Investigating the mutagenicity of DNA methyltransferase-induced DNA damage

Katie Hains

A thesis submitted for the degree of Doctor of Philosophy at the University of Oxford



Department of Biochemistry
Lincoln College
University of Oxford

Hilary Term 2026

I. Abstract

DNA methylation is an important epigenetic modification introduced by DNA methyltransferases (DNMTs). However, many lineages exhibit reduced or absent methylation. The propensity of DNMTs to damage DNA might contribute to this frequent evolutionary loss. 5mC spontaneously deaminates at an approximately two-fold greater rate than unmethylated cytosine and thus results in increased rate of C > T mutations. Interestingly, previous work showed that DNMTs alkylate DNA, introducing the toxic lesion 3mC. Furthermore, it was found that the mammalian DNA alkylation damage repair enzyme ALKB2 co-evolved with DNMTs, likely to mitigate 3mC-induced damage. However, the contribution of 3mC to DNMT-induced mutagenesis remains unknown. Here, I introduced CpG methylation into *E. coli,* which possess the single ALKB2 homologue AlkB. I used a rifampicin resistance assay to quantitatively determine the contribution of DNMT-induced DNA damage to mutagenesis in WT and Δ*alkB E. coli*. DNMT activity significantly increased rifampicin-resistant mutation frequency in AlkB mutants compared to WT (ΔalkB * M.SssI interaction $p<0.05$, two-way Anova), and induced C>T mutations. Surprisingly, constitutive DNMT activity in AlkB mutants led to additional TA > CG mutations. This signature was dependent on strain background; however, it was unlikely to be RecA-dependent as it was unaffected either by RecA depletion or overexpression. Finally, DNMT activity increased ROS independent of exogenous oxidative stress *in vivo*. Together, this suggested that DNMT-induced mutagenesis may have contributed to the factors driving repeated loss of DNA methylation across evolution.

II. Acknowledgements

Firstly, I would like to thank my supervisor Peter Sarkies for the opportunity to do my PhD in the EpiEvo lab. I have learnt a huge amount over the past three and a half years and have developed immensely as a scientist. None of this would have been possible without your guidance, patience and support. Thank you for giving me a chance and pushing me to the best of my ability.

I would like to thank all the members of the Sarkies lab. To Jo and Yimo, my day ones. I couldn't have picked two better people to embark on this journey with. I'm so grateful for your support and friendship over the years. To Manon, Max, Mariana, Joseph, Nabeel, Alex, Dan, Tina, Zoe, Poppy, Chris and all part IIs and summer students. Thank you for making the lab such a nice place to work and for the fun times outside of the lab too! I wish you all the best in your careers. You're some very smart and talented people.

I would also like to thank my thesis committee, Neil Brockdorff and Lidia Vasilieva, and secondary supervisor Nick Lakin for their support and useful discussions over the years. Similarly, I would like to thank the Uphoff lab, particularly Valentine Lagage, for help with all things lambda red recombination, and Ben Carty for his help with Nanopore sequencing, even if it didn't make it into this thesis.

A huge thank you must of course go to my friends. To Anna, Abbie and Grace, for being in my life for so many years and us always being close despite the fact we live all over the country. To Abi, Zoe, Molly W, Gian and Molly B, I'm so glad we all

decided Bristol in 2018 was the place to be. Those were the best of times and I look forward to many more. To Lucy, Flo, Tabs and Esme for being some of the best friends and flatmates a girl could wish for. You've seen me through the highs and lows for sure! Genuinely couldn't have done it without you and I'm so grateful for all the laughs, Hinksey swims and sweet treats. To all the girls at OURFC, you've made this experience so much more than just a degree. It's a privilege to be part of something so special. Rugby has been my constant and I'm so lucky to have had such a wonderful group of strong women around me. To the Johnstons, thank you welcoming me into your family and for all your love and support. I wouldn't have made it through the first few years without you and I'll always cherish the time we spent together.

Finally, the biggest thank you goes to my family. Thank you for always believing in me even if you still don't understand exactly what it is that I do. To Mum and Dad who have sacrificed so much for Anna and I. You're my rocks and my biggest inspirations – I can't thank you enough for all you've done for me. This thesis is for you. To Anna, for being my best friend since we were little balls of cells. Blessed to have a sister like you.

III. Abbreviations

1mA	1-methyl adenine
3mC	3-methyl cytosine
4mC	4-methyl cytosine
5caC	5-carboxy cytosine
5fC	5-formyl cytosine
5hmC	5-hydroxymethyl cytosine
5hmU	5-hydroxymethyl uracil
5mC	5-methyl cytosine
6mA	6-methyl adenine
6mG	6-methyl guanine
7mG	7-methyl guanine
8oxoG	8-oxoguanine
AmpR	Ampicillin resistance cassette
AP	Apurinic site
BER	Base excision repair
CamR	Chloramphenicol resistance cassette
CAP	cAMP receptor protein
DDR	DNA damage repair
DNMT	DNA methyltransferase
DSB	Double strand break
dsDNA	Double stranded DNA
ESC	Embryonic stem cell

gDNA	Genomic DNA
HR	Homologous recombination
IQR	Interquartile range
KanR	Kanamycin resistance cassette
KO	Knockout
MGMT	O^6-methylguanine methyltransferase
NER	Nucleotide excision repair
OGG1	8-oxoguanine DNA glycosylase 1
ONT	Oxford Nanopore Technologies
PPP	Pentose phosphate pathway
ROS	Reactive oxygen species
SAM	S-adenosyl methionine
SMRT	Single molecule real time
ssDNA	Single stranded DNA
TDG	Thymine DNA glycosylase
TE	Transposable element
TET	Ten-eleven translocase
TF	Transcription factor
TLS	Translesion synthesis
UNG	Uracil-DNA glycosylase
WGS	Whole genome sequencing
WT	Wild type

IV. Contents

Chapter 1 - Introduction and aims **11**

1.1. Introduction 11

1.1.1. Preface: the complex balance of maintaining genome stability 11

1.1.2. DNA methylation and epigenetics 13

1.1.2.1. The basis of DNA methylation 13

1.1.2.2. The role of DNA methylation in mammalian epigenetics 17

1.1.2.3. Evolutionary variability in DNA methylation 19

1.1.3. DNA methylation and genome stability 21

1.1.3.1. Promotion of genome stability by DNA methylation 21

1.1.3.2. Deamination of 5mC and other contributions to mutagenesis 22

1.1.3.3. Contributions of methylation to cancer mutational signatures 23

1.1.4. Aberrant DNA methylation, damage and repair 26

1.1.4.1. DNA alkylation lesion toxicity 26

1.1.4.2. Direct repair by ALKB2/3 27

1.1.4.3. Base excision repair 30

1.1.4.4. Bypass of alkylation damage by translesion synthesis 34

1.1.5. How DNA damage might influence the loss of DNA methylation 38

1.2. Aims and objectives 39

Chapter 2 – Materials and methods **42**

2.1. *In vitro* work 42

2.1.1. Oligonucleotide design 42

2.1.2. Ligation 43

2.1.3. ROS detection in a cell-free system 44

2.2. *In vivo* work 45

2.2.1. Bacterial strains 45

2.2.2. Plasmids 45

2.2.3. Bacterial media and culture conditions 48

2.2.4. Rifampicin resistance assay 49

2.2.5. Colony PCR 50

2.2.6. Sanger sequencing 51

2.2.7. Lambda red recombination 52

2.2.8. ROS detection assay 54

Chapter 3 – Methylation lesions reduced number of viable transformants in WT and Δ*alkB E. coli* **56**

3.1. Introduction 56

3.2. Results 56

3.2.1. DNA damage construct design and optimisation 56
3.2.2. Transformation and ligation efficiency 58
3.3. Specific discussion 65

Chapter 4 – M.SssI introduces alkylation damage that is specifically repaired by AlkB 67
4.1. General introduction to rifampicin resistance assay 67
4.2. Results 69
4.2.1. Constitutive system 69
4.2.1.1. Introduction to constitutive system 69
4.2.1.2. M.SssI activity and loss of AlkB increased rifampicin-resistant mutation frequency in C2523 *E. coli* 71
4.2.1.3. Identification of TA > CG mutations dependent on M.SssI activity and loss of AlkB 74
4.2.2. Inducible system 76
4.2.2.1. Introduction to inducible system 76
4.2.2.2. Knock out of *alkB* in Top10 by lambda red recombination 78
4.2.2.3. M.SssI activity and loss of AlkB caused synthetic increase in rifampicin-resistant mutation frequency in Top10 *E. coli* 83
4.2.2.4. Non-cytosine mutations were not observed with inducible DNMT activity in Top10 *E. coli* 87
4.2.3. Investigating potential involvement of TLS in the non-cytosine *rpoB*-specific mutation spectrum 89
4.2.3.1. Attempting to knock out *recA* in WT and Δ*alkB* C2523 by lambda red recombination 89
4.2.3.2. Inhibition of RecA with PsiB had no notable effect on rifampicin-resistant mutation frequency 97
4.2.3.3. Inhibition of RecA did not eliminate the non-cytosine *rpoB*-specific mutation spectrum in C2523 *E. coli* 100
4.2.3.4. Introduction of RecA into Top10 *E. coli* 102
4.2.3.5. Introduction of RecA into Δ*alkB* Top10 did not rescue non-cytosine mutations 107
4.2.4. The non-cytosine mutations were not specific to constitutively active M.SssI 109
4.2.4.1. Constitutive M.SssI activity in Top10 *E. coli* did not yield substantial non-cytosine mutations 109
4.2.4.2. Weakly-methylation M.MpeI did not increase Δ*alkB*-dependent non-cytosine mutations 113
4.2.4.3. Increasing the length of time for which inducible M.SssI was active did not rescue the non-cytosine *rpoB*-specific mutation spectrum in Top10 *E. coli* 118

4.2.4.4. Could both RecA activity and constitutive M.SssI activity be required for non-cytosine mutations in Δ*alkB* Top10 *E. coli* 122
4.2.4.4.1. Re-engineering MP1 resistance to be compatible with constitutively active pM.SssI 122
4.2.4.4.2. Induction of RecA and constitutive M.SssI activity in Top10 *E. coli* did not rescue non-cytosine mutations 124
4.3. Specific discussion 128
4.3.1. Limitations of rifampicin resistance assay 131
4.3.1.1. Jackpotting 131
4.3.1.2. Rifampicin resistant mutation distribution 132
4.3.1.3. Alternative analysis methods 141
4.3.2. Potential genetic and cellular contributions to the non-cytosine *rpoB*-specific mutation spectrum 144
4.3.2.1. Non-cytosine mutations were independent of RecA-mediated recombination 144
4.3.2.2. SulA is required for regulation of cell division during DNA damage bypass 145
4.3.2.3. Single-stranded DNA resulting from DNMT-induced alkylation damage could contribute to mutagenesis 146

Chapter 5 – DNA methylation contributes to reactive oxygen species production in *E. coli* 149
5.1. General introduction 149
5.2. Results 151
5.2.1. *In vivo* evidence for the introduction of oxidative stress by M.SssI 151
5.2.1.1. Induction of M.SssI activity alone increased ROS signal in *E. coli* 151
5.2.1.2. Investigating the contributions of experimental conditions to baseline ROS signal 155
5.2.1.3. M.SssI induction led to increased sensitivity to oxidative stress by H_2O_2 164
5.2.2. Investigating how M.SssI activity led to an increase in ROS *in vitro* 167
5.3. Specific discussion 175
5.3.1. Preferential sugar metabolism in *E. coli* 176
5.3.2. How does induction of M.SssI activity increase cellular ROS? 178
5.3.3. Contributions of additional reaction components to ROS levels 179

Chapter 6 – Discussion **182**
6.1. Summary of key findings 182
6.2. Overall insights and remaining questions 184
6.2.1. DNMT-induced mutagenesis and the evolutionary loss of DNA methylation 184
6.2.2. The relationship between DNMT activity and oxidative stress 186
6.3. Future work 188
6.3.1. Direct detection of alkylation lesions 188
6.3.2. Further characterisation of DNMT-induced mutagenesis 191
6.3.2.1. Do DNMTs introduce mutations in mammalian systems? 191
6.3.2.2. Further characterisation of DNMT-induced mutagenesis in *E. coli* 193
6.4. Clinical implications of DNMT-induced genome instability 194
6.5. Overall conclusion 197

V. References **199**

Chapter 1 - Introduction and aims

1.1. Introduction

1.1.1. Preface: The complex balance of maintaining genome stability

One of the biggest challenges faced by all organisms is the faithful propagation of DNA throughout cell division. It is vital to cell survival that the genetic code is copied with minimal error, in order to preserve the important information that it encodes. The double helix structure protects the organic bases and provides a template mechanism for replication, which ensures high fidelity during DNA replication (Franklin & Gosling, 1953; Watson & Crick, 1953). However, there are many endogenous and exogenous factors that can damage DNA and threaten genomic stability. Some DNA damage is inherently mutagenic, for example spontaneous deamination of cytosine that results in a CG > TA point mutation (Lindahl, 1996). Other types of DNA damage have the potential to be mutagenic if not repaired prior to DNA replication. Stalling of the replisome at lesions can lead to incorporation of an incorrect nucleotide opposite the damage (Du et al., 2019; McCulloch & Kunkel, 2008; Tissier et al., 2000). This can become fixed as a mutation after the next round of repair. Alternatively, failure to bypass bulky lesions can result in double strand breaks (DSBs) which are associated with large scale mutations such as translocations and deletions (Sedletska et al., 2013). Even if damage is detected and repaired before DNA replication, not all repair processes are error-free and may also introduce mutations (Bishop & Schiestl, 2000; Kishkevich et al., 2022;

McCulloch & Kunkel, 2008). However, some mutagenesis is necessary to ensure genetic diversity within a population. Some mutations will be advantageous and improve the organism's chances of survival within the environment, therefore driving adaptive evolution (Blundell et al., 2019; Luria & Delbrück, 1943). Conversely, too much mutagenesis is detrimental to genome stability and can be tumorigenic (Alexandrov et al., 2015; Pfeifer, 2010; Stratton et al., 2009).

Clearly, maintenance of genome stability is a complex balance between DNA replication, mutagenesis and repair. Interestingly, we are also starting to recognise the contributions of the epigenome to genome stability. The epigenome consists of a series of mitotically and meiotically stable chemical modifications to DNA that link genotype to phenotype without changing the underlying nucleotide sequence (Berger et al., 2009; Waddington, 2012). The epigenome is central to regulation of gene expression; allows for rapid response to internal and external stimuli; controls stage-specific gene expression during development and ensures dosage compensation via X chromosome inactivation (Brockdorff et al., 1992; Feinberg et al., 2006; Jones, 2012; Li et al., 1992; Waterland & Jirtle, 2003). Changes to the epigenome, termed epimutations, can affect chromatin structure and the ability of repair enzymes to access DNA. Therefore, epimutations can have mutagenic consequences for DNA and pose a substantial threat to genome stability (Besselink et al., 2023; Putiri & Robertson, 2011; Suter et al., 2004). One of the most important epigenetic modifications is DNA methylation, which is essential for mammalian regulation of gene expression (Bird, 1986; Li & Zhang, 2014) but is also associated with CpG depletion via CG > TA mutations (Lindahl, 1996). This thesis aims to

address the mutagenicity of off-target activity of DNA methyltransferases (DNMTs) and how this damage is repaired in order to maintain genome stability.

1.1.2. DNA methylation and epigenetics

1.1.2.1. The basis of DNA methylation

DNA methylation is the product of catalytic activity of DNA methyltransferases (DNMTs), which utilise the co-factor S-adenosylmethionine (SAM) to covalently link a methyl group to a DNA base (Klimasauskas et al., 1994). DNMTs are an ancient superfamily of enzymes found across eukaryotes and prokaryotes, although methylation levels vary considerably between species depending on the types of DNMTs that they retain (Jurkowski & Jeltsch, 2011). The type of DNA methylation is dependent on the specificity of the DNMT catalytic domain, while sequence specificity is determined by the target recognition domain (Lin et al., 2020). In mammals, the most common form of DNA methylation is cytosine-5 methylation (5mC) at CpG dinucleotides. Cytosine-5 DNMTs contain several highly conserved catalytic motifs that have been used to identify evolutionary relationships between cytosine-5 DNMTs across species (Ponger & Li, 2005; Pósfai et al., 1989). However, 5mC can also occur in non-CpG contexts, for example, plants display elevated levels of CHH and CHG methylation (where H = A, C or T) compared to eukaryotes, due to the activity of CMT2 and 3 (Henderson & Jacobsen, 2007). DNA methylation can also occur in non-cytosine contexts. For example, prokaryotes display

increased levels of 6-methyladenine (6mA) compared to eukaryotes (Dunn & Smith, 1958). Here, the focus will predominantly be on eukaryotic cytosine methylation.

The number and function of DNMTs varies across species, even within domains such as eukaryotes. In humans, there are five DNMTS: DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. DNMT1, DNMT3A and DNMT3B are canonical family members that catalyse the addition of a methyl group onto the 5' position of cytosine. DNMT3L does not possess catalytic activity, although it does retain some sequence similarity with the catalytically active enzymes and serves important regulatory and cellular functions (Ooi et al., 2007). DNMT2 retains sequence similarity with canonical DNMTs but has minimal activity on dsDNA. Its primary confirmed substrate is tRNA where it catalyses the methylation of cytosine 38 in $tRNA^{Asp}$ in humans (Goll et al., 2006). This has been shown to stabilise the tRNA under stress and promote protein translation (Jeltsch et al., 2017).

Furthermore, functionality differs within the canonical DNMT group. DNMT1 is a maintenance CpG DNMT, meaning that it preferentially methylates cytosine in a CpG dinucleotide where one cytosine is already methylated (Pradhan et al., 1999). Following DNA replication, the newly synthesised unmethylated DNA strand is annealed to the original, methylated DNA strand, resulting in hemi-methylated DNA. DNMT1 is recruited to the replication fork, where it recognises the hemi-methylated CpG dinucleotide (Leonhardt et al., 1992). DNMT1 then methylates the newly synthesised unmethylated cytosine, resulting in cytosine methylation on both strands. Both strands can then act as templates for future methylation following

DNA replication (Hermann et al., 2004). Therefore, it is the palindromic nature of CpG dinucleotides coupled with DNMT1 activity that is responsible for the persistence of 5mC after DNA replication. This makes it a stable epigenetic modification, the functions of which will be discussed further in 1.1.2.2. DNMT3A/B are *de novo* CpG DNMTs, so can methylate cytosines in non-methylated dsDNA. DNMT3A/B expression is developmentally and proliferatively regulated, with tissue-specific patterns of expression for each methyltransferase (Okano et al., 1999; Tóth et al., 2025; Xie et al., 1999). Knockout (KO) of DNMT1 and DNMT3B in mice are embryonic lethal, while DNMT3A KO leads to smaller mice that survive for a short period after birth (Li et al., 1992). This suggests that DNMT3B is required for embryonic development, whereas DNMT3A is more important for later cellular differentiation.

As well as DNA methylation 'writers' (DNMTs), there are several DNA methylation 'erasers'. Removal of 5mC typically involves addition of a hydroxyl group onto 5mC to form 5-hydroxymethyl-cytosine (5hmC) by ten-eleven translocase (TET) enzymes (Ito et al., 2010; Tahiliani et al., 2009). These belong to the α-ketoglutarate, Fe(II)-dependent dioxygenase family of enzymes (Tahiliani et al., 2009). Processing of 5hmC then proceeds via two pathways: further sequential oxidisation by TET enzymes, or deamination by AID/APOBEC. In the first, TET enzymes oxidise 5hmC to 5-formyl cytosine (5fC) and then 5-carboxy cytosine (5caC) (Ito et al., 2010). In the second, AID/APOBEC deaminate 5hmC to form 5-hydroxymethyl uracil (5hmU) (Guo et al., 2011), although this has been disputed (Nabel et al., 2012; Rangam et al., 2012). The products of both pathways are recognised and removed by base

excision repair (BER) components, specifically thymine DNA glycosylase (TDG), to yield unmodified cytosine (Guo et al., 2011). These will be reviewed in 1.1.4.

Demethylation is particularly important during mammalian embryonic development, where the germ-line genome undergoes widespread demethylation, resulting in a clean slate ready for cell-type specific methylation (Seki et al., 2005). Unlike DNMT KOs, an AID KO study conducted in mice resulted in viable offspring (Cortesao et al., 2013; Muramatsu et al., 2000). This suggests redundancy between demethylation pathways, perhaps indicating a developmentally critical role of DNA demethylation alongside methylation. As well as enzymatic demethylation, the genome can also experience passive demethylation. This occurs when DNMT1 fails to methylate the un-methylated strand in hemi-methylated DNA, leading to dilution over time following DNA replication (Kimura & Shiota, 2003; Li et al., 1992). This hypomethylation has been observed in fast dividing cancers (Wild & Flanagan, 2010). Finally, the combination of active and passive DNA de-methylation further results in dilution of DNA methylation. Here, DNMT1 is unable to recognise TET-oxidised 5mC intermediates in hemi-methylated DNA, therefore the unmethylated strand remains unmethylated and is unable to act as a template for DNA methylation in subsequent rounds of DNA replication (Seiler et al., 2018; Valinluck & Sowers, 2007).

1.1.2.2. The role of DNA methylation in mammalian epigenetics

There are four main mammalian epigenetic pathways: DNA methylation, histone modifications, chromatin re-modellers and non-coding RNAs, although the latter is considered more of a regulatory mechanism depending on the definition of epigenetics being used. Undoubtedly the most well characterised modification is DNA methylation, of which 5mC at CpG dinucleotides is the most predominant. Broadly, cytosine DNA methylation is most often associated with gene silencing, but its role is actually far more complex. DNA methylation regulates gene expression in three ways: first, the presence of DNA methylation changes the structure of the major groove, which affects the binding affinity of proteins such as transcription factors (TFs) to promoters. Approximately 70% of gene promoters are found within CpG islands (Saxonov et al., 2006). These are defined as >200bp regions of the genome that contain above average CpG content (Bird et al., 1985). CpG islands are often associated with constitutively active genes such as housekeeping genes, which are involved in crucial cellular processes such as cell division and DNA damage repair (DDR) (Gardiner-Garden & Frommer, 1987). In order to maintain a transcriptionally active state, CpG islands often contain fewer nucleosomes than other regions of DNA, thereby increasing DNA accessibility for TFs. Any nucleosomes that are present often have histone modifications that increase transcription (Tazi & Bird, 1990). Despite this, during early development CpG island promoters often become methylated, thereby repressing transcription. This also occurs during imprinting (Caspary et al., 1998; Wutz et al., 1997). Evidently, more

research is needed into the complex nature of CpG island methylation and its effects on gene expression.

The second way in which DNA methylation regulates gene expression is by modulating the binding of methylation-specific proteins and transcriptional repressors such as HDACs to both gene bodies and promoters (Nan et al., 1998). A key example of this is the methyl-CpG-binding domain (MBD) family of proteins, which includes MeCP2 and MBD1-4, which bind methylated CpG dinucleotides and block TF access (Lopez-Serra et al., 2006). Mutation of MeCP2 has been implicated in neurological disorders such as Rett syndrome (Adachi et al., 2009). Finally, methylated DNA promotes heterochromatin formation by chromatin re-modellers, therefore limiting the spatial accessibility of TFs and binding proteins (Karymov et al., 2001). The exact manner in which DNA methylation compacts DNA has been disputed. Studies in KO mouse embryonic stem cells (ESCs) indicated that DNA methylation affected modification of core and linker histones, but not bulk chromatin (Gilbert et al., 2007). Interestingly, some studies show preferential binding of TFs to methylated DNA (Hu et al., 2013), while some TFs appear to bind DNA recognition sequences irrespective of methylation status (Buck-Koehntop & Defossez, 2013). The specific effect of DNA methylation is therefore dependent on the genomic location of DNA methylation and the effect of the binding protein. It is unsurprising that changes to DNA methylation patterns have been implicated in the onset of diseases such as cancer. This will be discussed further in 1.1.3.3.

1.1.2.3. Evolutionary variability in DNA methylation

While cytosine DNA methylation is central to mammalian epigenetic regulation (Bird, 1986; Li & Zhang, 2014), the occurrence and extent of DNA methylation varies hugely across species and over evolutionary timescales (Sarkies, 2022). For example, an ancestor of the nematode *C. elegans* lost DNMT1 and DNMT3 around 350 million years ago, and so *C. elegans* now lack DNA methylation entirely (Rošić et al., 2018). However, other nematodes such as *T. spiralis* have retained *de novo* DNA methylation which regulates genes associated with parasitism (Gao et al., 2012). DNA methylation can also vary considerably within phylum. Studies of individual arthropod species revealed a range of DNA methylation levels from 1 to 30% of CpG sites (Bewick et al., 2016; de Mendoza, Pflueger, et al., 2019; Lyko et al., 2010; Wang et al., 2006). Another study identified ancestral gene body and transposable element (TE) DNA methylation in arthropods (Lewis et al., 2020). However, crustaceans have lost gene body methylation while the myriapod *S. maritima* and insect *P. citri* have gained both promoter and TE methylation (de Mendoza, Hatleberg, et al., 2019; de Mendoza, Pflueger, et al., 2019; Lewis et al., 2020). Interestingly, the sea sponge *A. queenslandica* possesses DNA methylation patterns closely resembling that of vertebrates, demonstrating convergent evolution towards a hypermethylated state from a sparsely methylated common ancestor (de Mendoza, Hatleberg, et al., 2019). Finally, notable model organism species such as *D. melanogaster, S. pombe* and *S. cerevisiae* all lack cytosine methyltransferases and cytosine DNA methylation (Jurkowski & Jeltsch, 2011; Ponger & Li, 2005).

This repeated yet independent evolutionary loss of DNMTs begs the question: is there a property of DNA methylation that is responsible for this negative selection? One theory by Zemach and Zilberman suggested that DNA methylation exists as a 'genomic immune system' that protects the genome from instability caused by TE expression. They proposed that the cost of TE expression is greater in organisms that reproduce sexually than asexually, therefore the cost of losing TE silencing by DNA methylation is also greater. This gives asexual organisms time to acquire alternative methods for TE silencing before going extinct, which they say explains the lack of DNA methylation in fungi and green algae. Similarly, this would explain why land plants and vertebrates, that reproduce sexually, rely on TE silencing by DNA methylation. Loss of DNA methylation would lead to dormant TE activation and mutagenesis which would result in extinction (Zemach & Zilberman, 2010). While this is possible, DNA methylation varies enormously between species as described above and cannot be neatly generalised within phyla. More recently, investigation revealed DNMT activity increased levels of the toxic lesion 3mC and therefore has the propensity to damage DNA (Dukatz et al., 2019; Rošić et al., 2018). This may be part of the reason why DNA methylation has been lost in some lineages, which will be discussed further in 1.1.5.

1.1.3. DNA methylation and genome stability

1.1.3.1. Promotion of genome stability by DNA methylation

Despite the presence and extent of DNA methylation varying considerably across species, DNA methylation can serve to stabilise the genome, which could contribute to its retention in some species. Most of the human genome is comprised of non-coding DNA, of which approximately half is repeat elements. These can be further categorised into TEs and tandem repeats (Ahmed & Liang, 2012). TEs can move and sometimes copy themselves at random into the genome (Boeke et al., 1985; Greenblatt & Brink, 1963; McClintock, 1950). This transposition leads to inter- and intrachromosomal rearrangements which threaten genome stability by increasing instances of deletions, duplications and translocations (Collins & Rubin, 1984; Geurts et al., 2006; Page et al., 2004; Yu et al., 2011). DNA methylation promotes genome stability by silencing the promoters of some TEs and repeats (Schulz et al., 2006). For example, in mouse embryos, transcriptional repression of the retrovirus IAP by DNA methylation of the long terminal repeat element is maintained even throughout global hypomethylation during development (Gaudet et al., 2004; Walsh et al., 1998). Furthermore, DNA methylation promotes formation of tightly packaged heterochromatin, which inhibits recombination of TEs and spurious transcription (Zhou & Robertson, 2016). Loss of TE methylation in mice has been shown to inhibit meiosis due to transcriptional activation of previously silenced TEs and subsequent promotion of recombination (Bourc'his & Bestor, 2004). DNA methylation is also central to

maintaining the integrity of peri-centromeric heterochromatin structure, which ensures correct chromosome condensation and segregation (Geiman et al., 2004; Monier et al., 2007). Finally, DNA methylation allows MMR proteins to distinguish between parent and newly synthesised DNA during DNA replication and correct any erroneous nucleotide incorporation (Hare & Taylor, 1985). DNMT1 also associates with PCNA and MLH1 at DNA damage sites to facilitate MMR (Mortusewicz et al., 2005; Ruzov et al., 2009). Indeed, DNMT1 KO in immortalised human fibroblasts saw an increase in mutations the CA_{17} reporter gene and conferred resistance to the chemotherapy drug 6-thioguanine (Loughery et al., 2011).

1.1.3.2. Deamination of 5mC and other contributions to mutagenesis

While DNA methylation is central to epigenetic signalling and contributes to genome stability, it is also well known that it is mutagenic. This was first suggested following the observation that highly methylated regions in mammalian genomes were depleted of CpG sequences (Bird, 1980). Over time, it became clear that methylation increased the chances of mutation at cytosine. Now, we know that the most common form of spontaneous DNA damage is cytosine deamination (Lindahl, 1996). Deamination of cytosine produces uracil, resulting in a UG mismatch. This can become fixed as a CG > TA mutation if unrepaired prior to replication. In mammals, uracil-DNA glycosylase (UNG) recognises uracil as foreign and excises it. This leaves an abasic site that can be repaired by BER to return DNA to its original sequence (see 1.1.4.1.3.) (Visnes et al., 2009). However, methylated cytosines behave slightly differently. Instead of uracil, 5mC deaminates to thymine, resulting

in a TG mismatch. Mammalian cells possess two thymine DNA glycosylases (TDG and MBD4) to directly excise thymine from the DNA backbone (Bellacosa & Drohat, 2015). Interestingly, 5mC deaminates approximately two times faster than unmethylated cytosine (Shen et al., 1994) and the CG > TA transition rate at methylated cytosines is higher than unmethylated (Duncan & Miller, 1980). The increased mutagenicity of 5mC suggests that uracil excision by UNG is more efficient than thymine excision by TDGs; a hypothesis that is supported by biochemical and structural evidence (Maiti et al., 2012; Tarantino et al., 2018). Furthermore, analysis of cancer genome data suggested that the mammalian replicative DNA Polymerase Pol ε is more error-prone when replicating 5mC than unmodified cytosine, therefore contributing to the methylation-dependent CG > TA mutational signature (Tomkova et al., 2018; Tomkova et al., 2024). As a result, this excess mutagenicity is likely a combination of a higher initial rate of deamination and differences in downstream processing of deamination products (De Bont & van Larebeke, 2004). Despite the many essential roles that DNA methylation plays in the cell, there is clearly a trade-off with the threat it poses to genome stability.

1.1.3.3. Contributions of methylation to cancer mutational signatures

Somatic mutations are the basis of all cancers, arising from a combination of cellular processes including infidelity during DNA replication, exposure to exogenous mutagens and defective DNA repair (Peña-Diaz et al., 2012; Pfeifer, 2010; Stratton et al., 2009). Mutational burden reflects the intensity of any endogenous or exogenous mutagen exposure, despite the fact that not all DNA damage will result

in mutation. Sequencing of many cancer samples builds a data base of commonly found mutations specific to different cancer types (Hudson et al., 2010) Crucially, not all mutations drive the progression of cancer and are therefore not under selection. This allows us to observe a spectrum of mutations that reflect the mutational processes occurring in the cell (Stratton et al., 2009). In some cancers, several mutational processes may have contributed to the observed mutations. Through sequencing advances, it has become possible to identify mutations arising from specific mutational processes and therefore extract unique mutational signatures that may be prevalent in multiple different cancers (Alexandrov et al., 2013; Nik-Zainal et al., 2012). In 2013, Alexandrov et al. found 22 distinct mutational signatures in human cancers, each representing a specific mutational process (Figure 1.1) (Alexandrov et al., 2013), although the number of validated mutational signatures has since increased (Alexandrov et al., 2020). Signatures are comprised of the percentage of mutations in each of the six substitution types: C>A, C>G, C>T, T>A, T>C and T>G. They are further defined by the sequence context that the mutation is found in; specifically, the adjacent 3' and 5' bases. This gives 96 possible unique mutation classifications and allows for differentiation between mutations of the same substitution. Cancer mutational signatures 1A and 1B are of particular relevance to DNA methylation. Together, they were observed in 25/30 cancer types, including breast, head and neck, and melanoma (Alexandrov et al., 2013). Signatures 1A and 1B are characterised by an enrichment of C > T mutations at NpCpG trinucleotides, likely reflecting increased spontaneous deamination of 5mC (Pfeifer, 2006). They are also the only validated signatures that are strongly correlated with age. Given that spontaneous deamination of 5mC occurs at a

relatively similar rate in somatic tissues, it follows that these signatures accumulate over time (Alexandrov et al., 2013). This highlights the significant threat to genome stability that is posed by endogenous DNA damage due to DNA methylation.

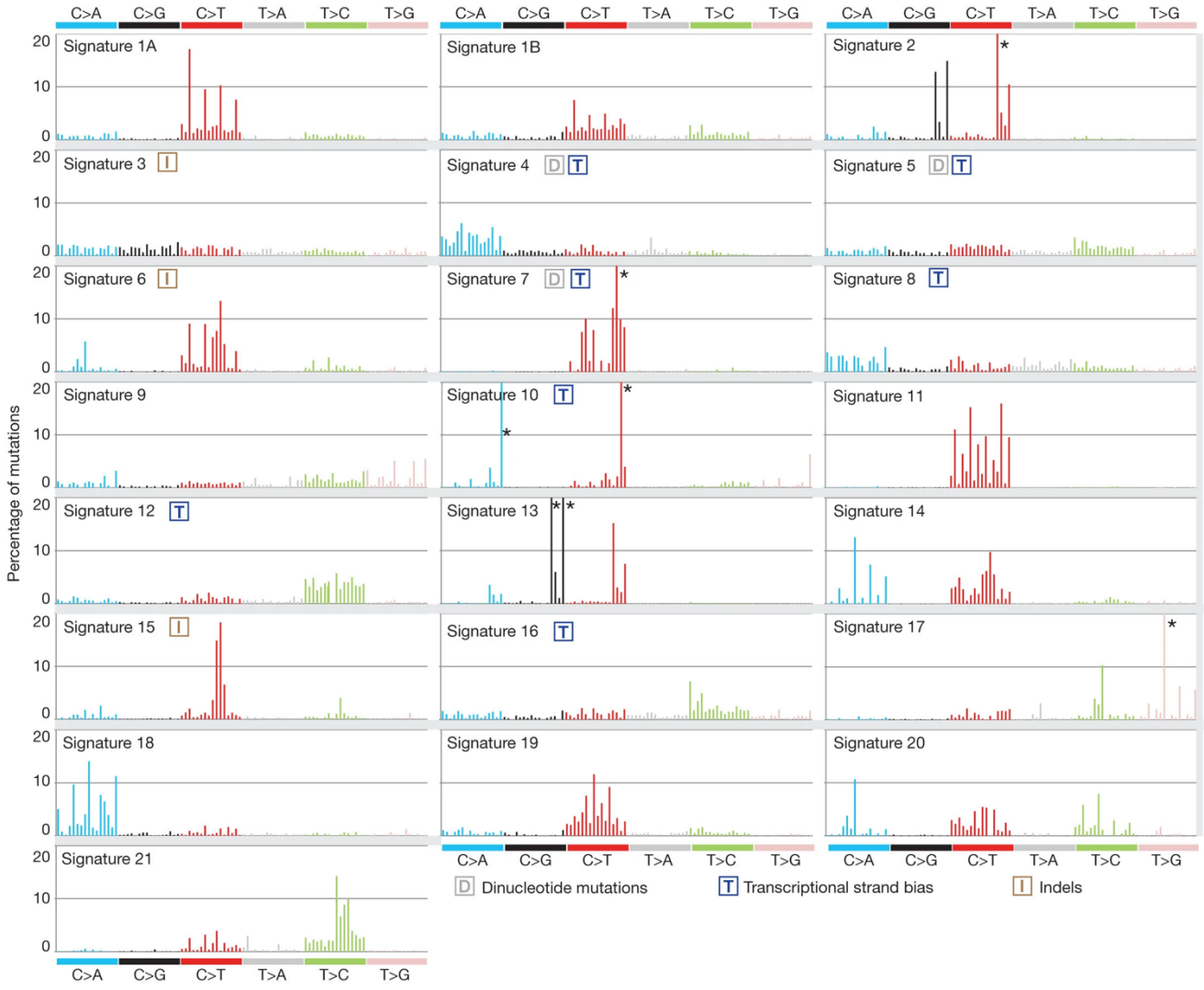


Figure 1.1: The original 22 validated mutational signatures from human cancer. Signatures are displayed according to the 96 substitution classification (x axis), comprised of the substitution type (by colour) and the 3' and 5' bases adjacent to the mutated base. The percentage of mutations attributed to each substitution type is depicted on the y axis with * indicating bars exceeding 20%. Figure taken directly from (Alexandrov et al., 2013). It should be noted that the set of validated mutational signatures has expanded since 2013 (Alexandrov et al., 2020).

1.1.4. Aberrant DNA methylation, damage and repair

1.1.4.1. DNA alkylation lesion toxicity

Despite its prevalence in epigenetics, 5mC is not the only possible base modification. Cytosine can also be methylated on 3' and 4' carbons, resulting in 3-methyl cytosine (3mC) and 4-methyl cytosine (4mC) respectively (Janulaitis et al., 1983; Singer & Kuśmierek, 1982). Methyl groups can also be added onto other organic bases, forming products such as 1-methyl adenine (1mA), 3-methyl adenine (3mA) and 7-methyl guanine (7mG) depending on the nucleophilic substitution mode of action (Sedgwick, 2004). Such methylation is often the product of alkylating agents rather than DNMTs, although there is evidence to suggest that some cytosine DNMTs possess off target activity and can introduce 3mC as well as 5mC (Rošić et al., 2018). This will be discussed further in 1.1.5. However, not all modified bases are conducive to bypass by replicative DNA polymerases such as Pol III in bacteria and Pol δ/ε in eukaryotes. In order to achieve high fidelity of replication, the active site of the replicative polymerase is highly specific and cannot tolerate lesions that change the conformation of the DNA double helix. Therefore, the passage of DNA Polymerase is blocked by alkylation lesions such as 3mC and 1mA that flip the base out of the helix and disrupt canonical Watson-Crick base pairing (Boiteux & Laval, 1982; Larson et al., 1985). Replication fork stalling can lead to collapse and DSBs if not resolved, which are potentially mutagenic or cytotoxic (Sedletska et al., 2013). Error-prone bypass of lesions may also be mutagenic (see 1.1.4.4). As a result, alkylation lesions pose a threat to genome stability.

Unsurprisingly, all organisms possess DDR pathways to deal with alkylation damage. Indeed, some lineages display co-evolution between DNA methyltransferases and specific repair pathways that target methylation lesions (Rošić et al., 2018). These will be discussed more below.

1.1.4.2. Direct repair by ALKB2/3

The most straightforward way of repairing alkylation lesions is via direct reversal. There are two types of repair systems capable of direct reversal in mammals: MGMT repair and ALKB repair. O^6-methylguanine methyltransferase (MGMT) was first discovered in *E. coli,* where it was found to primarily return O6-methylguanine (6mG) to guanine (Olsson & Lindahl, 1980; Samson & Cairns, 1977). Similarly, AlkB was first discovered in *E. coli* in a screen for methyl methanesulfonate (MMS) sensitivity (Kataoka et al., 1983). ALKB2/3 were then identified as human homologues of AlkB (Duncan et al., 2002). ALKB2/3 and AlkB are members of the α-ketoglutarate (αKG), Fe(II)-dependent family that directly reverse alkylation lesions such as 1mA, 3mC and 1-methylguanine (1mG), as well as other ethano-adducts, via oxidative demethylation (Aas et al., 2003; Aravind & Koonin, 2001). ALKB2/3 and AlkB use Fe(II) to bind molecular oxygen and generate a reactive Fe(IV)-oxo intermediate. Simultaneously, αKG undergoes oxidative decarboxylation which releases succinate and CO_2. This allows the reactive Fe(IV)-oxo intermediate to hydroxylate the alkyl group, which is then released as formaldehyde, thereby returning cytosine to its original, unmodified form (Duncan et al., 2002; Trewick et al., 2002) (Figure 1.2).

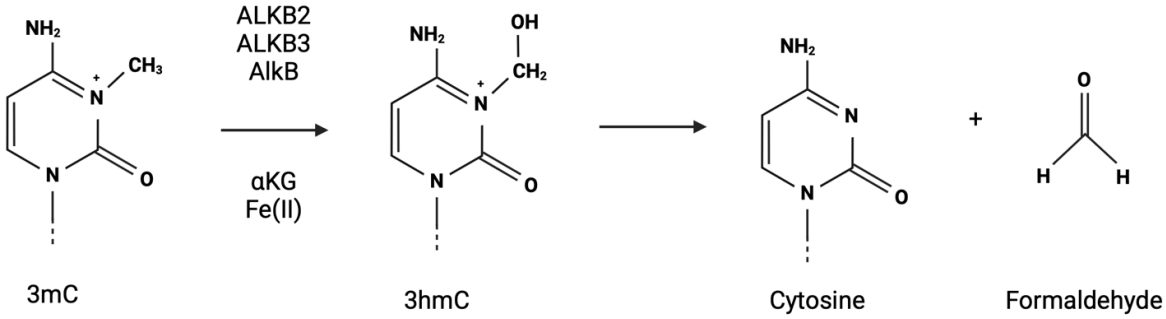


Figure 1.2: Schematic mechanism for direct repair of alkylation lesion 3mC by ALKB2/3 and AlkB. Oxidative demethylation by ALKB2/3 and AlkB using O_2, αKG and Fe(II) yields original unmodified cytosine. Created in Biorender.

In mammals, there are nine ALKB homologues, ALKB1-8 and FTO. Of the nine homologues, only ALKB2/3 are capable of directly reversing some alkyl adducts back to unmodified cytosine (Aas et al., 2003; Duncan et al., 2002). ALKB1 displays mitochondrial DNA and histone demethylase activity (Westbye et al., 2008). The others are involved in protein and RNA de-methylation. For example, ALKB5 and FTO primarily target 6mA in RNA (Jia et al., 2011; Xu et al., 2014), while ALKB4 demethylates actin filaments (Li et al., 2013). The ALKB family is highly conserved across all metazoans, with the exception of ALKB5 and FTO which evolved more recently and are only found in vertebrates (Robbens et al., 2008).

Prokaryotic AlkB is under the control of the *ada* operon. The *ada* operon is named due to the adaptive response that it induces when the cell experiences alkylation damage, conferring future resistance to higher doses of damage. The operon consists of *ada, alkA, alkB* and *aidB* (Samson & Cairns, 1977). Elevated levels of alkylation damage trigger increased production of Ada, which serves as a broad

inducer for the entire operon (Landini & Volkert, 1995; Lemotte & Walker, 1985). AlkB is capable of direct reversal of 3mC and 1mA via the same mechanism as ALKB2/3 (Falnes et al., 2002; Trewick et al., 2002) (Figure 1.3). AlkA is a DNA glycosylase that targets 3mA. Repair proceeds via hydrolysis of the 3mA glycosidic bond and repair of the resulting apurinic (AP) site by BER (see 1.1.4.1.3.). The constitutively expressed DNA glycosylase Tag also works alongside AlkA to remove 3mA. AlkA has broader substrate specificity than Tag and can remove additional lesions such as 3mG (O'Brien & Ellenberger, 2004). It is also 10-20 times more efficient at removal of 3mA in single stranded DNA (ssDNA) (Bjelland & Seeberg, 1996).

Efficiency of alkylation lesion repair may be affected by whether the lesion arises on double or single stranded DNA. Stalling of the replisome at dsDNA lesions could lead to re-priming of the replicative polymerase downstream, leaving behind a ssDNA gap at the lesion site (Heller & Marians, 2006). Alternatively, the replication fork may reverse to await dsDNA repair, or the replicative polymerase may be switched out for a back-up polymerase capable of lesion bypass (see 1.1.4.1.4.). The methyl groups in 3mC and 1mA are positioned on the nitrogen rings of cytosine and adenine respectively. Therefore, in dsDNA, these are better protected from the action of enzymes such as AlkB. However, these lesions are exposed in ssDNA, and it has been postulated that AlkB activity is higher in ssDNA as a result (Trewick et al., 2002). Indeed, prokaryotic AlkB and mammalian ALKB3 typically favour ssDNA or RNA. ALKB2 interacts with both strands in dsDNA by flipping the alkylated base out of the DNA helix. It then supports the undamaged bases by inserting a β-hairpin loop into the resulting gap (Yang et al., 2008). ALKB3 lacks this loop and so has weaker

base flipping ability, although this makes it better suited to ssDNA substrates (Dango et al., 2011). AlkB is functionally more similar to ALKB3, but does still possess base flipping capacity and activity on dsDNA (Yang et al., 2008; Yu et al., 2006). There is also evidence to suggest that AlkB has higher affinity for 1mA than 3mC due to its larger size, although it achieves equal repair capacity by altering enzyme kinetics to ensure increased 3mC turnover rate (Yu & Hunt, 2009).

1.1.4.3. Base excision repair

In 1974, Lindahl proposed that cells must be able to remove uracil from DNA, arising either from erroneous incorporation of deoxyuridine monophosphate (dUMP) during replication or via cytosine deamination. He then went on to isolate UNG from *E. coli* which cleaves the glycosidic bond between uracil and deoxyribose (Lindahl, 1974). This was the first key discovery that unearthed the now extremely well characterised BER pathway. BER is a highly conserved DDR pathway that is found across eukaryotes, prokaryotes and archaea. It is a broad repair pathway designed to repair DNA damage arising from processes such as alkylation, oxidation and deamination (Lindahl, 1993). These forms of DNA damage can disrupt canonical base pairing and introduce mutations if not repaired prior to DNA replication (Lindahl & Wood, 1999).

Core BER relies on the action of four enzymes: a DNA glycosylase, an AP endonuclease, a DNA Polymerase and a DNA Ligase (Figure 1.3). The DNA glycosylase is central to BER repair and is specific to the type of DNA damage and the base on which it occurs (Mitra, 2001). Following damage recognition, the DNA

glycosylase cleaves the N-glycosidic bond that attaches the damaged base to the sugar-phosphate backbone. This leaves an apurinic or apyrimidinic (AP) site, collectively known as an abasic site (Lindahl, 1974). DNA glycosylases are either monofunctional (possess only glycosylase activity) or bifunctional (possess both glycosylase and AP lyase activity). Following abasic site formation by monofunctional DNA glycosylases, the DNA backbone is then cleaved and processed by an AP endonuclease, resulting in a DNA nick located 5' to the original site of damage (O'Connor & Laval, 1989; Robson & Hickson, 1991). These include APE1 in mammals (Kane & Linn, 1981) and EndoIV in bacteria (Dianov & Lindahl, 1994). This nick has both a 3' hydroxyl group and a 5' phosphate group, so is ready for gap filling by DNA Polymerase (O'Connor & Laval, 1989; Robson & Hickson, 1991). In the case of bifunctional DNA glycosylases, abasic site formation is followed by AP lyase activity (Boiteux et al., 1987). This occurs to the 3' side of the abasic site and results in blocked 3' ends that are not suitable for DNA Polymerase activity. Therefore, bifunctional DNA glycosylase and AP lyase activity is always followed by further processing by endonucleases and/or polynucleotide kinase-phosphatase to create a 3' hydroxyl group (Bernstein et al., 2005; Suh et al., 1997; Wilson, 2007). Finally, after DNA Polymerase inserts a new nucleotide according to Watson-Crick base pairing, DNA Ligase seals the nick in the DNA backbone (Kim & Wilson, 2012; Tomkinson et al., 2006). Additional proteins may be involved depending on the organism, for example mammalian BER requires the scaffold protein XRCC1 to bring together core BER proteins such as Pol β and DNA Ligase III (Caldecott et al., 1995; Cappelli et al., 1997).

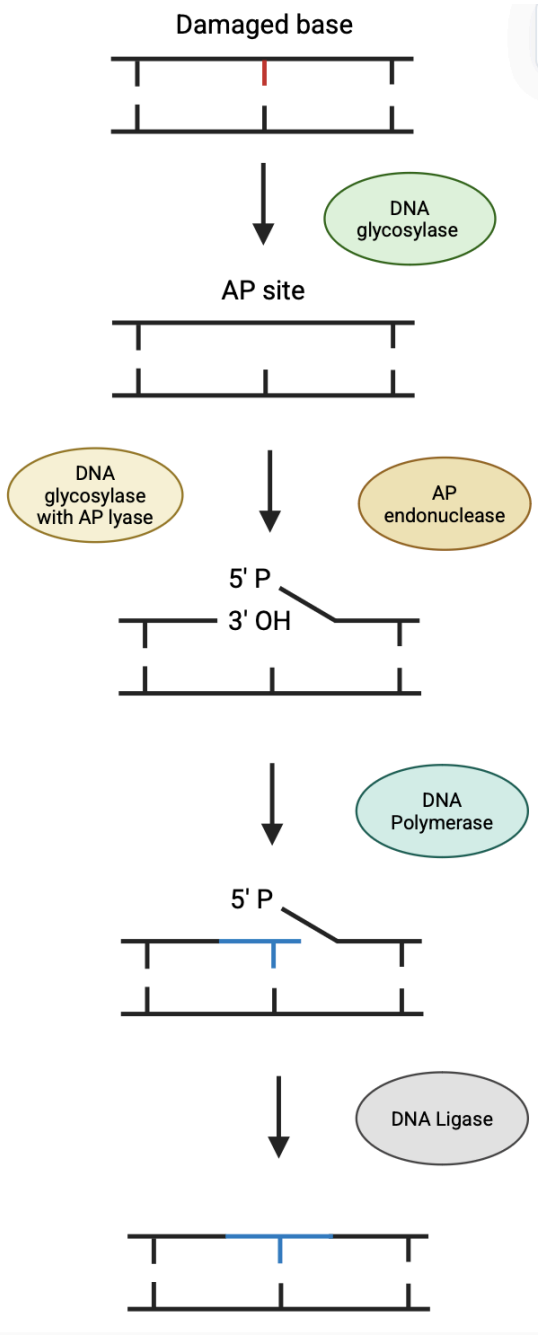


Figure 1.3: Schematic overview of short patch BER. A DNA glycosylase cleaves the glycosidic bond holding the damaged base to the sugar phosphate backbone. This leaves an abasic site. Depending on the functionality of the DNA glycosylase, the DNA backbone is either nicked by an AP endonuclease ready for gap filling or processed by AP lyase and additional enzymes (not shown). DNA Polymerase inserts one nucleotide for short-patch BER and the backbone is sealed by DNA Ligase. Created in Biorender.

BER can be further divided into two categories: short-patch and long-patch repair. Short-patch repair occurs when a single nucleotide is removed and is favoured for simple, single base lesions. Long-patch repair involves the removal of 2-10 nucleotides and is typically used for larger, more complex lesions and blocked 3' or 5' ends that interfere with short-patch BER (Fortini & Dogliotti, 2007; Frosina et al., 1996). Therefore, methylation lesions are typically repaired by short-patch BER given that they affect just one base and induce minimal helix distortion (Sedgwick, 2004).

There are many alkylation lesion-specific glycosylases, including monofunctional AlkA in *E. coli* and structurally unrelated N-methylpurine DNA glycosylase (MPG) in mammals, which excise lesions such as 3mA and 7mG (Denver et al., 2003; O'Connor & Laval, 1990). Another important DNA glycosylase is mammalian 8-oxoguanine DNA glycosylase 1 (OGG1), which was originally discovered in yeast (Lu et al., 1997; Nash et al., 1996). Bacteria do not possess OGG1; instead, they have the functional analogue Fpg (Morland et al., 2002). These bifunctional DNA glycosylases repair 8-oxoguanine (8oxoG), a prevalent form of oxidative damage caused by reactive oxygen species (ROS) (Bruskov et al., 2002). Fpg shares sequence similarity with mammalian DNA glycosylases NEIL1 and 2, which repair oxidised pyrimidines as well as 8-oxoG (Dou et al., 2003; Hazra et al., 2002). There are often many different glycosylases involved in the repair of DNA damage, and indeed many different repair pathways that can act on the same lesions (Jacobs & Schär, 2012). DDR pathways such as BER and nucleotide excision repair (NER) are often favoured initially as they are quick (milliseconds to seconds per lesion) and do

not inhibit the cell cycle, unless the cell is experiencing high levels of oxidative stress or alkylation damage (Horton et al., 2007; Sobol et al., 1996). More complex lesions or downstream products of unrepaired DNA damage such as DSBs require repair by pathways such as homologous recombination (HR), non-homologous end joining or transcription coupled repair. It is not relevant to this thesis to introduce these pathways in detail, but they have been comprehensively reviewed elsewhere (Burma et al., 2006; Hanawalt & Spivak, 2008; Jasin & Rothstein, 2013; Li & Heyer, 2008; Mellon, 2005; San Filippo et al., 2008; Selby et al., 2023; van Gent & van der Burg, 2007; Weterings & Chen, 2008).

1.1.4.4. Bypass of alkylation damage by translesion synthesis

Sometimes it is not possible for cells to directly reverse or excise alkylation lesions. Similarly, DNA damage levels may be too high for all adducts to be repaired error-free. Instead, the cell prioritises bypass of the lesion and resumption of replication fork passage. While this does not result in lesion repair, it is more favourable for DNA replication to resume and the lesion to persist than for the cell to experience replication fork collapse and potential mutagenic DSB formation (Walker, 1984) . There are several lesion bypass mechanisms in mammalian cells, primarily HR, template switching and translesion synthesis (TLS) (Lehmann, 2006; Li & Heyer, 2008; Lovett, 2017; Patel et al., 2010; San Filippo et al., 2008). Mammalian TLS is mediated by four Y-family polymerases: Pol κ, Pol η, Pol ι and REV1, while bacterial TLS is mediated by Pol IV and V (Goodman, 2002; Masutani et al., 1999). These polymerases have larger active sites that are able to encompass helix-distorting

DNA lesions such as 3mC (Furrer & van Loon, 2014). However, these more permissive active sites, combined with no proof-reading activity, result in far lower DNA replication fidelity (McCulloch & Kunkel, 2008). As a result, they are highly likely to insert an erroneous base opposite the damaged base, which becomes fixed as a mutation following DNA replication. Template switching and HR are less mutagenic, although can cause genetic rearrangements and crossover products if acting in repetitive regions (Bishop & Schiestl, 2000; Kishkevich et al., 2022).

In bacteria, TLS is part of the SOS response (Figure 1.4), a transcriptional response that inhibits cell division and upregulates specific DNA repair factors. Exposure to high levels DNA damage, for example by UV radiation, results in the accumulation of ssDNA (Witkin, 1976). This induces overexpression of the DNA-dependent ATPase RecA, which coats ssDNA as a nucleofilament. RecA nucleofilament formation induces auto-proteolysis of the LexA repressor protein (Little, 1984; Schlacher et al., 2006). LexA is normally bound to SOS boxes located around the promoters of target genes. Auto-proteolysis of LexA derepresses these target genes, resulting in induction of gene expression of over 40 SOS response genes, including those involved in error-free repair via BER and NER, as well as error-prone repair by TLS (Goodman, 2002). Essential upregulated TLS genes include *umuC* and *umuD*, which encode Pol V, *dinB*, which encodes PolIV, and *sulA*, which inhibits cell division during lesion bypass (Courcelle et al., 2001; Fernández De Henestrosa et al., 2000; Sommer et al., 1998; Tang et al., 1998).

In addition to formation of the nucleofilament that triggers cleavage of LexA and induces upregulation of otherwise lowly expressed SOS response genes, RecA triggers the cleavage of UmuD to UmuD' (Schlacher et al., 2006; Shinagawa et al., 1988). This is essential for assembly into the PolV complex (UmuD'$_2$C). RecA must also leave the nucleofilament and bind to PolV along with ATP, forming UmuD'$_2$C-RecA-ATP. This activates the complex and promotes bypass (Jiang et al., 2009; Reuven et al., 1999) (Figure 1.4). Crucially, it is the 3' RecA-ATP that is necessary for activation, implying a conformational dependence (Jiang et al., 2009). Additionally, ATP hydrolysis leads to filament disassembly (Bork et al., 2001). In this way, RecA is essential for promoting error-prone bypass of DNA lesions.

Obviously, promoting error-prone bypass also poses a substantial threat to genome stability. As a result, there are several processes that work to curb the amount of bypass that occurs. For example, continuous proteolysis of UmuC/D molecules by Lon protease keeps levels low even after SOS induction (Frank et al., 1996). Additionally, each active PolV complex can only catalyse one round of bypass. It then dissociates from the DNA and is deactivated by loss of RecA. It is only by association with new 3' RecA-ATP that the PolV complex is re-activated and can resume bypass of another lesion (Jiang et al., 2009). Loss of the RecA nucleofilament also reduces the SOS response signal, leading to reduced proteolysis of LexA and increased LexA SOS box binding. Finally, the β clamp is required for PolV function at the replication fork. Pol III competes very effectively for the clamp and quickly replaces PolV, resuming normal high fidelity DNA replication (Patel et al., 2010). Overall, these regulatory mechanisms allow cells to use TLS as

an extreme measure under alkylation stress only when no other bypass or repair mechanisms are available, while minimising excess mutagenesis.

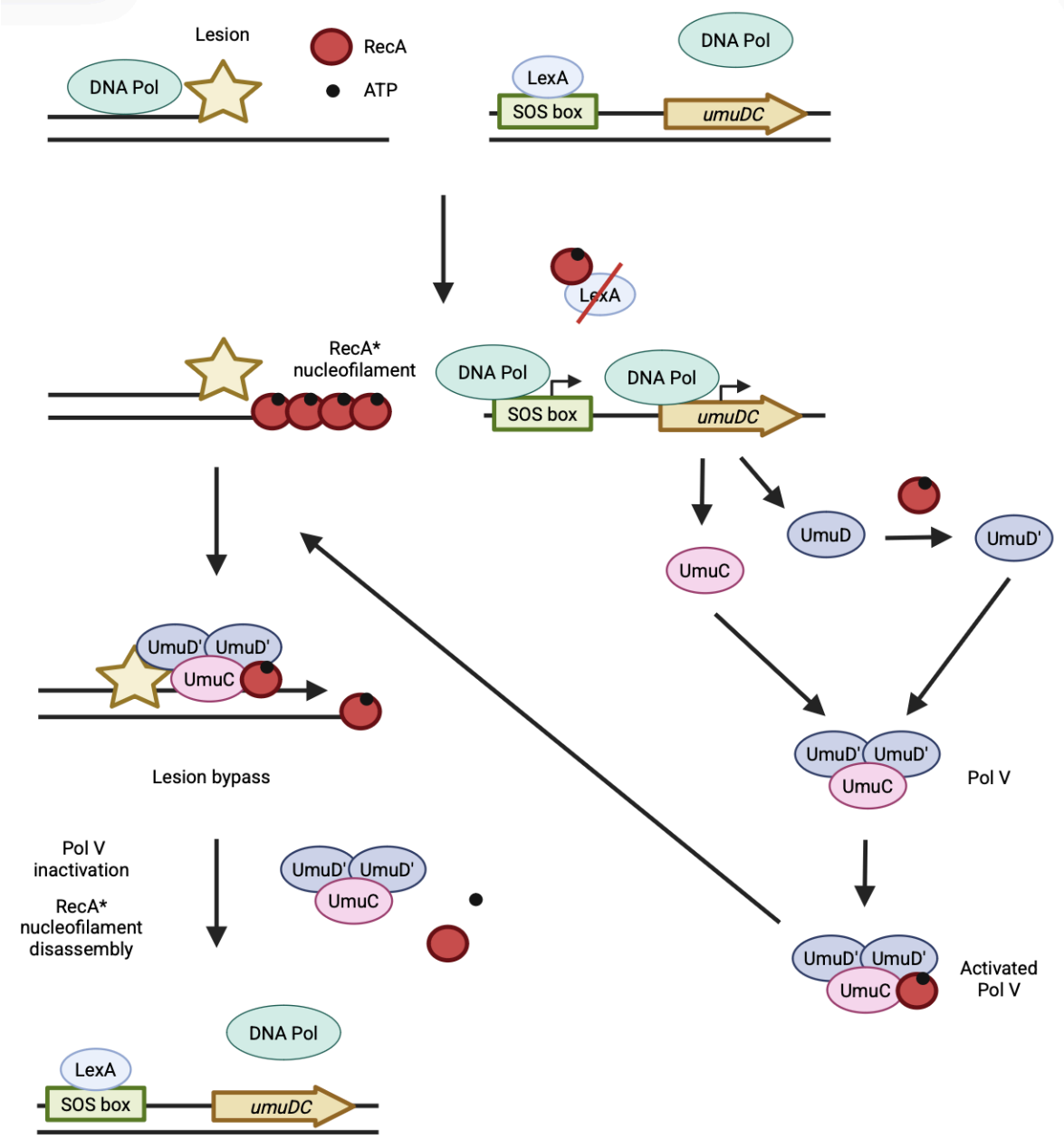


Figure 1.4: Schematic of SOS response and TLS induction in *E. coli*. DNA Pol stalling leads to ssDNA accumulation and RecA* nucleofilament assembly. This triggers LexA autocleavage and UmuC/D expression. Formation of activated PolV allows for lesion bypass, followed by Pol V inactivation and RecA* nucleofilament disassembly which re-instates SOS box and TLS gene repression by LexA. Created in Biorender.

1.1.5. How DNA damage might influence the loss of DNA methylation

The propensity of DNMTs to damage DNA via introduction of toxic 3mC could partly explain the widespread, independent evolutionary loss of DNA methylation across species. Interestingly, researchers found that the mammalian DDR genes ALKB2/3 had co-evolved with DNMTs in many lineages (Bewick et al., 2019; Lewis et al., 2020), as well as broader DDR genes involved in pathways such as BER. This suggested that DNMTs were associated with the introduction of DNA damage (Rošić et al., 2018). Indeed, eukaryotic loss of ALKB2 has been shown to correlate with a loss of CpG methylation (Lewis et al., 2020). As previously mentioned, DNMT activity was found to increase levels of toxic 3mC which was exacerbated by KO of ALKB2/3 in mice (Dukatz et al., 2019; Rošić et al., 2018). Together, this suggested that organisms possessing DNA methylation required ALKB2/3 repair activity in order to preserve genome integrity. While 3mC is introduced by DNMTs at levels less than 0.05% of 5mC, it equals those of apurinic (AP) sites (Chastain et al., 2010; Rošić et al., 2018). AP sites have been shown to contribute to changes in nucleotide levels over evolutionary time (Alexandrov et al., 2015). Therefore, DNMT-induced 3mC represents an abundant form of endogenous DNA damage that poses a threat to genome stability. As a result, it is feasible that the introduction of 3mC by DNMTs may have contributed to the evolutionary loss of DNMTs in some lineages.

1.2. Aims and objectives

This research aims to investigate the potential mutagenic effects of DNA lesions arising from off-target activity of DNMTs. It is known that alkylation adducts such as 1mA and 3mC block the progression of replicative polymerases (Boiteux et al., 1987; Larson et al., 1985) and are repaired by pathways such as BER or by direct reversal by ALKB2/3 (Aas et al., 2003; Duncan et al., 2002; O'Connor & Laval, 1990). This suggests that if left unrepaired, these lesions would be detrimental to genome stability. Furthermore, previous work identified an evolutionary link between cytosine DNMTs and ALKB2/3 (Rošić et al., 2018). This led to the hypothesis that DNMTs introduced lesions, specifically 3mC, that are substrates for repair by ALKB2/3. It was confirmed by mass spectrometry that DNMT activity increased 3mC levels *in vitro* and *in vivo* (Dukatz et al., 2019; Rošić et al., 2018). Therefore, it was hypothesised that the propensity of DNMTs to introduce toxic 3mC might be partly responsible for the widespread evolutionary loss of DNA methylation in many lineages (Sarkies, 2022). However, not all DNA damage results in mutagenesis. This is primarily due to error-free repair prior to replication (Alberts, 2002). Therefore, the overarching aim of this research was to investigate the effect of CpG DNMT activity on mutation frequency and mutation spectra in WT and repair-deficient backgrounds. The objectives of this work were as follows:

1) To determine the effects of different DNA alkylation lesions on genome stability.

Rationale:

Before investigating any mutagenic effects of alkylation lesions, I must first address whether they inhibit DNA replication and therefore impact genome stability in the relevant *E. coli* strains. If not, then I would have to question whether it would be worthwhile continuing with investigation into CpG methyltransferase-dependent mutagenesis. I hypothesised that lesions such as 3mC and 1mA would have a detrimental effect on DNA replication in Δ*alkB* backgrounds, suggesting that these lesions could be mutagenic.

2) To determine the effect of CpG methylation on mutagenesis in WT and Δ*alkB* *E. coli*, both quantitatively via rifampicin-resistant mutation frequency and qualitatively via *rpoB*-specific mutation spectrum.

Rationale:

Studying the mutagenic effects of mammalian CpG methylation can be challenging due to the essential but confounding role that it plays in epigenetic regulation of gene expression, as well as other roles in the cell. Therefore, I modelled mammalian CpG methylation by transforming *E. coli* lacking CpG methylation with a CpG DNMT. This would allow for the direct identification of any changes in mutagenesis due to CpG DNMT activity and AlkB repair deficiency. Given that *E. coli* does not rely on cytosine methylation for regulation of gene expression (Casadesús & Low, 2006), I could be confident that changes to mutagenesis were not due to secondary effects

arising from changes to epigenetic regulation. Identification of any DNMT-dependent mutagenesis might provide further insight into the burden that DNMT-induced damage places on the cell. This relates back to the original hypothesis; that such DNA damage and mutagenesis might have contributed to evolutionary loss of DNMTs.

3) To further investigate the relationship between CpG methylation and other forms of DNA damage using an inducible CpG DNMT.

Rationale:

An unexpected finding in the lab during the course of this work was that DNMTs induce sensitivity to ROS in *E. coli* (Krwawicz et al., 2024). A key question arising from this was whether DNMT activity alone increased ROS levels. I aimed to explore this using an inducible CpG DNMT system, and to further investigate how the DNMT was contributing to increased ROS levels. This objective presented another, unexpected, way in which DNMT activity might contribute to genomic instability and therefore evolutionary loss of DNMTs.

Chapter 2 - Materials and methods

2.1. *In vitro* work

2.1.1. Oligonucleotide design

A series of short DNA oligos were designed to investigate the effects of specific DNA lesions on DNA replication and genomic stability. Each forward oligo was annealed to an undamaged reverse oligo by heating at 95°C for ten minutes, then being allowed to cool. The exceptions to this were Oligo_8_abasic_KH which was annealed to Oligo_8_abasic_KH_R, which contained an abasic site at the same position, and Oligo_6_1mA_KH which was annealed to Oligo_6_1mA_KH_R. The 5' ends of the dsDNA oligos were then phosphorylated by T4 Polynucleotide Kinase (NEB, Cat. No. M0201S) according to their protocol. Phosphorylated oligos were purified using the Zymo Oligo Clean and Concentrator Kit (Cat. No. D4060). pBR322 (NEB, Cat. No. N3033) was digested with Fast Digest BamHI (ThermoFisher Scientific, Cat. No. FD005) and Fast Digest HindIII (ThermoFisher Scientific, Cat. No. FD050), according to their protocol. This produced complementary sticky ends to the annealed insert. Each dsDNA oligo was then ligated into pBR322 (see 1.2.) for transformation into *E. coli* as detailed in 2.3. Oligo sequences in Table 1 as follows:

Table 1: Oligonucleotides used in transformation assay.

Name	Sequence (5' to 3')	Modification
Oligo_3_undamaged	AAT TCA TGC AAG TCG CGA AGC TTA	Undamaged (forward)
Oligo_4_5mC	AAT TCA TGC AAG T/5mC/G CGA AGC TTA	5mC
Oligo_5_3mC	AAT TCA TGC AAG TCG /3mC/GA AGC TTA	3mC
Oligo_6_1mA	AAT TCA TGC /1mA/GT CGC GAA GCT TA	1mA
Oligo_8_abasic	AAT TCA TGC AAG TCG /Abasic/GA AGC TTA	Abasic site (forward)
Oligo_9_reverse	GAT CTA AGC TTC GCG ACT TGC ATG	Undamaged (reverse)
Oligo_10_1mA_reverse	GAT CTA AGC TC GCG ACT GCA TG	Undamaged (1mA-specific reverse)
Oligo_11_abasic_reverse	GAT CTA AGC TC /idSp/CGA CT GCA TG	Abasic (reverse)

2.1.2. Ligation

This project required the ligation of dsDNA oligos or PCR-amplified fragments into various plasmids. For ligation of above oligos (2.1.1. Table 1) into pBR322, plasmid DNA was simultaneously linearised and dephosphorylated using FastDigest BamHI (ThermoScientific, Cat. No. FD0054), EcoRI (ThermoScientific, Cat. No. FD0274) and FastAP Thermosensitive Alkaline Phosphatase (ThermoScientific, Cat. No. EF0654) according to their protocol. All ligation was performed using T4 DNA Ligase (NEB, Cat. No. M0202) according to their protocol, with additional 10mM ATP supplemented. 20-100ng vector DNA was used per ligation depending on the experiment requirements, with a 3:1 molar ratio of insert:vector unless otherwise specified. Ligation was allowed to proceed overnight at 16°C. In some cases, a ligation titration was performed. Here, ligations of various molar ratios of insert:vector were carried out using 20ng vector DNA.

2.1.3. ROS detection in a cell-free system

In order to investigate whether the activity of the CpG methyltransferase M.SssI was able to directly introduce ROS, a cell-free reaction was set up. Variable amounts of DNA were added depending on the experiment. DNA was either pBR322 plasmid DNA (NEB Cat. No. N3033) or dsDNA oligos as found in Materials and Methods Table 1. Forward oligo 3 (undamaged) or oligo 5 containing a single site of 5mC were annealed to oligo 9 (undamaged reverse) as in 1.1. Reactions were supplemented with 5µl 1600uM SAM, 1mM H_2O_2 and/or variable volumes of M.SssI (NEB, Cat. No. M0226) depending on the specific experiment. All reactions contained 5µl 10X supplied buffer and were made up to 50µl with dH_2O. Reactions were pipette mixed and incubated at 37°C for 1 hour. Then, 1µl of 50X ROS reagent (Total Reactive Oxygen Species (ROS) Assay Kit 520nm, Cat. No. 88-5930; ThermoScientific) was added to 50µl reaction in a 96-well plate. The supplied ROS reagent contained the widely used fluorescent probe DCFH-DA which is oxidised by a range of ROS to DCF which emits bright green fluorescence (Afri et al., 2004). The plate reader was heated to 37°C, then programmed to measure fluorescence by excitation at 475nm and emission at 500-550nm. Then, to shake for ten minutes, 200 cycles/min (orbital, 2mm shaking diameter). This loop was repeated 100 times.

2.2. *In vivo* work

2.2.1. Bacterial strains

This study used the *E. coli* strains C2523 (New England Biolabs, Cat. No. C2523, genotype *fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10*) and Top10 (ThermoFisher Scientific, Cat. No. C4040, genotype *F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(StrR) endA1 nupG*). Both are *mcrA* deficient and therefore do not cleave methylated DNA. The study required a WT and Δ*alkB* version of each strain. Previously, P1 phage transduction was used to knock out *alkB* in C2523 and replace it with a kanamycin resistant cassette (Baba et al., 2006). In Top10, *alkB* was knocked out via lambda red recombination (see 2.8.), using a kanamycin resistance cassette from a donor plasmid, with flanking homology sequences to *alkB* (Datsenko & Wanner, 2000). Successful KO of *alkB* was confirmed via PCR and selection on LB High Salt Agar + 50μg/ml kanamycin.

2.2.2. Plasmids

In order to introduce CpG methylation into *E. coli,* the bacterial CpG methyltransferase M.SssI from *Spiroplasma sp.* was cloned into pAIT2. This is a low copy number plasmid with both kanamycin and chloramphenicol resistance cassettes. In order to create a non-functional truncated control, the first 570bp of

M.SssI were removed, which include the SAM binding domain, and cloned into pAIT2. Both the full length and truncated versions of M.SssI were under the control of a constitutive promoter. The *Mycoplasma penetrans* CpG methyltransferase, M.MpeI, was a gift from Matthias Bochtler (IIMCB, Warsaw, Poland). Briefly, the methyltransferase gene was cloned into pET28a (KanR) and differs from UNIPROT protein (accession no. Q8EVR5) by Q68R, K71R and S295P. These do not affect activity (Wojciechowski et al., 2013). pM.MpeI was transformed into *alkB* KO C2523 not *alkB*::Kan C2523 to enable selection for M.MpeI activity on kanamycin media.

To allow for more precise control over induction of methylation, the full-length M.SssI was cloned into a pBAD plasmid with ampicillin resistance and the arabinose-inducible P_{BAD} promoter. This promoter is repressed in the presence of glucose, but activated upon arabinose addition, allowing for the induction of methyltransferase expression (Guzman et al., 1995). The inducible system was not functional in C2523, so was used solely in Top10 *E. coli*. Presence of methylation was determined by restriction digest with HpaII (NEB, Cat. No. R0171), which can only cleave unmethylated CCGG, and agarose gel electrophoresis.

Top10 have a point mutation in *recA* (see genotype in 1.1). In order to introduce RecA into Top10, the high copy number plasmid pMP1 was used. Here, *recA* was under the control of the arabinose-inducible P_{BAD} promoter, with chloramphenicol resistance. pMP1 was a gift from David Liu (Addgene plasmid #69627; http://n2t.net/addgene:69627 ; RRID: Addgene_69627). To change the resistance

cassette in pMP1 from chloramphenicol to ampicillin, the plasmid was linearised with PasI (ThermoScientific, Cat. No. ER1861) in supplied buffer under suggested conditions. Linearised backbone was gel-excised with the Zymoclean Gel DNA Recovery Kit (Cat. No. D4001), dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (ThermoScientific, Cat. No. EF0654) and purified by Zymo Clean & Concentrator-5 (Cat. No. D4004) according to their respective protocols. AmpR was amplified out of pBR322 using primers containing PasI restriction sites. PCR conditions were as follows: 1µl pBR322 DNA, 12.5µl 2x MasterMix (MedChemExpress, Cat. No. HY-K0531), 1.25µl each of 10µM AmpR_PasI_F_6: 5' GCGCCCCTGGGAGCTCTTGATCCGGCAAACA 3' and AmpR_PasI_R_6: 5' GCGCCCCTGGGACGAAAGGGCCTCGTGATAC 3', 9µl dH_2O. Cycles: 98°C 30 sec, then 25 x 98°C 10 sec, 66°C 30 sec, 72°C 40 sec, then 72°C for 2 minutes and 5°C for 10 minutes. PCR-amplified AmpR was purified using Invitrogen PureLink PCR Purification Kit (Cat. No. K310001), then digested with PasI as above to yield PasI sticky ends and purified using Zymo Clean & Concentrator-5 (Cat. No. D4004). The purified AmpR fragment was ligated into pMP1 as in 2.1.2. in a 3:1 molar ratio and transformed into Top10 as in 2.2.3.

In order to inhibit the activity of RecA in C2523, a plasmid encoding *psiB* under the control of an IPTG-inducible P_{lac} promoter with ampicillin resistance was used, referred to as pPsiB. *PsiB* encodes an inhibitor of RecA that directly binds to RecA and inhibits binding to ssDNA, therefore preventing initiation of the SOS response.

This was a gift from the Uphoff lab, Department of Biochemistry, University of Oxford.

2.2.3. Bacterial media and culture conditions

All bacterial strains were cultured in High Salt Luria-Bertani (LB) media, with appropriate antibiotic selection (kanamycin 50μg/ml, chloramphenicol 30μg/ml, ampicillin 100μg/ml) at 37°C and 220rpm. Strains were plated on LB High Salt Agar plates with antibiotics as above and incubated at 37°C overnight. In order to make competent cells for transformation, WT and Δ*alkB* strains were grown to OD600 = 0.6, then were centrifuged at 4^0C and the pellet re-suspended in ice cold $CaCl_2$. They were incubated on ice for 30-60 minutes and then centrifuged again under the same conditions. The pellet was re-suspended in 500μl ice cold $CaCl_2$ ready for transformation. For transformations, 1-10ng of plasmid DNA was added to 100μl competent cells and incubated on ice for 30 minutes. Cells were heat shocked at 42^0C for 1 minute 15 seconds and placed back on ice for 2 minutes. The cells were then recovered in SOC at 37^0C for 45 minutes with gentle shaking. The transformed bacteria were plated on LB High Salt Agar with appropriate antibiotic selection and incubated overnight at 37^0C. Additionally, in the case of *E. coli* containing the inducible methyltransferase and/or MP1, LB High Salt Agar and media were supplemented with 0.5% glucose to maintain promoter repression and suppress M.SssI and MP1 expression. Briefly, in order to induce expression, glucose-

supplemented cultures were centrifuged and the pellet re-suspended in LB High Salt media + 0.5% arabinose.

2.2.4. Rifampicin resistance assay

A rifampicin resistance assay was used to determine the difference in mutational burden caused by bacterial strains with and without methyltransferase activity. Rifampicin is an antibiotic that binds to the β subunit of bacterial RNA Polymerase. The gene *rpoB* encodes this subunit, and there are documented SNP mutations in *rpoB* that cause resistance to rifampicin (Garibyan et al., 2003). This assay enabled calculation of rifampicin resistant mutation frequency and mutation spectra arising from M.SssI activity.

WT and Δ*alkB* strains were transformed with either the full length or truncated form of M.SssI, M.MpeI or the inducible M.SssI. Where the introduction of RecA into Top10 was required, cells were transformed with both the inducible M.SssI and MP1 and plated on LB High Salt Agar + antibiotics supplemented with 0.8% glucose. Where inhibition of RecA was required, cells were transformed with both constitutive active or inactive M.SssI and PsiB and plated on LB High Salt Agar + antibiotics - IPTG. Individual colonies were inoculated overnight in LB High Salt media + antibiotics depending on plasmid content and strain background at 37°C, 220rpm. The next day, cultures were diluted 100x. For the constitutive system, cultures were allowed to grow to OD600 = 0.8 before centrifugation. For the inducible system and MP1, cultures were grown in 0.5% glucose to OD600 = 0.4.

Then, cultures were centrifuged and the pellet re-suspended in LB High Salt media + 0.5% arabinose or glucose (control). For experiments involving pPsiB, cultures were grown to OD600 = 0.4, then split in two and treated with 1mM or 0mM IPTG. All cultures were then allowed to grow to OD600 = 0.8-1.0 before centrifugation at room temperature. The pellet was then re-suspended in 300ul LB High Salt media. 100µl of undiluted culture was plated on 100µg/ml rifampicin plates and the remaining culture used to make ten-fold serial dilutions. An appropriate dilution was plated on LB High Salt Agar + antibiotic plates. The non-rifampicin plates were incubated overnight, while rifampicin plates were incubated for a total of 48 hours, both at 37^0C. The rifampicin-resistant mutation frequency is approximated by dividing the number of rifampicin-resistant colonies by the number of colonies on LB media without rifampicin per independent replicate clone, each a mean average of two technical replicates. Mean rifampicin-resistant mutational frequencies can then be averaged to give an overall mean rifampicin-resistant mutational frequency per condition.

2.2.5. Colony PCR

Colony PCR was used following both the transformation of *E. coli* with damaged oligo ligations and the rifampicin resistance assay to isolate genomic DNA (gDNA) and amplify a specific region, with the aim of identifying mutations via Sanger sequencing (see 2.7.). To liberate gDNA from inside the cell wall, individual colonies were inoculated in dH_2O and heated for 10 mins at 99^0C. These were then centrifuged and the supernatant (DNA template) moved to a separate tube. To

determine the mutation(s) present in *rpoB*, a ~500bp region of interest in *rpoB* was amplified by PCR. PCR primers were adapted from Kurepina *et al*. that flanked a 510bp region of *rpoB*: rpoB_F_short: 5' GTAGAGCGTGCGGTGAAAG 3', rpoB_R_short: 5' GCCTTTGCTACGGCAAGTTAC 3' (Kurepina et al., 2022). To determine the mutations arising from DNA lesions in ligated constructs, an ~500bp region overlapping the oligo insertion site was amplified using pBR322-specific primers: F1_PBR: 5' AAGGGAATAAGGGCGACACG 3' and R1_PBR: 5' GTAGGTTGAGGCCGTTGAGC 3'. All PCR reactions were set up as follows: 1µl DNA template, 12.5µl 2x MasterMix (MedChemExpress, Cat. No. HY-K0531), 1.25µl each of 10µM forward and reverse primers, 9µl dH_2O. Cycles: 98°C 30 sec, then 25 x 98°C 10 sec, 68°C 30 sec, 72°C 30 sec, then 72°C for 2 minutes and 5°C for 10 minutes. PCR products were run on a 1.5% agarose gel to determine product size and specificity.

2.2.6. Sanger sequencing

Sanger sequencing was used to identify mutations arising in the fragments amplified by colony PCR (see 2.6.). 12µl of each PCR reaction (10ng/µl) was sent for clean-up and sequencing with Source Bioscience using the rpoB_R_short primer (3.2pmol/µl). Resulting FASTQ reads were aligned to an *rpoB* reference sequence in Benchling using MAFTT v7 (Katoh & Standley, 2013) to identify SNPs.

2.2.7. Lambda red recombination

Lambda red recombination was used to knock out *alkB* and replace it with KanR in Top10 *E. coli*. Primers are specifically designed to amplify an insert gene, while containing 50nt homology arms that flank the target gene. The plasmid pKD46 contains lambda red recombination genes *exo*, β and γ. Exo is a 5' > 3' dsDNA-dependent exonuclease, Beta protects the subsequent ssDNA products and promotes annealing to homologous ssDNA, and Gam prevents endogenous nucleases from degrading the insert fragment (Yu et al., 2000). Altogether, these genes facilitate the replacement of a target gene with a PCR-amplified insert fragment, dependent on homology between the insert and target gene (Datsenko & Wanner, 2000). The kanamycin resistance cassette was amplified out of pET28a using the following primers: AlkBKanR_F_2.1: 5' tgtcgcgtaaagcgcaac tgctgcgccgcgaagctgaaaatgaggagaggtGTTTGCAAGCAGCAGATTACGC 3' and AlkBKanR_R_2.1: 5' cagcccgcagtttaaacatcttcgcgcgcacagcaataataattcttattt ATGTGCGCGGAACCCCTATT 3'. Homology arms are shown in lower case and KanR-specific sequences in capitals. PCR reactions were set up as follows: 1µl pET28a template, 12.5µl 2X MasterMix (MedChemExpress, Cat. No. HY-K0531), 1.25µl each of 10µM forward and reverse primer, 9µl dH_2O. PCR amplification was as follows: 98°C 30 sec, then 25 x 98°C 10 sec, 68°C 30 sec, 72°C 30 sec, then 72°C for 2 minutes and 5°C for 10 minutes. The desired PCR product was gel purified using the Zymoclean Gel DNA Recovery Kit (Cat. No. D4008) and run on the TapeStation to determine DNA concentration. WT Top10 were then transformed with pKD46, with recovery at 30°C to maintain the plasmid and plated on LB High Salt Agar + 100µg/ml

ampicillin. These were incubated overnight at 30°C. The following day, transformants were inoculated in LB High Salt media + ampicillin and incubated overnight at 30°C. The following morning, the overnight culture was diluted 100X in LB High Salt media + 50µg/ml ampicillin + 0.2% arabinose (to induce expression of lambda red genes) in a baffled flask. This was incubated at 30°C until OD600 = 0.6. The flask was submerged in ice for ten minutes and transferred to a pre-chilled 50ml falcon tube. Cells were then spun down for ten minutes at 3000 x g at 4°C and the pellet re-suspended in 35ml ice cold dH_2O. This was repeated three times, with the pellet re-suspended in 35ml again, then 2ml, then finally 500µl ice cold dH_2O. 5µl of the gel purified insert fragment was added to 90µl aliquots of cells, vortexed briefly and incubated on ice for ten minutes. They were then transferred to pre-chilled 2mm electroporation cuvettes and electroporated at 2.5kV (~5ms pulse). Cells were immediately recovered in 1ml SOC and incubated without shaking at 37°C for 3 hours. Cells were then spun down, re-suspended in 300µl SOC and plated on LB High Salt agar plates + 50µg/ml kanamycin. These were incubated overnight at 37°C. The following day, individual resistant colonies were picked, re-streaked on selective plates and incubated overnight at 37°C. The following day, each colony was streaked onto a new selective plate and also inoculated in LB High Salt media + 50µg/ml kanamycin at 37°C overnight. The following day, gDNA was extracted using the Qiagen DNeasy Blood and Tissue kit (Cat. No. 69504) according to their protocol for *E. coli* and DNA concentration determined via Qubit dsDNA HS Assay Kit (Cat. No. Q32851). Genotype was confirmed by PCR using primers that bound either side of the target site: alkB_F1: 5' TGTCGCGTAAAGCGCAACT 3' and alkB_R1: 5' CCCGGACAATGAATTACAGCC 3'. PCR products were run on a 1% agarose gel to

check for band size and successful *alkB* replacement with KanR. Replacement of *oxyR* with a kanamycin resistance cassette was used as a control for lambda red recombination efficiency. gDNA was extracted from the Keio collection OxyR::Kan strain using the Qiagen DNeasy Blood and Tissue kit (Cat. No. 69504) according to their protocol for *E. coli* and then used as a template for amplification of the KanR cassette. Primers used for KanR amplification: SU328_OxyR-Lreddel: 5' ctattctacctatcgccatgaactatcgtggcgatggaggatggata ATGGGAATTAGCCATGGTC 3' and SU329_oxyR-Lreddel: 5' gacgatggcggaagcctatcgggtagctgcgttaaacggt GTGTAGGCTGGAGCTGCTTCG 3' (Lagage et al., 2023). PCR conditions were as follows: 12.5µl 2X MedChemExpress Master Mix, 1.25µl each 10M primer, 1µl gDNA, 9µl dH2O. Cycles: 98°C 30 sec, then 25 x 98°C 10 sec, 67°C 30 sec, 72°C 30 sec, then 72°C for 2 minutes and 5°C for 10 minutes. Genotyping primers as follows: SU231_OxyRdel_fwd: 5' TCTCGAAACGGGCAGTGA 3' and SU232_OxyRdel_rev: 5' TCGGTCAGGCGATTATGG 3'. Genotyping was as follows: 12.5µl Q5 Hot Start High-Fidelity 2X Master Mix (Cat. No. M0494S), 1.25µl each 10M primer, 1µl gDNA, 9µl dH2O. Cycles: 98°C 30 sec, then 25 x 98°C 10 sec, 65°C 30 sec, 72°C 2 minutes, then 72°C for 2 minutes and 5°C for 10 minutes. The OxyR::Kan Keio collection strain and all OxyR primers were a gift from the Uphoff lab, Department of Biochemistry, University of Oxford.

2.2.8. ROS detection assay

A fluorescence-based assay was used to measure cellular ROS levels in real time. Here, we used the commercially available fluorescent probe DCFH-DA, which is

oxidised by a range of reactive oxygen species to highly fluorescent DCF (Afri et al., 2004). The density of bacterial cultures was also measured via absorbance at 600nm. Top10 *E. coli* were transformed with the inducible M.SssI system as detailed in 2.3 and inoculated overnight at 37°C, 200rpm in LB High Salt media + 0.5% glucose (+100μg/ml ampicillin for cells + inducible pM.SssI). The following morning, cultures were diluted 20X in LB High Salt media + 0.5% glucose and grown to OD600 = 0.4. Cultures were then split in two, centrifuged and re-suspended in LB High Salt media, supplemented with either 0.5% glucose or arabinose. Cultures were then grown to OD600 = 1.0. In a 96-well plate, 3 x 100μl of cells was plated and 2μl of 50X reagent (Total Reactive Oxygen Species (ROS) Assay Kit 520nm, ThermoFisher Scientific, Cat. No. 88-5930) was added to each well. The plate reader was heated to 37°C and programmed to measure absorbance and fluorescence, then shake for ten minutes, 200 cycles/min (orbital, 2mm shaking diameter). Fluorescence was measured by excitation at 475nm and emission at 500-550nm and absorbance was measured at 600nm. This loop was repeated 100 times. At least two independent clones (biological replicates) were used per condition with each plated twice to give two technical replicates per biological replicate. Mean fluorescence and absorbance were calculated over technical and biological replicates per condition to give a mean at each time point. Mean fluorescence was normalised by dividing by corresponding mean absorbance to give fluorescence/cell/ml at each time point. Non-plasmid-bearing Top10 were used as a negative control, as well as Top10 + inducible M.SssI under glucose repression.

Chapter 3 – Methylation lesions reduced number of viable transformants in WT and Δ*alkB E. coli*

3.1. Introduction

The initial question I sought to answer was whether alkylation lesions inhibited DNA replication and therefore threatened genome stability in *E. coli*. Furthermore, I sought to determine whether such lesions were mutagenic, particularly in the absence of AlkB. I hypothesised that alkylation lesions would indeed inhibit DNA replication in *E. coli,* as there is evidence for stalling of the replisome by small alkylation lesions in eukaryotic cells (Groth et al., 2010; Tercero & Diffley, 2001), as well as resulting mutagenicity (Anderson et al., 2024; Nieminuszczy et al., 2009; Volkova et al., 2020). Similar findings in *E. coli* would support more in depth investigation into mutagenicity arising from DNMT activity.

3.2. Results

3.2.1. DNA damage construct design and optimisation

In order to assess the effect of alkylation lesions on genome stability, I designed an assay whereby I transformed WT and Δ*alkB* Top10 and C2523 *E. coli* with plasmids containing a single site of DNA damage. The effect of the lesion on genome stability was inferred from the number of surviving colonies from damaged vs undamaged transformants. I designed a series of short oligos of the same sequence, each

containing a single lesion at a specific location. Each damaged oligo was annealed to an undamaged reverse oligo and ligated into a non-functional region of the plasmid backbone, expressed ampicillin resistance. WT and Δ*alkB* Top10 and C2523 *E. coli* were then transformed with the damaged plasmids. Cells that were able to repair or replicate over the lesion would pass on ampicillin resistance to daughter cells, therefore increasing the number of viable colonies on selective media (Figure 3.1). The oligos used in this assay are shown in Methods, Table 1. Types of modification included alkylation lesions 3mC and 1mA, as well as controls containing an abasic site or 5mC. Note that the forward oligo containing 1mA is one nucleotide shorter than the rest due to an error during the oligo design phase, therefore I re-designed the undamaged reverse oligo so that it fully annealed to the damaged forward. I also designed a reverse oligo containing an abasic site for annealing to the abasic forward oligo, to result in a double abasic site. This served as a strong negative control, as neither strand could be used as a template during DNA replication leading to replication fork stalling and collapse (Mehta et al., 2020; Thompson & Cortez, 2020).

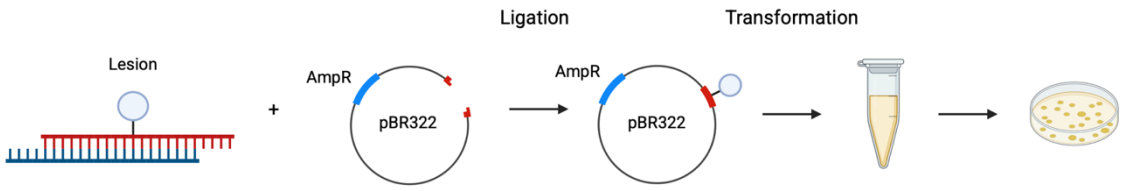


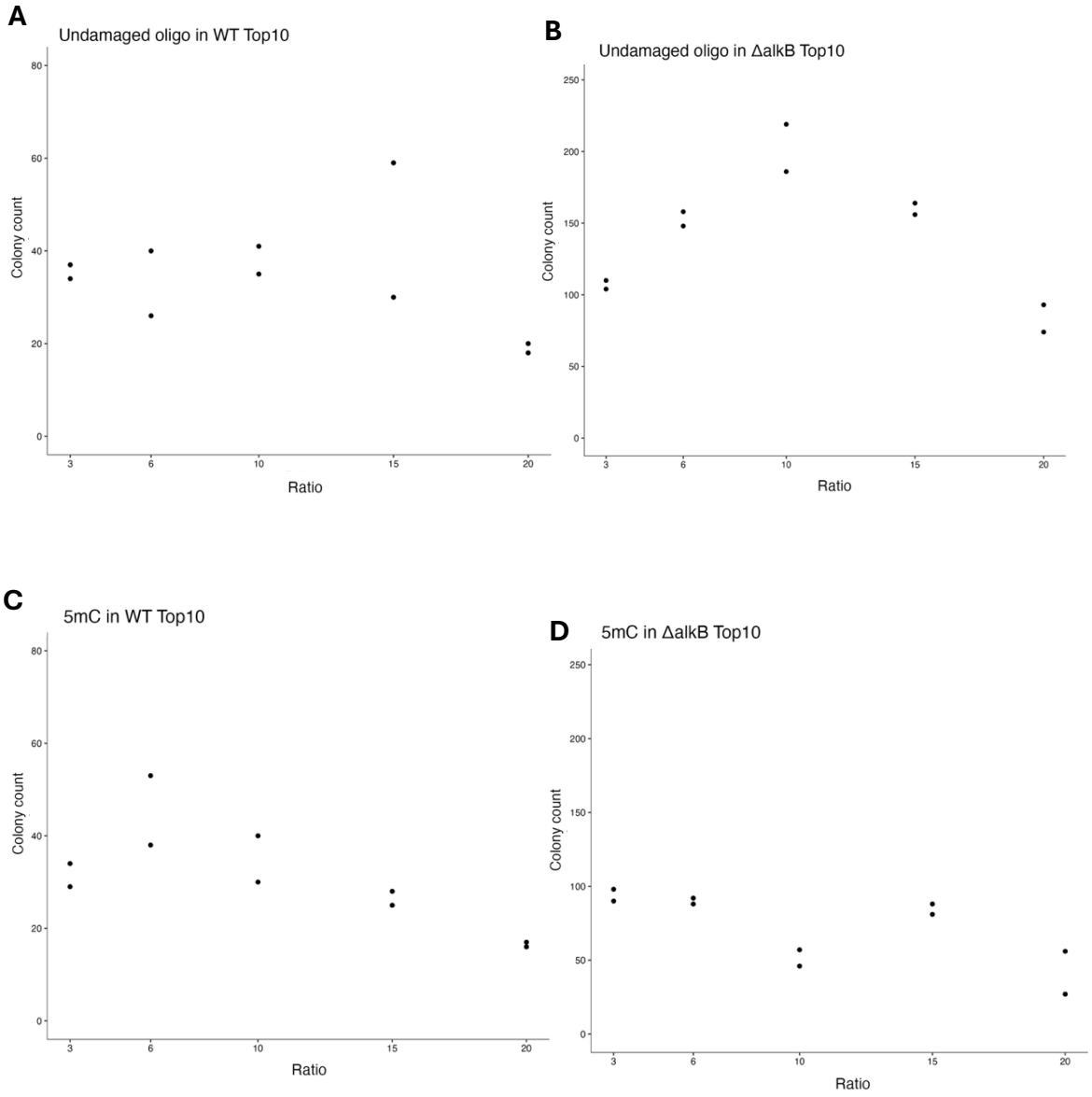
Figure 3.1: Schematic of cloning strategy used to determine the effect of DNA lesions on transformation efficiency and DNA replication. 24 base pair dsDNA oligos containing various DNA modifications were ligated into linearised pBR322 at 16°C overnight. The following day, ligation products were transformed into *E. coli* by heat shock at 42°C and recovery at 37°C. Viable transformants were selected for on LB High Salt media + 100μg/ml ampicillin. Created in Biorender.

3.2.2. Transformation and ligation efficiency

It was crucial that ligation efficiency was optimised in order to fully observe the effects of the modification on number of viable transformants. Typically when ligating an oligo into a plasmid, a 3:1 molar ratio of insert to vector is used to ensure high ligation efficiency. However, the optimum ratio may vary with factors such as the length of insert. For example, a higher ratio of insert to vector is sometimes recommended for short oligos, although very high amounts of insert can impede ligation (Bercovich et al., 1992). Given that the oligos above were very short, I decided to perform a ligation titration to find the optimum ratio of insert to vector. Here, I annealed each oligo to the undamaged reverse and ligated them into 20ng pBR322 DNA in a range of insert:vector molar ratios. I transformed each ligation into WT (Figure 3.2 A, C, E, G, I) and Δ*alkB* (Figure 3.2 B, D, F, H, J) Top10 *E. coli* and counted the number of viable colonies. I then normalised the number of viable colonies for transformants containing each modified oligo to the undamaged control (Figure 3.3).

Mean colony counts varied depending on the type of modification present (Figure 3.2). Increasing the insert:vector molar ratio generally decreased the mean number of viable colonies in both WT and Δ*alkB* cells. Normalisation of mean colony counts for plasmids containing modifications to the undamaged plasmid control revealed that 3mC, 1mA and a single abasic site had a negative effect on number of viable transformants in both WT and Δ*alkB* cells (Figure 3.3 B, C, D). Surprisingly, 5mC also reduced viable transformants compared to the undamaged control in WT cells and

to a greater extent in Δ*alkB* cells (Figure 3.3A). Overall, 3:1 and 6:1 molar ratios of insert to vector gave the highest ligation efficiency. I decided to use a 6:1 ratio, based on these data as well as the literature.



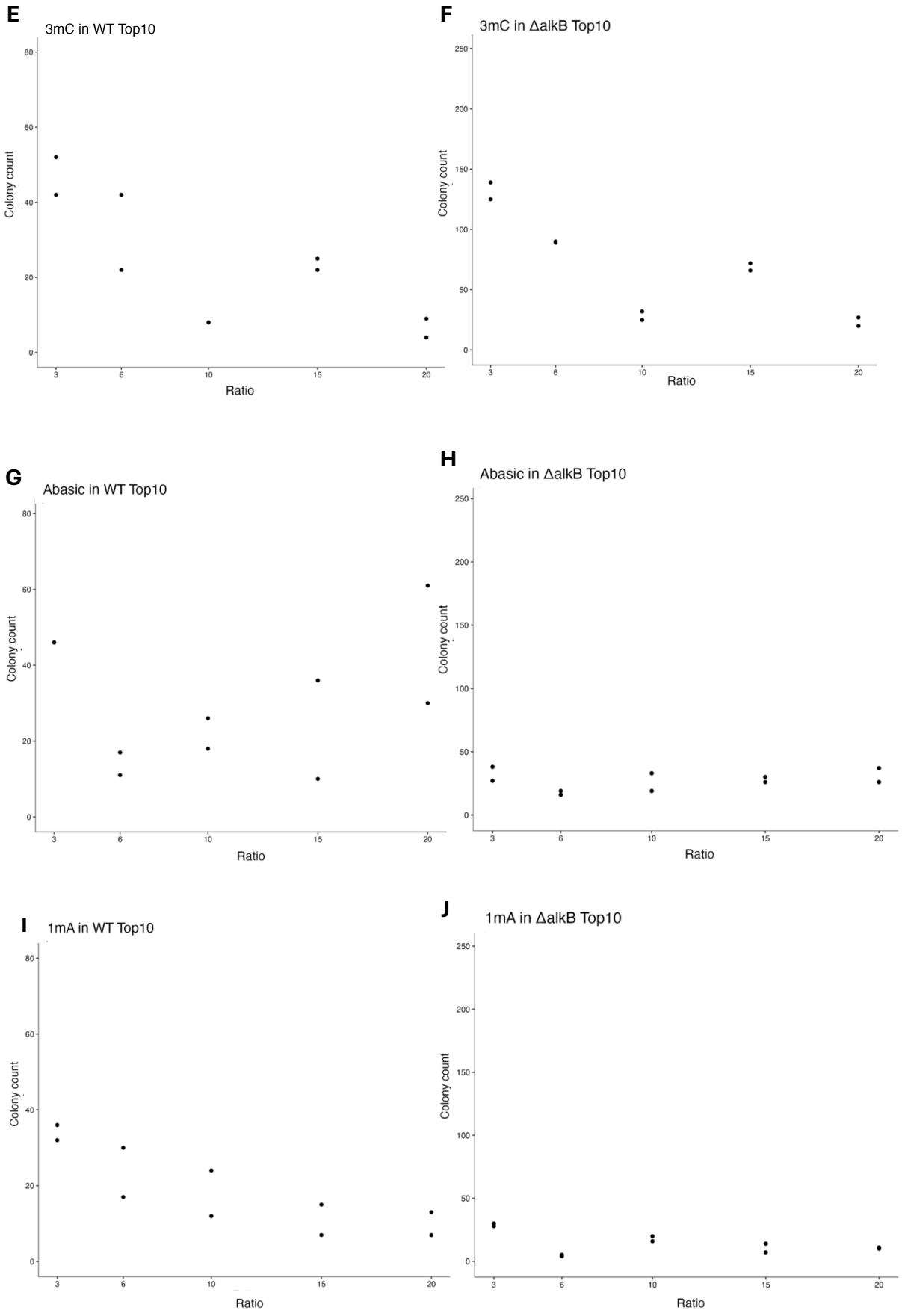

E
3mC in WT Top10
Colony count
Ratio
F
3mC in ΔalkB Top10
Colony count
Ratio
G
Abasic in WT Top10
Colony count
Ratio
H
Abasic in ΔalkB Top10
Colony count
Ratio
I
1mA in WT Top10
Colony count
Ratio
J
1mA in ΔalkB Top10
Colony count
Ratio

Figure 3.2: Optimisation of molar ratio of insert:vector via ligation titration. A 24 base pair oligo was ligated into 20ng pBR322. The oligo contained a single site on the forward strand of either no damage **(A, B)**, 5mC **(C, D)**, 3mC **(E, F)**, an abasic site **(G, H)** or 1mA **(I, J)**. The ligations were performed at an insert:vector molar ratio of 3:1, 6:1, 10:1, 12:1, 15:1 or 20:1. These constructs were transformed into WT (left hand side) or Δ*alkB* (right hand side) Top10 *E.* coli, plated and incubated overnight at 37°C. Each point represents the mean number of viable colonies following transformation of an independent replicate clone with 1µl of the ligation reaction, averaged over two technical (plating) replicates.

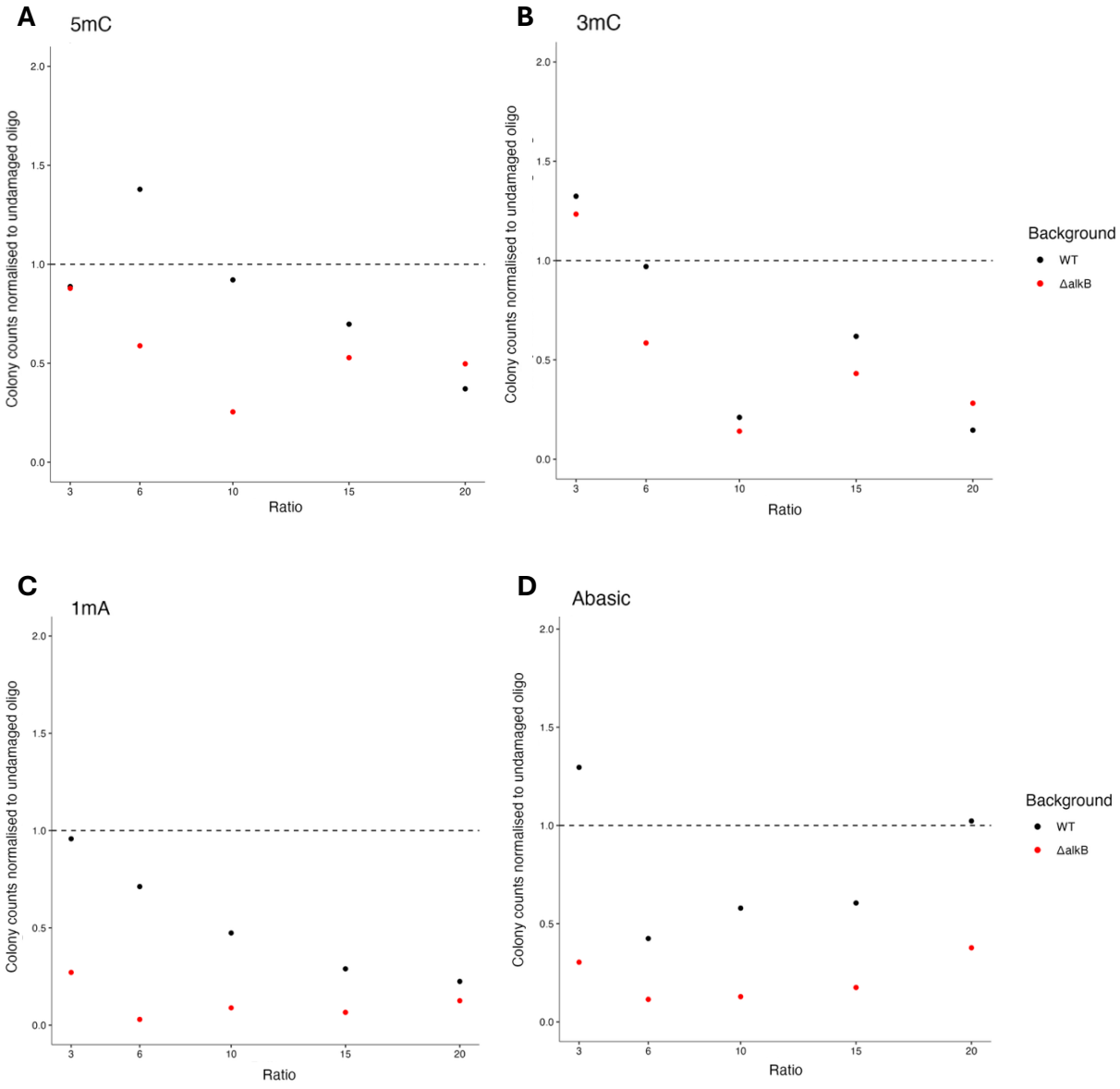
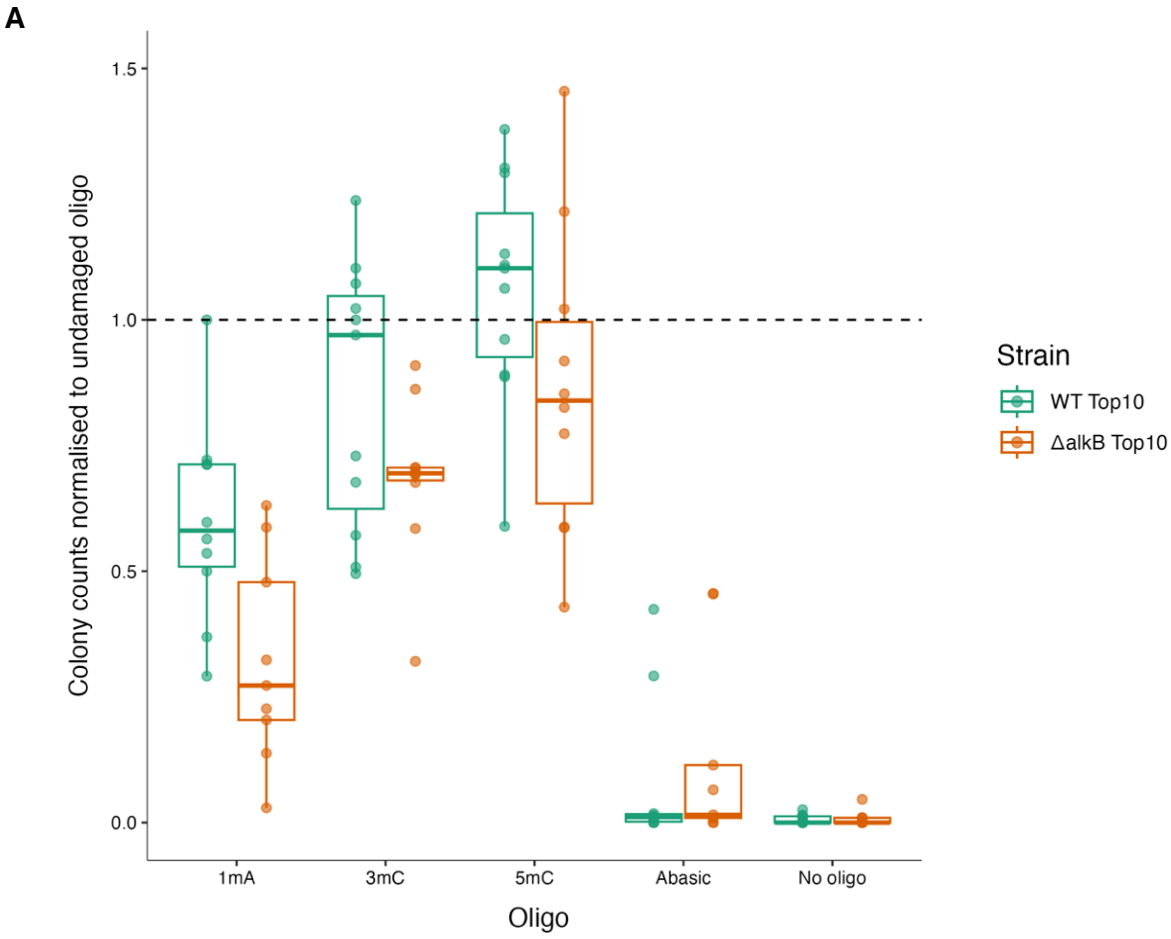


Figure 3.3: Normalisation of ligation titration colony counts to undamaged oligo control. Mean colony counts for transformants containing each modified oligo (5mC **(A)**, 3mC **(B)**, 1mA **(C)**, a single abasic site **(D)**). Averaged over two transformations of independent replicate clones, each the average of two technical replicates. The mean of the independent replicate clones for each modification was normalised to the mean of the undamaged control oligo in either WT (black) or Δ*alkB* (red) Top10 *E. coli*.

I then tested the effects of 5mC, 3mC and 1mA at a 6:1 molar ratio. I annealed the forward oligo containing an abasic site to the reverse containing an abasic site at the same location, providing a strong negative control. As before, I normalised the colony counts for transformants containing each modification type to the undamaged control. The effects of these modifications on number of viable transformants in WT and Δ*alkB* Top10 and C2523 are shown in Figure 3.4. Alkylation lesions 1mA and 3mC had a negative impact on number of viable transformants in all genetic backgrounds. 1mA in particular had a strong detrimental effect on colony number even in WT Top10 and C2523, and an even stronger effect in Δ*alkB* cells. As expected, the double abasic site and un-ligated control produced almost no colonies in any of the backgrounds. Surprisingly, 5mC slightly reduced number of viable transformants compared to the undamaged control in WT and Δ*alkB* C2523 and Δ*alkB* Top10. 5mC is typically tolerated by replicative polymerases (Huber et al., 2016) and so it was unexpected that presence of 5mC would impede replication and therefore transformation efficiency.

In order to investigate the potential mutagenic effects of these modifications, I extracted plasmid DNA from WT and Δ*alkB* cells from the assay described above and sequenced a portion of the plasmid containing the oligo insertion site. Each

oligo contained the modification at a defined site, allowing for the accurate identification of modification-specific mutagenesis. Surprisingly, none of the colonies derived from lesion-containing oligos in either the WT or Δ*alkB* background showed any mutations at the site of damage. This suggested that surviving transformants containing alkylation lesions were able to repair damage error-free, for example by template switching.



B

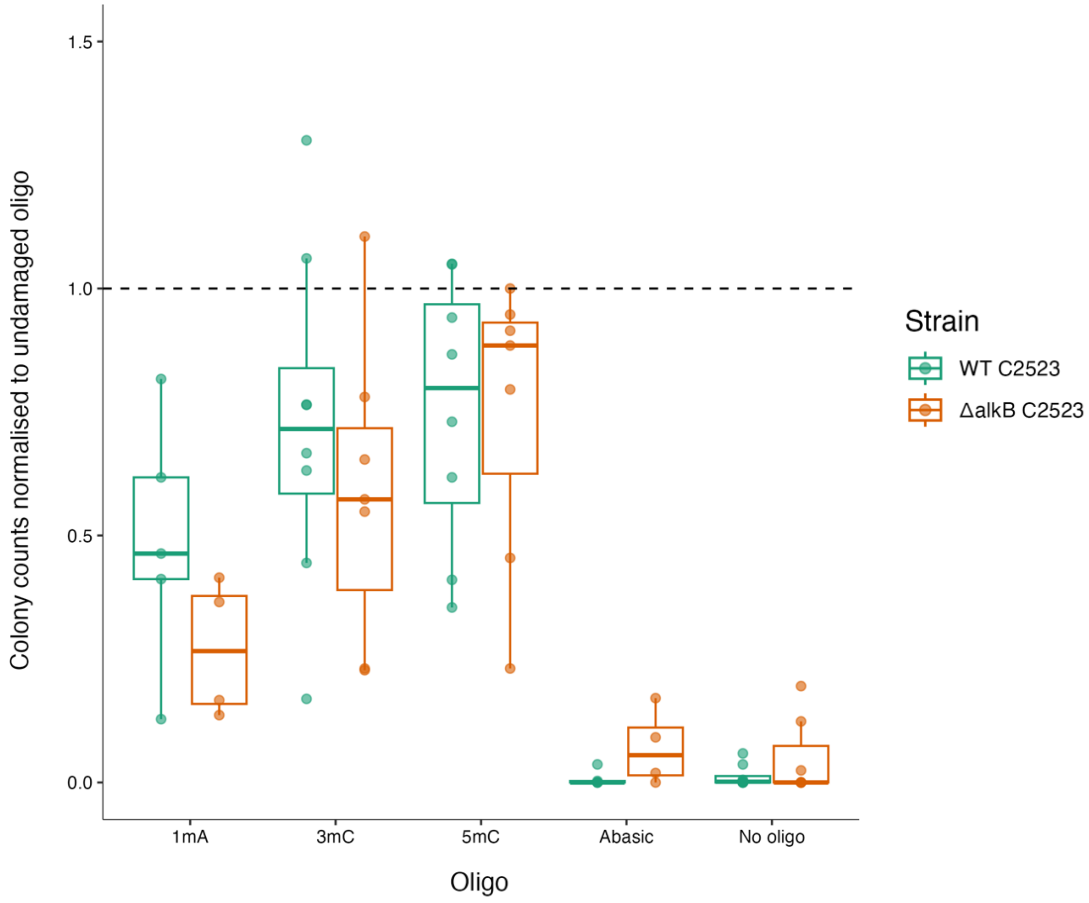


Figure 3.4: The effect of DNA modifications on transformation efficiency in WT and repair-deficient *E. coli*. Oligos containing a single site of 5mC, 3mC, 1mA, or a double stranded abasic site were ligated into pBR322 and transformed into WT and Δ*alkB* Top10 **(A)** and C2523 **(B)** *E. coli*. This included a negative control ligation containing the linearised plasmid with no oligo insertion. Colony counts for transformants containing modified oligos were normalised to the undamaged control. Each point represents mean colony counts for an independent replicate clone averaged over two technical replicates. Data is shown as box plots: the median is shown as the thick line within the box. The distance between the median and upper boundary of the box is the upper quartile, while the distance between the median and the lower boundary is the lower quartile. The box therefore shows the interquartile range (IQR). The whiskers show the highest and lowest data points within 1.5X IQR.

3.3. Specific discussion

Overall, these data suggested that the presence of alkylation lesions 3mC and 1mA reduced the number of viable transformants, particularly in the absence of AlkB. Interestingly, 1mA was a greater burden in all strains compared to 3mC. This suggested that 1mA was harder for cells to repair even when AlkB was present and was less efficiently repaired by backup repair pathways when AlkB was absent.

I hypothesise that the effect of lesions on number of viable transformants was due to stalling of the replisome and subsequent replication stress. Failure to resume DNA replication on the other side of the lesion could lead to replication fork collapse and cytotoxic DSBs (Sedletska et al., 2013). Alternatively, DNA replication may slow to accommodate repair (Iyer & Rhind, 2017; Lopes et al., 2001). Either of these possibilities or a combination of both could reduce genome stability and the number of viable transformants. However, it is possible that presence of DNA modifications also affected transformation efficiency. In this case, the modification could reduce the number of viable transformants without impacting DNA replication and genome stability.

Presence of mutagenesis at modification sites would have supported the hypothesis that modifications reduced number of transformants by increasing DNA replication stress. However, there was a distinct lack of mutagenesis at damage sites, particularly in Δ*alkB* cells. This was surprising, as the small alkylation lesion 3mA has been associated with T > M (M = A/C) mutations (Volkova et al., 2020),

suggesting that other forms of small alkylation damage may also be mutagenic. Another possibility is that surviving transformants were able to repair damage error-free, for example by template switching. This DNA damage tolerance mechanism is largely error-free and there is some evidence to suggest that it is promoted over error-prone TLS in eukaryotes (Zhang & Greenberg, 2022). Template switching uses an undamaged template to incorporate the correct nucleotide via Watson-Crick base pairing. In order to reduce the possibility of error-free repair, dsDNA oligonucleotides could be designed that have the alkylation lesion, for example 3mC, on both strands in the CpG dinucleotide.

Despite this, these data suggested that *E. coli* lacking AlkB could not tolerate small alkylation damage as well as WT. Therefore, potential introduction of alkylation damage, including 3mC, by DNMTs could also pose a threat to genome stability in the absence of AlkB. This supported further investigation into the potential mutagenic consequences of DNMT activity.

Chapter 4 – M.SssI introduces alkylation damage that is specifically repaired by AlkB

4.1. General introduction to rifampicin resistance assay

The work shown in the previous chapter demonstrated that alkylation lesions negatively impacted the number of viable transformants, potentially via inhibition of DNA replication, particularly in the absence of AlkB. Therefore, it was possible that induction of 3mC by DNMTs could pose a threat to genome stability (Dukatz et al., 2019; Rošić et al., 2018). If this were the case, then such evidence would support the wider hypothesis that the propensity of DNMTs to damage DNA contributed to the evolutionary loss of DNA methylation.

The next question was does this damage lead to mutagenesis? Previous investigation into 3mC induction by DNMTs was conducted in mouse ESCs (Rošić et al., 2018). However, studying the mutagenic effects of CpG DNMT activity in mammalian systems is challenging, mainly due to the complex role that CpG DNA methylation plays in epigenetic regulation. One can imagine that modifying existing DNMT activity would influence gene expression in an unpredictable manner, potentially resulting in mutagenesis not as a direct consequence of lesion induction by DNMTs, but instead due to secondary effects caused by changes to epigenetic regulation of gene expression.

To better isolate the mutagenic consequences of DNMT activity in WT and repair-deficient backgrounds, I modelled mammalian CpG methylation in *E. coli* by transforming WT and Δ*alkB* cells with a CpG methyltransferase. This will be detailed further in 4.2.1.1. Then, to study the mutagenic effects of introducing CpG methylation into *E. coli*, I needed a way of selecting for cells experiencing mutagenesis. For this purpose, I utilised a rifampicin resistance-based selection assay. Rifampicin is an antibiotic that binds to the β subunit of bacterial RNA Polymerase and inhibits catalytic activity. The β subunit is encoded by *rpoB* and many mutations in *rpoB* have been documented that give rise to resistance to rifampicin (Garibyan et al., 2003). Most of these mutations occur within three mutational hotspot regions (mutational clusters I, II and III). They change the structure of the β subunit such that it is still catalytically active, but rifampicin can no longer bind (Molodtsov et al., 2017).

The premise of the rifampicin resistance assay was that DNMT activity would potentially result in 3mC deposition, some of which would result in mutations. Some of these mutations would occur in *rpoB* and a proportion of these would give rise to rifampicin resistance. These mutants could then be selected for on media containing rifampicin (Figure 4.1). The number of rifampicin-resistant mutant colonies compared to total colonies on media without rifampicin could then be used to quantitatively determine a rifampicin-resistant mutation frequency. An approximately 250bp region of *rpoB* from mutant colonies could also be sequenced to give a *rpoB*-specific mutation spectrum, although it should be noted this is not an unbiased genome-wide mutational signature. This enables qualitative

determination of the effects of DNMT activity and repair capacity on mutagenesis. It is important to note that only a subset of mutations give rise to rifampicin resistance, therefore this assay does not give a true approximation of mutation rate or mutational signature. However, it did enable comparison of rifampicin-resistant mutation frequencies and *rpoB*-specific mutation spectra across test conditions, allowing for qualitative analysis of the effects of DNMT expression on mutagenesis.

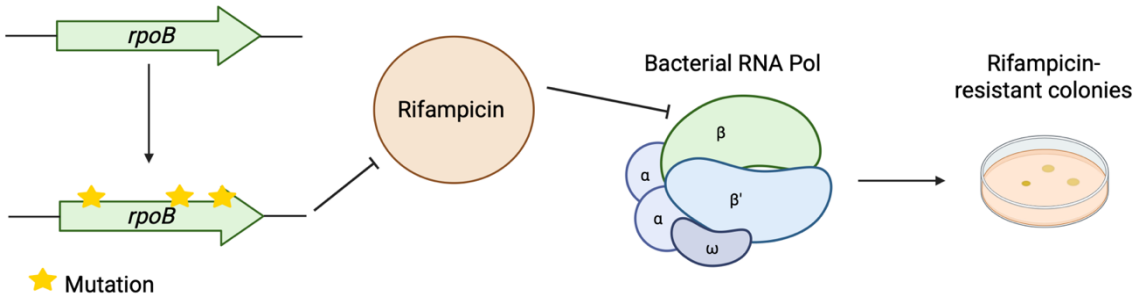


Figure 4.1: Schematic of rifampicin resistance assay. Resistance to the antibiotic rifampicin was used to select for mutant *E. coli*. Off-target activity of M.SssI can result in mutagenic lesions such as 3mC. Mutations in *rpoB* alter the structure of the β subunit of RNA Polymerase and prevent the binding of rifampicin. This renders *E. coli* resistant to rifampicin and allows for selection of mutant colonies on media containing rifampicin. Created in BioRender.

4.2. Results

4.2.1. Constitutive system

4.2.1.1. Introduction to the constitutive M.SssI system

The first question I aimed to answer was: how does the presence of CpG DNA methylation affect mutagenesis in *E. coli*? In order to address this, I modelled

mammalian CpG methylation in bacteria by transforming WT and Δ*alkB* C2523 *E. coli* with a plasmid containing the CpG DNMT *m.sssI* (pM.SssI) under the control of a constitutively active promoter. M.SssI specifically methylates cytosine within CpG dinucleotides in dsDNA. The *m.sssI* was either a full length, catalytically active form that methylated DNA, or a truncated, inactive form that did not methylate DNA (Figure 4.2). In this way, DNA methylation was either always 'on' or 'off'. Crucially, C2523 *E. coli* lacks the mcrABC restriction system that digests methylated cytosines such as 5mC, 5hmC and 4mC (Sukackaite et al., 2012) in RmC sequence context, where R is either A or G (Raleigh & Wilson, 1986). This allows stable introduction of CpG methylation via pM.SssI. Additionally, C2523 lacks *dcm,* therefore does not possess cytosine methylation and further, does not rely on cytosine methylation for epigenetic regulation. While Dcm methylates the internal cytosine in 5' CCWGG 3' (May & Hattman, 1975) and does not methylate in the CpG context, it was still important that strains used in this study lacked any cytosine methylation. This enabled accurate interpretation of the effect of introducing CpG methylation on mutagenesis.

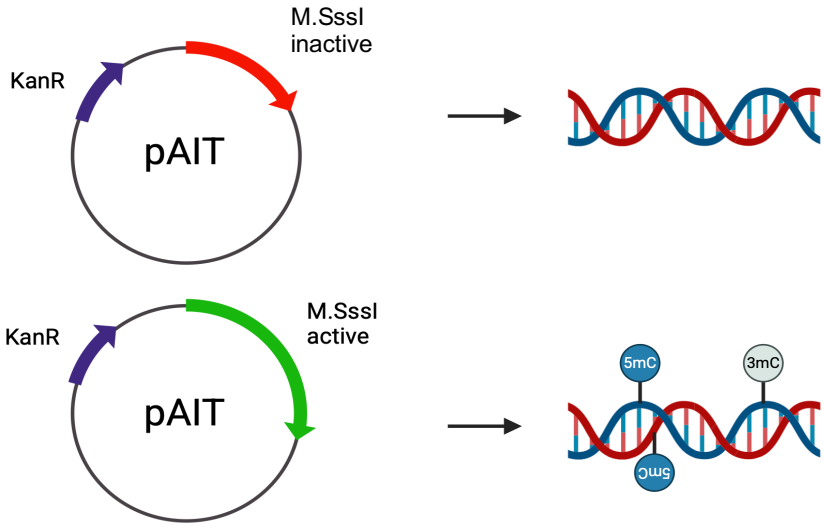


Figure 4.2: Schematic of plasmids containing either full length active or truncated inactive versions of the methyltransferase M.SssI for transformation into *E. coli*, introducing genome-wide CpG methylation. Created in BioRender.

4.2.1.2. M.SssI activity and loss of AlkB increased rifampicin-resistant mutation frequency in C2523 *E. coli*

I hypothesised that M.SssI activity and loss of AlkB would significantly increase mutagenesis due to the introduction of unrepaired 3mC. This was supported by previous work assessing the sensitivity of AlkB mutants with M.SssI activity to the alkylating agent MMS (Krwawicz et al., 2024). MMS predominantly introduces the alkylation adducts 7mG and 3mA, as well as other adducts such as 1mA and 3mC (Sedgwick, 2004). Therefore, it was predicted that the propensity of M.SssI to introduce off-target alkylation damage in the form of 3mC (Rošić et al., 2018) would increase sensitivity to MMS, particularly in the absence of AlkB. Results showed that *alkB* mutants were more sensitive to MMS than WT C2523, and the combination of AlkB loss and M.SssI activity led to a synthetic increase in sensitivity (DNMT * Δ*alkB* interaction p=0.0044, three-way Anova) (Krwawicz et al., 2024). This suggested that M.SssI introduced alkylation damage that went unrepaired, supporting further investigation into mutagenesis.

To determine the effect of Δ*alkB* and M.SssI activity on mutagenesis, I used the rifampicin resistance assay described in 4.1 in conjunction with the constitutive pM.SssI system. WT and Δ*alkB* C2523 *E. coli* transformed with active or inactive

pM.SssI were grown and plated on LB media with and without rifampicin. The mean mutation frequency was determined by dividing the mean number of rifampicin-resistant mutant colonies by mean number of colonies on media that did not contain rifampicin per condition. The mean was calculated over at least three independent replicate clones per condition, each an average of two technical replicates. This gave a rifampicin-resistant mutation frequency that could be compared across conditions to assess the effect of M.SssI activity on mutagenesis. M.SssI activity was confirmed semi-quantitatively via HpaII restriction digest. HpaII cuts DNA between cytosines in 5' CCGG 3' sites and is blocked by 5mC. M.SssI activity resulted in a high band following gel electrophoresis suggesting presence of DNA methylation, while inactive M.SssI resulted in a lower band pattern, indicating a lack of DNA methylation (Figure 4.3A). It is worth noting that HpaII digests do not report on CpG methylation levels specifically but provide an estimate of global DNA methylation.

Loss of AlkB significantly increased rifampicin-resistant mutation frequency compared to WT C2523 (Δ*alkB* p=5.02e-06 two-way Anova). M.SssI activity, independent of background, exacerbated this effect (M.SssI p=0.0052, two-way Anova) (Figure 4.3B). Indeed, the increase in rifampicin-resistant mutation frequency for Δ*alkB* + active M.SssI was greater than the additive increase expected for M.SssI activity and loss of AlkB (Δ*alkB* * M.SssI interaction p=0.00037, two-way Anova) (Figure 4.3B). This synthetic increase in rifampicin-resistant mutation frequency was consistent with AlkB-repaired alkylation damage, likely including 3mC introduced by M.SssI.

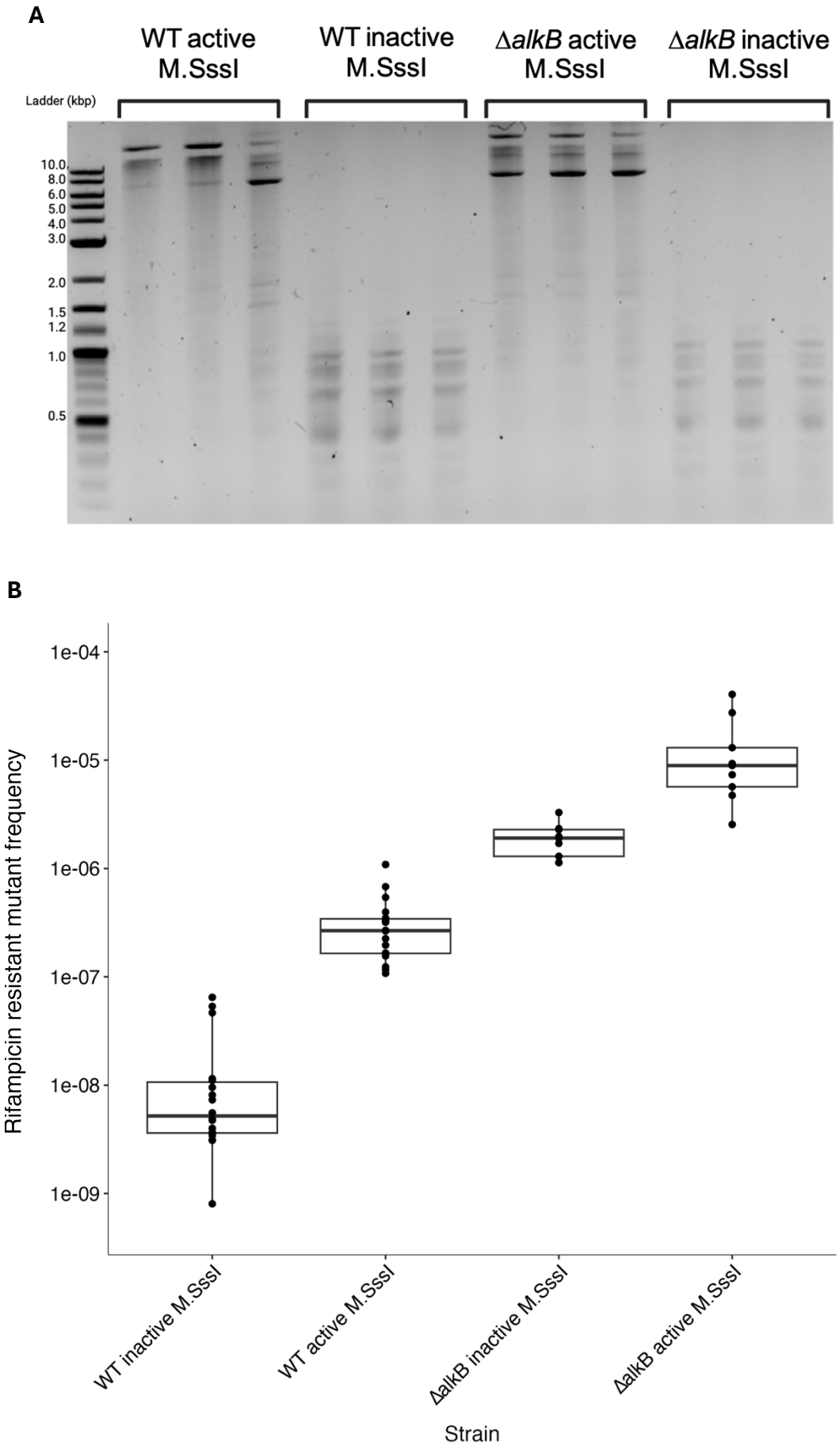

A
WT active M.SssI
WT inactive M.SssI
ΔalkB active M.SssI
ΔalkB inactive M.SssI
Ladder (kbp)
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.2
1.0
0.5
B
Rifampicin resistant mutant frequency
1e-04
1e-05
1e-06
1e-07
1e-08
1e-09
WT inactive M.SssI
WT active M.SssI
ΔalkB inactive M.SssI
ΔalkB active M.SssI
Strain

Figure 4.3: M.SssI activity and loss of AlkB increased rifampicin-resistant mutation frequency in C2523 *E. coli*. (A) Methylation-sensitive HpaII restriction enzyme digest of pM.SssI (active or inactive) DNA from *E. coli* containing either active or inactive pM.SssI. Each lane shows pM.SssI (active or inactive) DNA extracted from individual replicate clones. HpaII is blocked by methylation in CCGG context. **(B)** WT and Δ*alkB* C2523 *E. coli* were transformed with active and inactive pM.SssI and plated on LB media ± rifampicin. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates.

4.2.1.3. Identification of TA > CG mutations dependent on M.SssI activity and loss of AlkB

In order to determine the exact mutations that were causing this increase in rifampicin-resistant mutation frequency, *rpoB* from resistant colonies was sequenced and the frequencies of different mutations tabulated. Equal numbers of colonies were sequenced that arose from each original independent replicate clone for each condition. While not all mutations give rise to rifampicin resistance, it was still possible to compare *rpoB*-specific mutation spectra between conditions to infer the mutagenic effects of M.SssI expression and Δ*alkB*. In WT C2523 with M.SssI activity, observed mutations were primarily CG > TA (Figure 4.4A). This was expected due to introduction of 5mC by M.SssI and subsequent deamination to thymine. This was mostly observed at CpG dinucleotides, which again was consistent with M.SssI activity. However, when M.SssI was expressed in Δ*alkB* cells, there was a marked increase in CG > TA and TA > CG mutations (Figure 4.4B). This was surprising, as the

TA > CG mutations were dependent on both M.SssI activity and loss of AlkB, but occurred in a non-CpG context. Previous work suggested that M.SssI introduced alkylation damage at cytosine in the form of 3mC (Rošić et al., 2018), therefore it was unexpected that M.SssI would increase AT mutation frequency.

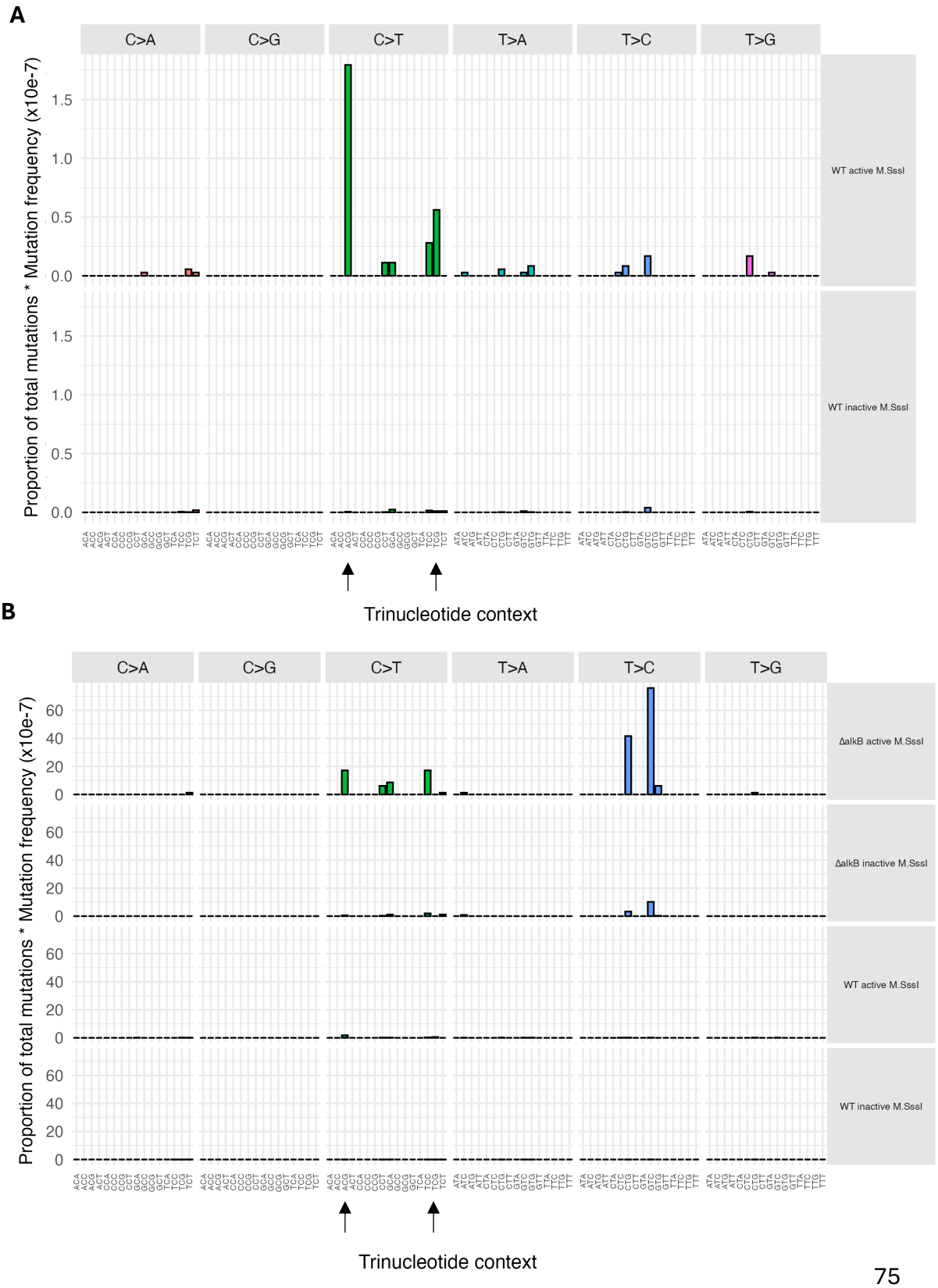


Figure 4.4: Introduction of CpG methylation into C2523 *E. coli* led to an increase in non-cytosine mutations dependent on M.SssI activity and loss of AlkB. DNA from mutant colonies was isolated and a ~500bp region of *rpoB* amplified by PCR. PCR products were sequenced and aligned to a reference genome to identify *rpoB*-specific mutations from individual replicate clones. Mutations were then pooled to give rpoB-specific mutation spectra in WT (**A**) and Δ*alkB* (**B**) C2523 *E. coli* transformed with active and inactive pM.SssI. WT results are also shown in **(B)** for easier comparison of rifampicin-resistant mutation frequency. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows**.**

4.2.2. Inducible system

4.2.2.1. Introduction to the inducible system

Collectively, the rifampicin-resistant mutation frequency and *rpoB*-specific mutation spectrum data shown above indicated that active M.SssI increased mutations in C2523 *E. coli*. One can imagine that this increased mutational burden might have a detrimental effect on cell fitness. Further, constitutive expression of M.SssI over the time course of the assay (~24 hours) could lead to a selective advantage for cells with lower M.SssI activity. This would ultimately hinder interpretation of the results. Evidently, there was need for finer control over induction of methylation during the assay. Previously, another lab member developed an inducible pM.SssI system, where the same full length, active *m.sssI* as used in the constitutive system was put under control of an arabinose-inducible

P_{BAD} promoter (inducible pM.SssI) (Figure 4.7A). Presence of glucose resulted in a lower band pattern indicative of repression of M.SssI expression and inhibition DNA methylation. Arabinose treatment resulted in a single, high band at 3.0kbp indicating induction of M.SssI expression and DNA methylation (Figure 4.7B). The inducible pM.SssI system was not stably expressed in C2523 so instead was transformed into WT Top10 *E. coli*. As with C2523, Top10 lacks the *mcrABC* restriction system and *dcm* so was suitable for stable maintenance of CpG methylation (ThermoFisherScientific, 2025).

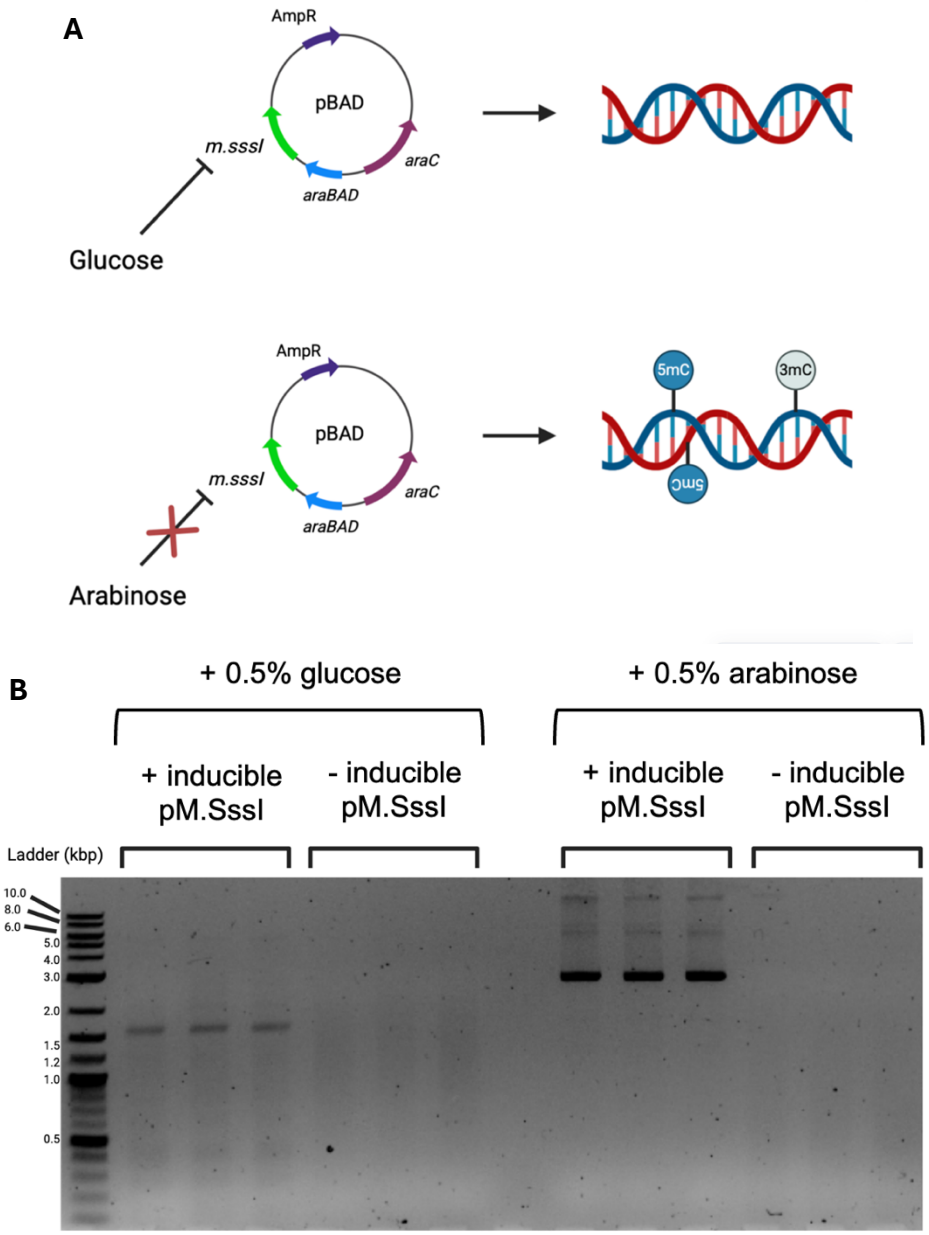


Figure 4.7: DNA methylation by inducible pBAD M.SssI. (A) Schematic of engineered plasmid containing the full length, active *m.sssI* under the control of an arabinose-inducible P_{BAD} promoter. The promoter is repressed and M.SssI expression inhibited in the presence of glucose. Removal of glucose and addition of arabinose alleviates repression and induces M.SssI expression, leading to DNA methylation. Created in BioRender. **(B)** Methylation-sensitive HpaII restriction enzyme digest of inducible pM.SssI DNA extracted from WT Top10 *E. coli* containing inducible pM.SssI under 0.5% glucose or arabinose respectively. Each lane shows inducible pM.SssI DNA extracted from individual replicate clones.

4.2.2.2. Knock out of *alkB* in Top10 by lambda red recombination

In order to fully assess the effects of inducible M.SssI expression on mutagenesis, an Δ*alkB* Top10 strain was required. This was not commercially available nor already engineered in the lab. Therefore, the first step was to make an Δ*alkB* Top10 strain. Crucially, Top10 contains the *recA1* allele which severely impairs RecA activity. RecA is a key component in DNA recombination (Del Val et al., 2019; Kowalczykowski, 2015; Smith & Wang, 1989). This meant that it was not possible to use common gene KO techniques such as P1 phage transduction, which relies on RecA-mediated recombination (Thomason et al., 2007). Instead, I used lambda red recombination, which is RecA-independent (Datsenko & Wanner, 2000) (Figure 4.8). Firstly, WT Top10 cells were transformed with a plasmid encoding *exo,* β and γ recombination genes. These facilitated the replacement of the target gene, in this case *alkB,* with a PCR-amplified kanamycin resistance cassette. KanR was amplified out of pET28a using primers that contained ~50bp homology arms that flanked the insertion site, enabling the localisation of the cassette to the target gene (Figure 4.8 – top). Then, this fragment was gel purified and electroporated into WT

Top10 *E. coli* (Figure 4.8 – bottom). Mutants were then propagated and screened for successful replacement of the target gene with the resistance cassette.

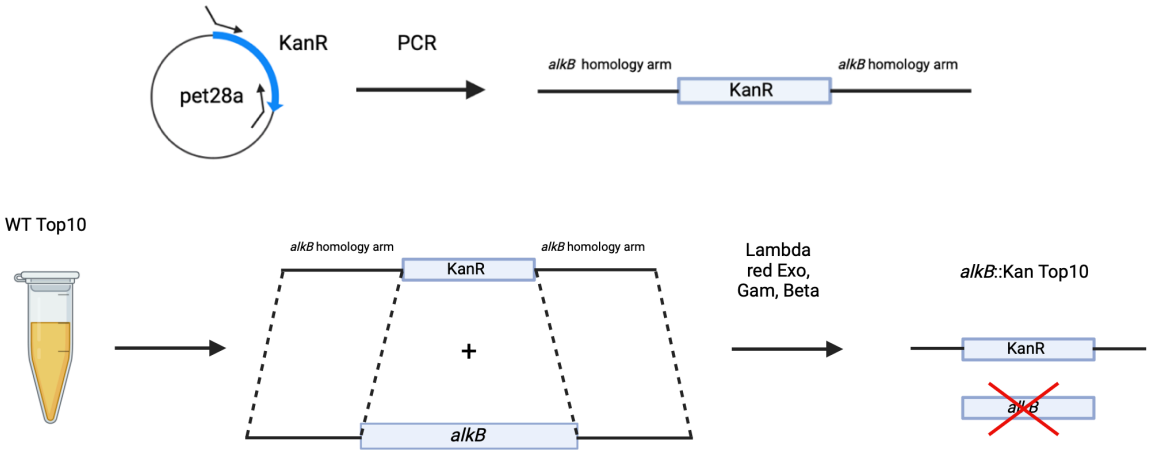
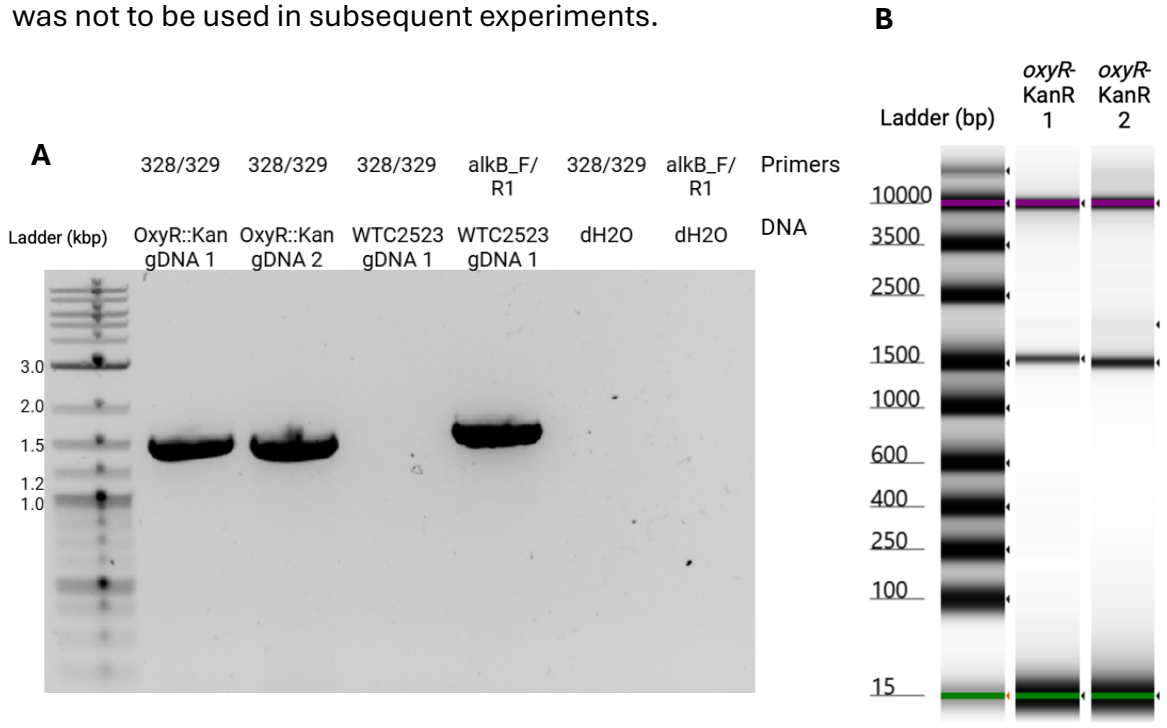


Figure 4.8: Schematic of lambda red recombination. Amplification of insertion cassette and incorporation into WT Top10 *E. coli,* resulting in replacement of *alkB* with KanR. Created in BioRender.

As a positive control for technical success of recombination, I knocked out *oxyR* in Top10 using primers that were previously used by the Uphoff lab for successful recombination (Lagage et al., 2023). The primers used were designated 328 (forward) and 329 (reverse) and were designed such that they amplified KanR out of the existing *oxyR*::Kan Keio collection mutant strain gDNA. Amplification of *oxyR*::Kan gDNA resulted in a single band of ~1.5kbp (Figure 4.9A), suggesting that the primers specifically targeted the KanR cassette. No band was present for WT C2523 gDNA, which was expected as this did not contain the resistance cassette. Note that WT C2523 and Top10 have very similar genotypes and therefore C2523 gDNA was suitable as a WT control despite recombination occurring in Top10. The ~1.5kbp bands were then extracted and run on the TapeStation to more accurately

determine band size and DNA concentration (Figure 4.9B). WT Top10 cells were transformed with pKD46 containing the lambda red recombination genes and electroporated with either *oxyR*-KanR fragment 1 or 2. Following recovery and plating, gDNA from kanamycin-resistant clones was isolated and screened to determine whether *oxyR* had been successfully replaced by the KanR cassette. The screening primers came from the Uphoff lab where they had been used previously for the same purpose (Figure 4.9C) (Lagage et al., 2023). All mutant clones (lanes 3-6, where lane 1 is the ladder) displayed a strong band at ~1.7kbp, as well as a fainter band at ~1.2kbp (Figure 4.9D). Positive control *oxyR*::Kan gDNA (lane 2) from the original Keio collection strain also displayed this band pattern, while WT C2523 gDNA displayed a single, strong band at ~1.2kbp. This suggested that the Keio collection strain was not a full mutant, and the extracted gDNA may have contained some intact WT *oxyR* with KanR inserted elsewhere. Nonetheless, the strong band at 1.7kbp for the mutant clones indicated at least partially successful *oxyR* deletion. This was sufficient for the purpose of this technical positive control, since this strain was not to be used in subsequent experiments.



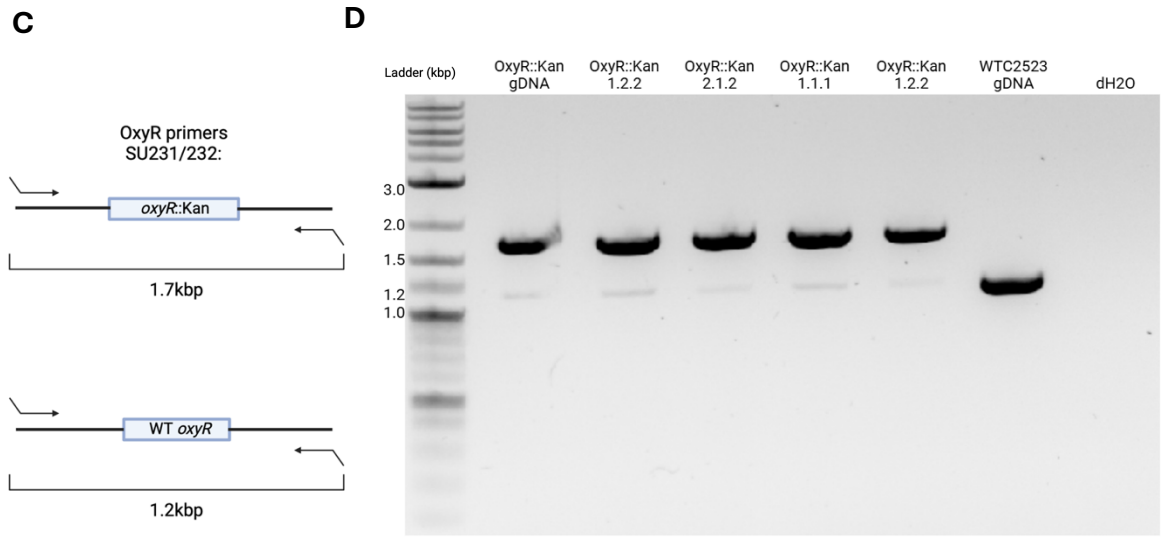
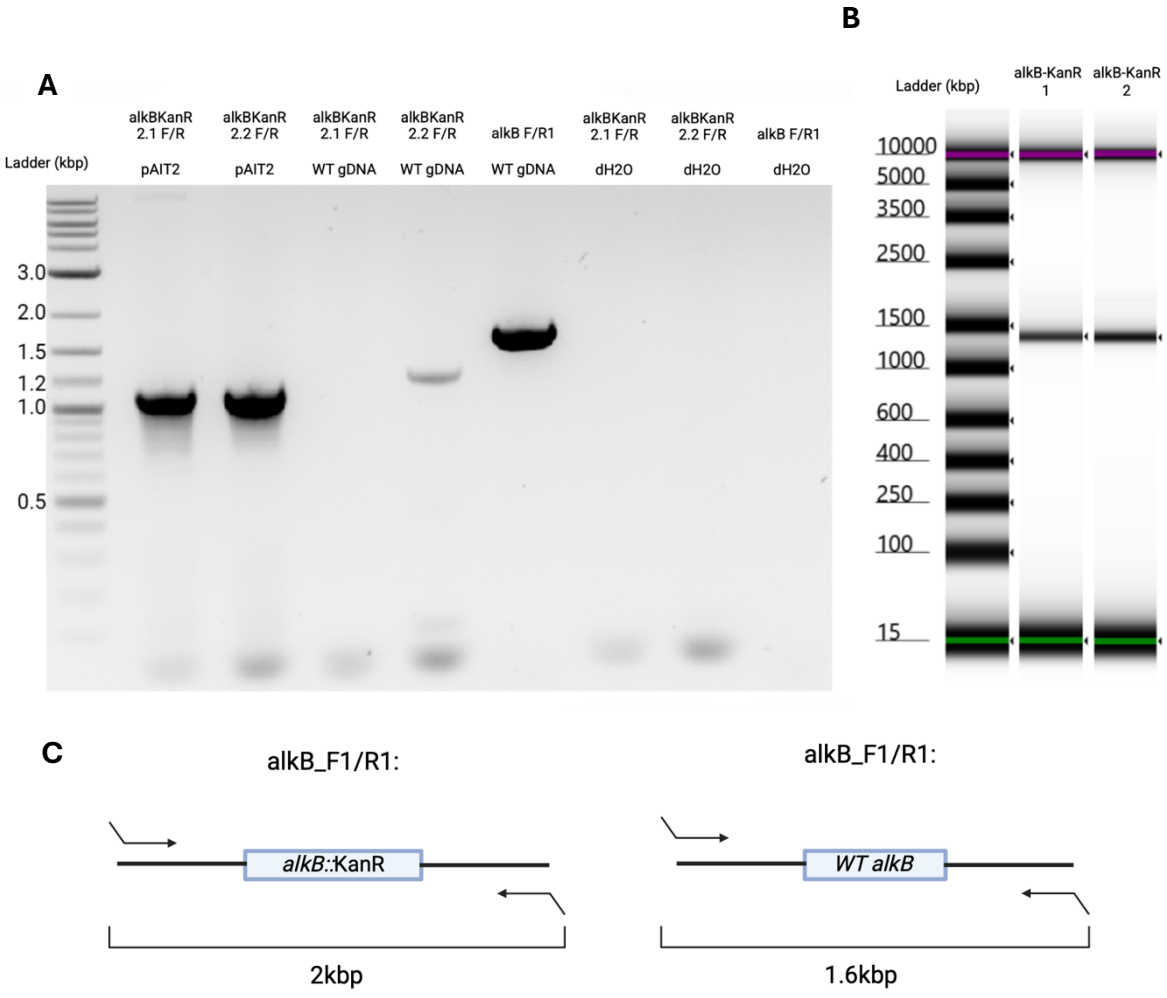


Figure 4.9: Knock out of *oxyR* in WT Top10 *E. coli* by lambda red recombination as a positive control. (A) PCR amplification of KanR from gDNA extracted from the pre-existing *oxyR*::Kan Top10 Keio collection strain. Primers 328/329 amplified KanR and were flanked by *oxyR* homology arms for future lambda red recombination. Primers alkB_F/R1 amplified *alkB* as a control for gDNA quality. **(B)** *oxyR*-KanR PCR products from lanes 2 & 3 of the gel in (B) were gel extracted and run on the TapeStation to assess band size and concentration. **(C)** Schematics of *oxyR*::Kan and WT Top10 gDNA screening by primers SU231/232 that flank *oxyR* with expected band sizes. Created in BioRender. **(D)** gDNA from kanamycin-resistant clones was extracted for PCR screening of mutants. Primers used were SU231 (forward) and SU232 (reverse), which amplified around the insertion site.

Having confirmed that the workflow was technically successful, I proceeded to replace *alkB* in Top10 with KanR, while also repeating independent replacement of *oxyR* with KanR as a positive control. Primers were designed that amplified KanR out of pAIT2 and contained *alkB* homology arms for correct localisation into the Top10 genome. These were designated alkBKanR2.1 F/R and alkBKanR2.2 F/R. Both primer pairs resulted in a band at ~1.0kbp as expected for the KanR cassette in pAIT2 (Figure 4.10A). Primer pair alkBKanR2.2 F/R also amplified a 1.2kbp band in the WT Top10 gDNA negative control. While both 1.0kbp *alkB*-KanR bands from lanes 2 & 3 were gel extracted and purified (Figure 4.10B), only the *alkB*-KanR fragment

amplified by primer pair 2.1 was used for electroporation, due to the non-specific band in the negative control with primer pair 2.2. Following electroporation, gDNA was extracted from kanamycin-resistant colonies and screened using primer pair alkB_F1/R1 (Figure 4.10C), which amplified around the insertion site. Amplification of gDNA from Δ*alkB* clones 1, 2 and 4 all yielded a band of ~2kbp as expected for successful KO of *alkB* and insertion of KanR (Figure 4.10D). Amplification by *oxyR* screening primers SU231/232 yielded a band of ~1.2kbp, indicating that WT *oxyR* remained. This served as a control for gDNA and PCR quality. Overall, these data showed that all three Δ*alkB* clones were correct KOs with complete replacement of *alkB* with KanR.



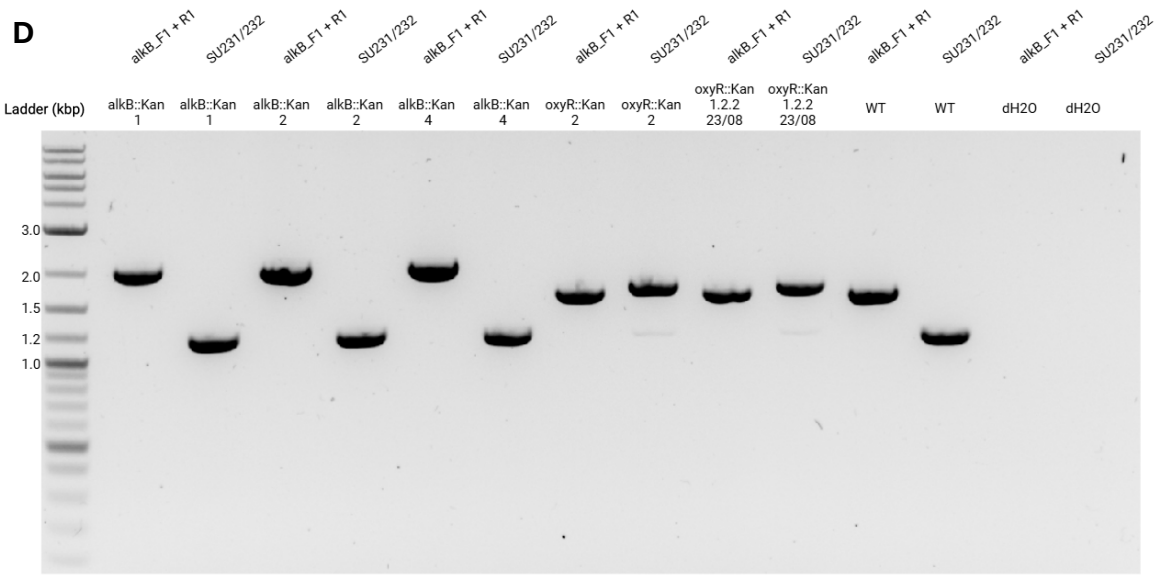


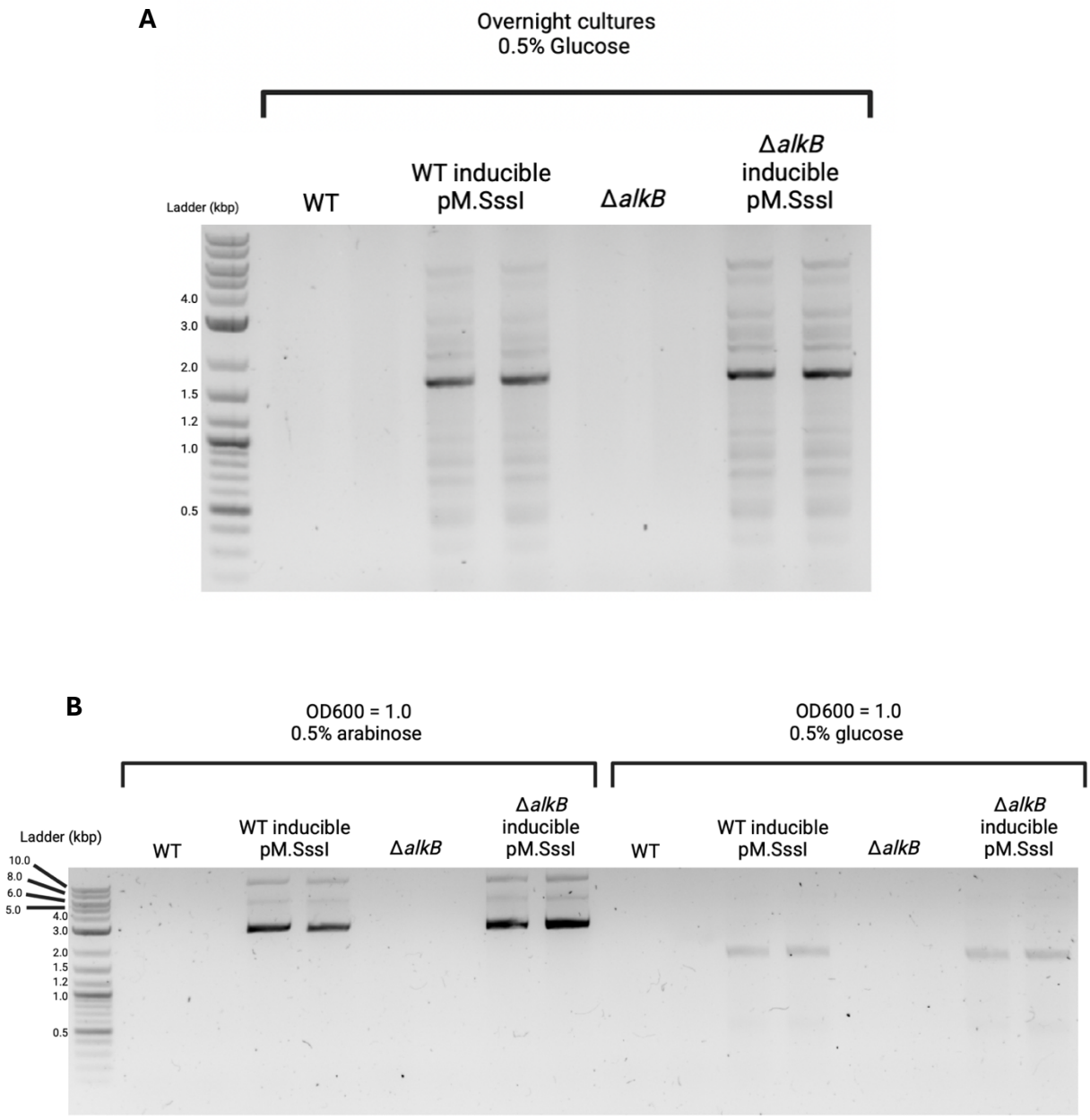
Figure 4.10: Knock out of *alkB* in WT Top10 *E. coli* by lambda red recombination. (A) PCR amplification of KanR from pAIT2. Primers alkBKanR2.1 F/R and alkBKanR2.2 F/R amplified KanR and were flanked by *alkB* homology arms for future lambda red recombination. Primers alkB_F/R1 amplified *alkB* as a control for gDNA quality. **(B)** *alkB*-KanR PCR products from lanes 2 & 3 of the gel in **(A)** were gel extracted and run on the TapeStation to assess band size and concentration. **(C)** Schematics of *alkB*::Kan and WT Top10 gDNA screening by primers alkB_F1/R1 with expected band sizes. Created in BioRender. **(D)** gDNA from kanamycin-resistant clones was extracted for PCR screening of mutants. Primers used were alkB_F1 (forward) and alkB_R1 (reverse), which amplified around the insertion site. PCR products were run on a 1% agarose gel for visualisation.

4.2.2.3. M.SssI activity and loss of AlkB caused synthetic increase in rifampicin-resistant mutation frequency in Top10 *E. coli*

With both WT and Δ*alkB* Top10 strains, it was then possible to perform the rifampicin resistance assay with the inducible pM.SssI system. WT and Δ*alkB* Top10 *E. coli* were transformed with inducible pM.SssI and plated on LB media supplemented with glucose to maintain M.SssI repression. Assay cultures were also grown up in glucose to prevent DNA methylation and limit any fitness costs. At OD600 = 0.4,

cultures were split and re-suspended in LB with either arabinose or glucose. Cultures were then grown to OD600 = 0.8-1.0 and plated on media with and without rifampicin. Rifampicin-resistant mutation frequency was calculated as described above for the constitutive system.

Plasmid DNA was extracted from overnight cultures under glucose repression, as well as from end-stage cultures treated with either glucose or arabinose. Methylation-sensitive restriction digest with HpaII showed that overnight treatment with glucose fully repressed DNA methylation, as indicated by a strong band at ~1.7kbp and multiple bands below (Figure 4.11A). This pattern was also observed for end-stage cultures grown in glucose (Figure 4.11B). DNA methylation was successfully induced by arabinose pre-plating, as indicated by a single, high band at 3.0kbp (Figure 4.11B). Rifampicin-resistant mutation frequencies are shown in Figure 4.11C. Induction of M.SssI expression significantly increased rifampicin-resistant mutation frequency (M.SssI * arabinose interaction p=2.17e-6, three-way Anova). Furthermore, the combination of Δ*alkB* and M.SssI activity resulted in a clear, synthetic increase in rifampicin-resistant mutation frequency (Δ*alkB* * M.SssI * arabinose interaction p=8.91e-5, three-way Anova). This supported the hypothesis that AlkB was responsible for the repair of M.SssI-induced damage, likely including 3mC.


A
Overnight cultures
0.5% Glucose
Ladder (kbp)
WT
WT inducible pM.SssI
ΔalkB
ΔalkB inducible pM.SssI
4.0
3.0
2.0
1.5
1.2
1.0
0.5
B
OD600 = 1.0
0.5% arabinose
OD600 = 1.0
0.5% glucose
Ladder (kbp)
WT
WT inducible pM.SssI
ΔalkB
ΔalkB inducible pM.SssI
WT
WT inducible pM.SssI
ΔalkB
ΔalkB inducible pM.SssI
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.2
1.0
0.5

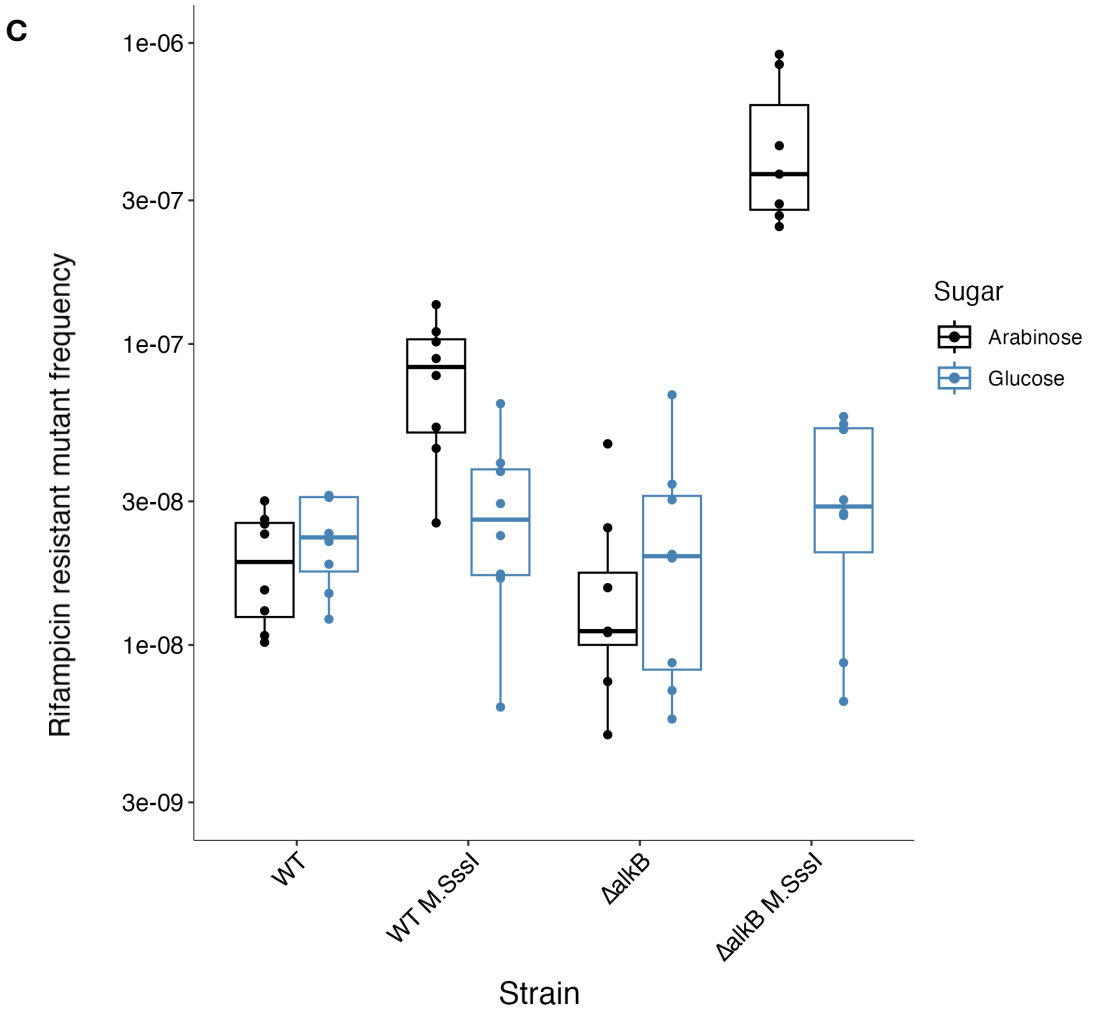
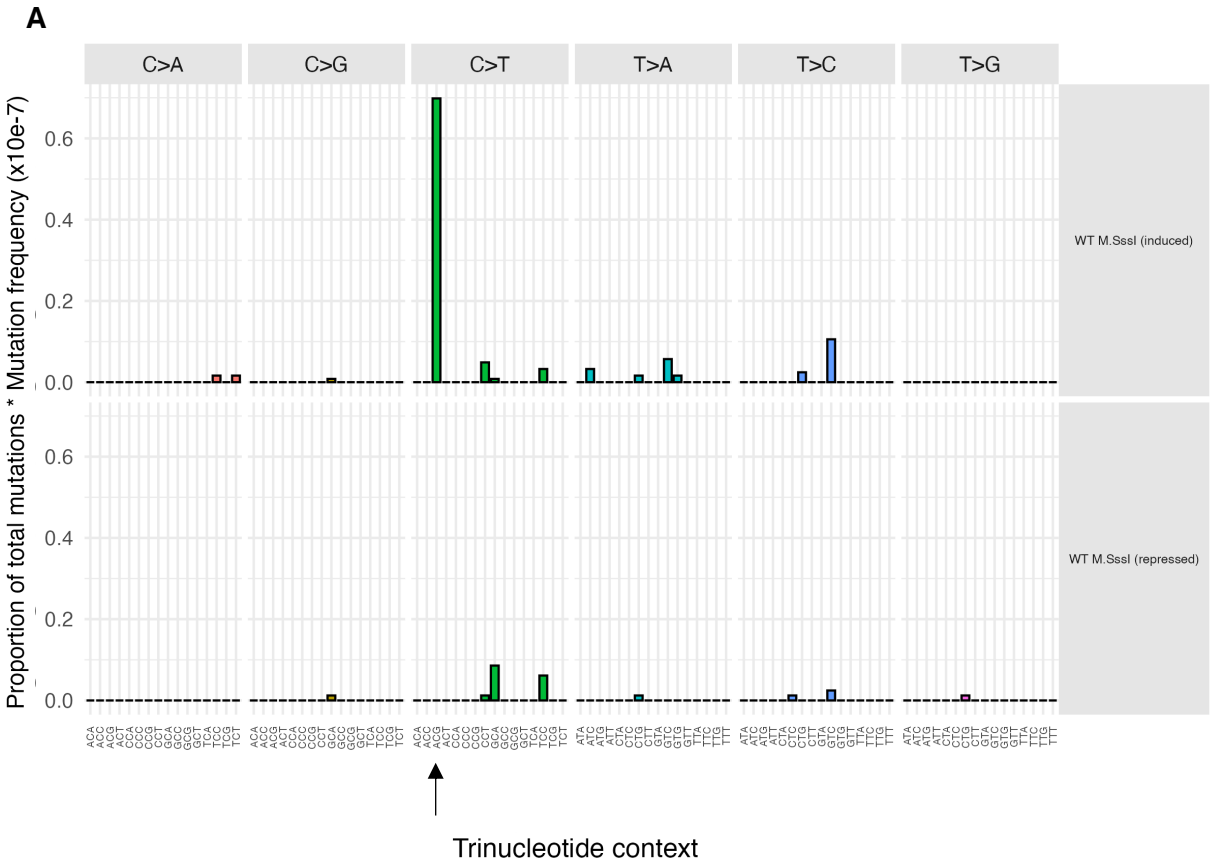


Figure 4.11: Induction of M.SssI activity in Δ*alkB* Top10 significantly increased rifampicin-resistant mutation frequency in a synthetic manner. Methylation-sensitive HpaII restriction enzyme digest of inducible pM.SssI DNA from *E. coli* extracted from overnight cultures **(A)** or endpoint cultures at OD600 = 1.0 **(B)**. Each lane shows an individual replicate clone. HpaII is blocked by methylation in CCGG context. **(C)** WT and Δ*alkB* Top10 *E. coli* were transformed with inducible pM.SssI, treated with either 0.5% arabinose (black) or glucose (blue) and plated on LB media ± rifampicin. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates.

4.2.2.4. Non-cytosine mutations were not observed with inducible DNMT activity in Top10 *E. coli*

Induction of DNA methylation in the absence of AlkB increased rifampicin-resistant mutation frequency, suggesting that there was a subset of damage induced by M.SssI that was dependent on AlkB for repair. I hypothesised that this would lead to mutations not observed in WT cells, most likely TA > CG as observed in Δ*alkB* C2523 with active pM.SssI. DNA from mutant colonies was isolated and sequenced as with the constitutive system. Induction of M.SssI activity in both WT and Δ*alkB* backgrounds increased CG > TA mutations as expected for deamination of 5mC to thymine (Figures 4.12 A & B). However, absence of AlkB did not increase non-cytosine mutations as in C2523 (Figure 4.12B). Evidently, M.SssI activity and lack of AlkB alone were not sufficient to induce this signature. The resulting question was, therefore, what was responsible for the TA > CG signature?


A
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
0.0
0.2
0.4
0.6
WT M.SssI (induced)
WT M.SssI (repressed)
Trinucleotide context

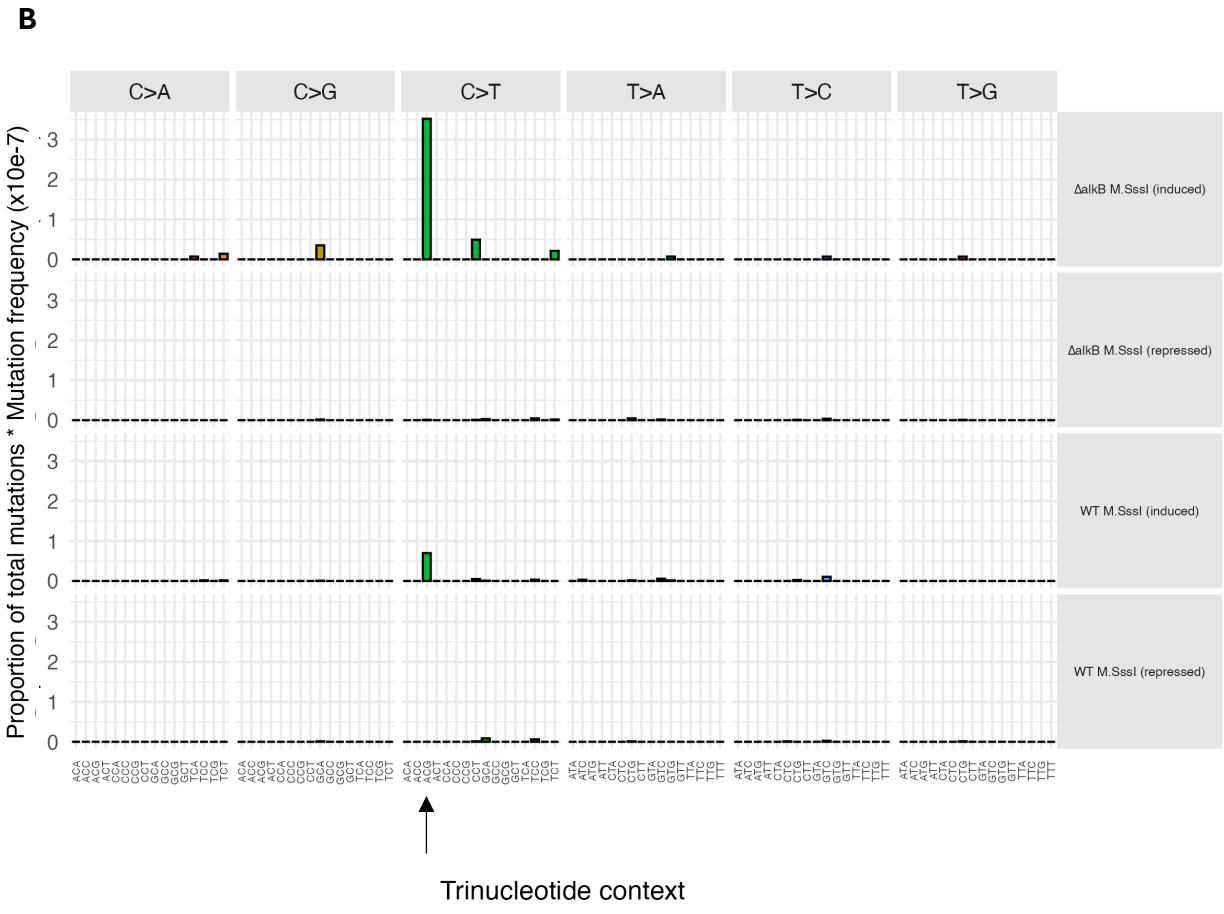

B
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
0
1
2
3
ΔalkB M.SssI (induced)
ΔalkB M.SssI (repressed)
WT M.SssI (induced)
WT M.SssI (repressed)
Trinucleotide context

Figure 4.12: Introduction of inducible CpG methylation into Top10 *E. coli* led to an increase in cytosine mutations dependent on M.SssI activity, but did not introduce previously observed Δ*alkB* and M.SssI-dependent non-cytosine mutations. DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in WT (**A**) and WT and Δ*alkB* (**B**) Top10 *E. coli* with induced and repressed M.SssI. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows**.**

4.2.3. Investigating potential involvement of TLS in the non-cytosine *rpoB*-specific mutation spectrum

4.2.3.1. Attempting to knock out *recA* in WT and Δ*alkB* C2523 by lambda red recombination

One possible explanation for the TA > CG signature appearing in C2523 but not Top10 *E. coli* is that C2523 possess functional RecA, whereas Top10 have the genotype *recA1* which impairs RecA activity. As mentioned in 1.1.4.4., RecA nucleofilament formation initiates the SOS response and subsequent error-prone bypass of lesions by TLS (Patel et al., 2010; Schlacher et al., 2006). This involves the replacement of the stalled replicative polymerase with a lower fidelity Y-family polymerase Pol IV or V that can replicate over the lesion (Goodman, 2002; Tang et al., 1998). Typically, this results in localised mutagenesis around the damage site (Napolitano et al., 2000; Tang et al., 2000). Therefore, the non-cytosine mutations

observed in Δ*alkB* C2523 might have been the result of Pol IV or V activity distal to sites of damage. It is also important to note that not all mutations would have resulted in rifampicin resistance, so only a subset would have been observed. Some mutations may also have had associated fitness costs if they hindered the activity of the polymerase. Those with a lower fitness cost may have conferred a selective advantage, particularly in conjunction with the fitness cost associated with pM.SssI, which may also have affected the distribution of observed mutations. This will be discussed further in 4.3.

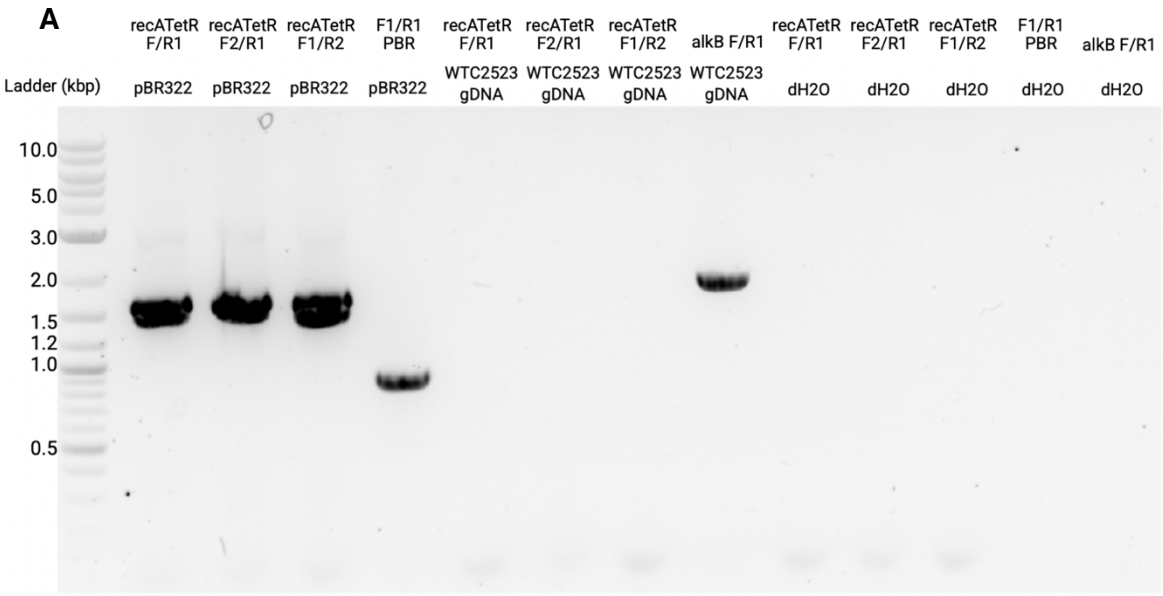
The involvement of RecA in repair of small alkylation damage is also supported by the literature. A study into DNA damage tolerance mechanisms in *D. melanogaster* revealed that REV1 mutants were hypersensitive to MMS. REV1 is a Y-family polymerase that is responsible for the recruitment of multiple translesion polymerases to damage sites in response to alkylation damage. Therefore, this work implicated TLS in bypass of alkylation damage in flies (Khodaverdian et al., 2024). Furthermore, a systematic quantification of the contributions of genotoxin exposure and DNA repair deficiency to mutational signatures in *C. elegans* found that error-prone TLS was responsible for the majority of genotoxin-induced mutagenesis. In particular, error-prone bypass following MMS treatment increased TA > CG and TA > AT mutations (Volkova et al., 2020). Finally, a study into how repair was influenced by the DNA strand on which damage occurred found that translesion polymerases were recruited to small alkylation damage on both strands, again implicating TLS in tolerance of alkylation lesions (Anderson et al., 2024). Overall, the involvement of

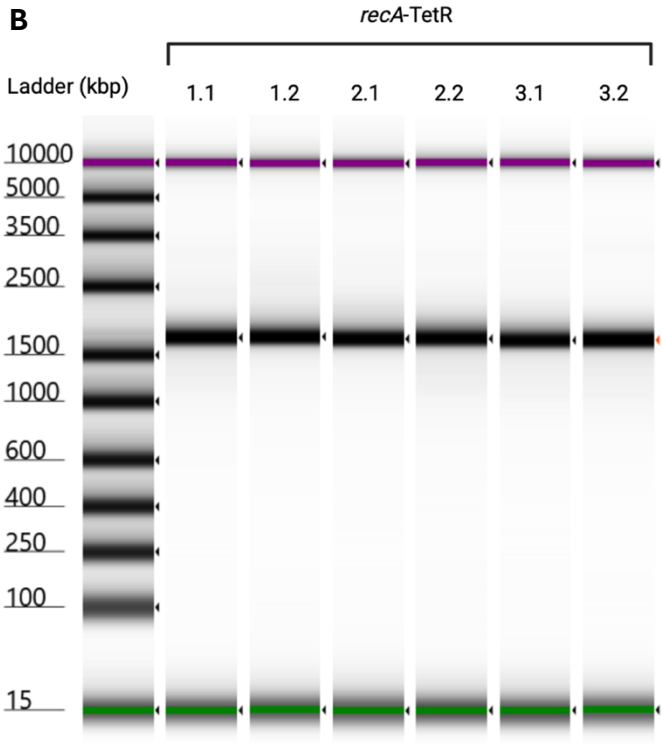
RecA seemed a well-supported place to start the investigation into these Δ*alkB* and M.SssI-dependent non-cytosine mutations.

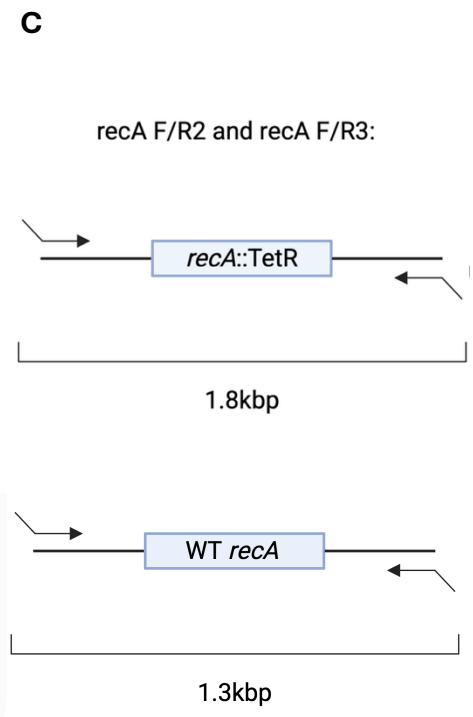
I hypothesised that if RecA was removed from C2523 *E. coli*, then they would no longer be able to initiate bypass of damage by TLS and would lose the TA > CG signature. To remove RecA from WT and Δ*alkB* C2523 I again designed two sets of primers for lambda red recombination (recATetR F1/R1 and F2/R2). These primers were designed to amplify the tetracycline resistance cassette TetR out of pBR322 with RecA homology arms for insertion into C2523. Three different combinations of primers were possible for amplification of TetR. All combinations resulted in a single band of expected 1.6kbp size (Figure 4.13A, lanes 2-4). It did appear as though there were two bands for each TetR amplification at 1.5kbp and 1.6kbp, but TapeStation analysis following gel extraction confirmed the presence of a single band at 1.6kbp (Figure 4.13B).

RecA-TetR fragment 1.1 (extracted from lane 2 of Figure 4.13A and analysed on lane 2 of Figure 4.13B) was used for electroporation of WT C2523 *E. coli*. Tetracycline-resistant colonies were screened by PCR, using the screening primers shown in Figure 4.13C which were designed to amplify around *recA*. All Δ*recA* clones yielded a single band at ~1.3kbp (Figure 4.13D) with both sets of primers. Screening of WTC2523 gDNA with the same primer pairs also resulted in a band of ~1.3kbp. Due to bending of bands on the gel, PCR products from screening were also run on the TapeStation to better resolve band size. TapeStation analysis confirmed presence of WT *recA* in tetracycline-resistant clones (Figure 4.13E). Expected band size for

correct mutants was ~1.8kbp, which was a sufficient difference to WT *recA* such that it would be visible on a gel or TapeStation. This suggested that the TetR cassette was not correctly localised to *recA* and had instead inserted elsewhere in the genome.







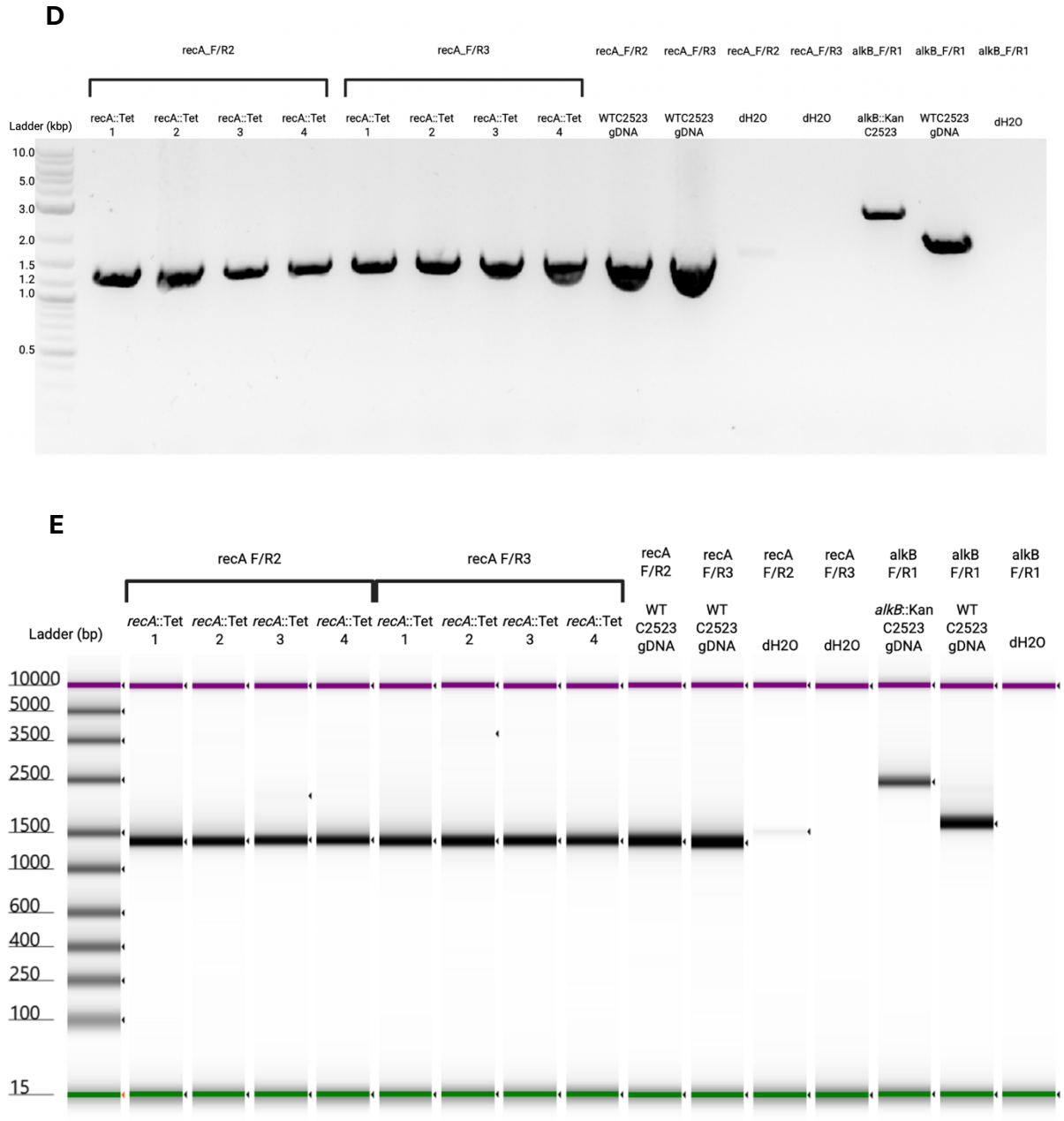


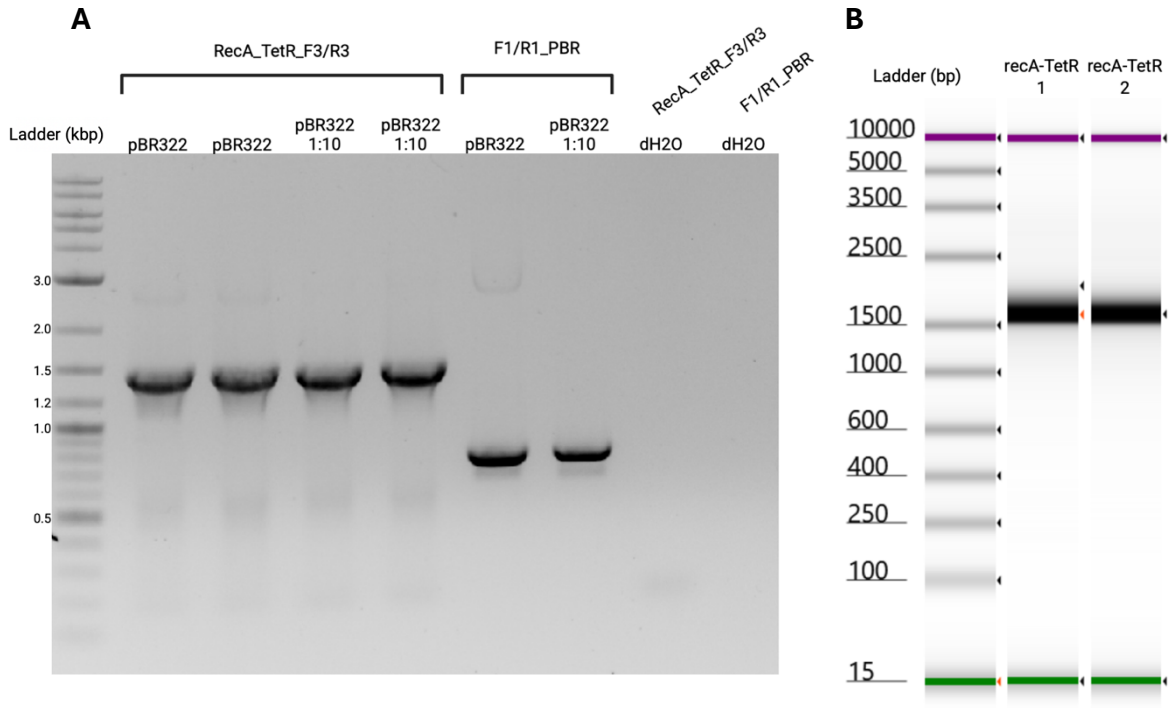
Figure 4.13: KO of *recA* in WT C2523 *E. coli* by lambda red recombination. (A) PCR amplification of TetR from pBR322. Primers recATetR F/R1 and recATetR F/R2 amplified TetR and were flanked by *recA* homology arms for future lambda red recombination. Primers F/R1 PBR amplified part of pBR322 as a control for plasmid DNA quality. Amplification primers were also tested on WT C2523 gDNA for off target binding. **(B)** recA-TetR PCR products from lanes 1-3 (excluding ladder) of the gel in (A) were gel extracted and run on the TapeStation to assess band size and concentration. Each gel extraction was run over two lanes on the TapeStation. **(C)** Schematics of theoretical *recA*::Tet and WT C2523 gDNA screening by primers recA F/R2 and recA F/R3 with expected band sizes. Created in

BioRender. **(D)** gDNA from tetracycline-resistant clones was extracted for PCR screening of mutants with primer pairs recA F/R2 and recA F/R3, which amplified around the insertion site. PCR products were run on a 1% agarose gel for visualisation. **(E)** PCR products from mutant screening as in **(D)** were run on a genomic DNA TapeStation to more accurately determine band sizes.

I re-designed the amplification primers so that the *recA* homology arms were longer (68bp instead of 50bp), to increase the likelihood that the recA-TetR fragment was inserted correctly into *recA*. I then repeated the lambda red protocol as before. TetR amplification out of pBR322 by new primer pair RecATetRF/R3 resulted in a band of ~1.4kbp as expected (Figure 4.14A). This band was gel extracted from lanes 2 – 5; lanes 2 & 3 and 4 & 5 respectively were pooled. TapeStation analysis indicated a slightly larger band than expected at 1.6kbp (Figure 4.14B). RecA-TetR fragment 2 was used for electroporation of WT C2523 *E. coli*, with an *alkB* KO performed alongside as a positive control. While the alkB-KanR fragment was originally intended for *alkB* KO in WT Top10 *E. coli,* the highly conserved nature of the *ada* operon meant that it should have been possible to use the fragment for *alkB* KO in C2523 as well.

In order to more accurately determine the incorporation of TetR into *recA*, the mutant screening protocol was also adjusted. As well as primers that amplified around the insertion site, a primer pair was used where one primer bound outside of the insertion site (recA F2), and the other within the TetR cassette (R1 PBR) (Figure 4.14C). Amplification around the target insertion site resulted in a single band of ~1.2kbp for all four tetracycline-resistant mutant clones (Figure 4.14D). This was

close to the expected band size of 1.3kbp for WT *recA,* which was observed in WT and Δ*alkB* C2523 gDNA. Additionally, none of the clones gave a band with the recA F2 and R1 PBR primer pair, indicating that the TetR cassette was not inserted into *recA* and was likely too far away in the genome to allow for DNA amplification between these two primers. The Δ*alkB* positive control gave a band at ~1.8kbp. While expected band size for mutants was 2kbp, it appeared to be slightly higher than the WT band, which sat at 1.5kbp compared to expected 1.66kbp. Despite bands running slightly lower than expected on this gel, this indicated successful insertion of KanR in *alkB*. Therefore, the failure of TetR to insert into *recA* was likely not due to faulty recombination but mis-localisation in the genome. It was also possible that knocking out *recA* was lethal for cells, therefore clones were only viable if TetR incorporated elsewhere. Indeed, Top10 possess a point mutation in *recA* , not a full KO. Given that KO of *recA* failed in WT C2523, I did not attempt KO in Δ*alkB* C2523.



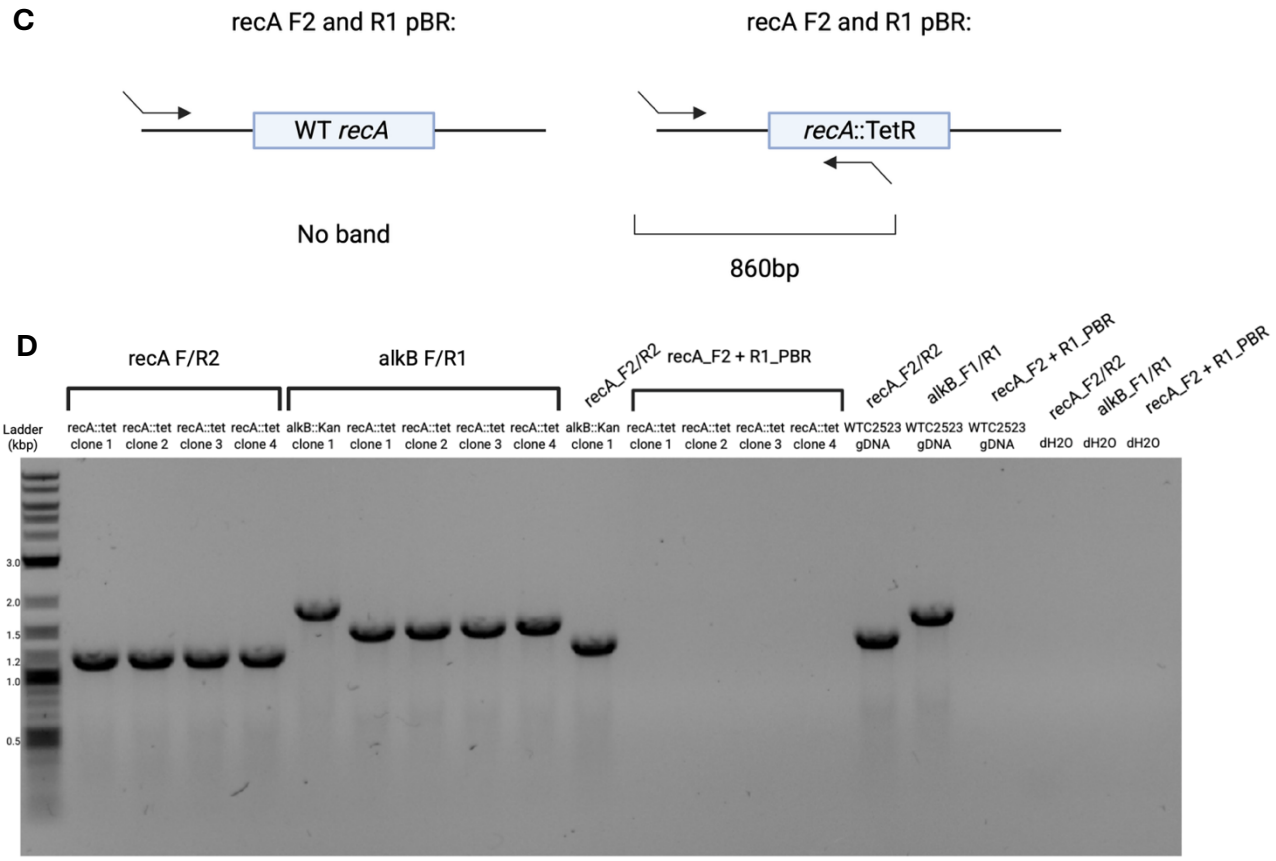


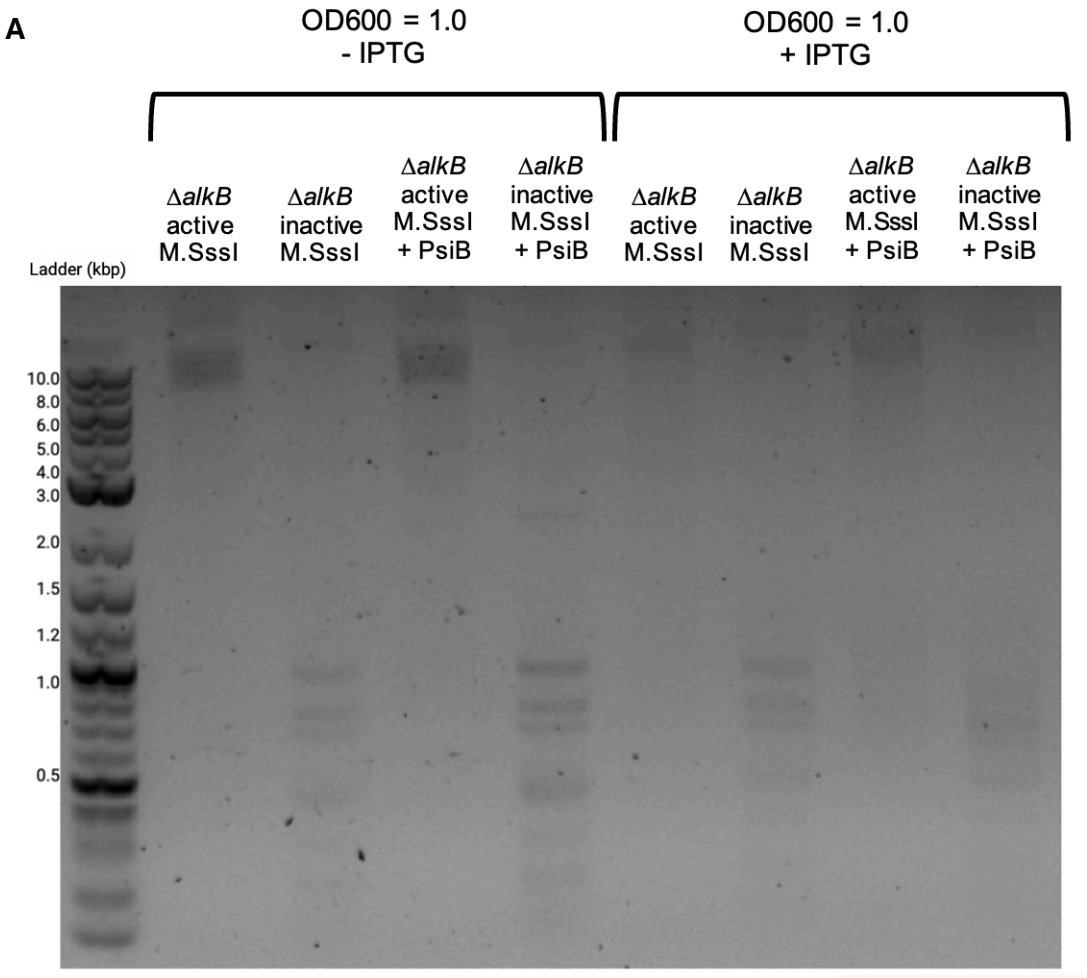
Figure 4.14: Repeat of KO of *recA* in WT C2523 *E. coli* by lambda red recombination. (A) PCR amplification of TetR from pBR322. Primers RecATetR F/R3 amplified TetR and were flanked by 68bp *recA* homology arms for future lambda red recombination. Primers F/R1 PBR amplified part of pBR322 as a control for plasmid DNA quality. **(B)** recA-TetR PCR products from lanes 2 & 3 and 4 & 5 respectively of the gel in (A) were gel extracted and run on the TapeStation to assess band size and concentration. **(C)** Schematics of theoretical *recA*::Tet and WT C2523 gDNA screening by primers recA F/R2 and recA F2 + R1 PBR with expected band sizes. Created in BioRender. **(D)** gDNA from tetracycline-resistant clones was extracted for PCR screening of mutants with primer pairs recA F/R2, which amplified around the insertion site, and recA F2 + R1 PBR, where R1 PBR binds inside the TetR cassette. PCR products were run on a 1% agarose gel for visualisation.

4.2.3.2. Inhibition of RecA with PsiB had no discernable effect on rifampicin-resistant mutation frequency

Following unsuccessful KO of *recA,* the next logical option was to inhibit RecA activity. To do this, I utilised a plasmid encoding *psiB* (pPsiB). PsiB inhibits activation of RecA, therefore preventing SOS response induction (Petrova et al., 2009). The gene was under the control of a P_{LAC} promoter, allowing inducible expression of PsiB with IPTG. WT C2523 were transformed with both the constitutively active or inactive pM.SssI and pPsiB. Cultures were grown in the absence of IPTG until D600 = 0.4, then cultures were split and treated with either 1mM IPTG or dH_2O. At OD600 = 1.0, cultures were plated on LB media with and without rifampicin without IPTG.

Rifampicin resistant mutation frequencies are shown for WT and Δ*alkB* C2523 with active pM.SssI ± pPsiB (Figure 4.15B). I hypothesised that induction of PsiB activity by IPTG and subsequent inhibition of RecA activation would decrease rifampicin-resistant mutation frequency as error-prone TLS was inhibited. HpaII digests confirmed that induction of PsiB had no effect on the ability of active M.SssI to methylate DNA (Figure 4.15A). However, induction of PsiB had no discernible effect on rifampicin-resistant mutation frequency in either WT or Δ*alkB* C2523 (Figure 4.15B). One possible explanation for this was that RecA was not involved in mutagenesis of M.SssI-induced DNA damage. Interestingly, treatment with either 0mM or 1mM IPTG did not affect rifampicin-resistant mutation frequency for samples containing pPsiB. This was surprising given that *psiB* was under the control of an IPTG-inducible promoter. However, the Uphoff lab previously noted some

constitutive expression of PsiB, which may explain why IPTG addition seemingly had no effect on rifampicin-resistant mutation frequency. These data supported this hypothesis. To further investigate the role of RecA in DNMT-induced mutagenesis, I proceeded to determine qualitative changes in *rpoB*-specific mutation spectra as a result of pPsiB addition.



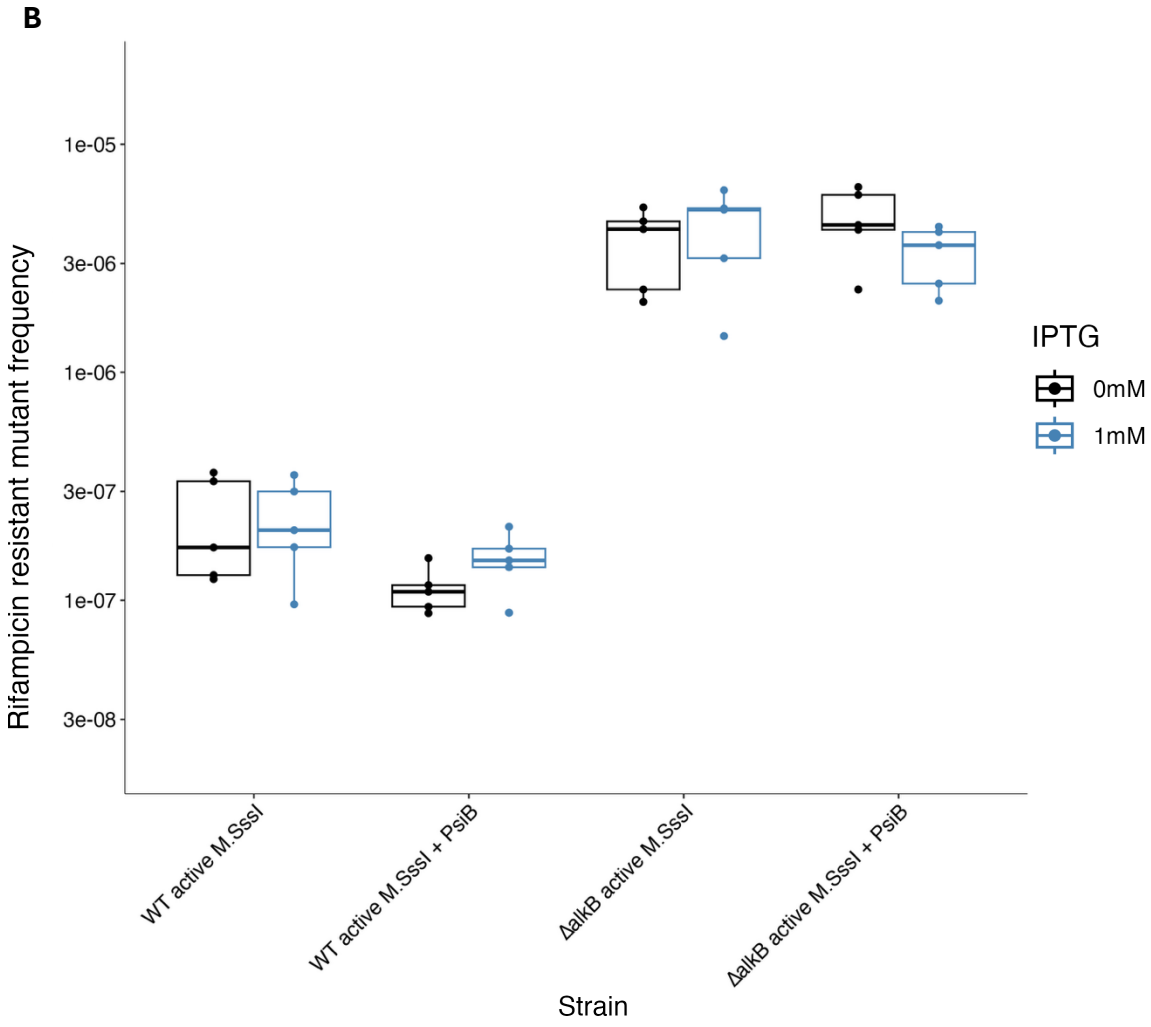
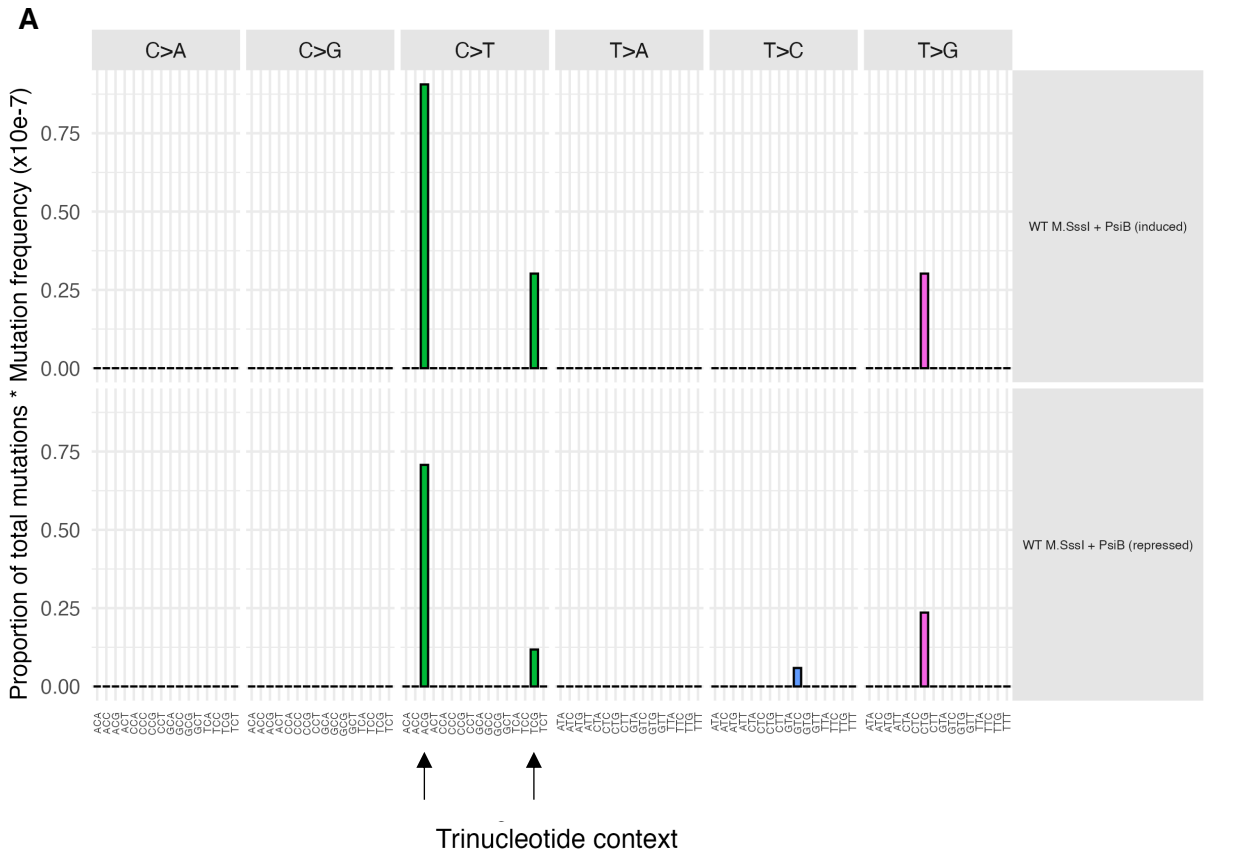


Figure 4.15: Inhibiting RecA in C2523 *E. coli* with constitutive M.SssI expression had no discernible effect on rifampicin-resistant mutation frequency. (A) Methylation-sensitive HpaII restriction enzyme digest of plasmid DNA from *E. coli* containing active or inactive pM.SssI ± pPsiB, which encodes an inhibitor of RecA. Plasmid DNA was extracted at OD600 = 1.0 from samples treated with either 1mM or 0mM (dH_2O) IPTG. HpaII is blocked by methylation in CCGG context. **(B)** WT and Δ*alkB* C2523 *E. coli* were transformed with active or inactive pM.SssI ± pPsiB. Samples were treated with either 1mM or 0mM (dH_2O) IPTG. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates. Rifampicin-resistant mutation frequencies for transformants with active M.SssI only are shown here for simplicity.

4.2.3.3. Inhibition of RecA did not eliminate the non-cytosine *rpoB*-specific mutation spectrum in C2523 *E. coli*.

Given that inhibition of RecA had no effect on rifampicin-resistant mutation frequencies, I hypothesised that it would also have no effect on *rpoB*-specific mutation spectra. Therefore, the I expected to still observe the TA > CG signature originally observed in Δ*alkB* C2523 *E. coli* (Figure 4.4). Mutant colonies were picked and *rpoB* amplified by PCR. Sequencing results of WT C2523 and Δ*alkB* C2523 *E. coli* with active pM.SssI ± pPsiB are shown in Figure 4.16. As expected, WT C2523 with active M.SssI displayed predominantly CG > TA mutations, with signatures unaffected by presence of pPsiB or IPTG treatment (Figure 4.16A). Δ*alkB* C2523 also displayed CG > TA mutations, as well as strong TA > CG peaks (Figure 4.16B). The intensity and distribution of these mutations was unaffected by presence or induction of PsiB. This was consistent with the proposed constitutive expression of PsiB, but more importantly with the hypothesis that these non-cytosine mutations were RecA-independent.


A
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
0.75
0.50
0.25
0.00
WT M.SssI + PsiB (induced)
WT M.SssI + PsiB (repressed)
Trinucleotide context

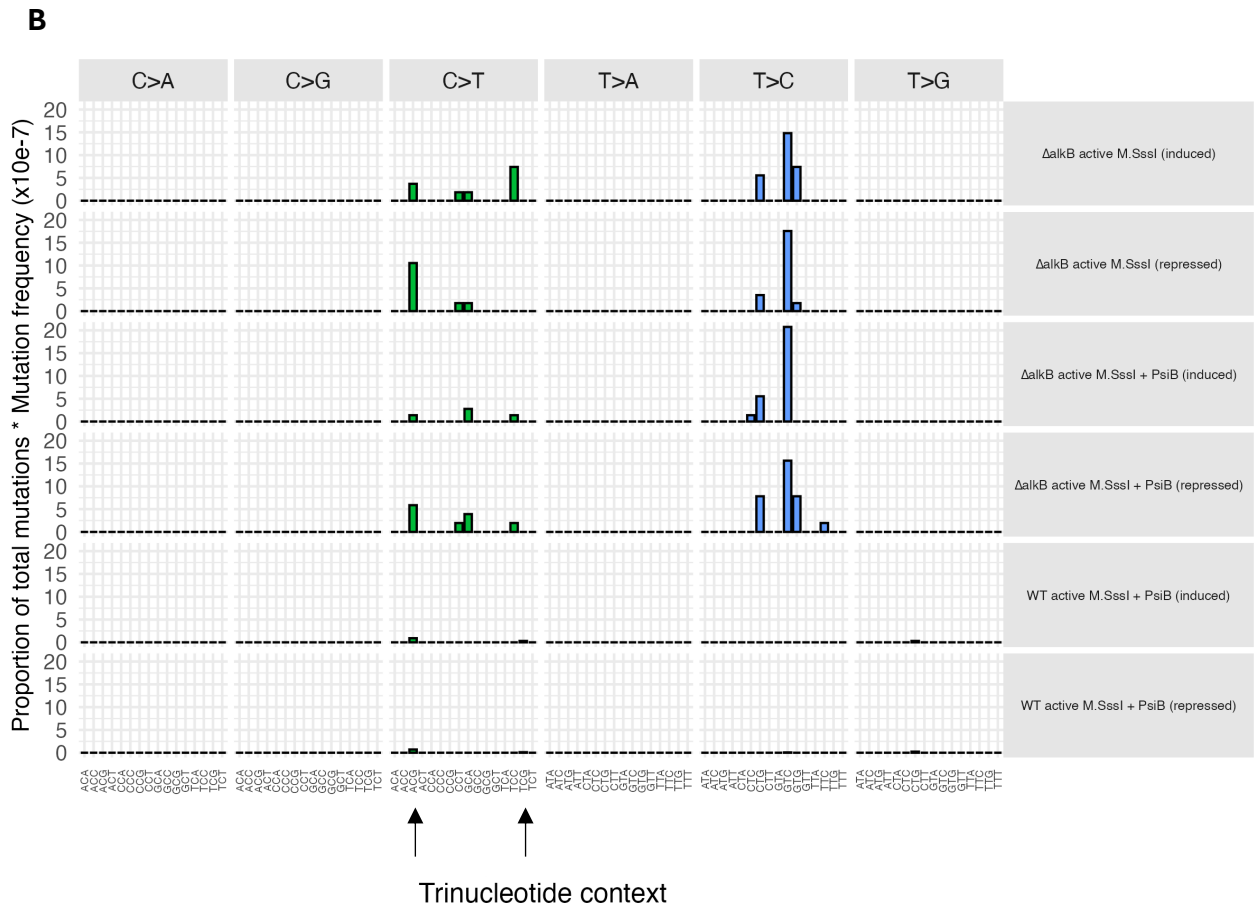

B
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
20
15
10
5
0
ΔalkB active M.SssI (induced)
ΔalkB active M.SssI (repressed)
ΔalkB active M.SssI + PsiB (induced)
ΔalkB active M.SssI + PsiB (repressed)
WT active M.SssI + PsiB (induced)
WT active M.SssI + PsiB (repressed)
Trinucleotide context

Figure 4.16: Inhibition of RecA in C2523 *E. coli* with constitutive M.SssI expression did not eliminate the non-cytosine *rpoB*-specific mutation spectrum. DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in WT (**A**) and WT and Δ*alkB* (**B**) C2523 *E. coli* with constitutively active pM.SssI ± pPsiB. Separate signatures are shown for +IPTG (induced) and –IPTG (repressed) conditions given that *psiB* is under the control of an IPTG-inducible P_{LAC} promoter. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows**.**

4.2.3.4. Introduction of RecA into Top10 *E. coli*

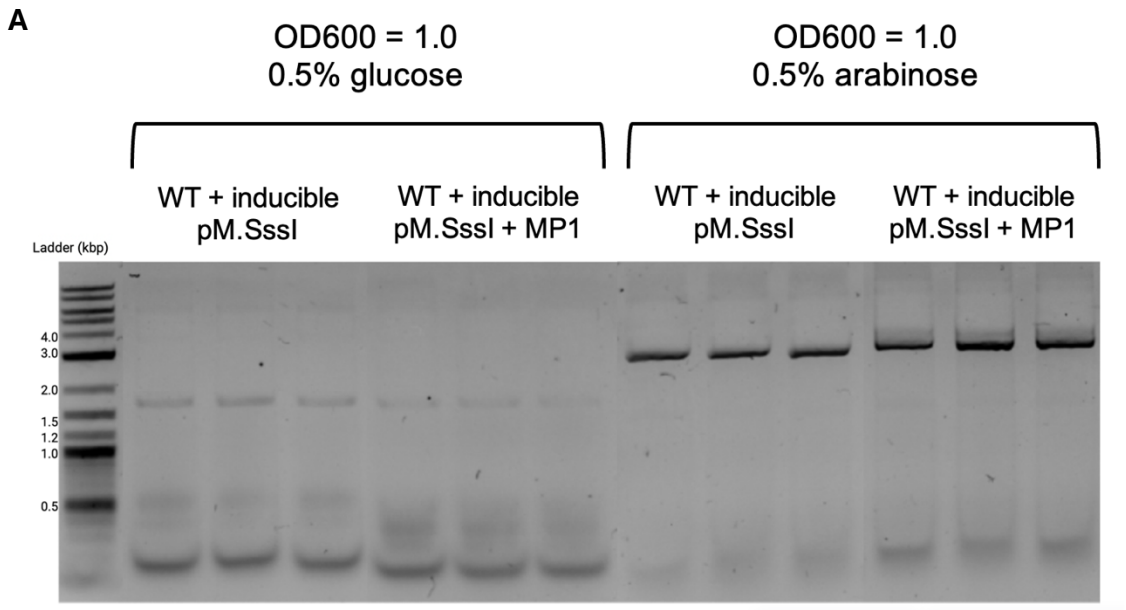
The data above point to the fact that RecA was not responsible for the TA > CG *rpoB*-specific mutation spectrum. However, it was also possible that PsiB expression was not sufficient to fully inhibit RecA, which would explain why there was no change in *rpoB*-specific mutation spectra with addition of IPTG. To fully test the involvement of RecA, I introduced RecA back into Top10 on a plasmid called pMP1. This was obtained from Addgene (see Materials and Methods). pMP1 encodes *recA* under the control of an arabinose-inducible promoter, as well as *umuDC* which are required for TLS bypass of lesions (Goodman, 2002). Therefore, pMP1 could be used alongside the inducible M.SssI system for induction of TLS in Top10, essentially re-creating the C2523 genotype.

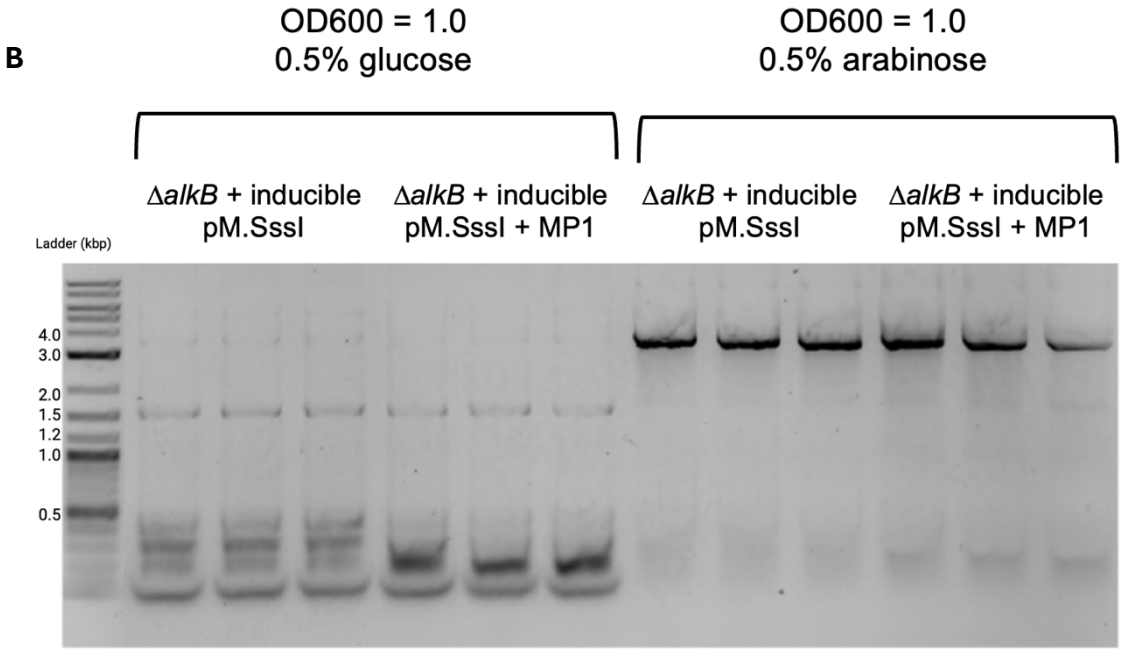
WT and Δ*alkB* Top10 were transformed with inducible pM.SssI and pMP1 and grown in 0.5% glucose to ensure repression of both M.SssI and RecA. At OD600 = 0.4,

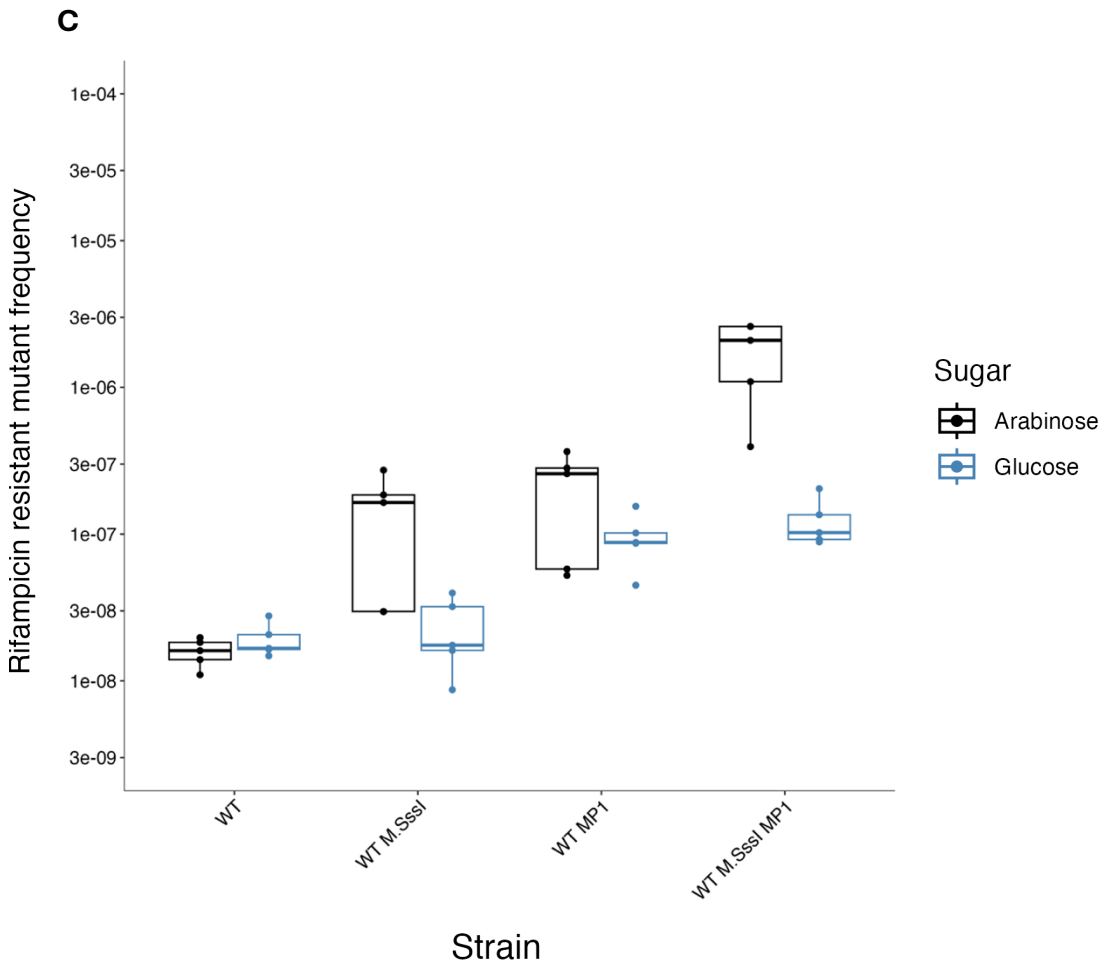
cultures were split, centrifuged and re-suspended in LB media + 0.5% arabinose or glucose. Cultures were plated on LB media ± rifampicin and the number of mutant colonies compared to total colonies. Induction of methylation by arabinose was assessed via HpaII digest of plasmid DNA. Arabinose successfully induced DNA methylation in both WT and Δ*alkB* Top10 + inducible pM.SssI (Figure 4.17A, B). Further, arabinose-induced pM.SssI methylated DNA when co-transformed with pMP1 (Figure 4.17A, B). This suggested that pMP1 did not affect the ability of M.SssI to methylate DNA.

Rifampicin-resistant mutation frequencies were calculated for WT and Δ*alkB* Top10 ± inducible pM.SssI ± pMP1 (Figure 4.17C and D respectively). Induction of M.SssI expression by arabinose in both WT and Δ*alkB* Top10 led to an increase in rifampicin-resistant mutation frequency, with the increase in Δ*alkB* exceeding that observed in WT. This was consistent with the synthetic effect of M.SssI induction on rifampicin-resistant mutation frequency in Δ*alkB* Top10 observed in Figure 4.11B. The presence of pMP1 caused an increase in rifampicin-resistant mutation frequency regardless of M.SssI expression or arabinose treatment in both WT and Δ*alkB* Top10. This suggested that expression of RecA and UmuC/D was largely constitutive, although it could be argued that induction by arabinose caused a small increase in rifampicin-resistant mutation frequency compared to glucose repression. Of course, the key rifampicin-resistant mutation frequencies to look at were those of WT and Δ*alkB* Top10 transformed with both inducible pM.SssI and pMP1. In both backgrounds, the glucose-repressed rifampicin-resistant mutation frequency was higher than that for Top10 ± inducible pM.SssI, but was equal to that of Top10 + pMP1 only. This

suggested that constitutive expression of RecA and UmuC/D from pMP1 raised baseline rifampicin-resistant mutation frequency, potentially due to error-prone bypass of spontaneous damage. For the double transformants in both backgrounds, induction by arabinose led to an increase in rifampicin-resistant mutation frequency compared to repression under glucose. However, this increase was no greater than that observed for induction of pM.SssI alone. It is possible that the combination of pM.SssI and pMP1 induction in WT and Δ*alkB* Top10 was additive and did not cause a synthetic increase in rifampicin-resistant mutation frequency, although this was not tested statistically. This suggested that RecA was not involved in tolerance of M.SssI-induced DNA damage in the absence of AlkB.




B
OD600 = 1.0
0.5% glucose
OD600 = 1.0
0.5% arabinose
ΔalkB + inducible pM.SssI
ΔalkB + inducible pM.SssI + MP1
ΔalkB + inducible pM.SssI
ΔalkB + inducible pM.SssI + MP1
Ladder (kbp)
4.0
3.0
2.0
1.5
1.2
1.0
0.5


C
Rifampicin resistant mutant frequency
1e-04
3e-05
1e-05
3e-06
1e-06
3e-07
1e-07
3e-08
1e-08
3e-09
WT
WT M.SssI
WT MP1
WT M.SssI MP1
Strain
Sugar
Arabinose
Glucose

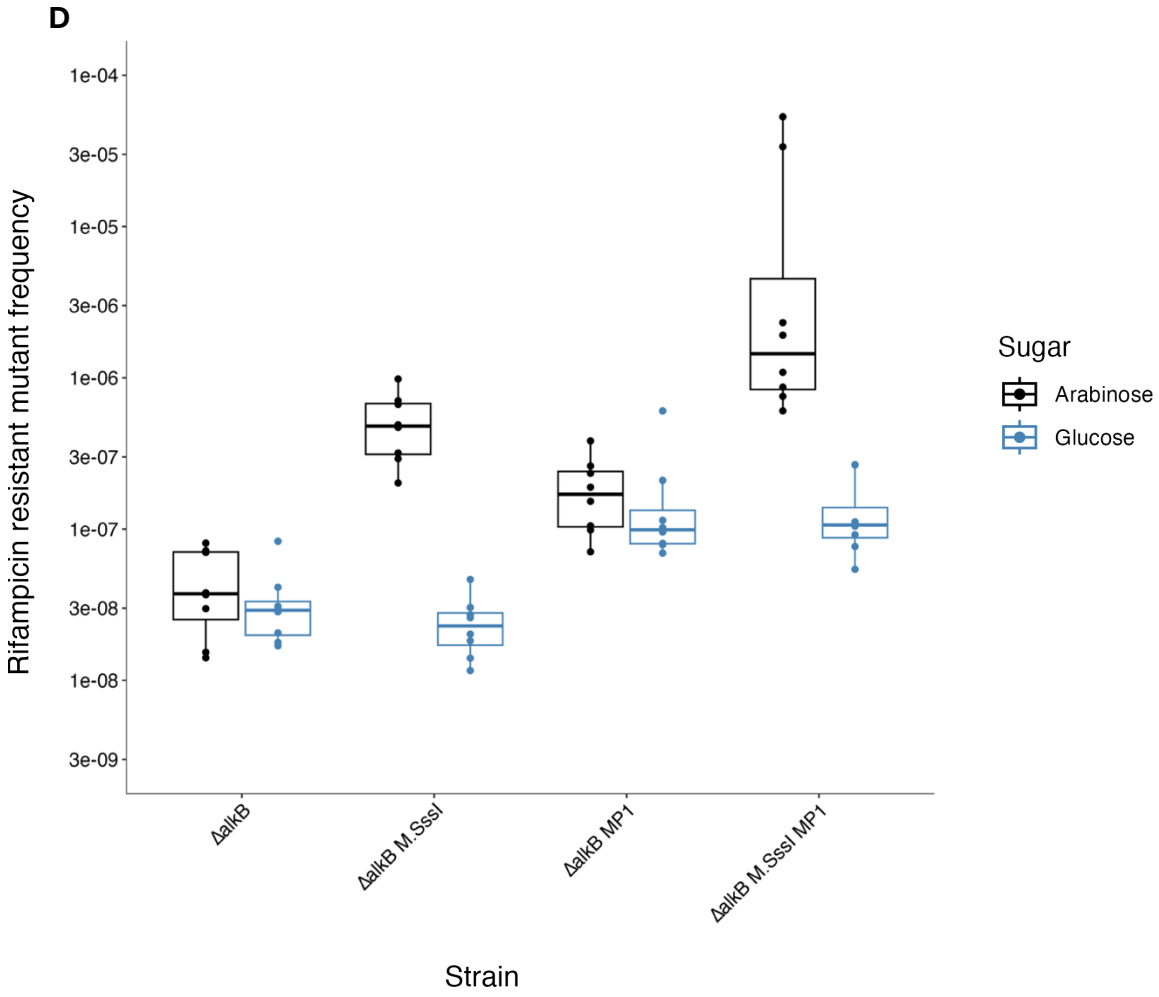
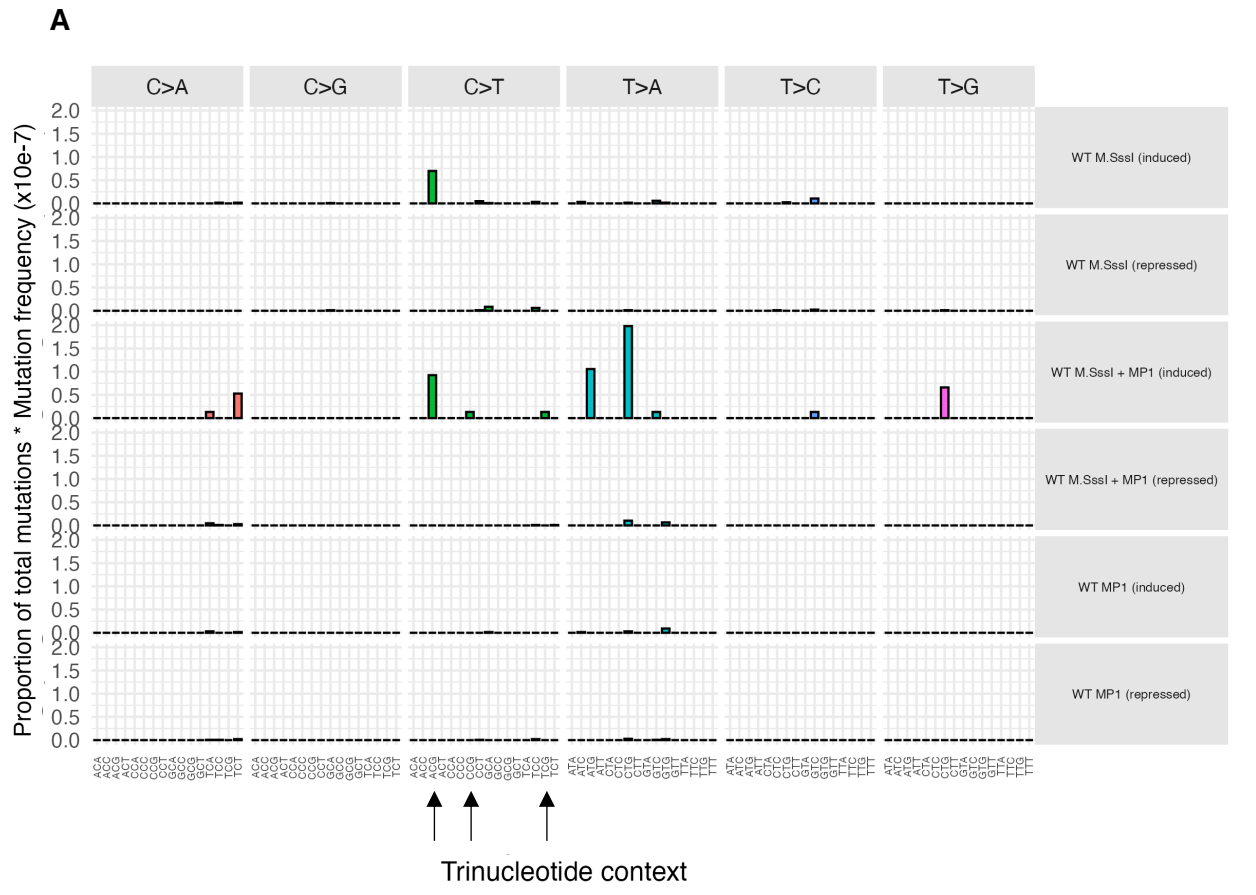


Figure 4.17: Introduction of RecA into Top10 *E. coli* with inducible M.SssI additively increased rifampicin-resistant mutation frequency. Methylation-sensitive HpaII restriction enzyme digest of plasmid DNA from WT **(A)** or Δ*alkB* **(B)** Top10 E. *coli* containing inducible pM.SssI ± pMP1, which encodes *recA* and *umuC/D*. Plasmid DNA was extracted at OD600 = 1.0 from samples treated with either 0.5% arabinose or glucose. HpaII is blocked by methylation in CCGG context. Each lane shows inducible pM.SssI ± pMP1 DNA extracted from individual replicate clones. WT **(C)** and Δ*alkB* **(D)** Top10 *E. coli* were transformed with inducible pM.SssI ± pMP1. Samples were treated with either 0.5% arabinose or glucose. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates.

4.2.3.5. Introduction of RecA into Δ*alkB* Top10 did not rescue non-cytosine mutations

Given that induction of genes in MP1 did not cause a marked increase in rifampicin-resistant mutation frequency in Δ*alkB* Top10 with inducible M.SssI activity, I suspected that it would not rescue the TA > CG signature either. DNA from rifampicin-resistant mutant colonies was isolated and sequenced as before. *RpoB*-specific mutation spectra for WT and Δ*alkB* Top10 are shown in Figures 4.18A and B respectively. Induction of methylation led to an increase in CG > TA mutations as expected. Interestingly, the combination of induced pM.SssI and pMP1 led to an increase in CG > AT, TA > AT and TA > GC mutations in an Δ*alkB*-independent manner. These were not observed to the same extent in WT or Δ*alkB* Top10 – pM.SssI + pMP1. This suggested that these mutations arose as a result of TLS bypass over M.SssI-induced damage that was not the target of AlkB. Most notably, there was no evidence of TA > CG mutations in Δ*alkB* Top10 with induced pM.SssI and pMP1, suggesting that this signature was independent of RecA. This supported previous data where inhibition of RecA in C2523 did not eliminate non-cytosine mutations.


A
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
WT M.SssI (induced)
WT M.SssI (repressed)
WT M.SssI + MP1 (induced)
WT M.SssI + MP1 (repressed)
WT MP1 (induced)
WT MP1 (repressed)
Trinucleotide context

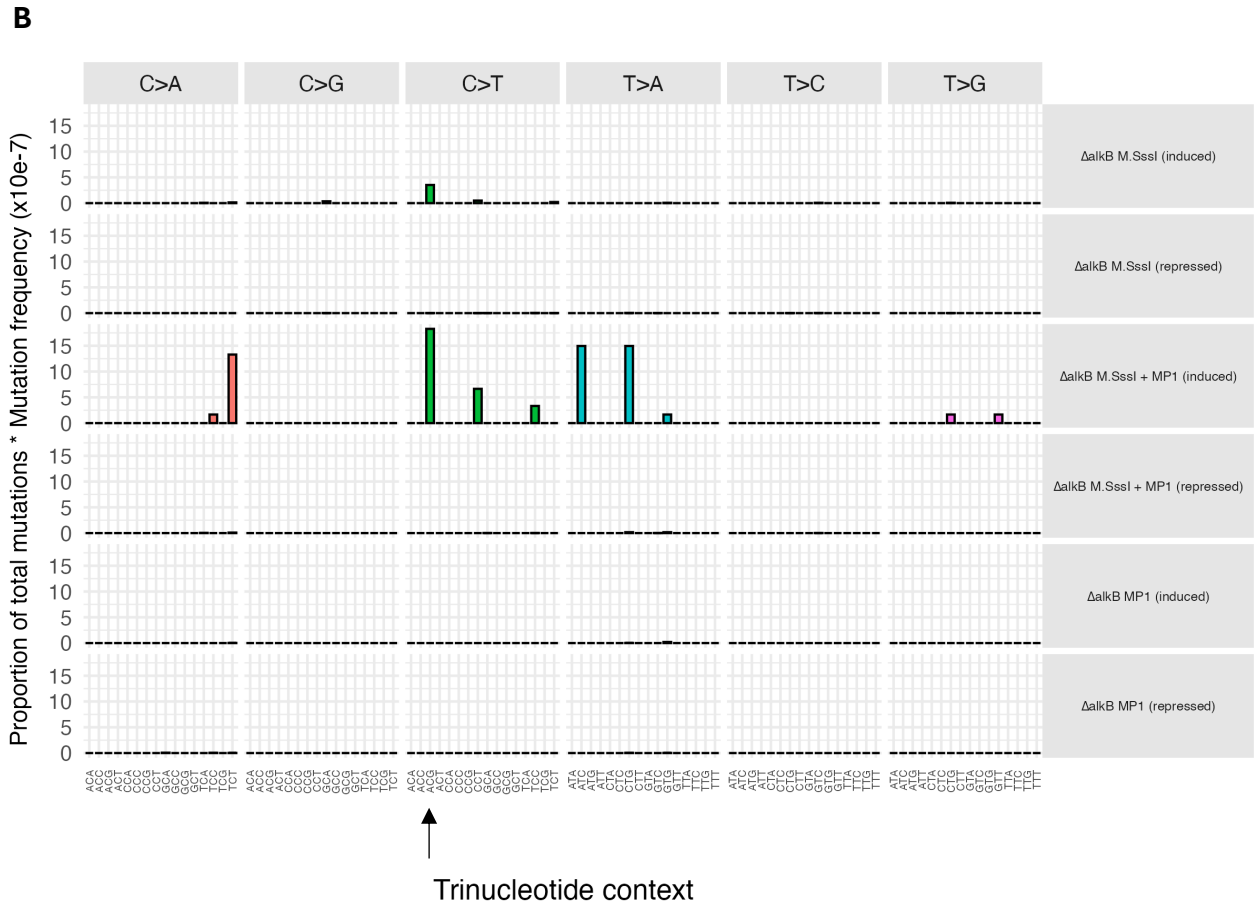

B
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
ΔalkB M.SssI (induced)
ΔalkB M.SssI (repressed)
ΔalkB M.SssI + MP1 (induced)
ΔalkB M.SssI + MP1 (repressed)
ΔalkB MP1 (induced)
ΔalkB MP1 (repressed)
Trinucleotide context

Figure 4.18: Introduction of RecA into Top10 *E. coli* with inducible M.SssI expression did not rescue the non-cytosine *rpoB*-specific mutation spectrum. DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in WT (**A**) and WT and Δ*alkB* (**B**) Top10 *E. coli* with inducible pM.SssI ± pMP1. Separate signatures are shown for arabinose (induced) and glucose (repressed) conditions given that *recA* is under the control of an arabinose-inducible promoter. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows.

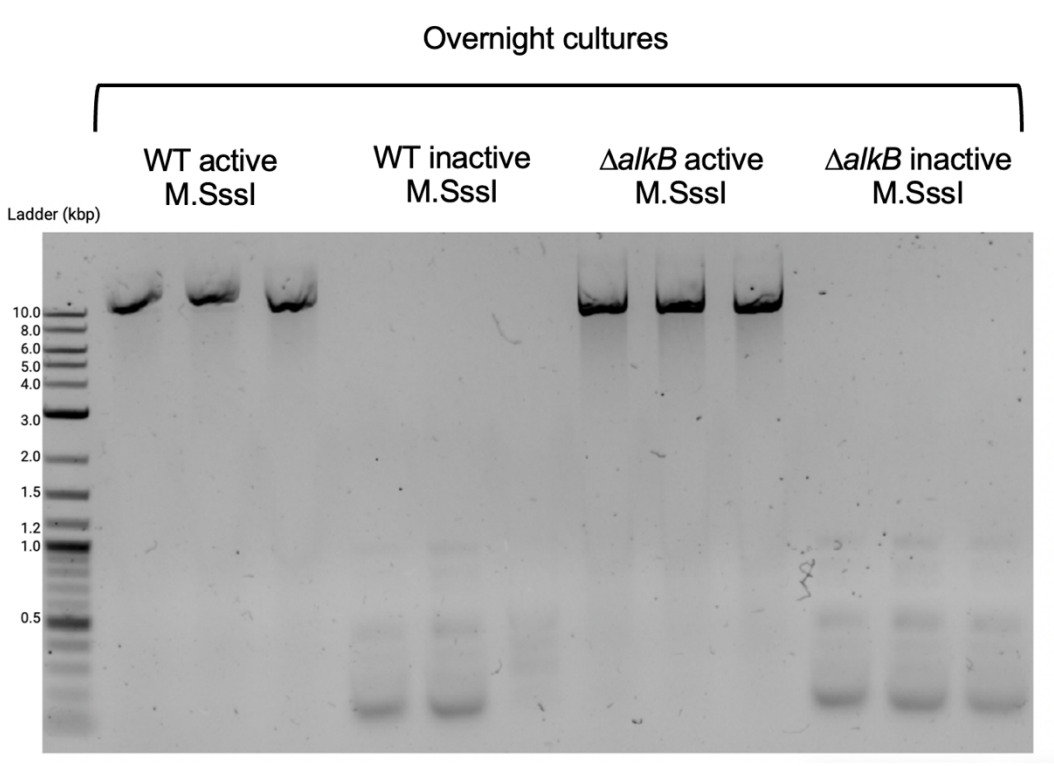
4.2.4. The non-cytosine mutations were not specific to constitutively active M.SssI

4.2.4.1. Constitutive M.SssI activity in Top10 *E. coli* did not yield non-cytosine mutations

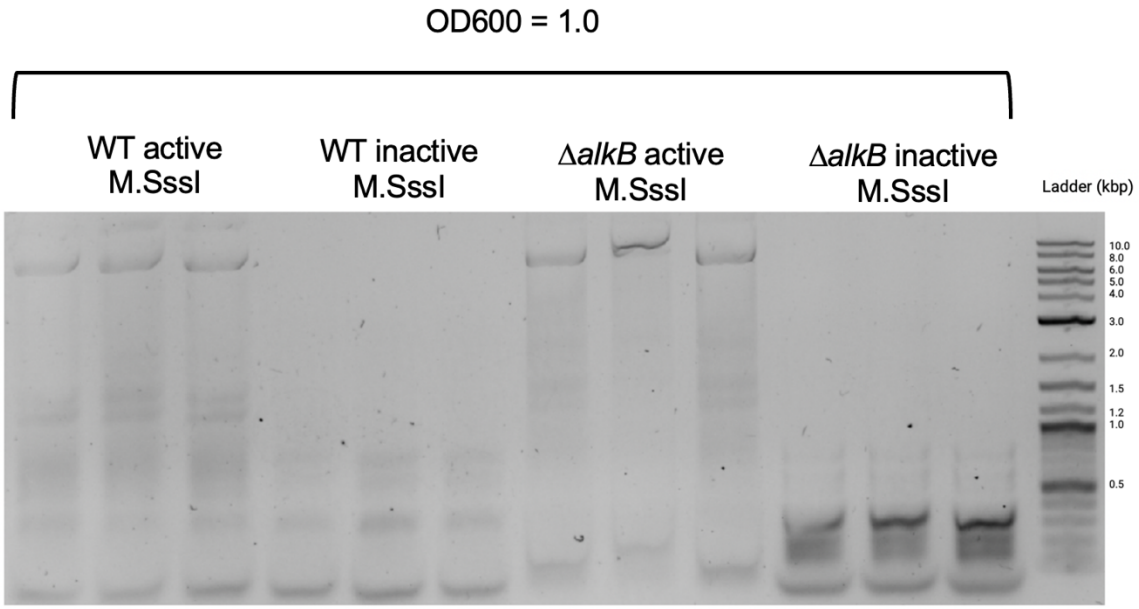
All the evidence so far suggested that the strong Δ*alkB* and M.SssI-dependent TA > CG signature was not caused by RecA activity. This had been shown robustly, both by trying to eliminate and rescue the signature in different backgrounds. So, what other explanations could have been behind this signature? The first was that there were other genetic differences between C2523 and Top10 that were responsible. This is something that will be discussed in Chapter 6. Another was that the mutations were caused by the constitutively active M.SssI itself. While the inducible M.SssI system was engineered from the constitutive, perhaps something had changed in the process that affected the methylating ability of M.SssI.

The first approach to testing this was to put the constitutive system into Top10. If the non-cytosine mutations were due to the constitutively active M.SssI and not the genetic background, then the signature would also appear in Top10. Therefore, WT and Δ*alkB* Top10 were transformed with active or inactive pM.SssI and rifampicin-resistant mutation frequencies were determined. Active M.SssI increased rifampicin-resistant mutation frequency in both WT and Δ*alkB* Top10 compared to inactive M.SssI (Figure 4.19C). While rifampicin-resistant mutation frequencies were lower than their C2523 counterparts (Figure 4.3B), M.SssI activity and loss of AlkB still increased rifampicin-resistant mutation frequency as expected. HpaII digests also suggested that active M.SssI was able to methylate DNA in Top10 (Figure 4.19A), although the lower bands observed for WT and Δ*alkB* Top10 with active M.SssI at OD600 = 1.0 might have suggested some reduction in DNA methylation capacity by M.SssI over the course of the assay (Figure 4.19B).

A



B


OD600 = 1.0
WT active M.SssI
WT inactive M.SssI
ΔalkB active M.SssI
ΔalkB inactive M.SssI
Ladder (kbp)
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.2
1.0
0.5

C

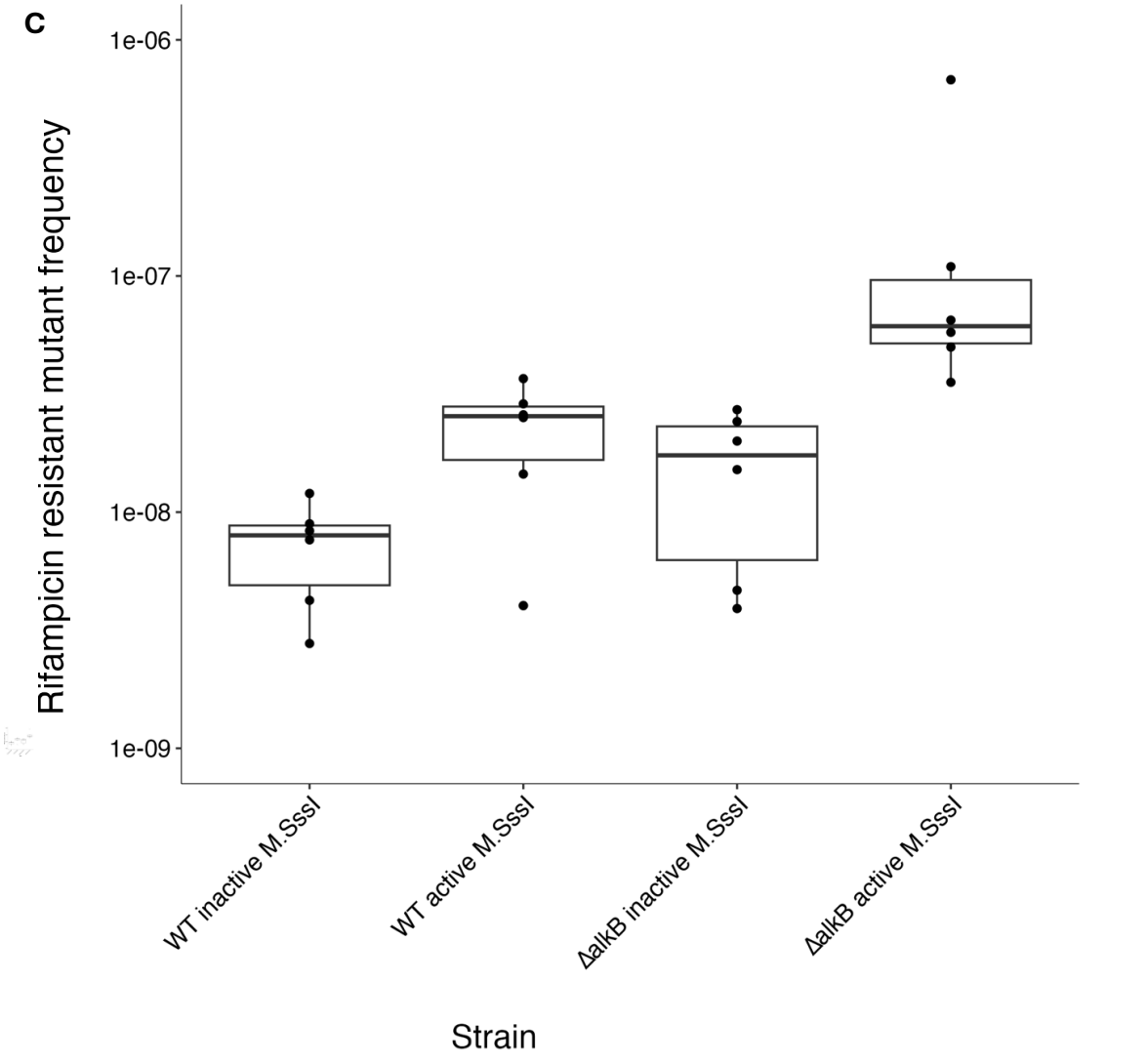

Rifampicin resistant mutant frequency
1e-06
1e-07
1e-08
1e-09
WT inactive M.SssI
WT active M.SssI
ΔalkB inactive M.SssI
ΔalkB active M.SssI
Strain

Figure 4.19: Constitutive M.SssI activity and loss of AlkB increased rifampicin-resistant mutation frequency in Top10 *E. coli*. Methylation-sensitive HpaII restriction enzyme digests of plasmid DNA from *E. coli* containing either active or inactive pM.SssI at beginning **(A)** and endpoint **(B)** of assay. HpaII is blocked by methylation in CCGG context. Each lane shows constitutively active or inactive pM.SssI DNA extracted from individual replicate clones. **(C)** WT and Δ*alkB* Top10 *E. coli* were transformed with active and inactive pM.SssI and plated on LB media ± rifampicin. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates.

RpoB-specific mutation spectra for Δ*alkB* Top10 with constitutively active or inactive pM.SssI are shown in Figure 4.20. Active M.SssI led to a predominant CG > TA signature, as expected. Δ*alkB* Top10 with active M.SssI displayed low levels of all other substitution types, including TA > CG. However, the contribution of these non-cytosine mutations to the overall *rpoB*-specific mutation spectrum was very low. Additionally, Δ*alkB* Top10 with inactive M.SssI also displayed low levels of TA > CG, albeit slightly lower than active M.SssI, suggesting that these mutations were the result of spontaneous damage rather than M.SssI activity specifically.

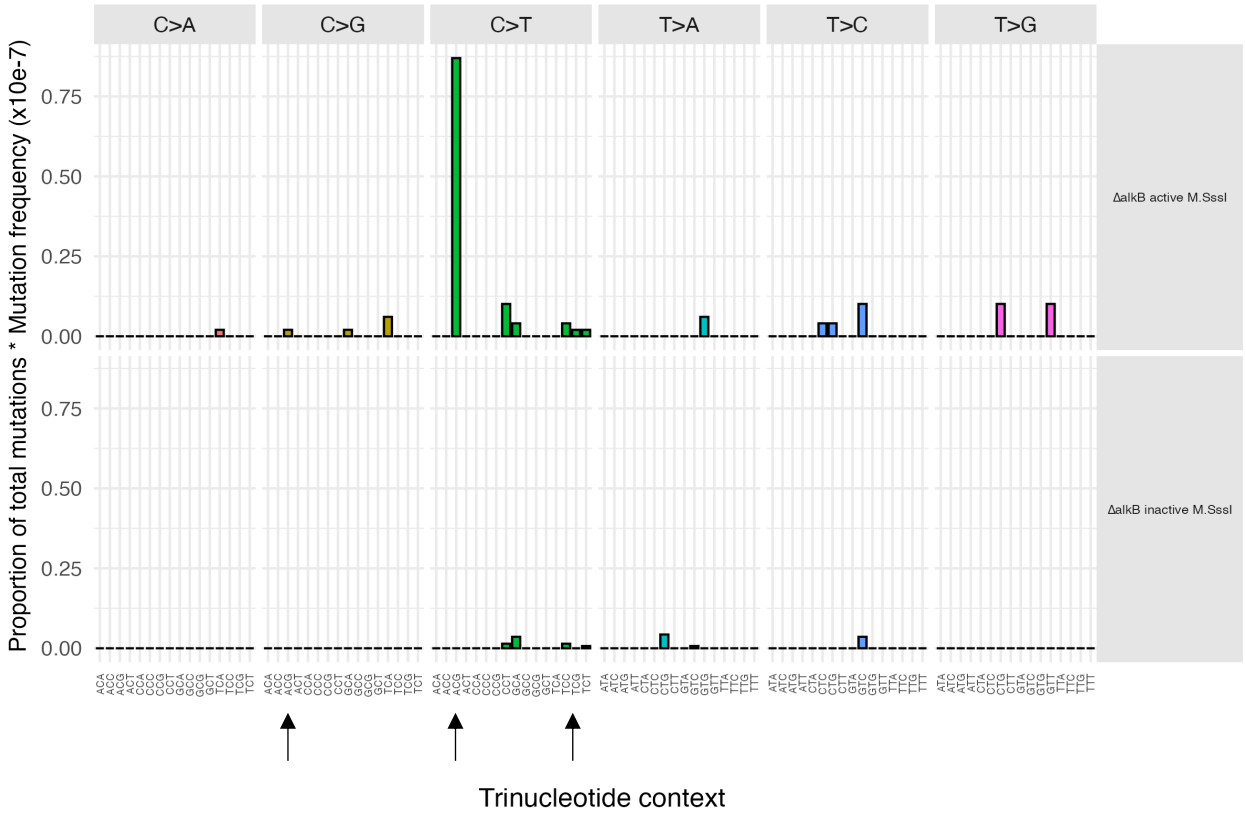


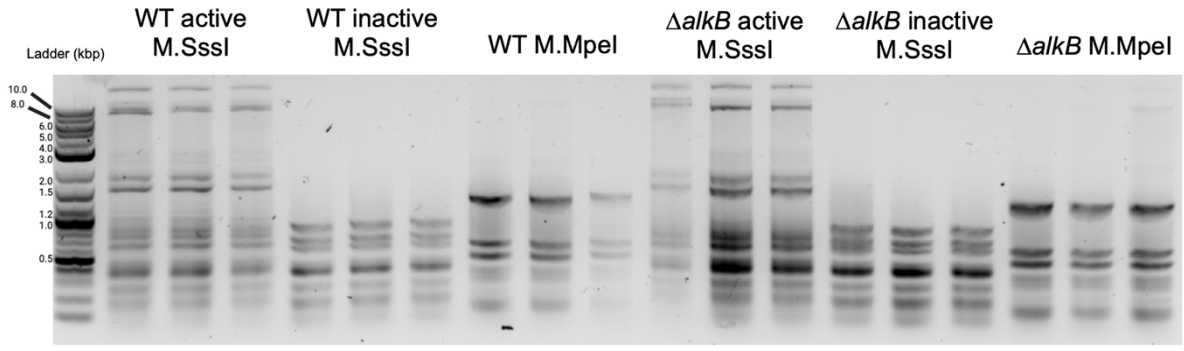
Figure 4.20: Constitutive M.SssI activity in Top10 *E. coli* did not rescue the non-cytosine *rpoB*-specific mutation spectrum. DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in Δ*alkB* Top10 *E. coli* with active or inactive pM.SssI. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows.

4.2.4.2. Weakly-methylating M.MpeI did not increase Δ*alkB*-dependent non-cytosine mutations

In order to fully test whether the non-cytosine mutations were M.SssI-specific, I repeated the rifampicin resistance assay in C2523 *E. coli* with an alternative CpG

methyltransferase M.MpeI. As with M.SssI, M.MpeI is constitutively active and results in cytosine methylation in the CpG context. WT and Δ*alkB* C2523 were transformed with pM.MpeI, or active or inactive pM.SssI as controls. HpaII digests of plasmid DNA suggested that M.MpeI only partially methylated DNA (Figure 4.21A). Active M.SssI controls also displayed lower band patterns, suggesting some loss of DNA methylation capacity over the assay. Indeed, mutation frequencies were slightly lower for WT and Δ*alkB* C2523 with active pM.SssI (Figure 4.21B) than previously observed (Figure 4.3B). Additionally, pM.MpeI rifampicin-resistant mutation frequency in Δ*alkB* C2523 was lower than the inactive pM.SssI control. There is evidence to suggest that M.MpeI has lower catalytic activity than M.SssI and does not methylate DNA to the same extent (Krwawicz et al., 2024). This could explain the lower rifampicin-resistant mutation frequency, although it was unexpected that it would be lower than the catalytically inactive M.SssI.

A



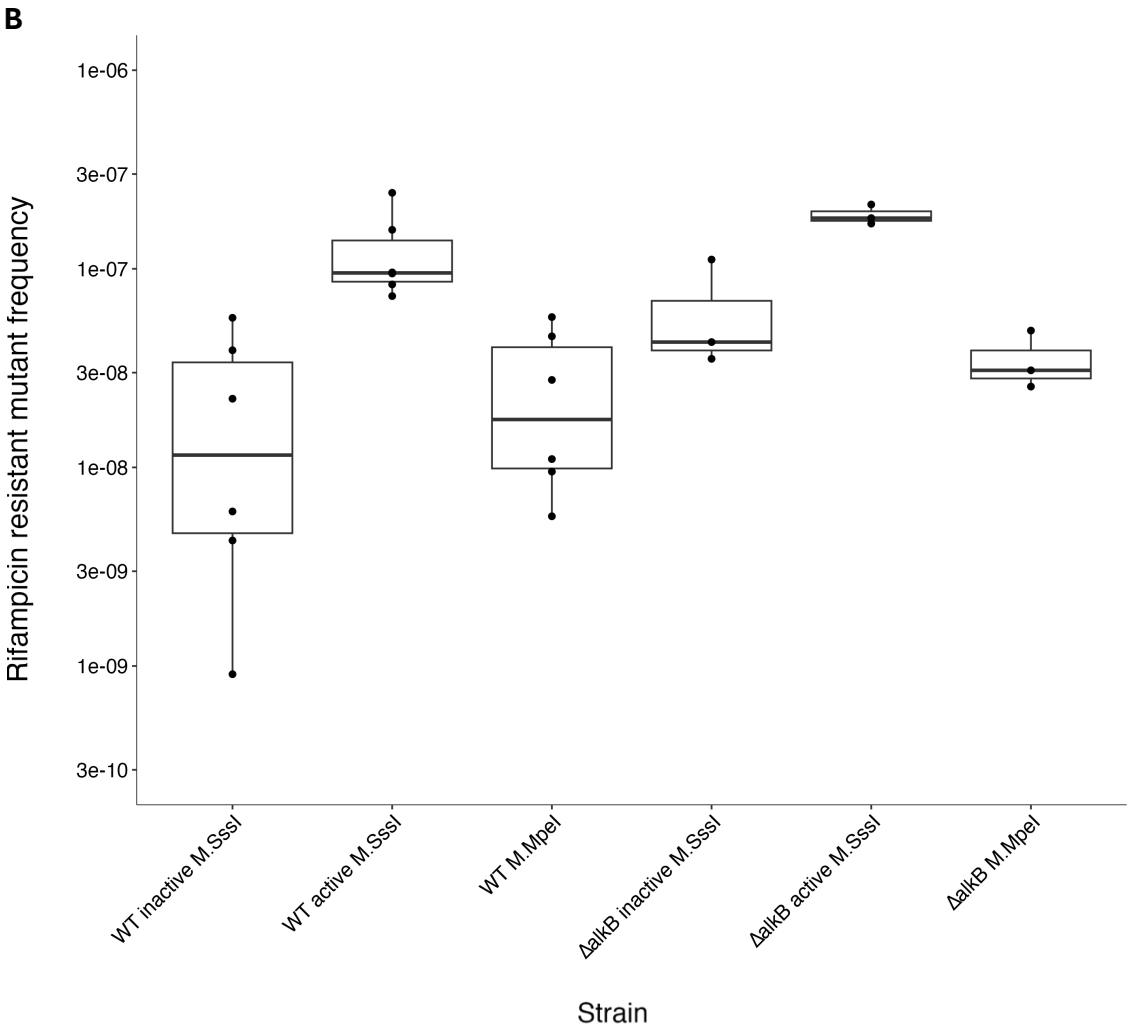


Figure 4.21: M.MpeI was weakly mutagenic compared to active M.SssI. (A) Methylation-sensitive HpaII restriction enzyme digest of plasmid DNA from WT and Δ*alkB* C2523 *E. coli* containing either active or inactive pM.SssI or pM.MpeI. HpaII is blocked by methylation in CCGG context. Each lane shows constitutively active or inactive pM.SssI or pM.MpeI DNA extracted from individual replicate clones. **(B)** WT and Δ*alkB* C2523 *E. coli* were transformed with active or inactive pM.SssI or pM.MpeI and plated on LB media ± rifampicin. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates.

Despite the low rifampicin-resistant mutation frequency, I proceeded to sequence some rifampicin-resistant colonies. Interestingly, both WT and Δ*alkB* C2523 with pM.MpeI displayed prominent CG > TA peaks, suggesting that at least some DNA methylation was occurring (Figure 4.22). While there is no direct negative control to compare it to, previous sequencing data from WT and Δ*alkB* C2523 with inactive pM.SssI displayed very low levels of CG > TA (Figure 4.4). Therefore, despite the similar rifampicin-resistant mutation frequencies for inactive cultures with pM.SssI and pM.MpeI, this suggested that M.MpeI methylated DNA to some extent. Additionally, both Δ*alkB* and WT C2523 displayed TA > CG signatures. This might have suggested that these were not Δ*alkB*-dependent, although it is worth noting that the WT rifampicin-resistant mutation frequency was only slightly lower than that of Δ*alkB*, so contributions of substitutions to *rpoB*-specific mutation spectra were very similar. Overall, it is difficult to say whether these data supported the hypothesis that the TA > CG signature was M.SssI-dependent. It is possible that the TA > CG mutations observed with pM.MpeI were due to spontaneous damage, not the action of the methyltransferase in the absence of AlkB. This would support the hypothesis that the signature was M.SssI-specific. However, the low methylation and rifampicin-resistant mutation frequency of M.MpeI make it difficult to properly interpret the sequencing results.

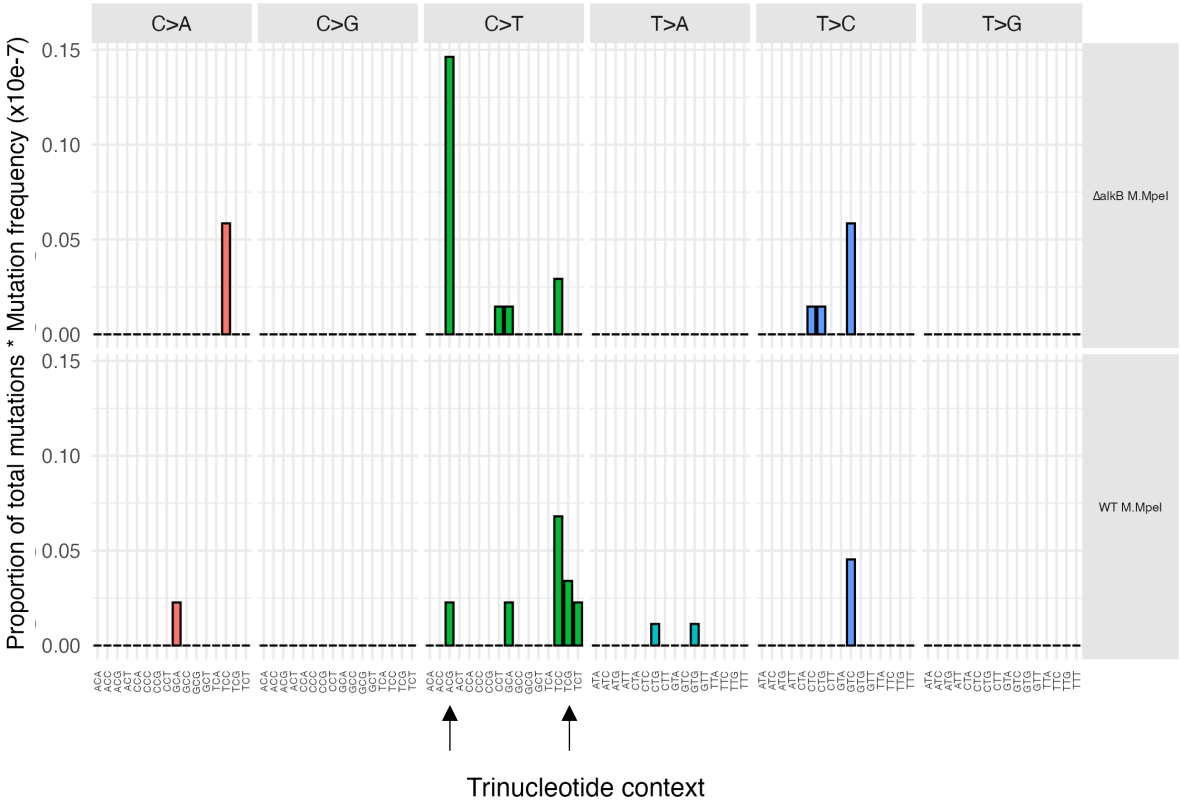


Figure 4.22: Constitutive M.MpeI activity in C2523 *E. coli* did not rescue the Δ*alkB*-dependent non-cytosine *rpoB*-specific mutation spectrum. DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in WT and Δ*alkB* C2523 *E. coli* with active pM.MpeI. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows.

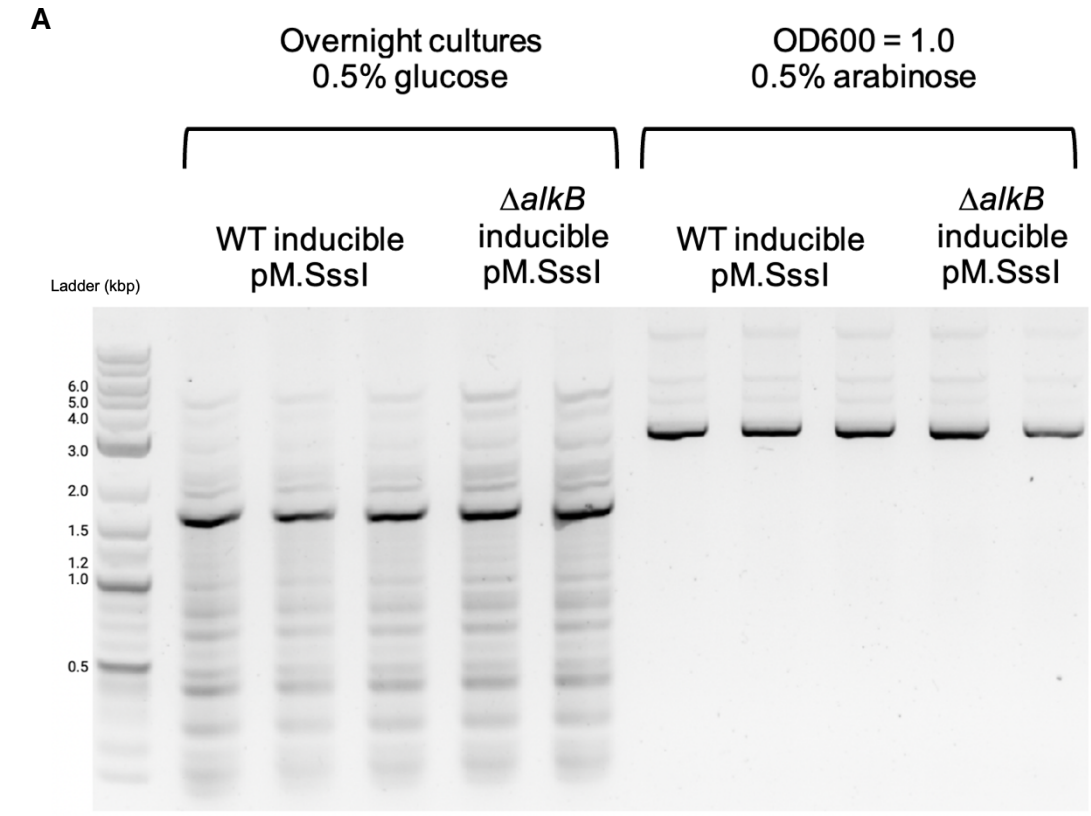
4.2.4.3. Increasing the length of time for which inducible M.SssI was active did not rescue the non-cytosine *rpoB*-specific mutation spectrum in Top10 *E. coli*

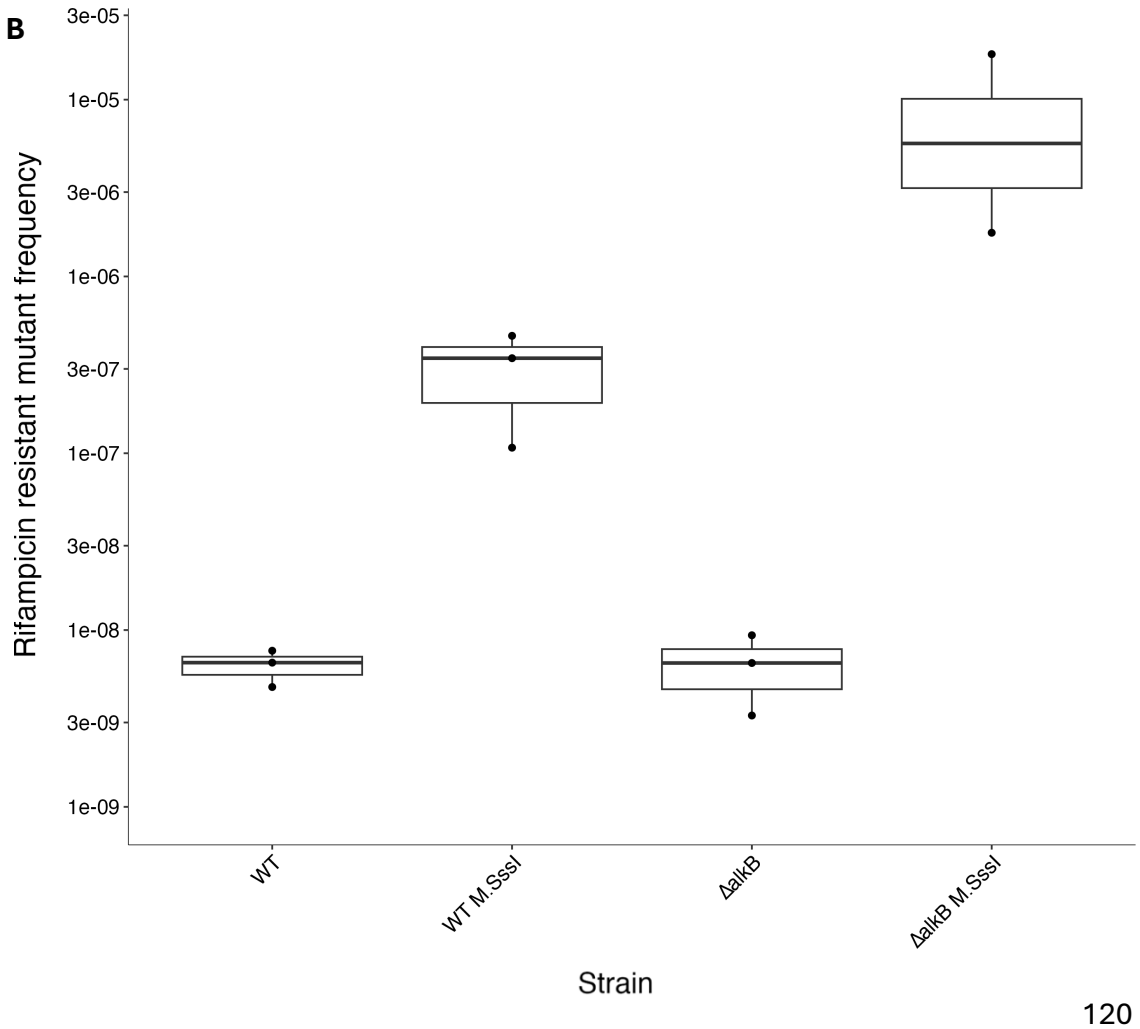
Another possibility was that the mutations were dependent on a technicality of the assay itself. The only difference between the constitutive and inducible systems was the length of time during which M.SssI was active. With the constitutive system, M.SssI was either always active or inactive throughout the assay. With the inducible however, M.SssI was expressed for a much shorter amount of time, approximately 2 hours compared to 24 hours for the constitutive. One can imagine that this would affect the amount of DNA methylation and therefore methylation lesions. It is logical to assume that the number of mutations was proportional to the amount of DNA damage. Therefore, perhaps the reason that non-cytosine mutations dominated in C2523 was not the capacity for bypass, but instead a direct reflection of the constitutive M.SssI system.

It would follow that even with the re-introduction of RecA into Top10, a short M.SssI induction time and therefore reduced window for methylation would mean less DNA damage and fewer mutations. In order to test the influence of active methylation time on mutagenesis, M.SssI activity was induced from the beginning of the assay, thereby increasing active methylation time to around 6 hours. To better replicate the constitutive system, M.SssI would have also been induced during the overnight culture phase. However, when tested, this had a substantial cell fitness cost resulting in poor to no growth of overnight cultures. Therefore, induction at the start

of the assay following overnight culture growth in glucose was optimal for maximising induction time while maintaining cell fitness.

Long induction of M.SssI activity resulted in widespread DNA methylation, as shown by the HpaII digest of plasmid DNA extracted from induced cultures at the end point of the assay (Figure 4.23A). Rifampicin-resistant mutation frequencies were equal in WT and Δ*alkB* Top10 – pM.SssI (Figure 4.23B), which was consistent with previous data (Figure 4.11C). Induced M.SssI activity caused an increase in rifampicin-resistant mutation frequency that was greater in Δ*alkB* than WT Top10, as expected. However, despite the longer duration of M.SssI activity, Δ*alkB* Top10 did not display any TA > CG mutations (Figure 4.23C). Interestingly, there was a marked increase in CG > AT mutations which has been intermittently observed in other assays and may have been the result of increased DNA methylation. Crucially, however, the clear lack of non-cytosine mutations suggested that the TA > CG signature did not arise solely from increased duration of DNMT activity.


A
Overnight cultures
0.5% glucose
OD600 = 1.0
0.5% arabinose
WT inducible pM.SssI
ΔalkB inducible pM.SssI
WT inducible pM.SssI
ΔalkB inducible pM.SssI
Ladder (kbp)
6.0
5.0
4.0
3.0
2.0
1.5
1.2
1.0
0.5


B
Rifampicin resistant mutant frequency
3e-05
1e-05
3e-06
1e-06
3e-07
1e-07
3e-08
1e-08
3e-09
1e-09
WT
WT M.SssI
ΔalkB
ΔalkB M.SssI
Strain

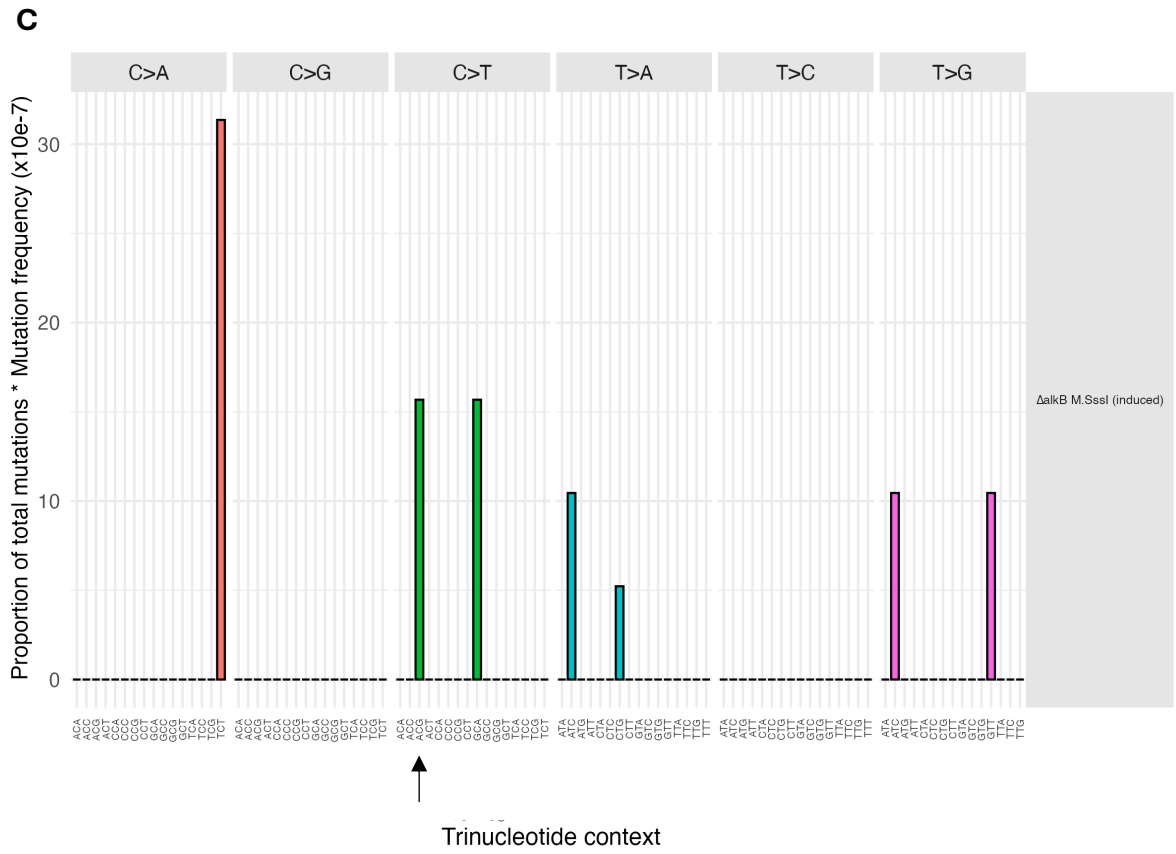


Figure 4.23: Long induction of inducible M.SssI did not rescue TA > CG signature in Top10. (A) Methylation-sensitive HpaII restriction enzyme digest of plasmid DNA from WT and Δ*alkB* Top10 *E. coli* containing inducible pM.SssI. Overnight cultures were grown in 0.5% glucose, while assay cultures were grown in 0.5% arabinose. HpaII is blocked by methylation in CCGG context. Each lane shows inducible pM.SssI DNA extracted from individual replicate clones. **(B)** WT and Δ*alkB* Top10 *E. coli* were transformed with inducible pM.SssI and plated on LB media ± rifampicin. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates. **(C)** DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in Δ*alkB* Top10 *E. coli* with induced pM.SssI. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows.

4.2.4.4. Could both RecA activity and constitutive M.SssI activity be required for non-cytosine mutations in Δ*alkB* Top10 *E. coli*?

4.2.4.4.1. Re-engineering MP1 resistance to be compatible with constitutively active pM.SssI

So far, the source of the non-cytosine mutations remained elusive. They did not appear to be dependent on constitutive M.SssI activity specifically, given that introducing the constitutive system into Top10 did not rescue the signature. RecA expression also did not appear to be responsible, given that neither re-introducing RecA into Top10 nor inhibiting RecA in C2523 *E. coli* rescued the signature. Increasing the duration of M.SssI expression also did not yield Δ*alkB*-dependent non-cytosine mutations. However, none of the experiments above addressed the combination of RecA activity and a long duration of M.SssI activity. Hypothetically, if there is error-prone bypass facilitated by RecA but few lesions, the mutational burden will be low. Similarly, lesion burden could be high, but if there is no bypass then those lesions will be less mutagenic and potentially cytotoxic. Therefore, the key experiment was to combine RecA activity and a long M.SssI expression time to test whether this rescued the non-cytosine mutations. As previously mentioned, the inducible system was not functional in C2523, so instead I transformed WT and Δ*alkB* Top10 with pMP1 and constitutive pM.SssI. However, both pMP1 and constitutive pM.SssI expressed chloramphenicol resistance. In order to select for and maintain both plasmids, I had to replace the chloramphenicol resistance cassette (CamR) in pMP1 with an ampicillin resistance cassette (AmpR).

To do this, I linearised pMP1 with PasI which cut once within the CamR cassette. I ligated in a PCR-amplified and gel extracted AmpR cassette. This was amplified using primers that contained PasI restriction sites, enabling seamless ligation into pMP1. WT Top10 and C2523 *E. coli* were transformed with the ligation mixture and plasmid DNA was extracted from resulting chloramphenicol-resistant colonies. Plasmid DNA was linearised with HindIII, which cut once in the plasmid, and band sizes were compared to unmodified pMP1. Modified pMP1, termed pMP1.2, was expected to be ~8kbp due to the insertion of AmpR into CamR. Clones 6 and 7 extracted from WT Top10 and clones 1, 3, 6 and 7 from WT C2523 displayed a single band at ~8kbp and were sent for sequencing (Figure 4.24). Sequencing confirmed successful integration of AmpR into clones 3, 6 and 7 from WT C2523. Clone 3 was taken forward for use in the following rifampicin resistance assay.

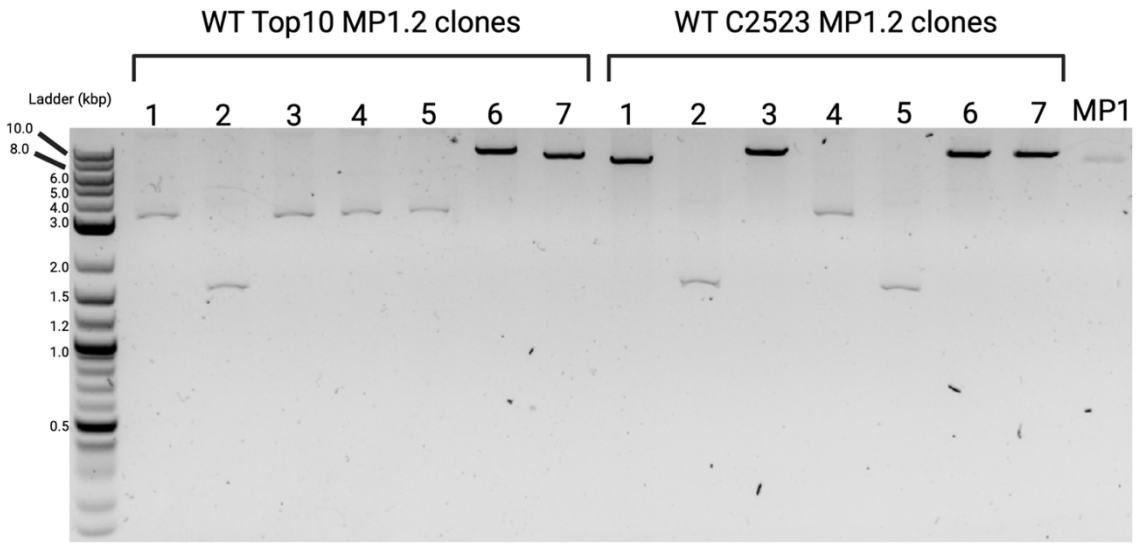


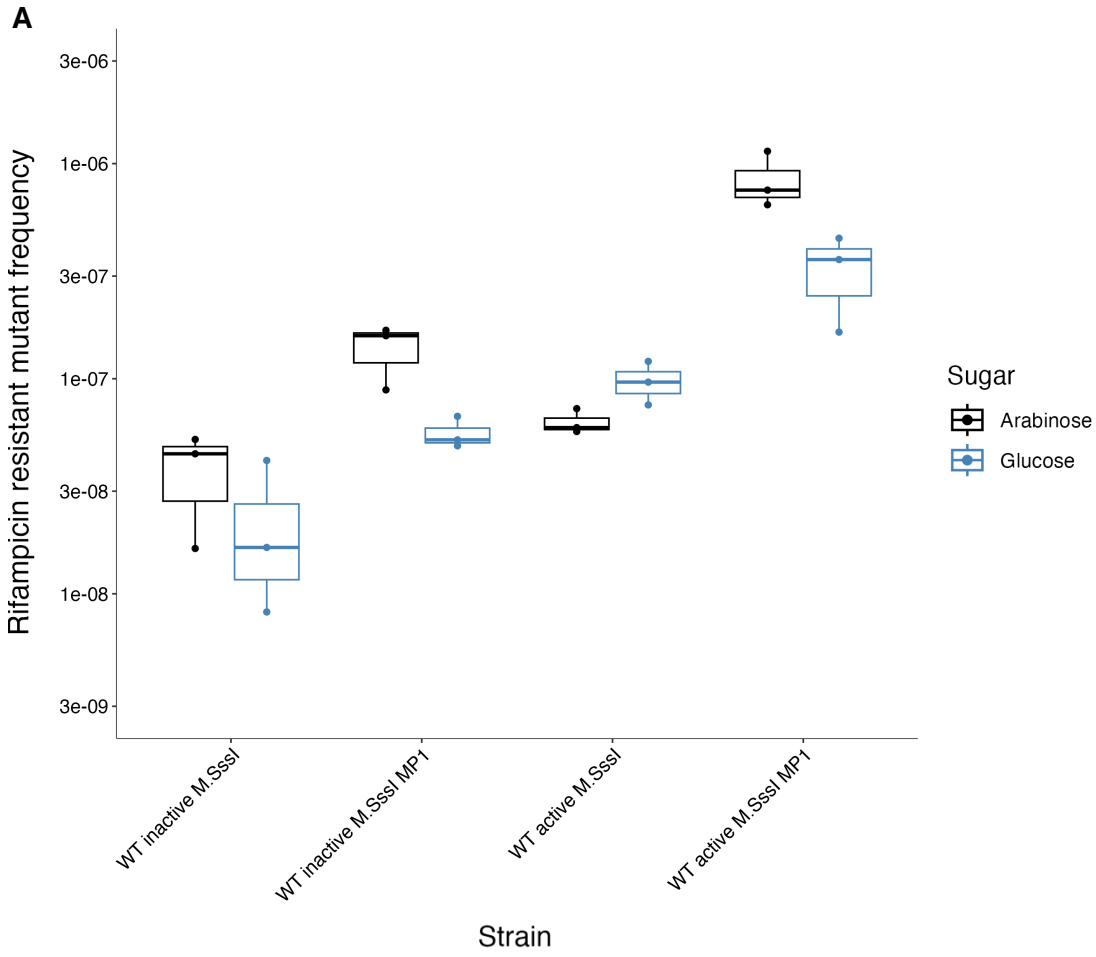
Figure 4.24: Screening of modified MP1 plasmids containing AmpR instead of CamR. Plasmid DNA was extracted from WT Top10 and C2523 transformants resistant to ampicillin. Plasmid DNA was linearised with HindIII and run on 1% agarose gel against an unmodified MP1 plasmid control (lane 16).

4.2.4.4.2. Induction of RecA and constitutive M.SssI activity in Top10 *E. coli* did not rescue non-cytosine mutations

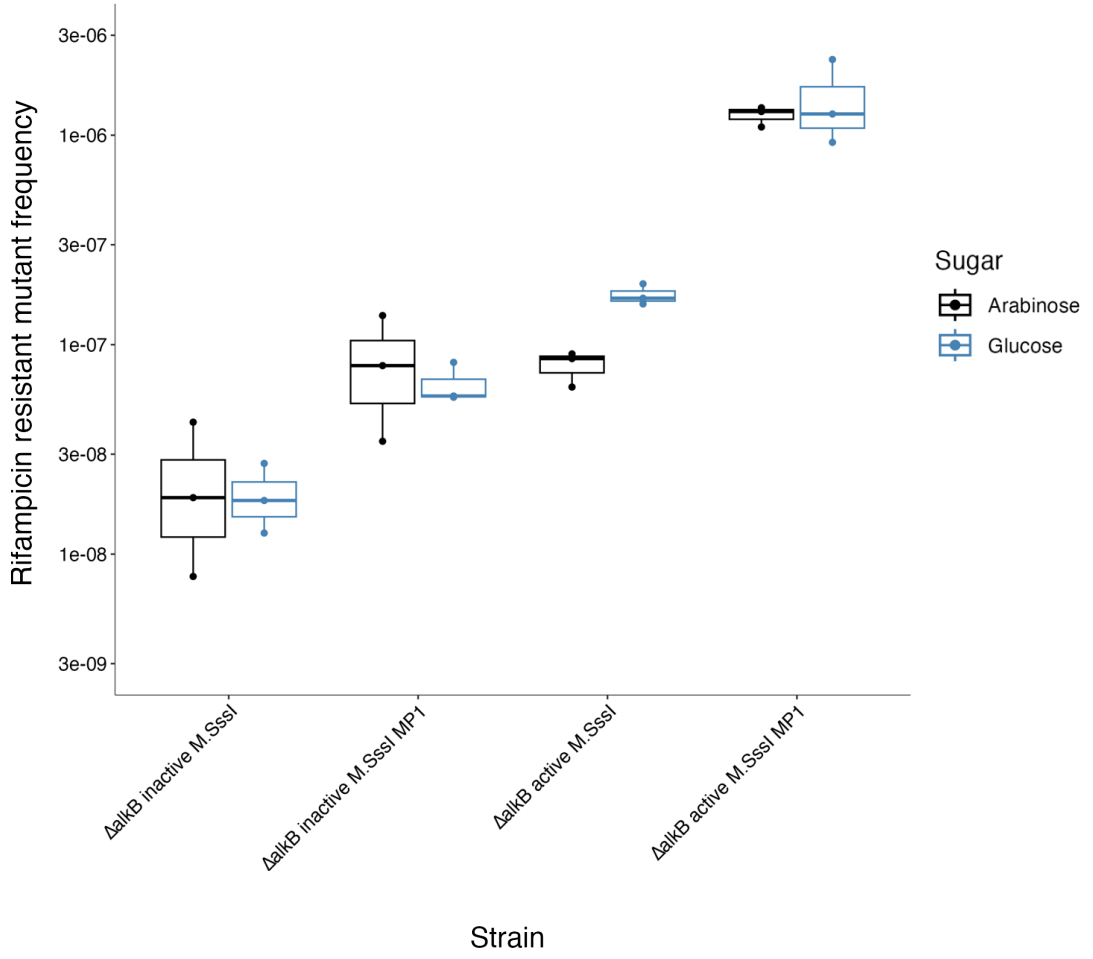
WT and Δ*alkB E. coli* were transformed with either constitutively active or inactive pM.SssI ± pMP1.2. Cultures were grown under glucose repression, then split and treated with either 0.5% glucose or arabinose to induce RecA expression. Rifampicin-resistant mutation frequencies are shown in Figure 4.25 A (WT) and B (Δ*alkB*) respectively. I hypothesised that a long M.SssI expression time and RecA activity would result in increased bypass of alkylation lesions. This would increase rifampicin-resistant mutation frequency in Δ*alkB* Top10 to a greater extent than WT. Expression of RecA via pMP1.2 increased 'baseline' rifampicin-resistant mutation frequency with inactive M.SssI in both WT and Δ*alkB* Top10, as observed previously in Figure 4.17C, D. Also consistent with previous results, arabinose treatment of cells transformed with pMP1.2 did not consistently increase rifampicin-resistant mutation frequency compared to those treated with glucose in either background. This was consistent with suspected constitutive expression of genes in pMP1. However, there was no notable increase in rifampicin-resistant mutation frequency in Δ*alkB* Top10 with active pM.SssI and pMP1.2 compared to the WT equivalent (Figure 4.25C). This suggested that there was not any M.SssI and RecA-dependent mutagenesis (Figure 4.25C).

Analysis of WT and Δ*alkB rpoB*-specific mutation spectra revealed no difference in mutation type or distribution for cells transformed with constitutive pM.SssI and pMP1.2 and treated with either arabinose or glucose (Figure 4.25D, E). WT Top10

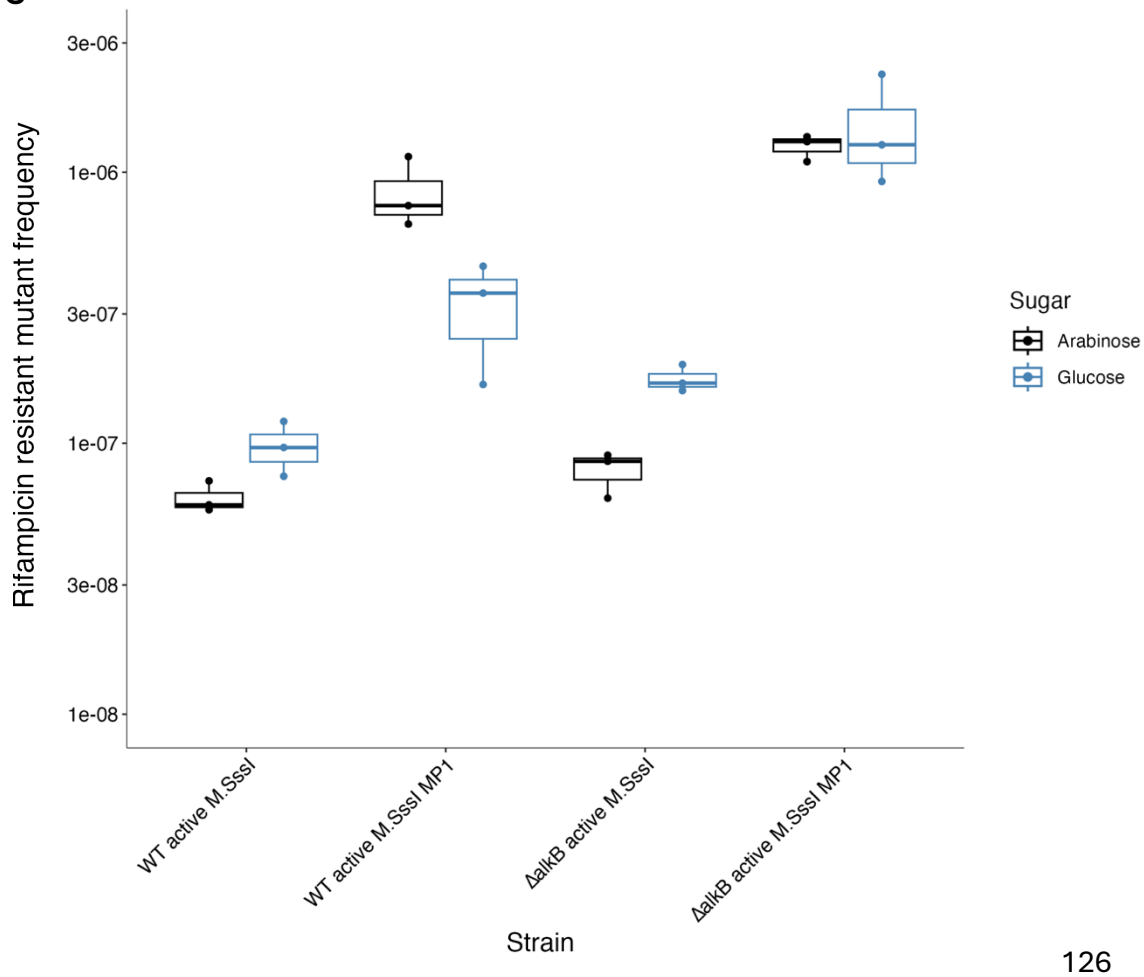
with active pM.SssI yielded predominantly CG > TA mutations and addition of pMP1.2 increased levels of CG > AT, TA > AT and TA > CG (Figure 4.25D). This was consistent with *rpoB*-specific mutation spectra observed in WT Top10 with induced pM.SssI + pMP1 (Figure 4.18A). A similar pattern was observed for Δ*alkB* Top10, albeit at a higher and more consistent rifampicin-resistant mutation frequency regardless of arabinose or glucose treatment (Figure 4.25E). Notably, low levels of TA > CG mutations occurred in both WT and Δ*alkB* backgrounds. Crucially however, levels of TA > CG in Δ*alkB* Top10 were not higher than in WT. This suggested than any TA > CG observed was a result of mutagenesis of spontaneous damage and was not caused by RecA-mediated bypass of M.SssI-induced damage.

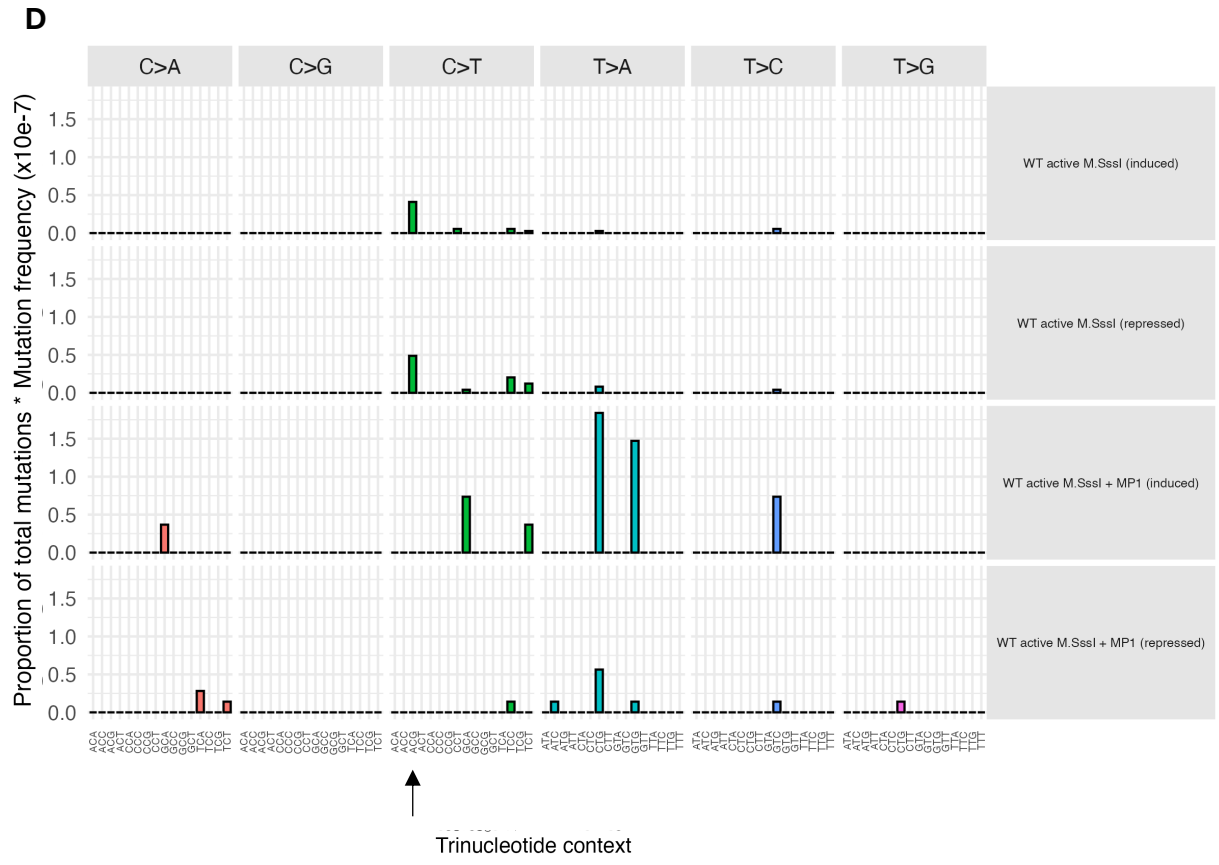


B


Rifampicin resistant mutant frequency
3e-06
1e-06
3e-07
1e-07
3e-08
1e-08
3e-09
ΔalkB inactive M.SssI
ΔalkB inactive M.SssI MP1
ΔalkB active M.SssI
ΔalkB active M.SssI MP1
Strain
Sugar
Arabinose
Glucose

C


Rifampicin resistant mutant frequency
3e-06
1e-06
3e-07
1e-07
3e-08
1e-08
WT active M.SssI
WT active M.SssI MP1
ΔalkB active M.SssI
ΔalkB active M.SssI MP1
Strain
Sugar
Arabinose
Glucose


D
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
WT active M.SssI (induced)
WT active M.SssI (repressed)
WT active M.SssI + MP1 (induced)
WT active M.SssI + MP1 (repressed)
Trinucleotide context

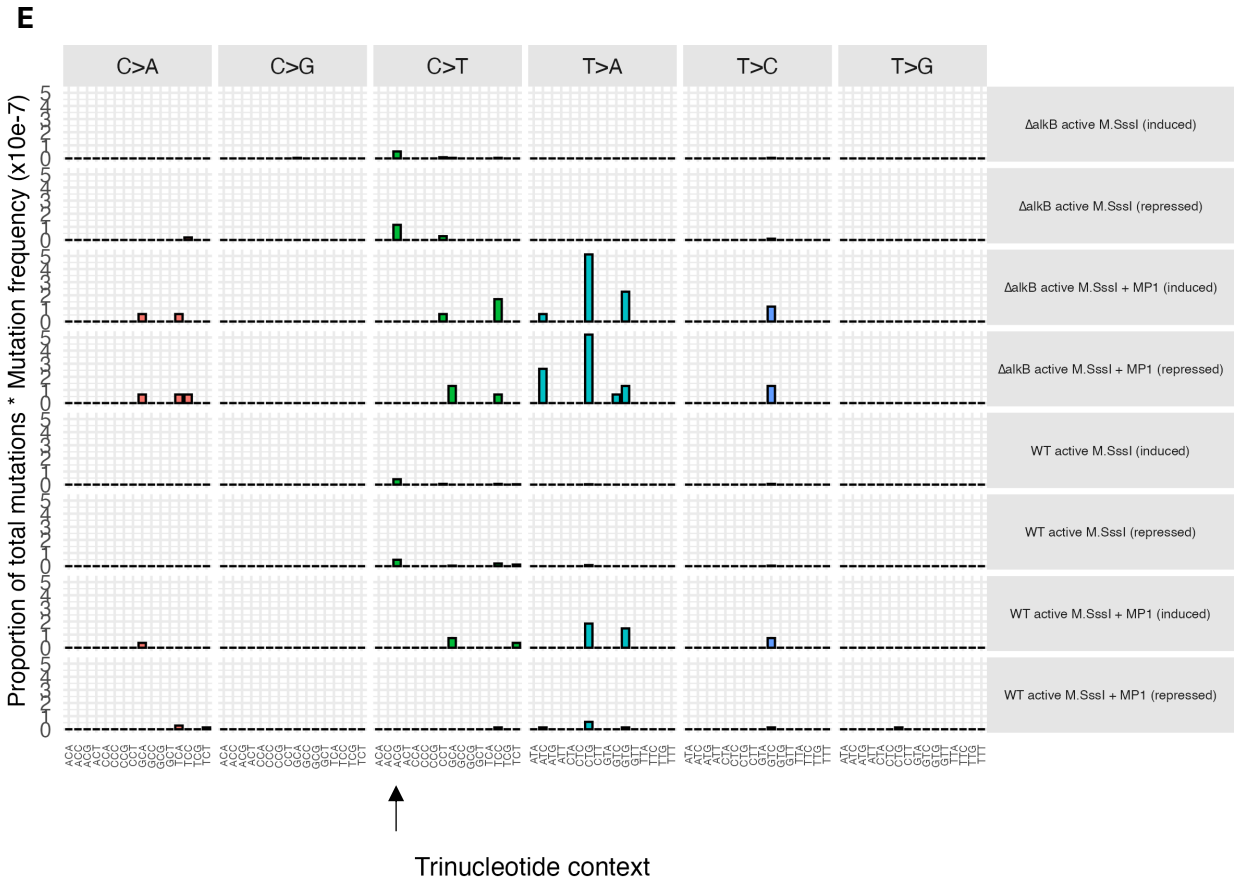

E
C>A
C>G
C>T
T>A
T>C
T>G
Proportion of total mutations * Mutation frequency (x10e-7)
ΔalkB active M.SssI (induced)
ΔalkB active M.SssI (repressed)
ΔalkB active M.SssI + MP1 (induced)
ΔalkB active M.SssI + MP1 (repressed)
WT active M.SssI (induced)
WT active M.SssI (repressed)
WT active M.SssI + MP1 (induced)
WT active M.SssI + MP1 (repressed)
Trinucleotide context

Figure 4.25: Introduction of RecA into Top10 *E. coli* with constitutively active M.SssI additively increased rifampicin-resistant mutation frequency and did not rescue TA > CG *rpoB*-specific mutation spectrum. WT **(A)** and Δ*alkB* **(B)** Top10 *E. coli* were transformed with constitutively active or inactive pM.SssI ± pMP1.2. Samples were treated with either 0.5% arabinose or glucose. The number of rifampicin-resistant mutant colonies was divided by number of colonies on LB media - rifampicin to calculate a mean rifampicin-resistant mutation frequency per independent replicate clone. Each point represents mean rifampicin-resistant mutation frequency for an independent replicate clone, averaged over two technical (plating) replicates. **(C)** Rifampicin-resistant mutation frequencies for WT and Δ*alkB* Top10 with active pM.SssI ± pMP1.2. DNA from mutant colonies was isolated and a 500bp region of *rpoB* amplified by PCR. PCR products were sequenced to determine *rpoB*-specific mutation spectra in WT **(D)** and Δ*alkB* **(E)** Top10 *E. coli* with active pM.SssI ± pMP1.2. Signatures are categorised by substitution type and trinucleotide context, comprised of the 3' and 5' bases adjacent to the mutated base. This gives 96 possible signatures. The percentage of unique signature was multiplied by mean rifampicin-resistant mutation frequency for each strain to give a qualitative contribution to overall mutation signature. Observed signatures occurring in the CpG context are indicated by black vertical arrows.

4.3. Specific discussion

The work presented in this chapter demonstrated that damage induced by DNMT activity was mutagenic in the absence of AlkB. Previous work found that off-target activity of DNMTs catalysed the deposition of 3mC, which is a potentially cytotoxic DNA lesion (Dukatz et al., 2019; Rošić et al., 2018). Further work found that DNMT activity sensitised Δ*alkB* cells to MMS, suggesting that the increased levels of alkylation damage were sufficient to overwhelm back-up repair capacity (Krwawicz et al., 2024). Therefore, this work supported the original hypothesis that DNMT

activity was mutagenic, which may have contributed in part to the evolutionary loss of DNA methylation in multiple independent lineages.

There was an increase in CG > TA mutations with M.SssI activity which was exacerbated by loss of AlkB. This suggested that in WT conditions, AlkB prevented some CG > TA mutations. Therefore, the CG > TA signature could not be attributed solely to deamination of 5mC. In the event that M.SssI activity resulted in 3mC deposition, it is possible that this could stall the replisome and be bypassed by error-prone TLS polymerases. This could lead to incorporation of adenine opposite 3mC and CG > TA mutations. Intriguingly, the Δ*alkB* and M.SssI-dependent TA > CG *rpoB*-specific mutation spectrum was only observed in C2523 *E. coli* with the constitutively active M.SssI. M.SssI is a CpG DNA methyltransferase with both *de novo* and maintenance DNA methyltransferase activity (Renbaum et al., 1990). Therefore, a non-cytosine M.SssI-dependent *rpoB*-specific mutation spectrum was unexpected. One possibility was that M.SssI was able to introduce 1mA or 3mA as off-target alkylation lesions alongside 3mC and canonical 5mC. 1mA and 3mA both stall the passage of Pol III and could therefore be mutagenic (Boiteux & Laval, 1982). AlkB is capable of directly de-methylating 1mA but cannot directly reverse 3mA back to un-methylated cytosine (Falnes et al., 2002; O'Brien & Ellenberger, 2004; Trewick et al., 2002), making 1mA the more likely candidate. However, this was unlikely to be the case as the active site of M.SssI does not encompass adenine and so cannot catalyse adenine methylation (Renbaum et al., 1990).

What, then, could be the cause of these mutations? In this chapter I investigated three possible factors: RecA activity, length of M.SssI expression time, and type of DNMT. The possibility of RecA involvement was first identified when comparing the genotypes of C2523 and Top10 *E. coli,* given that the non-cytosine mutations were only observed in C2523. One of the key differences is that C2523 contain RecA while Top10 have the *recA1* allele, which severely impairs but does not abolish RecA activity. This investigation was well supported by the literature, which found RecA and TLS to be involved in bypass of small alkylation damage in both *C. elegans* (Volkova et al., 2020) and *D. melanogaster* (Khodaverdian et al., 2024).

I then considered the second factor, length of M.SssI expression time, which was a notable technical difference between constitutive and inducible assays. The duration of M.SssI activity is much greater with the constitutive than inducible system. This could affect the amount of methylation on DNA, which would in turn influence the amount of DNA damage that those cells are subjected to. Logically, mutagenesis will be proportional to the amount of DNA damage. One could imagine that this may not be directly proportional, as mutagenesis might depend on whether a threshold level of DNA damage is reached and availability of non-canonical repair or bypass mechanisms. It is also possible that the approximately two hours during which the inducible M.SssI was active was sufficient to fully methylate the genome. Therefore, any additional expression time may have had no effect on DNA damage levels and mutagenesis.

Surprisingly, neither RecA activity nor increased M.SssI expression time alone rescued the TA > CG signature. Therefore, in section 4.2.4.4, I investigated how the presence of RecA and duration of M.SssI expression may have affected mutagenesis in the absence of AlkB. I hypothesised that both RecA activity and a long expression time were necessary for sufficient alkylation damage to occur and subsequent mutagenesis by error-prone RecA-mediated TLS. However, this still did not rescue the signature. Below, I will discuss technical limitations of the rifampicin resistance assay that may have contributed to this result, as well as genetic and cellular factors.

4.3.1. Limitations of rifampicin resistance assay

4.3.1.1. Jackpotting

In section 4.1 I touched on the main limitation of the rifampicin resistance assay – that not all mutations give rise to resistance, therefore only a subset of mutations can be observed via the assay. However, given that the aim was not to observe all mutations but instead to compare mutations across conditions, the assay was still suitable despite this limitation. Another possible limitation of the rifampicin resistance assay is jackpotting, where a particular mutation arises early in the population and is then disproportionately represented in the final population due to clonal expansion. In this case, clones would not be the result of independent mutational events and sequencing of jackpotted clones would skew representation of individual mutations in *rpoB*-specific mutation spectra. The rifampicin resistance

assay was designed to prevent this from occurring; selection by rifampicin was applied only on plates at the end point of the assay and not in the cultures, therefore removing any selective advantage and giving all mutations an equal chance of propagation. Of course, it was possible that a mutation conferring resistance to rifampicin could have arisen early in the population and become abundant due to drift. To counteract this, each assay was repeated multiple times with at least three independent replicate clones. Therefore, if jackpotting or accumulation due to drift was to occur in one replicate, a particular mutation would be elevated compared to spectra from the other replicates. Only mutations observed consistently across all replicates were pooled for *rpoB*-specific mutation spectra. Assays were also carefully conducted to limit the time that cultures spent in stationary phase, as levels of replication stress typically increase during this time (Yamanaka et al., 2001), therefore increasing the likelihood of mutations independent of DNMT activity.

4.3.1.2. Rifampicin resistant mutation distribution

Given that M.SssI is a CpG DNMT, I expected to see an enrichment of mutations at CpG dinucleotides for conditions with active compared to inactive M.SssI. To investigate this, I tabulated the number of CG > TA mutations occurring in either CpG or CpNpG sequence contexts for WT and Δ*alkB* C2523 with constitutively active or inactive pM.SssI (Table 2A). There was a higher frequency of CG > TA mutations for cells with active compared to inactive M.SssI, although this could have been due to higher rifampicin-resistant mutation frequencies in these conditions. To allow for

this, I then normalised the total number of CG > TA mutations in CpG or CpNpG contexts to mean rifampicin-resistant mutation frequency for each condition (Table 2B). Both WT and Δ*alkB* C2523 displayed an enrichment in CG > TA mutations with active M.SssI. WT C2523 had a higher fold change in CG > TA mutations than Δ*alkB* C2523. This might have been due to the fact that Δ*alkB* with active M.SssI displayed increased levels of other mutations such as TA > CG. Each *rpoB* sequence only contained a single mutation, so an increase in one type of mutation, for example TA > CG, would have resulted in a decrease in CG > TA.

Table 2A: Frequency of CG > TA mutations occurring at CpG or CpNpG sequence contexts for WT and Δ*alkB* C2523 with active or inactive pM.SssI. Total CG > TA mutations for each condition are shown in the right-most column.

	Frequency of mutation on C of CpG or CpNpG						
Sequence context	CpNpG	CpG	CpG	CpNpG	CpNpG	CpNpG	
Mutation at base pair	C>T 1565	C>T 1585	G>A 1586	G>A 1592	C>T 1691	C>T 1768	Total
WT active M.SssI	0	20	64	7	4	1	96
WT inactive M.SssI	0	4	2	5	1	0	12
Δ*alkB* active M.SssI	1	0	14	3	5	0	23
Δ*alkB* inactive M.SssI	1	0	2	0	1	0	4

Table 2B: Normalisation of CG > TA mutations occurring at CpG or CpNpG sequence contexts for WT and Δ*alkB* C2523 with active or inactive pM.SssI. The proportion of CG > TA mutations for each strain was calculated as a percentage of all CG > TA mutations across conditions. This proportion was then normalised by multiplying by mean rifampicin-resistant mutation frequency for each condition. Fold change was calculated by dividing the active/inactive normalised value. The inactive contribution was also subtracted from the active for each strain to remove background and identify CG > TA arising as a result of DNMT activity.

	Total frequency of CG > TA mutations	Proportion of CG > TA mutations * rifampicin-resistant mutation frequency	Active/inactive M.SssI fold change	Active – inactive M.SssI normalisation
WT active M.SssI	96	1.576	56.531	1.548
WT inactive M.SssI	12	0.028		
Δ*alkB* active M.SssI	23	17.221	32.560	16.692
Δ*alkB* inactive M.SssI	4	0.529		

Given that the observed TA > CG mutations were M.SssI and Δ*alkB*-dependent, I then considered whether there was an enrichment of TA > CG mutations downstream of CpG or CpNpG sites in cells with active M.SssI. I tabulated the frequency of TA > CG mutations occurring near CpNpG sites as either N+1, N+3 or N+4 (Table 3A). I then normalised to mean rifampicin-resistant mutation frequency and calculated the fold change between active and inactive M.SssI as above (Table 3B). WT and Δ*alkB* C2523 with active pM.SssI were enriched for TA > CG mutations near CpNpG sites at N+1, N+3 and N+4. Fold changes for N+1 and N+3 were higher

in WT than Δ*alkB* cells, although this was probably due to skewing of the data by low mutational frequencies (Table 3A).

However, in order to determine whether TA > CG mutations were truly enriched near CpNpG sites, one would have to compare these data to TA > CG mutations occurring away from CpNpG sites. To do this, I compared mutations occurring at N+1 (near) to those occurring at N+3 and N+4 (far). In Δ*alkB* cells with active M.SssI, the fold change in TA > CG mutations was higher at N+1 sequence context compared to N+3/4 contexts. Therefore, it was possible that TA > CG mutations were linked to proximity to methylation at CpNpG sites. Of course, defining a sequence context as either 'near' or 'far' is subjective. Another possibility was that TA > CG mutations across N+1, 3 and 4 contexts were all linked to methylation at CpNpG sites and occurred in a patch around the initial damage site. Given that all TA > CG mutations occurred no further than N+4 base pairs away from a CpNpG site, it is hard to say whether these mutations were enriched at these sequence contexts or not. It is important to note that not all TA > CG mutations occurred in CpNpG N+1, 3 or 4 contexts; some occurred in the CpG C + 2 context. However, these occurred at very low frequency and therefore were not included in this analysis. Overall, it is not possible to accurately conclude whether TA > CG mutations were enriched near methylation at CpNpG sequences. The distribution of TA > CG mutations would also have been limited to mutations conferring resistance to rifampicin. Therefore, accurate determination of enrichment at CpNpG sites would require sequencing data of all mutations, not just those conferring rifampicin resistance.

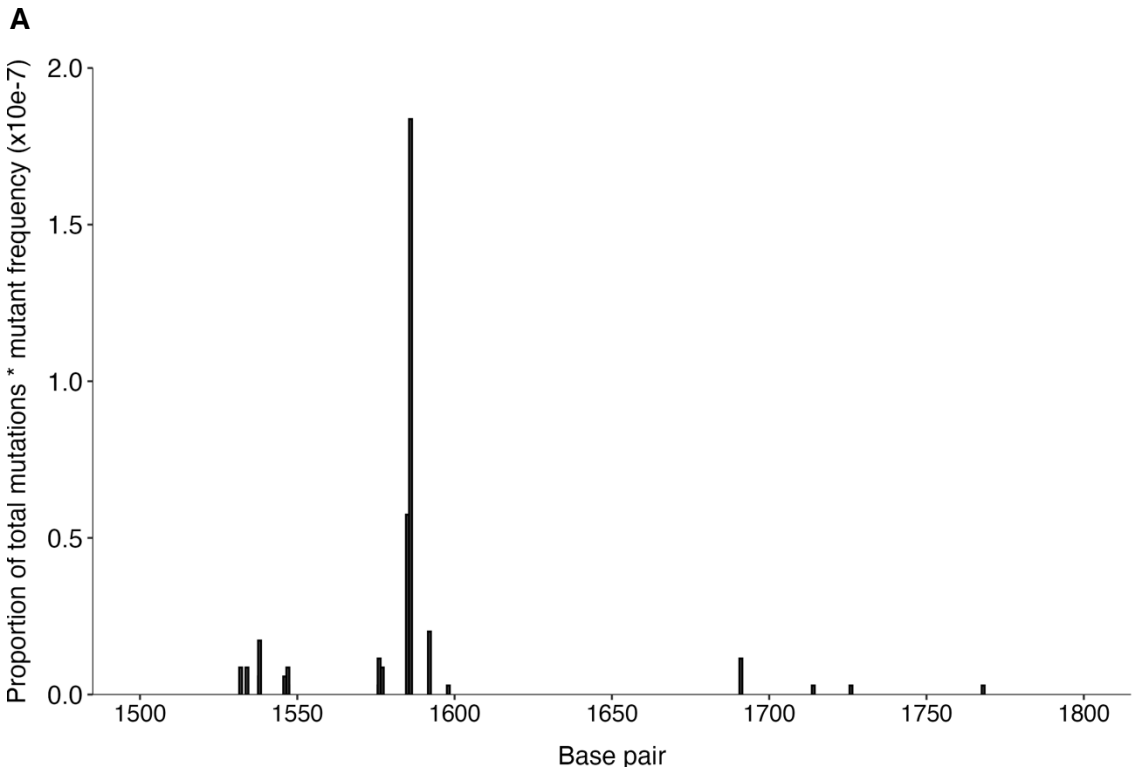
Table 3A: Frequency of TA > CG mutations occurring near CpNpG sequence contexts for WT and Δ*alkB* C2523 with active or inactive pM.SssI. Total TA > CG mutations for each condition are shown in the right-most column of each section.

	N+1			N+3			N+4	
	T>C 1532	A>G 1538	Total	T>C 1534	T>C 1598	Total	A>G 1547	Total
WT active M.SssI	3 0	0	3	3	1	4	3	3
WT inactive M.SssI	1 0	0	1	3	0	3	12	12
Δ*alkB* active M.SssI	16	18	34	29	0	29	33	33
Δ*alkB* inactive M.SssI	5 7	7	12	14	0	14	23	23

Table 3B: Normalisation of TA > CG mutations occurring near CpNpG sequence contexts for WT and Δ*alkB* C2523 with active or inactive pM.SssI. Mutations are defined by the base pair in *rpoB* where they occurred. They are categorised as N+1, N+3 or N+4 depending on how far they occurred from the central nucleotide in the CpNpG context. The proportion of CG > TA mutations for each strain was calculated as a percentage of all CG > TA mutations across conditions. This proportion was then normalised by multiplying by mean rifampicin-resistant mutation frequency for each condition. Fold change was calculated by dividing the active/inactive normalised value.

	N+1		N+3		N+4	
	Proportion * rifampicin-resistant mutation frequency	Active / inactive M.SssI fold change	Proportion * rifampicin-resistant mutation frequency	Active / Inactive M.SssI fold change	Proportion * rifampicin-resistant mutation frequency	Active / inactive M.SssI fold change
WT active M.SssI	0.049	21.199	0.066	9.422	0.049	1.767
WT inactive M.SssI	0.002		0.007		0.028	
Δ*alkB* active M.SssI	25.457	16.044	21.713	11.730	24.708	8.215
Δ*alkB* inactive M.SssI	1.587		1.851		3.041	

This potential bias posed the question: were the observed mutations a result of M.SssI preference or because they occurred at a location that caused rifampicin resistance? To answer this, I considered the distribution of mutations across *rpoB*. Figure 4.26 displays the distribution of all mutations, regardless of substitution type, across the sequenced region of *rpoB* normalised to mean rifampicin-resistant mutation frequency, for WT (A) and Δ*alkB* (B) C2523 with constitutively active pM.SssI. Mutations in WT cells predominantly occurred at the CpG site at 1585bp. This was expected, given that the main source of mutations in WT cells with active M.SssI would be spontaneous deamination of 5mC resulting in CG > TA mutations. Mutational distribution in Δ*alkB* cells was more spread out across the region, with a prominent peak at 1585bp owing to deamination of 5mC, as well as a cluster of peaks further upstream. It makes sense that Δ*alkB* cells might have had a broader distribution of mutations given that they were repair deficient .



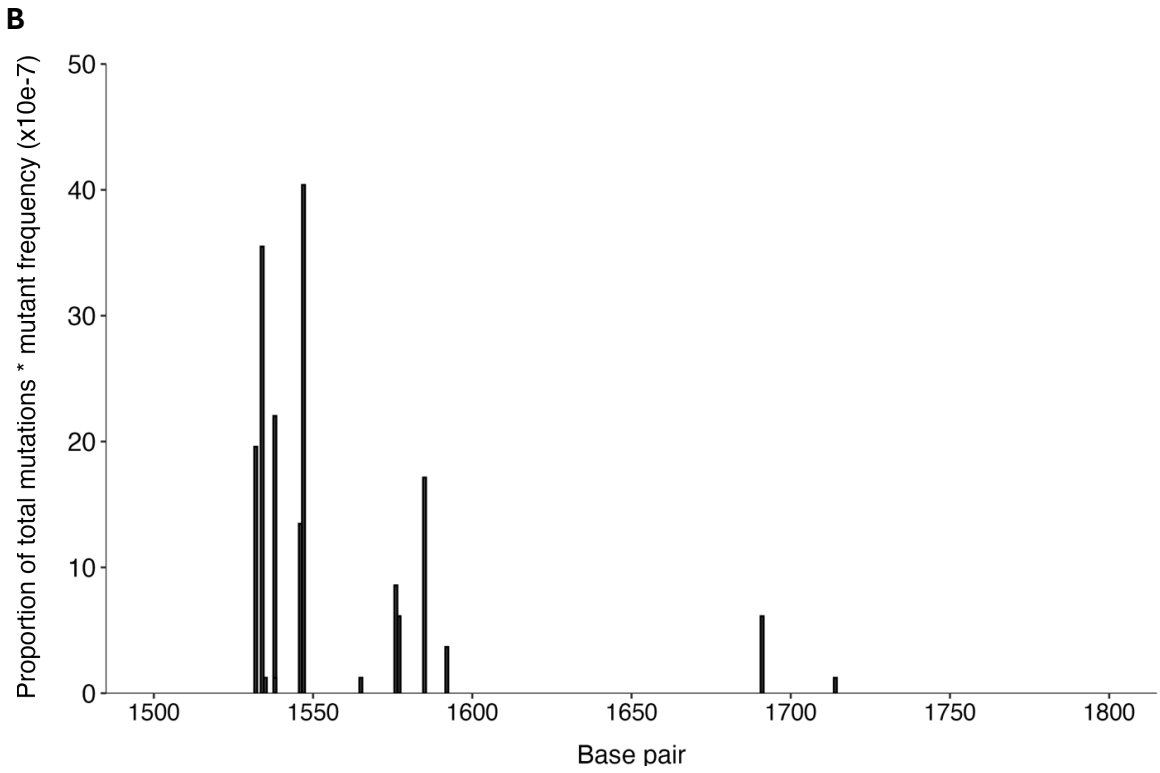
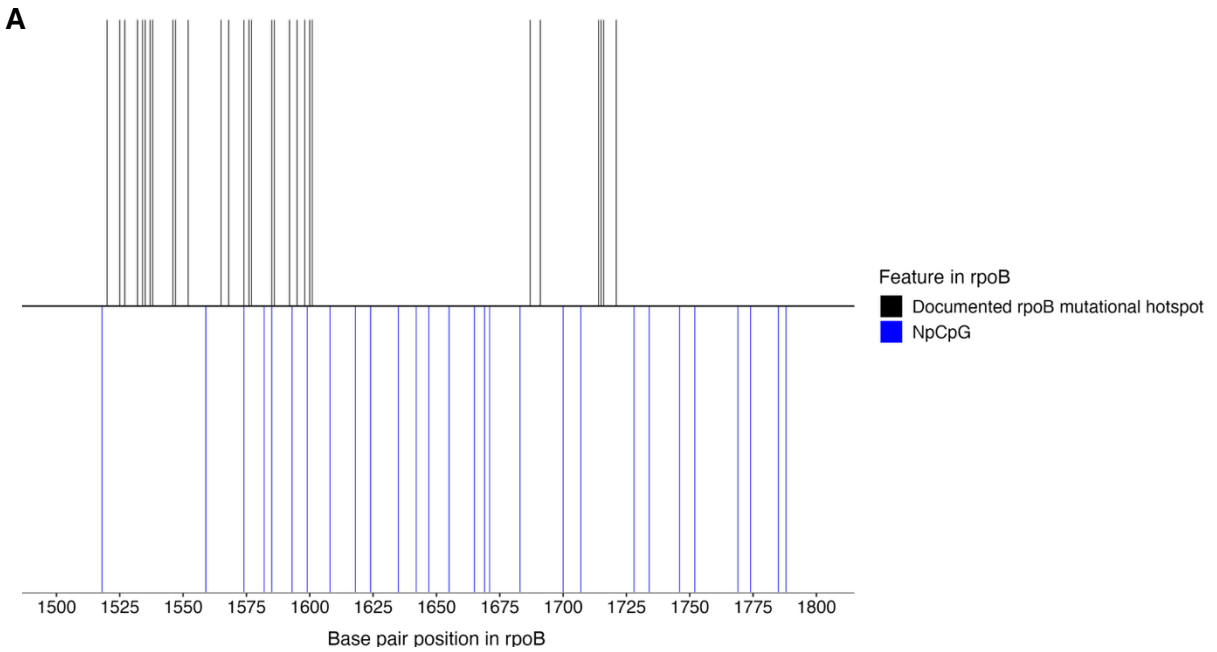


Figure 4.26: Distribution of all observed mutations in sequenced region of *rpoB* for C2523 with active pM.SssI. All mutations across substitution types were pooled for WT **(A)** and Δ*alkB* **(B)** C2523 with constitutively active pM.SssI. The proportion of each mutation at each base pair position was normalised to mean rifampicin-resistant mutation frequency for that either WT **(A)** or Δ*alkB* **(B)** C2523 + active pM.SssI calculated from rifampicin resistance assays.

Evidently, there was an uneven distribution of mutations across *rpoB*, suggesting a bias of either M.SssI or rifampicin resistance. To investigate this further, I compared all documented mutations (shown in black) (Garibyan et al., 2003) that give rise to resistance over the sequenced ~300bp region of *rpoB*, as well as all NpCpG sites within that region (shown in blue) (Figure 4.27A). There is a single CpG site at 1585bp that, when mutated, gives rise to resistance. Unsurprisingly, this CpG dinucleotide was the most prevalent location for mutations from samples with DNMT activity (Figure 4.26). This suggested that the uneven distribution of mutations was due to a

bias in mutations conferring resistance to rifampicin, not a preference of M.SssI. Sequencing data from the rifampicin resistance assay revealed other, less frequently mutated CpG dinucleotides. Interestingly, these occurred at CpG sites that had not previously been identified as sites for rifampicin resistance mutations to emerge.

Figure 4.27B shows NpCpG sites (blue) compared to sites where TA > CG mutations were observed in the sequencing data (black). The percentage of observed TA > CG mutations at each site was normalised to the mean rifampicin-resistant mutation frequency for Δ*alkB* C2523 + active pM.SssI to show bias for particular locations. TA > CG mutations were clustered in mutational cluster I of *rpoB* (base pairs 1519-1599). Most of these sites were between 1531-1554bp and were not proximal to any CpG dinucleotides. Two mutated sites at 1577bp and 1598bp were within 3 bp of a cytosine in a CpG dinucleotide. However, proximity to CpG did not increase the mutagenicity of the TA base pair, in fact, TA base pairs further from CpG sites were more frequently mutated in Δ*alkB* C2523 with active and inactive pM.SssI. This suggested that mutagenicity did not arise directly from methylated CpG sites. However, we cannot rule out the possibility that methylation at CpG sites further upstream influenced mutagenicity.


A
Feature in rpoB
Documented rpoB mutational hotspot
NpCpG
1500
1525
1550
1575
1600
1625
1650
1675
1700
1725
1750
1775
1800
Base pair position in rpoB

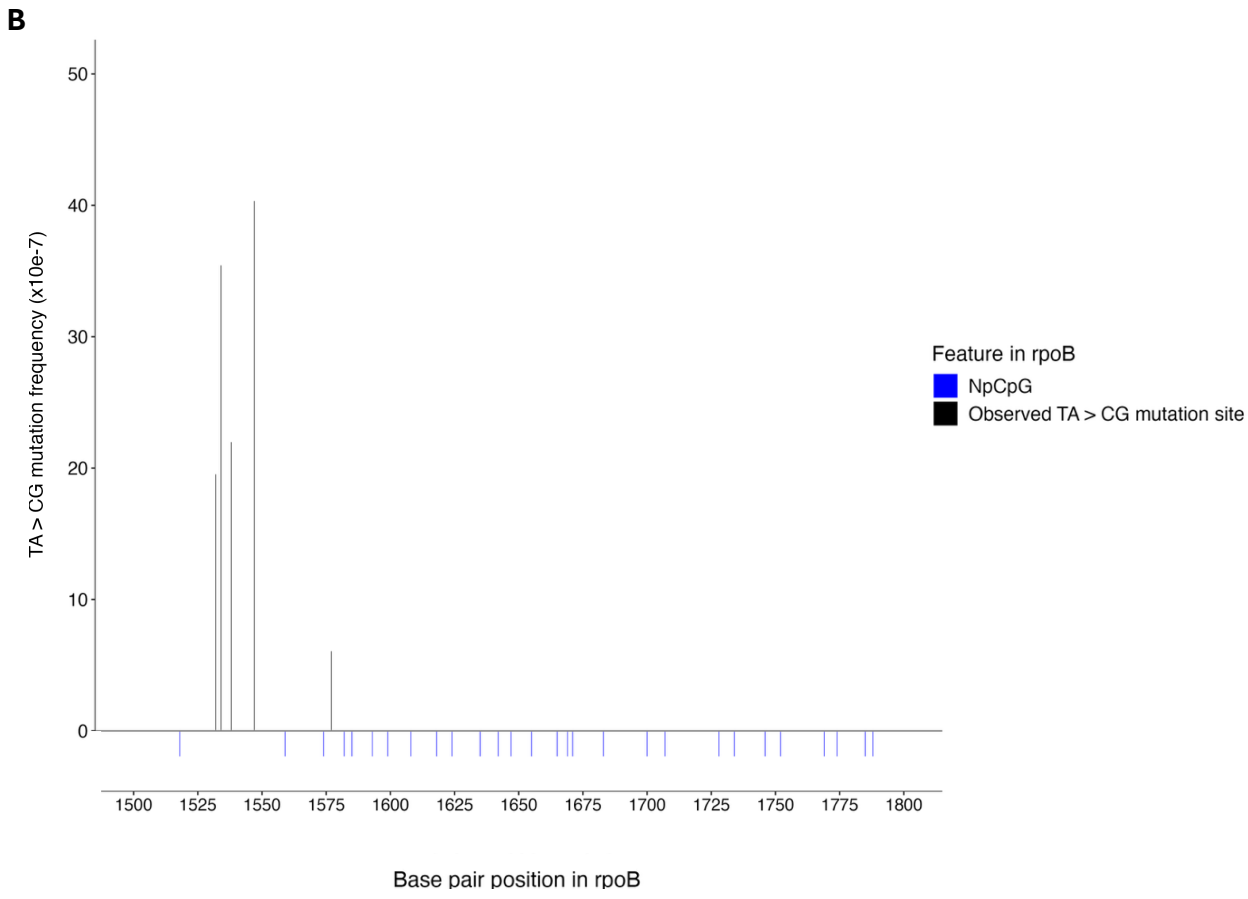

B
TA > CG mutation frequency (x10e-7)
50
40
30
20
10
0
Feature in rpoB
NpCpG
Observed TA > CG mutation site
1500
1525
1550
1575
1600
1625
1650
1675
1700
1725
1750
1775
1800
Base pair position in rpoB

Figure 4.27: Location of CpG dinucleotides in sequenced region of *rpoB* vs mutational hotspots. (A) Location of NpCpG sites (blue) where base pair corresponds to cytosine in NpCpG, compared to documented mutational hotspots that give rise to rifampicin resistance (black). **(B)** Location of NpCpG dinucleotides (blue) compared to observed TA > CG mutational hotspots (black) in sequencing data from rifampicin resistance assays. The percentage contribution of each TA > CG mutation to the overall *rpoB*-specific mutation spectrum has been normalised to mean rifampicin-resistant mutation frequency for Δ*alkB* + active pM.SssI in C2523 *E. coli*.

4.3.1.3. Alternative analysis methods

Broadly, there are two main experimental procedures for determination of mutation rate: mutation accumulation and fluctuation tests. During a mutation accumulation experiment, a single mutant from each generation is picked and allowed to proliferate before the same population bottleneck process is repeated (Mukai, 1964; Shaw et al., 2000). This minimises selection and allows for propagation of random, spontaneous mutations, including deleterious mutations that would normally be eliminated by selection. However, these experiments are time-consuming and rely on the assumption that growth rates of all mutants are the same. This is often not the case, as often resistant mutations confer a fitness effect so mutants will accumulate at a slower rate than the parent (Billington et al., 1999). As a result, fluctuation tests are often preferred over mutation accumulation experiments. The first fluctuation test was conducted by Luria and Delbruck in 1943. This experiment was designed to distinguish between pre- and post-adaptive mutation; put simply, whether mutagenesis occurs pre- or post- exposure to selection pressure. Here, a large number of parallel, independent cultures were set

up and allowed to grow. Then, a 'challenge' was introduced in the form of a bacteriophage. Following exposure, bacteria were plated and the number of surviving colonies was counted. If a spontaneous mutation occurred during growth that conferred resistance, then the cells containing that mutation would survive. This would result in highly variable or 'fluctuating' numbers of mutants between independent cultures. On the other hand, if mutation was induced by the bacteriophage, then one would expect roughly equal numbers of mutants for all cultures, given that their exposure was equal. This would result in a Poisson distribution of survivors, where mean was equal to variance. However, they found that the variance was higher than the mean, suggesting that mutations were pre-adaptive. This led to the creation of a new distribution, described by the Luria-Delbruck distribution, which described the theoretical distribution of mutant clones in a generation (Luria & Delbrück, 1943).

Fluctuation tests allow for the calculation of the number of mutations per culture (m), which can be converted to mutation rate depending on the number of cells in the population and the number of generations that have passed. There are several methods for estimation of mutation rate, many of which are based on the original Luria-Delbruck distribution (Rosche & Foster, 2000). Choice of analysis method depends on an initial estimation of m; p_0 is commonly chosen for low mutation rates (m = 0.3-2.3), while the method of the median suits intermediate mutation rates (m = 1-15). The MSS maximum-likelihood method is more mathematically complex but is the superior method for estimation of all mutation rates. When designing a fluctuation test, experimental conditions should be optimised to return m within the

suitable range of the desired analysis method. Other experimental considerations include the number of parallel cultures (C); a higher number of cultures ensures a more accurate estimation of *m*. It is also important that all cultures contain the same number of cells at the mutant selection stage (N_t) and that this is much higher than the initial number of cells (N_0). Furthermore, fluctuation assays rely on a number of assumptions, for example mutation rate per cell lifetime remains constant and does not vary throughout the growth of the culture. Growth rate is also assumed to remain constant throughout the assay (Pope et al., 2008; Rosche & Foster, 2000).

The rifampicin resistance assay used in this thesis was experimentally based upon a fluctuation test, although differed in a few key ways. Each condition, for example WT C2523 with active pM.SssI, was made up of several independent replicates, essentially forming a fluctuation test. However, the number of initial cultures was much lower than would be used in a fluctuation test. With both WT and Δ*alkB* backgrounds and two forms of pM.SssI, as well as other plasmids and culture conditions, a higher number of cultures would not have been feasible. Furthermore, the assumptions concerning mutation and growth rate would certainly not have been satisfied in the case of the inducible system, where induction of M.SssI would change mutation and subsequently growth rate. This may also be true for the constitutive system, where prolonged M.SssI activity slowed growth. Controlling for changes in growth rate throughout the rifampicin resistance assay would have been complex and may have affected accuracy of final rifampicin-resistant mutation frequency estimates. While these differences meant that it was not possible to

estimate the mutation rate via any associated analysis methods, using the fluctuation test as an experimental framework provided a basis for studying mutagenesis via selection. The aim of this thesis was not to estimate absolute mutation rate, but instead to compare mutagenesis between conditions. Therefore, estimation of a relative rifampicin-resistant mutation frequency was sufficient. However, designing a rifampicin resistance assay that more accurately mirrors a fluctuation test, combined with downstream analysis by the p_o method, could be a useful future experiment to further investigate the observed changes to mutagenesis.

4.3.2. Potential genetic and cellular contributions to the non-cytosine *rpoB*-specific mutation spectrum

4.3.2.1. Non-cytosine mutations were independent of RecA-mediated recombination

Several publications indicated the importance of TLS in bypass of small alkylation damage in various model organisms (Anderson et al., 2024; Khodaverdian et al., 2024; Volkova et al., 2020). Comparison of C2523 and Top10 genotypes highlighted RecA as a potential factor in mutagenesis outside of the CpG context. However, extensive investigation suggested that RecA alone was not responsible for the non-cytosine *rpoB*-specific mutation spectrum, regardless of the bacterial strain or duration of M.SssI activity. Before considering alternative explanations, it is important to note that expression levels of RecA were not quantified during the

assay. Therefore, we cannot be certain that RecA expression from pMP1 was sufficient for nucleofilament formation and induction of TLS. However, presence of pMP1 alone did raise rifampicin-resistant mutation frequency (Figure 4.17C, D) and produce a distinctive *rpoB*-specific mutation spectrum (Figure 4.18A, B), so RecA was almost certainly expressed in some capacity. Similarly, inhibition of RecA by PsiB was not quantified. PsiB directly binds to and inhibits RecA (Petrova et al., 2009), therefore neither gene expression nor protein levels would have decreased as a result of PsiB expression. In order to quantify inhibition, it would be necessary to either directly measure the catalytic activity of RecA or indirectly assess the effect of PsiB on TLS induction. However, the presence of pPsiB had no effect on mutagenesis. It is not possible to say whether this was because RecA was not responsible for the TA > CG signature, or because PsiB was not sufficiently expressed. However, taken together with the pMP1 experiments, it seems likely that RecA and subsequently TLS was not responsible.

4.3.2.2. SulA is required for regulation of cell division during DNA damage bypass

If not RecA, then what else could be responsible for the non-cytosine *rpoB*-specific mutation spectrum? Re-examination of C2523 and Top10 genotypes revealed another potential candidate. SulA is a cell division inhibitor that is induced when cells experience DNA damage and during the SOS response, to allow time for bypass of DNA damage. *SulA,* otherwise known as *sfiA,* is normally repressed by LexA. Activated RecA induces autocleavage of LexA and derepresses SOS response genes including *sulA* (Bi & Lutkenhaus, 1993; Huisman & D'Ari, 1981). Top10 possess

WT *sulA*, while C2523 have *sulA11* and therefore have reduced ability to halt cell division during DNA damage repair or bypass. As a result, C2523 may have a dysregulated SOS response that could increase mutagenesis, potentially causing the strong TA > CG signature in C2523 alone. Furthermore, C2523 also have a mutation in their Lon protease. This regulates SOS response proteins and SulA turnover (Higashitani et al., 1997; Mizusawa & Gottesman, 1983). Therefore, non-functional Lon protease may further contribute to SOS response dysregulation and genomic instability under stress. Perhaps DNMT-induced damage in C2523 triggers the SOS response, but cells struggle to finish TLS properly, leading to an 'abnormal' *rpoB*-specific mutation spectrum that is not seen with WT *sulA*. Further investigation into the contribution of SulA could be difficult, as this would require either KO of *sulA* in Top10 in combination with *alkB* KO and RecA supplementation, or re-introducing SulA and Lon protease back into C2523. Overall, the genotypic differences between C2523 and Top10 are likely more nuanced than just RecA alone and may have contributed to observed mutagenesis.

4.3.2.3. Single-stranded DNA resulting from DNMT-induced alkylation damage could contribute to mutagenesis

In section 1.1.4.2. I introduced the influence of ssDNA and dsDNA on repair of alkylation lesions by AlkB. In summary, ssDNA is inherently more mutagenic than dsDNA as the organic bases are unprotected, therefore, they are more susceptible to attack by genotoxic agents (Trewick et al., 2002). For example, it has been shown that MMS is approximately 20000 times more mutagenic at DSBs, which are

associated with extensive ssDNA, than at non-DSB locations. Furthermore, the primary lesion responsible for this mutagenesis was single strand-specific 3mC and subsequent bypass by TLS (Yang et al., 2010). In terms of repair, AlkB is biased towards repair of ssDNA than dsDNA, as dsDNA requires an extra step where the damaged base must be flipped out of the helix. It is worth noting that AlkB does still possess dsDNA repair capacity (Yang et al., 2008; Yu & Hunt, 2009). Given the increased mutagenicity of ssDNA and preferential repair of ssDNA by AlkB, it was possible that ssDNA contributed to DNMT-dependent mutagenesis.

One of the limitations of the approach in this thesis is that it was not always possible to identify the original strand on which the DNA damage occurred. For example, CG > TA mutations associated with DNMT activity always occurred on cytosine within a CpG dinucleotide due to the specificity of M.SssI, allowing for identification of the damaged strand. However, it was not possible to tell which strand DNMT-dependent TA > CG mutations occurred on, given that these occurred in a non-CpG context. It was also not possible to know whether that damage occurred on dsDNA or ssDNA, although this is not specifically a limitation of the rifampicin resistance assay and would likely require some sort of imaging or tagging approach. According to the literature, M.SssI possesses both *de novo* and maintenance DNA methyltransferase activity, although it is not clear whether it can methylate CpG dinucleotides in ssDNA as well as dsDNA (Renbaum et al., 1990). Therefore, I assume that alkylation lesions predominantly occurred on dsDNA. Any lesions on ssDNA were likely the result of DNA Polymerase stalling on dsDNA, then reversing upstream of the lesion or re-priming downstream, both of which would leave a ssDNA gap containing the

lesion (Heller & Marians, 2006). Given that AlkB is more efficient at repair of lesions on ssDNA (Yang et al., 2008), these may have been preferentially repaired in WT cells. In the absence of AlkB, alkylation lesions on dsDNA would typically be repaired by BER or other strand excision pathways, resulting in a ssDNA gap (Admiraal et al., 2019; Drabløs et al., 2004; Lindahl, 1993). If the dsDNA lesion resulted in stalling of the replisome and subsequent ssDNA gap formation prior to repair, then strand excision could result in a DSB. Together, M.SssI activity would increase ssDNA levels in the absence of AlkB. Therefore, M.SssI and Δ*alkB*-dependent mutagenesis may have been an indirect effect of ssDNA accumulation caused by M.SssI-induced alkylation lesions. This may explain why we saw an increase in non-cytosine mutations in some backgrounds; these could be the result of erroneous incorporation of nucleotides around the lesion site in stretches of ssDNA. Further, BER may be the preferred backup repair pathway in the absence of AlkB, which would explain why RecA involvement yielded negative results. The precise mechanism by which M.SssI-induced lesions were mutagenic and their relationship with ssDNA remains to be elucidated.

Chapter 5 – DNA methylation contributes to reactive oxygen species production in *E. coli*

5.1. General introduction

The work in the previous chapter showed that DNMT activity was potentially mutagenic in the absence of AlkB, which may be part of the reason why DNMTs have been lost in some lineages. DNMT-induced endogenous DNA damage and mutagenicity might be particularly costly if the organism were also to be exposed to exogenous DNA damaging agents. Investigation found that *E. coli* with constitutive M.SssI expression were hypersensitive to the alkylating agent MMS, particularly in Δ*alkB* mutants (Krwawicz et al., 2024). Interestingly, WT cells with constitutive M.SssI activity were also hypersensitive to oxidative stress caused by hydrogen peroxide (H_2O_2). Oxidative stress occurs when there is a greater level of oxidants than antioxidants present in the cell. Oxidants include ROS, which are highly reactive, oxygen-based free radicals that possess one or more spare electron, for example hydroxyl and superoxide radicals. There are also non-radical ROS, which include oxygen derivatives H_2O_2 and O_3 (Sies et al., 2017). H_2O_2 is relatively stable, however, can generate free radicals via reaction with metal ions (Halliwell & Gutteridge, 2015) or light (Buxton et al., 1988; Pignatello et al., 2006). Furthermore, superoxide dismutase catalyses the formation of H_2O_2 from O_2-. While H_2O_2 is a common metabolic byproduct and plays a role in cell signalling (Holmström & Finkel, 2014), disruptions to this reaction equilibrium can result in increased ROS production (Wang et al., 2018). Free radicals pose a significant threat to DNA as they

oxidise organic bases and generate several toxic DNA lesions, including oxidised purines (Collins, 1999). These are excised by the BER glycosylase Fpg in *E. coli* (Boiteux et al., 1987; Chetsanga & Lindahl, 1979). M.SssI activity combined with *fpg* deletion resulted in a synthetic increase in sensitivity to oxidative stress, suggesting that M.SssI expression itself might increase oxidative stress levels in the cell (Δ*fpg* * DNMT interaction p = 0.032, two-way Anova) (Krwawicz et al., 2024). This presented another way in which CpG methyltransferase activity could potentially damage DNA and cause genomic instability.

Oxidative stress caused by constitutive M.SssI activity in Krwawicz et al. was determined by measuring ROS levels using a commercially available fluorescent sensor that binds to a range of ROS. Given previously observed differences in mutagenicity between constitutive and inducible M.SssI systems, I decided to measure ROS levels as a result of inducible expression of M.SssI. Furthermore, I was particularly interested in the prior observation that constitutive M.SssI activity increased ROS levels independent of H_2O_2-induced oxidative stress. Could this be replicated using the inducible system. If so, was this a direct or indirect consequence of M.SssI activity?

5.2. Results

5.2.1. *In vivo* evidence for the introduction of oxidative stress by M.SssI

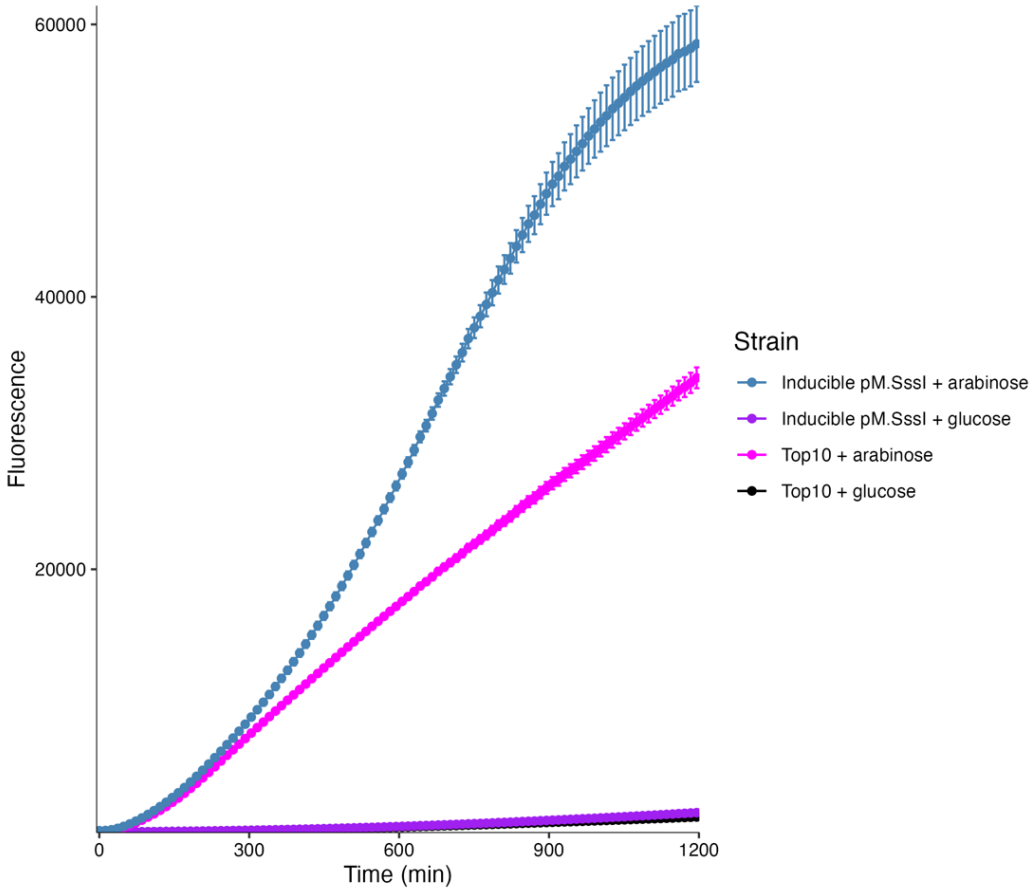
5.2.1.1. Induction of M.SssI activity alone increased ROS signal in *E. coli*

WT Top10 ± inducible pM.SssI were grown overnight in 0.5% glucose, then diluted and grown to OD600 = 0.4. Then, cultures were split and pellets re-suspended in either 0.5% arabinose or glucose. Cultures were grown to OD600 = 1.0 then aliquoted into a 96-well plate. Fluorescence and absorbance were measured every ten minutes for approximately eight hours. Each well contained a fluorescent sensor that is oxidised by a range of ROS molecules and reports the non-specific redox-sensitive signal in the population of cells. It should be noted that this sensor does not report the concentration of specific types of ROS nor does it measure oxidative damage. However, it was suitable for this assay given that we were not testing for a specific ROS molecule and were interested in overall ROS levels as a result of M.SssI activity.

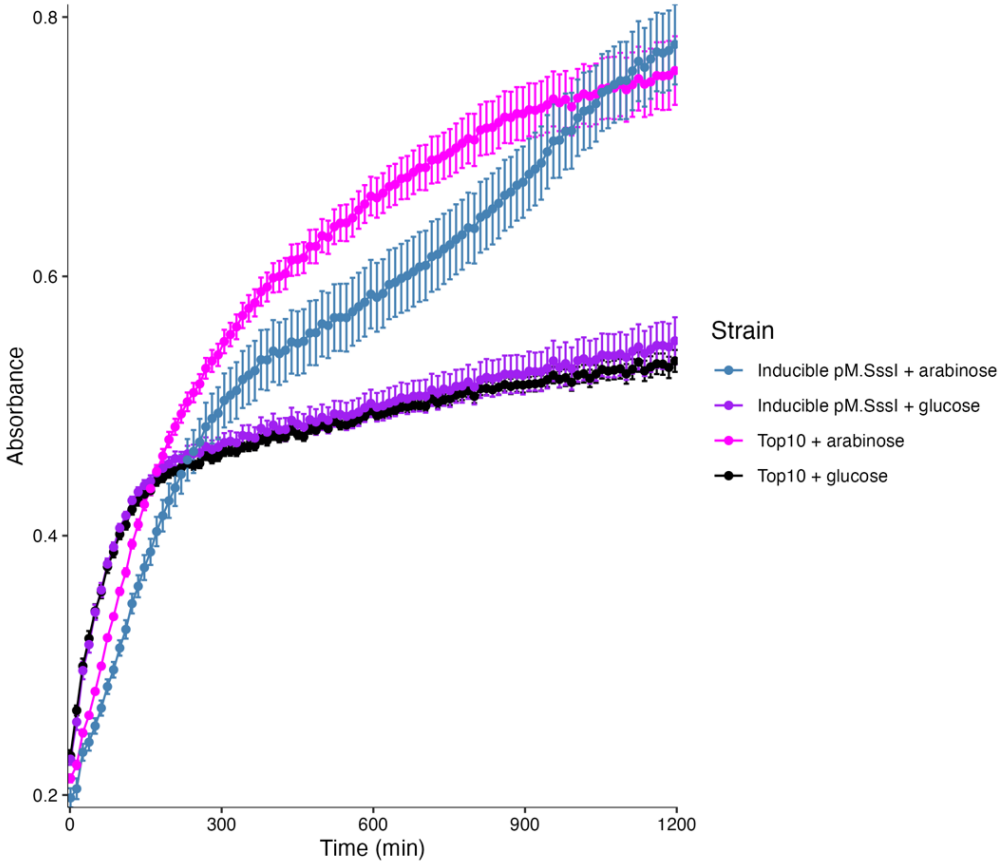
Mean fluorescence over time for WT Top10 ± inducible pM.SssI is shown in Figure 5.1A and mean absorbance in Figure 5.1B. Mean fluorescence was normalised to the number of cells present at each time point (calculated from mean absorbance) to give fluorescence per cell per ml (Figure 5.1D). Presence of DNA methylation was also determined via HpaII digest of plasmid DNA (Figure 5.1C). WT Top10 ± inducible pM.SssI treated with glucose displayed very low, constant fluorescence throughout

the assay (Figure 5.1D). This was expected as M.SssI was not present, so there was no driver of ROS signal. Induction of pM.SssI with arabinose increased fluorescence and therefore ROS levels throughout the assay as expected. This was interesting, as previous work by Krwawicz et al. suggested that M.SssI exacerbated sensitivity to oxidative stress, indicating that M.SssI might have contributed to ROS in the cell (Krwawicz et al., 2024). The data in Figure 5.1D suggested that M.SssI activity alone without additional oxidative stress was sufficient to increase ROS levels. Unexpectedly, treatment of WT Top10 - inducible pM.SssI with arabinose also increased ROS levels throughout the assay, albeit at a lesser extent than WT Top10 + inducible pM.SssI treated with arabinose, despite not expressing M.SssI and therefore lacking a driver of ROS.

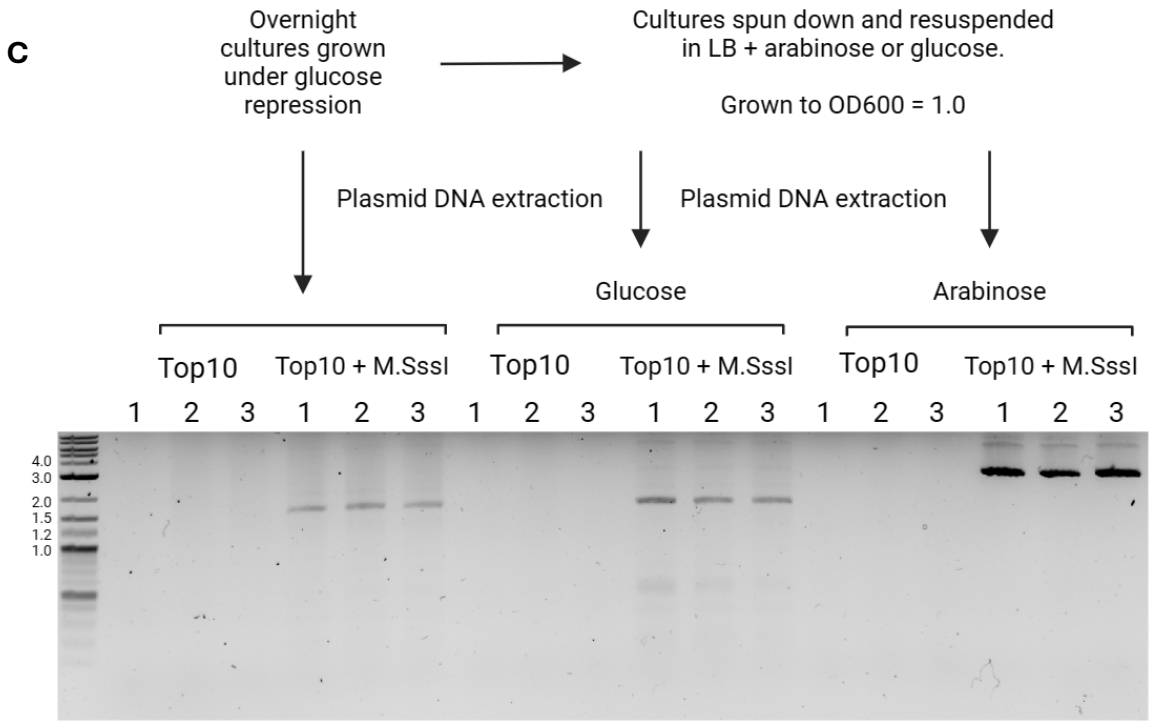
A



B


0.8
0.6
0.4
0.2
Absorbance
0
300
600
900
1200
Time (min)
Strain
Inducible pM.SssI + arabinose
Inducible pM.SssI + glucose
Top10 + arabinose
Top10 + glucose

C


Overnight cultures grown under glucose repression
Cultures spun down and resuspended in LB + arabinose or glucose.
Grown to OD600 = 1.0
Plasmid DNA extraction
Plasmid DNA extraction
Glucose
Arabinose
Top10
Top10 + M.SssI
Top10
Top10 + M.SssI
Top10
Top10 + M.SssI
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
4.0
3.0
2.0
1.5
1.2
1.0

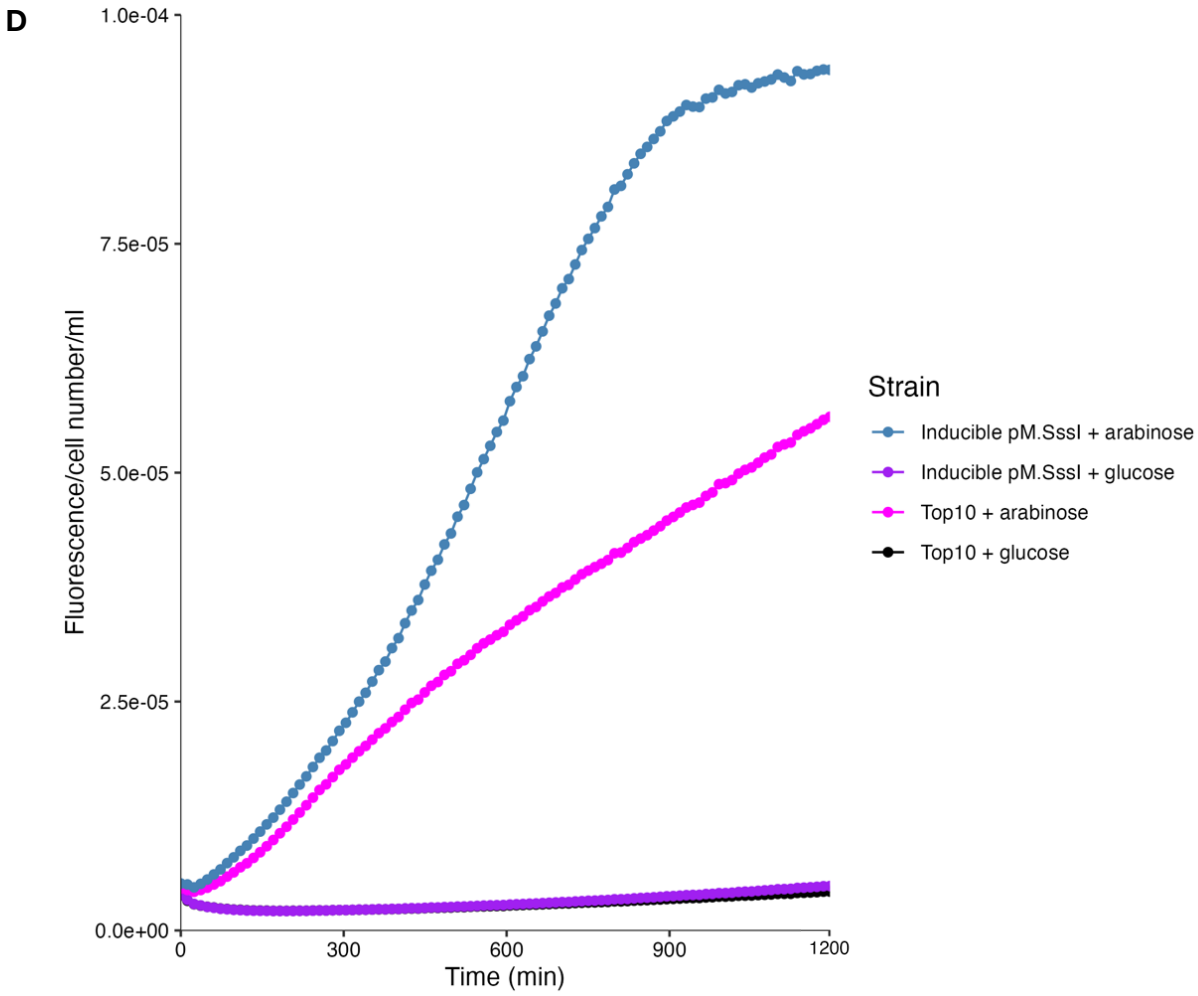
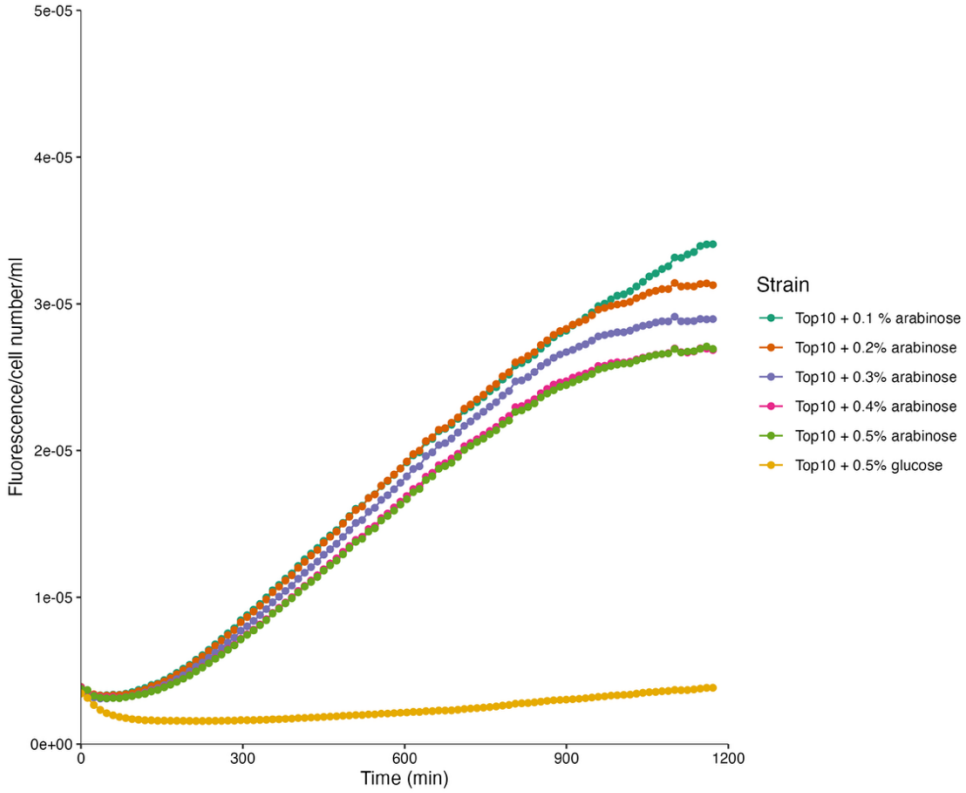


Figure 5.1: Induction of M.SssI activity increased ROS levels in WT Top10. (A) WT Top10 ± inducible pM.SssI treated with either 0.5% arabinose or glucose were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated to fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. Mean fluorescence for each strain was calculated over three independent replicate clones, each averaged over two technical replicates. Bars represent standard deviation at each time point. **(B)** Absorbance at OD600 was measured over time for WT Top10 ± inducible pM.SssI + 0.5% glucose/arabinose. Mean absorbance was calculated over three independent replicate clones, each averaged over two technical replicates. Bars represent standard deviation at each time point. **(C)** Plasmid DNA was extracted from WT Top10 ± inducible pM.SssI + 0.5% arabinose/glucose at the start (glucose-treated overnight culture) and end (OD600 = 1.0) of the assay. Plasmid DNA was digested with HpaII and run on a 1% agarose gel. **(D)** Mean fluorescence was normalised to mean absorbance at each time point.

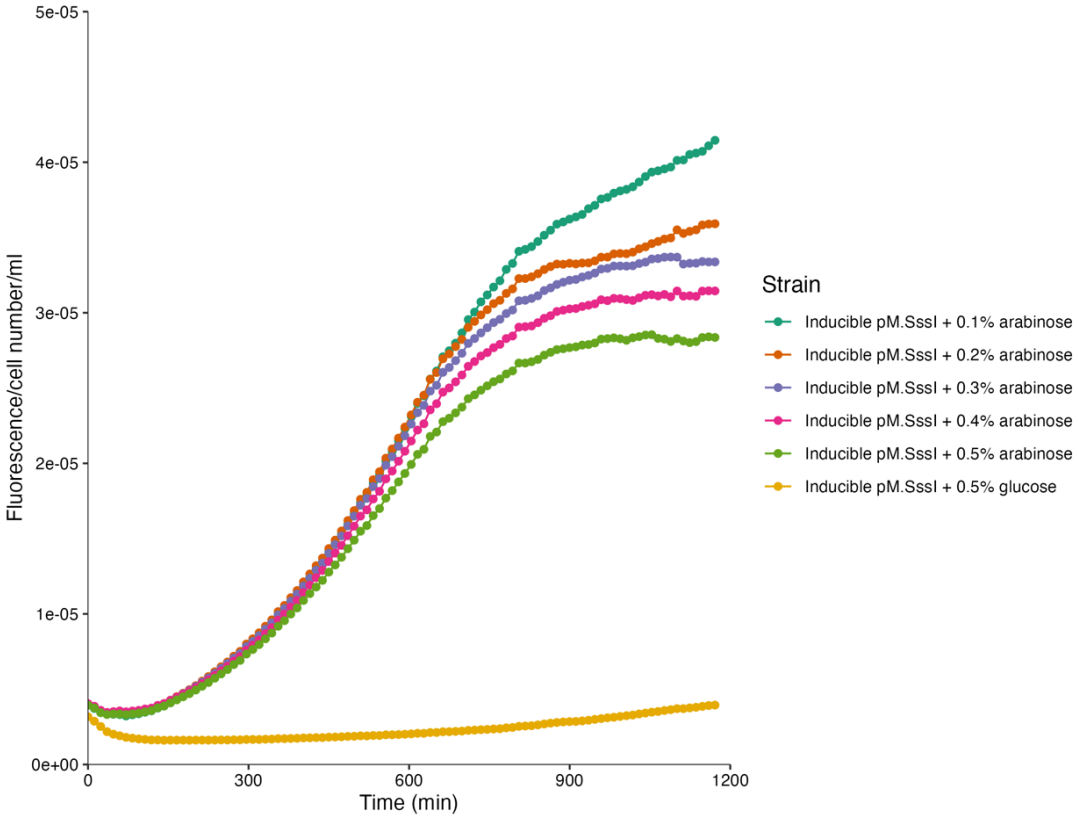
5.2.1.2. Investigating the contributions of experimental conditions to baseline ROS signal

While it was exciting that M.SssI activity alone increased ROS levels, it was concerning that a smaller but still notable increase in ROS levels was also observed for WT Top10 + arabinose. One explanation could be that arabinose itself led to an increase in ROS levels, which would be difficult to distinguish from the increase in ROS caused by M.SssI activity. To identify ROS caused by M.SssI activity only, I needed to find a concentration of arabinose that induced M.SssI expression but also reduced background ROS. Therefore, I conducted an arabinose titration. Here, WT Top10 ± inducible pM.SssI were treated with 0.1 - 0.5% arabinose, as well as a 0.5% glucose only control, and ROS measured over time as above. Fluorescence was normalised to absorbance to account for any differences in growth rate. Unexpectedly, increasing concentration of arabinose in WT Top10 ± inducible pM.SssI (Figure 5.2A, B) decreased ROS signal. HpaII digests suggested that DNA methylation was induced across all concentrations of arabinose in samples containing inducible pM.SssI (Figure 5.2C).

A


5e-05
4e-05
3e-05
2e-05
1e-05
0e+00
Fluorescence/cell number/ml
0
300
600
900
1200
Time (min)
Strain
Top10 + 0.1 % arabinose
Top10 + 0.2% arabinose
Top10 + 0.3% arabinose
Top10 + 0.4% arabinose
Top10 + 0.5% arabinose
Top10 + 0.5% glucose

B


5e-05
4e-05
3e-05
2e-05
1e-05
0e+00
Fluorescence/cell number/ml
0
300
600
900
1200
Time (min)
Strain
Inducible pM.SssI + 0.1% arabinose
Inducible pM.SssI + 0.2% arabinose
Inducible pM.SssI + 0.3% arabinose
Inducible pM.SssI + 0.4% arabinose
Inducible pM.SssI + 0.5% arabinose
Inducible pM.SssI + 0.5% glucose

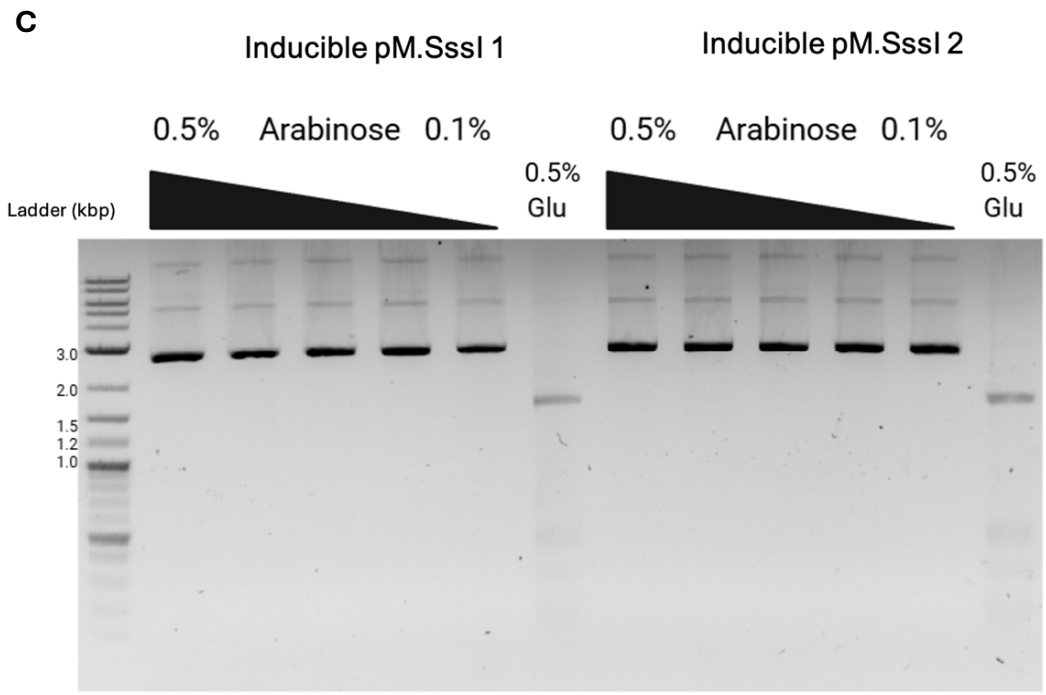


Figure 5.2: Increasing concentration of arabinose decreased background ROS levels. Two independent replicate clones of WT Top10 **(A)** or WT Top10 + inducible pM.SssI **(B)** were treated with 0.1%, 0.2%, 0.3%, 0.4% or 0.5% arabinose, or 0.5% glucose (control). Samples were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated to fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. Mean fluorescence and absorbance were calculated for each pair of replicates, each a mean average of two technical replicates, then mean fluorescence was normalised by mean absorbance. **(C)** Plasmid DNA was extracted from all WT Top10 ± inducible pM.SssI samples at the end point (OD600 = 1.0) of the assay. Plasmid DNA was digested with HpaII and run on a 1% agarose gel.

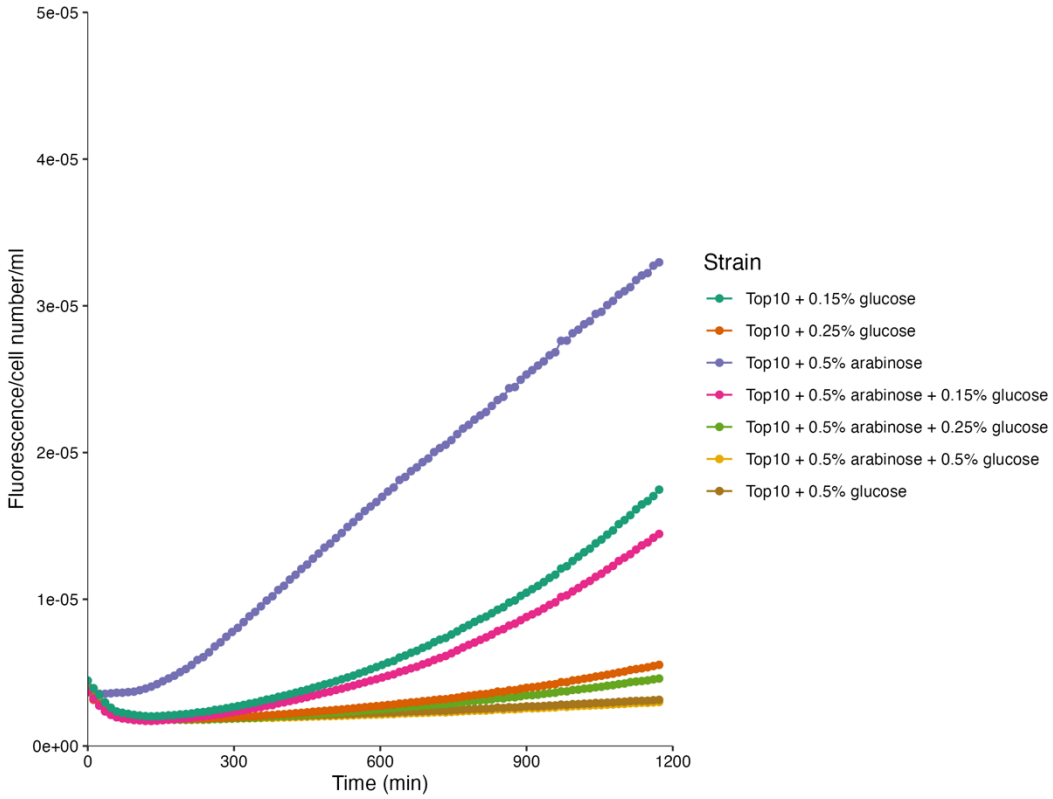
E. coli preferentially metabolise glucose (Loomis & Magasanik, 1967). Therefore, it was possible that a lack of glucose was responsible for the increase in ROS due to changes in cellular metabolism. I hypothesised that adding a small amount of glucose in combination with arabinose would reduce the background ROS signal while maintaining induction of M.SssI expression. To identify this combination, I conducted a glucose titration. This followed the same premise as the arabinose titration above. WT Top10 ± inducible pM.SssI were treated with 0.15%, 0.25% and

0.5% glucose ± 0.5% arabinose, as well as a 0.5% arabinose-only control. Crucially, each arabinose and glucose combination had an equivalent glucose-only control. This enabled the increase in ROS as a result of arabinose addition to be calculated.

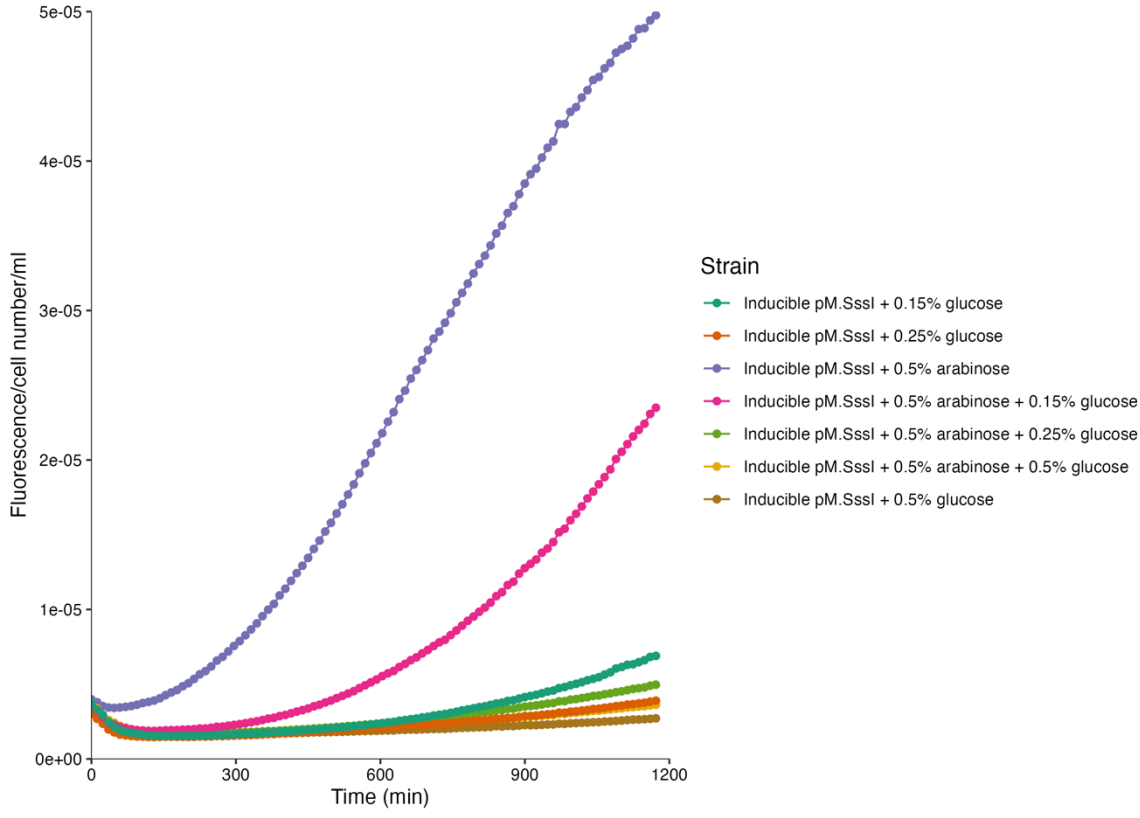
Figure 5.3 shows normalised fluorescence for Top10 (A) + inducible pM.SssI (B). Interestingly, HpaII digest of plasmid DNA suggested that DNA methylation was induced in all arabinose-treated samples regardless of glucose addition (Figure 5.3C). This suggested that presence of glucose did not override induction by arabinose, as suggested in the literature (Desai & Rao, 2010; Loomis & Magasanik, 1967). In Top10 ± inducible pM.SssI, 0.5% arabinose resulted in the highest ROS levels as previously observed. All glucose-only samples displayed negligible ROS levels regardless of presence of inducible pM.SssI as expected. In Top10, adding 0.15%, 0.25% or 0.5% glucose to 0.5% arabinose decreased ROS to a similar level to the equivalent glucose only control. This was expected as there was no driver of ROS present. This was also the case in Top10 + inducible pM.SssI treated with 0.25% or 0.5% glucose ± 0.5% arabinose. This suggested that these concentrations of glucose were too high in combination with arabinose to see the effect of induction of M.SssI activity on ROS levels. However, in Top10 + inducible pM.SssI, 0.5% arabinose + 0.15% glucose resulted in higher ROS levels than 0.15% glucose only. This difference is more clearly visualised in Figure 5.3D, where each combination was normalised to glucose-only background. It was clear that there was no change in ROS levels over time for all Top10 only samples, regardless of arabinose and glucose combination. In Top10 + inducible pM.SssI samples, the only combination that led to a marked increase in ROS was 0.5% arabinose + 0.15% glucose. This

indicated that 0.5% arabinose induced M.SssI expression and increased ROS levels, while the low concentration of glucose reduced background ROS.

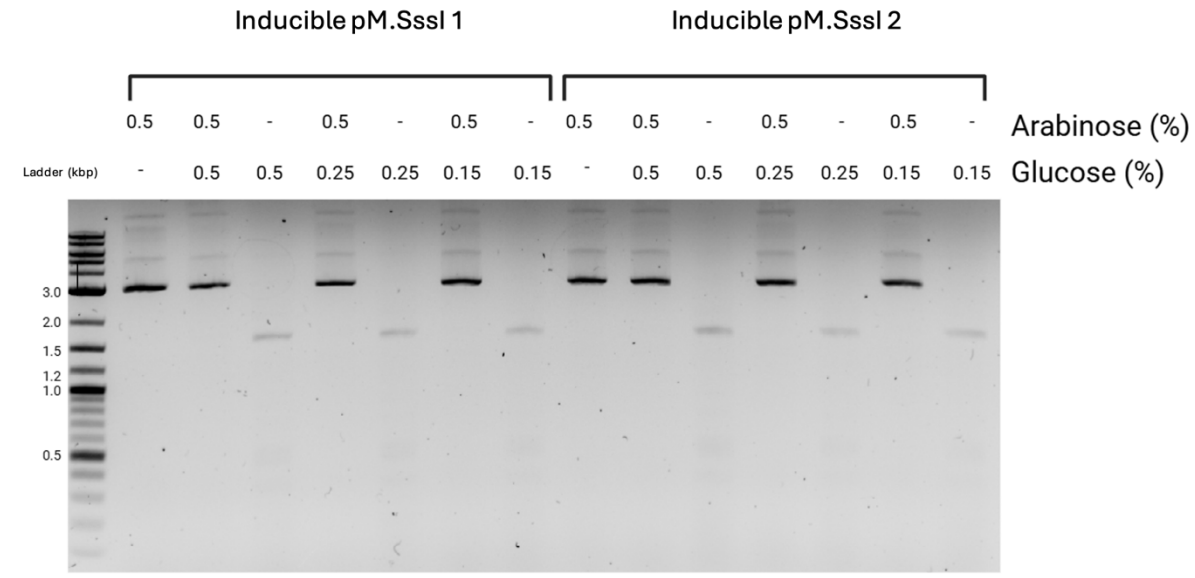
A



B



C


Inducible pM.SssI 1
Inducible pM.SssI 2
0.5 0.5 - 0.5 - 0.5 - 0.5 0.5 - 0.5 - 0.5 -
Arabinose (%)
Ladder (kbp) - 0.5 0.5 0.25 0.25 0.15 0.15 - 0.5 0.5 0.25 0.25 0.15 0.15
Glucose (%)
3.0
2.0
1.5
1.2
1.0
0.5

D

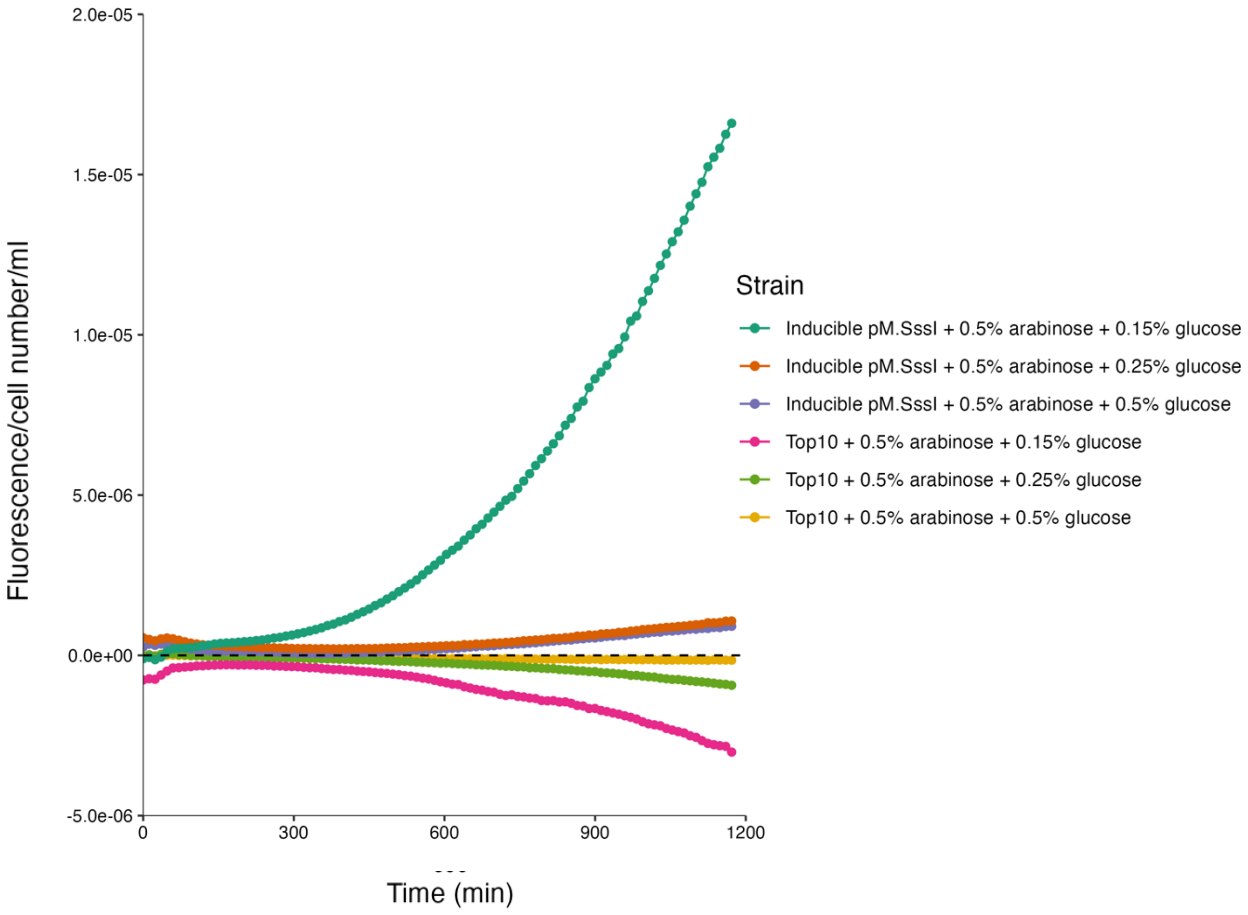
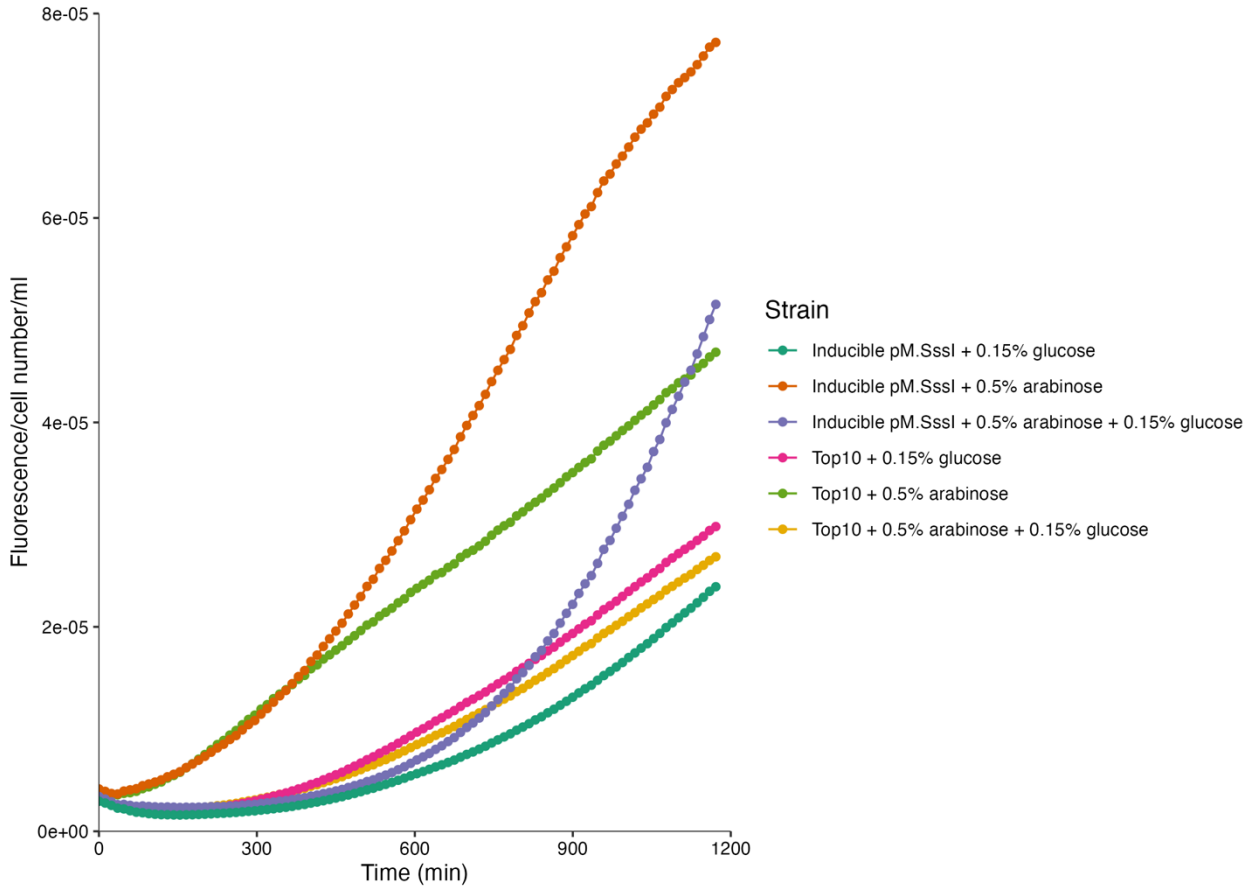

Fluorescence/cell number/ml
2.0e-05
1.5e-05
1.0e-05
5.0e-06
0.0e+00
-5.0e-06
0
300
600
900
1200
Time (min)
Strain
Inducible pM.SssI + 0.5% arabinose + 0.15% glucose
Inducible pM.SssI + 0.5% arabinose + 0.25% glucose
Inducible pM.SssI + 0.5% arabinose + 0.5% glucose
Top10 + 0.5% arabinose + 0.15% glucose
Top10 + 0.5% arabinose + 0.25% glucose
Top10 + 0.5% arabinose + 0.5% glucose

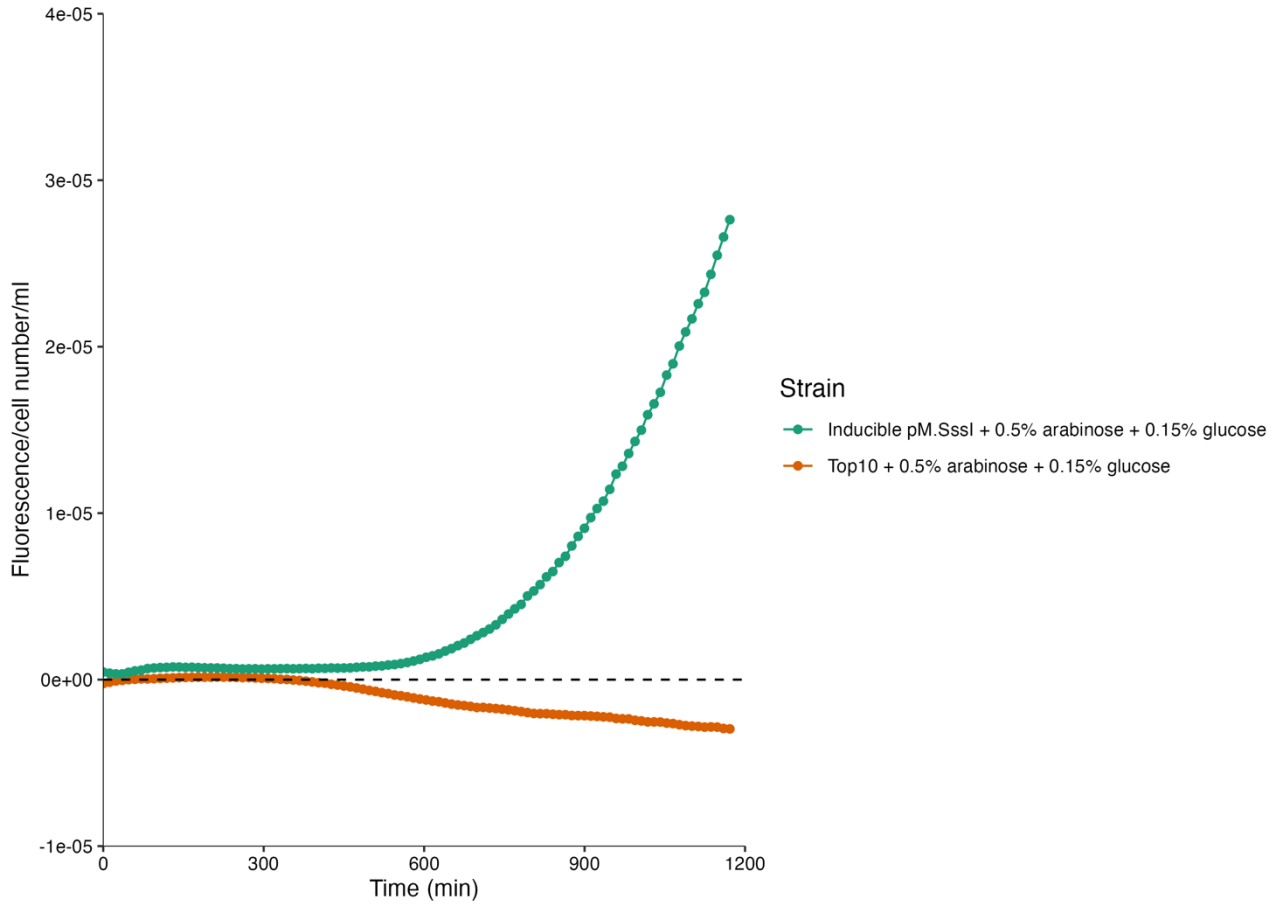
Figure 5.3: A combination of arabinose and glucose decreased background ROS while maintaining induction of DNA methylation. Two independent replicate clones of WT Top10 **(A)** or WT Top10 + inducible pM.SssI **(B)** were treated with 0.15%, 0.25% or 0.5% glucose ± 0.5% arabinose, including 0.5% arabinose-only control. Samples were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated to fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. Mean fluorescence and absorbance were calculated over each pair of replicates, each a mean average of two technical replicates, then mean fluorescence was normalised by mean absorbance. **(C)** Plasmid DNA was extracted from all WT Top10 ± inducible pM.SssI samples at OD600 = 1.0. Plasmid DNA was digested with HpaII and run on a 1% agarose gel. **(D)** WT Top10 ± inducible pM.SssI treated with combinations of arabinose and glucose were normalised by subtracting fluorescence/cell/ml for the glucose-only control from fluorescence/cell/ml for the corresponding arabinose + glucose condition at each time point.

I then repeated the arabinose titration with just 0.15% glucose ± 0.5% arabinose in WT Top10 ± inducible pM.SssI. Again, samples treated with 0.5% arabinose only resulted in high ROS levels over the course of the assay even in the absence of inducible pM.SssI (Figure 5.4A). The combination of 0.15% glucose and 0.5% arabinose dropped ROS levels in WT Top10 – inducible pM.SssI to those of 0.15% glucose only. In WT Top10 + inducible pM.SssI, 0.5% arabinose + 0.15% glucose led to much higher ROS levels than 0.15% glucose only due to induction of M.SssI activity. Removal of glucose-only background ROS displayed no change in ROS over time in WT Top10 – inducible pM.SssI, but a clear increase in ROS for WT Top10 + inducible pM.SssI with this combination of arabinose and glucose (Figure 5.4B). Furthermore, HpaII digests suggested that DNA methylation was induced by 0.5% arabinose even with 0.15% glucose added (Figure 5.4C).

A


8e-05
6e-05
4e-05
2e-05
0e+00
Fluorescence/cell number/ml
0
300
600
900
1200
Time (min)
Strain
Inducible pM.SssI + 0.15% glucose
Inducible pM.SssI + 0.5% arabinose
Inducible pM.SssI + 0.5% arabinose + 0.15% glucose
Top10 + 0.15% glucose
Top10 + 0.5% arabinose
Top10 + 0.5% arabinose + 0.15% glucose

B


4e-05
3e-05
2e-05
1e-05
0e+00
-1e-05
Fluorescence/cell number/ml
0
300
600
900
1200
Time (min)
Strain
Inducible pM.SssI + 0.5% arabinose + 0.15% glucose
Top10 + 0.5% arabinose + 0.15% glucose

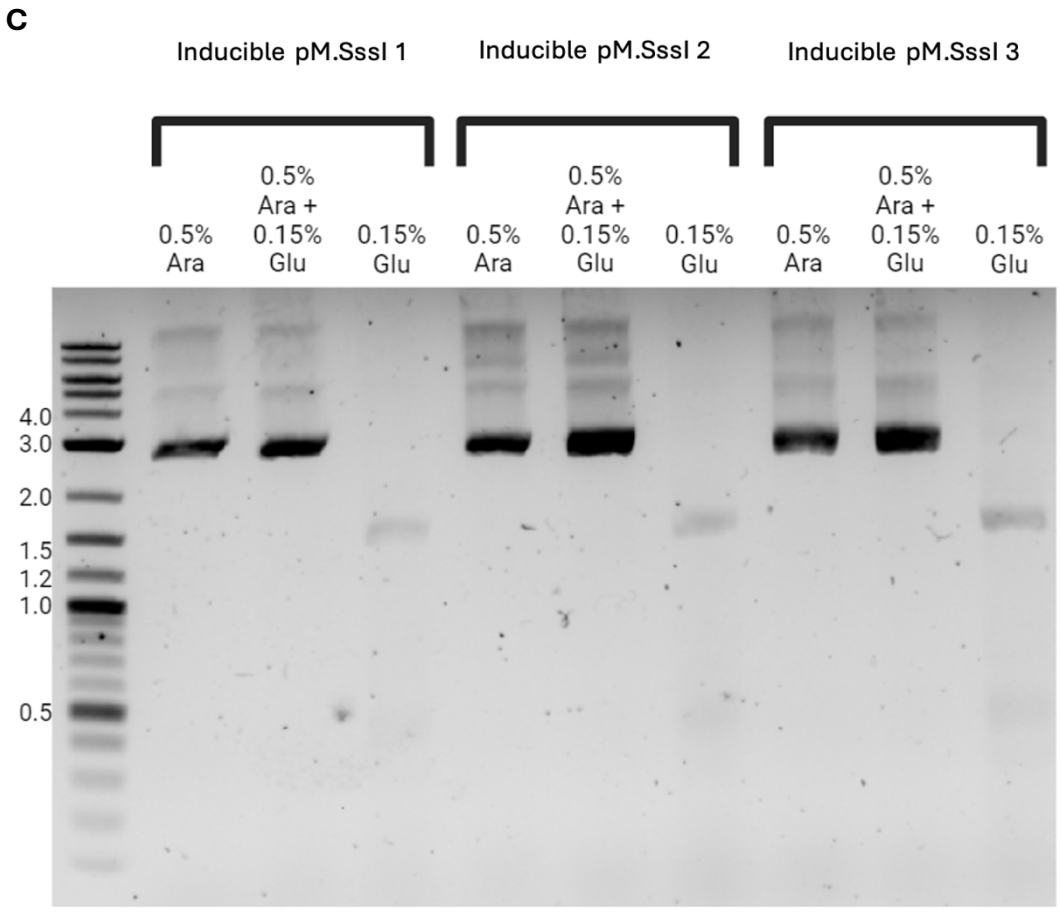


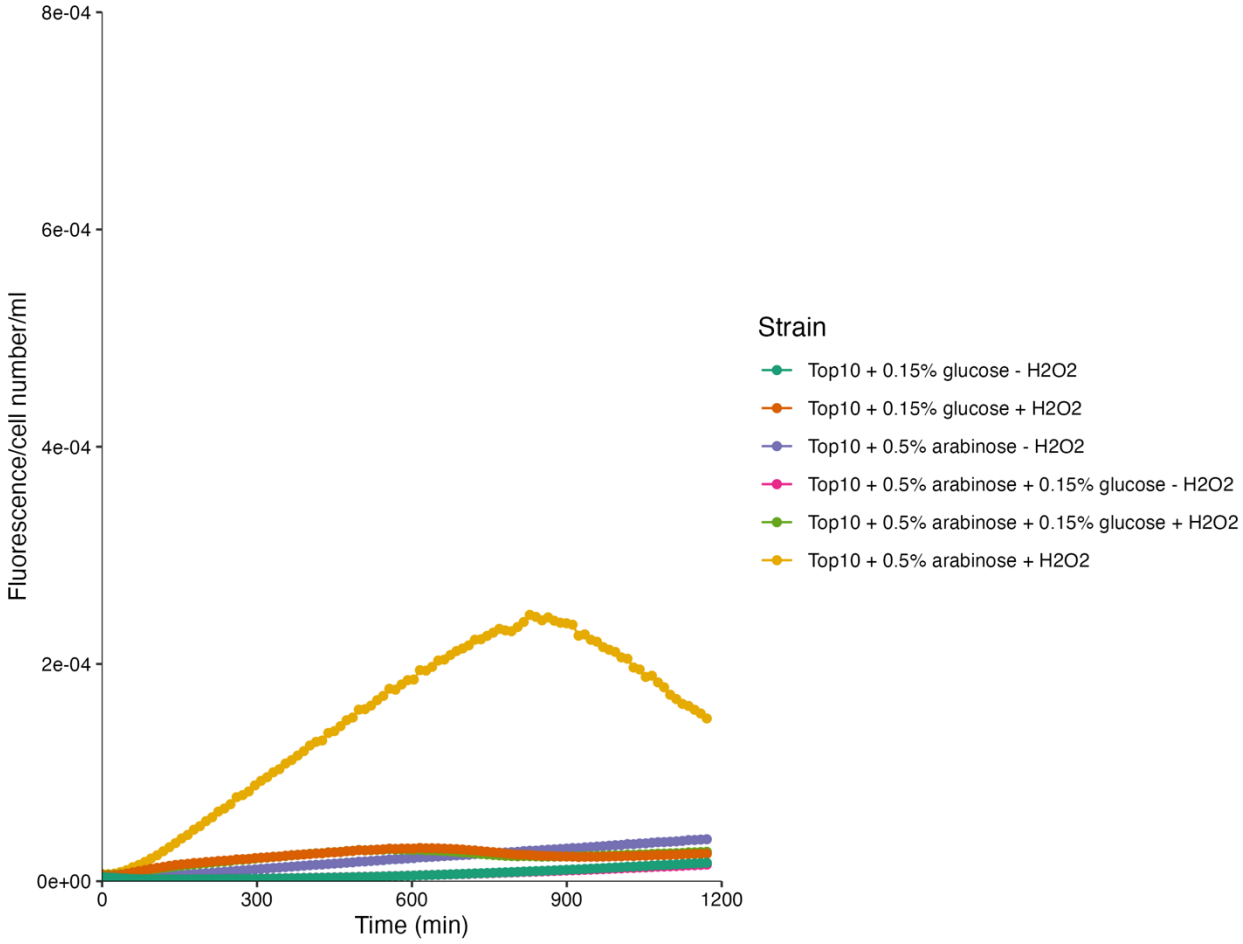
Figure 5.4: 0.5% arabinose and 0.15% glucose was sufficient to decrease background ROS while maintaining induction of DNA methylation. (A) Three independent replicate clones of WT Top10 ± inducible pM.SssI were treated with 0.15% glucose ± 0.5% arabinose, or 0.5% arabinose-only (control). Samples were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated to fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. Mean fluorescence and absorbance were calculated for each triplicate, each a mean average of two technical replicates, then mean fluorescence was normalised by mean absorbance. **(B)** ROS attributed to M.SssI activity in WT Top10 ± inducible pM.SssI was found by subtracting ROS (fluorescence/cell number/ml) under 0.15% glucose only from ROS under 0.5% arabinose + 0.15% glucose at each time point. **(C)** Plasmid DNA was extracted from all WT Top10 + inducible pM.SssI samples at the end point (OD600 = 1.0) of the assay. Plasmid DNA was digested with HpaII and run on a 1% agarose gel.

5.2.1.3. M.SssI induction led to increased sensitivity to oxidative stress by H_2O_2

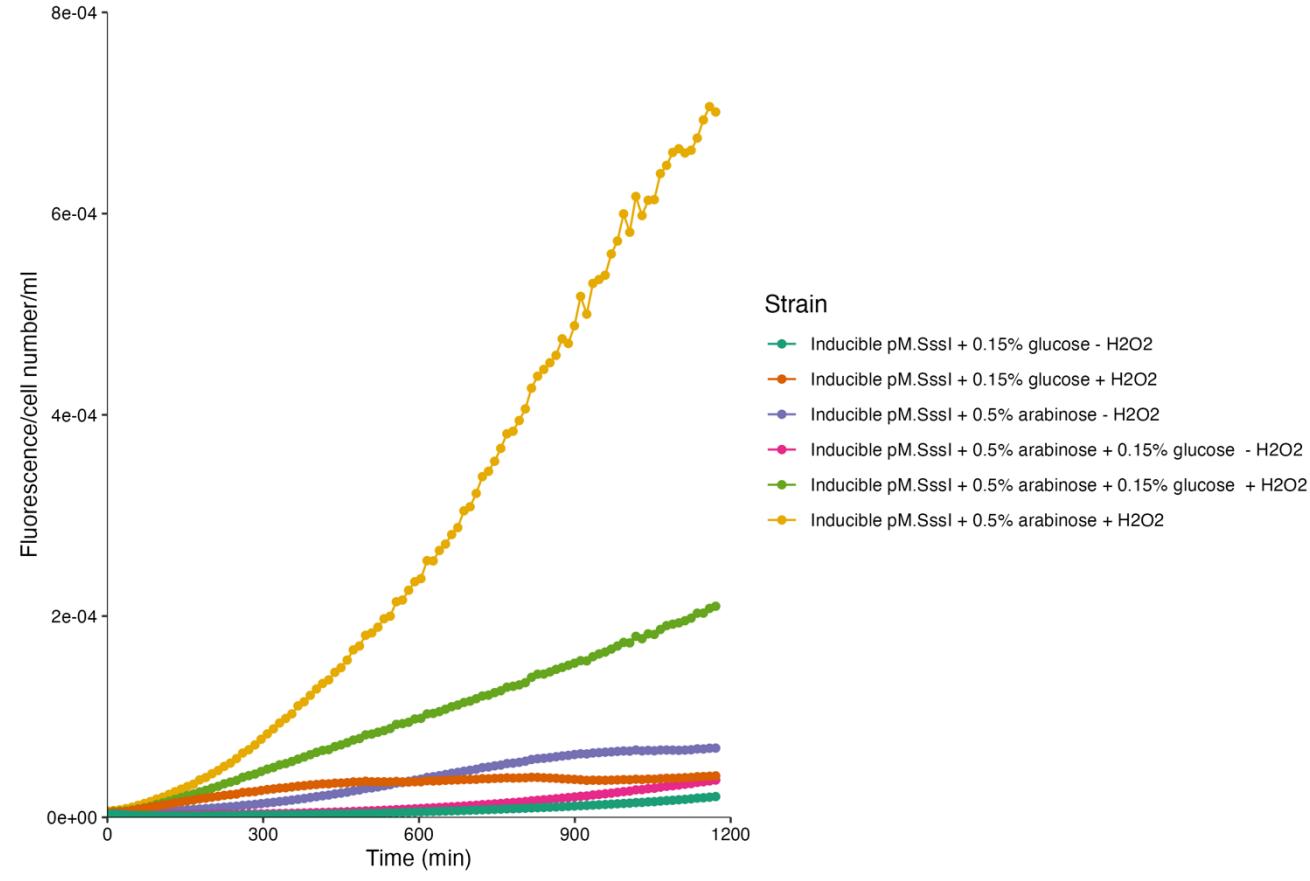
Having optimised the experimental conditions for measuring ROS signal caused by to M.SssI activity, I returned to the original aim: to determine whether induction of M.SssI expression increased sensitivity to oxidative stress. I hypothesised that this would be the case, given that constitutive expression of M.SssI previously increased sensitivity to H_2O_2.

WT Top10 ± inducible pM.SssI were treated with 0.5% arabinose ± 0.15% glucose, including a 0.15% glucose only control. Then cultures were split and treated with 0mM or 10mM H_2O_2. ROS levels were assayed as above. Addition of H_2O_2 only markedly increased ROS in WT Top10 – inducible pM.SssI cells treated with 0.5% arabinose (Figure 5.5A). ROS levels for all other H_2O_2 treated conditions, including 0.5% arabinose + 0.15% glucose, remained low. However, for WT Top10 + inducible pM.SssI, addition of H_2O_2 increased ROS in cells treated with 0.5% arabinose + 0.15% glucose, albeit to a lesser extent than those treated with 0.5% arabinose only (Figure 5.5B). Furthermore, treatment with H_2O_2 did not increase ROS in WT Top10 + inducible pM.SssI cells grown in 0.15% glucose. Together, this suggested that induction of M.SssI resulted in an increased sensitivity to oxidative stress. Removal of the 0.15% glucose-only background confirmed that H_2O_2 has no effect on ROS levels in WT Top10 – inducible pM.SssI (Figure 5.5C). This was expected, as a lack of M.SssI activity meant that there was no driver of oxidative stress and therefore ROS levels. Induction of M.SssI activity clearly increased ROS; an effect that was exacerbated by treatment with H_2O_2.

A


8e-04
6e-04
4e-04
2e-04
0e+00
Fluorescence/cell number/ml
0
300
600
900
1200
Time (min)
Strain
Top10 + 0.15% glucose - H2O2
Top10 + 0.15% glucose + H2O2
Top10 + 0.5% arabinose - H2O2
Top10 + 0.5% arabinose + 0.15% glucose - H2O2
Top10 + 0.5% arabinose + 0.15% glucose + H2O2
Top10 + 0.5% arabinose + H2O2

B


8e-04
6e-04
4e-04
2e-04
0e+00
Fluorescence/cell number/ml
0
300
600
900
1200
Time (min)
Strain
Inducible pM.SssI + 0.15% glucose - H2O2
Inducible pM.SssI + 0.15% glucose + H2O2
Inducible pM.SssI + 0.5% arabinose - H2O2
Inducible pM.SssI + 0.5% arabinose + 0.15% glucose - H2O2
Inducible pM.SssI + 0.5% arabinose + 0.15% glucose + H2O2
Inducible pM.SssI + 0.5% arabinose + H2O2

C

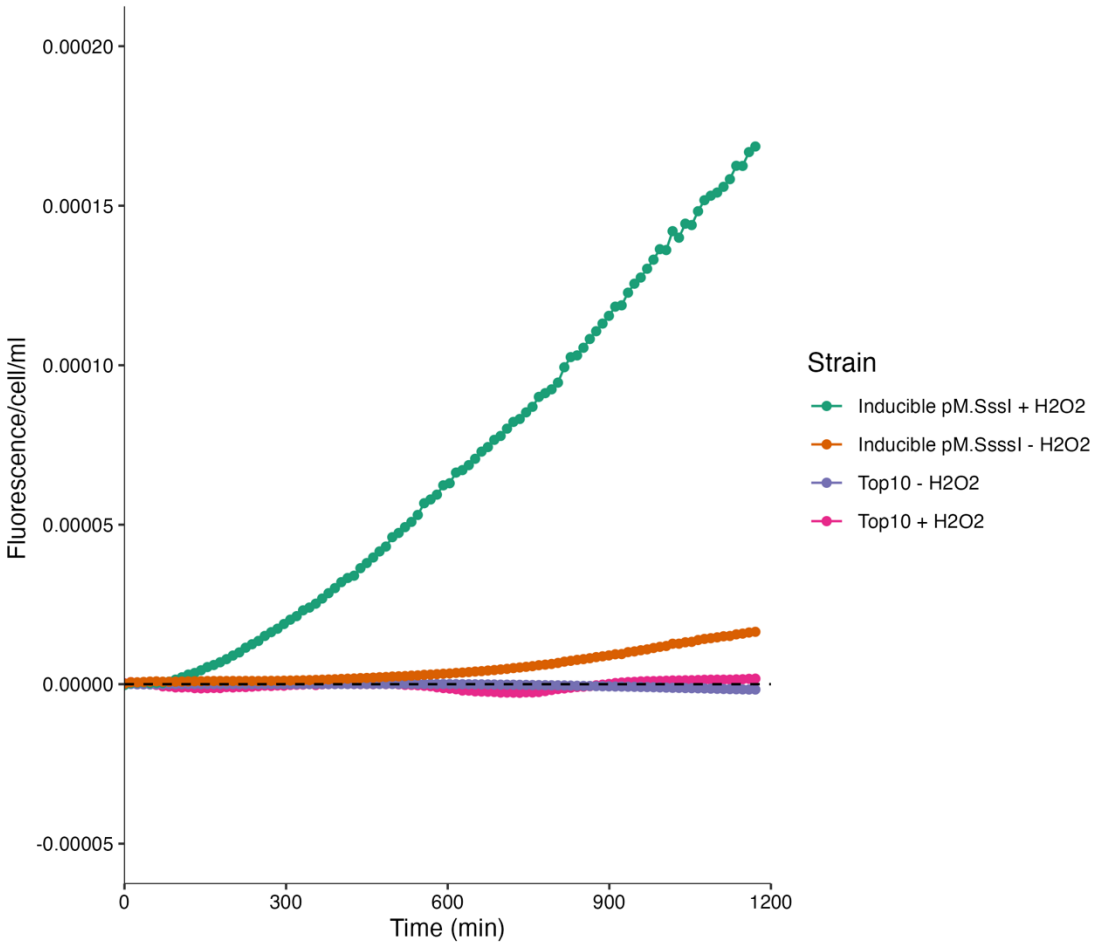


Figure 5.5: Induction of M.SssI activity increased sensitivity to oxidative stress. Two independent replicate clones of WT Top10 - inducible pM.SssI **(A)** or WT Top10 + inducible pM.SssI **(B)** were treated with 0.15% glucose ± 0.5% arabinose, or 0.5% arabinose-only (control) ± H_2O_2. Samples were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated to fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. Mean fluorescence and absorbance were calculated for each pair of replicates, each a mean average of two technical replicates, then mean fluorescence was normalised by mean absorbance. **(C)** WT Top10 ± inducible pM.SssI treated with 0.5% arabinose + 0.15% glucose ± H_2O_2 were normalised by subtracting fluorescence/cell/ml for the glucose-only control from fluorescence/cell/ml for the corresponding arabinose + glucose condition at each time point.

5.2.2. Investigating how M.SssI activity led to an increase in ROS *in vitro*

The *in vivo* experiments above showed that induction of M.SssI activity alone was sufficient to increase ROS levels in the cell independent of a source of oxidative damage such as H_2O_2. This replicated previous results with the constitutive system (Krwawicz et al., 2024), suggesting that this was a consequence of CpG DNMT activity and not a technical artefact. However, it wasn't clear whether ROS were produced as a direct byproduct of M.SssI catalysis, or whether reactions between DNA methylation and cellular components produced ROS and therefore raised ROS levels indirectly (Figure 5.6).

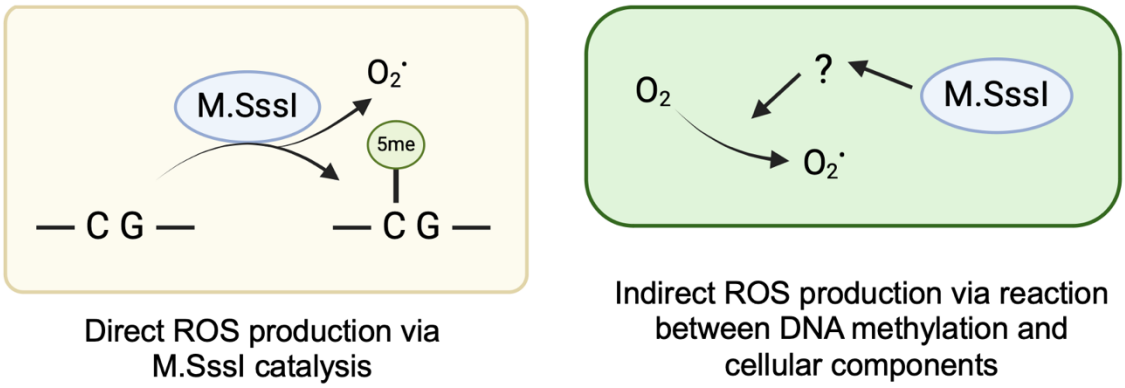
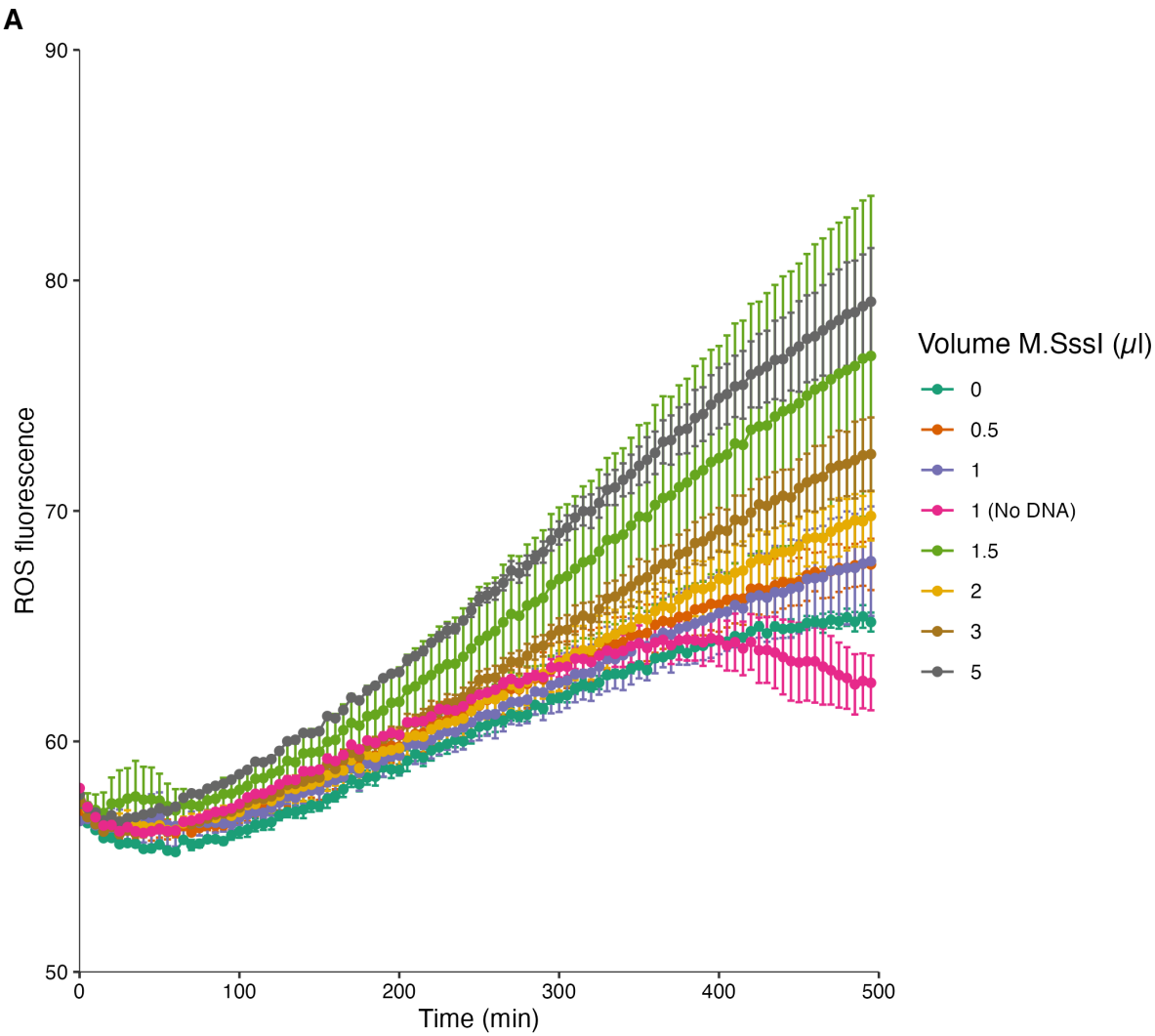


Figure 5.6: Schematic of theoretical direct (LHS) or indirect (RHS) ROS production due to introduction of M.SssI *in vivo*. Created in Biorender.

To investigate this, I decided to test the influence of M.SssI on ROS levels *in vitro*. I hypothesised that any changes to ROS in a cell free system would be due to the activity of M.SssI itself. To clearly identify changes due to M.SssI activity, I conducted an M.SssI titration. Here, a range of volumes of M.SssI were added to plasmid DNA and ROS levels were measured using the same fluorescent dye as above. If ROS

production was a direct consequence of M.SssI catalysis, I expected that increasing the volume of M.SssI would lead to a proportional increase in ROS.

Results of the titration are shown in Figure 5.7A. ROS fluorescence clearly increased with volume of M.SssI. HpaII digests also suggested that DNA methylation increased with volume of M.SssI (Figure 5.7B). This supported the hypothesis that M.SssI activity directly produces ROS as a result of catalysis. However, a catalytically dead M.SssI control would be required to isolate enzyme-specific ROS production.



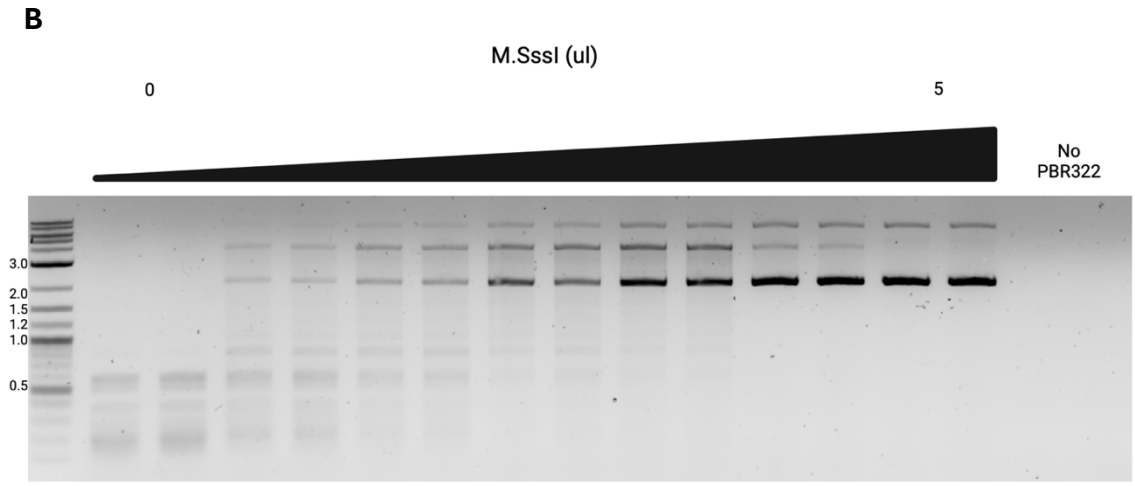
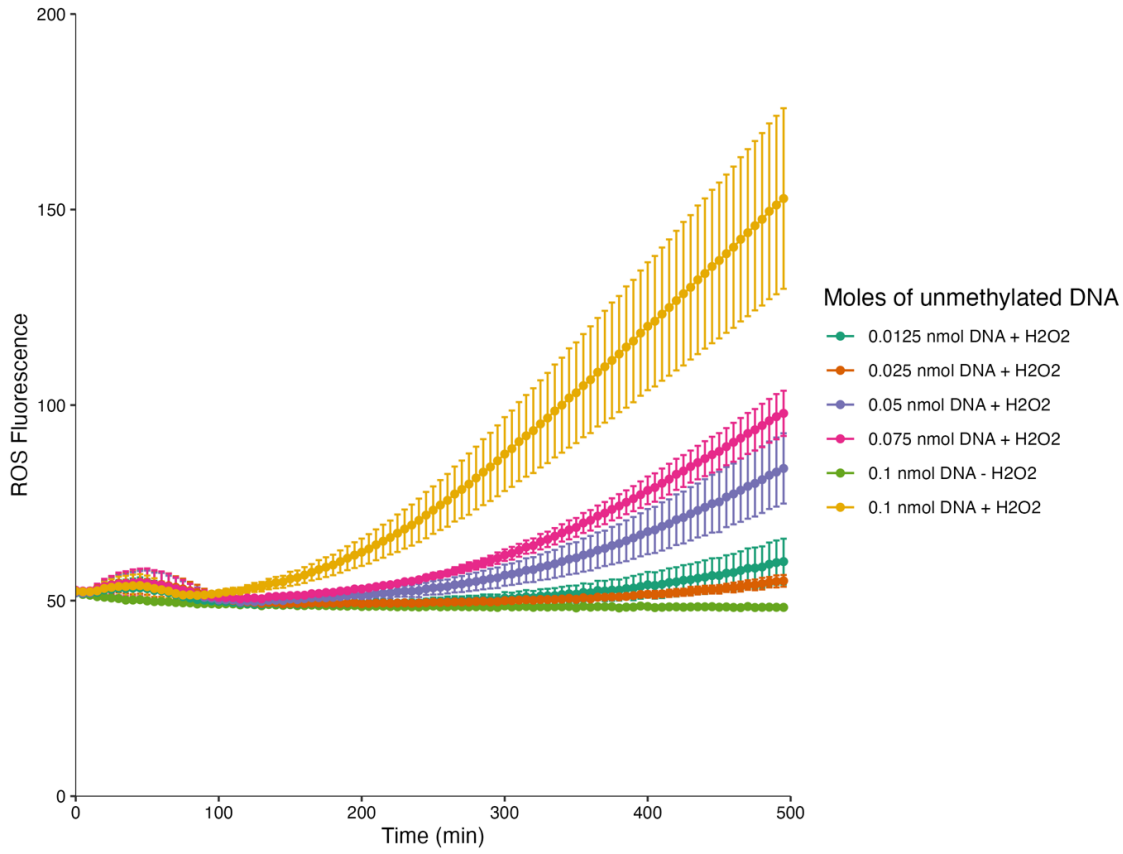


Figure 5.7: ROS fluorescence increased with volume of M.SssI. (A) 0 – 5ul of M.SssI were added to pBR322 plasmid DNA in supplied buffer, supplemented with 1600μM SAM. Two technical replicates were performed for each volume of M.SssI. Reactions were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated by fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. **(B)** Plasmid DNA was digested with HpaII and reaction products were run on a 1% agarose gel.

While the result of the titration was very clear, it was still possible that it was not directly due to M.SssI catalysis but was instead caused by other components of the *in vitro* reaction. One possibility was that it was not M.SssI activity itself, but presence of DNA methylation that was increasing ROS levels. To test this, I measured ROS levels for *in vitro* reactions containing either unmethylated or methylated dsDNA + H_2O_2 but lacking M.SssI or SAM. DsDNA was comprised of a 24 base pair forward oligo, either undamaged or containing 5mC, both annealed to an undamaged reverse (see Methods). If presence of DNA methylation was sufficient to increase ROS signal, then addition of H_2O_2 should lead to a substantial increase in ROS due to excess oxidative stress. Again, this was performed as a titration over a molar range of DNA to clearly identify any effects on ROS levels.

Increasing the amount of unmethylated DNA also increased ROS levels with consistent concentration of H_2O_2 (Figure 5.8A). This effect was also observed for methylated DNA (Figure 5.8B). Crucially, the increase in ROS with moles of DNA was no greater with methylated DNA than unmethylated. This suggested that it was the amount of DNA itself that was responsible for the increase in ROS, not the presence of methylation.

A



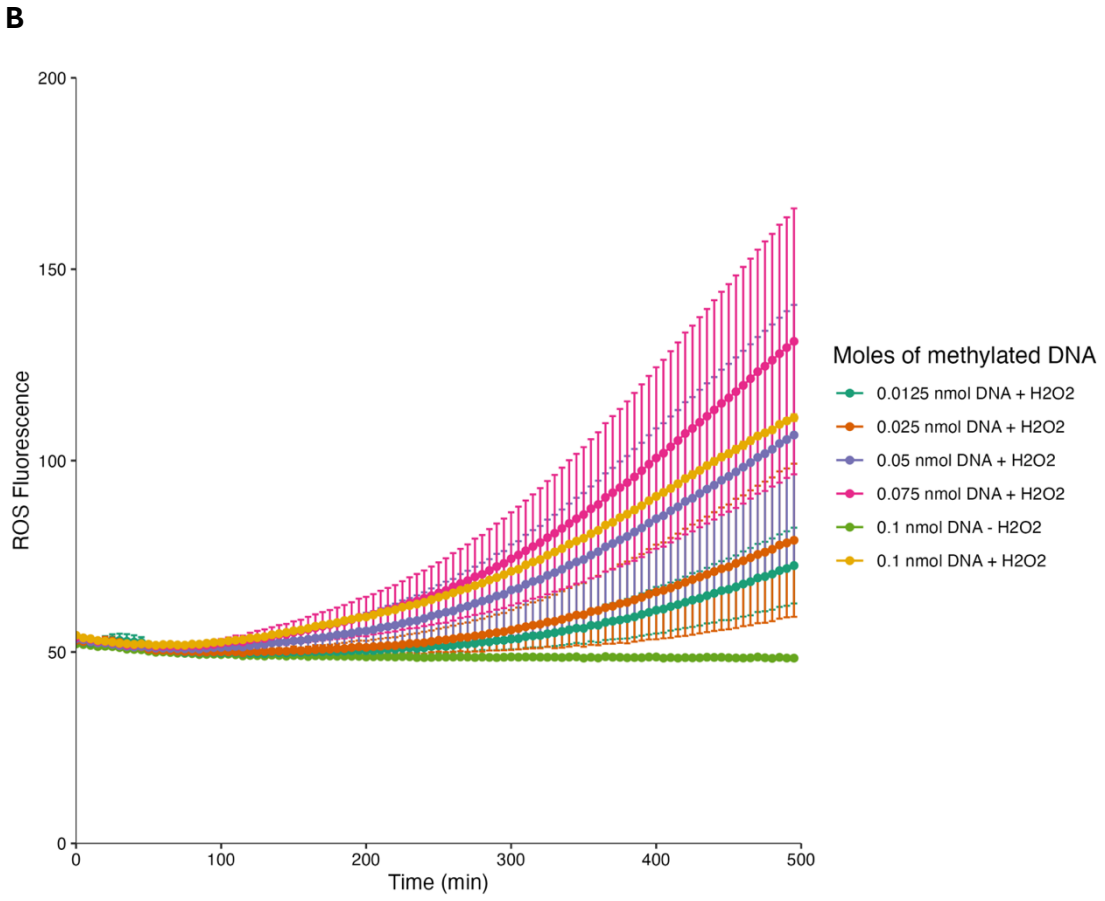


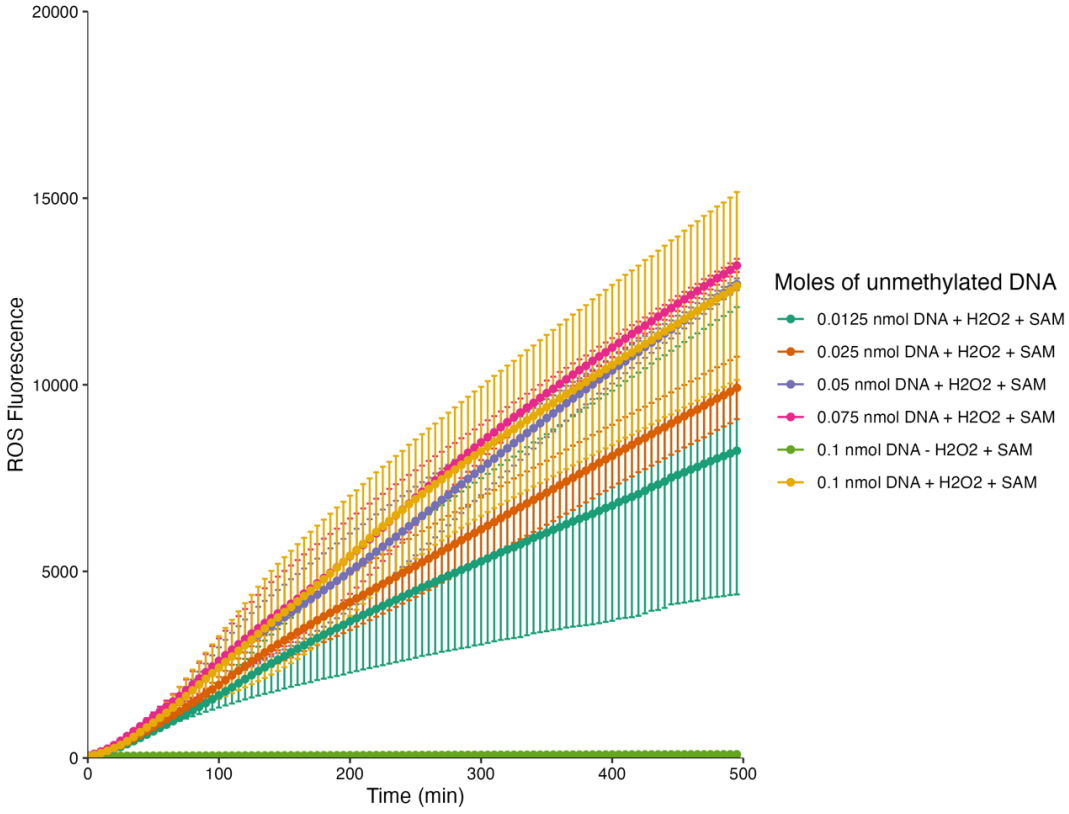
Figure 5.8: Presence of DNA methylation had no effect on sensitivity to oxidative stress. A titration of dsDNA constructs containing a CpG site that was either unmethylated **(A)** or methylated on one strand **(B)** were treated with 10μM H_2O_2 in the absence of M.SssI and SAM. dsDNA was added at 0.0125, 0.025, 0.05, 0.075 or 0.1nmol + 10μM (0.25 pmol) H_2O_2. 0.1nmol DNA - H_2O_2 was included as a negative control. Two technical replicates were conducted for each volume of DNA. Reactions were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated by fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm.

If the presence of DNA methylation was not responsible for the increase in ROS observed in the original M.SssI titration (Figure 5.7A), then what could be? Increasing the volume of M.SssI would also have increased the volume of SAM, as this was a component of the buffer M.SssI was stored in. SAM is the methyl group donor that

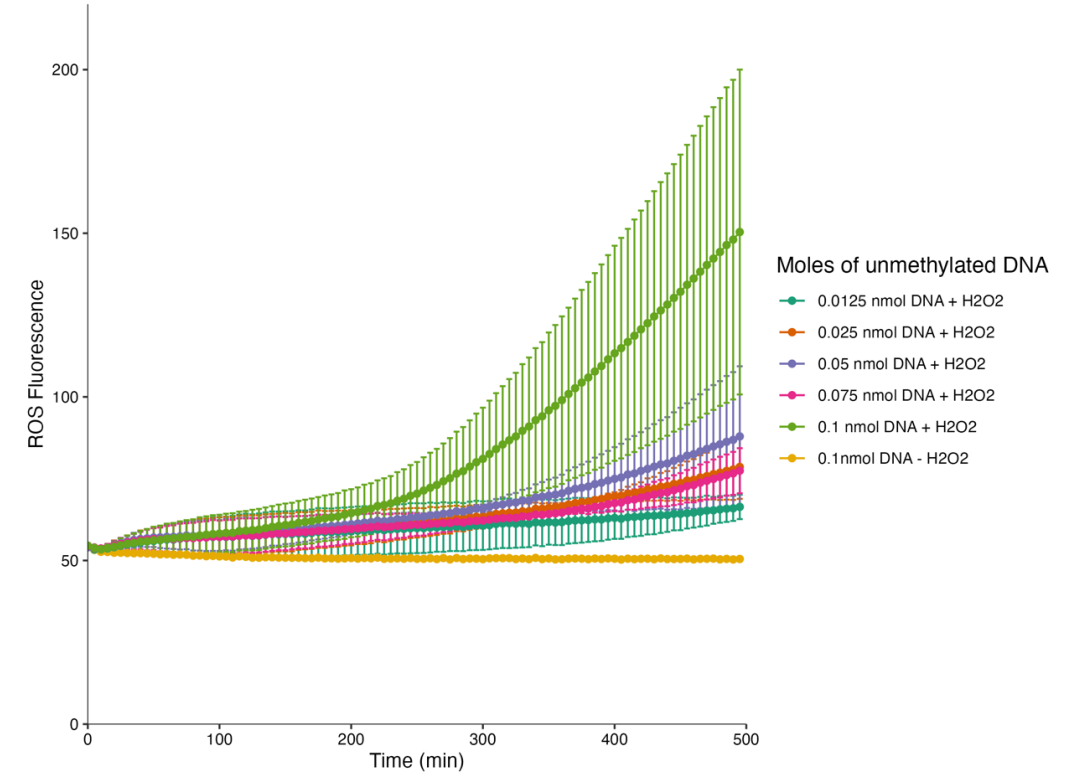
M.SssI uses to methylate DNA (Klimasauskas et al., 1994). Therefore, it was possible that the breakdown of SAM released ROS and contributed to the increase in ROS observed with increasing volume of M.SssI.

To test this, I repeated the previous experiment, but this time included ± SAM for each reaction. I hypothesised that if the reaction between H_2O_2 and SAM yielded ROS, then ROS would increase independent of DNA methylation. There was no increase in ROS over time for any condition in the absence of H_2O_2, suggesting that any increase in ROS was dependent on H_2O_2 (Figure 5.9). In the absence of SAM, ROS levels increased marginally with volume of DNA as previously observed, although again there was no difference between methylated and unmethylated constructs (Figure 5.9B, D). However, the combination of SAM and H_2O_2 led to an approximately 100-fold increase in ROS for both constructs, again independent of presence of methylation (Figure 5.9A, C). Together, this suggested that combination of SAM with H_2O_2 was primarily responsible for the increase in ROS observed in this assay and DNA methylation itself did not contribute to ROS levels.

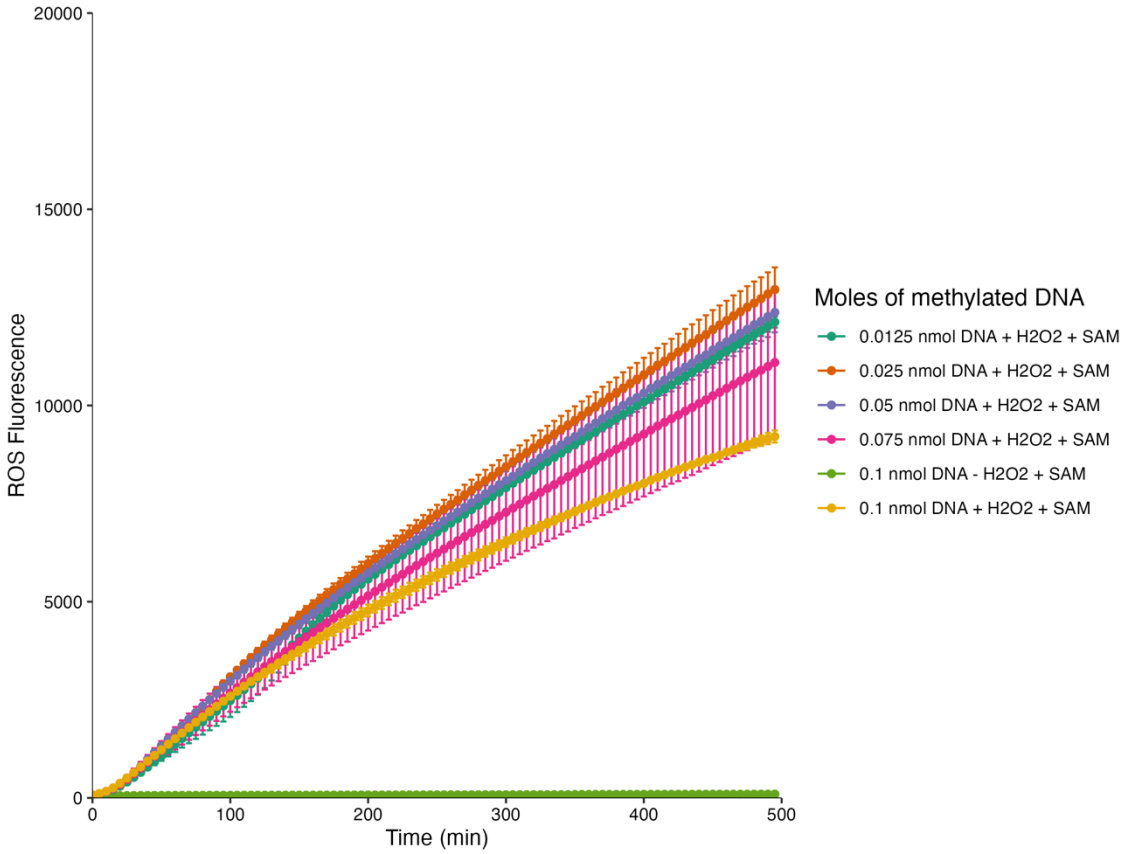
A


20000
15000
10000
5000
0
ROS Fluorescence
0
100
200
300
400
500
Time (min)
Moles of unmethylated DNA
0.0125 nmol DNA + H2O2 + SAM
0.025 nmol DNA + H2O2 + SAM
0.05 nmol DNA + H2O2 + SAM
0.075 nmol DNA + H2O2 + SAM
0.1 nmol DNA - H2O2 + SAM
0.1 nmol DNA + H2O2 + SAM

B


200
150
100
50
0
ROS Fluorescence
0
100
200
300
400
500
Time (min)
Moles of unmethylated DNA
0.0125 nmol DNA + H2O2
0.025 nmol DNA + H2O2
0.05 nmol DNA + H2O2
0.075 nmol DNA + H2O2
0.1 nmol DNA + H2O2
0.1nmol DNA - H2O2

C


20000
15000
10000
5000
0
ROS Fluorescence
0
100
200
300
400
500
Time (min)
Moles of methylated DNA
0.0125 nmol DNA + H2O2 + SAM
0.025 nmol DNA + H2O2 + SAM
0.05 nmol DNA + H2O2 + SAM
0.075 nmol DNA + H2O2 + SAM
0.1 nmol DNA - H2O2 + SAM
0.1 nmol DNA + H2O2 + SAM

D

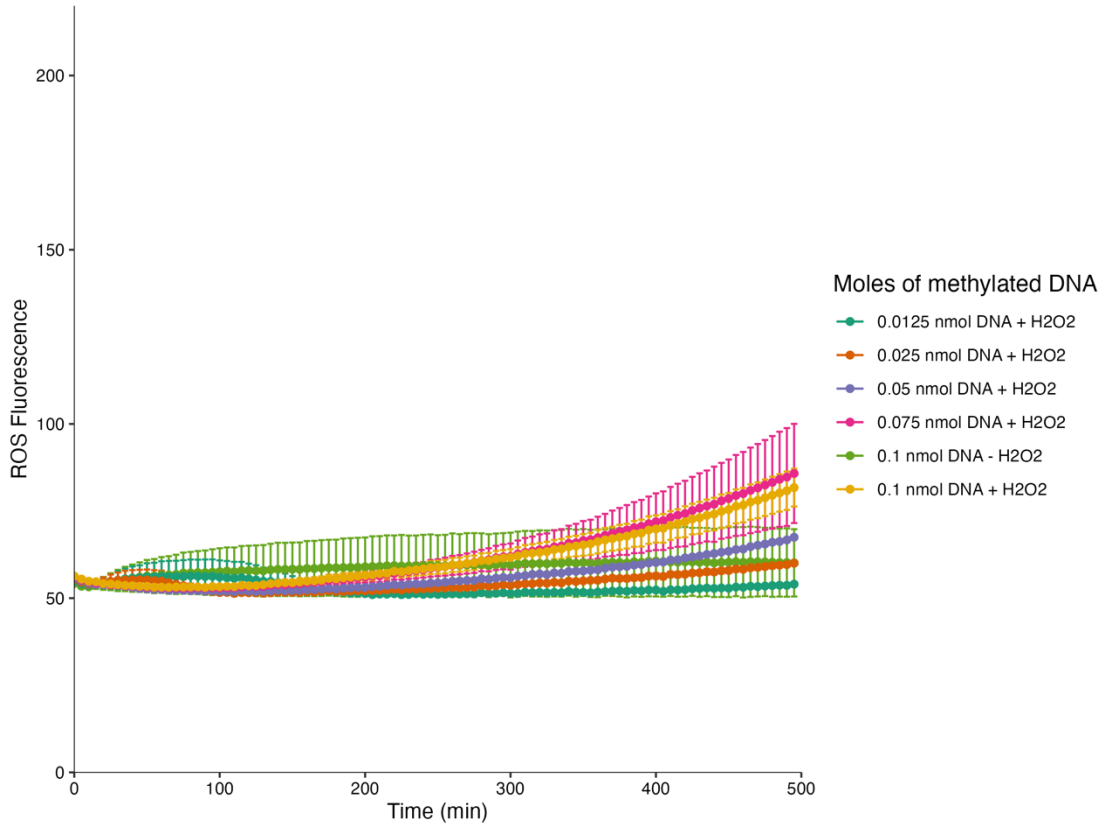

200
150
100
50
0
ROS Fluorescence
0
100
200
300
400
500
Time (min)
Moles of methylated DNA
0.0125 nmol DNA + H2O2
0.025 nmol DNA + H2O2
0.05 nmol DNA + H2O2
0.075 nmol DNA + H2O2
0.1 nmol DNA - H2O2
0.1 nmol DNA + H2O2

Figure 5.9: The combination of SAM and H_2O_2 substantially increased ROS levels independent of DNA methylation. 0.0125, 0.025, 0.05, 0.075 or 0.1 nmol of unmethylated DNA was treated with 10μM H_2O_2 + 1600 μM SAM **(A)** or 0μM SAM **(B)**. 0.1 nmol of unmethylated DNA was also treated with 0μM H_2O_2 ± 1600μM SAM as a control. The same titration was performed for a dsDNA construct containing 5mC on one strand only at a CpG dinucleotide, with + SAM **(C)** and – SAM **(D)**. Reactions were mixed with a fluorescent dye that binds to ROS, enabling ROS levels to be approximated by fluorescence over time. Every ten minutes, fluorescence was measured by excitation at 475nm and emission at 500-550nm. Each reaction was performed in duplicate, with standard deviation error bars calculated for each time point.

5.3. Specific discussion

The work presented in this chapter demonstrated that DNMT activity *in vivo* increased the level of total ROS independently of exogenous oxidative stress. Addition of H_2O_2 led to a substantial increase in ROS, suggesting that DNMT activity itself either directly or indirectly contributed to ROS levels. To probe this further, I tested how DNMT activity and associated experimental conditions affected total ROS levels *in vitro*. While initial results suggested that ROS levels were directly proportional to M.SssI activity, further investigation showed that a combination of SAM and H_2O_2 increased ROS levels independent of DNA methylation. However, this did not account for the H_2O_2-independent increase in ROS observed *in vivo*. This was most likely caused by secondary effects of DNA methylation and SAM breakdown in the cell, which will be discussed further in 5.3.2.

5.3.1. Preferential sugar metabolism in *E. coli*

One of the most striking results of this chapter was the increase in ROS with induction of M.SssI activity independent of exogenous oxidative stress (Figure 5.1D). This was confounded, however, by a smaller but substantial increase in ROS in cells lacking inducible pM.SssI and treated with arabinose. Arabinose and glucose titrations revealed that this increase in 'background' ROS was not caused by the presence of arabinose as first thought, but instead a lack of glucose in the system. *E. coli* possesses a carbon catabolite repression system, whereby presence of the preferred substrate, glucose, prevents expression of genes required for metabolism of other carbon sources (Görke & Stülke, 2008). Presence of glucose maintains low levels of cAMP, which prevents activation of genes involved in the metabolism of non-glucose sugars. In the presence of arabinose, catabolite repression is lifted and cAMP levels rise. cAMP binds cAMP receptor protein (CAP) and the cAMP-CAP complex activates the promoters of catabolic genes (Busby & Ebright, 1999; Malan et al., 1984; Tagami & Aiba, 1998) and regulates expression of small non-coding regulatory RNAs (Polayes et al., 1988). Additionally, arabinose binds AraC, the main regulatory protein of the *araBAD* operon, resulting in the activation of arabinose-specific metabolism genes (Schleif, 2000). Interestingly, the literature reports a hierarchy of sugars where glucose is metabolised over arabinose even if both are present (Desai & Rao, 2010). This suggests that M.SssI should have been repressed when treated with combinations of arabinose and glucose. However, 0.15% glucose may not have been sufficient to fully repress M.SssI expression, particularly in combination with 0.5% arabinose. Despite this, it may have contributed to regular

E. coli metabolism, therefore reducing background ROS. It was also possible that low concentrations of glucose reduced ROS by curbing M.SssI expression in Top10 with inducible pM.SssI. However, the combination of 0.5% arabinose and 0.15% glucose also reduced ROS in Top10 without inducible pM.SssI compared to 0.5% arabinose alone. This suggested that the decrease in ROS was not caused by a reduction in M.SssI activity and was instead due to metabolic changes.

Another interesting result was that low concentrations of arabinose increased ROS levels in Top10 ± inducible pM.SssI. This could be due to how arabinose is metabolised in *E. coli*. The enzymes in the *araBAD* operon convert L-arabinose in D-xyulose-5-phosphate, which is fed into the pentose phosphate pathway (PPP) (Gross & Englesberg, 1959; Schleif, 2010). The PPP produces NADPH, among other things, which combats oxidative stress by converting H_2O_2 into H_2O (TeSlaa et al., 2023). Perhaps the low concentrations of arabinose resulted in decreased NADPH levels and therefore reduced alleviation of oxidative stress. This does not explain why arabinose alone resulted in increased ROS compared to glucose. Given that glucose is preferentially metabolised over arabinose (Desai & Rao, 2010), it could be that glucose is fed more efficiently into the PPP. Therefore, a loss of glucose would lead to a reduction in NADPH, which would be exacerbated by low concentrations of arabinose.

5.3.2. How does induction of M.SssI activity increase cellular ROS?

A key question arising from this work was how does induction of M.SssI activity increase ROS? We know from previous work that this increase was independent of alkylation damage induction because sensitivity to oxidative stress did not increase with loss of AlkB (Krwawicz et al., 2024). It must, therefore, have proceeded via a different mechanism. Broadly, there are two possible options: firstly, the catalytic activity of M.SssI directly produced ROS. Secondly, products of M.SssI activity were involved in reactions in the cell that indirectly contributed to an increase in ROS. To this end, one possibility was that M.SssI-induced 5mC reacted with existing ROS to form radical cytosine derivatives. Other reactions of these radical derivatives with oxygen or water could also increase ROS (Madugundu et al., 2014). Presence of H_2O_2 would further exacerbate this, as breakdown of H_2O_2 releases ROS that can oxidise 5mC to 5hmC and 5fC via radical derivatives (Cadet & Wagner, 2013). 5hmC and 5fC are recognised as DNA damage in mammals and are excised by TDG and NEIL1/2 (Bochtler et al., 2017), which are related to FpG in *E. coli* (Hwang et al., 2025; Jacobs & Schär, 2012). Indeed, Krwawicz et al. showed that introducing mammalian TET enzymes into *E. coli* oxidised 5mC to 5hmC and 5fC. Further, this increased sensitivity to oxidative stress in combination with M.SssI activity, suggesting that oxidation products from 5mC may have contributed to the increase in ROS (Krwawicz et al., 2024). In this way, presence of DNA methylation would increase both DNA damage and oxidative stress. Furthermore, the propensity of 5mC to increase cellular ROS may contribute to mutagenesis at CpG dinucleotides (Madugundu et al., 2014). Overall, evidence of oxidative damage resulting from

M.SssI activity supports the hypothesis that DNMTs have been lost from some lineages due to the threat they pose to genome stability.

5.3.3. Contributions of additional reaction components to ROS levels

The work conducted *in vitro* aimed to address whether the catalytic activity of M.SssI could directly produce ROS. This was based on the premise that any ROS arising in a cell-free system would be the result of M.SssI activity. However, measuring total ROS in the reaction was not the same as directly measuring ROS arising from M.SssI catalysis. This would require structural and enzymatic analysis that is beyond the scope of this thesis. However, it is worth noting that current understanding of the catalysis of 5mC induction by DNMTs does not involve ROS generation (Jeltsch, 2006). It is also worth noting that the fluorescent DCFH-DA sensor used in this assay can be oxidised by a wide range of ROS and therefore does not provide a quantitative measure of ROS concentration. Instead, it provides a relative redox-sensitive readout which is suitable for identifying changes in total ROS arising from M.SssI activity.

Initial M.SssI titration results showed an increase in ROS with increasing units of M.SssI. This suggested that perhaps M.SssI activity was directly increasing ROS. Of course, another possibility was that the presence of DNA methylation, not necessarily catalysis, led to the increase in ROS. If this were true, then increasing the amount of pre-existing 5mC in an M.SssI-free reaction would correlate with an increase in ROS. This increase that would be exacerbated in the presence of H_2O_2

due to an overload of oxidative stress. However, while there was a small, proportional increase in ROS with increasing moles of DNA, there was no difference in total ROS between unmethylated and methylated DNA constructs in the presence of H_2O_2. This suggested that DNA methylation itself did not contribute to ROS levels.

What, then, could be responsible for this increase *in vitro*? Crucially, the unmethylated and methylated DNA titrations shown in Figure 5.7 lacked M.SssI, but also SAM. For M.SssI to catalyse the deposition of a methyl group onto cytosine, a methyl donor such as SAM is required (Klimasauskas et al., 1994). Where an increase in ROS was observed with M.SssI previously (Figure 5.7A), each reaction contained SAM, both in the buffer and supplemented additionally. Therefore, it was possible that the substantial increase in ROS in Figure 5.8 was caused by a reaction between H_2O_2 and SAM that released ROS. This operated independently of DNA methylation.

However, the SAM-dependent increase in ROS was only observed when H_2O_2 was added and did not explain how ROS levels increased in the initial M.SssI titration which had no exogenous oxidative stress present (Figure 5.7A). It is worth noting that total ROS levels in the initial M.SssI titration were very low, in the range of 50-90 arbitrary units. Due to the recombinant production of M.SssI from *E. coli*, the final product would likely have contained small amounts of DNA. Therefore, increasing the volume of M.SssI would have increased the volume of DNA in the reaction. In Figures 5.8 and 5.9, ROS levels were roughly proportional to volume of

unmethylated and methylated DNA, albeit over a very small range in the absence of SAM. Taken together, this suggested that volume of DNA resulted in a small, proportional increase in ROS and may have been responsible for the initial increase in ROS observed in Figure 5.7A. However, this was not sufficient to explain the increase in ROS observed upon addition of SAM and H_2O_2. So, while initial data suggested a direct role of M.SssI activity in ROS generation, it was far more likely that any increase in ROS was due to secondary reactions occurring between other assay components and not M.SssI catalysis itself.

Chapter 6 – Discussion

6.1. Summary of key findings

In Chapter 3, I demonstrated that introducing specific DNA lesions into *E. coli* reduced the number of viable transformants. For alkylation lesions 1mA and 3mC , this effect was exacerbated by the loss of AlkB. This was consistent with evidence in the literature of replisome stalling by alkylation adducts, which are typically reversed back to undamaged DNA by AlkB (Boiteux & Laval, 1982; Larson et al., 1985). Importantly, this suggested a direct effect of alkylation lesions on genome stability which supported further investigation into the effect of DNMT-induced alkylation lesions on mutagenesis.

In Chapter 4, I investigated the effect of CpG-specific DNMT activity on mutagenesis in WT and Δ*alkB E. coli*. I utilised a rifampicin resistance assay to select for mutants resulting from DNMT activity and/or loss of AlkB. Constitutive DNMT activity in Δ*alkB* C2523 *E. coli* resulted in an increase in CG > TA mutations. This M.SssI- and Δ*alkB*-dependent increase was consistent with AlkB-repaired alkylation damage, likely including 3mC. Additionally, a strong non-cytosine *rpoB*-specific mutation spectrum was observed at TA base pairs. This signature was dependent on both DNMT activity and a lack of AlkB, indicating that it was the result of DNMT-induced alkylation damage normally repaired by AlkB. However, it was unusual in that it did not occur in the CpG context, despite the documented specificity of the DNMT (Dukatz et al., 2019). Furthermore, induction of DNMT activity in Top10 *E. coli* did not

recapitulate the TA > CG signature. Further investigation into this revealed that the strain-dependent signature was not dependent on presence of RecA alone; it remains mechanistically unresolved. Nonetheless, this provided strong evidence that damage induction by DNMT activity caused mutagenesis. This supported the overarching hypothesis that the propensity of DNMTs to damage DNA and cause genome instability may have contributed to their frequent evolutionary loss.

In Chapter 5, I used the inducible DNMT system to build on previous work demonstrating increased sensitivity of DNMT-expressing cells to oxidative stress (Krwawicz et al., 2024). I found that glucose concentration affected background ROS signal. Therefore, I optimised the experimental conditions to reduce this effect. Following this, I demonstrated that induction of DNMT activity resulted in an increase in ROS even in the absence of exogenous oxidative stress. Addition of H_2O_2 caused a substantial increase in ROS, suggesting that DNMT activity itself either directly or indirectly increase ROS levels. I tested this hypothesis by assessing the effect of DNMT activity *in vitro*. While initial results suggested direct ROS production via DNMT catalysis, subsequent investigation revealed ROS levels were highly dependent on presence of SAM and H_2O_2. ROS levels were not affected by DNMT activity or pre-existing DNA methylation. Therefore, DNMT-associated ROS *in vivo* likely arose indirectly due to cellular reactions with radical methylation intermediates. This work demonstrated another unexpected way in which DNMT activity may damage DNA, again perhaps contributing to the evolutionary loss of DNMTs.

6.2. Overall insights and remaining questions

6.2.1. DNMT-induced mutagenesis and the evolutionary loss of DNA methylation

The work in Chapter 4 demonstrated the capacity of DNMTs to induce DNA damage, particularly in the absence of AlkB, resulting in elevated CG > TA and strain-dependent TA > CG mutagenesis. While the exact cause of this conditional signature is yet to be determined, the overall increase in mutagenesis supported the hypothesis that DNMTs have the propensity to introduce DNA damage. Therefore, I propose that the work in this chapter, and this thesis in general, supports a model in which DNMT-induced DNA damage and mutagenesis may have contributed to the evolutionary loss of DNMTs.

DNA methylation is an essential epigenetic mechanism for gene silencing in mammals (Greenberg & Bourc'his, 2019; Jones & Takai, 2001; Li & Zhang, 2014; Schulz et al., 2006). This may be in part due to their large genomes that have a high density of TEs (Ahmed & Liang, 2012), which could otherwise cause mutagenesis by random insertion into the genome (Boeke et al., 1985; Greenblatt & Brink, 1963; McClintock, 1950). Furthermore, DNA methylation evolved early in evolutionary timescales (Al Adhami et al., 2022; Keller et al., 2016) and became critical to development (Li et al., 1992), for example in processes such as genomic imprinting and X chromosome inactivation (Brockdorff et al., 1992; Li et al., 1993). Despite this, many organisms such as *C. elegans, D. melanogaster* and *S. pombe* lack DNA methylation entirely (Jurkowski & Jeltsch, 2011; Ponger & Li, 2005). For an organism

to maintain the DNA methylation machinery through purifying selection, the advantages conferred by DNA methylation must be greater than the disadvantages. Mutagenesis induced by DNMTs is likely to be detrimental to genomic stability but may not be sufficient to outweigh the selective advantage of DNMTs, particularly if initial damage levels are low or repair is efficient and largely error-free. Indeed, the fact that DNMTs and DDR pathways often co-evolve (Rošić et al., 2018) reinforces the importance of DNA methylation while also highlighting the severity of the threat that unrepaired DNMT-induced damage poses to genome stability. This suggests that presence of DNMTs could be selected against in organisms less dependent on DNA methylation.

Additionally, the work presented in this thesis suggested that DNMT-induced DNA damage is greater than previously thought. Currently, DNMT-induced mutagenesis is understood to be predominantly CG > TA mutations caused by deamination of 5mC to thymine (Lindahl, 1996), as well as mutagenicity arising from the error-prone activity of DNA Polymerase over 5mC (Tomkova et al., 2024). I propose that DNMT-induced mutagenesis may also include mutations resulting from alkylation damage, likely including 3mC. Consequently, the potential contribution of DNMT-induced DNA damage to the evolutionary loss of DNA methylation may be more influential than previously thought.

However, there are several outcomes which would falsify this model. These include absence of 3mC as a result of M.SssI activity, which could be tested via LC-MS; no increase in mutagenesis as a result of direct 3mC incorporation; and rescue of

mutagenesis by re-introduction of catalytically dead AlkB. There are also possible alternatives to this model, whereby mutagenesis arose due to increased metabolic stress caused by M.SssI overexpression burden, SAM depletion or plasmid replication stress. It seems unlikely that these could solely account for the observed increase in mutagenesis, given the statistically significant interactions between M.SssI activity and *alkB* loss which suggested that M.SssI-induced mutagenesis was linked to AlkB-repaired alkylation damage. Furthermore, one might expect broader mutational spectra arising from metabolic stress, rather than an increase in specific types of mutation as observed in this thesis. However, increases in specific mutation types may be biased due to selectivity of the rifampicin resistance assay and so mutagenesis caused by metabolic stress cannot be ruled out in this way. It is possible that M.SssI activity had secondary effects on transcription and metabolism which may have contributed to mutagenesis alongside the primary effect consistent with AlkB-repaired alkylation damage. Further omics work would provide clarity on this.

6.2.2. The relationship between DNMT activity and oxidative stress

The work in Chapter 5 revealed an interesting consequence of DNMT activity: an increase in ROS *in vivo*. While the validity of these results would be strengthened by inclusion of a catalytically dead M.SssI, glucose-repressed inducible pM.SssI was always included as a sufficient negative control. Following subsequent *in vitro* experiments, I proposed that the increase in ROS *in vivo* was due to reaction of methylated cytosine and radical cytosine intermediates with existing cellular ROS.

This builds on previous work whereby DNMT activity increased sensitivity to oxidative stress independent of loss of AlkB (Krwawicz et al., 2024). The evidence in this thesis for DNMT-induced oxidative stress supports a model in which DNMT-induced DNA damage may have contributed to evolutionary loss of DNA methylation, although it should be noted that DNA damage resulting from oxidative stress was not measured. Furthermore, organisms often experience oxidative stress from their environment (Samet & Wages, 2018). Therefore, induction of oxidative stress via DNMT activity could be selectively disadvantageous in certain environments.

Induction of DNMT activity led to an increase in ROS that greatly increased upon addition of H_2O_2. However, the mutagenic consequences of DNMT-induced oxidative stress remain unknown. In order to assess the mutagenicity of DNMT-induced oxidative damage, the rifampicin resistance assay could be repeated in WT and Δ*fpg E. coli* with and without H_2O_2. Hydroxyl radicals from breakdown of H_2O_2 typically attack guanine due to its low redox potential (Steenken & Jovanovic, 1997), resulting in 8oxoG which is excised by the DNA glycosylase FpG/MutM (Boiteux et al., 1992; Tchou et al., 1991). Unrepaired 8oxoG commonly mispairs with adenine, resulting in CG > AT mutations (Moriya & Grollman, 1993; Shibutani et al., 1991). I hypothesise that induction of oxidative stress by H_2O_2 in cells with DNMT activity would overwhelm repair, particularly in Δ*fpg* cells, leading to a synthetic increase in rifampicin-resistant mutation frequency. Sequencing of rifampicin resistant colonies would likely yield an increase in CG > AT mutations due to increased ROS production and subsequent 8oxoG. It is important to note that the M.SssI and Δ*alkB*-

dependent TA > CG signature observed in the rifampicin resistance experiment is likely not the result of M.SssI-induced oxidative stress as Δ*alkB* mutants did not display increased sensitivity to oxidative stress in combination with M.SssI activity (Krwawicz et al., 2024). However, it is possible that there would be an increase in other mutations at or around CpG sites due to radical cytosine intermediates and methylated cytosine derivatives. For example, while 5fC itself is not strongly mutagenic, it does stall DNA Polymerase and so could cause mutations via abasic sites or replication fork collapse (Kamiya et al., 2002). This would be more evident in Δ*fpg* cells. The rifampicin resistance assay would also enable the determination of any mutagenesis dependent on the combination of sugars used for DNMT induction.

6.3. Future work

6.3.1. Direct detection of alkylation lesions

A fundamental issue with investigating mutagenesis is that one will never be able to see the full picture, given that not all damage becomes fixed as mutations. A far more effective way of assessing DNMT-induced damage would be to detect it directly. This would allow for a more accurate measure of initial damage induction rates and could allow for determination of repair capacity by comparing initial damage to final mutagenesis.

There has been significant progress made into direct detection of DNA damage via third-generation sequencing techniques, such as those pioneered by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). While both next-generation and third-generation sequencing offer high-throughput analysis of millions of fragments simultaneously, third-generation techniques can sequence much longer reads from 10-100kbp. Nanopore sequencing provides some mega base reads, albeit with a higher error rate (Amarasinghe et al., 2020) (Scarano et al., 2024).

While these sequencing methods are technically very different, both have been used for the direct detection of DNA damage. PacBio developed single-molecule, real-time (SMRT) DNA sequencing that measures kinetic changes resulting from the movement of DNA polymerase along the DNA strand to identify unmodified and modified bases. Each base results in a specific kinetic profile, comprised of the length of time the polymerase is bound to a nucleotide (pulse width) and the time in between nucleotide-bound states (interpulse duration) (Rhoads & Au, 2015). The polymerase binds over an approximately 11 base region (Berman et al., 2007) which influences kinetic changes and provides information about the local sequence context (Flusberg et al., 2010). This technique has been used to successfully identify multiple types of DNA damage at single nucleotide resolution, including 8oxoG and 1mA (Clark et al., 2011). PacBio SMRT sequencing has also been used in a technique called HiDEF-seq. This aimed to identify the original strand on which ssDNA damage occurred by sequencing single DNA molecules without template amplification with high fidelity. HiDEF-seq successfully identified the original strand on which cytosine

deamination occurred, as well as cancer mutational signatures arising from ssDNA damage events (Liu et al., 2024). Crucially, this technique allowed real-time profiling of ssDNA lesions, which addressed a previously unstudied gap in mutagenesis. Therefore, PacBio SMRT-seq would be suitable for the direct detection of DNMT-induced alkylation damage. This would also address a major limitation of determining mutagenesis via Sanger sequencing in that it could identify the original lesion-containing strand, which would help to deconvolute *rpoB*-specific mutation spectra observed outside of the CpG context.

ONT sequencing uses nanoscale protein pores known as nanopores embedded in a membrane to sequence ssDNA and RNA. Negatively charged nucleic acid molecules translocate through the nanopore via electrophoresis and cause changes to the ionic current across the pore. These electrical changes are decoded to give the sequence of modified and unmodified bases that passed through the pore at single-nucleotide resolution (van Dijk et al., 2018; Wang et al., 2021). In a similar way to PacBio SMRT-seq, modified bases produce unique traces that can be distinguished from the corresponding unmodified base trace. Currently, ONT sequencing can detect modified bases such as 5mC, 5hmC and 6mA (Kovaka et al., 2025; Liu et al., 2019; Rand et al., 2017; Stoiber et al., 2017; Xu & Seki, 2020). However, neither PacBio SMRT-seq nor ONT sequencing currently detect alkylation lesions such as 3mC and 1mA. This would require synthetic DNA constructs containing these lesions at defined sites, which could be used to generate reference traces. They could then be used for the identification of alkylation lesions in DNA extracted from cells with DNMT activity. This concept has been performed

successfully for the identification of 8oxoG via ONT sequencing (Pagès-Gallego et al., 2025). It is feasible that the DNA damage plasmid constructs used in this thesis could be suitable for such experiments.

6.3.2. Further characterisation of DNMT-induced mutagenesis

6.3.2.1. Do DNMTs introduce mutations in mammalian systems?

The work in this thesis was conducted in *E. coli* because they lack regulation of gene expression via cytosine DNA methylation. This enabled investigation into mutagenesis that arose directly from damage induced by cytosine DNMT activity, not secondary mutagenesis arising from changes to gene expression. While mutagenesis arising from aberrant methylation is clinically relevant to cancer research, it is already well studied (Feinberg & Vogelstein, 1983; Flatau et al., 1983; Hanahan & Weinberg, 2011; Lee et al., 2002; Lin et al., 2001; Muggerud et al., 2010; Rauch et al., 2008; Watts et al., 2008) and was not the purpose of this thesis.

Having said that, introducing CpG methylation into *E. coli* was not a perfect replica of mammalian DNA methylation. M.SssI is a bacterial cytosine DNMT that possess both *de novo* and maintenance methyltransferase capability. Conversely, mammals possess three cytosine DNMTs: DNMT1, DNMT3A and DNMT3B, with either maintenance or *de novo* activity (Pradhan et al., 1999; Xie et al., 1999). Furthermore, M.SssI has high processivity and does not require additional regulatory domains, resulting in methylation of up to 100% of CpG sites *in vitro* (Hemeon et al., 2011).

While also highly processive, DNMT3A/B require additional regulatory proteins such as DNMT3L (Jia et al., 2007; Ooi et al., 2007) and activity is restricted by active chromatin domains and histone modifications such as H3K4me3 (Cedar & Bergman, 2009; Fu et al., 2020; Lehnertz et al., 2003). The complex genome architecture in mammals results in more variable methylation patterns, for example unmethylated CpG islands, with only 60-80% of total CpG sites methylated (Lister et al., 2009; Schultz et al., 2016). It follows that DNMT-induced damage and subsequent repair or mutagenesis in mammals may be similarly influenced by factors restricting DNA methylation. Perhaps an alternative would be to introduce mammalian DNMT1 and DNMT3A/B into *E. coli,* to determine whether the observed mutagenesis was M.SssI-specific.

Ideally, the DNMT-induced mutagenesis observed in this thesis would be explored further in mammalian models such as cancer cell lines. However, this would be difficult due to the confounding influence of DNA methylation on gene expression and wider epigenetic regulation. Cancer cell lines are also typically highly mutated due to their origin and immortalisation, which could affect response to synthetic induction of methylation. Other mammalians models such as mouse ESCs could be used instead; indeed, previous work identified an increase in DNMT-induced 3mC following ALKB2 KO in mouse ESCs (Rošić et al., 2018). The work in this thesis suggested that there could have been a subsequent increase in mutagenesis in these mutants. Therefore, future work could involve whole genome sequencing (WGS) of mouse ESC ALKB2 KO mutants to identify DNMT-induced mutational signatures.

6.3.2.2. Further characterisation of DNMT-induced mutagenesis in *E. coli*

A pertinent question is why not perform WGS of *E. coli* genomes to identify mutagenesis? As previously mentioned, one of the main limitations of the rifampicin resistance assay was that not all mutations conferred resistance to rifampicin, therefore not all mutations were observed. This could be overcome by WGS of *E. coli* gDNA without rifampicin selection. However, given the low induction rate of 3mC by M.SssI (approximately 0.35% of 5mC (Rošić et al., 2018)), resulting mutation rate would likely be very low. Therefore, without rifampicin selection of mutants, a vast quantity of WGS would likely return no mutations and would waste time and resources. Perhaps having demonstrated mutagenesis due to DNMT activity in Δ*alkB E. coli* using the rifampicin resistance assay and targeted Sanger sequencing of *rpoB*, future work could include WGS of rifampicin-resistant mutant gDNA. This would ensure that all DNA sequenced would contain at least one mutation, as well as additional mutations not under rifampicin selection.

It was assumed that introducing cytosine DNA methylation into *E. coli* would cause minimal changes to gene expression. However, substantial transcriptional changes were observed as a result of synthetic cytosine DNMT activity in previous work (Krwawicz et al., 2024). This suggested that 5mC may have affected regulation of gene expression. Understanding the transcriptional effects of synthetic DNMT activity in *E. coli* could prove useful in elucidating the conditional origin of the non-cytosine *rpoB*-specific mutation spectrum. However, it is worth noting that some of the changes in (Krwawicz et al., 2024) were attributed to secondary effects of

introducing DNMT activity, namely those associated with repair of alkylation damage. Therefore, changes to gene expression would need to be carefully analysed to determine whether they were directly due to DNA methylation or secondary consequences.

Overall, perhaps the best course of action would be to develop sequencing standards for direct detection of DNMT-induced DNA damage in either mouse ESCs or *E. coli*, then to look for evidence of this damage and associated mutations in existing cancer sequencing data. This would be particularly relevant to cancer types with hypermethylation (Esteller, 2000; Esteller et al., 2001; Paz et al., 2003) or those typified by extensive 5mC deamination resulting from DNMT activity over time (Alexandrov et al., 2013). While it would be interesting to know the mechanistic or genetic causes behind the M.SssI and Δ*alkB*-dependent non-cytosine signature observed in this thesis, I suggest that it would be more important to further investigate the wider mutagenic effects of DNMT activity in a more clinically relevant context.

6.4. Clinical implications of DNMT-induced genome instability

Genome instability is now considered an enabling characteristic of cancer; it facilitates the acquisition of typical capabilities or hallmarks of cancer cells (Hanahan & Weinberg, 2011). Genome instability can be caused by a number of exogenous and endogenous processes (Cannataro et al., 2022; Irigaray & Belpomme, 2009; Jackson & Loeb, 2001; Kumari et al., 2022; Tubbs & Nussenzweig,

2017), including aberrant DNA methylation (Esteller, 2000; Esteller et al., 2001; Gonzalgo & Jones, 1997). Aberrant DNA methylation can promote tumorigenesis in two distinct ways: 1) Changes in DNA methylation at promoters can alter epigenetic control of tumour suppressor and oncogenes and 2) CG > TA mutations caused by deamination of 5mC to thymine in coding regions of genes disrupt tumour suppressor function or produce dominant negative proteins which antagonise the WT (Esteller et al., 2000; Gonzalgo & Jones, 1997).

Changes to epigenetic modifications such as DNA methylation are referred to as epimutations: heritable changes in gene expression that do not alter the underlying DNA sequence. There are many examples of DNA methylation epimutations that give rise to cancer (Dobrovic & Kristensen, 2009; Hesson et al., 2010; Peltomäki, 2012); in fact, cancers generally exhibit promoter hypermethylation coupled with gene body hypomethylation (Feinberg & Vogelstein, 1983; Wilson et al., 2007). For example, MGMT repairs pro-mutagenic 6mG, preventing G > A mutations (Pegg et al., 1995). Loss of MGMT activity has been observed in many cancer types without any mutations to the gene. Instead, it was found that *MGMT* was transcriptionally silenced due to promoter hypermethylation (Esteller et al., 1999; Qian & Brent, 1997; Watts et al., 1997). Similarly, loss of *BRCA1* expression in sporadic breast and ovarian cancers was associated with promoter hypermethylation but a lack of somatic mutations (Catteau et al., 1999; Dobrovic & Simpfendorfer, 1997; Esteller et al., 2000).

Despite the prevalence of DNA methylation epimutations in cancer and increased understanding of their importance, the work presented in this thesis pertains to the second way in which DNA methylation can promote tumorigenesis. Deamination of 5mC to thymine and resulting CG > TA mutations have been well documented across various cancer types, so much so that deamination of 5mC is now recognised as the mechanism responsible for cancer mutational signatures 1A and B. These are the only cancer mutational signatures associated with age and are found in over 60% of cancer samples, including those from breast, prostate, stomach and pancreas (Alexandrov et al., 2015). Furthermore, almost a quarter of point mutations in the tumour suppressor gene *TP53* are CG > TA mutations found at CpG sites (Greenblatt et al., 1994). These are likely the result of 5mC deamination, as all CpG mutational hotspots in *TP53* have been shown to be methylated (Magewu & Jones, 1994; Rideout et al., 1990; Tornaletti & Pfeifer, 1995). This results in defective or dominant negative forms of the protein (de Vries et al., 2002). This represents a significant contribution of DNMT-induced mutagenesis to tumour development without affecting gene expression, especially since *TP53* is mutated in over half of all solid human tumours (Greenblatt et al., 1994).

I propose that DNMT-induced alkylation damage and potentially oxidative damage may also cause mutagenesis that could contribute to tumorigenesis. The observed increase in CG > TA mutations dependent on DNMT activity and loss of AlkB suggested that not all CG > TA mutations arose due to deamination of 5mC. Some were likely the result of misincorporation opposite 3mC lesions, potentially by TLS polymerases (Furrer & van Loon, 2014). Therefore, it is possible that DNMT-induced

alkylation lesions may contribute to cancer mutational signatures, particularly 1A and B. The contribution of DNMT-induced alkylation lesions to mutagenesis could be particularly relevant in cancers deficient in DDR pathways such as BER (Farmer et al., 2005; Markkanen et al., 2015; Smal et al., 2018). ALKB2/3 are not tumour suppressors and so are not typically associated with cancer progression. However, there is some evidence for epigenetic silencing of the *ALKB3* promoter by DNA methylation in breast cancer leading to the accumulation of alkylation damage (Stefansson et al., 2017). It should also be noted that many cancers including prostate and lung exhibit increased ALKB2/3 expression potentially in order to facilitate DNA repair in tumour cells (Konishi et al., 2005; Pilžys et al., 2019; Tasaki et al., 2011). Therefore, further investigation into the contribution of DNMT-induced alkylation damage to tumorigenesis should perhaps be focused on cancer types with increased CpG methylation in coding regions of tumour suppressors as well as related DDR deficiencies. A significant contribution to tumorigenesis could also indicate cancer types that would be particularly sensitive to treatment with alkylating agents that would overwhelm repair capacity in a synthetic lethal manner.

6.5. Overall conclusion

In conclusion, this work demonstrated that CpG DNMT activity increased CG > TA mutagenesis; an effect that was exacerbated by KO of *alkB* in *E. coli*. Therefore, this was likely the result of alkylation damage normally repaired by AlkB, such as 3mC. However, there was also evidence of non-cytosine mutations dependent on M.SssI activity, loss of AlkB and strain background. This was not dependent on the presence

of RecA alone, although the data presented in this thesis did not rule out potential contributions of other TLS components, SOS induction or sulA-dependent cell cycle control. Further work would be required to elucidate the mechanism behind this signature. This work also demonstrated the propensity of DNMT activity to introduce oxidative stress in the form of ROS, likely arising indirectly from radical cytosine intermediates following DNA methylation. Collectively, this work provides two additional ways in which DNMT activity may damage the genome and cause mutagenesis, outside of canonical deamination of 5mC and associated replication errors (Lindahl, 1996; Tomkova et al., 2024). This is consistent with the hypothesis that DNMT-induced DNA damage and mutagenesis may have contributed to the evolutionary loss of DNMTs (Rošić et al., 2018). Additionally, such mutagenesis could play a previously unacknowledged role in the onset of cancer, implicating DNA methylation further in human health and disease, as well as potentially providing much needed therapeutic avenues.

V. References

Aas, P. A., Otterlei, M., Falnes, P. O., Vågbø, C. B., Skorpen, F., Akbari, M.,...Krokan, H. E. (2003). Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, *421*(6925), 859-863. https://doi.org/10.1038/nature01363

Adachi, M., Autry, A. E., Covington, H. E., & Monteggia, L. M. (2009). MeCP2-mediated transcription repression in the basolateral amygdala may underlie heightened anxiety in a mouse model of Rett syndrome. *J Neurosci*, *29*(13), 4218-4227. https://doi.org/10.1523/JNEUROSCI.4225-08.2009

Admiraal, S. J., Eyler, D. E., Baldwin, M. R., Brines, E. M., Lohans, C. T., Schofield, C. J., & O'Brien, P. J. (2019). Expansion of base excision repair compensates for a lack of DNA repair by oxidative dealkylation in budding yeast. *J Biol Chem*, *294*(37), 13629-13637. https://doi.org/10.1074/jbc.RA119.009813

Afri, M., Frimer, A. A., & Cohen, Y. (2004). Active oxygen chemistry within the liposomal bilayer. Part IV: Locating 2',7'-dichlorofluorescein (DCF), 2',7'-dichlorodihydrofluorescein (DCFH) and 2',7'-dichlorodihydrofluorescein diacetate (DCFH-DA) in the lipid bilayer. *Chem Phys Lipids*, *131*(1), 123-133. https://doi.org/10.1016/j.chemphyslip.2004.04.006

Ahmed, M., & Liang, P. (2012). Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp Funct Genomics*, *2012*, 947089. https://doi.org/10.1155/2012/947089

Al Adhami, H., Bardet, A. F., Dumas, M., Cleroux, E., Guibert, S., Fauque, P.,...Weber, M. (2022). A comparative methylome analysis reveals conservation and divergence of DNA methylation patterns and functions in vertebrates. *BMC Biol*, *20*(1), 70. https://doi.org/10.1186/s12915-022-01270-x

Alberts, B. (2002). *Molecular biology of the cell* (4th ed.). Garland Science.

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nat Genet*, *47*(12), 1402-1407. https://doi.org/10.1038/ng.3441

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y.,...Consortium, P. (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94-101. https://doi.org/10.1038/s41586-020-1943-3

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V.,...PedBrain, I. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415-421. https://doi.org/10.1038/nature12477

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*, *21*(1), 30. https://doi.org/10.1186/s13059-020-1935-5

Anderson, C. J., Talmane, L., Luft, J., Connelly, J., Nicholson, M. D., Verburg, J. C.,...Consortium, L. C. E. (2024). Strand-resolved mutagenicity of DNA damage and repair. *Nature*, *630*(8017), 744-751. https://doi.org/10.1038/s41586-024-07490-1

Aravind, L., & Koonin, E. V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*, *2*(3), RESEARCH0007. https://doi.org/10.1186/gb-2001-2-3-research0007

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M.,...Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, *2*, 2006.0008. https://doi.org/10.1038/msb4100050

Bellacosa, A., & Drohat, A. C. (2015). Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA Repair (Amst)*, *32*, 33-42. https://doi.org/10.1016/j.dnarep.2015.04.011

Bercovich, J. A., Grinstein, S., & Zorzopulos, J. (1992). Effect of DNA concentration on recombinant plasmid recovery after blunt-end ligation. *Biotechniques*, *12*(2), 190, 192-193.

Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev*, *23*(7), 781-783. https://doi.org/10.1101/gad.1787609

Berman, A. J., Kamtekar, S., Goodman, J. L., Lázaro, J. M., de Vega, M., Blanco, L.,...Steitz, T. A. (2007). Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J*, *26*(14), 3494-3505. https://doi.org/10.1038/sj.emboj.7601780

Bernstein, N. K., Williams, R. S., Rakovszky, M. L., Cui, D., Green, R., Karimi-Busheri, F.,...Glover, J. N. (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell*, *17*(5), 657-670. https://doi.org/10.1016/j.molcel.2005.02.012

Besselink, N., Keijer, J., Vermeulen, C., Boymans, S., de Ridder, J., van Hoeck, A.,...Kuijk, E. (2023). The genome-wide mutational consequences of DNA hypomethylation. *Sci Rep*, *13*(1), 6874. https://doi.org/10.1038/s41598-023-33932-3

Bewick, A. J., Hofmeister, B. T., Powers, R. A., Mondo, S. J., Grigoriev, I. V., James, T. Y.,...Schmitz, R. J. (2019). Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol*, *3*(3), 479-490. https://doi.org/10.1038/s41559-019-0810-9

Bewick, A. J., Vogel, K. J., Moore, A. J., & Schmitz, R. J. (2016). Evolution of DNA Methylation across Insects. *Molecular Biology and Evolution*, *34*(3), 654-665. https://doi.org/10.1093/molbev/msw264

Bi, E., & Lutkenhaus, J. (1993). Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *J Bacteriol*, *175*(4), 1118-1125. https://doi.org/10.1128/jb.175.4.1118-1125.1993

Billington, O. J., McHugh, T. D., & Gillespie, S. H. (1999). Physiological cost of rifampin resistance induced in vitro in Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, *43*(8), 1866-1869. https://doi.org/10.1128/AAC.43.8.1866

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91-99. https://doi.org/10.1016/0092-8674(85)90312-5

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res*, *8*(7), 1499-1504. https://doi.org/10.1093/nar/8.7.1499

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, *321*(6067), 209-213. https://doi.org/10.1038/321209a0

Bishop, A. J., & Schiestl, R. H. (2000). Homologous recombination as a mechanism for genome rearrangements: environmental and genetic effects. *Hum Mol Genet*, *9*(16), 2427-2334. https://doi.org/10.1093/hmg/9.16.2427

Bjelland, S., & Seeberg, E. (1996). Different efficiencies of the Tag and AlkA DNA glycosylases from Escherichia coli in the removal of 3-methyladenine from single-stranded DNA. *FEBS Lett*, *397*(1), 127-129. https://doi.org/10.1016/s0014-5793(96)01166-0

Blundell, J. R., Schwartz, K., Francois, D., Fisher, D. S., Sherlock, G., & Levy, S. F. (2019). The dynamics of adaptive genetic diversity during the early stages of clonal evolution. *Nat Ecol Evol*, *3*(2), 293-301. https://doi.org/10.1038/s41559-018-0758-1

Bochtler, M., Kolano, A., & Xu, G. L. (2017). DNA demethylation pathways: Additional players and regulators. *Bioessays*, *39*(1), 1-13. https://doi.org/10.1002/bies.201600178

Boeke, J. D., Garfinkel, D. J., Styles, C. A., & Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. *Cell*, *40*(3), 491-500. https://doi.org/10.1016/0092-8674(85)90197-7

Boiteux, S., Gajewski, E., Laval, J., & Dizdaroglu, M. (1992). Substrate specificity of the Escherichia coli Fpg protein (formamidopyrimidine-DNA glycosylase): excision of purine lesions in DNA produced by ionizing radiation or photosensitization. *Biochemistry*, *31*(1), 106-110. https://doi.org/10.1021/bi00116a016

Boiteux, S., & Laval, J. (1982). Mutagenesis by alkylating agents: coding properties for DNA polymerase of poly (dC) template containing 3-methylcytosine. *Biochimie*, *64*(8-9), 637-641. https://doi.org/10.1016/s0300-9084(82)80103-x

Boiteux, S., O'Connor, T. R., & Laval, J. (1987). Formamidopyrimidine-DNA glycosylase of Escherichia coli: cloning and sequencing of the fpg structural gene and overproduction of the protein. *EMBO J*, *6*(10), 3177-3183. https://doi.org/10.1002/j.1460-2075.1987.tb02629.x

Bork, J. M., Cox, M. M., & Inman, R. B. (2001). RecA protein filaments disassemble in the 5' to 3' direction on single-stranded DNA. *J Biol Chem*, *276*(49), 45740-45743. https://doi.org/10.1074/jbc.M109247200

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, *431*(7004), 96-99. https://doi.org/10.1038/nature02886

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J.,...Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, *71*(3), 515-526. https://doi.org/10.1016/0092-8674(92)90519-i

Bruskov, V. I., Malakhova, L. V., Masalimov, Z. K., & Chernikov, A. V. (2002). Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker

of damage to DNA. *Nucleic Acids Res*, *30*(6), 1354-1363. https://doi.org/10.1093/nar/30.6.1354

Buck-Koehntop, B. A., & Defossez, P. A. (2013). On how mammalian transcription factors recognize methylated DNA. *Epigenetics*, *8*(2), 131-137. https://doi.org/10.4161/epi.23632

Burma, S., Chen, B. P., & Chen, D. J. (2006). Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)*, *5*(9-10), 1042-1048. https://doi.org/10.1016/j.dnarep.2006.05.026

Busby, S., & Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J Mol Biol*, *293*(2), 199-213. https://doi.org/10.1006/jmbi.1999.3161

Buxton, G. V., Greenstock, C. L., Helman, W. P., & Ross, A. B. (1988). Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals (·OH/·O− in Aqueous Solution. *Journal of Physical and Chemical Reference Data*, *17*(2), 513-886. https://doi.org/10.1063/1.555805

Cadet, J., & Wagner, J. R. (2013). DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb Perspect Biol*, *5*(2). https://doi.org/10.1101/cshperspect.a012559

Caldecott, K. W., Tucker, J. D., Stanker, L. H., & Thompson, L. H. (1995). Characterization of the XRCC1-DNA ligase III complex in vitro and its absence from mutant hamster cells. *Nucleic Acids Res*, *23*(23), 4836-4843. https://doi.org/10.1093/nar/23.23.4836

Cannataro, V. L., Mandell, J. D., & Townsend, J. P. (2022). Attribution of Cancer Origins to Endogenous, Exogenous, and Preventable Mutational Processes. *Molecular Biology and Evolution*, *39*(5). https://doi.org/10.1093/molbev/msac084

Cappelli, E., Taylor, R., Cevasco, M., Abbondandolo, A., Caldecott, K., & Frosina, G. (1997). Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. *J Biol Chem*, *272*(38), 23970-23975. https://doi.org/10.1074/jbc.272.38.23970

Casadesús, J., & Low, D. (2006). Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev*, *70*(3), 830-856. https://doi.org/10.1128/MMBR.00016-06

Caspary, T., Cleary, M. A., Baker, C. C., Guan, X. J., & Tilghman, S. M. (1998). Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol Cell Biol*, *18*(6), 3466-3474. https://doi.org/10.1128/MCB.18.6.3466

Catteau, A., Harris, W. H., Xu, C.-F., & Solomon, E. (1999). Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics. *Oncogene*, *18*(11), 1957-1965.

Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*, *10*(5), 295-304. https://doi.org/10.1038/nrg2540

Chastain, P. D., Nakamura, J., Rao, S., Chu, H., Ibrahim, J. G., Swenberg, J. A., & Kaufman, D. G. (2010). Abasic sites preferentially form at regions undergoing DNA replication. *FASEB J*, *24*(10), 3674-3680. https://doi.org/10.1096/fj.09-145276

Chetsanga, C. J., & Lindahl, T. (1979). Release of 7-methylguanine residues whose imidazole rings have been opened from damaged DNA by a DNA glycosylase from Escherichia coli. *Nucleic Acids Res*, *6*(11), 3673-3684. https://doi.org/10.1093/nar/6.11.3673

Clark, T. A., Spittle, K. E., Turner, S. W., & Korlach, J. (2011). Direct detection and sequencing of damaged DNA bases. *Genome Integr*, *2*, 10. https://doi.org/10.1186/2041-9414-2-10

Collins, A. R. (1999). Oxidative DNA damage, antioxidants, and cancer. *Bioessays*, *21*(3), 238-246. https://doi.org/10.1002/(SICI)1521-1878(199903)21:3<238::AID-BIES8>3.0.CO;2-3

Collins, M., & Rubin, G. M. (1984). Structure of chromosomal rearrangements induced by the FB transposable element in Drosophila. *Nature*, *308*(5957), 323-327. https://doi.org/10.1038/308323a0

Cortesao, C. S., Freitas, R. F., & Barreto, V. M. (2013). Activation-Induced Cytidine Deaminase Does Not Impact Murine Meiotic Recombination. *G3 (Bethesda)*, *3*(4), 645-655. https://doi.org/10.1534/g3.113.005553

Courcelle, J., Khodursky, A., Peter, B., Brown, P. O., & Hanawalt, P. C. (2001). Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics*, *158*(1), 41-64. https://doi.org/10.1093/genetics/158.1.41

Dango, S., Mosammaparast, N., Sowa, M. E., Xiong, L. J., Wu, F., Park, K.,...Shi, Y. (2011). DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. *Mol Cell*, *44*(3), 373-384. https://doi.org/10.1016/j.molcel.2011.08.039

Datsenko, K. A., & 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. https://doi.org/10.1073/pnas.120163297

De Bont, R., & van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, *19*(3), 169-185. https://doi.org/10.1093/mutage/geh025

de Mendoza, A., Hatleberg, W. L., Pang, K., Leininger, S., Bogdanovic, O., Pflueger, J.,...Lister, R. (2019). Convergent evolution of a vertebrate-like methylome in a marine sponge. *Nat Ecol Evol*, *3*(10), 1464-1473. https://doi.org/10.1038/s41559-019-0983-2

de Mendoza, A., Pflueger, J., & Lister, R. (2019). Capture of a functionally active methyl-CpG binding domain by an arthropod retrotransposon family. *Genome Res*, *29*(8), 1277-1286. https://doi.org/10.1101/gr.243774.118

de Vries, A., Flores, E. R., Miranda, B., Hsieh, H.-M., van Oostrom, C. T. M., Sage, J., & Jacks, T. (2002). Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function. *Proceedings of the National Academy of Sciences*, *99*(5), 2948-2953. https://doi.org/doi:10.1073/pnas.052713099

Del Val, E., Nasser, W., Abaibou, H., & Reverchon, S. (2019). RecA and DNA recombination: a review of molecular mechanisms. *Biochem Soc Trans*, *47*(5), 1511-1531. https://doi.org/10.1042/BST20190558

Denver, D. R., Swenson, S. L., & Lynch, M. (2003). An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol Biol Evol*, *20*(10), 1603-1611. https://doi.org/10.1093/molbev/msg177

Desai, T. A., & Rao, C. V. (2010). Regulation of arabinose and xylose metabolism in Escherichia coli. *Appl Environ Microbiol*, *76*(5), 1524-1532. https://doi.org/10.1128/AEM.01970-09

Dianov, G., & Lindahl, T. (1994). Reconstitution of the DNA base excision-repair pathway. *Curr Biol*, *4*(12), 1069-1076. https://doi.org/10.1016/s0960-9822(00)00245-1

Dobrovic, A., & Kristensen, L. S. (2009). DNA methylation, epimutations and cancer predisposition. *The International Journal of Biochemistry & Cell Biology*, *41*(1), 34-39. https://doi.org/https://doi.org/10.1016/j.biocel.2008.09.006

Dobrovic, A., & Simpfendorfer, D. (1997). Methylation of the BRCA1 gene in sporadic breast cancer. *Cancer research*, *57*(16), 3347-3350.

Dou, H., Mitra, S., & Hazra, T. K. (2003). Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J Biol Chem*, *278*(50), 49679-49684. https://doi.org/10.1074/jbc.M308658200

Drabløs, F., Feyzi, E., Aas, P. A., Vaagbø, C. B., Kavli, B., Bratlie, M. S.,...Krokan, H. E. (2004). Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)*, *3*(11), 1389-1407. https://doi.org/10.1016/j.dnarep.2004.05.004

Du, H., Wang, P., Li, L., & Wang, Y. (2019). Repair and translesion synthesis of O6-alkylguanine DNA lesions in human cells. *J Biol Chem*, *294*(29), 11144-11153. https://doi.org/10.1074/jbc.RA119.009054

Dukatz, M., Requena, C. E., Emperle, M., Hajkova, P., Sarkies, P., & Jeltsch, A. (2019). Mechanistic Insights into Cytosine-N3 Methylation by DNA Methyltransferase DNMT3A. *J Mol Biol*, *431*(17), 3139-3145. https://doi.org/10.1016/j.jmb.2019.06.015

Duncan, B. K., & Miller, J. H. (1980). Mutagenic deamination of cytosine residues in DNA. *Nature*, *287*(5782), 560-561. https://doi.org/10.1038/287560a0

Duncan, T., Trewick, S. C., Koivisto, P., Bates, P. A., Lindahl, T., & Sedgwick, B. (2002). Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A*, *99*(26), 16660-16665. https://doi.org/10.1073/pnas.262589799

Dunn, D. B., & Smith, J. D. (1958). The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochem J*, *68*(4), 627-636. https://doi.org/10.1042/bj0680627

Esteller, M. (2000). Epigenetic lesions causing genetic lesions in human cancer: promoter hypermethylation of DNA repair genes. *Eur J Cancer*, *36*(18), 2294-2300. https://doi.org/10.1016/s0959-8049(00)00303-8

Esteller, M., Corn, P. G., Baylin, S. B., & Herman, J. G. (2001). A gene hypermethylation profile of human cancer. *Cancer Res*, *61*(8), 3225-3229.

Esteller, M., Hamilton, S. R., Burger, P. C., Baylin, S. B., & Herman, J. G. (1999). Inactivation of the DNA repair gene O 6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer research*, *59*(4), 793-797.

Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E.,...Repasky, E. A. (2000). Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *JNCI: Journal of the National Cancer Institute*, *92*(7), 564-569.

Falnes, P., Johansen, R. F., & Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature*, *419*(6903), 178-182. https://doi.org/10.1038/nature01048

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N., Johnson, D. A., Richardson, T. B.,...Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, *434*(7035), 917-921. https://doi.org/10.1038/nature03445

Feinberg, A. P., Ohlsson, R., & Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nat Rev Genet*, *7*(1), 21-33. https://doi.org/10.1038/nrg1748

Feinberg, A. P., & Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, *301*(5895), 89-92. https://doi.org/10.1038/301089a0

Fernández De Henestrosa, A. R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H., & Woodgate, R. (2000). Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol Microbiol*, *35*(6), 1560-1572. https://doi.org/10.1046/j.1365-2958.2000.01826.x

Flatau, E., Bogenmann, E., & Jones, P. A. (1983). Variable 5-methylcytosine levels in human tumor cell lines and fresh pediatric tumor explants. *Cancer Res*, *43*(10), 4901-4905.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A.,...Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*, *7*(6), 461-465. https://doi.org/10.1038/nmeth.1459

Fortini, P., & Dogliotti, E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst)*, *6*(4), 398-409. https://doi.org/10.1016/j.dnarep.2006.10.008

Frank, E. G., Ennis, D. G., Gonzalez, M., Levine, A. S., & Woodgate, R. (1996). Regulation of SOS mutagenesis by proteolysis. *Proc Natl Acad Sci U S A*, *93*(19), 10291-10296. https://doi.org/10.1073/pnas.93.19.10291

Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, *171*(4356), 740-741. https://doi.org/10.1038/171740a0

Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L. S.,...Dogliotti, E. (1996). Two pathways for base excision repair in mammalian cells. *J Biol Chem*, *271*(16), 9573-9578. https://doi.org/10.1074/jbc.271.16.9573

Fu, K., Bonora, G., & Pellegrini, M. (2020). Interactions between core histone marks and DNA methyltransferases predict DNA methylation patterns observed in human cells and tissues. *Epigenetics*, *15*(3), 272-282. https://doi.org/10.1080/15592294.2019.1666649

Furrer, A., & van Loon, B. (2014). Handling the 3-methylcytosine lesion by six human DNA polymerases members of the B-, X- and Y-families. *Nucleic Acids Res*, *42*(1), 553-566. https://doi.org/10.1093/nar/gkt889

Gao, F., Liu, X., Wu, X. P., Wang, X. L., Gong, D., Lu, H.,...Liu, M. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode Trichinella spiralis. *Genome Biol*, *13*(10), R100. https://doi.org/10.1186/gb-2012-13-10-r100

Gardiner-Garden, M., & Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol*, *196*(2), 261-282. https://doi.org/10.1016/0022-2836(87)90689-9

Garibyan, L., Huang, T., Kim, M., Wolff, E., Nguyen, A., Nguyen, T.,...Miller, J. H. (2003). Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. *DNA Repair (Amst)*, *2*(5), 593-608. https://doi.org/10.1016/s1568-7864(03)00024-7

Gaudet, F., Rideout, W. M., Meissner, A., Dausman, J., Leonhardt, H., & Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. *Mol Cell Biol*, *24*(4), 1640-1648. https://doi.org/10.1128/MCB.24.4.1640-1648.2004

Geiman, T. M., Sankpal, U. T., Robertson, A. K., Chen, Y., Mazumdar, M., Heale, J. T.,...Robertson, K. D. (2004). Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res*, *32*(9), 2716-2729. https://doi.org/10.1093/nar/gkh589

Geurts, A. M., Collier, L. S., Geurts, J. L., Oseth, L. L., Bell, M. L., Mu, D.,...Largaespada, D. A. (2006). Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. *PLoS Genet*, *2*(9), e156. https://doi.org/10.1371/journal.pgen.0020156

Gilbert, N., Thomson, I., Boyle, S., Allan, J., Ramsahoye, B., & Bickmore, W. A. (2007). DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J Cell Biol*, *177*(3), 401-411. https://doi.org/10.1083/jcb.200607133

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X.,...Bestor, T. H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science*, *311*(5759), 395-398. https://doi.org/10.1126/science.1120976

Gonzalgo, M. L., & Jones, P. A. (1997). Mutagenic and epigenetic effects of DNA methylation. *Mutation Research/Reviews in Mutation Research*, *386*(2), 107-118. https://doi.org/https://doi.org/10.1016/S1383-5742(96)00047-6

Goodman, M. F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem*, *71*, 17-50. https://doi.org/10.1146/annurev.biochem.71.083101.124707

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*, *20*(10), 590-607. https://doi.org/10.1038/s41580-019-0159-6

Greenblatt, I. M., & Brink, A. R. (1963). Transpositions of Modulator in Maize into Divided and Undivided Chromosome Segments. *Nature*, *197*(4865), 412-413. https://doi.org/10.1038/197412a0

Greenblatt, M. S., Bennett, W. P., Hollstein, M., & Harris, C. C. (1994). Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res*, *54*(18), 4855-4878.

Gross, J., & Englesberg, E. (1959). Determination of the order of mutational sites governing L-arabinose utilization in Escherichia coli B/r bv transduction with phage Plbt. *Virology*, *9*, 314-331. https://doi.org/10.1016/0042-6822(59)90125-4

Groth, P., Ausländer, S., Majumder, M. M., Schultz, N., Johansson, F., Petermann, E., & Helleday, T. (2010). Methylated DNA causes a physical block to replication forks independently of damage signalling, O(6)-methylguanine or DNA single-strand breaks and results in DNA damage. *J Mol Biol*, *402*(1), 70-82. https://doi.org/10.1016/j.jmb.2010.07.010

Guo, J. U., Su, Y., Zhong, C., Ming, G. L., & Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, *145*(3), 423-434. https://doi.org/10.1016/j.cell.2011.03.022

Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol*, *177*(14), 4121-4130. https://doi.org/10.1128/jb.177.14.4121-4130.1995

Görke, B., & Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, *6*(8), 613-624. https://doi.org/10.1038/nrmicro1932

Halliwell, B., & Gutteridge, J. M. C. (2015). *Free Radicals in Biology and Medicine*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198717478.001.0001

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674. https://doi.org/10.1016/j.cell.2011.02.013

Hanawalt, P. C., & Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol*, *9*(12), 958-970. https://doi.org/10.1038/nrm2549

Hare, J. T., & Taylor, J. H. (1985). One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. *Proc Natl Acad Sci U S A*, *82*(21), 7350-7354. https://doi.org/10.1073/pnas.82.21.7350

Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P.,...Mitra, S. (2002). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc Natl Acad Sci U S A*, *99*(6), 3523-3528. https://doi.org/10.1073/pnas.062053799

Heller, R. C., & Marians, K. J. (2006). Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol*, *7*(12), 932-943. https://doi.org/10.1038/nrm2058

Hemeon, I., Gutierrez, J. A., Ho, M. C., & Schramm, V. L. (2011). Characterizing DNA methyltransferases with an ultrasensitive luciferase-linked continuous assay. *Anal Chem*, *83*(12), 4996-5004. https://doi.org/10.1021/ac200816m

Henderson, I. R., & Jacobsen, S. E. (2007). Epigenetic inheritance in plants. *Nature*, *447*(7143), 418-424. https://doi.org/10.1038/nature05917

Hermann, A., Goyal, R., & Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem*, *279*(46), 48350-48359. https://doi.org/10.1074/jbc.M403427200

Hesson, L. B., Hitchins, M. P., & Ward, R. L. (2010). Epimutations and cancer predisposition: importance and mechanisms. *Current Opinion in Genetics & Development*, *20*(3), 290-298. https://doi.org/https://doi.org/10.1016/j.gde.2010.02.005

Higashitani, A., Ishii, Y., Kato, Y., & Koriuchi, K. (1997). Functional dissection of a cell-division inhibitor, SulA, of Escherichia coli and its negative regulation by Lon. *Mol Gen Genet*, *254*(4), 351-357. https://doi.org/10.1007/s004380050426

Holmström, K. M., & Finkel, T. (2014). Cellular mechanisms and physiological consequences of redox-dependent signalling. *Nat Rev Mol Cell Biol*, *15*(6), 411-421. https://doi.org/10.1038/nrm3801

Horton, J. K., Stefanick, D. F., Kedar, P. S., & Wilson, S. H. (2007). ATR signaling mediates an S-phase checkpoint after inhibition of poly(ADP-ribose) polymerase activity. *DNA Repair (Amst)*, *6*(6), 742-750. https://doi.org/10.1016/j.dnarep.2006.12.015

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H. N.,...Zhu, H. (2013). DNA methylation presents distinct binding sites for human transcription factors. *Elife*, *2*, e00726. https://doi.org/10.7554/eLife.00726

Huber, C., von Watzdorf, J., & Marx, A. (2016). 5-methylcytosine-sensitive variants of Thermococcus kodakaraensis DNA polymerase. *Nucleic Acids Res*, *44*(20), 9881-9890. https://doi.org/10.1093/nar/gkw812

Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., Bell, C., Bernabé, R. R.,...Consortium, I. C. G. (2010). International network of cancer genome projects. *Nature*, *464*(7291), 993-998. https://doi.org/10.1038/nature08987

Huisman, O., & D'Ari, R. (1981). An inducible DNA replication-cell division coupling mechanism in E. coli. *Nature*, *290*(5809), 797-799. https://doi.org/10.1038/290797a0

Hwang, Y., Kang, S. J., Kang, J., Choi, J., Kim, S. J., & Jang, S. (2025). DNA repair and disease: insights from the human DNA glycosylase NEIL family. *Exp Mol Med*, *57*(3), 524-532. https://doi.org/10.1038/s12276-025-01417-0

Irigaray, P., & Belpomme, D. (2009). Basic properties and molecular mechanisms of exogenous chemical carcinogens. *Carcinogenesis*, *31*(2), 135-148. https://doi.org/10.1093/carcin/bgp252

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129-1133. https://doi.org/10.1038/nature09303

Iyer, D. R., & Rhind, N. (2017). Replication fork slowing and stalling are distinct, checkpoint-independent consequences of replicating damaged DNA. *PLoS Genet*, *13*(8), e1006958. https://doi.org/10.1371/journal.pgen.1006958

Jackson, A. L., & Loeb, L. A. (2001). The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *477*(1), 7-21. https://doi.org/https://doi.org/10.1016/S0027-5107(01)00091-4

Jacobs, A. L., & Schär, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma*, *121*(1), 1-20. https://doi.org/10.1007/s00412-011-0347-4

Janulaitis, A., Klimasauskas, S., Petrusyte, M., & Butkus, V. (1983). Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. *FEBS Lett*, *161*(1), 131-134. https://doi.org/10.1016/0014-5793(83)80745-5

Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol*, *5*(11), a012740. https://doi.org/10.1101/cshperspect.a012740

Jeltsch, A. (2006). Molecular enzymology of mammalian DNA methyltransferases. *Curr Top Microbiol Immunol*, *301*, 203-225. https://doi.org/10.1007/3-540-31390-7_7

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S.,...Helm, M. (2017). Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biology*, *14*(9), 1108-1123. https://doi.org/10.1080/15476286.2016.1191737

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, *449*(7159), 248-251. https://doi.org/10.1038/nature06146

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y.,...He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol*, *7*(12), 885-887. https://doi.org/10.1038/nchembio.687

Jiang, Q., Karata, K., Woodgate, R., Cox, M. M., & Goodman, M. F. (2009). The active form of DNA polymerase V is UmuD'(2)C-RecA-ATP. *Nature*, *460*(7253), 359-363. https://doi.org/10.1038/nature08178

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, *13*(7), 484-492. https://doi.org/10.1038/nrg3230

Jones, P. A., & Takai, D. (2001). The role of DNA methylation in mammalian epigenetics. *Science*, *293*(5532), 1068-1070. https://doi.org/10.1126/science.1063852

Jurkowski, T. P., & Jeltsch, A. (2011). On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS One*, *6*(11), e28104. https://doi.org/10.1371/journal.pone.0028104

Kamiya, H., Tsuchiya, H., Karino, N., Ueno, Y., Matsuda, A., & Harashima, H. (2002). Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells. *J Biochem*, *132*(4), 551-555. https://doi.org/10.1093/oxfordjournals.jbchem.a003256

Kane, C. M., & Linn, S. (1981). Purification and characterization of an apurinic/apyrimidinic endonuclease from HeLa cells. *J Biol Chem*, *256*(7), 3405-3414.

Karymov, M. A., Tomschik, M., Leuba, S. H., Caiafa, P., & Zlatanova, J. (2001). DNA methylation-dependent chromatin fiber compaction in vivo and in vitro: requirement for linker histone. *FASEB J*, *15*(14), 2631-2641. https://doi.org/10.1096/fj.01-0345com

Kataoka, H., Yamamoto, Y., & Sekiguchi, M. (1983). A new gene (alkB) of Escherichia coli that controls sensitivity to methyl methane sulfonate. *J Bacteriol*, *153*(3), 1301-1307. https://doi.org/10.1128/jb.153.3.1301-1307.1983

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. In (pp. 772-780): Oxford University Press

Keller, T. E., Han, P., & Yi, S. V. (2016). Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate-Vertebrate Boundary. *Mol Biol Evol*, *33*(4), 1019-1028. https://doi.org/10.1093/molbev/msv345

Khodaverdian, V., Sano, T., Maggs, L. R., Tomarchio, G., Dias, A., Tran, M.,...McVey, M. (2024). REV1 coordinates a multi-faceted tolerance response to DNA alkylation damage and prevents chromosome shattering in Drosophila melanogaster. *PLoS Genet*, *20*(7), e1011181. https://doi.org/10.1371/journal.pgen.1011181

Kim, Y. J., & Wilson, D. M. (2012). Overview of base excision repair biochemistry. *Curr Mol Pharmacol*, *5*(1), 3-13. https://doi.org/10.2174/1874467211205010003

Kimura, H., & Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J Biol Chem*, *278*(7), 4806-4812. https://doi.org/10.1074/jbc.M209923200

Kishkevich, A., Tamang, S., Nguyen, M. O., Oehler, J., Bulmaga, E., Andreadis, C.,...Whitby, M. C. (2022). Rad52's DNA annealing activity drives template switching associated with restarted DNA replication. *Nat Commun*, *13*(1), 7293. https://doi.org/10.1038/s41467-022-35060-4

Klimasauskas, S., Kumar, S., Roberts, R. J., & Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, *76*(2), 357-369. https://doi.org/10.1016/0092-8674(94)90342-5

Konishi, N., Nakamura, M., Ishida, E., Shimada, K., Mitsui, E., Yoshikawa, R.,...Tsujikawa, K. (2005). High Expression of a New Marker PCA-1 in Human Prostate Carcinoma. *Clinical Cancer Research*, *11*(14), 5090-5097. https://doi.org/10.1158/1078-0432.ccr-05-0195

Kovaka, S., Hook, P. W., Jenike, K. M., Shivakumar, V., Morina, L. B., Razaghi, R.,...Schatz, M. C. (2025). Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nat Methods*, *22*(4), 681-691. https://doi.org/10.1038/s41592-025-02631-4

Kowalczykowski, S. C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb Perspect Biol*, *7*(11). https://doi.org/10.1101/cshperspect.a016410

Krwawicz, J., Sheeba, C., Hains, K., McMahon, T., Zhang, Y., Kriaucionis, S., & Sarkies, P. (2024). Introduction of cytosine-5 DNA methylation sensitizes cells to oxidative damage. *13*. https://doi.org/https://doi.org/10.7554/eLife.103432.3

Kumari, S., Sharma, S., Advani, D., Khosla, A., Kumar, P., & Ambasta, R. K. (2022). Unboxing the molecular modalities of mutagens in cancer. *Environmental Science and Pollution Research*, *29*(41), 62111-62159. https://doi.org/10.1007/s11356-021-16726-w

Kurepina, N., Chudaev, M., Kreiswirth, B. N., Nikiforov, V., & Mustaev, A. (2022). Mutations compensating for the fitness cost of rifampicin resistance in Escherichia coli exert pleiotropic effect on RNA polymerase catalysis. *Nucleic Acids Res*, *50*(10), 5739-5756. https://doi.org/10.1093/nar/gkac406

Lagage, V., Chen, V., & Uphoff, S. (2023). Adaptation delay causes a burst of mutations in bacteria responding to oxidative stress. *EMBO Rep*, *24*(1), e55640. https://doi.org/10.15252/embr.202255640

Landini, P., & Volkert, M. R. (1995). Transcriptional activation of the Escherichia coli adaptive response gene aidB is mediated by binding of methylated Ada protein. Evidence for a new consensus sequence for Ada-binding sites. *J Biol Chem*, *270*(14), 8285-8289. https://doi.org/10.1074/jbc.270.14.8285

Larson, K., Sahm, J., Shenkar, R., & Strauss, B. (1985). Methylation-induced blocks to in vitro DNA replication. *Mutat Res*, *150*(1-2), 77-84. https://doi.org/10.1016/0027-5107(85)90103-4

Lee, S., Kim, W. H., Jung, H. Y., Yang, M. H., & Kang, G. H. (2002). Aberrant CpG island methylation of multiple genes in intrahepatic cholangiocarcinoma. *Am J Pathol*, *161*(3), 1015-1022. https://doi.org/10.1016/S0002-9440(10)64262-9

Lehmann, A. R. (2006). Translesion synthesis in mammalian cells. *Exp Cell Res*, *312*(14), 2673-2676. https://doi.org/10.1016/j.yexcr.2006.06.010

Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S.,...Peters, A. H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol*, *13*(14), 1192-1200. https://doi.org/10.1016/s0960-9822(03)00432-9

Lemotte, P. K., & Walker, G. C. (1985). Induction and autoregulation of ada, a positively acting element regulating the response of Escherichia coli K-12 to methylating agents. *J Bacteriol*, *161*(3), 888-895. https://doi.org/10.1128/jb.161.3.888-895.1985

Leonhardt, H., Page, A. W., Weier, H. U., & Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, *71*(5), 865-873. https://doi.org/10.1016/0092-8674(92)90561-p

Lewis, S. H., Ross, L., Bain, S. A., Pahita, E., Smith, S. A., Cordaux, R.,...Sarkies, P. (2020). Widespread conservation and lineage-specific diversification of genome-wide DNA methylation patterns across arthropods. *PLoS Genet*, *16*(6), e1008864. https://doi.org/10.1371/journal.pgen.1008864

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature*, *366*(6453), 362-365. https://doi.org/10.1038/366362a0

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915-926. https://doi.org/10.1016/0092-8674(92)90611-f

Li, E., & Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harb Perspect Biol*, *6*(5), a019133. https://doi.org/10.1101/cshperspect.a019133

Li, M. M., Nilsen, A., Shi, Y., Fusser, M., Ding, Y. H., Fu, Y.,...Yang, Y. G. (2013). ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nat Commun*, *4*, 1832. https://doi.org/10.1038/ncomms2863

Li, X., & Heyer, W. D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res*, *18*(1), 99-113. https://doi.org/10.1038/cr.2008.1

Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K., & Yuan, H. S. (2020). Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res*, *48*(7), 3949-3961. https://doi.org/10.1093/nar/gkaa111

Lin, C. H., Hsieh, S. Y., Sheen, I. S., Lee, W. C., Chen, T. C., Shyu, W. C., & Liaw, Y. F. (2001). Genome-wide hypomethylation in hepatocellular carcinogenesis. *Cancer Res*, *61*(10), 4238-4243.

Lindahl, T. (1974). An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc Natl Acad Sci U S A*, *71*(9), 3649-3653. https://doi.org/10.1073/pnas.71.9.3649

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*(6422), 709-715. https://doi.org/10.1038/362709a0

Lindahl, T. (1996). The Croonian Lecture, 1996: endogenous damage to DNA. *Philos Trans R Soc Lond B Biol Sci*, *351*(1347), 1529-1538. https://doi.org/10.1098/rstb.1996.0139

Lindahl, T., & Wood, R. D. (1999). Quality control by DNA repair. *Science*, *286*(5446), 1897-1905. https://doi.org/10.1126/science.286.5446.1897

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J.,...Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315-322. https://doi.org/10.1038/nature08514

Little, J. W. (1984). Autodigestion of lexA and phage lambda repressors. *Proc Natl Acad Sci U S A*, *81*(5), 1375-1379. https://doi.org/10.1073/pnas.81.5.1375

Liu, M. H., Costa, B. M., Bianchini, E. C., Choi, U., Bandler, R. C., Lassen, E.,...Evrony, G. D. (2024). DNA mismatch and damage patterns revealed by single-molecule sequencing. *Nature*, *630*(8017), 752-761. https://doi.org/10.1038/s41586-024-07532-8

Liu, Q., Fang, L., Yu, G., Wang, D., Xiao, C. L., & Wang, K. (2019). Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat Commun*, *10*(1), 2449. https://doi.org/10.1038/s41467-019-10168-2

Loomis, W. F., & Magasanik, B. (1967). Glucose-lactose diauxie in Escherichia coli. *J Bacteriol*, *93*(4), 1397-1401. https://doi.org/10.1128/jb.93.4.1397-1401.1967

Lopes, M., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M.,...Foiani, M. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. *Nature*, *412*(6846), 557-561. https://doi.org/10.1038/35087613

Lopez-Serra, L., Ballestar, E., Fraga, M. F., Alaminos, M., Setien, F., & Esteller, M. (2006). A profile of methyl-CpG binding domain protein occupancy of hypermethylated promoter CpG islands of tumor suppressor genes in human cancer. *Cancer Res*, *66*(17), 8342-8346. https://doi.org/10.1158/0008-5472.CAN-06-1932

Loughery, J. E., Dunne, P. D., O'Neill, K. M., Meehan, R. R., McDaid, J. R., & Walsh, C. P. (2011). DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response. *Hum Mol Genet*, *20*(16), 3241-3255. https://doi.org/10.1093/hmg/ddr236

Lovett, S. T. (2017). Template-switching during replication fork repair in bacteria. *DNA Repair (Amst)*, *56*, 118-128. https://doi.org/10.1016/j.dnarep.2017.06.014

Lu, R., Nash, H. M., & Verdine, G. L. (1997). A mammalian DNA repair enzyme that excises oxidatively damaged guanines maps to a locus frequently lost in lung cancer. *Curr Biol*, *7*(6), 397-407. https://doi.org/10.1016/s0960-9822(06)00187-4

Luria, S. E., & Delbrück, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*, *28*(6), 491-511. https://doi.org/10.1093/genetics/28.6.491

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., & Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol*, *8*(11), e1000506. https://doi.org/10.1371/journal.pbio.1000506

Madugundu, G. S., Cadet, J., & Wagner, J. R. (2014). Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA. *Nucleic Acids Res*, *42*(11), 7450-7460. https://doi.org/10.1093/nar/gku334

Magewu, A. N., & Jones, P. A. (1994). Ubiquitous and Tenacious Methylation of the CpG Site in Codon 248 of the p53 Gene May Explain Its Frequent Appearance as a Mutational Hot Spot in Human Cancer. *Molecular and Cellular Biology*, *14*(6), 4225-4232. https://doi.org/10.1128/mcb.14.6.4225-4232.1994

Maiti, A., Noon, M. S., MacKerell, A. D., Pozharski, E., & Drohat, A. C. (2012). Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc Natl Acad Sci U S A*, *109*(21), 8091-8096. https://doi.org/10.1073/pnas.1201010109

Malan, T. P., Kolb, A., Buc, H., & McClure, W. R. (1984). Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. *J Mol Biol*, *180*(4), 881-909. https://doi.org/10.1016/0022-2836(84)90262-6

Markkanen, E., Fischer, R., Ledentcova, M., Kessler, B. M., & Dianov, G. L. (2015). Cells deficient in base-excision repair reveal cancer hallmarks originating from adjustments to genetic instability. *Nucleic Acids Research*, *43*(7), 3667-3679. https://doi.org/10.1093/nar/gkv222

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M.,...Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature*, *399*(6737), 700-704. https://doi.org/10.1038/21447

May, M. S., & Hattman, S. (1975). Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *J Bacteriol*, *123*(2), 768-770. https://doi.org/10.1128/jb.123.2.768-770.1975

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A*, *36*(6), 344-355. https://doi.org/10.1073/pnas.36.6.344

McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*, *18*(1), 148-161. https://doi.org/10.1038/cr.2008.4

Mehta, K. P. M., Lovejoy, C. A., Zhao, R., Heintzman, D. R., & Cortez, D. (2020). HMCES Maintains Replication Fork Progression and Prevents Double-Strand Breaks in Response to APOBEC Deamination and Abasic Site Formation. *Cell Rep*, *31*(9), 107705. https://doi.org/10.1016/j.celrep.2020.107705

Mellon, I. (2005). Transcription-coupled repair: a complex affair. *Mutat Res*, *577*(1-2), 155-161. https://doi.org/10.1016/j.mrfmmm.2005.03.016

Mitra, S. (2001). DNA glycosylases: specificity and mechanisms. *Prog Nucleic Acid Res Mol Biol*, *68*, 189-192. https://doi.org/10.1016/s0079-6603(01)68099-1

Mizusawa, S., & Gottesman, S. (1983). Protein degradation in Escherichia coli: the lon gene controls the stability of sulA protein. *Proc Natl Acad Sci U S A*, *80*(2), 358-362. https://doi.org/10.1073/pnas.80.2.358

Molodtsov, V., Scharf, N. T., Stefan, M. A., Garcia, G. A., & Murakami, K. S. (2017). Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in Mycobacterium tuberculosis. *Mol Microbiol*, *103*(6), 1034-1045. https://doi.org/10.1111/mmi.13606

Monier, K., Mouradian, S., & Sullivan, K. F. (2007). DNA methylation promotes Aurora-B-driven phosphorylation of histone H3 in chromosomal subdomains. *J Cell Sci*, *120*(Pt 1), 101-114. https://doi.org/10.1242/jcs.03326

Moriya, M., & Grollman, A. P. (1993). Mutations in the mutY gene of Escherichia coli enhance the frequency of targeted G:C-->T:a transversions induced by a single 8-oxoguanine residue in single-stranded DNA. *Mol Gen Genet*, *239*(1-2), 72-76. https://doi.org/10.1007/BF00281603

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjørås, M., & Seeberg, E. (2002). Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res*, *30*(22), 4926-4936. https://doi.org/10.1093/nar/gkf618

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M. C., & Leonhardt, H. (2005). Recruitment of DNA methyltransferase I to DNA repair sites. *Proc Natl Acad Sci U S A*, *102*(25), 8905-8909. https://doi.org/10.1073/pnas.0501034102

Muggerud, A. A., Rønneberg, J. A., Wärnberg, F., Botling, J., Busato, F., Jovanovic, J.,...Tost, J. (2010). Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. *Breast Cancer Res*, *12*(1), R3. https://doi.org/10.1186/bcr2466

Mukai, T. (1964). The genetic structure of natural populations of drosophila melanogaster. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics*, *50*(1), 1-19. https://doi.org/10.1093/genetics/50.1.1

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, *102*(5), 553-563. https://doi.org/10.1016/s0092-8674(00)00078-7

Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T.,...Kohli, R. M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol*, *8*(9), 751-758. https://doi.org/10.1038/nchembio.1042

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, *393*(6683), 386-389. https://doi.org/10.1038/30764

Napolitano, R., Janel-Bintz, R., Wagner, J., & Fuchs, R. P. (2000). All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced

mutagenesis. *EMBO J*, *19*(22), 6259-6265. https://doi.org/10.1093/emboj/19.22.6259

Nash, H. M., Bruner, S. D., Schärer, O. D., Kawate, T., Addona, T. A., Spooner, E.,...Verdine, G. L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr Biol*, *6*(8), 968-980. https://doi.org/10.1016/s0960-9822(02)00641-3

Nieminuszczy, J., Mielecki, D., Sikora, A., Wrzesiński, M., Chojnacka, A., Krwawicz, J.,...Grzesiuk, E. (2009). Mutagenic potency of MMS-induced 1meA/3meC lesions in E. coli. *Environ Mol Mutagen*, *50*(9), 791-799. https://doi.org/10.1002/em.20497

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K.,...Consortium, B. C. W. G. o. t. I. C. G. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell*, *149*(5), 979-993. https://doi.org/10.1016/j.cell.2012.04.024

O'Brien, P. J., & Ellenberger, T. (2004). The Escherichia coli 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J Biol Chem*, *279*(26), 26876-26884. https://doi.org/10.1074/jbc.M403860200

O'Connor, T. R., & Laval, F. (1990). Isolation and structure of a cDNA expressing a mammalian 3-methyladenine-DNA glycosylase. *EMBO J*, *9*(10), 3337-3342. https://doi.org/10.1002/j.1460-2075.1990.tb07534.x

O'Connor, T. R., & Laval, J. (1989). Physical association of the 2,6-diamino-4-hydroxy-5N-formamidopyrimidine-DNA glycosylase of Escherichia coli and an activity nicking DNA at apurinic/apyrimidinic sites. *Proc Natl Acad Sci U S A*, *86*(14), 5222-5226. https://doi.org/10.1073/pnas.86.14.5222

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247-257. https://doi.org/10.1016/s0092-8674(00)81656-6

Olsson, M., & Lindahl, T. (1980). Repair of alkylated DNA in Escherichia coli. Methyl group transfer from O6-methylguanine to a protein cysteine residue. *J Biol Chem*, *255*(22), 10569-10571.

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z.,...Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*(7154), 714-717. https://doi.org/10.1038/nature05987

Page, D. R., Köhler, C., Da Costa-Nunes, J. A., Baroux, C., Moore, J. M., & Grossniklaus, U. (2004). Intrachromosomal excision of a hybrid Ds element induces large genomic deletions in Arabidopsis. *Proc Natl Acad Sci U S A*, *101*(9), 2969-2974. https://doi.org/10.1073/pnas.0400089101

Pagès-Gallego, M., van Soest, D. M. K., Besselink, N. J. M., Straver, R., Keijer, J. P., Vermeulen, C.,...de Ridder, J. (2025). Direct detection of 8-oxo-dG using nanopore sequencing. *Nat Commun*, *16*(1), 5236. https://doi.org/10.1038/s41467-025-60391-3

Patel, M., Jiang, Q., Woodgate, R., Cox, M. M., & Goodman, M. F. (2010). A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V. *Crit Rev Biochem Mol Biol*, *45*(3), 171-184. https://doi.org/10.3109/10409238.2010.480968

Paz, M. F., Fraga, M. F., Avila, S., Guo, M., Pollan, M., Herman, J. G., & Esteller, M. (2003). A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res*, *63*(5), 1114-1121.

Pegg, A. E., Dolan, M. E., & Moschel, R. C. (1995). Structure, function, and inhibition of O6-alkylguanine-DNA alkyltransferase. *Progress in nucleic acid research and molecular biology*, *51*, 167-223.

Peltomäki, P. (2012). Mutations and epimutations in the origin of cancer. *Experimental Cell Research*, *318*(4), 299-310. https://doi.org/https://doi.org/10.1016/j.yexcr.2011.12.001

Petrova, V., Chitteni-Pattu, S., Drees, J. C., Inman, R. B., & Cox, M. M. (2009). An SOS inhibitor that binds to free RecA protein: the PsiB protein. *Mol Cell*, *36*(1), 121-130. https://doi.org/10.1016/j.molcel.2009.07.026

Peña-Diaz, J., Bregenhorn, S., Ghodgaonkar, M., Follonier, C., Artola-Borán, M., Castor, D.,...Jiricny, J. (2012). Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol Cell*, *47*(5), 669-680. https://doi.org/10.1016/j.molcel.2012.07.006

Pfeifer, G. P. (2006). Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol*, *301*, 259-281. https://doi.org/10.1007/3-540-31390-7_10

Pfeifer, G. P. (2010). Environmental exposures and mutational patterns of cancer genomes. *Genome Med*, *2*(8), 54. https://doi.org/10.1186/gm175

Pignatello, J. J., Oliveros, E., & MacKay, A. (2006). Advanced Oxidation Processes for Organic Contaminant Destruction Based on the Fenton Reaction and Related Chemistry. *Critical Reviews in Environmental Science and Technology*, *36*(1), 1-84. https://doi.org/10.1080/10643380500326564

Pilžys, T., Marcinkowski, M., Kukwa, W., Garbicz, D., Dylewska, M., Ferenc, K.,...Wołosz, D. (2019). ALKBH overexpression in head and neck cancer: potential target for novel anticancer therapy. *Scientific reports*, *9*(1), 13249.

Polayes, D. A., Rice, P. W., Garner, M. M., & Dahlberg, J. E. (1988). Cyclic AMP-cyclic AMP receptor protein as a repressor of transcription of the spf gene of Escherichia coli. *J Bacteriol*, *170*(7), 3110-3114. https://doi.org/10.1128/jb.170.7.3110-3114.1988

Ponger, L., & Li, W. H. (2005). Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol*, *22*(4), 1119-1128. https://doi.org/10.1093/molbev/msi098

Pope, C. F., O'Sullivan, D. M., McHugh, T. D., & Gillespie, S. H. (2008). A practical guide to measuring mutation rates in antibiotic resistance. *Antimicrob Agents Chemother*, *52*(4), 1209-1214. https://doi.org/10.1128/AAC.01152-07

Pradhan, S., Bacolla, A., Wells, R. D., & Roberts, R. J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem*, *274*(46), 33002-33010. https://doi.org/10.1074/jbc.274.46.33002

Putiri, E. L., & Robertson, K. D. (2011). Epigenetic mechanisms and genome stability. *Clin Epigenetics*, *2*(2), 299-314. https://doi.org/10.1007/s13148-010-0017-z

Pósfai, J., Bhagwat, A. S., Pósfai, G., & Roberts, R. J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res*, *17*(7), 2421-2435. https://doi.org/10.1093/nar/17.7.2421

Qian, X. C., & Brent, T. P. (1997). Methylation hot spots in the 5' flanking region denote silencing of the O 6-methylguanine-DNA methyltransferase gene. *Cancer research*, *57*(17), 3672-3677.

Raleigh, E. A., & Wilson, G. (1986). Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc Natl Acad Sci U S A*, *83*(23), 9070-9074. https://doi.org/10.1073/pnas.83.23.9070

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods*, *14*(4), 411-413. https://doi.org/10.1038/nmeth.4189

Rangam, G., Schmitz, K. M., Cobb, A. J., & Petersen-Mahrt, S. K. (2012). AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One*, *7*(8), e43279. https://doi.org/10.1371/journal.pone.0043279

Rauch, T. A., Zhong, X., Wu, X., Wang, M., Kernstine, K. H., Wang, Z.,...Pfeifer, G. P. (2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc Natl Acad Sci U S A*, *105*(1), 252-257. https://doi.org/10.1073/pnas.0710735105

Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G. G., Rottem, S., & Razin, A. (1990). Cloning, characterization, and expression in Escherichia coli of the gene coding for the CpG DNA methylase from Spiroplasma sp. strain MQ1(M.SssI). *Nucleic Acids Res*, *18*(5), 1145-1152. https://doi.org/10.1093/nar/18.5.1145

Reuven, N. B., Arad, G., Maor-Shoshani, A., & Livneh, Z. (1999). The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication. *J Biol Chem*, *274*(45), 31763-31766. https://doi.org/10.1074/jbc.274.45.31763

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*, *13*(5), 278-289. https://doi.org/10.1016/j.gpb.2015.08.002

Rideout, W. M., Coetzee, G. A., Olumi, A. F., & Jones, P. A. (1990). 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science*, *249*(4974), 1288-1290. https://doi.org/10.1126/science.1697983

Robbens, S., Rouzé, P., Cock, J. M., Spring, J., Worden, A. Z., & Van de Peer, Y. (2008). The FTO gene, implicated in human obesity, is found only in vertebrates and marine algae. *J Mol Evol*, *66*(1), 80-84. https://doi.org/10.1007/s00239-007-9059-z

Robson, C. N., & Hickson, I. D. (1991). Isolation of cDNA clones encoding a human apurinic/apyrimidinic endonuclease that corrects DNA repair and mutagenesis defects in E. coli xth (exonuclease III) mutants. *Nucleic Acids Res*, *19*(20), 5519-5523. https://doi.org/10.1093/nar/19.20.5519

Rosche, W. A., & Foster, P. L. (2000). Determining mutation rates in bacterial populations. *Methods*, *20*(1), 4-17. https://doi.org/10.1006/meth.1999.0901

Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., Emperle, M., Beltran, T.,...Sarkies, P. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nat Genet*, *50*(3), 452-459. https://doi.org/10.1038/s41588-018-0061-8

Ruzov, A., Shorning, B., Mortusewicz, O., Dunican, D. S., Leonhardt, H., & Meehan, R. R. (2009). MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development*, *136*(13), 2277-2286. https://doi.org/10.1242/dev.032227

Samet, J. M., & Wages, P. A. (2018). Oxidative Stress from Environmental Exposures. *Curr Opin Toxicol*, *7*, 60-66.

Samson, L., & Cairns, J. (1977). A new pathway for DNA repair in Escherichia coli. *Nature*, *267*(5608), 281-283. https://doi.org/10.1038/267281a0

San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem*, *77*, 229-257. https://doi.org/10.1146/annurev.biochem.77.061306.125255

Sarkies, P. (2022). Encyclopaedia of eukaryotic DNA methylation: from patterns to mechanisms and functions. *Biochem Soc Trans*, *50*(3), 1179-1190. https://doi.org/10.1042/BST20210725

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A*, *103*(5), 1412-1417. https://doi.org/10.1073/pnas.0510310103

Scarano, C., Veneruso, I., De Simone, R. R., Di Bonito, G., Secondino, A., & D'Argenio, V. (2024). The Third-Generation Sequencing Challenge: Novel Insights for the Omic Sciences. *Biomolecules*, *14*(5). https://doi.org/10.3390/biom14050568

Schlacher, K., Pham, P., Cox, M. M., & Goodman, M. F. (2006). Roles of DNA polymerase V and RecA protein in SOS damage-induced mutation. *Chem Rev*, *106*(2), 406-419. https://doi.org/10.1021/cr0404951

Schleif, R. (2000). Regulation of the L-arabinose operon of Escherichia coli. *Trends Genet*, *16*(12), 559-565. https://doi.org/10.1016/s0168-9525(00)02153-3

Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol Rev*, *34*(5), 779-796. https://doi.org/10.1111/j.1574-6976.2010.00226.x

Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D.,...Ecker, J. R. (2016). Corrigendum: Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, *530*(7589), 242. https://doi.org/10.1038/nature16179

Schulz, W. A., Steinhoff, C., & Florl, A. R. (2006). Methylation of endogenous human retroelements in health and disease. *Curr Top Microbiol Immunol*, *310*, 211-250. https://doi.org/10.1007/3-540-31181-5_11

Sedgwick, B. (2004). Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol*, *5*(2), 148-157. https://doi.org/10.1038/nrm1312

Sedletska, Y., Radicella, J. P., & Sage, E. (2013). Replication fork collapse is a major cause of the high mutation frequency at three-base lesion clusters. *Nucleic Acids Res*, *41*(20), 9339-9348. https://doi.org/10.1093/nar/gkt731

Seiler, C. L., Fernandez, J., Koerperich, Z., Andersen, M. P., Kotandeniya, D., Nguyen, M. E.,...Tretyakova, N. Y. (2018). Maintenance DNA Methyltransferase Activity in the Presence of Oxidized Forms of 5-Methylcytosine: Structural Basis for Ten Eleven Translocation-Mediated DNA Demethylation. *Biochemistry*, *57*(42), 6061-6069. https://doi.org/10.1021/acs.biochem.8b00683

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., & Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol*, *278*(2), 440-458. https://doi.org/10.1016/j.ydbio.2004.11.025

Selby, C. P., Lindsey-Boltz, L. A., Li, W., & Sancar, A. (2023). Molecular Mechanisms of Transcription-Coupled Repair. *Annu Rev Biochem*, *92*, 115-144. https://doi.org/10.1146/annurev-biochem-041522-034232

Shaw, R. G., Byers, D. L., & Darmo, E. (2000). Spontaneous mutational effects on reproductive traits of arabidopsis thaliana. *Genetics*, *155*(1), 369-378. https://doi.org/10.1093/genetics/155.1.369

Shen, J. C., Rideout, W. M., & Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res*, *22*(6), 972-976. https://doi.org/10.1093/nar/22.6.972

Shibutani, S., Takeshita, M., & Grollman, A. P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*, *349*(6308), 431-434. https://doi.org/10.1038/349431a0

Shinagawa, H., Iwasaki, H., Kato, T., & Nakata, A. (1988). RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc Natl Acad Sci U S A*, *85*(6), 1806-1810. https://doi.org/10.1073/pnas.85.6.1806

Sies, H., Berndt, C., & Jones, D. P. (2017). Oxidative Stress. *Annu Rev Biochem*, *86*, 715-748. https://doi.org/10.1146/annurev-biochem-061516-045037

Singer, B., & Kuśmierek, J. T. (1982). Chemical mutagenesis. *Annu Rev Biochem*, *51*, 655-693. https://doi.org/10.1146/annurev.bi.51.070182.003255

Smal, M. P., Kuzhir, T. D., Savina, N. V., Nikitchenko, N. V., Rolevich, A. I., Krasny, S. A., & Goncharova, R. I. (2018). BER gene polymorphisms associated with key molecular events in bladder cancer. *Experimental Oncology*, *40*(4), 288-298.

Smith, K. C., & Wang, T. C. (1989). recA-dependent DNA repair processes. *Bioessays*, *10*(1), 12-16. https://doi.org/10.1002/bies.950100104

Sobol, R. W., Horton, J. K., Kühn, R., Gu, H., Singhal, R. K., Prasad, R.,...Wilson, S. H. (1996). Requirement of mammalian DNA polymerase-beta in base-excision repair. *Nature*, *379*(6561), 183-186. https://doi.org/10.1038/379183a0

Sommer, S., Boudsocq, F., Devoret, R., & Bailone, A. (1998). Specific RecA amino acid changes affect RecA-UmuD'C interaction. *Mol Microbiol*, *28*(2), 281-291. https://doi.org/10.1046/j.1365-2958.1998.00803.x

Steenken, S., & Jovanovic, S. V. (1997). How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *Journal of the American Chemical Society*, *119*(3), 617-618. https://doi.org/10.1021/ja962255b

Stefansson, O. A., Hermanowicz, S., van der Horst, J., Hilmarsdottir, H., Staszczak, Z., Jonasson, J. G.,...Sigurdsson, S. (2017). CpG promoter methylation of the ALKBH3 alkylation repair gene in breast cancer. *BMC Cancer*, *17*(1), 469. https://doi.org/10.1186/s12885-017-3453-8

Stoiber, M., Quick, J., Egan, R., Eun Lee, J., Celniker, S., Neely, R. K.,...Brown, J. (2017). De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv*, 094672. https://doi.org/10.1101/094672

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719-724. https://doi.org/10.1038/nature07943

Suh, D., Wilson, D. M., & Povirk, L. F. (1997). 3'-phosphodiesterase activity of human apurinic/apyrimidinic endonuclease at DNA double-strand break ends. *Nucleic Acids Res*, *25*(12), 2495-2500. https://doi.org/10.1093/nar/25.12.2495

Sukackaite, R., Grazulis, S., Tamulaitis, G., & Siksnys, V. (2012). The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res*, *40*(15), 7552-7562. https://doi.org/10.1093/nar/gks332

Suter, C. M., Martin, D. I., & Ward, R. L. (2004). Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet*, *36*(5), 497-501. https://doi.org/10.1038/ng1342

Tagami, H., & Aiba, H. (1998). A common role of CRP in transcription activation: CRP acts transiently to stimulate events leading to open complex formation at a diverse set of promoters. *EMBO J*, *17*(6), 1759-1767. https://doi.org/10.1093/emboj/17.6.1759

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y.,...Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, *324*(5929), 930-935. https://doi.org/10.1126/science.1170116

Tang, M., Bruck, I., Eritja, R., Turner, J., Frank, E. G., Woodgate, R.,...Goodman, M. F. (1998). Biochemical basis of SOS-induced mutagenesis in Escherichia coli: reconstitution of in vitro lesion bypass dependent on the UmuD'2C mutagenic complex and RecA protein. *Proc Natl Acad Sci U S A*, *95*(17), 9755-9760. https://doi.org/10.1073/pnas.95.17.9755

Tang, M., Pham, P., Shen, X., Taylor, J. S., O'Donnell, M., Woodgate, R., & Goodman, M. F. (2000). Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature*, *404*(6781), 1014-1018. https://doi.org/10.1038/35010020

Tarantino, M. E., Dow, B. J., Drohat, A. C., & Delaney, S. (2018). Nucleosomes and the three glycosylases: High, medium, and low levels of excision by the uracil DNA glycosylase superfamily. *DNA Repair (Amst)*, *72*, 56-63. https://doi.org/10.1016/j.dnarep.2018.09.008

Tasaki, M., Shimada, K., Kimura, H., Tsujikawa, K., & Konishi, N. (2011). ALKBH3, a human AlkB homologue, contributes to cell survival in human non-small-cell lung cancer. *British Journal of Cancer*, *104*(4), 700-706. https://doi.org/10.1038/sj.bjc.6606012

Tazi, J., & Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell*, *60*(6), 909-920. https://doi.org/10.1016/0092-8674(90)90339-g

Tchou, J., Kasai, H., Shibutani, S., Chung, M. H., Laval, J., Grollman, A. P., & Nishimura, S. (1991). 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proc Natl Acad Sci U S A*, *88*(11), 4690-4694. https://doi.org/10.1073/pnas.88.11.4690

Tercero, J. A., & Diffley, J. F. (2001). Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature*, *412*(6846), 553-557. https://doi.org/10.1038/35087607

TeSlaa, T., Ralser, M., Fan, J., & Rabinowitz, J. D. (2023). The pentose phosphate pathway in health and disease. *Nat Metab*, *5*(8), 1275-1289. https://doi.org/10.1038/s42255-023-00863-2

ThermoFisherScientific. (2025). *One Shot™ TOP10 Chemically Competent E. coli*. https://www.thermofisher.com/order/catalog/product/C404010

Thomason, L. C., Costantino, N., & Court, D. L. (2007). E. coli genome manipulation by P1 transduction. *Curr Protoc Mol Biol*, *Chapter 1*, 1.17.11-11.17.18. https://doi.org/10.1002/0471142727.mb0117s79

Thompson, P. S., & Cortez, D. (2020). New insights into abasic site repair and tolerance. *DNA Repair (Amst)*, *90*, 102866. https://doi.org/10.1016/j.dnarep.2020.102866

Tissier, A., Frank, E. G., McDonald, J. P., Iwai, S., Hanaoka, F., & Woodgate, R. (2000). Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase iota. *EMBO J*, *19*(19), 5259-5266. https://doi.org/10.1093/emboj/19.19.5259

Tomkinson, A. E., Vijayakumar, S., Pascal, J. M., & Ellenberger, T. (2006). DNA ligases: structure, reaction mechanism, and function. *Chem Rev*, *106*(2), 687-699. https://doi.org/10.1021/cr040498d

Tomkova, M., McClellan, M., Kriaucionis, S., & Schuster-Böckler, B. (2018). DNA Replication and associated repair pathways are involved in the mutagenesis of methylated cytosine. *DNA Repair (Amst)*, *62*, 1-7. https://doi.org/10.1016/j.dnarep.2017.11.005

Tomkova, M., McClellan, M. J., Crevel, G., Shahid, A. M., Mozumdar, N., Tomek, J.,...Kriaucionis, S. (2024). Human DNA polymerase ε is a source of C>T mutations at CpG dinucleotides. *Nat Genet*, *56*(11), 2506-2516. https://doi.org/10.1038/s41588-024-01945-x

Tornaletti, S., & Pfeifer, G. P. (1995). Complete and tissue-independent methylation of CpG sites in the p53 gene: implications for mutations in human cancers. *Oncogene*, *10*(8), 1493-1499.

Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., & Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature*, *419*(6903), 174-178. https://doi.org/10.1038/nature00908

Tubbs, A., & Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*, *168*(4), 644-656. https://doi.org/10.1016/j.cell.2017.01.002

Tóth, D. M., Szeri, F., Ashaber, M., Muazu, M., Székvölgyi, L., & Arányi, T. (2025). Tissue-specific roles of de novo DNA methyltransferases. *Epigenetics Chromatin*, *18*(1), 5. https://doi.org/10.1186/s13072-024-00566-2

Valinluck, V., & Sowers, L. C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res*, *67*(3), 946-950. https://doi.org/10.1158/0008-5472.CAN-06-3123

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet*, *34*(9), 666-681. https://doi.org/10.1016/j.tig.2018.05.008

van Gent, D. C., & van der Burg, M. (2007). Non-homologous end-joining, a sticky affair. *Oncogene*, *26*(56), 7731-7740. https://doi.org/10.1038/sj.onc.1210871

Visnes, T., Doseth, B., Pettersen, H. S., Hagen, L., Sousa, M. M., Akbari, M.,...Krokan, H. E. (2009). Uracil in DNA and its processing by different DNA glycosylases. *Philos Trans R Soc Lond B Biol Sci*, *364*(1517), 563-568. https://doi.org/10.1098/rstb.2008.0186

Volkova, N. V., Meier, B., González-Huici, V., Bertolini, S., Gonzalez, S., Vöhringer, H.,...Gerstung, M. (2020). Mutational signatures are jointly shaped by DNA damage and repair. *Nat Commun*, *11*(1), 2169. https://doi.org/10.1038/s41467-020-15912-7

Waddington, C. H. (2012). The epigenotype. 1942. *Int J Epidemiol*, *41*(1), 10-13. https://doi.org/10.1093/ije/dyr184

Walker, G. C. (1984). Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli. *Microbiol Rev*, *48*(1), 60-93. https://doi.org/10.1128/mr.48.1.60-93.1984

Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet*, *20*(2), 116-117. https://doi.org/10.1038/2413

Wang, Y., Branicky, R., Noë, A., & Hekimi, S. (2018). Superoxide dismutases: Dual roles in controlling ROS damage and regulating ROS signaling. *J Cell Biol*, *217*(6), 1915-1928. https://doi.org/10.1083/jcb.201708007

Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M.,...Robinson, G. E. (2006). Functional CpG methylation system in a social insect. *Science*, *314*(5799), 645-647. https://doi.org/10.1126/science.1135213

Wang, Y., Zhao, Y., Bollas, A., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*, *39*(11), 1348-1365. https://doi.org/10.1038/s41587-021-01108-x

Waterland, R. A., & Jirtle, R. L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol*, *23*(15), 5293-5300. https://doi.org/10.1128/MCB.23.15.5293-5300.2003

Watson, J. D., & Crick, F. H. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, *171*(4361), 964-967. https://doi.org/10.1038/171964b0

Watts, G. S., Futscher, B. W., Holtan, N., Degeest, K., Domann, F. E., & Rose, S. L. (2008). DNA methylation changes in ovarian cancer are cumulative with disease progression and identify tumor stage. *BMC Med Genomics*, *1*, 47. https://doi.org/10.1186/1755-8794-1-47

Watts, G. S., Pieper, R. O., Costello, J. F., Peng, Y.-M., Dalton, W. S., & Futscher, B. W. (1997). Methylation of discrete regions of the O 6-methylguanine DNA methyltransferase (MGMT) CpG island is associated with heterochromatinization of the MGMT transcription start site and silencing of the gene. *Molecular and cellular biology*, *17*(9), 5612-5619.

Westbye, M. P., Feyzi, E., Aas, P. A., Vågbø, C. B., Talstad, V. A., Kavli, B.,...Krokan, H. E. (2008). Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J Biol Chem*, *283*(36), 25046-25056. https://doi.org/10.1074/jbc.M803776200

Weterings, E., & Chen, D. J. (2008). The endless tale of non-homologous end-joining. *Cell Res*, *18*(1), 114-124. https://doi.org/10.1038/cr.2008.3

Wild, L., & Flanagan, J. M. (2010). Genome-wide hypomethylation in cancer may be a passive consequence of transformation. *Biochim Biophys Acta*, *1806*(1), 50-57. https://doi.org/10.1016/j.bbcan.2010.03.003

Wilson, A. S., Power, B. E., & Molloy, P. L. (2007). DNA hypomethylation and human diseases. *Biochim Biophys Acta*, *1775*(1), 138-162. https://doi.org/10.1016/j.bbcan.2006.08.007

Wilson, D. M. (2007). Processing of nonconventional DNA strand break ends. *Environ Mol Mutagen*, *48*(9), 772-782. https://doi.org/10.1002/em.20346

Witkin, E. M. (1976). Ultraviolet mutagenesis and inducible DNA repair in Escherichia coli. *Bacteriol Rev*, *40*(4), 869-907. https://doi.org/10.1128/br.40.4.869-907.1976

Wojciechowski, M., Czapinska, H., & Bochtler, M. (2013). CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.MpeI. *Proc Natl Acad Sci U S A*, *110*(1), 105-110. https://doi.org/10.1073/pnas.1207986110

Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., & Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature*, *389*(6652), 745-749. https://doi.org/10.1038/39631

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W.,...Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, *236*(1), 87-95. https://doi.org/10.1016/s0378-1119(99)00252-8

Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C. J., & Min, J. (2014). Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single-stranded N6-methyladenosine RNA demethylation. *J Biol Chem*, *289*(25), 17299-17311. https://doi.org/10.1074/jbc.M114.550350

Xu, L., & Seki, M. (2020). Recent advances in the detection of base modifications using the Nanopore sequencer. *J Hum Genet*, *65*(1), 25-33. https://doi.org/10.1038/s10038-019-0679-0

Yamanaka, K., Zheng, W., Crooke, E., Wang, Y. H., & Inouye, M. (2001). CspD, a novel DNA replication inhibitor induced during the stationary phase in Escherichia coli. *Mol Microbiol*, *39*(6), 1572-1584. https://doi.org/10.1046/j.1365-2958.2001.02345.x

Yang, C. G., Yi, C., Duguid, E. M., Sullivan, C. T., Jian, X., Rice, P. A., & He, C. (2008). Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, *452*(7190), 961-965. https://doi.org/10.1038/nature06889

Yang, Y., Gordenin, D. A., & Resnick, M. A. (2010). A single-strand specific lesion drives MMS-induced hyper-mutability at a double-strand break in yeast. *DNA Repair (Amst)*, *9*(8), 914-921. https://doi.org/10.1016/j.dnarep.2010.06.005

Yu, B., Edstrom, W. C., Benach, J., Hamuro, Y., Weber, P. C., Gibney, B. R., & Hunt, J. F. (2006). Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, *439*(7078), 879-884. https://doi.org/10.1038/nature04561

Yu, B., & Hunt, J. F. (2009). Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc Natl Acad Sci U S A*, *106*(34), 14315-14320. https://doi.org/10.1073/pnas.0812938106

Yu, C., Zhang, J., & Peterson, T. (2011). Genome rearrangements in maize induced by alternative transposition of reversed ac/ds termini. *Genetics*, *188*(1), 59-67. https://doi.org/10.1534/genetics.111.126847

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., & Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*, *97*(11), 5978-5983. https://doi.org/10.1073/pnas.100127597

Zemach, A., & Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol*, *20*(17), R780-785. https://doi.org/10.1016/j.cub.2010.07.007

Zhang, T., & Greenberg, R. A. (2022). The inner workings of replisome-dependent control of DNA damage tolerance. *Genes Dev*, *36*(3-4), 103-105. https://doi.org/10.1101/gad.349408.122

Zhou, D., & Robertson, K. D. (2016). Chapter 24 - Role of DNA Methylation in Genome Stability. In I. Kovalchuk & O. Kovalchuk (Eds.), *Genome Stability* (pp. 409-424). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-803309-8.00024-0