absence of RecG, divergent replication is initiated. If recombination triggers replication and, in the absence of 3' exonucleases, RecG limits initiation at recombination intermediates, then this might well explain the differences in peak height observed. A high incidence of origin-independent over-replication initiation would present as a larger peak in the replication profile, but the problems caused by such high levels of recombination might result in fewer cells able survive in the absence of origin firing, as might be the case in *ΔrecG* cells. In contrast, there is no indication that 3' exonucleases have a role in managing recombination intermediates and so perhaps, with RecG present to fulfil this role, there are fewer initiation events but also fewer problems to manage and so there is a similar level of viability of *dnaA Δtus rpo* ΔxonA ΔxseA* and *dnaA Δtus rpo* ΔrecG* cells at 42°C despite the fact that levels of synthesis appear to be quite significantly different.

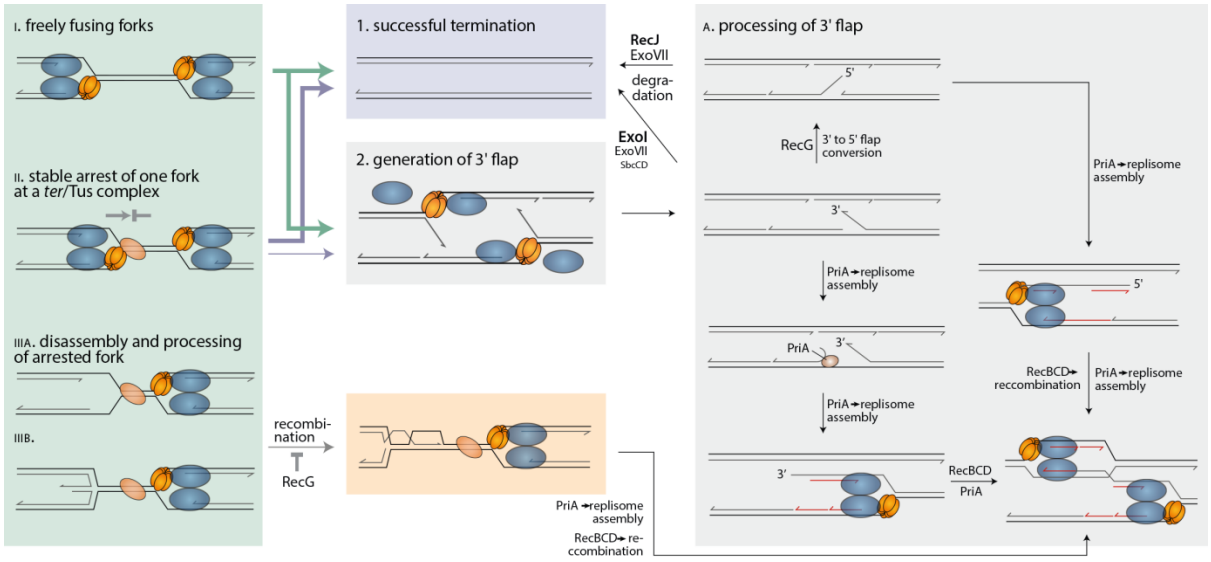

Figure 40: Schematic representation of events associated with the fusion of two replication forks in the termination area. Three different fork fusion scenarios are shown on the left (green box), resulting in a number of different intermediates (middle), which are then processed by a variety of different proteins (right). It was necessary to present some of the diagrams in a different orientation to the preceeding intermediate. This is the case for the processing of the recombination intermediates resulting from fork fusion scenario iii by either PriA or RecBCD, and also for the processing of the double-stranded end that results from PriA-mediated replisome assembly at a 5' flap structure. The double-stranded end in each case is processed and leads to recombination intermediates. See text for further details.

The data presented show that if there is an effect of the presence of an ectopic replication origin on the over-replication in cells lacking 3' exonucleases, it is only mild (Figure 38), in contrast to the situation in *ΔrecG* cells (Midgley-Smith et al., 2018; Rudolph et al., 2013; Wendel et al., 2014). In line with this, cells lacking 3' exonucleases that carry an ectopic replication origin can tolerate the deletion of the entire *oriC* region (Figure 39) whereas we have recently reported that *ΔoriC oriZ⁺ ΔrecG* cells are synthetically lethal (Midgley-Smith et al., 2018). If prolonged arrest of a replication fork leads to an increase in recombination events (Figure 40), this will cause problems in cells lacking RecG. In contrast, in the absence of 3' exonucleases, RecG is still present and will limit any effect of replication fork arrest by processing recombination intermediates, which explains the different levels of over-replication observed, especially in cells with an asymmetric replichore arrangement, and also explains the similar viabilities.

General discussion

Both initiation and elongation of DNA replication in bacteria have been studied for many years and are understood in quite some detail. In contrast, the events associated with termination of DNA replication have remained surprisingly elusive, as highlighted by Dewar and Walter in a recent review about termination, where they state: "Unlike initiation and elongation, which have been extensively studied, replication termination has received relatively little attention, especially in eukaryotic cells" (Dewar and Walter, 2017). However, even this statement is deceptive. Indeed, the Tus/*ter* replication fork trap system was found in some bacteria including *E. coli* many years ago and has been quite extensively studied, a fact that has led to the impression that termination is better understood in bacteria than it is in eukaryotic cells. However, it was shown previously that the replication fork trap system contributes to termination under normal growth conditions relatively infrequently (Bouché et al., 1982; Dimude et al., 2016; Duggin and Bell, 2009; Ivanova et al., 2015; Rudolph et al., 2013). This, in turn, suggests that the fusion of freely moving replication forks is the most common event, and until recently we knew hardly anything about the fusion of two moving forks, both in pro- and eukaryotic cells. Here I present data that clarify some of the events associated with replication fork fusion and highlight that the events associated with replication termination are surprisingly complex.

Data recently reported by the Rudolph laboratory already suggested that the fusion of forks can have surprisingly severe pathological consequences (Rudolph et al., 2013). As a result of a number of studies, a hypothesis of the events associated with the fusion of two replisomes was developed (Dimude et al., 2015; Hiasa and Marians, 1994; Krabbe et al., 1997; Rudolph et al., 2009b, 2010b, 2010a, 2013). As two replication forks converge, the DnaB helicase of one fork can displace the leading strand polymerase of the opposing fork, resulting in the formation of a 3' flap structure. If allowed to persist, 3' flaps can be processed by PriA restart protein to establish a new replication fork. The resulting duplex DNA on the leading strand will have a dsDNA end and so can trigger RecA-RecBCD-dependent recombination with the re-replicated DNA behind the newly established fork or with the sister chromosome, generating a substrate from which PriA can establish another replication fork that moves in the opposite direction. The fact that this origin-independent over-replication of the chromosome has the potential to sustain cell growth without the need for *oriC* firing illustrates the severity of these events. RecG has been identified as one of the key players in this process by normally processing any formed 3' flaps very rapidly (Dimude et al., 2015; Lloyd and Rudolph, 2016; Rudolph et al., 2013). Similarly, 3' exonucleases have also been implicated to be involved in the processing of 3' flaps (Rudolph et al., 2013; Sinha et al., 2018; Wendel et al., 2014, 2018).

The data presented here strongly support the importance of RecG in replication termination. Cells carrying a second, ectopic replication origin, *oriZ*, have an ectopic fork fusion region in addition to the native termination area. It has been identified previously that in *oriC*$^{+}$ *oriZ*$^{+}$ cells lacking RecG, there is

a broad amplification of this region (Midgley-Smith et al., 2018; Rudolph et al., 2013). When *ter* sites were introduced to create a replication fork trap around this region, the resulting peak seen in the replication profile of cells lacking RecG demonstrated that origin-independent over-replication is triggered in this region as well as in the native termination region (Figure 15) (Midgley-Smith et al., 2018). My data support that the proposed processing of R-loops by RecG helicase (Fukuoh et al., 1997; Hong et al., 1995; Lloyd and Rudolph, 2016; Vincent et al., 1996) is not involved in the initiation of this origin-independent synthesis, much in contrast to the situation in cells lacking RNase HI (Dimude et al., 2015; Kogoma, 1997; Maduike et al., 2014; Usongo et al., 2016). Over-replication of the chromosome causes a variety of different problems for cells and the over-replication that occurs in the termination region of *ΔrecG* cells, if exacerbated, can have lethal consequences, as illustrated by the lethality of *ΔoriC oriZ⁺ ΔrecG* cells, which require a *tus* deletion in order to be able to grow (Midgley-Smith et al., 2018).

In addition, the presented data provide strong support not only for a role of 3' exonucleases in termination, but also for 3' flap structures being key intermediates when replication forks fuse. As is seen in *ΔrecG* cells (Dimude et al., 2015; Rudolph et al., 2013), origin-independent over-replication of the chromosome is able to sustain robust cell growth in cells lacking 3' exonucleases once the replication fork trap is inactivated and the DNA synthesis initiated within the termination region is able to escape (Figure 24, Figure 27). Importantly, while RecG is capable of processing a variety of different substrates (Briggs et al., 2004; Rudolph et al., 2010b), the substrate specificity of 3' exonucleases is much more defined, thereby essentially confirming the presence of 3' flap structures as one of the major intermediates as replication forks fuse. The growth of *dnaA Δtus rpo* ΔxonA ΔxseA* cells at 42 °C despite the presence of the *srgA1* allele of PriA demonstrated that in the absence of 3' exonucleases, origin-independent DNA synthesis is on occasion able to initiate at structures other than 3' flaps. The subsequent finding that this growth is dependent on the absence of ExoVII, which has 5' – 3' exonuclease activity, and that the absence of RecJ, also a 5' exonuclease, improves origin-independent growth of *dnaA Δtus rpo* ΔxonA* cells, demonstrated that the alternative structure exploited in cells lacking 3' exonucleases is a 5' flap. The data presented in this thesis currently do not allow any conclusion of which proteins might be involved in the conversion of 3' flaps into 5' flaps. However, RecG helicase certainly can do this reaction (Bianco, 2015; Briggs et al., 2004; McGlynn et al., 2001; Tanaka and Masai, 2006). In addition, it was shown before that cells lacking the three main 3' exonucleases ExoI, ExoVII and SbcCD are inviable if *recG* is deleted (Rudolph et al., 2010a), making RecG an excellent candidate for carrying out this conversion.

There are several genetic approaches that could be used to confirm the involvement of RecG in the conversion, based on the principle that in the absence of RecG, the conversion would no longer take place and so if a *recG* deletion was introduced to cells lacking 3' exonucleases, the origin-independent synthesis seen in these cells would become entirely dependent on the 3' flap processing ability of PriA. Origin-independent growth seen in the absence of ExoI in a *dnaA Δtus rpo** background is abolished

when *srgA1* is introduced to these cells (Figure 34), as is seen in the absence of RecG (Dimude et al., 2015; Rudolph et al., 2013) and in cells lacking both RecG and ExoI (SLM-S and CJR, unpublished data), whereas *dnaA Δtus rpo* ΔxonA srgA1* cells can grow if the 5' flap processing capability of the cell is reduced by deleting *xseA*. The prediction is that by deleting *recG* from these cells, growth at 42 °C would be abolished. In any event, the work presented here highlights not only several protein activities involved in processing fork fusion intermediates, but also various pathways that are used for the processing of replication fork fusion intermediates.

While in the initial versions of the working model of replication termination RecG and 3' exonucleases appeared to have an equivalent role, the data presented here reveal distinct differences between the over-replication seen in *ΔrecG* cells and strains lacking 3'exonucleases, which highlight the complexity of the fusion process itself. In strains that have a second, ectopic copy of the replication origin inserted half way around the right hand replichore, the clockwise fork from the ectopic origin *oriZ* will arrive at the termination region before the anticlockwise fork coming from *oriC* and so will be blocked by Tus/*ter* complexes for some time (Dimude et al., 2018a; Ivanova et al., 2015; Midgley-Smith et al., 2018). This asymmetric replication results in a significant increase in the peak of over-replication in the termination area of *ΔrecG* cells in an *oriC*$^{+}$ *oriZ*$^{+}$ background (Rudolph et al., 2013), as was confirmed in this study (Figure 15). This was not observed in *oriC*$^{+}$ *oriZ*$^{+}$ cells lacking 3' exonucleases (Figure 38), highlighting a significant difference in the effect of RecG or 3' exonuclease absence. The fact that the growth of both *dnaA ΔrecG* (Figure 10) (Dimude et al., 2015; Rudolph et al., 2013) and *dnaA ΔxonA ΔxseA* cells (Figure 27) at 42 °C explicitly relies on the absence of Tus and not its presence indicates that the stalling of a replication fork specifically at a Tus/*ter* complex cannot be an essential event in the initiation of origin-independent over-replication.

One of the most apparent differences between *ΔrecG* cells with two origins compared to *ΔrecG* cells with a single replication origin, beside the fact that one replication fork is always held at a Tus/*ter* complex, is the length of time that forks are blocked before being met by the converging fork. Despite a recent report that a replication fork held at a Tus/*ter* complex for an extended period does not disassemble and instead remains stably bound (Moolman et al., 2016), *in vitro* and *in vivo* measurements of fork stability at obstacles including nucleoprotein roadblocks suggest a limited half-life of 4–6 min (Marians et al., 1998; McGlynn and Guy, 2008; Mettrick and Grainge, 2016), which suggests that after a relatively limited period of time, the replication fork complexes might start to disassemble. Forks permanently stalled at obstacles are processed by recombination proteins RecBCD (Dimude et al., 2018a; Michel and Leach, 2012). Thus, if forks stalled for a longer period at Tus/*ter* complexes are processed by recombination proteins, the opposing fork would not meet an arrested fork but a complex recombination intermediate. This suggests that at least three different type of fork fusion events have to be considered: the fusion of freely moving forks, the fusion of one free and one arrested fork and the fusion of one free fork with a fork that is at least partially disassembled and processed by recombination proteins. It was suggested that a fork blocked for an extended time at a

Tus/*ter* complex, as is the case in double origin strain, could trigger the over-replication seen in cells lacking RecG (Azeroglu et al., 2016). This cannot be essential for the over-replication seen in cells lacking RecG, as outlined above. However, if RecG was involved in both processing 3' flap structures as forks fuse and also destabilisation of recombination intermediates such as D-loops, this would provide an explanation for the differences observed between *ΔrecG* and 3' exonucleases-deficient cells. Thus, once over-replication is started at a 3' flap, in the absence of RecG the problem will become exacerbated by recombination, in particular in a situation where one fork is already processed by recombination proteins. Indeed, while the tandem repeat deletion studies presented here did not show increased levels of reversion events in the fork fusion area (page 96), they did highlight that RecG is able to prevent recombination events leading to tandem repeat deletions (page 94), as demonstrated before (Lovett et al. 2003; Lovett 2006; Lloyd and Rudolph 2016). There is no indication in the current literature that 3' exonucleases might have such a role, indicating that the lower levels of over-replication observed in the absence of 3' exonucleases stem mainly from the direct consequences of fork fusion reactions. Despite the low level of origin-independent over-replication seen in *ΔxonA ΔxseA* cells (Figure 26), this is clearly enough to sustain robust origin-independent growth.

Once origin-independent synthesis is established within the termination region, the replication forks will proceed until they are arrested at a Tus/*ter* complex. In this scenario, when the next round of replication is initiated at *oriC*, the replication forks will encounter a different situation compared to the previous round of replication when they reach the termination region as they will converge with replication forks held at the Tus/*ter* complexes. This idea adds an even great layer of complexity on the events that occur at replication termination and we have not yet begun to understand the consequences of this.

The work presented also clarifies our understanding of the ability of *ΔrecG* cells to grow in the absence of *oriC* firing. The ability of *E. coli* cells to grow if either the DnaA initiator protein or *oriC* is inactivated was described as Stable DNA Replication (SDR) (Kogoma, 1997). SDR was first observed in cells lacking RNase HI (Horiuchi et al., 1984; Ogawa et al., 1984) and then extended to a variety of mutations, including *ΔrecG* (Kogoma, 1997). Given the rather defined substrate specificity of RNase HI, which degrades RNA in RNA:DNA hybrids (Kogoma, 1997; Tadokoro and Kanaya, 2009), Kogoma and co-workers suggested that the origin-independent replication seen in *ΔrnhA* cells initiates at persisting R-loops (Kogoma et al., 1985; Meyenburg et al., 1987; Kogoma, 1997). Given that RecG was also found to remove R-loops by unwinding (Fukuoh et al., 1997; Vincent et al., 1996), it was assumed that the over-replication detected in *ΔrecG* cells might arise via a similar mechanism (Kogoma, 1997). However, the work presented here suggests otherwise. While the results strongly support the idea that over-replication in *ΔrnhA* cells is initiated at persisting R-loops, R-loops are unlikely to contribute towards the origin-independent synthesis observed in cells lacking RecG (Dimude et al., 2015; Lloyd and Rudolph, 2016; Midgley-Smith et al., 2018; Rudolph et al., 2013).

At least three different forms of SDR, namely constitutive, damage-induced and stationary SDR, have already been described (Kogoma, 1997), and processes such as persisting R-loops and recombination-dependent replication were suggested to contribute (Kogoma, 1997). My data highlight that in addition, termination of DNA replication is likely to have an important place as well and a variety of proteins, including RecG, ExoI, ExoVII, SbcCD and RecJ, contribute via this pathway (Figure 40) (Dimude et al., 2015; Midgley-Smith et al., 2018; Rudolph et al., 2010a, 2013; Wendel et al., 2014, 2018). Thus, while the term SDR is well suited to describe the resulting phenomenon of cells being able to grow without a functional replication origin, it obscures the underlying molecular mechanisms, because it "summarises" a number of unrelated and very different processes. Whilst it no doubt made sense from the historical perspective to use a defined name for what was observed, the common term suggests similarities where there are none, which hinders clarity rather than allowing us to understand phenomena better. With modern experimental techniques, we are now gaining increased levels of insight in to the various phenomena that were described as SDR, revealing an increasing number of processes involved and the molecular mechanisms behind these processes. Going forward it will be helpful to focus on the individual processes of nucleic acid metabolism rather than using descriptive terms such as SDR.

References

Åberg, C., Duderstadt, K.E., and van Oijen, A.M. (2016). Stability versus exchange: a paradox in DNA replication. *Nucleic Acids Res.* 44, 4846–4854.

Adhikari, S., and Curtis, P.D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol*. Rev. 40, 575–591.

Adiciptaningrum, A., Osella, M., Moolman, M.C., Cosentino Lagomarsino, M., and Tans, S.J. (2015). Stochasticity and homeostasis in the *E. coli* replication and division cycle. *Sci. Rep*. 5.

Aguilera, A., and Gómez-González, B. (2008). Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet*. 9, 204–217.

Al-Deib, A.A., Mahdi, A.A., and Lloyd, R.G. (1996). Modulation of recombination and DNA repair by the RecG and PriA helicases of *Escherichia coli* K-12. *J. Bacteriol*. 178, 6782–6789.

Amado, L., and Kuzminov, A. (2006). The replication intermediates in *Escherichia coli* are not the product of DNA processing or uracil excision. *J. Biol. Chem*. 281, 22635–22646.

Andersen, P.A., Griffiths, A.A., Duggin, I.G., and Wake, R.G. (2000). Functional specificity of the replication fork-arrest complexes of *Bacillus subtilis* and *Escherichia coli*: significant specificity for Tus–Ter functioning in *E. coli*. *Mol. Microbiol*. 36, 1327–1335.

Anderson, D.G., and Kowalczykowski, S.C. (1997). The recombination hot spot chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev*. 11, 571–581.

Arnold, D.A., and Kowalczykowski, S.C. (2001). RecBCD Helicase/Nuclease. **In** Encyclopedia of Life Sciences, John Wiley & Sons, Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd).

Atlung, T., Clausen, E.S., and Hansen, F.G. (1985). Autoregulation of the *dnaA* gene of *Escherichia coli* K12. *Mol. Gen. Genet*. 200, 442–450.

Au, K.G., Welsh, K., and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *J. Biol. Chem*. 267, 12142–12148.

Aune, T.E.V., and Aachmann, F.L. (2010). Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl. Microbiol. Biotechnol*. 85, 1301–1313.

Azeroglu, B., and Leach, D.R.F. (2017). RecG controls DNA amplification at double-strand breaks and arrested replication forks. *Febs Lett*. 591, 1101–1113.

Azeroglu, B., Mawer, J.S.P., Cockram, C.A., White, M.A., Hasan, A.M.M., Filatenkova, M., and Leach, D.R.F. (2016). RecG directs DNA synthesis during double-strand break repair. *PLoS Genet*. 12.

Bachmann, B.J. (1996). Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. **In** *Escherichia coli and Salmonella Cellular and Molecular Biology.* ASM Press

Balakrishnan, L., and Bambara, R.A. (2013). Okazaki Fragment Metabolism. *Cold Spring Harb. Perspect. Biol*. 5:a010173.

Barras, F., and Marinus, M.G. (1988). Arrangement of Dam methylation sites (GATC) in the *Escherichia coli* chromosome. *Nucleic Acids Res*. 16, 9821–9838.

Barre, F.-X., Søballe, B., Michel, B., Aroyo, M., Robertson, M., and Sherratt, D. (2001). Circles: The replication-recombination-chromosome segregation connection. *Proc. Natl. Acad. Sci*. 98, 8189–8195.

Bastia, D., Zzaman, S., Krings, G., Saxena, M., Peng, X., and Greenberg, M.M. (2008). Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. *Proc. Natl. Acad. Sci*. U. S. A. 105, 12831–12836.

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

Bates, D., Epstein, J., Boye, E., Fahrner, K., Berg, H., and Kleckner, N. (2005). The *Escherichia coli* baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle. *Mol. Microbiol*. 57, 380–391.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 21, 3329–3330.

Beattie, T.R., and Reyes-Lamothe, R. (2015). A Replisome's journey through the bacterial chromosome. *Front. Microbiol*. 6.

Beattie, T.R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A.J., Leake, M.C., and Reyes-Lamothe, R. (2017). Frequent exchange of the DNA polymerase during bacterial chromosome replication. *ELife*. 6.

Berghuis, B.A., Raducanu, V.-S., Elshenawy, M.M., Jergic, S., Depken, M., Dixon, N.E., Hamdan, S.M., and Dekker, N.H. (2018). What is all this fuss about Tus? Comparison of recent findings from biophysical and biochemical experiments. *Crit. Rev. Biochem. Mol. Biol*. 53, 49–63.

Bernhardt, T.G., and de Boer, P.A.J. (2003). The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol. Microbiol*. 48, 1171–1182.

Bernhardt, T.G., and de Boer, P.A.J. (2004). Screening for synthetic lethal mutants in *Escherichia coli* and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity. *Mol. Microbiol*. 52, 1255–1269.

Bétous, R., Couch, F.B., Mason, A.C., Eichman, B.F., Manosas, M., and Cortez, D. (2013). Substrate-selective repair and restart of replication forks by DNA translocases. *Cell Rep*. 3, 1958–1969.

Bianco, P.R. (2015). I came to a fork in the DNA and there was RecG. *Prog. Biophys. Mol. Biol*. 117, 166–173.

Bianco, P.R., and Kowalczykowski, S.C. (1997). The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5′-GCTGGTGG-3′. *Proc. Natl. Acad. Sci*. 94, 6706–6711.

Bianco, P.R., and Lyubchenko, Y.L. (2017). SSB and the RecG DNA helicase: an intimate association to rescue a stalled replication fork. *Protein Sci*. 26, 638–649.

Bichara, M., Meier, M., Wagner, J., Cordonnier, A., and Lambert, I.B. (2011). Postreplication repair mechanisms in the presence of DNA adducts in *Escherichia coli*. *Mutat. Res. Mutat. Res*. 727, 104–122.

Bidnenko, V., Ehrlich, S.D., and Michel, B. (2002). Replication fork collapse at replication terminator sequences. *EMBO J*. 21, 3898–3907.

Bigot, S., Saleh, O.A., Lesterlin, C., Pages, C., El Karoui, M., Dennis, C., Grigoriev, M., Allemand, J.-F., Barre, F.-X., and Cornet, F. (2005). KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. *EMBO J*. 24, 3770–3780.

Bigot, S., Sivanathan, V., Possoz, C., Barre, F.-X., and Cornet, F. (2007). FtsK, a literate chromosome segregation machine. *Mol. Microbiol*. 64, 1434–1441.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462.

Bouché, J.P., Gélugne, J.P., Louarn, J., Louarn, J.M., and Kaiser, K. (1982). Relationships between the physical and genetic maps of a 470 × 103 base-pair region around the terminus of *Escherichia coli* K12 DNA replication. *J. Mol. Biol*. 154, 21–32.

Boye, E., and Nordström, K. (2003). Coupling the cell cycle to cell growth. *EMBO Rep*. 4, 757–760.

Breier, A.M., Weier, H.-U.G., and Cozzarelli, N.R. (2005). Independence of replisomes in *Escherichia coli* chromosomal replication. *Proc. Natl. Acad. Sci. U. S. A*. 102, 3942–3947.

Brewer, B.J. (1988). When polymerases collide: Replication and the transcriptional organization of the *E. coli* chromosome. *Cell* 53, 679–686.

Briggs, G.S., Mahdi, A.A., Weller, G.R., Wen, Q., and Lloyd, R.G. (2004). Interplay between DNA replication, recombination and repair based on the structure of RecG helicase. *Philos. Trans. R. Soc. B Biol. Sci*. 359, 49–59.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S.T., and Modrich, P. (2001). *In vivo* requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc. Natl. Acad. Sci. U. S. A*. 98, 6765–6770.

Burgers, P.M., Kornberg, A., and Sakakibara, Y. (1981). The *dnaN* gene codes for the beta subunit of DNA polymerase III holoenzyme of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A*. 78, 5391–5395.

Buss, J.A., Kimura, Y., and Bianco, P.R. (2008). RecG interacts directly with SSB: implications for stalled replication fork regression. *Nucleic Acids Res*. 36, 7029–7042.

Bzymek, M., and Lovett, S.T. (2001). Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci*. 98, 8319–8325.

Campbell, J.L., and Kleckner, N. (1990). *E. coli* oriC and the *dnaA* gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell*. 62, 967–979.

Cardon, L.R., Burge, C., Schachtel, G.A., Blaisdell, B.E., and Karlin, S. (1993). Comparative DNA sequence features in two long *Escherichia coli* contigs. *Nucleic Acids Res*. 21, 3875–3884.

Castel, A.L., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol*. 11, 165–170.

Cebrián, J., Castán, A., Martínez, V., Kadomatsu-Hermosa, M.J., Parra, C., Fernández-Nestosa, M.J., Schaerer, C., Hernández, P., Krimer, D.B., and Schvartzman, J.B. (2015). Direct evidence for the formation of precatenanes during DNA replication. *J. Biol. Chem*. 290, 13725–13735.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in Eukaryotes. *FEBS J*. 276, 1494–1505.

Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem*. 70, 369–413.

Chase, J.W., and Richardson, C.C. (1974a). Exonuclease VII of *Escherichia coli*. Mechanism of action. *J. Biol. Chem*. 249, 4553–4561.

Chase, J.W., and Richardson, C.C. (1974b). Exonuclease VII of *Escherichia coli*. Purification and properties. *J. Biol. Chem*. 249, 4545–4552.

Chase, J.W., Rabin, B.A., Murphy, J.B., Stone, K.L., and Williams, K.R. (1986). *Escherichia coli* exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (*xseA*). *J. Biol. Chem*. 261, 14929–14935.

Cherepanov, P.P., and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158, 9–14.

Connelly, J.C., and Leach, D.R.F. (1996). The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells*. 1, 285–291.

Connelly, J.C., Leau, E.S. de, Okely, E.A., and Leach, D.R.F. (1997). Overexpression, purification, and characterization of the SbcCD protein from *Escherichia coli*. *J. Biol. Chem*. 272, 19819–19826.

Connelly, J.C., Kirkham, L.A., and Leach, D.R.F. (1998). The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc. Natl. Acad. Sci*. 95, 7969–7974.

Connelly, J.C., de Leau, E.S., and Leach, D.R. (1999). DNA cleavage and degradation by the SbcCD protein complex from *Escherichia coli*. *Nucleic Acids Res*. 27, 1039–1046.

Cooper, G.M. (2000). The Eukaryotic Cell Cycle. *Cell Mol. Approach* 2nd Ed.

Coskun-Ari, F.F., and Hill, T.M. (1997). Sequence-specific interactions in the Tus-Ter complex and the effect of base pair substitutions on arrest of DNA replication in *Escherichia coli. J. Biol. Chem*. 272, 26448–26456.

Coskun-Ari, F.F., Skokotas, A., Moe, G.R., and Hill, T.M. (1994). Biophysical characteristics of Tus, the replication arrest protein of *Escherichia coli*. *J. Biol. Chem*. 269, 4027–4034.

Costa, A., Hood, I.V., and Berger, J.M. (2013). Mechanisms for initiating cellular DNA replication. *Annu. Rev. Biochem*. 82, 25–54.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K.J. (2000). The importance of repairing stalled replication forks. *Nature*. 404, 37–41.

Cromie, G.A., Connelly, J.C., and Leach, D.R.F. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol. Cell*. 8, 1163–1174.

Cui, T., Moro-oka, N., Ohsumi, K., Kodama, K., Ohshima, T., Ogasawara, N., Mori, H., Wanner, B., Niki, H., and Horiuchi, T. (2007). *Escherichia coli* with a linear genome. *EMBO Rep*. 8, 181–187.

Darmon, E., Eykelenboom, J.K., Lincker, F., Jones, L.H., White, M., Okely, E., Blackwood, J.K., and Leach, D.R. (2010). *E. coli* SbcCD and RecA control chromosomal rearrangement induced by an interrupted palindrome. *Mol. Cell*. 39, 59–70.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A*. 97(12): 6640–6645.

Deneke, J., Ziegelin, G., Lurz, R., and Lanka, E. (2000). The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. *Proc. Natl. Acad. Sci. U. S. A*. 97, 7721–7726.

Deneke, J., Ziegelin, G., Lurz, R., and Lanka, E. (2002). Phage N15 telomere resolution: target requirements for recognition and processing by the protelomerase. *J. Biol. Chem*. 277, 10410–10419.

Dennis, P.P., Ehrenberg, M., Fange, D., and Bremer, H. (2009). Varying rate of RNA chain elongation during *rrn* transcription in *Escherichia coli*. *J. Bacteriol*. 191, 3740–3746.

Dewar, J.M., and Walter, J.C. (2017). Mechanisms of DNA replication termination. *Nat. Rev. Mol. Cell Biol*. 18, 507–516.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev*. 72, 642–671.

Dimude, J.U., Stockum, A., Midgley-Smith, S.L., Upton, A.L., Foster, H.A., Khan, A., Saunders, N.J., Retkute, R., and Rudolph, C.J. (2015). The consequences of replicating in the wrong orientation: bacterial chromosome duplication without an active replication origin. *MBio*. 6, e01294-15.

Dimude, J.U., Midgley-Smith, S.L., Stein, M., and Rudolph, C.J. (2016). Replication termination: containing fork fusion-mediated pathologies in *Escherichia coli*. *Genes*. 7(8): 40

Dimude, J.U., Stein, M., Andrzejewska, E.E., Khalifa, M.S., Gajdosova, A., Retkute, R., Skovgaard, O., and Rudolph, C.J. (2018a). Origins left, right, and centre: increasing the number of initiation sites in the *Escherichia coli* chromosome. *Genes*. 9(8): 376

Dimude, J.U., Midgley-Smith, S.L., and Rudolph, C.J. (2018b). Replication-transcription conflicts trigger extensive DNA degradation in *Escherichia coli* cells lacking RecBCD. *DNA Repair*. 70, 37–48.

Donachie, W.D. (1993). The Cell Cycle of *Escherichia coli*. *Annu. Rev. Microbiol*. 47, 199–230.

Donaldson, J.R., Courcelle, C.T., and Courcelle, J. (2004). RuvAB and RecG are not essential for the recovery of DNA synthesis following UV-induced DNA damage in *Escherichia coli*. *Genetics*. 166, 1631–1640.

Duggin, I.G., and Bell, S.D. (2009). Termination structures in the *Escherichia coli* chromosome replication fork trap. *J. Mol. Biol*. 387, 532–539.

Duggin, I.G., Wake, R.G., Bell, S.D., and Hill, T.M. (2008). The replication fork trap and termination of chromosome replication. *Mol. Microbiol*. 70, 1323–1333.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability. *Cell*. 146, 533–543.

Ede, C., Rudolph, C.J., Lehmann, S., Schürer, K.A., and Kramer, W. (2011). Budding yeast Mph1 promotes sister chromatid interactions by a mechanism involving strand invasion. *DNA Repair*. 10, 45–55.

Ellwood, M., and Nomura, M. (1982). Chromosomal locations of the genes for rRNA in *Escherichia coli* K-12. *J. Bacteriol*. 149, 458–468.

Evans, E., and Alani, E. (2000). Roles for mismatch repair factors in regulating genetic recombination. *Mol. Cell. Biol*. 20, 7839–7844.

Eykelenboom, J.K., Blackwood, J.K., Okely, E., and Leach, D.R.F. (2008). SbcCD causes a double-strand break at a DNA palindrome in the *Escherichia coli* chromosome. *Mol. Cell*. 29, 644–651.

Fang, L., Davey, M.J., and O'Donnell, M. (1999). Replisome assembly at *oriC*, the replication origin of *E. coli*, reveals an explanation for initiation sites outside an origin. *Mol. Cell*. 4, 541–553.

Ferullo, D.J., Cooper, D.L., Moore, H.R., and Lovett, S.T. (2009). Cell cycle synchronization of *E. coli* using the stringent response, with fluorescence labeling assays for DNA content and replication. *Methods San Diego Calif*. 48, 8–13.

Fijalkowska, I.J., Schaaper, R.M., and Jonczyk, P. (2012). DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. *FEMS Microbiol. Rev*. 36, 1105–1121.

Foster, P.L. (2006). Methods for determining spontaneous mutation rates. *Methods Enzymol*. 409, 195–213.

French, S. (1992). Consequences of replication fork movement through transcription units *in vivo*. *Science*. 258, 1362–1365.

Frey, J., Chandler, M., and Caro, L. (1981). The initiation of chromosome replication in a *dnaAts46* and a *dnaA+* strain at various temperatures. *Mol. Gen. Genet*. 182, 364–366.

Fuchs, R.P. (2016). Tolerance of lesions in E. coli: Chronological competition between translesion synthesis and damage avoidance. *DNA Repair*. 44, 51–58.

Fukuoh, A., Iwasaki, H., Ishioka, K., and Shinagawa, H. (1997). ATP-dependent resolution of R-loops at the ColE1 replication origin by *Escherichia coli* RecG protein, a Holliday junction-specific helicase. *EMBO J*. 16, 203–209.

Gabbai, C.B., and Marians, K.J. (2010). Recruitment to stalled replication forks of the PriA DNA helicase and replisome-loading activities is essential for survival. *DNA Repair*. 9, 202–209.

Gandon, S. (2016). Why be temperate: lessons from bacteriophage λ. *Trends Microbiol*. 24, 356–365.

Gao, D., and McHenry, C.S. (2001a). *tau* binds and organizes *Escherichia coli* replication proteins through distinct domains. Domain IV, located within the unique C terminus of *tau*, binds the replication fork, helicase, DnaB. *J. Biol. Chem*. 276, 4441–4446.

Gao, D., and McHenry, C.S. (2001b). *tau* binds and organizes *Escherichia coli* replication through distinct domains. Partial proteolysis of terminally tagged *tau* to determine candidate domains and to assign domain V as the alpha binding domain. *J. Biol. Chem*. 276, 4433–4440.

Georgescu, R.E., Yao, N.Y., and O'Donnell, M. (2010). Single-molecule analysis of the *E. coli* replisome and use of clamps to bypass replication barriers. *FEBS Lett*. 584, 2596–2605.

Goldmark, P.J., and Linn, S. (1972). Purification and properties of the *recBC* DNase of *Escherichia coli* K-12. *J. Biol. Chem*. 247, 1849–1860.

Gotta, S.L., Miller, O.L., and French, S.L. (1991). rRNA transcription rate in *Escherichia coli*. *J. Bacteriol*. 173, 6647–6649.

Gottlieb, P.A., Wu, S., Zhang, X., Tecklenburg, M., Kuempel, P., and Hill, T.M. (1992). Equilibrium, kinetic, and footprinting studies of the Tus-Ter protein-DNA interaction. *J. Biol. Chem*. 267, 7434–7443.

Gowrishankar, J. (2015). End of the beginning: elongation and termination features of alternative modes of chromosomal replication initiation in bacteria. *PLOS Genet*. 11, e1004909.

Graham, J.E., Marians, K.J., and Kowalczykowski, S.C. (2017). Independent and stochastic action of DNA polymerases in the replisome. *Cell*. 169, 1201-1213.e17.

Grant, M.A., Saggioro, C., Ferrari, U., Bassetti, B., Sclavi, B., and Cosentino Lagomarsino, M. (2011). DnaA and the timing of chromosome replication in *Escherichia coli* as a function of growth rate. *BMC Syst. Biol*. 5, 201.

Gregg, A.V., McGlynn, P., Jaktaji, R.P., and Lloyd, R.G. (2002). Direct rescue of stalled DNA replication forks via the combined action of PriA and RecG helicase activities. *Mol. Cell*. *9*, 241–251.

Grilley, M., Griffith, J., and Modrich, P. (1993). Bidirectional excision in methyl-directed mismatch repair. *J. Biol. Chem*. 268, 11830–11837.

Gupta, M.K., Guy, C.P., Yeeles, J.T.P., Atkinson, J., Bell, H., Lloyd, R.G., Marians, K.J., and McGlynn, P. (2013). Protein–DNA complexes are the primary sources of replication fork pausing in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A*. 110, 7252–7257.

Gupta, S., Yeeles, J.T.P., and Marians, K.J. (2014). Regression of replication forks stalled by leading-strand template damage: I. Both RecG and RuvAB catalyze regression, but RuvC cleaves the holliday junctions formed by RecG preferentially. *J. Biol. Chem*. 289, 28376–28387.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol*. 177, 4121–4130.

Haeusser, D.P., and Levin, P.A. (2008). The Great Divide: Coordinating cell cycle events during bacterial growth and division. *Curr. Opin. Microbiol*. 11, 94–99.

Han, E.S., Cooper, D.L., Persky, N.S., Sutera, V.A., Whitaker, R.D., Montello, M.L., and Lovett, S.T. (2006). RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Res*. 34, 1084–1091.

Hansen, F.G., and Atlung, T. (2018). The DnaA Tale. *Front. Microbiol*. 9.

Hansen, F.G., Christensen, B.B., and Atlung, T. (2007). Sequence characteristics required for cooperative binding and efficient *in vivo* titration of the replication initiator protein DnaA in *E. coli*. *J. Mol. Biol*. 367, 942–952.

Harfe, B.D., and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. *Annu. Rev. Genet*. 34, 359–399.

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H., et al. (2006). Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol*. 2, 2006.0007.

Heller, R.C., and Marians, K.J. (2006). Replisome assembly and the direct restart of stalled replication forks. *Nat. Rev. Mol. Cell Biol*. 7, 932–943.

Hénaut, A., Rouxel, T., Gleizes, A., Moszer, I., and Danchin, A. (1996). Uneven distribution of GATC motifs in the *Escherichia coli* chromosome, its plasmids and its phages. *J. Mol. Biol*. 257, 574–585.

Henderson, T., Nilles, A., Valjavec-Gratian, M., and Hill, T. (2001). Site-directed mutagenesis and phylogenetic comparisons of the *Escherichia coli* Tus protein: DNA-protein interactions alone can not account for Tus activity. *Mol. Genet. Genomics*. 265, 941–953.

Hiasa, H., and Marians, K.J. (1992). Differential inhibition of the DNA translocation and DNA unwinding activities of DNA helicases by the *Escherichia coli* Tus protein. *J. Biol. Chem*. 267, 11379–11385.

Hiasa, H., and Marians, K.J. (1994). Tus prevents overreplication of *oriC* plasmid DNA. *J. Biol. Chem*. 269, 26959–26968.

Hill, C.W., and Harnish, B.W. (1981). Inversions between ribosomal RNA genes of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A*. 78, 7069–7072.

Hill, T.M., Tecklenburg, M.L., Pelletier, A.J., and Kuempel, P.L. (1989). *tus*, the trans-acting gene required for termination of DNA replication in *Escherichia coli*, encodes a DNA-binding protein. *Proc. Natl. Acad. Sci*. 86, 1593–1597.

Hong, X., Cadwell, G.W., and Kogoma, T. (1995). *Escherichia coli* RecG and RecA proteins in R-loop formation. *EMBO J*. 14, 2385–2392.

Horiuchi, T., Maki, H., and Sekiguchi, M. (1984). RNase H-defective mutants of *Escherichia coli*: A possible discriminatory role of RNase H in initiation of DNA replication. *Mol. Gen. Genet*. 195, 17–22.

Horiuchi, T., Fujimura, Y., Nishitani, H., Kobayashi, T., and Hidaka, M. (1994). The DNA replication fork blocked at the Ter site may be an entrance for the RecBCD enzyme into duplex DNA. *J. Bacteriol*. 176, 4656–4663.

Hou, Y., Song, H., Croteau, D.L., Akbari, M., and Bohr, V.A. (2017). Genome instability in Alzheimer disease. *Mech. Ageing Dev*. 161, 83–94.

Hsieh, P., and Yamane, K. (2008). DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mech. Ageing Dev*. 129, 391–407.

Ivančić-Baće, I., Cass, S.D., Wearne, S.J., and Bolt, E.L. (2015). Different genome stability proteins underpin primed and naïve adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res*. 43, 10821–10830.

Ivanova, D., Taylor, T., Smith, S.L., Dimude, J.U., Upton, A.L., Mehrjouy, M.M., Skovgaard, O., Sherratt, D.J., Retkute, R., and Rudolph, C.J. (2015). Shaping the landscape of the *Escherichia coli* chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res*. 43, 7865–7877.

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

Jin, D.J., Cagliero, C., and Zhou, Y.N. (2012). Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev*. 36, 269–287.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem*. 74, 283–315.

Jung, H., Liang, J., Jung, Y., and Lim, D. (2015). Characterization of cell death in *Escherichia coli* mediated by XseA, a large subunit of exonuclease VII. *J. Microbiol*. 53, 820–828.

Kaguni, J.M. (2006). DnaA: controlling the initiation of bacterial DNA replication and more. *Annu. Rev. Microbiol*. 60, 351–371.

Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K. (1996). Structure of a replication-terminator protein complexed with DNA. *Nature*. 383, 598–603.

Katayama, T. (2001). Feedback controls restrain the initiation of *Escherichia coli* chromosomal replication. *Mol. Microbiol*. 41, 9–17.

Katayama, T., Kubota, T., Kurokawa, K., Crooke, E., and Sekimizu, K. (1998). The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. *Cell*. 94, 61–71.

Katayama, T., Fujimitsu, K., and Ogawa, T. (2001). Multiple pathways regulating DnaA function in *Escherichia coli*: Distinct roles for DnaA titration by the *datA* locus and the regulatory inactivation of DnaA. *Biochimie*. 83, 13–17.

Katayama, T., Kasho, K., and Kawakami, H. (2017). The DnaA cycle in *Escherichia coli*: activation, function and inactivation of the initiator protein. *Front. Microbiol*. 8.

Kato, J., and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J*. 20, 4253–4262.

Kaul, S., Mohanty, B.K., Sahoo, T., Patel, I., Khan, S.A., and Bastia, D. (1994). The replication terminator protein of the gram-positive bacterium *Bacillus subtilis* functions as a polar contrahelicase in gram-negative *Escherichia coli*. *Proc. Natl. Acad. Sci*. 91, 11143–11147.

Killen, M.W., Stults, D.M., Wilson, W.A., and Pierce, A.J. (2012). *Escherichia coli* RecG functionally suppresses human Bloom syndrome phenotypes. *BMC Mol. Biol*. 13, 33.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nat. Rev. Genet*. 13, 204–214.

Kisker, C., Kuper, J., and Houten, B.V. (2013). Prokaryotic nucleotide excision repair. *Cold Spring Harb. Perspect. Biol*. 5, a012591.

Kitagawa, R., Mitsuki, H., Okazaki, T., and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol. Microbiol*. 19, 1137–1147.

Kitagawa, R., Ozaki, T., Moriya, S., and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev*. 12, 3032–3043.

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol*. Rev. 61, 212–238.

Kogoma, T., and Lark, K.G. (1970). DNA replication in *Escherichia coli*: Replication in absence of protein synthesis after replication inhibition. *J. Mol. Biol*. 52, 143–164.

Kogoma, T., and Lark, K.G. (1975). Characterization of the replication of *Escherichia coli* DNA in the absence of protein synthesis: Stable DNA replication. *J. Mol. Biol*. 94, 243–256.

Kogoma, T., Skarstad, K., Boye, E., Meyenburg, K. von, and Steen, H.B. (1985). RecA protein acts at the initiation of stable DNA replication in *rnh* mutants of Escherichia coli K-12. *J. Bacteriol*. 163, 439–444.

Kono, N., Arakawa, K., Sato, M., Yoshikawa, H., Tomita, M., and Itaya, M. (2014). Undesigned selection for replication termination of bacterial chromosomes. *J. Mol. Biol*. 426, 2918–2927.

Kowalczykowski, S.C. (2000). Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci*. 25, 156–165.

Krabbe, M., Zabielski, J., Bernander, R., and Nordström, K. (1997). Inactivation of the replication-termination system affects the replication mode and causes unstable maintenance of plasmid R1. *Mol. Microbiol*. 24, 723–735.

Kunkel, T.A. (2004). DNA replication fidelity. *J. Biol. Chem*. 279, 16895–16898.

Kunkel, T.A., and Erie, D.A. (2005). Dna Mismatch Repair. *Annu. Rev. Biochem*. 74, 681–710.

Kurth, I., and O'Donnell, M. (2009). Replisome dynamics during chromosome duplication. *EcoSal Plus*. 3.

Kuzminov, A., Schabtach, E., and Stahl, F.W. (1994). Chi sites in combination with RecA protein increase the survival of linear DNA in *Escherichia coli* by inactivating *exoV* activity of RecBCD nuclease. *EMBO J*. 13, 2764–2776.

LaDuca, R.J., Crute, J.J., McHenry, C.S., and Bambara, R.A. (1986). The beta subunit of the *Escherichia coli* DNA polymerase III holoenzyme interacts functionally with the catalytic core in the absence of other subunits. *J. Biol. Chem*. 261, 7550–7557.

Lambert, S., Watson, A., Sheedy, D.M., Martin, B., and Carr, A.M. (2005). Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell*. 121, 689–702.

Lang, K.S., and Merrikh, H. (2018). The clash of macromolecular titans: replication-transcription conflicts in bacteria. *Annu. Rev. Microbiol*. 72, 71–88.

Langston, L.D., Indiani, C., and O'Donnell, M. (2009). Whither the replisome: Emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle*. 8, 2686–2691.

Larsen, N.B., Sass, E., Suski, C., Mankouri, H.W., and Hickson, I.D. (2014). The *Escherichia coli* Tus–*Ter* replication fork barrier causes site-specific DNA replication perturbation in yeast. *Nat. Commun*. 5, 3574.

Lau, I.F., Filipe, S.R., Søballe, B., Økstad, O.-A., Barre, F.-X., and Sherratt, D.J. (2003). Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol. Microbiol*. 49, 731–743.

Lea, D.E., and Coulson, C.A. (1949). The distribution of the numbers of mutants in bacterial populations. *J. Genet*. 49, 264.

Leach, D.R.F., Okely, E.A., and Pinder, D.J. (1997). Repair by recombination of DNA containing a palindromic sequence. *Mol. Microbiol*. 26, 597–606.

LeBowitz, J.H., and McMacken, R. (1986). The *Escherichia coli dnaB* replication protein is a DNA helicase. *J. Biol. Chem*. 261, 4738–4748.

Lee, E.H., and Kornberg, A. (1991). Replication deficiencies in *priA* mutants of *Escherichia coli* lacking the primosomal replication n' protein. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3029–3032.

Lee, E.H., and Kornberg, A. (1992). Features of replication fork blockage by the *Escherichia coli* terminus-binding protein. *J. Biol. Chem*. 267, 8778–8784.

Lee, E.H., Kornberg, A., Hidaka, M., Kobayashi, T., and Horiuchi, T. (1989). *Escherichia coli* replication termination protein impedes the action of helicases. *Proc. Natl. Acad. Sci. U. S. A*. 86, 9104–9108.

Lehman, I.R., and Nussbaum, A.L. (1964). The deoxyribonucleases of *Escherichia coli* V. on the specificity of Exonuclease I (phosphodiesterase). *J. Biol. Chem*. 239, 2628–2636.

Lehmann, A.R., and Fuchs, R.P. (2006). Gaps and forks in DNA replication: Rediscovering old models. *DNA Repair*. 5, 1495–1498.

Lemon, K.P., Kurtser, I., and Grossman, A.D. (2001). Effects of replication termination mutants on chromosome partitioning in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A*. 98, 212–217.

Leonard, A.C., and Grimwade, J.E. (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Mol. Microbiol*. 55, 978–985.

Leonard, A.C., and Grimwade, J.E. (2011). Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol*. 65, 19–35.

Lesterlin, C., Barre, F.-X., and Cornet, F. (2004). Genetic recombination and the cell cycle: what we have learned from chromosome dimers. *Mol. Microbiol*. 54, 1151–1160.

Lewis, J.S., Spenkelink, L.M., Jergic, S., Wood, E.A., Monachino, E., Horan, N.P., Duderstadt, K.E., Cox, M.M., Robinson, A., Dixon, N.E., et al. (2017). Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *ELife*. 6, e23932.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*. 362, 709–715.

Lloyd, R.G. (1991). Conjugational recombination in resolvase-deficient *ruvC* mutants of *Escherichia coli* K-12 depends on *recG*. *J. Bacteriol*. 173, 5414–5418.

Lloyd, R.G., and Buckman, C. (1991). Genetic analysis of the *recG* locus of *Escherichia coli* K-12 and of its role in recombination and DNA repair. *J. Bacteriol*. 173, 1004–1011.

Lloyd, R.G., and Rudolph, C.J. (2016). 25 years on and no end in sight: a perspective on the role of RecG protein. *Curr. Genet*. 62, 827–840.

Lloyd, R.G., and Sharples, G.J. (1991). Molecular organization and nucleotide sequence of the *recG* locus of *Escherichia coli* K-12. *J. Bacteriol*. 173, 6837–6843.

Lloyd, R.G., and Sharples, G.J. (1993). Dissociation of synthetic Holliday junctions by *E. coli* RecG protein. *EMBO J*. 12, 17–22.

Løbner-Olesen, A., Hansen, F.G., Rasmussen, K.V., Martin, B., and Kuempel, P.L. (1994). The initiation cascade for chromosome replication in wild-type and Dam methyltransferase deficient *Escherichia coli* cells. *EMBO J*. 13, 1856–1862.

Loeb, L.A., and Kunkel, T.A. (1982). Fidelity of DNA synthesis. *Annu. Rev. Biochem*. 51, 429–457.

Lovett, S.T. (2004). Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol*. 52, 1243–1253.

Lovett, S.T. (2006). Replication arrest-stimulated recombination: Dependence on the RecA paralog, RadA/Sms and translesion polymerase, DinB. *DNA Repair*. 5, 1421–1427.

Lovett, S.T. (2011). The DNA exonucleases of *Escherichia coli*. *EcoSal Plus*. 4.

Lovett, S.T. (2017). Template-switching during replication fork repair in bacteria. *DNA Repair*. 56, 118–128.

Lovett, S.T., Drapkin, P.T., Sutera-Jr., V.A., and Gluckman-Peskind, T.J. (1993). A sister-strand exchange mechanism for *recA*-independent deletion of repeated DNA sequences in *Escherichia Coli*. *Genetics*. 135, 631–642.

Lu, D., and Keck, J.L. (2008). Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of Exonuclease I. *Proc. Natl. Acad. Sci. U. S. A*. 105, 9169–9174.

Lu, D., Myers, A.R., George, N.P., and Keck, J.L. (2011). Mechanism of Exonuclease I stimulation by the single-stranded DNA-binding protein. *Nucleic Acids Res*. 39, 6536–6545.

Lu, M., Campbell, J.L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell*. 77, 413–426.

Maduike, N.Z., Tehranchi, A.K., Wang, J.D., and Kreuzer, K.N. (2014). Replication of the *Escherichia coli* chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. *Mol. Microbiol*. 91, 39–56.

Mahdi, A.A., Buckman, C., Harris, L., and Lloyd, R.G. (2006). Rep and PriA helicase activities prevent RecA from provoking unnecessary recombination during replication fork repair. *Genes Dev*. 20, 2135–2147.

Maisnier-Patin, S., Nordström, K., and Dasgupta, S. (2001). Replication arrests during a single round of replication of the *Escherichia coli* chromosome in the absence of DnaC activity. *Mol. Microbiol*. 42, 1371–1382.

Manosas, M., Perumal, S.K., Bianco, P., Ritort, F., Benkovic, S.J., and Croquette, V. (2013). RecG and UvsW catalyse robust DNA rewinding critical for stalled DNA replication fork rescue. *Nat. Commun*. 4.

Marbach, A., and Bettenbrock, K. (2012). *lac* operon induction in *Escherichia coli*: systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J. Biotechnol*. 157, 82–88.

Marians, K.J., Hiasa, H., Kim, D.R., and McHenry, C.S. (1998). Role of the core DNA polymerase III subunits at the replication fork: α is the only subunit required for processive replication. *J. Biol. Chem*. 273, 2452–2457.

Marinus, M.G. (2012). DNA Mismatch Repair. *EcoSal Plus*. 5.

Marinus, M.G., and Løbner-Olesen, A. (2014). DNA Methylation. *EcoSal Plus*. 6.

Markovitz, A. (2005). A new *in vivo* termination function for DNA polymerase I of *Escherichia coli* K12. *Mol. Microbiol*. 55, 1867–1882.

Matson, S.W., and Robertson, A.B. (2006). The UvrD helicase and its modulation by the mismatch repair protein MutL. *Nucleic Acids Res*. 34, 4089–4097.

McGlynn, P., and Guy, C.P. (2008). Replication forks blocked by protein–DNA complexes have limited stability *in vitro*. *J. Mol. Biol*. 381, 249–255.

McGlynn, P., and Lloyd, R.G. (2000). Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell*. 101, 35–45.

McGlynn, P., and Lloyd, R.G. (2002a). Recombinational repair and restart of damaged replication forks. *Nat. Rev. Mol. Cell Biol*. 3, 859–870.

McGlynn, P., and Lloyd, R.G. (2002b). Genome stability and the processing of damaged replication forks by RecG. *Trends Genet*. 18, 413–419.

McGlynn, P., Al-Deib, A.A., Liu, J., Marians, K.J., and Lloyd, R.G. (1997). The DNA replication protein PriA and the recombination protein RecG bind D-loops. *J. Mol. Biol*. 270, 212–221.

McGlynn, P., Mahdi, A.A., and Lloyd, R.G. (2000). Characterisation of the catalytically active form of RecG helicase. *Nucleic Acids Res*. 28, 2324–2332.

McGlynn, P., Lloyd, R.G., and Marians, K.J. (2001). Formation of Holliday junctions by regression of nascent DNA in intermediates containing stalled replication forks: RecG stimulates regression even when the DNA is negatively supercoiled. *Proc. Natl. Acad. Sci*. 98, 8235–8240.

McGlynn, P., Savery, N.J., and Dillingham, M.S. (2012). The conflict between DNA replication and transcription. *Mol. Microbiol*. 85, 12–20.

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

McLean, M.J., Wolfe, K.H., and Devine, K.M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol*. 47, 691–696.

Meddows, T.R., Savory, A.P., and Lloyd, R.G. (2004). RecG helicase promotes DNA double-strand break repair. *Mol. Microbiol*. 52, 119–132.

Merrikh, H., Machón, C., Grainger, W.H., Grossman, A.D., and Soultanas, P. (2011). Co-directional replication–transcription conflicts lead to replication restart. *Nature*. 470, 554–557.

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev*. 26, 355–374.

Mettrick, K.A., and Grainge, I. (2016). Stability of blocked replication forks *in vivo*. *Nucleic Acids Res*. 44, 657–668.

Meyenburg, K. von, Boye, E., Skarstad, K., Koppes, L., and Kogoma, T. (1987). Mode of initiation of constitutive stable DNA replication in RNase H-defective mutants of *Escherichia coli* K-12. *J. Bacteriol*. 169, 2650–2658.

Meyer, R.R., and Laine, P.S. (1990). The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol. Rev*. 54, 342–380.

Michel, B., and Leach, D. (2012). Homologous recombination—enzymes and pathways. *EcoSal Plus*. 5.

Michel, B., and Sandler, S.J. (2017). Replication restart in bacteria. *J. Bacteriol*. 199, e00102-17.

Michel, B., Flores, M.-J., Viguera, E., Grompone, G., Seigneur, M., and Bidnenko, V. (2001). Rescue of arrested replication forks by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A*. 98, 8181–8188.

Midgley-Smith, S.L., Dimude, J.U., Taylor, T., Forrester, N.M., Upton, A.L., Lloyd, R.G., and Rudolph, C.J. (2018). Chromosomal over-replication in *Escherichia coli recG* cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed. *Nucleic Acids Res*. 46, 7701–7715.

Midgley-Smith, S.L., Dimude, J.U., and Rudolph, C.J. (2019). A role for 3′ exonucleases at the final stages of chromosome duplication in *Escherichia coli*. *Nucleic Acids Res*. 47, 1847–1860.

Mirkin, E.V., and Mirkin, S.M. (2005). Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell. Biol*. 25, 888–895.

Mirkin, E.V., and Mirkin, S.M. (2007). Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev*. 71, 13–35.

Modrich, P. (2016). Mechanisms in *E. coli* and human mismatch repair. *Angew. Chem. Int. Ed Engl*. 55, 8490–8501.

Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem*. 65, 101–133.

Mohanty, B.K., Sahoo, T., and Bastia, D. (1998). Mechanistic studies on the impact of transcription on sequence-specific termination of DNA replication and vice versa. *J. Biol. Chem*. 273, 3051–3059.

Moolman, M.C., Krishnan, S.T., Kerssemakers, J.W.J., Berg, A. van den, Tulinski, P., Depken, M., Reyes-Lamothe, R., Sherratt, D.J., and Dekker, N.H. (2014). Slow unloading leads to DNA-bound β_2-sliding clamp accumulation in live *Escherichia coli* cells. *Nat. Commun*. 5, 5820.

Moolman, M.C., Tiruvadi Krishnan, S., Kerssemakers, J.W.J., de Leeuw, R., Lorent, V., Sherratt, D.J., and Dekker, N.H. (2016). The progression of replication forks at natural replication barriers in live bacteria. *Nucleic Acids Res*. 44, 6262–6273.

Moreau, M.J.J., and Schaeffer, P.M. (2012). Differential Tus-Ter binding and lock formation: implications for DNA replication termination in *Escherichia coli*. *Mol. Biosyst*. 8, 2783–2791.

Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol*. 5, 343–354.

Mulcair, M.D., Schaeffer, P.M., Oakley, A.J., Cross, H.F., Neylon, C., Hill, T.M., and Dixon, N.E. (2006). A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*. *Cell*. 125, 1309–1319.

Müller, C.A., Hawkins, M., Retkute, R., Malla, S., Wilson, R., Blythe, M.J., Nakato, R., Komata, M., Shirahige, K., de Moura, A.P.S., et al. (2014). The dynamics of genome replication using deep sequencing. *Nucleic Acids Res*. 42, e3.

Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K., and Bastia, D. (2001). Mechanism of termination of DNA replication of *Escherichia coli* involves helicase–contrahelicase interaction. *Proc. Natl. Acad. Sci. U. S. A*. 98, 9569–9574.

Murphy, K.C. (1991). Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme. *J. Bacteriol*. 173, 5808–5821.

Murphy, K.C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol*. 180, 2063–2071.

Nassonova, E.S. (2008). Pulsed field gel electrophoresis: Theory, instruments and application. *Cell Tissue Biol*. 2, 557.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol*. 11, 220–228.

Neylon, C., Kralicek, A.V., Hill, T.M., and Dixon, N.E. (2005). Replication termination in *Escherichia coli*: structure and antihelicase activity of the Tus-Ter complex. *Microbiol. Mol. Biol. Rev*. 69, 501–526.

Nielsen, O., and Løbner-Olesen, A. (2008). Once in a lifetime: strategies for preventing re-replication in prokaryotic and eukaryotic cells. *EMBO Rep*. 9, 151–156.

Nievera, C., Torgue, J.J.-C., Grimwade, J.E., and Leonard, A.C. (2006). SeqA blocking of DnaA-*oriC* interactions ensures staged assembly of the *E. coli* pre-RC. Mol. *Cell*. 24, 581–592.

Nöllmann, M., Crisona, N.J., and Arimondo, P.B. (2007). Thirty years of *Escherichia coli* DNA gyrase: From *in vivo* function to single-molecule mechanism. *Biochimie*. 89, 490–499.

Nordstrom, K., Bernander, R., and Dasgupta, S. (1991). The *Escherichia coli* cell cycle: one cycle or multiple independent processes that are co-ordinated? *Mol. Microbiol*. 5, 769–774.

Nurse, P., Liu, J., and Marians, K.J. (1999). Two modes of PriA binding to DNA. *J. Biol. Chem*. 274, 25026–25032.

O'Donnell, M. (2006). Replisome architecture and dynamics in *Escherichia coli*. *J. Biol. Chem*. 281, 10653–10656.

Ogawa, T., and Okazaki, T. (1980). Discontinuous DNA replication. *Annu. Rev. Biochem*. 49, 421–457.

Ogawa, T., Pickett, G.G., Kogoma, T., and Kornberg, A. (1984). RNase H confers specificity in the *dnaA*-dependent initiation of replication at the unique origin of the *Escherichia coli* chromosome *in vivo* and *in vitro*. *Proc. Natl. Acad. Sci. U. S. A*. 81, 1040–1044.

Ogawa, T., Yamada, Y., Kuroda, T., Kishi, T., and Moriya, S. (2002). The *datA* locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in *Escherichia coli*. *Mol. Microbiol*. 44, 1367–1375.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci. U. S. A*. 59, 598–605.

Osella, M., Tans, S.J., and Lagomarsino, M.C. (2017). Step by step, cell by cell: quantification of the bacterial cell cycle. *Trends Microbiol*. 25, 250–256.

Pham, T.M., Tan, K.W., Sakumura, Y., Okumura, K., Maki, H., and Akiyama, M.T. (2013). A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Mol. Microbiol*. 90, 584–596.

Phillips, G.J., and Kushner, S.R. (1987). Determination of the nucleotide sequence for the exonuclease I structural gene (*sbcB*) of *Escherichia coli* K12. *J. Biol. Chem*. 262, 455–459.

Plank, M., Wadhams, G.H., and Leake, M.C. (2009). Millisecond timescale slimfield imaging and automated quantification of single fluorescent protein molecules for use in probing complex biological processes. *Integr. Biol*. 1, 602–612.

Poleszak, K., Kaminska, K.H., Dunin-Horkawicz, S., Lupas, A., Skowronek, K.J., and Bujnicki, J.M. (2012). Delineation of structural domains and identification of functionally important residues in DNA repair enzyme exonuclease VII. *Nucleic Acids Res*. 40, 8163–8174.

Pomerantz, R.T., and O'Donnell, M. (2007). Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol*. 15, 156–164.

Poveda, A.M., Le Clech, M., and Pasero, P. (2010). Transcription and replication. *Transcription*. 1, 99–102.

Ralf, C., Hickson, I.D., and Wu, L. (2006). The Bloom's Syndrome helicase can promote the regression of a model replication fork. *J. Biol. Chem*. 281, 22839–22846.

Ravin, N.V. (2011). N15: The linear phage–plasmid. *Plasmid*. 65, 102–109.

Ravin, N.V. (2015). Replication and maintenance of linear phage-plasmid N15. *Microbiol. Spectr*. 3.

Ravin, N.V., Strakhova, T.S., and Kuprianov, V.V. (2001). The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form. *J. Mol. Biol*. 312, 899–906.

Reyes-Lamothe, R., Possoz, C., Danilova, O., and Sherratt, D.J. (2008). Independent positioning and action of *Escherichia coli* replisomes in live cells. *Cell*. 133, 90–102.

Reyes-Lamothe, R., Sherratt, D.J., and Leake, M.C. (2010). Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science*. 328, 498–501.

Reyes-Lamothe, R., Nicolas, E., and Sherratt, D.J. (2012). Chromosome replication and segregation in bacteria. *Annu. Rev. Genet*. 46, 121–143.

Riber, L., and Løbner-Olesen, A. (2005). Coordinated replication and sequestration of *oriC* and *dnaA* are required for maintaining controlled once-per-cell-cycle initiation in *Escherichia coli*. *J. Bacteriol*. 187, 5605–5613.

Robinson, A., and van Oijen, A.M. (2013). Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol*. 11, 303–315.

Rocha, E.P.C. (2008). The organization of the bacterial genome. *Annu. Rev. Genet*. 42, 211–233.

Rocha, E.P.C., Cornet, E., and Michel, B. (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLOS Genet*. 1, e15.

Roecklein, B., Pelletier, A., and Kuempel, P. (1991). The *tus* gene of *Escherichia coli*: autoregulation, analysis of flanking sequences and identification of a complementary system in *Salmonella typhimurium*. *Res. Microbiol*. 142, 169–175.

Rosche, W.A., and Foster, P.L. (2000). Determining mutation rates in bacterial populations. *Methods San Diego Calif*. 20, 4–17.

Rowen, L., and Kornberg, A. (1978). Primase, the *dnaG* protein of *Escherichia coli*. An enzyme which starts DNA chains. *J. Biol. Chem*. 253, 758–764.

Rudolph, C.J., Dhillon, P., Moore, T., and Lloyd, R.G. (2007a). Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair*. 6, 981–993.

Rudolph, C.J., Upton, A.L., and Lloyd, R.G. (2007b). Replication fork stalling and cell cycle arrest in UV-irradiated *Escherichia coli*. *Genes Dev*. 21, 668–681.

Rudolph, C.J., Upton, A.L., and Lloyd, R.G. (2008). Maintaining replication fork integrity in UV-irradiated *Escherichia coli cells*. *DNA Repair*. 7, 1589–1602.

Rudolph, C.J., Upton Amy L., Harris Lynda, and Lloyd Robert G. (2009a). Pathological replication in cells lacking RecG DNA translocase. *Mol. Microbiol*. 73, 352–366.

Rudolph, C.J., Upton, A.L., and Lloyd, R.G. (2009b). Replication fork collisions cause pathological chromosomal amplification in cells lacking RecG DNA translocase. *Mol. Microbiol*. 74, 940–955.

Rudolph, C.J., Mahdi, A.A., Upton, A.L., and Lloyd, R.G. (2010a). RecG protein and single-strand DNA exonucleases avoid cell lethality associated with PriA helicase activity in *Escherichia coli*. *Genetics*. 186, 473–492.

Rudolph, C.J., Upton, A.L., Briggs, G.S., and Lloyd, R.G. (2010b). Is RecG a general guardian of the bacterial genome? *DNA Repair*. 9, 210–223.

Rudolph, C.J., Upton, A.L., Stockum, A., Nieduszynski, C.A., and Lloyd, R.G. (2013). Avoiding chromosome pathology when replication forks collide. *Nature*. 500, 608–611.

Rybchin, V.N., and Svarchevsky, A.N. (1999). The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol. Microbiol*. 33, 895–903.

Sandigursky, M., and Franklin, W.A. (1994). *Escherichia coli* single-stranded DNA binding protein stimulates the DNA deoxyribophosphodiesterase activity of Exonuclease I. *Nucleic Acids Res*. 22, 247–250.

Sandler, S.J. (2005). Requirements for replication restart proteins during constitutive stable DNA replication in *Escherichia coli* K-12. *Genetics*. 169, 1799–1806.

Sandler, S.J., and Marians, K.J. (2000). Role of PriA in replication fork reactivation in *Escherichia coli*. *J. Bacteriol*. 182, 9–13.

Sandler, S.J., Marians, K.J., Zavitz, K.H., Coutu, J., Parent, M.A., and Clark, A.J. (1999). *dnaC* mutations suppress defects in DNA replication- and recombination-associated functions in *priB* and *priC* double mutants in *Escherichia coli* K-12. *Mol. Microbiol*. 34, 91–101.

Sandler, S.J., McCool, J.D., Do, T.T., and Johansen, R.U. (2001). PriA mutations that affect PriA–PriC function during replication restart. *Mol. Microbiol*. 41, 697–704.

Sasaki, M., Lange, J., and Keeney, S. (2010). Genome destabilization by homologous recombination in the germ line. *Nat. Rev. Mol. Cell Biol*. 11, 182–195.

Scheuermann, R.H., and Echols, H. (1984). A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. U. S. A*. 81, 7747–7751.

Schofield, M.J., and Hsieh, P. (2003). DNA mismatch repair: molecular mechanisms and biological function. *Annu. Rev. Microbiol*. 57, 579–608.

Shammas, M.A., Shmookler Reis, R.J., Koley, H., Batchu, R.B., Li, C., and Munshi, N.C. (2009). Dysfunctional homologous recombination mediates genomic instability and progression in myeloma. *Blood*. 113, 2290–2297.

Sharma, B., and Hill, T.M. (1995). Insertion of inverted Ter sites into the terminus region of the *Escherichia coli* chromosome delays completion of DNA replication and disrupts the cell cycle. *Mol. Microbiol*. 18, 45–61.

Sharples, G.J., Ingleston, S.M., and Lloyd, R.G. (1999). Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA. *J. Bacteriol*. 181, 5543–5550.

Shen, Z. (2011). Genomic instability and cancer: an introduction. *J. Mol. Cell Biol*. 3, 1–3.

Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., and Keck, J.L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol*. 43, 289–318.

Sherratt, D.J., Arciszewska, L.K., Crozat, E., Graham, J.E., and Grainge, I. (2010). The *Escherichia coli* DNA translocase FtsK. *Biochem. Soc. Trans*. 38, 395–398.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*. 122, 19–27.

Singleton, M.R., Scaife, S., and Wigley, D.B. (2001). Structural analysis of DNA replication fork reversal by RecG. *Cell*. 107, 79–89.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem*. 76, 23–50.

Sinha, A.K., Possoz, C., Durand, A., Desfontaines, J.-M., Barre, F.-X., Leach, D.R.F., and Michel, B. (2018). Broken replication forks trigger heritable DNA breaks in the terminus of a circular chromosome. *PLoS Genet*. 14.

Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol*. 5.

Skarstad, K., Boye, E., and Steen, H.B. (1986). Timing of initiation of chromosome replication in individual *Escherichia coli* cells. *EMBO J*. 5, 1711–1717.

Skokotas, A., Hiasa, H., Marians, K.J., O'Donnell, L., and Hill, T.M. (1995). Mutations in the *Escherichia coli* Tus protein define a domain positioned close to the DNA in the Tus-Ter complex. *J. Biol. Chem*. 270, 30941–30948.

Skovgaard, O., Bak, M., Løbner-Olesen, A., and Tommerup, N. (2011). Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res*. 21, 1388–1393.

Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., and Kleckner, N. (1995). *E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell*. 82, 927–936.

Smith, G.R. (2012). How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiol. Mol. Biol. Rev*. 76, 217–228.

Sousa, C., de Lorenzo, V., and Cebolla, A. (1997). Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology*. 143, 2071–2078.

Southern, E.M., Anand, R., Brown, W.R., and Fletcher, D.S. (1987). A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res*. 15, 5925–5943.

Speck, C., Weigel, C., and Messer, W. (1999). ATP- and ADP-DnaA protein, a molecular switch in gene regulation. *EMBO J*. 18, 6169–6176.

Spies, M., and Fishel, R. (2015). Mismatch repair during homologous and homeologous recombination. *Cold Spring Harb. Perspect. Biol*. 7.

Stockum, A., Lloyd, R.G., and Rudolph, C.J. (2012). On the viability of *Escherichia coli* cells lacking DNA topoisomerase I. *BMC Microbiol*. 12, 26.

Storm, P.K., Hoekstra, W.P.M., De Haan, P.G., and Verhoef, C. (1971). Genetic recombination in *Escherichia coli*: IV. Isolation and characterization of recombinaion-deficient mutants of *Escherichia coli* K12. *Mutat. Res. Mol. Mech. Mutagen*. 13, 9–17.

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

Su, S.S., and Modrich, P. (1986). *Escherichia coli mutS*-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci*. 83, 5057–5061.

Su, S.S., Lahue, R.S., Au, K.G., and Modrich, P. (1988). Mispair specificity of methyl-directed DNA mismatch correction *in vitro*. *J. Biol. Chem*. 263, 6829–6835.

Tadokoro, T., and Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J*. 276, 1482–1493.

Taft-Benz, S.A., and Schaaper, R.M. (2004). The θ subunit of *Escherichia coli* DNA polymerase III: a role in stabilizing the ε proofreading subunit. *J. Bacteriol*. 186, 2774–2780.

Tanaka, T., and Masai, H. (2006). Stabilization of a stalled replication fork by concerted actions of two helicases. *J. Biol. Chem*. 281, 3484–3493.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. 329, 533–538.

Thomas, K.R., and Olivera, B.M. (1978). Processivity of DNA exonucleases. *J. Biol. Chem*. 253, 424–429.

Tomasetti, C., Li, L., and Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*. 355, 1330–1334.

Toro, E., and Shapiro, L. (2010). Bacterial chromosome organization and segregation. *Cold Spring Harb. Perspect. Biol*. 2, a000349.

Trautinger, B.W., Jaktaji, R.P., Rusakova, E., and Lloyd, R.G. (2005). RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol. Cell*. 19, 247–258.

Turner, J., Hingorani, M.M., Kelman, Z., and O'Donnell, M. (1999). The internal workings of a DNA polymerase clamp-loading machine. *EMBO J*. 18, 771–783.

Upton, A.L., Grove, J.I., Mahdi, A.A., Briggs, G.S., Milner, D.S., Rudolph, C.J., and Lloyd, R.G. (2014). Cellular location and activity of *Escherichia coli* RecG proteins shed light on the function of its structurally unresolved C-terminus. *Nucleic Acids Res*. 42, 5702–5714.

Usongo, V., Martel, M., Balleydier, A., and Drolet, M. (2016). Mutations reducing replication from R-loops suppress the defects of growth, chromosome segregation and DNA supercoiling in cells lacking topoisomerase I and RNase HI activity. *DNA Repair*. 40, 1–17.

Vales, L.D., Rabin, B.A., and Chase, J.W. (1982). Subunit structure of *Escherichia coli* Exonuclease VII. *J. Biol. Chem*. 257, 8799–8805.

Vales, L.D., Rabin, B.A., and Chase, J.W. (1983). Isolation and preliminary characterization of *Escherichia coli* mutants deficient in Exonuclease VII. *J. Bacteriol*. 155, 1116–1122.

Vincent, S.D., Mahdi, A.A., and Lloyd, R.G. (1996). The RecG branch migration protein of *Escherichia coli* dissociates R-loops. *J. Mol. Biol*. 264, 713–721.

Viswanathan, M., and Lovett, S.T. (1998). Single-strand DNA-specific exonucleases in *Escherichia coli*. Roles in repair and mutation avoidance. *Genetics*. 149, 7–16.

Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P., and Lovett, S.T. (2001). Redundant exonuclease involvement in *Escherichia coli* methyl-directed mismatch repair. *J. Biol. Chem*. 276, 31053–31058.

Wang, J.C. (1996). DNA Topoisomerases. *Annu. Rev. Biochem*. 65, 635–692.

Wang, T.-C.V. (2005). Discontinuous or semi-discontinuous DNA replication in *Escherichia coli*? *BioEssays*. 27, 633–636.

Wang, J.D., Berkmen, M.B., and Grossman, A.D. (2007). Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A*. 104, 5608–5613.

Wang, X., Lesterlin, C., Reyes-Lamothe, R., Ball, G., and Sherratt, D.J. (2011). Replication and segregation of an *Escherichia coli* chromosome with two replication origins. *Proc. Natl. Acad. Sci*. 108, E243–E250.

Welsh, K.M., Lu, A.L., Clark, S., and Modrich, P. (1987). Isolation and characterization of the *Escherichia coli mutH* gene product. *J. Biol. Chem*. 262, 15624–15629.

Wendel, B.M., Courcelle, C.T., and Courcelle, J. (2014). Completion of DNA replication in *Escherichia coli*. *Proc. Natl. Acad. Sci*. 111, 16454–16459.

Wendel, B.M., Cole, J.M., Courcelle, C.T., and Courcelle, J. (2018). SbcC-SbcD and ExoI process convergent forks to complete chromosome replication. *Proc. Natl. Acad. Sci. U. S. A*. 115, 349–354.

West, S.C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet*. 31, 213–244.

Whitby, M.C. (2010). The FANCM family of DNA helicases/translocases. *DNA Repair*. 9, 224–236.

Wiktor, J., van der Does, M., Büller, L., Sherratt, D.J., and Dekker, C. (2018). Direct observation of end resection by RecBCD during double-stranded DNA break repair *in vivo*. *Nucleic Acids Res*. 46, 1821–1833.

Wimmer, K., and Etzler, J. (2008). Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? *Hum. Genet*. 124, 105–122.

Windgassen, T.A., Wessel, S.R., Bhattacharyya, B., and Keck, J.L. (2018). Mechanisms of bacterial DNA replication restart. *Nucleic Acids Res*. 46, 504–519.

Wold, S., Skarstad, K., Steen, H.B., Stokke, T., and Boye, E. (1994). The initiation mass for DNA replication in *Escherichia coli* K-12 is dependent on growth rate. *EMBO J*. 13, 2097–2102.

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

Yamaki, H., Ohtsubo, E., Nagai, K., and Maeda, Y. (1988). The *oriC* unwinding by *dam* methylation in *Escherichia coli*. *Nucleic Acids Res*. 16, 5067–5073.

Yao, N.Y., and O'Donnell, M. (2010). SnapShot: The Replisome. *Cell*. 141, 1088-1088.e1.

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

Yeeles, J.T.P., and Marians, K.J. (2011). The *Escherichia coli* replisome is inherently DNA damage tolerant. *Science*. 334, 235–238.

Yu, C., Tan, H.Y., Choi, M., Stanenas, A.J., Byrd, A.K., Raney, K., Cohan, C.S., and Bianco, P.R. (2016). SSB binds to the RecG and PriA helicases *in vivo* in the absence of DNA. *Genes Cells Devoted Mol. Cell. Mech*. 21, 163–184.

Zahradka, K., Buljubašić, M., Petranović, M., and Zahradka, D. (2009). Roles of ExoI and SbcCD nucleases in "reckless" DNA degradation in *recA* mutants of *Escherichia coli*. *J. Bacteriol*. 191, 1677–1687.

Zavitz, K.H., and Marians, K.J. (1992). ATPase-deficient mutants of the *Escherichia coli* DNA replication protein PriA are capable of catalyzing the assembly of active primosomes. *J. Biol. Chem*. 267, 6933–6940.

Zechiedrich, E.L., and Cozzarelli, N.R. (1995). Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in *Escherichia coli*. *Genes Dev*. 9, 2859–2869.

Zechiedrich, E.L., Khodursky, A.B., and Cozzarelli, N.R. (1997). Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*. *Genes Dev*. 11, 2580–2592.

Zhang, J., Mahdi, A.A., Briggs, G.S., and Lloyd, R.G. (2010). Promoting and avoiding recombination: contrasting activities of the *Escherichia coli* RuvABC Holliday junction resolvase and RecG DNA translocase. *Genetics*. 185, 23–37.

Zhivotovsky, B., and Kroemer, G. (2004). Apoptosis and genomic instability. *Nat. Rev. Mol. Cell Biol*. 5, 752–762.

Appendix

Linearisation of the E. coli chromosome

Although the chromosomes of most prokaryotes, including *E. coli*, are circular, some bacteriophages as well as some bacterial species have linear chromosomes like eukaryotes. Cui and colleagues (Cui et al., 2007) exploited the linearization mechanism of the *E. coli* bacteriophage N15 in order to investigate whether the *E. coli* chromosome could be linearised or not and to assess any effects on prokaryotic cells that result from having a linear chromosome. Bacteriophage N15 is a temperate λ-like phage, which means that it is capable of both a lytic and lysogenic pathway upon infection of host cells (Gandon, 2016). Unlike other phage with lysogenic capability, N15 is unusual in that it does not insert its genetic material in to the host chromosome during lysogenic infection. Instead, the N15 prophage (the viral genome within the host cell) exists as a plasmid molecule in the cytoplasm of host *E. coli* cells (reviewed in Ravin, 2011).

The DNA of mature N15 phage is linear and has a 12 bp single-stranded DNA over-hang at each end of the molecule, called *cosL* and *cosR*. These ends are complimentary to one another (cohesive ends) and anneal following infection of *E. coli* cells, resulting in a circular phage DNA molecule (Ravin, 2011; Ravin et al., 2001). In a lysogenic infection, this circular DNA molecule undergoes linearization resulting in a linear N15 prophage molecule with covalently closed hairpin ends. The bacteriophage telomerase TelN (encoded by the *telN* gene) is responsible for this process (Deneke et al., 2000; Ravin et al., 2001; Rybchin and Svarchevsky, 1999). The TelN target sequence on the N15 circular phage DNA is *telRL*, which is located within *tos* (telomerase occupancy site). *tos* consists of a central 22-bp palindromic sequence, *telO*, which is flanked by a series of inverted repeat sequence pairs, R1/L1, R2/L2 and R3/L3. The region spanning from R3 to L3 with *telO* at the centre forms *telRL*, the binding site for TelN protelomerase (Deneke et al., 2000, 2002) and the 56-bp sequence an almost perfect palindrome. TelN binds to *telRL* and introduces a staggered double stranded break within *telO*, yielding a linear molecule. Because of the palindromic sequence of *telRL*, the single-stranded ends are self-complementary and form the left and right hairpin ends of the molecule. TelN completes telomere resolution by covalently closing the ends (Ravin, 2011, 2015; Rybchin and Svarchevsky, 1999).

Ravin and colleagues (Ravin et al., 2001) cloned a DNA fragment containing *telN* and the TelN target site *telRL* from *tos* in to various plasmids and used them to demonstrate that these two components alone are sufficient to achieve linearization of a non-N15 circular DNA molecule, although the full *tos* sequence results in the most efficient binding of TelN to *telRL* (Deneke et al., 2002). The autonomy of this linearization system was exploited and enabled the development of a simple and effective method to linearise the chromosome in *E. coli* cells (Cui et al., 2007). The *tos*

sequence was inserted near to *dif* in the termination area of the *E. coli* chromosome and a kanamycin resistance marker (*kan*) was inserted directly downstream of the *tos* sequence in order to create a construct that was transferrable by P1*vir* transduction. The *telN* gene was cloned in to pBAD24 under the control of the arabinose promoter and *tos-kan* cells were then transformed with the resulting plasmid, enabling telomere resolution by TelN to produce a linear DNA molecule. Although the *telN* gene lies adjacent to *tos* in the N15 DNA and could have been included in the cloned DNA fragment to make the linearization protocol simpler, the separation of the *tos* sequence and TelN allows for more robust control of the system in an experimental setting as it allows control in having cells with circular and cells with linearised chromosomes. However, using a plasmid to achieve this generated cells that contained both linear (chromosome) and circular (plasmid) replicons, which caused problems when investigating separation of chromosomes following replication. Another way of providing the cell with TelN was lysogenic infection of tos-*kan* cells with N15, resulting in expression of *telN* by the N15 prophage. Once the strains were made, it was confirmed that the *E. coli* cells remained viable and the linearised chromosome structure was stable. Linearisation did not affect the growth of wild type cells (Cui et al., 2007), a finding that was subsequently independently confirmed (Rudolph et al., 2013).

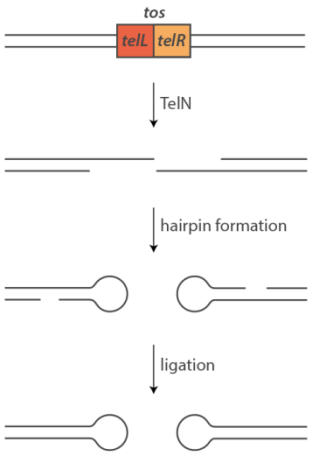

Figure 41: Schematic outlining the mechanism of chromosome linearisation.

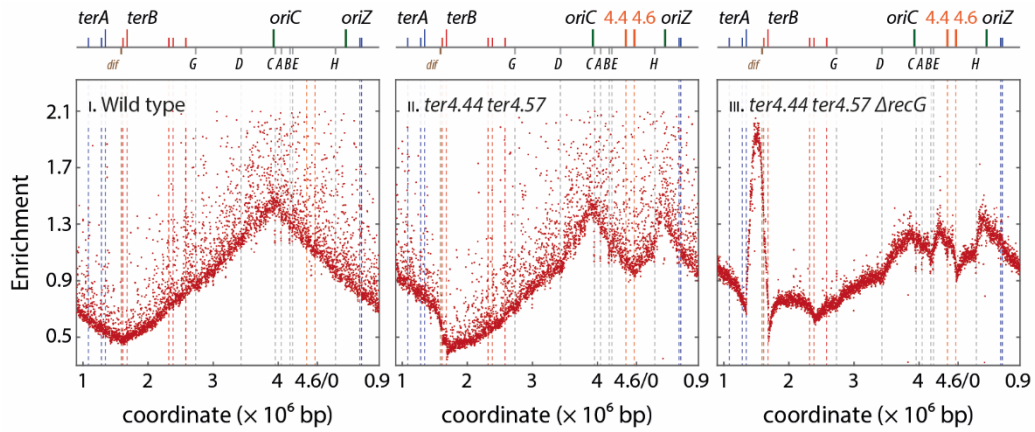

Figure 42: Marker frequency analysis of exponential phase *E. coli* cells with two replication origins and an ectopic termination area, in the presence and absence of RecG. Read numbers (normalised against a stationary phase wild type control) are plotted against the chromosomal coordinates, which are offset to start at 0.9 Mb to make it easier to see both replication origins. The schematic above the graphs is a representation of the *E. coli* chromosome, showing the position of *oriC* and *oriZ* (green lines) and *ter* sites (red lines for those in the left hand replichore and blue lines for those in the right hand replichore) above and the seven *rrn* operons and *dif* below. The strains used were MG 1655 (wild type), SLM 1197 (*oriZ ter 4.44 Mb><ter 4.57 Mb*) and SLM 1205 (*oriZ ter 4.44 Mb><ter 4.57 Mb recG*).

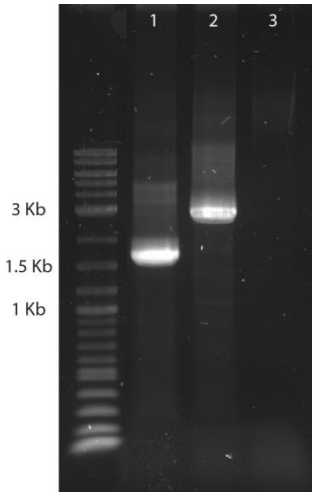

Figure 43: Linearisation of the *E. coli* chromosome in *dnaA Δtus rpo* ΔxonA ΔxseA* cells. PCR products generated with the linearisation verification primers for wild type (lane 1), *tos-kan* cells (lane 2) and N15 lysogen of *tos-kan* with a linearised chromosome (lane 3). The increase in product size between lane 1 and lane 2 indicates the presence of the *tos-kan* cassette. The N15 telomerase TelN is present in N15 lysogens. TelN cleaves its target sequence within *tos*, resulting in linearisation of the chromosome in *tos-kan* N15 lysogens. The linearisation site is located between the primer binding sites and so prevents the formation of a PCR product. For primer binding sites, see Figure 28. The chromosomal DNA was prepared from SLM1232 (*dnaA46 Δtus rpo* ΔxonA ΔxseA* N15), SLM1225 (*dnaA46 Δtus rpo* ΔxonA ΔxseA tos-kan*) and SLM1230 (*dnaA46 Δtus rpo* ΔxonA ΔxseA tos-kan* N15 lysogen).

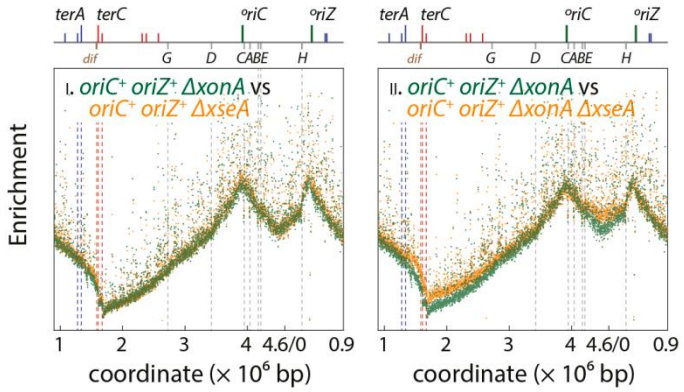

Figure 44: Replication profiles of *oriC*⁺ *oriZ*⁺ cells in the absence of ExoI and ExoVII. Data in II are replotted from Figure 38. The overlay of the replication profile for *oriC*⁺ *oriZ*⁺ Δ*xonA* cells with that of *oriC*⁺ *oriZ*⁺ Δ*xseA* cells demonstrates that there are no significant differences between the two profiles, which validates the use of the *oriC*⁺ *oriZ*⁺ Δ*xonA* profile as a baseline to compare the profile of double origin cells lacking both ExoI and ExoVII (II). The profiles were aligned according to peak height, which is explained previously (page 127).

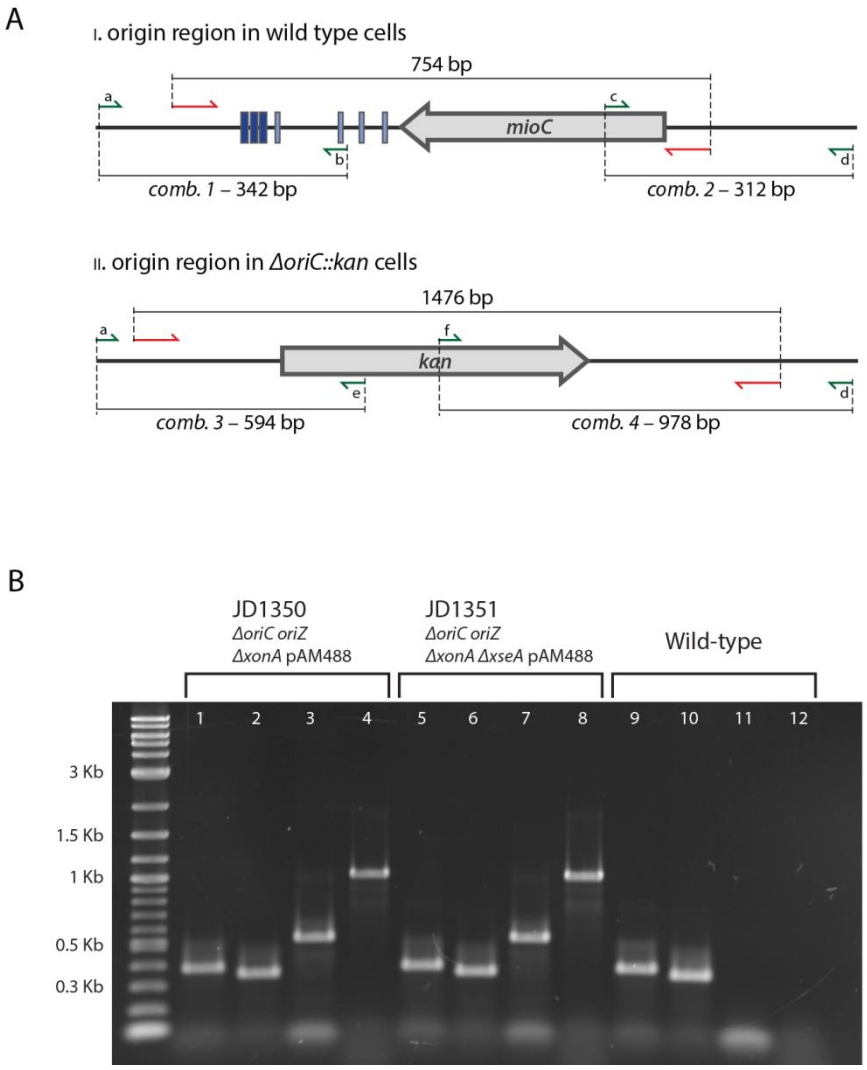

Figure 45: Deletion of the native replication origin, *oriC* in an *oriZ*$^{+}$ strain background. **A**) Schematic representation of the origin region in wild type cells and cells with the origin replaced by a gene for kanamycin resistance. The primers used to generate the *oriC*-disruption DNA are shown in red. The verification primers used to confirm the deletion of *oriC* are shown in green (Rudolph et al., 2013). **B**) PCR products generated via Hot Start PCR (page 54) with the *oriC*-deletion verification primers are shown for primer combination a/b (lanes 1, 5 and 9), c/d (lanes 2,6 and 10), a/e (lanes 3,7 and 11) and f/d (lanes 4, 8 and 12).

To generate an ectopic replication origin, a copy of a 5 kb region spanning *oriC* was generated and integrated into the chromosome at 344 kb (Wang et al., 2011), (Figure 45, AI; Rudolph et al., 2013). This fragment spans all of the binding sites for the verification primers shown in Figure 45 A. The presence of PCR products with the a/b and c/d primer combinations in the *ΔoriC* strains is due to the presence of *oriZ*, the sequence of which is identical to the *oriC* region. The verification of the *kan* flanks in the *ΔoriC* samples, combined with the absence of PCR products with these primers in the wild type control sample, confirms that *oriC* in the native location has been replaced by the kanamycin resistance cassette.