Investigating double-strand break formation and repair in meiosis

A thesis submitted for the degree of Doctor of Philosophy in Biochemistry

Dominic Johnson

Genome Damage and Stability Centre

University of Sussex

January 2017

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree

Signature:.....

Acknowledgements

Most importantly, I would like to thank Dr Matt Neale. As a PhD student I could not have asked for a better supervisor. Matt has been extremely supportive throughout my time at Sussex, not only from a research point of view but also from a personal one. His expert advice and encouragement has been a major contributing factor to the completion of this thesis.

I would also like to thank all other members of the Neale laboratory, past and present. They have all aided me with support and advice over the past four years and provide a friendly, enjoyable working atmosphere. Special thanks goes to Valerie Garcia and Tim Cooper for their help and contribution to various parts of this work. Additionally, special thanks goes to Rachal Allison for putting up with me in our lab bay and for providing amazing technical support.

I am grateful to the European Research Council for funding my PhD studentship. Without this financial support, I would not have been able to carry out this research project.

Thanks goes out to all members of the GDSC who have created a stimulating working environment and incredible advice and guidance across a vast range of subject areas. A special thanks goes out to all members of the lunch time/tea club, past and present, for keeping me sane and with whom I have made some special friendships.

Finally, I would like to thank Stephanie Bower and my Mum and Dad for their love and support throughout my PhD. Their faith in me has kept me going.

Summary

University of Sussex
Dominic Johnson
Ph.D. Biochemistry

Investigating DNA double-strand break formation and repair in meiosis

Meiotic recombination is a complex process that requires tight regulation to ensure accurate chromosomal segregation and to prevent DNA double-strand breaks (DSBs), introduced to initiate meiotic recombination, from becoming damaging. Spo11 introduces DSBs via a topoisomerase-like reaction during meiosis. In this thesis I present work investigating the mechanisms that regulate the formation and repair of the protein-linked DSBs created by Spo11 and topoisomerase II (Top2).

Initiation of Spo11-DSB resection is conducted by the Mre11-Rad50-Xrs2 (MRX) complex and Sae2 protein, which nucleolytically removes Spo11 covalently bound to oligonucleotides via a phosphotyrosine bond. Sae2 activity is controlled by post-translational modifications and regulation of its oligomeric state. Here I present data characterising the phenotype of Sae2 proteins mutated at putative Mec1/Tel1 phosphorylation sites **(Chapter 3)**.

The human TDP2 protein, hydrolytically removes proteolysed topoisomerase II (Top2) from the 5′ end of Top2-DSBs. Here I show that TDP2 is also active upon the phosphotyrosine bond between Spo11 and DNA *in vitro* **(Chapter 4)**. Removal of Spo11 from the 5′ end of double-stranded DNA by TDP2 permits resection by lambda exonuclease but no resection is observed by the primary meiotic 5′ to 3′ exonuclease, Exo1, *in vitro*. This suggests an evolutionary benefit of Spo11-DSB processing by the MRX complex and Sae2 at generating a substrate that permits Exo1 resection instead of the hydrolytic Spo11 removal mechanism by TDP2 **(Chapter 4)**.

Utilising TDP2 activity, I have developed a novel method to map Spo11-DSBs genome-wide with single nucleotide resolution **(Chapter 5)**.

The spatial patterning of meiotic DSBs is controlled in yeast by the ATM/ATR homologs Tel1/Mec1. Results from this new genome-wide DSB mapping method suggest that the kinase activity of Tel1 regulates hyper-local repression of coincident Spo11-DSBs **(Chapter 6)**.

Utilising TDP2 and the nucleotide resolution mapping procedure for Spo11, *Saccharomyces cerevisiae* topoisomerase II (Top2) was also mapped genome-wide, which indicated that there are preferential sites for Top2 cleavage *in vivo* **(Chapter 7)**. In the future this procedure can be adapted to map other protein-DNA complexes *in vivo* in a wide range of organisms.

Collectively the work presented in this thesis further elucidates the mechanisms underpinning the spatial patterning of Spo11-DSBs in meiosis, the subsequent repair of meiotic DSBs, and also contributes to our understanding of the location of Top2 cleavage sites *in vivo*.

Table of contents

List of figures

Abbreviations

BSA	Bovine serum albumin
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CDK	Cyclin-dependent kinase
CEN	Centromeric
CO	Crossover
CPT	Camptothecin
CsCl	Caesium chloride
DDR	DNA-damage response
dHJ	Double Holliday junction
DMSO	Dimethyl sulfoxide
dNTPs	Deoxynucleotide triphosphates
DSB	Double-strand break
DSBR	Double-strand break repair
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
gDNA	Genomic DNA
HJ	Holliday junction
HR	Homologous recombination
HRP	Horse-radish peroxidase
HYG	Hygromycin
IP	Immunoprecipitation
IR	Ionising radiation
λexo	Lambda exonuclease
LB	Liquid broth
MMS	Methyl methanesulfonate
NCO	Non-crossover
NHEJ	Non-homologous end joining
ORF	Open reading frame
PK	Proteinase K
rDNA	Ribosomal DNA
SC	Synaptonemal complex
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SDSA	Synthesis dependent strand annealing
SPM	Sporulation medium
SSB	Single-stranded break
ssDNA	Single-stranded DNA
TdT	Terminal deoxynucleotidyl transferase
TSS	Translation start site
UV	Ultraviolet

CHAPTER 1:

INTRODUCTION

Chapter 1: Introduction

2.1 Mitosis and meiosis

Cells need to divide to enable an organism to grow as well as to replace dying or damaged cells. They do this via a process known as mitosis, which, consists of DNA replication followed by one round of chromosome segregation resulting in two daughter cells identical to the parent cell. The actual mitotic divisional phase is relatively short, with the preceding interphase taking up the majority of the cycle. During interphase the cell initially grows and prepares for DNA replication through expression of proteins and replicating organelles (G1 phase). This is followed by replication of the cell's DNA to produce sister chromatids (S phase), and finally more growth and preparation for division (G2) before mitosis begins. The mitotic pathway condenses the chromosomes (prophase), aligns them at the metaphase plate (metaphase), separates sister chromatids (anaphase), with the chromatids arriving at opposite poles of the cell and decondensing (telophase), before the cell physically divides in two, generating the two identical daughter cells (cytokinesis) (see Figure 1.1 for schematic). During the creation of gametes a specialised mode of cell division occurs known as meiosis. A diploid cell undergoes one round of DNA replication (in pre-meiotic S phase) followed by two rounds of chromosome segregation to produce cells with half the chromosome complement. The first round of chromosomal segregation is a reductional division (meiosis I), where homologous chromosomes segregate. The second round is an equational division (meiosis II), where sister chromatids segregate, similar to mitotic sister chromatid segregation. This specialised cell division of meiosis is controlled by regulation of the cell cycle proteins, cyclin-dependent kinases (CDKs).

2.1.1 Cell cycle regulation in mitosis and meiosis

In mitotic cells, CDKs control the tight regulation required to ensure a cell cycle stage has been completed correctly before progressing to the next. CDKs are present in all known eukaryotes and are characterised by the need for a cyclin subunit for their enzymatic activities. In mammals CDKs evolved into three cell cycle-related subfamilies and five transcriptional subfamilies, whereas in *S. cerevisiae* there are five CDKs with the master regulator being Cdc28, an ortholog of mammalian Cdk1 (reviewed in Mendenhall & Hodge 1998; Malumbres 2014). In *S. cerevisiae*, Cdc28 levels are consistent and in excess throughout the cell cycle. The binding of inactivating proteins, as well as activating/inactivating phosphorylation events, control CDK activity to some degree but mainly it is the pairing of particular cyclins, of which

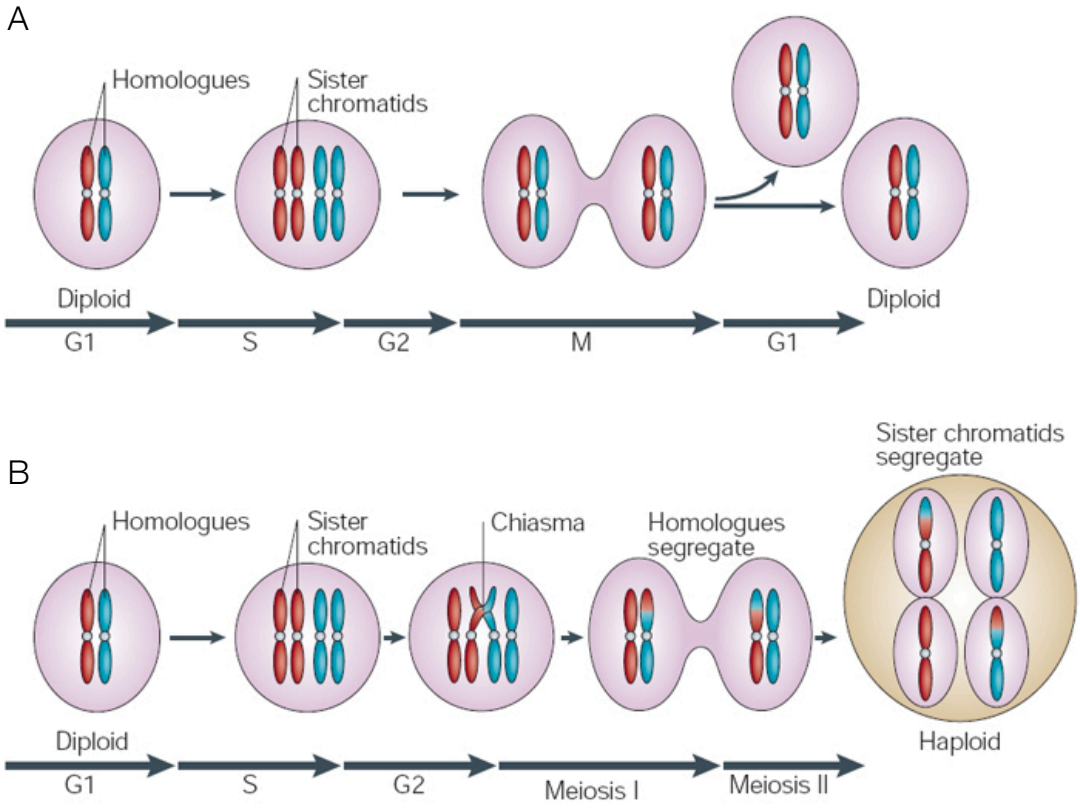


Figure 1.1: Mitotic and meiotic cell division.
A. In mitosis, DNA replication occurs during S phase to produce sister chromatids. Sister chromatids then condense, align and separate during the mitotic (M) phase of the cell cycle generating diploid daughter cells. **B.** Meiosis consists of one round of DNA replication in S phase (pre-meiotic S phase) followed by two rounds of segregation (meiosis I and II). During prophase I homologs combine physically via the formation of chiasmata before being segregated to opposite poles (reductional division). During meiosis II sister chromatids then segregate to opposite poles resulting in the formation of non-identical haploid gametes. Lengths of the cell-cycle stages are not drawn to scale.
Adapted from Marston & Amon, 2004

there are nine in *S. cerevisiae*, whose levels and binding to CDKs determine the cell cycle stage and its progression (Barral et al., 1995; Suryadinata et al., 2010). These cyclins consist of: Cln1-3 that sequentially interact with Cdc28 in G1 phase, Clb5 and Clb6 in S phase and Clb1-4 in mitosis (Nasmyth, 1996).

CDKs are serine/threonine kinases, and analysis of a large number of their substrates has revealed that CDK phosphorylation requires a proline residue immediately adjacent (C-terminally) to the serine or threonine site to be phosphorylated (S/T-P). This event can be promoted by, but is not dependent on, a positively charged arginine or lysine three amino acids downstream (S/T-P-X-R/K) (Songyang et al., 1994, 1996; Srinivasan et al., 1995). This consensus sequence is conserved for CDKs, but substrate specificity can be regulated via the activating cyclin-subunit, which also has specific residues that they recognise and interact with (cyclin-binding motif). Mutation of this cyclin-binding motif on the substrate can prevent phosphorylation by CDK (Adams et al., 1996; Furstenthal et al., 2001; Schulman et al., 1998).

The mitotic cycle consists of one round of DNA replication followed by one round of segregation in order to produce two identical daughter cells. This process is coordinated by a two-step regulation of CDK activity (Tanaka and Araki, 2010). First, in the G1 stage, CDK activity is kept low, which allows licensing to occur. This is where the pre-replicative complex assembles at origins of replication, a process essential for DNA replication to occur but is inhibited by CDK activity. Then, in late G1, CDK activity increases, triggering DNA replication from the origins. These origins of replication cannot be licensed again until exit from the mitotic cycle when CDK levels drop (Ohkura, 2015). This control of DNA replication is manipulated during the specialised cell division of meiosis.

One round of DNA replication followed by two rounds of segregation with no intervening S phase is essential for a reduction in ploidy. The inhibition of S phase is carried out in a meiotic cell by maintaining high CDK activity levels between the two divisions (Ohkura, 2015). Meiosis utilises many of the same cyclins from mitosis to control CDK activity. However in contrast, it has an essential requirement for Clb5 for pre-meiotic S phase and meiotic initiation, with the Clb2 cyclin not required. Therefore, *CLB2* expression is switched off at the start of meiosis. This natural silencing of the *CLB2* gene in meiosis, by suppression of its promoter, can be used to create meiotic nulls of proteins. Placing a gene under control of the *CLB2* promoter switches off expression of the protein during meiosis, with expression during the rest of the cell cycle unaffected. This allows the study and manipulation of meiotic proteins that may be essential for normal cycling cells (Jessop et al., 2006).

In diploid *Saccharomyces cerevisiae* cells the absence of a nitrogen source combined with a non-fermentable carbon source causes diploid cells to enter meiosis (Freese et al., 1982). This

entry into meiosis is initiated by the expression of the Ime1 transcription factor, which irreversibly directs the cell to undertake meiosis. Activation of Ime1 leads to the transcription of the 'early' meiotic genes, which play roles in entry into pre-meiotic S phase, recombination and chromosomal pairing in meiotic prophase (Primig et al., 2000). ImeI and another kinase, Ime2, also promotes suppression of replication between divisions and primes the cell for entry into the meiotic divisions (Guttmann-Raviv et al., 2001). Ime1 and Ime2 activate CDK and also initiate expression of the transcription factor Ndt80 (Holt et al., 2007; Shin et al., 2010). Ndt80 activation induces genes required for the exit from prophase (Tung et al., 2000).

Reductional division in meiosis I requires the homologs to pair in order to be segregated, a unique situation to meiosis. This process takes place in the first meiotic prophase whereby homologs recombine and generate a physical connection to each other to aid orientation and division. Accurate segregation of homologs is critical to prevent nondisjunction, a situation where gametes contain extra or missing chromosomes, which can, in humans, lead to genetic diseases such as Down's syndrome (three copies of chromosome 21). The process of recombining the homologs is achieved by utilising the homologous recombination repair pathway of DNA double-strand breaks (DSBs).

2.2 Double-strand breaks

External and internal forces continuously damage the genome of a cell with as many as 1 million individual lesions per cell per day (Lodish and Berk, 2003). There are many types of DNA damage with the most critical type being a DSB whereby both strands of the DNA duplex are broken. DSBs are formed from external sources such as ionising radiation (IR). Internally, DSBs can be generated by reactive oxygen species (from cellular metabolism) and during DNA replication when replication forks collide with a lesion (Bosco et al., 2004; Limoli et al., 2002). DSBs are also programmed by the cell during processes such as mating-type switching in yeast (Pueyo et al., 1993), V(D)J recombination (Franco et al., 2006) and during meiosis (Keeney and Neale, 2006). If left unrepaired DSBs can lead to cell death, and if not repaired correctly DSBs can cause translocations, loss of heterozygosity and deletions in the DNA, as well as chromosome loss. All these factors are associated with cancerous cells (Aplan, 2006). Therefore, the cell has evolved groups of proteins that sense DSBs and act as a checkpoint to arrest the cell and activate DNA repair pathways (more detail in later sections) (Su, 2006). Primarily the two pathways that repair DSBs are non-homologous end joining (NHEJ) and homologous recombination (HR). The differences between NHEJ and HR are described in more detail below and are represented as a schematic in Figure 1.2.

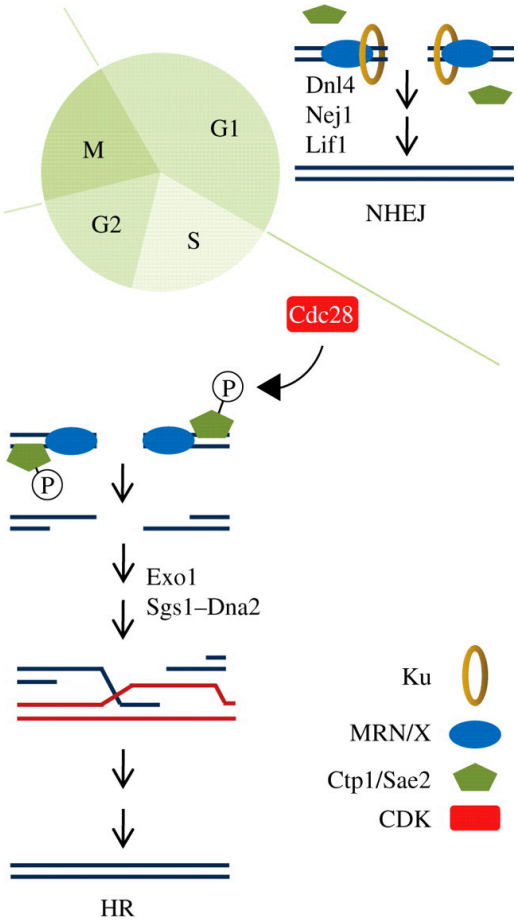


Figure 1.2: DSB repair pathways: Non-homologous end joining and homologous recombination. DSBs can be repaired by non-homologous end joining (NHEJ) or homologous recombination (HR) depending on the cell cycle stage. In G1, the Ku heterodimer binds DSBs and promotes NHEJ alongside Dnl4 and Lif1 with MRX tethering the ends. When Cdc28 is active, such as in G2/S or during meiotic prophase, Sae2 gets phosphorylated and activates the nuclease activity of Mre11 resulting in resection of the 5′ strand. Exo1 and/or Sgs1-Dna2 resection machinery performs long-range resection to generate long ssDNA tails capable of strand invasion, a step necessary for efficient homology search and repair of the DSB.
Adapted from Langerak & Russell 2011

The balance between DSB repair pathway choice differs between species, cell type and cell cycle phase. For example, in *S. cerevisiae,* unlike in mammalian cells, HR predominates over NHEJ as the DSB pathway of choice outside of G1 (Friedl et al., 1998). NHEJ does play a role, and predominates in G0 and G1, whereas HR predominates after DNA replication when an undamaged sister chromatid is present to use as a template (Veuger et al., 2003). The main process that shifts repair from NHEJ to HR is the 5′ resection of the DSB. The primary cyclin-dependent kinase in *S. cerevisiae*, Cdc28, is a major regulator of this pathway choice, and its activity is required for the 5′ resection of the DSB, thus shifting repair towards HR. Cdc28 is active in all cell cycle stages except G0/G1, the stage where NHEJ predominates. This therefore provides the link between DSB repair pathway choice and the cell cycle stage (Aylon et al., 2004). Corroborating this, CDK is active in meiosis and DSBs repair exclusively via HR (Huertas et al., 2008).

2.2.1 NHEJ

NHEJ is classically known as an error-prone method of DSB repair. However, in the case of clean-ended DSBs, such as those created by nucleases leaving complementary overhangs, NHEJ can repair accurately (Rassin, 2003). When DSBs are formed with non-complementary ends, or with complex adducts or structures, NHEJ cannot precisely re-join the ends. In this situation NHEJ involves micro-homology to direct repair whereby some complementary bases align and flaps are removed. This leads to the occurrence of insertions and deletions giving rise to its error-prone reputation.

S. cerevisiae NHEJ initiates with the binding of the highly conserved Mre11-Rad50-Xrs2 (MRX) complex (MRN (Nbs1) in mammalian cells), which tethers the DSB (Lisby et al., 2004). The MRX complex is also involved in HR, described later. The current mechanism is still unclear but either prior to, or following MRX binding, the Ku complex binds the DSB end (Mari et al., 2006). Ku is conserved from bacteria to humans and in *S. cerevisiae* is a heterodimer of Yku70 and Yku80. Ku binds using its ring structure to slide onto the DSB end and it orients itself with the Yku80 C-terminus positioned towards the DSB end to allow contact with Dln4 (Daley et al., 2005).

Dnl4 and Lif1 (mammalian DNA ligase IV and XRCC4) bind to Yku80 and Xrs2 respectively, permitting the recruitment of DNA ligase IV (Chen et al., 2001). The Nej1 protein also promotes NHEJ through controlling the subcellular localisation of Lif1 (Valencia et al., 2001). Alignment and complementary base pairing at the DSB now occurs and ligation is attempted. Failure to ligate the DSB results in processing of the end in some manner before attempting ligation again. This processing is done by a variety of different proteins such as Pol4, which

fills in ends (Wilson and Lieber, 1999), and Rad27 (mammalian FEN-1), which cleaves 5′ flaps (Liou et al., 2001).

2.2.2 An overview of homologous recombination

Homologous recombination (HR, Figure 1.3) utilises homologous sequences elsewhere in the genome as a template for repair. These sequences can include the sister chromatid, homologous chromosome or repeated regions either close by or on different chromosomes (Stahl, 1996; Szostak et al., 1983). However, in order for error-free repair, the homologous sequence must be a perfect complement, a situation only found on the sister chromatid after DNA replication (Figure 1.1A). Repair from a non-perfect homolog can lead to loss of heterozygosity. This is due to gene conversion, a process where the unbroken template sequence is copied non-reciprocally to the broken locus (Luo et al., 2000). Additionally, when HR repairs from a non-allelic/ectopic region, often due to repetitive sequences, damaging genomic rearrangements can occur (Stankiewicz and Lupski, 2002).

Briefly, in mitotic cells, with more detail on specific topics and meiotic events in later sections, the double-strand break repair (DSBR) model describes HR initiating from DSBs, which are recognised by the MRX complex. Nucleolytic end-processing of the 5′ strand occurs by MRX and Sae2 (CtIP in mammalian cells), generating short 3′ ssDNA tails (Mimitou and Symington, 2009; Pueyo et al., 1993). In mitotic cells the resection of the 5′ strand is continued by exonuclease 1 (Exo1), and/or the Sgs1-Top3-Rmi1 (STR)-Dna2 complex, to generate long 3′ ssDNA tails (Krogh and Symington, 2004). These ssDNA tails are bound by replication protein A (RPA), minimising the formation of secondary structures and facilitating loading of the recombinase Rad51 by the mediator protein Rad52 or human BRCA2 (New et al., 1998). Rad51 forms a nucleoprotein filament on ssDNA, which is capable of homology search, strand invasion and pairing with an undamaged homologous duplex (San Filippo et al., 2008). Here the invading strand generates a displacement loop (D-loop) and is used to prime DNA synthesis from the template strand. The 3′ ending strand from the other side of the DSB anneals to the displaced strand from the donor duplex and primes a second round of leading strand synthesis. A double Holliday junction (dHJ) intermediate is formed after ligation of the newly synthesised strand to the resected 5′ strand. These branched structures can be resolved in different ways resulting in different outcomes of HR. A process known as dissolution involves helicases and topoisomerases, which migrate the branched HJ structures towards each other and cleave the inner strands (decatenate) resulting in non-crossovers (NCOs) (Cejka et al., 2010; Ellis et al., 1995; Wu and Hickson, 2003). In contrast, cleavage of the outer strands of one HJ and the inner strands of the other HJ generates crossover (CO) products (reviewed in

Symington et al., 2014). Mitotic DSBs repairing via HR generally do not resolve via the CO pathway, unlike in meiosis where half of DSBs repair as COs in *S. cerevisiae* (Chen et al., 2008; Mancera et al., 2008). Another model, alongside the DSBR model, has been proposed known as the synthesis-dependent strand annealing (SDSA) model (Nassif et al., 1994; Pâques et al., 1998). The SDSA model states that one 3′ end of the DSB invades the homologous template duplex, which is displaced by helicases after limited DNA synthesis. The nascent complementary strand anneals and, following filling-in synthesis and ligation, NCO products are formed (Nassif et al., 1994; Symington et al., 2014).

2.3 The DNA damage response checkpoint

DSBs are critical lesions to the cell, therefore there are checkpoint pathways in place to ensure that they are dealt with in the correct manner before cell cycle progression can continue. The DNA damage response (DDR) signal transduction pathway, as the name suggests, detects DNA damage and couples the prevention of cell cycle progression to the activation of DNA repair pathways. It has at its heart two main players: Ataxia-telangiectasia mutated (ATM), and Ataxia-telangiectasia and RAD3-related (ATR) proteins (reviewed in Ciccia & Elledge 2010). ATM and ATR are members of the phosphoinositide 3-kinase-related protein kinase (PIKK) family and are highly conserved through all eukaryotes, with the *S. cerevisiae* homologs being Tel1 (Greenwell et al., 1995; Morrow et al., 1995) and Mec1 (Kato and Ogawa, 1994; Weinert et al., 1994) respectively.

To invoke their DDR checkpoint functions they phosphorylate a known consensus sequence on proteins, hydrophobic-X-hydrophobic-[S/T]-Q sites, also named S/T-Q or SQ/TQ sites (Kim et al., 1999). Some of these S/T-Q sites have been found to cluster within 100 amino acids, with these regions named S/T-Q cluster domains (SCD), and are considered to be a structural hallmark for DNA damage response proteins (Traven and Heierhorst, 2005).

To give time for DNA repair to occur before replication or mitosis begins, ATM and ATR phosphorylate CHK1 and CHK2, which all act together to reduce CDK activity. This inhibition of CDK arrests cell-cycle progression at G1/S and G2/M checkpoints (Bartek and Lukas, 2007; Kastan and Bartek, 2004). There are numerous other phosphorylation substrates of $Mec1^{ATR}$ and $Tel1^{ATM}$, including DNA repair proteins, which have been identified in yeast and mammalian systems using different techniques (Chen et al., 2010; Matsuoka et al., 2007; Smolka et al., 2007; Stokes et al., 2007).

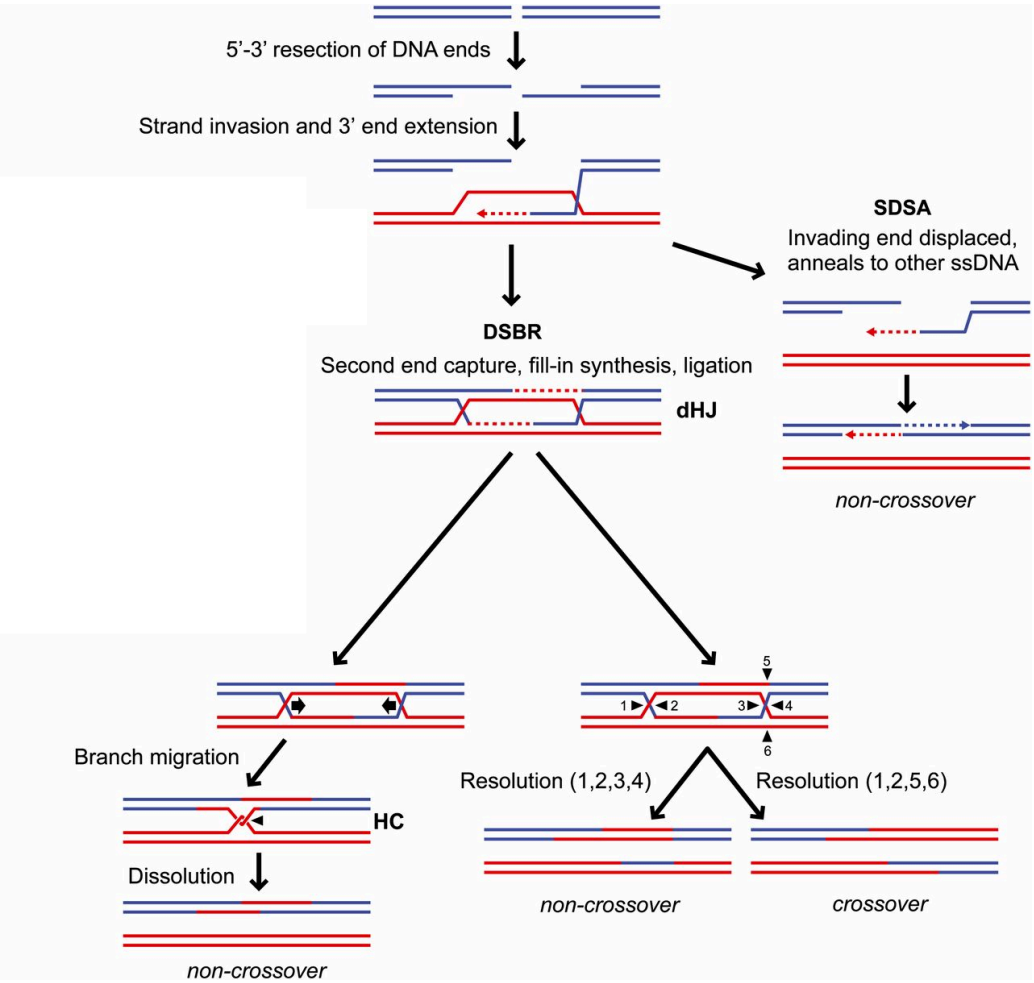


Figure 1.3: Models for homologous recombination resolution.
DSBs undergo 5′ to 3′ resection as detailed in the main text. The long ssDNA 3′ end invades a homologous sequence and primes leading strand synthesis. The SDSA pathway involves the invading strand being displaced after some limited synthesis, which then reanneals with the 3′ssDNA end of the other side of the DSB. Fill-in synthesis and ligation then always results in NCO products forming. The DSBR pathway involves the 3′ ssDNA tail from the other side of the break annealing to the displaced strand from the donor duplex (known as second-end capture), priming a second round of leading strand synthesis. A double-Holliday junction is formed after ligation of the newly synthesised DNA to the resected 5′ strands. Two resolution pathways are possible. The dHJ can branch migrate into a hemicatenane (HC) and be dissolved or they can cleaved by endonuclease at positions 1, 2, 3 and 4, both leading to NCO products. CO products form when the joint molecules are resolved by cleavage of the junctions at positions 1, 2, 5 and 6.
Adapted from Symington et al., 2014

Tel1^{ATM}

Mec1 and Tel1 are activated in different ways in response to DNA damage. ATM (Tel1) is mutated in ataxia-telangiectasia (A-T) disease, a disease characterised by cancer predisposition, defective cell cycle checkpoint and neurodegeneration (Anheim et al., 2012). These characteristics are similar to those seen in diseases linked to mutations in the individual members of the MRN complex (Shiloh, 2014), suggesting an overlap in the functions of ATM and the MRN complex. The MRN complex recruits ATM to the site of DNA damage, and they interact via the Nbs1/Xrs2 subunit (Nakada et al., 2003; You et al., 2005). Once localised to the DSB by MRN, ATM is activated by monomerisation and autophosphorylation (Bakkenist and Kastan, 2003). If any component of the MRX complex is defective, Tel1 activation is prevented, although the mechanism behind this is still relatively unknown (Berkovich et al., 2007; Falck et al., 2005; Lee and Paull, 2005; Nakada et al., 2003; You et al., 2005). Once the ends of DSBs are subjected to 5′ to 3′ resection, Tel1 signalling activity is disrupted (Mantiero et al., 2007), but the long ssDNA tails produced from extensive 5′ to 3′ resection go on to active the Mec1 kinase (Gobbini et al., 2013).

Mec1^{ATR}

Mec1 is also promoted by DSBs but its recruitment is reliant on single-stranded DNA (ssDNA), specifically RPA-coated ssDNA 3′overhangs via an ATR/Mec1 interacting protein, ATRIP (Ddc2 in *S. cerevisiae*). When situated at these ssDNA sites, the 9-1-1 complex (RAD9-RAD1-HUS1 in mammals and Ddc1-Rad17-Mec3 in *S. cerevisiae*), which was loaded at the ds-ssDNA junction by the RFC-like clamp loader (Rad24-Rfc2-5 in *S. cerevisiae*), directly stimulates Mec1 kinase activity (Chen et al., 2007; Cortez, 2001; Mains et al., 1990; Paciotti et al., 2000; Zou and Elledge, 2003). Therefore, early DSB formation (or unresected breaks) recruit and activate Tel1 through the MRX complex, and after end processing, resection and binding of RPA to the 3′ ssDNA produced, Mec1 is recruited and activated.

Many studies on these two proteins have concerned mitotic DSBs. However, as meiotic HR is reliant on DSB formation by Spo11, it is unsurprising that Mec1 and Tel1 are also active and even have unique roles in meiotic DSB repair (discussed in later sections).

2.4 Meiosis and HR

Meiosis in most organisms utilises the strand invasion ability of HR for the pairing of homologs and for the crossing over of genetic material, resulting in the production of genetically diverse haploids (Page, 2003; Petronczki et al., 2003). As previously stated, mitotic DSBs repaired by HR preferentially utilise the sister chromatid to minimize genome

rearrangements and loss of heterozygosity (Johnson and Jasin, 2001; Kadyk and Hartwell, 1992). Meiotic recombination shares similarities with mitotic recombination at the early stages; however, there are many differences between the two. More detail on meiotic DSB formation, end processing and resection will be discussed in later sections. The following sections outline briefly some other differences between meiotic HR and mitotic HR.

2.4.1 RPA

In both meiotic and mitotic cells the 3′ ssDNA tails generated from the action of MRX/Sae2 and Exo1 are bound by the abundant ssDNA binding protein, RPA (Lisby et al., 2004; Wang and Haber, 2004). RPA is composed of three subunits: RPA1, RPA2 and RPA3, and prevents formation of secondary structures, whilst protecting ssDNA from nucleolytic degradation. RPA is essential for mitotic growth and meiotic recombination, with mutants defective in the repair of Spo11-DSBs (Soustelle et al., 2002). However, recently, meiosis-specific RPA homologs have been shown to be involved. Studies in metazoans identified a paralog of RPA1, known as MEIOB, which interacts with RPA2 and is essential for proper meiotic recombination in mice (Luo et al., 2013; Souquet et al., 2013).

2.4.2 Inter-homolog bias

The aim for meiotic recombination is to join homologs and exchange genetic material. Recent studies indicate that 70-90% of meiotic DSBs are repaired via the homolog and the rest via the sister chromatid (Goldfarb and Lichten, 2010), a process known as inter-homolog bias (Haber et al., 1984; Jackson and Fink, 1985; Schwacha and Kleckner, 1994, 1997). This bias is driven by a number of factors. During meiosis the Mec1/Tel1 kinases are activated and phosphorylate the Hop1 protein (Carballo et al., 2008), which in turn promotes activation of Mek1, a protein required to promote this inter-homolog bias (Kim et al., 2010).

In mitosis, the Rad51 protein conducts homology search to direct repair from the sister chromatid. Rad51 is still expressed in meiosis but another meiosis-specific nucleofilament-producing protein is also expressed, Dmc1 (Bishop et al., 1992). Rad51 cannot substitute for the loss of Dmc1 in meiosis and cells accumulate resected DSBs (Bishop et al., 1992; Usui et al., 2001; Xu et al., 1997). However, Rad51 is still required for inter-homolog bias (Schwacha and Kleckner, 1997). Experiments looking at a separation-of-function mutant of Rad51 (*rad51-I13A*) revealed that Rad51 joint molecule activity is dispensable in meiosis with its important function being to act as an accessory factor to Dmc1 (Cloud et al., 2012).

2.4.3 Non-crossovers vs. crossovers

Like mitotic HR, meiotic DSBs can repair via either a NCO or CO pathway. However, COs are of critical importance to meiosis as they establish chiasmata, physical links between homologs tethering them together. Approximately half of DSBs repair as COs in *S. cerevisiae* meiosis with 5-6 COs per chromosome (Chen et al., 2008; Mancera et al., 2008). In other species, however, CO numbers can be lower with only a single 'obligate' crossover per chromosome (Jones, 1967) being essential for correct segregation (reviewed in Jones and Franklin, 2006).

The differentiation between a NCO and a CO is thought to be conducted at the formation of the D-loop stage by the proteinacious Rad51-Dmc1 filament (Bishop and Zickler, 2004; Börner et al., 2004; Hunter and Kleckner, 2001; Zickler and Kleckner, 1999). The subsequent dHJ formation via the DSBR pathway has the potential to produce either a CO or NCO; however it was reported that these joint molecules are predominantly resolved as COs (Allers and Lichten, 2001). Instead, NCOs are predominantly formed via the SDSA pathway (Pueyo et al., 1993). One way in which NCOs are promoted is via the Sgs1-Top3-Rmi1 complex. Even though Sgs1 (Dna2) does not seem to play a role in the resection of the DSB during meiosis (Zakharyevich et al., 2010), it is still a key player at the later stages of recombination by disrupting nascent D-loops and preventing multi-chromatid joint molecules, promoting NCOs (Oh et al., 2007).

2.4.4 Synaptonemal complex

In meiosis, the axis of homologous chromosomes are synapsed via a proteinacious structure known as the synaptonemal complex (SC). Assembly of this meiosis-specific scaffold, which stretches the entire length of the paired chromosomes, is essential for the formation of meiotic COs. The homologs initially pair via the process of homology search and strand invasion, mediated by Dmc1/Rad51 nucleoprotein filaments. Synapsis is then initiated and a tripartite structure consisting of Zip1, Hop1 and Red1 begins to form, with Zip1 polymerising the length of the chromosomes (Sym et al., 1993). Zip1 is part of a group of proteins known as ZMMs (Zip1-4, Msh4/5, Mer3, Spo16, Pph3), which have various roles in the formation and regulation of the SC. These proteins are key for the final stages of recombination. Their absence results in accumulation of DSBs and dHJs, and the homologs fail to synapse. Unsynapsed homologs results in checkpoint activation and defects in meiotic progression (Börner et al., 2004; Shinohara et al., 2008; Sym et al., 1993).

2.4.5 Meiotic phases

Compared to mitosis, meiosis has a prolonged prophase I that can be divided cytologically into five distinct stages: (i) Leptonema, chromosomes condense and axial elements begin to form on sister chromatids. (ii) Zygonema, homologs pair and begin to synapse along the axis via the SC. (iii) Pachynema, homologs align and synapse along the entire axis. (iv) Diplonema, dissolution of synapsis occurs but homologs remain linked via chiasmata. (v) Diakinesis, the majority of SC is lost and chromosomes begin to condense. The prolonged prophase I enables accumulation of DSBs to allow for the increased chance of chiasmata forming (Merino et al., 2000; Møens and Pearlman, 1988; Zickler and Kleckner, 1998, 1999). The pre-meiotic S phase is also prolonged compared to mitotic S phase and it is thought that during this extended period of replication other meiosis-specific events take place. This rationale is due to deletion of the *SPO11* gene (essential for meiotic recombination) reducing the length of S phase by a quarter, suggesting that the recombination machinery assembles alongside DNA replication (Cha et al., 2000). Additionally, the time that a DSB is made at a certain site is rigidly connected to the time of DNA replication at the same locus (Borde, 2000). A recent study showed that the replisome-associated components Tof1 and Csm3 associate and recruit Dbf4-dependent Cdc7 kinase (DDK) to the replication fork. Here DDK phosphorylates chromatin-bound Mer2 in replicating regions (Murakami and Keeney, 2014). Phosphorylation of Mer2 promotes recruitment of other DSB-forming proteins thus linking DNA replication timing and DSB formation (Henderson et al., 2006; Panizza et al., 2011; Sasanuma et al., 2008).

2.5 Meiotic DSB formation

Even though numerous spontaneous DSBs occur in a cell, during meiosis a consistent and efficient way to utilise HR to form COs is required and is achieved by the generation of DSBs via the Spo11 protein in prophase I (Szostak et al. 1983; Sun et al. 1989; Cao et al. 1990; Bergerat et al. 1997; Keeney et al. 1997). To ensure accurate homolog segregation, each cell forms numerous DSBs to ensure the formation of at least one chiasmata/CO per chromosome, as well as the numerous DSBs promoting homologue pairing and alignment (Henderson and Keeney, 2004; Kauppi et al., 2013; Tessé et al., 2003). It is estimated that 160 DSBs are formed per cell in *S. cerevisiae* (Pan et al., 2011), 15-30 in *Caenorhabditis elegans* and more than 200 in mammalian cells (Martinez-Perez and Colaiácovo, 2009; Rosu et al., 2011).

2.5.1 Spo11

Spo11 was first seen to be essential for meiotic recombination in *S. cerevisiae* in 1985 (Klapholz et al., 1985), but its precise role was unknown until 1997. Initial studies into meiotic recombination observed that in a background where DSBs accumulated (*sae2*Δ*, mre11S and rad50S*), a covalent protein linkage to the 5′ end either side of the DSB was present (de Massy et al., 1995; Keeney and Kleckner, 1995; Liu et al., 1995). This sort of covalent linkage is reminiscent of topoisomerase-like transesterification reactions where the protein forms a transient covalent bond with the DNA at an intermediary step (reviewed in Champoux, 2001).

Genetic and sequencing studies on the archaeal TopoVIA subunit described similarities with four proteins of unknown function, one being the *S. cerevisiae* protein Spo11 (Bergerat et al., 1997). Topoisomerases cleave DNA by attacking the phosphodiester bond with a catalytic tyrosine residue, forming a covalent bond. Spo11 contains such a tyrosine at Tyr135. Upon mutating this residue to a phenylalanine, preventing any catalytic activity, DSBs in meiosis were no longer formed and no meiotic recombination took place (Bergerat et al., 1997). Spo11-DNA species were isolated using a biochemical method in *rad50s* meiotic cells. Nuclei were isolated from meiotic cells and gDNA extracted using guanidine-HCl plus ionic detergent at 65 °C, with the bulk protein removed using a caesium chloride gradient. These harsh conditions isolate gDNA away from all weakly bound proteins but still retain any covalently bound protein on the DNA. Protein-bound DNA was isolated away from free DNA using a glass fibre filter to which proteins specifically adsorb to and protein-DNA species eluted using SDS. Non-specific nucleases were used to remove any bound DNA and the sample analysed by mass spectrometry. A 45 kDa protein, Spo11, was enriched in the *rad50s* strains and proceeding experiments immunoprecipitated the Spo11-HA protein, which coprecipitated with DNA from the known strong *HIS4::LEU2* DSB hotspot (Keeney et al. 1997). These studies indicated that Spo11 was the responsible protein for creating DSBs in meiosis and accumulates covalently bound to the end in *rad50S* and *sae2*Δ cells (de Massy et al., 1995; Keeney et al., 1997; Keeney and Kleckner, 1995; Liu et al., 1995).

The essential function of Spo11 in *S. cerevisiae* extends to its orthologs in other species. The Spo11 protein itself is highly conserved and studies in *C. elegans*, *Drosophila melanogaster* and *Mus musculus* have all demonstrated its ability to initiate meiotic recombination (Baudat et al., 2000; Dernburg et al., 1998; Hayashi-Hagihara, 1998; Romanienko and Camerini-Otero, 2000). Interestingly, in *C. elegans,* ionising radiation-induced DSBs alleviated the dependence on Spo11 for meiotic recombination. This indicates that whilst cells are dependent on Spo11 for meiotic recombination, it is primarily the DSB-forming ability of Spo11 that is required for this function (Dernburg et al., 1998).

2.5.2 The Spo11 complex

Topoisomerase II (Top2), which cleaves DNA in a similar manner to Spo11, acts as a dimer (Shelton et al., 1983). Experiments where Spo11 was tagged with two different affinity tags, the two tagged Spo11 forms coprecipitated suggesting dimeric assembly or a larger oligomeric structure. This dimerisation of Spo11 is essential for DSB formation and dependent on two other proteins, Rec102 and Rec104 (Sasanuma et al., 2007). The Rec102-Rec104 complex is also required for the nuclear localisation and activity of Spo11 (Kee et al., 2004). These two proteins are not the only proteins essential for Spo11 DSB formation. In fact there are at least nine proteins required for DSB formation in meiosis. Of these nine proteins, yeast two-hybrid experiments identified four sub-complexes (Rec102-Rec104, Spo11-Ski8, Rec114-Mei4-Mer2 and Mre11-Rad50-Xrs2), which also interact between themselves and are all required for DSB formation (Arora et al., 2004; Maleki et al., 2007; Petrini, 1999; Uetz et al., 2000).

Ski8 plays a role in RNA metabolism in vegetative cells (Masison et al., 1995), but it is also essential for meiotic DSB formation (Gardiner et al., 1997). It has been described as a scaffold protein that aids in the nuclear localisation of Spo11 and in the recruitment of other DSB proteins to the chromosome. Yeast two-hybrid assays also indicated Ski8 is required for the interaction of Spo11 and Rec104 (Arora et al., 2004). Mer2 interacts with itself, Mei4, Xrs2 and Rec114 (Arora et al., 2004; Henderson et al., 2006; Li et al., 2006; Maleki et al., 2007). Its phosphorylation by CDK in meiosis is essential for meiotic recombination (Henderson et al., 2006). The Rec114-Mei4-Mer2 sub-complex binds axial sites which anti-correlate with DSB sites. Mer2 is recruited to these axial sites by Red1 and Hop1 (Panizza et al., 2011). The role of Rec114 itself is relatively unknown but phosphorylation by Mec1/Tel1 is important for its function. Mutating the S/T-Q motifs of Rec114 to alanines resulted in a subtle increase in DSB formation, whereas a phosphomimetic mutation, and over-expression of the wild type protein, decreased DSB formation (Bishop et al., 1999; Carballo et al., 2013). The evolutionarily conserved MRX complex, as described earlier and in more detail later, has a role in the nucleolytic processing of DSBs in HR. However, it is also required for DSB formation in *S. cerevisiae* (reviewed in Keeney 2001) and in *C. elegans* (Chin and Villeneuve, 2001), but not for DSB formation in *Schizosaccharomyces pombe* or *Arabidopsis thaliana* (Bleuyard et al., 2004; Puizina, 2004; Young et al., 2004). In *S. cerevisiae* Mre11 requires all the components of the Spo11 complex, except Rad50, for localisation to the DSB site. However, the binding of Mre11 does not require DSB formation itself and Mre11 still transiently associates with the DNA in a *spo11-Y135F* (catalytic-dead) mutant, indicating that MRX plays a role in the pre-DSB complex. This tightly couples DSB formation with the end-processing of Spo11-DSBs (Borde et al., 2004). However, in mammalian cells these DSB-forming proteins are essential

during early development, preventing a determination of their role in mammalian meiosis (Cole et al., 2010).

2.5.3 Interference & spatial patterning of DSBs

When and where DSBs are made in the genome during meiosis is under tight control by numerous different pathways. As previously stated, DSBs only form in a narrow time window of prophase I (Padmore et al., 1991). DSB formation is tightly coupled to pre-meiotic replication, occurring 1.5-2 hours after, presumably ensuring that DSBs achieve their correct function and are not toxic lesions at inappropriate times. It was first seen that blocking DNA replication prevented DSB formation in a replication checkpoint-independent manner (Borde, 2000), but later it was documented that pre-meiotic replication is not a prerequisite for DSB formation. Rather, the coordination of the two processes is regulated by CDK and Dbf4-dependent kinase Cdc7 (DDK) phosphorylation of Mer2 (Blitzblau et al., 2012; Murakami and Keeney, 2008, 2014).

Spo11 has some bias towards DNA sequence (Murakami and Nicolas, 2009; Pan et al., 2011), yet the sites of the genome where Spo11 creates DSBs are not driven towards this weak site preference. Instead DSBs are spread non-randomly throughout the genome and are often clustered at, what are designated, recombination hotspots, a region of approximately 70-250 bp (reviewed in Lichten & Goldman 1995). In *S. cerevisiae* there doesn't seem to be one responsible pathway for designating hotspot usage, instead numerous pathways affect DSB position, as discussed below.

One well-recognised factor in the location of DSB hotspots is that they always form in micrococcal nuclease (MNase)-sensitive/nucleosome-depleted regions (NDRs), a feature commonly linked with promoter regions (Fan and Petes, 1996; Kaplan et al., 2009; Pan et al., 2011) (Figure 1.4C). However, even though the majority of DSBs occur in NDRs, not all NDRs contain DSBs, which therefore suggests a further level of control in Spo11 recruitment and/or activation at these sites (Pan et al., 2011). DSB hotspots may be targeted in different ways between organisms. In *S. cerevisiae* it has been published that methylation of histone H3 on lysine 4 (H3K4) correlates with Spo11 hotspots, with the methyltransferase Set1 required for this modification. In the absence of Set1, the frequency of DSBs at 84% of meiotic sites was severely reduced, correlating with a reduction of local levels of H3K4 methylation (Borde et al., 2009; Sollier et al., 2004). Similarly, in mammalian cells a H3K4 methyltransferase known as PRDM9 mediates hotspot selection. The DNA binding zinc finger domain of PRDM9 contains a variable amino acid sequence thereby giving rise to unique hotspot maps

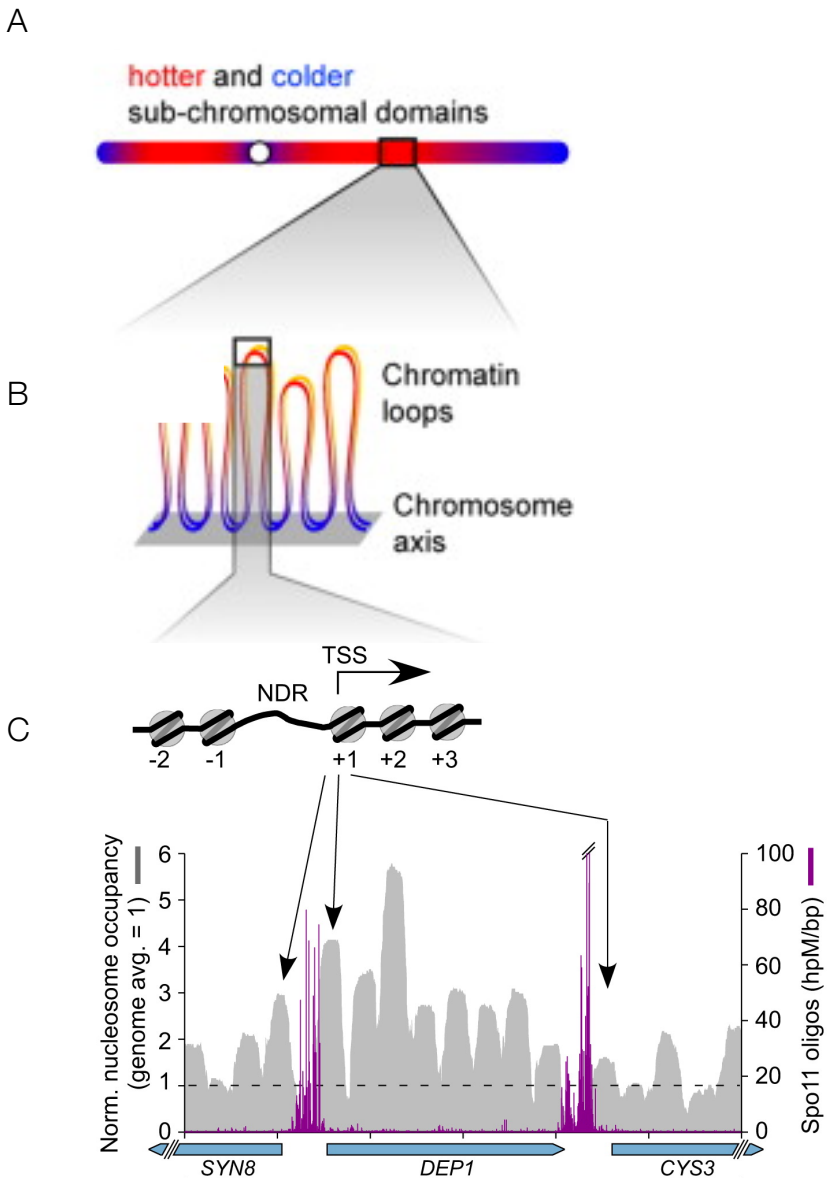


Figure 1.4: Spo11-DSB hotspot distribution.
A. Spo11 creates DSBs non-randomly throughout the genome, enriched at what are designated recombination hotspots. Telomere-proximal and pericentric regions are suppressed for DSB formation. **B.** Meiotic chromosomes are organised into protruding loops, attached to a proteinaceous axis with DSBs predominantly occuring on these loops. **C.** Within these loops Spo11-oligos (processed Spo11-DSBs) cluster in promoter nucleosome-depleted regions adjacent to the +1 nucleosome containing the transcription start site (TSS).
Adapted from Pan et al., 2011

within individuals bearing differing *PRDM9* alleles (Baudat et al., 2010; Parvanov et al., 2010).

In meiosis, sister chromatids condense by being organised into protruding loops, attached to a proteinaceous axis (as shown in Figure 1.4B), a feature distinct from mitotic chromosomes. The axis comprises of axial elements; the cohesin complex containing the meiosis-specific subunit, Rec8, and the meiosis-specific Hop1 and Red1 proteins (Blat and Kleckner, 1999; Hollingsworth et al., 1990; Klein et al., 1999; Panizza et al., 2011; Smith and Roeder, 1997; Zickler and Kleckner, 2015). DSB hotspots predominantly occur within these loops, however many of the proteins responsible for the activation and regulation of Spo11-induced DSBs reside at the axis (Blat et al., 2002; Borde and de Massy, 2013; Kleckner, 2006). A model known as the tethered-loop axis model (as depicted in Figure 1.5) has been proposed to counteract this discrepancy. This model states that the Spp1 subunit of the Set1 complex simultaneously associates with axis-associated protein Mer2 and with H3K4me3 sites, which are enriched at hotspots on the loop. Spp1 therefore acts as a bridge to tether the protruding loop to the axis where the DSB can occur (Sommermeyer et al., 2013).

In *S. cerevisiae,* in principle any base pair can be a site of a DSB in meiosis. However, there are approximately 3600 regions, termed hotspots, where DSBs more frequently occur (Lichten and Goldman, 1995). However, in a single cell only 150-200 DSBs are made (Pan et al., 2011). Therefore, mechanisms must be in place that limit and evenly distribute DSBs throughout the genome (reviewed in Keeney et al. 2014; Cooper et al. 2014). As DSBs in meiosis repair off a homologous chromatid, it would seem unwise to break the homologous sequence after the first DSB has been made. A DSB on one chromosome prevents a DSB forming in the same location on the sister chromatid, and has been seen to decrease the frequency of a DSB forming on its homolog. This is known as *trans* inhibition and is dependent on the DNA damage response (DDR) kinases $Mec1^{ATR}$ and $Tel1^{ATM}$ (Fukuda et al., 2008; Rocco and Nicolas, 1996; Xu and Kleckner, 1995; Zhang et al., 2011). Inhibition of further DSB formation in close proximity to an initial DSB on the same molecule (*cis* inhibition) was previously reported (Fan et al., 1997; Fukuda et al., 2008; Wu and Lichten, 1995; Xu and Kleckner, 1995), with a recent study finding this inhibition on the same chromatid, dubbed DSB interference, is regulated by Tel1 (Garcia et al., 2015). Tel1 has a localised, suppressive effect, which spans over ~70-100 kb on the same chromatid, reducing the frequency of coincident DSB formation below that expected by chance. As well as DSBs now forming independently of one another, DSBs also group within ~7.5 kb at a greater coincidence than random independent DSBs would. This indicates that the DSBs are somehow being formed in a concerted manner, with the authors explaining this phenomenon by stating that these local DSBs form together due to them being on the same loop of chromatin with

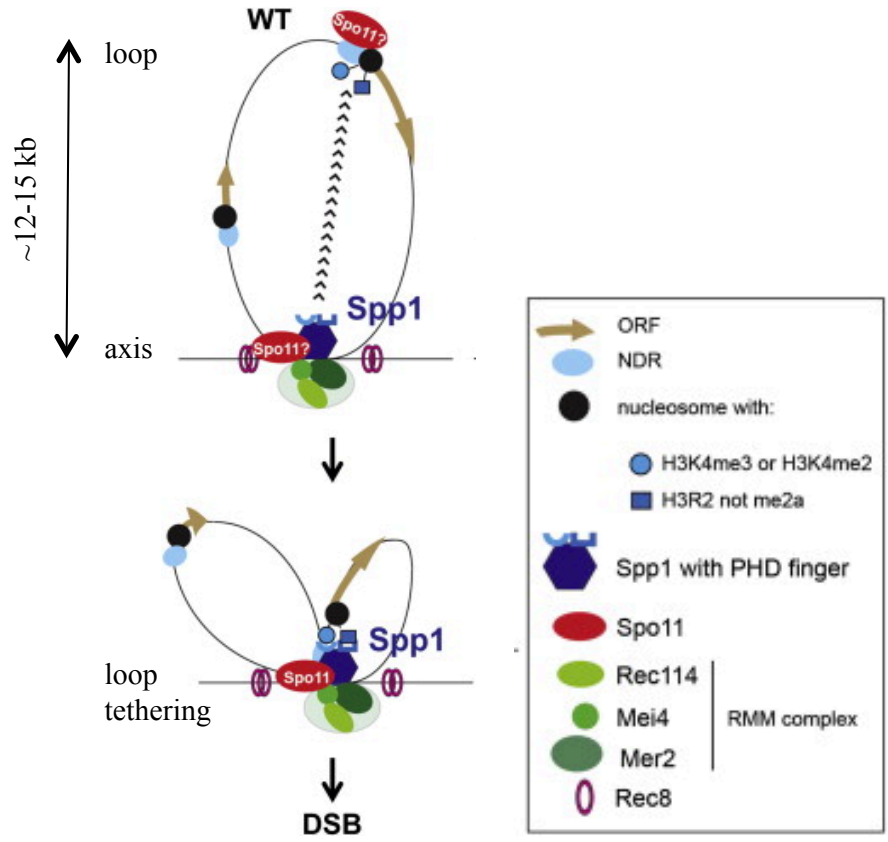


Figure 1.5: Meiotic loops and axis and the tethered-loop axis model.
Spo11-DSBs predominantly within the protruding loops that are formed in meiosis. The PHD domain of Spp1, which is located on the axis with Mer2, identifies the H3K4me3 on the protruding loop and bridges the site down to the axis where DSB forming proteins are found. Due to this sporadic long-range interaction, as assessed by ChIP, Spo11 may appear to be located at either the hotspot site or the axis.
Adapted from Sommermeyer et al., 2013

pre-tethering of the loop, upstream of DSB formation, 'activating' it and allowing DSBs to form close by on the same loop (Garcia et al., 2015).

2.5.4 Genome-wide mapping of Spo11 DSBs

Genome-wide mapping of Spo11-DSBs has given us vast amounts of information on where Spo11 creates breaks. Initially, DNA samples enriched for meiosis-specific DSBs from *rad50S* strains were used as hybridisation probes on DNA microarrays and it was observed that hotspots were nonrandomly associated in G/C rich regions and certain transcriptional profiles, whilst coldspots were more associated with telomeres and centromeres (Gerton et al., 2000). Later studies observed that in a *dmc1*Δ background, more DSBs were formed than observed in *rad50S* mapping data and that some DSBs were artificially lower in some regions in *rad50S* (Blitzblau et al., 2007; Buhler et al., 2007). In these studies the ssDNA produced from resection of Spo11-DSBs was used to map hotspots in the *dmc1*Δ background. Here, Spo11-DSBs had been processed and formed in a wild type scenario but ssDNA accumulates due to *dmc1*Δ cells lacking strand invasion and thus completion of HR. Double the number of hotspots was observed compared to the *rad50S* mapping data and a lot of the previously described cold-regions now had detectable ssDNA from DSBs (Blitzblau et al., 2007; Buhler et al., 2007). However, these studies still had limited quantitative and spatial resolution due to the design of the microarrays, large sizes of DNA used as probes and low dynamic range of hybridisation signal. They were also performed in mutant backgrounds where DSBs are not fully repaired, which may influence later forming DSBs due to various checkpoint signals from the large amounts of ssDNA present. Therefore, the Keeney lab designed a method to map the Spo11-oligonucleotides produced in wild type cells (more details on Spo11-oligos later) (Pan et al., 2011c). Their map confirmed DSBs form in nucleosome-depleted regions, mainly in intergenic regions, but a significant number within ORFs. Their DSB sites correlated with the previous mapping work done in *rad50S* and *dmc1*Δ cells but gave much higher resolution. For example, suppressed regions for formation of DSBs, such as telomeres and pericentric regions (see Figure 1.4A), were shortened compared to previous studies. Spo11-oligos were also observed to negatively correlate with Rec8 sites, a meiosis-specific cohesin subunit at the axis (Pan et al., 2011). However, due to the G-tailing required for preparing the Spo11-oligonucleotides for sequencing there is still a potential discrepancy for any Spo11 cleavage sites where there is a 5′ C present. Additionally, the short oligonucleotide could not be mapped due to potential multi-mapping that would occur with reads this short. This issue also arises for *S. pombe* when trying to map the short (13-14 nt) Rec12 (Spo11)-oligos (Fowler et al., 2014). This leaves a requirement for further high-resolution methods for mapping of Spo11 (Rec12)-DSBs.

2.6 DSB end-processing

2.6.1 Spo11-oligonucleotide formation

In order for Spo11-DSBs to be repaired by HR, the Spo11 moiety has to be removed to allow proper homology search and strand invasion. Two mechanisms of Spo11 removal from the 5′ ends of the DSB are possible: (i) direct hydrolysis of the covalent bond between Spo11 and the 5′ end or (ii) single-stranded nucleolytic cleavage of Spo11 bound to a short oligonucleotide (Keeney 2001; Keeney et al. 1997). Radioactive end-labelling of DNA in an immunoprecipitated Spo11-HA protein sample, with species resolved on an SDS-PAGE gel, revealed two bands, specific for Spo11-HA. These species were not present in the catalytic-dead Spo11 mutant (*spo11-Y135F*) or in *rad50S* or *sae2*Δ mutants, indicating that these species were Spo11-formation and DSB processing dependent (Neale et al., 2005). Protease digestion of Spo11 and the resolution of the attached oligonucleotides on a urea-PAGE gel revealed an upper band ~24-40 nt in length and a lower band at ~10-15 nt (Neale et al., 2005). These two discrete sizes of Spo11-oligonucleotide species have been observed in a variety of other species, including mouse, which indicates this mechanism for processing of Spo11-DSBs is evolutionarily conserved (Neale et al., 2005). However, in *S. pombe,* Rec12 (Spo11) is endonucleolytically released by Rad32 (Mre11) nuclease activity, in cooperation with Ctp1 (Sae2/CtIP), attached to only one length of oligonucleotide (~13-29 nt), suggesting a different nucleolytical regulatory mechanism to that seen in mammalian and *S. cerevisiae* cells (Milman et al., 2009).

As the size of these Spo11-oligos are shorter than the average length of the ssDNA tails produced via resection, this indicates that there is an initial clipping off of Spo11 attached to the oligonucleotide prior to extensive resection of the 5′ strand. However, the reason behind the production of two sizes of Spo11-oligo rather than one is still unknown, but it has been postulated to be due to asymmetric processing of the DSB with one side endonucleolytically cleaved further away than the other by Mre11 endonuclease activity (Neale et al., 2005). In support of this model a large number of hotspots were observed with the long oligonucleotides predominantly mapping to the Crick strand and not to the Watson strand (presumably the missing Spo11-oligos from the Watson strand consisted of the shorter length oligonucleotides which were not mapped in this study) (Pan et al., 2011). Additionally, Spo11-oligonucleotides are resistant to degradation by DNase, which may be explained by the multi-protein Spo11 complex forming an asymmetric barrier to resection and complete degradation of Spo11-oligos (Garcia et al., 2011).

2.6.2 The Mre11-Rad50-Xrs2 complex

The components of the MRX complex, whose structure can be seen in Figure 1.6, have been subject to numerous studies. The MRX complex has been implicated in multiple roles in the cell including DSB formation in meiosis (reviewed in Lam and Keeney, 2015), DSB resection at mitotic and meiotic breaks (Mimitou and Symington, 2008; Zhu et al., 2008), and $Tel1^{ATM}$ signalling through the Xrs2/Nbs1 subunit (Nakada et al., 2003). Its role in DSB formation, NHEJ and Tel1 signalling has been discussed in earlier sections. This section will delve into the DSB end processing activity of the MRX complex and its interplay between NHEJ and HR.

The MRX complex is composed of Mre11, Rad50 and Xrs2 proteins (Johzuka and Ogawa, 1995; Usui et al., 1998). The Mre11 and Rad50 subunits are highly conserved through all domains of life. Mre11 contains five conserved phosphodiesterase motifs, all of which are essential for both the exo- and endonuclease functions (Bressan et al., 1999; Moreau et al., 1999; Paull and Gellert, 1998; Trujillo et al., 1998; Usui et al., 1998). Rad50 is a large 153kDa protein which contains ATP-dependent binding to DNA (Alani et al., 1989; Raymond and Kleckner, 1993; Sharples and Leach, 1995). The MRX complex structure contains a head domain that consists of a dimer of Mre11 interacting with two Rad50 ATP-binding cassettes. This domain is responsible for the DNA-binding and the ATP-dependent nuclease activity of Mre11. MR complexes can interact inter-molecularly via the coiled-coil domain of Rad50, which emanates from this head domain and allows tethering of the DSB ends (Chen et al., 2001; De Jager et al., 2001; Hopfner et al., 2002), a process essential for meiotic DSB formation (Wiltzius et al., 2005). Loss of any component of the MRX complex, as well as Rad50 zinc hook mutants, exhibit an increase in separation of the two ends of an induced I-*Sce*I break, whereas Mre11 nuclease mutants did not, indicating these functions are separable (Lobachev et al., 2004).

As stated, Mre11 and Rad50 are conserved from bacteria to humans, however the third component of this complex (Xrs2) is more diverged in eukaryotes and doesn't exist in prokaryotic bacteria, archaea and the T4 bacteriophage. Yeast Mre11 and Rad50 proteins complex with Xrs2 whilst the mammalian component is Nbs1. Xrs2 and Nbs1 are less well conserved, with similarity limited to two regions of the proteins: the N-terminal fork-head associated (FHA) domain, and a conserved region in the C-terminus (Chahwan et al., 2003). Xrs2 displays some DNA binding activity (Trujillo et al., 2003), interacts with Mre11 via a C-terminal region but not Rad50 (Usui et al., 1998), and translocates Mre11 into the nucleus (Tsukamoto, 2004). All the components of the MRX complex, and these activities of Xrs2, are essential for all functions of the MRX/N complex *in vivo* (Tsukamoto, 2004). However, it

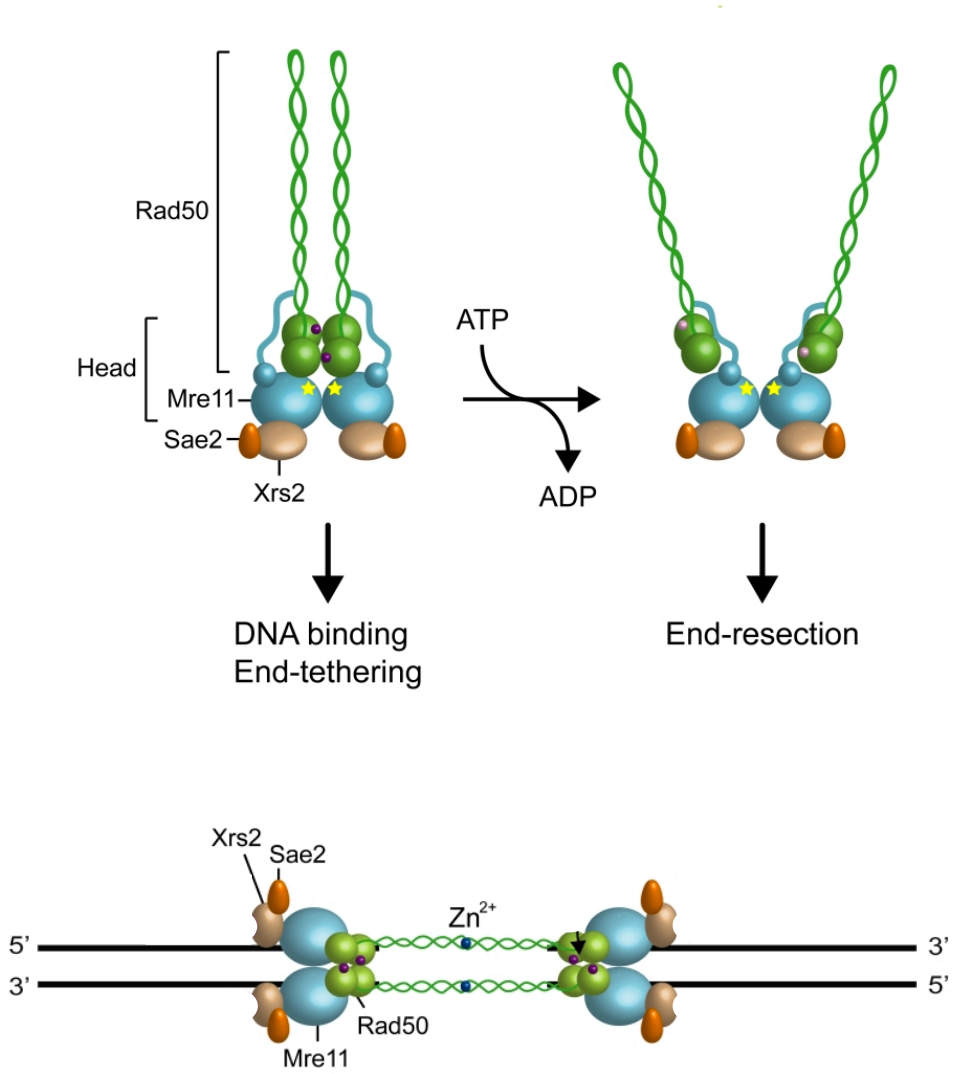


Figure 1.6: Structural organisation of the MRX complex and its DSB end tethering ability.
The MRX complex consists of a dimer of Mre11 interacting with two Rad50 ATP-binding cassettes forming a head domain. When ATP (purple dots) is bound to Rad50, it alters the conformational shape of the complex into a closed state that allows DNA binding and end-tethering but occludes the Mre11 nuclease domains (stars) (top left). Upon ATP hydrolysis to ADP (pink dots), the Rad50 conformational is altered leading to disengagement of the Rad50 dimer and an exposure of the Mre11 nuclease domains, promoting end-resection. In the closed state, the coiled-coil domain of Rad50 that emanates from the head domain allows tethering of two MRX complexes via the zinc hook of Rad50 thus tethering two DNA ends together.
Adapted from Gobbini et al 2016

seems that the translocation of Mre11 into the nucleus by Xrs2, as well as its Tel1 interaction, are the major roles of the Xrs2 protein. Mutating the Mre11 interaction motif of Xrs2 resulted in an *xrs2*Δ phenotype (Tsukamoto, 2004). A recent study demonstrated that tethering a nuclear localisation signal (NLS) to Mre11 rescued *xrs2*Δ defects in DNA end resection, meiosis and hairpin resolution. This strain did however retain defects in $Tel1^{ATM}$ signalling and repair of DSBs by NHEJ (Oh et al.,, 2016). Mutations in the conserved regions of the *XRS2* gene elucidated various functional domains of the Xrs2 protein: (i) an Mre11 interaction domain, essential for the roles of the MRX complex (Falck et al., 2005; Nakada et al., 2003; Tsukamoto, 2004), (ii) a Tel1-interaction domain, which when mutated resulted in a *tel1*Δ-like phenotype of defective DNA damage signalling and shortened telomeres (Nakada et al., 2003; Shima et al., 2005), and (iii) an FHA domain, which in contrast to the mammalian Nbs1 protein was not essential for any of the major roles of the MRX complex (Shima et al., 2005), but binds phosphorylated Sae2 (CtIP) (Liang et al., 2015). In mammalian cells the Xrs2/Nbs1 subunit is a critical component as mutations within the *NBS1* gene are responsible for the rare autosomal recessive disease, Nijmegen breakage syndrome, which increases cancer predisposition (Tauchi et al., 2002; Williams et al., 2002; Zhu et al., 2001).

2.6.3 MRX nuclease functions

It was first seen that meiotic recombination and DSB resection is inhibited by *rad50S* and *mre11* nuclease mutants, so it was suggested that the MRX complex is involved in the processing of Spo11-DSBs (Ajimura et al., 1993; Alani et al., 1990; Keeney and Kleckner, 1995; Liu et al., 1995; Moreau et al., 1999). Indeed, in the *rad50S* background, Spo11-oligo species are not generated. Moreover, in wild type cells the Spo11-oligos generated contained a free 3′ hydroxyl group indicative of endonuclease cleavage (Neale et al., 2005). Mre11 contains endonuclease activity as well as 3′ to 5′ exonuclease activity, and in a nuclease-dead mutant of Mre11 (*mre11-H125N/mre11-nd*), although DNA binding of the Mre11 mutant was unaffected (Nicolette et al., 2011), Spo11-oligos were no longer generated (Garcia et al., 2011; Moreau et al., 1999; Stracker and Petrini, 2011). Likewise, in *S. pombe* the nuclease activity of Rad32 (Mre11) is essential for the removal of Rec12 (Spo11) (Hartsuiker et al., 2009). In these *mre11*-nuclease and *rad50S* mutants, DSBs were still present; indicating the role MRN plays in promoting DSB formation is distinct from its nuclease activities.

The nuclease activity of Mre11 was first postulated because Mre11 has sequence homology with the bacterial SbcD nuclease (Moreau et al., 1999). The 3′-5′ exonuclease activity of recombinant Mre11 is stimulated by Rad50 and is dependent on the divalent manganese cation, whilst being inhibited by magnesium (Paull and Gellert, 1998). A weak endonuclease

activity was also seen on 5′-terminated DNA strands and other structures (Hopkins and Paull, 2008; Trujillo and Sung, 2001). Through structural studies it was seen that the catalytic domain of the protein could be disrupted through mutation of *D56N* and/or *H125N*, resulting in a separation of function mutant (Moreau et al., 1999). This mutant, unlike the null mutant, was proficient in mating-type switching, telomere maintenance, classical NHEJ, DSB formation in meiosis and repair of HO-DSBs (Krogh et al., 2005; Moreau et al., 1999). However, the *mre11*-nuclease mutant accumulates DSBs in meiosis (like the *rad50S* allele), is sensitive to topoisomerase poisons and is mildly sensitive to methyl methane sulfonate (MMS) (Krogh et al., 2005).

For many years there was an inherent contradiction in models for DSB repair and the observed exonuclease activity of Mre11. Upon repair of a DSB, long ssDNA tails are generated which are 3′ ending. This means that 5′ to 3′ resection had occurred from the DSB but Mre11 possesses 3′ to 5′ exonuclease activity. This contradiction was solved from work in the Neale lab using a mutant of Mre11 that was deficient in exonuclease activity but proficient in endonuclease activity (*mre11-H59S*), a mutant discovered in *Pyrococcus furiosus* (Williams et al., 2008). This mutant still generated Spo11-oligonucleotides and seemed to have no obvious resection defect of its own. However, combining the *mre11-H59S* mutant with deletion of Exo1, the major 5′ to 3′ resector of DSBs in meiosis (Hodgson et al., 2011; Keelagher et al., 2011; Manfrini et al., 2010; Tsubouchi and Ogawa, 2000; Zakharyevich et al., 2010), caused DSBs to accumulate for longer than in either of the single mutants and resection was reduced and delayed. Spo11-oligonucleotides, which still formed in *mre11-H59S* cells, had altered size distribution with an increase in the levels of longer Spo11-oligonucleotides up to 300 nt in length. These observations supported a hypothesis that Mre11 creates a nick via its endonuclease activity then the 5′ strand is resected bidirectionally with the inherent 3′ to 5′ Mre11 exonuclease resecting towards the DSB and Exo1 in a 5′ to 3′ direction to extend the 3′ ssDNA tail required for meiotic recombination (Garcia et al., 2011) (see Figure 1.7).

A similar bidirectional model has been proposed in mammalian cells for DSBs generated from IR (Shibata et al., 2014). In this study it was found that the endonuclease function of Mre11 was a control point for the choice between HR repair and NHEJ repair. Inhibition of Mre11 endonuclease function, with specific inhibitors, increased repair of DSBs via NHEJ, suggesting that activation of Mre11 endonuclease, potentially through Sae2 (CtIP) (discussed later), begins resection of the 5′ strand, inhibiting NHEJ repair and promoting repair by HR (Shibata et al., 2014).

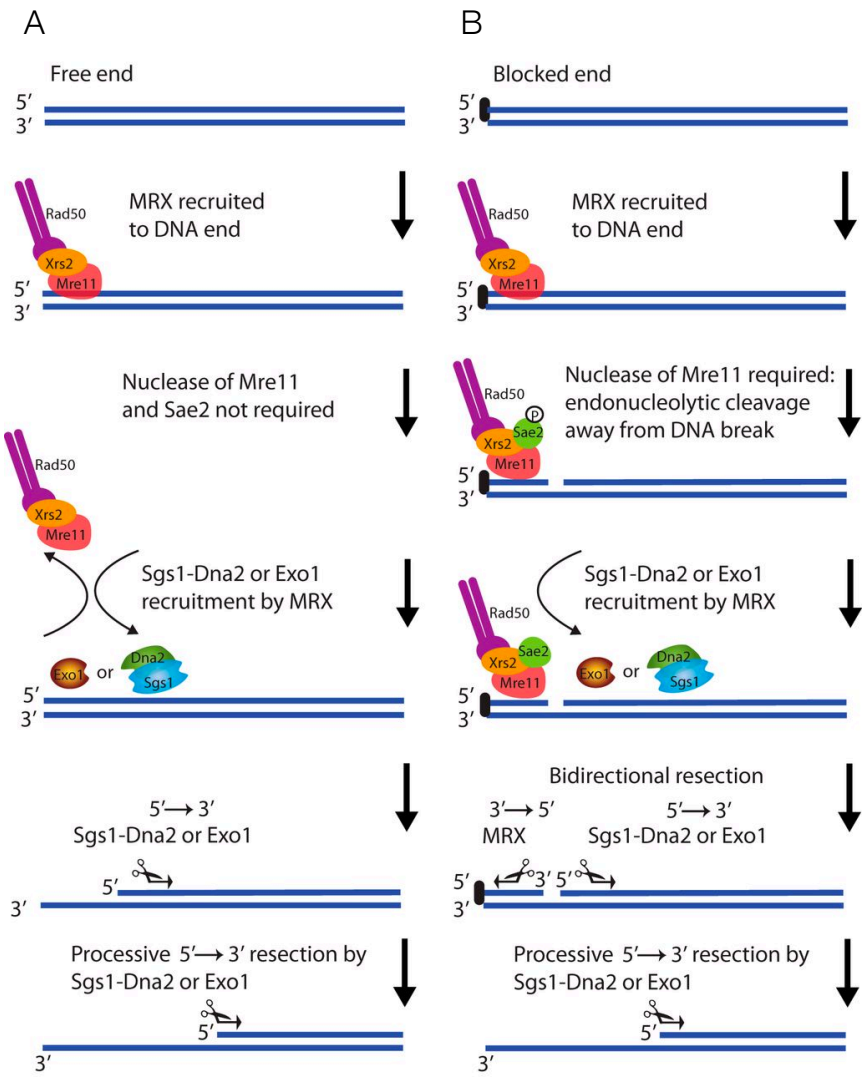


Figure 1.7: The dependency on MRX nuclease activity at free or blocked DSB ends.
Even though the MRX complex is recruited to most DSB ends the requirement for MRX nuclease activity differs depending on the complexity of the DSB ends. **A.** When the DSB end is free (clean) MRX is rapidly recruited but its nuclease activity is not essential. However, it does speed up the process of resection, potentially through creating a more efficient DNA substrate for binding of Exo1. **B.** When the end of a DSB is complex (blocked), such as those containing a chemical modification, stalled topoisomerase protein or covalently bound Spo11, the nuclease activity of MRX is essential. MRX and Sae2 are rapidly recruited, Sae2 is phosphorylated (in G2/M or meiosis) and Mre11 endonuclease activity is stimulated. The 5′ strand is cleaved and from this nick bidirectional resection occurs through the 3′ to 5′ exonuclease activity of Mre11 and 5′ to 3′ exonuclease activity of Exo1 or Sgs1-Dna2.
Figure taken from Cejka, 2015

All these experiments indicate that the nuclease function of Mre11 is essential for blocked or complex DSB ends like those created by Spo11 in meiosis and by stalled Top2 proteins. These ends need to be nucleolytically processed by Mre11 before other DSB repair machinery, such as long range 5′ to 3′ exonucleases, can access the DSB. However, at 'clean' ended DSBs, such as at those created by the I-*Sce*I endonuclease or the HO endonuclease (an unrepairable single DSB generated by the HO-endonuclease at the MAT locus of JKM139 derivative (Lee et al., 1998)), Mre11 nuclease function is not as important with initiation of resection being simply delayed in *mre11*Δ*, rad50*Δ and *xrs2*Δ mutants (Mimitou and Symington, 2009). This suggests these ends lacking chemical modifications are more liable/accessible to 5′ to 3′ exonucleases to allow repair (Llorente and Symington, 2004; Moreau et al., 1999). Therefore, the necessity for Mre11 nuclease action is dependent on how well the long-range resection machinery, Exo1 or Sgs1-Dna2, can access the DNA (see Figure 1.7 for a schematic comparison between MRX activity on 'complex' and 'clean' ends). The nuclease functions of Mre11 are not dependent on Rad50 or Nbs1/Xrs2 but they are enhanced when in a holocomplex (Williams et al., 2007).

At clean-ended breaks, the interplay between MRX and Ku can also control the choice of DSB repair down either the NHEJ or HR route. Both MRX and Ku display almost simultaneous, but independent, binding to the DSB end (Wu et al., 2008). Ku can only bind to free DNA ends and, once there, protects them from degradation and promotes repair by NHEJ. Ku has been shown to block free ends from resection by Exo1 when the MRX complex is defective, suggesting the importance of MRX in promoting HR. Likewise, MRX was shown to dissociate Ku from the end of the DSB (Shim et al., 2010; Wu et al., 2008). Deletion of *YKU70* led to increased resection, even in the absence of functional MRX (Mimitou and Symington, 2010). Similarly, over-expression of *EXO1* suppressed sensitivity to end-processing mutants (*sae2*Δ *sgs1*Δ) (Lewis et al., 2002; Moreau et al., 2001). These results suggest that Ku inhibits HR by blocking access of Exo1 to the DSB and that MRX is required to remove Ku to allow HR to occur (Bressan et al., 1999; Clerici et al., 2008; Mimitou and Symington, 2010).

2.6.4 Rad50 roles

Rad50 activity, and its ability to regulate MRX activity in DNA damage signalling and resection, are dependent on ATP binding and hydrolysis, which induces conformational changes (see Figure 1.6) (Deshpande et al., 2014; Lammens et al., 2011). The ATPase activity of Rad50 is essential for its function, as mutations in the conserved Walker A-type ATPase domain results in a *rad50*Δ phenotype (Chen et al., 2005). Similarly, the *rad50S* mutant, whose phenotype was identical to *mre11-nd* and *sae2*Δ*,* has a mutation close to this ATPase

domain; however, the exact biochemistry of the *rad50S* mutation has not been characterised (Cao et al., 1990). Recent work on the MRX complex has shown Rad50 to be essential for Mre11 endonuclease activity and that the *rad50S* mutant lacks the capacity to stimulate this activity *in vitro* (Cannavo and Cejka, 2014). This endonuclease activity is also dependent on ATP suggesting that the *rad50S* allele may be defective in ATP hydrolysis and/or at stimulating the Mre11 endonuclease activity, potentially via modulating Sae2 interaction with the rest of the MRX complex (Cannavo and Cejka, 2014). Similarly, in *S. pombe* the equivalent *rad50S* mutant is also defective in removing Rec12 (Spo11) (Hartsuiker et al., 2009).

2.6.5 Mre11 association at the DSB

MRX localises first to the site of damage, tethers the ends and acts as a sensor, enabling recruitment and activation of the DNA damage checkpoint through $Tel1^{ATM}$ (D'Amours and Jackson, 2002; Dolganov et al., 1996; Hopfner et al., 2000; Petrini et al., 1995). Preventing end-processing of DSB ends by disrupting MRX function through either *rad50S* or *sae2*Δ mutations resulted in a delay in the disassembly of Mre11 foci. In meiosis, Mre11 foci accumulate in the *rad50S* background but are not dependent on Tel1, suggesting that this accumulation of Mre11 at the DSB is Tel1-independent (Usui et al., 2001). Interestingly, the nuclease-deficient *mre11-58* (*His213Tyr)* mutant and *mre11-58 rad50S* double-mutant are capable of forming DSBs in meiosis but do not have persistent Mre11 foci, unlike wild type and other nuclease mutants (Usui et al., 1998, 2001). The phenotypes associated with the *mre11-58* mutant has been linked to the loss of interaction with Rad50 (Moreau et al., 1999), but, as DSBs still form, a loss of this interaction is unlikely (Usui et al., 2001).

The absence of Sae2 also causes prolonged binding of MRX at DSB sites leading to persistent DNA-damage checkpoint activation by Tel1 and Rad53 (Clerici et al., 2006; Lisby et al., 2004). *MRE11* nuclease-dead and *sae2*Δ mutants are defective in dissociation of Mre11 from unprocessed DSBs in meiosis (Borde et al., 2004), as well as DSBs from IR (Lisby et al., 2004). This prolonged binding and accumulation of Mre11 (MRX) at the DSB site is thus thought to be due to the lack of Mre11 nuclease activity itself and also lack of its stimulation by Sae2. However, Mre11 persists for longer at DSBs in *sae2*Δ cells compared to *mre11* nuclease-deficient cells, as well as the DNA damage checkpoint being hyperactivated in *sae2*Δ. (Clerici et al., 2006; Lisby et al., 2004; Mimitou and Symington, 2011). Prolonged retention of Mre11 in *sae2*Δ mutants, as well as the defective checkpoint deactivation in *sae2*Δ cells, can be rescued by some Mre11 mutations in its N-terminal region. For example, two separate Mre11 mutants, *H37R* and *P110L*, decrease both DNA binding and retention to DSBs

in a *sae2*Δ background. These mutants retain their meiotic DSB-forming and nuclease abilities, indicating these activities are separable (Chen et al., 2015; Puddu et al., 2015). These studies suggest that it is the accumulation of Mre11 (MRX) at the DSB sites, in a *sae2*Δ background where ends are not processed, that cause activation of the DNA damage checkpoint and sensitivity to DNA damaging agents. Alongside this, over-expression of *SAE2* results in a faster turnover of Mre11 at DSBs and correlates with a reduction in Rad53 phosphorylation (Clerici et al., 2006). These results suggest that a major function of Sae2 is to remove the MRX complex from break ends, an independent function from its MRX nuclease-stimulating ability.

TEL1 inactivation does not suppress the DNA damage hypersensitivity of *sae2*Δ (Puddu et al., 2015). It does, however, reduce the ability of the *mre11-H37R* mutant to suppress *sae2*Δ CPT hypersensitivity. Furthermore, the *mre11-H37R* mutant, which does not accumulate at the DSB, now forms foci in *TEL1*-deficient cells suggesting Tel1 cooperates with Sae2 to promote the removal of Mre11 from DSBs (Puddu et al., 2015). A recent genetic screen for mutants that require Tel1 for survival to DNA damaging agents isolated the *rad50-V1269M* mutant (Cassani et al., 2016). This mutation resulted in hypersensitivity in *tel1*Δ cells to phleomycin and CPT, and impaired MRX association at the DSB, which is further reduced in the absence of Tel1. These results suggest that loss of MRX, and the repair defect, are due to loss of end-tethering and that this tethering effect is necessary for repair of DSBs by either NHEJ or HR. Thus, Tel1 may promote/stabilise the MRX complex at the DSB in a positive feedback loop. Intriguingly, this role of Tel1 was independent of its kinase activity (Cassani et al., 2016).

The MRX complex also plays a role alongside Tel1 in maintenance of telomere length (Hirano et al., 2009; Martina et al., 2012), whereas the protein Rif2 negatively regulates telomere length (Wotton and Shore, 1997). Rif2 appears to compete with Tel1 for binding the Xrs2 subunit *in vitro,* and Rif2 interferes with MRX-Tel1 interaction, preventing Tel1-dependent lengthening of telomeres (Hirano et al., 2009). Rif2 also interacts with Xrs2 at DSBs *in vivo* (Hirano et al., 2009), preventing the promotion of MRX retention at the break site by Tel1, and, similar to a *tel1*Δ, impedes on the MRX ability to end-tether and repair DSBs (Cassani et al., 2016). Rif2 also enhances ATP hydrolysis by Rad50 with *rif2*Δ cells showing an increase in end-tethering and NHEJ (Cassani et al., 2016), indicating that Rif2 counteracts these processes and that Rif2 may play a role in the choice between NHEJ and HR by modulating the open and closed state of Rad50 (see Figure 1.6). Specifically, it is proposed that an open state would increase end resection thereby favouring HR over NHEJ (Cassani et al., 2016).

2.6.6 Phosphorylation of the MRX complex by Tel1

All three members of the MRX/N complex are phosphorylated by ATM/ATR in response to DNA damage (D'Amours and Jackson, 2001; Grenon et al., 2001). Phosphorylation of each component has different cellular effects. Phosphorylation of Nbs1 at two sites (Ser278 and Ser343) in response to DNA damage mediates control of the S phase checkpoint and plays a role in maintenance of telomeres (Lavin et al., 2015). Only one site of mammalian Rad50 has been observed to be phosphorylated in response to DNA damage, Ser635. Phosphorylation of this site doesn't affect MRN complex formation but it was important for MRN function because the *S635G* phospho-mutant is not able to correct the S phase checkpoint in Rad50-deficient cells (Gatei et al., 2011). Finally, exposure of cells to IR, and other agents that induce DSBs in mammalian cells, identified phosphorylation sites on Mre11 at Ser676 and Ser678, which, when mutated, lead to decreased cell survival and increased chromosomal aberrations (Kijas et al., 2015). Further investigation found that hyperphosphorylation of Mre11 by ATM led to inactivation of the MRN complex via decreasing its association from chromatin, enabling down-regulation of the DNA damage checkpoint (Di Virgilio et al., 2009)

The MRX complex and Sae2

The idea that Sae2 modulates Mre11 activity and disassembly from the DSB is supported by the observation that Sae2 foci form at the precise time Mre11 foci disassemble, and that Mre11 foci persist similarly in the *mre11-H125N* (nuclease deficient) mutant as to that seen in *sae2*Δ cells (Lisby et al., 2004). However, Mre11 foci persist longer in *sae2*Δ and *mre11-H125N* cells at DSBs with apparent 'dirty/complex' ends compared to clean-ended I-*Sce*I DSBs, suggesting that the nuclease activity of Mre11 is more important for the processing of complex ends, such as those generated in meiosis by Spo11. This also suggests that alternative nucleases can compensate for Mre11 nuclease activity when the DSB end is clean (Lisby et al., 2004).

2.7 Sae2

The screens that identified an accumulation of Spo11-DSBs in meiosis in *rad50S* and *mre11S* mutants also identified *SAE2/COM1* deficient cells with the same phenotype of unresected DSBs, deficient HR and homolog synapsis, and weak sensitivity to MMS (McKee and Kleckner, 1997; Prinz et al., 1997). This suggested that the role of Sae2 intersected with that of the MRX complex.

The *SAE2* gene is not well conserved in eukaryotes but there have been related proteins identified in humans (CtIP), *S. pombe* (Ctp1), *A. thaliana* (COM1), and *C. elegans* (COM-1) (Deveaux et al., 2000; Gönczy et al., 1999; Limbo et al., 2007; Penkner et al., 2007). Sae2

contains a short, C-terminal region that displays sequence alignment to these functional orthologs (reviewed in Tsutsui et al. 2011). This region includes sites for CDK and $Mec1^{ATR}$/$Tel1^{ATM}$ phosphorylation as well as an oligomerisation motif (Kim et al., 2008; Sartori et al., 2007; Wang et al., 2012). The essential nature of Sae2 and its orthologs in meiosis is replicated through *S. pombe* to higher eukaryotes (Hartsuiker et al., 2009; Penkner et al., 2007; Uanschou et al., 2007). For example, the distantly related *S. cerevisiae* and *S. pombe* species show functional conservation between Sae2 and Ctp1, as seen by the requirement for Ctp1 in Rec12 (Spo11) removal in meiosis (Hartsuiker et al., 2009).

2.7.1 Sae2 protein domains

The Sae2 protein C-terminus contains a conserved region, which is essential for the stimulation of DSB end-processing by MRX and is where many essential phosphorylation sites reside (Cannavo and Cejka, 2014; Kim et al., 2008). The N-terminus seems to be somewhat dispensable as overexpression of *sae2*-ΔN_{120} or *sae2*-ΔN_{170} can support sporulation in *sae2*Δ cells, whereas the overexpression of *sae2*-ΔC_{170} cannot. However, the overexpression of *sae2*-ΔN_{170} does not suppress MMS sensitivity, and the self-interacting mutant of Sae2 (*L25P-E171G*) confers only partial suppression, suggesting that the N-terminal region has a function other than just self-interaction (Kim et al., 2008).

2.7.2 Interaction with MRX

The similarity between *SAE2* mutants and *MRE11* nuclease mutants suggest that Sae2 functionally integrates with the MRX complex (Keeney and Kleckner, 1995). To investigate this in detail the Cejka lab directly demonstrated that recombinant Sae2 stimulates Mre11 endonuclease activity on *in vitro* blocked-end DNA substrates, whilst Mre11 exonuclease activity was independent of Sae2. This Mre11 endonuclease activity showed a preference for the 5′ blocked end of dsDNA substrates and cleaved 15-25 nt from the end (Cannavo and Cejka, 2014). This observation may explain why Spo11-oligonucleotides are the observed lengths – due to endonucleolytic clipping of the 5′ strand flanking the DSB. The 3′ to 5′ exonuclease function of Mre11 may therefore only be required for further resection of the 5′ strand by multiple nicking, up to 300 nt from the DSB, followed by resection towards the DSB (Garcia et al., 2011).

With mutants of Sae2 and the MRX complex conveying similar phenotypes, it was thought that they should interact physically. It was initially reported that via yeast two-hybrid and *in vitro* assays that they did not physically interact in solution unless DNA was present

(Lengsfeld et al., 2007; Uetz et al., 2000). More recently it was shown that recombinant Sae2 is able to pull-down the MRX complex via the Mre11 and Xrs2 components (Cannavo and Cejka, 2014). Conversely, in mammalian cells, CtIP interacts with all three members of the MRN complex (Chen et al., 2008; Sartori et al., 2007; Yuan and Chen, 2009). Over-expression of *SAE2* is able to suppress the SSA defect in *rad50S* cells (Clerici et al., 2005), suggesting that the *rad50S* mutant prevents interaction of Sae2 with the MRX complex, which can be overcome with high concentrations of Sae2 (reviewed in Paull, 2010). However, although the *rad50S* mutant has a similar phenotype to *sae2*Δ*,* the hypothesis that the *rad50S* mutation prevents Sae2 association with the MRX complex and/or the DSB is not supported by experiments demonstrating that Sae2 foci are not disrupted in *mre11*Δ or *rad50S* mutants (Lisby et al., 2004).

Sae2, although essential for processing of meiotic breaks, is dispensable for some mitotic HR processes (Bressan et al., 1999; Ivanov et al., 1994; Malkova et al., 1996; Tsubouchi and Ogawa, 1998). Sae2 deficiency, similar to MRX deficiency, causes a delay of approximately 30 minutes before resection initiates at HO-endonuclease induced DSBs, resulting in an increased frequency of NHEJ repair. This indicates that the initial processing by MRX and Sae2 is a driving force towards HR rather than NHEJ (Lee et al., 2008; Mimitou and Symington, 2008). Loss of Sae2 activity also confers hypersensitivity to MMS (McKee and Kleckner, 1997; Rattray et al., 2001), and enhanced Tel1-mediated Rad53 phosphorylation after DNA damage (Usui et al., 2001). By contrast, high levels of Sae2 cause telomere lengthening in a Tel1-dependent manner (Perrone,et al., 2003). Sae2 is likely to have additional roles in mitotic cells as *sae2*Δ mutants are more sensitive to DNA damaging agents than nuclease-deficient Mre11 mutants (Usui et al., 2001).

2.7.3 Hairpin processing

In Sae2-deficient (and also *rad50S* and *mre11* nuclease-deficient) cells, large palindromic duplications arise within inverted repeats due to misrepair of DSBs, thought to have arisen from fold-back and replication of 3′ ssDNA intermediates (Rattray et al., 2001; 2005). Unprocessed hairpin structures also arise at such inverted repeats in *sae2*Δ cells (Lobachev et al., 2002). Subsequent work demonstrated that Sae2 itself is an endonuclease, responsible for cleaving 5′ flap structures and at single-stranded regions adjacent to hairpin DNA (Lengsfeld et al., 2007). However, there is no discernable nuclease domain on the Sae2 protein, and more recently other laboratories have not detected this same endonuclease activity (Cannavo and Cejka, 2014; Niu et al., 2010).

2.7.4 CtIP/Ctp1

As previously stated, Sae2 and CtIP/Ctp1, although functional orthologs, only have a short stretch of sequence conservation found in the C-terminus. Both proteins seemingly play a similar role in activation of the Mre11 nuclease and are both required for DSB resection, but their biochemical interactions differ. For example, CtIP and Ctp1 require MRN for localisation to a DSB (Limbo et al., 2007), whereas Sae2 can localise independently of the MRX complex (Lisby et al., 2004). CtIP has also been shown to directly interact with the MRN complex (Sartori et al., 2007), and it was not until recently that Sae2 was found to interact weakly with the MRX complex (Cannavo and Cejka, 2014). CtIP is also much larger than Sae2 (897 a.a. vs. 345 a.a.), with the only defined region of Sae2 being a coiled-coil region in the N-terminus that mediates homodimerisation (Dubin et al., 2004).

2.7.5 Regulation of Sae2 activity

Even though Sae2 is essential for processing meiotic DSBs and that is has important roles in other pathways, its actual mechanistic role is still largely unknown. To elucidate the functional role of Sae2, a substantial body of work has investigated the role of post-translational modification and the impacts of oligomerisation, with Sae2 being phosphorylated and acetylated after MMS treatment on numerous residues (Fu et al., 2014), as discussed below.

CDK phosphorylation of Sae2

As previously mentioned, Cdc28, the main CDK in *S. cerevisiae*, increases in activity in S/G2 phases of the cell cycle. Such activation leads to a wide range of substrates being phosphorylated, one of which is Sae2 (Huertas et al., 2008). HR repair, in contrast to NHEJ repair, is the more accurate repair method of DSBs, but it requires the presence of a sister-chromatid. In *S. cerevisiae*, HR therefore predominates in S and G2 phases of the cell cycle, whereas NHEJ is generally restricted to G1 (as described earlier) (Aylon et al., 2004; Caspari et al., 2002; Esashi et al., 2005; Grzegorz et al., 2004). CDK activity governs which repair pathway a DSB should take by phosphorylating substrates during S and G2 to favour HR repair, with a major control step being at the point of resection initiation. Promotion of resection of a DSB drives repair down the HR pathway because resected DNA is refractory to repair by NHEJ (Aylon et al., 2004; Grzegorz et al., 2004).

The Sae2 protein contains three potential CDK phosphorylation sites: Ser134, Ser179 and Ser267. Of these, Ser267 maps to the only region on Sae2 that is conserved with mammalian CtIP and fission yeast Ctp1 (Huertas et al., 2008). Mutation of the two sites not in this conserved region to unphosphorylatable alanines (*S134A* and *S179A*) did not alter the

sensitivity of strains to the topoisomerase I poison, camptothecin (CPT). By contrast, mutation of the conserved S267 site to alanine caused sensitivity almost to *sae2*Δ levels. Additionally, mutation of the cyclin-binding motif (Arg-X-Leu (RXL)) on Sae2, a motif essential for phosphorylation of substrates by CDK, also caused CPT sensitivity. In contrast, mutating the Ser267 site to glutamic acid to mimic phosphorylation (*S267E*) showed no sensitivity to CPT, and combining this *S267E* mutant with the cyclin-binding motif mutant reduced the sensitivity of the cyclin-binding mutant strains (Huertas et al., 2008). At a HO-DSB, *sae2-S267A* confers *sae2*Δ-like absence of end resection. By contrast, the *sae2-S267E* mutant displays wild type resection kinetics that are only slightly affected by inhibiting *cdc28-as1* (an inhibitable form of Cdc28), unlike wild type Sae2. However, the *sae2-S267E* mutant displays sensitivity to IR when the cells are in the G1 phase of the cell cycle (something that is not seen in *sae2*Δ or *sae2-S267A)*. One explanation for this observation is that the DSBs generated are resected and therefore inhibitory to repair by NHEJ, but the lack a sister chromatid to repair from results in aberrant attempts to repair by HR. The analogous CDK site in human CtIP (Thr847) when mutated to alanine also displays CPT hypersensitivity, indicating that CDK-dependent control of Sae2 activity and resection initiation may be conserved through other eukaryotes (Huertas et al., 2008). However, in fission yeast the Sae2 homolog, Ctp1, does not contain a CDK phosphorylation site analogous to Ser267. Instead CDK controls protein expression of Ctp1, such that it is cell-cycle regulated (Limbo et al., 2007).

CDK activity in meiosis rises during pre-meiotic S phase and increases through prophase I (Marston and Amon, 2004). Sae2 is also phosphorylated by CDK on Ser267 in meiosis, and the *sae2-S267A* mutant fails to repair meiotic DSBs similar to *sae2*Δ mutants (Manfrini et al., 2010). Phosphorylation at S134 may also have a role during meiosis because the double *sae2-S134A-S267A* mutant has lower spore viability than the *sae2-S267A* single mutant (Huertas et al., 2008; Manfrini et al., 2010). Similarly, a recent study noted a hypomorphic phenotype in the *S267A* mutant compared to a nearly null phenotype in the double *S134A-S267A* mutant, indicating that this S134 site may also be an, albeit minor, CDK phosphorylation site that regulates Sae2 activity (Fu et al., 2014). *In vitro*, phosphorylation of Sae2 on S267 is critical for its ability to stimulate MRX endonuclease activity (Cannavo and Cejka, 2014). The *S267A* mutant has no change in its ability to bind DNA, but has impaired interaction with the MRX complex, suggesting that the conserved C-terminal region (and S267 site) of Sae2 either directly interacts with the MRX complex or that this mutant protein is improperly folded (Cannavo and Cejka, 2014).

Mec1/Tel1 phosphorylation of Sae2

As discussed, Tel1 is recruited to, and binds to, DSBs via the Xrs2 protein of the MRX complex (Nakada et al., 2003), where it (and Mec1) phosphorylate many proteins associated

with the DSB including MRX and Sae2 (Baroni et al., 2004). Sae2 contains five S/T-Q sites, which are a consensus sequence for Mec1/Tel1 phosphorylation: Ser73, Thr90, S249, Thr279 and Ser289. Mutation of all five sites to prevent putative phosphorylation eliminated the majority of Sae2 phosphorylation when assessed by gel mobility assays (Baroni et al., 2004), and conferred phenotypes similar to *sae2*Δ in respect to MMS sensitivity, mitotic recombination at inverted repeats, and restart of cell cycle progression once the checkpoint is activated (Baroni et al., 2004). A more recent study noted that mutation of Thr90, Ser249 and Thr279 eliminates the bulk of slower migrating species via Western blot upon phleomycin treatment (Liang et al., 2015). Whilst S249 mutation had no effect on Sae2 function, the two threonine residues appear to act redundantly to regulate Sae2 function. The *S. cerevisiae* Thr279 site (found in the conserved region of Sae2) aligns with CtIP Thr289, a site phosphorylated by ATR/ATM and required for HR (Liang et al., 2015; Wang et al., 2013). Collectively it is proposed that these residues mediate specific interactions between Sae2/CtIP and multiple FHA domain-containing proteins including Xrs2, and that such interactions may link Sae2 to its role in regulation of the MRX complex (Liang et al., 2015).

Sae2 is phosphorylated periodically during the meiotic cycle in a Mec1- and Tel1-dependent manner (Cartagena-Lirola et al., 2006). This phosphorylation appears at the onset of pre-meiotic S phase, increases when DSBs are formed and then decreases upon DSB repair. Mutation of these five residues to prevent phosphorylation causes an accumulation of the DSBs formed by Spo11 in a similar manner to that of *sae2*Δ cells, indicating a defect in DSB resection (Cartagena-Lirola et al., 2006).

Oligomerisation

Although phosphorylation of Sae2 alters its activity *in vitro* and *in vivo,* the actual biochemical mechanism of Sae2 activity and how these modifications physically affect the protein are still relatively unknown. Recent work has focused on the way in which post-translational modifications of Sae2 affects its oligomeric state (Cejka lab - personal communication; Fu et al., 2014). Sae2 exists as a multimer *in vivo* (Kim et al., 2008). However, whilst mutation of the L25 self-interacting site (*L25P*) abolishes oligomerisation of Sae2 (Kim et al., 2008), a Sae2 mutant lacking the first 169 amino acids can still promote MRX endonuclease activity *in vitro* (Cannavo and Cejka, 2014), suggesting that self-interaction is not essential for Sae2 activity but instead may regulate its activity *in vivo*. Mutation of the CDK site to alanine (*S267A*) reduces the active/monomeric fraction of Sae2, whereas mimicking phosphorylation at this site, alongside phosphorylation mimics of the Mec1/Tel1 phosphorylation sites, increased the active/monomeric fraction (Fu et al., 2014).

2.8 DSB resection

DSB resection in cycling cells is a process that involves three pathways: MRX/Sae2, Exo1 and Sgs1/Dna2. Resection of the 5′ strand can proceed 2000-4000 nt at mitotic DSBs for allelic recombination, 3000-6000 nt for ectopic recombination (Chung et al., 2010), and 850 nt at meiotic DSBs (Zakharyevich et al. 2010). Mre11 is a 3′ to 5′ exonuclease that acts by creating a nick and resecting towards the DSB (Garcia et al., 2011). With the length resection proceeds, resection by the action of the Mre11 endo- and exonuclease alone would be very inefficient and thus, there is a requirement for more proficient, 5′ to 3′ exonucleases. However, the MRX complex and Sae2 are important for the initiation of resection at DSBs. Resection at both HO-endonuclease (Llorente and Symington, 2004) or I-*Sce*I DSBs (Clerici et al., 2005) is greatly delayed in cells deficient for any component of the MRX complex or Sae2, but resection length is unaffected. Therefore other nucleases can substitute for the initial resection, performed by the MRX complex alongside Sae2, on clean-ended DSBs (Mimitou and Symington, 2009). The 5′ to 3′ nucleases involved in DSB resection are the Exo1 protein and the Sgs1-Dna2 pathway. Resection tract lengths are reduced to 100-300 nt in mitotic cells and 270 nt in meiotic cells containing the *sgs1*Δ *exo1*Δ mutant (Zakharyevich et al. 2010; Chung et al. 2010). The 100-300 nt of resection by the MRX complex can enable the repair of DSBs in the *sgs1*Δ *exo1*Δ double mutant, but a triple mutant of *exo1*Δ *sgs1*Δ *sae2*Δ (or Mre11 mutation) causes cell lethality and no resection occurs (Mimitou and Symington, 2008).

At clean-ended DSBs in mitotic cells the Exo1 protein and the Sgs1-Dna2 pathway act as redundant nucleases required for this extensive resection (Zhu et al., 2008). However, the resection length of DSBs in meiosis is not affected in *SGS1* mutants, suggesting that Exo1 is the sole long range resection mechanism in meiosis (Zakharyevich et al. 2010). In *exo1*Δ meiotic cells, resection proceeds ~270 nt compared to ~850 nt in wild type meiotic cells. In a *sgs1*Δ mutant the length of wild type resection is unchanged, and in the *sgs1*Δ *exo1*Δ double mutant the resection length is still 270 nt (Zakharyevich et al., 2010), consistent with findings that Mre11 and Sae2 cooperate to initially cleave the 5′ strand up to ~270 nt from the DSB and resect 3′ to 5′ (Garcia et al., 2011). Despite this consensus view, it has been reported that Sgs1 may function at late stages of meiosis, but only in the absence of Exo1 (Manfrini et al., 2010).

Quite why Exo1 is the sole 5′ to 3′ resector in meiosis, unlike mitotic cells, is still unknown. However, unpublished work in the Neale laboratory has determined that in a checkpoint-deficient background (*rad24*Δ), Sgs1-Dna2 and Mre11/MRX both perform substantial resection. This suggests that the Mre11 and Sgs1/Dna2 pathways are usually suppressed in a Rad24-dependent manner in meiosis. In mitosis and meiosis the limited resection, dependent on MRX and Sae2 activity, is enough for homology search and efficient joint molecule

formation (Chung et al., 2010; Zhu et al., 2008), suggesting long range resection by Exo1 and Sgs1-Dna2 might only be important for the activation of the DNA-damage checkpoint via ssDNA/RPA stimulation of $Mec1^{ATR}$ (reviewed in Cejka 2015).

2.8.1 MRX and Sae2 interactions with Exo1

Over-expression of *EXO1* can partially suppress the resection initiation defect of Mre11 and Sae2 mutants, but only at clean-ended breaks (Tsubouchi and Ogawa, 2000). This indicates that the nucleolytic action of Mre11 and Sae2 are important for the repair of blocked (protein bound/chemically modified) DSB ends, but are not essential for clean-ended ones. Instead it is thought that MRX simply increases the efficiency of the long-range resection enzymes at clean-ended DSBs (Moreau et al., 2001; Tsubouchi and Ogawa, 2000). As previously discussed, Ku may play a role in preventing Exo1 recruitment or access to the DSB, as deletion of Yku70/80 in Mre11 mutant or *sae2*Δ background suppresses the initiation of resection defect seen in these mutants (Langerak and Russell, 2011; Limbo et al., 2007; Mimitou and Symington, 2011; Shim et al., 2010).

In vitro the favoured substrate for Exo1 is a 3′ overhang duplex DNA substrate (Cannavo et al., 2013). Thus, the nick and 3′ ending ssDNA tail created by MRX/Sae2 *in vivo* creates a more efficient substrate for Exo1 to act upon. MRX/N stimulates Exo1 resection *in vitro* but is not required for its activity as long as Ku is absent, corroborating the *in vivo* data (Cannavo et al., 2013). Stimulation of resection by Exo1 by MRX/Sae2 is not dependent on Mre11 nuclease activity (Nicolette et al., 2010). However, to date, no direct physical interaction between Exo1 and MRX has been detected. Because MRX and Exo1 only interact in the presence of DNA it has been postulated that MRX and Sae2 may alter the conformation of the DNA, rendering it more accessible to Exo1 through unwinding (Nicolette et al., 2010).

MRX also stimulates the Sgs1-Dna2 pathway, as demonstrated by the observation that *exo1*Δ *rad50*Δ cells have decreased resection compared to *exo1*Δ cells alone (Zhu et al., 2008). Additionally, the nuclease-deficient mutant of Mre11 confers sensitivity to IR, which is heightened in the absence of Dna2 nuclease function despite Exo1 still being present (Budd and Campbell, 2009). Thus, the function of Dna2 is partially redundant with MRX with respect to processing of IR-induced damage, and this function is distinct from that of Exo1. Expression of a helicase-proficient, nuclease-deficient form of Dna2 is lethal in cells expressing an Mre11 nuclease-deficient allele, suggesting that structures generated by Dna2 specifically require Mre11-dependent processing in the presence of a helicase+/nuclease- form of Dna2 (Budd and Campbell, 2009; reviewed in Paull, 2010).

2.8.2 Regulation of Exo1

The ssDNA binding protein, RPA, is not required for Exo1 binding to DNA, but is required to promote its extensive resection potentially by preventing the formation of secondary DNA structures and non-specific binding of Exo1 (Chen et al., 2013; Myler et al., 2016). In mitotic cells, Exo1 is phosphorylated in a Rad53-dependent manner, which has been proposed to inhibit its resection activity, a process thought to limit the amount of ssDNA produced and subsequently limit the DNA damage checkpoint (Morin et al., 2008). Conversely, human EXO1 is phosphorylated by CDK, which is required for its activity. Phosphorylation of *S. cerevisiae* Exo1 by CDK has not been demonstrated and may not be conserved. In meiosis, resection is limited to ~850 nt, much less than the ~2000 nt seen at mitotic DSBs even though *EXO1* is upregulated in meiosis in multiple organisms (reviewed in Symington, 2016). This suggests that Exo1 may be subject to negative regulation, possibly similar to the Rad53-dependent regulation proposed in mitotic cells (Morin et al., 2008). The upregulation of Exo1 may contribute as to why resection is so limited in an *exo1*Δ mutant, with Exo1 normally out-competing any alternative resection machinery in wild type cells.

2.9 Meiotic recombination checkpoint

Creation of Spo11-DSBs in meiosis activates the DDR checkpoint, transiently pausing progression to the first meiotic nuclear division (reviewed in Ciccia and Elledge, 2010). This allows time for the induced DSBs to be repaired before the cell attempts to segregate chromosomes. As in mitotic cells, $Mec1^{ATR}$ and $Tel1^{ATM}$ are the main players in the DDR pathway (Morrow et al., 1995; Weinert et al., 1994). The MRX complex, which, as previously discussed, is essential for DSB formation, is responsible for the initial activation of this pathway (Lisby et al., 2004). The MRX complex, through its Xrs2 subunit, recruits and interacts with $Tel1^{ATM}$ at the DSB through a motif in its C-terminus. Deletion of this motif in *XRS2* results in a defect in DNA damage signalling, similar to that of a *tel1*Δ mutant (Nakada et al. 2003). The presence of MRX at a DSB site with a covalently bound protein, such as Spo11, stimulates Tel1 activity (Fukunaga et al., 2011), and this stimulation is hyperactivated when Spo11-DSBs accumulate, such as in a nuclease deficient Mre11 mutant or in a *sae2*Δ background (which are unable to nucleolytically remove Spo11 from the DSB ends). Tel1 recruitment to the DSB site leads to phosphorylation of Mre11, Xrs2 and Sae2, which, as previously described, is essential for Sae2 activity (Cartagena-Lirola et al., 2006; Usui et al., 2001).

In wild type cells, Spo11 removal and resection generates ~1500 nt of ssDNA around each DSB, activating $Mec1^{ATR}$ via Rad24 (reviewed in Navadgi-Patil and Burgers, 2011). Activation of Mec1 results in phosphorylation of the axial element protein Hop1 (Carballo et al., 2008). This in turn results in recruitment and autophosphorylation of the meiosis-specific kinase, Mek1. Activation of Hop1 and Mek1 promotes repair of the DSB via the intact homologous non-sister chromatid rather than the sister chromatid, thus aiding in homolog pairing and preventing Dmc1-independent repair of meiotic DSBs (Carballo et al., 2008). In the case of defective recombination and repair, Mec1 phosphorylation of Mek1 prevents exit of cells from meiotic prophase. In *S. cerevisiae,* this meiotic prophase arrest also requires members of the 9-1-1 checkpoint clamp and loader, Rad17 and Rad24 (Lydall et al., 1996), in a situation similar to the mitotic DDR (Weinert, 1998). Checkpoint activation triggers Mek1 hyperphosphorylation of Swe1 (the *S. cerevisiae* homolog of Wee1), which in turn phosphorylates Cdc28. This phosphorylation inactivates Cdc28 causing the observed delay in meiotic prophase, referred to as pachytene arrest (Leu and Roeder, 1999; Pérez-Hidalgo et al., 2002).

Another downstream target of the DDR is the Ndt80 transcription factor (reviewed in Winter, 2012). Ndt80 is activated by phosphorylation in wild type cells, which triggers a positive autoregulatory loop that leads to the induction of genes required for exit from prophase. An unknown target of the DDR causes Ndt80 hypophosphorylation preventing exit from meiotic prophase (Tung et al., 2000).

2.10 Topoisomerases

DNA undergoes a lot of conformational changes throughout its metabolism. It is densely packaged into the nucleus but requires relaxation and alteration for processes such as transcription, replication and proper chromosomal segregation. Torsional stress in the form of supercoiling is generated from these processes, and therefore the cell requires a mechanism to relieve this tension. This mechanism comes in the form of topoisomerases with genetic and cell biological studies providing the essential nature of these proteins in various cellular processes (Almouzni and Méchali, 1988; Mondal et al., 2003).

Topoisomerases essentially cut and religate DNA, allowing DNA strand passage, strand decatenation and unwinding, and are able to do this unassisted (see Figure 1.8 for schematic). Topoisomerases work by utilising a reversible transesterification reaction (Champoux, 1977). The active site of these enzymes contains a catalytic tyrosine, which acts as a nucleophile to the phosphate backbone of the DNA (Tse-Dinh et al., 1984). This first transesterification reaction breaks the DNA backbone, creating a covalently bound protein adduct. This transient

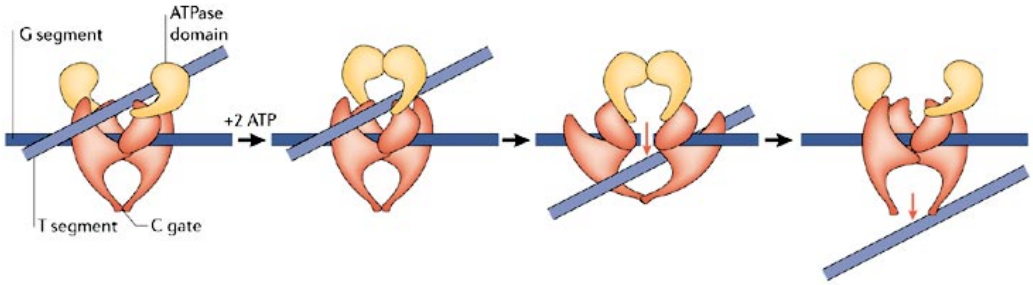


Figure 1.8: A model for the mechanism of strand passage for yeast Top2.
Top2 binds to the G segment of DNA and cleaves it through a transesterification reaction, generating a covalent bond to the 5′ end of the scissle phosphate on both strands. This creates a double-strand break in the G segment. A second DNA segment known as the T segment can now pass through the DSB. It is first captured by the Top2 dimer in an ATP-dependent manner whereby the N-terminal (ATPase) domains of the two Top2 monomers form a clamp around the T segment. Strand passage of the T segment enables it to exit through the bottom of the dimer. ATP hydrolysis releases the clamp and allows the Top2 enzyme to reset.
Figure taken from Cozzarelli et al., 2006

break allows for the various transformations of the DNA structure by the topoisomerases. The second transesterification reaction reseals the DNA backbone and regenerates the catalytic tyrosine of the topoisomerase (Wang, 2002).

In mammalian cells there are six genes coding for topoisomerases. The equivalent proteins are split into two categories, type I and type II topoisomerases, depending on their mechanism of action (Champoux, 2001). Type II topoisomerases are classified into two distinct families IIA and IIB, with IIA family members including eukaryotic Top2, bacterial DNA gyrase and topoIV. The six genes are also categorised into three groups, each containing two isoforms: Top1 and Top1mt, Top2α and Top2β, and finally Top3α and Top3β (reviewed in Wang, 2002). Topoisomerases differ by their structure and mechanism of this cleaving and religating. Topoisomerase I (Top1) and Topoisomerase III (Top3) cleave and religate a single strand of DNA, whereas Topoisomerase II (Top2) cleaves both strands generating a four base pair overhang before religating (reviewed in Pommier et al. 2014).

Top1 first creates a single-stranded break through creation of a covalent link to the 3′ strand of the DNA phosphate backbone, that permits dissipation of supercoiling through rotation of the broken strand around the intact one (Chen et al. 2013). Top2 can also reduce supercoiling but, as well as Top3, can also work through strand passage, allowing a single or double strand of DNA through the cleavage complex of another DNA strand (Soret et al., 2003). Both create covalent adducts with opposite polarity to Top1, thus creating a bond with the 5′ phosphate. (Pommier et al., 2010, 2014; Schoeffler and Berger, 2008). Top3α allows passage of DNA molecules such as those formed in double Holliday junctions (Plank et al., 2006; Pommier, 2012; Wu et al., 2000), whereas with Top3β the substrate can be a single-stranded RNA molecule (Stoll et al., 2013; Xu et al., 2013).

2.10.1 Topoisomerase II

Eukaryotic topoisomerase II proteins are largely homodimers, and whereas *S. cerevisiae* has only one type of type II topoisomerase (Top2), vertebrates have two isoforms, Top2α and Top2β. *S. cerevisiae* Top1 is not essential (but is essential in multicellular eukaryotes for early development), thought to be due to Top2 being able to substitute Top1 function (Champoux, 2001; Soret et al., 2003; Uemura and Yanagida, 1984).

A study in fission yeast using ChIP-chip showed that Top1 and Top2 sites map to intergenic regions, with 65% of sites mapping both Top1 and Top2 (Salceda et al., 2006). However, Top1 sites positively correlated with mRNA levels, whereas in highly transcribed ORFs they noted reduced Top2 binding. This indicates that Top1 activity is linked to transcriptionally active regions (Durand-Dubief et al., 2010). Top2 localises more than Top1 to the coding

(nucleosomal) regions, consistent with Top2 being the main modulator of DNA topology on nucleosomal DNA (Salceda et al., 2006).

2.10.2 Stalled/poisoned Top2

There are many anti-cancer drugs that target topoisomerases, with two most commonly used in the laboratory setting being camptothecin (CPT) and etoposide. These drugs bind to Top1 and Top2 respectively and increase the half-life of the covalently bound protein (Top1 and Top2 cleavable complex – Top1cc/Top2cc) to the DNA end (Baldwin and Osheroff, 2005). The two subunits of Top2cc can religate their broken DNA strand independently of one another, which means that etoposide exposure, which prevents religation, can trap Top2ccs as a single-stranded nick (etoposide only affects one monomer) or double-stranded break (etoposide affects both monomers) (Bromberg et al., 2003), in a concentration dependent manner. However, the Top2 lesions are not directly sensed as damage because the bound, intact Top2 has a strong interaction between the two Top2 subunits, holding either side of the break together (Tennyson and Lindsley, 1997). True DSBs can arise when a DNA replication fork collides with a nick or a DSB at the Top2cc (Kuzminov, 1999). Alternatively, collision with the transcription machinery can disrupt the Top2cc, or it can stimulate proteolysis of the Top2cc by the 26S proteasome (Mao et al., 2001; Tammaro et al., 2013).

Genome-wide mapping of topoisomerase II cleavage following etoposide treatment of mammalian cells has recently been reported (Baranello et al., 2014). Single-stranded DNA breaks (SSBs) were labelled by nick translation using DNA polymerase I and digoxigenin-modified nucleotides, whilst DSBs were 3′ end tailed with biotin labelled nucleotides. Labelled DNA molecules were enriched for after shearing. Etoposide was found to generate mostly SSBs compared to DSBs, potentially due to etoposide inhibiting the religation of each broken strand by the two Top2 monomers independently (Bromberg et al., 2003). Top2 lesions were enriched at promoter regions at the transcriptional start site (TSS) compared to the rest of the genome, and this was positively correlated with the levels of transcription (weakly expressed genes had lower Top2 breaks and higher expressed genes had more). One explanation for this possible correlation may be that the increase in gene expression increased the level of transcription-generated torsional stress, which required increased action of Top2 (Baranello et al., 2014).

2.10.3 Repairing stalled topoisomerase complexes

Top2-DNA complexes, formed after inhibition by drugs such as etoposide (described above), are very similar to DSBs formed by Spo11, in that both generate 5′ covalent phosphotyrosyl bonds on both sides of the DSB. MRN null mutants and *rad50S* or *mre11* nuclease-dead mutants in *S. pombe* are hypersensitive to both Top1 and Top2 poisons, suggesting a role for the MRX/N complex in processing topoisomerase lesions (Hartsuiker et al., 2009; Malik and Nitiss, 2004). The mechanism for repairing DSBs generated by Top2 is still far from clear. MRX/Sae2-dependent release of Spo11 generates Spo11-oligos (Neale et al., 2005). For Top2, a similar mechanism has been proposed, yet the apparent size of Top2-oligo complexes was much shorter (Hartsuiker et al., 2009; Neale et al., 2005). However, in *S. pombe,* Ctp1 (Sae2) is only directly required for processing Top2 lesions and not Top1 ones (Hartsuiker et al., 2009), correlating with recent studies on Sae2 demonstrating that it promotes Mre11 endonuclease activity on the 5′ strand adjacent to 5′ blocked DNA ends (Cannavo and Cejka, 2014).

In mammalian cells, Top2-DSBs can also be repaired via the action of tyrosyl-DNA phosphodiesterase 2 (TDP2), a protein that contains 5′ phosphodiesterase activity capable of directly hydrolysing the phosphotyrosyl bond between proteolysed Top2 and the 5′ end of the DSB allowing repair. As such, TDP2-defective cells are highly sensitive to etoposide (Cortes-Ledesma et al., 2009). TDP2 has since been characterised to be specific for hydrolysing an array of 5′ phosphotyrosyl bonds as well as an, albeit weak, activity for 3′ phosphotyrosyl bonds, only in the absence of Tdp1 (Zeng et al., 2012). The mechanism of repairing Top2 DSBs via the MRX complex results in ssDNA generation, which is refractory to NHEJ and instead drives repair of DSBs down the HR pathways (Shibata et al., 2014). With NHEJ being the predominant repair pathway in higher eukaryotes, it would suggest that TDP2 is the main mechanism of repair at Top2-DSBs (Mårtensson et al., 2003).

In *S. cerevisiae*, there is no obvious TDP2 ortholog. Tdp1, a protein that specifically hydrolyses 3′ phosphotyrosyl bonds, is present and orthologs are found in all organisms. Tdp1 activity against 3′-tyrosyl substrates correspond to Top1 lesions and not 5′ linked Top2 or Top3 substrates (Pouliot et al., 1999, 2001). However, deletion of *TDP1* in *S. cerevisiae* results in hypersensitivity to both Top2 targeting agents and over-expression of *TOP2* (Nitiss et al., 2006), suggesting that Tdp1 does play a role in processing Top2-5′ lesions. Consistent with this, recombinant Tdp1 was able to hydrolyse a covalently bound, Top2 derived, peptide from the 5′ end of a DNA substrate *in vitro* (Nitiss et al., 2006). Therefore, it is possible that, in *S. cerevisiae,* Tdp1 acts on both types of phosphotyrosyl bonds, whereas in humans two proteins have evolved to tackle the two different types of lesion. Alternatively, in *S. cerevisiae*,

Tdp1 plays a minor role in processing Top2 lesions with other pathways, such as the MRX/Sae2 pathway, responsible for the processing of these lesions. This latter hypothesis may indicate why HR is favoured over NHEJ in *S. cerevisiae* for the repair of DSBs, as the increased activity of MRX/Sae2 would generate ssDNA tails, which are inhibitory to repair by NHEJ.

2.11 Questions to answer

DSBs are critical lesions that occur frequently in cells of all organisms. Pathways have evolved to repair DSBs to prevent deletions, translocations and fusions in the DNA. The HR and NHEJ pathways are the two main mechanisms by which DSBs can be repaired. The choice between the two pathways depends on a variety of factors, including the cell-cycle stage, coordinated by CDKs. The proteins that have been proposed to play a role at the critical decision point between the two pathways are the MRX and Sae2 complex. Activation of Sae2 by CDKs, which stimulates the Mre11 endonuclease, drives repair of DSBs down the HR route. Therefore, study into the role of Sae2 is of importance to understand how Sae2 and the MRX complex are regulated, as aberrant repair by NHEJ or HR can lead to genomic instability. The balance between NHEJ and HR also differs between species, with mammalian cells predominantly repairing DSBs via NHEJ, in contrast to *S. cerevisiae.* The presence or absence of specific repair pathways may be responsible for this difference. For example, mammalian cells contain the TDP2 protein, which hydrolyses covalently bound TOP2 from DSBs. This process generates a clean-ended DSB, which is easily repaired by NHEJ. In *S. cerevisiae* a TDP2 ortholog does not exist. An ortholog of mammalian TDP1 protein, which hydrolyses 3′ bound TOP1 from DNA, does exist (Tdp1) and may have dual polarity of removing 3′ and 5′ bound proteins from DNA. However, as HR predominates in *S. cerevisiae*, it is more likely that other repair pathways, such as the MRX/Sae2 pathway, act instead and repair DSBs in a manner that promotes HR. It is therefore of interest as to how DSBs in *S. cerevisiae* repair if a pathway such as the mammalian TDP2 repair pathway could be simulated. This is of particular interest at meiotic Spo11-DSBs, which always repair via HR due to the action of MRX/Sae2, for accurate chromosome alignment and pairing. Spo11-DSBs can occur anywhere in the genome but appear more frequently at defined hotspots. The mechanism and regulation of where Spo11 creates DSBs has been studied, but there is still a requirement for further mapping of Spo11-DSB sites. Development of a method for mapping Spo11 may also be applicable for mapping Top2-DSBs genome-wide because Spo11 and Top2 generate similar DSBs. Therefore, there are numerous questions in the field that still need answering. Some specific questions that will be looked at in this thesis are as follows:

- How is Sae2 regulated in meiosis in direct respect to processing the ends of Spo11-DSBs?
- As TDP2 has been shown to be able to remove proteolysed Top2 and other covalently bound substrates, can it also act on Spo11-bound DSBs? If so is there any role for TDP2 in meiosis?
- If Spo11-DSBs are processed in a non-nucleolytical manner, does this affect how the resection machinery and HR proteins act upon the break?
- How does Tel1 regulate interference of DSBs during meiosis? Are the effects Tel1 has on DSB formation dependent on its kinase activity?
- Can we develop a way to map, with single nucleotide resolution, where Spo11 creates DSBs during meiosis in a *sae2*Δ background where DSBs accumulate?
- If we can develop a method for this, can we extrapolate this method to look into other DNA-protein adducts and where they form?

CHAPTER 2:

MATERIALS AND METHODS

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Strains

Table 2.1: ***S. cerevisiae*** **strains used in this study. Note for the *SAE2* mutagenesis strains see plasmid table for corresponding strain numbers and genotypes**

Parent: **VG296** lys2/", ura3/", arg4-nsp or bgl?/", leu2::hisG or leu2Δ?/", his4X::LEU2,/" nuc1::LEU2/", SPO11-His6-FLAG3-loxP-KanMX-loxP/"			
Strain #	**Name**	**Genotype**	**Notes**
VG303	*sae2*Δ	sae2Δ::KanMX6/"	Constructed by VG
DJ14	*SAE2 +HYG*	pVG19 (CEN)::HYG	Empty CEN plasmid expressing hygromycin resistance as a control
DJ15	*sae2*Δ *+ HYG*	sae2Δ::KanMX6/", pVG19 (CEN)::HYG	Empty CEN plasmid expressing hygromycin resistance as a control
DJ20	*CEN:SAE2*	sae2Δ::KanMX6/", pDJ8: 0.5kb+-Sae2 (HYG)	CEN plasmid: Untagged Sae2
DJ21	*CEN:SAE2-His10*	sae2Δ::KanMX6/", pDJ11: 0.5kb+-Sae2-His10 (HYG)	CEN plasmid: Sae2-His10
DJ22	*CEN:SAE2-3HA*	sae2Δ::KanMX6/", pDJ10: 0.5kb+-Sae2-HA (HYG)	CEN plasmid: Sae2-3HA
DJ160	*CEN: pADH1-TDP2-His6*	sae2Δ::KanMX6/", pDJ85: p*ADH1-TDP2-His6* (HYG)	CEN plasmid: Tdp2-His6 constitutively expressed
DJ70	*CEN:GAL1 -TDP2*	sae2Δ::KanMX6/", pDJ62: p*GAL1-TDP2* (HYG)	CEN plasmid: Tdp2 induced expression

Table 2.2: **Strains for *SAE2* over-expression**

Parent: **MJ548xMJ555** (ho::LYS2/', lys2/', arg4-bgl/', leu2/', ura3::P_{GPD1}GAL4(848)-ER::URA3, SPO11-His6-FLAG3-loxP-KanMX-loxP/', HIS4::LEU2-(BamHI;+ori)/his4X::LEU2-(NgoMIV;+ori)--URA3, SAE2/ sae2Δ::KanMX6) – MJ548 was transformed with integrating plasmid as stated before being crossed with MJ555			
Strain	**Construction**	**Genotype**	**Notes**
DJ74	MJ548+pMN118	sae2_5'UTR::natNT2::P_GAL1::SAE2[+1 to	*GAL1-SAE2 (cryptic SK1 ATG)*

	x MJ555	+1192]/sae2Δ::KanMX6	
DJ75	MJ548+pMN119 x MJ555	sae2_5'UTR::natNT2::P_GAL1::SAE2[+1 to +1192]/sae2Δ::KanMX6	*GAL1-SAE2 (cryptic S288C ATG)*
DJ132	MJ552+pDJ79 x MJ555	sae2_5'UTR::natNT2::P_GAL1::SAE2(S267A)[+1 to +1192]/sae2Δ::KanMX6	*GAL1-SAE2 (S267A) (cryptic SK1 ATG)*
DJ134	MJ552+pDJ80 x MJ555	sae2_5'UTR::natNT2::P_GAL1::SAE2(S278A-T279A)[+1 to +1192]/sae2Δ::KanMX6	*GAL1-SAE2 (S278A-T279A) (cryptic SK1 ATG)*
DJ157	MJ552+pDJ84 x MJ555	sae2_5'UTR::natNT2::P_GAL1::SAE2(L25P)[+1 to +1192]/sae2Δ::KanMX6	*GAL1-SAE2 (L25P) (cryptic SK1 ATG)*

Table 2.3: **Strains for Spo11-DSB Mapping**

Parent: **MJ6 (**ho::LYS2/", lys2/", ura3/", arg4-nsp/", leu2::hisG/", his4X::LEU2/", nuc1::LEU2/")			
Strain	**Name**	**Genotype**	**Notes**
VG377	*exo1*Δ *dmc1*Δ	dmc1Δ::HphMX4/", exo1Δ::KanMX4/"	Alias: KW1
MJ315	Untagged	sae2Δ::KanMX6/'	'Wild type' for Spo11 Mapping but actually *sae2*Δ
MJ319	Spo11-YF	sae2Δ::KanMX6/', spo11(Y135F)-HA3His6::KanMX4/'	Spo11-catalytic dead (*Y135F*), *sae2*Δ
VG402	*tel1*Δ	sae2Δ::KanMX4/", tel1Δ::HphMX4/"	
VG431	*tel1kd*	sae2Δ::KanMX6/", tel1-D2612A, N2617A, D2631A/"	*tel1-kinase dead (tel1kd), sae2*Δ

Table 2.4: **Strains for topoisomerase mapping**

Parent:	**MJ429 (**ura3Δ0, leu2Δ0, his3Δ1, met15Δ0, pdr1Δ::PDR1-DBD-CYC8::LEU2) - haploid		
Strain	**Name**	**Genotype**	**Notes**
MJ429	*Wild type*	As above	Wild type with sensitivity cassette
MJ475	*sae2*Δ	sae2Δ::KanMX6	*sae2*Δ with sensitivity cassette
MJ551	*mre11*Δ	mre11Δ::KanMX4	*mre11*Δ with sensitivity cassette

2.1.2 Plasmids

Centromeric plasmids

Table 2.2: **Centromeric plasmids (all containing hygromycin (HYG) drug-resistance marker)**

Plasmid#	Name	Construction	Strain #
pVG19	Empty HYG resistance	Constructed by VG. Used as a control for HYG resistance	DJ15
pVG24	Sae2	Constructed by VG. Untagged wild type Sae2	DJ20
pDJ9	pVG24+*Nhe*I	Mutagenesis of pVG24 using primers DJ#63&64 to incorporate *Nhe*I site at C-terminal of SAE2	N/A
pDJ10	Sae2-HA/Wild type	Ligated in –HA3 tag in frame to C-terminal of *SAE2*	DJ22
pDJ16	*SAE2 (S267A)*	Mutagenesis of pDJ10 with primers VG#22&23	DJ23
pDJ17	*SAE2 (S267E)*	Mutagenesis of pDJ10 with primers VG#24&25	DJ24
pDJ18	*SAE2 (S73A)*	Mutagenesis of pDJ10 with primers DJ#24&26	DJ25
pDJ19	*SAE2 (S73E)*	Mutagenesis of pDJ10 with primers DJ#25&27	DJ26
pDJ20	*SAE2 (T90A)*	Mutagenesis of pDJ10 with primers DJ#28&30	DJ27
pDJ21	*SAE2 (T90E)*	Mutagenesis of pDJ10 with primers DJ#29&31	DJ28
pDJ22	*SAE2 (S249A)*	Mutagenesis of pDJ10 with primers DJ#32&34	DJ29
pDJ23	*SAE2 (S249E)*	Mutagenesis of pDJ10 with primers DJ#33&35	DJ30
pDJ26	*SAE2 (S289A)*	Mutagenesis of pDJ10 with primers DJ#44&45	DJ33
pDJ27	*SAE2 (S289E)*	Mutagenesis of pDJ10 with primers DJ#46&49	DJ34
pDJ32	*SAE2 (S73A-T90A)*	Mutagenesis of pDJ18 with primers DJ#28&30	DJ41
pDJ33	*SAE2 (S73E-T90E)*	Mutagenesis of pDJ19 with primers DJ#29&31	DJ42
pDJ34	*SAE2 (S249A-S289A)*	Mutagenesis of pDJ22 with primers DJ#44&45	DJ43

pDJ40	*SAE2 (K239R-K266R)*	Mutagenesis of pDJ28 with primers DJ#60&62	DJ39
pDJ41	*SAE2 (K239Q-K266Q)*	Mutagenesis of pDJ29 with primers DJ#59&61	DJ40
pDJ43	*SAE2 (S73E-T90E-S249E)*	Mutagenesis of pDJ33 with primers DJ#33&35	DJ80
pDJ48	*SAE2 (S134A)*	Mutagenesis of pDJ10 with primers DJ#71&72	DJ78
pDJ49	*SAE2 (S134E)*	Mutagenesis of pDJ10 with primers DJ#73&74	DJ79
pDJ53	*SAE2 (T279A)*	Mutagenesis of pDJ10 with primers DJ#79&80	DJ54
pDJ57	*SAE2 (N123A-R127A)*	Mutagenesis of pDJ55 with primers DJ#81&82	DJ56
pDJ58	*SAE2 (T279E)*	Mutagenesis of pDJ10 with primers DJ#77&78	DJ76
pDJ59	*SAE2 (E24Q-L25P)*	Mutagenesis of pDJ10 with primers RA#1&2	DJ82
pDJ63	*TDP2*	Sub-cloned in *TDP2* in place of *SAE2* on *GAL1-SAE2* plasmid	DJ70
pDJ64	*SAE2 (E131V)*	Mutagenesis of pDJ10 with primers RA#7&8	DJ83
pDJ65	*SAE2 (T90S)*	Mutagenesis of pDJ10 with primers DJ#150&151	DJ97
pDJ66	*SAE2 (T279S)*	Mutagenesis of pDJ10 with primers DJ#172&173	DJ98
pDJ69	*SAE2 (S289D)*	Mutagenesis of pDJ10 with primers DJ#174&175	DJ99
pDJ75	*EXO1*	pVG25 digested with *Xho*I and *Not*I to remove *SAE2+-0.5kb. EXO1+-*0.5kb amplified with *Xho*I and *Not*I flanking ends. Insert and vector digested and *EXO1* ligated into centromeric plasmid.	DJ110
pDJ78	*SAE2 (S179A)*	Mutagenesis of pDJ10 with primers DJ#160&161	DJ126
pDJ83	*SAE2 (L25P)*	Mutagenesis of pDJ10 with primers DJ#224&225	DJ82
pDJ86	*SAE2 (S249E-S267E)*	Mutagenesis of pDJ23 with primers VG#24&25	DJ176
pDJ87	*SAE2 (T279E-S267E)*	Mutagenesis of pDJ58 with primers VG#24&25	DJ177

pDJ88	*SAE2 (S289E-S267E)*	Mutagenesis of pDJ27 with primers VG#24&25	DJ178
pDJ89	*SAE2 (S73E-T90E-S249E-T279E)*	Mutagenesis of pDJ43 with primers DJ#77&78 Alias: 4E	DJ174
pDJ91	*SAE2 (S179E)*	Mutagenesis of pDJ10 with primers DJ#162&163	DJ172
pDJ92	*SAE2 (S73E-T90E-S249E-T279E-S267E)*	Mutagenesis of pDJ89 with primers VG#24&25 Alias: 4E+S267E	DJ175
pDJ94	*SAE2 (S73A-T90A-S249A-T279A)*	Mutagenesis of pDJ93 with primers DJ#28&30 Alias: 4A	DJ173
pDJ85	*pADH1-TDP2-His6*	Subcloned in hygromycin resistance marker in place of *URA3* marker on pMN110	DJ160
pMN118	*GAL1-SAE2 (SK1 ATG)*	*SAE2* under control of *GAL1* promoter with SK1 cryptic ATG (constructed by NL)	DJ74
pMN119	*GAL1-SAE2 (S288C ATG)*	*SAE2* under control of *GAL1* promoter with S288C cryptic ATG (constructed by NL)	DJ75
pDJ79	*GAL1-SAE2 (S267A) (SK1 ATG)*	pMN118 mutagenised using primers VG22+23	DJ132
pDJ80	*GAL1-SAE2 (S278A-T279A) (SK1 ATG)*	pMN118 mutagenised using primers DJ36+38	DJ134
pDJ84	*GAL1-SAE2 (L25P) (SK1 ATG)*	pMN118 mutagenised using primers DJ224+225	DJ157

All *SAE2* plasmids listed here are Sae2-HA unless otherwise stated.

The base *SAE2-HA* centromeric plasmid (pDJ10) was created as depicted in Figure 2.1.

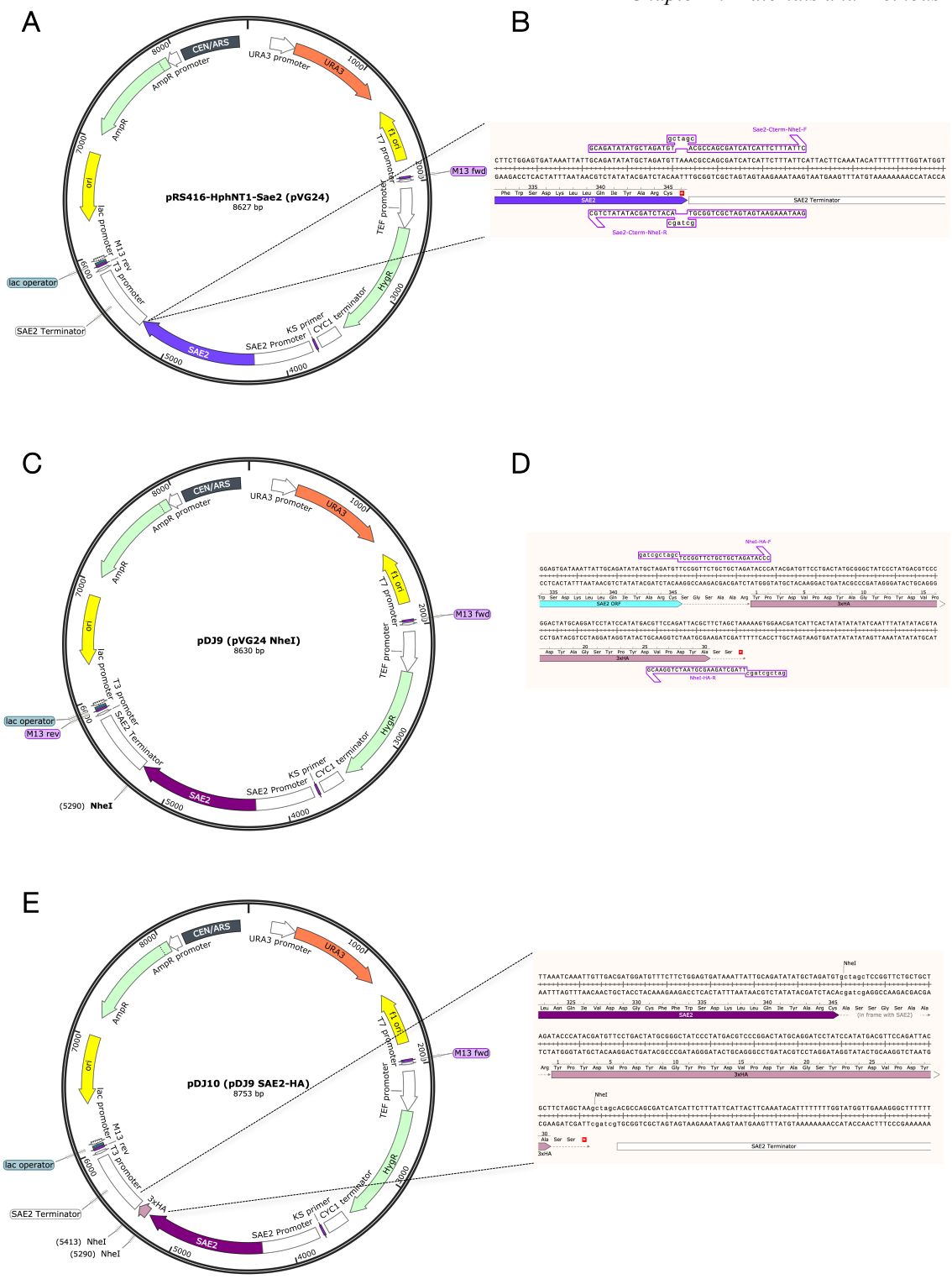


Figure 2.1: Schematic of cloning steps to generate centromeric *SAE2-HA* plasmid for mutagenesis.

A. pVG24, a centromeric plasmid cloned by Valerie Garcia containing the *SAE2* open reading frame (ORF) +/- 0.5 kb placing *SAE2* under it's natural promoter and terminator. The plasmid also contains the hygromycin drug-resistance marker for transformed yeast selection. **B.** Overlapping primers (DJ#63&#64) were designed and site-directed mutagenesis performed to incorporate a *Nhe*I site a the C-terminal of *SAE2*. **C.** Insertion of the *Nhe*I site at the C-terminal of *SAE2*, removing the stop codon and permitting ligation of a *HA3* tag in frame with the *SAE2* ORF. **D.** *HA3* tag was amplified with *Nhe*I flanking ends using colony PCR of MJ552 using primers DJ#65&#66. **E.** Schematic of final plasmid for mutagenesis, pDJ10. pDJ9 and the amplified HA3 insert were digested with *Nhe*I at 37 ° C for 1 hour, PCR clean-up performed and the insert ligated using T4 ligase. Zoomed in region shows *SAE2* ORF in frame with the *HA3* tag.

2.1.3 Oligonucleotides

Oligonucleotide Synthesis

All oligonucleotides were synthesised by Fisher Scientific UK Ltd. or Integrated DNA Technologies (IDT, *)

Table 2.3: **Oligonucleotides used in this study**

Oligo #	Name	Sequence 5′- 3′
24	SAE2-S73A-F	CTCCTCAACAATCCGCCCAGACGTCTGCGG
25	SAE2-S73E-F	CTCCTCAACAATCCGAACAGACGTCTGCGG
26	SAE2-S73A-R	CCGCAGACGTCTGGGCGGATTGTTGAGGAG
27	SAE2-S73E-R	CCGCAGACGTCTGTTCGGATTGTTGAGGAG
28	SAE2-T90A-F	GAAGATTTCATCCTTGCCCAGTTTGATGAG
29	SAE2-T90E-F	GAAGATTTCATCCTTGAACAGTTTGATGAG
30	SAE2-T90A-R	CTCATCAAACTGGGCAAGGATGAAATCTTC
31	SAE2-T90E-R	CTCATCAAACTGTTCAAGGATGAAATCTTC
32	SAE2-S249A-F	CAGTAGTTATAGAAGCCCAAAATTCGGACT
33	SAE2-S249E-F	CAGTAGTTATAGAAGAACAAAATTCGGACT
34	SAE2-S249A-R	AGTCCGAATTTTGGGCTTCTATAACTACTG
35	SAE2-S249E-R	AGTCCGAATTTTGTTCTTCTATAACTACTG
44	SAE2-S289A-F	GAGGACAAAAAGAAAGCCCAGGAAATCATC
45	SAE2-S289A-R	GATGATTTCCTGGGCTTTCTTTTTGTCCTC
46	SAE2-S289E-R	GATGATTTCCTGTTCTTTCTTTTTGTCCTC
49	SAE2-S289E-F	GAGGACAAAAAGAAAGAACAGGAAATCATC
55	SAE2 K239-acetyl-mimic_F	GCCAGAGGATTCTCAACACAGATCATTGTC
56	SAE2 K239-acetyl-mimic_R	GACAATGATCTGTGTTGAGAATCCTCTGGC
57	SAE2 K239-acetyl-mutant_F	GCCAGAGGATTCTAGACACAGATCATTGTC
58	SAE2 K239-acetyl-mutant_R	CAAAACCTGGGGGGGATCTTGATCTATTCCTC
59	SAE2 K266-acetyl-mimic_R	CAAAACCTGGGGGGGATTGTGATCTATTCCTC
60	SAE2 K266-acetyl-mutant_F	GAGGAATAGATCAAGATCCCCCCCAGGTTTTG
61	SAE2 K266-acetyl-mimic_F	GAGGAATAGATCACAATCCCCCCCAGGTTTTG
62	SAE2 K239-acetyl-mutant_R	GACAATGATCTGTGTCTAGAATCCTCTGGC

63	SAE2-CTERM-NHEI-F	GCAGATATATGCTAGATGTGCTAGCACGCCAGCGATCATCATTC
64	SAE2-CTERM-NHEI-R	GAATGATGATCGCTGGCGTGCTAGCACATCTAGCATATATCTGC
71	CDK S134A F	CGGAGTTTAGTGCCCCTTTAAATGGCC
72	CDK S134A R	GGCCATTTAAAGGGGCACTAAACTCCG
73	CDK S134E F	CGGAGTTTAGTGAACCTTTAAATGGCC
74	CDK S134E R	GGCCATTTAAAGGTTCACTAAACTCCG
75	N123A R	CGTTTATGCCTGGCAGGTGGCATAGTAAC
76	N123A F	GTTACTATGCCACCTGCCAGGCATAAACG
77	T279E F	GGATTTTCCCTCCGAACAGGAAGGGAACG
78	T279E R	CGTTCCCTTCCTGTTCGGAGGGAAAATCC
79	T279A R	CGTTCCCTTCCTGGGCGGAGGGAAAATCC
80	T279A F	GGATTTTCCCTCCGCCCAGGAAGGGAACG
81	R127A R	ACTCCGAAATTTTGGCTTTATGCCTATTAG
82	R127A F	CTAATAGGCATAAAGCCAAAATTTCGGAGT
148	EXO1-0.5KB (XHOI)_F	AAGTCTCGAGGTGACAAATCACTGGAAGA
149	EXO1+0.5KB (NOTI)_R	CCGTGCGGCCGCTTTGATTAACAAAATTCTCG
150	SAE2-T90S_F	GATTTCATCCTTTCTCAGTTTGATG
151	SAE2-T90S_R	CATCAAACTGAGAAAGGATGAAATC
160	SAE2-S179A_F	CTGAATCTACAGCGCCAAATTTATAC
161	SAE2-S179A_R	GTATAAATTTGGCGCTGTAGATTCAG
162	SAE2-S179E_F	CTGAATCTACAGAGCCAAATTTATAC
163	SAE2-S179E_R	GTATAAATTTGGCTCTGTAGATTCAG
172	SAE2-T279S_F	GATTTTCCCTCCTCTCAGGAAGGGAAC
173	SAE2-T279S_R	GTTCCCTTCCTGAGAGGAGGGAAAATC
174	SAE2-S279D_F	CAAAAAGAAAGACCAGGAAATCATC
175	SAE2-S279D_R	GATGATTTCCTGGTCTTTCTTTTTG
220*	ADAPTOR_P5_(1)_TOP	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
221*	ADAPTOR_P5_(1)_BOTTOM	/5PHOS/GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/INVDT/
222*	ADAPTOR_P7_(2)_TOP	/5PHOS/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC/INVDT/
223*	ADAPTOR_P7_(2)_BOTTOM	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

224	SAE2: L25P_F	CAGTCTCGATGAACCACTAAATGTGCA
225	SAE2: L25P_R	TGCACATTTAGTGGTTCATCGAGACTG

2.1.4 Reagents

Any specific chemicals, buffers and solutions used are detailed in the appropriate method. All chemicals were purchased from Sigma Aldrich and all enzymes from New England Biolabs (NEB), unless otherwise stated.

2.1.5 Media

Bacteria Media

Luria Broth (LB): 1% Bacto-tryptone, 0.5% Yeast Extract, 1% NaCl pH 7.0 .

For LB plates, 2% Bacto Agar is added and solution microwaved to dissolve the agar.

LB-Amp plates: Ampicillin is added to a final concentration of 100 μg ml^{-1} after microwaving when solution has cooled to holding temperature.

Yeast Media

YPDAU: 1% BD Bacto Yeast Extract, 2% BD Bacto Peptone, 0.5 mM Adenine, 0.4 mM Uracil, 20% D-Glucose. Yeast extract, peptone, adenine and uracil are added to 90% volume double distilled water and autoclaved. D-Glucose is added to a final concentration of 2% bringing the total final volume to 100%.

For YPDAU plates, 2% Bacto Agar is added before autoclaving.

For drug selection, YPDAU is prepared as documented and drugs are added to the final concentration as listed below:

Table 2.4: **Drug concentrations in yeast media**

Drug	Final Concentration
G418-200	200 μg ml^{-1}
G418-400	400 μg ml^{-1}
Hygromycin (HYG)	300 μg ml^{-1}
Nourseothricin-Dihydrogen Sulfate (Nat)	100 μg ml^{-1}
Methyl Methanesulfonate (MMS)	0.025%
Camptothecin (CPT**)**	20-60 μM

YPA: 1% BD Bacto Yeast Extract, 2% BD Bacto Peptone, 1% Potassium Acetate, 0.001% Antifoam 204. Antifoam 204 added after autoclaving.

Sporulating media (SPM): 2% Potassium Acetate, 5 μg ml^{-1} Adenine, 5 μg ml^{-1} Arginine, 5 μg ml^{-1} Histidine, 15 μg ml^{-1} Leucine, 5 μg ml^{-1} Tryptophan, 5 μg ml^{-1} Uracil, 0.001% Sigma Antifoam 204. Antifoam 204 and amino acid supplements added after autoclaving.

2.2 **Methods**

2.2.1 Nucleic Acid Manipulation

PCR Amplification

A solution was made containing 10.5 μl water, 0.5 μl DNA template, 1 μl 10 μM mix of forward and reverse primers, and 12.5 μl 2× Phusion High-Fidelity PCR Master Mix (#M0531L; NEB). A PCR program was run consisting of: 95 °C for 5 minutes initial denaturation; 30 cycles of denaturation at 95 °C for 30 seconds, primer annealing at 56 °C (as standard, but was altered depending on primer annealing Tm °C) for 1 minute, extension at 72 °C (1 minute per kb of mutating plasmid); final extension at 72 °C for 15 minutes.

Colony PCR

1-2 mm of a yeast colony was taken using a p200 tip and mixed into 50 μl of double distilled water. The mix was boiled at 100 °C for 5 minutes, left to cool to room temperature for 10 minutes then spun at 20,000 × *g* for 1 minute to pellet cellular debris. 0.5-1 μl of supernatant was taken as DNA template for PCR amplification.

Restriction Digests

DNA was restriction digested as per NEB restriction enzyme guidelines. As standard, 1 μg DNA in 1× NEBuffer with 1 μl restriction enzyme was placed at 37 °C for 1-2 hours.

Annealing Oligonucleotides

Complementary oligonucleotides were mixed together in a 1× TE annealing solution (15 μM top primer, 15 μM bottom primer, 50 mM NaCl) and placed at 100 °C for 5 minutes in a hot block. The metal block was then removed and placed on the bench for 3-4 hours to cool the sample slowly to room temperature allowing efficient annealing of the oligonucleotides. Annealed oligonucleotides were stored at -20 °C.

Dephosphorylation of Vector DNA

Desphosphorylation of DNA after digestion using restriction enzymes was conducted by adding 1 µl of Alkaline Phosphatase or Calf Intestinal Phosphatase (CIP) directly to the digestion mix and placing at 37 °C for 1 hour.

Ligation of DNA

As per NEB's guidelines the following reaction was set up on ice: 2 µl 10× T4 DNA Ligase Buffer, 50 ng (0.020 pmol), Vector DNA (4 kb), 37.5 ng (0.060 pmol), Insert DNA (1 kb), 1 µl T4 DNA Ligase, double distilled water to 20 µl. The sample was mixed and incubated at 16 °C overnight. The sample was heat inactivated at 65 °C for 10 minutes.

PCR Clean-up

DNA was cleaned after PCR using a QIAquick PCR Purification Kit (#28106, QIAGEN). The manufacturer's guidelines were followed.

Gel-electrophoresis

Typically a 1% agarose 1× TAE (40 mM Tris Base·HCl, 20 mM glacial acetic acid, 1 mM EDTA pH 8.0) gel containing 50 µg ml^{-1} ethidium bromide was cast. DNA was mixed with NEB purple loading dye to 1× (2.5% Ficoll, 10 mM EDTA, 3.3 mM Tris Base·HCl pH 8.0, 0.08% SDS, 0.02% Dye 1, 0.001% Dye 2) and DNA was fractionated in 1× TAE at 100 V for 40 minutes. Gels were imaged using a Syngene InGenius bioimaging system.

Gel Extraction and Purification of DNA

Gel extraction was conducted using QIAquick Gel Extraction Kits (QIAGEN) as per the manufacturer's guidelines. A supplied buffer (QG) solubilises the agarose gel slice when heated at 50 °C for 10 minutes. Isopropanol is added to increase the yield of fragments less than 500 bp and greater than 4 kb. The solution is then bound to the silica membrane on the column as detailed in *Plasmid Extraction from Escherichia coli*. DNA is eluted from the column in 20-50 µl 1× TE.

DNA Sequencing

DNA sequencing up to 1 kb was performed by GATC Biotech.

Nucleic Acid Quantification

Nucleic acid quantification was performed using a Nanodrop 1000 Spectrophotometer (Thermo Scientific) or by a Qubit 2.0 Fluorometer (Thermo Scientific) as per manufacturer's guidelines.

Site-directed mutagenesis

Overlapping forward and reverse primers of approximately 30 bp were designed containing the desired sequence mutation (see Table 2.3 for strain specific primers). A solution was made containing 10.5 µl water, 0.5 µl plasmid to be mutagenised, 0.5 µl 25 µM forward mutating primer, 0.5 µl 25 µM reverse mutating primer and 12.5 µl 2× Phusion High-Fidelity PCR Master Mix (#M0531L; NEB). A PCR program was run consisting of: 95 °C for 5 minutes initial denaturation; 18 cycles of 95 °C for 30 seconds, 45 °C for 1 minute, 72 °C (1 minute per kb of mutating plasmid); 72 °C for 15 minutes. To remove unmutagenised template plasmid samples were incubated with 1 µl *Dpn*I restriction enzyme at 37 °C for 1 hour. To increase efficiency of bacterial transformation, excess reagents and primers were removed using a Nucleospin Gel and PCR Clean-up kit (#740609.50; Macherey-Nagel) as per the manufacturer's guidelines. The DNA binding buffer NT1, containing chaotropic salts to allow DNA to bind the silica column, was diluted 1 in 7 with water (a ratio which prevents DNA less than approximately 300 bp form binding the column) and 2 volumes added to size-select away from unwanted primers. Samples were washed using buffer NT3 (ethanol containing solution to wash away non-DNA molecules) and the DNA eluted off the column using 25 µl 70 °C pre-heated elution NE buffer (Tris·HCl pH 7.5) which was left incubating on the column for 5 minutes before spinning through at 11,000 × *g* for 1 minute. Samples were transformed into High Efficiency DH5α *E. coli* (#2987, NEB) as detailed in section *Bacterial Transformation*.

2.2.2 Bacterial transformation

An aliquot of bacterial DH5α cells (#C2987H, NEB) or homemade competent DH5α was removed from -80 °C storage and put on ice to defrost for 10 minutes. Between 5 and 50 µl of cells were transferred to a 1.5 ml tube and the transforming plasmid added to the tube at no more than 10% of the final volume. Mixing is done carefully using the pipette tip to swirl the cells. The cells were put on ice for 30 minutes, heat shocked at 42 °C for 40 seconds and then put back on ice for 5 minutes. 300 µl of 37 °C pre-warmed LB was added, the solution mixed and incubated at 37 °C for 30-60 minutes. The mix was then plated onto LB-Amp plates for selection of transformants.

2.2.3 Plasmid Extraction from *Escherichia coli*

Bacterial colonies were picked off LB-Amp plates using a p200 pipette tip and the tip placed inside a 50 ml centrifuge tube containing 5 ml LB with 100 μg ml^{-1} ampicillin. Tubes were incubated at 30 °C overnight in a shaking incubator. Tubes were centrifuged at 6,800 × *g* for 3 minutes and the supernatant aspirated off.

To obtain plasmid DNA from *E. coli,* QIAprep Spin Miniprep Kits (#27106; QIAGEN) were used. The protocol was used as described in the manufacturer's guidelines. Bacteria are lysed under alkaline conditions and the lysate is subsequently neutralized and adjusted to high-salt binding conditions. The lysate is cleared and applied to the silica membrane where the DNA is adsorbed, while RNA, cellular proteins and metabolites are not. DNA is washed using an ethanol containing solution to remove salts and eluted in 50 μl 1× TE. The eluate is frozen at -20 °C until use with 2 μl loaded onto a 1% agarose gel in × TAE buffer to check the plasmid size and purity.

2.2.4 Meiotic culture

Yeast strains were woken up from -80 °C storage, streaked onto YPD plates or YPD-HYG (for centromeric plasmid containing strains) plates and incubated at 30 °C for 3 days. A single colony was inoculated into 4 ml liquid YPD and incubated at 30 °C at 250 rpm for 24 hours. Cell density (OD_{600}) was measured and 50-250 ml YPA was inoculated to a starting density of 0.2 OD_{600} and the culture incubated at 30 °C at 250 rpm for 14 hours 30 minutes. The culture was spun at 4 °C at 6,000 × *g* for 5 minutes, resuspended in an equal volume (to the initial YPA volume) of double distilled water, re-spun and then resuspended in and equal volume of 30 °C pre-warmed sporulation media (SPM). The culture was incubated at 30 °C at 250 rpm for the duration of the time course with samples taken at relevant time points.

2.2.5 Protein induction during meiosis

Strains containing the *GAL4-ER* fusion-protein expression cassette and a gene under the expression of a galactose promoter had protein expression induced by addition of 1.5 μM β-estradiol (in 100% ethanol) into the sporulating media.

2.2.6 Tetrad dissection

50 µl of sporulated cells in sporulation media was added to 150 µl 200 mM sodium phosphate buffer in a 1.5 ml tube. 1 µl of 1 mg ml^{-1} zymolyase 100T solution (10 mM Sucrose, 0.7% Glucose, 1 mM HEPES pH 7.5, 1 mg Zymolyase 100T) is added, the tube flicked to resuspend and incubated at 37 °C for 15 minutes. 10 µl is pipetted onto a YPD plate and left to dry before tetrads are dissected.

2.2.7 Yeast transformation

5 ml of YPD per transformation was inoculated to a cell density of 0.4 OD_{600} with overnight YPD culture and grown at 30 °C at 250 rpm for 4 hours to a cell density of approximately 1.5 OD_{600}. Cells were spun down at 1,200 × *g* for 3 minutes, 4 °C. The supernatant was poured off and the pellet resuspended in 1 ml of ice-cold 100 mM lithium acetate, transferred to 1.5 ml tubes and spun at 2,600 × *g* for 1 minute. The lithium acetate was removed and this process was repeated. The pellet was resuspended in 30 µl 100 mM lithium acetate per transformation with 40 µl of the mix then transferred into 1.5 ml tubes. A transformation mix (33.3% PEG, 100 mM lithium acetate, 0.28 mg ml^{-1} single stranded DNA (salmon sperm), between 0.1-10 µg transformation DNA fragment) was added to each tube. Tubes are vortexed until pellet is fully resuspended then incubated at 30 °C for 30 minutes. Cells are heat shocked at 42 °C for 30 minutes then 1 ml double distilled water is added. Cells are spun down at 25 × *g* for 1 minute, supernatant aspirated and pellet resuspended in 300 µl YPD. Cells are then plated onto a YPD plate, grown for 12 hours and replica plated onto drug-containing YPD plates using velvet. These replica-plated cells are then grown for 2-3 days before transformants are selected.

2.2.8 Spot tests

YPD cultures were diluted 40-fold into fresh YPD and grown vigorously at 30 °C for 4 hours. Cultures were then diluted to 0.2 OD_{600} in distilled water and a 10-fold dilutions series down to 0.00002 was made. 7 µl of each dilution was spotted in series onto the stated drug-containing YPD plates which were incubated at 30 °C for 4 days.

2.2.9 Genomic DNA preparation

Non-proteolysing gDNA extraction

Spheroplasts were prepared from frozen meiotic cell culture aliquots. Cell pellets were

resuspended in 300 µl spheroplasting buffer for every 10 ml of cells (1 M sorbitol, 50 mM $NaHPO_4$ pH 7.5, 10 mM EDTA, 1% β-mercaptoethanol, 200 µg ml^{-1} zymolyase 100T) and incubated at 37 °C for 20 minutes. Samples were mixed by inversion every 5 minutes during this incubation. Samples were immediately placed on ice and Sigma P8215 protease inhibitor cocktail and Pefabloc-SC (Roche) were added at 1× concentration. Samples containing greater than 10 ml of cells were split into eppendorfs to contain 10 ml of cells each. Spheroplasts were fixed with 1 ml ice-cold 100% ethanol, inverted to mix and placed on ice for 10 minutes. Fixed spheroplasts were spun down at 1,000 × *g* for 1 minute, ethanol aspirated off, spun again and residual ethanol completely removed. Pelleted material was resuspended in STE lysis buffer (2% SDS, 0.5 M Tris pH 8.1, 10 mM EDTA, 0.05% bromophenol blue) using a pestle, incubated on ice for 5 minutes then lysed at 65 °C for 10 minutes. After, samples were left to reach room temperature for 15 minutes before proceeding. Non-covalently bound proteins were removed with an equal volume (or smaller volumes detailed for specific experiments) of phenol:chloroform:isoamyl alcohol (25:24:1), mixing vigorously and left to settle for 5 minutes. Samples were vigorously mixed again and then spun at 20,000 × *g* for 5 minutes. 500 µl of the top aqueous phase was taken using a cut P1000 tip and added to a new eppendorf tube containing 500 µl of 100% ice-cold ethanol and mixed by inversion to precipitate the nucleic acids. Precipitates were pelleted by centrifugation for 1 minute at 16,000 × *g*, washed in 1 ml 70% ethanol and left to air dry for 15 minute before being dissolved in 400 µl 1× TE (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA) overnight at 4 °C. RNase was added at 100 µg ml^{-1}, incubated at 37 °C for 1 hour and reprecipitated with ice-cold 100% ethanol. DNA pellets were left to dissolve in 200 µl 1× TE overnight at 4 °C.

Proteolysing gDNA Extraction

The proteolysing form of genomic DNA extraction mirrors the non-proteolysing method except for the following changes: No protease inhibitors were added after spheroplasting; cells were not fixed in ethanol; cells were lysed by addition of lysis solution (3% SDS, 100 mM EDTA, 1 mg ml^{-1} Proteinase K) to the spheroplasting solution and placed overnight at 65 °C.

2.2.10 Southern blot

Approximately 2 µg of genomic DNA was digested at 37 °C overnight using *Pst*I restriction enzyme in NEBuffer 3.1 (100 mM NaCl, 50 mM Tris Base·HCl pH 7.9, 10 mM $MgCl_2$, 100 µg ml^{-1} BSA). Additional *Pst*I was added for 4 hours before the addition of NEB purple loading dye to 1×. Digested samples were proteolysed using 1 mg ml^{-1} Proteinase K at 60 °C for 30 minutes (unless the proteolysing gDNA extraction method was used), left to reach room

temperature before was 10 µg loaded on a 0.7% 1× TAE agarose gel (40 mM Tris Base·HCl, 20 mM glacial acetic acid, 1 mM EDTA pH 8.0) containing 50 µg ml^{-1} ethidium bromide. DNA was fractionated in 1× TAE at 60 V for 18 hours. The gel was imaged using Syngene InGenius bioimaging system to check migration and then exposed to 1800 J/m^2 UV in the Stratalinker. The gel was then soaked in three times its volume of denaturation solution (0.5 M NaOH, 1.5 M NaCl) for 30 minutes and then transferred to Bio-rad zetaprobe membrane by means of a vacuum at 55 mBar for 2 hours. After transfer the membrane was washed in water ten times and then cross-linked by exposing the membrane to 1880 J/m^2 UV in the Stratalinker. The membrane was incubated in 30 ml of hybridisation buffer (0.5 M $NaHPO_4$ pH 7.5, 7% SDS, 1 mM EDTA, 1% BSA) at 65 °C for 1 hour. The *MXR2* probe for looking at the *HIS4::LEU2* locus was created from 50 ng of template DNA, 0.1 ng of NEB Lambda DNA digested with *BstE*II, and water. The mix was denatured at 100 °C for 5 minutes then put on ice. Roche High Prime was added in addition to 0.5-3 mBq of α-^{32}P dCTP and incubated at 37 °C for 15 minutes. 30 µl 1× TE was added and the probe spun through a GE Healthcare G-50 spin column at 400 × *g* for 2 minutes. The probe was then denatured by incubating at 100 °C for 5 minutes and then put on ice before being added to 20 ml hybridisation mixture. The original 30 ml hybridisation buffer was discarded and the 20 ml containing the probe was added to the membrane and incubated overnight at 65 °C. After incubation, the membrane was washed five times with 100 ml pre-warmed southern wash buffer (1% SDS, 40 mM $NaHPO_4$ pH 7.5, 1 mM EDTA) and exposed to phosphor screen overnight.

2.2.11 Covalently linked protein-DNA molecule enrichment

This protocol describes the final procedure that was created after optimisations were carried out as detailed in *Chapter 5: Genome-wide mapping Spo11-DSBs*. Genomic DNA was purified as detailed in section *Genomic DNA preparation (non-proteolysing)* with phenol:chloroform:isoamyl alcohol volumes sometimes altered as stated in specific figures. DNA was fragmented using two different schemes. Firstly, sonication using a Covaris M220 machine (Settings: target BP range 200-700 bp, duty cycle 10%, intensity/peak power incidence 75W, cycles/burst 200, bath temperature 7 °C, time 24 minutes (time length of sonication varied between experiments as detailed in figures). Secondly, restriction enzyme digestion made with a mix of 125 µg DNA, 16 µl NEBuffer, 56 µl double distilled water, and 8 µl restriction enzyme of choice. The mix was placed at 37 °C for 4 hours. To either the sonicated or the digestion mix a column binding buffer was added (0.1% sarkosyl, 0.2% Triton-X100, 0.3 M NaCl) and samples pipetted onto QIAprep 2.0 Spin Columns (QIAGEN) at 20 µg of approximated starting genomic DNA per column. Under these conditions protein

molecules (and protein molecules with DNA covalently attached (Spo11-DNA)) bind to the silica membrane of these columns whereas DNA molecules do not. DNA was incubated at room temperature on the column for 5 minutes before being centrifuged at 16,000 × *g* for 1 minute. Flow-through was put back onto the column, incubated for 5 minutes and spun through again. Flow-through was discarded and wash buffer (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.3 M NaCl) was applied to the column and left to incubate for 1 minute. Columns were then either spun as before or applied to a QIAvac 24 Plus manifold and the vacuum switched on. These washes were repeated six times before all the columns were spun at 16,000 × *g* for 1 minute to remove any residual wash buffer. Protein-linked DNA was eluted from the column by applying 50 µl TES (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.5% SDS) directly to the silica membrane, incubating it for 5 minutes at room temperature and eluting the solution through centrifugation at 16,000 × *g* for 1 minute. Another 50 µl was added and eluted to increase yield of protein-DNA molecules. Protein-DNA molecules were then proteolysed using 1 mg ml^{-1} Proteinase K at 60 °C for 1 hour before being precipitated overnight (200 mM NaOAc, 20 µg glycogen, 2.5 volumes ice-cold ethanol) at -80 °C. Samples were centrifuged at 4 °C, 21,000 × *g* for 60 minutes, ethanol aspirated, 1 ml 70% ethanol added to wash samples, span at 21,000 × *g* for 10 minutes, 70% ethanol aspirated and left to air dry for 15 minutes. Pellets were resuspended in specific buffer for subsequent experiments.

2.2.12 Spo11-DNA mapping

This mapping protocol describes the final version used for mapping single-cut Spo11 and Top2 molecules. The optimisation experiments used to get to this final version can be found in Chapter 5.

Precipitated, column-enriched DNA fragments from a 50 ml starting cell volume were resuspended in 55 µl Tris Base·HCl pH 8.0 with 5 µl used to measure DNA concentration on a Qubit 2.0 machine. The majority of steps now follow the NEBNext Ultra II DNA Library Prep Kit for Illumina with a few alterations (Figure 2.2 – Library preparation workflow).

NEBNext End Prep

50 µl Fragmented DNA was mixed with 7 µl NEBNext Ultra II End Prep Reaction Buffer and 3 µl NEBNext Ultra II End Prep Enzyme Mix. The sample was placed in a thermocycler, with heated lid set to ≥75 °C, and run with the following program: 20 °C for 30 minutes, 65 °C for 30 minutes, hold at 4 °C. The sample should then be processed for adaptor ligation immediately.

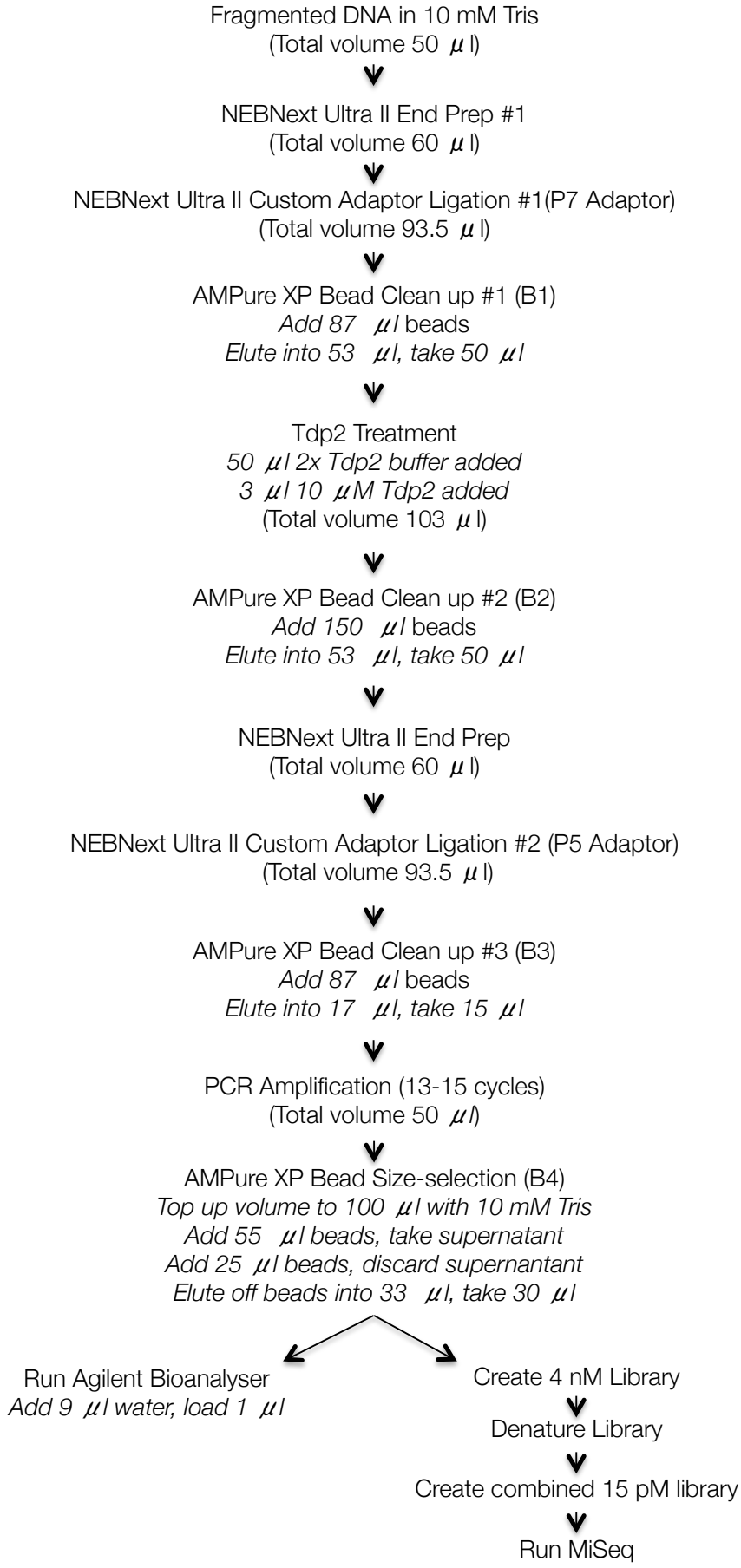


Figure 2.2: Workflow of single-cut Spo11 Mapping Protocol. A detailed method for mapping Spo11-DSBs is located within the text

Adaptor Ligation

The following components were added directly to the End Prep Reaction Mixture: 30 µl NEBNext Ultra II Ligation Master Mix, 1 µl NEBNext Ligation Enhancer, 2.5 µl Custom Adaptor P5 or P7*. The sample was placed in a thermocycler at 20 °C for 15 minutes with the heated lid off.

*Adaptors (stock is 15 µM) must be diluted as per Table 2.5 below:

Table 2.5 **Working adaptor concentration dilution table**

INPUT	Adaptor Dilution (Volume of Adaptor:Total Volume)	Working Adaptor Concentration
1 µg-101 ng	No dilution	15 µM
100 ng-5 ng	10-Fold (1:10)	1.5 µM
Less than 5 ng	25-Fold (1:25)	0.6 µM

AMPure XP Beads Clean-up of DNA

Volume/concentration of AMPure XP beads used for clean-up and size-selection varied depending on the stage of the protocol. See workflow (Figure 2.1) for volumes used at each specific stage. The following method describes a general clean-up (buffer swap) of fragmented DNA.

1.5× volume of AMPure XP beads was added to the sample and pipetted up and down ten times to thoroughly mix and incubated at room temperature for 7 minutes. The tube was briefly spun down to collect the liquid from the sides of the tube before placing on a magnetic stand. After the solution cleared (approximately 5 minutes) the supernatant was removed and discarded. 200 µl of 80% freshly prepared ethanol was added to the tube whilst in the magnetic stand. The tubes were incubated at room temperature for 30 seconds, and the ethanol removed and discarded. This step was repeated for a total of two washes. The beads were air dried for 3 minutes whilst on the magnetic stand with the lid open. The tube was removed from the magnetic stand, the buffer required for the next stage of the protocol added, the tube vortexed and pipette up and down until the beads are fully resuspended. The tubes were incubated at room temperature for 3 minutes, residual liquid from the sides of the tube spun down and the tube placed back onto the magnetic stand for 5 minutes until the solution cleared. The supernatant was removed minus 2 µl (to prevent beads being taken) and placed into a fresh tube ready for the next stage of the protocol.

AMPure XP Beads Size-Selection

For single-cut Spo11-DNA fragments post-PCR amplification size-selection of 150 to 600 bp was required. The volume of the sample was brought up to 100 µl with 10 mM Tris Base·HCl pH 8.0. 55 µl AMPure XP beads was added, the sample was mixed thoroughly, incubated at room temperature for 7 minutes and residual liquid spun down from the sides of the tube. The tubes were placed on a magnetic stand for 5 minutes until the solution cleared. The supernatant was transferred to a fresh tube not on the magnetic stand (beads discarded), 25 µl of fresh AMPure XP beads added, the tube mixed thoroughly, incubated at room temperature for 7 minutes and residual liquid from the sides of the tube spun down. The tube was placed back onto the magnetic stand for 5 minutes until the solution cleared. The supernatant was removed and discarded and beads washed twice with 80% ethanol as for the *AMPure XP Beads Clean-up of DNA* method (above). Size-selected, fragmented DNA was eluted in 33 µl 0.1× TE (1 mM Tris Base·HCl pH 8.0, 0.1 mM EDTA) and 30 µl transferred to a fresh tube. 1 µl of this final library was diluted 10-fold with double-distilled water and 1 µl ran on an Agilent Bioanalyser (high sensitivity chip) to check size distribution. The final library was stored at -20 °C.

TDP2 Treatment of fragmented DNA

To remove covalently bound 5′protein/peptide from DNA TDP2 (gift from Keith Caldecott) was used. Samples were resuspended and an equal volume of 2× TDP2 buffer (100 mM NaOAc, 100 mM TrisOAc, 2 mM MgOAc, 2 mM DTT, 200 µg ml^{-1} BSA) and TDP2 protein to 300 nM added. Samples were incubated in a thermocycler at 37 °C for 1 hour. Samples were subsequently cleaned up using AMPure XP beads before proceeding to the next step.

PCR Amplification of Adaptor Ligated DNA

15 µl of AMPure XP bead cleaned, fragmented, adaptor ligated DNA had the following components added: 25 µl NEBNext Ultra II Q5 Master Mix, 5 µl Index Primer/i7 Primer, 5 µl Universal PCR Primer/i5 Primer.

Within a batch of samples a unique Index Primer/i7 Primer was used for each individual sample. NEBNext Multiplex Oligos for Illumina Set 1 (NEB #E7335 lot 0091412) or Set 2 (NEB #E7500 lot 0071412).

The entire volume was pipetted up and down ten times to mix thoroughly and quickly spun to collect liquid from the side of the tube. The tube was placed in a thermo cycler and PCR amplification was performed using the following PCR cycling conditions (Table 2.6):

Table 2.6 **Overview of PCR amplification program**

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	13-15*
Annealing/Extension	65 °C	75 seconds	
Final Extension	65 °C	5 minutes	1
Hold	4 °C	∞	

*The number of cycles used depended on the input amount of DNA. 5 ng to 1 ng = 13 cycles, 1 ng to 0.5 ng = 14 cycles, ≤0.5 ng = 15 cycles.

\

Preparing Libraries for Sequencing on the MiSeq

Libraries were prepared as per the NEBNext manufacturer's guidelines. 1 ml 1 N NaOH was freshly prepared and a 0.2 N dilution made using double-distilled water. The DNA concentrations of the libraries were determined from the Bioanalyser trace and from the Qubit 2.0 machine. Libraries were diluted to 4 nM with double-distilled water. 2 µl of each library were pooled together in a fresh tube and an equal volume of 0.2 N NaOH was added. The samples were vortexed briefly, centrifuged at 280 × *g* for 1 minute and incubated at room temperature for 5 minutes to denature the DNA into single strands. 10 µl of the denatured DNA was added to 990 µl HT1 buffer to create a 20 pM library in 1 mM NaOH. For running the library on the MiSeq, the library was further diluted to 15 pM with HT1 buffer. 20 pM PhiX control was added to 1% total volume (see below for preparation).

Preparing PhiX Control for MiSeq library

2 µl of a 10 nM PhiX library was diluted to 4 nM with 3 µl of a solution of 10 mM Tris Base·HCl pH 8.5, 0.1% Tween 20. This had an equal volume of 0.2 N NaOH added, was vortexed briefly and centrifuged at 280 × *g* for 1 minute before being incubated at room temperature for 5 minutes. 10 µl was added to 990 µl HT1 buffer resulting in a 20 pM PhiX library which was stored at -20 °C for up to three weeks.

Sequencing

Sequencing was conducted on the Illumina MiSeq using the MiSeq Reagent Kit v3 (150 cycle) (Illumina MS-102-3001).

2.2.13 Western blot

Polyacrylamide gels were homemade using the following procedure. Usually an 8.75% polyacrylamide resolving phase (375 mM Tris Base·HCl pH 8.8, 0.1% SDS, 0.1% ammonium persulfate, 0.01% TEMED) was made, poured into a casting kit and 100% isopropanol used to level the top of the gel. After the gel had set (approximately 30 minutes), the isopropanol was poured off and the kit rinsed in distilled water. The 4% polyacrylamide stacking gel (125 mM Tris Base·HCl pH 6.8, 0.1% SDS, 0.1% ammonium persulfate, 0.02% TEMED) was made and poured on top of the set resolving phase gel and a comb placed on top of the stacking phase. Once set (approximately 30 minutes), the gel was placed into an SDS-PAGE gel running tank, the tank was filled with 1× SDS-PAGE Running buffer (25 mM Tris Base·HCl pH 7.5, 200 mM glycine, 0.5% SDS) and the comb was removed. Protein samples had 2× Laemmli loading buffer (4% SDS, 100 mM Tris pH 6.8, 20% glycerol, 2 mM EDTA, 0.005% bromophenol blue) added, boiled at 100 °C for 5 minutes, placed on ice for 5 minutes, spun at 9,400 × *g* and 20 µl loaded into the gel wells. The gel was run at a 1500 V until samples have migrated to the desired position. The gel apparatus is disassembled and the gel is soaked in 1× CAPS buffer (10 mM CAPS·NaOH pH 11, 10% methanol). PVDF membrane (Millipore) is activated in 100% methanol for 5 minutes then soaked in 1× CAPS buffer for 10 minutes. The gel is captured upon the activated PVDF membrane, placed between Whatmann paper and setup within a wet transfer tank. Protein transfer is done in 1× CAPS buffer at 0.65 mA for 1 hour. The transfer apparatus is dismantled and the PVDF washed in 1× TBST (25 mM Tris Base·HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20). The membrane is blocked in 5% non-fat dried milk in 1× TBST for 30 minutes on a rocking machine before the blocking solution being removed and the primary antibody (which binds the protein of interest) in 1× TBST added (antibody concentration and incubation time varies between antibodies). The primary antibody was removed and the membrane washed three times in 1× TBST for 5 minutes each. If the primary antibody was conjugated to horseradish peroxidase (HRP) protein then the membrane had enhanced chemiluminescence (ECL) reagent added and exposed to film (see below). For standard Western blots the secondary antibody (concentration and incubation time varies between antibodies), conjugated to HRP protein, was added in 1× TBST added (which binds the primary antibody). The secondary antibody is removed and the membrane washed three times in 1× TBST for 5 minutes. ECL reagent is poured onto the membrane and left to incubate for 2 minutes before being poured off. The ECL reacts with the HRP on the secondary antibody, emitting low intensity light, which can be detected using the ImageQuant LAS4000 machine for varying exposure times.

2.2.14 Spo11-oligonucleotide assay

Spo11-oligonucleotide complexes were detected by immunoprecipitation and end-labelling following established methods (Neale et al., 2005). Specifically, 10 ml of sporulating culture was lysed in 10% ice-cold TCA using zirconia beads and a BioSpec 24. Precipitated material was dissolved in 300 µl STE (2% SDS, 0.5 M Tris pH 8.1, 10 mM EDTA, 0.05% bromophenol blue), and boiled at 100 °C for 5 minutes. Extracts were centrifuged at 16,000 × *g* for 1 minute at 4 °C, and supernatant was diluted with one volume 2× IP buffer (2% Triton X100, 300 mM NaCl, 30 mM Tris Base·HCl pH 8.1, 2 mM EDTA) and further two-fold in 1× IP buffer. Anti-FLAG antibody (#F1804; Sigma-Aldrich) was added at 1 in 500 dilution, protein-G-agarose matrix (Roche) at 1 in 50 dilution, and then incubated with rotation overnight at 4 °C. Immune complexes were collected by low speed centrifugation, and washed twice with 1× IP buffer. This was followed by two washes in 1× TKAc (20 mM Tris·acetate pH 7.9, 50 mM K·acetate) before incubation in 25 µl labelling buffer (10-20 units terminal deoxynucleotidyl transferase (TdT) (Fermentas), 0.5-3 mBq of α-^{32}P dCTP or dATP 3`-[a-32P] (cordycepin) (Perkin Elmer), 1× TdT buffer (500 mM potassium cacodylate pH 7.2, 10 mM $CoCl_2$, 1 mM DTT)) at 37 °C for 1 hour. 20 µl of the labelling buffer was removed and 50 µl ice-cold 2× Laemmli loading buffer was added, samples boiled for 7 minutes, chilled on ice for 5 minutes and spun at 9,400 × *g* for 1 minute. 20 µl was loaded onto a 7.5% or 8.75% SDS-PAGE gel (detailed in section *Western Blot*) before fractionation at 150 V for 110 minutes. The gel was transferred to PVDF membrane in 1× CAPS buffer at 0.65 mA for 1 hour and exposed to a phosphor screen.

2.2.15 Oligonucleotide sequencing gel

For nucleotide resolution analysis of Spo11-oligonucleotide lengths, the Spo11-oligonucleotide protocol was followed up to the labelling step with TdT. Then, the labelling solution was completely removed using a gel loading tip, beads washed in 1 ml 1× IP buffer and 50 µl 1× TE added containing 1 mg ml^{-1} Proteinase K. Samples were proteolysed at 60 °C for 1 hour. 1 volume of Laemmli buffer was added and the mixture was boiled at 100 °C for 5 minutes, left on ice for 5 minutes and spun down at 9,400 × *g* for 1 minute. All the supernatant was taken and DNA precipitated with the addition of 20 µg glycogen (R0561; Thermo Scientific) and 10 volumes of 100% ice-cold ethanol and placed at -80 °C overnight. Precipitates were collected by centrifugation at 16,000 × *g* for 15 minutes and the pellet dissolved in 20-50 µl 1× TE. The oligonucleotides were mixed with 1 volume of sequencing gel loading dye (95% formamide, 10 mM EDTA, 0.01% xylene cyanol), and fractionated

through a 28 cm tall, 0.5 mm thick 19% polyacrylamide (19:1)/6 M urea gel (or different percentage polyacrylamide gel as stated) in 1× TBE (90 mM Tris Base·HCl pH 8.0, 90 mM boric acid, 2 mM EDTA) running buffer at approximately 1200 V for 80 minutes. Gels were fixed in fixing buffer (10% methanol, 7% acetic acid, 5% glycerol) and exposed to a phosphor screen for imaging.

2.2.16 Bioinformatics

All bioinformatics scripts can be found in the Appendix and were developed by either Matt Neale or Tim Cooper as stated.

CHAPTER 3:

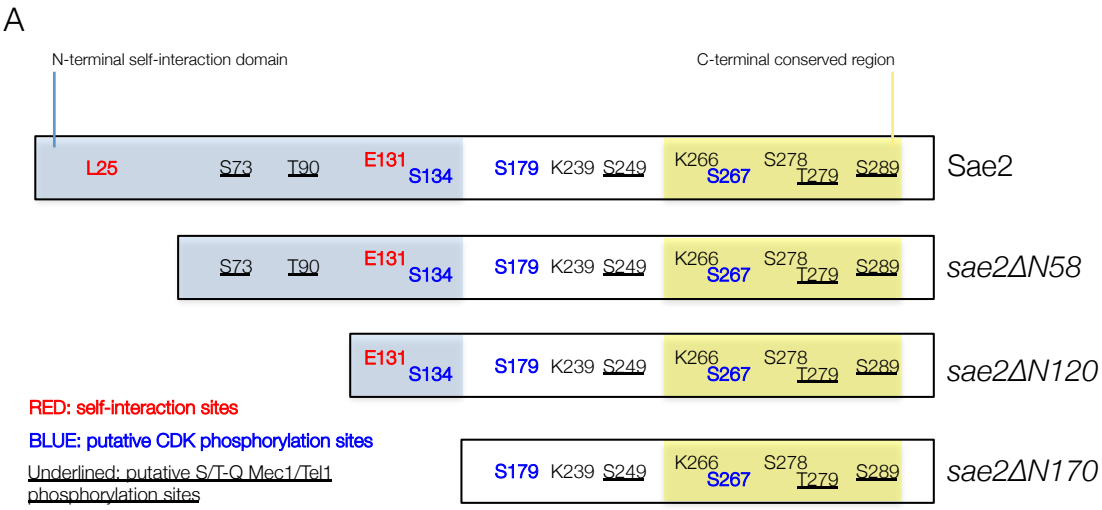
INVESTIGATING SAE2 ACTIVITY IN MEIOTIC DSB END PROCESSING

Chapter 3: Investigating Sae2 activity in meiotic DSB end processing

3.1 Introduction

Sae2 physically interacts with the MRX complex and is required for activation of Mre11 endonuclease activity, an activity essential for processing blocked-ended DSBs (Cannavo and Cejka, 2014). At Spo11-DSBs the combined action of Mre11 and Sae2 generates Spo11-oligonucleotides from nucleolytic end-processing of the DSB (Garcia et al., 2011; Moreau et al., 1999). Such clipping of the 5′ strand adjacent to the DSB followed by bidirectional resection generates a long 3′ ssDNA tail and the release of Spo11-oligonucleotides. Spo11-oligos are formed with two distinct length distributions and can be detected by immunoprecipitation of a tagged form of Spo11, end-labelling the oligonucleotide and separating the Spo11 protein-DNA species on an SDS-PAGE gel (as described in Materials and Methods) (Neale et al., 2005). This method of detecting Spo11-oligos can be utilised to analyse, semi-quantitatively, mutants of the end-processing machinery, such as Sae2, for their ability to process Spo11-DSBs in meiosis.

Mitotic cells deficient in Sae2 (or other members of the HR pathway) are sensitive to methyl methanesulfonate (MMS) (McKee and Kleckner, 1997; Rattray et al., 2001), which generates single-stranded breaks resulting in DSBs through replication fork collision (Ensminger et al., 2014; Lundin et al., 2005). MMS can therefore be used in spot tests to study any potential Sae2 mutant phenotypes in cycling cells. In mitotic cells the activation of Mre11 endonuclease by Sae2 is a crucial regulation point for the choice between repairing a DSB via NHEJ or HR, with the initial resection refractory to repair by NHEJ (Shibata et al., 2014). The Sae2 protein contains various important residues, some of which are phosphorylation sites, which control Sae2 activity (Figure 3.1A). Phosphorylation of Sae2 mediated by CDK activity, generates an active form of the protein, which promotes this endonuclease activity of Mre11 (Cannavo and Cejka, 2014; Huertas et al., 2008). Therefore Sae2 activity is the link between the cell cycle stage, CDK activity and the choice between NHEJ or HR repair. Other phosphorylation events and post-translational modifications have also been implicated in regulation of Sae2 activity (Baroni et al., 2004; Cartagena-Lirola et al., 2006; Fu et al., 2014). However, the role of post-translational modification state in the regulation of Sae2 activity in meiosis requires further elucidation.



B

```
ORF:                                    ATGGTGACTGGTGAAGAAAATGTGTATCTA...
S288C:                        ...TGTATGTGAG-ATGGTGACTGGTGAAGAAAATGTGTATCTA...
SK1:                          ...TGTATGTGGG-ATGGTGACTGGTGAAGAAAATGTGTATCTA...

GAL1:          ...GAGAAAAAACCCCGGATTCTAGA-      ATGGTGACTGGTGAAGAAAATGTGTATCTA...
GAL1(SK1):     ...GAGAAAAAACCCCGGATTCTAGATGTGGG-ATGGTGACTGGTGAAGAAAATGTGTATCTA...
GAL1(S288C): ...GAGAAAAAACCCCGGATTCTAGATGTGAG-ATGGTGACTGGTGAAGAAAATGTGTATCTA...
```

Figure 3.1: Schematic of Sae2 protein and alignment of *SAE2* in *S. cerevisiae* strains.
A. Residues discussed in this chapter are annotated. Self-interaction sites are written in red. Putative CDK phosphorylation sites (S/T-P) are written in blue. Putative Mec1/Tel1 sites (S/T-Q) are underlined. The N-terminal self-interaction domain is highlighted in blue. The yellow box indicates the C-terminal conserved region. The three truncation mutants used are displayed. **B.** The sequence of the *SAE2* ORF is indicated alongside the natural upstream sequences containing cryptic ORFs (highlighted) from both SK1 and S288C strains of *S. cerevisiae*. The sequence of *SAE2* and its upstream region from the different *GAL1* constructs are indicated.

3.2 The Spo11-end processing activity of Sae2 and the MRX complex can be measured using the high-sensitivity Spo11-oligo assay

In order to characterise the phenotype of *SAE2* point mutants during meiosis, the *SAE2* gene was placed upon a centromeric plasmid with 500 nt flanking either side of the ORF. The benefits of this were: (i) Site-directed mutagenesis was very easy to perform on the gene (as described in Materials and Methods section *Site-directed mutagenesis*). (ii) Sae2 is not essential for growth of *S. cerevisiae* cells so therefore the natural *SAE2* locus could be deleted. This construct permits for complementation of the *sae2*Δ strain with the centromeric plasmid containing the *SAE2* gene and the mutant forms. Plasmid retention is selected using a drug-resistance marker (hygromycin) contained on the plasmid. (iii) The 500 nt flanking sequence either side of the ORF ensures *SAE2* will still be controlled by its natural promoter and terminator, keeping transcriptional control of the plasmid-borne gene similar to that of the chromosomal locus. (iv) Tagging the Sae2 protein with different affinity tags to monitor its expression levels is much easier than attempting to incorporate a tag at the natural *SAE2* locus. Details of the construction of the centromeric *SAE2* plasmid can be found in Materials and Methods Figure 2.1. To determine whether expression of wild type *SAE2* off the centromeric (CEN) plasmid could rescue the *sae2*Δ meiotic defect (deficient in Spo11-oligo formation), *sae2*Δ cells containing the *SPO11-FLAG* construct were transformed with the *SAE2-CEN* plasmid. Samples from the *SAE2-CEN* strain, alongside *SPO11-FLAG* containing wild type and *sae2*Δ strains, were taken from synchronised meiotic cultures at mid-meiotic prophase and Spo11-FLAG immunoprecipitated from cellular extracts. Immunoprecipitated Spo11-oligonucleotides were radioactively end-labelled and resolved on an SDS-PAGE gel (Figure 3.2).

Wild type Spo11-oligos (two discrete bands 24-40 and 10-15 nt in length) are present in the wild type strain confirming the results of Neale et al., 2005 that Spo11 is released from DSB covalently attached to two oligonucleotide lengths (Figure 3.2A left hand panel). As previously reported, in *sae2*Δ cells these two discrete bands are absent indicating a defect in Spo11-DSB end processing (Figure 3.2A middle panel). Larger bands with a periodicity of 10 nt starting at 45 nt in length were seen in all strains including *sae2*Δ cells. Longer Spo11-oligo species were also observed in Garcia et al., 2011 and presumed to be due to nicking of the 5′ strand up to 300 nt away from the DSB. However, subsequent unpublished experiments have determined that these species arise in the *mre11-H125N* (nuclease dead) and *rad50S* backgrounds (data not shown, Matt Neale and Valerie Garcia). Therefore, because Sae2 is thought to be an essential activator of Mre11 endonuclease activity (Cannavo and Cejka, 2014), and because the *rad50S* mutation also renders the Mre11 endonuclease inactive, it would seem that these molecules are not due to nicking by the Mre11 endonuclease. Instead

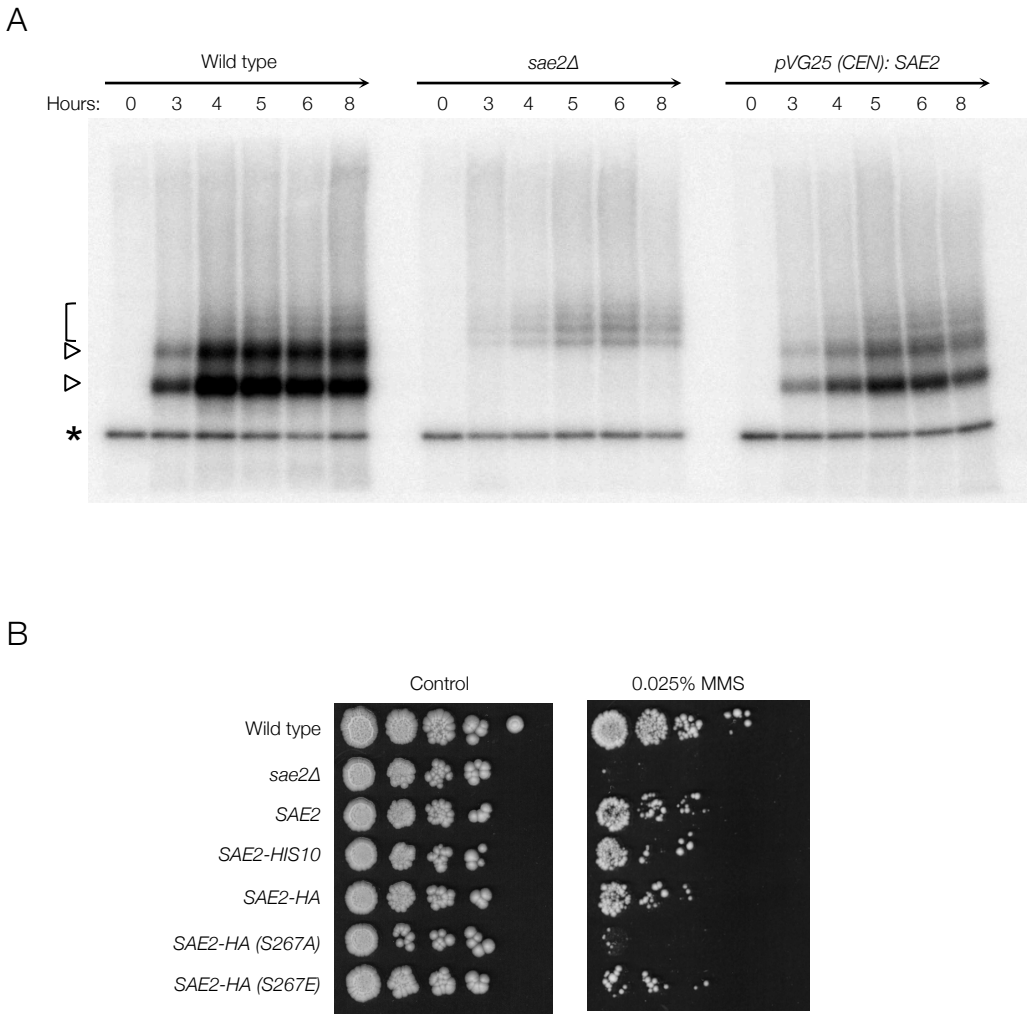


Figure 3.2: Two discrete lengths of Spo11-oligonucleotides are generated in wild type but not *sae2*Δ cells. *SAE2* expressed from a centromeric plasmid complements the *sae2*Δ deficiency.
A. Spo11-oligonucleotide blot from wild type, *sae2*Δ and *sae2*Δ cells transformed with a centromeric plasmid containing *SAE2* under control of its natural promoter and terminator (*pVG25*). Wild type and *sae2*Δ strains were also transformed with an empty hygromycin resistant plasmid for control purposes. 10 ml of cells were taken at the stated hour from the start of meiosis and Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in all strains. Asterisk marks non-specific TdT band. **B.** Spot tests. Wild type and *sae2*Δ strains were transformed with an empty hygromycin resistant plasmid for control purposes. All plates contained 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. MMS plates contained 0.025% MMS. Cultures were grown overnight in YPD (with 300 μg ml^{-1} hygromycin) then transferred to fresh YPD+HYG for 4 hours. Cultures were diluted to 0.2 OD_{600} and a 10-fold dilution series spotted onto the plate down to 0.00002, plates were incubated at 30 °C for 4 days and images taken using a Syngene InGenius bioimaging system.

these species are hypothesised to arise from the coincident formation of adjacent Spo11-DSBs on the same DNA molecule (this concept will be discussed in more detail in Chapter 6). *SAE2* expressed from the centromeric plasmid, under its natural promoter and terminator, is able to restore DSB end processing, as determined by the production of the two discrete lengths of wild type Spo11-oligonucleotides. However, the levels of wild type Spo11-oligos observed are less compared to cells expressing *SAE2* from its chromosomal locus (Figure 3.2A right hand panel).

To determine if the centromeric *SAE2* plasmid would also rescue *sae2*Δ mitotic DNA repair defects, cycling cells were serially diluted onto YPD and 0.025% MMS containing YPD plates and grown for four days. Centromeric-borne *SAE2* rescues the *sae2*Δ growth defect on MMS plates but, similar to the rescue of the meiotic phenotype, rescue was not to wild type levels (Figure 3.2B). Nevertheless, these assays will permit the analysis of mutant alleles of *SAE2*.

3.3 Determining the detection limit of wild type Spo11-oligonucleotide species

Because Sae2 is a critical component for meiotic DSB end processing it is expected that mutating the protein may alter its ability to generate Spo11-oligos. Because different mutations may have varying levels of effect on this process, the dynamic range of the Spo11-oligo assay was determined. To do this, different percentages of wild type meiotic cell extract were mixed with *sae2*Δ meiotic extract prior to Spo11-immunoprecipitation and end labelling to determine how little wild type Spo11-oligos can be detected when in the presence of cellular material (Figure 3.3). The detection limit was high, with wild type Spo11-oligos readily detected when only 10% of the meiotic cell extract volume originated from wild type cells (Figure 3.3 right hand panel). This enables the detection of Spo11-oligos across a range of hypomorphic *SAE2* alleles.

3.4 Affinity tagging Sae2 enables detection of protein levels from meiotic cultures

Mutating proteins for analytical purposes can often cause improper folding and/or expression resulting in a false negative result. To be able to determine expression levels of Sae2 once mutagenised, an epitope tagged version was used to monitor protein levels via Western blotting. The human influenza hemagglutinin (HA) tag is a commonly used epitope tag that facilitates in the detection, isolation and purification of proteins (Field et al., 1988). Combining three repeats of this peptide enhances its detectability and this 3HA sequence was cloned in

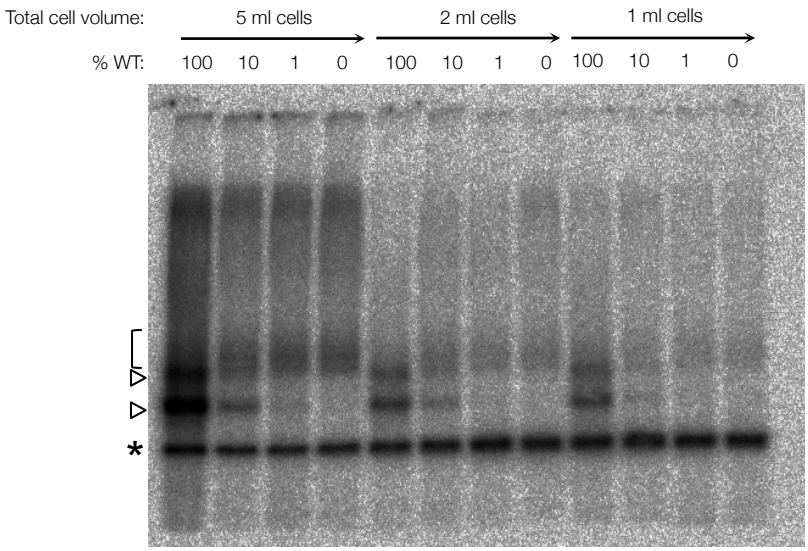


Figure 3.3: Wild type Spo11-oligonucleotide species are detectable down to a starting material of 0.1 ml cells.
Spo11-oligonucleotide blot. Stated amounts of wild type whole cell TCA extract were spiked into *sae2*Δ cell TCA extract and Spo11-oligonucleotides were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods (reagents were not scaled from the standard 10 ml starting material size). Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket (⌈) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

frame to the C-terminal end of the *SAE2* gene on the centromeric plasmid (Figure 2.1). A comparable His10 C-terminal tag was also generated.

Tagging Spo11 with a 3HA-tag generates a hypomorphic allele of the protein which alters DSB formation (Diaz et al., 2002; Gray et al., 2013). In order to determine whether the -3HA and -10His-tagged forms of Sae2 displayed hypomorphic phenotypes, meiotic cultures of these tagged, wild type forms of Sae2 were assayed for Spo11-oligo formation (Figure 3.4). The two wild type Spo11-oligo bands were present in all three strains expressing *SAE2* from a centromeric plasmid indicating untagged, -His10 and -3HA-tagged Sae2 are all capable of processing Spo11-DSBs (Figure 3.4). The levels of Spo11-oligos produced are similar between the *SAE2-3HA* and untagged constructs (Figure 3.4 left and middle panels), with the *SAE2-HIS10* construct being slightly hypomorphic (Figure 3.4 right hand panel). Cycling cells of these strains were also assayed for their sensitivity to MMS via spot testing as before (Figure 3.2B). Spo11-3HA is more sensitive to MMS compared to the untagged *SAE2* and Sae2-His10 constructs (Figure 3.2B). However, the *SAE2-HA* tagged form of Sae2 can be utilised to monitor expression levels of Sae2 mutants whilst having only a minor effect on wild type Sae2 activity *in vivo*.

To determine the relative expression level of tagged *SAE2* constructs, a Western blot was performed on extracts from meiotic cultures and probed for Sae2-HA or Sae2-His10 using anti-HA or anti-His antibodies. However, Sae2 protein was below the detection limit (data not shown). Expression levels of Sae2 from its natural chromosomal locus are only at ~100 molecules per cell (Fu et al., 2014). Such low level expression from the natural *SAE2* promoter, alongside the mild hypomorphic nature of expressing *SAE2* from the centromeric plasmid (Figure 3.2), suggests that Sae2 protein levels may be extremely low and therefore undetectable via Western blot from whole cell extract. An alternative approach that could be conducted on these cells would be to immunoprecipitate the Sae2 protein via their respective epitope tags and to repeat the Western, removing the potential inhibitory/masking effect the whole cell lysate has on detecting low level proteins. With this in mind, studying Sae2 function via mutagenesis was all conducted on the Sae2-3HA form of the protein to allow the potential to attempt to monitor the expression levels at a later date if required.

3.5 Over-expression of mutant alleles of *SAE2* permits observation of activity undetectable at low-level expression

In vitro work performed in the Cejka laboratory has demonstrated that some mutant forms of the Sae2 protein display *null*-like activity at low concentrations, but retain some residual activity when used at higher concentrations

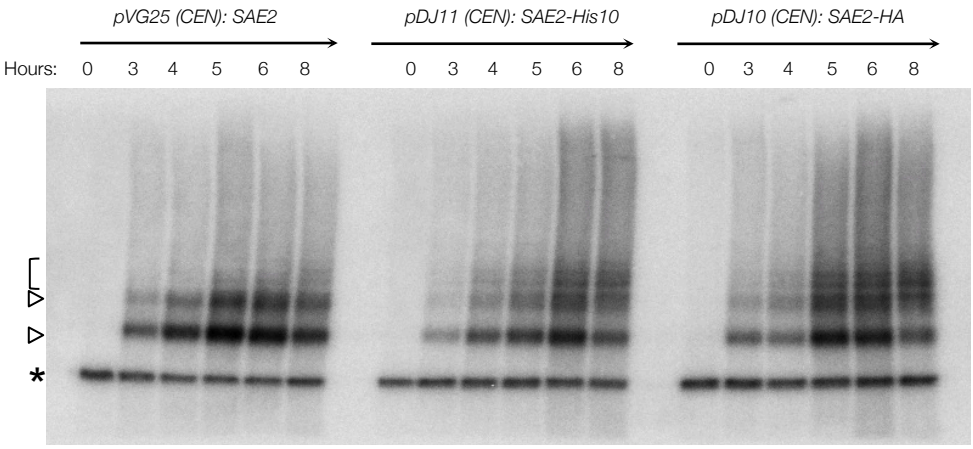


Figure 3.4: Spo11-oligonucleotides are generated in *sae2*Δ cells transformed with a centromeric plasmid expressing *SAE2*, *SAE2-HIS10* or *SAE2-HA*.
Spo11-oligonucleotide blot. The stated plasmids were transformed into *sae2*Δ cells with transformants selected using hygromycin resistance. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at the stated hour from the start of meiosis and Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket (⎡) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

(Cannavo and Cejka, 2014 - personal communication). To mimic this *in vivo,* a system capable of over-expression of *SAE2* alleles in meiosis was created. *SAE2-3HA* was placed under control of the *GAL1* promoter and integrated at the *SAE2* locus. Because addition of galactose inhibits meiosis, the strain contained a chimeric *GAL4-ER* cassette, which enables induction via the addition of β-estradiol to the media (Louvion et al., 1993). However, early induction also prevents sporulation, therefore meiosis was allowed to proceed for 4 hours to allow accumulation of Spo11-DSBs before induction by β-estradiol. Addition of β-estradiol led to an increase in the formation of Spo11-oligos (Figure 3.5A - compare 5, 6 and 8 hours with and without induction). This generation of Spo11-oligos correlated with the detectable expression of the Sae2-HA protein via Western blot (Figure 3.5B). This β-estradiol-controlled expression of *SAE2* was however leaky, with Spo11-oligo species visible at 3 and 4 hours before any induction of *SAE2* had occurred (Figure 3.5A – 3 and 4 hour lanes). These species were also present in the uninduced sample from 5 to 12 hours (Figure 3.5A - 5-12 h left hand panel).

To detect expression levels of Sae2-HA protein, Western blotting was used on whole cell extract from the meiotic time course (Figure 3.5B). The –HA tag fused to Sae2 was probed using anti-HA antibody. There was very little detectable protein from 3 hours onwards however, there is a relatively strong band observed at the 1 h period (Figure 3.5B). Expression of wild type *SAE2-3HA,* under control of its natural promoter from the centromeric plasmid, generated less Spo11-oligos (Figure 3.4 – right hand panel) compared to Spo11-oligos generated through leakiness of the *GAL1* promoter (Figure 3.5A – 3-5 h uninduced). Undetectable protein levels of Sae2-HA from the CEN plasmid (data not shown) suggest that at early time points (Figure 3.5B - 1-3 h) leakiness of expression from the *GAL1* promoter exceeds the basal expression levels of *SAE2* from its natural promoter. This leakiness allows for a large proportion of Spo11-DSBs to be processed. Nevertheless, further expression of *SAE2* did increase the levels of Spo11-oligos detected. This indicates that the leaky Sae2 protein levels are still somewhat limiting biochemically.

These results suggest expression of *SAE2* from a centromeric plasmid infers a hypomorphic phenotype for Spo11-oligo production potentially due to the cells containing a single copy of the centromeric plasmid and thus also the *SAE2* gene. Over-expression of *SAE2* using a *GAL4-ER* induction system results in leaky expression of *SAE2* prior to induction with β-estradiol permitting Spo11-oligo production but upon induction Spo11-oligo production is vastly increased.

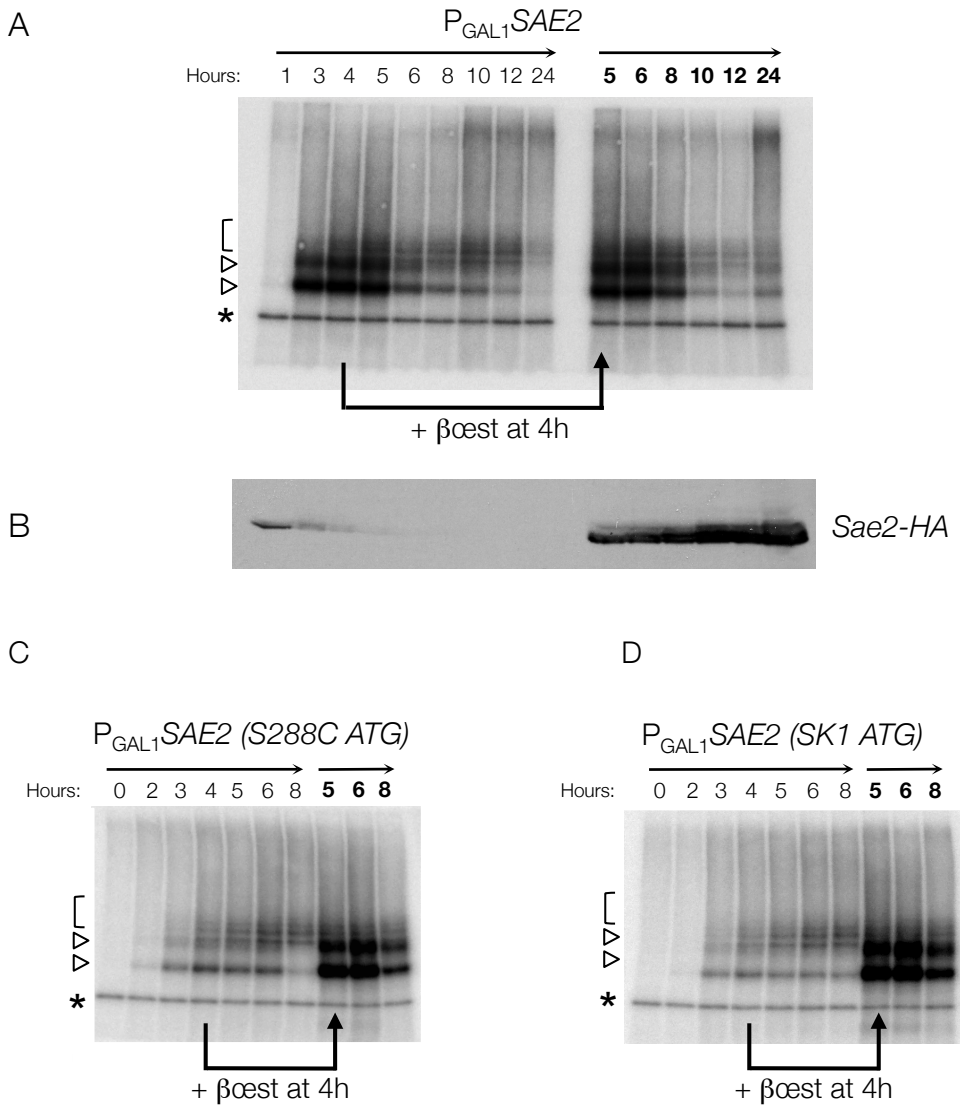


Figure 3.5: Galactose-inducible *SAE2* leakiness is decreased with the presence of a natural cryptic ATG from both S288C and SK1 strains.
10 ml of cells were taken at the stated hour from the start of meiosis and Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. *SAE2* expression was induced with the addition of 2 μM βœstradiol at 4 hours. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket (⟦) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band. **A.** *SAE2-HA* expression was put under control of a galactose-inducible promoter with the start site sequence reading ATGgtga. **B.** Western blot performed on the whole cell lysate and Sae2-HA detected using primary antibody anti-HA (mouse) at 1 in 5000 and secondary antibody anti-mouse (rabbit) at 1 in 5000. Chemiluminescence was detected using an ImageQuant LAS4000 machine **C.** *SAE2* expression was put under control of a galactose-inducible promoter with the natural cryptic ATG from S288C inserted, start site sequence reading ATGtgagATGgtga. **D.** *SAE2* expression was put under control of a galactose-inducible promoter with the natural cryptic ATG from SK1 inserted, start site sequence reading ATGtgggATGgtga.

3.6 A cryptic start site at the *SAE2* locus regulates expression of the protein

The relatively high level of basal (uninduced) expression from the *GAL1-SAE2* construct hinders the usage of this construct in a regulatable manner. From studying the sequence of the *GAL1-SAE2* construct used in Figure 3.5A-B and the sequence of the natural *SAE2* locus it was observed that the natural *SAE2* gene in SK1 and S288C variants of *S. cerevisiae* both contain an upstream ORF (cryptic ATG site) six base pairs upstream of the annotated start codon (Figure 3.1B). In *SAE2*, this cryptic start site has the potential to generate a two and a seven amino acid peptide in S288C and SK1 strains respectively, that always overlaps with the natural start site. Such upstream, overlapping ORFs may suppress expression/translation of the main ORF by limiting its access to the ribosome (Morris and Geballe, 2000). The cryptic ATG at the *SAE2* locus has the sequence ATGtgag**ATG**gtga in S288C strains and ATGtggg**ATG**gtga in SK1 strains with the second **ATG** coding for the annotated start of the *SAE2* gene. By contrast, the sequence of the *GAL1-SAE2* construct does not contain a cryptic ATG sequence because only the annotated ORF of the gene, without any upstream sequence, was cloned downstream of the *GAL1* promoter (Figure 3.1B). To investigate whether the inclusion of the cryptic ATG could help to reduce the (leaky) basal level of *SAE2* expression, constructs were generated with an additional six base pairs, either SK1- or S288C-like, between the *GAL1* promoter and the annotated *SAE2* ORF (Figure 3.1B) (Constructed by Nils Lambacher without the *-3HA* tag). Cells were sporulated as before, with addition of β-estradiol at 4 h to induce *SAE2* expression, and Spo11-oligos measured via IP, end labelling and separation via SDS-PAGE. Spo11-oligo production still occurred before and without induction (Figure 3.5C-D - see 3 and 4 hours before induction and 5-8 hours in uninduced cells for both strains), however, levels are significantly lower than in the original *GAL1-SAE2* expression strain (Figure 3.5A). After induction of *SAE2,* Spo11-oligo signal reached similar levels to those observed in the original *GAL1-SAE2* construct in both the S288C-like and SK1-like constructs indicating the expression system works as effectively as before. Of the two cryptic strains, the SK1-like construct had lower levels of Spo11-oligos before induction and in the uninduced cells compared to the S288C strain, with the induced signal slightly higher. Therefore, the SK1-like construct was utilised for inducible-expression of *SAE2* in subsequent experiments.

3.7 Over-expression of N-terminal truncations of Sae2 permits Spo11-end processing

The C-terminal region of Sae2 contains a conserved domain that is essential for most of Sae2 functions (Lengsfeld et al., 2007). The N-terminal region contains domains required for self-

interaction and oligomerisation, the function of which, have been studied in mitotic and meiotic cells (Kim et al., 2008; Lengsfeld et al., 2007). Deletion of the N-terminal region of Sae2, where the self-interaction residues (L25 and E131) reside, infer opposing phenotypes between mitotic and meiotic cells. Over-expression of the ΔN_{170} mutant of Sae2 supports sporulation in meiotic cells (Spo11 end processing was not directly assessed), but is unable to suppress MMS sensitivity (Kim et al., 2008). *In vitro,* the N-terminal truncation mutant is able to stimulate MRX endonuclease activity, although the concentration of Sae2 required was relatively high (Cannavo and Cejka, 2014).

Because stimulation of the MRX endonuclease for DSB end processing only requires the C-terminal domain of Sae2, these N-terminal truncations, when over-expressed, are expected to generate Spo11-oligos in meiotic cells. To test this prediction, inducible over-expression cassettes containing a range of N-terminal truncations were assessed for their ability to support Spo11-oligo production in meiotic cultures (Figure 3.6). All three truncation mutants (Δ*N58*, Δ*N120* and Δ*N170*) are capable of producing Spo11-oligos when over-expressed (Figure 3.6 MJ Neale – unpublished observations), supporting the conclusion that the N-terminal region is not essential for DSB end processing in meiotic cells. However, Spo11-oligo levels were progressively reduced the more truncated the Sae2 protein was. This reduction in Sae2 activity suggests that the N-terminal region influences the regulation of Sae2 activity. Self-interaction of Sae2 is thought to be critical for its activity (Kim et al., 2008), however, Fu et al., 2014 reported that the monomer form of Sae2 was the active component but did postulate that *in vivo* it was more likely that Sae2 acted as a dimer. Therefore, it may be that, *in vivo,* efficient Sae2 activity is dependent on dimerisation but when self-interacting mutants (N-terminal deletions or point mutations) are over-expressed the monomeric form of Sae2 has some residual activity capable of stimulating the endonuclease activity of the MRX complex.

As previously observed (Figure 3.5), expression of wild type *SAE2* via the *GAL4-ER* system still results in a leakiness of the promoter prior to induction causing Spo11-oligos to be observed before induction and in the uninduced cells (Figure 3.5). This observation suggests that even low levels of full-length Sae2 are capable of processing Spo11-DSBs. However, in all the N-terminal truncation mutants of Sae2 (Δ*N58*, Δ*N120* and Δ*N170*), there is no observable Spo11-oligos produced before induction or in the uninduced cells. This observation points towards a hypomorphic phenotype of Sae2 when it is unable to dimerise and/or oligomerise. Whilst monomeric Sae2 can still function, its ability to stimulate MRX is therefore drastically reduced.

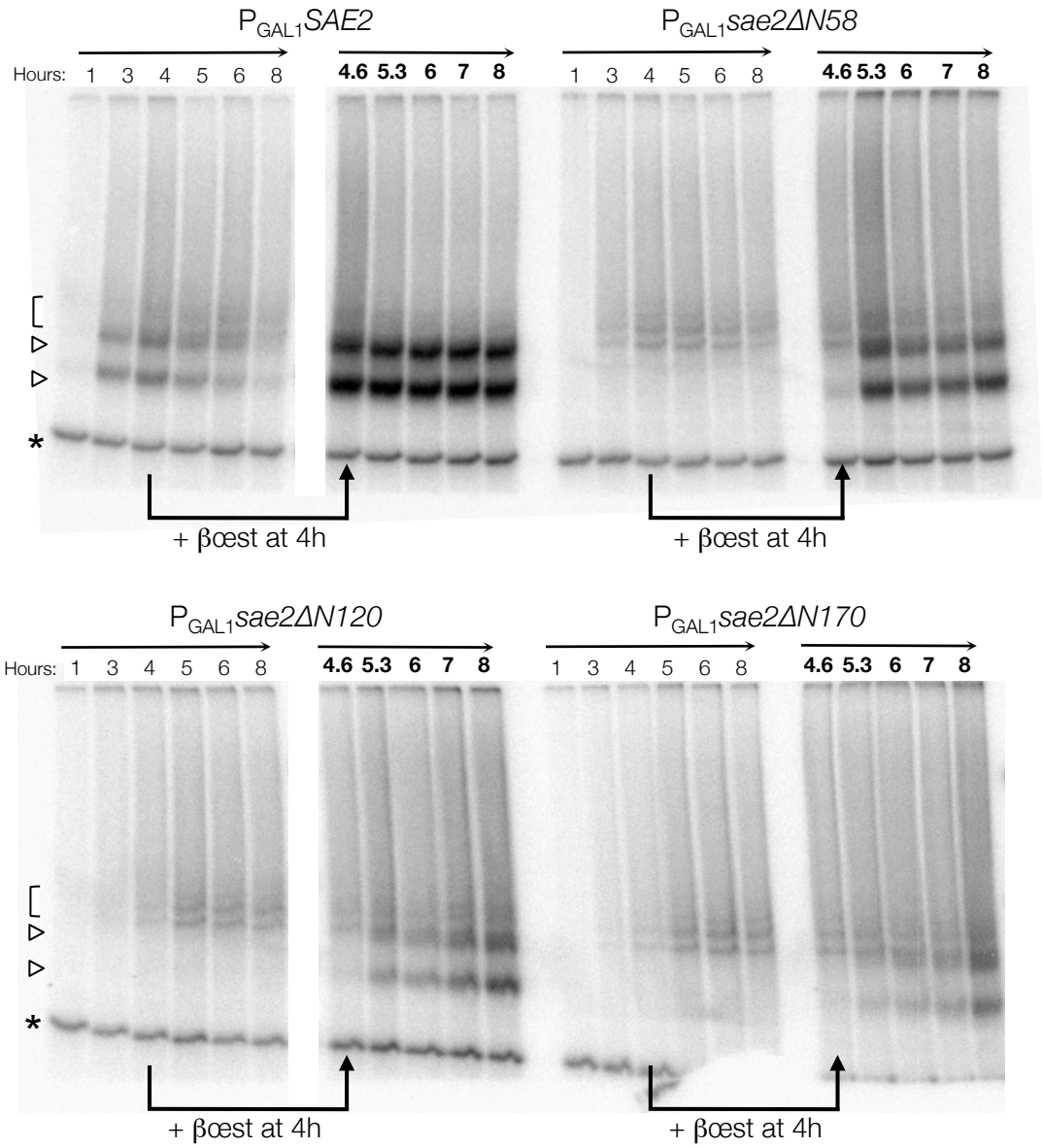


Figure 3.6: Increasing the length of an N-terminal truncation of Sae2 decreases Spo11-oligonucleotide production.
Sae2 was truncated as stated and placed under control of a galactose-inducible promoter. The strains contained the *GAL4-ER* hybrid cassette allowing for controlled induction of galactose-induced genes by addition of 2 μM βœstradiol at 4 hours. 10 ml of cells were taken at the stated hour from the start of meiosis and Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket (⟦) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

3.8 Mutation of the self-interaction residue L25 of Sae2 prevents Spo11-oligo formation

Mutation of L25 and E131 in the N-terminal region of *SAE2* causes MMS sensitivity and a reduction in sporulation efficiency (Kim et al., 2008). The L25 residue is important for inter-molecular interaction between itself and E131 (or a domain in the vicinity of E131). E131-E131 interactions were not observed (Kim et al., 2008). However, over-expression of the N-terminal truncation mutants (Figure 3.6) would suggest this self-interaction is not essential for Sae2 activity in meiosis, or at least the requirement can be overcome by over-expression. To test the requirement of these self-interaction residues for the ability of Sae2 to promote Spo11-oligo formation in meiosis, the *L25P* or *E131V* point mutants, that prevent self-interaction, were incorporated onto the centromeric plasmid and transformed into *sae2*Δ strains. Meiotic cultures were assessed for Spo11-oligo production as before (Figure 3.7). The *E131V* mutant had little effect on Sae2 activity with Spo11-oligo levels only slightly reduced (Figure 3.7A right hand panel). However, Spo11-oligos were completely absent in the *L25P* mutant - identical to the *sae2*Δ phenotype (Figure 3.7A - compare middle two panels). These observations support the conclusion that the loss of leakiness (observed Spo11-oligos) in the truncation mutants (Figure 3.6) stems from a reduction in the activity of the truncated protein due to loss of self-interaction. Likewise when expressed from the centromeric plasmid, which appears to reduce expression relative to the chromosomal locus (Figure 3.2A), the *L25P* mutant is unable to produce Spo11-oligos, therefore the low level of presumably monomeric Sae2 in this mutant is unable to activate MRX endonuclease activity.

If this was the case, then over-expressing the *L25P* mutant might be expected to result in a similar outcome to over-expression of the truncated forms of Sae2. To test this idea, the *L25P* mutant was incorporated into the over-expression cassette (SK1-like ATG) with *SAE2* expression of the meiotic culture induced with β-estradiol at 4 h (Figure 3.7B). However, unlike wild type *SAE2* (left hand panel), the *L25P* mutant failed to produce any Spo11-oligos after induction. This result suggests that the *L25P* mutant is deficient for any Spo11 end-processing ability during meiosis and that it can't be rescued by over-expression.

3.9 Phosphorylation of Sae2 at S267 by CDK is essential for Sae2 activity in meiosis

CDK phosphorylation of Sae2 at S267 is important for all of Sae2 cellular functions *in vivo* (Huertas et al., 2008; Manfrini et al., 2010) and *in vitro* (Cannavo and Cejka, 2014). The S267 site was identified by looking for putative consensus sequences for CDK phosphorylation (S/T-P) (Songyang et al., 1994, 1996; Srinivasan et al., 1995). Sae2 contains three such

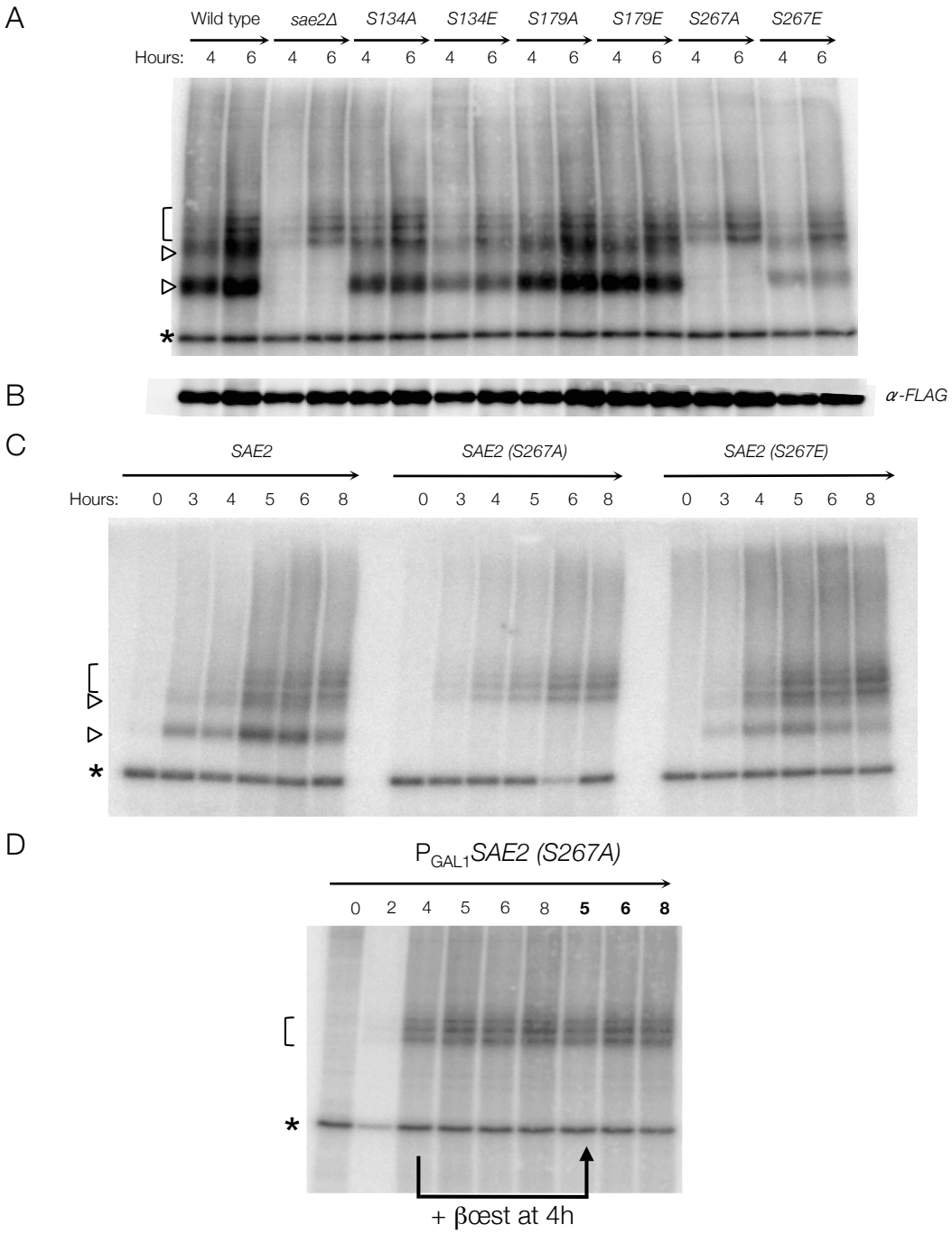


Figure 3.8: Only phosphorylation at S267 by CDK is essential for Sae2 DSB end-processing activity and not the other potential CDK phosphorylation sites.
A. The stated putative CDK phosphorylation site mutants were created on a centromeric *SAE2-HA* plasmid and transformed into *sae2*Δ cells with transformants selected using hygromycin resistance. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at 4 h and 6 h points. **B.** Western blot of the Spo11-oligonucleotide membrane. 1:4000 α-FLAG antibody (Santa Cruz) with HRP conjugate was used against the FLAG epitope tag of Spo11 and chemiluminescence measured using an ImageQuant LAS 4000 machine. **C.** 10 ml of cells were taken at the stated time points. **D.** The *sae2-S267A* mutant was placed under control of a galactose inducible promoter containing a SK1 cryptic ATG site and expression induced with the addition of 2 μM βœstradiol at 4 hours. 10 ml of cells were taken at the stated hour from the start of meiosis. **A,C&D.** Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 8.75% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

consensus sites at S134, S179 and S267. *S267A* mutants, that prevent CDK phosphorylation at this residue, accumulate DSBs via Southern blotting, indicating a repair defect (Manfrini et al., 2010). The *sae2-S134A-S267A* double mutant has a further reduction in spore viability than the *S267A* mutant alone (Manfrini et al., 2010), suggesting that the S134 site is also a target for CDK phosphorylation, which also stimulates Sae2 activity. The *S179A* mutant had no defect in spore viability (Manfrini et al., 2010).

In order to test these mutants directly for Spo11-DSB processing, meiotic cultures were assayed for Spo11-oligo formation in strains containing centromeric plasmid-borne mutant alleles of the three putative CDK consensus sites (Figure 3.8). Spo11-oligos were not observed at any time point of meiosis (up to 8 h) in the *S267A* mutant (Figure 3.8A&C). Spo11-oligos in the *S134A* mutant were slightly reduced compared to wild type, whilst the *S179A* mutant observed similar Spo11-oligo levels to wild type (Figure 3.8A). These results suggest that phosphorylation of the S267 putative CDK phosphorylation site is essential for Spo11-DSB end processing. The S134 site, whilst hypomorphic when mutated to alanine, can still promote Spo11-oligo formation in meiosis, suggesting a more minor role compared to the S267 putative phosphorylation site. In agreement with previous reports, mutating the S179 site infers no phenotype, indicating phosphorylation of this residue has no effect on Sae2 activity. Cycling *sae2-S267A* cells were also assessed for sensitivity to MMS via spot testing as before. The *S267A* mutant rendered the cells more sensitive to MMS compared to wild type *SAE2* (Figure 3.2B), in agreement with previous reports (Huertas et al., 2008).

To determine whether the *S267A* mutant was active when over-expressed, the mutant allele was cloned downstream of the *GAL1* (SK1-like) promoter and introduced at the chromosomal locus. Meiotic cultures were assayed for Spo11-oligo formation as before, with addition of β-estradiol at 4 h to induce *sae2-S267A*. Over-expression of the *sae2-S267A* mutant (Figure 3.8D) resulted in no production of Spo11-oligos, supporting the view that this site is the critical phosphorylation site for an active Sae2 protein in meiosis (Huertas et al., 2008; Manfrini et al., 2010).

Mimicking phosphorylation by replacing a serine/threonine residue with glutamic (E) or aspartic (D) acid enables the ability to study the effects of having an artificial always-phosphorylated state of a protein (negatively charged aspartic/glutamic acid residues are similar to the negative charge created by phosphorylation). Such mutants also allow us to appraise whether mutating a specific residue (e.g. serine to alanine mutation) disrupts protein folding rather than affecting its activity by preventing phosphorylation. The putative CDK phosphorylation sites of *SAE2* were individually mutated to glutamic acid, placed onto a centromeric plasmid and transformed into *sae2*Δ cells. Meiotic cultures were assayed for Spo11-oligo production (Figure 3.8A&C). Mimicking phosphorylation at all three putative

CDK sites of Sae2 still permitted Spo11-oligo production. However, both the S267 and S134 sites had slightly lower levels of Spo11-oligos compared to wild type. This suggests that any mutation of these residues may slightly reduce Sae2 activity but the protein still effectively folds to generate an active protein. To determine the effects of mimicking phosphorylation at the S267 site in mitotic cells, cycling cells containing the *S267E* mutant was assayed for MMS sensitivity via spot testing (Figure 3.2B). Consistent with the slightly hypomorphic nature of this mutation in meiosis, *sae2-S267E* cells were slightly more sensitive to MMS compared to wild type *SAE2* but not as sensitive as the *S267A* mutant (Figure 3.2B). These results suggest phosphorylation of the S267 site is essential for processing of Spo11-DSBs.

3.10 Phosphorylation of the putative Mec1/Tel1 site at T279 is important for the role of Sae2 in meiosis

Phosphorylation by CDK is not the only phosphorylation event that occurs on Sae2. After DNA damage, and during the cell cycle, Sae2 becomes phosphorylated in a Mec1/Tel1 dependent manner (Baroni et al., 2004). Mec1/Tel1 phosphorylate substrates on SQ or TQ residues (Kim et al., 1999). Sae2 contains five putative acceptor sites: S73, T90, S249, T279, S289 (Figure 3.1A). Mutation of all five residues to alanine (5A) causes MMS sensitivity, defective recombination at inverted repeats and defective meiotic recombination, phenotypes that are similar to *sae2*Δ (Baroni et al., 2004; Cartagena-Lirola et al., 2006). Mutation of the putative Mec1/Tel1 phosphorylation sites found at in the C-terminus of Sae2 (S249A, T279A and S289A, in combination with a S278A mutant) indicated these putative phosphorylation sites are more important for Sae2 function in meiosis than those found in the N-terminal region (S73 and T90) (Cartagena-Lirola et al., 2006). Because of the subtle differences in the phenotypes of the previously characterised mutants it was of interest to determine which residues were important for Spo11-DSB processing in meiosis. Single point mutants of each of the putative S/T-Q sites were generated by site-directed mutagenesis and assayed for meiotic Spo11-oligo production and for sensitivity to MMS in cycling cells (Figure 3.9). The S278 residue on Sae2, whilst not a putative S/T-Q site for Mec1/Tel1 phosphorylation, has frequently been mutated alongside the actual putative Mec1/Tel1 phosphorylation site T279 (Baroni et al., 2004; Cartagena-Lirola et al., 2006; Cejka lab - personal communication). Therefore, the double *sae2-S278A-T279A* mutant was also assayed for meiotic and mitotic DSB repair defects (Figure 3.9).

The Spo11-DSB end-processing activity and the sensitivity to MMS had comparable outcomes. The *sae2-S278A-T279A* mutant is defective in Spo11-oligo formation (Figure 3.9A – right hand panel), as well as being sensitive to MMS (Figure 3.9C). The single mutation of

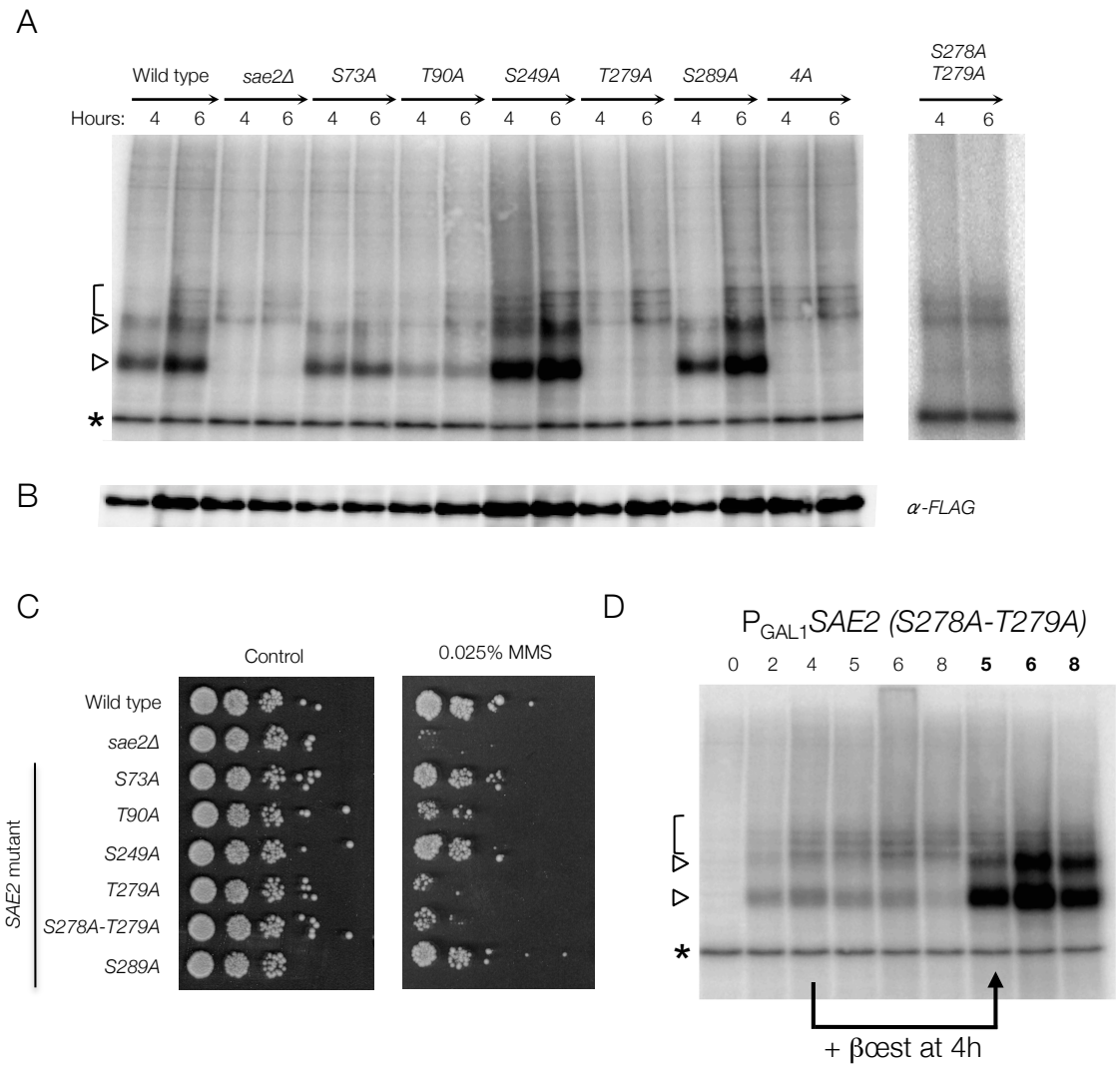


Figure 3.9: Mutation of individual potential Mec1/Tel1 phosphorylation sites shows T90 and T279 to be critical sites of phosphorylation for Sae2 activity with a combination mutant of the other four sites also abolishing Sae2 activity. The stated potential Mec1/Tel1 phosphorylation site mutants were created on a centromeric Sae2-HA plasmid and transformed in *sae2*Δ cells with transformants selected for using hygromycin resistance. The *4A* mutant consists of *S73A*, *T90A*, *S249A* and *S289A*. **A.** YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. **B.** Western blot of the Spo11-oligonucleotide membrane. 1:4000 α-FLAG antibody (Santa Cruz) with HRP conjugate was used against the FLAG epitope tag of Spo11 and chemiluminescence measured using an ImageQuant LAS4000 machine. **C.** Spot tests. All plates contained 300 μg ml^{-1} hygromycin to maintain selection of the plasmid and the MMS plate contained 0.025% MMS. Cultures were grown overnight in YPD (with 300 μg ml^{-1} hygromycin) then transferred to fresh YPD+HYG for 4 hours. Cultures were diluted to 0.2 OD_{600} and a 10-fold dilution series spotted onto the plate down to 0.00002, plates incubated at 30 °C for 4 days and images taken using a Syngene InGenius bioimaging system. **D.** The *sae2-S278A-T279A* mutant was placed under control of a galactose inducible promoter containing a SK1-like (cryptic) ATG site and expression induced with the addition of 2 μM βœstradiol at 4 hours. **A and D:** 10 ml of cells were taken at stated time points from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 8.75% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket (⎡) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

T279A prevents any Spo11-oligo formation and confers a phenotype similar to *sae2*Δ on MMS plates, suggesting that the T279 site is a critical phosphorylation site to promote both mitotic and meiotic DSB repair. Mutation of the T90 site to alanine caused slight hypersensitivity to MMS and a reduction in Spo11-oligo levels compared to wild type. The other sites (S73, S249 and S289) had very little effect on MMS sensitivity or Spo11-oligo formation suggesting that T279 and T90 are the most important of all the putative Mec1/Tel1 sites for regulation of Sae2 activity. These observations agree with a recent study that also identified T90 and T279 to be important during mitotic DSB repair compared to the other S/T-Q sites (Liang et al., 2015).

To determine whether the S/T-Q sites, other than T279, might have a redundant role, a *sae2-4A* mutant was generated that contained the *S73A-T90A-S249A-S289A* mutations. The *T279A* mutation was left out, as it was known prior to this experiment that mutating T279 prevents any activity of Sae2 (see above). The *sae2-4A* mutant was as deficient in Spo11-DSB processing as the *sae2*Δ (Figure 3.9A right hand panel). This observation suggests that either there is some residual function for phosphorylation on these sites that only becomes apparent when all four sites are mutated or that the physical nature of this many mutations of Sae2 may have caused improper folding, unrelated to its phosphorylation state and renders the protein inactive.

3.11 Over-expression of the *sae2-S278A-T279A* mutant overcomes the low level expression defects of this mutant

A collaboration with the Cejka lab (Zurich) has demonstrated residual stimulation of MRX endonuclease activity *in vitro* by the *sae2-S278A-T279A* mutant protein, suggesting that this mutant may be just strongly hypomorphic, rather than an inactive allele. To test this idea, the *sae2-S278A-T279A* mutant was placed downstream of the *GAL1* (SK1-like) promoter (Figure 3.9D). Expression of the *sae2-S278A-T279A* was induced in a meiotic culture at 4 h with β-estradiol and Spo11-oligo production assayed as before. Upon induction, a substantial production of Spo11-oligos was observed (Figure 3.9D). This result indicates that whilst defective at low-level expression, the *sae2-S278A-T279A* mutant is not fully inactive – in agreement with observations from the Cejka laboratory. Fu et al., 2014 proposed that Sae2 activity is regulated via phosphorylation, which alters its oligomeric state. If this idea is correct, over-expression of a hypomorphic allele may generate enough monomeric (active) protein. Unlike the *S278A-T279A* mutant the *S267A* (CDK) mutant was not rescued by over-expression (Figure 3.8D), which suggests that even if S267 phosphorylation does affect oligomeric state it may also have a major role in activating the Sae2 protein biochemically.

3.12 Mimicking Mec1/Tel1 phosphorylation on Sae2 at different residues renders differing outcomes in meiotic DSB processing

As previously discussed with CDK phosphorylation, mimicking phosphorylation on putative Mec1/Tel1 phosphorylation sites is also informative as to how these residues function. Such mutants are also useful to study the effects of mimicking phosphorylation when the kinases responsible for the phosphorylation are absent. Mimicking Sae2 phosphorylation at all five S/T-Q sites caused a higher fraction of the soluble Sae2 to be in the active monomer/dimer form, which is thought to be the active form (Fu et al., 2014).

To determine the impact of mimicking phosphorylation, individual and multiple phosphomimicking constructs were generated by site-directed mutagenesis and cloned into the CEN vector (Figure 3.10). Mimicking phosphorylation at T279, a site where phosphorylation is important (Figure 3.9), enabled production of Spo11-oligos, although at a reduced level, suggesting that mutation of this site does not affect protein folding negatively and instead supports the requirement for Mec1/Tel1 phosphorylation. Conversely, the T90 site, seen to be hypomorphic when mutated to alanine, is also hypomorphic with mutation to glutamic acid, suggesting mutation at this site may be interfering with proper folding of Sae2 rather than the a putative phosphorylation event itself having an effect on Sae2 activity. The *T90E* mutation, when coupled with *S73E, S249E* and *T279E* mutants (*4E mutant)* to mimic phosphorylation on all S/T-Q sites except S289 (more detail later) resulted in no Spo11-oligos being formed (Figure 3.10A). This could potentially be due to mis-folding of to the *T90E* mutation coupled with three other mutations that could disrupt the *4E* mutant structure, thereby preventing any active Sae2. Alternatively, this result may suggest that hyperphosphorylation of Sae2 may inhibit its activity - in contrast to Fu et al., 2014.

Intriguingly, despite the *sae2-S289A* mutant being fully proficient for Spo11-oligo formation (Figure 3.9A) and no more sensitive to MMS than wild type cells (Figure 3.9B), mutation of S289 to glutamic acid (*S289E*) resulted in no Spo11-oligo formation in meiosis (MMS sensitivity was not tested) (Figure 3.10 – *S289E*). To further investigate this phenotype, the *S289E* mutant was compared alongside another type of phospho-mimetic mutation of the site, *sae2-S289D.* Both these mimetic mutations abolish Spo11-oligo formation (Figure 3.10C). These results support unpublished, collaborative work from Petr Cejka's lab who have observed no stimulation of MRX endonuclease activity *in vitro* even with high concentrations of the *S289E* mutant protein (personal communication). One hypothesis is that hyper-phosphorylation of Sae2 at this site switches off the protein. In this model, the initial phosphorylation by Tel1 (whose activity is stimulated by the unprocessed DSB and may be less active then Mec1) at T279 may activate Sae2 to process the DSB ends allowing 5′ to 3′

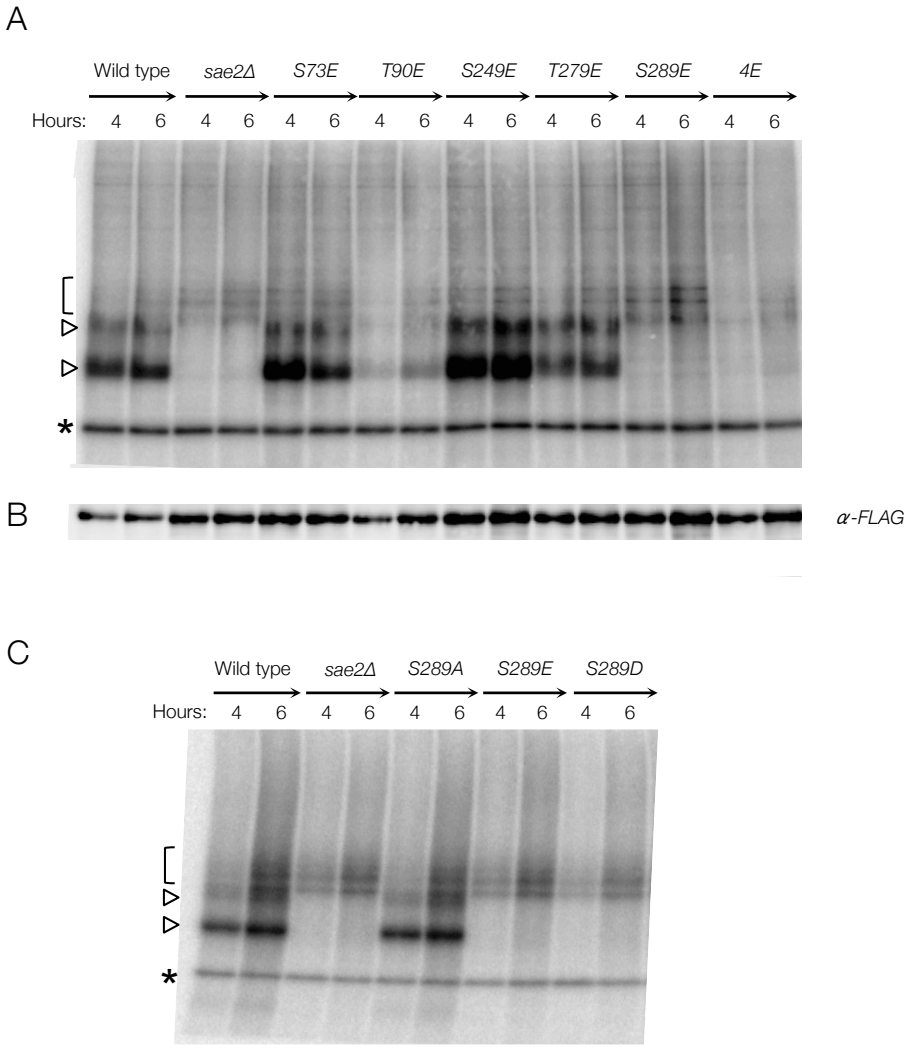


Figure 3.10: Mimicking phosphorylation of individual putative Mec1/Tel1 phosphorylation sites suggests that the *T90E* and *4E* mutation are hypomorphic whilst *S289E* abolishes Sae2 activity. The stated putative Mec1/Tel1 phosphorylation site mutants were created on a centromeric *SAE2-HA* plasmid and transformed in *sae2*Δ cells with transformants selected using hygromycin resistance. The *4E* mutant consists of *S73E*, *T90E*, *S249E* and *T279E*. **A&C.** YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at 4 and 6 hours from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 8.75% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band. **B.** Western blot of the Spo11-oligonucleotide membrane. 1:4000 α-FLAG antibody (Santa Cruz) with HRP conjugate was used against the FLAG epitope tag of Spo11 and chemiluminescence measured using an ImageQuant LAS4000 machine.

resection to generate long ssDNA tails. These ssDNA tails then hyper-stimulate Mec1, and/or Mec1 is naturally more active, resulting in phosphorylation of Sae2 on the S289 residue, thereby switching Sae2 activity off.

3.13 Combining CDK phospho-mimetics with S/T-Q site phospho-mimetics of Sae2 reduces the production of Spo11-oligos

Sae2 phosphorylation at S267 by CDK is critical for its activity (Huertas et al., 2008; Manfrini et al., 2010). This site is also the priming phosphorylation site required for subsequent phosphorylation by Mec1/Tel1 (Fu et al., 2014). It was therefore of interest to determine the impact of combining phospho-mimetics on S267 and other S/T-Q sites. As described earlier (Figure 3.8), the *S267E* mutation has a slightly hypomorphic phenotype for Spo11-oligo formation suggesting some potential mis-folding of the protein. Meiotic cultures were assessed for Spo11-oligo formation as before and when in combination with *S249E* and *T279E, S267E* reduces Spo11-oligo formation (Figure 3.11A). The *S267E* mutant does not rescue the defect in Spo11-DSB processing in the *S289E* or the *4E* mutants (Figure 3.11A). These observations suggest that either the *S267E* destabilises these already hypomorphic alleles, or that the S267 phosphorylation of Sae2 is not an over-riding activator, and instead the other phosphorylation sites play an important role.

3.14 Mec1 and Tel1 are required for Spo11-DSB end processing in meiosis

In the absence of Mec1 and Tel1 activity Spo11-DSBs accumulate, an identical phenotype to *sae2*Δ cells (Cartagena-Lirola et al., 2006). Therefore, neither kinase is required for DSB formation but they are for processing of Spo11-DSBs and initiation of resection. However, using the Spo11-oligo assay for direct analysis of Spo11-DSB processing has not been reported. To investigate this further, *tel1*Δ*, CLB2-MEC1* and *CLB2-MEC1 tel1*Δ strains were made and assayed for Spo11-oligo production (Figure 3.12). The *CLB2* gene is switched off in meiosis, therefore, placing a gene under the control of the *CLB2* promoter allows for a meiosis-specific knockdown of the protein. The *CLB2-MEC1* construct allows assessment of a *mec1*-null like phenotype, whilst having no effect on the mitotic roles of Mec1. Meiotic cultures of the mutant strains were assessed for Spo11-oligo production as before. Spo11-oligo levels were slightly raised in the *tel1*Δ background with an increase in higher molecular weight bands (see Chapter 6 for more detail about these molecules). The *CLB2-MEC1* strain had a very low level of Spo11-oligos whereas the *tel1*Δ *CLB2-MEC1* mutant had an even lower residual level of Spo11-oligos, that were just above the detection limit (Figure 3.12). These results support the conclusion that Mec1 and Tel1 are important for the activation of Spo11-

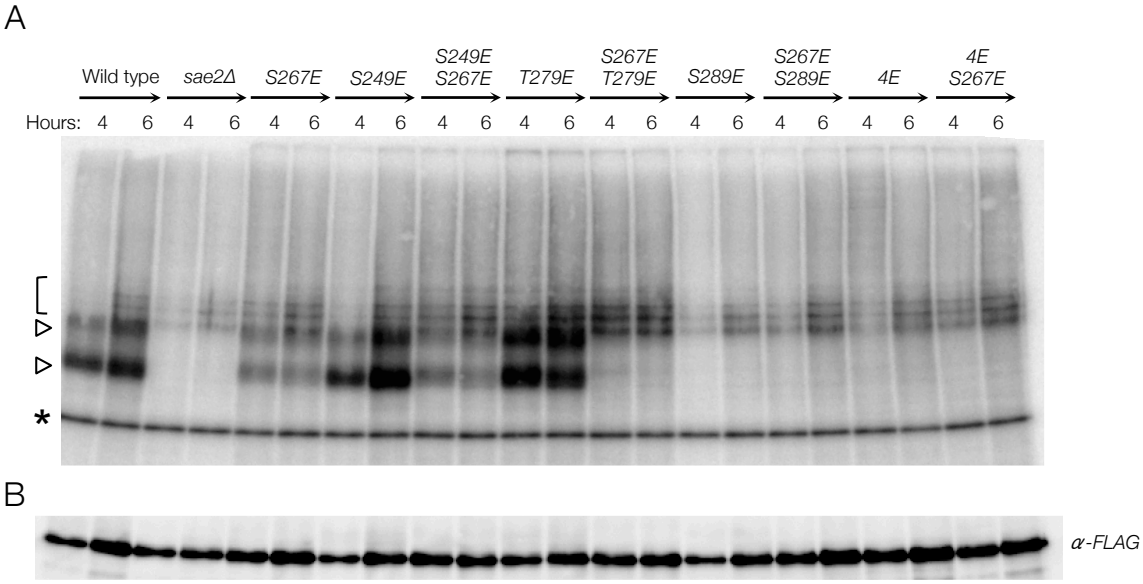


Figure 3.11: Mimicking phosphorylation of a combination of Mec1/Tel1 and/or CDK putative phosphorylation sites decreases the levels of Spo11-oligos.
The stated putative Mec1/Tel1 phosphorylation site mutants were created on a centromeric *SAE2-HA* plasmid and transformed in *sae2*Δ cells with transformants selected using hygromycin resistance. The *4E* mutant consists of *S73E*, *T90E*, *S249E* and *T279E*. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at 4 and 6 hours from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 8.75% SDS-PAGE gel at 150 V for 1 hour 30 minutes and transferred to a PVDF membrane. **A.** The membrane was exposed to a phosphor screen overnight. **B.** Western blot of the Spo11-oligonucleotide membrane. 1:4000 α-FLAG antibody (Santa Cruz) with HRP conjugate was used against the FLAG epitope tag of Spo11 and chemiluminescence measured using an ImageQuant LAS4000 machine. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

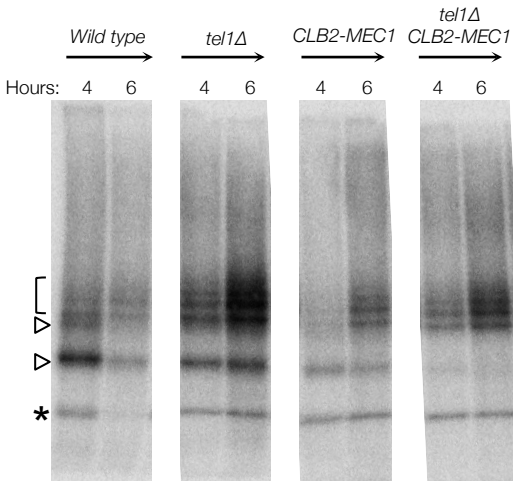


Figure 3.12: Spo11-oligonucleotide production is decreased in the absence of both Mec1 and Tel1 kinases.
The centromeric *SAE2-3HA* plasmid was transformed into the stated Mec1/Tel1 mutant cells with transformants selected using hygromycin resistance. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at 4 and 6 hours from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 8.75% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

DSB processing. The low residual level of Spo11-oligos seen in the double mutant may indicate that *SAE2* can function even without Mec1/Tel1 activation. Alternatively, the residual Spo11-oligos formed may be due to a low level of residual Mec1 activity in the *CLB2-MEC1* strain.

3.15 Mimicking Sae2 phosphorylation on putative Mec1/Tel1 residues cannot rescue the Spo11-DSB processing defect of Mec1/Tel1 deficient cells

If Sae2 is a major regulator of end processing, controlled by phosphorylation state, then mimicking phosphorylation of Sae2 may rescue this end-processing defect present in the Mec1/Tel1 strain. Mec1 and Tel1 deficient cells were transformed with centromeric plasmids containing the *sae2-4E*(*S73E-T90E-S249E-T279E)*, *sae2-S267E* and *sae2-4E(S267E)* mutants. Meiotic samples were processed for Spo11-oligos as before (Figure 3.13). In the *SAE2+ CLB2-MEC1 tel1*Δ strain, the previously observed basal level of Spo11-oligos is present (Figure 3.12A). Introduction of the *sae2-4E* mutant to the *sae2*Δ *CLB2-MEC1 tel1*Δ strain failed to increase Spo11-oligo formation (Figure 3.13). This negative result may be due to potential mis-folding of the *sae2-4E* mutant because the *4E* mutant was defective in a wild type background (see Figure 3.10A). To further investigate whether mimicking phosphorylation on Sae2 can bypass the requirement for Mec1 and Tel1, it would be best to test the *T279E* allele alone in *CLB2-MEC1 tel1*Δ background because it is still proficient in producing Spo11-oligos in an otherwise wild type background and looks to be the main phospho-site for Mec1/Tel1. Thus it is possible that the *T279E* allele may rescue the end-processing defect of Mec1/Tel1 deficient cells.

3.16 Preventing or mimicking acetylation of Sae2 has no effect on its ability to stimulate Spo11-DSB processing

Sae2 is acetylated after DNA damage on sites K239 and K266 (Fu et al., 2014). This modification targets Sae2 for degradation by the autophagy pathway. Mutation of these acetylation sites to block this modification (*sae2-K239R-K266R)* has no effect on DNA damage survival however; the double mutation mimicking acetylation (*sae2-K239Q-K266Q*) causes a growth defect even in the absence of damage, correlating with a decrease in the levels of Sae2 protein (Fu et al., 2014). To determine whether acetylation may have a role in meiotic or mitotic DSB end processing, the *sae2-K239R-K266R* mutant was created by site-directed mutagenesis of the CEN vector and meiotic cultures assayed for Spo11-oligo production and cycling cells assessed for sensitivity to MMS. Mutation of these acetylation acceptor sites had no effect on Spo11-oligo production (Figure 3.14A) or sensitivity to MMS in cycling cells

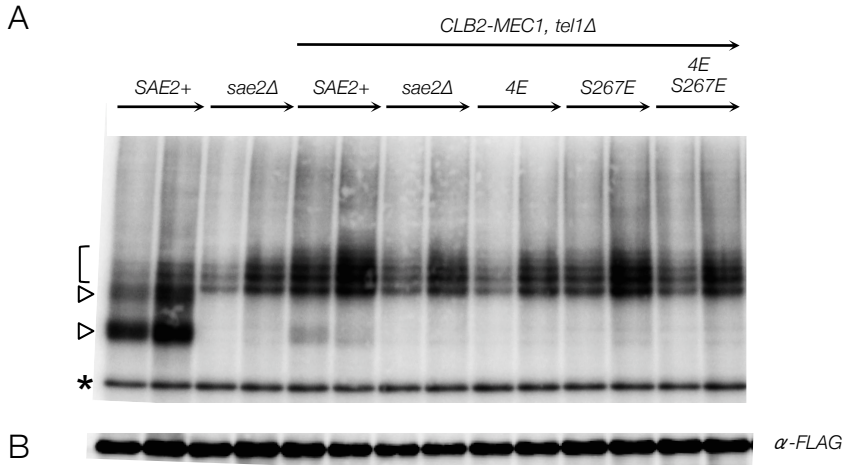


Figure 3.13: Mimicking phosphorylation of putative Mec1/Tel1 phosphorylation sites in a *CLB2-MEC1, tel1*Δ background does not rescue the Spo11-oligo production deficiency of *CLB2-MEC1, tel1*Δ cells.

The stated potential Mec1/Tel1 and CDK phosphorylation site mutants were created on a centromeric *SAE2-HA* plasmid and transformed into *sae2*Δ or *sae2*Δ, *CLB2-MEC1, tel1*Δ cells as stated with transformants selected using hygromycin resistance. The *4E* mutant consists of *S73E*, *T90E*, *S249E* and *T279E*. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at 4 and 6 hours from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 8.75% SDS-PAGE gel at 150 V for 1 hour 30 minutes and transferred to a PVDF membrane. **A.** The membrane was exposed to a phosphor screen overnight. **B.** Western blot of the Spo11-oligonucleotide membrane. 1:4000 α-FLAG antibody (Santa Cruz) with HRP conjugate was used against the FLAG epitope tag of Spo11 and chemiluminescence measured using an ImageQuant LAS4000 machine. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

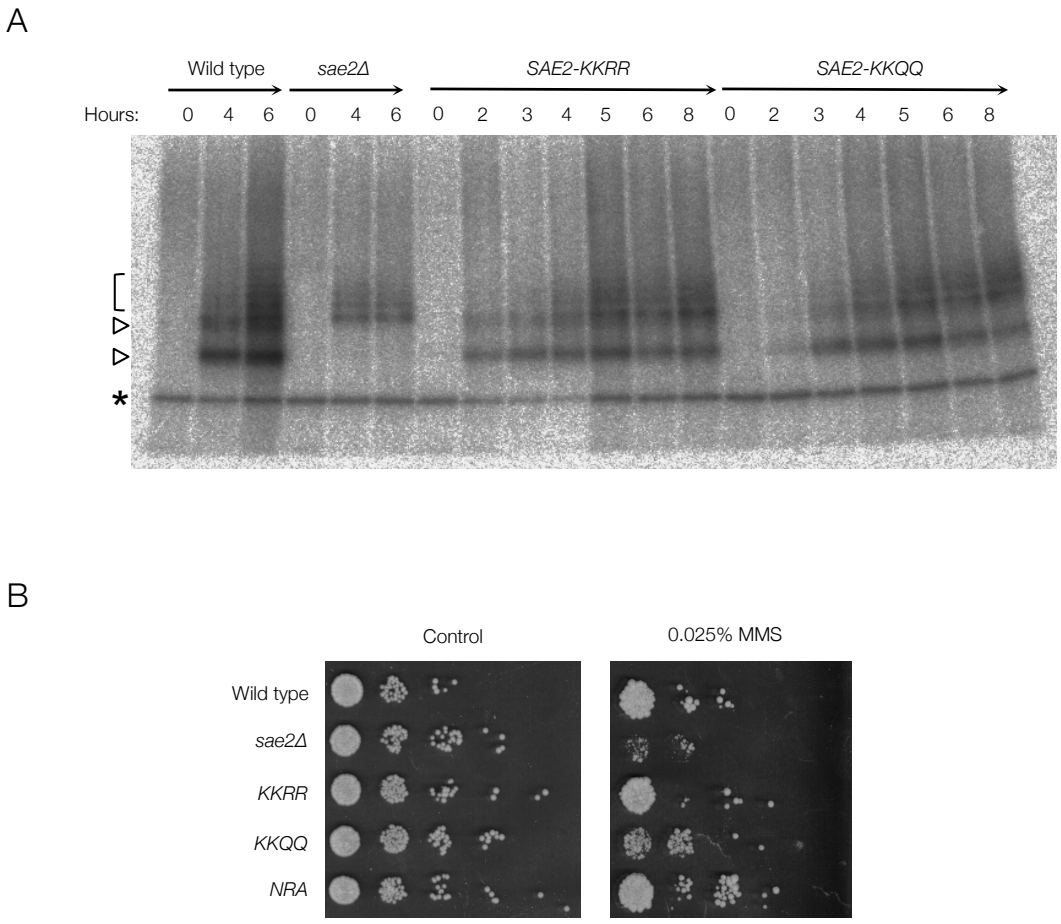


Figure 3.14: Acetylation mutants and mimetics of Sae2 do not have a defect in end-processing and Spo11-oligo production.
The acetylation mutants (*K239R-K266R*) and mimics (*K239Q-K266Q*) were created on a centromeric *SAE2-HA* plasmid and transformed into *sae2*Δ cells with transformants selected using hygromycin resistance. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at the time points stated from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

(Figure 3.14B). Mutation of the putative acetylation sites on Sae2 to glutamine (*sae2-K239R-K266R*) mimics acetylation by neutralising the charge (Dormeyer, 2005). As reported, the acetylation mimic had a growth defect on MMS plates (Figure 3.14B), however, there was no effect on Spo11-oligo production in meiosis (Figure 3.14A). From these observations it is possible to conclude that whilst the acetylation pathway may have some impact on Sae2 function, it is not relevant for Spo11-DSB processing in meiosis.

3.17 The reported endonuclease activity of Sae2 has no role in meiosis

Sae2 has been reported to display endonuclease activity capable of cleaving 5′ flap structures and single-stranded regions at the base of hairpin loops (Lengsfeld et al., 2007). It was postulated that this Sae2 endonuclease activity is responsible for cleaving the 5′ strand during DSB repair alongside, but independently of, MRX (Lengsfeld et al., 2007). A subsequent study determined that mutation of N123 and R127 to alanine resulted in camptothecin sensitivity and defective hairpin processing (Wang et al., 2014), suggesting that these residues were part of the Sae2 catalytic domain.

To test the putative role of the Sae2 endonuclease during meiosis, the *sae2-N123A-R127A* mutant was generated on the CEN vector and transformed into *sae2*Δ cells. Meiotic cultures were assessed for Spo11-oligo production (Figure 3.15). This putative endonuclease mutant generated Spo11-oligos at levels comparable to wild type cells (Figure 3.15). This observation suggests that either the endonuclease activity of Sae2 plays no role in meiotic Spo11-DSB processing or, as reported in other labs, Sae2 does not contain endonuclease activity (Cannavo and Cejka, 2014; Niu et al., 2010).

3.18 Discussion

The precise role of Sae2 in DSB processing has been intensively studied but there are still questions to be answered. Sae2 has an essential role in meiosis: Spo11-DSBs accumulate in the absence of the protein and resection is prevented (McKee and Kleckner, 1997; Prinz et al., 1997). In mitotic cells however, the role of Sae2 can differ depending on the type of DNA damage that is present. The established model for DSB repair identifies Sae2 as an essential component when the end(s) of DNA damage are complex, be that due to covalently bound proteins, adducts or chemical modifications. When the end of the damage site is clean, for example, due to endonuclease activity, then the requirement for Sae2 is removed. Instead, at clean-ended breaks, Sae2 plays a role in enhancing the repair process rather than being essential. Sae2 function is linked to the nuclease function of Mre11 (MRX complex),

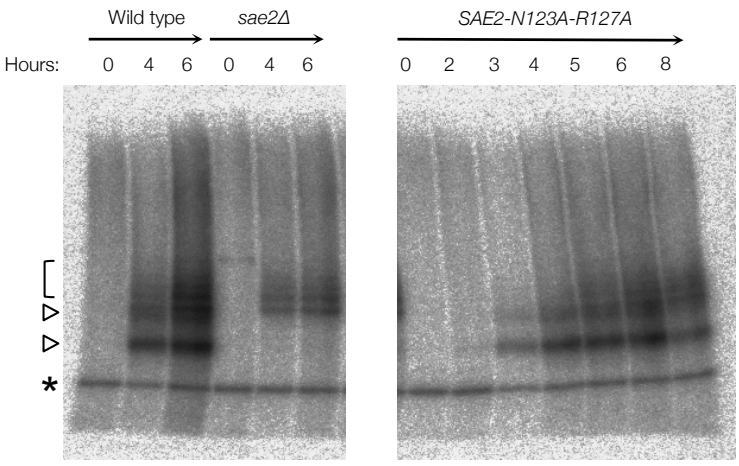


Figure 3.15: The endonuclease mutant of Sae2 does not have a defect in end-processing and Spo11-oligo production.
The endonuclease mutant (*N123A-R127A*) was created on a centromeric *SAE2-HA* plasmid and transformed into *sae2*Δ cells with transformants selected using hygromycin resistance. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at the time points stated from the start of meiosis. Spo11-oligos were enriched using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. The open bracket ([ ) marks the double-cut Spo11-oligonucleotide species seen in wild type and *sae2*Δ. Asterisk marks non-specific TdT band.

specifically, Sae2 stimulates the endonuclease activity of Mre11 through a direct interaction with members of the MRX complex (Cannavo and Cejka, 2014). However, interaction domains have not been identified. In other studies, Sae2 has been reported to contain an endonuclease which is active on hairpin structures (Lengsfeld et al., 2007; Lobachev et al., 2002). However, the lack of a nuclease domain on Sae2 and conflicting reports from other labs suggest that this is not the case. In meiosis, mutation of the putative endonuclease domain has no effect on the ability to generate Spo11-oligos and initiate HR (Figure 3.15).

Sae2 has also been reported to be regulated by post-translational modifications. Phosphorylation of Sae2 by CDK has been studied in both mitotic and meiotic cells and is essential for all Sae2 processes (Huertas et al., 2008; Manfrini et al., 2010). In support of these findings Figure 3.8 demonstrates that the S267 site, and to a lesser extent, the S134 site, are essential for the initiation of Spo11-DSB repair. Control by CDK links activation of the HR pathway (via stimulation of the Mre11 endonuclease activity by Sae2) with the cell cycle stage and the presence of a homologous template that HR can repair from.

The other pathway that regulates Sae2 activity by phosphorylation is the Mec1/Tel1 pathway. This pathway is activated by DNA damage and when activated leads to cell cycle arrest. Sae2 phosphorylation by these kinases has also been studied in both mitotic and meiotic cells. It has been previously observed that preventing phosphorylation on all five of the potential Mec1/Tel1 S/T-Q sites at the same time created a defect in mitotic and meiotic recombination and sensitivity to MMS (Baroni et al., 2004; Cartagena-Lirola et al., 2006). Here I have shown that mutation of four of the five S/T-Q (S73, T90, S249 and S289) sites individually has no effect on Spo11-DSB processing (Figure 3.9), but a combination of the four (*S73A, T90A, S249A* and *S289A*) completely prevents Spo11-oligo formation. In contrast, mutation of the fifth S/T-Q site (T279) alone prevents any Spo11-oligo formation suggesting that this site is the crucial phosphorylation site for Sae2 activity in meiosis. This site, alongside T90, was recently reported to be crucial in Sae2 activity in mitotic cells with these residues linked to interaction with the MRX complex through the FHA domain of Xrs2 (Liang et al., 2015). Consistent with this report, the *T90A* mutant is hypomorphic for Spo11-oligo formation in meiosis (Figure 3.9).

Interestingly, the *sae2-S289E* mutant (Figure 3.10) has an opposite effect to that of other phospho-mimetics with Spo11-DSB processing completely prevented. This result has been replicated in the Cejka lab's *in vitro* work (personal communication). An exciting possibility is that phosphorylation at this site may switch off Sae2 stimulation of the MRX complex. With this in mind, the multiple potential Mec1/Tel1 phosphorylation sites on the Sae2 protein may act as a molecular switch. For example, DSB formation activates Tel1, which phosphorylates a subset (T279) of residues on Sae2 activating the protein. The now active Sae2 stimulates DSB

processing and resection, enabling the generation of long ssDNA tails. These DNA structures may then (hyper-) activate Mec1, which in turn hyper-/further phosphorylates Sae2 on the S289 site switching activity off (Figure 3.16).

Mec1/Tel1 activity are required for DSB repair in meiosis with *mec1*Δ *tel1*Δ strains accumulating unresected DSBs (Cartagena-Lirola et al., 2006). Consistent with this observation, in a background deficient for these two kinases, initiation of DSB repair/Spo11-oligo formation is suppressed (Figure 3.12). It has been suggested that it is the phosphorylation of Sae2 by these kinases that prevents the initiation of resection. If phosphorylation of Sae2 on T279 is the only role of Mec1/Tel1 at Spo11-DSB end processing then mimicking phosphorylation might be expected to bypass the requirement for Mec1/Tel1. However, the combination of four phospho-mimetic mutations resulted in an inactive form of Sae2, and thus this idea remains untested (Figure 3.13). Future work to answer this question would be to incorporate the *sae2-T279E* mutation alone in this Mec1/Tel1 deficient background and investigate the effect on Spo11-oligo formation and subsequent resection. However, it is also highly likely that phosphorylation of Sae2 is not the only Mec1/Tel1-dependent phosphorylation event that plays a role in control of DSB resection.

Sae2 forms a higher order oligomer, which regulates its activity *in vivo* (Fu et al., 2014; Kim et al., 2008). Conflicting reports suggest either oligomerisation is critical for Sae2 activity (Kim et al., 2008) or that Sae2 is inactive when restrained in an oligomer, with phosphorylation (at all five S/T-Q and CDK putative sites) releasing Sae2 into an active monomeric form (Fu et al., 2014). Therefore, in meiosis, phosphorylation of Sae2 may also regulate the protein via alteration of its oligomeric state. The *T279A* mutation is deficient for stimulation of Spo11-oligo formation at low-level expression (natural promoter) (Figure 3.9A) but, when over-expressed, in combination with a mutation at S278 (*S278A-T279A),* was proficient for initiation of Spo11-DSB repair (Figure 3.8D). This suggests that Mec1-Tel1-dependent phosphorylation of T279 may regulate its oligomeric state as previously suggested (Fu et al., 2014), with increased expression generating an increase in the 'free' monomeric (active) form of Sae2. In contrast, over-expressing the *sae2-S267A* CDK mutant, could not rescue the Spo11-oligo formation defect seen at low expression levels (CEN plasmid) of the allele (Figure 3.8), suggesting CDK-dependent phosphorylation of S267 may active Sae2 in another manner besides regulation of oligomerisation.

These phosphorylation sites are not the only sites on Sae2 that control the higher order structure of Sae2. The N-terminus of Sae2 promotes oligomerisation via the L25 residue (Kim et al., 2008), yet *in vitro* the *L25P* mutant is still proficient for stimulation of MRX endonuclease activity when assayed at high concentration (Cannavo and Cejka, 2014). Similarly, over-expression of N-terminal truncation mutants also allow processing of Spo11-

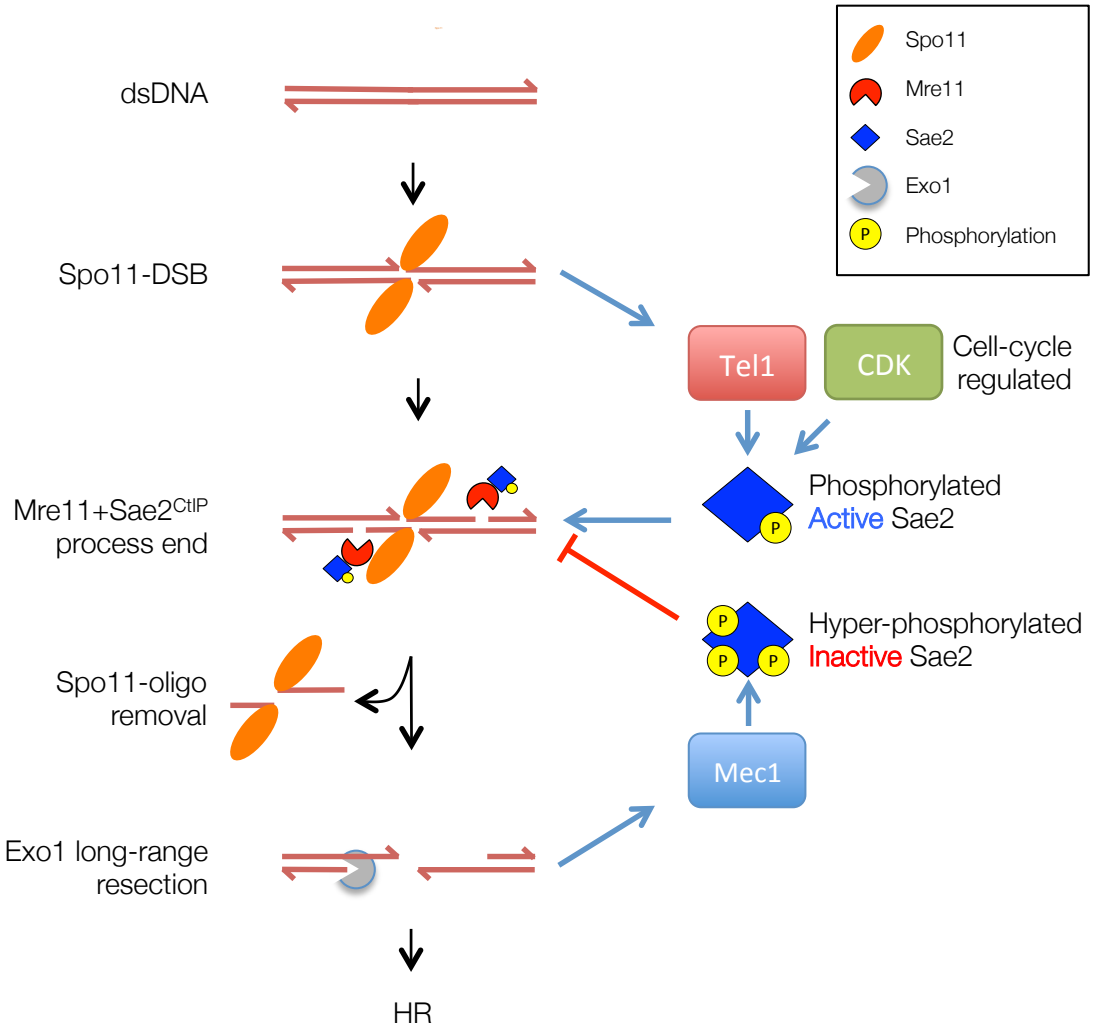


Figure 3.16: Model for Mec1/Tel1 regulation of Sae2 activity in meiosis.
Sae2 is essential for Spo11-DSB end-processing. The MRX complex, located at the Spo11-DSB activates the Tel1 kinase. Tel1 phosphorylates Sae2 at T279, which, alongside phosphorylation by the cell-cycle regulated CDK at S267, activates Sae2. Active Sae2 stimulates the Mre11 endonuclease, which nucleolytically removes Spo11 bound to an oligonucleotide (Spo11-oligo). The resulting 3′ overhang is further resected by the long-range resection enzyme, Exo1. The generation of single-stranded DNA by MRX/Sae2 and Exo1 activates the Mec1 kinase. Mec1 (hyper-)phosphorylates Sae2 on S289 (and possibly other S/T-Q residues) inactivating Sae2 to prevent aberrant activation of the Mre11 endonuclease.

DSBs *in vivo* (Figure 3.6). Together, these observations suggest self-interaction is not essential for Sae2 activity but that it does promote the activity of Sae2. However, neither low nor over-expression of the *L25P* mutant could restore Spo11-DSB processing *in vivo*. This suggests that this site may have another function in regulation of Sae2, or that when mutated it causes aberrant interactions of Sae2 causing the formation of an inactive protein.

Collectively, this investigation presented here reveals novel mechanistic insight into the regulation of the Sae2 protein during meiosis with a variety of Sae2 modification sites found to be essential for Spo11-oligo removal during meiosis. Importantly, a single Mec1/Tel1 phosphorylation site was found to be most important for meiotic Sae2 activity compared to the previously reported five sites. Additionally, previously reported Sae2 endonuclease activity and acetylation state of the protein plays no role in Sae2 activity in meiosis.

CHAPTER 4:

INVESTIGATING TDP2 BIOCHEMISTRY

Chapter 4: Investigating TDP2 biochemistry

4.1 Introduction

In meiosis the evolutionary conserved topoisomerase-like protein, Spo11, generates programmed DSBs throughout the genome from which recombination initiates (Szostak et al., 1983). The Spo11 dimer forms these DSBs by creating a 5′ covalent linkage with either side of the DSB via its catalytic tyrosine residue (Bergerat et al., 1997). The Spo11 moiety needs to be removed in order for the DSB to be repaired. There are two potential ways in which this could occur: direct hydrolysis of the covalent bond between Spo11 and the 5′-end leaving a 2 nt 5′-overhang, or nucleolytic cleavage of the 5′ strand removing Spo11 covalently attached to single-stranded oligonucleotides. Current evidence indicates that Spo11-DSBs are in fact repaired solely via the nucleolytic pathway in *S. cerevisiae* via the action of the MRX complex and Sae2 (Neale et al., 2005). Acting together, these proteins endonucleolytically clip the 5′ strand adjacent to the DSB and from this nick exonucleases can further resect the 5′ strand, releasing Spo11 still covalently bound to single-stranded oligonucleotides of set length via the 5′ phosphotyrosine bond (Garcia et al., 2011).

In mitotic cells, DSBs are also generated in a programmed manner although more transiently through the action of Top2. Top2, in order to regulate the topological state of the DNA and remove catenanes, creates DSBs to allow strand passage to occur (see Figure 1.7). It does so in a similar manner to Spo11, creating 5′ covalent phosphotyrosine linkages with either side of the DSB. A slight difference between the two types of DSB is that Top2 creates a 4 nt 5′-overhang compared to Spo11 creating a 2 nt 5′ overhang (Burden and Osheroff, 1998). These DSBs are usually benign because the Top2 dimer reseals the break after use. However, during the normal catalytic cycle of Top2 and in response to chemical poisons, Top2-DSBs can become stabilised resulting in a toxic lesion (Corbett and Osheroff, 1993). Unlike Spo11-DSBs in *S. cerevisiae*, in mammalian cells, TOP2-DSBs can be repaired by both nucleolytic and hydrolytic pathways. MRN and CtIP can act upon TOP2-DSBs to nucleolytically process them although in yeast the oligonucleotides covalently bound to Top2 are shorter than those created during Spo11-DSB processing (Hartsuiker et al., 2009; Neale et al., 2005). The direct method of hydrolysing the phosphotyrosyl covalent bond between TOP2 and the 5′ end is conducted by the recently characterised tyrosine phosphodiesterase 2 (TDP2) protein in humans (Cortes-Ledesma et al., 2009). TDP2 cleaves the phosphotyrosyl bond after TOP2 has been proteolysed and can also hydrolyse an array of other phosphotyrosyl bonds (Gao et al., 2012). However, even though the bond between Spo11 and the 5′ end of the DSB is such a linkage

there have been no reports of TDP2 acting on Spo11-DSBs or that TDP2 has any role in meiosis at all. In *S. cerevisiae*, there is no known yeast TDP2 ortholog, however yeast Tdp1 does contain some residual 5′ phosphodiesterase activity (Murai et al., 2012; Nitiss et al., 2006), but as in humans, has not been shown to act within yeast meiosis.

If TDP2 were able to cleave the bond between Spo11 and the DSB end, this activity could be utilised to monitor DSB repair in meiosis on a DSB substrate with clean ends, which are potentially religatable via the NHEJ pathway. Additionally, direct removal of the protein moiety and direct ligation of a sequencing adaptor specifically to the Spo11 break end could be utilised to map Spo11-DSBs with nucleotide resolution (see Chapter 5). To determine if this is possible, TDP2 activity on Spo11 substrates requires verification.

4.2 Can TDP2 cleave Spo11 from the end of a single-stranded oligonucleotide?

The simplest Spo11 substrate that TDP2 activity could be tested on is the Spo11-oligo complexes generated during meiosis by the nucleolytic action of MRX/Sae2. These single-stranded oligonucleotides are 5′ covalently linked to Spo11 with a phosphotyrosyl bond and can be enriched by immunoprecipitation (see Materials and Methods, and Chapter 3). Because TDP2 has been reported to remove TOP2 from the end of a DSB only after TOP2 proteolysis, it is also possible that Spo11 may also need to be proteolysed in order for TDP2 to work. Two different proteases, Proteinase K (Prot K, leaving 3 amino acids attached to the DNA) and trypsin (leaving 12 amino acids) can be used to generate differently sized peptides attached to the oligos for TDP2 to act upon. A schematic of the proposed action of TDP2 on various Spo11-oligo substrates is presented in Figure 4.1.

4.3 TDP2 can cleave proteolysed Spo11 from the end of a single-stranded oligonucleotide

Separation of Spo11-oligo species on an SDS-PAGE gel generates two bands that correspond to Spo11 attached to differing lengths of single-stranded oligonucleotide (Figure 4.1). The actual length of the oligonucleotides attached to Spo11 can be more accurately sized using a denaturing urea-PAGE (sequencing) gel. These gels separate DNA species by length, however, protein association to DNA molecules retards migration significantly. Therefore, in order to size these oligonucleotides in this manner, Spo11 must first be proteolysed. Typically Proteinase K is used to proteolyse Spo11 resulting in a bound peptide of only three amino acids. This short length of peptide has only a minor effect on migration of the DNA species but does still lead to slightly inaccurate sizing (Neale et al., 2005). To determine whether TDP2 is able to remove the protein component of Spo11-oligo substrates,

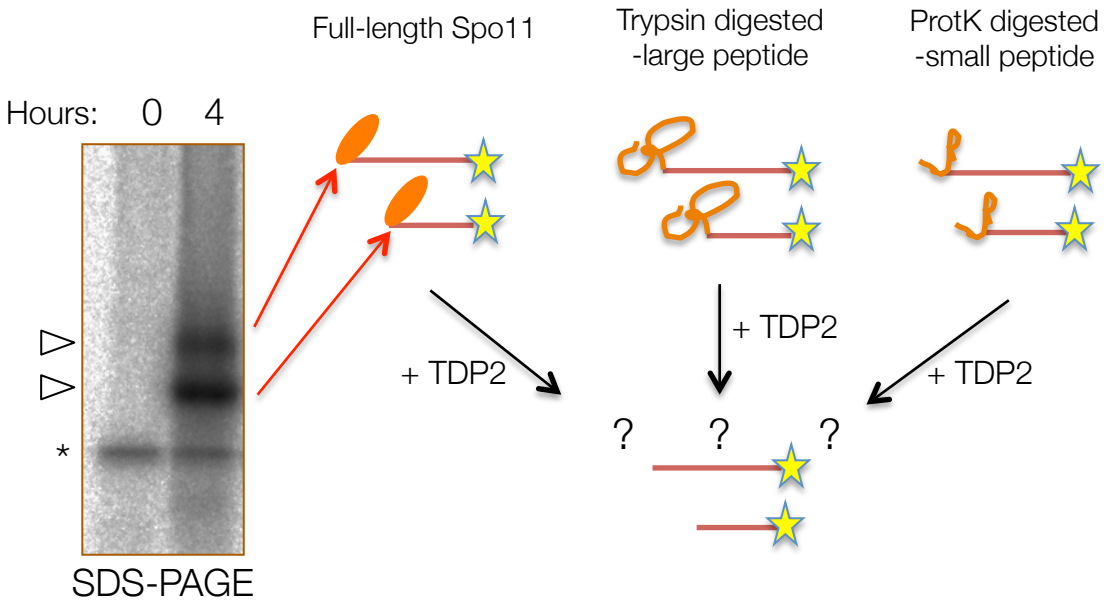


Figure 4.1: Schematic of potential TDP2 activity on Spo11-oligos and their digested derivatives. 10 ml of cells were taken at the stated hour from the start of meiosis and Spo11-oligos were enriched from wild type cells using the Spo11-oligonucleotide assay as stated in Materials and Methods. Samples were resolved on a 7.5% SDS-PAGE gel at 150 V for 1 hour 30 minutes, transferred to a PVDF membrane and exposed to a phosphor screen overnight. Open triangles (▷) mark the long and short Spo11-oligo species generated in wild type cells. Asterisk marks non-specific TdT band. Tdp2 can cleave covalently bound, proteolysed Topoisomerase-II from the 5′ end of DNA. Can TDP2 cleave 5′ covalently bound Spo11, or its proteolysed forms, from the end of single-stranded oligonucleotides?

Spo11-oligos were enriched by IP from meiotic cultures at mid-meiotic prophase, end-labelled and reacted with increasing concentrations of TDP2 (gift from Keith Caldecott) (Figure 4.2). If TDP2 can act upon Spo11 substrates it was expected to be most active upon proteolysed forms of Spo11, as it does for Top2 *in vivo* (Cortes-Ledesma et al., 2009). Therefore, Spo11-oligo complexes were first pre-digested with either Proteinase K or trypsin before incubation with TDP2. Reactions were then separated on a denaturing acrylamide gel (Figure 4.2). Without TDP2 treatment, Proteinase K treated Spo11-oligos migrate as a heterogeneous bimodal distribution, ranging from 10-15 bp and 25-40 bp (Figure 4.2 – 0 nM PK lane), consistent with previous reports (Neale et al., 2005). Incubation with TDP2 caused this distribution to migrate more rapidly, equivalent to a shift of ~3 nt. For the trypsin-digested samples, digestion leaves a twelve amino acid peptide still bound, which severely retards migration (Figure 4.2 - 0 nM trypsin lane) resulting in a heterogeneous smear. Incubation of these molecules with TDP2 caused a dramatic shift in migration pattern, which matched that of the Proteinase K/TDP2 treated samples.

In DSB end-processing deficient mutants, such as *sae2*Δ, Spo11-oligonucleotides are still observed on SDS-PAGE gels but have much longer lengths (Chapter 3 – Figure 3.2). To determine whether TDP2 can also act upon these Spo11-DNA derived species, Spo11-oligos were purified as before and reacted with increasing concentrations of TDP2 (Figure 4.3). Resolution of these proteolysed species on a denaturing gel migrate with a 10 bp periodicity and after TDP2 treatment also migrate faster (Figure 4.3). These results are consistent with TDP2-dependent hydrolysis of the covalent phosphotyrosine bond between the Spo11 peptide and the single-stranded DNA.

4.4 TDP2-dependent shift in the size of Spo11-derived oligonucleotides is not due to nuclease contamination

It is assumed that the shift in migration of the Spo11-oigo complexes is due to the action of TDP2. However, a contaminant nuclease in the TDP2 preparation may also cause degradation and faster migration of the product. To test this, a radioactively labelled primer was reacted with increasing concentrations of TDP2 to observe if there was any shift in migration. No shift, or release of single-radioactive nucleotides, was observed with any concentration of TDP2 (Figure 4.3), indicating that the shift observed in TDP2 treated Spo11-oligo species is not due to a contaminant nuclease and instead points to TDP2 being active on the Spo11 phosphotyrosyl bond, something that has not been previously reported.

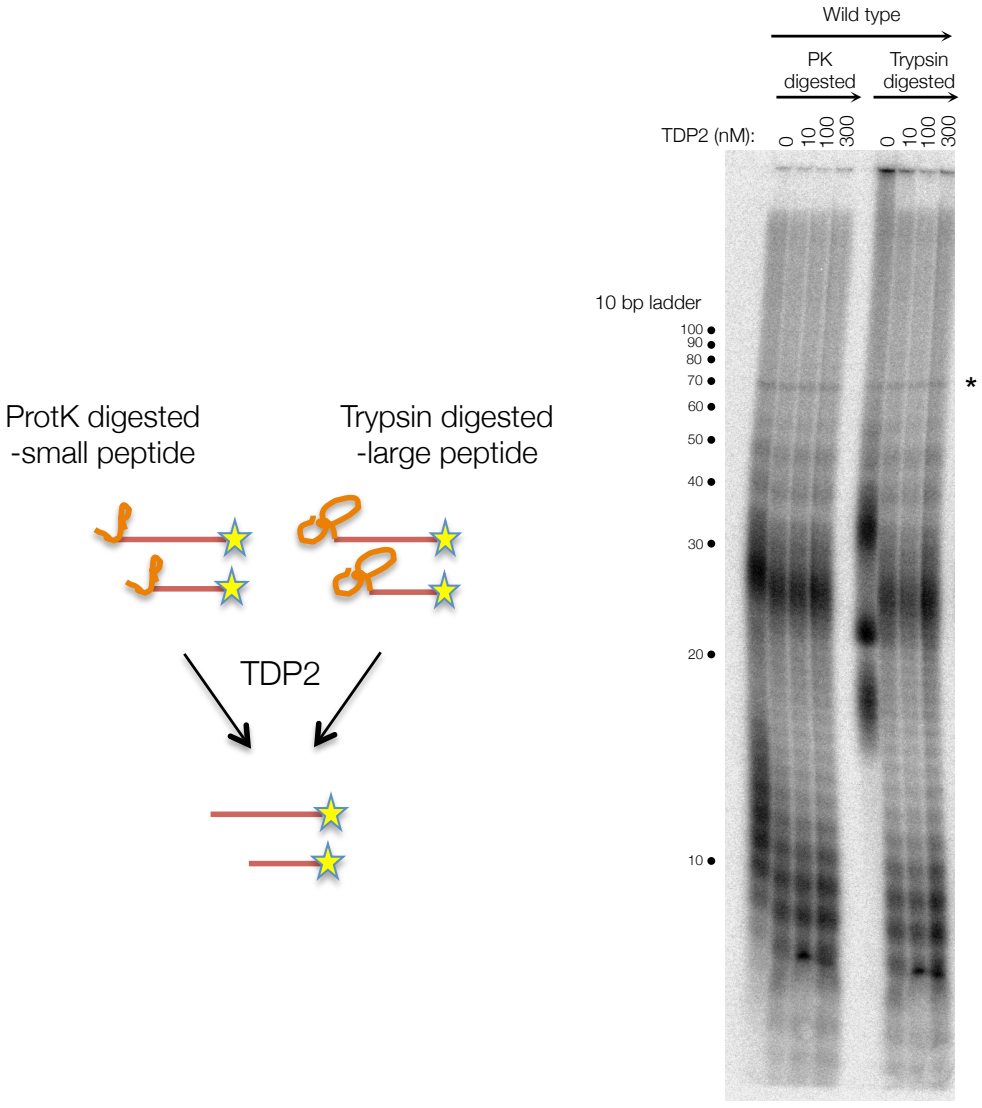


Figure 4.2: TDP2 can cleave the covalent bond between proteolysed Spo11 and a single-stranded oligonucleotide.
Spo11-oligonucleotide detection in wild type cells during meiosis. Spo11 complexes were 3′ end labelled using TdT and dCTP radionucleotide. Spo11-oligos were digested with Proteinase K at 60 °C for 1 hour or with Trypsin at 37 °C for 1 hour and DNA precipitated in ethanol overnight at -80 °C. DNA precipitates were resuspended in TDP2 reaction buffer and reacted with stated concentrations of TDP2 at 37 °C for 1 hour. Species were fractionated by denaturing urea-PAGE (19% acrylamide). The 10 bp ladder was labelled using TdT and dCTP radionucleotide. Asterisks are non-specific labelled contaminants.

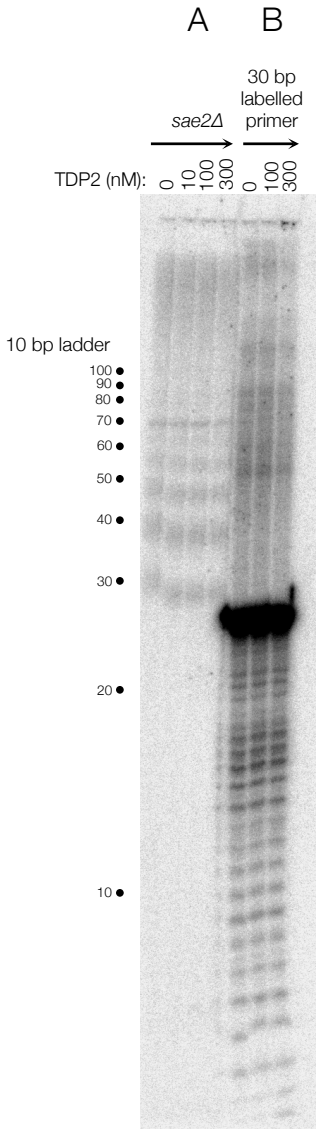


Figure 4.3: Faster migration of Spo11-oligo complexes following TDP2 treatment is not due to contaminating nucleases.
A. Spo11-oligonucleotide detection in *sae2*Δ cells during meiosis. Spo11 complexes were 3′ end labelled using TdT and dCTP radionucleotide. Spo11-oligos were digested with Proteinase K at 60 °C for 1 hour. Complexes were reacted with stated concentrations of TDP2 at 37 °C for 1 hour. **B.** A random 30 bp primer was labelled using TdT and dCTP radionucleotide and reacted with the stated concentrations of TDP2 at 37 °C for 1 hour. The 10 bp ladder was labelled using TdT and dCTP radionucleotide. Asterisks are non-specific labelled contaminants. All samples were fractionated by denaturing urea-PAGE (19% acrylamide).

4.5 TDP2 can cleave full-length Spo11 from the end of a single-stranded oligonucleotide

TDP2 is reported to act upon proteolysed TOP2 protein, hydrolysing the phosphotyrosyl bond between the degraded protein and the DNA (Cortes-Ledesma et al., 2009). TDP2 cannot process full-length native TOP2 (Gao et al., 2012). To determine whether TDP2 can remove unproteolysed, full-length Spo11 from the 5′-end of the oligonucleotide, intact radiolabelled (denatured) Spo11-oligo species were incubated directly with increasing concentrations of TDP2 and the reactions resolved on a denaturing acrylamide gel (Figure 4.4). Under these assay conditions full-length Spo11-oligonucleotide species do not migrate on these gels and are frequently lost from the well during fixation (Figure 4.4, lane 0 nM of unproteolysed Spo11-oligos). Increasing concentrations of TDP2 resulted in radiolabelled DNA species entering the gel (Figure 4.4 left hand panel) with lengths indistinguishable from the TDP2-treated samples that had first been incubated with Proteinase K (Figure 4.4 right hand panel) or trypsin (Figure 4.2 right hand panel). However, compared to the samples pre-treated with Proteinase K or trypsin, the efficiency of TDP2 against full-length Spo11-oligos was much lower. Spo11 is a 45 kDa protein. Thus the reduction in total oligonucleotide signal compared to Prot K and TDP2 treated samples is likely to be explained by the larger protein molecule occluding access of the catalytic site of TDP2 to the phosphotyrosyl bond. Therefore, similar to previous observations of TDP2 activity on Top2 substrates, proteolysis of Spo11 enhances the ability of TDP2 to hydrolyse the covalent bond between the 5′-end of the ssDNA species and the protein adduct.

4.6 TDP2 can remove proteolysed Spo11 from the end of double-stranded genomic DNA to enable λ exonuclease to resect but not Exo1

In strains containing Spo11-DSB end processing mutants such as *sae2*Δ*, mre11-nd* and *rad50S,* Spo11-DSBs accumulate. These DSBs contain Spo11 covalently bound to the 5′ end, preventing resection and repair. To determine whether TDP2 is able to remove such covalently attached Spo11 from duplex DNA a novel assay system was developed. Extracting genomic DNA in a proteolysing manner (with Proteinase K) from mid-meiotic prophase generates a population of genomic DNA fragments, some of which are derived from DSBs and thus have 5′ ends covalently bound to a short (~3 aa) peptide. One of the most frequent Spo11-DSB sites in the genome is the *HIS4::LEU2* meiotic recombination hotspot (Xu and Kleckner, 1995). This site contains two strong DSBs, that can be visualised via Southern blotting (Figure 4.5). In a *sae2*Δ mutant these two DSBs migrate as two tight bands on the gel because no *in vivo*

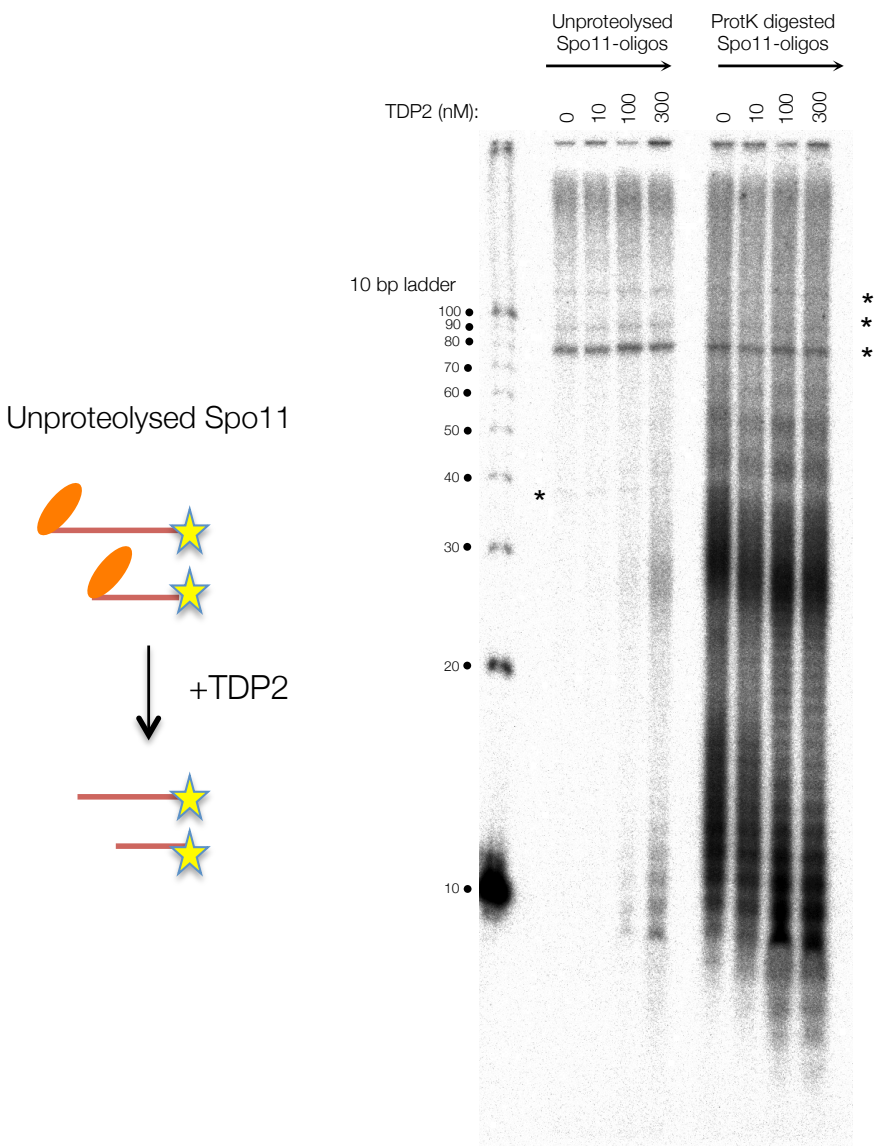


Figure 4.4: TDP2 can cleave the covalent bond between unproteolysed/full-length Spo11 and a single-stranded oligonucleotide.
Spo11 complexes were 3′ end labelled using TdT and dCTP radionucleotide. Spo11-oligos were left unproteolysed or were digested with Proteinase K at 60 °C for 1 hour. Complexes were reacted with stated concentrations of TDP2 at 37 °C for 1 hour and fractionated by nucleotide resolution urea-PAGE (19% acrylamide). The 10 bp ladder was labelled using TdT and dCTP radionucleotide. Asterisks are non-specific labelled contaminants.

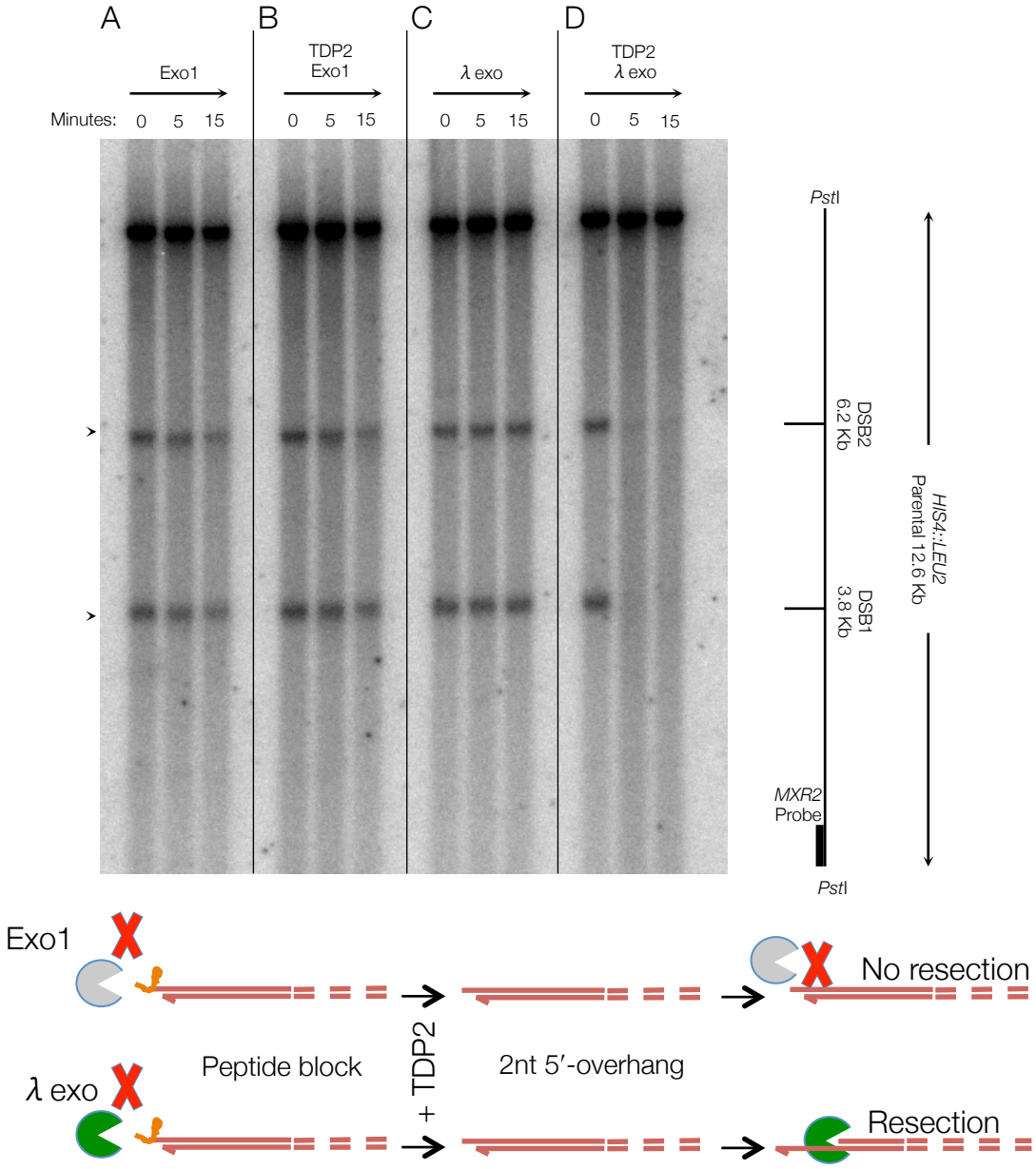


Figure 4.5: TDP2 can remove proteolysed Spo11 from the end of double-stranded DNA to enable λ exonuclease, but not Exo1, to resect DSBs.
Southern blot of DSBs at the *HIS4::LEU2* meiotic hotspot. A meiotic time course was performed for a *sae2*Δ strain and samples taken at 6 hours. Proteolysed genomic DNA was extracted and split into four sub-samples. **A** ('Exo1') and **C** ('λexo') were incubated at 37 °C for 1 hour in Exo1 reaction buffer without the addition of TDP2 protein. **B** ('TDP2 Exo1') and **D** ('TDP2 λexo') were incubated at 37 °C for 1 hour in Exo1 reaction buffer with the addition of 300 nM TDP2 protein. **A** and **B** were then incubated at 30 °C with the addition of 20 nM Exo1 and 680 nM RPA proteins and **C** and **D** were supplemented with λ exonuclease buffer and 5 units of λ exonuclease (NEB) and reacted at 37 °C. Time points were taken at the stated times. All reactions were stopped and DNA extracted with the addition of 1 volume of water and 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1). DNA was precipitated using ethanol and digested with *Pst*I overnight before being separated on a 0.7% TAE agarose gel for 18 hours at 60 V. The gel was transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membrane was exposed to a phosphor screen and image taken using a Fuji phosphor scanner. DSBs are marked with an arrowhead (>). Dotted lines on the schematic indicate gDNA.

end resection has occurred. Such gDNA substrates were incubated with TDP2 under the same conditions that were shown to remove Spo11 from Spo11-oligo complexes (Figure 4.2). To determine whether Spo11 had been removed by TDP2, samples were subsequently incubated with lambda exonuclease (λexo). The processive 5′ to 3′ λexonuclease requires a free, unmodified 5′ phosphate group at the DNA end (Subramanian et al., 2003). Thus, any 5′ covalently bound peptide is expected to block resection by λexo, just as it blocks resection *in vivo* by 5′ to 3′ exonucleases such as Exo1. Without TDP2 treatment, λexo is unable to resect the Spo11-DSB molecules (Figure 4.5C). By contrast, pre-treatment of the gDNA with TDP2 before λexo treatment enables λexo to resect both DSBs at the *HIS4::LEU2* hotspot (Figure 4.5D). Therefore, TDP2 can remove a short Spo11-derived peptide from the end of dsDNA even in the presence of many-fold excess competitor gDNA.

λexo is a commercial enzyme from *E. coli* and thus, although it allows for a positive read out for TDP2 treatment on Spo11-DSBs, it has little biological context for meiosis. The main 5′ to 3′ exonuclease in meiosis is Exo1 (Zakharyevich et al., 2010). Spo11 creates DSBs with a 2 bp 5′-overhang but the action of the MRX complex and Sae2 *in vivo* generates a short 3′ ssDNA tail (~270 nt) onto which, Exo1 loads and catalyses further 5′ to 3′ resection (Zakharyevich et al., 2010). The importance of this initial end-processing and resection by MRX/Sae2 has been linked to the promotion of HR at the expense of NHEJ (Shibata et al., 2014).

Therefore, it was of interest to determine whether, if Spo11 were to be removed directly through hydrolysis of the covalent bond by TDP2, recombinant Exo1 would be able to resect the resulting DNA substrate. Normally the bound Spo11 protein would be a block itself to the 5′ to 3′ exonuclease but with TDP2 pre-treatment this would free the 5′ end, as it did for λexo, for the potential access to Exo1. In collaboration with Petr Cejka's lab, recombinant Exo1 was purified from the baculovirus system (Cannavo et al., 2013). As expected, the Spo11 peptide bound to the 5′ end of the DSBs prevents any resection by Exo1 (Figure 4.5A). The question was then asked as to whether Exo1 could resect the DSBs once this peptide was removed. Pre-treatment with TDP2 followed by reaction of the substrate with Exo1 for up to 15 min, however, still resulted in no resection by Exo1. This intriguing result suggests that even when the Spo11 peptide block is removed, Exo1 is still incapable of initiating resection.

To verify these results and to further test the activity of TDP2, the experiment was conducted on genomic DNA from mid-meiotic prophase, extracted in the absence of proteolysis (see Materials and Methods). This procedure extracts gDNA and any covalently attached protein whilst removing all non-covalently bound proteins. Under these conditions, full-length Spo11 will be bound to the 5′ ends of any DSB molecules rather than a short peptide. This material was then incubated with either λexo or Exo1 for up to 15 min with or without pre-treatment of

TDP2 (Figure 4.6). As with proteolysed gDNA substrate, λexo resects DSB molecules only after pre-treatment with TDP2 (Figure 4.6C&D), whilst Exo1 is unable to resect with or without pre-treatment with TDP2 (Figure 4.6A&B).

4.7 Recombinant Exo1 can resect genomic DNA containing DSBs with 3′ overhangs

Because of the failure of Exo1 to degrade *sae2*Δ-derived DSBs that had been deprotected by TDP2, it was important to validate the activity of the recombinant protein preparation. One possible explanation for the negative results when using Exo1 may be due to a buffer incompatibility. To remove any potential inhibitory effects of any of the components of the TDP2 buffer on Exo1 activity, TDP2 activity in the Exo1 reaction buffer was tested (Figure 4.7). Incubation of Spo11-oligos, enriched from mid-meiotic prophase, with TDP2 in the Exo1 reaction buffer resulted in faster migration of the oligonucleotides as before (Figure 4.7A). This observation indicates that TDP2 is active in the Exo1 buffer, and means that the prior assays can all be repeated using the Exo1 buffer, thus avoiding any buffer swaps or potential carry over of inhibitory components from the TDP2 reaction buffer.

Next, genomic DNA was purified from an *exo1*Δ *dmc*Δ strain at mid-meiotic prophase. This strain accumulates DSBs that have been processed *in vivo* and potentially resected by the MRX complex and Sae2. The *dmc1*Δ mutation prevents any strand invasion and repair from occurring thus allowing DSBs to accumulate. The resulting DNA substrate contains DSBs with 3′ ssDNA tails up to 270 nt in length due to no further 5′ to 3′ resection by Exo1 (*exo1*Δ) (Zakharyevich et al., 2010). This substrate is expected to be similar to the substrate that Exo1 acts upon *in vivo*. Thus if resection is observed, Exo1 activity of the recombinant protein can be verified. Incubation of *exo1*Δ *dmc*Δ gDNA, extracted from mid-meiotic prophase, with recombinant Exo1 leads to resection of these 3′ overhang DSB substrates even though they are in the presence of excess competitor gDNA. Additionally, and importantly pre-treatment with TDP2 does not negatively affect Exo1 activity on this type of DNA substrate (Figure 4.7B – final column).

These results confirm that despite recombinant Exo1 being active it is unable to resect a 2 nt 5′ overhang generated by TDP2-dependent removal of Spo11 (Figure 4.5 & 4.6). These observations agree with a report that *in vitro,* Exo1 resection has varying levels of efficiency depending on the composition of the DNA end (Cannavo et al., 2013). In particular, it was noted that the preferred substrate of Exo1 is dsDNA with a 3′ extension, similar to that created by MRX/Sae2 *in vivo* (Garcia et al., 2011), whereas Exo1 is less efficient at resecting blunt

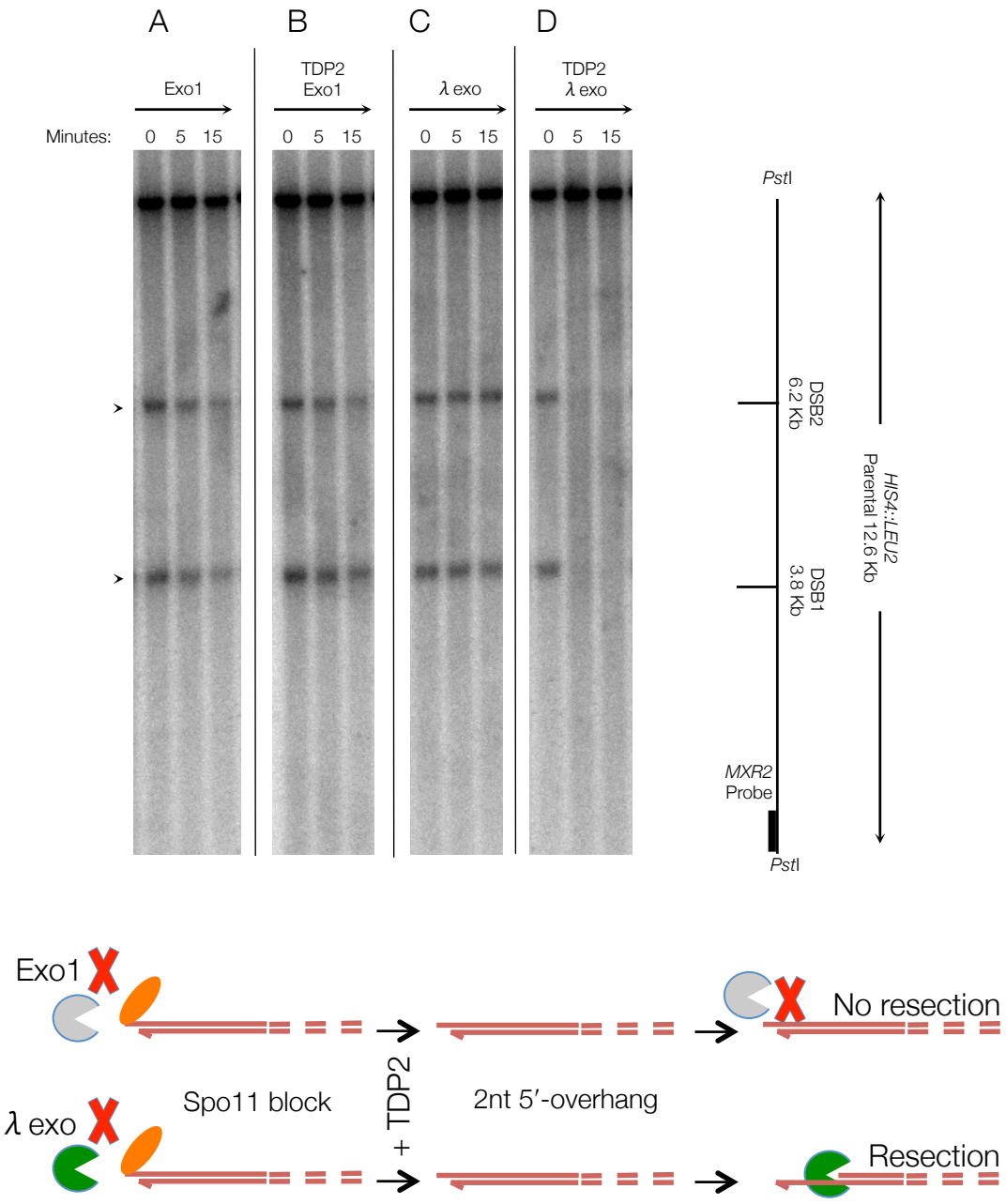


Figure 4.6: TDP2 can remove unproteolysed/full-length Spo11 from the end of double-stranded DNA to enable λ exonuclease, but not Exo1, to resect DSBs.
Southern blot of DSBs at the *HIS4::LEU2* meiotic hotspot. Experiments were performed identically to Figure 4.5, except unproteolysed genomic DNA (rather than proteolysed genomic DNA) was used as a substrate. DSBs are marked with an arrowhead (>).

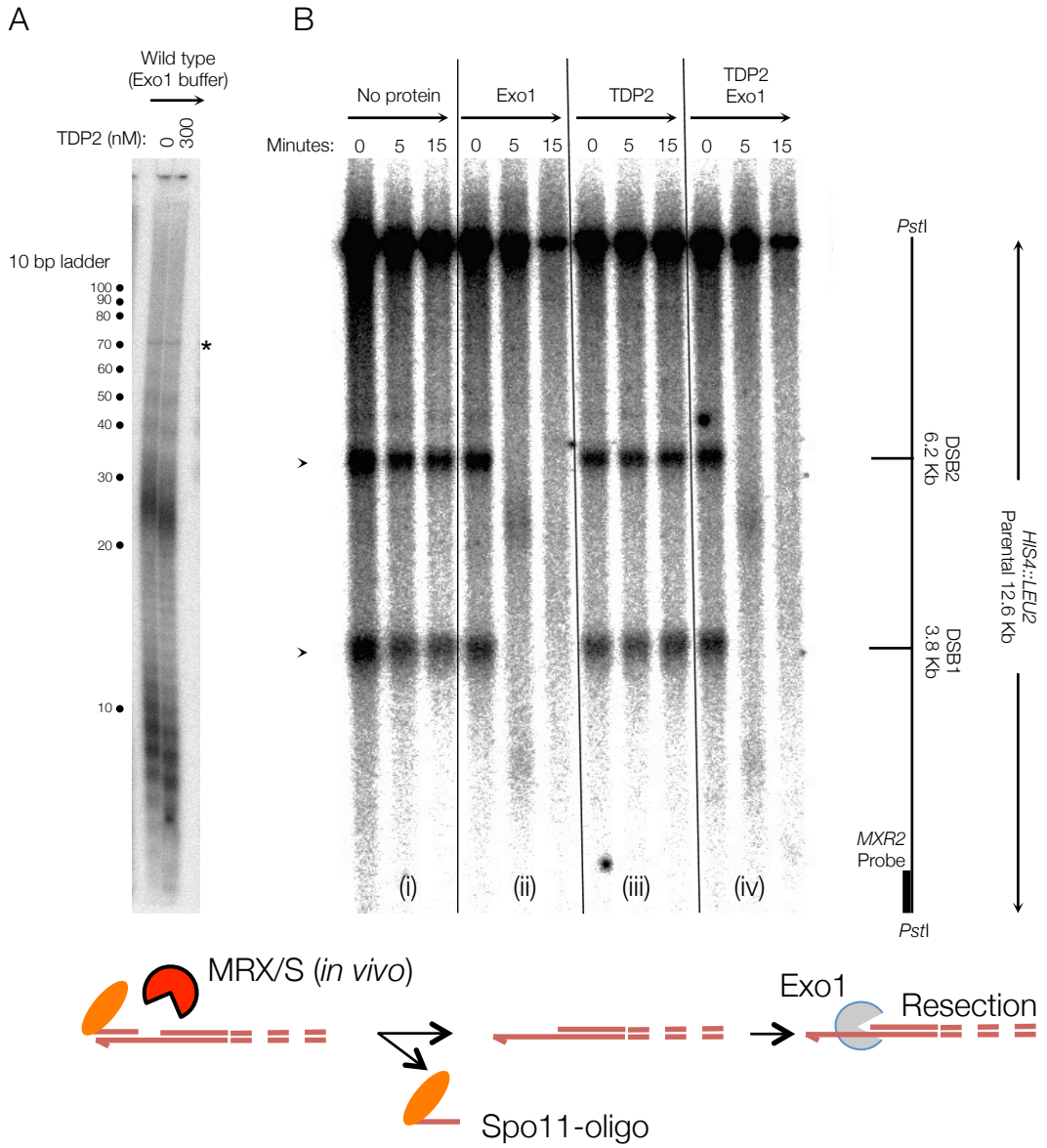


Figure 4.7: Recombinant Exo1 can resect purified genomic DNA from an *exo1*Δ *dmc1*Δ background.

A. Spo11 complexes were 3′ end labelled using TdT and dCTP radionucleotide. Spo11-oligos were digested with Proteinase K at 60 °C for 1 hour and DNA precipitated in ethanol overnight at -80 °C. DNA precipitates were resuspended in Exo1 reaction buffer, reacted with stated concentrations of TDP2 at 37 °C for 1 hour and fractionated by nucleotide resolution urea-PAGE (19% acrylamide). The 10 bp ladder was labelled using TdT and dCTP radionucleotide. Asterisks are non-specific labelled contaminants. **B.** Southern blot of DSBs at the *HIS4::LEU2* meiotic hotspot. A meiotic time course was performed for an *exo1*Δ *dmc1*Δ strain and samples taken at 6 hours. Proteolysed genomic DNA was extracted and split into four sub-samples. (i) and (ii) were incubated at 37 °C for 1 hour in TDP2/Exo1 reaction buffer without the addition of TDP2 protein. (i) was incubated at 30° C without the addition of Exo1 protein ('No protein') and (ii) with 20 nM Exo1 ('Exo1') and 680 nM RPA proteins. (iii) and (iv) were incubated at 37 ° C for 1 hour in TDP2/Exo1 reaction buffer with the addition of 300 nM TDP2 protein. (iii) was incubated at 30° C for the stated time without the addition of Exo1 protein ('TDP2') and (iv) with 20 nM Exo1 ('TDP2 Exo1') and 680 nM RPA proteins. Time points were taken at the stated times. All reactions were stopped and DNA extracted with the addition of 1 volume of water and 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1). DNA was precipitated using ethanol and digested with *Pst*I overnight before being resolved on a 0.7% TAE agarose gel for 18 hours at 60 V. The gel was transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membrane was exposed to a phosphor screen and image taken using a Fuji phosphor scanner. DSBs are marked with an arrowhead (➤).

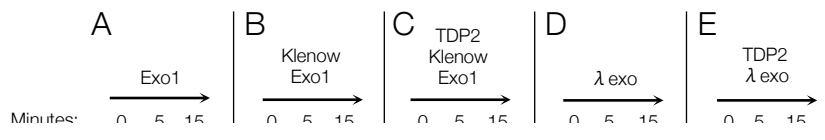
ended substrates and significantly impeded by dsDNA with a 5′ overhang (Cannavo et al., 2013).

4.8 Exo1 cannot resect DSBs with 5′ overhangs or ends blunted with Klenow

To determine whether blunting the 2 nt 5′ overhang would be sufficient to stimulate Exo1 activity, Klenow fragment and dNTPs were used to blunt the 2 nt 5′ overhang after TDP2 removal of the covalently bound peptide (Figure 4.8). This genomic DNA substrate with blunted DSBs was reacted with Exo1 as in prior experiments. As a positive control to confirm that TDP2 was removing the bound peptide, parallel reactions were also incubated with λexo (Figure 4.8D-E). Whereas λexo was still able to resect the DSBs after TDP2 pre-treatment, Exo1 remained unable to initiate resection (Figure 4.8A-C). One possibility to explain the inability of Exo1 to initiate resection at 2 nt 5′ overhangs, or blunt ends used in this system, is because of the excess of competitor genomic DNA negatively affecting Exo1 compared to published *in vitro* experiments using linear plasmid substrates (Cannavo et al., 2013).

4.9 Ectopic expression of TDP2 is unable to cleave Spo11 from DSBs in *S. cerevisiae* meiosis

The results presented above indicate that TDP2 is capable of removing denatured, full-length Spo11 from the ends of DSBs *in vitro*. *S. cerevisiae* does not contain an ortholog of the human TDP2 protein but a previous study saw that ectopic expression of human TDP2 in yeast cells rescued *tdp1*Δ *rad1*Δ sensitivity to CPT (Cortes-Ledesma et al., 2009) indicating that human TDP2 is active within this environment. To determine if TDP2 could perform this activity *in vivo* during meiosis the *TDP2* gene was placed under the control of the *GAL1* promoter on a centromeric plasmid and transformed into a *sae2*Δ strain. This strain will accumulate unresected DSBs during meiosis due to the *sae2*Δ mutant. *TDP2* induction was initiated at 4 hours after the start of meiosis and Spo11-DSBs were monitored using Southern blotting of the *HIS4::LEU2* hotspot as in previous figures (Figure 4.9A). However, no difference was observed between induced, uninduced or control (*sae2*Δ only) samples at any time points (Figure 4.9A right hand panel). DSBs accumulated in all strains with levels, or electrophoretic mobility of DSBs, unchanged. To test whether TDP2 was being effectively induced in these experiments *TDP2* expression was monitored using Western blotting on whole cell lysate using a specific anti-TDP2 antibody (gifted from the Caldecott lab). Bands indicative of TDP2 were present only in the induced samples. As expected, these species migrated slightly faster than the TDP2-His12 recombinant protein (right hand most lanes), which was used as a control for the specificity of the antibody.



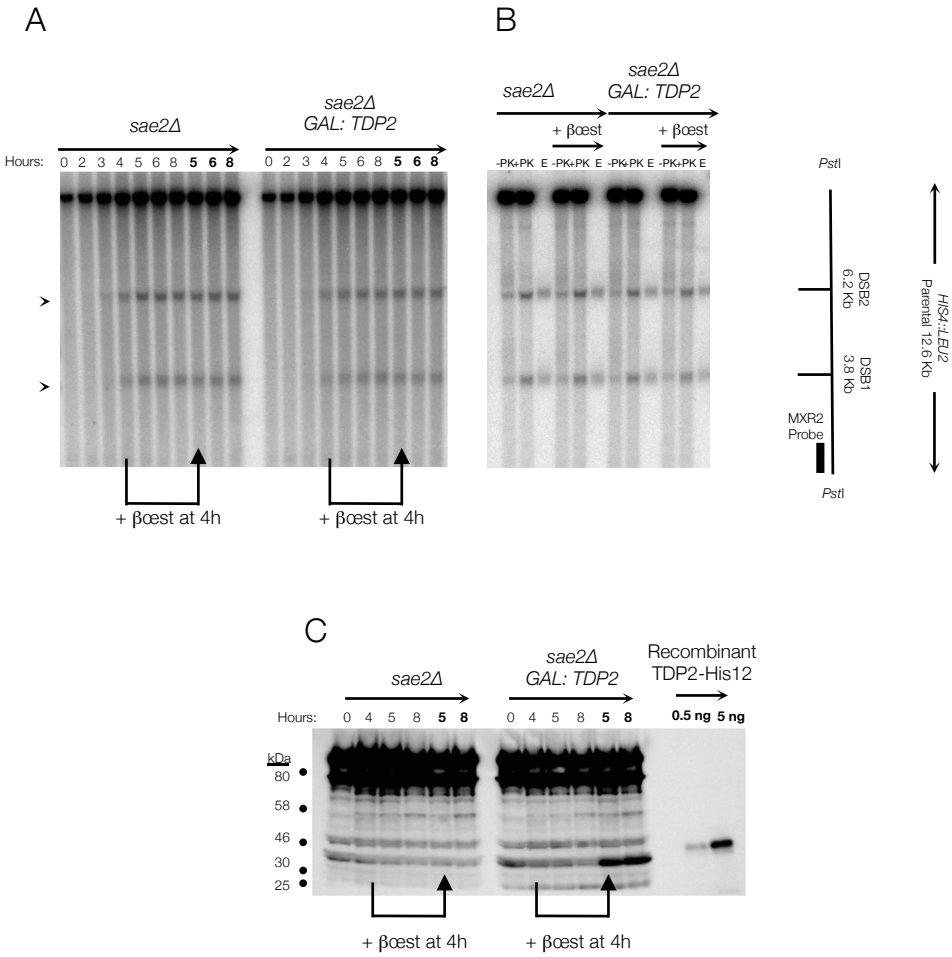


Figure 4.9: Induction of *TDP2* expression *in vivo* in *sae2*Δ cells has no effect on Spo11-DSBs at *HIS4::LEU2*.

TDP2 expression was placed under the control of a galactose inducible promoter on a centromeric plasmid (hygromycin resistance). *sae2*Δ cells were transformed with an empty hygromycin resistance plasmid or the *TDP2* expression plasmid. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmid. 10 ml of cells were taken at stated time points from the start of meiosis. *TDP2* expression was induced with the addition of 2 μM βœstradiol at 4 hours. **A.** DSBs were detected at the *HIS4::LEU2* locus, using Southern blotting as described in previous figures, in the presence and absence of *TDP2* induction. **B.** 8 hour time points from both *sae2*Δ and *sae2*Δ *GAL:TDP2* stains, with and without *TDP2* expression, were taken and the Spo11-DNA enrichment protocol performed. '–PK' is without Proteinase K digestion, '+PK' is with Proteinase K digestion 65 °C for 30 minutes, 'E' is the eluate from the column enrichment. **C.** Western blot on whole cell lysate from stated time points. 0.5 ng and 5 ng of recombinant TDP2-His12 loaded. TDP2 was probed with anti-TDP2 antibody at 1:5000 and anti-mouse secondary antibody used at 1:5000.

If TDP2 can remove Spo11 *in vivo*, the outcome of the DSBs in meiosis is unknown. In principle, DSBs could be repaired as normal, with Exo1 able to resect the now clean end – however, evidence presented in Figure 4.7 would indicate this may not be possible. Alternatively DSBs may now be repaired by NHEJ because the clean ends would be capable of simple religation. Both these explanations though would be visualised on the Southern blots by the DSB bands at *HIS4::LEU2* disappearing over time in a TDP2-dependent manner, which is not the case. Alternatively, TDP2 may have removed Spo11 from the DSB ends but both Exo1 is incapable of resecting these 5′ overhangs (as indicated by Figure 4.7) and the NHEJ machinery is inhibited in meiosis by factors such as CDK (Hentges et al., 2014). In this scenario unrepaired DSBs would accumulate and not disappear from the Southern blots even if TDP2 were functional. To test for this scenario, glass filter columns were used to enrich for protein-associated DNA molecules in samples prepared from the *sae2*Δ samples with and without induction of *TDP2* (see Materials and Methods and Chapter 5 for more details). Protein-associated DSB eluates were then resolved on an agarose gel and visualised by Southern blotting (Figure 4.9B). No reduction in protein associated DNA molecules were observed in the TDP2-induced samples suggesting that TDP2 is not capable of removing Spo11 from the ends of the DSBs in meiosis.

4.10 Constitutive expression of *TDP2* cannot remove Spo11 from DSB ends

TDP2 expression by the meiotic *GAL1* induction system did not produce large amounts of TDP2 (Figure 4.9C), suggesting that it may have been limiting for the Spo11 removal reaction *in vivo*. As an alternative, a constitutive expression system was made by cloning *TDP2* downstream of the highly expressed, constitutive *ADH1* promoter in a centromeric plasmid. The plasmid containing this construct was transformed into *sae2*Δ cells and DSB formation around the *HIS4::LEU2* hotspot analysed as before (Figure 4.10). DSBs from strains with or without the TDP2 expression plasmid did not differ again suggesting that TDP2 is not capable of removing Spo11 *in vivo*. To test whether TDP2 had removed Spo11 but the DSBs were unrepairable, the genomic DNA from these samples were extracted and reacted with λexo. λexo should be able to resect the DSBs if Spo11 has been removed *in vivo*. However, no resection was observed with or without TDP2 expression *in vivo* (Figure 4.10B). Therefore, it is concluded that TDP2 cannot hydrolyse the phosphotyrosyl bond between Spo11 and the DSB ends *in vivo*, contrary to its activity in an *in vitro* system.

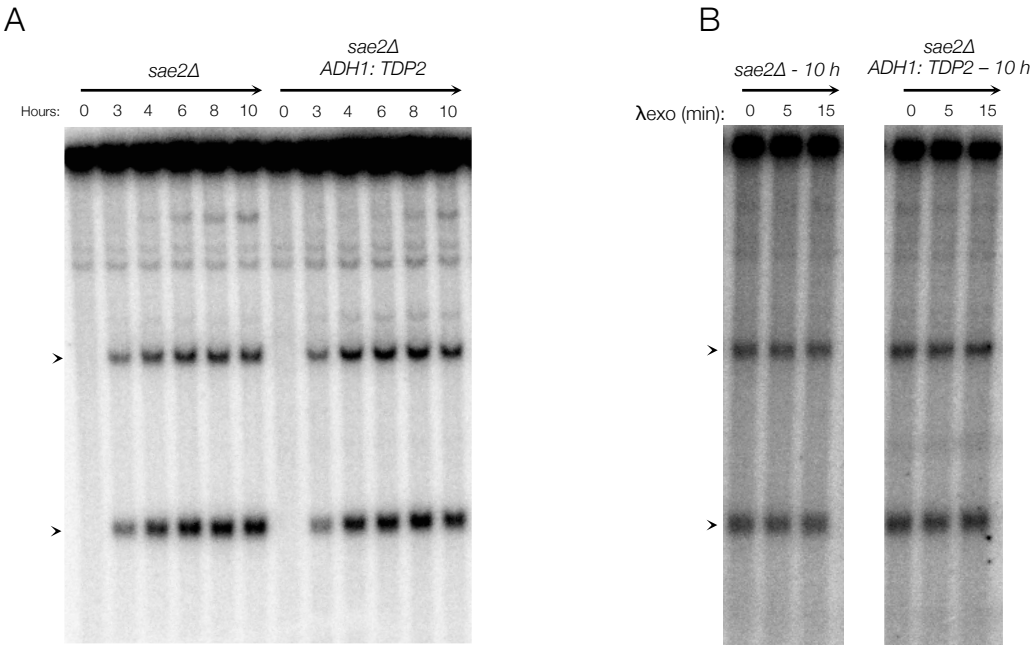


Figure 4.10: Constitutive expression of *TDP2* cannot remove Spo11 from DSB ends.
TDP2 expression was placed under control of the highly expressed *ADH1* promoter on a multi-copy plasmid, marked with hygromycin resistance. *sae2*Δ cells were transformed with an empty hygromycin resistance plasmid or the *ADH1*:*TDP2* plasmid and transformants selected for hygromycin resistance. YPD and YPA cultures were grown in the presence of 300 μg ml^{-1} hygromycin to maintain selection of the plasmids. 10 ml of cells were taken at stated time points from the start of meiosis and genomic DNA extracted. **A.** Southern blot at *HIS4::LEU2* hotspot. Genomic DNA was digested with *Pst*I overnight before being separated on a 0.7% agarose gel for 18 hours at 60 V. **B.** Genomic DNA from both strains at 10 h into meiosis were reacted with λexo at 37 °C for stated time to monitor potential Spo11 removal by TDP2. Reactions were stopped with addition of Proteinase K and incubated at 60 °C for 30 min. DNA was digested with *Pst*I overnight before being separated on a 0.7% agarose gel for 18 hours at 60 V. The gels (**A** & **B**) were transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membranes were exposed to phosphor screens and images taken using a Fuji phosphor scanner. DSBs are marked with an arrowhead (>).

4.11 Discussion

TDP2 has previously been shown to cleave phosphotyrosyl bonds on synthetic substrates as well as *in vivo* on proteolysed TOP2-DSB complexes (Cortes-Ledesma et al., 2009; Gao et al., 2012). There is no yeast ortholog of human TDP2, enabling the study of its activity on Spo11-DSBs in meiosis. Interestingly, TDP2 is expressed in mouse and human testes tissue (Pype et al., 2000) as well as in the *C. elegans* germline (Shi et al., 2012), which could indicate that TDP2 plays a role on a subset of Spo11-DSBs, but so far there has been no reports demonstrating this. Here, TDP2 is shown to act upon the phosphotyrosyl bond of Spo11 when bound to ssDNA oligonucleotides (Figure 4.2&4.4) or on dsDNA ends (Figure 4.5-4.6) *in vitro*. Similarly to TOP2-DSBs TDP2 is much more active when the bound protein has been proteolysed however, TDP2 can also act upon the bond with full-length Spo11 attached. The reason for reduced activity upon the full-length Spo11-DNA substrate may be because the catalytic site of TDP2 is buried within the core of the protein (Shi et al., 2012), which impedes the access of the phosphotyrosyl bond when a large protein structure is present. This may explain why ectopic expression of *TDP2* could not act upon Spo11-DSBs *in vivo* in meiosis (Figure 4.9 and 4.10), with the large native protein blocking access. Additionally, in these *in vitro* experiments on gDNA or Spo11-oligos, even though the Spo11 protein was full-length it was also denatured, which may allow access of TDP2 to the phosphotyrosine bond. Spo11 creates DSBs in meiosis as part of a 10-protein complex (Arora et al., 2004; Maleki et al., 2007; Petrini, 1999; Uetz et al., 2000). This complex itself may also restrict access of TDP2 to the phosphotyrosyl bond preventing removal of Spo11. In a Spo11-DSB processing deficient strain, such as *sae2*Δ, Mre11 foci accumulate at the DSB site (Usui et al., 2001). This accumulation may also restrict access of TDP2 to the phosphotyrosyl bond. Unfortunately, because the MRX complex is required for Spo11-DSB formation (reviewed in Lam and Keeney, 2015), deletion of *MRE11, RAD50* or *XRS2* to overcome this potential issue is not possible. However, recently two N-terminal mutants of Mre11 were reported that decreased binding and retention of Mre11 to DSBs in a *sae2*Δ background (Chen et al., 2015; Puddu et al., 2015). If this mutant prevents the accumulation of Mre11 at Spo11-DSBs, combining this mutant with a *sae2*Δ background and expressing TDP2 may enable TDP2 to gain access to the Spo11 phosphotyrosine bond. Other avenues of investigation would be to look at much later time points after induction of TDP2. This is because previous reports have observed a fraction of Spo11-DSBs behaving as if partially or fully proteolysed after prolonged meiotic arrest in a *sae2*Δ background (MJ Neale – unpublished observation).

If TDP2 were able to remove Spo11 *in vivo*, the outcome of the DSBs is unknown. Potentially three outcomes are possible: (i) The HR machinery begins to resect the DSBs and HR

proceeds as normal with only a delay. (ii) With the DSB now clean-ended and initial resection by MRX and Sae2 absent, NHEJ is free to repair the DSBs. (iii) Neither HR or NHEJ can repair the DSBs leading to accumulation of DSBs. At clean-ended mitotic DSBs created by the HO-endonuclease, loss of Sae2 or any component of the MRX complex only slows resection with HR still occurring but just somewhat delayed (Mimitou and Symington, 2009). However, the HO-endonuclease creates 4 nt 3′ overhangs (Jin et al., 1997), which is a preferred substrate for Exo1 compared to the 2 nt 5′ overhang present at a clean-ended Spo11-DSB. Similar to the HO-DSBs, I-*Sce*I also generates clean-ended DSBs with 3′ overhang ends (Colleaux et al., 1988). Therefore, comparison between DSBs containing clean 3′ overhangs compared to a potential clean 5′ overhang generated from TDP2 activity on Spo11-DSBs may not be an accurate reflection due the apparent 5′ overhang impeding resection by Exo1 (Figure 4.5). Exo1 activity on mitotic DSBs containing 5′ overhangs has yet to be demonstrated. The other aspect to consider is that the Sgs1-Dna2 pathway can act as a redundant resection pathway at mitotic breaks, yet *sgs1* mutants have no defect in resection in meiosis (Zakharyevich et al., 2010). Therefore, comparing differences between mitotic and meiotic 5′ overhang DSBs is difficult because Sgs1-Dna2 may be much more proficient at resecting these types of DSB end than Exo1. Preliminary data in the Neale lab attempting to reconstitute Sgs1-Dna2 dependent resection on Spo11-DSB ends have so far suggested that Sgs1-Dna2 is unable to act upon these short 5′ overhang DSB ends but further clarification is required.

Collectively, these observations point towards the essential nature of MRX and Sae2 activity on processing Spo11-DSBs (Figure 4.11). MRX and Sae2 activity drive repair of DSBs down the HR pathway by endonucleolytically cleaving off Spo11 and generating 3′ overhang DSB ends that are refractory to NHEJ. This is an efficient way of removing Spo11 as well as generating a favourable substrate for the sole meiotic resecting enzyme, Exo1, to resect 5′ to 3′. These observations may point to an evolutionary drive towards processing Spo11-DSBs in this manner as if a pathway was present that simply clips off Spo11 at the phosphotyrosyl bond (e.g. via TDP2), this would create inefficient substrates for long-range resection to occur from, a process that is necessary for efficient meiotic recombination.

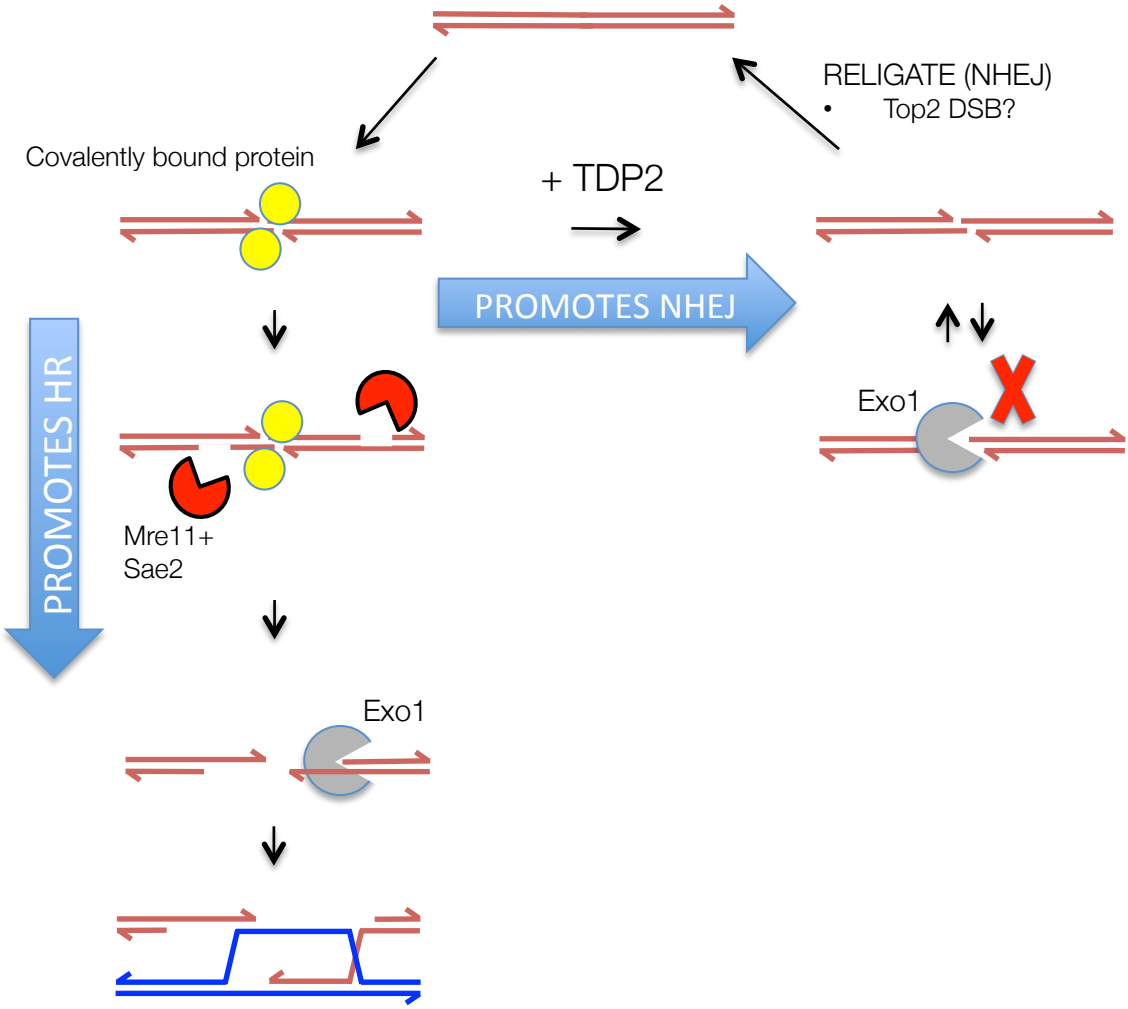


Figure 4.11: Model of DSB repair via HR or NHEJ depending on repair mechanism.
The activity of the DSB repair mechanism may alter the outcome of DSB repair. Covalently bound protein at DSBs has to be removed in order for DSB repair to occur. Mre11 and Sae2 initiate repair by nucleolytically removing the protein bound to an oligonucleotide. This process generates a 3′ ssDNA overhang, a substrate refractory to NHEJ and the preferred substrate for long-range resection by Exo1. This resection promotes homologous recombination by allowing homology search and strand invasion. Removal of covalently bound protein, for example, by hydrolytically cleaving the phoshotyrosyl bond between the protein and the 5′ end of the DSB (e.g. by TDP2) generates clean, short 5′ overhang ends. These ends are refectory to Exo1 resection, thus preventing repair by HR, but the complementary nature of the two ends potentially promotes repair by NHEJ, for example at Top2-DSBs.

CHAPTER 5:

GENOME-WIDE MAPPING OF SPO11 DSBs

Chapter 5: Genome-wide mapping of Spo11-DSBs

5.1 Introduction

Genome-wide mapping of protein interactions with DNA has been conducted in a variety of different ways. ChIP-seq requires affinity tags or synthesis of ChIP-seq grade, protein-specific antibodies to identify regions of DNA a protein associates with *in vivo*. Specific protocols to enrich, and directly sequence, protein-DNA species have been developed, such as the Spo11-oligonucleotide mapping procedure, which uses immunopreciptiation of affinity-tagged Spo11 (see below) (Pan et al., 2011). These methods have produced a wealth of data for protein-DNA interactions, with the Spo11-oligo mapping data generating near nucleotide-resolution of the sites where Spo11 cleaves the DNA phosphate backbone to create DSBs in meiosis.

These methods have limitations on the data they produce. ChIP-seq generates data detailing regions of DNA where proteins reside *in vivo*. However, the specific interaction/binding sites, within nucleotide resolution, are not achievable. Whilst this is not always necessary, generating sequence bias and high-resolution base pair interactions between other proteins, is not possible. Protocols such as the Spo11-oligonucleotide mapping assay (Pan et al., 2011) allows much higher resolution maps to be generated. Spo11-oligo mapping involves purification of Spo11 covalently bound to oligonucleotides and the ligation of an adaptor to the Spo11-bound end. From preparing samples in this manner Spo11-DSB sites, in a wild type cell, were mapped with near nucleotide resolution, the one caveat being that the rGTP-tailing used within this protocol generates some discrepancy with real 5′ cytosines. Mapping of Spo11-oligos require affinity tags to be incorporated onto the protein of interest, which, in the case of Spo11-HA, can alter the activity of the protein (Gray et al., 2013; Martini et al., 2006). For example, the Spo11-HA allele is a DSB formation hypomorph, variably reducing DSBs by 11-50% (Martini et al., 2006). Spo11-HA may also affect DSB formation at some sites compared to others, altering the distribution of DSBs genome-wide (Gray et al., 2013). To overcome these deficiencies, more recent work utilises Spo11-FLAG or Spo11-ProtA alleles, which are reported to display more wild type-like activity (Mohibullah and Keeney, 2016; Thacker et al., 2014). However, these affinity tags may also have some influence on Spo11 activity.

Spo11 generates a phosphotyrosyl bond with both 5′ ends of the DSB and these species accumulate in Spo11 end-processing deficient backgrounds, such as *rad50S*, *sae2*Δ or *mre11* -nuclease deficient mutants. Recombinant, human TDP2 protein is capable of directly hydrolysing the covalent bond between Spo11 and the end of dsDNA, freeing the 5′ phosphate

(Chapter 4). The ability to remove Spo11 from the 5′ end without any loss of nucleotides enables nucleotide-resolution mapping of the precise location of Spo11-DSBs (Figure 5.1). Previous studies in *rad50S* strains mapped Spo11-DSBs in this end-processing deficient background using ChIP-on-ChIP (Gerton et al., 2000); however, the resolution was low in comparison to the Spo11-oligo mapping data (Pan et al., 2011).

5.2 Enrichment of Spo11-bound DNA molecules using glass fibre membranes

Before being able to map Spo11 genome-wide in this manner (Figure 5.1), a way to enrich for Spo11-DSBs was required. Mapping untagged Spo11 removes any possible alterations in Spo11 activity created by an affinity tag, as previously mentioned. A method, adapted from Thomas et al., 1979, that enriched for Spo11-DSBs in a *rad50S* background, was used to originally identify Spo11 as the protein responsible for generating DSB in meiosis. This method involved mass-spectrometry of enriched protein-bound DNA molecules from a *rad50S* strain where Spo11-DSBs accumulate (Keeney et al., 1997). A two-step purification procedure to isolate protein-DNA species consisted of: (i) purifying genomic DNA away from the bulk of cellular proteins using a caesium chloride (CsCl) gradient, whilst retaining covalently bound proteins, and (ii) purifying protein-DNA species away from protein-free DNA by passing the CsCl-purified material over a glass fibre filter, to which proteins specifically adsorb (Thomas et al., 1979). Free-DNA was washed away and protein-DNA species eluted using SDS. The eluate was treated with non-specific nucleases and the sample analysed by mass-spectrometry to identify Spo11 as the protein responsible for DSB formation in meiosis (Keeney et al., 1997). To test for enrichment, gDNA was digested with the restriction enzyme *Pst*I, the two-step purification procedure conducted and a Southern blot used to probe the *HIS4::LEU2* hotspot. A Southern blot of *HIS4::LEU2* contains a parental, unbroken band at the top with the two common DSB sites at *HIS4::LEU2* migrating as two shorter species (Xu and Kleckner, 1995), with the DSBs accounting for ~10% of total lane signal. After protein-DSB enrichment, the two bands corresponding to the two DSBs were still present, however the unbroken parental band was absent (Keeney et al., 1997), indicating enrichment for protein-associated DNA molecules.

Before adapting this method for mapping Spo11-DSBs genome-wide, the verification of Spo11-DSB enrichment was required. The CsCl gradient used in the above method is technically time consuming and challenging. Therefore, the procedure was adapted to extract gDNA away from the bulk of cellular proteins (as detailed in Materials and Methods and schematically in Figure 5.2). A non-proteolysing phenol-chloroform gDNA preparation was

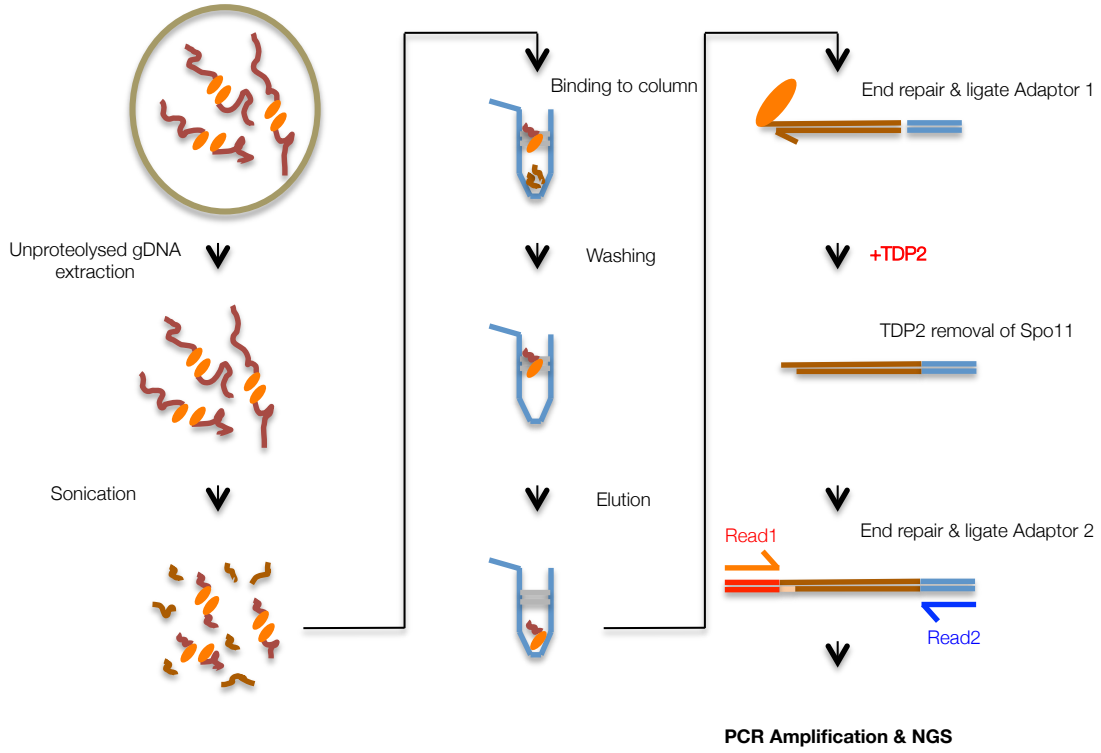


Figure 5.1: Schematic of Spo11 enrichment and mapping.
A meiotic time course is performed for a *sae2*Δ strain and cells harvested at 6 hours. Unproteolysed genomic DNA was extracted by fixing spheroplasts in ethanol, lysing with SDS, and extracting DNA and protein-bound DNA using phenol:chloroform:isoamyl alcohol (25:24:1). Molecules were precipitated with ethanol, resuspended in 1× TE and sonicated to fragment the DNA. The sample is bound to the glass fibre membrane of a QIAQuick spin column, centrifuged and the flow-through rebound to the column and centrifuged again to increase yield. The membrane is washed using TEN (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 300 mM NaCl) to remove any non-protein-bound DNA. Spo11-bound DNA is released from the column using two sequential elutions in 50 µl TES (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.5% SDS). The sonicated end of the DNA fragments are blunted using the NEBNext Ultra II end-repair kit and Adaptor 1 ligated on before removal of unligated adaptors using AMPure XP beads. The fragments are then reacted with TDP2 to remove the Spo11 peptide covalently bound to the 5′ end of the DNA before fill-in of the end that was Spo11-bound. Adaptor 2 is then ligated onto the Spo11-derived end of the fragment, and the now polar molecules are amplified by PCR and 2×75 bp paired-end sequencing conducted using a Illumina MiSeq machine. Read1 contains the 5′ base of where Spo11 cut. Read2 contains information regarding the sheared end of the fragment.

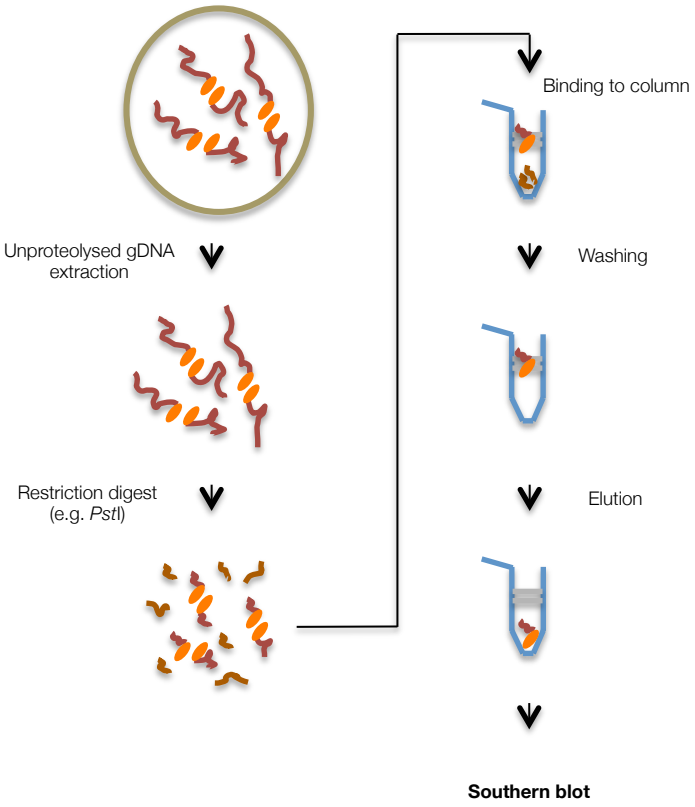


Figure 5.2: Schematic of Spo11 enrichment for Southern blotting.
A meiotic time course is performed for a *sae2*Δ strain and cells harvested at 6 hours. Unproteolysed genomic DNA was extracted by fixing spheroplasts in ethanol, lysing with SDS, and extracting DNA and protein-bound DNA using phenol:chloroform:isoamyl alcohol (25:24:1). Molecules were precipitated with ethanol, resuspended in 1× TE and restriction enzyme digested with *Pst*I for 1 h at 37 °C to fragment the DNA. The sample is bound to the glass fibre membrane of a QIAQuick spin column, centrifuged and the flow-through rebound to the column and centrifuged again to increase yield. The membrane is washed using TEN (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 300 mM NaCl) to remove any non-protein-bound DNA. Spo11-bound DNA is released from the column using two sequential elutions in 50 µl TES (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.5% SDS). Eluates are Proteinase K treated at 60 °C for 1 hour and the sample separated on a 0.7% agarose gel for 18 hours at 60 V. The gel was transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membrane was exposed to a phosphor screen and image taken using a Fuji phosphor scanner.

conducted where gDNA partitions to the aqueous phase and proteins to the phenol/interphase. Proteins bound covalently to the DNA should not prevent this partitioning from occurring. Once extracted, the gDNA was digested with *PstI* to fragment the gDNA. QIAGEN QIAprep 2.0 Spin Columns (referred to as 'columns' from here), contain a glass fibre filter and, under conditions stated in the Materials and Methods, allow specific adsorption of proteins (DNA doesn't bind). Washes with a high salt 1× TE (TEN) buffer removes unbound/free DNA whilst retaining protein-DNA species on the filter. Protein-DNA species are subsequently eluted using 1× TE containing SDS (TES buffer). Eluates were proteolysed using Proteinase K, DNA fragments resolved on an agarose gel and the *HIS4::LEU2* hotspot probed via Southern blotting (Figure 5.3). Spo11-DSBs at *HIS4::LEU2* were enriched, as demonstrated by the presence of the two DSB bands and the absence of the parental, unbroken band (Figure 5.3B lane 6 - ELUATE). The unbroken, non-protein bound parental fragment was observed in the 'wash' fractions (Figure 5.3B lane 4 - WASH 1). Not all Spo11-DSB molecules are enriched on the column, - indicated by the presence of the two DSB bands in the unbound lane (Figure 5.3B –UNBOUND), suggesting proteolysis of Spo11 *in vivo* or during the preparation of the sample, preventing DSB binding to the glass fibre filter. This hypothesis is supported by the presence of the two DSB bands in the non-Proteinase K sample (Figure 5.3B lane 1 - INPUT no PK), which was not column enriched or proteolysed with Proteinase K. Because protein species retard migration on agarose gels, unproteolysed/partially proteolysed Spo11-DSBs will not migrate, indicating that the two DSBs observed in the non-Proteinase K sample is due to proteolysis of some DSB molecules at *HIS4::LEU2*. Collectively these results suggest enrichment of protein-bound DSBs is achievable, with DSBs making up 99% of signal compared to unbroken, parental signal in the enriched eluate. Additionally, Proteinase K treatment of the sample before column purification resulted in no enrichment of Spo11-DSBs via Southern blotting or from sequencing (data not shown), indicating that the protein moiety is an essential factor for this enrichment process.

5.3 Optimisation: A higher concentration of SDS in the elution buffer enriches more Spo11-DNA

Enrichment of the Spo11-DSBs by the method described above, was not efficient with yield of the DSBs at approximately half of the total level (Figure 5.3 – compare INPUT and ELUATE x4). Addition of two sequential rounds of elution buffer (TES) was initially used to elute the Spo11-DNA species from the column. However, in order to increase the yield a series of tests were conducted (Figure 5.4). (i) A single volume of elution buffer was re-passed through the column in an attempt to solubilise more Spo11-DNA species in each subsequent flow-through.

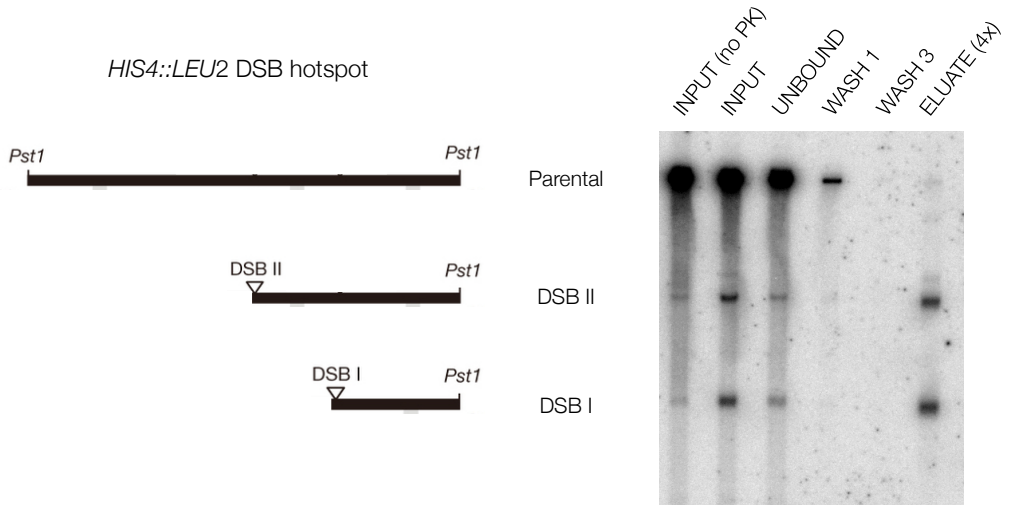


Figure 5.3: Southern blot of Spo11-DSB enrichment.
A meiotic time course is performed for a *sae2*Δ strain and cells harvested at 6 hours. Unproteolysed genomic DNA was extracted as described in Figure 5.2. Molecules were precipitated with ethanol, resuspended in 1× TE and restriction enzyme digested with *Pst*I for 1 h at 37 °C to fragment the DNA ('INPUT'). The sample is bound to the glass fibre membrane of a QIAQuick spin column and centrifuged. The flow-through was rebound to the column and centrifuged again to increase yield (second flow-through – 'UNBOUND'). The membrane is washed using TEN (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 300 mM NaCl) to remove any non-protein-bound DNA (fraction taken – 'WASH'). Spo11-bound DNA is released from the column using two sequential 50 µl TES buffer (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.5% SDS) ('ELUATE'). All samples were Proteinase K treated (unless stated) at 60 °C for 1 hour and the samples separated on a 0.7% agarose gel for 18 hours at 60 V. The gel was transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membrane was exposed to a phosphor screen and image taken using a Fuji phosphor scanner.

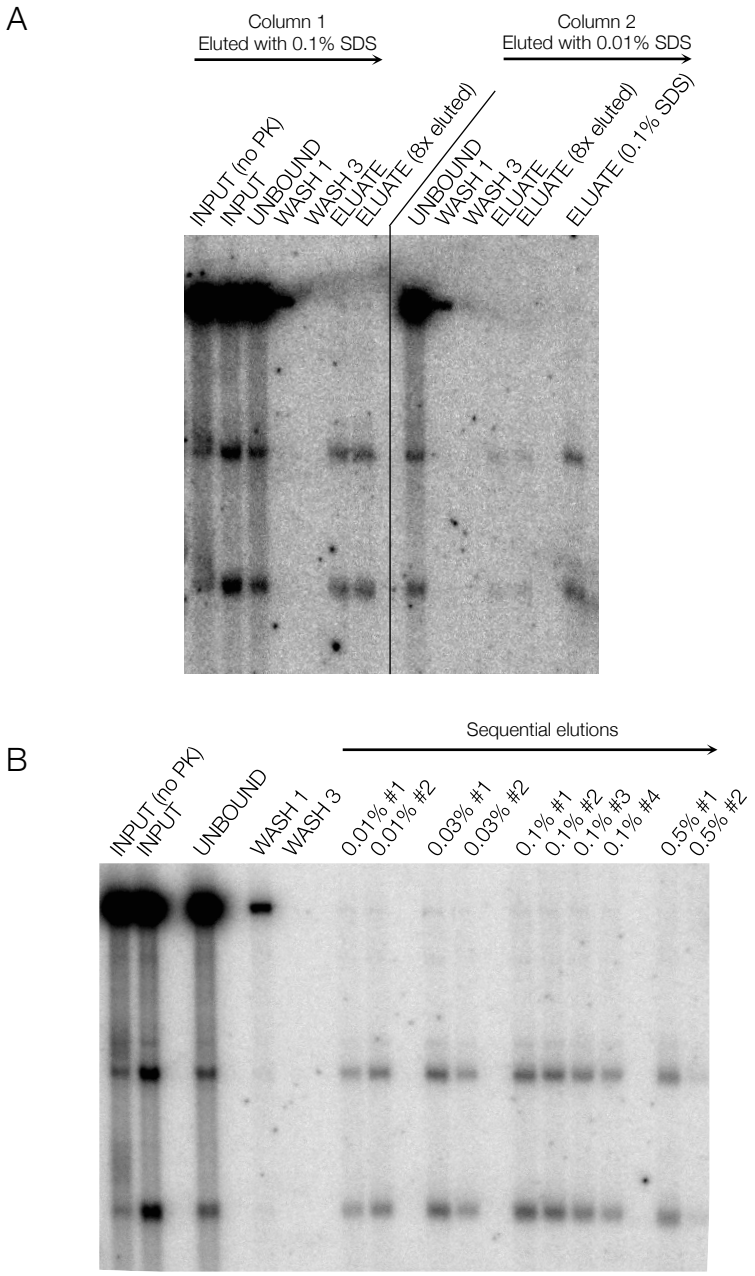


Figure 5.4: Southern blot of Spo11-DSB enrichment elution optimisation.
Southern blot of DSBs at the *HIS4::LEU2* meiotic hotspot. A meiotic time course is performed for a *sae2*Δ strain and cells harvested at 6 hours. Unproteolysed genomic DNA was extracted as described in Figure 5.2. Molecules were precipitated with ethanol, resuspended in 1× TE and restriction enzyme digested with *Pst*I for 1 h at 37 °C to fragment the DNA ('INPUT'). The sample is bound to the glass fibre membrane of a QIAQuick spin column as stated in Figure 5.2 (second flow-through – 'UNBOUND'). The membrane is washed using TEN as stated in Figure 5.2 (fraction taken – 'WASH'). **A.** Spo11-DNA was released from the column with 1× TE containing 0.1% SDS (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.1% SDS) (Column 1) or 1× TE containing 0.01% SDS (Column 2) and this eluate subsequently rebound to the column and eluted eight times. **B.** Spo11-DNA was released from the column with sequential elutions of fresh 1× TE with stated concentrations of SDS (eluates not rebound) . All samples were Proteinase K (PK) treated (unless otherwise stated) before being resolved on a 0.7% TAE agarose gel for 18 hours at 60 V. The gel was transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membrane was exposed to a phosphor screen and image taken using a Fuji phosphor scanner.

This was to check if more Spo11-DNA could be eluted and, if it could, could it be concentrated into a small volume. However, eight passages through the column did not increase the signal in the eluate compared to one passage (Figure 5.4A left hand panel), suggesting that either all the Spo11-DNA had been removed in the first passage or a saturation point had been reached in this volume of elution buffer. (ii) The presence of SDS in the elution buffer may be carried over into subsequent steps and inhibit any reactions required for mapping Spo11-DSBs. Therefore, a lower concentration of SDS was initially tested (0.1% reduced to 0.01%). Using this lower concentration of SDS resulted in a lower yield of Spo11-DNA (Figure 5.4A right hand panel). A subsequent elution with 0.1% SDS (TES) buffer on the same column eluted a large fraction of Spo11-DNA similar to what was eluted in previous elution attempts with 0.1% SDS, suggesting that there is a solubilisation limit of Spo11-DNA, which is dependent on the concentration of SDS. (iii) Subsequent elutions, all on the same column, with multiple increasing concentrations of SDS (0.01%, 0.03%, 0.1% and 0.5%) was conducted (i.e. one column was eluted ten times) (Figure 5.4B). Spo11-DSBs were observed from every elution. Subsequent elutions with the same concentration of SDS yielded progressively less Spo11-DSBs (except for 0.01%). A subsequent elution with a higher SDS percentage eluted more Spo11-DSBs than the last elution with a lower percentage, suggesting a saturation point for Spo11-DNA elution, dependent on SDS concentration. Additionally, there is a fraction of Spo11-DSB molecules that can only be eluted with a higher SDS concentration, suggesting that there may be protein aggregates. The final elution with 0.5% SDS, after the first 0.5% elution, had a low level of Spo11-DSBs, suggesting that the prior elution with 0.5% SDS stripped the glass fibre filter of the majority of Spo11-DSBs. For subsequent steps of the protocol, enriched Spo11-DSBs need to be in a small volume. Therefore, these multiple elutions steps are not possible to enrich for all available Spo11-DSBs. However, using 0.5% SDS in the elution buffer should allow for high percentage of the species to be eluted for mapping and will be used for future experiments.

5.4 Sonication of Spo11-bound genomic DNA generates an unbiased average fragment length for sequencing

In the experiments described so far, genomic DNA was fragmented using the *Pst*I restriction enzyme, which enables specific loci (*HIS4::LEU2*) to be probed to estimate Spo11-DSB enrichment (Figure 5.3-5.4). Restriction enzyme-mediated fragmentation could also be used for generating a genome-wide sequencing library, however, there will be areas of the genome where Spo11 creates DSBs close to, or far from, *Pst*I cut sites. This size bias of specific fragments of the genome may result in loss of information from mapping Spo11 cut sits due to too small, or too large fragment size. To avoid this, a different fragmentation strategy was

required. Sonication of gDNA allows the generation of a heterogeneous size range of DNA molecule, with little or no bias towards certain areas of the genome. This size range can be controlled by sonication time length, with a target size of DNA molecule needed for efficient PCR amplification and mapping being ~500 bp. To optimise the sonication step, a sonication time course was conducted with samples taken at stated time points (Figure 5.5). Sonication up to 24 min decreased the average DNA molecule size in an exponential manner. 24 min sonication resulted in a 500 bp average size of DNA molecule, and will be used for subsequent column enrichments.

5.5 Retaining polarity of the fragmented DNA to distinguish the Spo11-bound end from the sonicated end

Using restriction digest to fragment the genome enables the Spo11-bound end of the fragment to be distinguished from the other, due to the known consensus site for a restriction enzyme. To retain this knowledge of polarity of the molecules, the activity of recombinant TDP2 can be utilised. The first adaptor is ligated to the sheared end with excess adaptor washed away. This adaptor cannot be ligated to the Spo11-end of the fragment due to the presence of the bound Spo11 protein. Removal of Spo11 by TDP2 allows a second adaptor, with a different DNA sequence, to be selectively ligated to the Spo11-end, without loss of nucleotides, generating polarity of the molecules. However, proteolysed Spo11-DNA is a more efficient substrate for TDP2 cleavage compared to full-length Spo11 (Figure 4.4), with the remaining ~3 amino acid peptide, from proteolysis by Proteinase K, still blocking resection by λexo (Figure 4.5). Therefore, once Spo11-DNA has been enriched and eluted from the column, the large Spo11 moiety is now unnecessary and Proteinase K digestion of the eluate can be performed. Ligation of unique adaptors onto each end permits paired-end sequencing and enables single-base pair resolution of where Spo11 cleaves the DNA by bioinformatically extracting the 5′ end of the first read pair (Read1) generated from the Illumina MiSeq machine (Figure 5.1 right hand column). When these dual-adaptor ligated molecules are PCR amplified, they are also barcoded with unique sequences. This barcoding enables multiple samples to be multiplexed in the same sequencing run.

5.6 AMPure XP beads allows size-selection of DNA fragments for optimal library construction

To prepare the Spo11-DNA for sequencing the NEBNext Ultra II Library Prep Kit was used, as described in Materials and Methods. The two main alterations to the protocol included: (i) two custom adaptors (described above) used for ligation to generate polarity of the molecules,

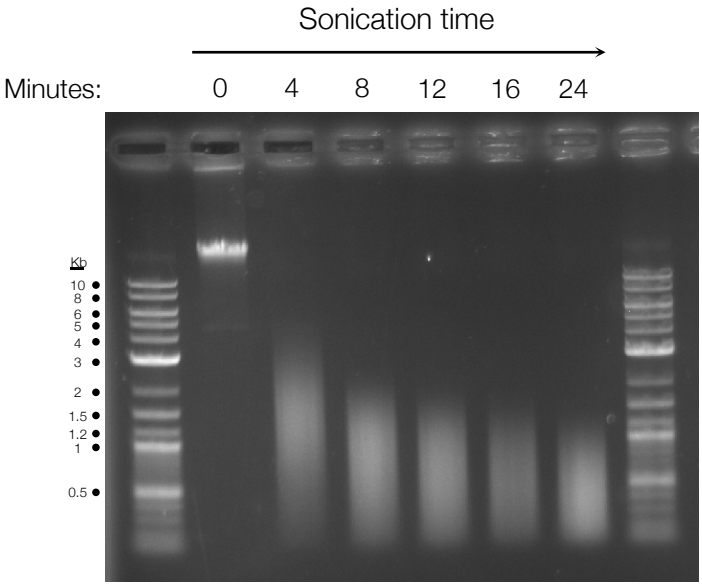


Figure 5.5: Agarose gel of sonicated genomic DNA.
Unproteolysed genomic DNA was extracted from 50 ml *sae2*Δ cells 6 hours into meiosis, dissolved in 1 ml of TE, and sonicated in a Covaris M220 machine. 5 µl was taken at the stated time points and 11.7 µl double-distilled water, 3.3 µl 6× purple loading dye (NEB) added to each. 10 µl was loaded on a 1× TAE agarose gel containing 50 µg ml^{-1} ethidium bromide alongside a 2-log ladder (NEB) and separated for 40 minutes at 100 V. Gels were imaged using a Syngene InGenius bioimaging system.

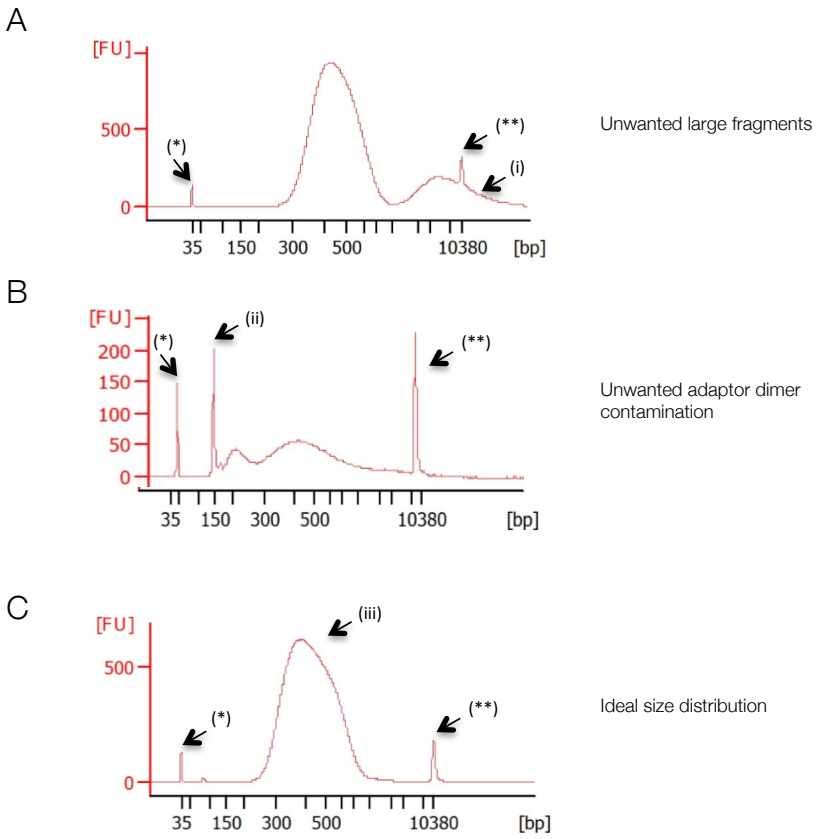


Figure 5.6: Bioanalyser traces to show contaminating small and large fragments compared to ideal size distribution.
Spo11-DNA libraries were diluted 1 in 10 with double-distilled water and 1 µl loaded on an Agilent High Sensitivity chip. Lower marker peak at 35 bp (*). Large marker peak at 10380 bp (**). **A.** Figure shows large DNA fragments that have been amplified by PCR co-migrating with the large marker peak (i). **B.** Figure shows contaminating adaptor dimer peak at approximately 150 bp (ii). **C.** Trace shows an ideal size distribution of DNA (~400 bp, (iii)) with no large or small contaminating peaks.

and (ii) size-selection of molecules was performed after PCR amplification rather than before due to the low concentration of input DNA enriched from the column. Size-selection was carried out using AMPure XP beads. This is a necessary step to remove amplified large molecules and adaptor dimers, both of which negatively affect mapping quality. Figure 5.6 demonstrates a set of example bioanalyser traces from samples where large fragments have been amplified that co-migrate with the large (10380 bp) marker (Figure 5.6A), samples that are contaminated by adaptor dimers, sharp peak at ~150 bp (Figure 5.6B), and a sample, which following AMPure size selection has the desired size range of amplified molecules ready for sequencing (Figure 5.6C).

5.7 The Illumina MiSeq allows multiple samples to be multiplexed, generating 3-7.5 million reads per sample

As described above, fragments have adaptors ligated and are PCR amplified, which also incorporates unique indexes. Double-stranded DNA fragments from different samples are mixed together, denatured and loaded onto a flow-cell for paired-end (2 x 75 bp) sequencing with an Illumina MiSeq machine. Sequencing generates .FASTQ output files containing specific base composition of the paired-end reads – Read1 (from Spo11-end) and Read2 (from sheared end). The individual strain sequences are demultiplexed within BaseSpace (Illumina) generating a pair of unique FASTQ files for each sample. Typically, sequencing libraries on the Illumina MiSeq generates 20-30 million reads per run. Mulitplexing with barcoded samples allows the mixing of 4-6 different samples, each generating 3-7.5 million reads per sample.

5.8 Sequencing reads are aligned to the S288C genome and the location of the 5′- Spo11-bound ends are extracted

Using bowtie2 the demulitplexed sequencing reads are aligned end-to-end to a custom version of the *S. cerevisiae* S288C published genome, that contains the *HIS4::LEU2* and *LEU2::HISG* regions incorporated (‘Cer3H4L2’) (‘Spo11Align.command’ – Appendix 10.1). The resulting .SAM files contain the globally aligned read coordinates, chromosome number, if they are Watson or Crick and their read quality. Using perl, correctly aligned 5′ ends of Read1 (the Spo11 end) are extracted. Reads containing ambiguous 5′ ends are filtered out, whilst reads with an unmapped mate (either Read1 or Read2) are further processed to trim non-5′ ends to attempt re-mapping. Any resulting mapped reads from the trimmed files are combined with the originally mapped reads. The output file (.txt) contains the 5′ co-ordinates of the Read1

sequences, and indicate the chromosome number and whether they come from the Watson or Crick strand ('Spo11Extract.pl' – Appendix 10.2).

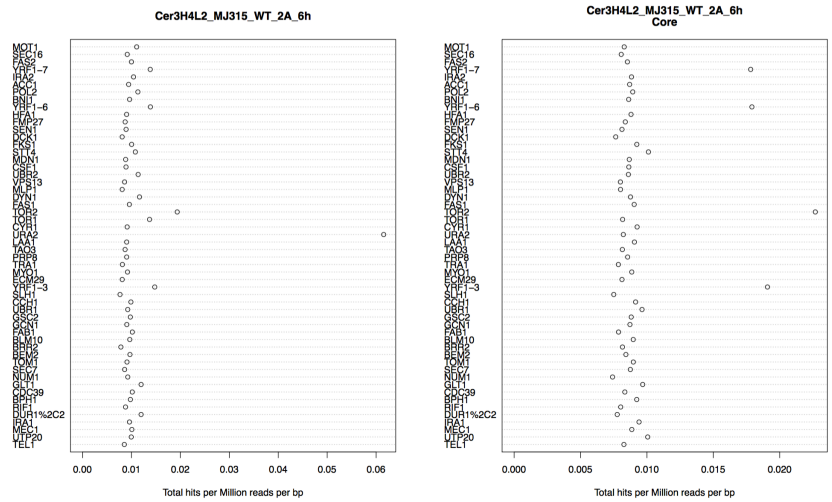
5.9 5′ Spo11 ends are converted into a histogram for analysis

To convert the aligned reads into useful data for analysis, the single nucleotide position of the Spo11 5′ end is aggregated into a 1 bp histogram using R Studio to calculate the number of times a certain base in the genome was broken by Spo11 ('Creating 1bp histograms v03 (H4L2).R' – Appendix 10.3). These histograms are utilised for analysis of the data and the file names annotated as follows: StrainNumber_Genotype_BiologicalRepeatNumber_ TechnicalRepeatLetter_TimePointFromMeiosis. Also note, the 'wild type (WT)' annotation refers to *sae2*Δ because all strains from this mapping procedure contain *sae2*Δ. For example, a *sae2*Δ strain (strain number MJ315) from a second biological repeat ('2') taken at six hours into meiosis ('6h'), from the first attempt at mapping from this material ('A') reads as follows: MJ315_WT_2A_6h.

5.10 Calculating background reads allows for correction and correlation between datasets

Any enrichment method will contain a certain level of contaminating non-specific DNA that may map to the reference genome. These contaminating/background reads need to be controlled for because variation in the background levels may have an effect on real signal levels (a decrease in the signal:noise ratio). Thus, adjusting for background reads enables a more accurate comparison between data sets. Background reads were calculated using any annotated gene ORF over 5.5 kb in length and calculating the hit rate per million reads per base pair ('Calculating background reads v03.R' – Appendix 10.4). ORFs of this length have been reported to contain very few Spo11-DSB sites (Pan et al., 2011); therefore the majority of hits in these regions are expected to be background. The *TEL1* ORF is one such gene greater than 5.5 kb in length, however, as future mapping studies involve *TEL1* mutants, this gene was excluded from the list. ORFs *NUM1, YRF1-7, YRF1-6, YRF1-3, URA2* and *TOR2* were empirically determined to be outliers (Figure 5.7A vs. 5.8B) and also removed, leaving 47 genes >5.5 kb used to estimate background (as seen in the list in Figure 5.7B). Excluding 1 kb proximal and terminal of the ORF ('Core') lowers the standard deviation (see Table 5.1), suggesting some real DSB hits reside in these regions (Figure 5 left hand columns) and thus, using the 'Core' is likely to generate a more accurate determination of background.

A. All ORFs >5.5 kb



B. ORFs >5.5 kb with exclusions

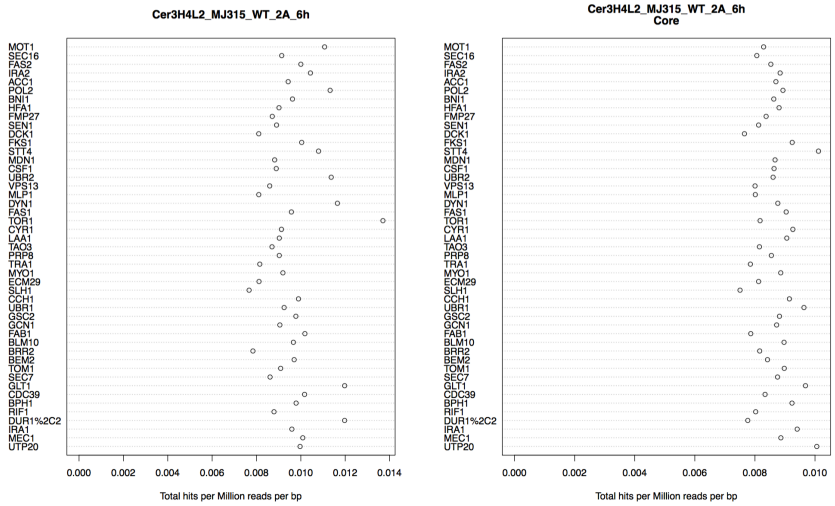


Figure 5.7: Background levels are calculated from the number of hits occurring in the core region of 47 of the largest ORFs.
The R script 'Calculating background reads v03.R' was used to calculate the level of background (putative non-specific reads) from individual samples. Any annotated gene greater then 5.5 kb in length was subset and the number of hits per million reads per base pair within these regions was calculated (left hand panels). Hits within 1 kb proximal and terminal of these ORFs were excluded (right hand panels). **A.** All 54 genes greater than 5.5 kb were used to calculate the background reads. **B.** *TEL1* was excluded from the calculation because the Tel1 protein function will be studied using this method and is not present in all strains sequenced. *NUM1, YRF1-7, YRF1-6, YRF1-3, URA2* and *TOR2* were empirically determined to be outliers and excluded. Overall estimated background reads ('Mean' or 'MeanCore') was calculated from the average number of total hits per million reads per base pair from each stated ORF.

Table 5.1 Background reads from wild type (*sae2*Δ), *spo11-YF* and Spo11-oligo data (Pan_HA)

Strain	Mean (hits/bp/MReads)	S.D. (hits/bp/MReads)	Mean Core (hits/bp/MReads)	S.D. Core (hits/bp/MReads)
All genes >5.5 kb				
MJ315_WT_2A_6h	0.01098	0.0073	0.00940	0.00296
MJ319_Y135F_1_6h	0.07431	0.0090	0.07391	0.00829
Pan_HA_1_4h	0.00489	0.0101	0.00363	0.00316
After removal of outliers				
MJ315_WT_2A_6h	0.00961	0.0012	0.00865	0.00060
MJ319_Y135F_1_6h	0.07664	0.0021	0.07610	0.00273
Pan_HA_1_4h	0.00364	0.0011	0.00331	0.00092

Calculating the background reads between *sae2*Δ (WT) and *spo11-Y135F* (catalytic dead) indicates the majority of hits in the *spo11-YF* mutant are non-specific background (0.07610 hits/bp/MReads*12.5 Mb (size of *S. cerevisiae* genome)/1,000,000 reads = 95% of reads are background – as expected). This figure, compared to the 12% generated from *sae2*Δ (WT), indicates the real signal to noise ratio is much higher in wild type Spo11 strains. Comparing the *sae2*Δ (WT) background to the Spo11-oligo mapping data (Pan_HA) (Pan et al., 2011), which has 4% background hits, suggests that the Spo11-oligo mapping data is more specific for real Spo11-DSB signal. Alternatively, 'real' hits in the long ORFs are not mapped in the Pan_HA data but are in the *sae2*Δ. From calculating the background reads, hits at specific locations can be corrected to visualise a more quantitative image of DSB frequency between strains/mapping procedure.

5.11 Reducing the amount of phenol used to extract gDNA and decreasing the sonication amount reduces the apparent background levels

Background in *sae2*Δ is relatively low (and not too different from the Spo11-oligo mapping data), however, during protocol development it was found that background levels in other strains varied dramatically (data not shown). In order to improve the signal:noise ratio, and to try and improve reproducibility, numerous steps were altered to try and optimise the method and decrease apparent background levels. Contaminating reads, or loss of real signal, could derive from a number of different steps each looked at in detail below.

Phenol

The separation of proteins from DNA by phenol works as follows: Phenol and water are immiscible, therefore when the two solutions are mixed two phases form: an aqueous phase (water) and a phenol phase, sometimes with a large interphase between the two. Thorough mixing of a sample with phenol forces the phenol to mix with the water, forming an emulsion of droplets throughout. Phenol denatures any proteins solubilised in the aqueous (water) phase and this partitions them into the phenol. DNA however, stays soluble in the aqueous phase.

Molecules of DNA with Spo11 covalently bound will be partitioned to either phase depending on the length of the DNA and the strength of the solvent (water/phenol). If DNA molecules are short, the protein being pulled into the phenol phase may overcome the nature of the DNA molecules to partition into the aqueous phase, therefore, these molecules would be lost (as the aqueous phase is taken and used for subsequent steps of the library preparation). Phenol extraction of DNA is conducted prior to fragmentation and therefore, in general, the length of DNA bound to Spo11 is sufficient to partition Spo11-bound DNA molecules to the aqueous phase (as shown by the enrichment at the *HIS4::LEU2* DSB hotspot – Figure 5.3). However, it was considered that lowering the volume of phenol added to the lysed cells may reduce the amount of Spo11-DNA molecules lost to the phenol phase. There is a balance between extracting all DNA with protein covalently attached (desired) and extracting DNA with potentially loosely bound proteins, or even some completely unbound proteins. The latter is not so much of an issue as the only negative effect it may cause is blocking the glass fibre filter and lowering the binding capacity of the column to real protein-bound DNA molecules. The extraction of non-covalently bound proteins that are loosely attached may be more of an issue as these may be retained on the column through the wash cycle and eluted off with non-specific DNA still attached. These DNA molecules would then pass through to the end-preparation and sequencing steps, generating background reads.

Therefore, to determine the effect phenol has on enrichment of Spo11-DNA (and subsequent background levels) a test was conducted using a range of saturated phenol volumes during gDNA extraction (Table 5.2). A decrease in the volume of phenol used decreased the apparent background observed (an increase in the signal to noise ratio) down to 100 µl of phenol (Table 5.2 – 'Mean Core': 0.05779 (250 µl) down to 0.01545 (100 µl) hits/bp/MReads). At 50 µl (a tenth of the aqueous volume), the apparent background increased (0.02549 hits/bp/MReads). This latter result may be due to the low volume of phenol increasing the difficulty of taking solely aqueous phase at the DNA extraction step.

Table 5.2: **Effect of phenol volume on background reads**

Strain	Mean (hits/bp/MReads)	S.D. (hits/bp/MReads)	Mean Core (hits/bp/MReads)	S.D. Core (hits/bp/MReads)
MJ315_WT 250 µl	0.05947	0.0056	0.05779	0.00542
MJ315_WT 175 µl	0.01589	0.0017	0.01496	0.00179
MJ315 WT 100 µl	0.01645	0.0012	0.01545	0.00142
MJ315_WT 50 µl	0.02631	0.0021	0.02549	0.00224

Sonication

To determine the effect varied sonication has on apparent background levels different times of sonication were tested (Table 5.3). An increase in sonication time increased apparent background reads (8 min sonication – 0.0501 hits/bp/MReads to 24 min sonication – 0.0659 hits/bp/MReads), suggesting that less sonication may be more favourable to lower background reads. This may be due to increased sonication generating more non-protein bound fragments (and adaptor ligatable ends) that are retained and eluted from the column and now sequenced.

Table 5.3: **Effect of sonication amount on background reads**

Strain	Mean (hits/bp/MReads)	S.D. (hits/bp/MReads)	Mean Core (hits/bp/MReads)	S.D. Core (hits/bp/MReads)
MJ315_WT 8 min sonication	0.0514	0.0024	0.0501	0.0023
MJ315_WT 12 min sonication	0.0602	0.0019	0.0594	0.0023
MJ315_WT 16 min sonication	0.0652	0.0026	0.0648	0.0031
MJ315_WT 24 min sonication	0.0659	0.0024	0.0659	0.0026

Due to the above findings (Table 5.2 and 5.3), the final protocol involves using a lower volume of phenol compared to the 1:1 ratio of soluble extract:phenol used initially. Less sonication also appears to be more favourable to decrease the apparent background and will be utilised for mapping Spo11-DSBs in future strains.

5.12 Mapping Spo11-DSBs using the *sae2Δ* DSB-enrichment procedure agrees with the Spo11-oligonucleotide mapping data

Deep sequencing *sae2*Δ strains generated between 2-5 million reads that were mapped to the genome of the S288C *S. cerevisiae* strain. Approximately 95% of total read pairs mapped to the S288C genome and of these total mapped pairs, 98.5% generated unambiguous (valid) 5′ hits. Using the 1 bp histogram file, which contains all 5′ hits for the Spo11-DSBs, the entire length of each chromosome was plotted ('Spo11 mapping v09 MC' – Appendix 10.5). Chromosome 1 and chromosome 7 were visualised to compare the Spo11-oligo map (Pan et al., 2011), the new *sae2*Δ data, and newly mapped data from a *spo11-Y135F* (catalytic dead) mutant as a control (Figure 5.8). Strong peaks from both the Spo11-oligo data and the *sae2*Δ are present in the same locations. The *spo11-YF* mutant fails to show significant enrichment at any position in the genome on a chromosome wide (Figure 5.8), local chromosome region (Figure 5.9) or at DSB hotspots on a fine scale (Figure 5.10). Rather, all the reads from the *spo11-YF* mutant average evenly across every bp of the genome, strongly supporting the view that enrichment of Spo11-DSBs at hotspots in the *sae2*Δ data is, as expected, dependent on the catalytic activity of Spo11.

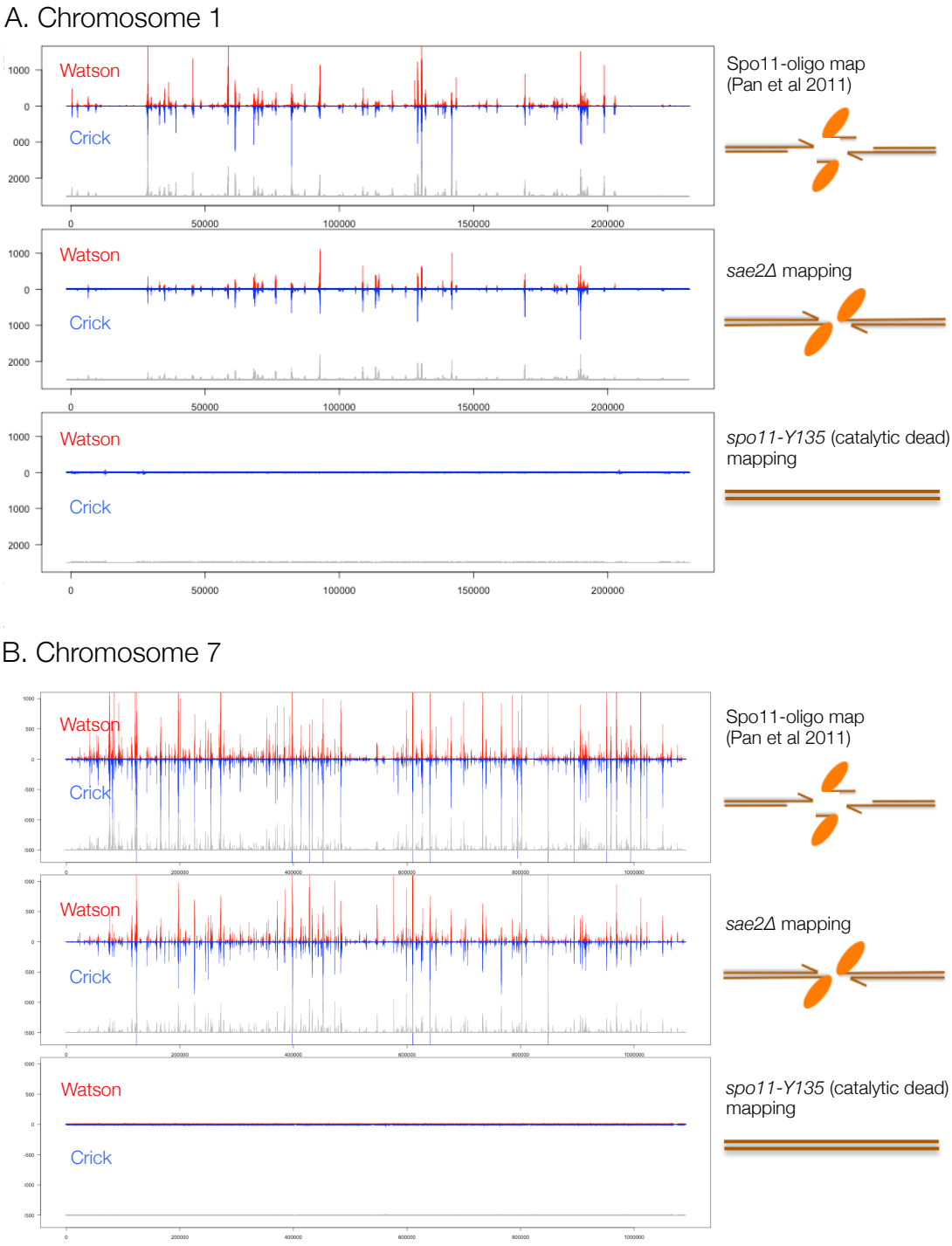


Figure 5.8: New *sae2*Δ genome-wide Spo11-DSB mapping reveals hotspots are located at similar positions to that characterised in Spo11-oligo map.
Spo11-DSBs were mapped in a *sae2*Δ background and compared to Spo11-oligo maps generated by Pan et al., 2011. Sequencing reads generated from a *sae2*Δ strains enriched for Spo11-DSBs were aligned against the *S. cerevisiae* S288C genome with 5′ Spo11 ends used to generate a 1 bp histogram. Chromosome 1 (**A**) and Chromosome 7 (**B**) were plotted with hits on the Watson strand (red), Crick strand (blue) and total hits (grey) indicated. Background reads were calculated for all strains and used to normalise signal intensity ('Spo11 Mapping v09 MC.R' – Appendix 10.5).

5.13 Spo11-DSBs are enriched at promoter regions and a 2 bp Watson/Crick offset is observed at the majority of sites

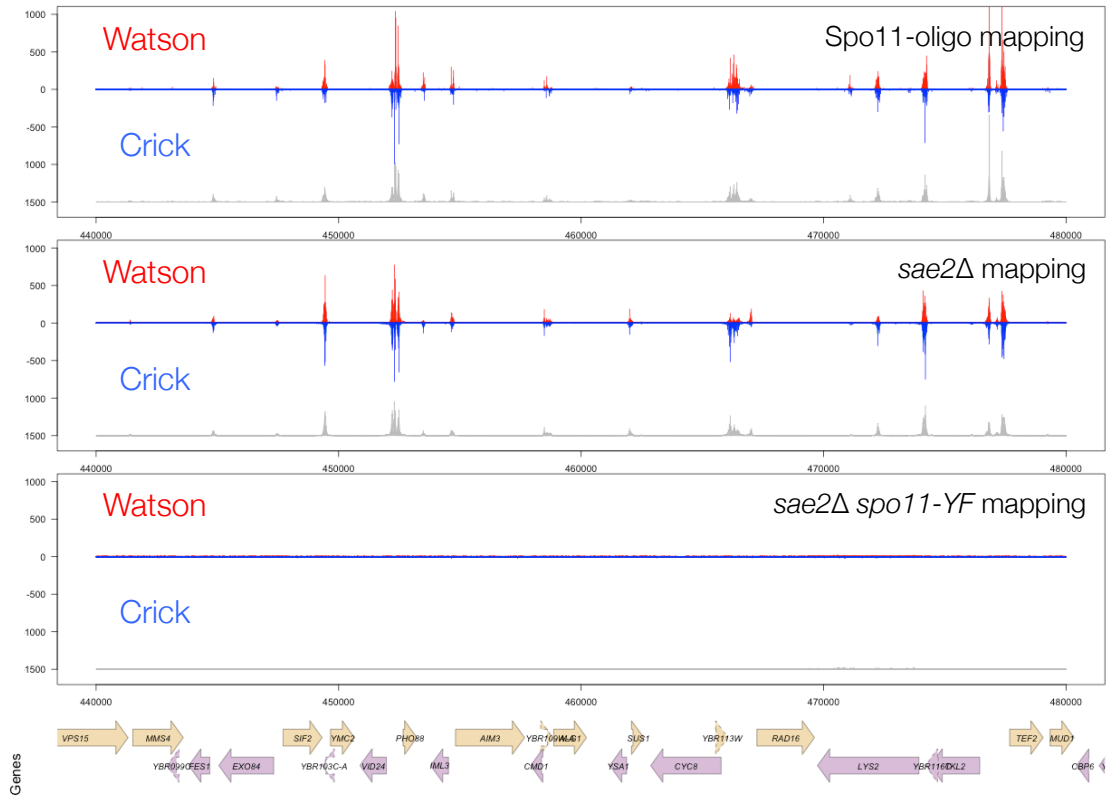
The Pan et al., 2011 data set, and studies prior to this (Baudat and Nicolas, 1997), observed that the majority of Spo11-DSBs form in intergenic regions containing promoters. To determine whether Spo11-DSBs from the *sae2*Δ DSB-enrichment procedure agree with these studies, a narrow 40 kb region of the genome was visualised for Spo11-DSB hits (Figure 5.9A)('Spo11 mapping 09 MC' – Appendix 10.5). In agreement with previous studies, Spo11-DSBs from the *sae2*Δ dataset are observed enriched at promoter regions (Figure 5.9A). This corroborates the view that an open chromatin structure is necessary for Spo11 to access the DNA. Both promoter and terminator regions are nucleosome-depleted (Fan et al., 2010; Lantermann et al., 2010), therefore if an open chromatin structure is all that is required then promoter and terminator regions should both contain Spo11-DSBs. To further determine whether the presence of a promoter, or just an open chromatin structure, increases the enrichment of Spo11-DSBs in an intergenic region, Spo11-DSBs were subset into intergenic region type (tandem (two promoters), convergent (two terminators), and divergent (two promoters) – as visualised in Figure 5.9B) and numbers aggregated (Figure 5.9B). Spo11-DSBs are enriched at tandem and divergent intergenic regions – regions that contain promoters, with divergent regions having increased Spo11-DSB hits compared to tandem regions. At convergent regions, Spo11-DSB levels are very low, indicating that Spo11-DSBs rarely form in terminator regions, even if these regions are nucleosome-depleted. At a closer high-resolution scale of visualisation, a 2 bp offset between the Watson and Crick strands is observed (Figure 5.9C), which correlates with the expected 2 bp overhang the Spo11 dimer generates when creating a DSB (Liu et al., 1995).

The sites where Spo11 generates DSBs in the genome have been previously mapped by sequencing the Spo11-oligonucleotide products formed from the initial processing of the Spo11-DSB end by the MRX complex and Sae2 (Pan et al., 2011). The Spo11-oligo mapping data identified and annotated DSB hotspots, areas of the genome where Spo11-oligos/DSBs clustered (Pan et al., 2011). 3599 hotspots were called from the Spo11-oligo data. Of these 3599 hotspots, the new *sae2*Δ mapping detects 3576 hotspots (99.4%) that are >2-fold over background and 3384 (94%) >5-fold over background, suggesting that these two methods correlate well.

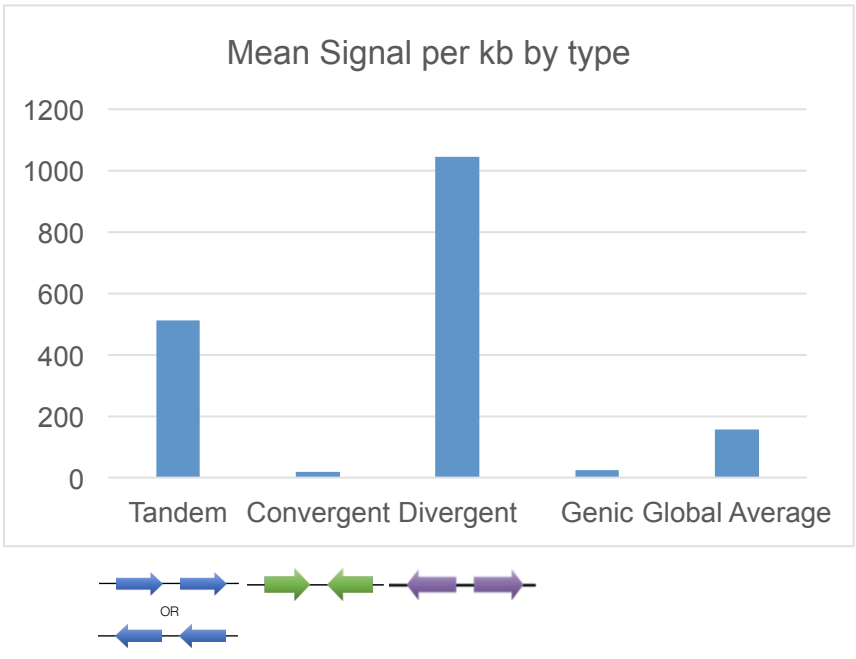
5.14 Spo11-DSBs mapped via the new *sae2Δ* method enrich at known hotspots

Historically, directly observing Spo11-DSBs at specific places in the genome is conducted via Southern blotting or pulse-field gel electrophoresis. These assays are limited to studying one to

A. 40 kb window



B. Spo11-DSBs subset for intergenic region type



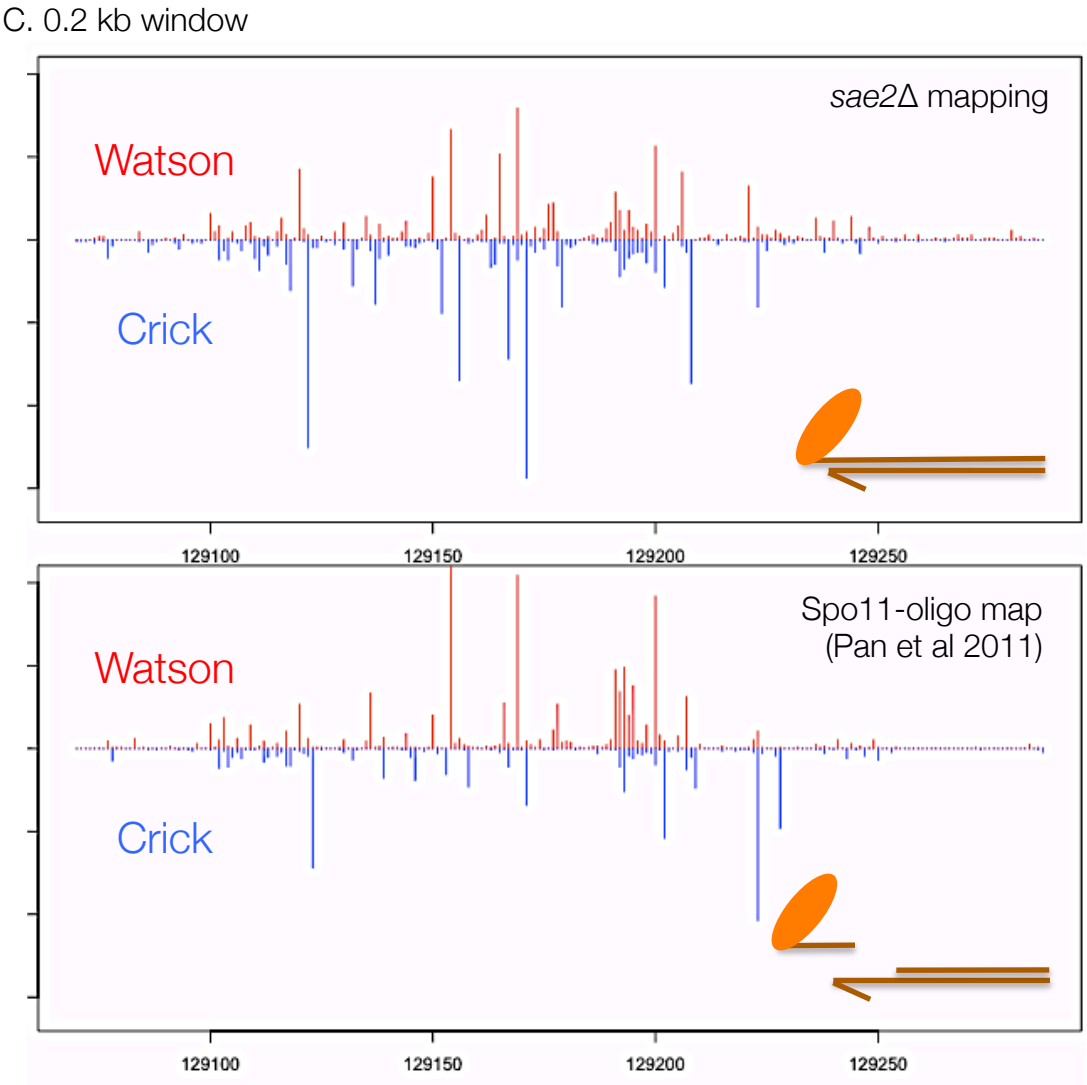


Figure 5.9: Spo11-DSBs are enriched at promoter regions and create a 2 bp 5′ overhang.
Spo11-DSBs were mapped in a *sae2*Δ background and compared to Spo11-oligo maps generated by Pan et al., 2011. Sequencing reads generated from a *sae2*Δ strains enriched for Spo11-DSBs were aligned against the *S. cerevisiae* S288C genome with 5′ Spo11 ends used to generate a 1 bp histogram. **A.** A 40 kb interval on chromosome 2 was visualised for Spo11-DSB hits. **B.** Spo11-DSB hits were subset into intergenic region type and aggregated. **C.** A 200 bp interval was plotted from Chromosome 1 with each base pair represented as a single line with Watson strand (red) hits, Crick strand (blue) and total (grey) hits indicated. Estimated background reads were calculated for all strains and used to normalise signal intensity ('Spo11 Mapping v09 MC.R' – Appendix 10.5).

a few hotspots at any one time due to the technical limitations. Utilising the new *sae2*Δ mapping procedure, once the strain has been sequenced, any given site of the genome can be visualised for Spo11-DSBs simultaneously. Because the procedure takes a similar time frame to conducting a Southern blot this method is beneficial for studying Spo11 hotspots genome-wide. Two hotspots (*ERG25* and *ARE1*) were chosen for visualisation (Figure 5.10), with the strong *HIS4::LEU2* hotspot not chosen due to the strains used for Spo11-oligo mapping not containing this strong artificial hotspot (Pan et al., 2011) ('Spo11 mapping v09 MC' – Appendix 10.5). These two hotspots are both enriched for Spo11-DSBs in both the wild type (Pan et al., 2011) data and the new *sae2*Δ mapping method. The peaks are similar between the Spo11-oligo data and the *sae2*Δ data, however differences can be noted with more defined peaks observed in the Spo11-oligo data, whereas the *sae2*Δ contains a more broad range of signal throughout the hotspot. This may be due to a *sae2*Δ-specific phenotype or due to the Spo11-oligo data only mapping the long oligonucleotide produced, thus losing signal that may have been contained within these hotspots. In both studies, signal is restricted mainly within the promoter region, as previously characterised (Pan et al., 2011), and described (Figure 5.9).

5.15 The measured strength of each hotspot is reproducible across multiple *sae2Δ* repeats and correlates with the Spo11-oligo hotspot data

The *sae2*Δ strain was mapped multiple times from biological duplicates (originating from separate colonies/cultures). To determine how reproducible these repeats were the strength of each annotated hotspot was compared. This comparison can also be made between the *sae2*Δ data sets and the Spo11-oligo data set. Hits per million reads, that have been normalised to the level of background observed (HpMNorm), were calculated for each hotspot ('Hotspot totals v09 MC.R' - Appendix 10.6). The fraction of hits per hotspot out of the total number of reads from all the hotspots (after background subtraction and filtering) were multiplied by 1 million (to generate an easy to use number scale) and plotted against each other (Figure 5.11) ('Plotting hotspot tables v03.R' – Appendix 10.7).

Comparison of the biological replicates of the *sae2*Δ maps generated an r^2 value between 0.93 and 0.99 indicating that the *sae2*Δ repeats are highly reproducible (Figure 5.11A). The data points between *sae2*Δ repeats correlate even at low hit strength hotspots (HpMNorm), producing signal above background level at these weak sites, indicating a high dynamic range. Therefore, these *sae2*Δ repeats were combined and averaged for comparison against the Spo11-oligo data (Figure 5.11B). When comparing an averaged dataset for five *sae2*Δ strains,

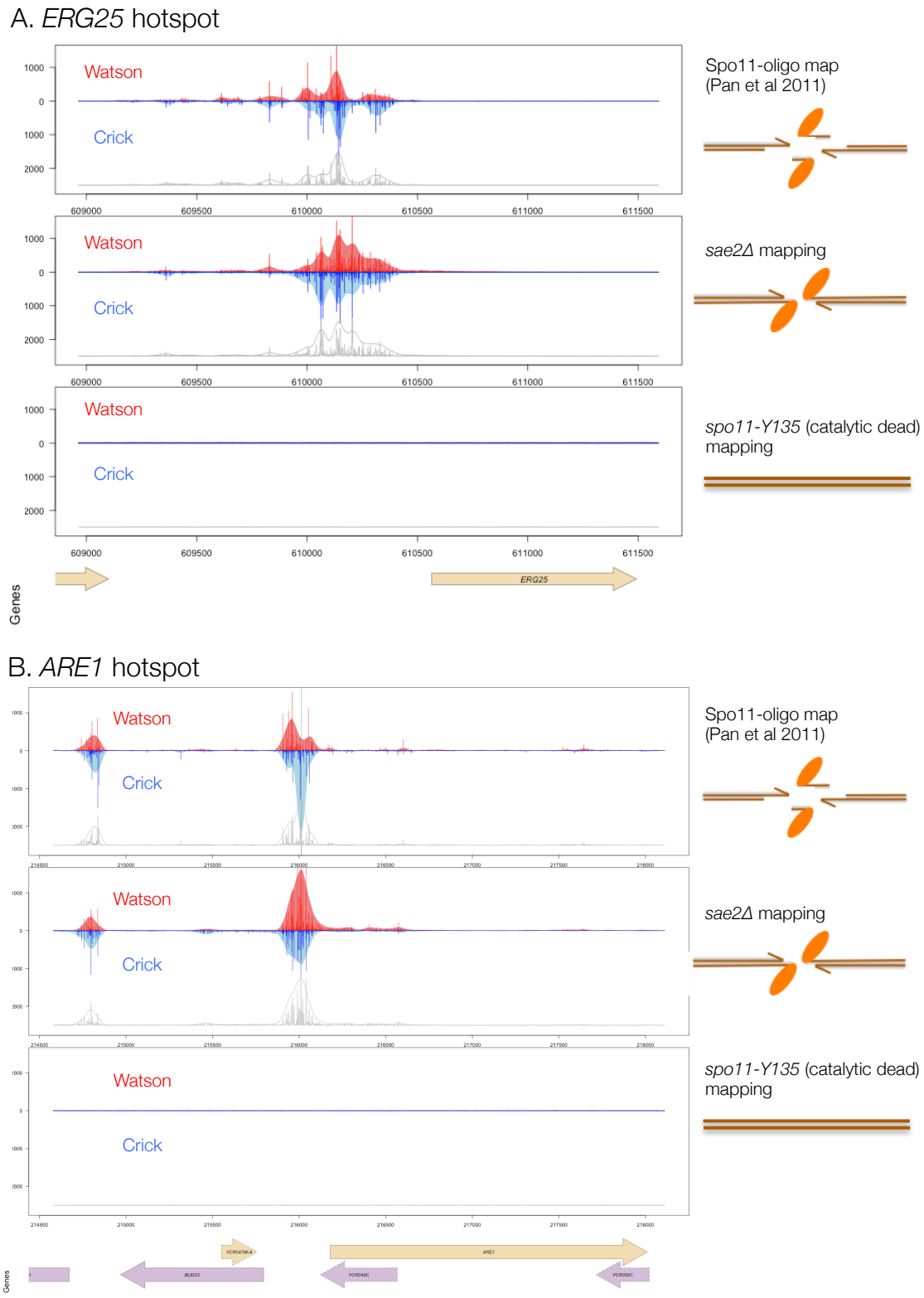


Figure 5.10: Spo11-DSB hits from *sae2*Δ mapping reside within promoter regions at known hotspots.
Spo11-DSBs were mapped in a *sae2*Δ background and compared to Spo11-oligo maps generated by Pan et al., 2011. Sequencing reads generated from a *sae2*Δ strains enriched for Spo11-DSBs were aligned against the *S. cerevisiae* S288C genome with 5′ Spo11 ends used to generate a 1 bp histogram. Hits on the Watson strand (red), Crick strand (blue) and total hits (grey) are indicated. Background reads were calculated for all strains and used to normalised signal intensity ('Spo11 Mapping v09 MC.R' – Appendix 10.5). **A.** *ERG25* hotspot. **B.** *ARE1* hotspot.

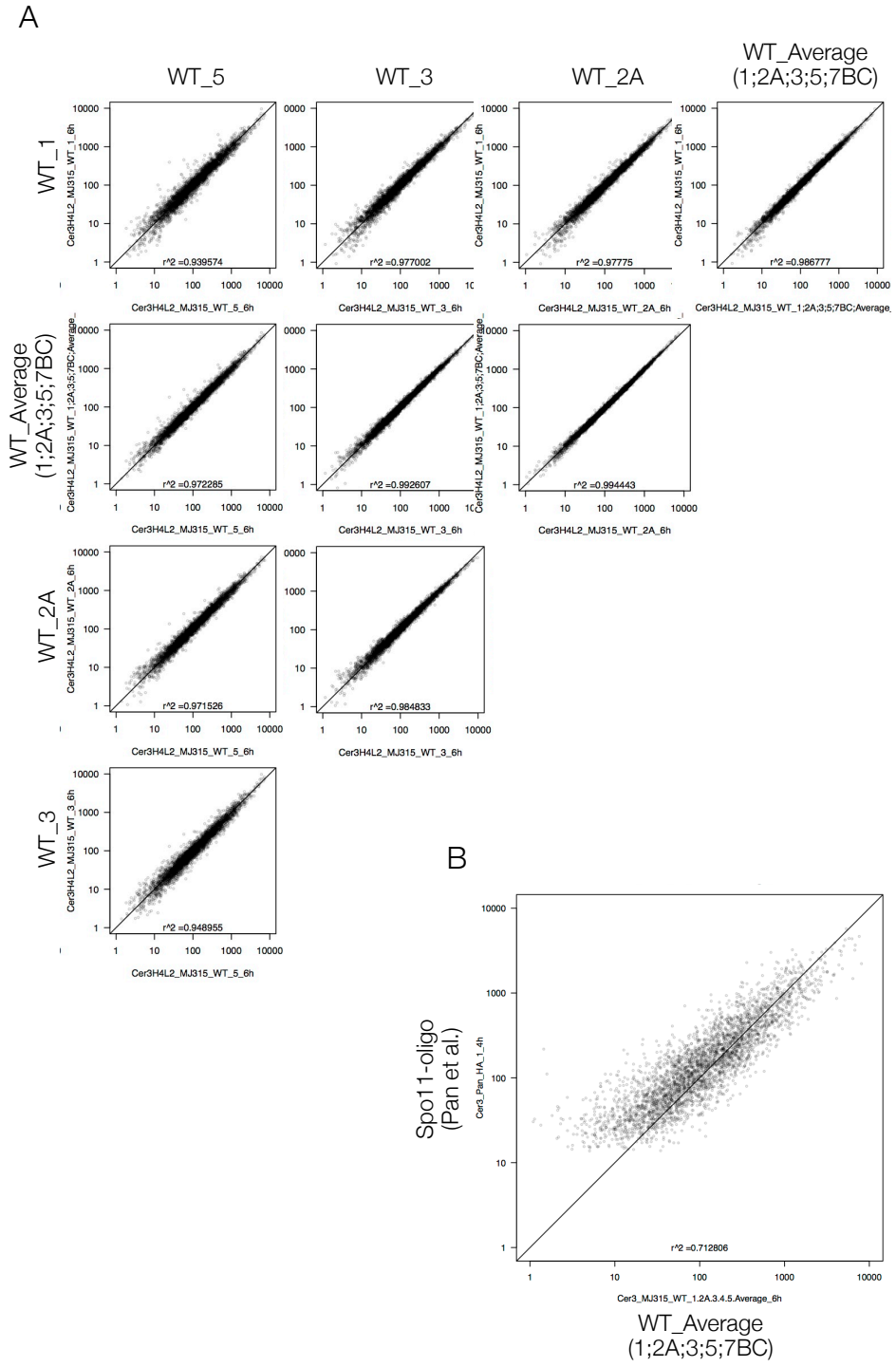


Figure 5.11: Multiple repeats of the wild type (*sae2*Δ) strain have highly correlated hotspot strengths. Correlation between wild type (*sae2*Δ) and the Spo11-oligo dataset is also observed. A. Quantitative reproducibility. Comparisons are shown for pairs of biological replicates (WT_1, WT_2A, WT_3, WT_5) and against an averaged data set (WT_Average). Total Spo11-DSB hits were summed up for all annotated hotspots from Pan et al., 2011 and the hotspot strengths (hits per million mapped reads (HpM)) compared between data sets (**A.**) and between an averaged sae*2*Δ data set and the Spo11-oligo data set (**B.**). Pearson's r^2 for pairwise comparisons ranged from 0.939-0.994 for all biological replicates and against the averaged data set. Pearson's r^2 for pairwise comparison between the averaged *sae2*Δ and the Spo11-oligo data set was calculated as 0.713. Script ('Plotting hotspot tables v03.R' – Appendix 10.7).

against the Spo11-oligo (Pan) data, there is strong correlation (r^2 value ~0.68). However, the correlation between *sae2*Δ data was much stronger compared to between the *sae2*Δ and the Spo11-oligo dataset, suggesting that there is a real biological difference between wild type and *sae2*Δ strains. This can be observed at medium strength hotspots (100-1000 HpMNorm) that are 10-fold greater in the Spo11-oligo data. There is however some major differences in the comparison between *sae2*Δ and the Spo11-oligo data at weak hotspots with a skew occurring at this lower end (Figure 5.11B). For the calling of hotspots in the Spo11-oligo dataset, a cut-off was used of 25 total mapped Spo11-oligos (Pan et al., 2011). Therefore, no hotspots were called below this threshold. By contrast, the *sae2*Δ dataset had no cut-off applied, apart from the correction for background levels. This may cause the skew of the data at the weak hotspots because at these designated sites the *sae2*Δ data could record less than the 25 hit cut-off, whereas all the Spo11-oligo hotspots at this weaker end would be 25 hits and above.

5.16 Chromosome size-correlated variation in DSB frequencies

The Spo11-oligo data (Pan et al., 2011), and older mapping studies (Blitzblau et al., 2007; Gerton et al., 2000; Martini et al., 2006), observed a negative correlation between chromosome size and hit density, potentially due to chromosome synapsis-dependent, Zip3-mediated shut-off of DSB formation (Thacker et al., 2014). For chromosomes to synapse DSB resection, ssDNA-dependent homology search and strand invasion is essential. Larger chromosomes have an increased chance for homolog engagement due to more DSBs forming on these chromosomes (Thacker et al., 2014). In a DSB processing mutant, such as *sae2*Δ, DSB resection, and therefore synapsis-dependent shut-off of DSB formation, does not occur. First, to determine whether Spo11-DSBs numbers increase with chromosome length in the *sae2*Δ data set, as seen in the Spo11-oligo data, chromosome length was plotted against total chromosome hits per million (HpM) (Figure 5.12A). As in the Spo11-oligo data, a positive linear correlation ($r^2 = 0.88$) is observed between chromosome length and numbers of Spo11-DSB hits. For unknown reasons, chromosome 12, is an outlier in this data, forming fewer DSBs relative to its size.

To determine whether homolog synapsis in a wild type background does cause larger chromosomes to contain a lower DSB density, the hit density of the *sae2*Δ data set was calculated (total Spo11 hits per million per bp) and plotted against chromosome length (Figure 5.12B). The *sae2*Δ data set also contains a negative correlation between hit density and chromosome length however, the r^2 value was much lower than the wild type data ($r^2 = 0.061$ in *sae2*Δ compared to $r^2 = 0.43$ in the Spo11-oligo data). This suggests that larger

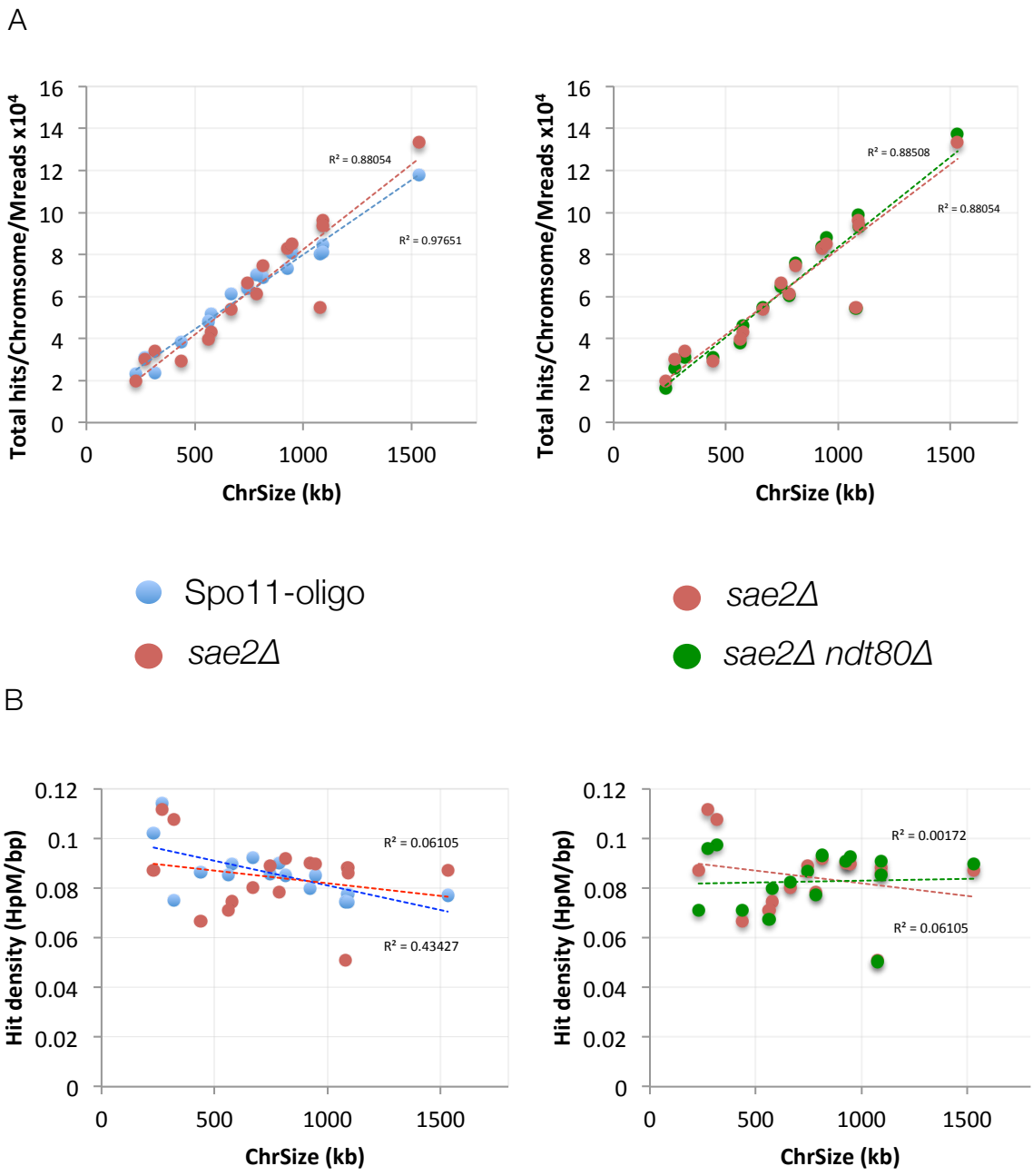


Figure 5.12: Spo11 DSBs increase with chromosome length but larger chromosomes contain a lower density of Spo11-DSBs, potentially due to synaptic shut-off.
A. Total chromosome hits per million reads (MReads) was calculated and plotted against the chromosome size (kb). The correlation between total chromosomal Spo11 hits per million reads and chromosome length was calculated and displayed on the graph (r^2). **B.** The Spo11 hit density (total chromosome hits per million reads per base pair) was calculated and plotted against chromosome size (kb). The correlation between hit density and chromosome size was calculated and displayed on the graph (r^2).

chromosomes in *sae2*Δ contain Spo11-DSBs more similar to smaller chromosomes, supporting the hypothesis that DSB formation is repressed by synapsis-dependent shut-off. In the *sae2*Δ data set small chromosomes still have a higher hit density to larger chromosomes, therefore, another potential factor is active that is suppressing DSB formation on larger chromosomes. One hypothesis is that *sae2*Δ cells have a shorter prophase due to a weakened prophase checkpoint response. This shortened prophase would decrease the time for larger chromosomes to form DSBs. Therefore, to determine whether prevention of prophase exit, through an *ndt80*Δ mutation, has an effect on DSB hit density, the *sae2*Δ hit density was compared to *sae2*Δ *ndt80*Δ (Figure 5.12B – right hand panel). The hit density of the *sae2*Δ *ndt80*Δ mutant was similar across all chromosome sizes ($r^2 = 0.0017$), suggesting that length of prophase has an effect on DSB formation on larger chromosomes in a *sae2*Δ background. In order to distinguish whether DSB formation suppression on larger chromosomes is synapsis-dependent or due to prophase timing, the Spo11-oligo mapping procedure would need to be conducted on wild type cells in an *ndt80*Δ background.

Potentially the shortened length of prophase in a *sae2*Δ mutant could cause chromosome 12 to display fewer Spo11 hits than expected for its size. To test the hypothesis, the total number of hits on chromosome 12 was calculated for a *sae2*Δ *ndt80*Δ mutant (Figure 5.12A – right hand size). However, preventing exit from prophase did not affect the number of hits on chromosome 12, suggesting that this factor is not the cause for chromosome 12 being an outlier.

5.17 Spo11 contains a weak sequence bias for cleavage

The Spo11-oligo mapping study determined that there is a weak sequence bias for Spo11 cleavage (Pan et al., 2011). However, due to the ambiguity of 5′ C residues (due to the rGTP-tailing conducted to generate sequenceable molecules), and this procedure only mapping the long Spo11-oligo produced, an incomplete picture was formed. Using the nucleotide-resolution *sae2*Δ mapping data, a more accurate sequence bias can be obtained that contains no ambiguity for the 5′ end, with both sides of the DSB enriched for (assuming the shorter Spo11-oligo that was unmapped arises from asymmetric processing of the Spo11-DSB) (Figure 5.13). To determine whether Spo11 displays a sequence bias for cleavage, the 5′ (Spo11) ends +-35 bp were aggregated, with any single 5′ site that occurred greater than five times (to remove any potential background effect) used to generate a sequence bias profile (Figure 5.13). This consensus sequence for Spo11 cleavage at DSB sites consists of biases of A at positions +10, G at positions +2 and +13, and T at positions +1, +4, +5 and +11. These biases are palindromic and therefore contain the complementary purine/pyrimidine on the opposing strand (negative

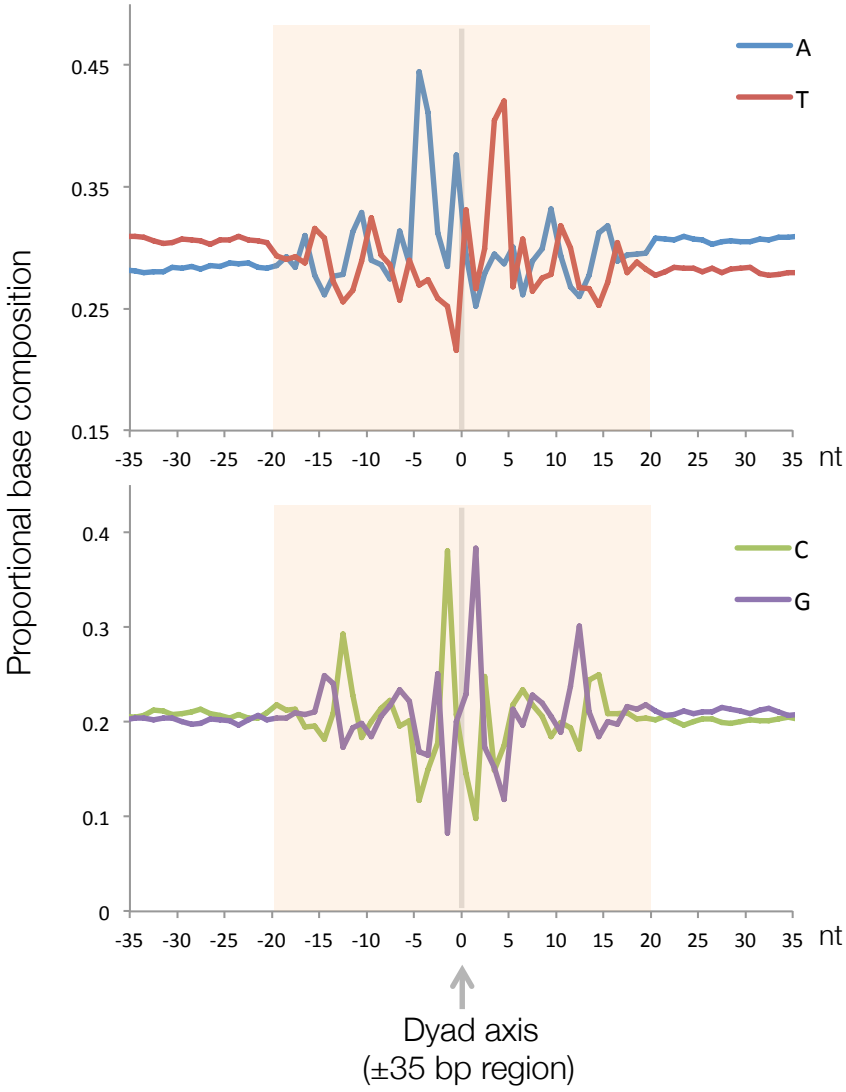


Figure 5.13: Spo11 cleavage sites genome-wide contain a sequence bias.
Sequences derived from the *sae2*Δ, Spo11 enrichment protocol were aligned to the *S. cerevisiae* S288C genome using bowtie2. 5′ Read1 (Spo11) ends were extracted and converted into a 1 bp histogram using R Studio. The sequence of each 5′ read +/- 35 bp were aggregated for sites where Spo11-DSBs were at least 5 to reduce the influence of non-specific signal. The sequence bias was plotted for A/T and G/C pairs, which are expected to show rotational symmetry consistent with Spo11 cleaving DNA as a dimer.

values). Notably, some positions were depleted below the average occupancy: A at +1, C at +2, and G at +5. Overall, a weak (palindromic) consensus for Spo11-DSBs can be read as AAGC*A|TGCTT (top-strand displayed only, dyad axis (|) indicated, Spo11 cleavage site (*) would be reciprocated on the right hand side of the axis, bottom strand). The palindromic nature and rotational symmetry of the bias is indicative of a Spo11 dimer cleaving the dsDNA, comparable with previous reports (Sasanuma et al., 2007).

5.18 Discussion

The mapping procedure detailed in this chapter is able to generate a single-nucleotide resolution genome-wide map of Spo11-DSBs in a *sae2*Δ (Spo11-end processing deficient) background, improving on the previous low-resolution *rad50S* mapping data (Gerton et al., 2000). In comparison to the Spo11-oligo mapping (Pan et al., 2011), this high resolution map also observed Spo11-DSB accumulation at the majority of the previously annotated hotspots. These hotspots reside in nucleosome-depleted regions at promoters (Figure 5.9-5.11). As expected from the Spo11-oligo data (Pan et al., 2011), the number of DSBs per chromosome increase with the length of chromosome (Figure 5.12A). The Spo11-oligo data set observed a slight negative correlation of DSB density with chromosome length attributed to synapsis-dependent shut-off of DSB formation, which, due to increased numbers of DSBs on larger chromosomes would decrease the DSB density the larger a chromosome is (Thacker et al., 2014). DSB resection-defective mutants (e.g. *sae2*Δ) do not permit homolog synapsis and the *sae2*Δ data set observes a more even DSB density between short and long chromosomes (Figure 5.12B). However, increasing the length of prophase using an *ndt80*Δ mutant resulted in short and long chromosomes containing almost identical DSB density, suggesting that synapsis-dependent regulation of DSBs on large chromosomes may not be the only factor involved in a *sae2*Δ background, and that length in prophase is also important.

A weak Spo11 sequence bias was previously observed, however an ambiguity for 5′ C molecules due to technical limitations skewed the data slightly (Pan et al., 2011). Using the *sae2*Δ data set that contains no ambiguity for the 5′ end of molecules, a more accurate consensus sequence could be generated (Figure 5.13). Potentially, the bias arises due to DNA properties that affect Spo11 binding or cleavage of the DNA. The enrichment or depletion of each base reflects that observed previously (Pan et al., 2011) and thus supports the hypothesis that whilst Spo11 is directed to specific regions of the genome, for example at nucleosome-depleted regions, it also is slightly directed to specific regions due to base composition.

This *sae2*Δ, Spo11-DSB mapping procedure involved the enrichment of Spo11-DNA molecules by utilising the covalently bound protein moiety as a substrate for specific

enrichment on glass fibre columns. This enables genome-wide DSB maps of untagged Spo11, avoiding any potential effects affinity tags have on Spo11 function. This is in contrast to the Spo11-oligo mapping data, which initially relied on HA-tagged Spo11 for enrichment (Pan et al., 2011). A caveat to this new method of enrichment is that any protein-bound molecules could potentially be enriched and thus, the signal observed in the *sae2*Δ maps could actually be due to another DNA-bound protein. However, sequencing of a catalytic dead Spo11 mutant (*spo11-Y135F)* mutant indicates that the signal observed is Spo11-DSB dependent because no specific signal was observed in this strain (Figure 5.8-5.11).

In principle, this procedure is not *S. cerevisiae* specific because all that is required is the presence of the protein covalently bound to the 5′ end of the DNA. The lack of requirement for an affinity tag has the advantage of removing any potential effect affinity tags have on protein activity and also aids in mapping in other organisms where incorporating a tagged form of a protein is technically difficult. Additionally, TDP2 shows no specificity to proteins (Chapter 4), rather it is capable of cleaving any 5′ phosphotyrosine covalent bond. Therefore, there is the potential to utilise this procedure to map Top2, which, like Spo11, generates a 5′ phosphotyrosyl bond at DSBs. In situations where the covalently bound protein is 3′ linked, such as Top1, other hydrolytic proteins, such as Tdp1, which cleaves 3′ phosphotyrosyl bonds, could be utilised to map these proteins. Thus, the protocol developed in this chapter has the potential to be used to create high-resolution, genome-wide maps of a variety of covalently-attached proteins in a range of organisms.

CHAPTER 6:

THE ROLE OF TEL1 KINASE ACTIVITY ON DSB FORMATION

Chapter 6: The role of Tel1 kinase activity on DSB formation

6.1 Introduction

Spo11-DSBs are essential for initiation of meiotic recombination (Szostak et al., 1983), however, they are still lesions that have the potential to be detrimental to genomic stability if not repaired correctly. In *S. cerevisiae* there are ~3600 hotspots, regions that are enriched for Spo11-DSBs. However, within a single cell only 150-200 DSBs form per meiosis. Therefore, meiotic cells must have evolved pathways that limit the number and distribution of DSBs (reviewed in Cooper et al., 2014; Keeney et al., 2014).

Two proteins have been implicated in the negative regulation of DSB formation in meiosis; the serine/threonine kinases Tel1^{ATM} and Mec1^{ATR}. Both proteins are involved in *trans* inhibition, a process that prevents DSBs from forming on the same location on sister chromatids and homologous chromosomes once a DSB has formed at the corresponding locus (Fukuda et al., 2008; Rocco and Nicolas, 1996; Xu and Kleckner, 1995; Zhang et al., 2011). DSB events along chromatids also exhibit non-random patterns via a process known as DSB interference (Garcia et al., 2015). Tel1 controls DSB interference, with coincident formation of DSBs adjacent to each other occurring less often than by chance. In meiosis, sister chromatids condense by being organised into protruding loops, attached to a proteinaceous axis. DSB hotspots predominantly occur within these loops, however many of the proteins responsible for the activation and regulation of Spo11-induced DSBs reside at the axis (Blat et al., 2002; Borde and de Massy, 2013; Kleckner, 2006). The tethered-loop axis model suggests that DSBs forming primarily on loops are tethered to the axis to initiate the DSB (Sommermeyer et al., 2013). In *tel1*Δ strains, a loss of interference occurs whereby DSBs form independently of one another but also now form more closely spaced than by chance (Garcia et al., 2015). This concerted DSB activity is restricted to ~15 kb windows that correlate with the location of singular meiotic loop domains. The data suggest that Tel1 is responsible for limiting a loop to contain a single DSB whilst also suppressing adjacent sites in the surrounding regions (Garcia et al., 2015). However, it is unknown whether the kinase activity of Tel1 is responsible for this action and if it is, what the relevant target(s) is.

Mutation to the kinase domain of Tel1 results in a loss of *in vitro* kinase activity and *in vivo* phenotypes associated with *tel1*Δ (Mallory and Petes, 2000). However, a more recent study identified Tel1-kinase independent functions in telomere maintenance, suggesting that Tel1 has two separate functions and that the Xrs2-dependent recruitment and association of Tel1 to telomeres plays an important role, independent of its kinase activity (Ma and Greider, 2009).

Therefore it was of interest to determine whether Tel1 kinase activity is required for DSB interference.

6.2 In a *sae2Δ* background the *tel1-kinase dead* mutant causes DSB signal to smear on Southern blots

Loss of *TEL1* causes loss of DSB interference (Garcia et al., 2015). To determine whether the kinase activity of Tel1 was responsible for meiotic DSB interference, the *tel1-kinase dead* mutant (*D2612A-N2617A-D2631A*) (Ma and Greider, 2009) (*tel1kd*) was generated in the SK1 background (V.Garcia). Genomic DNA extracted from meiotic cultures were assayed for DSB formation via Southern blotting. These experiments were conducted in a *sae2*Δ background to prevent resection and inhibit DSB repair, which otherwise causes heterogeneous migration of the DSBs and makes quantification challenging (Figure 6.1). At the *HIS4::LEU2* hotspot, two DSBs form (Xu and Kleckner, 1995) and, in a *TEL1+, sae2*Δ background, migrate as two tight bands (Figure 6.1A). In *sae2*Δ *tel1*Δ strains, migration of these two bands does not differ from *sae2*Δ (Figure 6.1B). However, in *sae2*Δ *tel1kd* the two DSBs form but now two sub-populations are visualised for each DSB site. The two tight DSBs bands are present alongside a heterogeneous, faster migrating, smear down the gel (initially observed by V.Garcia – unpublished work, replicated in Figure 6.1C), similar to that seen at resected DSBs.

In a *sae2*Δ background, unresected DSBs accumulate with Spo11 covalently bound to the 5′ ends, preventing any resection/repair from occurring (de Massy et al., 1995; Keeney et al., 1997; Keeney and Kleckner, 1995; Liu et al., 1995). Due to the resection-like smear observed (Figure 6.1C), it was hypothesised that the *tel1kd* mutant was somehow bypassing the Spo11-block to resection. However, subsequent experiments by Valerie Garcia have confirmed these smears are resistant to S1 nuclease, indicating the smear is not due to ssDNA being present (i.e. from resection). Additionally, the Spo11-DSB enrichment procedure (detailed in Figure 5.3) is able to confirm that the smear is covalently bound Spo11-DNA (V.Garcia – data not shown). Both these results indicate that the smear is not due to resection of the DSB. The smear increases in length over time and intriguingly only extends in the direction of transcription at *HIS4::LEU2* (V.Garcia – unpublished observation and Figure 6.1C).

6.3 Genome-wide, nucleotide-resolution mapping supports the spreading of DSBs in the direction of transcription in the *tel1kd* mutant

In order to determine whether the *tel1kd* mutant causes DSB spreading at all DSB hotspots, the *sae2*Δ Spo11 mapping protocol (developed in Chapter 5) was utilised. As before,

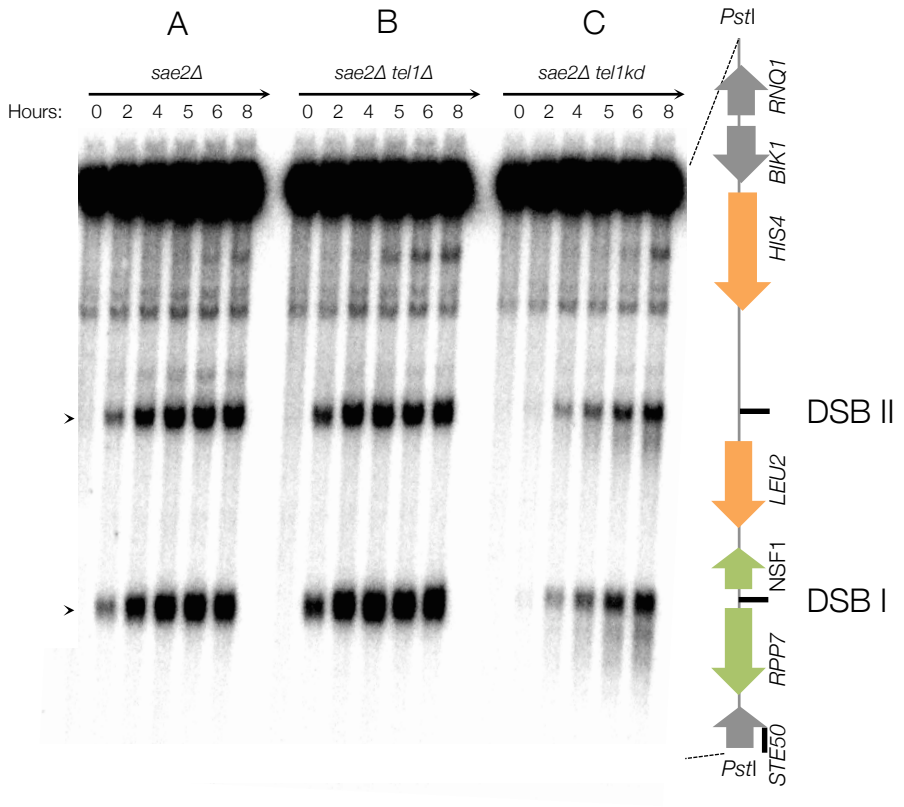


Figure 6.1: At the *HIS4::LEU2* hotspot the *tel1kd* mutant causes a heterogeneous smear of DSB signal in the direction of transcription.
Southern blot of DSBs at the *HIS4::LEU2* hotspot. A meiotic time course was performed for *sae2*Δ (**A**), *tel1*Δ (**B**) and *te1kd* (**C**), and samples taken at the indicated time points after the start of meiosis. Proteolysed genomic DNA was extracted from each time point and digested with *Pst*I overnight before being resolved on a 0.7% TAE agarose gel for 18 hours at 60 V. The gel was transferred to nylon membrane under denaturing conditions and hybridised with a radioactive probe for the *HIS4::LEU2* locus. The membrane was exposed to a phosphor screen overnight and an image taken using a Fuji phosphor scanner. DSBs are marked with an arrowhead (>). Annotated gene array (right hand side) indicates positions and orientations of genes and the locations of the two DSBs at the *HIS4::LEU2* locus.

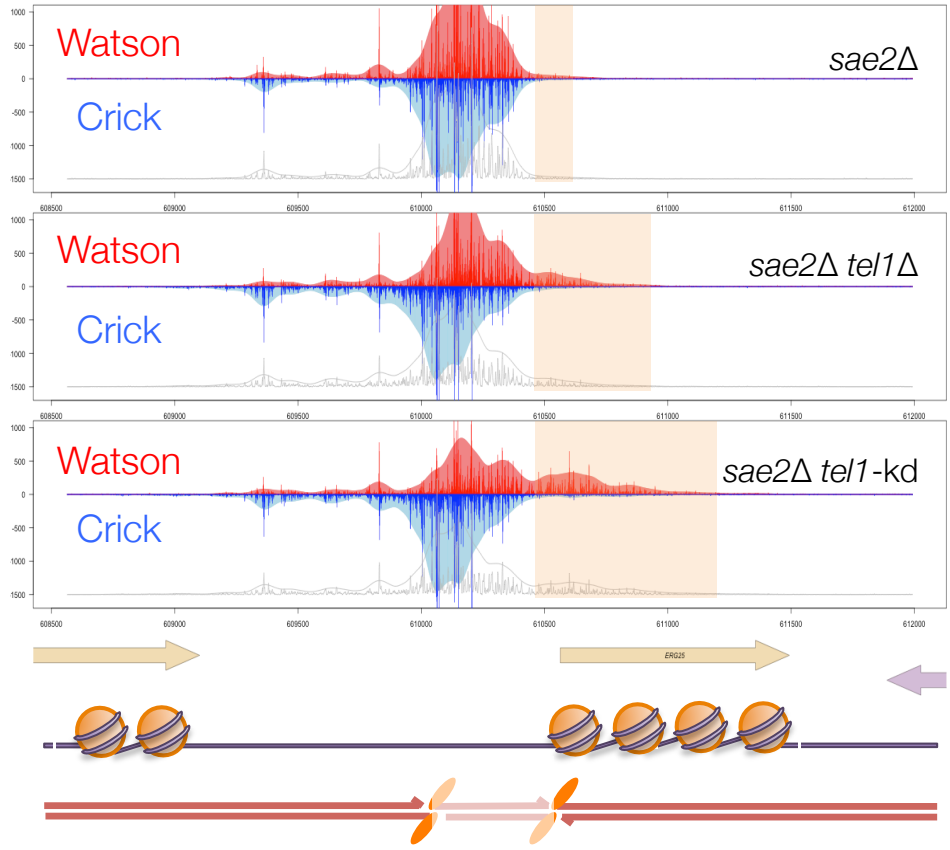
unproteolysed genomic DNA was extracted from meiotic cultures at mid-meiotic prophase from *sae2*Δ, *sae2*Δ *tel1*Δ and *sae2*Δ *tel1kd* strains. DNA was fragmented to ~500 bp using sonication and protein-associated molecules enriched on a glass fibre column. Free DNA molecules were removed using high salt washes and Spo11-DNA molecules eluted using SDS. Sonicated ends of molecules had Adaptor 1 (Read2) ligated on with free adaptor subsequently removed. Molecules were incubated with human TDP2 protein to hydrolytically remove Spo11 from the other 5′ end without any loss of nucleotides, which allowed specific ligation of Adaptor 2 (Read1) to the Spo11 bound end. Molecules were PCR amplified and sequences read on an Illumina MiSeq. Sequences were aligned to the *S. cerevisiae* S288C genome and the 5′ ends from Read1 (i.e. the 5′ base where Spo11 cut) sorted into a histogram for each specific base pair in the genome.

To determine whether the spreading observed via Southern blotting (Figure 6.1C) was visible in the mapping data, the specific sites of the genome (*HIS4::LEU2* and *ERG25*) were plotted (Figure 6.2) ('Spo11 mapping v09 MC.R' – Appendix 10.5). In *sae2*Δ, *sae2*Δ *tel1*Δ and *sae2*Δ *tel1kd*, the majority of the signal observed at these sites resides in the promoter regions of these genes. However, the *tel1kd* mutant also contains signal spreading from the promoter region into the 5′ end of the ORF in the direction of transcription at the *ERG25* locus (Figure 6.2A) and at both DSB sites of *HIS4::LEU2* (Figure 6.2B). The amount of spreading observed at *HIS4::LEU2* in the mapping data agrees with the Southern blots of the same locus, where the spreading at DSB-I is stronger than that observed at DSB-II (Figure 6.2B). At sites of the genome where hotspots aggregate in close proximity, such as *ARE1*, the *tel1-kd* mutant causes spreading of the hotspot DSB to overlap (Figure 6.2C). The spreading phenotype observed at these specific hotspots in the *tel1kd* mutant is also slightly present in in *tel1*Δ strains in comparison to wild type (*sae2*Δ) strains (Figure 6.2). By contrast, Southern blotting of the *tel1*Δ strain displays no indication of spreading at *ERG25* (data not shown) or *HIS4::*LEU2 (Figure 6.1B), suggesting the Southern blotting detection level is too low to observe this subtle spreading phenotype in *tel1*Δ.

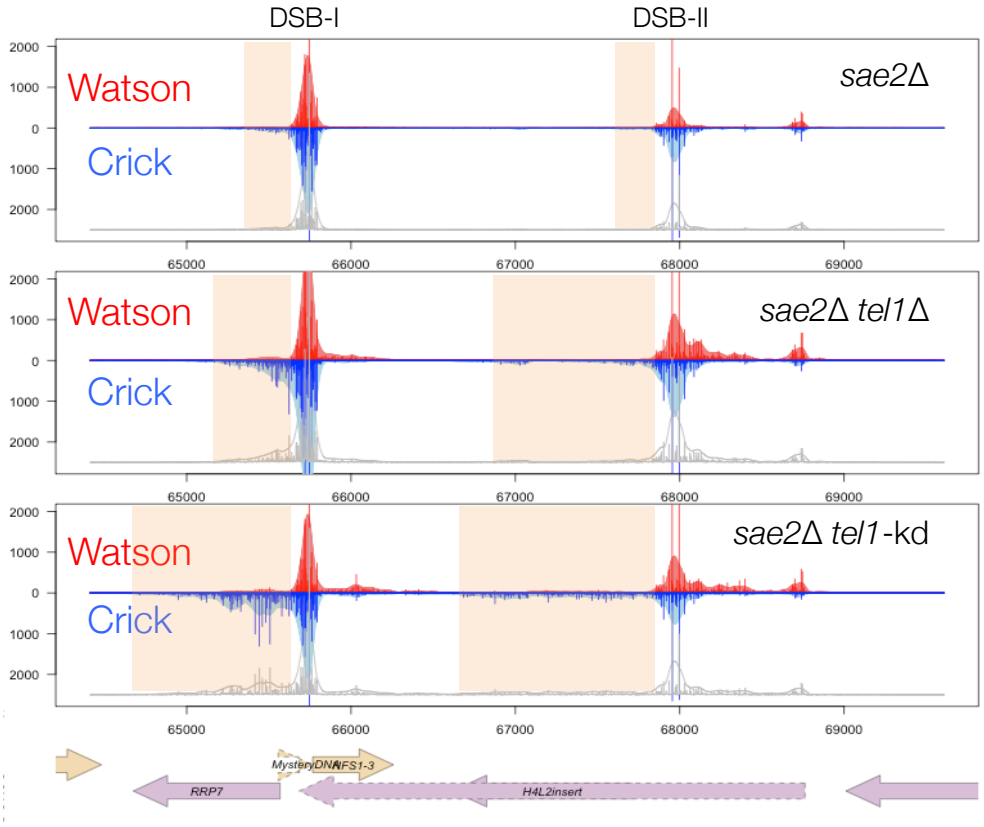
6.4 Spreading of DSB signal in *tel1kd* and *tel1Δ* mutants occurs genome-wide

This spreading phenomenon may be distinct to the *ERG25* and *HIS4::LEU2* loci. Therefore, to determine whether spreading of DSB signal into ORFs occurs across the genome, the promoter regions of all genes, centred on the translation start site (TSS), were piled-up for each strain ('Pileups around TSS v02.R' - Appendix 10.8). The orientations of genes located on the Crick strand were reversed to match the direction of the Watson genes for visualisation purposes

A. *ERG25* locus:



B. *HIS4::LEU2* locus:



C. *ARE1* locus:

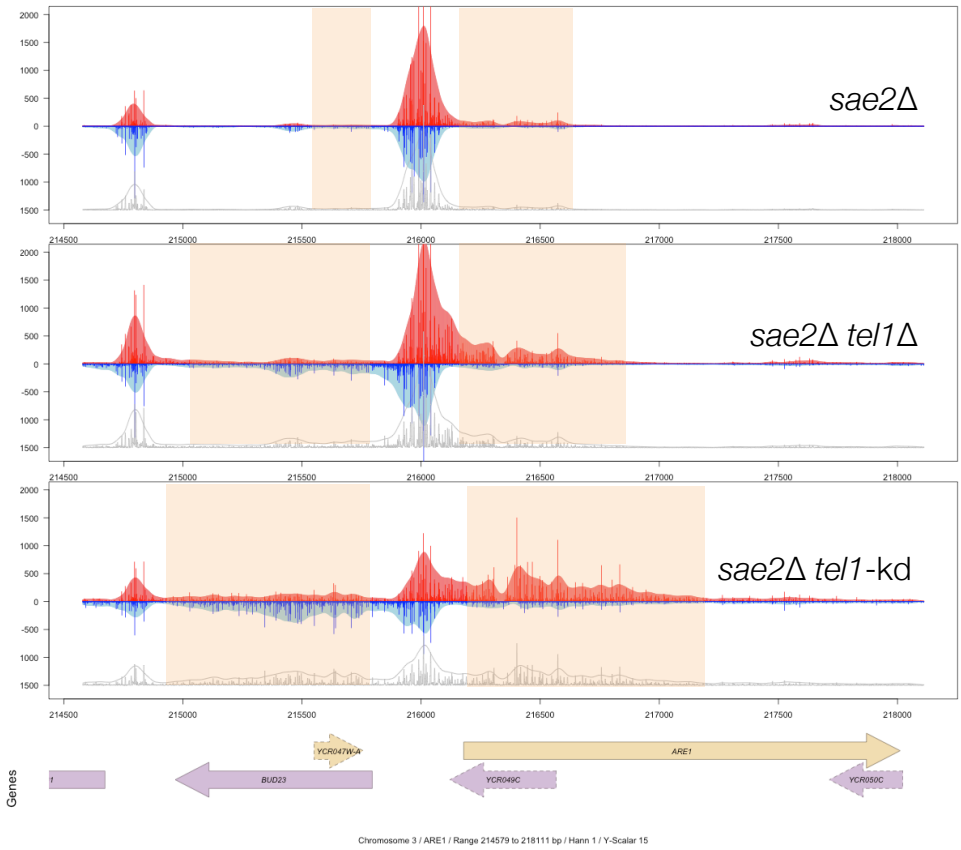


Figure 6.2: DSB signal spreads in the direction of transcription at various Spo11-DSB hotspots in *TEL1* mutants.
Unproteolysed genomic DNA was extracted from meiotic cultures at mid-meiotic prophase from *sae2*Δ, *sae2*Δ *tel1*Δ and *sae2*Δ *tel1kd* strains. DNA was fragmented using sonication and protein-associated molecules enriched on a glass fibre column. Samples were processed for next-generation sequencing as described in Chapter 4. Sequences were aligned to the *S. cerevisiae* S288C genome and the 5′ ends from Read1 (i.e. the 5′ base where Spo11 cut) sorted into a histogram for each specific base pair in the genome. Regions indicated were visualised in R Studio ('Spo11 Mapping v09 MC.R' – Appendix 10.5). Highlighted region indicates region of spread signal. **A.** *ERG25* locus. **B.** *HIS4::LEU2* locus. **C.** *ARE1* locus.

(Figure 6.3). As previously observed at *ERG25* and *HIS4::LEU2*, the majority of signal resides in the promoter region adjacent to the ORF. Spreading of Spo11-DSB signal into the ORF is observed slightly in all strains, however, the *tel1kd* mutant displays a vastly heightened signal spreading into the ORF from the promoter region for ~500 bp (Figure 6.3C), suggesting that this spreading phenotype is a genome-wide phenomena. The small amount of spreading seen at *ERG25* and *HIS4::LEU2* in the *tel1*Δ mutant (Figure 6.2B), compared to wild type (Figure 6.2A), is replicated in the genome-wide TSS pile-up data (Figure 6.3B). This suggests that the prevention of extensive DSB spreading in wild type cells is Tel1-kinase dependent, but yet independently, the Tel1 protein itself actually promotes DSB spreading via interaction with an unknown factor. Alternatively, the catalytically-dead Tel1 protein is obstructing a redundant pathway from compensating for the lack of Tel1 kinase activity.

6.5 Spo11-DSB sites in *tel1* mutants spread in the direction of the ORF

To further test the hypothesis that DSB signal spreads in the direction of the ORF, DSB hotspots were subset into tandem (one promoter, one terminator), divergent (two promoters) and convergent (two terminators) hotspot regions (as depicted in Figure 6.4), and Spo11-DSB signal piled-up for each individual hotspot type centred on the DSB hotspot ('Pileups around TSS v03 Gene expression' – Appendix 10.9). At tandem hotspots, the *tel1kd* mutant Spo11-DSB signal spreads only in the direction of the ORF (i.e. not backwards into the 3′ terminal end of the adjacent ORF) (Figure 6.4 - Tandem). At divergent hotspots, DSB signal spreads in both directions into both 5′ ends of the two adjacent ORFs (Figure 6.4 - Divergent). There are very few hotspots at converging genes, therefore the total hits found at these types is very low. For hotspots that are contained at convergent gene sites, spreading is minimal (Figure 6.4 – Convergent). These results suggest that a factor associated with the start of ORFs at promoter regions is a target for the Tel1 kinase with phosphorylation of this target suppressing hyper-localised coincident Spo11-DSB formation.

6.6 Spo11-DSB spreading correlates with disparity in recovery of Watson/Crick hits

From visualising specific hotspots in *tel1kd,* such as *ERG25* and *HIS4::LEU2* (Figure 6.2), it became clear that directional spreading was strongly correlated with disparity in recovery of Watson or Crick ends, depending on the direction of the ORF. For example, at the *ERG25* locus, the gene is encoded on the Watson strand, with the spread DSB signal enriched on the Watson strand with very little recovery of any spread DSB signal on the Crick strand (Figure 6.2A). This disparity can be explained by a second/coincident cleavage of Spo11 adjacent to

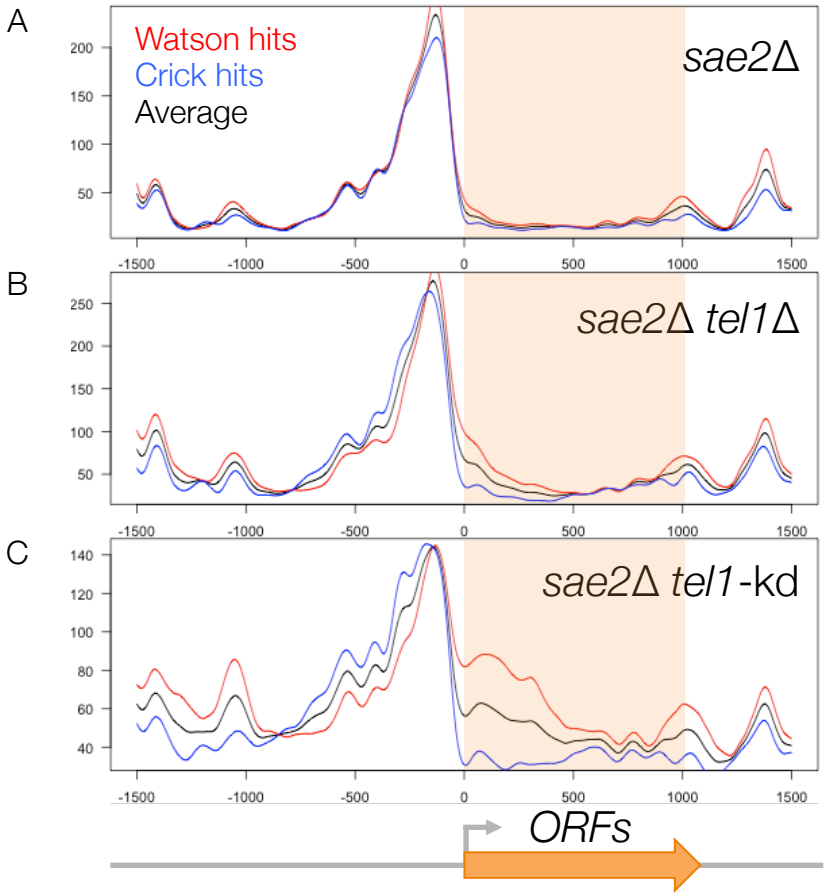


Figure 6.3: Spo11-DSB signal spreads from the NDR region into the ORF in the direction of transcription occurs genome-wide *tel1* mutants.
Unproteolysed genomic DNA was extracted from meiotic cultures at mid-meiotic prophase from *sae2*Δ (**A**), *sae2*Δ *tel1*Δ (**B**) and *sae2*Δ *tel1kd* (**C**) strains. DNA was fragmented using sonication and protein-associated molecules enriched on a glass fibre column. Samples were processed for next-generation sequencing as described in Chapter 4. Sequences were aligned to the *S. cerevisiae* S288C genome and the 5′ ends from Read1 (i.e. the 5′ base where Spo11 cut) sorted into a histogram for each specific base pair in the genome. 5′ Spo11-hits were piled-up centred on the translation start site (TSS) and visualised using R Studio ('Pileups around TSS v02.R' – Appendix 10.8). Highlighted region indicates spread region.

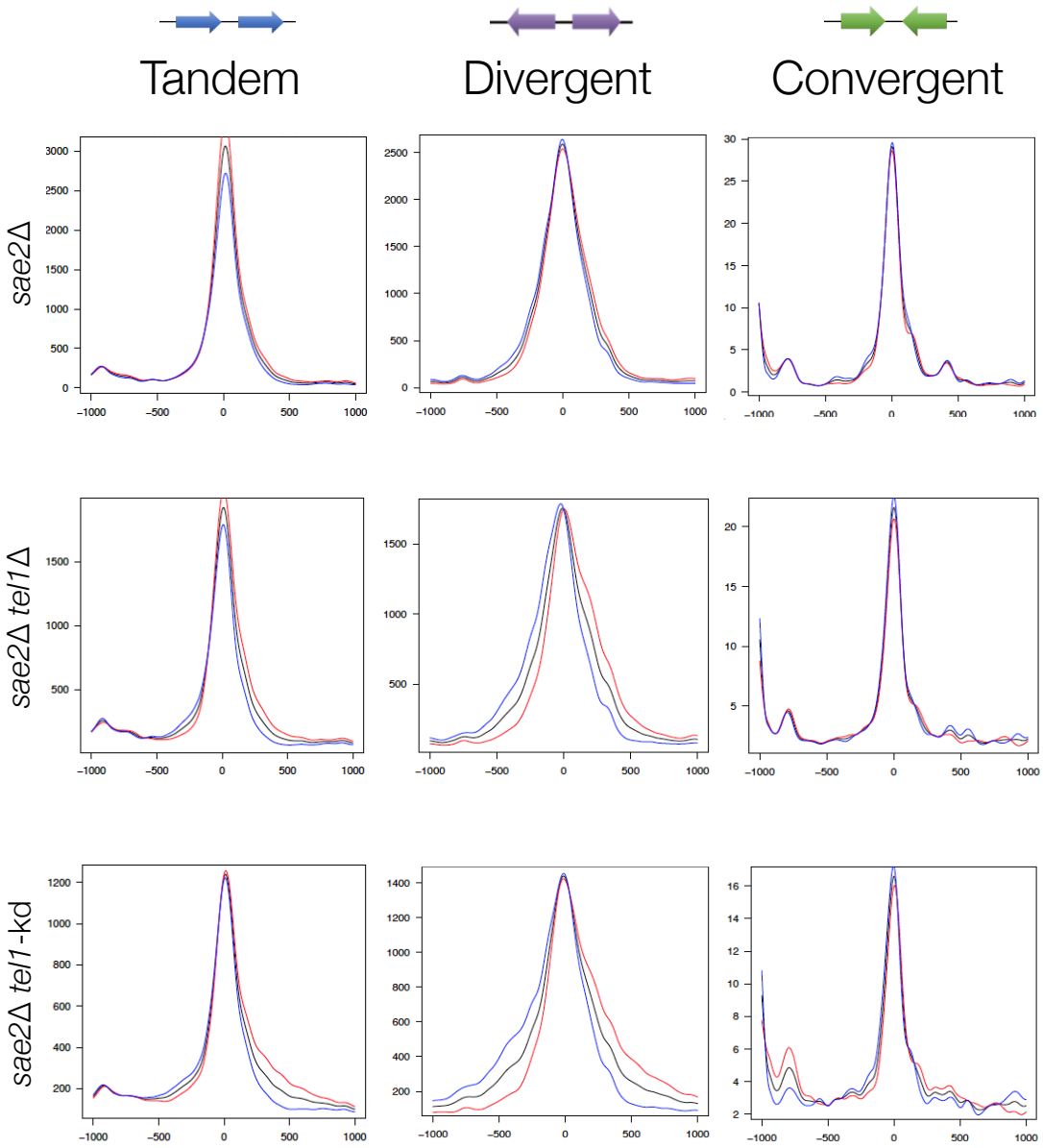


Figure 6.4: Spo11-DSB signal spreads from the NDR region into the ORF in the direction of transcription genome-wide in *tel1* mutants.
Unproteolysed genomic DNA was extracted from meiotic cultures at mid-meiotic prophase from *sae2*Δ, *sae2*Δ *tel1*Δ and *sae2*Δ *tel1kd* strains. DNA was fragmented using sonication and protein-associated molecules enriched on a glass fibre column. Samples were processed for next-generation sequencing as described in Chapter 4. Sequences were aligned to the *S. cerevisiae* S288C genome and the 5′ ends from Read1 (i.e. the 5′ base where Spo11 cut) sorted into a histogram for each specific base pair in the genome. 5′ Spo11-hits on Watson (red) and Crick (blue) strands at annotated hotspots (Pan et al., 2011) were subset into hotspots located at tandem, divergent and convergent genes and hits piled-up centred on the DSB hotspot and and visualised using R Studio ('Pileups around TSS v03 Gene expression.R' – Appendix 10.9).

an initial Spo11-DSB being formed (Figure 6.5A). This coincident cleavage creates hyper-localised 'double-cut' events. Coincident cutting would result in Spo11 bound to either end of a short fragment of dsDNA. In Spo11 end-processing mutants, such as *sae2*Δ, oligonucleotide species longer than the Spo11-oligos formed from normal Spo11-DSB end-processing are observed via the Spo11-oligo assay (Figure 3.2A – middle panel). These Spo11-oligo species are estimated to be 35 bp in length and increase in size with a 10 bp periodicity. The levels of these larger species increase in *TEL1* mutants (Mohibullah and Keeney, 2016), therefore, they could be attributable to Spo11 double-cuts. When duplex DNA is wrapped around nucleosomes, every 10.5 bp an exposed, DNase sensitive, region is displayed (Brogaard et al., 2012; Cockell et al., 1983) (Figure 6.5B). This exposed site may also be more liable for Spo11 cleavage generating this 10 bp periodicity observed. Therefore, if Spo11 mainly forms a DSB juxtaposed to the +1 nucleosome at a promoter, a 10 bp periodicity would be observed if Spo11 cleaved at positions where it could access (every 10 bp). For technical reasons, these double-cut molecules are unlikely to enter the sequencing library (Figure 6.5C) resulting in only the most distal Watson hit being recovered. Consistent with this hypothesis, direct physical analysis of hotspots via Southern blotting has revealed that spreading is only visible towards where the probe is located - but could not be detected distal to the location of the canonical DSB site (Valerie Garcia - data not shown).

6.7 Whole-genome mapping of Spo11 double-cut molecules reveals opposing disparity in Watson/Crick signal to single-cut libraries

To determine whether the disparity observed in the Watson/Crick hits in the spread signal is due to coincident cleavage by Spo11 of the same DNA molecule in close proximity, the *sae2*Δ Spo11-DSB mapping protocol was modified in order to enrich for DNA molecules where Spo11 was stably bound to both ends (schematic in Figure 6.6). The Spo11 moiety on the 5′ ends of the DSBs prevent degradation by λexo nuclease (Figure 4.5C). Therefore, λexo nuclease treatment of Spo11-DNA fragments eluted from the column enrichment procedure should degrade any molecules with a free 5′ end, i.e. any contaminating free DNA molecules and any single-cut Spo11-DNA molecules (molecules with Spo11 bound one end and a shear point at the other). Thus, this treatment will selectively degrade all non-double-cut molecules. TDP2 was utilised to hydrolyse Spo11 from both ends of the double-cut molecules and NEB (polar) adaptors were ligated simultaneously to both Spo11-derived ends. Double-cuts were PCR amplified and sequenced as before. Both 5′ ends (Read1 and Read2) are extracted from the sequencing data ('Spo11ExtractDC' – Appendix 10.10) and a 1 bp histogram made of the specific Spo11-cut sites in the genome ('Creating 1bp histograms v03 (H4L2).R' –

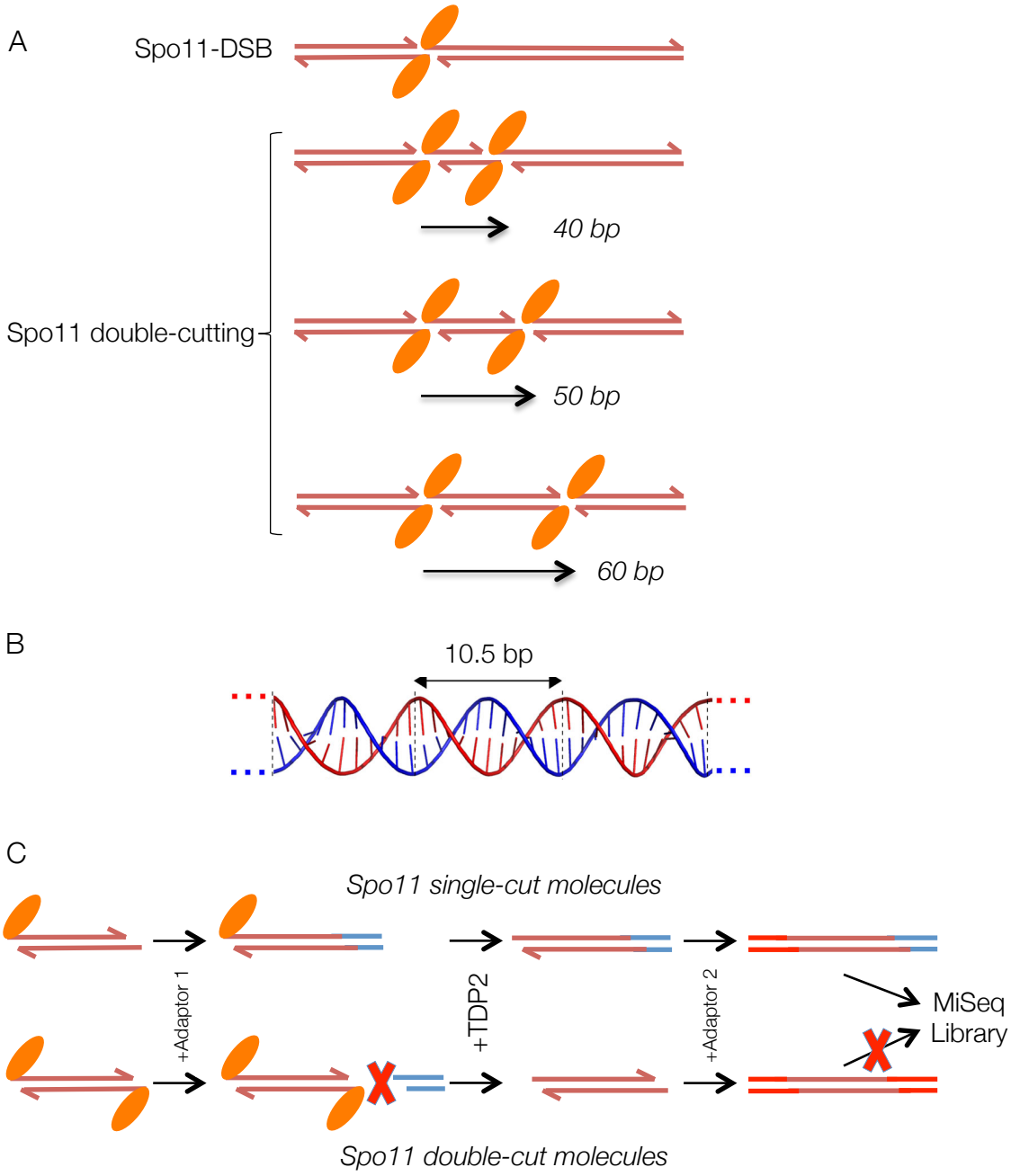


Figure 6.5: Hyperlocalised double-cutting of Spo11.
Model of proposed Spo11-double cutting mechanism. **A.** Spo11 generates multiple DSBs on the same molecule in a hyper-local area (within a hotspot) with a 10 bp periodicity, a phenotype exacerbated in a *TEL1* mutant backgrounds. **B.** Each helical turn of nucleosome bound duplex DNA contains a DNase-sensitive region 10.5 bp in periodicity (Brogaard et al., 2012; Cockell et al., 1983), comparable to the 10 bp periodicity observed in double-cut molecule size. **C.** For a DNA fragment to be sequenced it must contain polar adaptors to enter the MiSeq library. Using TDP2 to remove Spo11, two unique adaptors can be specifically ligated to the sonicated or Spo11 end generating a polar molecule with the Spo11-end distinguishable. These polar molecules can be sequenced via the paired-end Illumina MiSeq and enter the library for analysis. A molecule with Spo11 bound both ends does not have Adaptor 1 ligated due to Spo11 blocking access. Therefore, TDP2 removal of Spo11 generates a DNA fragment with two free ends that both have Adaptor 2 ligated to. These molecules with Adaptor 2 ligated to both ends are not sequencable by the Illumina MiSeq and are not present in the library.

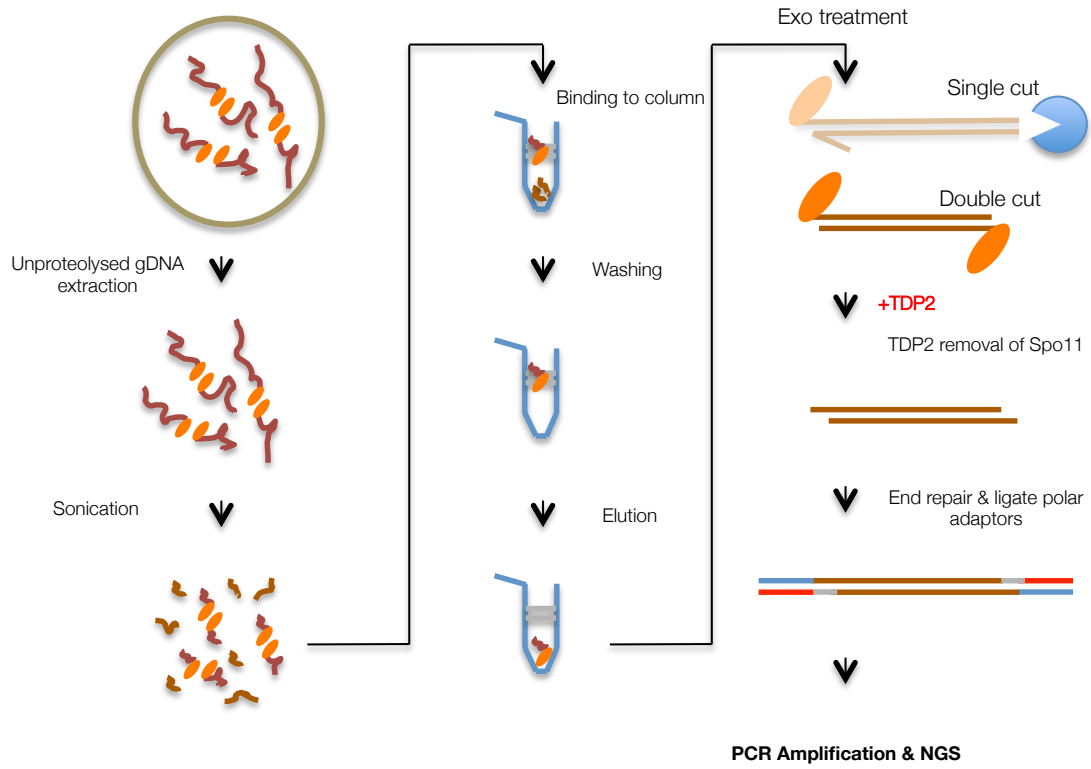


Figure 6.6: Schematic of Spo11 double-cut enrichment and mapping.
A meiotic time course is performed for a *sae2*Δ strain and cells harvested at 6 hours. Unproteolysed genomic DNA was extracted by fixing spheroplasts in ethanol, lysing with SDS, and extracting DNA and protein-bound DNA using phenol:chloroform:isoamyl alcohol (25:24:1). Molecules were precipitated with ethanol, resuspended in 1× TE and sonicated to fragment the DNA. The sample is bound to the glass fibre membrane of a QIAQuick spin column, centrifuged and the flow-through rebound to the column and centrifuged again to increase yield. The membrane is washed using TEN (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 300 mM NaCl) to remove any non-protein-bound DNA. Spo11-bound DNA is released from the column using two sequential elutions in 50 µl TES (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.5% SDS). Eluates are ethanol precipitated overnight and resuspended in 1x TE. Free-DNA and Spo11 single-cut molecules are degraded with lambda exonuclease followed by heat denaturation and AMPure XP beads clean-up. The fragments are reacted with TDP2 to remove the peptide covalently bound to the 5′ ends of the DNA before end-repair and ligation of NEBNext polar adaptors. The molecules are amplified by PCR and 2×75 bp paired-end sequencing conducted using a Illumina MiSeq machine.

Appendix 10.3). First, to determine whether these double-cut molecules were enriched at Spo11-hotspots, the *ERG25* hotspot was visualised for Spo11-DSB signal (Figure 6.7). At *ERG25* the Spo11-DSB signal spread into the ORF as before, but the disparity between Watson and Crick strand hits was reversed with Crick hits now predominating (Figure 6.7). These double-cut molecules also exist in *sae2*Δ (Figure 6.7 – *sae2*Δ) and *tel1*Δ (Figure 6.7 – *sae2*Δ *tel1*Δ), but are exacerbated in *tel1kd* (Figure 6.7 – *sae2*Δ *tel1kd)*. However, in *sae2*Δ these double-cut molecules reside within the characterised hotspot (the NDR promoter region) and not into the ORF (Figure 6.7 – *sae2*Δ), unlike *tel1kd* whose signal again spreads into the ORF in the direction of transcription.

6.8 A 10 bp periodicity of Spo11 double-cut molecule lengths is observed in *tel1* mutants, correlating with large Spo11-oligonucleotide lengths observed in *sae2Δ* cells

To determine whether these double-cut molecules have a set length or periodicity (as hypothesised – Figure 6.5), the length of double-molecule was tallied and the abundance of each length of molecule compared to the length (Figure 6.8A) (lengths of molecule calculated as part of 'Spo11ExtractDC' script). To smooth noise in the data, a 3-point moving average was calculated and plotted (Figure 6.8A) and molecules between 0-200 bp zoomed in to observe these smaller sized molecules (Figure 6.8B). Tallying the lengths of double-cuts, enriched in the modified Spo11-DSB mapping protocol, rendered a 10 bp periodicity starting at 100 bp up to 400 bp (Figure 6.8A).

6.9 Larger Spo11-oligonucleotide species increase in *tel1* mutants with a 10 bp periodicity

Cells deficient in *TEL1* generate longer Spo11-oligonucleotides when assessed via the Spo11-oligo assay (Mohibullah and Keeney, 2016). It was of interest to further characterise these larger species and determine the effect these *tel1* mutations had in a *sae2*Δ background (Figure 6.9). Following the Spo11-oligo assay as before, Spo11-oligo species enriched from wild type, *tel1*Δ and *tel1kd* (+-*sae2*Δ) meiotic cultures were end labelled with dCTP, then proteolysed with Proteinase K. In order to correctly size DNA species, TDP2 was used to remove the covalently bound peptide left after proteolysis. DNA species were resolved on a 19% denaturing urea-PAGE (sequencing) gel to resolve species from 5 nt to 100 nt (Figure 6.9A). The larger (>30 nt) species can be observed in all lanes, with and without *sae2*Δ, with a 10 nt periodicity. The precise sizes of the observable molecules are: 32 nt, 42 nt, 52nt, 62 nt, 72 nt, 82 nt, 92 nt and 102 nt. The abundance of these larger species increase in a *sae*Δ background

ERG25 locus:

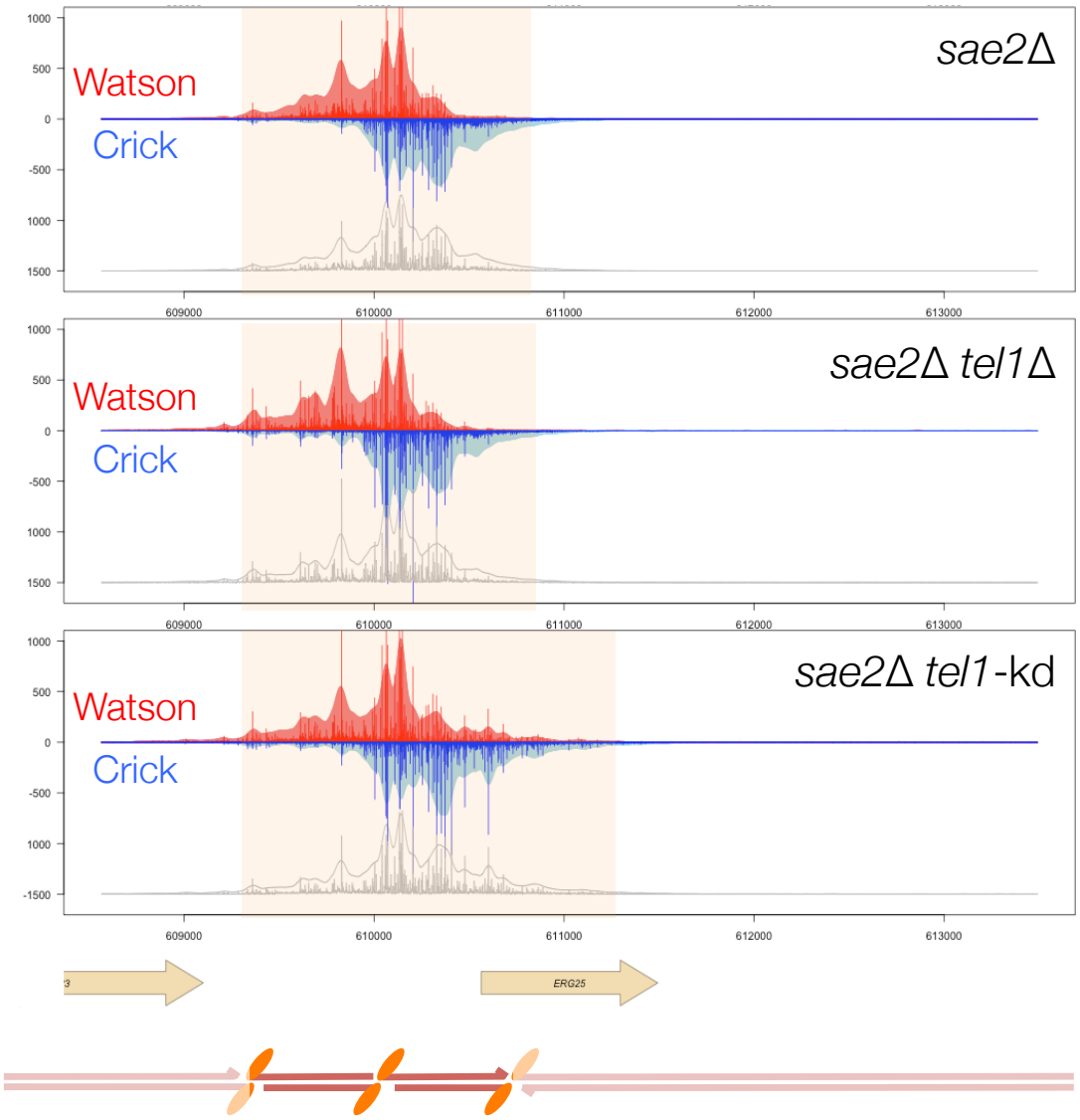


Figure 6.7: Mapping Spo11-double-cut molecules at specific hotspots.
Double-cut Spo11-DSB molecules were mapped in a *sae2*Δ background in the stated strains using the double-cut method described in Figure 6.6. Sequencing reads were generated from the stated strains enriched for Spo11-DSBs and aligned against the *S. cerevisiae* S288C genome with both 5′ Spo11 ends used to generate a 1 bp histogram. Hits on the Watson (red) and Crick (blue) strands are indicated. Putative background reads were calculated for all strains and used to normalise signal intensity. The *ERG25* locus is visualised as an example of the distribution of Spo11 double-cut molecules ('Spo11 Mapping v09 MC.R' – Appendix 10.5). Highlighted region indicates double-cut area.

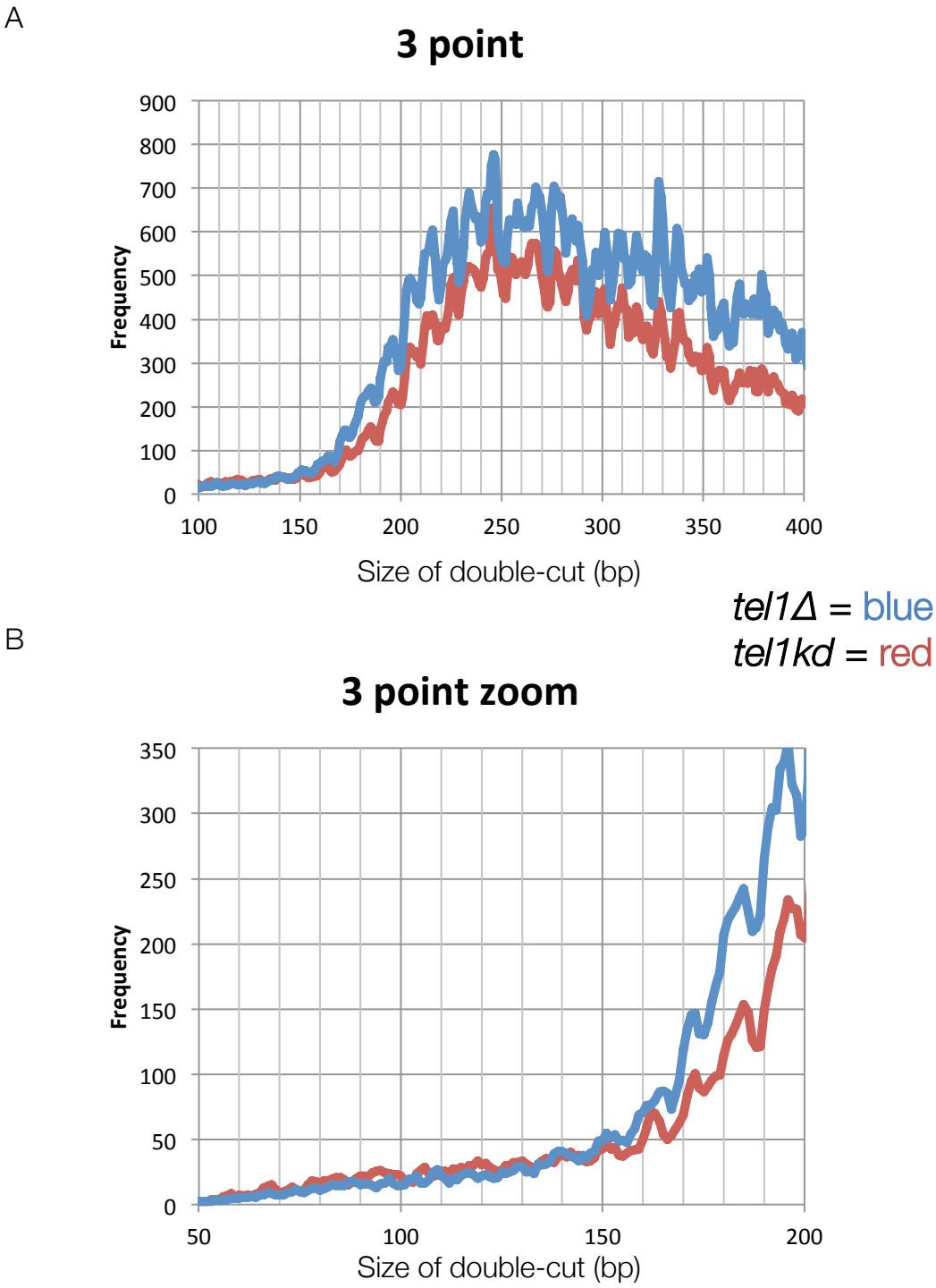


Figure 6.8: Analysing the length of the double-cut molecules reveals a 10 bp periodicity. Double-cut Spo11-DSB molecules were mapped in a *sae2*Δ *tel1*Δ and in a *sae2*Δ *tel1-kd* background. Sequencing reads generated from the double-cut enrichment strategy (Figure 6.6 and 6.7) were aligned against the *S. cerevisiae* S288C genome. **A.** A 3-point moving average was calculated and plotted for the distance between the two 5′ ends (size of double-cut) against the number of hits for each given length (frequency). **B.** To visualise short molecules, the axis were scaled to observe molecules less than 200 bp in length.

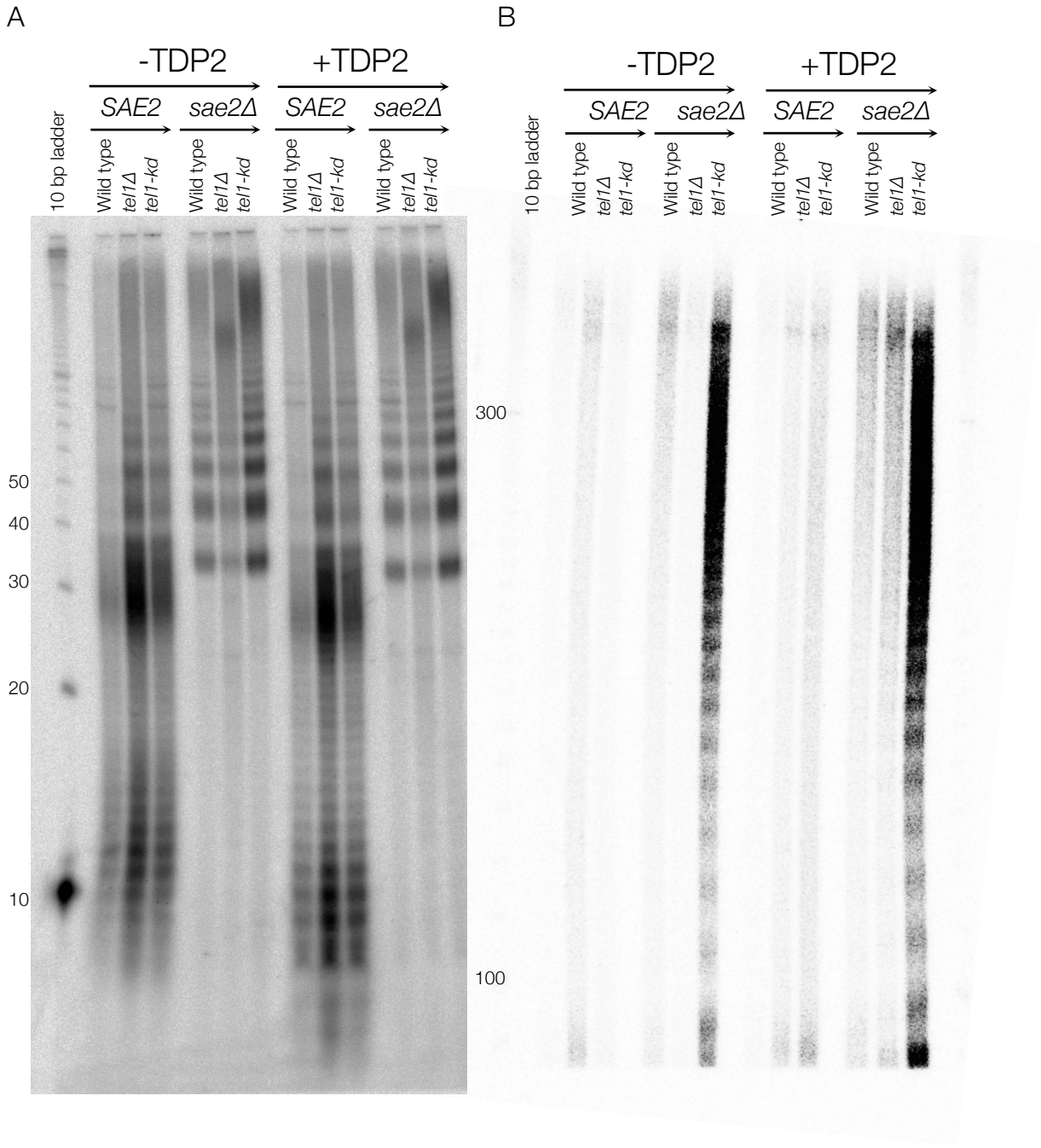


Figure 6.9: In both wild type and *sae2*Δ cells a 10 bp periodicity of species is observed from 35 bp. These species are exacerbated in *TEL1* mutants.
Spo11 oligonucleotides were enriched and 3′ end labelled using TdT and dCTP radionucleotide from stated mutants. Spo11-oligos were digested with Proteinase K at 60 °C for 1 hour and DNA precipitated overnight at -80 °C using 100% ethanol. Precipitate was resuspended in TDP2 reaction buffer and TDP2 protein added (unless otherwise stated) and incubated at 37 °C for 1 hour. The 10 bp ladder was labelled using TdT and dCTP radionucleotide. **A.** Spo11-oligos from 6 hours into meiosis were treated with and without TDP2 to accurately size Spo11-oligos from the stated mutants. Samples were separated on a 19% denaturing urea-PAGE gel. Putative Spo11 double-cut molecules are observed in *sae2*Δ (and wild type *SAE2*) with a 10 bp periodicity starting at 35 bp. **B.** Spo11 oligos from stated time points in meiosis were separated on a 10% denaturing urea-PAGE gel. Putative Spo11 double-cut molecules with a 10 bp periodicity are observed up to 150 bp in *tel1kd.* Asterisks indicate non-specific contaminants.

(Figure 6.9A right hand panel) and further increase in the *tel1kd* mutant. In both *tel1* mutants a smear is present at ~150 nt in *tel1*Δ and at 150 to 400 nt in *tel1kd*. The 10 bp periodicity observed when analysing the Spo11 double-cut molecule lengths from the genome-wide data (Figure 6.8A) started at 100 bp up to 400 bp, suggesting this smear observed could be the same double-cut molecules (as hypothesised – Figure 6.5).

To try and further resolve these molecules, Spo11-derived DNA molecules from *sae2*Δ, *sae2*Δ *tel1*Δ and *sae2*Δ *te1kd* cells were resolved on a 10% denaturing urea-PAGE gel (Figure 6.9B). The 10 bp periodicity of Spo11-DNA molecules can be observed up to 250 bp in length in the *tel1kd* strain with the smeared signal extending further but unable to be resolved into a ladder at longer lengths (Figure 6.9B left hand panel). This periodicity thus appears to directly correlate with the periodicity observed from the genome-wide mapping data (Figure 6.8A). This periodicity cannot be observed in the *tel1*Δ or *TEL1*+ strains above 80 bp (Figure 6.9B middle and right hand panels) potentially due to either low abundance or absence, supported by the lack of visible spreading in these strains (Figure 6.3).

6.10 Discussion

Tel1 is responsible for DSB interference in *cis* during meiosis (Garcia et al., 2015). To determine whether Tel1 kinase activity was required for DSB interference a kinase deficient mutant of Tel1 (*tel1kd*) was generated. Unexpectedly, the *tel1kd* mutant causes a smearing phenotype of DSB signal via Southern blotting (Figure 6.1C), preventing accurate determination of DSB interference in this strain. However, this phenotype was intriguing as this smearing of the DSB sites was not due to resection (VG – unpublished work). Utilising the *sae2*Δ Spo11-DSB mapping procedure developed in Chapter 5, genome-wide mapping of Spo11-DSBs in the *tel1kd* background was performed. The *tel1kd* mutant caused genome-wide spreading of DSBs from promoter regions into the ORFs of the downstream genes, in the direction of transcription (Figure 6.4). A disparity between Watson and Crick strand signal was noted in the spread region (Figure 6.2). This disparity was attributed to coincident cleavage of the same DNA molecule by Spo11, generating a double-cut Spo11 molecule (a molecule with Spo11 covalently bound to both 5′ ends – Figure 6.5). By altering the *sae2*Δ Spo11 mapping procedure, these double-cut molecules were enriched and mapped as before (Figure 6.6), generating maps with the opposing disparity between Watson and Crick signal, supporting the view these molecules are double-cut molecules.

Analysis of the lengths of these double-cut molecules revealed a 10 bp periodicity ranging from 100-400 bp in length (Figure 6.8A). This 10 bp periodicity correlated with a 10 bp periodicity of Spo11-oligonucleotides observed via the Spo11-oligo assay (Figure 3.2 and

Figure 6.9). This 10 bp periodicity ranged from 32 to 102 nt (Figure 6.9) and these larger Spo11-DNA molecules were observed to increase in *tel1*Δ and *tel1kd* mutants (Mohibullah and Keeney, 2016). Using denaturing urea-PAGE, the 10 bp periodicity of these Spo11-derived DNA molecules was further characterised (Figure 6.9). These species ranged from 32 nt to 202 nt in *tel1*Δ strains and from 32 nt to 402 nt in *tel1kd* (Figure 6.9B) agreeing with *sae2*Δ mapping data of Spo11 double-cut lengths (Figure 6.8A). However, molecules shorter than 100 bp in the *sae2*Δ mapping data are not detected via this procedure, potentially due to the size-selection steps required for efficient mapping (to remove contaminant adaptors). Therefore, whilst the molecules below 100 bp also display a 10 bp periodicity via urea-PAGE gels (Figure 6.9), the mapping data has yet to detect these very short double-cut molecules. The 10 bp periodicity observed could be attributed to exposed/DNase sensitive sites of nucleosomal duplex DNA that occur every 10.5 bp (Brogaard et al., 2012a; Cockell et al., 1983) (Figure 6.5B), as these exposed regions may also be more liable for Spo11 cleavage. Therefore, if Spo11 mainly forms a DSB juxtaposed to a nucleosome, a 10 bp periodicity would be observed if Spo11 cleaved at positions where it could access (sites with 10 bp increments).

The increased hyper-localised DSBs may occur as a consequence of aberrant control of DSB numbers, due to the *tel1* mutation, with multiple DSBs occurring within the same hotspot. The 10 bp periodicity observed within this localised region could be due to the cleavage of Spo11 on one side/turn of the DNA duplex potentially due to DNA-binding proteins. More intriguingly it could be due to adjacent nucleosomes. DNA duplexes wrap around nucleosomes and have an exposed site every 10 bp, observed via MNase hyper-sensitivity (Brogaard et al., 2012; Cockell et al., 1983). Therefore these sites may also be more favourable to Spo11-cleavage.

To account for the differences in spreading observed between the *tel1*Δ and *tel1kd* mutants it is possible that the presence of the Tel1 protein in the *tel1kd* mutant is blocking a redundant pathway from preventing hyper-localised spreading. This redundant pathway could potentially be Mec1 due to the similar substrates that both Mec1 and Tel1 phosphorylate (Baldo et al., 2008; Ritchie et al., 1999). Preliminary work in the Neale lab has observed in a *mec1-kinase dead* mutant, a reduction in Spo11-DSB spreading into the ORF compared to wild type strains (DJ/MJN – unpublished data). However, the target for phosphorylation by Tel1/Mec1 is still unknown.

One potential target for this spreading phenotype is the MRX complex. The MRX complex is essential not only for DSB end processing and resection (Mimitou and Symington, 2008; Zhu et al., 2008b), but the components are also members of the Spo11 complex and are required for DSB formation in meiosis (reviewed in Keeney, 2008). In Spo11 end processing mutants such

as *sae2*Δ*, rad50S* and *mre11* nuclease mutants, Mre11 proteins accumulate at the DSB site (Usui et al., 2001). Two Mre11 N-terminal mutations, *H33R* and *P110L* decrease both Mre11 binding and retention to DSBs in a *sae2*Δ background and suppress *sae2*Δ CPT sensitivity (Chen et al., 2015; Puddu et al., 2015). *TEL1* inactivation does not suppress the DNA damage hypersensitivity of *sae2*Δ but does reduce the ability of the *mre11-H37R* mutant to suppress *sae2*Δ CPT hypersensitivity. In *TEL1* deficient cells the *mre11-H37R* mutant now forms foci suggesting Tel1 cooperates with Sae2 to promote the removal of Mre11 from DSBs (Puddu et al., 2015). Hyper-phosphorylation of the mammalian MRX complex (MRN) by ATM (Tel1) leads to dissociation of the MRN complex from the DSB site. Therefore, if the *tel1kd* protein stabilised the MRX complex at the DSB but failed to phosphorylate it (and/or prevented Mec1 from phosphorylating it) then this may lead to increased DSB formation locally within the hotspot.

CHAPTER 7:

GENOME-WIDE MAPPING OF TOP2-DSBs

Chapter 7: Genome-wide mapping of Top2-DSBs

7.1 Introduction

The topoisomerase II (Top2) dimer introduces transient DSBs, with 4 bp overhangs, throughout the genome to allow strand passage and relaxation of superhelical tension (Burden and Osheroff, 1998; Nitiss, 2009; Schoeffler and Berger, 2005), issues generated during the normal cellular processes of DNA replication, transcription and chromatin remodelling (Bermejo et al., 2007; Fachinetti et al., 2010). The transient Top2-DSBs are formed through a similar mechanism to Spo11-DSBs. The catalytic tyrosine residue attacks the DNA phosphate backbone to create a covalent bond to the 5′ ends, either side of the DSB. Strand passage occurs through this gap and the Top2 dimer reseals the DSB once this is complete (Deweese and Osheroff, 2009; McClendon and Osheroff, 2007; Nitiss, 2009; Pommier et al., 2010b; Schoeffler and Berger, 2008; Schoeffler and Berger, 2005; Wang, 2002).

The DSBs generated by Top2 can become stabilised during its normal catalytic cycle or in response to poisons, resulting in a toxic lesion (Corbett and Osheroff, 1993). Etoposide binds the protein-DNA interface, misaligning the two DNA strands, resulting in prevention of religation and trapping of the Top2 dimer at the DSB site (Pommier and Marchand, 2011; Wu et al., 2011). These Top2-DSBs can be repaired by either the NHEJ or HR pathways, but first the 5′ adducts need to be removed (Cruz-García et al., 2014; Gómez-Herreros et al., 2013). Two pathways contribute to the processing of Top2-DSBs: (i) In mammalian, but not yeast cells, the TDP2 pathway can directly hydrolyse the phosphotyrosine bond between the 5′ linked adduct and the DNA (Cortes-Ledesma et al., 2009; Gao et al., 2014). (ii) In yeast and mammalian cells, the Sae2 (CtIP) and Mre11 pathway endonucleolytically processes Top2-DSBs (Aparicio et al., 2016; Hartsuiker et al., 2009b; Nakamura et al., 2010). *S. cerevisiae* do not contain an ortholog of mammalian TDP2 but do contain Tdp1, an ortholog of mammalian 3′-tyrosyl-DNA-phosphodiesterase (TDP1), which cleaves 3′ phosphotyrosyl bonds and has shown residual *in vitro* activity on 5′ phosphotyrosyl bonds (Murai et al., 2012; Nitiss et al., 2006).

Genome-wide mapping of stalled Top2 complexes in mammalian cells has been conducted at low resolution (Baranello et al., 2014). Single-stranded (SSBs) and double-stranded (DSBs) breaks were both labelled and enriched for from cells exposed to etoposide. Supporting earlier studies, etoposide was found to generate mostly SSBs over DSBs, potentially due to etoposide inhibiting the religation of each broken strand by the two Top2 monomers independently (Bromberg et al., 2003). Top2 lesions are enriched at promoter regions, with an anti-

correlation to nucleosome occupancy, compared to the rest of the genome (Baranello et al., 2014; Sperling et al., 2011) and this was positively correlated with the high levels of transcription (Baranello et al., 2014). Nucleotide-resolution mapping of Top2 lesions has yet to be conducted.

7.2 Deletion of *PDR1* allows exposure and sensitivity of *S. cerevisiae* cells to etoposide in liquid culture

Wild type *S. cerevisiae* are not sensitive to low doses of etoposide due to an array of drug extrusion pumps that are involved in actively pumping out these compounds, preventing toxic build-up (Balzi et al., 1987; Balzi and Goffeau, 1991; Stepanov et al., 2008). These pumps are encoded for by the Pleiotropic Drug Resistance (PDR) network of genes. This network consists of four transcriptional regulators (*PDR1, PDR3, PDR7* and *PDR9),* which regulate gene expression of the membrane drug extrusion pump *PDR5* (Balzi et al., 1994). The *PDR1* transcriptional regulator is also involved in regulation of other membrane pumps, which may be involved in multidrug resistance (Balzi and Goffeau, 1991). A sensitivity cassette has been created that fuses the potent transcriptional repressor *CYC8p* to Pdr1p (Stepanov et al., 2008). To determine whether the *CYC8-PDR1* cassette confers etoposide sensitivity to wild type, *sae2*Δ and *mre11*Δ cells spot tests were conducted (Figure 7.1, conducted by Holly Thomas (HT), MSc student). Serial 10-fold dilutions of exponentially growing cells were spotted onto YPD plates containing various concentrations of etoposide for each strain and grown for 3 days. Wild type cells that did not contain the sensitivity cassette were resistant to all concentrations of etoposide (Figure 7.1A). The sensitivity cassette, incorporated into a wild type strain (*pdr1*Δ) resulted in sensitivity to etoposide at 0.3 μM and 1.0 μM etoposide concentrations, whilst the *sae2*Δ and *mre11*Δ strains were sensitive at all concentrations. Therefore, this sensitivity cassette can be used to render *S. cerevisiae* sensitive to topoisomerase poisons under chronic exposure.

To determine the effects of the sensitivity cassette on acute exposure to etoposide, overnight cells (*sae2*Δ and *mre11*Δ both containing the sensitivity cassette) were diluted to an OD_{600} of 0.5 in fresh YPD and grown to an OD_{600} of 2.0 (Figure 7.1B, conducted by HT). Cells were taken, serially diluted and spotted (t0). The remaining culture had 1 μM etoposide or equivalent volume of DMSO added and left to incubate at 30°C. Cells were taken after 1, 2, 4 and 24 hours, serially diluted and spotted onto YPD plates and grown for 3 days. Both *mre11*Δ and *sae2*Δ strains displayed very little etoposide sensitivity after 1 h acute exposure. 2-24 h etoposide exposure caused increased sensitivity in both *sae2*Δ and *mre11*Δ strains, with

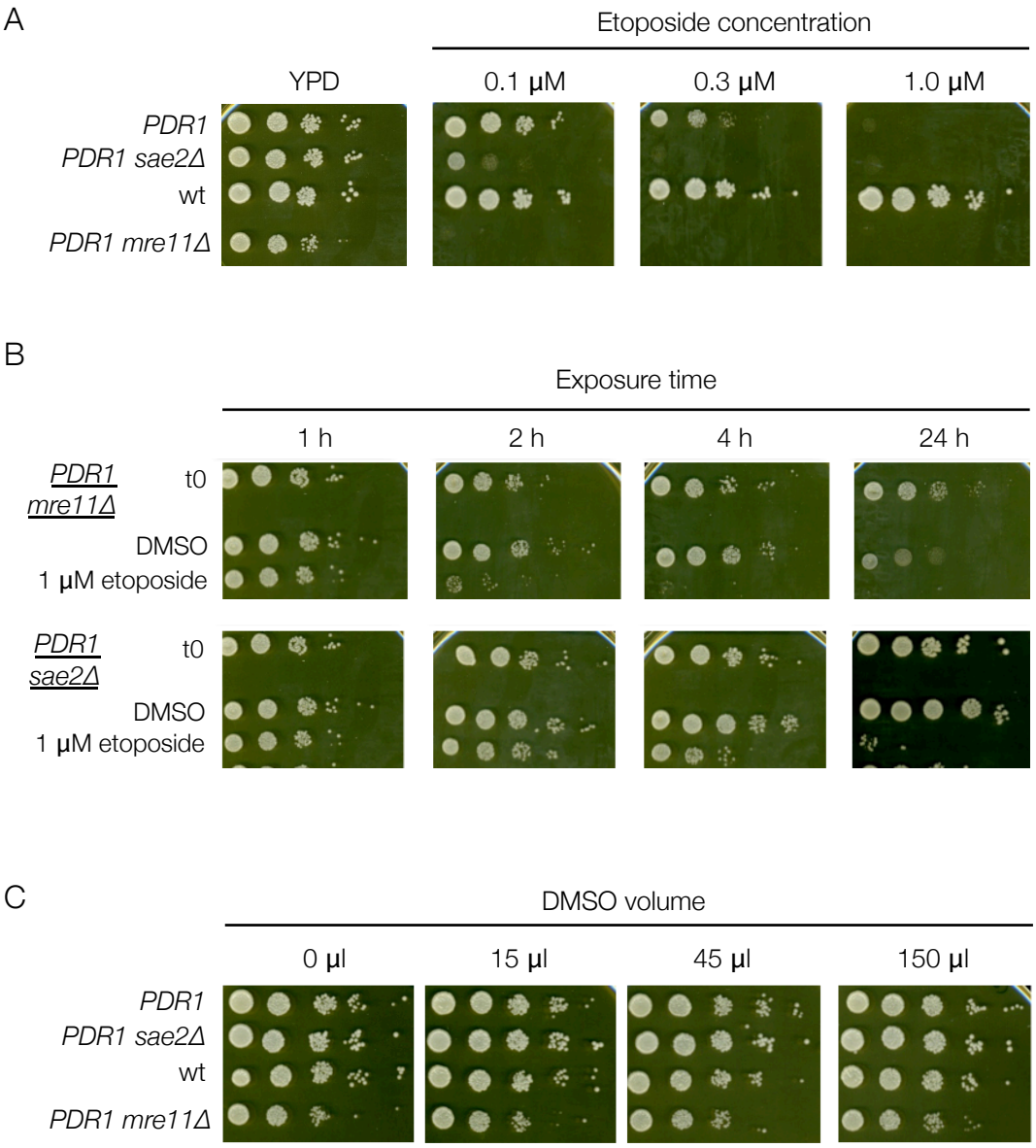


Figure 7.1: A sensitivity cassette renders wild type, *sae2*Δ and *mre11*Δ cells sensitive to chronic and acute etoposide exposure.
Serial spot dilution assays of sensitivity to etoposide of stated *S. cerevisiae* mutants. The sensitivity cassette ('*PDR1'*) consists of a deletion of the *PDR1* gene and a fusion of the transcriptional repressor CYC8p to the DNA binding domain of Pdr1p. **A&C.** Cells were grown overnight at 30 °C, diluted to an OD_{600} of 0.5 in fresh YPD and grown to an OD_{600} of 2.0. A serial 10-fold dilution series, starting at OD_{600} 2.0, were spotted onto YPD plates containing the stated final concentration of etoposide (**A**) or DMSO (**C**) and grown at 30 °C for 3 days. **B.** Overnight cells were diluted to an OD_{600} of 0.5, grown to an OD_{600} of 2.0 and the culture split into four. Cells were taken at time 0 ('t0') and a serial dilution series was spotted. 1 µM Etoposide or DMSO was added and incubated for the stated exposure time. A serial 10-fold dilution series was spotted and cells grown at 30 °C for 3 days. Images were taken using a Syngene InGenius bioimaging system. Experiment was performed by Holly Thomas.

*mre11*Δ having very little cell grow at all time points from 2 h. The *mre11*Δ cells had reduced growth not only to the DMSO control (growth defect) but also to t0 suggesting cell death had occurred for 2-24 h. The *sae2*Δ cells had a similar phenotype at 4-24 h (Figure 7.1B). Etoposide is poorly soluble in aqueous solution and was therefore dissolved in DMSO. To determine whether the DMSO solvent affects cell growth, wild type, *sae2*Δ and *mre11*Δ strains were spotted on YPD plates containing varying levels of DMSO. All three strains observed no growth defect to DMSO solvent controls (Figure 7.1C), suggesting effects observed with etoposide exposure are due to the compound itself.

Collectively these results from Figure 7.1 suggest the sensitivity cassette incorporated into *S. cerevisiae* render the cells sensitive to 1 μM etoposide exposure for 4 h in liquid culture. These conditions were used for acute exposure of strains to etoposide in liquid culture to accumulate Top2-DSBs on a large scale.

7.3 Exposure to etoposide generates Top2 hotspots spread across the entire genome

Genome-wide mapping of meiotic Spo11-DSBs from *sae2*Δ cells utilised the covalently bound Spo11-protein to enrich for protein-linked DNA molecules (Chapter 5). TDP2, a human protein that *in vivo* cleaves the phosphotyrosine bond between proteolysed Top2 and the 5′ end of DNA was also capable of removing Spo11 (Chapter 4). Exposure of cycling cells to etoposide (Figure 7.1) causes accumulation of Top2-DSBs, complexes that consist of Top2 covalently bound to the 5′ end either side of the DSB, a situation similar to Spo11-DSBs. Therefore, the *sae2*Δ mapping procedure (Chapter 5) could potentially be utilised for mapping of Top2-DSBs genome-wide with nucleotide resolution (as depicted in Figure 7.2).

To determine whether Top2-DSBs could be mapped genome-wide in this manner, haploid wild type, *sae2*Δ and *mre11*Δ strains, all containing the sensitivity cassette, were grown overnight in YPD (Figure 7.3, conducted alongside HT). Cultures were diluted to an OD_{600} of 0.5 in 100 ml YPD and grown to an OD_{600} of 2.0. Cultures were split into 2 × 50 ml and a final concentration of 1 μM etoposide ('E' – etoposide), or an equivalent volume of DMSO for a control ('U' – untreated), was added and cultures grown for a further 4 h. Cells were pelleted and unproteolysed genomic DNA extracted. DNA was fragmented to an average size of 500 bp and protein-DNA species enriched for on a glass fibre filter. The sonicated ends of the fragments were end-repaired and Adaptor 1 (Read2) ligated. Samples were incubated with TDP2 to remove the covalently bound Top2 protein, a second round of end-repaired conducted and Adaptor 2 (Read1) ligated onto the Top2-bound ends. Samples were PCR amplified and

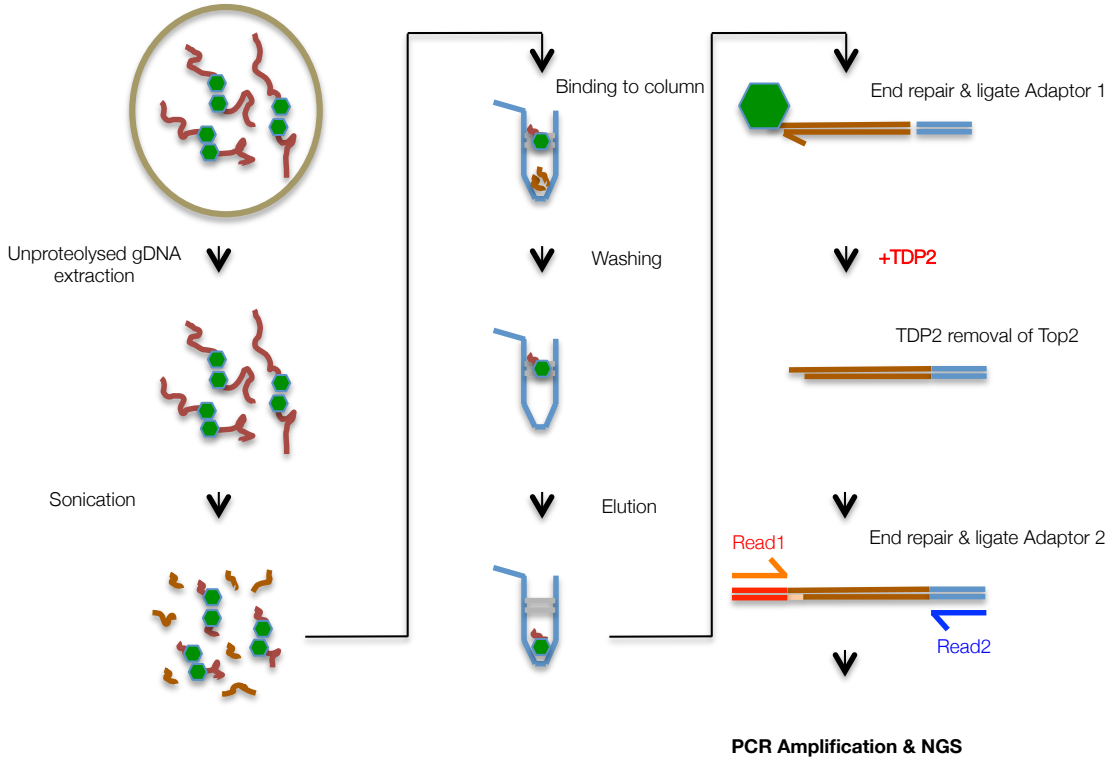
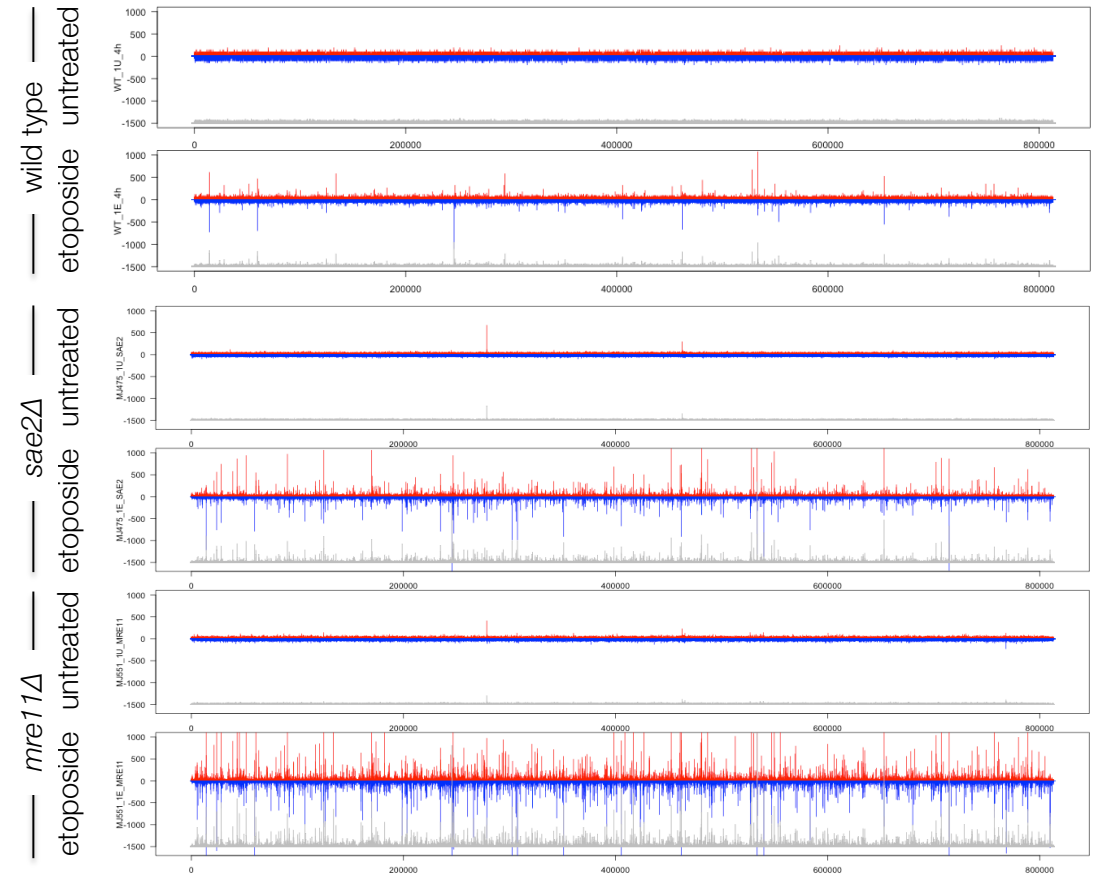
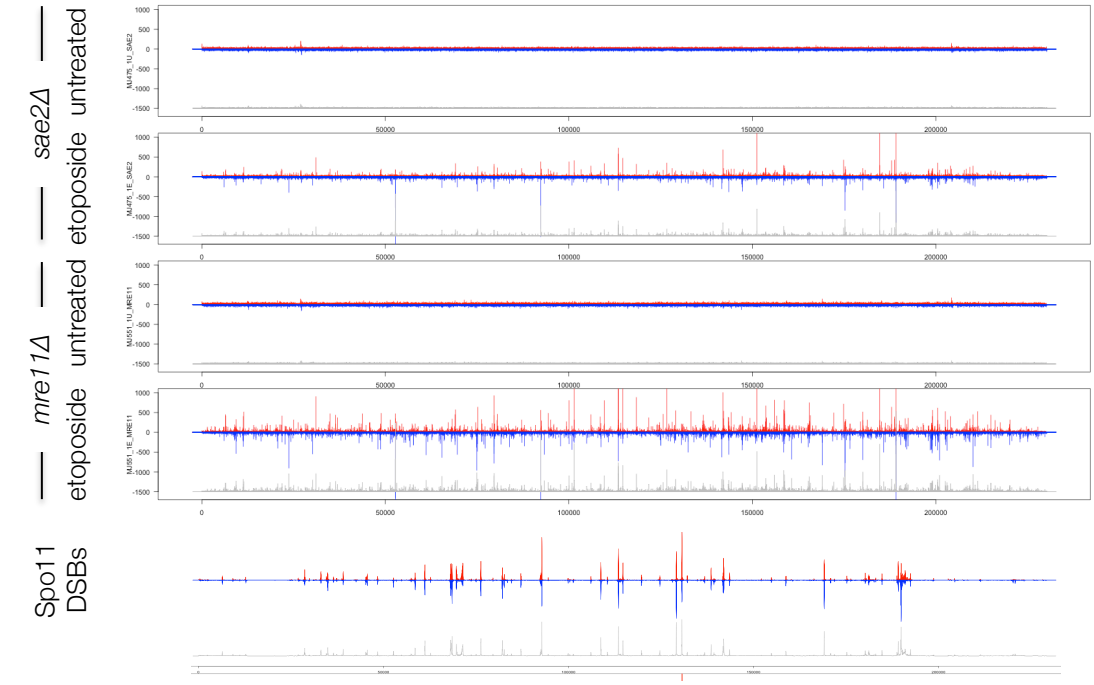


Figure 7.2: Schematic of Top2 enrichment and mapping.
Overnight YPD cultures were diluted to an OD_{600} of 0.5 in 50 ml YPD, in duplicate, and grown to an OD_{600} of 2.0. A final concentration of 1 µM etoposide ('E' – etoposide), or an equivalent volume of DMSO for a control ('U' – untreated), was added and cultures grown for a further 4 h and cells pelleted. Unproteolysed genomic DNA was extracted by fixing spheroplasts in ethanol, lysing with SDS, and extracting DNA and protein-bound DNA using phenol:chloroform:isoamyl alcohol (25:24:1). Molecules were precipitated with ethanol, resuspended in 1× TE and sonicated to fragment the DNA. The sample is bound to the glass fibre membrane of a QIAQuick spin column, centrifuged and the flow-through rebound to the column and centrifuged again to increase yield. The membrane is washed using TEN (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 300 mM NaCl) to remove any non-protein-bound DNA. Top2-bound DNA is released from the column using two sequential elutions in 50 µl TES (10 mM Tris Base·HCl pH 8.0, 1 mM EDTA, 0.5% SDS). The sonicated end of the DNA fragments are blunted using the NEBNext Ultra II end-repair kit and Adaptor 1 ligated on before removal of unligated adaptors using AMPure XP beads. The fragments are then reacted with TDP2 to remove the Top2 peptide covalently bound to the 5′ end of the DNA before fill-in of the end that was Top2-bound. Adaptor 2 is then ligated onto the Top2-derived end of the fragment, and the now polar molecules are amplified by PCR and 2×75 bp paired-end sequencing conducted using a Illumina MiSeq machine. Read1 contains the 5′ base of where Top2 cut. Read2 contains information regarding the sheared end of the fragment.

A. Chromosome 2 (813 kb)



B. Chromosome 1 (230 kb)



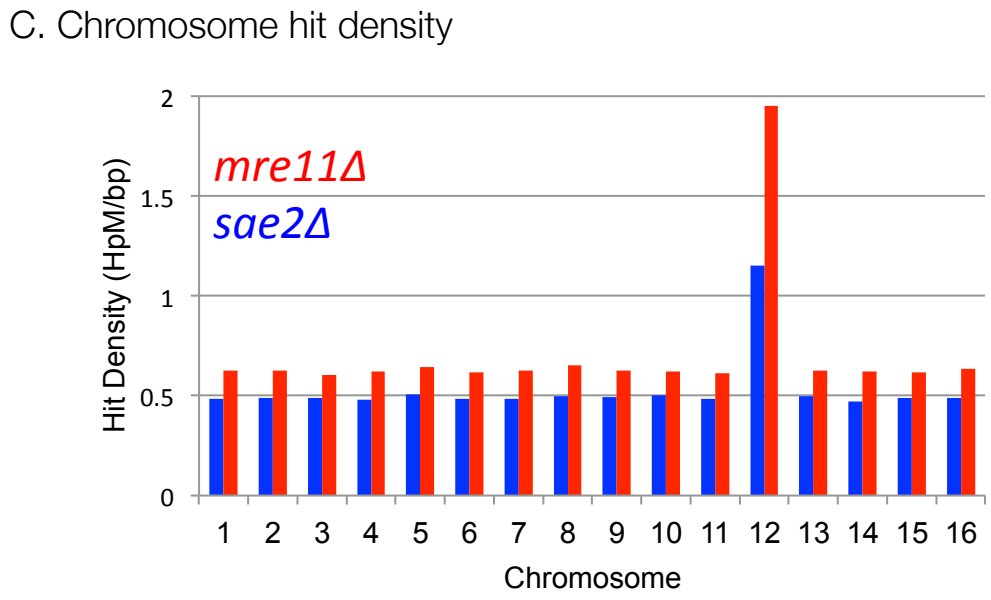


Figure 7.3: Genome-wide mapping of Top2 lesions and the Top2 hit density per chromosome. Haploid wild type, *sae2*Δ and *mre11*Δ strains were grown overnight in YPD. Cultures were diluted to an OD_{600} of 0.5 and grown to an OD_{600} of 2.0. Cells were split into two 50 ml cultures and 1 µM etoposide ('etoposide') or equivalent volume of DMSO ('untreated') added. Cells were incubated for an additional 4 h and cells harvested by centrifugation. Next-generation sequencing libraries were prepared as stated in Figure 7.2. Sequences were aligned to the *S. cerevisiae* S288C genome using bowtie2. 5′ Read1 (Top2) ends were extracted an converted into a 1 bp histogram using R Studio. Top2 hits along Chromosome 2 (**A**) and Chromosome 1 (**B**) were compared between strains and between treatments using R Studio ('Topo mapping v01.R' – Appendix 10.11). Top2 hits were also compared against Spo11 hits on Chromosome 1 (**B**). **C.** The hit density (hits per million reads (HpM) per base pair) was calculated and compared for each chromosome and between *mre11*Δ and *sae2*Δ strains.

uniquely barcoded, enabling mixing/multiplexing of the samples. Samples were sequenced on an Illumina MiSeq, demultiplexed and aligned to the *S. cerevisiae* S288C genome using bowtie2 ('Spo11Align.command' – Appendix 10.1). 5′ Read1 reads (Top2 end) were extracted ('Spo11Extract.pl' – Appendix 10.2) and converted into a 1 bp histogram using R Studio ('Creating 1 bp histograms v03.R' – Appendix 10.3). Top2 hits were compared between wild type, *sae2*Δ and *mre11*Δ (untreated or +etoposide) and plotted for chromosome 2 with Watson and Crick hits differentiated (Figure 7.3A) ('topo mapping v01 mc.R' – Appendix 10.11). All untreated samples have no distinct peaks at this range with even coverage across the whole length of chromosome 2, suggesting this is background/non-specific hits. Treatment with etoposide in all three strains studied generated a large increase in signal with defined peaks. The etoposide treated wild type cells had less signal compared to the *sae2*Δ and *mre11*Δ mutants suggesting that Sae2 and Mre11 are both responsible for the repair of etoposide-induced Top2-DSBs. The *mre11*Δ strain had higher signal compared to *sae2*Δ, suggesting that Mre11 is more important for the repair of Top2-DSBs than Sae2. These results correlate with the sensitivity to etoposide via spot testing (Figure 7.1). Chromosome 1 was also visualised for Top2-DSBs, in untreated and etoposide treated, *sae2*Δ and *mre11*Δ cells and was compared against the Spo11-DSB hits over the same region (Figure 7.3B). Top2-DSBs are much more frequent than Spo11-DSBs with a broad coverage across most NDRs throughout the entire genome. In comparison Spo11-DSBs reside in much more defined regions (hotspots), mainly in promoters (Pan et al., 2011; Chapter 5). This suggests that Spo11 has a much tighter regulation for DSB formation than Top2 does, with factors restricting Spo11-DSB numbers and their location.

To determine whether the length of the chromosome has an effect on Top2-DSB formation the hit density between chromosomes, or between the *mre11*Δ and *sae2*Δ strains, was analysed (Figure 7.3C). Hit density (hits per kb) was unchanged across chromosomes, except for chromosome 12. This can be accounted to the rDNA that resides on chromosome 12. The rDNA consists of repetitive DNA, whose repeat number can differ between strains, normally between 100-200 repeats (Petes, 1979). The published *S. cerevisiae* S288C genome only contains two copies of the repeat and therefore, this site has amplified signal due to multiple hits in the repetitive region all aggregating at one site. Finally, the peaks between all three etoposide treated strains are in similar positions, indicating some specificity as to where Top2 is generating DSBs in the genome.

7.4 Top2-DSBs accumulate in NDRs with a 4 bp overhang observable

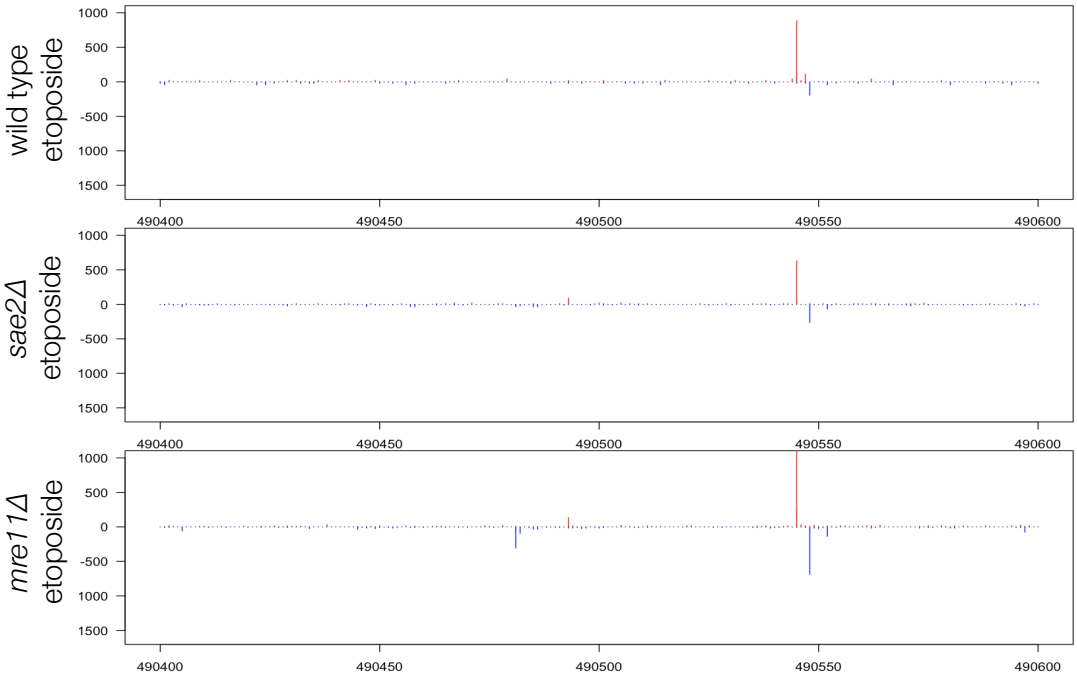
The Top2 dimer, when generating DSBs, does so with a 4 bp overhang (Dong and Berger, 2007). Spo11 generates similar DSBs to Top2 except with a 2 bp overhang (Liu et al., 1995), which can be visualised from the *sae2*Δ mapping (Figure 5.10). To determine whether the Top2 4 bp overhang could be observed, a 300 bp region was studied (Figure 7.4A). The 4 bp offset can be observed in the etoposide treated *sae2*Δ cells and in both the untreated and etoposide treated *mre11*Δ cells, suggesting that both sides of the Top2-DSBs are enriched from these DSBs.

At this scale, Top2-DSBs are observed enriched at nucleosome-depleted regions (NDRs), although there are hits within genic regions (Figure 7.4B). Top2 activity is positively correlated with actively transcribed regions, potentially due to an increase in transcription-generated torsional stress which requires increased action of Top2 (Baranello et al., 2014). An array of genes can be orientated in three different manners (as depicted in Figure 7.4C). Tandem genes are adjacent genes on the same strand with the intergenic region containing a terminator from the first gene and a promoter from the second gene. Divergent genes are genes on opposing strands with the intergenic region containing both gene promoters. Convergent genes are genes on opposing strands with the intergenic region containing both terminators. Therefore, to determine whether the orientation of genes, and thus promoter containing regions, affects the level of Top2-DSBs, the mean Top2 signal per kb was plotted for NDRs (split tandem, convergent and divergent gene regions) and genic regions (Figure 7.4B). Only 10% of Top2-DSBs occur within genic regions, suggesting that Top2-DSBs favourably form within intergenic/NDR regions. Top2 hits within NDRs form more favourably in tandem and divergent regions compared to convergent regions (Figure 7.4B), suggesting that Top2 is more active at promoter regions compared to terminator regions. This is observed in both *mre11*Δ and *sae2*Δ strains.

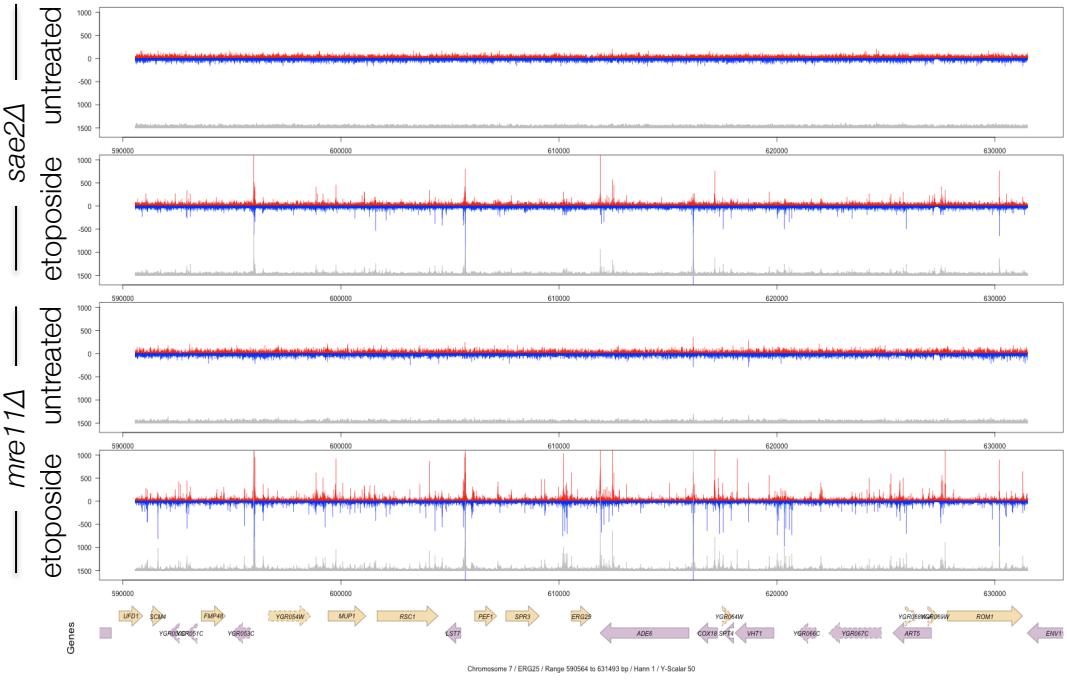
7.5 Top2-DSBs accumulate in NDRs but also occur in genic regions with an anti-correlation to nucleosome occupancy

Top2-DSBs form more in promoter regions than terminator regions (Figure 7.4B), suggesting a link between transcription initiation and Top2 activity. To determine whether there is any pattern of Top2-DSBs forming within the promoter region and the start of the transcribing gene/ORF, pile-ups of the 5′ Top2 hits in promoter regions, centred on the translation start site (TSS) and stretching 500 bp into the ORF, were made for wild type, *sae2*Δ and *mre11*Δ cells untreated and with etoposide (Figure 7.5A) (‘Pileups around TSS Topo stratify by transcription v02’ – Appendix 10.12). Top2 hits forming mainly within the promoter regions at

A. 200 bp region – 4 bp offset



B. 40kb region



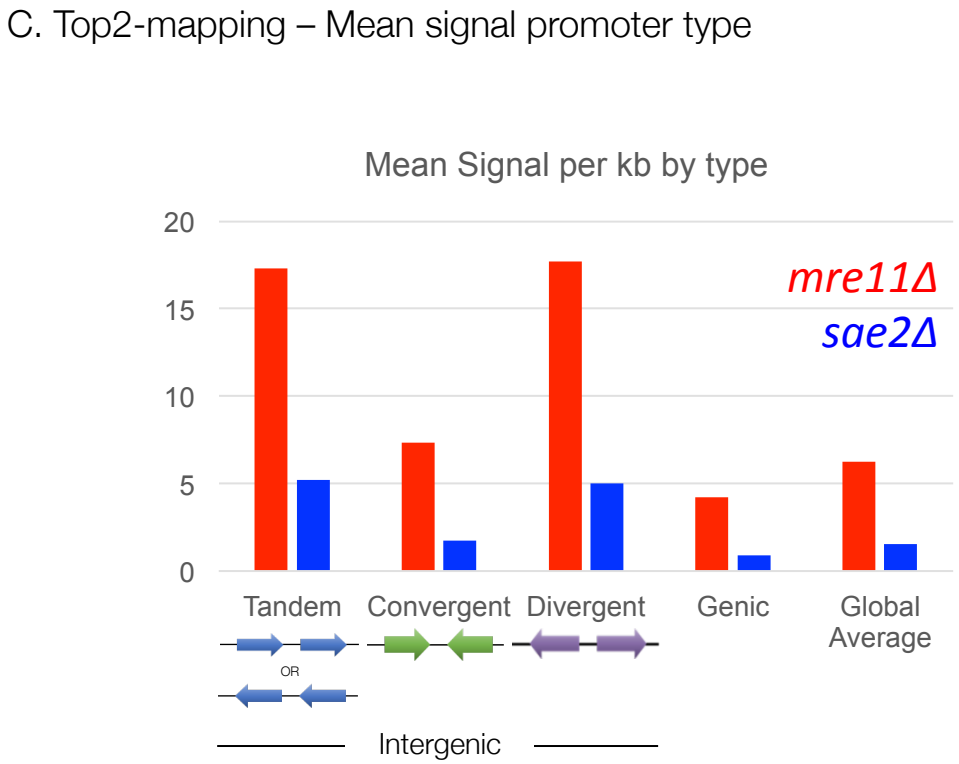
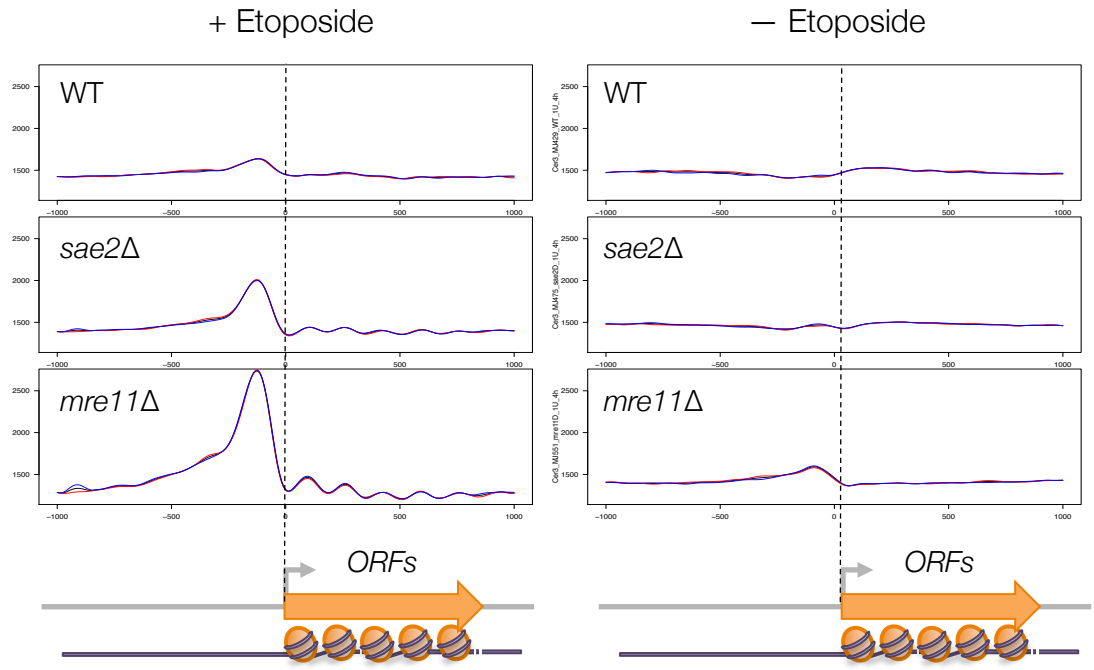
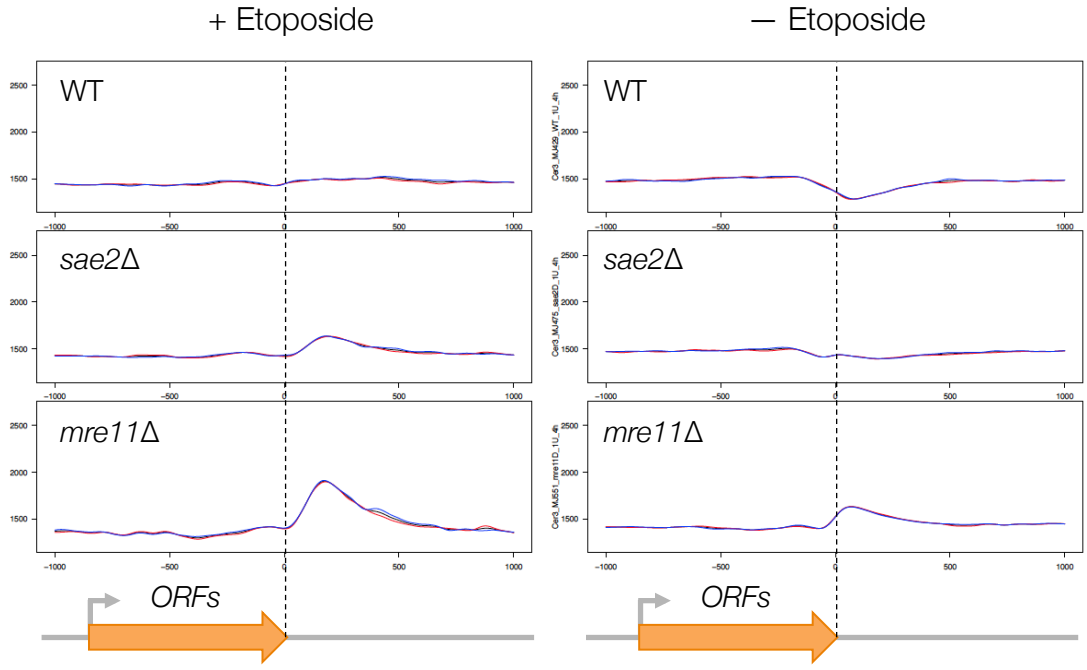


Figure 7.4: Top2-DSBs preferentially form within intergenic regions and display a 4 bp offset. Haploid wild type, *sae2*Δ and *mre11*Δ strains were grown overnight in YPD. Cultures were diluted to an OD_{600} of 0.5 and grown to an OD_{600} of 2.0. Cells were split into two 50 ml cultures and 1 µM etoposide ('etoposide') or equivalent volume of DMSO ('untreated') added. Cells were incubated for an additional 4 h and cells harvested by centrifugation. Next-generation sequencing libraries were prepared as stated in Figure 7.2. Sequences were aligned to the *S. cerevisiae* S288C genome using bowtie2. 5′ Read1 (Top2) ends were extracted an converted into a 1 bp histogram using R Studio. **A.** A 200 bp region from chromosome 8 was visualised to observed a 4 bp offset between Watson (red) and Crick (blue) Top2 hits. **B.** A 40 kb region was visualised to observed Top2-DSBs enriching at intergenic region with Watson hits (red), Crick hits (blue) and total hits (grey) displayed. **C.** The mean Top2 signal per kb was calculated for intergenic (split tandem, convergent and divergent) and genic regions of the genome and *mre11*Δ and *sae2*Δ data plotted. Script for mapping 'Topo mapping v01.R' – Appendix 10.11.

A. Pileups on START sites



B. Pileups on STOP sites



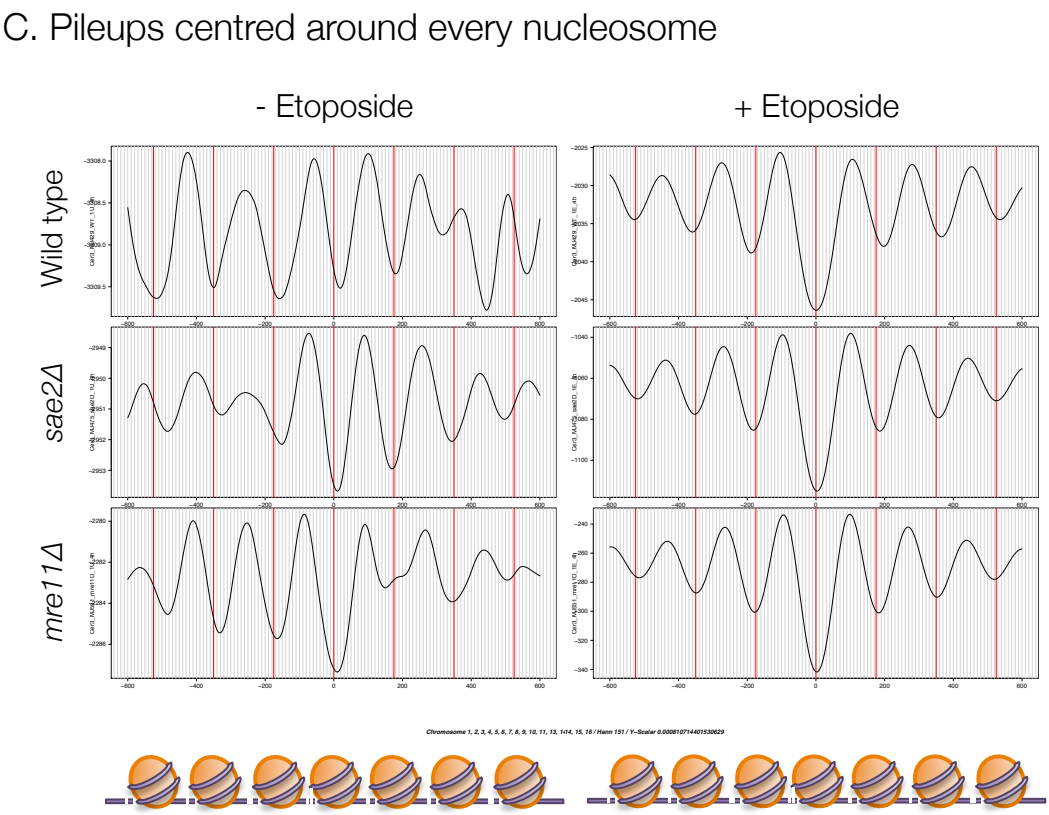


Figure 7.5: Top2-DSBs accumulate in NDRs, but also within genic regions, with an anti-correlation to nucleosome occupancy.
Next-generation sequencing libraries for wild type, *sae2*Δ and *mre11*Δ strains were prepared as stated in Figure 7.2. Sequences were aligned to the *S. cerevisiae* S288C genome using bowtie2. 5′ Read1 (Top2) ends were extracted an converted into a 1 bp histogram using R Studio. **A.** Top2 hits were piled up centred on the translation start site (TSS, at 0) +-1000 bp with Watson (red), Crick (blue) and averaged (black) hits annotated. **B.** Top2 hits were piled up centred on the translation stop site (at 0) +-1000 bp. **C.** Top2 hits were piled-up centred around every nucleosome in the genome +-600 bp. R Studio scripts: 'Pileups around TSS Topo stratify by transcription v02.R' – Appendix 10.12.

the TSS can be observed in wild type, *sae2*Δ and *mre11*Δ cells treated with etoposide. Enrichment at promoter regions can also be observed in *mre11*Δ untreated cells, suggesting that this observation is not an etoposide-dependent process but is amplified when Top2-complexes are poisoned. However, in wild type and *sae2*Δ untreated cells there is no enrichment, suggesting these lesions are not common, or are efficiently repaired in wild type cells by Mre11. These observations suggest that the role of Sae2 is not as important as Mre11 for Top2-DSB processing. To determine whether Top2-DSB enrichment is also observed around the translation stop sites, pile-ups were created at the terminal region of all the ORFs in wild type, *sae2*Δ and *mre11*Δ cells (Figure 7.5B). Enrichment of Top2 hits was seen at the terminal region in both *sae2*Δ and *mre11*Δ etoposide treated cells, but only observed in *mre11*Δ untreated cells - a similar situation to the TSS (Figure 7.5A). Both these observations suggest that Top2-DSBs are not common in wild type cells and only accumulate when DSB repair by Mre11 is abrogated (*mre11*Δ).

Intriguingly, in etoposide treated *sae2*Δ and *mre11*Δ cells repeating peaks with ~150 bp periodicity were observed stretching into the ORF/genic region from the TSS (Figure 7.5A). This 150 bp periodicity is reminiscent of MNase sensitive linker regions between nucleosomal DNA (Axel, 1975; Clark and Felsenfeld, 1971). To determine whether this periodicity observed is due to nucleosome free regions, smoothed pile-ups of Top2 hits were centred around every nucleosome location in the genome (Brogaard et al., 2012b) for wild type, *sae2*Δ and *mre11*Δ cells both untreated and etoposide treated (Figure 7.5C). An anti-correlation of Top2 hits with nucleosome position is observed in all strains. Untreated cells had a less of an anti-correlation but the repeating pattern is still observed, suggesting that even though anti-correlation was not observable for wild type and untreated cells in the TSS pile-up data, genome-wide Top2 hits generally form in linker regions and NDRs.

7.6 Top2 displays a sequence bias for generating DSBs

Mapping Top2-DSBs in the manner described generates nucleotide resolution maps of where Top2 cuts throughout the genome. To determine whether Top2 displays a sequence bias for cleavage, the 5′ (Top2) ends +-20 bp were aggregated, with any single 5′ site that occurred greater than five times (to remove any potential background effect) used to generate a sequence bias profile (Figure 7.6A). Recovery of both sides of a Top2-DSB only occurred in 10% of cases and produces a rotationally symmetrical cleavage sequence bias due to the Top2 dimer cleaving both the Watson and Crick strand. This consensus sequence for Top2 cleavage at DSB sites consist of biases of A at positions +8, +9 and +10, C at positions +1 and +2, G at position +3, and T at positions +4, +5 and +6. These biases are palindromic and therefore

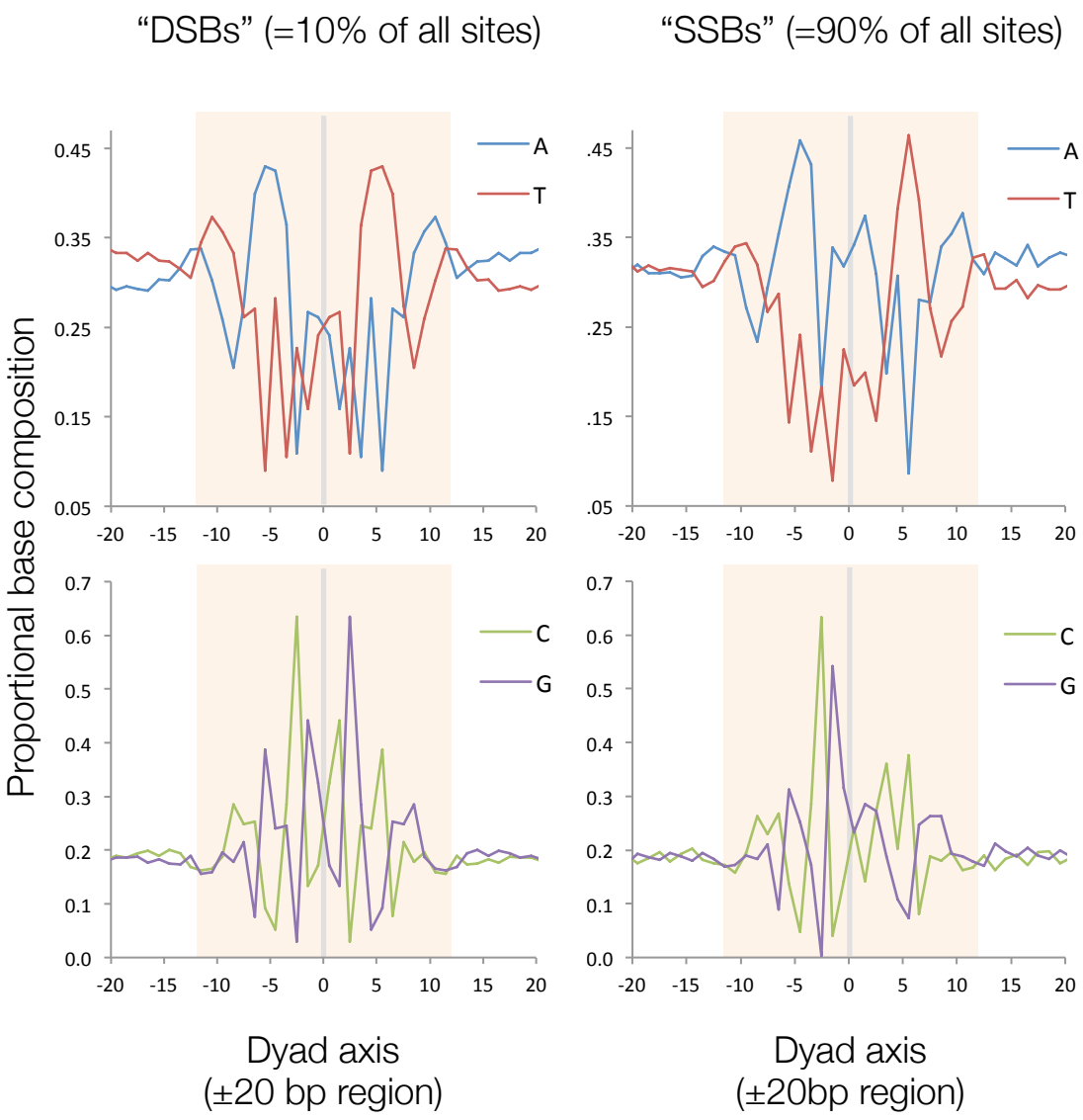


Figure 7.6: Top2 displays a sequence bias for generating DSBs.
Next-generation sequencing libraries were prepared as stated in Figure 7.2. Sequences were aligned to the *S. cerevisiae* S288C genome using bowtie2. 5′ Read1 (Top2) ends were extracted and converted into a 1 bp histogram using R Studio. The sequence of each 5′ read +/- 20 bp were aggregated for sites where Top2-DSBs were at least 5 to reduce the influence of non-specific signal. The sequence bias was plotted for A/T and G/C pairs, which are expected to show rotational symmetry consistent with Top2 cleaving DNA as a dimer. Top2 hits that contained a cognate pair (hit site on opposing strand 4 bp offset) were subset into DSBs. Top2 hits without a cognate pair were subset into SSBs.

contain the complementary purine/pyrimidine on the opposing strand (negative values). Overall, a weak palindromic consensus for Top2-DSBs can be read as GAAC*GG|CCGTTC (top-strand displayed only, dyad axis (|) indicated, Top2 cleavage site (*) would be reciprocated on the right hand side of the axis, bottom strand). In the other 90% of cases, only one side of the Top2-DSB is recovered and the rotational symmetry and palindromic sequence bias, is lost. This is indicative of single-stranded breaks (SSBs) occurring instead of DSBs. The high proportion of SSBs to DSBs agrees with previous mapping studies who attributed this observation to an etoposide-concentration dependent mechanism (Baranello et al., 2014). Etoposide acts by intercalating into the DNA within the cut site of the Top2 monomer and preventing religation (Pommier and Marchand, 2011; Wu et al., 2011). Etoposide acts independently on either Top2 monomer, suggesting a low concentration of etoposide would be more likely to only prevent religation of one monomer of Top2 and not both. Although this may account for these observations made (Figure 7.6B), a high concentration of etoposide was used (1 μM). Therefore, an alternative explanation may be that, due to DNA sequence preference, only one Top2 monomer of the dimer was able to cleave, with the preferred sequence not present on the other strand. This would result in only one monomer cleaving and therefore, when etoposide binds, only the one strand is blocked from religation and a SSB is formed.

7.7 Discussion

S. cerevisiae is not sensitive to low levels of etoposide due to active removal of the compound from the cell by the PDR network (Stepanov et al., 2008). Insertion of a sensitivity cassette, which removes *PDR1* activity and deletes *PDR1* itself, causes etoposide sensitivity of wild type cells (Figure 7.1). This construct has enabled the study of Top2 poisoning by etoposide and has enabled the enrichment of Top2-DSB fragments and nucleotide resolution, genome-wide mapping of Top2 cut sites via the methods established in Chapter 6.

In wild type cells without etoposide treatment very few Top2-DSBs are observed (Figure 7.3). Addition of etoposide causes an accumulation and amplification of Top2-DSB signal genome-wide (Figure 7.3). Deficiency in *MRE11* or *SAE2* causes an increase in Top2 signal with *mre11*Δ causing the most Top2 hits overall, suggesting that *MRE11* and *SAE2* are involved in the repair of etoposide-induced Top2-lesions. *MRE11* deficiency also causes an increase in Top2 hits at NDRs, in the absence of etoposide, however wild type and *sae2*Δ cells do not show this. This suggests that Top2-lesions occur in wild type cells and that Mre11 (and in part, Sae2) actively repairs any Top2-DSBs generated under normal conditions, preventing their accumulation and thus enrichment via this protocol. The locations of Top2 sites in untreated

and etoposide treated cells are consistent, although the levels are amplified with etoposide. This indicates that etoposide may simply be stabilising Top2-DSB complexes in locations it frequently cuts. Top2 displays a sequence bias for where it cleaves in the genome, however, even though Top2 acts as a dimer, symmetrically cleaved sites only account for 10% of all Top2 hits (Figure 7.6). The remainder are incidences where only one side of the DSB is recovered, suggesting this site is a single-stranded break (SSB). This high occurrence of SSBs could be etoposide-concentration dependent, inhibiting only one monomer as previously hypothesised (Baranello et al., 2014). By contrast, this high occurrence of SSBs could also be due to Top2 cutting a preferred DNA sequence that is only present on one strand at the majority of locations.

The majority of Top2 hits reside in nucleosome-depleted regions (NDRs) (Figure 7.3C). From these NDRs, a greater proportion of Top2 hits are located in promoter regions compared to terminator sites (Figure 7.3C), indicated by a large number occurring in tandem (one promoter) and divergent (two promoters) regions compared to convergent regions. For efficient transcription elongation to occur positive supercoils ahead of RNA polymerase and negative supercoils behind it need to be relaxed by topoisomerases (Pedersen et al., 2012; Sperling et al., 2011). Cells deficient in Top1 and Top2 have a global down-regulation of gene expression (Pedersen et al., 2012), suggesting Top2 plays an important role in promoting transcription. Additionally, the increased supercoiling generated by the transcriptional machinery may require relaxation by topoisomerases within the ORF of a transcribing gene. An access point for Top2 may therefore be at linker regions between nucleosomes generating the 150 bp wave observed when aggregating Top2 hits (Figure 7.5A).

Genome-wide mapping of *S. cerevisiae* nucleosomes demonstrates that promoters and sites of transcription termination are depleted in nucleosomes (Fan et al., 2010; Lantermann et al., 2010). Top2 hits were enriched at terminator regions (Figure 7.5B), supporting the link between NDRs and Top2 activity. However, terminator regions were less enriched than promoter regions (Figure 7.5). This could be linked to the requirement for topoisomerases to regulate supercoiling at promoter regions to aid transcription initiation (Pedersen et al., 2012). However, it may be due to nucleosome occupancy at these sites. Unlike promoter regions, nucleosome occupancy at terminator regions is strongly linked to growth conditions, with a reduction in transcription increasing the nucleosome occupancy (Fan et al., 2010). If the cycling cells mapped were nutrient-limited, then terminator regions would increase in nucleosome occupancy and thus decrease Top2 cleavage at these sites. Alternatively, a decrease in transcription in these cells decreases the supercoiling generated at the end of the ORF and the need for Top2 activity at terminator regions.

This work has confirmed that the procedure for genome-wide, high resolution mapping of Spo11 in a *sae2*Δ background can also be applied to map Top2-lesions in cycling cells. This suggests a wide potential for further investigation of Top2 function and other covalently bound proteins in *S. cerevisiae* and other organisms.

CHAPTER 8:

DISCUSSION

Chapter 8: Discussion

8.1 Summary

In this thesis I have presented work investigating the mechanisms that regulate the formation and repair of the protein-linked DSBs created by Spo11 and topoisomerase II (Top2). I have characterised the phenotype of Sae2 proteins mutated at putative CDK and Mec1/Tel1 phosphorylation sites **(Chapter 3)**, showed TDP2 is active upon the phosphotyrosine bond between Spo11 and DNA *in vitro* **(Chapter 4)**, and that removal of Spo11 in this manner permits resection by lambda exonuclease but no resection by the sole meiotic 5′ to 3′ exonuclease, Exo1 *in vitro*. Utilising TDP2 activity, I have developed a novel method to map Spo11-DSBs genome-wide with single nucleotide resolution **(Chapter 5)** and I used this procedure to further elucidate the regulation of Spo11 DSB formation by Tel1 **(Chapter 6)**, and also to map Top2 cleavage sites genome-wide **(Chapter 7)**.

8.2 Evolutionary pressures upon the use of the MRX complex and Sae2 for Spo11-DSB processing

The DSBs Spo11 generates throughout the genome during meiosis are utilised for homologue pairing via the homologous recombination repair pathway (Szostak et al., 1983). In *S. cerevisiae,* the initiation of Spo11-DSB repair is coordinated by the MRX complex (MRN in mammalian cells) and the Sae2 protein (CtIP in mammalian cells) (Mimitou and Symington, 2009; Pueyo et al., 1993). These two proteins nucleolytically process the Spo11-DSB end, clipping the 5′ strand adjacent to the DSB generating Spo11 covalently bound to the 5′ end of short oligonucleotides (Garcia et al., 2011; Neale et al., 2005). This process generates a short 3′ ssDNA overhang, a substrate favourable for Exo1 to bind to and further resect from in a 5′ to 3′ direction (Cannavo et al., 2013; Krogh and Symington, 2004). These long ssDNA tracts are important for efficient homology search and strand invasion, an essential step in the pairing of homologs during meiosis (San Filippo et al., 2008). Theoretically, Spo11 could also be removed via a different mechanism to the nucleolytic method. Directly hydrolysing the covalent bond formed between the Spo11 protein and the 5′ end of the DSB would remove the protein block, potentially permitting resection and thus homolog search. One such protein identified in humans, TDP2 (Cortes-Ledesma et al., 2009), contains a tyrosine phosphodiesterase capable of cleaving a range of 5′ phospho-tyrosine bonds *in vitro* (Gao et al., 2012), including the bond between Spo11 and the 5′ end of DSBs (Figures 4.3-4.8). However, here I have shown that *in vitro* removal of Spo11 in this manner does not permit

Exo1 resection of DSBs (Figures 4.5-4.8). These findings suggest that the mechanism for Spo11 removal by the MRX complex and Sae2, rather than a TDP2-like mechanism, has evolved to permit Exo1 resection and thus efficient HR to allow crossing over of homologues to enable gamete production and to produce genetic diversity. Another issue that may have arisen with a TDP2-like removal of Spo11 is the potential for repair via NHEJ. End-resection by the MRX complex and Sae2 promote HR and inhibit NHEJ through the production of ssDNA (Shibata et al., 2014). By contrast, repair by hydrolytic removal of Spo11 from the DSB end would generate a substrate capable of repair by NHEJ. This sort of repair is not proficient for the pairing homologs and thus needs to be avoided during meiosis. Although TDP2 is expressed in mouse and human testes tissue (Pype et al., 2000), as well as in the *C. elegans* germline (Shi et al., 2012), no role for TDP2 has been found during meiosis (Keith Caldecott – personal communication). Here, I have shown that expression of TDP2 during *S. cerevisiae* meiosis was not capable of removing Spo11 from the ends of DSBs *in vivo* (Figures 4.9-4.10), suggesting that TDP2 may not act during meiosis in any organism due to the evolutionary pressures for repair via the MRX/Sae2 pathway. In support of this conclusion, loss or mutation of MRN or CtIP in mammalian and *C. elegans* cells leads to unviable offspring even though, in *C. elegans,* a *COM-1*^{CtIP} mutant is capable of repair of meiotic DSBs via the NHEJ pathway (Lemmens et al., 2013). This suggests that there is no redundant pathway that can compensate for the Spo11-removal and HR promoting role of MRN/CtIP during meiosis (Buis et al., 2008; Chen et al., 2005).

Due to the importance of Sae2^{CtIP} during meiotic recombination there are numerous regulatory pathways that control activity. Sae2^{CtIP} is phosphorylated by CDK, a modification essential for Sae2 activity (Huertas et al., 2008; Wang et al., 2013). This phosphorylation event is regarded as the switch between HR and NHEJ in mitotic cells, with activation of Sae2 stimulating Mre11 endonuclease activity (Shibata et al., 2014). This nucleolytic event promotes HR and prevents NHEJ through generation of ssDNA (Shibata et al., 2014). Because CDK activity is linked to DNA replication, this ensures HR is only active when a sister chromatid template is available for repair (Aylon et al., 2004; Caspari et al., 2002; Esashi et al., 2005; Grzegorz et al., 2004). In meiosis, CDK is also active and therefore promotes HR through phosphorylation of Sae2 (Huertas et al., 2008; Marston and Amon, 2004). This is therefore another way by which the cell promotes HR repair of DSBs during meiosis to ensure efficient homolog pairing, and the formation of genetic variation.

8.3 Development of the protein-linked DNA mapping procedure (*sae2Δ* Spo11-DSB) has the potential for a wide range of other applications

The development of a nucleotide resolution mapping procedure for Spo11 not only improved our understanding of *S. cerevisiae* meiotic recombination but it has the potential to further our understanding of a variety of different pathways in multiple organisms. This procedure only requires the presence of a covalently bound protein to map where this protein forms this bond with the DNA *in vivo*. Therefore, this procedure cannot only be used to map Spo11 but as shown in Chapter 7, has the flexibility to map other covalently bound proteins such as topoisomerase-II. One caveat with mapping Spo11 in other species is that a *sae2*$^{CtIP}\Delta$ mutation, or another end-processing mutant such as a nuclease dead form of Mre11, is required to stabilise Spo11 at the 5′ ends of the DSB. Covalent Spo11-DSBs are transient complexes that are processed almost simultaneously with Spo11-DSB formation (Neale et al., 2005). Whilst, there is potential to identify stable Spo11-DSB complexes in a wild type background, this has yet to be conducted. This procedure can be utilised to map Spo11^{Rec12} in other organisms, such as *S. pombe*, where only oligonucleotides produced from end-processing by Mre11 and Ctp1 greater than 15 nucleotides in length could be mapped (Fowler et al., 2014). Creating a *ctp1*Δ or a *mre11-nd* mutation within this species would enable the mapping of Spo11^{Rec12} at nucleotide resolution, generating a map that would retain all the information about where Rec12 generated DSBs throughout the genome without the loss of information at the short oligonucleotide locations.

In the case of Top2, no mutation is required to stabilise the Top2-DNA complex. Instead, these complexes can be poisoned by drugs such as etoposide (Baldwin and Osheroff, 2005) (Figure 7.3). Although mutation of *MRE11* and *SAE2* further enriched for Top2-DSB complexes, mapping was possible in a wild type background with no affinity tags required, providing information on the function of Top2 without the potential disruption to its activity by affinity tags or mutations to downstream pathways (Chapter 7). Etoposide could be utilised to map Top2-DSB complexes within nucleotide resolution in a wide variety or organisms, including human cells. By monitoring where etoposide affects Top2 in humans, there is the potential to develop new drugs that may affect Top2 via different mechanisms, which may aid in the selective poisoning of Top2 in human cancer cells.

The procedure described in Chapter 5, relied on the human TDP2 protein to hydrolytically cleave the phospho-tyrosine bond to prevent any loss of nucleotides. Therefore, proteins that covalently link to the 3′ end of the DNA, such as Top1 (Chen et al. 2013), are not expected to be mapped by this method. However, TDP2 may contain some residual 3′ phosphodiesterase activity (Zeng et al., 2012). Alternatively, the *S. cerevisiae* Tdp1 protein, which contains 3′

and residual 5′ phosphodiesterase activity (Nitiss et al., 2006) or the human TDP1 3′ phosphodiesterase protein (Yang et al., 1996) could be utilised to map Top1 genome-wide within nucleotide resolution. Mutations in proteases, such as Wss1^{SPARTAN}, that degrade proteins covalently bound to the DNA (Stingele et al., 2015), could also be investigated for their effects on covalent DNA-protein binding genome-wide. Therefore, as long as there is a mechanism for the removal of the covalently bound protein from the DNA, this procedure has the ability to map numerous proteins in a wide-range of organisms without the requirement for affinity tags.

8.4 Tel1 kinase activity is responsible for hyper-localised repression of DSB formation

The Tel1 protein is responsible for DSB interference, a process of inhibiting DSBs on the same molecule in close proximity once a DSB is first formed (Garcia et al., 2015). To determine whether the Tel1 kinase activity was responsible for this action, a *TEL1* mutant was made that abolished the kinase activity of the protein. Whilst determination of DSB interference was not possible in the *tel1-kd* mutant, an interesting phenotype was observed. Spo11-DSB signal spread from the promoter hotspot in the direction of transcription (Figure 6.1). This spreading signal is attributed to a hyper-local loss of repression of DBS formation. This loss of repression resulted in the same molecule of DNA being cut multiple times by Spo11 within a small region (Figure 6.5), generating double-cut molecules, which are relatively short dsDNA molecules that contain Spo11 covalently bound to both 5′ ends. These molecules arose with a 10 bp periodicity (Figure 6.8), comparable to the 10 bp periodicity observed via the Spo11-oligo assay in Spo11 end-processing mutants such as *sae2*Δ (Figure 6.9). This periodicity may arise due to the structure of nucleosomal duplex DNA containing a DNase sensitive region every helical turn (10.5 nt), which is also potentially susceptible to cleavage by Spo11 (Figure 6.5). Studying the function of *TEL1* in *S. cerevisiae* may lead to further understanding of how the human ortholog *ATM* may function. Understanding the role of *ATM* is important due to its numerous roles in the cell. In particular, because mutation of *ATM* is linked to a variety of different disease states (van Os et al., 2016), understanding how it functions in the cell, both in mitotic and meiotic cells, may be important for the development of new therapeutics.

References

Adams, P.D., Sellers, W.R., Sharma, S.K., Wu, A.D., Nalin, C.M., Kaelin, W.G., 1996. Identification of a cyclin-cdk2 recognition motif present in substrates and p21-like cyclin-dependent kinase inhibitors. Mol. Cell. Biol. 16, 6623–33.

Ajimura, M., Leem, S.H., Ogawa, H., 1993. Identification of new genes required for meiotic recombination in Saccharomyces cerevisiae. Genetics 133, 51–66.

Alani, E., Padmore, R., Kleckner, N., 1990. Analysis of wild-type and rad50 mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. Cell 61, 419–436.

Alani, E., Subbiah, S., Kleckner, N., 1989. The yeast RAD50 gene encodes a predicted 153-kD protein containing a purine nucleotide-binding domain and two large heptad-repeat regions. Genetics 122, 47–57.

Allers, T., Lichten, M., 2001. Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106, 47–57.

Almouzni, G., Méchali, M., 1988. Assembly of spaced chromatin promoted by DNA synthesis in extracts from Xenopus eggs. EMBO J. 7, 665–72.

Anheim, M., Tranchant, C., Koenig, M., 2012. The Autosomal Recessive Cerebellar Ataxias. N. Engl. J. Med. 366, 636–646.

Aparicio, T., Baer, R., Gottesman, M., Gautier, J., 2016. MRN, CtIP, and BRCA1 mediate repair of topoisomerase II–DNA adducts. J. Cell Biol. 212.

Aplan, P.D., 2006. Causes of oncogenic chromosomal translocation. Trends Genet. 22, 46–55.

Arora, C., Kee, K., Maleki, S., Keeney, S., 2004. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. Mol. Cell 13, 549–559.

Axel, R., 1975. Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. Biochemistry 14, 2921–5.

Aylon, Y., Liefshitz, B., Kupiec, M., 2004. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. EMBO J. 23, 4868–4875.

Bakkenist, C.J., Kastan, M.B., 2003. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature 421, 499–506.

Baldo, V., Testoni, V., Lucchini, G., Longhese, M.P., 2008. Dominant TEL1-hy mutations

compensate for Mec1 lack of functions in the DNA damage response. Mol. Cell. Biol. 28, 358–75.

Baldwin, E.L., Osheroff, N., 2005. Etoposide, topoisomerase II and cancer. Curr. Med. Chem. Anticancer. Agents 5, 363–372.

Balzi, E., Chen, W., Ulaszewski, S., Capieaux, E., Goffeau, A., 1987. The multidrug resistance gene PDR1 from Saccharomyces cerevisiae. J. Biol. Chem. 262, 16871–9.

Balzi, E., Goffeau, A., 1991. Multiple or pleiotropic drug resistance in yeast. Biochim. Biophys. Acta 1073, 241–52.

Balzi, E., Wang, M., Leterme, S., Van Dyck, L., Goffeau, A., 1994. PDR5, a novel yeast multidrug resistance conferring transporter controlled by the transcription regulator PDR1. J. Biol. Chem. 269, 2206–14.

Baranello, L., Kouzine, F., Wojtowicz, D., Cui, K., Przytycka, T.M., Zhao, K., Levens, D., 2014. DNA break mapping reveals topoisomerase II activity genome-wide. Int. J. Mol. Sci. 15, 13111–13122.

Baroni, E., Viscardi, V., Cartagena-Lirola, H., Lucchini, G., Longhese, M.P., 2004. The functions of budding yeast Sae2 in the DNA damage response require Mec1- and Tel1-dependent phosphorylation. Mol. Cell. Biol. 24, 4151–4165.

Barral, Y., Jentsch, S., Mann, C., 1995. G1 cyclin turnover and nutrient uptake are controlled by a common pathway in yeast. Genes Dev. 9, 399–409.

Bartek, J., Lukas, J., 2007. DNA damage checkpoints: from initiation to recovery or adaptation. Curr. Opin. Cell Biol. 19, 238–245.

Baudat, F., Buard, J., Grey, C., de Massy, B., 2010. Identification d'une protéine-clé pour le contrôle des sites de recombinaison méiotique Prdm9, a key control of mammalian recombination hotspots. Med Sci 26, 468–470.

Baudat, F., Manova, K., Yuen, J.P., Jasin, M., Keeney, S., 2000. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Mol Cell 6, 989–998.

Baudat, F., Nicolas, A., 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. Proc. Natl. Acad. Sci. U. S. A. 94, 5213–8.

Bergerat, A, de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A, Forterre, P., 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature 386, 414–417.

Berkovich, E., Monnat, R.J., Kastan, M.B., 2007. Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. Nat. Cell Biol. 9, 683–690.

Bermejo, R., Doksani, Y., Capra, T., Katou, Y.-M., Tanaka, H., Shirahige, K., Foiani, M., 2007. Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. Genes Dev. 21, 1921–36.

Bishop, D.K., Nikolski, Y., Oshiro, J., Chon, J., Shinohara, M., Chen, X., 1999. High copy number suppression of the meiotic arrest caused by a dmc1 mutation: REC114 imposes an early recombination block and RAD54 promotes a DMC1-independent DSB repair pathway. Genes to Cells 4, 425–443.

Bishop, D.K., Park, D., Xu, L., Kleckner, N., 1992. DMC1: A meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. Cell 69, 439–456.

Bishop, D.K., Zickler, D., 2004. Early decision: Meiotic crossover interference prior to stable strand exchange and synapsis. Cell 117, 9–15.

Blat, Y., Kleckner, N., 1999. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. Cell 98, 249–259.

Blat, Y., Protacio, R.U., Hunter, N., Kleckner, N., 2002. Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. Cell 111, 791–802.

Bleuyard, J.Y., Gallego, M.E., White, C.I., 2004. Meiotic defects in the Arabidopsis rad50 mutant point to conservation of the MRX complex function in early stages of meiotic recombination. Chromosoma 113, 197–203.

Blitzblau, H.G., Bell, G.W., Rodriguez, J., Bell, S.P., Hochwagen, A., 2007. Mapping of Meiotic Single-Stranded DNA Reveals Double-Strand-Break Hotspots near Centromeres and Telomeres. Curr. Biol. 17, 2003–2012.

Blitzblau, H.G., Chan, C.S., Hochwagen, A., Bell, S.P., 2012. Separation of DNA replication from the assembly of break-competent meiotic chromosomes. PLoS Genet. 8, e1002643.

Borde, V., 2000. Direct Coupling Between Meiotic DNA Replication and Recombination Initiation. Science (80-.). 290, 806–809.

Borde, V., de Massy, B., 2013. Programmed induction of DNA double strand breaks during meiosis: Setting up communication between DNA and the chromosome structure. Curr. Opin. Genet. Dev. 23, 147–155.

Borde, V., Lin, W., Novikov, E., Petrini, J.H., Lichten, M., Nicolas, A., 2004. Association of Mre11p with Double-Strand Break Sites during Yeast Meiosis. Mol. Cell 13, 389–401.

Borde, V., Robine, N., Lin, W., Bonfils, S., Geli, V., Nicolas, A., 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. Embo J 28, 99–111.

Börner, G.V., Kleckner, N., Hunter, N., 2004. Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell 117, 29–45.

Bosco, E.E., Mayhew, C.N., Hennigan, R.F., Sage, J., Jacks, T., Knudsen, E.S., 2004. RB signaling prevents replication-dependent DNA double-strand breaks following genotoxic insult. Nucleic Acids Res. 32, 25–34.

Bressan, D.A., Baxter, B.K., Petrini, J.H., 1999. The Mre11-Rad50-Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in Saccharomyces cerevisiae. Mol. Cell. Biol. 19, 7681–7.

Brogaard, K., Xi, L., Wang, J.-P., Widom, J., 2012. A map of nucleosome positions in yeast at base-pair resolution. Nature 486, 496–501.

Bromberg, K.D., Burgin, A.B., Osheroff, N., 2003. A two-drug model for etoposide action against human topoisomerase IIα. J. Biol. Chem. 278, 7406–7412.

Budd, M.E., Campbell, J.L., 2009. Interplay of Mre11 nuclease with Dna2 plus Sgs1 in Rad51-dependent recombinational repair. PLoS One 4, e4267.

Buhler, C., Borde, V., Lichten, M., 2007. Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in Saccharomyces cerevisiae. PLoS Biol. 5,

Buis, J., Wu, Y., Deng, Y., Leddon, J., Westfield, G., Eckersdorff, M., Sekiguchi, J.M., Chang, S., Ferguson, D.O., 2008. Mre11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation. Cell 135, 85–96.

Burden, D.A., Osheroff, N., 1998. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. Biochim. Biophys. Acta 1400, 139–54.

Cannavo, E., Cejka, P., 2014. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature 514, 122–125.

Cannavo, E., Cejka, P., Kowalczykowski, S.C., 2013. Relationship of DNA degradation by Saccharomyces cerevisiae exonuclease 1 and its stimulation by RPA and Mre11-Rad50-Xrs2 to DNA end resection. Proc. Natl. Acad. Sci. U. S. A. 110, E1661-8.

Cao, L., Alani, E., Kleckner, N., 1990. A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. Cell 61, 1089–101.

Carballo, J.A., Johnson, A.L., Sedgwick, S.G., Cha, R.S., 2008. Phosphorylation of the Axial Element Protein Hop1 by Mec1/Tel1 Ensures Meiotic Interhomolog Recombination. Cell

132, 758–770.

Carballo, J.A., Panizza, S., Serrentino, M.E., Johnson, A.L., Geymonat, M., Borde, V., Klein, F., Cha, R.S., 2013. Budding Yeast ATM/ATR Control Meiotic Double-Strand Break (DSB) Levels by Down-Regulating Rec114, an Essential Component of the DSB-machinery. PLoS Genet. 9, e1003545.

Cartagena-Lirola, H., Guerini, I., Viscardi, V., Lucchini, G., Longhese, M.P., 2006. Budding yeast Sae2 is an in vivo target of the Mec1 and Tel1 checkpoint kinases during meiosis. Cell Cycle 5, 1549–1559.

Caspari, T., Murray, J.M., Carr, A.M., 2002. Cdc2-cyclin B kinase activity links Crb2 and Rqh1-topoisomerase III. Genes Dev. 16, 1195–1208.

Cassani, C., Gobbini, E., Wang, W., Niu, H., Clerici, M., Sung, P., Longhese, M.P., 2016. Tel1 and Rif2 Regulate MRX Functions in End-Tethering and Repair of DNA Double-Strand Breaks. PLoS Biol. 14, e1002387.

Cerrone, C.A., Tritschler, D.P.T., Bower, R., Yoder, B.K., 2003. A Novel Dynein Light Intermediate Chain Colocalizes with the Retrograde Motor for Intraflagellar Transport at Sites of Axoneme Assembly in Chlamydomonas and Mammalian Cells. Mol. Biol. Cell 14, 2559–2569.

Cejka, P., 2015. DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination. J. Biol. Chem. 290, 22931–8.

Cejka, P., Plank, J.L., Bachrati, C.Z., Hickson, I.D., Kowalczykowski, S.C., 2010. Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. Nat. Struct. Mol. Biol. 17, 1377–82.

Cha, R.S., Weiner, B.M., Keeney, S., Dekker, J., Kleckner, N., 2000. Progression of meiotic DNA replication is modulated by interchromosomal interaction proteins, negatively by Spo11p and positively by Rec8p. Genes Dev. 14, 493–503.

Chahwan, C., Nakamura, T.M., Sivakumar, S., Russell, P., Rhind, N., 2003. The fission yeast Rad32 (Mre11)-Rad50-Nbs1 complex is required for the S-phase DNA damage checkpoint. Mol. Cell. Biol. 23, 6564–73.

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

Champoux, J.J., 1977. Strand breakage by the DNA untwisting enzyme results in covalent attachment of the enzyme to DNA. Proc. Natl. Acad. Sci. U. S. A. 74, 3800–4.

Chen, H., Donnianni, R.A., Handa, N., Deng, S.K., Oh, J., Timashev, L.A., Kowalczykowski,

S.C., Symington, L.S., 2015. Sae2 promotes DNA damage resistance by removing the Mre11-Rad50-Xrs2 complex from DNA and attenuating Rad53 signaling. Proc Natl Acad Sci U S A 112, E1880-7.

Chen, H., Lisby, M., Symington, L., 2013. RPA Coordinates DNA End Resection and Prevents Formation of DNA Hairpins. Mol. Cell 50, 589–600.

Chen, L., Nievera, C.J., Lee, A.Y.-L., Wu, X., 2008. Cell Cycle-dependent Complex Formation of BRCA1.CtIP.MRN Is Important for DNA Double-strand Break Repair. J. Biol. Chem. 283, 7713–7720.

Chen, L., Trujillo, K., Ramos, W., Sung, P., Tomkinson, A.E., 2001. Promotion of Dnl4-Catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. Mol. Cell 8, 1105–1115.

Chen, L., Trujillo, K.M., Van Komen, S., Roh, D.H., Krejci, L., Lewis, L.K., Resnick, M.A., Sung, P., Tomkinson, A.E., 2005. Effect of amino acid substitutions in the rad50 ATP binding domain on DNA double strand break repair in yeast. J. Biol. Chem. 280, 2620–7.

Chen, P.L., Liu, F., Cai, S., Lin, X., Li, A., Chen, Y., Gu, B., Lee, E.Y.H.P., Lee, W.-H., 2005. Inactivation of CtIP leads to early embryonic lethality mediated by G1 restraint and to tumorigenesis by haploid insufficiency. Mol. Cell. Biol. 25, 3535–42.

Chen, S.H., Albuquerque, C.P., Liang, J., Suhandynata, R.T., Zhou, H., 2010. A proteome-wide analysis of kinase-substrate network in the DNA damage response. J. Biol. Chem. 285, 12803–12812.

Chen, S.H., Chan, N.L., Hsieh, T., 2013. New mechanistic and functional insights into DNA topoisomerases. Annu. Rev. Biochem. 82, 139–70.

Chen, S.Y., Tsubouchi, T., Rockmill, B., Sandler, J.S., Richards, D.R., Vader, G., Hochwagen, A., Roeder, G.S., Fung, J.C., 2008. Global Analysis of the Meiotic Crossover Landscape. Dev. Cell 15, 401–415.

Chen, X., Zhao, R., Glick, G.G., Cortez, D., 2007. Function of the ATR N-terminal domain revealed by an ATM/ATR chimera. Exp. Cell Res. 313, 1667–1674.

Chin, G.M., Villeneuve, A.M., 2001. C. elegans mre-11 is required for meiotic recombination and DNA repair but is dispensable for the meiotic G2 DNA damage checkpoint. Genes Dev. 15, 522–534.

Chung, W.H., Zhu, Z., Papusha, A., Malkova, A., Ira, G., 2010. Defective resection at DNA double-strand breaks leads to de novo telomere formation and enhances gene targeting. PLoS Genet. 6, e1000948.

Ciccia, A., Elledge, S.J., 2010. The DNA Damage Response: Making It Safe to Play with Knives. Mol. Cell 40, 179–204.

Clark, R.J., Felsenfeld, G., 1971. Structure of chromatin. Nat. New Biol. 229, 101–6.

Clerici, M., Mantiero, D., Guerini, I., Lucchini, G., Longhese, M.P., 2008. The Yku70-Yku80 complex contributes to regulate double-strand break processing and checkpoint activation during the cell cycle. EMBO Rep. 9, 810–818.

Clerici, M., Mantiero, D., Lucchini, G., Longhese, M.P., 2006. The Saccharomyces cerevisiae Sae2 protein negatively regulates DNA damage checkpoint signalling. EMBO Rep. 7, 212–218.

Clerici, M., Mantiero, D., Lucchini, G., Longhese, M.P., 2005. The Saccharomyces cerevisiae Sae2 protein promotes resection and bridging of double strand break ends. J. Biol. Chem. 280, 38631–38638.

Cloud, V., Chan, Y.-L., Grubb, J., Budke, B., Bishop, D.K., 2012. Rad51 Is an Accessory Factor for Dmc1-Mediated Joint Molecule Formation During Meiosis. Science (80-.). 337, 1222–1225.

Cockell, M., Rhodes, D., Klug, A., 1983. Location of the primary sites of micrococcal nuclease cleavage on the nucleosome core. J. Mol. Biol. 170, 423–46.

Cole, F., Keeney, S., Jasin, M., 2010. Evolutionary conservation of meiotic DSB proteins: More than just Spo11. Genes Dev. 24, 1201–1207.

Colleaux, L., D'Auriol, L., Galibert, F., Dujon, B., 1988. Recognition and cleavage site of the intron-encoded omega transposase. Proc. Natl. Acad. Sci. U. S. A. 85, 6022–6.

Cooper, T.J., Wardell, K., Garcia, V., Neale, M.J., 2014. Homeostatic regulation of meiotic DSB formation by ATM/ATR. Exp. Cell Res. 329, 124–131.

Corbett, A.H., Osheroff, N., 1993. When good enzymes go bad: conversion of topoisomerase II to a cellular toxin by antineoplastic drugs. Chem Res Toxicol 6, 585–597.

Cortes-Ledesma, F., El Khamisy, S.F., Zuma, M.C., Osborn, K., Caldecott, K.W., 2009. A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. Nature 461, 674–678.

Cortez, D., 2001. ATR and ATRIP: Partners in Checkpoint Signaling. Science (80-.). 294, 1713–1716.

Cruz-García, A., López-Saavedra, A., Huertas, P., 2014. BRCA1 Accelerates CtIP-Mediated DNA-End Resection, Cell Reports.

D'Amours, D., Jackson, S.P., 2002. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. Nat. Rev. Mol. Cell Biol. 3, 317–327.

D'Amours, D., Jackson, S.P., 2001. The yeast Xrs2 complex functions in S phase checkpoint regulation. Genes Dev. 15, 2238–2249.

Daley, J.M., Palmbos, P.L., Wu, D., Wilson, T.E., 2005. Nonhomologous end joining in yeast. Annu. Rev. Genet. 39, 431–51.

De Jager, M., Van Noort, J., Van Gent, D.C., Dekker, C., Kanaar, R., Wyman, C., 2001. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. Mol. Cell 8, 1129–1135.

de Massy, B., Rocco, V., Nicolas, A., 1995. The nucleotide mapping of DNA double-strand breaks at the CYS3 initiation site of meiotic recombination in Saccharomyces cerevisiae. Embo J 14, 4589–4598.

Dernburg, A.F., McDonald, K., Moulder, G., Barstead, R., Dresser, M., Villeneuve, A.M., 1998. Meiotic recombination in C. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. Cell 94, 387–398.

Deshpande, R.A., Williams, G.J., Limbo, O., Williams, R.S., Kuhnlein, J., Lee, J.H., Classen, S., Guenther, G., Russell, P., Tainer, J.A., Paull, T.T., 2014. ATP-driven Rad50 conformations regulate DNA tethering, end resection, and ATM checkpoint signaling. EMBO J. 33, 482–500.

Deveaux, Y., Alonso, B., Pierrugues, O., Godon, C., Kazmaier, M., 2000. Molecular cloning and developmental expression of AtGR1, a new growth-related Arabidopsis gene strongly induced by ionizing radiation. Radiat. Res. 154, 355–64.

Deweese, J.E., Osheroff, N., 2009. The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing. Nucleic Acids Res. 37, 738–748.

Di Virgilio, M., Ying, C.Y., Gautier, J., 2009. PIKK-dependent phosphorylation of Mre11 induces MRN complex inactivation by disassembly from chromatin. DNA Repair (Amst). 8, 1311–1320.

Diaz, R.L., Alcid, A.D., Berger, J.M., Keeney, S., 2002. Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. Mol. Cell. Biol. 22, 1106–15.

Dolganov, G.M., Maser, R.S., Novikov, A., Tosto, L., Chong, S., Bressan, D.A., Petrini, J.H., 1996. Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. Mol. Cell.

Biol. 16, 4832–4841.

Dong, K.C., Berger, J.M., 2007. Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. Nature 450, 1201–1205.

Dormeyer, W., 2005. Probing Lysine Acetylation in Proteins: Strategies, Limitations, and Pitfalls of in Vitro Acetyltransferase Assays. Mol. Cell. Proteomics 4, 1226–1239.

Dubin, M.J., Stokes, P.H., Sum, E.Y.M., Williams, R.S., Valova, V.A., Robinson, P.J., Lindeman, G.J., Glover, J.N.M., Visvader, J.E., Matthews, J.M., 2004. Dimerization of CtIP, a BRCA1- and CtBP-interacting protein, is mediated by an N-terminal coiled-coil motif. J. Biol. Chem. 279, 26932–26938.

Durand-Dubief, M., Persson, J., Norman, U., Hartsuiker, E., Ekwall, K., 2010. Topoisomerase I regulates open chromatin and controls gene expression in vivo. Embo J 29, 2126–2134.

Eggleston, A.K., West, S.C., 1997. Recombination initiation: Easy as A, B, C, D… χ? Curr. Biol. 7, R745–R749.

Ellis, N.A., Groden, J., Ye, T.Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M., German, J., 1995. The Bloom's syndrome gene product is homologous to RecQ helicases. Cell 83, 655–666.

Ensminger, M., Iloff, L., Ebel, C., Nikolova, T., Kaina, B., Löbrich, M., 2014. DNA breaks and chromosomal aberrations arise when replication meets base excision repair. J. Cell Biol. 206, 29–43.

Esashi, F., Christ, N., Gannon, J., Liu, Y., Hunt, T., Jasin, M., West, S.C., 2005. CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. Nature 434, 598–604.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V.A., Foiani, M., 2010. Replication Termination at Eukaryotic Chromosomes Is Mediated by Top2 and Occurs at Genomic Loci Containing Pausing Elements. Mol. Cell 39, 595–605.

Falck, J., Coates, J., Jackson, S.P., 2005. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. Nature 434, 605–611.

Fan, Q.Q., Petes, T.D., 1996. Relationship between nuclease-hypersensitive sites and meiotic recombination hot spot activity at the HIS4 locus of Saccharomyces cerevisiae. Mol. Cell. Biol. 16, 2037–43.

Fan, Q.Q., Xu, F., White, M.A., Petes, T.D., 1997. Competition between adjacent meiotic recombination hotspots in the yeast Saccharomyces cerevisiae. Genetics 145, 661–670.

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., Struhl, K., 2010. Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. Proc. Natl. Acad. Sci. 107, 17945–17950.

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R.A., Wigler, M., 1988. Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method. Mol. Cell. Biol. 8, 2159–65.

Fowler, K.R., Sasaki, M., Milman, N., Keeney, S., Smith, G.R., 2014. Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. Genome Res. 24, 1650–1664.

Franco, S., Alt, F.W., Manis, J.P., 2006. Pathways that suppress programmed DNA breaks from progressing to chromosomal breaks and translocations. DNA Repair (Amst). 5, 1030–1041.

Freese, E.B., Chu, M.I., Freese, E., 1982. Initiation of Yeast sporulation by partial carbon, nitrogen, or phosphate deprivation. J. Bacteriol. 149, 840–851.

Friedl, A.A., Kiechle, M., Fellerhoff, B., Eckardt-Schupp, F., 1998. Radiation-induced chromosome aberrations in Saccharomyces cerevisiae: Influence of DNA repair pathways. Genetics 148, 975–988.

Fu, Q., Chow, J., Bernstein, K. a, Makharashvili, N., Arora, S., Lee, C.-F., Person, M.D., Rothstein, R., Paull, T.T., 2014. Phosphorylation-regulated transitions in an oligomeric state control the activity of the Sae2 DNA repair enzyme. Mol. Cell. Biol. 34, 778–93.

Fukuda, T., Kugou, K., Sasanuma, H., Shibata, T., Ohta, K., 2008. Targeted induction of meiotic double-strand breaks reveals chromosomal domain-dependent regulation of Spo11 and interactions among potential sites of meiotic recombination. Nucleic Acids Res. 36, 984–997.

Fukunaga, K., Kwon, Y., Sung, P., Sugimoto, K., 2011. Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends. Mol. Cell. Biol. 31, 1959–1971.

Furstenthal, L., Kaiser, B.K., Swanson, C., Jackson, P.K., 2001. Cyclin E uses Cdc6 as a chromatin-associated receptor required for DNA replication. J. Cell Biol. 152, 1267–1278.

Gao, R., Huang, S.Y.N., Marchand, C., Pommier, Y., 2012. Biochemical characterization of human tyrosyl-DNA phosphodiesterase 2 (TDP2/TTRAP): A Mg2+/Mn2+-dependent phosphodiesterase specific for the repair of topoisomerase cleavage complexes. J. Biol. Chem. 287, 30842–30852.

Gao, R., Schellenberg, M.J., Huang, S. -y. N., Abdelmalak, M., Marchand, C., Nitiss, K.C., Nitiss, J.L., Williams, R.S., Pommier, Y., 2014. Proteolytic Degradation of Topoisomerase II (Top2) Enables the Processing of Top2.DNA and Top2.RNA Covalent Complexes by Tyrosyl-DNA-Phosphodiesterase 2 (TDP2). J. Biol. Chem. 289, 17960–17969.

Garcia, V., Gray, S., Allison, R.M., Cooper, T.J., Neale, M.J., 2015. Tel1(ATM)-mediated interference suppresses clustered meiotic double-strand-break formation. Nature 520, 114–118.

Garcia, V., Phelps, S.E.L., Gray, S., Neale, M.J., 2011. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 479, 241–4.

Gardiner, J.M., Bullard, S.A., Chrome, C., Malone, R.E., 1997. Molecular and genetic analysis of REC103, an early meiotic recombination gene in yeast. Genetics 146, 1265–1274.

Gatei, M., Jakob, B., Chen, P., Kijas, A.W., Becherel, O.J., Gueven, N., Birrell, G., Lee, J.H., Paull, T.T., Lerenthal, Y., Fazry, S., Taucher-Scholz, G., Kalb, R., Schindler, D., Waltes, R., Dörk, T., Lavin, M.F., 2011. ATM protein-dependent phosphorylation of Rad50 protein Regulates DNA repair and cell cycle control. J. Biol. Chem. 286, 31542–31556.

Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O., Petes, T.D., 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 97, 11383–11390.

Gobbini, E., Cesena, D., Galbiati, A., Lockhart, A., Longhese, M.P., 2013. Interplays between ATM/Tel1 and ATR/Mec1 in sensing and signaling DNA double-strand breaks. DNA Repair (Amst). 12, 791–799.

Goldfarb, T., Lichten, M., 2010. Frequent and efficient use of the sister chromatid for DNA double-strand break repair during budding yeast meiosis. PLoS Biol. 8, e1000520.

Gómez-Herreros, F., Romero-Granados, R., Zeng, Z., Álvarez-Quilón, A., Quintero, C., Ju, L., Umans, L., Vermeire, L., Huylebroeck, D., Caldecott, K.W., Cortés-Ledesma, F., Mittelman, D., Pereira-Smith, O., Wilson, J., Gorbunova, V., 2013. TDP2–Dependent Non-Homologous End-Joining Protects against Topoisomerase II–Induced DNA Breaks and Genome Instability in Cells and In Vivo. PLoS Genet. 9, e1003226.

Gönczy, P., Schnabel, H., Kaletta, T., Amores, A.D., Hyman, T., Schnabel, R., 1999. Dissection of cell division processes in the one cell stage Caenorhabditis elegans embryo by mutational analysis. J. Cell Biol. 144, 927–46.

Gray, S., Allison, R.M., Garcia, V., Goldman, A.S.H., Neale, M.J., 2013. Positive regulation of meiotic DNA double-strand break formation by activation of the DNA damage

checkpoint kinase Mec1(ATR). Open Biol. 3, 130019.

Greenwell, P.W., Kronmal, S.L., Porter, S.E., Gassenhuber, J., Obermaier, B., Petes, T.D., 1995. TEL1, a gene involved in controlling telomere length in S. cerevisiae, is homologous to the human ataxia telangiectasia gene. Cell 82, 823–829.

Grenon, M., Gilbert, C., Lowndes, N.F., 2001. Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. Nat. Cell Biol. 3, 844–847.

Grzegorz, I., Pellicioli, A., Balijja, A., Wang, X., Fiorani, S., Carotenuto, W., Liberi, G., Bressan, D., Wan, L., Hollingsworth, N.M., Haber, J.E., Foiani, M., 2004. DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. Nature 431, 1011–1017.

Guttmann-Raviv, N., Boger-Nadjar, E., Edri, I., Kassir, Y., 2001. Cdc28 and ime2 possess redundant functions in promoting entry into premeiotic DNA replication in Saccharomyces cerevisiae. Genetics 159, 1547–1558.

Haber, J.E., Thorburn, P.C., Rogers, D., 1984. Meiotic and mitotic behavior of dicentric chromosomes in Saccharomyces cerevisiae. Genetics 106, 185–205.

Hartsuiker, E., Mizuno, K., Molnar, M., Kohli, J., Ohta, K., Carr, A.M., 2009a. Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions. Mol. Cell. Biol. 29, 1671–81.

Hartsuiker, E., Neale, M., Carr, A., 2009b. Distinct Requirements for the Rad32 Mre11 Nuclease and Ctp1 CtIP in the Removal of Covalently Bound Topoisomerase I and II. Mol. Cell 33, 117–123.

Henderson, K.A., Kee, K., Maleki, S., Santini, P.A., Keeney, S., 2006. Cyclin-Dependent Kinase Directly Regulates Initiation of Meiotic Recombination. Cell 125, 1321–1332.

Henderson, K. a, Keeney, S., 2004. Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks. Proc. Natl. Acad. Sci. U. S. A. 101, 4519–4524.

Hentges, P., Waller, H., Reis, C.C., Ferreira, M., Doherty, A.J., 2014. Cdk1 Restrains NHEJ through Phosphorylation of XRCC4-like Factor Xlf1, Cell Reports.

Hirano, Y., Fukunaga, K., Sugimoto, K., 2009. Rif1 and Rif2 Inhibit Localization of Tel1 to DNA Ends. Mol. Cell 33, 312–322.

Hodgson, A., Terentyev, Y., Johnson, R.A., Bishop-Bailey, A., Angevin, T., Croucher, A., Goldman, A.S.H., 2011. Mre11 and Exo1 contribute to the initiation and processivity of

resection at meiotic double-strand breaks made independently of Spo11. DNA Repair (Amst). 10, 138–148.

Hollingsworth, N.M., Goetsch, L., Byers, B., 1990. The HOP1 gene encodes a meiosis-specific component of yeast chromosomes. Cell 61, 73–84.

Holt, L.J., Hutti, J.E., Cantley, L.C., Morgan, D.O., 2007. Evolution of Ime2 Phosphorylation Sites on Cdk1 Substrates Provides a Mechanism to Limit the Effects of the Phosphatase Cdc14 in Meiosis. Mol. Cell 25, 689–702.

Hopfner, K.-P., Craig, L., Moncalian, G., Zinkel, R. a, Usui, T., Owen, B. a L., Karcher, A., Henderson, B., Bodmer, J.-L., McMurray, C.T., Carney, J.P., Petrini, J.H.J., Tainer, J.A., 2002. The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. Nature 418, 562–566.

Hopfner, K.P., Karcher, A., Shin, D., Fairley, C., Tainer, J.A., Carney, J.P., 2000. Mre11 and Rad50 from Pyrococcus furiosus: Cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine. J. Bacteriol. 182, 6036–6041.

Hopkins, B.B., Paull, T.T., 2008. The P. furiosus Mre11/Rad50 Complex Promotes 5??? Strand Resection at a DNA Double-Strand Break. Cell 135, 250–260.

Huertas, P., Cortés-Ledesma, F., Sartori, A., 2008. CDK targets Sae2 to control DNA-end resection and homologous recombination. Nature 455, 689–692.

Hunter, N., Kleckner, N., 2001. The single-end invasion: An asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. Cell 106, 59–70.

Ivanov, E.L., Sugawara, N., White, C.I., Fabre, F., Haber, J.E., 1994. Mutations in XRS2 and RAD50 delay but do not prevent mating-type switching in Saccharomyces cerevisiae. Mol. Cell. Biol. 14, 3414–3425.

Jackson, J.A., Fink, G.R., 1985. Meiotic recombination between duplicated genetic elements in Saccharomyces cerevisiae. Genetics 109, 303–332.

Jessop, L., Rockmill, B., Roeder, G.S., Lichten, M., 2006. Meiotic chromosome synapsis-promoting proteins antagonize the anti-crossover activity of sgs1. PLoS Genet. 2, 1402–1412.

Jin, Y., Binkowski, G., Simon, L.D., Norris, D., 1997. Ho endonuclease cleaves MAT DNA in vitro by an inefficient stoichiometric reaction mechanism. J. Biol. Chem. 272, 7352–7359.

Johnson, R.D., Jasin, M., 2001. Double-strand-break-induced homologous recombination in

mammalian cells. Biochem. Soc. Trans. 29, 196–201.

Johzuka, K., Ogawa, H., 1995. Interaction of Mre11 and Rad50: Two proteins required for DNA repair and meiosis-specific double-strand break formation in Saccharomyces cerevisiae. Genetics 139, 1521–1532.

Jones, G.H., 1967. The control of chiasma distribution in rye. Chromosoma 22, 69–90.

Jones, G.H., Franklin, F.C.H., 2006. Meiotic Crossing-over: Obligation and Interference. Cell 126, 246–248.

Oh, J., Al-Zain, A., Cannavo, E., Cejka, P., and Symington, L., 2016. Xrs2 dependent and independent functions of the Mre11-Rad50 complex. Mol. Cell.

Kadyk, L.C., Hartwell, L.H., 1992. Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. Genetics 132, 387–402.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Leproust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., Segal, E., 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458, 362–366.

Kastan, M.B., Bartek, J., 2004. Cell-cycle checkpoints and cancer. Nature 432, 316–323.

Kato, R., Ogawa, H., 1994. An essential gene, ESR1, is required for mitotic growth, DNA repair and meiotic recombination saccharomyces cerevisiae. Nucleic Acids Res. 22, 3104–3112.

Kauppi, L., Jasin, M., Keeney, S., 2013. How much is enough? Control of DNA double-strand break numbers in mouse meiosis. Cell Cycle 12, 2719–2720.

Kee, K., Protacio, R.U., Arora, C., Keeney, S., 2004. Spatial organization and dynamics of the association of Rec102 and Rec104 with meiotic chromosomes. EMBO J. 23, 1815–1824.

Keelagher, R.E., Cotton, V.E., Goldman, A.S.H., Borts, R.H., 2011. Separable roles for Exonuclease I in meiotic DNA double-strand break repair. DNA Repair (Amst). 10, 126–137.

Keeney, S., 2008. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. Genome Dyn Stab. 1, 81–123.

Keeney, S., Giroux, C.N., Kleckner, N., 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88, 375–384.

Keeney, S., Kleckner, N., 1995. Covalent protein-DNA complexes at the 5' strand termini of meiosis-specific double-strand breaks in yeast. Proc. Natl. Acad. Sci. U. S. A. 92, 11274–

Keeney, S., Lange, J., Mohibullah, N., 2014. Self-organization of meiotic recombination initiation: general principles and molecular pathways. Annu. Rev. Genet. 48, 187–214.

Keeney, S., Neale, M.J., 2006. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. Biochem. Soc. Trans. 34, 523–5.

Kevin Lewis, L., Karthikeyan, G., Westmoreland, J.W., Resnick, M.A., 2002. Differential suppression of DNA repair deficiencies of yeast rad50, mre11 and xrs2 mutants by EXO1 and TLC1 (the RNA component of telomerase). Genetics 160, 49–62.

Kijas, A.W., Lim, Y.C., Bolderson, E., Cerosaletti, K., Gatei, M., Jakob, B., Tobias, F., Taucher-Scholz, G., Gueven, N., Oakley, G., Concannon, P., Wolvetang, E., Khanna, K.K., Wiesm??ller, L., Lavin, M.F., 2015. ATM-dependent phosphorylation of MRE11 controls extent of resection during homology directed repair by signalling through Exonuclease 1. Nucleic Acids Res. 43, 8352–8367.

Kim, H.-S., Vijayakumar, S., Reger, M., Harrison, J.C., Haber, J.E., Weil, C., Petrini, J.H.J., 2008. Functional interactions between Sae2 and the Mre11 complex. Genetics 178, 711–23.

Kim, K.P., Weiner, B.M., Zhang, L., Jordan, A., Dekker, J., Kleckner, N., 2010. Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. Cell 143, 924–937.

Kim, S.T., Lim, D.S., Canman, C.E., Kastan, M.B., 1999. Substrate specificities and identification of putative substrates of ATM kinase family members. J. Biol. Chem. 274, 37538–37543.

Klapholz, S., Waddell, C.S., Easton Esposito, R., 1985. The role of the SPO11 gene in meiotic recombination in yeast. Genetics 110, 187–216.

Kleckner, N., 2006. Chiasma formation: Chromatin/axis interplay and the role(s) of the synaptonemal complex. Chromosoma 115, 175–194.

Klein, F., Mahr, P., Galova, M., Buonomo, S.B.C., Michaelis, C., Nairz, K., Nasmyth, K., 1999. A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell 98, 91–103.

Krogh, B.O., Llorente, B., Lam, A., Symington, L.S., 2005. Mutations in Mre11 phosphoesterase motif I that impair Saccharomyces cerevisiae Mre11-Rad50-Xrs2 complex stability in addition to nuclease activity. Genetics 171, 1561–70.

Krogh, B.O., Symington, L.S., 2004. Recombination proteins in yeast. Annu. Rev. Genet. 38, 233–71.

Kuzminov, A., 1999. Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. Microbiol. Mol. Biol. Rev. 63, 751–813,

Lam, I., Keeney, S., 2015. Mechanism and regulation of meiotic recombination initiation, Cold Spring Harbor Perspectives in Biology.

Lammens, K., Bemeleit, D.J., Möckel, C., Clausing, E., Schele, A., Hartung, S., Schiller, C.B., Lucas, M., Angermüller, C., Söding, J., Sträßer, K., Hopfner, K.P., 2011. The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair. Cell 145, 54–66.

Langerak, P., Russell, P., 2011. Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366, 3562–71.

Lantermann, A.B., Straub, T., Strålfors, A., Yuan, G.-C., Ekwall, K., Korber, P., 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. 17, 251–257.

Lavin, M., Kozlov, S., Gatei, M., Kijas, A., 2015. ATM-Dependent Phosphorylation of All Three Members of the MRN Complex: From Sensor to Adaptor. Biomolecules 5, 2877–2902.

Lee, J., Paull, T.T., 2005. ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex. Science (80-.). 308, 551–554.

Lee, K., Zhang, Y., Lee, S.E., 2008. Saccharomyces cerevisiae ATM orthologue suppresses break-induced chromosome translocations. Nature 454, 543–6.

Lee, S.E., Moore, J.K., Holmes, A., Umezu, K., Kolodner, R.D., Haber, J.E., 1998. Saccharomyces Ku70, Mre11/Rad50, and RPA proteins regulate adaptation to G2/M arrest after DNA damage. Cell 94, 399–409.

Lemmens, B.B.L.G., Johnson, N.M., Tijsterman, M., 2013. COM-1 Promotes Homologous Recombination during Caenorhabditis elegans Meiosis by Antagonizing Ku-Mediated Non-Homologous End Joining. PLoS Genet. 9, e1003276.

Lengsfeld, B.M., Rattray, A.J., Bhaskara, V., Ghirlando, R., Paull, T.T., 2007. Sae2 is an Endonuclease that Processes Hairpin DNA Cooperatively with the Mre11/Rad50/Xrs2 Complex. Mol. Cell 28, 638–651.

Leu, J.Y., Roeder, G.S., 1999. The pachytene checkpoint in S. cerevisiae depends on Swe1-mediated phosphorylation of the cyclin-dependent kinase Cdc28. Mol. Cell 4, 805–14.

Li, J., Hooker, G.W., Roeder, G.S., 2006. Saccharomyces cerevisiae Mer2, Mei4 and Rec114 form a complex required for meiotic double-strand break formation. Genetics 173, 1969–1981.

Liang, J., Suhandynata, R.T., Zhou, H., 2015. Phosphorylation of Sae2 mediates Forkhead-associated (FHA) domain-specific interaction and regulates its DNA repair function. J. Biol. Chem. 290, 10751–10763.

Lichten, M., Goldman, A.S.H., 1995. Meiotic Recombination Hotspots. Annu. Rev. Genet. 29, 423–444.

Limbo, O., Chahwan, C., Yamada, Y., 2007. Ctp1 is a cell cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. … cell 28, 134–146.

Limoli, C.L., Giedzinski, E., Bonner, W.M., Cleaver, J.E., 2002. UV-induced replication arrest in the xeroderma pigmentosum variant leads to DNA double-strand breaks, gamma - H2AX formation, and Mre11 relocalization. Proc. Natl. Acad. Sci. U. S. A. 99, 233–238.

Lisby, M., Barlow, J.H., Burgess, R.C., Rothstein, R., 2004. Choreography of the DNA damage response: Spatiotemporal relationships among checkpoint and repair proteins. Cell 118, 699–713.

Liu, J., Wu, T.C., Lichten, M., 1995. The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. Embo J 14, 4599–608.

Llorente, B., Symington, L.S., 2004. The Mre11 nuclease is not required for 5' to 3' resection at multiple HO-induced double-strand breaks. Mol. Cell. Biol. 24, 9682–9694.

Lobachev, K., Vitriol, E., Stemple, J., Resnick, M.A., Bloom, K., 2004. Chromosome fragmentation after induction of a double-strand break is an active process prevented by the RMX repair complex. Curr. Biol. 14, 2107–2112.

Lobachev, K.S., Gordenin, D. a, Resnick, M. a, 2002. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. Cell 108, 183–93.

Lodish, H., Berk, A., 2003. Molecular cell biology.

Longhese, M.P., Bonetti, D., Manfrini, N., Clerici, M., 2010. Mechanisms and regulation of DNA end resection. Embo J. 29, 2864–2874.

Louvion, J.F., Havaux-Copf, B., Picard, D., 1993. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast,

Gene.

Lundin, C., North, M., Erixon, K., Walters, K., Jenssen, D., Goldman, A.S.H., Helleday, T., 2005. Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. Nucleic Acids Res. 33, 3799–3811.

Luo, G., Santoro, I.M., McDaniel, L.D., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R. a, Bradley, a, 2000. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. Nat. Genet. 26, 424–429.

Luo, M., Yang, F., Leu, N.A., Landaiche, J., Handel, M.A., Benavente, R., La Salle, S., Wang, P.J., 2013. MEIOB exhibits single-stranded DNA-binding and exonuclease activities and is essential for meiotic recombination. Nat. Commun. 4, 2788.

Lydall, D., Nikolsky, Y., Bishop, D.K., Weinert, T., 1996. A meiotic recombination checkpoint controlled by mitotic checkpoint genes. Nature 383, 840–843.

Moens, P.B., Pearlman, R.E., 1988. Chromatin organization at meiosis. BioEssays 9, 151–153.

M, K.S., Hayashi-Hagihara, A., 1998. mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: Evidence that the mechanism for initiating meiotic recombination is conserved. Genes Dev. 12, 2932–2942.

Ma, Y., Greider, C.W., 2009. Kinase-independent functions of TEL1 in telomere maintenance. Mol. Cell. Biol. 29, 5193–202.

Mains, P.E., Sulston, I.A., Wood, W.B., 1990. Dominant maternal-effect mutations causing embryonic lethality in Caenorhabditis elegans. Genetics 125, 351–369.

Maleki, S., Neale, M.J., Arora, C., Henderson, K.A., Keeney, S., 2007. Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in Saccharomyces cerevisiae. Chromosoma 116, 471–486.

Malik, M., Nitiss, J.L., 2004. DNA Repair Functions That Control Sensitivity to Topoisomerase-Targeting Drugs. Eukaryot. Cell 3, 82–90.

Malkova, A., Ross, L., Dawson, D., Hoekstra, M.F., Haber, J.E., 1996. Meiotic recombination initiated by a double-strand break in rad50Δ yeast cells otherwise unable to initiate meiotic recombination. Genetics 143, 741–754.

Mallory, J.C., Petes, T.D., 2000. Protein kinase activity of Tel1p and Mec1p, two Saccharomyces cerevisiae proteins related to the human ATM protein kinase. Proc. Natl. Acad. Sci. 97, 13749–54.

Malumbres, M., 2014. Cyclin-dependent kinases. Genome Biol. 15, 122.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., Steinmetz, L.M., 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454, 479–485.

Manfrini, N., Guerini, I., Citterio, A., Lucchini, G., Longhese, M.P., 2010. Processing of meiotic DNA double strand breaks requires cyclin-dependent kinase and multiple nucleases. J. Biol. Chem. 285, 11628–11637.

Mantiero, D., Clerici, M., Lucchini, G., Longhese, M.P., 2007. Dual role for Saccharomyces cerevisiae Tel1 in the checkpoint response to double-strand breaks. EMBO Rep. 8, 380–387.

Mao, Y., Desai, S.D., Ting, C.Y., Hwang, J., Liu, L.F., 2001. 26 S Proteasome-mediated Degradation of Topoisomerase II Cleavable Complexes. J. Biol. Chem. 276, 40652–40658.

Mari, P.-O., Florea, B.I., Persengiev, S.P., Verkaik, N.S., Bruggenwirth, H.T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J.A.A., Luider, T.M., Houtsmuller, A.B., van Gent, D.C., 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. Proc. Natl. Acad. Sci. 103, 18597–18602.

Marston, A.L., Amon, A., 2004. Meiosis: cell-cycle controls shuffle and deal. Nat. Rev. Mol. Cell Biol. 5, 983–997.

Mårtensson, S., Nygren, J., Osheroff, N., Hammarsten, O., 2003. Activation of the DNA-dependent protein kinase by drug-induced and radiation-induced DNA strand breaks. Radiat. Res. 160, 291–301.

Martina, M., Clerici, M., Baldo, V., Bonetti, D., Lucchini, G., Longhese, M.P., 2012. A balance between Tel1 and Rif2 activities regulates nucleolytic processing and elongation at telomeres. Mol Cell Biol 32, 1604–1617.

Martinez-Perez, E., Colaiácovo, M.P., 2009. Distribution of meiotic recombination events: talking to your neighbors. Curr. Opin. Genet. Dev. 19, 105–112.

Martini, E., Diaz, R.L., Hunter, N., Keeney, S., 2006. Crossover Homeostasis in Yeast Meiosis. Cell 126, 285–295.

Masison, D.C., Blanc, A., Ribas, J.C., Carroll, K., Sonenberg, N., Wickner, R.B., 1995. Decoying the cap- mRNA degradation system by a double-stranded RNA virus and poly(A)- mRNA surveillance by a yeast antiviral system. Mol. Cell. Biol. 15, 2763–71.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S.P., Elledge, S.J., 2007. ATM and ATR substrate analysis reveals extensive protein networks

responsive to DNA damage. Science (80-.). 316, 1160–1166.

McClendon, A.K., Osheroff, N., 2007. DNA topoisomerase II, genotoxicity, and cancer. Mutat. Res. Mol. Mech. Mutagen. 623, 83–97.

McKee, A., Kleckner, N., 1997. A General Method for Identifying Recessive Diploid-Specific Mutations Saccharomyces cerevisiae, Its Application to the Isolation of Mutants Blocked at Intermediate Stages of Meiotic Prophase and Characterization of a New Gene SAE2. Genetics 146, 797–816.

Mendenhall, M.D., Hodge, a E., 1998. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. 62, 1191–1243.

Merino, S.T., Cummings, W.J., Acharya, S.N., Zolan, M.E., 2000. Replication-dependent early meiotic requirement for Spo11 and Rad50. Proc. Natl. Acad. Sci. U. S. A. 97, 10477–82.

Milman, N., Higuchi, E., Smith, G.R., 2009. Meiotic DNA double-strand break repair requires two nucleases, MRN and Ctp1, to produce a single size class of Rec12 (Spo11)-oligonucleotide complexes. Mol. Cell. Biol. 29, 5998–6005.

Mimitou, E.P., Symington, L.S., 2011. DNA end resection--unraveling the tail. DNA Repair (Amst). 10, 344–8.

Mimitou, E.P., Symington, L.S., 2010. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. EMBO J. 29, 3358–69.

Mimitou, E.P., Symington, L.S., 2009. DNA end resection: many nucleases make light work. DNA Repair (Amst). 8, 983–95.

Mimitou, E.P., Symington, L.S., 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature 455, 770–4.

Mohibullah, N., Keeney, S., 2016. Numerical and spatial patterning of yeast meiotic DNA breaks by Tel1. Genome Res. gr.213587.116.

Mondal, N., Zhang, Y., Jonsson, Z., Dhar, S.K., Kannapiran, M., Parvin, J.D., 2003. Elongation by RNA polymerase II on chromatin templates requires topoisomerase activity. Nucleic Acids Res. 31, 5016–5024.

Moreau, S., Ferguson, J.R., Symington, L.S., 1999. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. Mol. Cell. Biol. 19, 556–566.

Moreau, S., Morgan, E., Symington, L., 2001. Overlapping functions of the Saccharomyces

cerevisiae Mre11, Exo1 and Rad27 nucleases in DNA metabolism. Genetics 159, 1423–1433.

Morin, I., Ngo, H.-P., Greenall, A., Zubko, M.K., Morrice, N., Lydall, D., 2008. Checkpoint-dependent phosphorylation of Exo1 modulates the DNA damage response. EMBO J. 27, 2400–10.

Morris, D.R., Geballe, A.P., 2000. Upstream Open Reading Frames as Regulators of mRNA Translation MINIREVIEW Upstream Open Reading Frames as Regulators of mRNA Translation. Mol. Cell. Biol. 20, 8635–42.

Morrow, D.M., Tagle, D.A., Shiloh, Y., Collins, F.S., Hieter, P., 1995. TEL1, an S. cerevisiae homolog of the human gene mutated in ataxia telangiectasia, is functionally related to the yeast checkpoint gene MEC1. Cell 82, 831–40.

Murai, J., Huang, S.Y.N., Das, B.B., Dexheimer, T.S., Takeda, S., Pommier, Y., 2012. Tyrosyl-DNA phosphodiesterase 1 (TDP1) repairs DNA damage induced by topoisomerases I and II and base alkylation in vertebrate cells. J. Biol. Chem. 287, 12848–12857.

Murakami, H., Keeney, S., 2014. Temporospatial coordination of meiotic dna replication and recombination via DDK recruitment to replisomes. Cell 158, 861–873.

Murakami, H., Keeney, S., 2008. Regulating the formation of DNA double-strand breaks in meiosis. Genes Dev. 22, 286–292.

Murakami, H., Nicolas, A., 2009. Locally, meiotic double-strand breaks targeted by Gal4BD-Spo11 occur at discrete sites with a sequence preference. Mol. Cell. Biol. 29, 3500–16.

Myler, L.R., Gallardo, I.F., Zhou, Y., Gong, F., Yang, S.-H., Wold, M.S., Miller, K.M., Paull, T.T., Finkelstein, I.J., 2016. Single-molecule imaging reveals the mechanism of Exo1 regulation by single-stranded DNA binding proteins. Proc. Natl. Acad. Sci. U. S. A. 113, E1170-1179.

Nakada, D., Matsumoto, K., Sugimoto, K., 2003. ATM-related Tel1 associates with double-strand breaks through an Xrs2-dependent mechanism. Genes Dev. 17, 1957–1962.

Nakamura, K., Kogame, T., Oshiumi, H., Shinohara, A., Sumitomo, Y., Agama, K., Pommier, Y., Tsutsui, K.M., Tsutsui, K., Hartsuiker, E., Ogi, T., Takeda, S., Taniguchi, Y., Barber, S., Koyama, H., Adachi, N., So, S., Koyama, H., Hartsuiker, E., Neale, M., Carr, A., B., Zhang, C., Okada, T., Sonoda, E., Yamashita, Y., Koyoshi, S., Tateishi, S., Kikuchi, K., Taniguchi, Y., Hatanaka, A., Sonoda, E., Hochegger, H., 2010. Collaborative Action of Brca1 and CtIP in Elimination of Covalent Modifications from Double-Strand Breaks to Facilitate Subsequent Break Repair. PLoS Genet. 6, e1000828.

Nasmyth, K., 1996. At the heart of the budding yeast cell cycle. Trends Genet. 12, 405–412.

Nassif, N., Penney, J., Pal, S., Engels, W.R., Gloor, G.B., 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol. 14, 1613–25.

Navadgi-Patil, V.M., Burgers, P.M., 2011. Cell-cycle-specific activators of the Mec1/ATR checkpoint kinase. Biochem. Soc. Trans. 39, 600–5.

Neale, M., Pan, J., Keeney, S., 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature 436, 1053–1057.

New, J.H., Sugiyama, T., Zaitseva, E., Kowalczykowski, S.C., 1998. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. Nature 391, 407–410.

Nicolette, M.L., Lee, K., Guo, Z., Rani, M., Chow, J.M., Eun, S., Paull, T.T., 2011. Mre11–Rad50–Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks. Nat Struct Mol Biol 17, 1478–1485.

Nitiss, J.L., 2009. DNA topoisomerase II and its growing repertoire of biological functions. Nat. Rev. Cancer 9, 327–337.

Nitiss, K.C., Malik, M., He, X., White, S., Nitiss, J.L., 2006. Tyrosyl-DNA phosphodiesterase (Tdp1) participates in the repair of Top2-mediated DNA damage. Proc. Natl. Acad. Sci. U. S. A. 103, 8953–8958.

Niu, H., Chung, W.-H., Zhu, Z., Kwon, Y., Zhao, W., Chi, P., Prakash, R., Seong, C., Liu, D., Lu, L., Ira, G., Sung, P., 2010. Mechanism of the ATP-dependent DNA end-resection machinery from Saccharomyces cerevisiae. Nature 467, 108–11.

Oh, S.D., Lao, J.P., Hwang, P.Y.H., Taylor, A.F., Smith, G.R., Hunter, N., 2007. BLM Ortholog, Sgs1, Prevents Aberrant Crossing-over by Suppressing Formation of Multichromatid Joint Molecules. Cell 130, 259–272.

Ohkura, H., 2015. Meiosis: An overview of key differences from mitosis. Cold Spring Harb. Perspect. Biol. 7, 1–15.

Paciotti, V., Clerici, M., Lucchini, G., Longhese, M.P., 2000. The checkpoint protein Ddc2, functionally related to S. pombe Rad26, interacts with Mec1 and is regulated by Mec1-dependent phosphorylation in budding yeast. Genes Dev. 14, 2046–2059.

Padmore, R., Cao, L., Kleckner, N., 1991. Temporal comparison of recombination and synaptonemal complex formation during meiosis in S. cerevisiae. Cell 66, 1239–1256.

Page, S.L., 2003. Chromosome Choreography: The Meiotic Ballet. Science (80-.). 301, 785–789.

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H.G., Tischfield, S.E., Zhu, X., Neale, M.J., Jasin, M., Socci, N.D., Hochwagen, A., Keeney, S., 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 144, 719–31.

Panizza, S., Mendoza, M.A., Berlinger, M., Huang, L., Nicolas, A., Shirahige, K., Klein, F., 2011. Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. Cell 146, 372–383.

Pâques, F., Leung, W.Y., Haber, J.E., 1998. Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol. Cell. Biol. 18, 2045–2054.

Parvanov, E.D., Petkov, P.M., Paigen, K., 2010. Prdm9 controls activation of mammalian recombination hotspots. Science 327, 835.

Paull, T.T., 2010. Making the best of the loose ends: Mre11/Rad50 complexes and Sae2 promote DNA double-strand break resection. DNA Repair (Amst). 9, 1283–1291.

Paull, T.T., Gellert, M., 1998. The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. Mol. Cell 1, 969–79.

Pedersen, J.M., Fredsoe, J., Roedgaard, M., Andreasen, L., Mundbjerg, K., Kruhøffer, M., Persson, J., Norman, U., Hartsuiker, E., Ekwall, K., Garinther, W., Schultz, M., Gavin, I., Robinson, M., Grigull, J., Mohammad, N., Hughes, T., 2012. DNA Topoisomerases Maintain Promoters in a State Competent for Transcriptional Activation in Saccharomyces cerevisiae. PLoS Genet. 8, e1003128.

Penkner, A., Portik-Dobos, Z., Tang, L., Schnabel, R., Novatchkova, M., Jantsch, V., Loidl, J., 2007. A conserved function for a Caenorhabditis elegans Com1/Sae2/CtIP protein homolog in meiotic recombination. EMBO J. 26, 5071–82.

Pérez-Hidalgo, L., Moreno, S., San-Segundo, P.A., 2002. Regulation of meiotic progression by the meiosis-specific checkpoint kinase Mek1 in fission yeast. J. Cell Sci. 116, 259–271.

Petes, T.D., 1979. Yeast ribosomal DNA genes are located on chromosome XII. Proc. Natl. Acad. Sci. U. S. A. 76, 410–4.

Petrini, J.H., 1999. The mammalian Mre11-Rad50-nbs1 protein complex: integration of functions in the cellular DNA-damage response. Am. J. Hum. Genet. 64, 1264–1269.

Petrini, J.H., Walsh, M.E., DiMare, C., Chen, X.N., Korenberg, J.R., Weaver, D.T., 1995. Isolation and characterization of the human MRE11 homologue. Genomics 29, 80–86.

Petronczki, M., Siomos, M., Nasmyth, K., 2003. Un menage a quatre: the molecular biology of chromosome segregation in meiosis. Cell 112, 423–440.

Pincus, D.W., Neurosurg, C., Pouliot, J.J., Yao, K.C., Robertson, C.A., Nash, H.A., 1999. Yeast Gene for a Tyr-DNA Phosphodiesterase that Repairs Topoisomerase I Complexes. Science (80-.). 286, 552–556.

Plank, J.L., Wu, J., Hsieh, T.-S., 2006. Topoisomerase IIIalpha and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. Proc. Natl. Acad. Sci. U. S. A. 103, 11118–23.

Pommier, Y., 2012. DNA Topoisomerases and Cancer, in: Zhurnal Eksperimental'noi I Teoreticheskoi Fiziki. Springer New York, pp. 177–178.

Pommier, Y., Huang, S. yin N., Gao, R., Das, B.B., Murai, J., Marchand, C., 2014. Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). DNA Repair (Amst). 19, 114–129.

Pommier, Y., Leo, E., Zhang, H., Marchand, C., 2010. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. Chem. Biol. 17, 421–433.

Pommier, Y., Marchand, C., 2011. Interfacial inhibitors: targeting macromolecular complexes. Nat. Rev. Drug Discov. 11, 25–36.

Pouliot, J.J., Robertson, C.A., Nash, H.A., 2001. Pathways for repair of topoisomerase I covalent complexes in Saccharomyces cerevisiae. Genes to Cells 6, 677–687.

Primig, M., Williams, R.M., Winzeler, E.A., Tevzadze, G.G., Conway, A.R., Hwang, S.Y., Davis, R.W., Esposito, R.E., 2000. The core meiotic transcriptome in budding yeasts. Nat. Genet. 26, 415–423.

Prinz, S., Amon, A., Klein, F., 1997. Isolation of COM1, a new gene required to complete meiotic double-strand break-induced recombination in Saccharomyces cerevisiae. Genetics 146, 781–795.

Puddu, F., Oelschlaegel, T., Guerini, I., Geisler, N.J., Niu, H., Herzog, M., Salguero, I., Ochoa-Montaño, B., Viré, E., Sung, P., Adams, D.J., Keane, T.M., Jackson, S.P., 2015. Synthetic viability genomic screening defines Sae2 function in DNA repair. EMBO J. 34, 1509–22.

Pueyo, E., Dizy, M., Polo, M.C., 1993. Varietal differentiation of must and wines by means of protein fraction. Am. J. Enol. Vitic. 44, 255–260.

Puizina, J., 2004. Mre11 Deficiency in Arabidopsis Is Associated with Chromosomal Instability in Somatic Cells and Spo11-Dependent Genome Fragmentation during Meiosis. Plant Cell Online 16, 1968–1978.

Pype, S., Declercq, W., Ibrahimi, A., Michiels, C., Van Rietschoten, J.G., Dewulf, N., de Boer, M., Vandenabeele, P., Huylebroeck, D., Remacle, J.E., 2000. TTRAP, a novel protein

that associates with CD40, tumor necrosis factor (TNF) receptor-75 and TNF receptor-associated factors (TRAFs), and that inhibits nuclear factor-kappa B activation. J. Biol. Chem. 275, 18586–93.

Rassin, E., 2003. The White Bear Suppression Inventory (WBSI) Focuses on Failing Suppression Attempts. Eur. J. Pers. 17, 285–298.

Rattray, A., McGill, C., Shafer, B., Strathern, J., 2001. Fidelity of mitotic double-strand-break repair in Saccharomyces cerevisiae: a role for SAE2/COM1. Genetics 158, 109–122.

Rattray, A.J., McGill, C.B., Shafer, B.K., Strathern, J.N., 2001. Fidelity of mitotic double-strand-break repair in Saccharomyces cerevisiae: A role for SAE2/COM1. Genetics 158, 109–122.

Rattray, A.J., Shafer, B.K., Neelam, B., Strathern, J.N., 2005. A mechanism of palindromic gene amplification in Saccharomyces cerevisiae. Genes Dev. 19, 1390–1399.

Raymond, W.E., Kleckner, N., 1993. RAD50 protein of S.cerevisiae exhibits ATP-dependent DNA binding. Nucleic Acids Res. 21, 3851–3856.

Ritchie, K.B., Mallory, J.C., Petes, T.D., 1999. Interactions of TLC1 (which encodes the RNA subunit of telomerase), TEL1, and MEC1 in regulating telomere length in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. 19, 6065–75.

Rocco, V., Nicolas, A., 1996. Sensing of DNA non-homology lowers the initiation of meiotic recombination in yeast. Genes Cells 1, 645–61.

Romanienko, P.J., Camerini-Otero, R.D.D., 2000. The mouse Spo11 gene is required for meiotic chromosome synapsis. Mol. Cell 6, 975–987.

Rosu, S., Libuda, D.E., Villeneuve, A. M., 2011. Robust Crossover Assurance and Regulated Interhomolog Access Maintain Meiotic Crossover Number. Science (80-.). 334, 1286–1289.

Salceda, J., Fernández, X., Roca, J., 2006. Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA. EMBO J. 25, 2575–83.

San Filippo, J., Sung, P., Klein, H., 2008. Mechanism of eukaryotic homologous recombination. Annu. Rev. Biochem. 77, 229–257.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., Jackson, S.P., 2007. Human CtIP promotes DNA end resection. Nature 450, 509–14.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., Jackson, S.P., 2007. Human CtIP promotes DNA end resection. Nature 450, 509–514.

Sasanuma, H., Hirota, K., Fukuda, T., Kakusho, N., Kugou, K., Kawasaki, Y., Shibata, T., Masai, H., Ohta, K., 2008. Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. Genes Dev. 22, 398–410.

Sasanuma, H., Murakami, H., Fukuda, T., Shibata, T., Nicolas, A., Ohta, K., 2007. Meiotic association between Spo11 regulated by Rec102, Rec104 and Rec114. Nucleic Acids Res. 35, 1119–1133.

Schoeffler, A.J., Berger, J.M., 2008. DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. Q. Rev. Biophys. 41, 41–101.

Schoeffler, A.J., Berger, J.M., 2005. Recent advances in understanding structure–function relationships in the type II topoisomerase mechanism. Biochem. Soc. Trans. 33, 1465.

Schulman, B. a, Lindstrom, D.L., Harlow, E., 1998. Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. Proc. Natl. Acad. Sci. U. S. A. 95, 10453–10458.

Schwacha, A., Kleckner, N., 1997. Interhomolog bias during meiotic recombination: Meiotic functions promote a highly differentiated interhomolog-only pathway. Cell 90, 1123–1135.

Schwacha, A., Kleckner, N., 1994. Identification of joint molecules that form frequently between homologs but rarely between sister chromatids during yeast meiosis. Cell 76, 51–63.

Sharples, G.J., Leach, D.R., 1995. Structural and functional similarities between the SbcCD proteins of Escherichia coli and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. Mol. Microbiol. 17, 1215–1217.

Shelton, E.R., Osheroff, N., Brutlag, D.L., 1983. DNA topoisomerase II from Drosophila melanogaster. Purification and physical characterization. J Biol Chem 258, 9530–9535.

Shi, K., Kurahashi, K., Gao, R., Tsutakawa, S.E., Tainer, J.A., Pommier, Y., Aihara, H., 2012. Structural basis for recognition of 5′-phosphotyrosine adducts by Tdp2. Nat. Struct. Mol. Biol. 19, 1372–1377.

Shibata, A., Moiani, D., Arvai, A.S., Perry, J., Harding, S.M., Genois, M.M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., Ismail, A., Ismalaj, E., Petricci, E., Neale, M.J., Bristow, R.G., Masson, J.Y., Wyman, C., Jeggo, P.A., Tainer, J.A., 2014. DNA Double-Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities. Mol. Cell 53, 7–18.

Shiloh, Y., 2014. ATM: Expanding roles as a chief guardian of genome stability. Exp. Cell

Res. 329, 154–161.

Shim, E.Y., Chung, W.-H., Nicolette, M.L., Zhang, Y., Davis, M., Zhu, Z., Paull, T.T., Ira, G., Lee, S.E., 2010. Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. EMBO J. 29, 3370–80.

Shima, H., Suzuki, M., Shinohara, M., 2005. Isolation and characterization of novel xrs2 mutations in Saccharomyces cerevisiae. Genetics 170, 71–85.

Shin, M.E., Skokotas, A., Winter, E., 2010. The Cdk1 and Ime2 protein kinases trigger exit from meiotic prophase in Saccharomyces cerevisiae by inhibiting the Sum1 transcriptional repressor. Mol. Cell. Biol. 30, 2996–3003.

Shinohara, M., Oh, S.D., Hunter, N., Shinohara, A., 2008. Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. Nat. Genet. 40, 299–309.

Smith, A. V., Roeder, G.S., 1997. The yeast Red1 protein localizes to the cores of meiotic chromosomes. J. Cell Biol. 136, 957–967.

Smolka, M.B., Albuquerque, C.P., Chen, S., Zhou, H., 2007. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. Proc. Natl. Acad. Sci. U. S. A. 104, 10364–10369.

Sollier, J., Lin, W., Soustelle, C., Suhre, K., Nicolas, A., Geli, V., de La Roche Saint-Andre, C., 2004. Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. EMBO J. 23, 1957–1967.

Sommermeyer, V., Béneut, C., Chaplais, E., Serrentino, M.E., Borde, V., 2013. Spp1, a Member of the Set1 Complex, Promotes Meiotic DSB Formation in Promoters by Tethering Histone H3K4 Methylation Sites to Chromosome Axes. Mol. Cell 49, 43–54.

Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M.F., Piwnica-Worms, H., Cantley, L.C., 1994. Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr. Biol. 4, 973–982.

Songyang, Z., Lu, K.P., Kwon, Y.T., Tsai, L.-H., Filhol, O., Cochet, C., Brickey, D.A., Soderling, T.R., Bartleson, C., Graves, D.J., Demaggio, A.J., Hoekstra, M.F., Blenis, J., Hunter, T., Cantley, L.C., 1996. A Structural Basis for Substrate Specificities of Protein Ser/Thr Kinases: Primary Sequence Preference of Casein Kinases I and II, NIMA, Phosphorylase Kinase, Calmodulin- Dependent Kinase II, CDK5, and Erk1. Mol. Cell. Biol. 16, 6486–6493.

Soret, J., Gabut, M., Dupon, C., Kohlhagen, G., Stévenin, J., Pommier, Y., Tazi, J., 2003.

Altered Serine/Arginine-Rich Protein Phosphorylation and Exonic Enhancer-Dependent Splicing in Mammalian Cells Lacking Topoisomerase I. Cancer Res. 63, 8203–8211.

Souquet, B., Abby, E., Hervé, R., Finsterbusch, F., Tourpin, S., Le Bouffant, R., Duquenne, C., Messiaen, S., Martini, E., Bernardino-Sgherri, J., Toth, A., Habert, R., Livera, G., 2013. MEIOB Targets Single-Strand DNA and Is Necessary for Meiotic Recombination. PLoS Genet. 9, e1003784.

Soustelle, C., Vedel, M., Kolodner, R., Nicolas, A., 2002. Replication protein A is required for meiotic recombination in Saccharomyces cerevisiae. Genetics 161, 535–547.

Sperling, A.S., Jeong, K.S., Kitada, T., Grunstein, M., 2011. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. Proc. Natl. Acad. Sci. U. S. A. 108, 12693–8.

Srinivasan, J., Koszelak, M., Mendelow, M., Kwon, Y.G., Lawrence, D.S., 1995. The design of peptide-based substrates for the cdc2 protein kinase. Biochem J 309, 927–931.

Stahl, F., 1996. Meiotic recombination in yeast: Coronation of the double-strand-break repair model. Cell 87, 965–968.

Stankiewicz, P., Lupski, J.R., 2002. Genome architecture, rearrangements and genomic disorders. Trends Genet. 18, 74–82.

Stepanov, A., Nitiss, K.C., Neale, G., Nitiss, J.L., 2008. Enhancing Drug Accumulation in Saccharomyces cerevisiae by Repression of Pleiotropic Drug Resistance Genes with Chimeric Transcription Repressors. Mol. Pharmacol. 74, 423–431.

Stingele, J., Habermann, B., Jentsch, S., 2015. DNA–protein crosslink repair: proteases as DNA repair enzymes. Trends Biochem. Sci. 40, 67–71.

Stokes, M.P., Rush, J., Macneill, J., Ren, J.M., Sprott, K., Nardone, J., Yang, V., Beausoleil, S.A., Gygi, S.P., Livingstone, M., Zhang, H., Polakiewicz, R.D., Comb, M.J., 2007. Profiling of UV-induced ATM/ATR signaling pathways. Proc. Natl. Acad. Sci. U. S. A. 104, 19855–60.

Stoll, G., Pietiläinen, O.P.H., Linder, B., Suvisaari, J., Brosi, C., Hennah, W., Leppä, V., Torniainen, M., Ripatti, S., Ala-Mello, S., Plöttner, O., Rehnström, K., Tuulio- U., Freimer, N.B., Palotie, A., 2013. Deletion of TOP3β, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. Nat. Neurosci. 16, 1228–1237.

Stracker, T.H., Petrini, J.H.J., 2011. The MRE11 complex: starting from the ends. Nat. Rev. Mol. Cell Biol. 12, 90–103.

Su, T.T., 2006. Cellular responses to DNA damage: one signal, multiple choices. Annu. Rev.

Genet. 40, 187–208.

Subramanian, K., Rutvisuttinunt, W., Scott, W., Myers, R.S., 2003. The enzymatic basis of processivity in ?? exonuclease. Nucleic Acids Res. 31, 1585–1596.

Sun, H., Treco, D., Schultes, N.P., Szostak, J.W., 1989. Double-strand breaks at an initiation site for meiotic gene conversion. Nature 338, 87–90.

Suryadinata, R., Sadowski, M., Sarcevic, B., 2010. Control of cell cycle progression by phosphorylation of cyclin-dependent kinase (CDK) substrates. Biosci. Rep. 30, 243–55.

Sym, M., Engebrecht, J., Roeder, G.S., 1993. ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. Cell 72, 365–378.

Symington, L.S., Rothstein, R., Lisby, M., 2014. Mechanisms and regulation of mitotic recombination in saccharomyces cerevisiae. Genetics 198, 795–835.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., Stahl, F.W., 1983. The double-strand-break repair model for recombination. Cell 33, 25–35.

Tammaro, M., Barr, P., Ricci, B., Yan, H., 2013. Replication-dependent and transcription-dependent mechanisms of DNA double-strand break induction by the topoisomerase 2-targeting drug etoposide. PLoS One 8, e79202.

Tanaka, S., Araki, H., 2010. Regulation of the initiation step of DNA replication by cyclin-dependent kinases. Chromosoma 119, 565–574.

Tauchi, H., Kobayashi, J., Morishima, K., 2002. Nbs1 is essential for DNA repair by homologous recombination in higher vertebrate cells. Nature 420, 93–8.

Tennyson, R.B., Lindsley, J.E., 1997. Type II DNA topoisomerase from Saccharomyces cerevisiae is a stable dimer. Biochemistry 36, 6107–6114.

Tessé, S., Storlazzi, A., Kleckner, N., Gargano, S., Zickler, D., 2003. Localization and roles of Ski8p protein in Sordaria meiosis and delineation of three mechanistically distinct steps of meiotic homolog juxtaposition. Proc. Natl. Acad. Sci. U. S. A. 100, 12865–12870.

Thacker, D., Mohibullah, N., Zhu, X., Keeney, S., 2014. Homologue engagement controls meiotic DNA break number and distribution. Nature 510, 241–246.

Thomas, C.A., Saigo, K., McLeod, E., Ito, J., 1979. The separation of DNA segments attached to proteins. Anal. Biochem. 93, 158–166.

Traven, A., Heierhorst, J., 2005. SQ/TQ cluster domains: Concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins. BioEssays 27, 397–407.

Trujillo, K.M., Roh, D.H., Chen, L., Van Komen, S., Tomkinson, A., Sung, P., 2003. Yeast

xrs2 binds DNA and helps target rad50 and mre11 to DNA ends. J. Biol. Chem. 278, 48957–64.

Trujillo, K.M., Sung, P., 2001. DNA structure-specific nuclease activities in the Saccharomyces cerevisiae Rad50*Mre11 complex. J. Biol. Chem. 276, 35458–64.

Trujillo, K.M., Yuan, S.S.F., Lee, E.Y.H.P., Sung, P., 1998. Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. J. Biol. Chem. 273, 21447–21450.

Tse-Dinh, Y.C., Wong, T.W., Goldberg, A.R., 1984. Virus- and cell-encoded tyrosine protein kinases inactivate DNA topoisomerases in vitro. Nature 312, 785–786.

Tsubouchi, H., Ogawa, H., 2000. Exo1 roles for repair of DNA double-strand breaks and meiotic crossing over in Saccharomyces cerevisiae. Mol Biol Cell 11, 2221–2233.

Tsubouchi, H., Ogawa, H., 1998. A novel mre11 mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. Mol. Cell. Biol. 18, 260–268.

Tsukamoto, Y., 2004. Xrs2p Regulates Mre11p Translocation to the Nucleus and Plays a Role in Telomere Elongation and Meiotic Recombination. Mol. Biol. Cell 16, 597–608.

Tsutsui, Y., Kawasaki, A., Iwasaki, H., 2011. Human CtIP and Its Homologs: Team Players in DSB Resection Games, in: DNA Repair - On the Pathways to Fixing DNA Damage and Errors. InTech.

Tung, K.S., Hong, E.J., Roeder, G.S., 2000. The pachytene checkpoint prevents accumulation and phosphorylation of the meiosis-specific transcription factor Ndt80. Proc Natl Acad Sci U S A 97, 12187–12192.

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., Akimcheva, S., Woglar, A., Klein, F., Schlögelhofer, P., 2007. A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. EMBO J. 26, 5061–70.

Uemura, T., Yanagida, M., 1984. Isolation of type I and II DNA topoisomerase mutants from fission yeast: single and double mutants show different phenotypes in cell growth and chromatin organization. EMBO J. 3, 1737–1744.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. a, Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, a, Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M., 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403, 623–627.

Usui, T., Ogawa, H., Petrini, J.H.J., 2001. A DNA damage response pathway controlled by Tel1 and the Mre11 complex. Mol. Cell 7, 1255–1266.

Usui, T., Ohta, T., Oshiumi, H., Tomizawa, J., Ogawa, H., Ogawa, T., 1998. Complex formation and functional versatility of Mre11 of budding yeast in recombination. Cell 95, 705–716.

Valencia, M., Bentele, M., Vaze, M.B., Herrmann, G., Kraus, E., Lee, S.E., Schär, P., Haber, J.E., 2001. NEJ1 controls non-homologous end joining in Saccharomyces cerevisiae. Nature 414, 666–669.

van Os, N.J.H., Roeleveld, N., Weemaes, C.M.R., Jongmans, M.C.J., Janssens, G.O., Taylor, A.M.R., Hoogerbrugge, N., Willemsen, M.A.A.P., 2016. Health risks for ataxia-telangiectasia mutated heterozygotes: a systematic review, meta-analysis and evidence-based guideline. Clin. Genet. 90, 105–117.

Veuger, S.J., Curtin, N.J., Richardson, C.J., Smith, G.C.M., Durkacz, B.W., 2003. Radiosensitization and DNA repair inhibition by the combined use of novel inhibitors of DNA-dependent protein kinase and poly(ADP-ribose) polymerase-1. Cancer Res. 63, 6008–6015.

Wang, H., Li, Y., Truong, L.N., Shi, L.Z., Hwang, P.H., He, J., Do, J., Cho, M.J., Li, H., Negrete, A., Shiloach, J., Berns, M.W., Shen, B., Chen, L., Wu, X., 2014. CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. Mol. Cell 54, 1012–1021.

Wang, H., Shao, Z., Shi, L.Z., Hwang, P.Y.-H., Truong, L.N., Berns, M.W., Chen, D.J., Wu, X., 2012. CtIP protein dimerization is critical for its recruitment to chromosomal DNA double-stranded breaks. J. Biol. Chem. 287, 21471–80.

Wang, H., Shi, L.Z., Wong, C.C.L., Han, X., Hwang, P.Y.-H., Truong, L.N., Zhu, Q., Shao, Z., Chen, D.J., Berns, M.W., Yates, J.R., Chen, L., Wu, X., 2013. The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. PLoS Genet. 9, e1003277.

Wang, J.C., 2002. Cellular roles of DNA topoisomerases: a molecular perspective. Nat. Rev. Mol. Cell Biol. 3, 430–440.

Wang, X., Haber, J.E., 2004. Role of Saccharomyces single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. PLoS Biol. 2, E21.

Weinert, T., 1998. DNA damage and checkpoint pathways: Molecular anatomy and interactions with repair. Cell 94, 555–558.

Weinert, T.A., Kiser, G.L., Hartwell, L.H., 1994. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. Genes Dev. 8, 652–665.

Williams, B.R., Mirzoeva, O.K., Morgan, W.F., Lin, J., Dunnick, W., Petrini, J.H.J., 2002. A murine model of Nijmegen breakage syndrome. Curr. Biol. 12, 648–653.

Williams, R.S., Moncalian, G., Williams, J.S., Yamada, Y., Limbo, O., Shin, D.S., Groocock, L.M., Cahill, D., Hitomi, C., Guenther, G., Moiani, D., Carney, J.P., Russell, P., Tainer, J.A., 2008. Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair. Cell 135, 97–109.

Williams, R.S., Williams, J.S., Tainer, J.A., 2007. Mre11–Rad50–Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. Cell Biol. 85, 509–520.

Wilson, T.E., Lieber, M.R., 1999. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase (POL4)-dependent pathway. J. Biol. Chem. 274, 23599–23609.

Wiltzius, J.J.W., Hohl, M., Fleming, J.C., Petrini, J.H.J., 2005. The Rad50 hook domain is a critical determinant of Mre11 complex functions. Nat Struct Mol Biol 12, 403–407.

Winter, E., 2012. The Sum1/Ndt80 transcriptional switch and commitment to meiosis in Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. 76, 1–15.

Wotton, D., Shore, D., 1997. A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae. Genes Dev. 11, 748–760.

Wu, C.-C., Li, T.-K., Farh, L., Lin, L.-Y., Lin, T.-S., Yu, Y.-J., Yen, T.-J., Chiang, C.-W., Chan, N.-L., 2011. Structural Basis of Type II Topoisomerase Inhibition by the Anticancer Drug Etoposide. Science (80-.). 333, 459–462.

Wu, D., Topper, L.M., Wilson, T.E., 2008. Recruitment and dissociation of nonhomologous end joining proteins at a DNA double-strand break in Saccharomyces cerevisiae. Genetics 178, 1237–1249.

Wu, L., Davies, S.L., North, P.S., Goulaouic, H., Riou, J.F., Turley, H., Gatter, K.C., Hickson, I.D., 2000. The Bloom's syndrome gene product interacts with topoisomerase III. J. Biol. Chem. 275, 9636–9644.

Wu, L., Hickson, I.D., 2003. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature 426, 870–874.

Wu, T.C., Lichten, M., 1995. Factors that affect the location and frequency of meiosis-induced double- strand breaks in Saccharomyces cerevisiae. Genetics 140, 55–66.

Xu, D., Shen, W., Guo, R., Xue, Y., Peng, W., Sima, J., Yang, J., Sharov, A., Srikantan, S., Mohanty, S., Shaw, A.C., Lloyd, T.E., Brown, G.W., Ko, M.S.H., Gorospe, M., Zou, S., Wang, W., 2013. Top3β is an RNA topoisomerase that works with fragile X syndrome protein to promote synapse formation. Nat. Neurosci. 16, 1238–47.

Xu, L., Kleckner, N., 1995. Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. EMBO J. 14, 5115–5128.

Xu, L., Weiner, B.M., Kleckner, N., 1997. Meiotic cells monitor the status of the interhomolog recombination complex. Genes Dev. 11, 106–118.

Yang, S.W., Burgin, A.B., Huizenga, B.N., Robertson, C.A., Yao, K.C., Nash, H.A., Nash, H.A., 1996. A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc. Natl. Acad. Sci. U. S. A. 93, 11534–9.

You, Z., Chahwan, C., Bailis, J., Hunter, T., Russell, P., 2005. ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1. Mol. Cell. Biol. 25, 5363–5379.

Young, J.A., Hyppa, R.W., Smith, G.R., 2004. Conserved and nonconserved proteins for meiotic DNA breakage and repair in yeasts. Genetics 167, 593–605.

Yuan, J., Chen, J., 2009. N terminus of CtIP is critical for homologous recombination-mediated double-strand break repair. J. Biol. Chem. 284, 31746–31752.

Zakharyevich, K., Ma, Y., Tang, S., Hwang, P., 2010. Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. Mol. Cell 40, 1001–1015.

Zeng, Z., Sharma, A., Ju, L., Murai, J., Umans, L., Vermeire, L., Pommier, Y., Takeda, S., Huylebroeck, D., Caldecott, K.W., El-Khamisy, S.F., 2012. TDP2 promotes repair of topoisomerase I-mediated DNA damage in the absence of TDP1. Nucleic Acids Res. 40, 8371–8380.

Zhang, L., Kim, K.P., Kleckner, N.E., Storlazzi, A., 2011. Meiotic double-strand breaks occur once per pair of (sister) chromatids and, via Mec1/ATR and Tel1/ATM, once per quartet of chromatids. Proc. Natl. Acad. Sci. U. S. A. 108, 20036–41.

Zhu, J., Petersen, S., Tessarollo, L., Nussenzweig, A., 2001. Targeted disruption of the Nijmegen breakage syndrome gene NBS1 leads to early embryonic lethality in mice. Curr. Biol. 11, 105–109.

Zhu, Z., Chung, W.-H., Shim, E.Y., Lee, S.E., Ira, G., 2008. Sgs1 helicase and two nucleases

Dna2 and Exo1 resect DNA double-strand break ends. Cell 134, 981–94.

Zickler, D., Kleckner, N., 2015. Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harb. Perspect. Biol. 7, a016626-.

Zickler, D., Kleckner, N., 1999. Meiotic chromosomes: integrating structure and function. Annu. Rev. Genet. 33, 603–754.

Zickler, D., Kleckner, N., 1998. the Leptotene-Zygotene Transition of Meiosis. Annu. Rev. Genet. 32, 619–697.

Zou, L., Elledge, S.J., 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300, 1542–1548.

APPENDIX

Appendix

10.1 Spo11Align.command (created by Tim Cooper)

```
#!/usr/bin/env bash
#Version: 1.5
#################################################################################################
# Author(s): T.J.Cooper
# Updated: 13/9/2016
# Automates batch-processing of FASTQ/SAM files for genome-wide mapping
#################################################################################################
DIR=$( cd "$( dirname "${BASH_SOURCE[0]}" )" && pwd )
cd "$DIR" || exit
source "./Spo11Extract.config" || exit
if [[ ! -f $PERLDIR || ! -d $GENOMEDIR || -z $GENOMENAME || -z $READ1_EXT || -z $READ2_EXT || -z $TRIM || -z $TRIMLEN ]]; then
    printf "\nError: User parameters are missing and/or incorrectly specified.\n\n"; exit -1
fi
mkdir -p "./Logs"
printf %b "\n\n\n\n----------------------------------------------\n"
printf "FASTQ Alignment (1 of 2) (Paired End-To-End)\n"
printf %b "----------------------------------------------\n"
printf "Currently aligning:\n"
declare -A STR
for file in *$READ1_EXT.*; do
  STR["${file%*$READ1_EXT.*}"]=1
done
for i in "${!STR[@]}"; do
  printf "%s\n" "$i"
  printf "%s\n%s\n%s\n%s\n" "$i" "-----------------------" "GLOBAL" "-----------------------" > "$DIR/Logs/$i.txt"
  bowtie2 -X 1000 --end-to-end --dovetail --no-discordant --very-sensitive --mp 5,1 --rg-id 1 --rg PU:1 --rg LB:1 --rg SM:1 --rg PL:1 -p 4 -x $GENOMEDIR/$GENOMENAME -1 $i$READ1_EXT.fastq -2 $i$READ2_EXT.fastq -S $i"_Global".SAM 2>> "$DIR/Logs/$i.txt"
done
perl "$PERLDIR" "Global.SAM" $TRIM $TRIMLEN $READ1_EXT $READ2_EXT
if [ $TRIM = Y ]; then
  printf %b "\n\n----------------------------------------------\n"
  printf "FASTQ Alignment (2 of 2) (Trimmed Local)\n"
  printf %b "----------------------------------------------\n"
  printf "Currently realigning:\n"
  for n in "${!STR[@]}"; do
    printf "%s\n" "${n}"
    printf "\n%s\n%s\n%s\n" "-----------------------" "TRIMMED" "-----------------------" >> "$DIR/Logs/${n}.txt"
    bowtie2 -X 1000 --local --dovetail --no-discordant --very-sensitive --mp 5,1 --rg-id 1 --rg PU:1 --rg LB:1 --rg SM:1 --rg PL:1 -p 4 -x $GENOMEDIR/$GENOMENAME -1 ${n}$READ1_EXT"_unmapped_trimmed.fastq" -2 ${n}$READ2_EXT"_unmapped_trimmed.fastq" -S $n"_Trimmed".SAM 2>> "$DIR/Logs/${n}.txt"
  done
  perl "$PERLDIR" "Trimmed.SAM"
fi
printf %b "\n----------------------------------------------\n"
printf "Generating logs & cleaning directory....\n"
printf %b "----------------------------------------------\n\n"
mkdir -p "./FASTQ"
mkdir -p "./SAM"
mv *.fastq ./FASTQ
mv *.SAM ./SAM
cd "$DIR/Coordinates" || exit
printf "%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\n" "Strain" "Total Read Pairs (A)" "Total Mapped Pairs (B)" "% of (A)" "Multimapping Pairs" "% of (B)" "Valid 5' Hits" "% of (B)" "Ambiguous 5' Hits" "% of (B)" > "$DIR/Logs/LogSummary.txt"
for k in "${!STR[@]}"; do
  printf "\n%s\n%s\n%s\n" "-----------------------" "CALL STATS" "-----------------------" >> "$DIR/Logs/${k}.txt"
  TotalRead=$(awk '/reads/ && ! seen {print $1; seen=1}' < "$DIR/Logs/${k}.txt")
  MultiMap=$(awk '/concordantly >1/ {sum+=$1} END{print sum}' < "$DIR/Logs/${k}.txt")
  MappedRead=$(($(awk '/concordantly exactly/ {sum+=$1} END{print sum}' < "$DIR/Logs/${k}.txt") + $MultiMap))
  Ambig=$(($(wc -l < ${k}_Ambiguous.txt)-1))
  Global=$(($(wc -l < ${k}_Global.txt)-1))
  if [ $TRIM = Y ]; then
    Trimmed=$(($(wc -l < ${k}_Trimmed.txt)-1))
    ValidHits=$(($Global + $Trimmed))
    printf "%s\t%d\n%s\n%s\t%d\n%11s\t%d\n%11s\t%d\n%s\n%s\t%d\n" "Total Hits:" "$(($Global + $Trimmed + $Ambig))" "-----------------------" "Valid Hits:" "$ValidHits" "Global:" "$Global" "Trimmed:" "$Trimmed" "-----------------------" "Ambig Hits:" "$Ambig" >> "$DIR/Logs/${k}.txt"
    awk 'FNR==1 && NR!=1{next;}{print}' ${k}_Global.txt ${k}_Trimmed.txt > ${k}_Combined.txt
  elif [ $TRIM = N ]; then
    ValidHits=$Global
    printf "%s\t%d\n%11s\t%d\n%s\n%s\t%d\n" "Total Hits:" "$(($Global + $Ambig))" "Valid Hits:" "$Global" "-----------------------" "Ambig Hits:" "$Ambig" >> "$DIR/Logs/${k}.txt"
  fi
  printf "%s\t%d\t%d\t%.3f\t%d\t%.3f\t%d\t%.3f\t%d\t%.3f\n" "${k}" "$TotalRead" "$MappedRead" "$(bc -l <<< "($MappedRead/$TotalRead)*100")" "$MultiMap" "$(bc -l <<< "($MultiMap/$MappedRead)*100")" "$ValidHits" "$(bc -l <<< "($ValidHits/$MappedRead)*100")" "$Ambig" "$(bc -l <<< "($Ambig/$MappedRead)*100")" >> "$DIR/Logs/LogSummary.txt"
done
```

10.2 Spo11Extract.pl (created by Tim Cooper)

```
#!/usr/bin/env perl
#Version: 1.5

###################################################################################################################################################
# Author(s): T.J.Cooper
# Updated: 13/9/2016
# Processes paired-end .SAM files, extracting Watson + Crick coordinate information for single-cut Spo11 and Topo-II libraries
# Quality-control and filtering (atypical read-orientation, dubious ends)
# Two-step alignment (unmapped mate read-trimming, --local alignment)
###################################################################################################################################################

use strict;
use warnings;
use Cwd;
use List::Util qw(first);
my $outext = '.txt';          #Output .file-extension
my $inext = $ARGV[0];  #Input .file-extension
my @files = glob("*$inext");
my $chk = scalar(@files);
print "\nFailed to detect any .SAM files within the current directory.\n\n" if $chk == 0;
exit if $chk == 0;            #Stop script if no .SAM files are found
my $sub = cwd()."/Coordinates";
mkdir("$sub") unless $chk == 0;
my $trimmode = $ARGV[1];
my $trimlength = $ARGV[2];
print "---------------------------------------";
print "\nCalculating Coordinates....\n";
print "---------------------------------------\n";
print "Currently processing:\n";
for my $file (@files) {   #For-each input file
open my $IN, '<', $file or die "$!";      #Open and read input .SAM file(s)
(my $strain = $file) =~ s/_[^_]+$//;     #Strain-name
(my $mode = $ARGV[0]) =~ s/\.SAM//;                              #Alignment-mode
print "$strain\n";
my $outfile = $strain."_".$mode.$outext;                         #Output files
my $outfile2 = $strain."_Ambiguous".$outext;
my ($OUT,$OUT2,$OUT3, $OUT4);
open $OUT, '>', "$sub/$outfile" or die "$!";
open $OUT2, '>>', "$sub/$outfile2" or die "$!";
print $OUT "Strand\tChr\tPos\tReadLength\tCIGAR\tAdjustment\n";
if ($inext eq "Global.SAM" && $trimmode eq "Y")  {
        print $OUT2 "Strand\tChr\tPos\tReadLength\tCIGAR\tAdjustment\tMD-Tag\n";
        my $outfile3 = $strain.$ARGV[3]."_unmapped_trimmed.fastq";          #Unmapped R1 FASTQ file
        my $outfile4 = $strain.$ARGV[4]."_unmapped_trimmed.fastq";          #Unmapped R2 FASTQ file
        open $OUT3, '>', "$outfile3" or die "$!";
        open $OUT4, '>', "$outfile4" or die "$!";
}
while (<$IN>) {         #For-each .SAM record
        chomp $_;
        next if /^\s*@/; #Skip .SAM headerlines
        my @F = split("\t", $_);  #Split each tab-delimited field
        my $orientation = $F[3]-$F[7];          #Discard atypical read-orientations
        if ($F[1] == 99 && $orientation > 0 || $F[1] == 83 && $orientation < 0) {
                my $skipline = <$IN>;
                next;
        }
        if ($inext eq "Global.SAM" && $trimmode eq "Y") {          #Populate unmapped R1/R2 FASTQ files mapped-unmapped pairs
                if (grep {$_ == $F[1]} 73,137) {
                        print $OUT3 "\@$F[0] 1:N:0:1\n$F[9]\n+\n$F[10]\n" if $F[1] == 73;
                        print $OUT4 "\@$F[0] 1:N:0:1\n$F[9]\n+\n$F[10]\n" if $F[1] == 137;
                }
                if (grep {$_ == $F[1]} 89,153) {
                        $F[9] =~ tr/GATC/CTAG/;
                        my $revseq = reverse($F[9]);
                        my $revqual = reverse($F[10]);
                        print $OUT3 "\@$F[0] 1:N:0:1\n$revseq\n+\n$revqual\n" if $F[1] == 89;
                        print $OUT4 "\@$F[0] 1:N:0:1\n$revseq\n+\n$revqual\n" if $F[1] == 153;
                }
                if (grep {$_ == $F[1]} 69,133) {
                        my $trimseq = substr($F[9],0,$trimlength);
                        my $trimqual = substr($F[10],0,$trimlength);
                        print $OUT3 "\@$F[0] 1:N:0:1\n$trimseq\n+\n$trimqual\n" if $F[1] == 69;
                        print $OUT4 "\@$F[0] 1:N:0:1\n$trimseq\n+\n$trimqual\n" if $F[1] == 133;
                }
        }
        if (grep {$_ == $F[1]} 99,83) {
                my $index = first{/MD:Z/} @F;          #Obtain variable-column MD:Z: tag
                my @MDtag = $index =~ /\d+/g;                   #Remove non-numeric characters
                my @revMDtag = reverse(@MDtag);
                my %rules = (M => 1,D => 1,I => 0,S => 1);        #Rules to handle insertion/deletions/matches/soft-clipping
                my ($s,$LS,$RS) = (0)x3;
                while ($F[5] =~ /(\d+)([MDIS])/g) {    #Parse and interpret CIGAR code
```

```
                                        my ($n,$op)  = ($1,$2);
                                        $s += $n * $rules{$op} unless $op eq 'S';               #Calculate POS adjustment (insertions/deletions)
                                        $LS += $n * $rules{$op} if $op eq 'S' && $-[0]==0;   #(upstream soft-clip)
                                        $RS += $n * $rules{$op} if $op eq 'S' && $+[0]==length($F[5]);   #(downstream soft-clip)
                                }
                my $l = length($F[9]);    #Read-length
                my $wp = $F[3]-$LS;      #Adjusted 5' coordinate (Watson strand)
                my $cp = $F[3]+($RS+$s)-1;              #Adjusted 5' coordinate (Crick strand)
                if ($MDtag[0] == 0 && $MDtag[1] == 0 && $F[1] == 99 || $LS >1 && $F[1] == 99) {                 #Detect ambigious ends (Watson
strand)
                                printf($OUT2 "%s\t%d\t%d\t%d\t%s\t%d\t%s\n", "w",$F[2],$wp,$l,$F[5],0-$LS,$index);
                } elsif ($F[1] == 99 && $wp > 0) {
                                printf($OUT "%s\t%d\t%d\t%d\t%s\t%d\n", "w",$F[2],$wp,$l,$F[5],0-$LS);
                }
                if ($revMDtag[0] == 0 && $revMDtag[1] == 0 && $F[1] == 83 || $RS >1 && $F[1] == 83) {           #Detect ambigious ends (Crick
strand)
                                printf($OUT2 "%s\t%d\t%d\t%d\t%s\t%d\t%s\n", "c",$F[2],$cp,$l,$F[5],$RS+$s-1,$index);
                } elsif ($F[1] == 83 && $cp >0) {
                                printf($OUT "%s\t%d\t%d\t%d\t%s\t%d\n", "c",$F[2],$cp,$l,$F[5],$RS+$s-1);
                }
}
}
}
my $run_time = time() - $^T;
print "---------------------------------------";
print "\nRun Completed\n";
print "Processing Runtime: $run_time Seconds\n";
print "---------------------------------------\n\n";
```

10.3 Creating 1bp histograms v03 (H4L2).R (created by Matt Neale)

```
# Create new empty master dataframe
master=NULL
data=NULL
require(stringr)

# Sequentially imports each _Global.txt read file from working directory

files = list.files(pattern="*_Combined.txt") # import files names with ""*_Global.txt" string into variable "files"
files1 = length(files) # Count number of files
files2 = read.table(text = files, sep = "_", as.is = TRUE) #Split file names by "_" separator and create table "files2"


# Create 1bp histogram of data for each chromosome
ChrSize  =  c(230218,813184,320870,1531933,576874,270161,1090940,562643,439888,745751,666816,1078177,924431,784333,1091291,948066)
# Note not sure if these are correct since some reads more distal than expected
ChrSize=ChrSize+100 # Add padding at right end in case of mismapped 5' ends due to CIGAR encoding and repeats
master=NULL #create empty dataframe
master2=NULL #create empty dataframe
binsize=1

for (i in 1:files1){
 data=NULL
 master=NULL #create empty dataframe
 master2=NULL #create empty dataframe
 master_c=NULL #create empty dataframe
 data <- read.table(files[i], sep = "\t", header=TRUE) #Import datatable from files number 1 to j

 for (j in 1:16){
  chrlen=ChrSize[j]
  temp=subset(data, Chr==j & Strand=="w" & Pos>=0)
  w=hist((temp$Pos),breaks=seq(0,chrlen,l=chrlen+1), plot=F) #watson
  temp=subset(data, Chr==j & Strand=="c" & Pos>=0)
  c=hist((temp$Pos),breaks=seq(0,chrlen,l=chrlen+1), plot=F) #crick
  pos = w$mids+0.5

  #Create new table containing the histogram data for Watson, Crick and Total. Each in separate columns
  master2=data.frame(Chr=j, Pos = pos, Watson=w$counts, Crick=c$counts)
  master <- rbind(master, master2) # Combine data into master table
 }

 master_c=subset(master, Watson!=0 | Crick!=0) # Creates compressed version of data histogram discarding all zeros

 #Write out master files EDIT NAME OF STRING!
 wd = getwd()
 #out = paste(wd,"/","FullMap.Cer3H4L2_",files2[i,1],"_",files2[i,2],"_",files2[i,3],"_",files2[i,4],".txt", sep="")
 #write.table(master, out, col.names =TRUE, row.names = FALSE, quote = FALSE, sep="\t", append=F)
 out = paste(wd,"/","FullMap.Cer3H4L2_",files2[i,1],"_",files2[i,2],"_",files2[i,3],"_",files2[i,4],"_c.txt", sep="")
 write.table(master_c, out, col.names =TRUE, row.names = FALSE, quote = FALSE, sep="\t", append=F)

 }

#########################################################################################################################
```

10.4 Calculating background reads v03.R (created by Matt Neale)

```
require("e1071") # This pacakge permits smoothing functions (used later)
 require(stringr)
 options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
 require(doParallel)
 require(plyr)
#########################################################################################################################
 # Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
 Mreads=NULL; DSBList=list();DSBListNames=NULL

 library(doParallel)
 cl <- makeCluster(8)
 registerDoParallel(cl)
 #files=c("FullMap.sae2.S288C_MJ315_WT_1_6h_c.txt")

 #Read in all tables with string "Full.Map."
 files = list.files(pattern="FullMap.") # import files names with "FullMap." string into variable "files"
 DSBListNames = substr(files, 9, nchar(files)-6) # Shorten filename by 8 characters from beginning and 6 characters form end (i.e. remove
"FullMap." and "_c.txt")
 nfiles = length(files) # Count number of files

 DSBList=foreach (k = 1:nfiles) %dopar% { DSBList[[k]] = read.table(files[k], sep = "\t", header=TRUE) } #Import datatable
 #dflist=s
 stopCluster(cl)

 AllElementsDUB = read.table("AllElementsDUB_H4L2_2016.08.11b.txt", sep = "\t", header=TRUE) #Import datatable
 for (i in 1:nfiles){Mreads[i]=sum(DSBList[[i]]$Watson+DSBList[[i]]$Crick)/1000000} # Calculate Million reads per sample for conveting to HpM

 Pan = read.table("Pan.Hotspots.IGR.SacCer3_H4L2_2016.08.10a.txt", sep = "\t", header=TRUE) #Import hotspot datatable

#########################################################################################################################
########### START HERE ONCE DATAFRAMES ARE LOADED ###################
#########################################################################################################################
# New loop to plot multiple comparisions
strains=c(nfiles) # Analyse these numbered dataframes from the dflist
strains=1:nfiles
#########################################################################################################################
# MODULE for pulling out largest genes
AllElementsDUB$length=AllElementsDUB$stop-AllElementsDUB$start # Add length of feature column
genes=AllElementsDUB #First make a copy of the ALLElements table
genes=subset(genes, type=="gene" & genename!="Dubious_ORF" & chr !="chrmt")
genes=subset(genes, !genes$genename %in% c("TEL1","NUM1","YRF1-7","YRF1-6","YRF1-3","URA2","TOR2")) # Exclude these genes
(emprically determined to be outliers)
genes=subset(genes, length>=5500)
at=sum(genes$length)
genes$start2=genes$start+1000
genes$stop2=genes$stop-1000
genes$length2=genes$stop2-genes$start2
at2=sum(genes$length2)
#genes=genes[1:5,]
#########################################################################################################################
bg=list() # list of tables containing info on the backgrounds for each gene for each strain
library(doParallel)
cl <- makeCluster(8)
registerDoParallel(cl)

#writeLines(c(""), "log.txt")

r=NULL
r=foreach (k = strains) %dopar% { #step through sequentially each dataframe/strain. r collects within it a list of the foreach loops
 bg[[k]]=data.frame(NULL)
 bg[[k]][1:nrow(genes),"Genename"]=genes$genename
 #sink("log.txt", append=TRUE) # Send console output to text file to monitor run
 #print(dflistNames[k])

for (i in 1:nrow(genes)) {
 #cat("\r", i, "of", nrow(genes)); flush.console() # Keep track of progress. cat "\r" overprints to same line of console

 temp=subset(DSBList[[k]], Chr==(genes[i,"chr"]) & Pos>=(genes[i,"start"]) & Pos<=(genes[i,"stop"])) # Create temp vector with DSB hits across
each gene in the table
 bg[[k]][i,"Total"]=sum(temp$Watson+temp$Crick) # Calculate sum of hits within this region
 bg[[k]][i,"Density"]=bg[[k]][i,"Total"]/Mreads[k]/genes[i,"length"] # Calculate density/bp of hits within this region

 temp2=subset(DSBList[[k]], Chr==(genes[i,"chr"]) & Pos>=(genes[i,"start2"]) & Pos<=(genes[i,"stop2"])) # Create temp vector with DSB hits
across each gene in the table for "core" region
 bg[[k]][i,"TotalCore"]=sum(temp2$Watson+temp2$Crick) # Calculate sum of hits within this region
 bg[[k]][i,"DensityCore"]=bg[[k]][i,"TotalCore"]/Mreads[k]/genes[i,"length2"] # Calculate density/bp of hits within this region
}
 bg[[k]][1:5]=bg[[k]][1:5] # This line is essential in multicore loops for some reason
```

```
 #cat("\r", "Job", k, dflistNames[k], "COMPLETED"); flush.console() # Keep track of progress. cat "\r" overprints to same line of console

}
bg=r # Collect the foreach loop
stopCluster(cl)

#####################################################################
######### Drawing DotPlots of background counts ###################
#####################################################################

wd = getwd(); out = paste(wd,"/","Output_Files","/","BackgroundReads",Sys.time(),".pdf",sep=""); pdf(file=out, width=15,height=9);
layout(matrix(c(1,2), 1, 2, byrow = T))
for (k in strains) { #step through sequentially each dataframe/strain
#dotchart(bg[[k]]$Total, labels=bg[[k]]$Genename, main=dflistNames[k], xlim=c(0,max(bg[[k]]$Total)))
dotchart(bg[[k]]$Density, labels=bg[[k]]$Genename, main=DSBListNames[k], xlim=c(0,max(bg[[k]]$Density)), xlab="Total hits per Million reads per bp")
dotchart(bg[[k]]$DensityCore, labels=bg[[k]]$Genename, main=paste(c(DSBListNames[k],"Core")), xlim=c(0,max(bg[[k]]$DensityCore)), xlab="Total hits per Million reads per bp")
}
dev.off()

#####################################################################
######### Making background coutn table "BGreads" ###################
#####################################################################

BG=data.frame(NULL)
for (k in strains) {
BG[k,"Strain"]=DSBListNames[k]
BG[k,"Mean"]=mean(bg[[k]]$Density)
BG[k,"StDev"]=sd(bg[[k]]$Density)
BG[k,"StDev%"]=sd(bg[[k]]$Density)/BG[k,"Mean"]*100
BG[k,"MeanCore"]=mean(bg[[k]]$DensityCore)
BG[k,"StDevCore"]=sd(bg[[k]]$DensityCore)
BG[k,"StDevCore%"]=sd(bg[[k]]$DensityCore)/BG[k,"MeanCore"]*100
BG[k,"Mreads"]=Mreads[k]
BG[k,"BGreads"]=(BG[k,"MeanCore"]*12.071326*Mreads[k]) #Calculate TOTAL reads that are background. 12.071326=Mbp of genome
BG[k,"BGreadspM"]=(BG[k,"MeanCore"]*12.071326) #Calculate FRACTION of million reads that are background. 12.071326=Mbp of genome
BG[k,"Mreads_BGreads"]=Mreads[k]-BG[k,"BGreads"] #Calculate TOTAL reads that are NOT background. 12.071326=Mbp of genome
}
BG[2:11]=signif(BG[2:11], digits=4)
#Write out master files EDIT NAME OF STRING!
wd = getwd()
out = paste(wd,"/","BGreads6.txt", sep="")
write.table(BG, out, col.names =TRUE, row.names = FALSE, quote = FALSE, sep="\t", append=F)
```

10.5 Spo11 Mapping v09 MC.R (created by Matt Neale)

```
require("e1071") # This pacakge permits smoothing functions (used later)
require(stringr)
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
require(doParallel)
require(plyr)
#####################################################################################################################################
# Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
Mreads=NULL; DSBList=list();DSBListNames=NULL

library(doParallel)
cl <- makeCluster(8)
registerDoParallel(cl)
#files=c("FullMap.sae2.S288C_MJ315_WT_1_6h_c.txt")

#Read in all tables with string "Full.Map."
files = list.files(pattern="FullMap.") # import files names with "FullMap." string into variable "files"
DSBListNames = substr(files, 9, nchar(files)-6) # Shorten filename by 8 characters from beginning and 6 characters form end (i.e. remove "FullMap." and "_c.txt")
nfiles = length(files) # Count number of files

DSBList=foreach (k = 1:nfiles) %dopar% { DSBList[[k]] = read.table(files[k], sep = "\t", header=TRUE) } #Import datatable
#DSBList=s
stopCluster(cl)

for (i in 1:nfiles){Mreads[i]=sum(DSBList[[i]]$Watson+DSBList[[i]]$Crick)/1000000} # Calculate Million reads per sample for conveting to HpM

AllElementsDUB = read.table("AllElementsDUB_H4L2_2016.08.11b.txt", sep = "\t", header=TRUE) #Import datatable
BG = read.table("BGreads6.txt", sep = "\t", header=TRUE) #Import hotspot datatable
BGmean=NULL; for (i in 1:nfiles){BGmean[i]=unlist(subset(BG, Strain==DSBListNames[i], MeanCore))} # Ensure that background vector BGmean is using same indexe numbering as hit data
Nfactor=(1-(BGmean*12.01)) # Normalisation factor: based on number of reads that appear NOT to be background

#####################################################################################################################################
#######################
############ START HERE ONCE DATAFRAMES ARE LOADED ####################
```

```
#########################################################################################################################################
#######################
# MODULE for pulling out specific locus of interest
orf="7"
genes=AllElementsDUB #First make a copy of the ALLElements table
upstream=1600; downstream=100 # bp to extend by in either direction of ORF
genes=subset(genes, genename==orf | sysname==orf)
xl1=genes$start-upstream
xl2=genes$stop+downstream
chrom=genes$chr

#########################################################################################################################################
# MODULE for manually specifying chromosome and location
#chrom=8; xl1=400000; window.w=100000; xl2=xl1+window.w; #Plot range minimum (bp); # Plot range width (bp); #Plot range maximum (bp)

#########################################################################################################################################
# New loop to plot multiple comparisions
strains=c(8,9,3,5,1) # Plot these numbered dataframes from the DSBList
scalar.u=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1) # Unique scaling factor for each strain in the DSBList (default =1 is identical scaling)
scalar.u=1/Nfactor # Override scaling factor and use apparent hit reads instead

#########################################################################################################################################
#Plotting: first set up how the plots are organised. How many panes per image for example using the layout command
plotnumber=length(strains) # Number from 1 to 5
if (plotnumber==1) {layout(matrix(c(1,1,1,2,2),5, 1, byrow = T))}
if (plotnumber==2) {layout(matrix(c(1,1,1,2,2,2,3,3), 8, 1, byrow = T))}
if (plotnumber==3) {layout(matrix(c(1,1,1,2,2,2,3,3,3,4,4),11, 1, byrow = T))}
if (plotnumber==4) {layout(matrix(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5),14, 1, byrow = T))}
if (plotnumber==5) {layout(matrix(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6),17, 1, byrow = T))}

par(mar=c(1,5,1,0),oma = c(0, 1, 1, 1),las=1) # Sets margins per graph and outside margins per grouped set (order is bottom, left, top, right)
layout.show((length(strains)+1))

for (k in strains){ #step through sequentially each dataframe/strain

#Subset for region of interest
sae2.0=subset(DSBList[[k]], Chr==chrom & Pos>=xl1 & Pos <=xl2) #Make a sub-table of the sae2-DSB data that only contains those rows where chr = 1 in range of interest

#########################################################################################################################################
#Decompression code here
sae2.1 <- data.frame(Chr=chrom, Pos=(xl1:xl2)) # Creates expanded empty dataframe with Chr and Pos locations
sae2.1 <- merge(sae2.1,sae2.0, all=TRUE) # Merge expanded empty dataframe with compressed sae2.1 dataframe
sae2.1[is.na(sae2.1)] <- 0 # Convert all NA values to zero
sae2.1$Total=sae2.1$Watson+sae2.1$Crick

#########################################################################################################################################
# Smoothing function #### temp and smooth are just two temporary vectors. New version creates two smoothed plots for each profile for overlaying
win=1 # hanning window size [1]
win2=101 # hanning window size [2] for overlay
scalar=15 #scalar [1]
scalar2=2 # scalar [2] for overlay
sae2scalar=scalar.u[k]*scalar # adjust this if needed when adjusting hann window smoothing
sae2scalar2=scalar.u[k]*scalar2 # adjust this if needed when adjusting hann window smoothing
hw=hanning.window(win) #create hanning window (require package e1071 to be loaded)
hw2=hanning.window(win2) #create hanning window (require package e1071 to be loaded)

#########################################################################################################################################
temp=NULL
for (j in 3:5){
temp=c(rep(0,win),sae2.1[1:nrow(sae2.1),j], rep(0,win)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
smooth=filter(temp,hw) # smooth the temp vector using the hann window
smooth=smooth[(win+1):(length(smooth)-win)] # trim smooth to correct lengthsmooth2=smooth2[(win2+1):(length(smooth2)-win2)] # trim smooth to correct length
sae2.1[j+3]=smooth
}
temp2=NULL
for (j in 3:5){
  temp2=c(rep(0,win2),sae2.1[1:nrow(sae2.1),j], rep(0,win2)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
  smooth2=filter(temp2,hw2) # smooth the temp vector using the hann window
  smooth2=smooth2[(win2+1):(length(smooth2)-win2)] # trim smooth to correct length
  sae2.1[j+6]=smooth2
}

colnames(sae2.1)=c("Chr", "Pos", "Watson", "Crick", "Total", "watson.s", "crick.s", "total.s","watson.s2", "crick.s2", "total.s2")

#########################################################################################################################################
# Plot boundaries:
plot(sae2.1$Pos,sae2.1$total.s/Mreads[k], type="n", xlim=c(xl1,xl2), ylim=c(-1600,1000), ylab=paste(c(DSBListNames[k]," / Scalar ",round(scalar.u[k],digits=2)), collapse="")) #plot the start histogram
# Broad Overlays:
lines(sae2.1$Pos,sae2.1$watson.s2*sae2scalar2/Mreads[k], type="h", xlim=c(xl1,xl2), col="lightcoral") #plot the start histogram
lines(sae2.1$Pos,-sae2.1$crick.s2*sae2scalar2/Mreads[k], type="h", xlim=c(xl1,xl2), col="lightblue") #plot the start histogram
lines(sae2.1$Pos,0.5*sae2.1$total.s2*sae2scalar2/Mreads[k]-1500, type="l", xlim=c(xl1,xl2), col="grey") #plot the start histogram
```

```
#hi-res smoothed data:
lines(sae2.1$Pos,sae2.1$watson.s*sae2scalar/Mreads[k], type="h", xlim=c(xl1,xl2), col="red") #plot the start histogram
lines(sae2.1$Pos,-sae2.1$crick.s*sae2scalar/Mreads[k], type="h", xlim=c(xl1,xl2), col="blue") #plot the start histogram
lines(sae2.1$Pos,0.5*sae2.1$total.s*sae2scalar/Mreads[k]-1500, type="l", xlim=c(xl1,xl2), col="grey") #plot the start histogram
# overlay of smoothed W/C ratio :
#lines(sae2.1$Pos,100*log2((sae2.1$watson.s2*sae2scalar2/Mreads[k])/(sae2.1$crick.s2*sae2scalar2/Mreads[k])), type="l", xlim=c(xl1,xl2), col="black") #plot the ratio
}
#########################################################################################################################
#Now plot the gene datatrack
#First subset the relevant data
genes=AllElementsDUB #First make a copy of the ALLElements table
genes=subset(genes,chr==chrom & start>(xl1-10000) & stop<(xl2+10000)) #Make a sub-table of ALLElements where chr = 1 and has limits just beyond plot range
genes=subset(genes,type=="gene") #Make a sub-table of ALLElements
#Now perform the plot
plot(sae2.1$range,sae2.1$filtered, xaxt="n",yaxt="n",type="n", ylab=paste("Genes"),cex.lab=1.5,font=2, xlim=c(xl1,xl2), ylim=c(-100,120),axes=F) #set up empty plot
text((xl1+xl2)/2,-80, labels=paste("Chromosome",chrom, "/",orf,"/ Range",xl1,"to",xl2,"bp / Hann", win, "/ Y-Scalar",scalar), cex.lab=1.4)
#########################################################################################################################
########## STOP HERE IF YOU ARE PLOTTING WHOLE CHROMOSOMES!!! ############
#########################################################################################################################
# Following module draws arrows for each element
xrange=xl2-xl1
ahead=xrange/25 #make arrowhead length proportional to plot range
ahead[(ahead>500)]=500 #limit max length to 500
av=75 #arrow vertical location relative to plot dimensions
ahw=15 #arrow/head width
genesW=subset(genes,genename !="Dubious_ORF" & orientation =="+") #Make a sub-table of ALLElements
for (i in 1:nrow(genesW)){
  polygon(c(genesW[i,"start"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead,genesW[i,"stop"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead, genesW[i,"start"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="wheat", border="wheat4")
  text((genesW[i,"start"]+genesW[i,"stop"])/2,av, font=3, genesW[i,"genename"], cex=0.9) }
genesW=subset(genes,genename=="Dubious_ORF" & orientation =="+") #Make a sub-table of ALLElements
for (i in 1:nrow(genesW)){
  polygon(c(genesW[i,"start"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead,genesW[i,"stop"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead, genesW[i,"start"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="wheat", border="wheat4", lty=2)
  text((genesW[i,"start"]+genesW[i,"stop"])/2,av, font=3, genesW[i,"sysname"], cex=0.9) }

av=25 #arrow vertical location for Crick genes relative to plot dimensions
genesC=subset(genes,genename !="Dubious_ORF" & orientation =="-") #Make a sub-table of ALLElements
for (i in 1:nrow(genesC)){
  polygon(c(genesC[i,"stop"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead,genesC[i,"start"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead, genesC[i,"stop"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="thistle", border ="thistle4")
  text((genesC[i,"start"]+genesC[i,"stop"])/2,av, font=3, genesC[i,"genename"], cex=0.9) }
genesC=subset(genes,genename=="Dubious_ORF" & orientation =="-") #Make a sub-table of ALLElements
for (i in 1:nrow(genesC)){
  polygon(c(genesC[i,"stop"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead,genesC[i,"start"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead, genesC[i,"stop"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="thistle", border ="thistle4", lty=2)
  text((genesC[i,"start"]+genesC[i,"stop"])/2,av, font=3, genesC[i,"sysname"], cex=0.9) }
#########################################################################################################################
```

10.6 Hotspot totals v09 MC.R (created by Matt Neale)

```
#########################################################################################################################
require("e1071") # This pacakge permits smoothing functions (used later)
require(stringr)
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
require(doParallel)
library(doParallel)
require(plyr)
#########################################################################################################################

# YOU MUST CHANGE TO WORKING DIRECTORY CONTAINING THE INPUT and DATA MAPS

# First load required dataset files:
setwd("~/Dropbox/Documents/Work Docs/Research/Lab Notebooks/Computer scripting/R scripts/R datasets/Cer3H4L2")
AllElementsDUB = read.table("AllElementsDUB_H4L2_Brar_2016.08.16.txt", sep = "\t", header=TRUE) #Import datatable
Pan = read.table("Pan.Hotspots.IGR.SacCer3_H4L2_2016.08.10a.txt", sep = "\t", header=TRUE) #Import hotspot datatable
names(Pan) <- c("Chr","Start","End","Length","PanHits","Feature_name","Name","Midpoint", "Type","Direction","IGR","IGR.start","IGR.end")
Pan <- Pan[c("Chr","Start","End","Length","Midpoint","PanHits","Feature_name","Name","Type","Direction","IGR","IGR.start","IGR.end")] # reorder

# Now point at the data to be processed:
setwd("~/Dropbox/Aligned Files/Spo11 H4L2/Temp")

# Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
Mreads=NULL; DSBList=list();dflistNames=NULL

#Read in all tables with string "Full.Map."
files = list.files(pattern="FullMap.Cer3H4L2") # import files names with "FullMap." string into variable "files"
dflistNames = substr(files, 9, nchar(files)-6) # Shorten filename by 8 characters from beginning and 6 characters form end (i.e. remove "FullMap." and "_c.txt")
```

```
nfiles = length(files) # Count number of files

BG = read.table("BGreads8Average.txt", sep = "\t", header=TRUE) #Import background datatable
BGmean=NULL; for (i in 1:nfiles){BGmean[i]=unlist(subset(BG, Strain==dflistNames[i], MeanCore))} # Ensure that background vector BGmean
is using same indexe numbering as hit data

#########################################################################################################################
########### START HERE ONCE DATAFRAMES ARE LOADED ###################
#########################################################################################################################
#num=1
strains=c(1:nfiles) # Process these numbered dataframes from the DSBList
#strains=c(1:2) # Process these numbered dataframes from the DSBList
hotspots=Pan
extend=300
hs=list() # list of tables containing info on the hotspots for each gene for each strain

cl <- makeCluster(8)
registerDoParallel(cl)

writeLines(c(""), "log.txt")
hs=NULL
r=NULL
r=foreach (k = strains) %dopar% { #step through sequentially each dataframe/strain
  DSBList=NULL; Mreads=NULL
  # DSB map files now loaded in within each loop/instance
  DSBList = read.table(files[k], sep = "\t", header=TRUE) #Import datatable
  Mreads=sum(DSBList$Watson+DSBList$Crick)/1000000 # Calculate Million reads per sample for conveting to HpM

  hs[[k]]=Pan[c("Chr","Start","End","Length","Midpoint","PanHits","Feature_name","Name","Type","Direction")]
  hs[[k]]=hs[[k]][NULL,]
  row=1
  sink("log.txt", append=TRUE) # Send console output to text file to monitor run

  for (j in 1:16){
    hs1=NULL
    hs1=data.frame(subset(Pan[1:10], Chr==j)) # subsetting the datatables first by chromosome massively speeds the script up!
    hs1[,c("WatsonHpM","CrickHpM","TotalHpM","BGHpM","Total-
BGHpM","WatsonHpM300","CrickHpM300","TotalHpM300","BGHpM300","Total-
BGHpM300","NormHpM","NormHpM300","NormHpChr")]=NA #Fill in missing columns before rbind call
    temp1=subset(DSBList, Chr==j) # subsetting the datatables first by chromosome massively speeds the script up! I am sure this is most important
for the large DSBList tables!
    hs[[k]]=rbind(hs[[k]],hs1)
    cat("\r", "Job", k, dflistNames[k], "Chromosome", j, "Hotspot", row, "of", nrow(hotspots)); flush.console() # Keep track of progress. cat "\r"
overprints to same line of console

    for (i in 1:nrow(hs1)) {
      temp=subset(temp1, Pos>=(hs1[i,"Start"]) & Pos<=(hs1[i,"End"])) # Create temp vector with DSB hits across each hotspot in the table
      hs[[k]][row,"WatsonHpM"]=sum(temp$Watson)/Mreads # Calculate sum of hits within this region/Mreads[k]
      hs[[k]][row,"CrickHpM"]=sum(temp$Crick)/Mreads # Calculate sum of hits within this region/Mreads[k]

      temp=subset(temp1, Pos>=((hs1[i,"Start"])-extend) & Pos<=((hs1[i,"End"])+extend)) # Create temp vector with DSB hits across hotspot in table
+/-300bp
      hs[[k]][row,"WatsonHpM300"]=sum(temp$Watson)/Mreads # Calculate sum of hits within this region/Mreads[k]
      hs[[k]][row,"CrickHpM300"]=sum(temp$Crick)/Mreads # Calculate sum of hits within this region/Mreads[k]

      row=row+1 # Increment counter
    }
  }

  # Vectorised math:
  hs[[k]][1:nrow(hs[[k]]),"TotalHpM"]=hs[[k]][1:nrow(hs[[k]]),"WatsonHpM"]+hs[[k]][1:nrow(hs[[k]]),"CrickHpM"]

  hs[[k]][1:nrow(hs[[k]]),"TotalHpM300"]=hs[[k]][1:nrow(hs[[k]]),"WatsonHpM300"]+hs[[k]][1:nrow(hs[[k]]),"CrickHpM300"]

  hs[[k]][1:nrow(hs[[k]]),"BGHpM"]=BGmean[k]*hs[[k]][1:nrow(hs[[k]]),"Length"]
  hs[[k]][1:nrow(hs[[k]]),"BGHpM300"]=BGmean[k]*(hs[[k]][1:nrow(hs[[k]]),"Length"]+(extend*2))

  hs[[k]][1:nrow(hs[[k]]),"Total-BGHpM"]=hs[[k]][1:nrow(hs[[k]]),"TotalHpM"]-hs[[k]][1:nrow(hs[[k]]),"BGHpM"]
  hs[[k]][1:nrow(hs[[k]]),"Total-BGHpM300"]=hs[[k]][1:nrow(hs[[k]]),"TotalHpM300"]-hs[[k]][1:nrow(hs[[k]]),"BGHpM300"]

  hs[[k]][1:nrow(hs[[k]]),"NormHpM"]=hs[[k]][1:nrow(hs[[k]]),"Total-BGHpM"]
  hs[[k]][1:nrow(hs[[k]]),"NormHpM"][hs[[k]][1:nrow(hs[[k]]),"NormHpM"] < 0] = 0 # Convert all -ve values to zero
  FinalSum=sum(hs[[k]][1:nrow(hs[[k]]),"NormHpM"])
  hs[[k]][1:nrow(hs[[k]]),"NormHpM"]=hs[[k]][1:nrow(hs[[k]]),"NormHpM"]/FinalSum*1000000

  hs[[k]][1:nrow(hs[[k]]),"NormHpM300"]=hs[[k]][1:nrow(hs[[k]]),"Total-BGHpM300"]
  hs[[k]][1:nrow(hs[[k]]),"NormHpM300"][hs[[k]][1:nrow(hs[[k]]),"NormHpM300"] < 0] = 0 # Convert all -ve values to zero
  FinalSum=sum(hs[[k]][1:nrow(hs[[k]]),"NormHpM300"])
  hs[[k]][1:nrow(hs[[k]]),"NormHpM300"]=hs[[k]][1:nrow(hs[[k]]),"NormHpM300"]/FinalSum*1000000

  #Loop to calculate NormHpChr

  rowA=0
  for (i in 1:16){
    aChr=subset(hs[[k]],Chr==i)
    aChrDSBs=nrow(aChr) # Number of hotspots per chromosome (rows)
```

```
   aChrTotal=sum(aChr[1:aChrDSBs,"NormHpM"])
   hs[[k]][(rowA+1):(aChrDSBs+rowA),"NormHpChr"]=aChr[1:aChrDSBs,"NormHpM"]/aChrTotal*1000000
   rowA=rowA+aChrDSBs
  }

  #hs[[k]][1:nrow(hs[[k]]),"Total-BG"]=hs[[k]][1:nrow(hs[[k]]),"Watson"]+hs[[k]][1:nrow(hs[[k]]),"Crick"]-hs[[k]][1:nrow(hs[[k]]),"BGhits"]
  #hs[[k]][1:nrow(hs[[k]]),"Total300-BG"]=hs[[k]][1:nrow(hs[[k]]),"Watson300"]+hs[[k]][1:nrow(hs[[k]]),"Crick300"]-
hs[[k]][1:nrow(hs[[k]]),"BGhits300"]

  #hs[[k]][1:nrow(hs[[k]]),"Hit.Increase"]=hs[[k]][1:nrow(hs[[k]]),"Total300-BG"]-hs[[k]][1:nrow(hs[[k]]),"Total-BG"]
  #hs[[k]][1:nrow(hs[[k]]),"Fold.Increase"]=hs[[k]][1:nrow(hs[[k]]),"Total300-BG"]/hs[[k]][1:nrow(hs[[k]]),"Total-BG"]

  #hs[[k]][1:nrow(hs[[k]]),"WC.ratio"]=hs[[k]][1:nrow(hs[[k]]),"Watson"]/hs[[k]][1:nrow(hs[[k]]),"Crick"]
  #hs[[k]][1:nrow(hs[[k]]),"WC.ratio300"]=hs[[k]][1:nrow(hs[[k]]),"Watson300"]/hs[[k]][1:nrow(hs[[k]]),"Crick300"]

  cat("\r", "Job", k, "COMPLETED", dflistNames[k], "Chromosome", j, "Hotspot", row-1, "of", nrow(hotspots)); flush.console()
  hs[[k]][1:23]=hs[[k]][1:23] # For unknown reasons this code is ESSENTIAL to get the script to populate hs[[k]] with anything. Otherwise it returns
"NULL"
  hs[[k]][11:23]=round(hs[[k]][11:23], digits=2)

  ######### Write tables to text file ---- Now part of loop ##########
  wd = getwd()
  out = paste(wd,"/","Hotspot.Table.",dflistNames[k],".txt", sep="")
  write.table(hs[[k]], out, col.names =TRUE, row.names = FALSE, quote = FALSE, sep="\t", append=F)

  }
stopCluster(cl)

sink()

#a=hs[[3]]

#hs=r

#for (i in 1:nfiles) { hs[[i]][11:20]=round(hs[[i]][11:20], digits=2); hs[[i]][1:15,] }

#########################################################################################################################
STOP                                                                                                                HERE
#########################################################################################################################
######### Write tables to text file ###################

#for (i in 1:nfiles){
#wd = getwd()
#out = paste(wd,"/","Hotspot.Table.",dflistNames[i],".txt", sep="")
#write.table(hs[[i]], out, col.names =TRUE, row.names = FALSE, quote = FALSE, sep="\t", append=F)
#}

##############

#layout(matrix(c(1,2), 1, 2, byrow = T))
#for (k in strains) { #step through sequentially each dataframe/strain

  #dotchart(bg[[k]]$Total, labels=bg[[k]]$Genename, main=dflistNames[k], xlim=c(0,max(bg[[k]]$Total)))
  #dotchart(bg[[k]]$Density, labels=bg[[k]]$Genename, main=dflistNames[k], xlim=c(0,max(bg[[k]]$Density)), xlab="Total hits per Million reads
per bp")
  #dotchart(bg[[k]]$DensityCore, labels=bg[[k]]$Genename, main=paste(c(dflistNames[k],"Core")), xlim=c(0,max(bg[[k]]$DensityCore)),
xlab="Total hits per Million reads per bp")
#}
```

10.7 Plotting hotspot tables v03.R (created by Matt Neale)

```
setwd("~/Dropbox/Aligned Files/Spo11 H4L2/Temp")
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc

#Read in all tables with string "Hotspot.Table."
files = list.files(pattern="Hotspot.Table.") # import files names with ""Hotspot.Table." string into variable "files"
#files1 = substr(files, 1, nchar(files)-4) # SHorten filename by 4 characters (i.e. remove ".txt")
nfiles = length(files) # Count number of files
#files2 = read.table(text = files1, sep = "_", as.is = TRUE) #Split file names by "_" separator and create table "files2"
hs=list() # Initialise empty list
hsListNames=NULL # Initialise hsListNames
Nfactor = NULL
for (j in 1:nfiles) {
  hs[[j]] <- read.table(files[j], sep = "\t", header=TRUE) # Import datatable from files number 1 to j into hs[[X]]
  #hsListNames[j]=paste(files2[j,2],files2[j,3],files2[j,4],files2[j,5], sep="_")} # Extract identifiers from filename
 hsListNames[j] = substr(files[j], 15, nchar(files[j])-4) # Shorten filename by 14 characters from beginning and 6 characters form end (i.e. remove
"Hotspot.Table." and ".txt")
}

#BG = read.table("BGreads8.txt", sep = "\t", header=TRUE) #Import background datatable

#BGmean=NULL; for (i in 1:nfiles){BGmean[i]=unlist(subset(BG, Strain==hsListNames[i], MeanCore)) # Ensure that background vector BGmean
is using same indexe numbering as hit data
```

```
#  Nfactor[i]=(1-(BGmean[i]*12.01))} # Normalisation factor: based on number of reads that appear NOT to be background

# Other data:
CEN=c(151523.5, 238265, 114443, 449766,152045.5, 148568.5, 496979,105644.5, 355687, 436366,440187.5, 150887.5, 268090, 628816.5, 326643, 556015) # CENtromere positions
#RMM = read.table("RMMSubSamp_simple.txt", sep = "\t", header=TRUE) #Import RMM datatable
#RMM1=RMM[seq(1, NROW(RMM), by = 100),] # Subsample every 10th row to reduce resoltuion of RMM plot

#########################################################################################################################
#####Data-manipulation
#########################################################################################################################

# Create "averaged" WT
#i=nfiles+1 # Index of additional table
#hs[[i]]=hs[[7]] # EDIT! table 7 is currently Kayleigh's WT data
#hs[[i]][11:20]=round((hs[[7]][11:20]+hs[[10]][11:20]+hs[[11]][11:20])/3,digits=2) # EDIT! Tables 7,10,11 are currently the WT repeats
#hs[[i]][1:100,]
#BGmean[i]=mean(BGmean[c(7,10,11)])
#Nfactor[i]=(1-(BGmean[i]*12.01)) # Normalisation factor: based on number of reads that appear NOT to be background
#hsListNames[i]="WT1.3.4.average"


# Correlation plots
plots=c(2,1)
nplots=length(plots)
wd = getwd(); out = paste(wd,"/","Output_Files","/","HotspotCorrelations",Sys.time(),".pdf",sep=""); pdf(file=out, width=4*nplots,height=4*nplots);
par(mar=c(4,4,1,1),oma = c(1, 1, 1, 1),las=1) # Sets margins per graph and outside margins per grouped set (order is bottom, left, top, right)plots=nfiles
layout(matrix(c(1:nplots^2),nplots, nplots, byrow = T))
#layout.show(nplots^2)
for (n in plots){
  for (m in plots){
    y=subset(hs[[n]],Chr %in% (1:16)) # Subset data for particular chromosomes if desired
    y=y[,"NormHpM"]
    x=subset(hs[[m]],Chr %in% (1:16)) # Subset data for particular chromosomes if desired
    x=x[,"NormHpM"]
    #model1=lm(log2(y)~log2(x))
    model2=lm(y~x)

    #plot(log2(x),log2(y),              xlim=c(log2(0.1),log2(10000)),              ylim=c(log2(.1),log2(10000)),xlab=hsListNames[m],ylab=hsListNames[n],
col=rgb(0,0,0,0.2),
    #     text(log2(100),log2(0.1),paste(c("r^2 =", signif(summary(model1)$r.squared)),collapse="")), cex=0.5)
    #abline(lm(log2(y)~log2(x)))

    plot(x,y, log="xy", xlim=c(1,10000), ylim=c(1,10000),xlab=hsListNames[m],ylab=hsListNames[n], col=rgb(0,0,0,0.2),
        text(100,1,paste(c("r^2 =", signif(summary(model2)$r.squared)),collapse="")), cex=0.5)
    abline(0,1)
}
}

dev.off()
```

10.8 Pileups around TSS v02.R (created by Matt Neale)

```
setwd("~/Dropbox/Aligned Files/Spo11 H4L2/Temp")
DUB = read.table("AllElementsDUB_H4L2_Brar_2016.08.16.txt", sep = "\t", header=TRUE) #Import datatable
#########################################################################################################################
require("e1071") # This package permits smoothing functions (used later)
require(stringr)
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
require(doParallel)
library(doParallel)
require(plyr)
#########################################################################################################################
# Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
Mreads=NULL; dflist=list();dflistNames=NULL

#Read in all tables with string "Full.Map."
files = list.files(pattern="FullMap.Cer3H4L2") # import files names with "FullMap.Cer3H4L2" string into variable "files"
dflistNames = substr(files, 9, nchar(files)-6) # Shorten filename by 8 characters from beginning and 6 characters form end (i.e. remove "FullMap." and "_c.txt")
nfiles = length(files) # Count number of files

cl <- makeCluster(8)
registerDoParallel(cl)

dflist=foreach (k = 1:nfiles) %dopar% { dflist[[k]] = read.table(files[k], sep = "\t", header=TRUE) } #Import datatable
stopCluster(cl)

#########################################################################################################################

BG = read.table("BGreads8Average.txt", sep = "\t", header=TRUE) #Import background datatable
```

```
BGmean=NULL; for (i in 1:nfiles){BGmean[i]=unlist(subset(BG, Strain==dflistNames[i], MeanCore))} # Ensure that background vector BGmean is using same indexe numbering as hit data

# Calculate Mreads for each datatable
for (i in 1:nfiles){Mreads[i]=sum(dflist[[i]]$Watson+dflist[[i]]$Crick)/1000000}

####################################################################################################################
########### START HERE ONCE DATAFRAMES ARE LOADED #####################
####################################################################################################################
##################################
# New loop to plot multiple comparisions
strains=c(11,12,17,18,19,20,1,2,5,6,3,4,21,22,13,14,15,16,9,10,7,8) # Plot these numbered dataframes from the dflist (MAXIMUM of FIVE!)
columns=2 # How many columns for data output?
scalar.u=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1) # Unique scaling factor for each strain in the dflist (default =1 is identical scaling)
#scalar.u=1/Nfactor # Override unique caling factor and instead use apparnt non-background-read normalisation factor
#ymin=c(0,145,0,0,0)
#ymax=c(250,160,250,250,400)
####################################################################################################################
# Send output to PDF: height of page scales with number of plots
wd = getwd(); out = paste(wd,"/","Output_Files","/","TSSPileup_",Sys.time(),".pdf",sep=""); pdf(file=out, width=columns*4,height=((length(strains)/columns*3)+3));
#Plotting: first set up how the plots are organised. How many panes per image for example using the layout command
plotnumber=length(strains) # Number from 1 to 16
#layout(matrix(c(1:(plotnumber+length(quantiles))),(length(strains)+1), length(quantiles), byrow = T))
layout(matrix(c(1:plotnumber,rep((plotnumber+1),columns)),((length(strains)/columns)+1), columns, byrow = T))
#if (plotnumber>0) {layout(matrix(c(1,2,1,2,3,3),3, 2, byrow = T))}
#if (plotnumber>2) {layout(matrix(c(1,2,1,2,3,4,3,4,5,5), 5, 2, byrow = T))}
#if (plotnumber>4) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,7),7, 2, byrow = T))}
#if (plotnumber>6) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,9),9, 2, byrow = T))}
#if (plotnumber>8) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,11),11, 2, byrow = T))}
#if (plotnumber>10) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,13),13, 2, byrow = T))}
#if (plotnumber>12) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,14,13,14,15,15),15, 2, byrow = T))}
#if (plotnumber>14) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,14,13,14,15,16,15,16,17,17),17, 2, byrow = T))}
par(mar=c(1,4,1,2),oma = c(1, 1, 1, 1),las=1) # Sets margins per graph and outside margins per grouped set (order is bottom, left, top, right)
#layout.show((length(strains))+1)
####################################################################################################################
# MODULE for pulling out specific locus of interest
genes=DUB #First make a copy of the ALLElements table
genes=subset(genes, type=="gene" & genename!="Dubious_ORF") #ONly look at genes and exlcude dubious orfs
genesW=subset(genes, orientation=="+") # genesW = Watson strand genes
genesC=subset(genes, orientation=="-") # genesC = Crick strand genes

############################################ Start of Strain Loop ##########################################

for (k in strains) { #step through sequentially each dataframe/strain

width1=1000 # bp upstream and downstream of TSS to retrieve
padding=500
width=width1+padding

#Zero out TSS dataframe - ONLY DO THIS FOR EACH STRAIN! - This must be outside of chromosome loop!
TSS=data.frame(NULL)
TSS[1:((width*2)+1),"Pos"]=(1:((2*width)+1))

TSS$WatsonW=0 # Watson genes watson hits
TSS$CrickW=0 # Watson genes crick hits
TSS$WatsonC=0 # Crick genes watson hits
TSS$CrickC=0 # Crick genes crick hits

temp=NULL
#chroms=c(1:16)
#chroms=c(1:2,4:16) # Specify the chromosomes you want to plot
chroms=c(2,4,5,7:16) # Or just a single chromoosme. Hash this line out if you wnat to plot all of them specified above.

############################################ Start of Chromosome Loop ##########################################
for (chrom in chroms) { #Specify which chromosomes to process
df.1=subset(dflist[[k]],Chr==chrom) # Make copy of the datatable

genesW1=subset(genesW, chr==chrom)
genesC1=subset(genesC, chr==chrom)

#Watson genes
for (i in 1:nrow(genesW1)) {
   temp=subset(df.1[2:4], Pos>=(genesW1[i,"start"]-width) & Pos<=(genesW1[i,"start"]+width)) # Create temp matrix with DSB hits in this region
   temp$Pos=temp$Pos-genesW1[i,"start"]+width+1
   TSS[temp$Pos,"WatsonW"]=TSS[temp$Pos,"WatsonW"]+temp[,"Watson"]
   TSS[temp$Pos,"CrickW"]=TSS[temp$Pos,"CrickW"]+temp[,"Crick"]
}
#Crick genes
for (i in 1:nrow(genesC1)) { #NOTE: TSS is at the stop location in the table
  temp=subset(df.1[2:4], Pos>=(genesC1[i,"stop"]-width) & Pos<=(genesC1[i,"stop"]+width)) # Create temp matrix with DSB hits in this region
  temp$Pos=temp$Pos-genesC1[i,"stop"]+width+1
  TSS[temp$Pos,"WatsonC"]=TSS[temp$Pos,"WatsonC"]+temp[,"Watson"]
  TSS[temp$Pos,"CrickC"]=TSS[temp$Pos,"CrickC"]+temp[,"Crick"]
}
```

```
} # Repeat for each chromosome

# This section combines the Watson and Crick hits, reversing the order of the crick gene hits and adding the Watsons to the Crick hits and Crick to Watsons
# (i.e. reverse complements the data for the Crick genes).
# This allows both to be combined in the plot
TSS$revWc=rev(TSS$WatsonC) #Crick genes with Watson hits
TSS$revCc=rev(TSS$CrickC) #Crick genes with Crick hits
TSS$WatsonTotal=TSS$WatsonW+TSS$revCc # Add WatsonW hits to reversed CrickC hits
TSS$CrickTotal=TSS$CrickW+TSS$revWc # Add CrickW hits to reversed WatsonC hits
TSS$WCTotal=(TSS$WatsonTotal+TSS$CrickTotal)/2 #Create total column and halve it

#########################################################################################################################
# Smoothing function #### temp and smooth are just two temporary vectors
win=151 # hanning window size
scalar=.1
sae2scalar=scalar.u[k]*scalar # adjust this if needed when adjusting hann window smoothing
hw=hanning.window(win) #create hanning window (require package e1071 to be loaded)

temp=NULL
for (j in 8:10){ #Smooth columns 8 to 10 in the TSS table
  temp=c(rep(0,win),TSS[1:nrow(TSS),j], rep(0,win)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
  smooth=filter(temp,hw) # smooth the temp vector using the hann window
  smooth=smooth[(win+1):(length(smooth)-win)] # trim smooth to correct length
  TSS[j+3]=smooth
}
colnames(TSS)=c("Pos", "WatsonW", "CrickW", "WatsonC", "CrickC", "revWc", "revCc", "WatsonTotal", "CrickTotal", "WCTotal", "WatsonTotalS", "CrickTotalS","WCTotalS")
TSS=TSS[padding:(padding+width1+width1),]
#Unsmoothed data:
#plot(TSS$Pos-width, TSS$WCTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], xlim=c(-width1,width1))
#lines(TSS$Pos-width, TSS$WatsonTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="red")
#lines(TSS$Pos-width, TSS$CrickTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="blue")
#Smoothed data:
plot(TSS$Pos-width, TSS$WCTotalS*sae2scalar/Mreads[k], type="l", ylab=c(dflistNames[k]))
lines(TSS$Pos-width, TSS$WatsonTotalS*sae2scalar/Mreads[k], type="l", col="red")
lines(TSS$Pos-width, TSS$CrickTotalS*sae2scalar/Mreads[k], type="l", col="blue")

#########################################################################################################################

} # Strain loop
plot(TSS$Pos-width, TSS$WCTotal, xaxt="n",yaxt="n", type="n", ylab="", ylim=c(-100,120), axes=F)
text(0,0, font=3, labels=paste("Chromosome",toString(chroms), "/ Hann", win, "/ Y-Scalar",scalar))

dev.off()
```

10.9 Pileups around TSS v03 Gene expression.R (created by Matt Neale)

```
setwd("~/Dropbox/Aligned Files/Spo11 H4L2/Temp")
DUB = read.table("AllElementsDUB_H4L2_Brar_2016.08.16.txt", sep = "\t", header=TRUE) #Import datatable
DUB[is.na(DUB)]=0 # Convert all NA values to zero
#########################################################################################################################
require("e1071") # This pacakge permits smoothing functions (used later)
require(stringr)
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
require(doParallel)
library(doParallel)
require(plyr)
#########################################################################################################################
# Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
Mreads=NULL; dflist=list();dflistNames=NULL

#Read in all tables with string "Full.Map."
files = list.files(pattern="FullMap.Cer3H4L2") # import files names with "FullMap.Cer3H4L2" string into variable "files"
dflistNames = substr(files, 9, nchar(files)-6) # Shorten filename by 8 characters from beginning and 6 characters form end (i.e. remove "FullMap." and "_c.txt")
nfiles = length(files) # Count number of files

cl <- makeCluster(8)
registerDoParallel(cl)

dflist=foreach (k = 1:nfiles) %dopar% { dflist[[k]] = read.table(files[k], sep = "\t", header=TRUE) } #Import datatable
stopCluster(cl)

#########################################################################################################################

BG = read.table("BGreads8.txt", sep = "\t", header=TRUE) #Import background datatable
BGmean=NULL; for (i in 1:nfiles){BGmean[i]=unlist(subset(BG, Strain==dflistNames[i], MeanCore))} # Ensure that background vector BGmean is using same indexe numbering as hit data

# Calculate Mreads for each datatable
for (i in 1:nfiles){Mreads[i]=sum(dflist[[i]]$Watson+dflist[[i]]$Crick)/1000000}
```

```
#########################################################################################################################
########### START HERE ONCE DATAFRAMES ARE LOADED ######################
#########################################################################################################################
# New loop to plot multiple comparisions
strains=c(1,7,12) # Plot these numbered dataframes from the dflist (MAXIMUM of FIVE!)
quantiles=c(1:5) # Split the data up into this many fractions (Syntax is (1:5) or (1:10) etc)
scalar.u=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1) # Unique scaling factor for each strain in the dflist (default =1 is identical scaling)
#scalar.u=1/Nfactor # Override unique caling factor and instead use apparnt non-background-read normalisation factor
#ymin=c(0,145,0,0,0)
#ymax=c(250,160,250,250,400)
#########################################################################################################################
# Send output to PDF: height of page scales with number of plots
wd = getwd(); out = paste(wd,"/","Output_Files","/","TSSPileup_",Sys.time(),".pdf",sep=""); pdf(file=out,
width=length(quantiles)*4,height=(length(strains)*3+3));
#Plotting: first set up how the plots are organised. How many panes per image for example using the layout command
plotnumber=length(strains)*length(quantiles) # Number from 1 to 16
#layout(matrix(c(1:(plotnumber+length(quantiles))),(length(strains)+1), length(quantiles), byrow = T))
layout(matrix(c(1:plotnumber,rep((plotnumber+1),length(quantiles))),(length(strains)+1), length(quantiles), byrow = T))
#if (plotnumber>0) {layout(matrix(c(1,2,1,2,3,3),3, 2, byrow = T))}
#if (plotnumber>2) {layout(matrix(c(1,2,1,2,3,4,3,4,5,5), 5, 2, byrow = T))}
#if (plotnumber>4) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,7),7, 2, byrow = T))}
#if (plotnumber>6) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,9),9, 2, byrow = T))}
#if (plotnumber>8) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,11),11, 2, byrow = T))}
#if (plotnumber>10) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,13),13, 2, byrow = T))}
#if (plotnumber>12) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,14,13,14,15,15),15, 2, byrow = T))}
#if (plotnumber>14) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,14,13,14,15,16,15,16,17,17),17, 2, byrow = T))}
par(mar=c(1,4,1,2),oma = c(1, 1, 1, 1),las=1) # Sets margins per graph and outside margins per grouped set (order is bottom, left, top, right)
#layout.show((plotnumber+length(quantiles)))

#########################################################################################################################
# MODULE for pulling out specific locus of interest
genes=DUB #First make a copy of the ALLElements table
genes=subset(genes, type=="gene" & genename!="Dubious_ORF") #ONly look at genes and exlcude dubious orfs
genes=genes[order(genes$RNAseq_Brar_00_EXP),] ## Order datatable by relevant Brar expression column
genesW=subset(genes, orientation=="+") # genesW = Watson strand genes
genesC=subset(genes, orientation=="-") # genesC = Crick strand genes

############################################## Start of Strain Loop ##########################################
for (k in strains) { #step through sequentially each dataframe/strain
 width1=1000 # bp upstream and downstream of TSS to retrieve
 padding=500
 width=width1+padding

 ############################################## Start of Quantile Loop ##########################################
 quantE=nrow(genes)/max(quantiles) # calculate the range of each quantile
 quantE1=c(1,quantiles*quantE)
 for (kk in quantiles){
  genesQ=genes[quantE1[kk]:quantE1[kk+1],] #subset DUB table by expression quantile
  genesW=subset(genesQ, orientation=="+")
  genesC=subset(genesQ, orientation=="-")
  kkTotal=sum(genesQ$RNAseq_Brar_00_EXP) # Add up all the expression data for this quartile and store in variable - will be added to plot.

  #Zero out TSS dataframe - ONLY DO THIS FOR EACH STRAIN! - This must be outside of chromosome loop!
  TSS=data.frame(NULL)
  TSS[1:((width*2)+1),"Pos"]=(1:((2*width)+1))
  TSS$WatsonW=0 # Watson genes watson hits
  TSS$CrickW=0 # Watson genes crick hits
  TSS$WatsonC=0 # Crick genes watson hits
  TSS$CrickC=0 # Crick genes crick hits

  temp=NULL
  chroms=c(1:16) # Or just a single chromosome.

  ############################################## Start of Chromosome Loop ##########################################
  for (chrom in chroms) { #Specify which chromosomes to process
   df.1=subset(dflist[[k]],Chr==chrom) # Make copy of the datatable

   genesW1=subset(genesW, chr==chrom) #Subset for Watson genes
   genesC1=subset(genesC, chr==chrom) #Subset for Crick genes

   #Watson genes
   for (i in 1:nrow(genesW1)) {
    temp=subset(df.1[2:4], Pos>=(genesW1[i,"start"]-width) & Pos<=(genesW1[i,"start"]+width)) # Create temp matrix with DSB hits in this
region
    temp$Pos=temp$Pos-genesW1[i,"start"]+width+1
    TSS[temp$Pos,"WatsonW"]=TSS[temp$Pos,"WatsonW"]+temp[,"Watson"]
    TSS[temp$Pos,"CrickW"]=TSS[temp$Pos,"CrickW"]+temp[,"Crick"]
   }
   #Crick genes
   for (i in 1:nrow(genesC1)) { #NOTE: TSS is at the stop location in the table
    temp=subset(df.1[2:4], Pos>=(genesC1[i,"stop"]-width) & Pos<=(genesC1[i,"stop"]+width)) # Create temp matrix with DSB hits in this region
    temp$Pos=temp$Pos-genesC1[i,"stop"]+width+1
    TSS[temp$Pos,"WatsonC"]=TSS[temp$Pos,"WatsonC"]+temp[,"Watson"]
    TSS[temp$Pos,"CrickC"]=TSS[temp$Pos,"CrickC"]+temp[,"Crick"]
   }
```

```
  } # Repeat for each chromosome

  # This section combines the Watson and Crick hits, reversing the order of the crick gene hits and adding the Watsons to the Crick hits and Crick to Watsons
  # (i.e. reverse complements the data for the Crick genes).
  # This allows both to be combined in the plot
  TSS$revWc=rev(TSS$WatsonC) #Crick genes with Watson hits
  TSS$revCc=rev(TSS$CrickC) #Crick genes with Crick hits
  TSS$WatsonTotal=TSS$WatsonW+TSS$revCc # Add WatsonW hits to reversed CrickC hits
  TSS$CrickTotal=TSS$CrickW+TSS$revWc # Add CrickW hits to reversed WatsonC hits
  TSS$WCTotal=(TSS$WatsonTotal+TSS$CrickTotal)/2 #Create total column and halve it


#########################################################################################################################
  # Smoothing function #### temp and smooth are just two temporary vectors
  win=151 # hanning window size
  scalar=.1
  sae2scalar=scalar.u[k]*scalar # adjust this if needed when adjusting hann window smoothing
  hw=hanning.window(win) #create hanning window (require package e1071 to be loaded)

  temp=NULL
  for (j in 8:10){ #Smooth columns 8 to 10 in the TSS table
    temp=c(rep(0,win),TSS[1:nrow(TSS),j], rep(0,win)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
     smooth=filter(temp,hw) # smooth the temp vector using the hann window
     smooth=smooth[(win+1):(length(smooth)-win)] # trim smooth to correct length
     TSS[j+3]=smooth
  }
  colnames(TSS)=c("Pos", "WatsonW", "CrickW", "WatsonC", "CrickC", "revWc", "revCc", "WatsonTotal", "CrickTotal", "WCTotal", "WatsonTotalS", "CrickTotalS","WCTotalS")
  TSS=TSS[padding:(padding+width1+width1),]
  #Unsmoothed data:
  #plot(TSS$Pos-width, TSS$WCTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], xlim=c(-width1,width1))
  #lines(TSS$Pos-width, TSS$WatsonTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="red")
  #lines(TSS$Pos-width, TSS$CrickTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="blue")
  #Smoothed data:
  plot(TSS$Pos-width, TSS$WCTotalS*sae2scalar/Mreads[k], type="l", ylab=c(dflistNames[k],paste("Quantile=",kk,"of",length(quantiles),"/ Total=",kkTotal)))
  lines(TSS$Pos-width, TSS$WatsonTotalS*sae2scalar/Mreads[k], type="l", col="red")
  lines(TSS$Pos-width, TSS$CrickTotalS*sae2scalar/Mreads[k], type="l", col="blue")


#########################################################################################################################

 } # quantile loop
} # Strain loop
plot(TSS$Pos-width, TSS$WCTotal, xaxt="n",yaxt="n", type="n", ylab="", ylim=c(-100,120), axes=F)
text(0,0, font=3, labels=paste("Chromosome",toString(chroms), "/ Hann", win, "/ Y-Scalar",scalar,"/ Quantiles are : RNAseq_Brar_00_EXP"))

dev.off()
```

10.10 Spo11ExtractDC.pl (created by Tim Cooper)

```
#!/usr/bin/env perl
#Version: 1.5

#########################################################################################################################
# Author(s): T.J.Cooper
# Updated: 13/9/2016
# Processes paired-end .SAM files, extracting Watson + Crick coordinate information for double-cut Spo11 libraries
# Quality-control and filtering (atypical read-orientation, dubious ends)
# Two-step alignment (unmapped mate read-trimming, --local alignment)
# Calculates inter-event distances (between double-cut DSBs) and tallies instances of specific double-cuts
#########################################################################################################################

use strict;
use warnings;
use Cwd;
use List::Util qw(first);
my $outext = '.txt';          #Output .file-extension
my $inext = $ARGV[0];  #Input .file-extension
my @files = glob("*$inext");
my $chk = scalar(@files);
print "\nFailed to detect any .SAM files within the current directory.\n\n" if $chk == 0;
exit if $chk == 0;            #Stop script if no .SAM files are found
my $sub = cwd()."/Coordinates";
mkdir("$sub") unless $chk == 0;
my $sub2 = cwd()."/Analysis";
mkdir("$sub2") unless $chk == 0;
my $trimmode = $ARGV[1];
my $trimlength = $ARGV[2];
print "---------------------------------------";
print "\nCalculating Coordinates....\n";
```

```
print "---------------------------------------\n";
print "Currently processing:\n";
for my $file (@files) {   #For-each input file
	open my $IN, '<', $file or die "$!";	#Open and read input .SAM file(s)
	(my $strain = $file) =~ s/_[^_]+$//;	#Strain-name
	(my $mode = $ARGV[0]) =~ s/\.SAM//;				#Alignment-mode
	print "$strain\n";
	my $outfile = $strain."_".$mode.$outext;		#Output files
	my $outfile2 = $strain."_Ambiguous".$outext;
	my ($OUT, $OUT2, $OUT3, $OUT4);
	open $OUT, '>', "$sub/$outfile" or die "$!";
	open $OUT2, '>>', "$sub/$outfile2" or die "$!";
	print $OUT "PairID\tStrand\tChr\tPos\tReadLength\tCIGAR\tAdjustment\n";
	if ($inext eq "Global.SAM" && $trimmode eq "Y") {
		print $OUT2 "Strand\tChr\tPos\tReadLength\tCIGAR\tAdjustment\tMD-Tag\n";
		my $outfile3 = $strain.$ARGV[3]."_unmapped_trimmed.fastq";		#Unmapped R1 FASTQ file
		my $outfile4 = $strain.$ARGV[4]."_unmapped_trimmed.fastq";		#Unmapped R2 FASTQ file
		open $OUT3, '>', "$outfile3" or die "$!";
		open $OUT4, '>', "$outfile4" or die "$!";
	}
	my ($ID,$A,$B,$R1var,$R2var);
	while (<$IN>) {		#For-each .SAM record
		chomp $_;
		next if /^\s*@/; #Skip .SAM headerlines
		my @F = split("\t", $_);  #Split each tab-delimited field
		my $orientation = $F[3]-$F[7];		#Discard atypical read-orientations
		if ($F[1] == 99 && $orientation > 0 || $F[1] == 83 && $orientation < 0) {
			my $skipline = <$IN>;
			next;
		}
		if ($inext eq "Global.SAM" && $trimmode eq "Y") {		#Populate unmapped R1/R2 FASTQ files mapped-unmapped pairs
			if (grep {$_ == $F[1]} 73,137) {
				print $OUT3 "\@$F[0] 1:N:0:1\n$F[9]\n+\n$F[10]\n" if $F[1] == 73;
				print $OUT4 "\@$F[0] 1:N:0:1\n$F[9]\n+\n$F[10]\n" if $F[1] == 137;
			}
			if (grep {$_ == $F[1]} 89,153) {
				$F[9] =~ tr/GATC/CTAG/;
				my $revseq = reverse($F[9]);
				my $revqual = reverse($F[10]);
				print $OUT3 "\@$F[0] 1:N:0:1\n$revseq\n+\n$revqual\n" if $F[1] == 89;
				print $OUT4 "\@$F[0] 1:N:0:1\n$revseq\n+\n$revqual\n" if $F[1] == 153;
			}
			if (grep {$_ == $F[1]} 69,133) {
				my $trimseq = substr($F[9],0,$trimlength);
				my $trimqual = substr($F[10],0,$trimlength);
				print $OUT3 "\@$F[0] 1:N:0:1\n$trimseq\n+\n$trimqual\n" if $F[1] == 69;
				print $OUT4 "\@$F[0] 1:N:0:1\n$trimseq\n+\n$trimqual\n" if $F[1] == 133;
			}
		}
		sub parseSAM {		#Subroutine to interpret SAM-field data
			my @rcd = @_;
			my @read;
			my $index = first{/MD:Z/} @rcd;	#Obtain variable-column MD:Z tag
			my @MDtag = $index =~ /\d+/g;		#Remove non-numeric characters
			my %rules = (M => 1,D => 1,I => 0,S => 1);	#Rules to handle insertion/deletions/matches/soft-clipping
			my ($s,$LS,$RS) = (0)x3;
			while ($rcd[5] =~ /(\d+)([MDIS])/g) {		#Parse and interpret CIGAR code
				my ($n,$op)  = ($1,$2);
				$s += $n * $rules{$op} unless $op eq 'S';		#Calculate POS adjustment (insertions/deletions)
				$LS += $n * $rules{$op} if $op eq 'S' && $-[0]==0;		#(upstream soft-clip)
				$RS += $n * $rules{$op} if $op eq 'S' && $+[0]==length($rcd[5]);		#(downstream soft-clip)
			}
			my $l = length($rcd[9]);  #Read-length
			my $wp = $rcd[3]-$LS;	#Adjusted 5' coordinate (Watson strand)
			my $cp = $rcd[3]+($RS+$s)-1;		#Adjusted 5' coordinate (Crick strand)
			push(@read, $rcd[2],$wp,$cp,$l,$rcd[5],$s,$LS,$RS,$index);
			return(\@read, \@MDtag);
		}
		if (grep {$_ == $F[1]} 99,83) {		#For 99/147 or 83/163 read-pairs
			my $partner = <$IN>;
			my @F2 = split("\t", $partner);		#Split each tab-delimited field
			if ($F[1] == 99) {
				($A, $R1var) = parseSAM(@F); ($B, $R2var) = parseSAM(@F2);
			} else {
				($A, $R1var) = parseSAM(@F2); ($B, $R2var) = parseSAM(@F);
			}
			my @revMDtag = reverse(@{$R2var});
			my @Wat = @{$A}; my @Cri = @{$B};
			my ($chr, $pos, $rl, $cigar, $Lclip, $vtag) = @Wat[0,1,3,4,6,8];
			my ($chrp, $posp, $rlp, $cigarp, $cc, $Rclip, $vtagp) = @Cri[0,2,3,4,5,7,8];
			if ($R1var->[0] == 0 && $R1var->[1] == 0 && $pos > 0 || $revMDtag[0] == 0 && $revMDtag[1] == 0 && $pos > 0 || $Lclip > 1 && $pos > 0 || $Rclip > 1 && $pos > 0) {		#Detect ambigious ends (99/147 or 83/163 pairs)
```

```
                                                printf($OUT2                    "%s\t%d\t%d\t%d\t%s\t%d\t%s\n%s\t%d\t%d\t%d\t%s\t%d\t%s\n",
"w",$chr,$pos,$rl,$cigar,0-$Lclip,$vtag,"c",$chrp,$posp,$rlp,$cigarp,$Rclip+$cc-1,$vtagp);
                                        } else {
                                                $ID++;
                                                printf($OUT                     "%d\t%s\t%d\t%d\t%d\t%s\t%d\n%d\t%s\t%d\t%d\t%d\t%s\t%d\n",
$ID,"w",$chr,$pos,$rl,$cigar,0-$Lclip,$ID,"c",$chrp,$posp,$rlp,$cigarp,$Rclip+$cc-1);
                                        }
                                }
                        }
                        close $IN;
                        close $OUT;
                        close $OUT2;
                        if ($ARGV[0] eq "Global.SAM") {
                                open my $IN2, '<', "$sub/$outfile" or die "$!";
                                my $outfile5 = $file."_Distances".$outext;
                                my $outfile6 = $file."_DoubleCuts".$outext;
                                open my $OUT5, '>', "$sub2/$outfile5" or die "$!";
                                open my $OUT6, '>', "$sub2/$outfile6" or die "$!";
                                <$IN2> for (1..1);
                                my (%IED, %DoubleCut);
                                while (<$IN2>) {
                                        chomp $_;
                                        next if eof;
                                        my @F2 = split("\t", $_); #Split each tab-delimited field
                                        my $partner = <$IN2>;
                                        chomp $partner;
                                        my @F3 = split("\t", $partner);         #Split each tab-delimited field
                                        $IED{(abs($F2[3]-$F3[3])+1)}++;     #Calculates and tallies IEDs
                                        ($F2[3], $F3[3]) = ($F3[3], $F2[3]) if $F2[3] > $F3[3];
                                        $DoubleCut{$F2[2]}{$F2[3]}{$F3[3]}++;           #Tallies double-cut molecules
                                }
                                print $OUT5 "IED\tFreq\n";
                                print $OUT6 "Chr\tCoord-A\tCoord-B\tFreq\n";
                                foreach my $key (sort {$a <=> $b} keys %IED) {
                                        print $OUT5 "$key\t$IED{$key}\n";
                                }
                                foreach my $chr (sort {$a <=> $b} keys %DoubleCut) {
                                        foreach my $coord1 (sort {$a <=> $b} keys %{$DoubleCut{$chr}}) {
                                                foreach my $coord2 (sort {$a <=> $b} keys %{$DoubleCut{$chr}{$coord1}}) {
                                                        print $OUT6 "$chr\t$coord1\t$coord2\t$DoubleCut{$chr}{$coord1}{$coord2}\n"
                                                }
                                        }
                                }
                        }
}
my $run_time = time() - $^T;
print "---------------------------------------";
print "\nRun Completed\n";
print "Processing Runtime: $run_time Seconds\n";
print "---------------------------------------\n\n";
```

10.11 Topo mapping v01 mc.R (created by Matt Neale)

```
setwd("~/Dropbox/Documents/Work Docs/People/Holly Thomas/Top2/Top2_R_scripts")

#########################################################################################################################
require("e1071") # This pacakge permits smoothing functions (used later)
require(stringr)
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
require(doParallel)
require(plyr)
#########################################################################################################################
# Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
Mreads=NULL; dflist=list()
library(doParallel)
cl <- makeCluster(8)
registerDoParallel(cl)
files=c("FullMap.sae2.S288C_MJ475_1U_SAE2_c.txt",
      "FullMap.sae2.S288C_MJ475_1E_SAE2_c.txt",
      "FullMap.sae2.S288C_MJ551_1U_MRE11_c.txt",
      "FullMap.sae2.S288C_MJ551_1E_MRE11_c.txt",
      "FullMap.sae2.S288C_MJ319_1_c.txt")
num=length(files) # Number of files to load

s=foreach (k = 1:num) %dopar% { dflist[[k]] = read.table(files[k], sep = "\t", header=TRUE) } #Import datatable
dflist=s
stopCluster(cl)

AllElementsDUB = read.table("AllElementsDUB.txt", sep = "\t", header=TRUE) #Import datatable
for (i in 1:num){Mreads[i]=sum(dflist[[i]]$Watson+dflist[[i]]$Crick)/1000000} # Calculate Million reads per sample for conveting to HpM
dflistNames=c("MJ475_1U_sae2D",
          "MJ475_1E_sae2D",
          "MJ551_1U_mre11D",
```

```
        "MJ551_1E_mre11D",
        "MJ319_Y135F_1")

BG = read.table("BGreads5.txt", sep = "\t", header=TRUE) #Import hotspot datatable
BGmean=NULL; for (i in 1:num){BGmean[i]=unlist(subset(BG, Strain==dflistNames[i], MeanCore))} # Ensure that background vector BGmean is using same indexe numbering as hit data
#BGmean=Mreads/12.01/Mreads*0.8
Nfactor=(1-(BGmean*12.01)) # Normalisation factor: based on number of reads that appear NOT to be background
Nfactor=Mreads #Supercede previous calculation and just plot everythign based on fraction of million reads per library
Nfactor=BGmean

########################################################################################################################
########### START HERE ONCE DATAFRAMES ARE LOADED ####################
########################################################################################################################
# MODULE for pulling out specific locus of interest
orf="12"
genes=AllElementsDUB #First make a copy of the ALLElements table
upstream=500; downstream=000 # bp to extend by in either direction of ORF
genes=subset(genes, genename==orf | sysname==orf)
xl1=genes$start-upstream
xl2=genes$stop+downstream
chrom=genes$chr

########################################################################################################################
# MODULEor manually specifiying chromosome and location
chrom=12; xl1=460400; window.w=400; xl2=xl1+window.w; #Plot range minimum (bp); # Plot range width (bp); #Plot range maximum (bp)

########################################################################################################################
# New loop to plot multiple comparisions
strains=c(1:5) # Plot these numbered dataframes from the dflist
scalar.u=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1) # Unique scaling factor for each strain in the dflist (default =1 is identical scaling)
scalar.u=1/Nfactor/Mreads # Override scaling factor and use apparent hit reads instead

########################################################################################################################
#Plotting: first set up how the plots are organised. How many panes per image for example using the layout command
plotnumber=length(strains) # Number from 1 to 5
if (plotnumber==1) {layout(matrix(c(1,1,1,2,2),5, 1, byrow = T))}
if (plotnumber==2) {layout(matrix(c(1,1,1,2,2,2,3,3), 8, 1, byrow = T))}
if (plotnumber==3) {layout(matrix(c(1,1,1,2,2,2,3,3,3,4,4),11, 1, byrow = T))}
if (plotnumber==4) {layout(matrix(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5),14, 1, byrow = T))}
if (plotnumber==5) {layout(matrix(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6),17, 1, byrow = T))}

par(mar=c(1,5,1,0),oma = c(0, 1, 1, 1),las=1) # Sets margins per graph and outside margins per grouped set (order is bottom, left, top, right)
layout.show((length(strains)+1))

for (k in strains){ #step through sequentially each dataframe/strain

#Subset for region of interest
sae2.0=subset(dflist[[k]], Chr==chrom & Pos>=xl1 & Pos <=xl2) #Make a sub-table of the sae2-DSB data that only contains those rows where chr = 1 in range of interest

########################################################################################################################
#Decompression code here
sae2.1 <- data.frame(Chr=chrom, Pos=(xl1:xl2)) # Creates expanded empty dataframe with Chr and Pos locations
sae2.1 <- merge(sae2.1,sae2.0, all=TRUE) # Merge expanded empty dataframe with compressed sae2.1 dataframe
sae2.1[is.na(sae2.1)] <- 0 # Convert all NA values to zero
sae2.1$Total=sae2.1$Watson+sae2.1$Crick

########################################################################################################################
# Smoothing function #### temp and smooth are just two temporary vectors. New version creates two smoothed plots for each profile for overlaying
masterscalar=.2
win=1 # hanning window size [1]
win2=1 # hanning window size [2] for overlay
scalar=1*masterscalar #scalar [1]
scalar2=.25*masterscalar # scalar [2] for overlay
sae2scalar=scalar.u[k]*scalar # adjust this if needed when adjusting hann window smoothing
sae2scalar2=scalar.u[k]*scalar2 # adjust this if needed when adjusting hann window smoothing
hw=hanning.window(win) #create hanning window (require package e1071 to be loaded)
hw2=hanning.window(win2) #create hanning window (require package e1071 to be loaded)

########################################################################################################################
temp=NULL
for (j in 3:5){
temp=c(rep(0,win),sae2.1[1:nrow(sae2.1),j], rep(0,win)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
smooth=filter(temp,hw) # smooth the temp vector using the hann window
smooth=smooth[(win+1):(length(smooth)-win)] # trim smooth to correct lengthsmooth2=smooth2[(win2+1):(length(smooth2)-win2)] # trim smooth to correct length
sae2.1[j+3]=smooth
}
temp2=NULL
for (j in 3:5){
  temp2=c(rep(0,win2),sae2.1[1:nrow(sae2.1),j], rep(0,win2)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
  smooth2=filter(temp2,hw2) # smooth the temp vector using the hann window
  smooth2=smooth2[(win2+1):(length(smooth2)-win2)] # trim smooth to correct length
```

```
 sae2.1[j+6]=smooth2
}

colnames(sae2.1)=c("Chr", "Pos", "Watson", "Crick", "Total", "watson.s", "crick.s", "total.s","watson.s2", "crick.s2", "total.s2")

#########################################################################################################################
# Plot boundaries:
plot(sae2.1$Pos,sae2.1$total.s/Mreads[k], type="n", xlim=c(xl1,xl2), ylim=c(-160,100), ylab=paste(c(dflistNames[k]," / Scalar ",round(scalar.u[k],digits=2)), collapse="")) #plot the start histogram
# Broad Overlays:
#lines(sae2.1$Pos,sae2.1$watson.s2*sae2scalar2/Mreads[k], type="h", xlim=c(xl1,xl2), col="lightcoral") #plot the start histogram
#lines(sae2.1$Pos,-sae2.1$crick.s2*sae2scalar2/Mreads[k], type="h", xlim=c(xl1,xl2), col="lightblue") #plot the start histogram
#lines(sae2.1$Pos,0.5*sae2.1$total.s2*sae2scalar2/Mreads[k]-150, type="l", xlim=c(xl1,xl2), col="grey") #plot the start histogram
#hi-res smoothed data:
lines(sae2.1$Pos,sae2.1$watson.s*sae2scalar/Mreads[k], type="h", xlim=c(xl1,xl2), col="red") #plot the start histogram
lines(sae2.1$Pos,-sae2.1$crick.s*sae2scalar/Mreads[k], type="h", xlim=c(xl1,xl2), col="blue") #plot the start histogram
lines(sae2.1$Pos,0.5*sae2.1$total.s*sae2scalar/Mreads[k]-150, type="l", xlim=c(xl1,xl2), col="grey") #plot the start histogram
# overlay of smoothed W/C ratio :
#lines(sae2.1$Pos,100*log2((sae2.1$watson.s2*sae2scalar2/Mreads[k])/(sae2.1$crick.s2*sae2scalar2/Mreads[k])), type="l", xlim=c(xl1,xl2), col="black") #plot the ratio
}
#########################################################################################################################
#Now plot the gene datatrack
#First subset the relevant data
genes=AllElementsDUB #First make a copy of the ALLElements table
genes=subset(genes,chr==chrom & start>(xl1-10000) & stop<(xl2+10000)) #Make a sub-table of ALLElements where chr = 1 and has limits just beyond plot range
genes=subset(genes,type==!"CDS") #Make a sub-table of ALLElements ********* Hash out if you want to plot all elements! ********
genes=subset(genes,type=="rRNA_gene") #Make a sub-table of ALLElements ********* Hash out if you want to plot all elements! ********
#genes=subset(genes,type=="gene") #Make a sub-table of ALLElements ********* Hash out if you want to plot all elements! ********
#Now perform the plot
plot(sae2.1$range,sae2.1$filtered, xaxt="n",yaxt="n",type="n", ylab=paste("Genes"),cex.lab=1.5,font=2, xlim=c(xl1,xl2), ylim=c(-100,120),axes=F) #set up empty plot
text((xl1+xl2)/2,-80, labels=paste("Chromosome",chrom, "/",orf,"/ Range",xl1,"to",xl2,"bp / Hann", win, "/ Y-Scalar",scalar), cex.lab=1.4)
#########################################################################################################################
########## STOP HERE IF YOU ARE PLOTTING WHOLE CHROMOSOMES!!! #############
#########################################################################################################################
# Following module draws arrows for each element
xrange=xl2-xl1
ahead=xrange/25 #make arrowhead length proportional to plot range
ahead[(ahead>500)]=500 #limit max length to 500
av=75 #arrow vertical location relative to plot dimensions
ahw=15 #arrow/head width
genesW=subset(genes,genename !="Dubious_ORF" & orientation =="+") #Make a sub-table of ALLElements
for (i in 1:nrow(genesW)){
  polygon(c(genesW[i,"start"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead,genesW[i,"stop"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead, genesW[i,"start"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="wheat", border="wheat4")
 text((genesW[i,"start"]+genesW[i,"stop"])/2,av, font=3, genesW[i,"genename"], cex=0.9) }
genesW=subset(genes,genename=="Dubious_ORF" & orientation =="+") #Make a sub-table of ALLElements
for (i in 1:nrow(genesW)){
 polygon(c(genesW[i,"start"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead,genesW[i,"stop"], genesW[i,"stop"]-ahead, genesW[i,"stop"]-ahead, genesW[i,"start"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="wheat", border="wheat4", lty=2)
 text((genesW[i,"start"]+genesW[i,"stop"])/2,av, font=3, genesW[i,"sysname"], cex=0.9) }

av=25 #arrow vertical location for Crick genes relative to plot dimensions
genesC=subset(genes,genename !="Dubious_ORF" & orientation =="-") #Make a sub-table of ALLElements
for (i in 1:nrow(genesC)){
 polygon(c(genesC[i,"stop"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead,genesC[i,"start"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead, genesC[i,"stop"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="thistle", border ="thistle4")
 text((genesC[i,"start"]+genesC[i,"stop"])/2,av, font=3, genesC[i,"genename"], cex=0.9) }
genesC=subset(genes,genename=="Dubious_ORF" & orientation =="-") #Make a sub-table of ALLElements
for (i in 1:nrow(genesC)){
 polygon(c(genesC[i,"stop"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead,genesC[i,"start"], genesC[i,"start"]+ahead, genesC[i,"start"]+ahead, genesC[i,"stop"]),c(av+ahw,av+ahw,av+ahw+ahw,av,av-ahw-ahw,av-ahw, av-ahw), col="thistle", border ="thistle4", lty=2)
 text((genesC[i,"start"]+genesC[i,"stop"])/2,av, font=3, genesC[i,"sysname"], cex=0.9) }
```

10.12 Pileups around TSS Topo stratify by transcription v02.R (created by Matt Neale)

```
# Load secondary datasets:
setwd("~/Dropbox/Documents/Work Docs/Research/Lab Notebooks/Computer scripting/R scripts/R datasets/Cer3")

DUB = read.table("AllElementsDUB_Brar_2016.08.16.txt", sep = "\t", header=TRUE) #Import datatable
#Pan = read.table("Pan.Hotspots.IGR.SacCer3_H4L2_2016.08.10a.txt", sep = "\t", header=TRUE) #Import hotspot datatable

#names(Pan) <- c("Chr","Start","End","Length","PanHits","Feature_name","Name","Midpoint", "Type","Direction","IGR","IGR.start","IGR.end")
#Pan <- Pan[c("Chr","Start","End","Length","Midpoint","PanHits","Feature_name","Name","Type","Direction","IGR","IGR.start","IGR.end")] # reorder


#########################################################################################################################
```

```
# YOU MUST CHANGE TO WORKING DIRECTORY CONTAINING THE INPUT DATA MAPS
# Use a temporary directory that just contains COPIES of the files of interest:
setwd("~/Dropbox/Aligned Files/Top2 NEW/Temp")

#########################################################################################################################################
require("e1071") # This pacakge permits smoothing functions (used later)
require(stringr)
options(scipen=999) #Suppresses scientific notation appearing in plots/graphs etc
require(doParallel)
library(doParallel)
require(plyr)
#########################################################################################################################################
# Import histogram FullMap files for each strain in working directory and tally up the total number of Million mapped reads
Mreads=NULL; dflist=list();dflistNames=NULL

#Read in all tables with string "Full.Map."
files = list.files(pattern="FullMap.Cer3") # import files names with "FullMap.Cer32" string into variable "files"
dflistNames = substr(files, 9, nchar(files)-6) # Shorten filename by 8 characters from beginning and 6 characters form end (i.e. remove "FullMap."
and "_c.txt")
nfiles = length(files) # Count number of files

cl <- makeCluster(8)
registerDoParallel(cl)

dflist=foreach (k = 1:nfiles) %dopar% { dflist[[k]] = read.table(files[k], sep = "\t", header=TRUE) } #Import datatable
stopCluster(cl)

#########################################################################################################################################

BG = read.table("BGreadsTopo1.txt", sep = "\t", header=TRUE) #Import background datatable
BGmean=NULL; for (i in 1:nfiles){BGmean[i]=unlist(subset(BG, Strain==dflistNames[i], MeanCore))} # Ensure that background vector BGmean
is using same indexe numbering as hit data

# Calculate Mreads for each datatable
for (i in 1:nfiles){Mreads[i]=sum(dflist[[i]]$Watson+dflist[[i]]$Crick)/1000000}

#########################################################################################################################################
############ START HERE ONCE DATAFRAMES ARE LOADED ######################
#########################################################################################################################################
# Send output to PDF:
wd = getwd(); out = paste(wd,"/","Output_Files","/","TSSPileup_",Sys.time(),".pdf",sep=""); pdf(file=out, width=16,height=16);

# New loop to plot multiple comparisions
strains=c(1,2,3,4,5,6) # Plot these numbered dataframes from the dflist (MAXIMUM of SIX!)
scalar.u=c(1,1,1,1,1,1) # Unique scaling factor for each strain in the dflist (default =1 is identical scaling)
#scalar.u=1/Nfactor # Override unique caling factor and instead use apparnt non-background-read normalisation factor
#ymin=c(0,145,0,0,0)
#ymax=c(250,160,250,250,400)
#########################################################################################################################################
#Plotting: first set up how the plots are organised. How many panes per image for example using the layout command
plotnumber=length(strains) # Number from 1 to 5
if (plotnumber==1) {layout(matrix(c(1,2,1,2,3,3),3, 2, byrow = T))}
if (plotnumber==2) {layout(matrix(c(1,2,1,2,3,4,3,4,5,5), 5, 2, byrow = T))}
if (plotnumber==3) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,7),7, 2, byrow = T))}
if (plotnumber==4) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,9),9, 2, byrow = T))}
if (plotnumber==5) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,11),11, 2, byrow = T))}
if (plotnumber==6) {layout(matrix(c(1,2,1,2,3,4,3,4,5,6,5,6,7,8,7,8,9,10,9,10,11,12,11,12,13,13),13, 2, byrow = T))}
par(mar=c(1,4,1,2),oma = c(1, 1, 1, 1),las=1) # Sets margins per graph and outside margins per grouped set (order is bottom, left, top, right)
#layout.show((length(strains)*2)+1)

#########################################################################################################################################
# MODULE for pulling out specific locus of interest
genes=DUB #First make a copy of the ALLElements table
genes=subset(genes, type=="gene" & genename!="Dubious_ORF")
genesW=subset(genes, orientation=="+")
genesC=subset(genes, orientation=="-")
#########################################################################################################################################
k=1
for (k in strains) { #step through sequentially each dataframe/strain
temp=NULL
chroms=c(1:16)
#chroms=c(1:2,4:16) # Specify the chromosomes you want to plot
#chroms=c(2) # Or just a single chromosome. Hash this line out if you want to plot all of them specified above.

width1=1000 # bp upstream and downstream of TSS to retrieve
padding=500
width=width1+padding

#Initialise TSS dataframe
TSS=data.frame(NULL)
TSS[1:((width*2)+1),"Pos"]=(1:((2*width)+1))
TSS$WatsonW=0 # Watson genes watson hits
TSS$CrickW=0 # Watson genes crick hits
TSS$WatsonC=0 # Crick genes watson hits
TSS$CrickC=0 # Crick genes crick hits
```

```
#### Start of Chromosome loop
for (chrom in chroms) { #Specify which chromosomes to process
df.1=subset(dflist[[k]],Chr==chrom) # Make copy of the datatable
#sae2.1=sae2.0 #Second copy (no longer any need to decompress datatables)
#sae2.1$Total=sae2.1$Watson+sae2.1$Crick

temp=NULL

genesW1=subset(genesW, chr==chrom) #Subset for Watson genes
genesC1=subset(genesC, chr==chrom) #Subset for Crick genes

# The following lines superced the preceding ones in order to first subtract a mean level of background from the zeroed out array based on number of genes to be piled-up and BGmean
#TSS1500[1:((width1*2)+1),"WatsonW"]=0-(BGmean[k]/2*nrow(genesW1))                    #TSS1500[1:((width1*2)+1),"CrickW"]=0-(BGmean[k]/2*nrow(genesW1)) #TSS1500[1:((width1*2)+1),"TotalW"]=0-(BGmean[k]*nrow(genesW1))
#i=1

for (i in 1:nrow(genesW1)) {
   temp=subset(df.1[2:4], Pos>=(genesW1[i,"start"]-width) & Pos<=(genesW1[i,"start"]+width)) # Create temp matrix with DSB hits in this region
   temp$Pos=temp$Pos-genesW1[i,"start"]+width+1
   TSS[temp$Pos,"WatsonW"]=TSS[temp$Pos,"WatsonW"]+temp[,"Watson"]
   TSS[temp$Pos,"CrickW"]=TSS[temp$Pos,"CrickW"]+temp[,"Crick"]
}
for (i in 1:nrow(genesC1)) { #NOTE: TSS is at the stop location in the table
 temp=subset(df.1[2:4], Pos>=(genesC1[i,"stop"]-width) & Pos<=(genesC1[i,"stop"]+width)) # Create temp matrix with DSB hits in this region
 temp$Pos=temp$Pos-genesC1[i,"stop"]+width+1
 TSS[temp$Pos,"WatsonC"]=TSS[temp$Pos,"WatsonC"]+temp[,"Watson"]
 TSS[temp$Pos,"CrickC"]=TSS[temp$Pos,"CrickC"]+temp[,"Crick"]
}

} # Repeat for each chromosome

# This section combines the Watson and Crick hits, reversing the order of the crick gene hits and adding the Watsons to the Crick hits and Crick to Watsons (i.e. reverse complements the data for the Crick genes).
# This allows both to be combined in the plot
TSS$revWc=rev(TSS$WatsonC) #Crick genes with Watson hits
TSS$revCc=rev(TSS$CrickC) #Crick genes with Crick hits
TSS$WatsonTotal=TSS$WatsonW+TSS$revWc
TSS$CrickTotal=TSS$CrickW+TSS$revCc
TSS$WCTotal=(TSS$WatsonTotal+TSS$CrickTotal)/2 #Create total column and halve it

#########################################################################################################################################
# Smoothing function #### temp and smooth are just two temporary vectors
win=151 # hanning window size
scalar=.1
sae2scalar=scalar.u[k]*scalar # adjust this if needed when adjusting hann window smoothing
hw=hanning.window(win) #create hanning window (require package e1071 to be loaded)

temp=NULL
for (j in 8:10){ #Smooth columns 8 to 10 in the TSS table
 temp=c(rep(0,win),TSS[1:nrow(TSS),j], rep(0,win)) # Create vector length of chromosome and extend by the length of the slidign window with zeros at both ends
 smooth=filter(temp,hw) # smooth the temp vector using the hann window
 smooth=smooth[(win+1):(length(smooth)-win)] # trim smooth to correct length
 TSS[j+3]=smooth
}
colnames(TSS)=c("Pos", "WatsonW", "CrickW", "WatsonC", "CrickC", "revWc", "revCc", "WatsonTotal", "CrickTotal", "WCTotal", "WatsonTotalS", "CrickTotalS","WCTotalS")
TSS=TSS[padding:(padding+width1+width1),]
#Unsmoothed data:
#plot(TSS$Pos-width, TSS$WCTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], xlim=c(-width1,width1))
#lines(TSS$Pos-width, TSS$WatsonTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="red")
#lines(TSS$Pos-width, TSS$CrickTotal*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="blue")
#Smoothed data:
plot(TSS$Pos-width, TSS$WCTotalS*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], ylim=c(1200,2700))
lines(TSS$Pos-width, TSS$WatsonTotalS*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="red")
lines(TSS$Pos-width, TSS$CrickTotalS*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], col="blue")

#########################################################################################################################################
##TSS1500=subset(TSS1500, Pos>=padding & Pos <=(width*2)+padding)
#plot(TSS1500$Pos-width1, TSS1500$TotalWS*sae2scalar/Mreads[k], type="l", ylab=dflistNames[k], ylim=c(ymin[k],ymax[k]))
#lines(TSS1500$Pos-width1, TSS1500$WatsonWS*sae2scalar/Mreads[k]*2, type="l", ylab=dflistNames[k], col="red")
#lines(TSS1500$Pos-width1, TSS1500$CrickWS*sae2scalar/Mreads[k]*2, type="l", ylab=dflistNames[k], col="blue")

#plot(TSS1500$Pos-width1, TSS1500$TotalCS*sae2scalar/Mreads[k], type="l", ylab=paste(c("Watson", "\n", "Crick")), yaxt="n", ylim=c(ymin[k],ymax[k]))
#axis(4)
#lines(TSS1500$Pos-width1, TSS1500$WatsonCS*sae2scalar/Mreads[k]*2, type="l", ylab=dflistNames[k], col="red")
#lines(TSS1500$Pos-width1, TSS1500$CrickCS*sae2scalar/Mreads[k]*2, type="l", ylab=dflistNames[k], col="blue")
}

plot(TSS$Pos-width, TSS$WCTotal*sae2scalar/Mreads[k], xaxt="n",yaxt="n", type="n", ylab="", ylim=c(-100,120), axes=F)
text(0,0, font=3, labels=paste("Chromosome",toString(chroms), "/ Hann", win, "/ Y-Scalar",scalar))

dev.off()
```