Synthesis-Dependent Microhomology-Mediated End Joining: A Unifying Model For Multiple DNA Double Strand Break Repair Outcomes

A dissertation submitted by

Amy Marie Yu

In partial fulfillment of the requirements

For the degree of Doctor of Philosophy

in

Biology

TUFTS UNIVERSITY

May 2011

Advisor: Dr. Mitch McVey

Abstract

DNA double-strand breaks (DSBs) are repaired by homology-directed repair or non-homologous end joining. End-joining repair can be classified as classical non-homologous end joining, which requires DNA ligase 4, or "alternative" end joining (alt-EJ), which does not. The mechanism of alt-EJ is largely uncharacterized.

We show that *Drosophila melanogaster* DNA polymerase theta (pol theta), encoded by the *mus308* gene, plays a role in alt-EJ. In pol theta mutants, end joining is impaired, creating large deletions. Sequencing suggests that pol theta promotes the use of long microhomologies during alt-EJ. These results suggest a mechanistic link between alt-EJ and interstrand crosslink repair**.**

Alt-EJ repair frequently correlates with junctional microhomology but also produces junctions without microhomology, and the role of microhomology in alt-EJ remains unclear. To investigate this, we examined inaccurate repair of an I-*Sce*I DSB with few nearby microhomologies in Drosophila. Lig4 deficiency did not affect junctional microhomology, but significantly increased insertions. Many insertions appeared templated. Based on sequence analysis, we propose a model of synthesis-dependent microhomology-mediated end joining (SD-MMEJ), in which *de novo* synthesis by an accurate non-processive DNA polymerase creates microhomology. Repair junctions with apparent blunt joins, junctional microhomologies, and short indels (deletion with insertion) are often considered to reflect different repair mechanisms. However, a majority of each type had structures consistent with the predictions of our SD-MMEJ model, suggesting that a single underlying mechanism could be responsible for all three. SD-MMEJ is Ku70, Lig4, and Rad51-independent, but impaired in *mus308* (POLQ) mutants.

Although the SD-MMEJ model is a powerful conceptual tool, the question remained whether or not patterns associated with SD-MMEJ would be found at other DSBs. A meta-

analysis of end-joining repair junctions produced in the course of previous studies in Drosophila showed that substantial numbers of the junction sequences had SD-MMEJ consistent repeats. Patterns of repeats varied in a sequence specific manner consistent with the predictions of the SD-MMEJ model. Correlating the patterns of SD-MMEJ consistent repeats at breaks from C-NHEJ proficient and deficient backgrounds with sequence content and break end structure confirms a complex relationship among sequence content, genetic background, and end-joining repair outcome.

Acknowledgements

My advisor, **Dr. Mitch McVey**, and committee: **Dr. Susan Ernst, Dr. Catherine Freudenreich**, and **Dr. Sergei Mirkin**

The McVey Lab present and past research technicians **Alice Witsell** and **Justine Liepkalns** for making the fly food and doing all of the other ten thousand things that made these experiments possible, as well as our student fly caretakers **Lisa Netkowicz**, **Nicole Cardona**, **Mary Bruynell**, and **Lena Cantone**.

The Biology Department administrative staff: **Eileen Magnant**, **Liz Palmer**, **Karin Murphy**, **Michael Grossi**, **Anthony Keevan**, and **Phil Bibb**, who know everything about where everything is, how it works (or why it doesn't, and what to do about it), and have been so generous with their time, patience, and good cheer.

Department Chair and past Graduate Program Chair **Dr. Juliet Fuhrman** and current graduate chair **Dr. Kelly McLaughlin**.

My labmates: graduate students **Dan Kane**, **Adam Thomas**, and **Kelly Beagan**; postdoctoral fellow **Dr. Elyse Bolterstein**, and undergraduates **Amelia Sinkin**, **Ilana Traynis**, **Sarah Rubin**, **Greg Mayes**, **Endry Martinez**, **Karina Sanchez**, **Lisa Tran**, **Michael Shusterman**, **Lauren Watts**, **Amanda Albin**, **Caitlin Peterman**, and **Rachel Cox**. I would like to extend especially warm thanks to the undergraduates with whom I worked most closely: **Bena Chan**, **Camille Petersen**, and **Amy Baker**, who have taught me so much about how to learn. Thank you all for creating the heart and soul of the McVey Lab.

Dr. Jane Kim for insightful feedback on this manuscript.

All of my colleagues in the **Tufts Biology Department**, from whom I have learned that it really is true that all of us is smarter than any of us.

The organizers of and speakers at the **MIT Center For Environmental Health Sciences DNA Repair And Mutagenesis Group** series, who have provided a first-class education in DNA repair.

The people and resources that enabled my crash course in software engineering: **Tom Yu** and **MIT OpenCourseWare**.

Dr. Yikong Rong and **Dr. Bill Engels** for the generous gift of fly stocks.

Dedication

For my teacher,

Linda Walker Indek

and my fellow student,

Richard Aldous

Table of Contents

Abstract ii
Acknowledgements iv
Dedication v
List Of Abbreviations xi

Chapter 1 1

Introduction 1
DNA Double Strand Breaks: Definition And Biological Significance 2
DSBs are repaired by homologous recombination or by end joining 2
Eukaryotic cells possess multiple genetically distinct end joining mechanisms 3
Table 1-1. Differences in genetic requirements for the three major events comprising end joining via C-NHEJ versus alt-EJ in metazoans. 4
The mechanism and genetics of alt-EJ are poorly understood 5
Microhomology-Mediated End Joining: Basic Mechanism And Sequence Analysis 7
Figure 1-1. Sequence requirements and outcome of microhomology-mediated end joining (MMEJ) 7
Evidence for MMEJ is widely observed 8
Detailed consideration of possible MMEJ mechanisms helps interpret genetic data 9
Figure 1-2. Possible MMEJ mechanisms can be distinguished based on different enzymatic requirements. 10
Table 1-2. Nucleases involved in MMEJ 11
Table 1-3. Signaling molecules implicated in MMEJ 12
Table 1-4. Helicases implicated in MMEJ. 12
Table 1-5. Mismatch repair factors implicated in MMEJ 13
Table 1-6. DNA polymerases implicated in MMEJ 14
Table 1-7. DNA ligases implicated in MMEJ 15
MMEJ at microhomologies of different lengths may be genetically distinct 15
The Relationship Between alt-EJ And MMEJ Is Not Straightforward 16
Genetic requirements for MMEJ overlap those for both C-NHEJ and alt-EJ 17
Non-genetic factors affect frequency and length of junctional microhomology 18
Evidence for non-MMEJ alt-EJ 18
Patterns in observed microhomologies are not consistent with the "alt-EJ end glue" model 19
A Modeling Approach To Sequence Analysis Uncovers Hidden Assumptions About The MMEJ Model And Provides Alternative Interpretations 20
Statistical arguments assume specific mechanisms 20
Table 1-8. A formula commonly used for calculating the probability of observing junctional microhomology by random chance. 21
A modeling approach reveals alternative interpretations of shifts in average microhomology lengths 22
Table 1-9: MMEJ at repeated motifs consisting of a mononucleotide run. 23
Table 1-10: MMEJ with asymmetric nuclease activity at repeated motifs consisting of a mononucleotide run 23
Outline Of Subsequent Chapters 25

Chapter 2: Dual Roles For DNA Polymerase Theta In Alternative End-Joining Repair Of Double-Strand Breaks In Drosophila 26

Introduction 27
Results 32
Drosophila pol θ is important for both interstrand crosslink repair and end-joining repair of DNA double strand breaks 32
Figure 2-1. Schematic of the mus308 gene. 32
Figure 2-2. Drosophila pol θ is required for repair of IR–induced damage in the absence of Rad51 34
Figure 2-3. Pol θ is involved in end-joining repair. 36
Figure 2-4. Pol θ–dependent end joining acts independently of HR and C-NHEJ. 41
Figure 2-5. Flies lacking Rad51 and Pol θ have synergistic phenotypic defects 43
Pol θ-mediated end joining is distinct from C-NHEJ 45
Pol θ has two distinct functions in alternative end joining 46
Figure 2-6. Pol θ is required for two distinct classes of alt–EJ repair products. 47
Table 2-1. *P{w^a}* repair junctions recovered from *spn-A$^{093/057}$* mutants. 49
Table 2-2. *P{w^a}* repair junctions recovered from *spn-A$^{093/057}$ mus308$^{D5/2003}$* mutants. 50
Table 2-3. *P{w^a}* sequences recovered from *spn-A$^{093/057}$ mus308$^{D2/2003}$* mutants. 51
Pol θ operates in alternative end-joining of complementary-ended double-strand breaks 52
Figure 2-7. Model for pol θ function in alt–EJ 55
Discussion 56
Evidence for two functions of pol θ in alternative end joining 57
Potential roles of pol θ in DNA interstrand crosslink repair 59
Pol θ and alternative end joining: promoting genome (in)stability 60
Conclusions 62
Materials And Methods 63
Acknowledgements 66

Chapter 3: Synthesis-Dependent Microhomology-Mediated End Joining Accounts For Multiple Types Of Repair Junctions 67

Introduction 68
Results 69
Drosophila alt-EJ does not require long microhomologies 69
Figure 3-1. The *[lw]* construct. 69
C-NHEJ deficiency does not alter patterns of short junctional microhomology 71
Figure 3-2. C-NHEJ deficiency increases frequency of net insertions. 71
Alternative end joining produces insertions at repair junctions 72
Genotype specific effects of sequence context on repair outcome 73
Figure 3-3. Combined effects of sequence context and genotype on deletion boundary frequency 73
Templated insertions and deletion boundaries correlate with specific patterns of short repeats in the original sequence 75
Figure 3-4. Repair products contain repeated motifs with a consistent pattern. 75
Figure 3-5. Loop-out SD-MMEJ produces repair products with direct repeats. 77
Figure 3-6. Snap-back SD-MMEJ produces repair products with inverted repeats. 78
The SD-MMEJ model shows that potential templates can be determined for short as well as long insertions 79
SD-MMEJ repair can produce apparent blunt joins and microhomology junctions 80

Figure 3-7. SD-MMEJ repair can produce repair junctions without insertions, which may or may not have junctional microhomology. 80
SD-MMEJ consistency criteria for specific junction structures 81
Figure 3-8. An apparently templated complex insertion that is not SD-MMEJ consistent 83
Significant majorities of apparent blunt joins, microhomology junctions, and short insertions are SD-MMEJ consistent 84
Figure 3-9. A majority of 1-3 bp insertions, apparent blunt joins, and microhomologies are SD-MMEJ consistent. 84
Categorization of repair products according to SD-MMEJ parameters supports a single underlying mechanism 86
Table 3-1. Categorizing SD-MMEJ consistent repair junctions by primer repeat reveals structural similarities among multiple repair junction types. 87
Figure 3-10. Most SD-MMEJ consistent repair junctions accounted for by four primer repeats. 90
Figure 3-11. Multiple mechanisms for alternative end joining 91
Figure 3-12. SD-MMEJ Explains Patterns In Frequency Of Junctional Microhomologies. 92
Additional Genetic Studies 93
mus205 (DNA polymerase zeta) 93
nbs1 94
Discussion 95
Previous reports of apparently templated insertions 96
Antecedents and alternatives to the SD-MMEJ model 99
Materials And Methods 103

Chapter 4: Investigating The Generality Of The SD-MMEJ Model 106

Introduction 107
Results (I): Modeling And Analysis Of Original Sequences 108
Computational analysis of SD-MMEJ consistency criteria 108
Figure 4-1. Effect of minimum SD-MMEJ consistent repeat length and total base pairs searched on the frequency of SD-MMEJ consistent randomly generated apparent blunt joins 111
Table 4-1. Effect of base composition on probability of fortuitous SD-MMEJ consistency 112
Table 4-2. Effect of microhomology length on the probability of SD-MMEJ consistency. 114
Table 4-3. Estimates of expected frequency of fortuitously SD-MMEJ consistent apparent blunt joins and microhomology junctions for the DSB substrates used in this meta-analysis. 115
Properties Of The Double Strand Break Substrates 115
I-*Sce*I DSBs: the *[Iw]7* and Rr3 constructs 116
Table 4-4. Structures of molecularly defined I-*Sce*I DSBs 116
Table 4-5. Alignment of the sequences of the *[Iw]7* and Rr3 DSB substrates. 116
Zinc finger nuclease DSBs 118
Table 4-6. Sequences of the zinc finger nuclease (ZFN) DSB substrates. 118
Patterns of deletion boundaries at the Rr3, *yellow* and *rosy* DSBs are consistent with SD-MMEJ repair 120
Figure 4-2. Deletion boundary frequency histograms for deletions to the left and right of the Rr3 DSB 121
Figure 4-3. Deletion boundary frequency histogram for the *yellow* DSB 121
Figure 4-4. Deletion boundary frequency histograms by genotype for the *rosy* DSB 122

Gap repair subsequent to transposon (*P* element) excision ... 124
Results (II): Sequence Analysis Of Repair Junctions ... 126
Substantial variation in relative proportions of apparent blunt joins, microhomology junctions, and insertion junctions at different double strand breaks ... 127
Figure 4-5. Relative proportions of different junction types by DSB substrate ... 127
Table 4-7. Estimated proportions of junction types by random chance for DSB substrates ... 129
The SD-MMEJ model suggests that most short insertions in most systems are templated from surrounding sequence ... 130
Figure 4-6. Proportion of SD-MMEJ consistent 1-3 bp insertions by DSB and Lig4 status ... 130
Table 4-8. Examples of representative patterns of SD-MMEJ consistent repeats at repair junctions with 1-3 bp insertions. ... 132
Table 4-10. Non SD-MMEJ consistent short insertions recovered from the yellow DSB. 134
Table 4-11. Non SD-MMEJ consistent short insertions recovered from the *rosy* DSB ... 135
Microhomology junctions ... 136
Figure 4-7. Proportions of SD-MMEJ consistent microhomology junctions ... 136
Table 4-12. Representative patterns of SD-MMEJ consistent microhomology junctions. 138
Table 4-13. SD-MMEJ consistent microhomology junctions recovered from the *yellow* DSB. ... 140
Apparent blunt joins ... 142
Figure 4-8. Proportion of SD-MMEJ consistent apparent blunt joins by DSB and Lig4 status ... 142
Table 4-14. Representative sequences showing patterns of SD-MMEJ consistent apparent blunt joins ... 143
Discussion ... 146
Table 4-15. The number of possible SD-MMEJ events is large. ... 147
Materials And Methods ... 150
Table 4-16. Sample output from SD-MMEJ analysis software ... 153
Table 4-17. Sample random deletions output from SD-MMEJ analysis software ... 155
Table 4-18. Sample output from SD-MMEJ analysis software ... 156
Chapter 5 ... 159
Future Directions ... 159
Appendix A ... 163
Introduction ... 164
Figure A1-1. Experimental strategy for rapid screening of end-joining repair junctions in Drosophila embryos ... 165
Materials And Methods ... 165
Results ... 167
Figure A1-2. End-joining repair junctions obtained from initial studies in the plasmid embryo injection assay. ... 167
Discussion ... 168
Works Cited ... 169
References For Chapter 1 ... 170
References For Chapter 2 ... 177
References For Chapter 3 ... 183

References For Chapter 4 187
References for Chapter 5 189
References for Appendix A 190

List Of Abbreviations

alt-EJ: Alternative end joining

ATM: Ataxia telangiectasia mutated

ATR: Ataxia telangiectasia and Rad3-related protein

C-NHEJ: Canonical non-homologous end joining

CSR: Class switch recombination

DSB: (DNA) double strand break

FLT3: FMS-like tyrosine kinase-3 receptor

HR: Homologous recombination

ICL: Interstrand (DNA) crosslink

MMEJ: Microhomology-mediated end joining

PARP: Poly(ADP)-Ribose Polymerase

SD-MMEJ: Synthesis-dependent microhomology-mediated end joining

T-DNA: Transfer DNA (derived from the Ti plasmid present in plant tumor-inducing bacteria)

T-nucleotides: Apparently templated nucleotides inserted at a DSB repair junction

ZFN: Zinc finger nuclease

Chapter 1

Introduction

DNA Double Strand Breaks: Definition And Biological Significance

DNA double strand breaks (DSBs) are molecular lesions in which covalent bonds are broken in both phosphodiester backbones in sufficient proximity that the ends come apart. DSBs are a diverse set of lesions, ranging from "clean" breaks that can be directly re-ligated with no further processing, to "dirty" breaks that involving extensive chemical damage to the DNA. Even if the ends are chemically competent for direct re-ligation, however, all DSBs pose a serious threat to genome stability because both strands of the double helix are affected, incurring the potential for loss of genetic information, chromosomal rearrangements, or both. Even so, deliberate induction of DSBs is an essential component of essential biological processes such as meiosis and V(D)J recombination. The wide variety of possible DSB end structures, as well as differential requirements for speed and accuracy in repair in different situations, necessitates that DSB repair processes be highly flexible and tightly regulated.

DSBs are repaired by homologous recombination or by end joining

DSB repair mechanisms are broadly classified as homologous recombination (HR) or end joining (EJ) (reviewed in (1)). HR and EJ are distinguished both genetically and mechanistically. The relative importance of HR and EJ varies among organisms; budding yeast rely largely on HR, whereas end joining is more prominent in vertebrates.

Genetically, homologous recombination repair is distinguished from end joining by its dependence on members of the Rad52 epistasis group. HR requires an external template, usually a sister chromosome or homologous chromatid. Repair synthesis from this external template results in repair with a high probability of maintaining genomic integrity, even if the DSB involves damage to both DNA strands. Although it preserves genetic information, HR is limited in that it is relatively slow, and usually restricted to G2 or late S phase.

In contrast to homologous recombination repair, end-joining repair ligates the broken ends without provision for ensuring accurate restoration of the original sequence. Conceptually, end-joining repair involves three key tasks: synapsis and alignment of the DNA ends, rendering the ends chemically competent for re-ligation, and resealing the phosphodiester backbone.

The possibility of accurate repair via end joining depends on the structure of the break. End joining of blunt or complementary ended breaks can perfectly restore original sequence although, interestingly, inaccurate end-joining repair of complementary ends does occur.

Eukaryotic cells possess multiple genetically distinct end joining mechanisms

Eukaryotic cells possess multiple genetically distinct end-joining mechanisms. The best-characterized is commonly referred to as "canonical non-homologous end joining" (C-NHEJ) (2). C-NHEJ requires two conserved "core" complexes: the DNA Ligase 4 (Lig4)/XRCC4 and the Ku70/80 (Ku) heterodimers. In vertebrates, Ku70/80 complexes with the DNA-dependent protein kinase (DNA-PKcs), binds the break ends and recruits other repair factors, such as nucleases, kinases, or DNA polymerases, as dictated by the requirements of the specific DSB (3, 4). Lig4 carries out the final ligation. (5, 6) and likely plays a role in end bridging and alignment (5, 7). Although not strictly necessary, if complementary single stranded regions are present at break ends, base pairing can stimulate C-NHEJ (6).

Table 1-1. Differences in genetic requirements for the three major events comprising end joining via C-NHEJ versus alt-EJ in metazoans.

	end synapsis	end processing (polymerase)	end processing (nuclease)	re-ligation
C-NHEJ	**Ku70/80** Lig4	Pol lambda Pol mu	Artemis	**Lig4/XRCC4** Cernunnos/XLF
alt-EJ	**base pairing** MRN/X PARP	Pol mu	MRN/X CtIP	**Lig3** Lig1

Major contributors are indicated in bold.

Observation of end-joining activity in C-NHEJ deficient cells led to the discovery that eukaryotes also possess "alternative" end joining mechanisms (alt-EJ), which are broadly defined as any end-joining activity remaining in a C-NHEJ deficient genetic background (8, 9, 10, 11). Ku has been shown to recruit Lig4 to break ends (12, 13), suggesting that mutations in Ku and Lig4 should be largely equivalent. However, there is some evidence for different phenotypes in Ku versus Lig4-deficient cells (14, 15, 16). Our study described in Chapter 3 supports the notion that Ku and Lig4 deficient alt-EJ may be subtly different. Little is known about these differences, however; therefore, we will adhere to the conventional definition of alt-EJ as any end-joining pathway that does not require Ku or Lig4, making further distinctions where appropriate.

Alt-EJ has proven difficult to study in vertebrates, as C-NHEJ deficiency in vertebrates is either lethal or associated with profound immunodeficiency and cancer. Lig4 deficiency is embryonic lethal in mice (17). Lethality is rescued in a p53 mutant background, but *lig4 p53* double mutant mice exhibit a high rate of translocations and other chromosomal aberrations, and invariably develop cancers such as medulloblastoma (18). These severe phenotypes, along with kinetic data suggesting that C-NHEJ independent end joining is relatively slow (19, 20), led to the general assessment that alt-EJ is less efficient and less accurate than C-NHEJ. Evidence that

the C-NHEJ components actively inhibit alt-EJ suggested that it might not operate in wild-type cells (21).

However, a number of studies over the past several years have challenged the notion that alt-EJ is inefficient and irrelevant in C-NHEJ competent vertebrate cells. Unexpectedly efficient alt-EJ has been demonstrated in mice (22, 23, 24, 25), and evidence has surfaced indicating that alt-EJ may take place in C-NHEJ competent cells; in fact, some DSBs may be preferentially repaired by alt-EJ (26, 27). Interestingly, C-NHEJ deficient Drosophila are viable and fertile, rendering the fly one of the few multicellular model organisms in which alt-EJ can be readily observed.

The mechanism and genetics of alt-EJ are poorly understood

Substantial uncertainty still exists regarding the genetic requirements and mechanism of alt-EJ, up to and including the question of whether there are one or several mechanisms mediating alt-EJ. Genetic studies have identified a number of enzymes involved in alt-EJ (Table 1-1), including DNA ligases 1 and 3, poly-ADP-ribose polymerase 1 (PARP-1) (28), and the MRN/X complex (28). Lethality in *Ku86*-/- *XPF*-/- double mutant mice also implicates the XPF endonuclease in alt-EJ (29). Further work is required to connect these observations into a coherent whole.

Candidates exist for the end processing and ligation steps of alt-EJ, but the question of how break ends are synapsed in C-NHEJ deficient cells has less of a clear answer. One possibility is that protein factors could help tether the ends. There is evidence for PARP-dependent end synapsis (30), and the MRN complex has a structure suggesting that it could help tether break ends together (31). It has also been observed that joining of DSBs with long GC-rich overhangs is less dependent on C-NHEJ (32, 33), leading to the hypothesis that in alt-EJ,

break end synapsis could be mediated via base pairing between complementary sequences on opposite sides of the DSB, a mechanism termed microhomology-mediated end joining, or MMEJ (reviewed in (34)). There is indeed abundant evidence suggesting that cross-break pairing may be more important in C-NHEJ deficient backgrounds, but, as we will argue, the significance of these observations, and the precise relationship between the genetically defined alt-EJ and the mechanistically defined MMEJ, is unclear.

Microhomology-Mediated End Joining: Basic Mechanism And Sequence Analysis

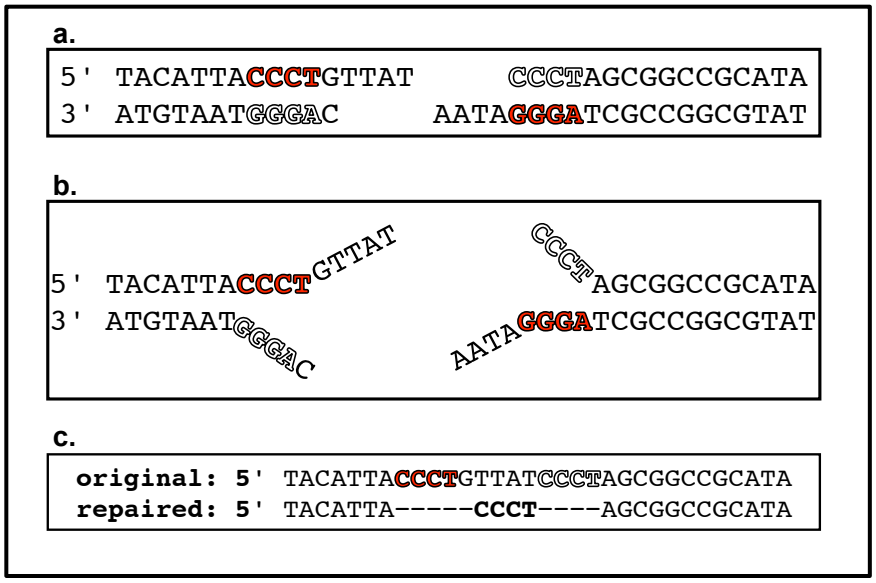


Figure 1-1. Sequence requirements and outcome of microhomology-mediated end joining (MMEJ)

a. MMEJ requires a direct repeat spanning the DSB. In this example, MMEJ takes place at the CCCT repeat.

b. Complementary single stranded regions are exposed. In this example, the DNA is unwound, but nuclease activity could also produce the necessary single stranded regions.

c. Removal of unpaired single stranded DNA and nick ligation results in net deletion of one copy of the repeated motif and all intervening sequence (here, GTTAT). The resulting junction has ambiguous nucleotides ("junctional microhomology") that cannot be definitively assigned to either side of the break when the original sequence and repair product are aligned.

In MMEJ, end joining proceeds via annealing complementary single stranded regions on either side of the DSB (Figure 1-1). Removal of the unpaired regions and ligation generates a repair product in which one motif of the repeat, along with the sequence intervening between the repeated motifs, has been deleted. Thus, when the MMEJ repair product is aligned with the original sequence, the junction contains bases that cannot be unambiguously assigned to either side of the deletion ("junctional microhomology").

There is no clear consensus regarding how many ambiguous bases need to be present at the junction before a repair event is deemed to be MMEJ. There is a nontrivial probability that at least one base of microhomology will exist purely by random chance at any given end-joining junction. MMEJ can be defined conceptually as end-joining that strictly requires base pairing across the break, but this is difficult to verify experimentally, especially if MMEJ coexists with other end-joining repair pathways that are enhanced by, but do not require, microhomology pairing. This difficulty is usually circumvented by scoring relative changes in the frequency and/or length of microhomology junctions in different genetic backgrounds. However, this does not address questions such as whether an increment of a base pair or two in the median junctional microhomology length is biologically meaningful, even if highly statistically significant. To address this question, we will evaluate the data with particular attention to inferring the *function* of microhomology pairing in the end-joining mechanism.

Evidence for MMEJ is widely observed

Evidence for MMEJ is documented in a wide range of organisms from yeasts through vertebrates. It is thus a general phenomenon, although details may vary between species. Microhomology annealing appears to play a role in a wide variety of mutagenic events, including telomere fusions (35), sequence capture at DSBs (36), retrovirus and retrotransposon integration (37, 38), T-DNA integration in plants (39), chromosomal translocations (40, 41), nonrecurrent microduplications and microdeletions (42), and interspecies horizontal gene transfer (43).

Since identification of microhomology junctions requires knowledge of the original sequence, the earliest studies of MMEJ were necessarily carried out with molecularly defined DSB substrates, and such substrates are still essential for carrying out controlled studies to isolate the contribution of sequence context and genetic background to MMEJ. Early reports of

MMEJ came from studies observing recircularization of restriction enzyme-digested plasmid substrates with a variety of end structures in SV40 cells (44) and Xenopus embryo extract systems (45).

Concern that plasmid recircularization might not accurately reflect all aspects of MMEJ at a chromosomal DSB led to development of chromosomally integrated MMEJ reporter substrates. Non-complementary ended DSBs induced by multiple HO endonuclease cuts have facilitated detailed studies of the genetics of MMEJ in *S. cerevisiae* (26, 46). In Drosophila, the long non-complementary 3' overhangs produced by P element transposon excision have been found to be good substrates for MMEJ (27, 47).

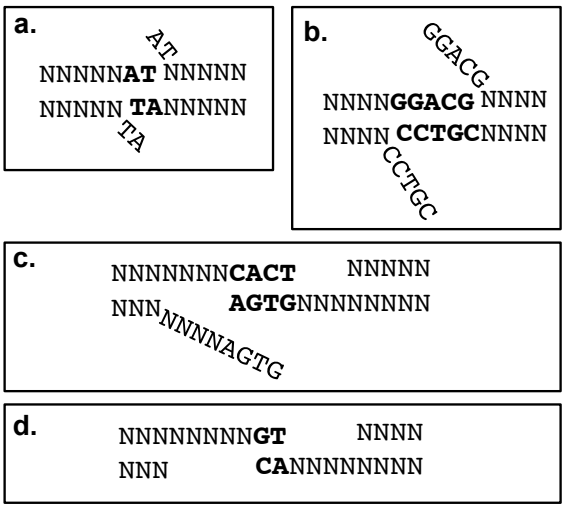
Molecularly defined DSBs induced by the yeast homing endonuclease I-*Sce*I are widely used to study MMEJ. I-*Sce*I recognizes a long (18 bp) target site not normally present in many metazoan genomes. I-*Sce*I produces complementary ended DSBs with 4 bp 3' overhangs; many systems employ multiple I-*Sce*I sites to produce incompatible break ends (48).

Genome sequencing has enabled study of MMEJ at end-joining repair junctions produced by repair of endogenous DSBs, including class switch recombination (CSR) junctions and chromosomal translocations (See Tables 1-2 through 1-7).

Detailed consideration of possible MMEJ mechanisms helps interpret genetic data

Since increased MMEJ is often observed in C-NHEJ deficient backgrounds, the genetics of MMEJ have been extensively studied in efforts to understand alt-EJ. To facilitate interpretation of the data regarding the genetics of MMEJ, it is helpful to first consider the mechanistic details of MMEJ in greater detail. It is possible to envision a number of variations on the basic MMEJ model presented above, with different genetic requirements, and in which microhomology pairing could serve different functions (Figure 1-2).

Figure 1-2. Possible MMEJ mechanisms can be distinguished based on different enzymatic requirements.



a. MMEJ at break ends where end "breathing" has uncovered complementary sequence. Requires flap endo- or 5'-3' exonuclease activity and a DNA ligase.

b. MMEJ at repeats uncovered by helicase activity. Also requires flap endo- or 5'-3' exonuclease activity and a DNA ligase. May require 3'-5' exonuclease activity if the microhomologies were not immediately at the break ends.

c. MMEJ via helicase or 5'-3' exonuclease activity. Requires gap filling by a DNA polymerase in addition to the requirements in (b).

d. MMEJ at a very short microhomology. May require additional end stabilization.

An examination of the genetic data suggests that there is abundant evidence supporting the existence of several variants of MMEJ mechanism, some of which may be genetically distinct. Genes implicated in MMEJ are summarized in Tables 1-2 through 1-7, grouped by biochemical function.

Table 1-2. Nucleases involved in MMEJ

Gene Product	Effect on MMEJ
Sae2p	MMEJ defect in *sae2*Δ *S. cerevisiae* at a noncomplementary ended HO DSB (46)
Mre11p	MMEJ defect in *mre11*Δ *S. cerevisiae* (46)
Exo1p	Overexpression: partial rescue of *mre11*Δ defect; survival defect exacerbated in *mre11*Δ *exo1*Δ . (46)
Exo1	Necessary for MMEJ in *S. pombe* (49)
Rad1p/Rad10p	MMEJ impaired in *rad1*Δ and *rad10*Δ single and double mutant *S. cerevisiae* (26)
Rad1p/Rad10p	Increase in MMEJ in *trans* consequent to irradiation reduced in mutant *S. cerevisiae*; microhomology length decreased (50)
CtIP	Increase in microhomology joins observed at class switch recombination junctions in *ku70* background dependent on CtIP (51)
CtIP	Required for MMEJ in G1 in the DT40 cell line (52)
Artemis	Increased microhomology at coding joints in Artemis-deficient B cells (53)
Fen-1	Depletion impaired MMEJ in vertebrate cell extract (54)
Fen-1	Necessary for end processing in MMEJ in *S. cerevisiae* (55)

A variety of exo- and endonucleases and associated signaling factors appear to be important for promoting or repressing MMEJ in both yeast and metazoan systems (Tables 1-2 and 1-3). MMEJ-associated nuclease activity appears to be promoted by the orthologous signaling kinases Tel1p and ATM in budding yeast and vertebrates, respectively, especially in p53 deficient backgrounds. MMEJ nucleases include Sae2p, Mre11p, Exo1, Rad1/10p, and Rad27 in budding yeast, and Artemis, CtIP, and Ercc1-XPF in vertebrates.

Interestingly, despite the involvement of flap endonucleases, there is little evidence supporting a requirement for helicase activity in MMEJ, suggesting that the MMEJ mechanism shown in Figure 1-2 b may not be especially common. In budding yeast, the Srs2 helicase promotes MMEJ, but its role is believed to be primarily antirecombinogenic.

Table 1-3. Signaling molecules implicated in MMEJ

Gene Product	Effect on MMEJ
Tel1p	reduced in *tel1*Δ
p53	expressing p53 in p53-deficient cells decreased MMEJ (56)
ATR	MMEJ at long microhomologies during class switch recombination may depend on ATR (57)
ATM	Regulates end resection (58)
ATM	Promotes by increasing end resection during class switch recombination (59)
DNA-PK	May participate in microhomology search via activation by compatible termini (60)
DNA-PKcs	Wortmannin disrupted MMEJ at an I-*Sce*I DSB (61)
Parp-1	May be involved in MMEJ-independent alt-EJ of I-*Sce*I DSBs (30)
Parp-1	Promotes MMEJ during class switch recombination (62)
Parp-1	Promotes a repair pathway that is most efficient at overhangs with G-C rich termini (63)
BRCA1	50-100 fold MMEJ defect at I-*Sce*I DSB in Brca1(-/-)p53(-/-) MEFs (64)
BRCA1	Suppresses MMEJ by restricting deletion at DSBs (65)
CHK2	Dominant negative kinase dead CHK2 promoted MMEJ (65).
Ku	MMEJ inhibited in *ku* mutants (66)
Ku	Inhibits MMEJ (54)
Ku	MMEJ at non-exposed 8bp microhomlogy not dependent on Ku in bladder carcinomas (67)
FLT3	Increased MMEJ, increased Lig3, decreased Ku in FLT3-duplication knock-in mice (68)

Note that contrasting different studies indicates that the genetic requirements for MMEJ do not support a simple equation of MMEJ with alt-EJ: impaired MMEJ has been observed in C-NHEJ deficient genetic backgrounds (Ku, DNA-PKcs). Additionally, although some studies indicate that PARP-1 promotes a microhomology-mediated alt-EJ pathway, it has also been observed that PARP-1 mediated break end synapsis does not depend on DNA sequence. FLT3: FMS-like tyrosine kinase-3 receptor.

Table 1-4. Helicases implicated in MMEJ.

Gene Product	Effect on MMEJ
Srs2p	impaired in *srs2*Δ*;* defect rescued in *srs2*Δ *rad51*Δ *S. cerevisiae*; (46)
Blm	"modest shift" towards microhomology junctions in Blm-deficient B cells (69)

Few helicases have been implicated in MMEJ, suggesting that the MMEJ mechanism pictured in Figure 1-2b may not accurately reflect the mechanism of most MMEJ.

Table 1-5. Mismatch repair factors implicated in MMEJ

Gene Product	Effect on MMEJ
Mlh1	Promotes MMEJ, extent of deletions smaller at *mlh1* mutant cells (70)
Msh2	Affected "pattern" of microhomologies in *S. pombe* (49)
Pms1	Affected "pattern" of microhomologies in *S. pombe* (49)

Differences in the patterns of observed microhomologies in *S. pombe* hint at the possibility of multiple genetically distinct end-joining mechanisms in that organism.

For effective MMEJ, some way of confirming that a microhomology has been located must be necessary. Some evidence has been collected suggesting that that this could be effected via components of the mismatch repair system (Table 1-5).

Participation of DNA polymerases in MMEJ has also been firmly established in both budding yeast and vertebrates (Table 1-6). Several distinct roles can be envisioned for DNA polymerases in MMEJ. If single-stranded gaps remain after microhomology pairing, a DNA polymerase will be required to fill these gaps. Pol mu, which has some terminal transferase activity, may also be able to create some terminal microhomology via untemplated addition of nucleotides to noncomplementary ends. DNA polymerase binding to a complementary 3' end and 5' single stranded region on opposite sides of a DSB can contribute to end synapsis (71). Finally, extension from base-paired microhomologies by a DNA polymerase could serve as a signal that cross-break pairing had been successfully completed. Interestingly, these last two possibilities suggest that the function of microhomology in alt-EJ may not be limited to simply holding the ends together purely via base pairing interactions.

In yeast, Pol4p, Pol32p, Rad30p, and Rev3p all appear to participate in MMEJ (46). DNA polymerases mu and lambda participate in gap filling in vertebrates, and the terminal transferase activity of pol mu may facilitate creation of overhangs in vertebrate MMEJ. In

mycobacterial end joining, the LigD polymerase domain has been shown to facilitate microhomology pairing (72).

Orthologs of the end-joining associated polymerases mu and lambda are not known to be present in the Drosophila genome (73). Nevertheless, inserted nucleotides have frequently been observed at end-joining repair junctions in Drosophila. The study detailed in Chapter 2 describes the identification of the gene *mus308*, which encodes a Drosophila ortholog of DNA polymerase theta, as a key player in an end-joining mechanism responsible for these insertions. Chapter 3 proposes a mechanism, which we call synthesis-dependent microhomology-mediated end joining (SD-MMEJ), to explain the role of this gene product in alternative end joining.

Table 1-6. DNA polymerases implicated in MMEJ

Gene Product	Effect on MMEJ
Pol4p	Combined NHEJ/MMEJ defect in *pol4*Δ *S. cerevisiae;* junction seqs and sep-fn mutants suggest MMEJ impairment specifically due to gap filling defect (46)
Pol4p	necessary for MMEJ at "discontinuous" microhomologies in *S. pombe* (49)
Pol32p	MMEJ defect in *pol32*Δ *S. cerevisiae* (46)
Rad30p	MMEJ defect in *rad30*Δ *S. cerevisiae* (46)
Rev3p	MMEJ defect in *rev3*Δ *S. cerevisiae* (46)
Polμ	Promotes microhomology pairing *in vitro* (74)
Pol μ	Can add terminal nucleotides in a non template dependent manner (6)
Pol λ	Necessary for MMEJ at partially complementary I-*Sce*I overhangs (75)
LigD polymerase domain	Promotes synapsis of mismatched ends via microhomology base pairing, base flipping, and hairpin formation in *M. tuberculosis* (72)
Pol θ	Promotes MMEJ at long microhomologies at *P* element DSBs and synthesis-dependent MMEJ at I-*Sce*I DSBs (76, 77)

The studies presented in Chapters 2 and 3 describe our finding that DNA polymerase theta participates in MMEJ-mediated alt-EJ in Drosophila.

Table 1-7. DNA ligases implicated in MMEJ

Gene Product	effect on MMEJ
LigI	siRNA against or hypomorphic mutation reduced MMEJ in a plasmid-based end-joining assay using a vertebrate-derived system (78)
Lig3	siRNA against or hypomorphic mutation reduced MMEJ (78)
Lig3/XRCC1	Required for MMEJ via PARP-1 pathway (63)
Dnl4p	MMEJ in *S. cerevisiae* partially dependent on Dnl4p (26)

The differential requirement for Lig4 between yeasts and vertebrates suggests that there may be multiple genetically distinct types of MMEJ.

MMEJ at microhomologies of different lengths may be genetically distinct

Some genetic studies suggest that there may be more than one type of MMEJ, or that genetic requirements for MMEJ may be context dependent. For example, BRCA1 has been found to both promote and suppress MMEJ (64, 65). MMEJ at long versus short microhomologies may have distinct genetic requirements. The observed lengths of junctional microhomologies are generally greater in yeasts than in metazoans, and the genetic requirements for MMEJ appear to differ between yeasts and metazoans. *S. cerevisiae* MMEJ differs from metazoan MMEJ by partial dependence on Dnl4p, the budding yeast ortholog of vertebrate Lig4 (26), and *S. pombe* MMEJ appears to be genetically related to the homologous recombination related single-strand annealing pathway, which proceeds by a mechanism thought to be similar to MMEJ, but involves considerably longer repeats (49). There is also evidence that within vertebrate systems, MMEJ involving microhomologies of different lengths may have different genetic requirements. In one study, MMEJ at a 10-bp microhomology was found to involve both DNA Ligase 1 and DNA Ligase 3, whereas MMEJ at 2-3 bp microhomologies required only DNA Ligase 3 (78). MMEJ specifically at longer microhomologies during class switch recombination may depend on ATR (57).

MMEJ at long versus short microhomologies may also be genetically distinct in Drosophila. For example, a hypomorphic mutation in *nbs* shifted MMEJ towards preferential use of longer microhomologies (79). The genetic distinction between "short-patch" and "long-patch" MMEJ is further supported by the results of our study in Chapter 2, in which we found that MMEJ at microhomologies greater than 4 bp in length is impaired in flies with mutations in *mus308,* which encodes an ortholog of DNA polymerase theta.

The Relationship Between alt-EJ And MMEJ Is Not Straightforward

A crucial unresolved question, which directly motivated the study described in Chapter 3, is whether or not all alt-EJ requires microhomology pairing across the DSB. The idea that microhomology pairing is required to hold break ends together in the absence of the C-NHEJ machinery is appealing, and many studies have documented an increase in the length and/or frequency of junctional microhomologies observed at end-joining repair junctions from C-NHEJ deficient cells. In fact, this correlation is sufficiently reliable that the terms "alt-EJ" and "MMEJ" are sometimes used interchangeably, and observations of increased junctional microhomology may be interpreted as evidence of alt-EJ repair (80, 81).

However, closer examination of the literature indicates that the relationship between alt-EJ and MMEJ is far from straightforward, for several reasons. First, the genetic requirements for MMEJ overlap those for both C-NHEJ and alt-EJ, and the length and frequency of junctional microhomologies is influenced by factors other than genetic background. Second, not all end-joining repair junctions recovered from C-NHEJ deficient cells exhibit junctional microhomology, indicating that base-pairing at pre-existing microhomologies is not an absolute requirement for all types of alt-EJ. Finally, the idea that the primary function of microhomology pairing in alt-EJ is to serve as "glue" to keep the break ends stuck together does not account for

observations of statistically significant increases in junctional microhomologies too short to effectively mediate stable end synapsis.

Genetic requirements for MMEJ overlap those for both C-NHEJ and alt-EJ

The genetic requirements for MMEJ appear to overlap those of both alt-EJ and C-NHEJ, suggesting that a simple equation of alt-EJ with MMEJ is premature. Supporting the link between alt-EJ and MMEJ is the observation that in vertebrate-derived systems, MMEJ is impaired by siRNA against DNA ligase 3, which is thought to be the primary end-joining ligase in Lig4-deficient genetic backgrounds (Table 1-7) (78), and enhanced by increased Lig3 expression (68). End joining in African trypanosomes, which lack clear orthologs of key C-NHEJ components, appears to proceed entirely by MMEJ (82). Increased MMEJ is often observed in Ku-deficient vertebrate cells (51). In tumors, decreased Ku-binding activity and increased MMEJ is associated with more aggressive, invasive phenotype (83). However, Ku deficiency (66) and disruption of DNA-PKcs by the drug wortmannin (61) have also been observed to impair MMEJ. In fact, evidence that DNA-PK is activated by compatible ends implicates DNA-PK in MMEJ (60). Another study documents decreased mutation rates in Ku-deficient mice, which is inconsistent with increased levels of MMEJ (84). Although Lig4 deficiency is often associated with increased junctional microhomology in vertebrates (85), this association is not always observed (41, 86). Further complicating the picture, although MMEJ tends to increase in Lig4-deficient vertebrate systems, budding yeast MMEJ appears to be partially dependent on Dnl4p.

Non-genetic factors affect frequency and length of junctional microhomology

Further complicating the question of whether alt-EJ and MMEJ can be considered equivalent is that factors other than genetic background can affect the occurrence of microhomology junctions. DNA sequence context and and break structure appear to influence MMEJ. Some studies suggest that the requirement for microhomology in Ku-deficient cells varies according to characteristics of the specific ends being joined (87). The presence of long microhomologies near a break has been observed to exert a "dominant" effect on repair, regardless of genetic background (88, 89). For both HO DSBs in budding yeast (26) and I-*Sce*I breaks in vertebrate cell culture (61), repair of incompatible ends appears to rely more heavily on MMEJ. Experiments both *in vivo* and *in vitro* suggest that increasing the concentration of DSBs present can shift the predominant mode of EJR substantially towards MMEJ (54, 90, 91). Use of MMEJ may be developmentally regulated in a tissue specific manner (92) and decreased capacity for MMEJ has been observed in senescent cells (93). Environmental factors also may play a role in the formation of microhomology-mediated oncogenic translocations (94).

Evidence for non-MMEJ alt-EJ

There is also evidence supporting the existence of non-MMEJ alt-EJ pathways. Not all end-joining repair junctions in C-NHEJ deficient cells have junctional microhomology (22), indicating that base pairing at pre-existing microhomologies is not an absolute requirement for all alt-EJ. In vertebrate cells, apparent blunt joins (end-joining junctions with no ambiguous nucleotides) could be consistent either with non-MMEJ alt-EJ, or with MMEJ at short microhomologous ends produced by pol mu terminal transferase activity. However, as will be shown in Chapter 3, non-microhomology junctions are observed in both *ku70* mutant and *lig4* mutant Drosophila, which are not known to possess any DNA polymerase with terminal

transferase activity (73). This suggests that there may indeed exist non-MMEJ alt-EJ mechanisms, at least in Drosophila. Finally, there is evidence for protein-mediated end tethering in alt-EJ. End synapsis in a form of PARP-1 mediated alt-EJ has been observed to not depend on microhomologies (30), and the MRN complex has end-bridging activity (31).

Patterns in observed microhomologies are not consistent with the "alt-EJ end glue" model

If the exclusive purpose of microhomology pairing in alt-EJ is to physically hold break ends together via base pairing, and if alt-EJ requires base pairing to mediate end synapsis, we would expect that alt-EJ at longer repeats would always be more efficient than alt-EJ at shorter repeats, and there should be a threshold lower bound for microhomology length consistent with a requirement for a certain amount of hydrogen bonding between break ends. This, however, is not observed. Especially in vertebrate studies, highly significant increases in frequency of microhomologies apparently too short to mediate stable end synapsis are often the observed phenotype in C-NHEJ deficient backgrounds. Furthermore, our study in Chapter 3, as well as the meta-analysis in Chapter 4, is congruent with previous studies that have observed the existence of microhomology "hot spots" that appear far more frequently at repair junctions than other nearby microhomologies of comparable length (26). We found, in fact, that microhomology length is a poor predictor of the frequency at which it will appear at repair junctions.

A Modeling Approach To Sequence Analysis Uncovers Hidden Assumptions About The MMEJ Model And Provides Alternative Interpretations

Together, these results indicate that although microhomology pairing seems to play a key role in end-joining repair, especially in C-NHEJ deficient backgrounds, the simple concept of MMEJ as glue for holding break ends are held together during alt-EJ does not furnish an adequate explanation for the observed patterns of microhomologies. This suggests that the generally accepted ideas about what mechanistic inferences can be drawn from junction sequence data ought to be re-examined. The results of the following thought experiments suggest that integrating the available genetic data with a computational modeling approach could bring to light unconsious assumptions about the mechanism of MMEJ, resulting in substantial progress towards resolving some of the current "MMEJ mysteries."

Statistical arguments assume specific mechanisms

Because it is not possible to conclude by examination of any given microhomology junction whether or not repair actually proceeded by MMEJ, interpretation of microhomology data generally involves a statistical argument. Interestingly, examination of some of these arguments reveals that assumptions about the mechanism of MMEJ are encoded in the method chosen to calculate the probability of fortuitous junctional microhomologies. A frequently used formula for calculating the expected probability P by random chance of observing a microhomology of length X is given in Table 1-8 (44):

Table 1-8. A formula commonly used for calculating the probability of observing junctional microhomology by random chance.

$P=(X+1)(1/4)^X(3/4)^2$	$(X+1)$: the number of ways that random identities could yield the specified microhomology
	$(1/4)^X$: the probability that X nucleotides match
	$(3/4)^2$: the probability that the flanking nucleotides do not match
note: can be adjusted for biased nucleotide compositions	

Formula is taken from (44). Examination of the logic behind derivation of this formula reveals that its accuracy depends on non-MMEJ events involving solely blunt ended deletions and complete pairing at all repeated nucleotides in MMEJ events.

This equation was derived based on a model in which all deletions result in blunt ends. This assumption is reflected in the $(X+1)$ term, which represents the number of ways that the X remaining nucleotides on one side could match the last X deleted on the other. This term assumes that there is only one way that that could happen on either side of the DSB, and if only blunt ends are present, this is indeed the case. However, MMEJ is usually conceptualized as a process involving production of long single stranded tails, not blunt ends. As will be shown in detail in the next section, a modeling approach reveals that if repair is not proceeding via a blunt-ended intermediate, the assumption that there is only one way that a certain number of fortuitous microhomology bases could appear on either side of the break is not valid in some sequence contexts. Thus, reliance on this equation to determine whether microhomology junctions are being observed more frequently than would be predicted by random chance without consideration of the mechanistic assumptions it encodes could easily lead to apparently contradictory results in different sequence contexts.

A modeling approach reveals alternative interpretations of shifts in average microhomology lengths

The conceptual difficulties inherent in attempting to calculate a closed-form solution for the probability of observing random junctional microhomologies are sometimes addressed by considering only changes in typical length or abundance of microhomology junctions in different genetic backgrounds. The latter metric, however, can be used as an example to show that without definite knowledge of the specific end-processing events comprising MMEJ repair, even relative shifts in microhomology phenotypes should be interpreted with caution: in some sequence contexts, factors other than annealing at longer microhomologies can account for an increase in the number of ambigous bases observed at a repair junction.

Consider the following blunt-ended DSB:

```
5'-NNC-3A-2A-1A-0          +0A+1A+2A+3GNN-3'
   NNG-3T-2T-1T-0          +0T+1T+2T+3CNN
```

5'-3' resection to the end of the 3 bp AAA repeat would produce the following ss3' ends:

```
5'-NNC-3A-2A-1A-0                    GNN-3'
   NNG                      +0T+1T+2T+3CNN
```

If no 3'-5' exo or flap endonuclease activity participates in repair, then the number of ambiguous nucleotides observed at the repair junction will accurately reflect the number of nucleotides that annealed during the MMEJ event (Table 1-9):

Table 1-9: MMEJ at repeated motifs consisting of a mononucleotide run.

bp del from L/R ss3' ends	bp annealed	repair event	repair product aligned with original sequence	bp mh
0/0	3	5'–NNC$_{-3}$**A**$_{-2}$**A**$_{-1}$**A**$_{-0}$GNN–3' NNG$_{+0}$**T**$_{+1}$**T**$_{+2}$**T**$_{+3}$CNN	NNCAAAAAAGNN NNC***AAA***---GNN	3
0/0	2	5'–NNC$_{-3}$A$_{-2}$**A**$_{-1}$**A**$_{-0}$A$_{-0}$GNN–3' NNC$_{-3}$T$_{+0}$**T**$_{+1}$**T**$_{+2}$T$_{+3}$GNN	NNCAAAAAAGNN NNCA***AA***--AGNN	2
0/0	1	5'–NNC$_{-3}$A$_{-2}$A$_{-1}$**A**$_{-0}$A$_{-}$ A$_{-0}$GNN–3' NNC$_{-3}$T$_{-3}$T$_{+0}$**T**$_{+1}$T$_{+2}$T$_{+3}$GNN	NNCAAAAAAGNN NNCAA***A***-AAGNN	1

In the absence of 3'-5' exo or flap nuclease activity, the number of ambiguous bases observed at the repair junction (bold italic underlined) corresponds to the actual number of base pairs annealed during MMEJ.

However, if a different number of nucleotides are removed from the left and right hand single stranded tails, then the number of ambiguous bases observed in the repair product can be greater than the number of bases that actually annealed during the MMEJ event (Table 1-10):

Table 1-10: MMEJ with asymmetric nuclease activity at repeated motifs consisting of a mononucleotide run

bp del from L/R ss3' ends	bp annealed	repair event	repair product aligned with original sequence	bp mh
1/0	1	5'–NNC$_{-3}$A$_{-2}$**A**$_{-1}$A$_{-}$ A$_{--}$GNN–3' NNG$_{-3}$T$_{+0}$**T**$_{+1}$T$_{+2}$T$_{+3}$CNN	NNCAAA AAAGNN NNCA***AA***---AGNN	2
2/0	1	5'–NNC$_{-3}$**A**$_{--}$A$_{--}$A$_{-0}$GNN–3' NNC$_{+0}$**T**$_{+1}$T$_{+2}$T$_{+3}$GNN	NNCAAAAAAGNN NNC***AAA***---GNN	3

In this case, asymmetrical removal of base pairs on the two sides of the DSB results in junctional microhomologies (bold italic underlined) that do not accurately reflect the mechanism of the repair event.

This modeling experiment reveals a possible alternative explanation for increase in the length of junctional microhomologies in this sequence context: increased 5'-3' nuclease activity

at the DSB. Thus, observations of an increase in junctional microhomology length do not necessarily reflect a need for especially stable end synapsis via base pairing.

In sum, arguments that appeal to patterns in sequence structure are essentially statistical arguments, but these arguments are weakened considerably without explicit consideration of the mechanism that is hypothesized to have produced these patterns. The overall message conveyed by these examples is that there is much to be gained by a more sophisticated mechanism-centered approach to sequence analysis of end-joining repair junctions. Instead of simply tabulating static patterns such as microhomologies, modeling the likely repair intermediates and determining what model parameters yield patterns that match the data is an effective way to develop novel hypotheses about the origins of junctional microhomologies. The SD-MMEJ model presented in Chapter 3 is the product of such as exercise.

Outline Of Subsequent Chapters

Our overall goal in carrying out the studies in the following chapters was to progress towards understanding the mechanism of alt-EJ. We had previously demonstrated that in Drosophila, end-joining repair within the 3' single stranded ends resulting from *P* element excision proceeds by MMEJ in both wild-type and *lig4* mutant genetic backgrounds (27). Interestingly, the resulting end-joining repair junctions often contain inserted nucleotides, even though Drosophila are not known to possess orthologs of the DNA polymerases implicated in end-joining repair in other organisms. The study in Chapter 2 describes how our efforts to identify the DNA polymerase responsible for these insertions led to identification of the *mus308* gene product, the Drosophila ortholog of DNA polymerase theta, as a novel mediator of alt-EJ. Chapter 3 focuses on experiments designed to investigate the requirement for microhomology pairing in Drosophila alt-EJ. The surprising result that the length and frequency of junctional microhomologies did not increase in *lig4* mutant flies, but insertions at repair junctions became more frequent, together with the results from the study in Chapter 2, inspired development of new strategies for sequence analysis that ultimately produced a model for synthesis-dependent microhomology-mediated end joining. The SD-MMEJ model demonstrates that one basic mechanism can account for the formation of apparent blunt joins, microhomology junctions, and insertion junctions, which have previously been thought to represent fundamentally different repair mechanisms. Finally, in Chapter 4, we demonstrate the generality of the SD-MMEJ model by showing that it explains patterns in end-joining repair observed at other DSBs.

Chapter 2

Dual Roles For DNA Polymerase Theta In Alternative End-Joining Repair Of Double-Strand Breaks In Drosophila

Sze Ham Chan[1,^], Amy Marie Yu[1,^], and Mitch McVey[*,1,2]

[1]Department of Biology, Tufts University, Medford, MA 02155
[2]Program in Genetics, Tufts Sackler School of Graduate Biomedical Sciences, Boston, MA 02111

[^] indicates these authors contributed equally to this work
[*]corresponding author: mitch.mcvey@tufts.edu

This chapter has been previously published as

PLoS Genet 6(7): e1001005. doi:10.1371/journal.pgen.1001005

Supplemental Figures and Tables are available at PLoS Genetics online.

Contribution: Graduate student mentor for S.H. Chan (undergraduate). Provided data for Figure 6; wrote text.

Introduction

DNA double-strand breaks (DSBs) and interstrand crosslinks pose serious threats to cell survival and genome stability. Because these lesions compromise both strands of the double helix, they impede DNA replication and transcription and therefore must be removed in a timely and coordinated manner. Interstrand crosslink repair has been shown to involve a DSB intermediate in some cases (reviewed in [1]). Therefore, there may be substantial mechanistic overlap in the processes used during repair of these two lesions.

Error-free repair of DSBs can be accomplished through homologous recombination (HR) with an undamaged homologous template (reviewed in [2]). However, in contexts where suitable templates for HR do not exist, error-prone repair mechanisms are also used. For example, non-homologous end joining (NHEJ) frequently creates small insertions and deletions during DSB repair, particularly in cases where the broken ends cannot be readily ligated (reviewed in [3]). Analogously, the use of translesion DNA polymerases during interstrand crosslink repair can result in mutations, due to the reduced fidelity of these polymerases [4,5].

Accumulating evidence suggests that NHEJ is composed of at least two genetically distinct mechanisms. Classical NHEJ (C-NHEJ) involves the sequential recruitment of two highly conserved core complexes (reviewed in [6]). First, the Ku70/80 heterodimer recognizes and binds to DNA ends in a sequence-independent manner, thereby protecting them from degradation. In many eukaryotes, Ku70/80 also recruits DNA-PK_{cs}, forming a synaptic complex that can recruit additional processing enzymes such as the Artemis nuclease and the X family DNA polymerases mu and lambda. These proteins expand the spectrum of broken ends that can be rejoined. The second core complex, composed of DNA ligase 4, XRCC4, and XLF/Cernunnos, catalyzes ligation of the processed ends. Depending on the substrate, C-NHEJ

can result in perfect repair of broken DNA, or it can result in small deletions of 1-10 nucleotides and/or insertions of 1-3 nucleotides [7]. Although C-NHEJ can repair blunt-ended substrates, a subset of C-NHEJ products appear to involve annealing at 1-4 nucleotide microhomologous sequences on either side of the break.

Alternative end-joining (alt-EJ) is defined as end-joining repair that is observed in cells or organisms lacking one or more C-NHEJ components (reviewed in [8]). Alt-EJ in yeast is associated with deletions larger than those typically created by C-NHEJ, together with an increased tendency to repair by annealing at microhomologous sequences. Ku and ligase 4-independent end joining observed in mammalian cells also displays an increased tendency towards use of short microhomologies compared to C-NHEJ [9,10]. Therefore, alt-EJ is sometimes called microhomology-mediated end joining (MMEJ) [11]. However, the relationship between MMEJ and alt-EJ is still unclear, and alt-EJ may comprise one or more C-NHEJ-independent repair mechanisms [8].

The importance of alt-EJ repair is highlighted by multiple studies that suggest it may promote chromosome instability and carcinogenesis. Alt-EJ produces chromosome translocations in mouse embryonic stem cells lacking Ku70 [12] and the use of alt-EJ during V(D)J recombination in C-NHEJ-deficient murine lymphocytes causes complex chromosome translocations and progenitor B cell lymphomas [13]. Furthermore, alt-EJ has been implicated in various translocations associated with chronic myeloid leukemia and human bladder cancer [14,15]. Importantly, alt-EJ also operates during V(D)J rejoining in C-NHEJ-proficient B lymphocytes [16], suggesting that its role in DSB repair is not limited to situations where C-NHEJ is defective. However, alt-EJ is frequently masked by more dominant repair processes

that are essential for vertebrate development, making it difficult to study. Therefore, its molecular mechanisms and the proteins involved remain largely unknown.

Several lines of evidence demonstrate that Drosophila is an excellent model system in which to study alt-EJ in a metazoan. The Drosophila genome lacks several mammalian C-NHEJ components, including DNA-PK_{cs} and Artemis. This may predispose flies towards non-C-NHEJ repair. Consistent with this, we have previously shown that a DSB caused by excision of a *P* element transposon in flies is readily repaired by a DNA ligase 4-independent end-joining process [17]. Interestingly, although Drosophila orthologs for the Pol X family DNA polymerases mu and lambda have not been identified [18], we and others have found evidence for polymerase activity in Drosophila end-joining repair [17,19,20]. Specifically, end joining in flies is often associated with the insertion of nucleotides at repair junctions, frequently involving imperfect repeats of 5-8 nucleotides. Full or partial templates for the insertions, occasionally possessing mismatches, can often be identified in adjacent sequences, suggesting the action of an error-prone polymerase. Similar templated nucleotides (T-nucleotides) have previously been identified at translocation breakpoints in human lymphomas [21-23]. Therefore, T-nucleotides could represent a signature of alt-EJ and may be informative regarding its molecular mechanisms.

Additional insight into alt-EJ is provided by recent reports suggesting a mechanistic link between alt-EJ and interstrand crosslink repair. For example, a study of two Chinese hamster ovary cell lines sensitive to the crosslinking agent mitomycin C found that they were also deficient in alt-EJ [24]. Furthermore, certain interstrand crosslink-sensitive cell lines from Fanconi Anemia patients are also impaired in DNA-PK_{cs}-independent rejoining of linearized plasmids [25]. Based on these reports, we hypothesized that additional mechanistic insight into

both interstrand crosslink repair and alt-EJ could be gained by searching for mutants defective in both processes. To test this, we have screened Drosophila mutants that are sensitive to DNA crosslinking agents for additional defects in alt-EJ repair. In this work, we describe our studies with one such mutant, *mus308*.

The *mus308* (mutagen sensitive 308) mutant was originally identified by its extreme sensitivity to interstrand crosslinking agents but normal resistance to alkylating agents [26]. Subsequently, *mus308* was found to code for DNA polymerase theta, which is most similar to A family DNA polymerases such as *Escherichia coli* Pol I [27]. Orthologs of polymerase theta (hereafter referred to as pol θ) are found in many metazoans, including *Caenorhabditis elegans*, *Arabidopsis thaliana,* and *Homo sapiens*, but not in unicellular eukaryotes, including the yeasts [28-30]. Several lines of evidence suggest that pol θ may play an important role in maintaining genome stability. Similar to flies, *C. elegans* with mutations in *POLQ-1* are defective in repair of interstrand crosslinks [28]. Mice lacking pol θ (*chaos1* mutants) have a high frequency of spontaneous and mitomycin C-induced micronuclei in erythrocytes, consistent with genomic instability [31]. In addition, vertebrate pol θ orthologs have been implicated in a wide range of repair processes, including base excision repair, bypass of abasic sites, and somatic hypermutation of immunoglobulin genes [32-36]. Finally, upregulation of pol θ is observed in a variety of human tumors and is associated with a poor clinical outcome, suggesting that its overexpression may contribute to cancer progression [37].

Pol θ is unusual in possessing an N-terminal helicase-like domain and a C-terminal polymerase domain. Although pol θ purified from human cell lines and Drosophila has error-prone polymerase activity and single-stranded DNA-dependent ATPase activity, helicase activity

has not been demonstrated *in vitro* [30,38,39]. Therefore, it remains unclear exactly how the structure of pol θ relates to its multiple functions in DNA repair in different organisms.

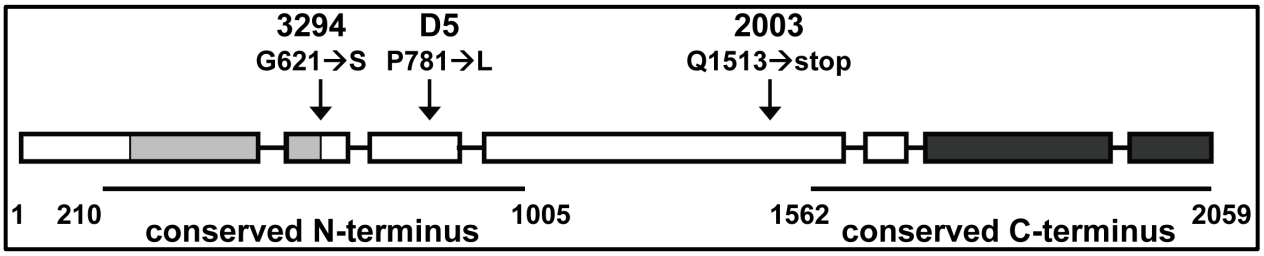
We report here that in addition to its role in DNA interstrand crosslink repair, Drosophila pol θ is involved in end-joining repair of DSBs. This alt-EJ mechanism operates independently of both Rad51-mediated HR and ligase 4-dependent C-NHEJ. Genetic analysis using separation-of-function alleles provides support for distinct roles of both the N- and C-terminal domains of pol θ in alt-EJ. Collectively, our data support a model in which helicase and polymerase activities of Drosophila pol θ cooperate to generate single-stranded microhomologous sequences that are utilized during end alignment in alt-EJ.

Results

Drosophila pol θ is important for both interstrand crosslink repair and end-joining repair of DNA double strand breaks

Drosophila *mus308* mutants were initially identified based on their sensitivity to low doses of chemicals that induce DNA interstrand crosslinks [26]. To confirm this phenotype, we assembled a collection of previously identified *mus308* mutant alleles [40,41] and measured the ability of hemizygous mutant larvae to survive exposure to the crosslinking agent mechlorethamine (nitrogen mustard). Of the mutants that we tested, four were unable to survive exposure to 0.005% mechlorethamine: *D2*, *D5*, *2003*, and *3294* (data not shown), consistent with their inability to repair interstrand crosslinks.

Figure 2-1. Schematic of the mus308 gene.

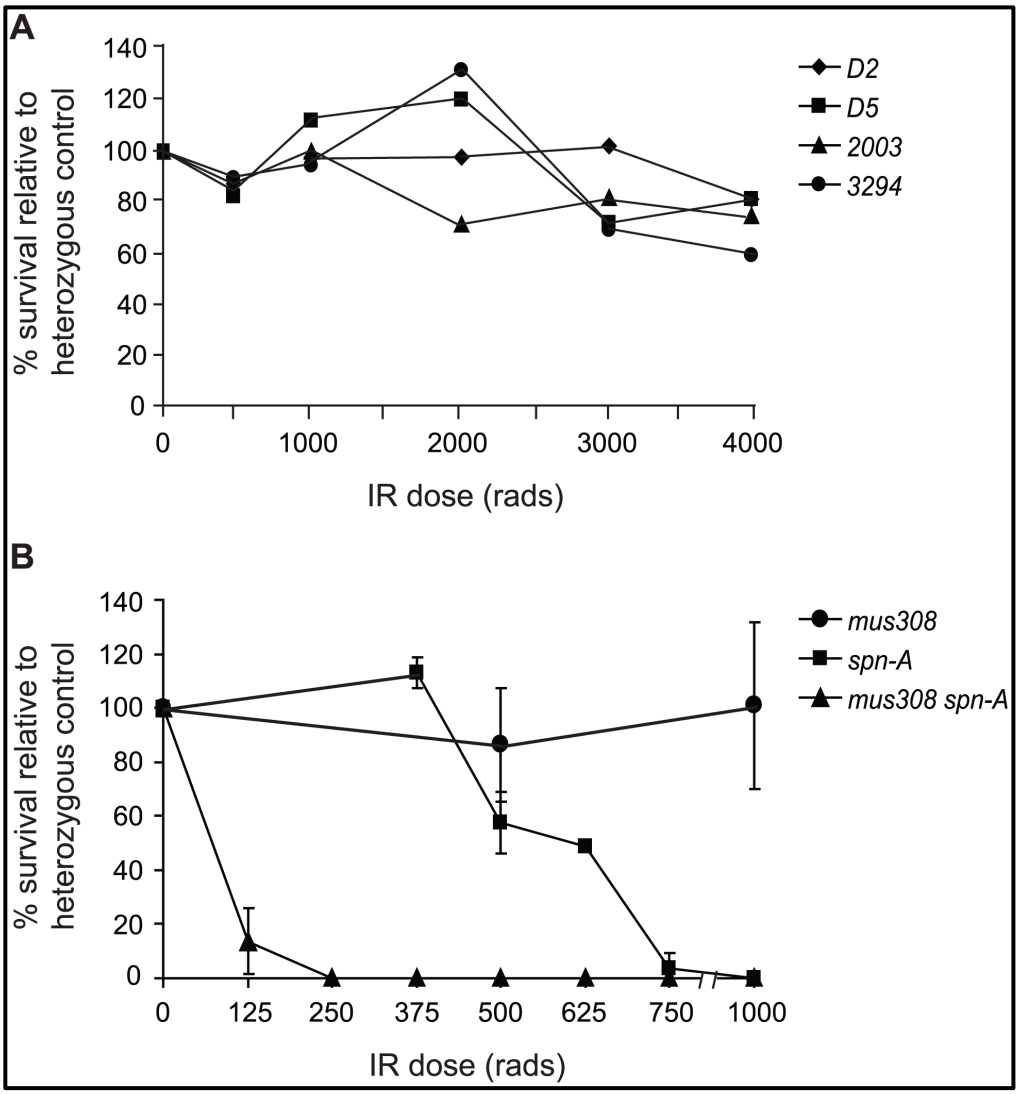


Exons are represented by boxes. The helicase-like domain (light gray shading) and polymerase domain (dark gray shading) are shown. Locations of the *D5*, *3294*, and *2003* point mutations are indicated with arrows (numbers correspond to amino acid positions in the protein). Not shown is the *D2* mutation, which results in severely reduced levels of pol θ.

To determine the molecular lesions responsible for mechlorethamine sensitivity, we sequenced the entire *mus308* coding region of flies hemizygous for each mutant allele. Pol θ possesses both a conserved N-terminal helicase-like domain and a C-terminal pol I-like polymerase domain (Figure 2-1, Figure S1) [30]. Three of the four alleles contain unique sequence changes that are predicted to affect pol θ primary structure (Figure 2-1, Figures S2 and

S3). The *2003* allele is a nonsense mutation upstream of the polymerase domain, while the *D5* allele is a missense mutation that alters a highly conserved proline in the conserved N-terminus. The *3294* allele changes an invariant glycine in the helicase domain to serine. Interestingly, this residue is conserved in the related *mus301* helicase, but not in other DNA helicases (data not shown). No mutations were found in the coding sequence of the *D2* allele. Because homozygous *D2* flies have undetectable levels of pol θ protein [38], the *D2* mutation may affect a regulatory region of *mus308*.

Figure 2-2. Drosophila pol θ is required for repair of IR–induced damage in the absence of Rad51.



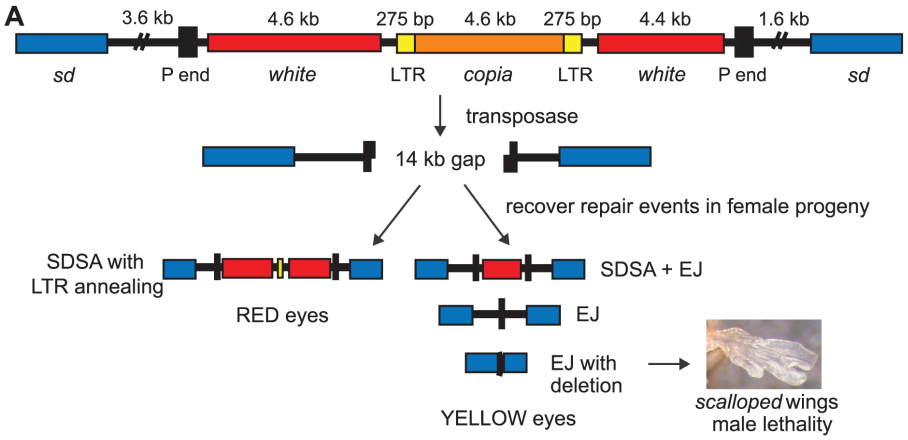
(A) *mus308* mutants are not sensitive to ionizing radiation.
Female flies heterozygous for the indicated *mus308* alleles were mated to *Df(3R)Sz29/TM3* males and third instar larval progeny were irradiated with increasing doses of IR. The percent survival of *mus308* hemizygous mutants was calculated relative to an unirradiated control. Each dose was repeated twice; the average of the two experiments is shown.

(B) *spn-A mus308* double mutants are extremely sensitive to ionizing radiation.
Heterozygous *spn-A*^{057}*mus308*^{D5} females were mated to heterozygous *spn-A*^{093}*mus308*^{D5} males and third instar larval progeny were irradiated with increasing doses of IR. The percent survival of *spn-A mus308* compound heterozygotes was calculated relative to an unirradiated control. Each dose was repeated at least twice. Error bars represent standard deviations.

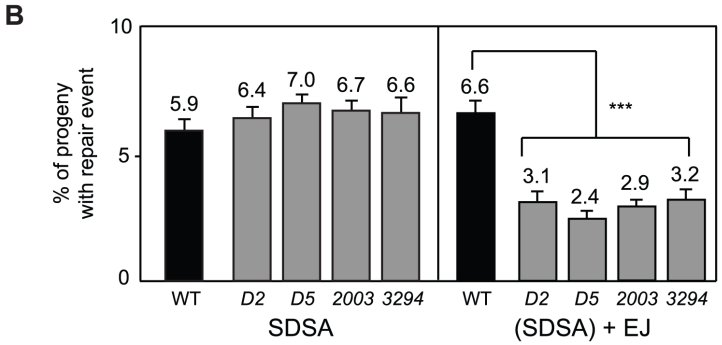
One explanation for the extreme sensitivity of *mus308* mutants to mechlorethamine could be a defect in the repair of certain types of DSB intermediates that are created during crosslink repair. To test this, we exposed flies hemizygous for each mutant allele to increasing doses of ionizing radiation (IR). Although IR creates many different types of lesions, unrepaired DSBs are thought to be the main cause of cell death following irradiation. All four *mus308* mutants survived IR exposures as high as 4000 rads (Figure 2-2A), although bristle and wing defects characteristic of apoptotic cell death were frequently observed at high doses. Drosophila *lig4* mutants, which are completely defective in C-NHEJ, also survive IR doses in excess of 4000 rads [17]. However, *spn-A* mutants, which lack the Rad51 recombinase required for strand invasion during homologous recombination initiation [42], are highly sensitive to IR [17]. Thus, in Drosophila, HR is the dominant mechanism used to repair IR-induced DSBs.

To test whether pol θ acts to repair IR damage in the absence of HR, we created *mus308 spn-A* double mutants and exposed them to doses of 125-1000 rads. Strikingly, doses as low as 125 rads resulted in almost complete killing of *mus308 spn-A* mutants (Figure 2-2B). In contrast, *lig4 spn-A* double mutants are only slightly more sensitive than *spn-A* single mutants to IR [17]. Thus, in the absence of HR, pol θ participates in a process crucial for repair of damage caused by ionizing radiation.

Figure 2-3. Pol θ is involved in end-joining repair.

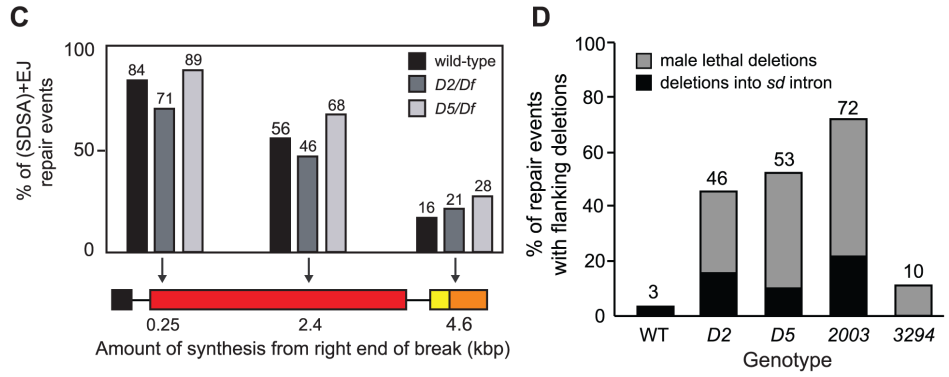


(A) A *P* element excision assay monitors DSB repair outcomes.
The *P{w^a}* transposon is inserted in a 5.2 kb intron of the essential *scalloped* (*sd*) gene. A copia retrotransposon flanked by long terminal repeats (LTR) is inserted into an exon of the *white* gene, causing reduced expression of *white*. Excision of *P{w^a}* results in a 14-kilobase gap relative to an intact sister chromatid. Following excision, repair events from the male pre-meiotic germline are recovered in female progeny and the method of repair is inferred from their eye color. SDSA = synthesis-dependent strand annealing, SDSA+EJ = SDSA+end joining, EJ = end joining. Accompanying deletion into *sd* exons (EJ with deletion) results in a scalloped-wing phenotype and/or male lethality.



(B) Pol θ mutants are defective specifically in end-joining repair of DSBs.
Each bar represents the mean percentage of repair events recovered from independent males possessing the indicated *mus308* allele in trans to *Df(3R)Sz29*. (SDSA)+EJ = repair involving HR initiation followed by end joining or repair involving only end joining. Number of independent males assayed: wild type = 101, *D2* = 155, *D5* = 161, *2003* = 155, *3294* = 100. Error bars represent SEM. (*** indicates $p<0.001$, Kruskal-Wallis non-parametric ANOVA). $p > 0.5$ for all comparisons between different *mus308* alleles.

Figure 2-3 (continued)



(C) Synthesis during HR is not impaired in pol θ mutants.
Each bar represents the percentage of repair events recovered from yellow-eyed females that showed evidence for repair synthesis of at least the indicated length. Number of independent repair events: wild type = 32, *D2* = 28, *D5* = 53.

(D) Pol θ mutants have an increased frequency of DSB repair events with flanking deletions.
For all genotypes, small deletions <3.6 kilobases upstream or <1.6 kilobases downstream of *P{w^a}* were detected by PCR. Large deletions resulted in scalloped-winged female progeny and/or male lethality.

Because interstrand crosslink repair and alternative end joining have been shown to have partially overlapping genetic requirements in mammals [24,25], we hypothesized that the extreme sensitivity of *mus308 spn-A* mutants to IR might relate to a role of pol θ in an alternative end-joining mechanism. To explore this hypothesis, we tested each *mus308* mutant allele using a site-specific double-strand break repair assay that can distinguish between synthesis-dependent strand annealing (SDSA, a specific type of HR) and end joining (EJ) (Figure 2-3A) [43]. We have previously shown that the majority of end joining observed in this assay occurs independently of DNA ligase 4, and is therefore a form of alt-EJ [17]. In this system, excision of a *P* element (*P{w^a}*) located on the *X* chromosome is catalyzed by *P* transposase, resulting in a 14-kilobase gap relative to an undamaged sister chromatid. The DNA ends remaining after excision each have 17-nucleotide non-complementary 3' single-stranded overhangs [44]. These ends are highly recombinogenic and repair by SDSA is initially favored. However, because

repair synthesis in this system is not highly processive, most repair products that are recovered from wild-type flies result from incomplete repair synthesis from one or both sides of the break, followed by end joining of the nascent DNA (SDSA+EJ events) [45]. To quantitate the percentage of repair events that derive from each mechanism, repair events are recovered from male pre-meiotic germline cells by mating individual males to females homozygous for the *P{w^a}* element. Each of the resulting female progeny represents a single repair event that can be classified by eye color. Red eyed-females inherit a repair event involving homology-dependent synthesis that generated complementary single-stranded regions that subsequently anneal (repair by SDSA). Yellow-eyed females inherit a chromosome that was repaired by EJ or SDSA+EJ mechanisms (these repair events are hereafter referred to as (SDSA)+EJ; for further details, see Materials and Methods).

Overall, the results from the *P{w^a}* assay indicated that *mus308* mutants are defective in end-joining repair of DSBs. We observed no decrease in the percentage of red-eyed progeny recovered from *mus308* mutant males (Figure 2-3B), suggesting that SDSA repair is not impaired when pol θ is missing or defective. In contrast, all four *mus308* mutant alleles resulted in a significantly decreased percentage of yellow-eyed progeny relative to wild type ($p<0.001$, Kruskal-Wallis test). Because yellow-eyed progeny can only result from a repair mechanism involving end joining, these data suggest that pol θ is involved in an end-joining process.

To further demonstrate that pol θ is not involved in DNA synthesis during SDSA, we recovered independent (SDSA)+EJ events in males, isolated genomic DNA, and used PCR to estimate the approximate amount of DNA repair synthesis that occurred prior to end joining. The amount of repair synthesis in (SDSA)+EJ repair products did not differ significantly between wild-type and *mus308* mutant flies (Figure 2-3C). We conclude that pol θ is not

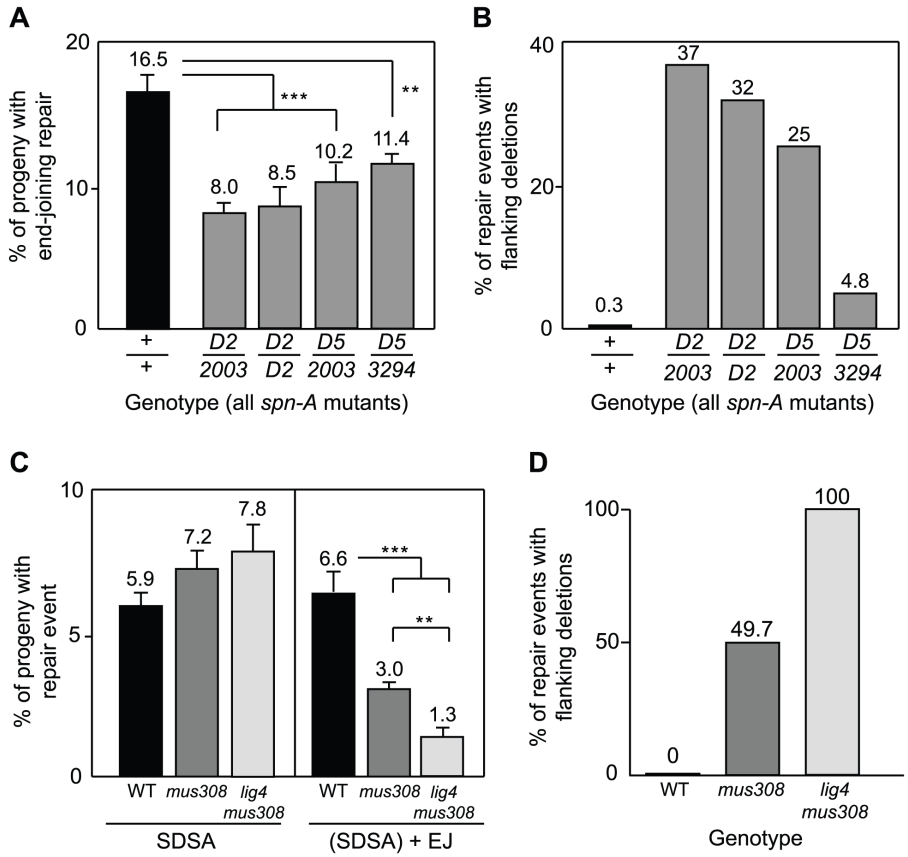
required for DNA synthesis during SDSA, but plays an important role in end-joining repair following aborted SDSA.

DSB repair in *mus308* mutants frequently results in rad51-independent genomic deletions

Mutations that abolish end joining in flies cause an increased frequency of genomic deletions during repair of site-specific DSBs [46,47]. To determine whether mutation of *mus308* also results in repair-associated deletions, we took advantage of the fact that deletions can be easily scored in the *P* element excision assay. Because *P{w^a}* is inserted in the essential *scalloped* (*sd*) gene, repair events that delete into *sd* exons cause a scalloped-wing phenotype when recovered in heterozygous females and lethality in hemizygous males [48,49]. We observed a substantial increase in the percentage of deletion-associated repair events isolated from *mus308* mutant males relative to wild type (Figure 2-3D). Overall, the total percentage of end-joining repair events involving deletions recovered from *mus308* mutants was elevated from 3- to 26-fold over wild type, depending on the *mus308* allele tested.

Previously, we observed a similar deletion-prone phenotype in flies lacking the DmBlm protein, which is involved in repair of DSBs by SDSA [48]. Because our data did not support a role for pol θ in homologous recombination, we expected the deletion-prone phenotype of *mus308* mutants to persist even in SDSA-deficient flies. To confirm this, we assayed repair following *P{w^a}* excision in *mus308* mutants lacking the Rad51 protein, which renders them unable to carry out HR repair [42,45].

Figure 2-4. Pol θ–dependent end joining acts independently of HR and C-NHEJ.



(A) Decreased end joining in *mus308* mutants does not depend on homologous recombination.
Each bar represents the mean percentage of end-joining events recovered from independent males of each genotype. Number of independent males assayed: *spn-A* = 63; *spn-A*^{057}*mus308*^{D2}/*spn-A*^{057}*mus308*^{2003} = 216; *spn-A*^{057}*mus308*^{D2}/*spn-A*^{057}*mus308*^{D2} = 33; *spn-A*^{093}*mus308*^{D5}/*spn-A*^{057}*mus308*^{2003} = 80; *spn-A*^{093}*mus308*^{D5}/*spn-A*^{093}*mus308*^{3294} = 128. Error bars represent SEM. ** indicates p<0.01, *** indicates P<0.001 (Kruskal-Wallis non-parametric ANOVA).

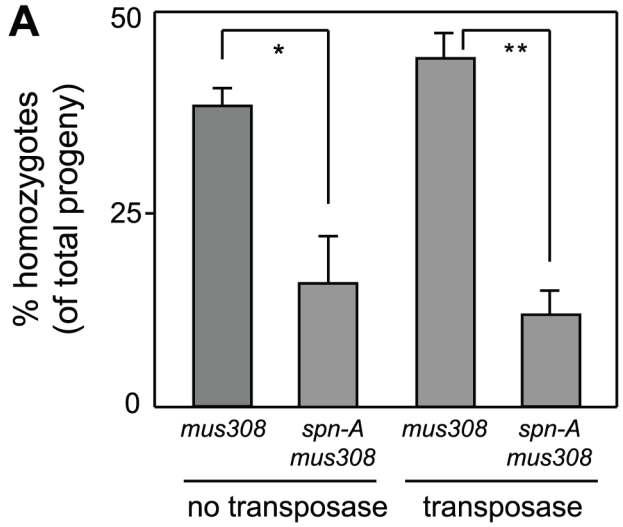
(B) Genomic deletions during DSB repair in *mus308* mutants do not depend on initiation of HR.
Each bar represents the percentage of yellow-eyed female progeny with scalloped wings (representing large flanking deletions) recovered from males of indicated genotype.

(C) Mutation of *mus308* in a *lig4*^{169a} background further reduces end-joining repair relative to *mus308* single mutants.
Each bar represents the mean percentage of repair events recovered from wild type, *mus308*^{2003}/*Df(3R)kar3l*, or *lig4*^{169a}; *mus308*^{2003}/*Df(3R)kar3l* males. Number of independent males assayed: wild type = 101, *mus308* = 182, *lig4 mus308* = 57. Error bars represent SEM. ** indicates p<0.01, *** indicates P<0.001 (Kruskal-Wallis non-parametric ANOVA).

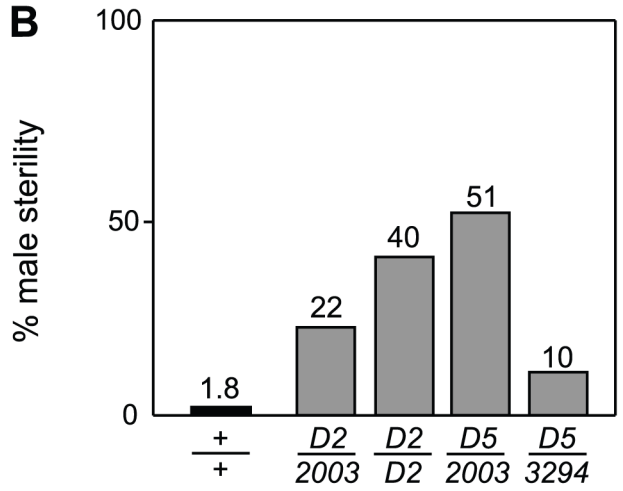
(D) Genomic deletions always occur during DSB repair in *lig4 mus308* mutants.
All yellow-eyed repair events recovered from *lig4*^{169a}; *mus308*^{2003}/*Df(3R)kar3l* males had large deletions that resulted in male lethality and/or female progeny with scalloped wings. Number of independent repair events assayed: wild type = 101, *mus308* = 167, *lig4 mus308* = 22.

As expected, PCR analysis of repair products showed that SDSA was abolished in both *spn-A* and *spn-A mus308* mutants (data not shown). Approximately 17% of *P{w^a}* chromosomes recovered from *spn-A* mutant males showed evidence for end joining at the 17-nucleotide overhangs that are created by *P* transposase (Figure 2-4A and Table 2-1); the other 83% of *P{w^a}* chromosomes recovered were presumably uncut. We observed a 30-50% decrease in end-joining repair products in *spn-A mus308* double mutants compared to *spn-A* mutants ($p<0.001$, Kruskal-Wallis test), confirming a unique role for polθ in end joining when HR is absent. Importantly, mutation of *mus308* still caused an increased percentage of deletions in the absence of Rad51 (Figure 2-4B). From these data, we conclude that the deletions formed during break repair in *mus308* mutants are not the result of aborted SDSA. Rather, they are a consequence of a deletion-prone repair mechanism that operates in the absence of both SDSA and pol θ-dependent end joining.

Figure 2-5. Flies lacking Rad51 and Pol θ have synergistic phenotypic defects



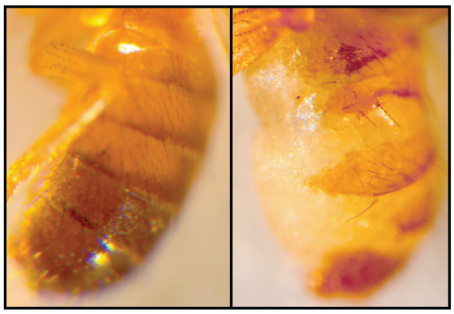
(A) *spn-A mus308* mutants are sub-viable.
Females and males with the *P{w^a}* transposon and heterozygous for the *spn-A^{057}* or *spn-A^{093}* and *mus308^{D2}* or *mus308^{D5}* alleles were interbred and the percentage of viable compound heterozygous progeny was determined. The crosses were repeated in the presence of *P* transposase. Each bar represents three independent experiments with at least 300 progeny per experiment. Error bars represent SEM. * indicates p<0.05, ** indicates P<0.01 (unpaired T test).



(B) Increased male sterility in *spn-A mus308* double mutant males.
Each bar represents the percentage of *spn-A* or *spn-A mus308* males that produced no progeny. Number of independent males assayed: *spn-A* = 57; *spn-A^{057}mus308^{D2}/spn-A^{057}mus308^{2003}* = 305; *spn-A^{057}mus308^{D2}/spn-A^{057}mus308^{D2}* = 50; *spn-A^{093}mus308^{D5}/spn-A^{057}mus308^{2003}* = 102; *spn-A^{093}mus308^{D5}/spn-A^{093}mus308^{3294}* = 100.

Figure 2-5 (continued)

C



(C) Mutation of both *spn-A* and *mus308* results in external morphological defects.
Phenotypes include abdominal cuticle mispatterning (note the severe disruption of normal segmental banding patterns). Pictured are representative wild type and *spn-A*^{057}*mus308*^{D2}/*spn-A*^{057}*mus308*^{2003} males in which *P* element excision is occurring.

During the course of these experiments, we made a number of observations suggesting that Rad51 and pol θ act in parallel and distinct DSB repair mechanisms. First, we recovered fewer *spn-A mus308* double mutant males than would be predicted from Mendelian ratios. For example, in crosses between *mus308*^{D2} and *mus308*^{D5} heterozygotes, 38% of the progeny were *mus308*^{D2}/*mus308*^{D5} compound heterozygotes. In contrast, only 16% of progeny recovered from parallel matings between *spn-A*^{057}*mus308*^{D2} and *spn-A*^{093}*mus308*^{D5} mutants were *spn-A mus308* compound heterozygotes ($P<0.05$, Fisher's exact test; Figure 2-5A). This difference in viability between *mus308* and *spn-A mus308* mutants was even more extreme in flies in which excision of *P{w*^{a}*}* was occurring ($P<0.01$; Figure 2-5A). In addition, we observed heightened male sterility in various combinations of *spn-A mus308* mutants undergoing *P{w*^{a}*}* transposition, with 51% of the double mutant males unable to produce viable progeny in the most severe allele combination (Figure 2-5B). Finally, we observed morphological abnormalities, specifically abdominal closure defects and aberrant cuticle banding patterns, in 100% of *spn-A*^{093}*mus308*^{D5}/*spn-*

$A^{057}mus308^{D2}$ double mutants (Figure 2-5C). These defects were more severe in the double mutants experiencing *P{w^a}* transposition, but were not apparent in either *mus308* or *spn-A* single mutants. From these data, we conclude that Rad51 and pol θ participate in independent pathways required for repair of DSBs that arise during both endogenous developmental processes and during *P* element transposition.

Pol θ-mediated end joining is distinct from C-NHEJ

P element ends are good substrates for DNA ligase 4-independent end joining [17]. Based on the results presented above, it seemed likely that pol θ is involved in an end-joining process different from C-NHEJ. To formally test this, we repeated the *P{w^a}* assay in *lig4 mus308* double mutants that lack DNA ligase 4 and are unable to repair DSBs by C-NHEJ. Unlike *spn-A mus308* mutants, we observed no viability, fertility, or morphological defects in *lig4 mus308* mutants. We also observed no defect in HR repair in the double mutants (Figure 2-4C), consistent with results obtained using *mus308* single mutants. In contrast, we observed a further decrease in the percentage of end joining repair products recovered from *lig4 mus308* double mutants relative to *mus308* mutants, from 3.0% to 1.3% ($P<0.01$, Kruskal-Wallis test) (Figure 2-4C). Previously, we have shown that end-joining repair of DSBs induced by *P{w^a}* excision is unaffected in *lig4* mutants [17]. Therefore, the removal of pol θ-mediated end joining reveals a previously hidden role for DNA ligase 4 in the repair of DSBs created by *P* transposase. Strikingly, although only 50% of end-joining products isolated from *mus308* mutants involved large, male-lethal deletions, 100% of end-joining products recovered from *lig4 mus308* mutant males were associated with large deletions (Figure 2-4D).

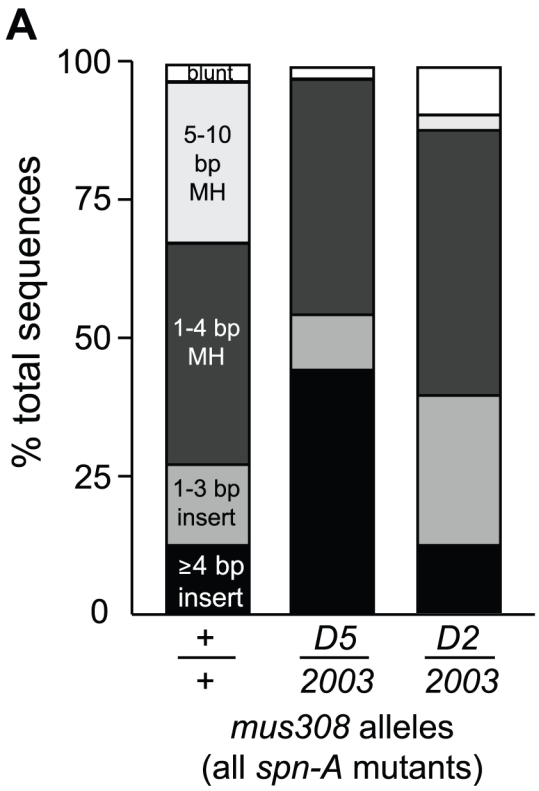
From these results, we conclude that at least three distinct mechanisms for end-joining repair exist in Drosophila. One, which corresponds to C-NHEJ, requires DNA ligase 4 and other

canonical NHEJ proteins, including XRCC4, Ku70, and Ku80 [46,47,50]. Another mechanism, which is at least partially independent of DNA ligase 4, is defined by a requirement for pol θ and corresponds to alt-EJ. Interestingly, alt-EJ appears to be used more frequently than C-NHEJ, at least for the repair of *P* element-induced breaks. In the absence of these two repair processes, a Rad51-independent backup mechanism characterized by extensive genomic deletions operates at low efficiency.

Pol θ has two distinct functions in alternative end joining

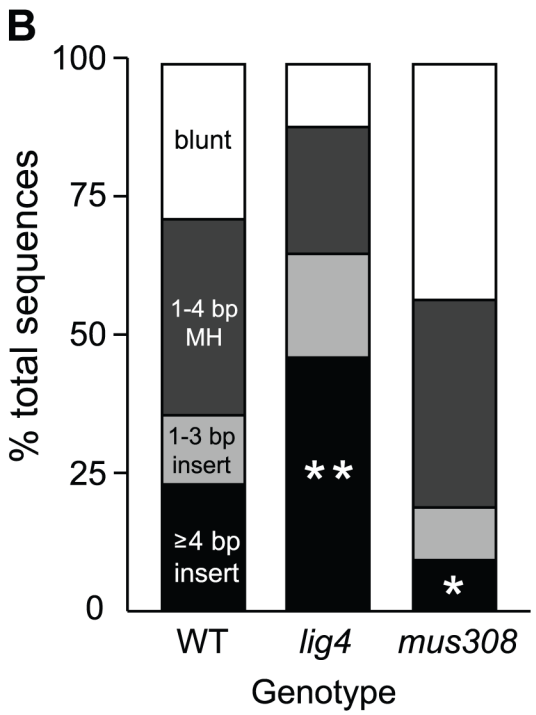
Alt-EJ repair in Drosophila is frequently associated with annealing at microhomologous sequences of more than four nucleotides and with long DNA insertions at repair junctions [8]. To determine whether pol θ-dependent end joining involves either of these types of repair, we sequenced repair junctions obtained from *spn-A* and *spn-A mus308* double mutants following *P{w^a}* excision. Because we sequenced only one junction per male germline, each junction analyzed represents an independent repair event. Five distinct junction types were identified. Three of these types are characteristic of junctions arising from C-NHEJ in mammalian systems [7]: junctions involving small, 1-3 base pair insertions, junctions involving annealing at 1-3 nucleotide microhomologies, and junctions for which no microhomologies can be identified (apparent blunt end junctions). The other two types of junctions, characteristic of alt-EJ [8], involve annealing at 5-10 nucleotide microhomologous sequences or insertions of more than three base pairs.

Figure 2-6. Pol θ is required for two distinct classes of alt–EJ repair products.



(A) Annealing at long microhomologies is greatly decreased in pol θ mutants.
Repair junctions isolated following *P{w^a}* excision were amplified from *spn-A* and *spn-A mus308* females of the indicated genotypes and sequences were aligned to the original *P{w^a}* chromosome. White, blunt joins; hatched, 5–10 bp microhomologies; light gray, 1–4 bp microhomologies; dark gray, 1–3 bp insertions; black, ≥4 bp insertions. Number of junctions sequenced: *spn-A* = 68, *spn-A^{093}mus308^{D5}/spn-A^{057}mus308^{2003}* = 23, *spn-A^{057}mus308^{D2}/spn-A^{057}mus308^{2003}* = 25. The percentage of 5–10 bp microhomologies was significantly decreased in both *spn-A mus308* mutants (P<0.01, Fisher's exact test).

Figure 2-6 (continued)



(B) Reduced frequency of ≥4 bp insertions during inaccurate end-joining repair of a chromosomal I-*Sce*I DSB in *mus308* mutants.
Complementary ended DSBs were induced by expression of I-*Sce*I endonuclease in the pre-meiotic germline of wild type (n = 70), *lig4*^{169a} (n = 83), and *mus308*^{2003}*/Df(3R)kar-Sz29* (n = 57) males, and independent inaccurate end-joining repair junctions were sequenced. White, blunt joins; light gray, 1–4 bp microhomologies; dark gray, 1–3 bp insertions; black, ≥4 bp insertions. No >4 bp microhomologies are available near the DSB in this system. * indicates $P = 0.03$, ** indicates $P<0.01$, relative to wild type (Fisher's exact test).

Table 2-1. *P{w[a]}* repair junctions recovered from *spn-A[093/057]* mutants.

Type of repair event	Sequence 5′ of break[a]	Microhomology/inserted sequence	Sequence 3′ of break[a]	Number of isolates
Original sequence	acccagacCATGATGAAATAACATA	-	TATGTTATTTCATCATGacccagac	-
Long microhomology[b]				
	acccagac	(CATgATGA)[c]	cccagac	14
	acccagac	(CATGA)	cccagac	4
	none[d]	(TGACCCAGAC)	-	2
Short microhomology				
	acccagacCATG	(ATG)	TTATTTCATCATGacccagac	1
	acccagacCA	(TGA)	cccagac	1
	acccagacCATGATGAAATAA	(CAT)	Gacccagac	3
	acccagacCATGATGAAATAAC	(AT)	GTTATTTCATCATGacccagac	14
	acccagacCATGATGAAATAACA	(TA)	TGTTATTTCATCATGacccagac	5
	acccagacCATGATGAAA	(TA)	TGTTATTTCATCATGacccagac	1
	acccagacCATGATGAA	(AT)	GTTATTTCATCATGacccagac	1
	none[d]	(T)	TCATGacccagac	1
Blunt join				
	acccagacCATGATGAAATA	-	TTATTTCATCATGacccagac	1
Insertion[e]				
	acccagacCATGATGAAATAACATA	A	CATCATGacccagac	1
	acccagacCATGATGA	G	CATCATGacccagac	1
	acccagacCATGATGAAATAACATA	AC	ATGTTATTTCATCATGacccagac	3
	acccagacCATGATGAAATAACAT	GT	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGA	AC	ATGTTATTTCATCATGacccagac	1
	acccagacCA	AG	ATGacccagac	1
	acccagacCATGATGAAATAACATA	TTC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	GTTA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TTTAT	TGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	TTATCA	TGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	TTAACATAAC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TTATTATTATA	TTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	GAAATAATAAC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	GTATTACATAAC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAA	TAATAATAATATAA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATA	TCATGAAATATCATA	TCATCATGacccagac	1

a Uppercase letters represent the 17-nucleotide 3′ single-stranded tails that remain following transposase action, lowercase letters correspond to the 8 base pair target sequence duplicated upon *P* element insertion.
b Microhomologies (in parentheses) are sequences that could have been derived from either side of the break site. Long microhomologies were five or more nucleotides, while short microhomologies were three or fewer nucleotides.
c represents an 8-nucleotide imperfect microhomology.
d indicates a deletion that extends past the 8 base pair target sequence.
e Insertions were identified as any sequence not present at the original break site. Templated insertions and corresponding potential templates in flanking sequences are underlined.
doi:10.1371/journal.pgen.1001005.t001

Table 2-2. *P{w^a}* repair junctions recovered from *spn-A$^{093/057}$ mus308$^{D5/2003}$* mutants.

Type of repair event	Sequence 5′ of break	Microhomology/inserted sequence	Sequence 3′ of break	Number of isolates
Original sequence	acccagacCATGATGAAATAACATA	-	TATGTTATTTCATCATGacccagac	-
Short microhomology[a]				
	acccagacCATGATGAAATAA	(CAT)	CATGacccagac	2
	acccagacCATG	(ATG)	TTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAAC	(AT)	GTTATTTCATCATGacccagac	5
	acccagacCATGATGAAATAACA	(TA)	TGTTATTTCATCATGacccagac	2
Blunt join				
	acccagacCATGATGAAATAACATA	-	TTATTTCATCATGacccagac	1
Insertion[b]				
	acccagacCATGATGAAATAACAT	G	TGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	AC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	GTTA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TGTA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACA	GTGAA	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	GTTATGT	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	GTTATACA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TATGTTATAACA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAA	TCATGTTATTTC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TAACATGAATAAC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TATAATGTTATAACATAT-AACATATGTTATGAAATAATAACA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	CATCATTTATCATTTATTATTATTA-TTATTTATTATTATTTATTATTTA	TTATTTCATCATGacccagac	1

a Microhomologies (in parentheses) are sequences that could have been derived from either side of the break site.
b Insertions were identified as any sequence not present at the original break site. Templated insertions and corresponding potential templates in flanking sequences are underlined.
doi:10.1371/journal.pgen.1001005.t002

Table 2-3. *P{w[a]}* sequences recovered from *spn-A*$^{093/057}$ *mus308*$^{D2/2003}$ mutants.

Type of repair event	Sequence to left of break	Microhomology/inserted sequence	Sequence to right of break	Number of isolates
Original sequence	acccagacCATGATGAAATAACATA	-	TATGTTATTTCATCATGacccagac	-
Long microhomology[a]				
	acccagac	(CATgATGA)	cccagac	1
Short microhomology				
	acccagacCATGATGAAATAAC	(AT)	GTTATTTCATCATGacccagac	7
	acccagacCATGATGAAATAACA	(TA)	TGTTATTTCATCATGacccagac	3
	acccagacCATGATGA	(AT)	GTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACA	(T)	TATTTCATCATGacccagac	1
Blunt join				
	acccagacCATGATGAAATAACATA	-	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACAT	-	TATGTTATTTCATCATGacccagac	1
Insertion[b]				
	acccagacCATGATGAAATAACAT	T	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	CA	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	AC	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TA	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAA	TGT	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TGT	TATGTTATTTCATCATGacccagac	2
	acccagacCATGATGAAATAACATA	ACATAA	ATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TATACCG	TATGTTATTTCATCATGacccagac	1
	acccagacCATGATGAAATAACATA	TGTTATAAC	ATGTTATTTCATCATGacccagac	1

a Microhomologies (in parentheses) are sequences that could have been derived from either side of the break site.
b Insertions were identified as any sequence not present at the original break site. Templated insertions and corresponding potential templates in flanking sequences are underlined.
doi:10.1371/journal.pgen.1001005.t003

Approximately 58% of junctions from *spn-A* mutants showed structures considered typical of C-NHEJ repair, while 29% involved annealing at 5-10 nucleotide microhomologies and 13% had insertions of greater than three base pairs (Figure 2-6A and Table 2-1). Potential templates for the larger insertions could almost always be identified in flanking sequences. These insertions may be analogous to the apparently templated insertions sometimes referred to as T-nucleotides that have been observed at translocation breakpoint junctions isolated from certain human cancers [21-23].

When we sequenced repair junctions from *spn-A mus308* mutants, we observed two distinct patterns, depending on the *mus308* alleles used. For both the *D2/2003* and *D5/2003* allele combinations, the percentage of junctions involving annealing at long microhomologies

was significantly decreased ($P<0.01$, Fisher's exact test; Figure 2-6A, Tables 2-2 and 2-3). Only 12% of *D2/2003* junctions possessed an insertion greater than three base pairs, compared to 44% of junctions recovered from males with the *D5* and *2003* alleles. In addition, most insertions isolated from *D5/2003* males were highly complex and had multiple copies of imperfect repeats of T-nucleotides (underlined sequences in tables 2-2 and 2-3). Similar results were obtained with the *D5/3294* allele combination (data not shown). An overall comparison of insert length showed that flies with wild-type *mus308* alleles had an average insert length of 5.5 nucleotides, compared to 3.8 nucleotides for *D2/2003* mutants and 13.3 nucleotides for *D5/2003* mutants.

In summary, both *mus308* mutant combinations significantly abrogated annealing at long microhomologies during alt-EJ repair. However, we observed a distinct difference in repair junctions recovered from males harboring the *D2* allele, which greatly reduces overall pol θ protein levels [38], compared to flies with the *D5* allele, which alters a conserved residue near the helicase-like domain. These results suggest that pol θ has two distinct functions in alt-EJ: one that promotes the annealing of long microhomologous sequences during end alignment, and another that is responsible for complex T-nucleotide insertions. Flies with the *D2* allele are impaired in their ability to carry out both the annealing and insertion functions, whereas flies possessing the *D5* separation-of-function allele cannot perform the microhomology annealing function but can still produce complex insertions.

Pol θ operates in alternative end-joining of complementary-ended double-strand breaks

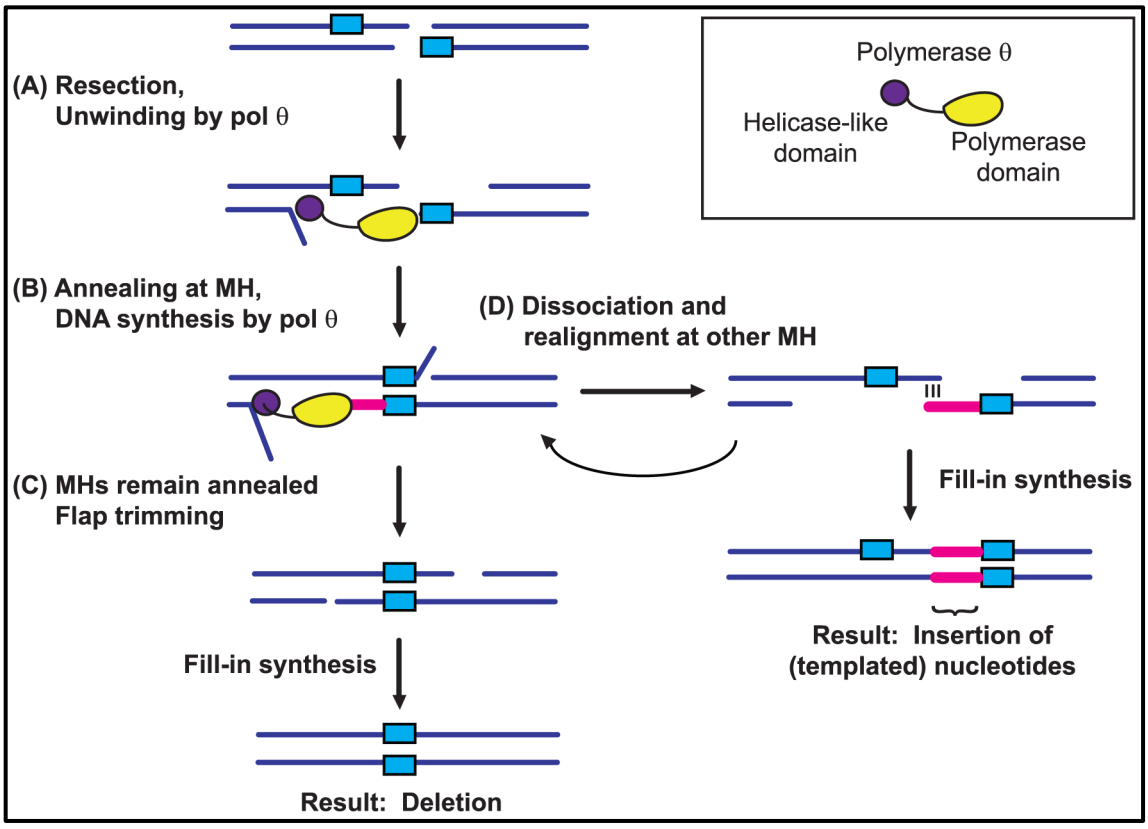
P element-induced breaks are unique in that they possess 17-nucleotide non-complementary ends that are poor substrates for C-NHEJ. To test whether the results obtained with *P* elements can be generalized to other types of breaks, we used the I-*Sce*I endonuclease and the previously characterized *[Iw]7* reporter construct [50] to create site-specific DSBs in wild-

type flies and flies lacking either DNA ligase 4 or pol θ. I-*Sce*I produces a DSB with 4-nucleotide complementary overhangs that can be directly ligated through a C-NHEJ mechanism [50,51]. Accurate repair regenerates the original I-*Sce*I recognition sequence, which can then be cut again, while inaccurate end-joining repair abolishes further cutting. We utilized an *hsp70* or ubiquitin-driven I-*Sce*I construct integrated on chromosome *2* to drive high levels of I-*Sce*I expression [50,52]. Nearly 100% of repair events that we recovered involved gene conversion (HR repair from the homologous chromosome) or inaccurate end joining (data not shown). In the *[Iw]7* system, both gene conversion events and large deletions that remove the *white* marker gene are phenotypically indistinguishable. PCR analysis confirmed that many repair events recovered from *mus308* mutants involved large deletions (>700 base pairs, data not shown). Our subsequent analysis focused on the characterization of repair events involving smaller deletions.

Twenty-three percent of I-*Sce*I repair junctions isolated from wild-type flies possessed insertions of more than 3 base pairs (Figure 2-6B). This percentage was significantly increased to 46% in *lig4* mutants ($P<0.01$, Fisher's exact test), consistent with increased use of alt-EJ in the absence of C-NHEJ. If pol θ plays a general role in insertional mutagenesis during alt-EJ repair, one would predict that the frequency and length of insertions following I-*Sce*I cutting should decrease in *mus308* mutants. Indeed, the percentage of large insertions decreased to 9% in *mus308* mutants ($P=0.03$, Fisher's exact test). Wild-type flies had an average insertion length of 7.6 base pairs, compared to 4.2 base pairs for *mus308* mutants. Strikingly, no *mus308* insertion was longer than twelve base pairs, while insertions of more than twenty base pairs occurred in both wild type and *lig4* mutants. Because microhomologies of greater than four base pairs are not present near the I-*Sce*I cut site in this construct, repair involving annealing at long microhomologies was not observed.

Surprisingly, the total percentage of repair junctions with short, 1-3 base pair insertions was not decreased in *lig4* mutants relative to wild type (17% vs. 13%, respectively). Furthermore, the percentage of junctions involving annealing at 1-3 nucleotide microhomologies was also similar between the two genotypes (25% for *lig4* mutants vs. 34% for wild type). These two types of junctions have historically been associated with ligase 4-dependent C-NHEJ repair. Our results suggest that this may not be the case. Indeed, a fine-level sequence analysis of I-*Sce*I repair junctions that we have recently conducted suggests that alt-EJ may produce C-NHEJ-like junctions in certain sequence contexts [53]. Nevertheless, our data obtained using two independent site-specific DSB repair assays strongly suggest that C-NHEJ and alt-EJ represent at least partially independent mechanisms for the repair of DSBs and that pol θ plays an important role in the generation of T-nucleotide insertions during alt-EJ repair of both *P* element and I-*Sce*I-induced breaks.

Figure 2-7. Model for pol θ function in alt–EJ



Pol θ is drawn as a bipartite protein, with a helicase domain and a polymerase domain separated by a flexible linker region.

(A) Initially, the helicase activity of pol θ unwinds short stretches of double-stranded DNA to expose pre-existing microhomologous sequences.

(B) These microhomologies (MH) are used to align the broken ends to provide a template for pol θ polymerase activity. The unwinding activity could also serve to make the polymerase more processive.

(C) Processing of the ends and ligation results in repair accompanied by a deletion.

(D) In cases where the ends do not remain stably aligned, annealing at other microhomologies closer to the break site could result in the insertion of T-nucleotides. Multiple rounds of unwinding, synthesis, and alignment could result in the complex insertions that are often observed in alt-EJ in flies.

Discussion

Several studies have identified proteins important for end-joining repair of DSBs in the absence of C-NHEJ factors in yeasts [11, 54, 55]. More recently, the Mre11 protein has been identified as an important alt-EJ component in vertebrate systems [10, 56-59]. However, much remains uncertain about the genetics or mechanisms of alt-EJ in metazoans. Our previous work demonstrated that end-joining repair of *P* element-induced breaks can occur independently of DNA ligase 4, suggesting the presence of a highly active alternative end-joining mechanism in Drosophila [17]. We have now identified the *mus308* gene, encoding DNA polymerase theta, as a critical component of alt-EJ. Pol θ-dependent alt-EJ operates in parallel to C-NHEJ to promote repair of both *P* element and I-*Sce*I-induced breaks. Because we observed similar alt-EJ defects with four different *mus308* mutant alleles, several of which were studied *in trans* to at least two independently-derived deficiencies, we consider it highly unlikely that the phenotypes we observed are due to second-site mutations or other differences in genetic background.

Importantly, pol θ does not appear to participate in homology-directed repair. HR repair of DSBs following *P{w^a}* excision is thought to proceed largely through synthesis-dependent strand annealing (SDSA) [60]. We observed that SDSA frequencies in the *P{w^a}* assay were similar in wild type and *mus308* mutants. Although we did not formally test for a role of pol θ in single-strand annealing (SSA), a non-conservative HR pathway that involves annealing at direct repeats larger than 25 nucleotides, a separate study demonstrated that pol θ has no effect on SSA repair [47]. In addition, comparison of the DNA sequences located 3 kilobases to either side of *P{w^a}* by BLAST does not reveal any significant similarities of more than 20 nucleotides. Therefore, it seems unlikely that the repair observed in *mus308* mutants arose through an SSA mechanism.

Three findings suggest that *mus308*-dependent alt-EJ is an important repair option for both cell and organism survival in flies, particularly in the absence of homologous recombination. First, *spn-A mus308* double mutants are sub-viable and have severe defects in adult abdominal cuticle closure, consistent with a high level of apoptosis in rapidly proliferating histoblasts during pupariation. Second, *spn-A mus308* double mutant males undergoing *P{w^a}* excision have up to 30-fold increased sterility relative to *spn-A* mutants. Third, *mus308* mutant males undergoing I-*Sce*I cutting show premature sterility and produce few progeny. The few germline repair events that are recovered from each male are frequently clonal, suggesting extensive germline apoptosis (A. Yu, unpublished data).

Evidence for two functions of pol θ in alternative end joining

Pol θ orthologs characterized from a variety of metazoans possess both helicase-like and DNA polymerase domains [27-31, 35]. Pol θ purified from both Drosophila and human cells has a Pol I-like polymerase activity and single-stranded DNA-dependent ATPase activity [30, 38]. However, DNA helicase activity of the purified protein remains to be demonstrated. Although our experiments did not formally test for helicase activity of pol θ, our results are consistent with pol θ having a DNA unwinding or strand displacement function. Flies with the *D5* and *3294* mutations (located in or near the conserved helicase domain) produce repair products with complex T-nucleotide insertions but not products involving annealing at long microhomologies. The *D5* and *3294* alleles may therefore encode proteins that retain polymerase activity but lack unwinding activity, resulting in an inability to expose internal microhomologous sequences. Because the microhomologies used in repair following *P* element excision are often located in the 17-nucleotide 3' single-stranded tails, pol θ may also be important for the unwinding of secondary structures that form in single-stranded DNA. Alternatively, the DNA-dependent

ATPase activity demonstrated by pol θ might represent an annealing function of the protein that is required during alt-EJ. Such an annealing activity was recently described for the human HARP protein, which is able to displace stably bound replication protein A and rewind single-stranded DNA bubbles [61].

One notable aspect of alt-EJ in Drosophila is the large percentage of repair junctions with templated insertions. These insertions may be "synthesis footprints" that are formed during the cell's attempt to create microhomologous sequences that can be used during the annealing stage of alt-EJ when suitable endogenous microhomologies are not present or are not long enough to allow for stable end alignment. Indeed, analysis of the insertions from I-*Sce*I repair junctions suggests a model involving local unwinding of double-stranded DNA and iterative synthesis of 3-8 nucleotide runs [53]. The *P{w^a}* repair junctions isolated from *mus308^{D2}/mus308^{2003}* mutants are consistent with an important (but not exclusive) role for the polymerase domain of pol θ in the synthesis of T-nucleotides.

We speculate that pol θ may be involved in both DNA unwinding and repair synthesis during alt-EJ (Figure 7). Linking these two activities in one protein would provide a convenient mechanism for creating longer microhomologies that could increase the thermodynamic stability of aligned ends prior to the action of a DNA ligase. Studies based on the crystal structure of a dual function NHEJ polymerase-ligase protein found in *Mycobacterium tuberculosis* suggest that a synaptic function for an NHEJ polymerase is plausible [62]. Because ligase 4 is not involved in alt-EJ in Drosophila, another ligase must be involved in the ligation step. Studies from mammalian systems have identified DNA ligase 3 as a likely candidate [63, 64].

Potential roles of pol θ in DNA interstrand crosslink repair

Pol θ was originally identified in Drosophila based on the inability of *mus308* mutants to survive exposure to chemicals that induce DNA interstrand crosslinks. A crucial question posed by our findings is whether pol θ performs a common function during the repair of both DSBs and interstrand crosslinks. The *C. elegans* pol θ ortholog, POLQ-1, is also required for resistance to interstrand crosslinks and acts in a pathway that is distinct from HR but depends on *CeBRCA1* [28]. In *S. cerevisiae*, several repair mechanisms are utilized during interstrand crosslink repair, including nucleotide excision repair (NER), HR, and translesion synthesis [65, 66]. Given our results and the findings from *C. elegans*, it seems unlikely that the role of pol θ in interstrand crosslink repair involves a function in HR.

In human cells, exposure to agents that induce interstrand crosslinks causes a shift in repair mechanisms that leads to increased use of non-conservative pathways associated with complex insertions and deletions [67]. Furthermore, interstrand crosslinks can cause frequent recombination between direct repeats [68, 69], suggesting that single-strand annealing may provide a viable mechanism for interstrand crosslink repair. The single-strand annealing model of interstrand crosslink repair posits that NER-independent recognition and processing of the crosslinked DNA is followed by generation of single-stranded regions flanking the crosslink and annealing at repeated sequences. Because alt-EJ frequently proceeds through annealing at short direct repeats, it is tempting to speculate that the role of pol θ in interstrand crosslink repair might be to expose and/or promote the annealing of microhomologous single-stranded regions that flank the crosslinked DNA. Consistent with this model, the initial incision step made after recognition of the interstrand crosslink remains normal in *mus308* mutants [26]. Alternatively, pol θ might utilize its polymerase activity and nearby flanking sequences as a template to

synthesize short stretches of DNA that could be used to span a single-stranded gap opposite of a partially excised crosslink. Such a model has been proposed to explain the formation of microindels in human cancers [70]. We are currently testing these two models using helicase- and polymerase-specific *mus308* mutant alleles.

Pol θ and alternative end joining: promoting genome (in)stability

Although it seems counterintuitive, alt-EJ likely functions in some situations to promote genome stability. As evidence of this, we found that DSB repair following *P* element excision in *mus308* mutant flies frequently results in genomic deletions of multiple kilobases. A similar deletion-prone phenotype was previously observed in *mus309* mutants, which lack the Drosophila BLM ortholog [43,71]. Epistasis analysis demonstrated that the *mus309* deletion phenotype depends on Rad51, implying that DmBlm acts after strand invasion during HR and that the deletions observed in *mus309* mutants are likely a result of failed SDSA [48]. In contrast, the deletions observed in *mus308* mutants do not depend on Rad51, demonstrating that the function of pol θ in DSB repair is independent of HR. The deletion phenotype is exacerbated in *lig4 mus308* double mutants, suggesting that C-NHEJ and alt-EJ represent two parallel mechanisms to prevent deletions. In the absence of these two end-joining options, resection at the broken ends may continue unchecked, resulting in extensive genomic deletions that are generated by an unknown Rad51-independent repair mechanism. Therefore, both C-NHEJ and alt-EJ function to prevent overprocessing of broken DNA ends and extreme degradation of the genome. Microhomology-mediated end joining, which shares many features with alt-EJ, has been proposed to perform a similar function in urothelial cells [72].

Nonetheless, alternative end-joining repair can also be genome destabilizing, as demonstrated by an increasing number of reports linking it to cancer. We have shown that

complex insertions observed in alternative end-joining products are more frequent in flies possessing pol θ. These insertions, which are often combinations of nucleotides derived from several templates inserted in both direct and reverse-complement orientations, are remarkably similar to T-nucleotide insertions found in translocation breakpoints reconstructed from follicular and mantle cell lymphomas (reviewed in [73]). Therefore, if pol θ also functions in alternative end joining and T-nucleotide generation in mammals, it might be an important factor involved in translocation formation.

A recent study suggests that pol θ levels are tightly regulated in humans and that loss of this regulation may promote cancer progression [37]. The protein is primarily found in lymphoid tissues but is upregulated in lung, stomach, and colon cancers. Furthermore, high levels of pol θ expression correlate with poorer clinical outcomes. Intriguingly, pol θ is regulated by endogenous siRNAs in Drosophila [74, 75], although the significance of this regulation is currently unclear. We suggest that polθ-mediated alt-EJ serves as a medium-fidelity repair option used by cells when precise repair cannot be carried out for any number of reasons. As such, it prevents extreme loss of genetic information. However, its error-prone nature requires tight regulation, which, when lost, may lead to excessive inaccurate repair and ultimately, carcinogenesis.

Conclusions

The results described here establish that Drosophila pol θ plays two distinct roles in an alternative end-joining mechanism operating in parallel to canonical DNA ligase 4-mediated C-NHEJ. This novel finding lays the groundwork for future studies focusing on the specific roles of the pol θ helicase-like and polymerase domains in alt-EJ and DNA interstrand crosslink repair. Whether pol θ plays a similar role in alt-EJ in other organisms, including mammals, remains to be determined. Regardless, these studies reveal an unexpected role for DNA polymerase θ that is required for genomic integrity in Drosophila and possibly other metazoans.

Materials And Methods

Drosophila stocks and mus308 alleles: All flies were maintained on standard cornmeal-based agar food and reared at 25° C. The *mus308 D2* and *D5* stocks were obtained from the Bloomington Stock Center and the *2003* and *3294* stocks were from the Zuker collection [76]. To identify mutations in these stocks, genomic DNA was isolated from flies harboring each allele in trans to *Df(3R)Exel6166* and PCR was performed with primers specific to overlapping regions of the entire coding sequence. PCR products were sequenced and the sequence was compared to the Drosophila reference sequence release 5.10. Sequence changes unique to each allele were verified by sequencing in both orientations. The $lig4^{169a}$ [17], $spn\text{-}A^{093}$ and $spn\text{-}A^{057}$ [42] stocks harbor null alleles of DNA ligase 4 and Rad51, respectively.

Mutagen sensitivity studies: For mechlorethamine sensitivity assays, balanced, heterozygous parents were crossed to *Df(3R)Exel6166* and allowed to lay eggs in vials containing 10mL of food for three days, after which they were moved to new vials for two additional days. The first vials were treated with 250μL of 0.005% mechlorethamine dissolved in ddH_2O, while the second vials were treated only with ddH_2O. Survival was calculated as the number of homozygous mutant adults divided by the total number of adults that eclosed within 10 days of treatment. Ratios were normalized to untreated controls for each set of vials (five to eight sets of vials were counted for each experiment). For ionizing radiation sensitivity assays, heterozygous parents laid eggs on grape-juice agar plates for 12 hr. Embryos developed at 25°C until larvae reached third-instar stage, at which point they were irradiated in a Gammator 1000 irradiator at a dose rate of 800 rads/min and larvae were transferred to food-containing bottles. Relative survival rates were calculated as above.

P{w^a} assay: Repair of DNA double-strand breaks was monitored after excision of the *P{w^a}* transposon as described previously [43,77]. *P{w^a}* was excised in males using a second chromosome transposase source (*CyO, H{w$^+$,D2-3}*) and individual repair events were recovered in female progeny over an intact copy of *P{w^a}*. Females with two copies of *P{w^a}* have apricot eyes [78]. Progeny with red eyes possess a repair event involving HR with annealing of the *copia* LTRs. A fraction of apricot-eyed females also possess HR repair events, but these cannot be distinguished from chromosomes in which no excision event occurred (using the *CyO, H{w$^+$,D2-3}* transposase source , ~80% of apricot-eyed female progeny inherit a non-excised *P{w^a}* element). Yellow-eyed females harbor a repair event in which repair is completed by end joining.

For each genotype, at least 50 individual male crosses were scored for eye color of female progeny lacking transposase. The percentage of progeny from each repair class was calculated on a per vial basis, with each vial representing a separate experiment. Statistical comparisons were done with a Kruskal-Wallis non-parametric ANOVA followed by Dunn's multiple comparisons test using InStat3 (GraphPad).

For analysis of HR synthesis tract lengths, genomic DNA was purified and PCR reactions were performed as in [43], using primer pairs with the internal primer located 250, 2420, and 4674 base pairs from the cut site at the 5' end of *P{w^a}*.

For deletion analysis, the percentage of females with scalloped wings was calculated relative to all yellow-eyed females counted. The percentage of male lethal and small (0.1-3.6 kb) deletions was calculated based on a subset of yellow-eyed females (one from each original male parent) that were individually crossed to males bearing the *FM7w* balancer. Vials for which no white-eyed male progeny were recovered were scored as male lethal. Some of the male lethal

events also caused a scalloped-wing phenotype in heterozygous females. For those that did not, testing to ensure that the male lethality was due to deletion of *scalloped* coding sequence was performed by recovering the repaired chromosomes *in trans* to the hypomorphic sd^{1} mutation [79] and scoring for a scalloped-wing phenotype. Repair events which could be recovered in males were subjected to PCR analysis, using primers internal to $P\{w^{a}\}$ [43], to detect small deletions into one or both introns of *sd*.

I-SceI break repair assay: Repair of I-*Sce*I mediated DNA double strand breaks was studied in the context of the chromosomally integrated *[Iw]7* construct [52], which contains a single target site for the I-*Sce*I endonuclease. DSBs were induced in the male pre-meiotic germline by crossing females harboring *[Iw]7* to males expressing the I-*Sce*I endonuclease from a second chromosomal location under the control of either the hsp70 promoter (70[I-*Sce*I]1A) [52] or the ubiquitin promoter (UIE[I-*Sce*I]2R) [50]. Independent inaccurate end-joining repair events from the male pre-meiotic germline were recovered in male progeny and DNA was isolated for analysis [80]. PCR was performed using primers PE5' (GATAGCCGAAGCTTACCGAAGT) and jn3'b (GGACATTGACGCTATCGACCTA) to amplify a 1.3 kb fragment of the *[Iw]7* construct including the I-*Sce*I target site. Products were gel purified (GenScript) and sequencing of PCR products was performed using the PE5' primer. Sequences were aligned using ClustalW or by manual inspection against sequence obtained from an uncut *[Iw]7* construct. Statistical comparisons were done using Excel and SPSS.

Acknowledgements

We are grateful to members of the McVey, Mirkin, and Freudenreich labs for discussions that informed this work and to Jeff Sekelsky and Graham Walker for helpful comments. We would also like to thank the three anonymous reviewers whose suggestions greatly improved the paper. We thank the Bloomington stock center and Zuker labs for *mus308* mutant alleles, Bill Engels for the UIE-I-*Sce*I stock, and Yikang Rong for the *hsp70*-I-*Sce*I and *[Iw]7* stocks.

Chapter 3

Synthesis-Dependent Microhomology-Mediated End Joining Accounts For Multiple Types Of Repair Junctions

Amy Marie Yu[1] and Mitch McVey*[,1, 2]

[1]Department of Biology, Tufts University, Medford, MA 02155
[2]Program in Genetics, Tufts Sackler School of Graduate Biomedical Sciences, Boston, MA

This chapter is an updated and expanded version of

Nucl. Acids Res. (2010) 38 (17): 5706-5717. (doi: 10.1093/nar/gkq379)

Supplemental Materials not included in this document are available at Nucleic Acids Research Online.

Introduction

The fruit fly, *Drosophila melanogaster*, is one of the few model organisms in which alt-EJ can be readily observed in wild-type individuals. We have previously shown that end-joining repair of *P*-element-induced DSBs in Drosophila typically proceeds by MMEJ and is unaffected in *lig4* mutant flies (1). Conversely, studies using the I-*Sce*I endonuclease to induce DSBs support a role for Lig4 in Drosophila end joining (2, 3). This raises the question of whether alt-EJ repair of an I-*Sce*I DSB in flies uses an MMEJ mechanism similar to that observed at *P*-element DSBs, or a different type of alt-EJ.

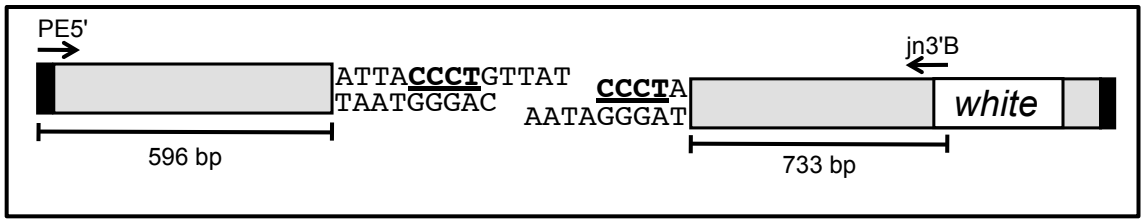
To address this question, we examined inaccurate repair junctions of a chromosomal I-*Sce*I DSB in C-NHEJ proficient and deficient flies. Annealing at pre-existing microhomologies did not appear to be important for end-joining repair in any genetic background. Sequence analysis of repair junctions suggested a model we term synthesis-dependent MMEJ (SD-MMEJ). We show that end-joining junctions with and without microhomology, as well as junctions with insertions, have structures consistent with predictions made by our SD-MMEJ model. Thus, in this system, a single mechanism accounts for multiple classes of end-joining repair products usually thought to be produced by different mechanisms.

Results

Drosophila alt-EJ does not require long microhomologies

End-joining repair of *P*-element-induced DSBs in Drosophila often produces repair products with microhomologies similar to those seen in Ku-independent MMEJ in budding yeast. To investigate microhomology use in alt-EJ of a non *P*-element DSB in flies, we used transgenic flies with one copy of the *[Iw]* DSB substrate to examine Ku-independent repair of a complementary ended I-*Sce*I DSB (4).

Figure 3-1. The *[Iw]* construct.



The *[Iw]* construct (4) is a chromosomally integrated *P*-element containing the 18 bp I-*Sce*I site and a mini-*white* gene 733 bp downstream of the I-*Sce*I site. I-*Sce*I expression induces a single complementary ended DSB with 4 bp 3' overhangs. Iterative cycles of cutting and repair continue until inaccurate end-joining repair or gene conversion destroys the I-*Sce*I site. *white* expression facilitates identification of end-joining repair events. A 4 bp microhomology (CCCT, underlined) is present at the break site. No other ≥ 4 bp microhomologies are available near the DSB. Primers used to amplify repair products are indicated by arrows.

The *[Iw]*/ I-*Sce*I system (Figure 3-1) induces a single complementary ended chromosomal DSB. High levels of I-*Sce*I expression ensure repeated cycles of cutting and repair until inaccurate end joining or gene conversion destroys the I-*Sce*I site.

Unlike *P*-element DSBs, the *[Iw]* DSB lacks ≥5 bp microhomologies near the break. In the *[Iw]* system, MMEJ at ≥5 bp of pre-existing microhomology results in deletion of over 70 bp. This enables a PCR-based assay to rule out use of ≥5 bp microhomologies.

In wild-type flies (n=90), 87.8% of independently recovered repair events yielded a PCR product inconsistent with use of ≥5 bp microhomologies (data not shown). Such products

constituted 82.9% of repair events in *ku70* mutants (n=82), which is not significantly different from wild-type (p=0.39, Fisher's Exact Test). Similar results were obtained for *lig4* mutants. Of 469 junctions sequenced in this study (Supplemental Tables S3-S8; available at NAR Online), only one involved annealing at a 5 bp microhomology, and junctional microhomologies greater than 5 bp were not observed.

C-NHEJ deficiency does not alter patterns of short junctional microhomology

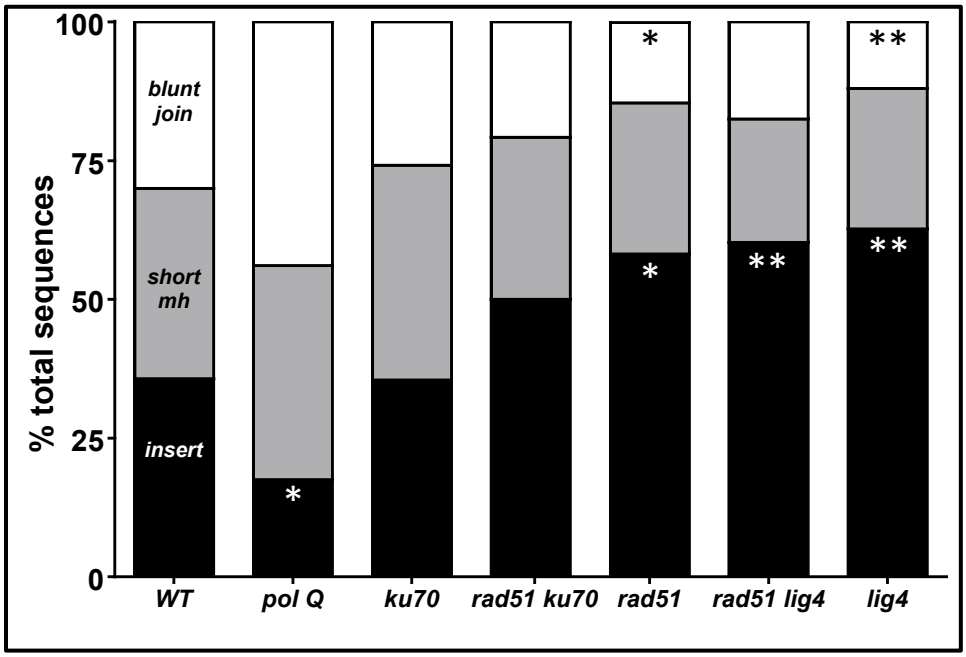


Figure 3-2. C-NHEJ deficiency increases frequency of net insertions.

Relative proportions of end-joining junction types (insertion, microhomology, apparent blunt join) by genotype. The relative proportion of junctions with insertions is significantly decreased relative to wild-type in *polQ* mutants (white asterisk; p=0.03, Fisher's Exact Test) and significantly increased relative to wild type (WT) in *rad51* (p=0.02), *rad51 lig4* (p<0.01), and *lig4* (p<0.01) mutants. The relative proportion of apparent blunt joins is significantly decreased relative to WT in *rad51* (black asterisk; p=0.05) and *lig4* (p<0.01) flies (Fisher's Exact Test). The relative proportion of repair products with junctional microhomology is not significantly different from wild-type in any genotype assayed. WT n=70; *polQ* n=57; *ku70* n=62; *rad51 ku70* n=48; *rad51* n=55; *rad51 lig4* n=63; *lig4* n=83.

In vertebrates, alt-EJ is often characterized by larger deletions and more frequent short junctional microhomologies. Comparison of *[Iw]7* repair junctions from C-NHEJ proficient and deficient flies showed no differences in mean or median deletion size (Figure S1a), frequency of junctional microhomologies (Figure 3-2), or length of junctional microhomologies (Table S1). Together, these results indicate that alt-EJ of this DSB is not characterized by the increased junctional microhomology typically observed in budding yeast and higher eukaryotes. Interestingly, this result is congruent with a recent study in which microhomology at end-joining junctions subsequent to a zinc finger nuclease induced DSB was not increased in *lig4* mutant

flies (5), suggesting that the failure to observe microhomology at inaccurate end-joining repair junctions is not unique to the *[Iw]* DSB (See also Chapter 4).

Alternative end joining produces insertions at repair junctions

A large number of *[Iw]* repair junctions had insertions (Figure S1b; (6)). The frequency of insertions was significantly increased relative to wild-type in *lig4* mutants, and decreased in *mus308* mutants, which lack DNA polymerase theta (PolQ) (Figure 3-2; (6)). PolQ is an A-family DNA polymerase known to be important for DNA crosslink repair (7, 8). PolQ has a unique structure: in addition to a polymerase domain, it also possesses an SF2 helicase-like domain attached to the polymerase domain by a highly conserved flexible linker region.

Many inserted sequences appeared to have been templated from sequence near the DSB. In Drosophila, *P*-element DSBs can be repaired by a hybrid HR/EJ pathway in which some Rad51-dependent templated repair synthesis takes place before repair is completed by end joining (9). To ask if these insertions were produced by a similar process, we sequenced repair junctions from *rad51*, *rad51 ku70*, and *rad51 lig4* flies. In all HR-deficient and C-NHEJ/HR double mutant genetic backgrounds, insertion frequency was comparable to or higher than wild-type (Figure 3-2). Therefore, the insertions result from an alternative end-joining mechanism that requires neither Ku nor Lig4, but is impaired in *polQ* mutants.

Genotype specific effects of sequence context on repair outcome

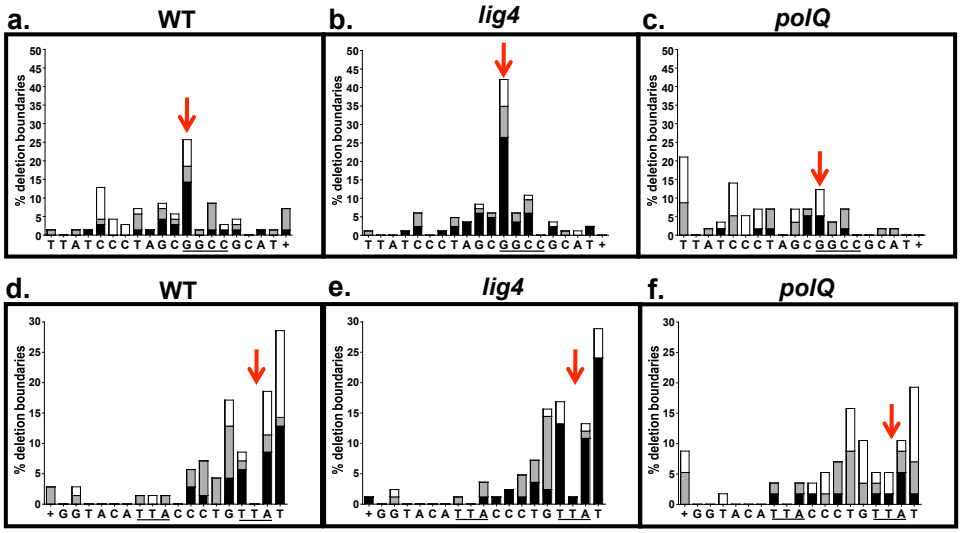


Figure 3-3. Combined effects of sequence context and genotype on deletion boundary frequency

Frequency as percentage of total repair products of deletion boundaries associated with net insertions (black), junctional microhomologies (gray) and apparent blunt joins (white) in wild-type, *lig4*, and *polQ* genetic backgrounds. "Deletion boundary" is defined as the position (including any junctional microhomologies) where an uninterrupted match between the original sequence and the repair product resumes. X axis indicates top strand sequence. Deletions extending beyond the sequence shown are represented by a plus sign (+).

a-c: Boundaries of deletions to the right of the bottom strand nick. The underlined sequence (GGCC) is the left half of a GGCC repeat. Red arrows show clustering of deletion boundaries at this repeat in wild-type and *lig4*, but not *polQ* sequences.

d-f: Deletion boundaries to the left of the top strand nick. The underlined sequence (TTA) is a 3 bp direct repeat. Note different overall shape of the histograms of deletion boundaries to the right and left of the DSB in all genotypes (compare a-c, d-f). Red arrows show virtual absence of 2 bp deletions in wild-type and *lig4*, but not *polQ* sequences. WT: n=70; *lig4:* n=83; *polQ:* n=57.

To gain insight into the mechanisms responsible for the insertions, we carried out fine-level sequence analysis of the repair junctions. This revealed patterns suggesting that genetic background and sequence composition interact to influence repair outcome. Deletion boundary frequency histograms (Figure 3-3; Figure S2) showed that to the right of the DSB, deletion boundaries cluster at a 4 bp GGCC repeat in wild-type and *lig4*, but not *polQ* flies (Figure 3-3 panels a-c, red arrows). The sequence to the left of the DSB lacks a comparable repeat, and

similar clustering was not observed (Figure 3-3 panels d-f). However, deletions of 3 bp to the left of the top strand nick (red arrow) were virtually absent in wild-type and *lig4* flies. This bias was also attenuated in *polQ* flies. Together, these observations suggest that flanking sequence influences repair outcome, and that PolQ contributes to this sequence specific effect.

Templated insertions and deletion boundaries correlate with specific patterns of short repeats in the original sequence

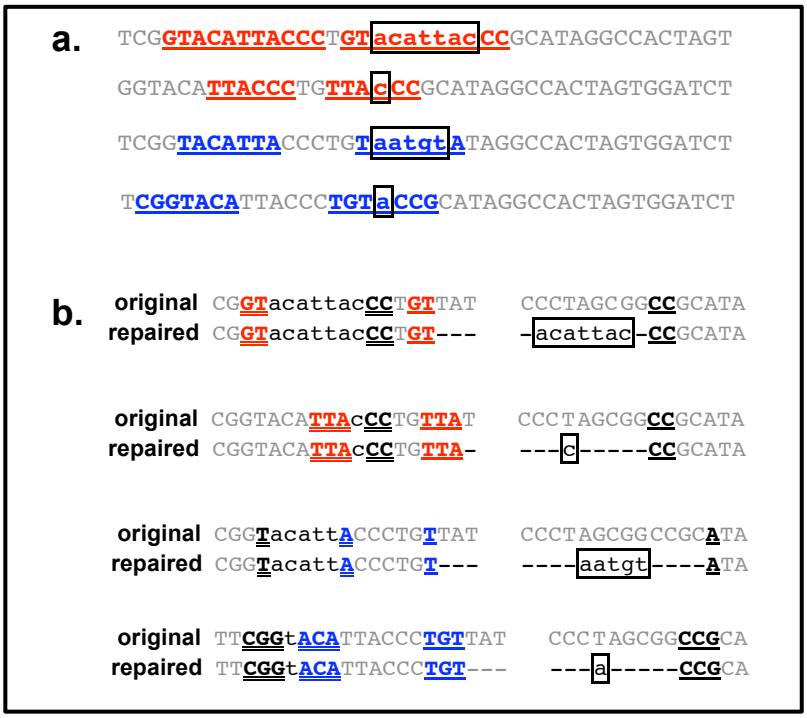


Figure 3-4. Repair products contain repeated motifs with a consistent pattern.

a. Sequences of representative repair products with insertion/deletion (indel) events. In each case, the repair event contains a direct (red) or inverted (blue) repeat. The repeated motif (underlined) comprises the insertion (boxed, lowercase) and flanking sequence on each side. Repeats are found both in repair products with insertions of sufficient length to be identified as templated and in repair products with apparently random short insertions. Sequences are (top to bottom) Table S8 sequence 83; Table S5 sequences 1-10; Table S8 sequence 125; Table S5 sequence 15. Additional similar examples are given in Figure S3.

b. Repair events from panel **a** above, aligned with the original *[Iw]* sequence. Only top strands are shown. Whitespace indicates the top strand nick. Dashes indicate bases deleted in the repair product. Net insertions are boxed and in lowercase. The apparent insertion template is in lowercase, not boxed. *Red/blue, underlined:* Short direct (red) or inverted (blue) repeat near but not spanning the DSB. *Black, underlined:* short direct or inverted repeat spanning the DSB.

Many inserted sequences formed direct or inverted repeats with nearby sequence unaffected by the repair event, suggesting that the insertions may have been templated. Sequence analysis to identify potential insertion templates revealed that repeated motifs included

not only the insertion itself, but also flanking sequence on both sides of the insertion (Figure 3-4a; Figure S3). Notably, this repeat pattern was present not only in insertions of sufficient length to show obvious similarity to surrounding sequence, but also in short 1-3 bp insertions.

Alignment of these repair products with the original *[Iw]* sequence (Figure 3-4b; Figure S3) showed a consistent relationship between two short repeats present in the original sequence, the insertion template, and the repair product. The insertion template and the deletion were each bounded by two short repeats. One of the short repeats was situated entirely to one side of the DSB; the other short repeat spanned the DSB. The location of these short repeats relative to each other and to the putative insertion template suggested that mispairing at these short repeats could prime synthesis at the insertion template.

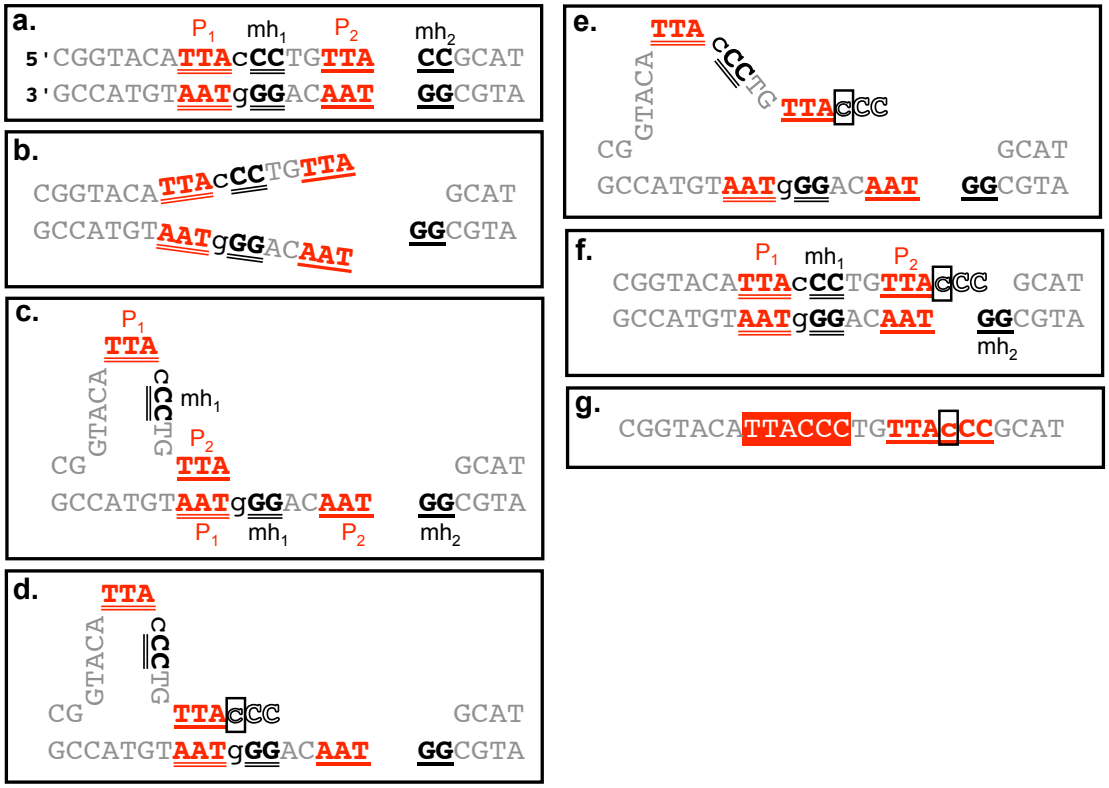


Figure 3-5. Loop-out SD-MMEJ produces repair products with direct repeats.

Example is taken from Figure 4. Both strands are shown. Deleted bases are not shown.

a. Loop-out SD-MMEJ requires two short direct repeats. The "primer repeat" (P1, P2; red) is located completely to one side of the DSB. The "microhomology repeat" (mh1, mh2; black) spans the DSB. Notational convention is that P2 and mh2 are closest to the DSB.

b. The DNA double helix is unwound through P1.

c. P2 on the top strand base-pairs with P1 on the bottom strand. mh1 and P1 on the top strand are looped out.

d. A template-dependent DNA polymerase extends through the end of mh1. Newly synthesized bases are in white type outlined in black. The base that will form the net insertion is boxed and in lowercase.
e. The loop-out structure dissociates or is unwound.

f. P1 and P2 re-anneal as originally. This DSB end now has a single stranded 3' overhang that is complementary to mh2.

g. The resulting repair product showing the net insertion and the direct repeat spanning the junction created by loop-out SD-MMEJ repair. (Compare with Figure 4b.) The part of the repeat that was present in the original sequence (white text on red) comprises P1, mh1 and the sequence between them. The part of the repeat containing the insertion (underlined, red) comprises P2, mh2, and a copy of the sequence between P1 and mh1. This convention is used throughout the remaining figures/tables.

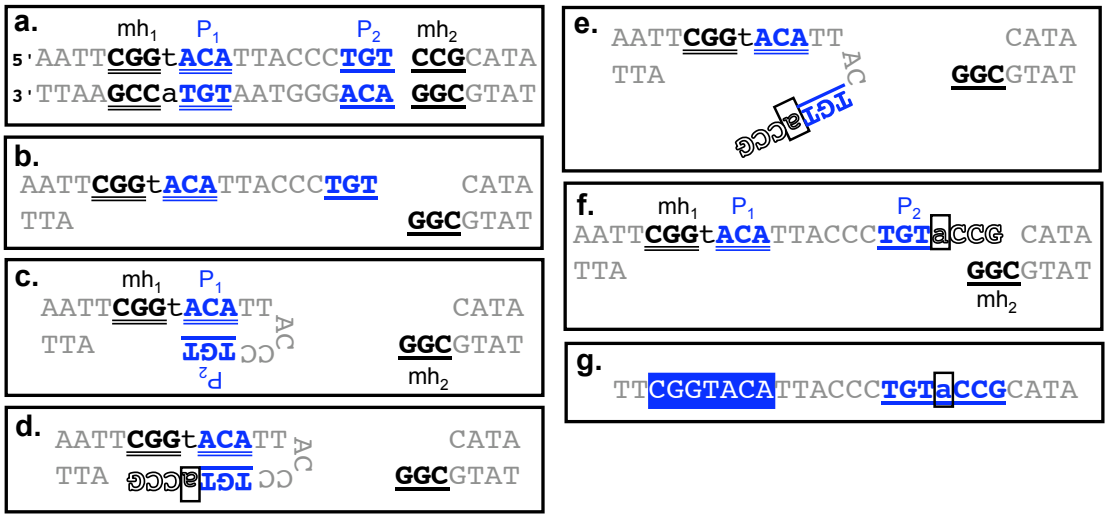


Figure 3-6. Snap-back SD-MMEJ produces repair products with inverted repeats.

Example is taken from Figure 4. Both strands are shown. Deleted bases are not shown.

a. Snap-back SD-MMEJ requires two short direct repeats. The "primer repeat" (P1, P2; blue) is located completely to one side of the DSB. The "microhomology repeat" (mh1, mh2; black) spans the DSB. Notational convention is that P2 and mh2 are closest to the DSB.

b. The DNA double helix is resected or unwound unwound through mh1. In contrast to loop-out SD-MMEJ (compare Figure 5, panels b and c), ssDNA may be generated by either nuclease or helicase activity.

c. P2 base-pairs with P1, forming a hairpin with a 3' end.

d. A template-dependent DNA polymerase extends through the end of mh1. Newly synthesized bases are in white type outlined in black. The base that will form the net insertion is boxed and in lowercase.
e. The hairpin dissociates or is unwound.

f. P1 and P2 re-anneal as originally. This DSB end now has a single stranded 3' overhang complementary to mh2.

g. The resulting repair product showing the net insertion and the inverted repeat spanning the junction created by snap-back SD-MMEJ repair. (Compare Figure 4b). The part of the repeat that was present in the original sequence (white text on blue) comprises mh1, P1 and the sequence between them. The part of the repeat containing the insertion (underlined, blue) comprises the reverse complements of P2, mh2, and the reverse complement of the sequence intervening between mh1 and P1. This convention is used throughout the remaining figures/tables.

The SD-MMEJ model shows that potential templates can be determined for short as well as long insertions

Evidence that PolQ may be involved in an alt-EJ mechanism that generates templated insertions, together with the patterns that emerged from our sequence analysis, suggested the following model, which we term synthesis-dependent MMEJ (SD-MMEJ).

In SD-MMEJ, a short direct or inverted repeat (Figures 3-5a and 3-6a) primes synthesis of *de novo* microhomologies at DSB ends. Resection or unwinding exposes short direct or inverted repeats near the DSB (Figures 3-5b and 3-6b). These short "primer repeats" mediate formation of transient secondary structures (Figures 3-5c and 3-6c), which prime synthesis by a template dependent non-processive DNA polymerase (Figures 3-5d and 3-6d). If the newly synthesized 3' end is complementary to sequence on the other side of the DSB, dissociation or unwinding of the structure (Figure 3-5e and 3-6e) allows annealing across the break. The result of this process is creation of a repeated motif that contains the insertion and flanking sequence on both sides (Figures 3-5g and 3-6g).

SD-MMEJ repair can produce apparent blunt joins and microhomology junctions

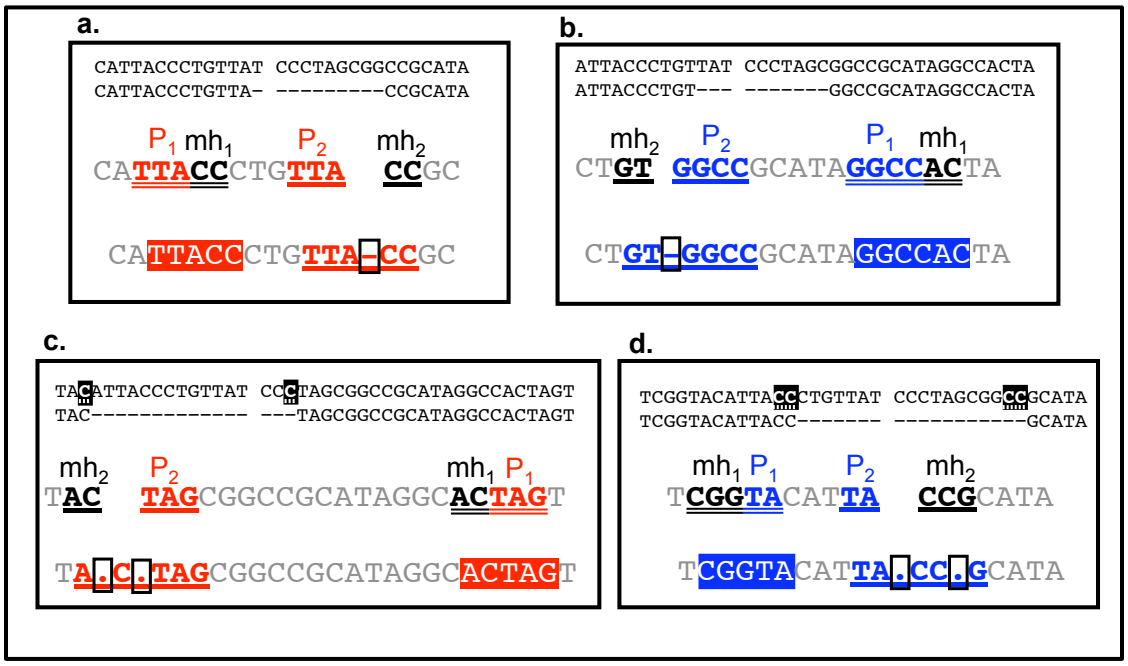


Figure 3-7. SD-MMEJ repair can produce repair junctions without insertions, which may or may not have junctional microhomology.

If P1 and mh1 are directly adjacent, SD-MMEJ will not produce an insertion. To efficiently summarize the proposed SD-MMEJ mechanism, this "insertion" of length zero is indicated by a box containing no sequence. In the case of ambiguous deletion breakpoints (junctional microhomologies), the ambiguous bases are flanked by boxes containing dots. Each panel shows alignment of the original sequence and repair product, the primer and microhomology repeats, and the longer repeat created in the repair product. Notational conventions are as in Figures 5 and 6.

a, b: Apparent blunt joins consistent with SD-MMEJ repair. Box with dash indicates the point of ligation (zero bp insertion). **a**: sequence H, Table S7. **b**: sequence B, table S7.

c, d: SD-MMEJ consistent repair products with junctional microhomologies. In both of these cases, SD-MMEJ results in annealing at longer microhomologies than would MMEJ at the microhomologies present in the repair product. **c**: sequence O, Table S8. **d**: sequence F, Table S8.

A thorough evaluation of the properties of the SD-MMEJ mechanism suggests that although SD-MMEJ is a microhomology-mediated process, it can produce junctions with no insertion and no junctional microhomology.

Consider that SD-MMEJ as illustrated in Figures 3-5 and 3-6 necessarily produces a repair product with a direct or inverted repeat. The repeated motif comprises the net insertion

and flanking sequence on both sides of the insertion. The size of the net insertion is equal to the distance between P1 (the primer) and mh1 (the template for the newly synthesized microhomology) (Figure 3-5a and 3-6a). It directly follows that if P1 (the primer) and mh1 (the template for the newly synthesized microhomology) are directly adjacent (ie, the distance between them is zero), SD-MMEJ repair will produce a junction with no net insertion (Figure 3-7). This "zero-insertion" SD-MMEJ repair event will not necessarily have any ambigious nucleotides at the repair junction (Figure 3-7 panels a and b). The net result is a microhomology-mediated event that produces a repair junction that superficially resembles a blunt join. Nevertheless, "zero-insertion" SD-MMEJ products with no junctional microhomology can be distinguished from true blunt joins because they will exhibit the same repeat structure as SD-MMEJ products with insertions. This criterion can be used to classify repair junctions without net insertions as consistent or not consistent with repair by SD-MMEJ.

Microhomology junctions can likewise be SD-MMEJ consistent. Under the convention of assigning ambiguous nucleotides to the left side of the break, a zero-insertion SD-MMEJ repair event will show junctional microhomology when aligned with the original sequence if, as in Figure 3-7c and 3-7d, one or more of the deleted nucleotides immediately adjacent to P2 is repeated in mh2. Consequently, there is at least one base that it is not possible to unambiguously assign to primer repeat P2 or to microhomology repeat mh2.

SD-MMEJ consistency criteria for specific junction structures

The essential criterion of SD-MMEJ consistency is that the repair junction has a repeat structure that is consistent with templated synthesis of overhangs complementary to sequence across the DSB. The specific guidelines for classifying repair junctions that follow from this criterion vary slightly in the context of different repair junction structures.

An apparent blunt join is SD-MMEJ consistent if it has a repeated motif near the repair junction such that the apparent point of ligation is inside (not on the border of) one of the two repeated motifs. Interestingly, the motifs in a valid SD-MMEJ consistent repeat can partially overlap.

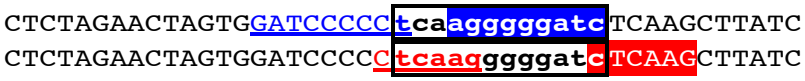
Microhomology junctions differ from apparent blunt joins in not having a single unique point of apparent ligation. It can be shown that an SD-MMEJ mechanism can be reconstructed for any microhomology junction with an SD-MMEJ consistent repeat intersecting any one of the possible ligation points at a microhomology junction.

Evaluation of insertion junctions for SD-MMEJ consistency is complicated by the possibility that multiple rounds of synthesis can take place before the final ligation event. In addition to the simple templated insertions described in Figure 3-4, we also recovered a substantial number of complex insertions comprising multiple overlapping copies of nearby sequence, including some with evidence for snap-back self-templating on newly synthesized 3' ends. A subset of this type of complex insertion is consistent with SD-MMEJ involving multiple rounds of synthesis and dissociation from one or both sides of the DSB (Figure S4).

Complex insertions are SD-MMEJ consistent if there is a series of overlapping repeats such that, starting from the endpoints of the flanking sequence remaining after any deletion event, the final repair product can be constructed by iterative rounds of synthesis from the free 3' ends, culminating in base-pairing across the break. From this description, it initially appears that complex insertions will be SD-MMEJ consistent simply if they can be contiguously tiled by a series of overlapping repeats; however, this criterion turns out to be insufficient, due to the possibility that newly synthesized sequence can itself serve as a template. Because not all synthesis templates are necessarily located in sequence that existed at the beginning of the repair

event, to confirm that a complex insertion is indeed SD-MMEJ consistent, it must be verified at each step of the mechanism that the postulated template actually exists. An illustration of this is provided in Figure 3-8 below.

Figure 3-8. An apparently templated complex insertion that is not SD-MMEJ consistent



This repair junction fits the criteria of being fully tiled by overlapping repeats, and all the inserted sequence appears to be templated from surrounding sequence. Nevertheless, this repair product is *not* SD-MMEJ consistent because the sequence that would be necessary for the initial priming events does not yet exist at the initiation of repair. After the deletion, the sequence that exists before any repair synthesis has taken place is in bold/black; the inserted sequence which has not yet been synthesized is grayed out.

CTCTAGAACTAGTGGATCCCCCtcaaggggatc**TCAAGCTTAC**

It is not possible to begin synthesis of the blue inverted repeat on the left side, because the t-a pair that would prime that event does not yet exist. Likewise, the red direct repeat cannot be synthesized because right-hand motif of the primer repeat does not yet exist. It is of course possible to speculate about the process beginning on each side with 2-3 bp repeated motifs, but such speculation is not readily distinguishable from pandering to the whim of the investigator. Therefore, although this repair junction appears to contain a templated insertion, it is not SD-MMEJ consistent. This example serves to illustrate that any procedure for evaluating the SD-MMEJ consistency status of a repair junction with an insertion must verify that there is a full primer repeat and template available for every proposed round of synthesis. Repair junction taken from (10).

A consequence of the requirement that the synthesis template be verified at each step is that evaluation of SD-MMEJ consistency for junctions with insertions requires algorithmically intricate direct modeling rather than simple pattern matching. Extensive characterization of complex longer insertions was therefore not pursued further in this study.

Significant majorities of apparent blunt joins, microhomology junctions, and short insertions are SD-MMEJ consistent

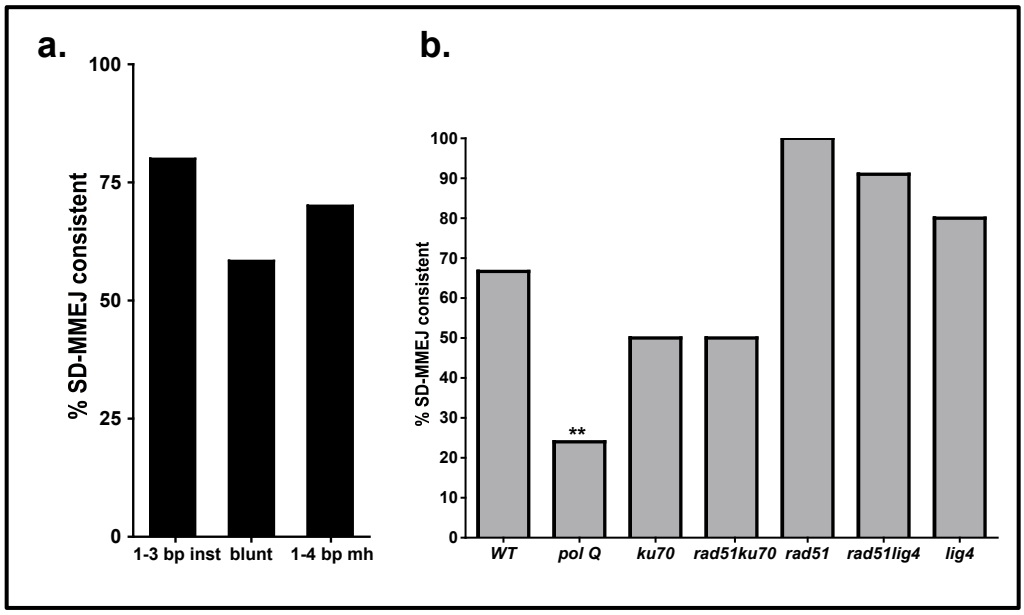


Figure 3-9. A majority of 1-3 bp insertions, apparent blunt joins, and microhomologies are SD-MMEJ consistent.

a. 80.0% of all repair products with 1-3 bp insertions (inst), 58.4% of all apparent blunt joins, and 70.1% of all junctions with microhomologies are SD-MMEJ consistent (a ≥4 bp direct or inverted repeat that contains the breakpoint(s) is present within ±20 bp of the junction). For all three junction types, the observed frequency of SD-MMEJ consistent repair products is significantly higher than the estimated frequency expected by random chance for this sequence: 1-3 bp insertions (n=71), $p<0.01$; apparent blunt joins (n=101), $p<0.05$; microhomologies (n=134) $p=0.05$; Fisher's Exact Test.

b. The proportions by genotype of SD-MMEJ consistent blunt joins follow a pattern similar to the proportions of net insertions by genotype (compare Figure 3-2).
The proportion of *polQ* blunt joins consistent with the SD-MMEJ model is significantly decreased relative to wild-type ($p<0.01$, Fisher's Exact Test). This is consistent with a role for PolQ in SD-MMEJ. WT n=21; *polQ* n=25; *ku70* n=16; *rad51 ku70* n=10; *rad51* n=8; *rad51 lig4* n=11; *lig4* n=10.

We analyzed all repair product sequences with 1-3 bp insertions, apparent blunt joins, and junctional microhomologies to determine if they had repeat motifs consistent with SD-MMEJ within ±20 bp of the junction. A repair product was classified as consistent with repair by SD-MMEJ if the junction, plus at least one base on either side, was contained within half of a direct or inverted repeat of at least 4 bp. By these criteria, the expected frequencies of SD-

MMEJ consistent 1-3 bp insertions, apparent blunt joins, and microhomologies by random chance alone were estimated to be, respectively, 18, 43, and 54 percent (see Materials and Methods).

The observed frequencies of SD-MMEJ consistent repair junctions with 1-3 bp insertions, apparent blunt joins, and junctions with microhomology were, respectively, 80.0, 58.4, and 70.1 percent (Figure 3-9a). All of these are significantly greater than predicted by random chance ($p<0.01$; $p<0.05$; $p=0.05$; Fisher's Exact Test). Further supporting the hypothesis that repair products with and without net insertions could have been formed by the same mechanism, the proportion by genotype of SD-MMEJ consistent apparent blunt joins followed the same genotype specific patterns as net insertion frequency (Figure 3-9b; compare Figure 3-2). Importantly, relative to wild-type, significantly fewer apparent blunt joins and junctional microhomologies recovered from *polQ* flies were SD-MMEJ consistent (Figure 3-9b and data not shown).

Categorization of repair products according to SD-MMEJ parameters supports a single underlying mechanism

A significant majority of repair junctions from each of three apparently distinct structural classes (short insertions, apparent blunt joins, and junctional microhomologies) had structures consistent with repair by SD-MMEJ (Figure 3-9). If SD-MMEJ is in fact producing all three types of repair junctions, grouping of SD-MMEJ consistent repair products by primer repeat should result in groups of structurally similar repair products comprising more than one of the above junction types. Side by side comparison of SD-MMEJ consistent 1-3 bp insertions, apparent blunt joins, and microhomologies with the same SD-MMEJ primer repeat in fact demonstrates such similarities (Table 3-1, panels a-d).

Table 3-1. Categorizing SD-MMEJ consistent repair junctions by primer repeat reveals structural similarities among multiple repair junction types.

Representative examples of SD-MMEJ-consistent sequences with the specified primer repeats (top row) are shown. Superscripts cross-reference supplemental tables S3-S8.

original sequence	P_2 P_1 5'GTTAT CCCTAGCGGCCGCATAGGCC 5'GTTAT CCCTAGCGGCCGCATAGGCC
short insert	GTTAqtGGCCGCATAGGCCACTAGT S6-B ATTAqtGGCCGCATAGGCCACTAGT S6-F TGTaqtGGCCGCATAGGCCACTAGT S7-D TATaqtGGCCGCATAGGCCACTAGT S7-E CCCataGGCCGCATAGGCCACTAGT S7-C GTTATaGGCCGCATAGGCCACTAGT S5-B
apparent blunt join	GTTAT–GGCCGCATAGGCCACTAG S3-A CCTGT–GGCCGCATAGGCCACTAG S3-B TCGGT–GGCCGCATAGGCCACTAG S3-G TGTTA–GGCCGCATAGGCCACTAG S3-D
junctional mh	CCCT.G.GCCGCATAGGCCACTA S3-A

Table 3-1a. Primer repeat group I.
Corresponds to SD-MMEJ at a 4 bp direct and inverted repeat. Insertion junctions, apparent blunt joins, and junctional microhomologies are all represented. The deletion resulting from SD-MMEJ at this primer repeat corresponds to the cluster of deletion boundaries to the right of the DSB (Figure 3 a-c). This primer group contains the most frequently recovered junctional microhomology (Table S2, top row).

original sequence	P_1 P_2 GTACATTACCCTGTTAT CCCTAG
short insert	ACATTACCCTGTTAcCCGCA S5-A ACATTACCCTGTTAccCGGC S6-C ACATTACCCTGTTAcCGGCC S5-F
apparent blunt join	ACATTACCCTGTTA–CCCTAGC S3-C ACATTACCCTGTTA–CCGCATA S3-H ACATTACCCTGTTA–CCACTAG S3-N ACATTACCCTGTTA–CGGCCGC S3-J

Table 3-1b. Primer repeat group II.
Corresponds to loop-out SD-MMEJ at a 3 bp TTA direct repeat to the left of the DSB. Both short inserts and apparent blunt joins are represented. Use of this repeat for SD-MMEJ at this repeat is consistent with the absence of deletion boundaries 2 bp to the left of the top strand nick (Figure 2d-f).

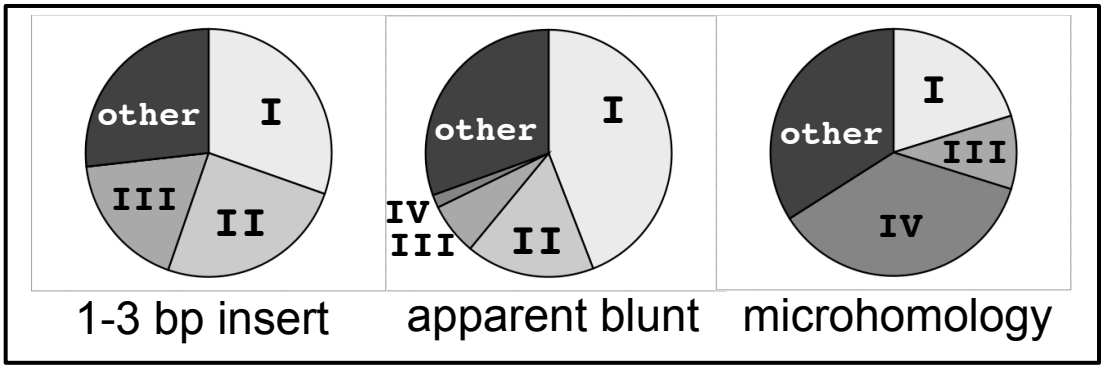
original sequence	P_2 P_1 5'GTTAT CCCTAGCGGCCGCATAGGCC
short insert	TACCCTatGCGGCCGCATAGGCCA S6-J CCCTGTatGCGGCCGCATAGGCCA S6-G TGTTATatGCGGCCGCATAGGCCA S6-H TTACCtatGCGGCCGCATAGGCCA S7-A
apparent blunt join	GTTAT–GCGGCCGCATAGGC S3-F CCTGT–GCGGCCGCATAGGC S3-W CTGTT–GCGGCCGCATAGGC S3-X ACATT–GCGGCCGCATAGGC S3-Y
junctional mh	ACCCT.G.CGGCCGCATAGG S4-E

Table 3-1c. Primer repeat group III.
Corresponds to SD-MMEJ at a 4 bp inverted repeat to the right of the DSB. Short inserts, apparent blunt joins, and junctional microhomologies are all represented.

original sequence	P_2 P_1 5'GTTAT CCCTAGCGGCCGCATAGGCC
short insert	TACCCTatGCGGCCGCATAGGCCA **S6-J** CCCTGTatGCGGCCGCATAGGCCA **S6-G** TGTTATatGCGGCCGCATAGGCCA **S6-H** TTACCtatGCGGCCGCATAGGCCA **S7-A**
apparent blunt join	GTTAT-GCGGCCGCATAGGC **S3-F** CCTGT-GCGGCCGCATAGGC **S3-W** CTGTT-GCGGCCGCATAGGC **S3-X** ACATT-GCGGCCGCATAGGC **S3-Y**
junctional mh	ACCCT.G.CGGCCGCATAGG **S4-E**

Table 3-1d.
Primer repeat group IV corresponds to a 3 bp inverted repeat to the left of the DSB. Junctional microhomologies and an apparent blunt join are represented.

Figure 3-10. Most SD-MMEJ consistent repair junctions accounted for by four primer repeats.



Quantification of the proportion of SD-MMEJ consistent 1-3 bp insertions, apparent blunt joins, and junctional microhomologies with primer repeats I-IV as described in Table 1, or a different primer repeat (other). Chi-square contingency tables used to compare the observed ratios of the five categories with the ratios derived from randomly generated blunt joins and microhomologies showed that, for both SD-MMEJ consistent blunt joins and microhomologies, the observed proportions of primer repeats in repair products was significantly different from random chance ($p<0.01$ and $p=0.01$, respectively). Ratios of primer repeats expected from random chance was estimated from 200 randomly generated SD-MMEJ consistent blunt joins and directly calculated by enumeration for microhomology junctions. Similar methods did not generate a sufficient number of random SD-MMEJ consistent 1-3 bp insertions to populate a contingency table.

Many of the most frequently used primer repeats are the same for multiple junction types. A small set of frequently used primer repeats accounts for a substantial portion of SD-MMEJ consistent repair products (Figure 3-10). Chi-square contingency table analysis comparing experimentally observed ratios of repair junctions in the five classes in Figure 3-10 with ratios obtained from randomly generated blunt joins and microhomologies indicates that the observed frequencies of specific SD-MMEJ primer repeats are unlikely due to random chance ($p<0.01$ and $p=0.01$, respectively). Together, these results suggest that SD-MMEJ may be responsible for a wide variety of superficially different end-joining repair junction structures.

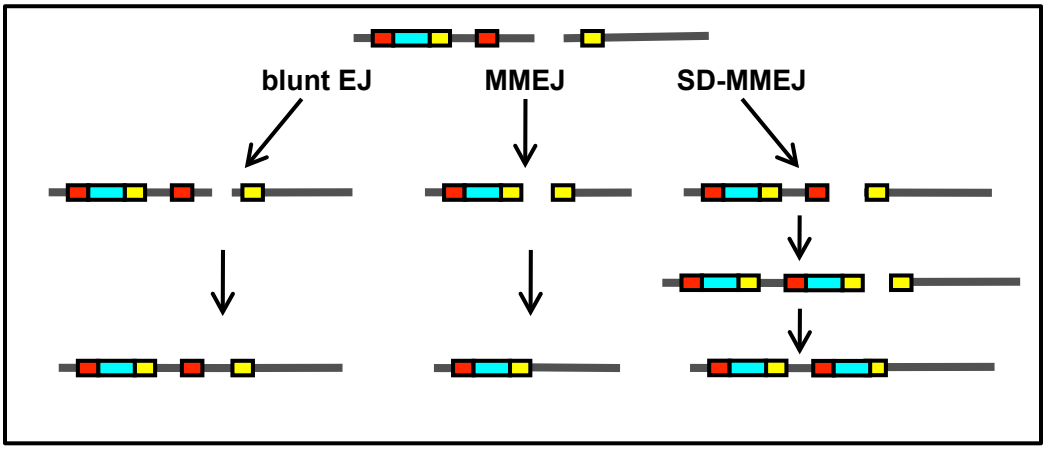
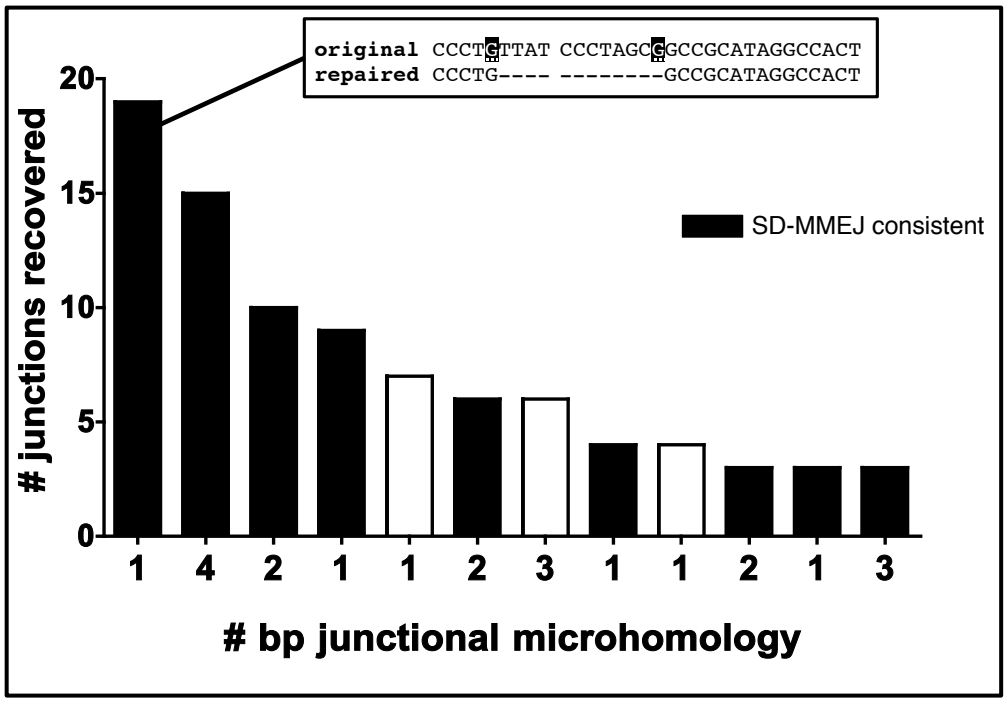


Figure 3-11. Multiple mechanisms for alternative end joining

Our results suggest that alternative end joining can proceed by a variety of mechanisms: direct ligation of non-complementary blunt ends (left), annealing at pre-existing microhomologies (center), or synthesis of *de novo* microhomologies via SD-MMEJ (right). Deletions during direct ligation of blunt ends are not influenced by repeated sequences. MMEJ at pre-existing microhomologies (yellow) involves deletion of one of the two repeated motifs as well as all sequence intervening between the repeated motifs. SD-MMEJ synthesizes *de novo* microhomologies via priming at a repeat situated entirely to one side of the DSB (red). Any sequence intervening between the primer and the template for the new microhomology (blue) is copied and inserted at the break site. Identically colored boxes represent direct repeats.

Figure 3-12. SD-MMEJ Explains Patterns In Frequency Of Junctional Microhomologies.



Histogram shows the relative frequency at which specific junctional microhomologies were observed, ordered by the number of ambigous nucleotides at the junction. Contrary to what would be expected if repeat length were the primary factor in determining which junctional microhomologies ought to be most frequently observed, no clear relationship is observed between length of a particular microhomology (x axis labels) and the frequency at which it was observed; a 1 bp junctional microhomology was recovered most frequently (see inset for alignment with original sequence). However, of the 13 junctional microhomologies recovered 3 or more times, 12 were SD-MMEJ consistent.

Additional Genetic Studies

mus205 (DNA polymerase zeta)

SD-MMEJ consistent junctions and insertion junctions were significantly reduced, but not eliminated in *mus308* mutants. This could be due to presence of maternal effect PolQ in the early embryo, or due to action of an additional DNA polymerase in SD-MMEJ. The few insertions observed in *polQ* mutants were distinct from those seen in PolQ competent flies in that they never exceeded 12 bp in length and appeared to have a less complex structure. Other preliminary results in the lab had hinted at a potential role for DNA polymerase zeta (PolZ) in end-joining repair of large chromosomal gaps (Daniel Kane, unpublished data). We therefore asked if SD-MMEJ might be impaired in flies with a null mutation in *mus205*, which encodes the rev3 catalytic subunit of DNA polymerase zeta (*polz* mutants).

The relative proportion of apparent blunt joins, microhomology junctions, and insertion junctions did not differ between junctions recovered from WT and *polz* mutant flies. 17 apparent blunt joins were recovered; of these, 5 (29%) were SD-MMEJ consistent. This is a statistically significant decrease relative to WT blunt joins ($P<0.05$, Fisher's Exact Test). Nine short *polz* insertion junctions were recovered; of these, 6 (67%) were SD-MMEJ consistent, representing a modest nonsignificant decrease ($P=0.21$, Fisher's Exact Test) relative to wild-type. 14 of the 20 total *polz* microhomology junctions (70%) observed were SD-MMEJ consistent, which is not different from wild-type. Together, these results are suggestive that DNA polymerase zeta may play a minor role in SD-MMEJ, especially in cases where nuclease activity has not uncovered a junctional microhomology. It will be interesting to carry out further studies in a *mus308 polz* double mutant background.

nbs1

A number of recent studies suggest that the MRN complex may play a role in microhomology-mediated alternative end joining (11, 12, 13). In a recent study (14), end-joining repair junctions from *nbs* mutant Drosophila were characterized by longer microhomologies than typically found in wild-type. Given evidence that the MRN complex may have end-bridging activity (15), we wondered if hypomorphic mutations in MRN complex components could result in increased SD-MMEJ.

25 apparent blunt joins were recovered from nbs flies; of these, 9 (36%) were SD-MMEJ consistent, a decrease relative to wild-type which approaches, but does not reach statistical significance (P=0.075, Fisher's Exact Test). 8 of 9 short insertions and 18 of 22 junctional microhomologies were SD-MMEJ consistent, which is not different from wild-type. We conclude that if the MRN protein plays a role in Drosophila SD-MMEJ, this role does not depend on presence of fully functional nbs1.

Discussion

In this study, we provide experimental evidence that end-joining junction structures typically considered diagnostic of fundamentally different repair mechanisms can be most parsimoniously explained by a single mechanism, which we term synthesis-dependent MMEJ (SD-MMEJ). Genetic evidence suggests that SD-MMEJ in Drosophila is a form of Ku and Lig4 independent alternative end joining that involves *mus308*, the Drosophila ortholog of vertebrate DNA polymerase theta (PolQ).

Historically, one of the most difficult to explain aspects of end-joining repair has been the diverse spectrum of repair events produced, and the variability of this spectrum both within and between studies. The SD-MMEJ model accounts for many apparently idiosyncratic patterns observed in repair junction sequences. For example, the frequency of particular junctional microhomologies in this study did not correlate with microhomology length or distance from the DSB, but correlated strongly with whether or not the resulting repair product was SD-MMEJ consistent (Figure 3-12).

The SD-MMEJ model suggested by our results provides a fresh perspective on the interplay of sequence context and genetic background during end-joining repair. Previous evidence suggests that sequence context can modulate the effect of C-NHEJ deficiency in potentially complex ways. For example, studies comparing different class switch recombination junctions from individual C-NHEJ deficient patients have found increased microhomology at some junctions, but not others (Reviewed in (16)).

End joining and homologous recombination (HR) repair have usually been considered to be entirely distinct pathways, rendering the observed increase in insertions in *rad51* mutants quite surprising. When present in a single copy, as in this study, the *[Iw]* DSB can be repaired

by gene conversion, although at a low frequency (<10% observed progeny). One possible explanation is that in the absence of HR, DSBs that would be repaired by gene conversion are redirected to an alt-EJ pathway. This would be consistent with reports that nucleases associated with HR participate in creation of single stranded 3' DNA tails during alt-EJ (17).

Previous reports of apparently templated insertions

Apparently templated insertions, or "T-nucleotides" are reported in a wide variety of systems, from plants to vertebrates. The frequency at which they are reported varies from species to species, possibly indicating species-specific differences in DNA repair or differences in the type of DSB substrates investigated in different species. Correlation of the observation that T-nucleotides are more frequently reported in Drosophila (1, 5, 10, 18) than in vertebrate-derived systems with genome data indicating that no orthologs of the NHEJ-associated X family polymerases are present in Drosophila (19) supports the existence of species-specific differences in the type and purpose of DNA synthesis during end-joining repair.

Very short insertions and/or sequence capture are observed at molecularly defined DSBs in vertebrate-derived systems. There have been comparatively few reports of T-nucleotides in vertebrate systems, although exceptions do exist, for example (20) and (21), which found complex templated insertions strikingly similar to those observed in our study at inaccurate I-*Sce*I repair junctions in the U937 lymphoblastoid cell line. Similar apparently templated insertions were also found in a bioinformatic survey of cancer genomes (22), although in this case the nature of the lesion precipitating the templated insertion event could not be determined. Increased insertions are occasionally reported as a phenotype in C-NHEJ deficient cells (23, 24). Since the length of a typical insertion in vertebrate-derived systems is often too short to be amenable to the type of statistical analysis usually employed to determine the probability that the

insertions are nonrandom, and template-independent DNA polymerase activity is known to be involved in vertebrate C-NHEJ, short insertions have generally been assumed to be nontemplated. It will be interesting to revisit these data to ask if short insertions found in vertebrate-derived systems are SD-MMEJ consistent. Initial sequence analysis of examples from the literature presented in Chapter 4 show that this possibility merits further consideration.

In vertebrate-derived systems, templated nucleotides are most frequently reported in situations associated with abnormal DNA repair - specifically, at translocation junctions and in cancer genomes. Interestingly, PolQ expression is increased in several cancers (25), suggesting the possibility that templated SD-MMEJ could contribute to genome instability in cancer. Templated insertions are frequently observed in vertebrate-derived systems is at translocation junctions in lymphomas (26, 27, 28), and casual inspection of these sequence data suggests that some of these templated insertions may be consistent with an SD-MMEJ mechanism.

If a process similar to the templated SD-MMEJ that we observe to be impaired in *mus308* mutants contributes to DSB repair in humans or other vertebrates, it may not necessarily be mediated primarily or exclusively by POLQ. Structural differences between the POLQ and *mus308* polymerase domains suggest that the functions of these two proteins in their respective organisms may not be completely analogous (29). Although POLQ is usually considered to be the human *mus308* ortholog, human DNA polymerase nu (POLN) and the HEL308 helicase also have substantial homology to the polymerase and helicase-like domains, respectively, of *mus308* (30, 31), suggesting the possibility that they may serve similar or redundant functions to the cognate POLQ domains. POLN may, in fact, be more closely related than the POLQ polymerase domain to the *mus308* polymerase doman at the sequence level in that the polymerase domain of

human POLQ possesses three "insertions," two of which are required for bypass of thymine glycols and abasic sites, which are not found in *mus308* (29).

Overall, the roles described to date for HEL308 and POLN in DSB repair support a role for these proteins primarily in homologous recombination rather than end-joining repair. As well as being similar to the *mus308* helicase-like domain, HEL308 also possesses significant homology to the Drosophila *mus301 (spn-C)* gene product, which, unlike *mus308*, has a demonstrated role in homologous recombination repair (32). It has recently been shown that human POLN participates in homologous recombination and, together with HEL308, mediates repair of interstrand crosslinks (33). Consistent with a role in homologous recombination, HEL308 demonstrates easily detectable helicase activity *in vitro* (31), whereas although the helicase-like domain of POLQ shows modest DNA-dependent ATPase activity (34), actual unwinding by the POLQ helicase domain has not been observed, suggesting that if it has helicase activity at all, it is relatively weak or non-processive. This would, however, be consistent with a role in destabilizing already transient secondary structures during SD-MMEJ. Therefore, we speculate that although HEL308 and POLN may be structurally homologous to *mus308*, and are on that basis certainly possible candidates for SD-MMEJ mediators in vertebrate cells, they are unlikely to be functional orthologs in terms of SD-MMEJ repair. This possibility, however, remains to be tested.

Notably, the SD-MMEJ model provides an alternative explanation for insertions appearing to be a single imperfect copy of flanking sequence, with implications for hypotheses about what DNA polymerases are likely to play a role in end-joining repair. Such insertions could instead be built up by iterative rounds of synthesis by an accurate non-processive DNA polymerase. The SD-MMEJ model suggests that if "imperfect copies" can be broken down into

overlapping repeats consistent with an SD-MMEJ mechanism, activity of accurate template dependent as well as error prone polymerases should be considered a possibility.

Antecedents and alternatives to the SD-MMEJ model

A number of models have been proposed to account for apparently templated mutations, and the SD-MMEJ model shares both shares features and contrasts with many of these models. One recurring general theme, which is shared by the SD-MMEJ model, is the hypothesis that transient secondary structures formed at very short repeated sequences can prime DNA synthesis. Models sharing this basic premise generally differ from the SD-MMEJ model in one or both of two ways. First, many similar models describe events originating at repeats of sufficient length to form secondary structures stabilized primarily by base pairing (35, 36, 37). In contrast with these models, some intermediate structures repeatedly suggested by our sequence analysis, such as snap-back priming between two adjacent nucleotides, are unlikely to form and persist spontaneously. We favor the idea that in SD-MMEJ, otherwise thermodynamically unfavorable secondary structures are stabilized by the binding of a DNA polymerase. The notion that such structures could be stabilized by protein binding is supported by crystallographic data showing that *M. tuberculosis* LigD with its ancillary polymerase domain can mediate a hairpin between two adjacent nucleotides (38).

Other models are similar to the SD-MMEJ model in positing an event mediated by extremely short repeats. Of particular note is an analysis of mutations at the maize *waxy* locus (39), which describes repeat structures essentially identical to SD-MMEJ "footprints." Also in this category are the "Tarzan" model proposed in (22) to explain templated insertions in cancer genomes, and the FoSTeS model based on the structure of nonrecurrent microindels (40). In fact, the structures of more complex SD-MMEJ events bear remarkable resemblance to miniature

FoSTeS events. These models, while being similar to the SD-MMEJ model in that they involve very short repeats, differ from SD-MMEJ in that they generally postulate that the that the initiating event is not a DSB, but a fault in DNA replication, arising either spontaneously or due to a fork-blocking lesion. Loop-out SD-MMEJ also superficially resembles slippage during DNA replication, but differs in that SD-MMEJ is not specific to highly repetitive DNA sequences (41).

The "Paired Priming" model suggested in (42) comes very close to the SD-MMEJ model in describing the relationship between DSB repair and mutation via templated synthesis at short repeats, but again postulates a round of DNA synthesis to fix the mutation. Some events reported in that series of papers are SD-MMEJ consistent (20).

Although the SD-MMEJ model is highly successful in accounting for a compelling proportion of the insertions observed in this study, there are nevertheless insertions that are not consistent with the SD-MMEJ model under the parameters used in this analysis. Depending on the structure of the specific junction in question, several alternative explanations are possible. Some junctions not consistent under the conditions specified in this analysis do show short repeats spanning the breakpoints. In such cases, it is possible that repair did proceed by an SD-MMEJ mechanism, but some iterations involved net synthesis of too few nucleotides to result in creation of a 4-bp repeat. Future computational work will include simulations to determine whether or not there are conditions under which shorter repeats representing potential synthesis tracts can be distinguished from fortuitously SD-MMEJ consistent short repeats.

Other insertion junctions appear to have no identifiable junction-spanning repeats on one or both sides. In these cases, a viable alternative to SD-MMEJ as an explanation for apparently templated insertions is a sequence-capture model such as those presented in (26, 43). Capture of

foreign DNA at DSBs is well documented in both Drosophila and vertebrate-derived systems (5). Two types of sequence-capture models can be envisioned. Short single or double stranded oligonucleotides produced by endonuclease activity during end processing could be inserted into the break. These insertions would appear to be "discontinuous deletions," and would be characterized by lack of template in the surrounding sequence and being present in only a single copy. Another possibility is that short direct or snap-back templated oligonucleotides produced by "run-off synthesis" (26) near the break could be inserted in the DSB. Insertions produced by run-off synthesis would start as single stranded DNA, which could potentially oligomerize, accounting for overlapping copies of surrounding sequence. These oligomers could be inserted into the DSB via short microhomologies or blunt end ligation.

The SD-MMEJ model is distinguished in part by its emphasis on a crucial role for coupled DNA polymerase and helicase activity. We speculate that in SD-MMEJ, transient base pairing interactions are stabilized by DNA polymerase binding. This addresses the apparently paradoxical observation that some microhomologies that appear to have been created by SD-MMEJ are not longer than pre-existing microhomologies available near the DSB. It seems likely that annealing at a short microhomology together with polymerase binding could provide better end stabilization than base pairing alone. This is supported by recent evidence from budding yeast that a DNA polymerase can have end-bridging activity and can mediate stabilizing interactions via base stacking (44). A role for polymerase-mediated end synapsis in vertebrate MMEJ would explain why some C-NHEJ deficient systems exhibit increased frequency of microhomologies apparently too short to mediate effective end synapsis via base pairing alone. It is also consistent with observation of longer microhomologies in *S. cerevisiae* MMEJ, because budding yeast does not have an ortholog of POLQ.

Our results support the hypothesis that alt-EJ may comprise multiple pathways or mechanisms with varying requirements for microhomology (Figure 3-11; Figure S5; Figure S6). Previous studies have documented differences between Ku and Lig4 independent alt-EJ (45, 46, 47). Our SD-MMEJ model reveals striking differences between Ku and Lig4 independent alt-EJ. In *ku70* mutants, approximately half of the repair junctions without insertions or microhomology did not fit the SD-MMEJ model and are likely true blunt joins. This suggests that in *ku70* mutants, other proteins, such as the MRN complex, may tether break ends. However, nearly all of the rare apparent blunt joins recovered from *lig4* mutants were SD-MMEJ consistent (Figure 3-9b). We speculate that this may reflect a need for especially stable break end synapsis to maintain the ends in proximity during recruitment of an alternative DNA ligase.

In sum, our results contribute to the emerging idea that end-joining repair is best conceptualized not as a set of distinct pathways, but as a flexible network of interacting repair strategies (48, 49). Such flexibility could encompass the ability to take advantage of local sequence features to facilitate repair. Thus, characterization of sequence specific effects will likely provide important new insights into the fundamental nature of end-joining repair.

Materials And Methods

Fly stocks and crosses

Flies were maintained according to standard methods on cornmeal-agar medium in a 25 degree incubator on a 12 hour light-dark cycle.

The *[Iw]7* construct (4) was used in all experiments. The *[Iw]7* construct is a *P*-element integrated on chromosome 2 at 53F8 containing the 18 bp I-*Sce*I site (not normally present in the Drosophila genome) and a mini-*white* gene 733 bp downstream (Figure 3-1). Location of the integrated *[Iw]7* construct was confirmed by inverse PCR. Alignments are based on our sequencing data using the PE5' and 3jnB primers (see below).

The I-*Sce*I sources used were either the *UIE[I-SceI]2R* (50) expression construct on the right arm of chromosome 2, which is under the control of a ubiquitin promoter, or the *hsp70*[I-*Sce*I]1A construct (4) integrated on chromosome 3, which is under the control of a constitutive *hsp70* promoter fragment.

Mutant alleles used in this study were $lig4^{169a}$ (1), $spnA^{057}$ (51), $spnA^{093}$ (51), $ku70^{7B2.2}$ (2), $ku70^{ex8}$ (2), $mus308^{2003}$ (52), $mus308^{Sz\text{-}2\ 9}$ (53) nbs^{1}, nbs^{2}, nbs^{SM9} (14), and $mus205^{DK3B}$ (Daniel Kane, personal communication). Transheterozygotes were used whenever possible to avoid second site effects.

Male flies undergoing germline DSB induction were generated by mating females of appropriate genotype harboring the *[Iw]* construct to males of appropriate genotype harboring an I-*Sce*I expression construct. Germline repair events were recovered in the next generation by appropriate individual matings. Virtually all chromosomes that had been cut had undergone a repair event altering or eliminating the I-*Sce*I cut site. Thus, although an indeterminate number of accurate repair events may have occurred prior to inaccurate repair, accurate repair is not

quantified in this assay. Progeny that inherited a chromosome repaired by end joining were identified by having pigmented eyes. Since the *[Iw]* construct was present on only one chromosome, DSB repair by gene conversion or end joining with a deletion to the right >700 bp results in the loss of *white* expression. On average, >90% of progeny flies inheriting *[Iw]* chromosomes from each mating had pigmented eyes or were the results of gene conversion (determined by PCR), indicating that end joining events with large deletions is not a major repair pathway in this system. Only independent events were sequenced.

Repair junction analysis

DNA was extracted from individual flies according to (54). PCR was carried out using the primers PE5' (GATAGCCGAAGCTTACCGAAGT) and jn3'B (GGACATTGACGCTATCGACCTA), which amplify a 1.3 kb region and are approximately equidistant from the I-*Sce*I site. A ~400 bp secondary PCR product was used as an internal PCR control. Amplified products were gel purified (GenScript) and sequenced using the PE5' primer.

Sequence analysis and modeling

Alignments were done with ClustalW software and/or manual inspection. Repeated sequences were identified via the EMBOSS PALINDROME software (55) or by a suffix-tree based repeat finding program. Computer programs written in our lab were used to estimate the frequency of SD-MMEJ consistent junctions by random chance alone. All simulations and calculations used the sequence interval shown in Figure 3-3. The probability of SD-MMEJ consistent blunt joins occurring by random chance at this DSB was estimated by a Monte Carlo simulation run three separate times on 10,000 randomly generated apparent blunt joins. The probability of SD-MMEJ consistent junctional microhomologies by random chance alone was calculated by enumeration of microhomology junctions. The probability of SD-MMEJ consistent indels with 1-3 bp insertions was estimated by analysis of 110 randomly generated junctions with 1-3 bp insertions. Randomly generated SD-MMEJ consistent repair junctions were used to estimate the expected frequency of SD-MMEJ consistent repair junctions in primer groups I-IV. Statistical analyses were carried out in Excel, SPSS, or GraphPad Prism.

Chapter 4

Investigating The Generality Of The SD-MMEJ Model: A Meta-Analysis

Introduction

The success of the SD-MMEJ model in explaining patterns observed in the spectrum of end-joining repair junctions at the *[Iw]7* double strand break (DSB) raises questions concerning the model's generality and predictive power. The critical question is whether evidence can be found for SD-MMEJ repair at other molecularly defined DSBs. Initial surveys of end-joining repair junction sequences published in the literature suggested that the SD-MMEJ model might indeed be applicable outside the context of the *[Iw]7* construct and, potentially, in other organisms.

Given that the SD-MMEJ model could have broader relevance, we wanted to better understand what specific sequence contexts and genetic backgrounds facilitate or inhibit SD-MMEJ. This information is also necessary to establish a biologically relevant basis for choosing the values that define an SD-MMEJ consistent repeat. These questions are best addressed by examining end-joining DSB repair junctions from a variety of different contexts in light of the SD-MMEJ model. Therefore, the logical first step was to carry out a meta-analysis of end-joining repair junction sequences produced in the course of previously published studies.

Ultimately, our primary goal is to refine the SD-MMEJ model so that a suite of SD-MMEJ reporter constructs can be designed and constructed to facilitate efficient screening in any desired number of different genetic backgrounds. We have developed and validated a novel Drosophila embryo plasmid injection assay that yields repair junctions very similar to those observed in the chromosomal I-*Sce*I DSB assay, and is far more efficient (for details, see Appendix A).

A concurrent, complementary goal is development of computational tools for studying SD-MMEJ. It became apparent in the course of the meta-analysis that although the SD-MMEJ

model is potentially a powerful tool for predicting the types of mutations likely to be generated by inaccurate end-joining repair, it is not feasible to carry out the sequence analysis necessary to generate these predictions by hand. A survey of currently available repeat finding and DNA modeling software tools indicated that software optimized for the type of fine structural analysis necessary to analyze SD-MMEJ events is not readily available. We therefore undertook development of the necessary computer programs.

This chapter details the first phase of the meta-analysis. The first section briefly discusses the issues inherent in choosing parameters for SD-MMEJ consistency and presents the results of simulations directed towards better understanding the factors that influence the expected frequency of fortuitous SD-MMEJ consistent repeats. The second section presents the results of analyzing sequences representing repair of a wide variety of specific DSBs for SD-MMEJ consistency.

Results (I): Modeling And Analysis Of Original Sequences

Computational analysis of SD-MMEJ consistency criteria

Previous analyses of apparently templated insertions have been constrained by probabilistic considerations (1, 2). Short runs of apparently templated synthesis are widely observed at end-joining repair junctions. However, it is in most sequence contexts difficult to determine statistically whether a short run of apparently templated synthesis is likely to be nonrandom. To minimize this difficulty, we would like to generate DSB substrates with sequences that are expected to generate few SD-MMEJ consistent junctions by random chance.

Since DNA sequences contain only four different bases, it seems reasonable to suspect that the rate of fortuitously SD-MMEJ consistent junctions observed by random chance could be quite high. We would therefore like to know how frequently we would expect to observe SD-

MMEJ consistent junctions if SD-MMEJ repair never occurred. Since the genetic requirements for SD-MMEJ repair are not known, it is not possible to determine this frequency experimentally. Thus, the expected frequency of fortuitously SD-MMEJ consistent junctions must be estimated from examination of the original sequence flanking the DSB.

The expected frequency of SD-MMEJ consistent repeats by random chance depends both on parameters determined by the investigator and by the properties of the DNA sequence itself. The parameters for which appropriate values must be chosen to estimate the expected frequency of fortuitously SD-MMEJ consistent junctions in the context of a specific DNA sequence include:

- The specific set of deletions to be applied to the original sequence
- The minimum acceptable length for an SD-MMEJ consistent repeat
- How much sequence flanking the junction to search for consistent repeats

Since it is not possible to determine what the deletion spectrum would be in the absence of SD-MMEJ repair, a straightforward approach to the first parameter is to simply choose the set of deletions to be used in the estimation to be similar to the deletion spectrum observed in the actual sequence data. This method is problematic, however, given that the SD-MMEJ model predicts that deletion boundaries will cluster at strong primer repeats. If SD-MMEJ is indeed a major contributor to repair, restricting the deletion spectrum considered in estimation of the frequency of consistent junctions by random chance to the range of deletions observed in the data may overestimate the probability of consistent junctions by random chance.

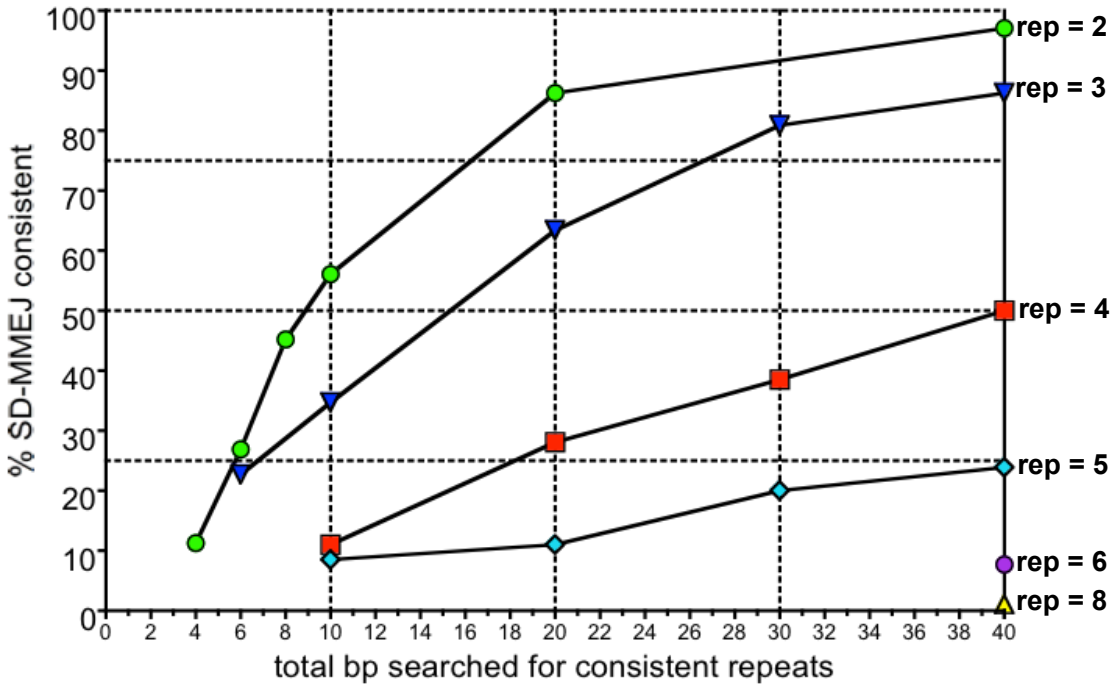
If the frequency of fortuitously SD-MMEJ consistent junctions depends strongly on the properties of the sequence near the DSB, the choice of which deletions to apply to the original sequence will dictate the values chosen for the other two parameters. It is therefore necessary to

determine the degree to which specific sequence context may cause the expected frequency of fortuitously SD-MMEJ consistent junctions to deviate from the expected overall frequency for completely random DNA sequences.

In the Chapter 3 study, simulations were carried out in the context of the *[Iw]7* DSB to estimate the frequency of fortuitously SD-MMEJ consistent repair junctions and choose appropriate minimum lengths for SD-MMEJ consistent repeats. To design efficient SD-MMEJ reporters, it is necessary to develop a more general understanding of the relationship between minimum consistent repeat length, sequence content, and the expected frequency of SD-MMEJ consistent repair junctions by random chance. We therefore developed simulations to investigate the expected frequency of SD-MMEJ consistent repair junctions in randomly generated sequences.

Figure 4-1 shows the results of a simulation in which large numbers of different randomly generated apparent blunt joins were tested for SD-MMEJ consistency for a range of minimum consistent repeat lengths and extents of sequence examined to either side of the repair junction. Surprisingly, under the conditions used to determine SD-MMEJ consistency in the study in chapter 3 (at least one 4 bp repeat within 20 bp on either side of the apparent point of ligation), 50% of randomly generated apparent blunt joins contained at least one SD-MMEJ consistent repeat. The estimated frequency of fortuitously SD-MMEJ consistent apparent blunt joins at the *[Iw]7* DSB (~43%) is thus somewhat lower than would be predicted for completely random DNA of unbiased base pair composition.

Figure 4-1. Effect of minimum SD-MMEJ consistent repeat length and total base pairs searched on the frequency of SD-MMEJ consistent randomly generated apparent blunt joins



Each data point represents the percentage of 1000 different randomly generated apparent blunt joins with unbiased base pair composition that were SD-MMEJ consistent with the indicated minimum consistent repeat lengths (rep). "bp searched" is the sum of base pairs searched to the left and right of the apparent point of ligation (ie, a value of 40 corresponds to examining sequence 20 bp to both sides of the junction). The parameters used to determine SD-MMEJ consistency in the NAR 2010 study correspond to rep=4, bp searched=40. Trend lines are drawn for each individual repeat length as labeled in the figure. For minimum consistent repeat lengths of 6 and 8, the simulation was run for only one value for base pairs searched due to the considerable period of time required for the simulation to run to completion for long minimum consistent repeats.

The discrepancy between the expected proportion of fortuitously SD-MMEJ consistent sequences in randomly generated apparent blunt joins and near the *[Iw]7* DSB led us to hypothesize that biased base pair composition can affect the expected frequency of SD-MMEJ consistent repeats. To test this, we tested sets of randomly generated apparent blunt joins with biased base pair composition for SD-MMEJ consistency. The results are presented in Table 4-1.

		base composition								
trial	**# runs**	**left of DSB**				**right of DSB**				**% consistent**
		A	T	G	C	A	T	G	C	
A	10,000	25	25	25	25	25	25	25	25	**50.03**
B	1,000	70	5	5	20	70	5	5	20	**51.5**
C	1,000	10	5	5	80	85	5	5	5	**59.0**
D	100	20	30	20	30	30	20	30	20	**47**
E	100	70	10	10	10	10	10	10	70	**49**
F	100	30	10	40	20	50	30	10	10	**57**
G	100	95	0	5	0	0	0	5	95	**2**
H	100	90	5	5	0	5	5	0	90	**32**
I	100	85	5	5	5	5	5	5	85	**52**

Table 4-1. Effect of base composition on probability of fortuitous SD-MMEJ consistency

Results of a simulation investigating the effect of DNA base composition on the frequency of SD-MMEJ consistency in randomly generated apparent blunt joins under the criteria used in Yu and McVey, NAR 2010 (minimum repeated motif length 4, maximum motif separation 15). Base compositions of the DNA to the left and right of the DSB were determined independently. No *a priori* knowledge of the original sequence was assumed.
Trial A shows that given unbiased base composition, 50% of randomly generated apparent blunt joins can be expected to be SD-MMEJ consistent under the NAR 2010 criteria. Subsequent runs show that substantially biased base composition may or may not have a large effect on the proportion of SD-MMEJ consistent blunt joins (compare trials C and I, D and F).

It is apparent that relationship between biased base composition and the expected frequency of fortuitously SD-MMEJ consistent apparent blunt joins is not straightforward. The overall impression yielded by the random simulation data is that biased base composition generally does not have a substantial effect on the probability that randomly generated apparent blunt joins will be SD-MMEJ until the bias becomes quite extreme (compare trials G, H, and I). Therefore, the difference between the expected frequency of SD-MMEJ consistent apparent blunt joins in the context of a completely random sequence versus the *[Iw]7* construct likely is largely due to the specific ordering of the bases relative to each other, not the relatively slight bias in base pair composition.

The results of this simulation demonstrate that it is indeed possible to manipulate base composition to create sequences with low to intermediate probability of fortuitous SD-MMEJ

consistency, but that the base composition of sequences generated using a strategy of base pair content manipulation alone to ensure a low frequency of fortuitously SD-MMEJ consistent junctions is unlikely to be representative of most naturally occurring DNA sequences. Future investigations into developing a strategy for generating SD-MMEJ reporter substrates will therefore focus on identifying the relationship between the ordering of specific bases relative to each other and the expected frequency of fortuitously SD-MMEJ consistent junctions.

These results also have important implications for choice of method to estimate the expected frequency of fortuitously SD-MMEJ consistent junctions in the context of a particular sequence. If, as these results indicate, base pair composition alone plays a minor role in determining the expected frequency of fortuitously SD-MMEJ consistent junctions, it is not appropriate to estimate the expected frequency of SD-MMEJ consistent repeats in specific sequence contexts by scrambling the original sequence and applying the observed spectrum of deletion lengths.

In light of these results, we decided to estimate the expected frequency of fortuitously SD-MMEJ consistent apparent blunt joins and microhomology junctions by examining the frequency of SD-MMEJ consistent blunt joins and microhomology junctions in the interval containing most of the deletion boundaries, since although this would likely lead to an overestimate of the expected frequency of SD-MMEJ consistent junctions, the results would nevertheless probably be closer to the true value than results obtained from applying the observed deletion spectrum to randomly scrambled flanking sequence.

We also carried out simulations to better understand how the length of pre-existing microhomologies near the break site might affect the frequency of fortuitously SD-MMEJ consistent repair junctions. The analysis of the SD-MMEJ mechanism presented in Chapter 3

indicates that fortuitously SD-MMEJ consistent microhomology junctions are likely to be more common than fortuitously SD-MMEJ consistent apparent blunt joins, and that this frequency ought to increase with increasing length of junctional microhomology. The simulation results presented in Table 4-2 bear out these predictions, but also reveal the interesting trend that the rate of increase in the frequency of randomly SD-MMEJ consistent microhomology junctions is not constant with microhomology length increases.

Trial	# runs	mh length	% consistent
A	1000	1	60.0
B	1000	2	70.5
C	500	3	76.0

Table 4-2. Effect of microhomology length on the probability of SD-MMEJ consistency

Results of a simulation investigating the effect of junctional microhomology length on the frequency of SD-MMEJ consistency in randomly generated sequences with microhomology of a given length under the criteria used in the study in Chapter 3 (minimum repeated motif length 4, maximum motif separation 15). All trials used unbiased base compositions. No *a priori* knowledge of the original sequence was assumed. Details of the computer programs used to generate these data are given in the Materials and Methods for this chapter.

The results of simulations with randomly generated sequences confirmed that it is important to determine expected frequencies of fortuitously SD-MMEJ consistent end-joining junctions in the sequence context of the specific DSB under consideration. The results of simulations to determine the expected frequencies of randomly SD-MMEJ consistent apparent blunt and microhomology junctions at the DSBs we examined in this study appear in Table 4-3.

	% consistent	
System	**blunt**	**mh**
[Iw]7 (I-*Sce*I)	43	54
Rr3 (I-*Sce*I)	52	40
FokI ZFN (*yellow* locus)	42	57
FokI ZFN (*rosy* locus)	51	63

Table 4-3. Estimates of expected frequency of fortuitously SD-MMEJ consistent apparent blunt joins and microhomology junctions for the DSB substrates used in this meta-analysis.

Frequencies for *[Iw]7* calculated as described in NAR 2010. Frequencies for other systems estimated from analysis of 100 randomly generated junctions in the interval encompassing >90% of observed deletions. Details of the structures of these DSBs are given in the following section. ZFN: zinc finger nuclease.

Properties Of The Double Strand Break Substrates

In this phase of the meta-analysis, we wished to focus specifically on the relationship between DSB end sequence and structure and repair outcome, and to avoid confounding factors due to interspecies differences in DNA repair. Therefore, we carried out direct comparisons only between data sets generated in Drosophila. We selected for study two different chromosomally integrated I-*Sce*I DSB reporter constructs (the *[Iw]7* construct and the Rr3 construct) (3) and two zinc finger nuclease DSBs at distinct chromosomal loci (*yellow* and *rosy*) (4, 5, 6), as well as repair junctions produced by end-joining repair of a large chromosomal gap subsequent to *P* element excision (7). All DSB induction and repair events in all systems take place in the male pre-meiotic germline.

I-*Sce*I DSBs: the *[Iw]7* and Rr3 constructs

[Iw]7	AGCAGCAGCCGAATTCGGTACATTACCCTG***TTAT*** CCCTAGCGGCCGCATAGGCCACTAGTGGAT TCGTCGTCGGCTTAAGCCATGTAATGGGAC ***AATA***GGGATCGCCGGCGTATCCGGTGATCACCTA
Rr3	ACGGTGTTGTGGCGGCCGTATATTACCCTG***TTAT*** CCCTAGCCGGGGAAGGACAGCTTCTTGTAG TGCCACAACACCGCCGGCATATAATGGGAC ***AATA***GGGATCGGCCCCTTCCTGTCGAAGAACATC

Table 4-4. Structures of molecularly defined I-*Sce*I DSBs

Overhanging single stranded ends are shown in bold italic.

The Rr3 reporter, like the *[Iw]7* reporter, is a chromosomally integrated *P* element containing a single I-*Sce*I target site (Table 4-4). Thus, although the sequence immediately adjacent to the DSB is identical between the two reporters, the sequences flanking the I-*Sce*I site differ substantially in their potential to form particular transient secondary structures (Table 4-5).

```
Iw7 TAGAGCAGCAGCCGAATTCGGTACATTACCCTGTTAT CCCTAGCGGCCGCATAGGCCACTAGTGGATCTGG
Rr3 TTCACGGTGTTGTGGCGGCCGTATATTACCCTGTTAT CCCTAGCCGGGGAAGGACAGCTTCTTGTAGTCGG
```

Table 4-5. Alignment of the sequences of the *[Iw]7* and Rr3 DSB substrates.

Matching bases are shaded gray. The GGCC repeat associated with a large fraction of the SD-MMEJ consistent repair events at the *[Iw]7* DSB is underlined. Note that this repeat is absent in the Rr3 sequence. The most substantial repeat to the right of the Rr3 DSB, a CCC/GGG inverted repeat, is underlined. Also notable is that due to the first sequence difference to the left of the DSB (C-T mismatch, red), junctions with the CCCT microhomology are SD-MMEJ consistent in the *[Iw]7* system, but not consistent in the Rr3 system.

The SD-MMEJ model predicts that although these two DSBs have identical end structures and sequences, they should yield distinct repair junction sequence spectra. Furthermore, the specific junction sequences recovered should be consistent with synthesis primed by the specific repeats available near the DSB. Of particular note is that the GGCC repeat to the right of the DSB that appears to mediate a substantial proportion of SD-MMEJ consistent repair events at the *[Iw]7* DSB is absent in the Rr3 construct (Table 4-5). We predict that a substantial number of SD-

MMEJ consistent events in the Rr3 construct will likely be associated with the CCC/GGG inverted repeat to the right of the DSB.

All Rr3 sequences examined were obtained from a wild-type genetic background. *[Iw]7* sequences were obtained from a variety of genetic backgrounds, including wild-type, C-NHEJ deficient, and homologous recombination deficient. To facilitate comparisons with the other data sets available, *[Iw]7* sequences from Lig4 competent and deficient backgrounds were divided into two groups and analyzed separately.

Zinc finger nuclease DSBs

ZFN (*yellow*)	GCGCCAACAGTATTACCACTGCCTACCGCA ***TTAA***AGTGGATGAGTGTGGTCGGCTGTGGG CGCGGTTGTCATAATGGTGACGGATGGCGT***AATT*** TCACCTACTCACACCAGCCGACACCC
ZFN (*rosy*)	ATCCCAATGCCCGCACCTATAGCTACTACA ***CGAA***TGGCGTGGGAGTCACTGTGGTAGAGA TAGGGTTACGGGCGTGGATATCGATGATGT***GCTT*** ACCGCACCCTCAGTGACACCATCTCT

Table 4-6. Sequences of the zinc finger nuclease (ZFN) DSB substrates.

Single stranded overhangs are in bold italic. Terminal base pairs of the recessed 3' ends are underlined. Note opportunities for slipped mispairing between the terminal 3' base and the single stranded overhang present at the *yellow* DSB are absent at the *rosy* DSB.

To ask whether efficient SD-MMEJ requires the presence of single stranded 3' overhangs at the DSB, as well as gain more information about the effects of local sequence context on the qualitative outcomes of repair, we analyzed two distinct site-specific chromosomal DSBs with 4 bp 5' overhangs produced by engineered zinc finger nucleases (ZFNs) (8). Zinc finger motifs recognize specific triplet DNA sequences with high specificity and can be assembled with catalytic domains to target the enzyme activity of choice to a specific DNA sequence. The ZFNs that produced the DSBs comprise triplets of zinc fingers fused to the *FokI* nuclease domain, targeted to sites within the Drosophila *yellow* and *rosy* loci, which are nonessential genes with easily recognizable mutant phenotypes.

DSBs with 5' overhangs are different from DSBs with 3' overhangs in that the recessed 3' end allows the single-stranded overhang to be filled in by a template-dependent DNA polymerase. This could result in both ZFN DSBs being better substrates for blunt-ended ligation than the I-*Sce*I DSBs, and poorer substrates for SD-MMEJ.

The sequences around the *yellow* and *rosy* DSBs differ in a number of ways that could affect the way DNA synthesis proceeds during repair. For example, one of the terminal 3' base pairs at the *yellow* DSB, but not the *rosy* DSB, is complementary to the adjacent base in the overhang, potentially increasing the likelihood of addition of a few bases to the 3' end by slipped

mispairing with the single stranded end. It is therefore possible that mispairing between the terminal 3' base pair and the single stranded 5' ends could result in very short templated additions to the 3' end. Such additions would probably not be long enough to form SD-MMEJ consistent repeats. Because the ends are AT-rich, it is easy to speculate that if the slipped mispairing of the 3' end occurs once, it could "stutter" multiple times. The end structure leads to speculation that there may be a substantial number of short AT-rich inserts that will not be SD-MMEJ consistent. The *yellow* DSB is also notable for a lack of substantial perfect microhomologies immediately distal to the DSB.

All repair junctions at the *yellow* DSB were obtained from a wild-type genetic background. Repair junctions at the *rosy* DSB were obtained both from wild-type and from genetic backgrounds deficient in critical components for C-NHEJ (*lig4*), homologous recombination repair (*spn-A* or *okra*, which encode Drosophila orthologs of Rad51 and Rad54, respectively), or both. This analysis, as well as the original study that generated these sequences, considers the sequences derived from Lig4 competent and deficient backgrounds separately. Data from the Chapter 3 study have been accordingly reclassified to allow comparison of the effects of Lig4 deficiency in different sequence contexts.

Patterns of deletion boundaries at the Rr3, *yellow* and *rosy* DSBs are consistent with SD-MMEJ repair

A striking result obtained from the *[Iw]7* study was the asymmetry of deletions to the left and to the right of the DSB (Figure 3-3). Although this was interpreted as evidence of resection to SD-MMEJ consistent repeats, an alternative explanation is differential protection of the two ends of an I-*Sce*I DSB from nuclease attack (9). If this is the case, the deletion profiles from the two systems should be similar. However, if SD-MMEJ repair is influencing the resection events, the larger deletions at the *[Iw]7* and Rr3 events should be on opposite sides of the break: the more substantial deletions should be to the left of the DSB in the Rr3 system. Examination of the data (Figure 4-2) showed that the deletions to the left were comparatively shorter at the Rr3 junctions in a manner consistent with SD-MMEJ repair at the CCC/GGG repeat rather than nuclease protection.

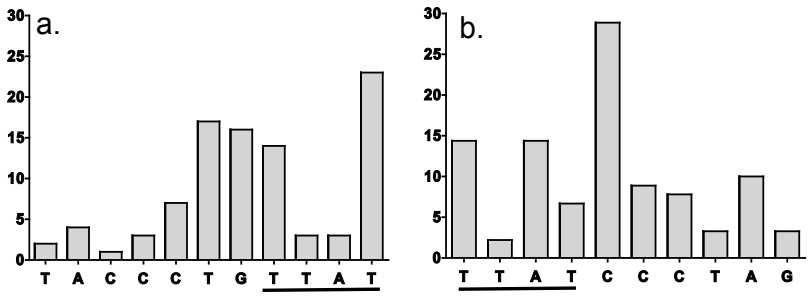


Figure 4-2. Deletion boundary frequency histograms for deletions to the left and right of the Rr3 DSB

Bars indicate the remaining nucleotide nearest the DSB. The 4 bp 3' overhang is shown on both sides of the break (underlined). The intervals shown encompass >90% of total deletions.

a. Deletion boundaries to the left of the DSB. Comparison with Figure 3-3 shows differences consistent with the predictions of the SD-MMEJ model: the rapid falling-off in deletion boundary frequency observed to the left of the DSB in *[lw]7* (Figure 3-3) is attenuated in Rr3.

b. Deletion boundaries to the right of the DSB. A majority of deletions are much shorter than at the *[lw]7* DSB (compare Figure 3-3). The clustering of deletion boundaries at the CCC is consistent with repair by snap-back SD-MMEJ at the CCC/GGG repeat shown in Table 4-5.

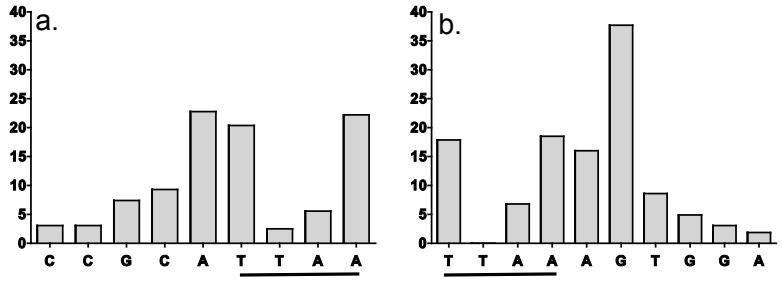


Figure 4-3. Deletion boundary frequency histogram for the *yellow* DSB

Bars indicate the remaining nucleotide nearest the DSB. The 4 bp 5' overhang is shown on both sides of the break (underlined). The intervals shown encompass >90% of total deletions.

a. Deletion boundaries to the left of the DSB.

b. Deletion boundaries to the right of the DSB are consistent with partial or complete fill-in of the 5' overhangs or SD-MMEJ at the GTGG repeat to the right of the DSB.

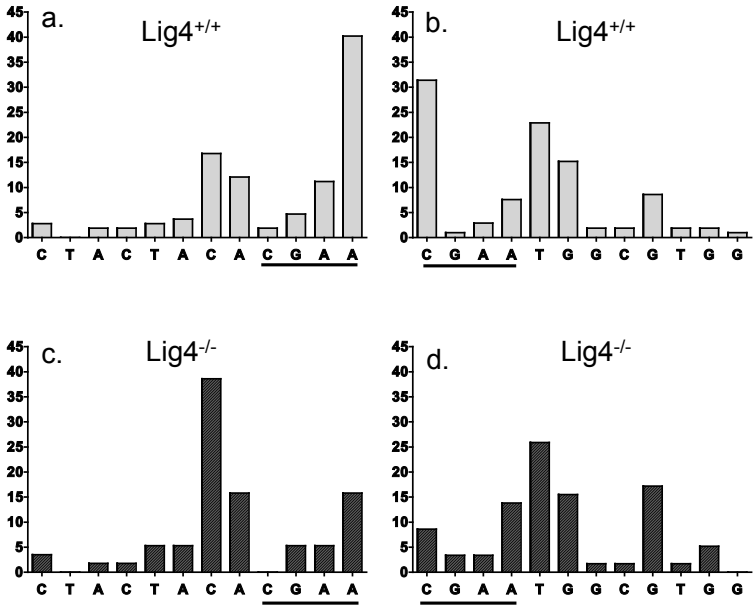


Figure 4-4. Deletion boundary frequency histograms by genotype for the *rosy* DSB

Bars indicate the remaining nucleotide nearest the DSB. The 4 bp 5' overhang is shown on both sides of the break (underlined). The intervals shown encompass >90% of total deletions.

a,b: Deletion boundaries to the left (a) and right (b) of the *rosy* DSB in a Lig4 competent genetic background (light gray; n=108)

c,d: Deletion boundaries to the left (c) and right (c) of the *rosy* DSB in a Lig4 deficient genetic background (dark hatched gray; n=58).

C-NHEJ deficiency shifts the deletion boundaries in a manner consistent with SD-MMEJ repair using the TAC direct repeat to the left of the DSB (compare a and c) and the GTGG repeat to the right of the DSB (compare b and d).

Comparison of the *yellow* and *rosy* zinc finger DSB repair junctions (Figures 4-3 and 4-4) showed that the distributions of deletion boundaries varied with sequence and genotype in a way consistent with SD-MMEJ repair. At the *yellow,* DSB, the distribution of deletion boundaries is consistent with repair synthesis occurring at the AT-rich ends and via the GTGG repeat to the right of the DSB. At the *rosy* DSB, deletions from a Lig4-competent genetic background show evidence of repair via filling in the left and right 5' overhangs, with evidence of minor contribution from SD-MMEJ at the TGG direct repeat to the right of the DSB. In the Lig4-deficient background, a dramatic shift in deletion boundaries is observed consistent with SD-MMEJ at the TAC direct repeat to the left of the DSB and increased reliance on SD-MMEJ at the GT-rich repeats to the right of the DSB.

Gap repair subsequent to transposon (*P* element) excision

To ask if SD-MMEJ might operate during end-joining repair of a large gap as well as at a complementary ended DSB, we examined end-joining repair junctions consequent to *P* element transposon excision in Drosophila (7). End-joining at this DSB in homologous recombination proficient flies is typically carried out by a hybrid HR/EJ pathway in which some amount of repair synthesis is carried out before HR is aborted and repair is finished by end joining (10). Unlike the repair events examined at the other DSBs in this study, it involves processive nuclease, helicase, and DNA polymerase activity. It is therefore possible that the typical complement of repair factors recruited to this DSB could be distinct from that recruited to a restriction endonuclease DSB. The structure of the DNA ends are also almost certainly long 3' single stranded tails, rather than the blunt ended deletions or short overhangs presupposed by many models of end-joining without prior attempts at homologous recombination. Therefore, it is interesting to ask if evidence for SD-MMEJ appears at this DSB as well.

The repair junctions produced by repair of this DSB differ from the others in that large and variable amounts of resection take place in the course of repair. Therefore, a specific sequence interval for calculating the probability of fortuitous SD-MMEJ consistent repair junctions cannot be meaningfully identified. Therefore, in calculating the probability of SD-MMEJ consistent junctions by random chance, we used values produced by the random simulation in Table 4-1. The lack of a consistent sequence context for potential SD-MMEJ events also limits potential for commentary on the relationship between sequence context and SD-MMEJ repair; hence, in the following sections, notes on the specific patterns of SD-MMEJ consistency of the transposon excision DSBs will be relatively brief.

The *P{w^a}* repair junctions in this study were obtained in either wild-type or *blm* (Bloom helicase) deficient backgrounds.

Results (II): Sequence Analysis Of Repair Junctions

Before considering the detailed results of the sequence analysis, we present summary data showing that, consistent with the predictions of the SD-MMEJ model, the relative proportions of different junction types differ dramatically in a sequence and genotype specific manner, independently of the end structure of the DSB. We then present three sections reporting the overall proportion of SD-MMEJ consistent insertions, microhomology junctions, and blunt joins for each DSB considered, and discussing the fine structure of the junctions in light of the predictions of the SD-MMEJ model. Overall, the SD-MMEJ model performs well in predicting both the overall proportion of consistent junctions, as well as the sequence details of the particular consistent junctions observed. We also discuss patterns apparent in non SD-MMEJ consistent junctions, and suggest alternative explanations for their origins.

Substantial variation in relative proportions of apparent blunt joins, microhomology junctions, and insertion junctions at different double strand breaks

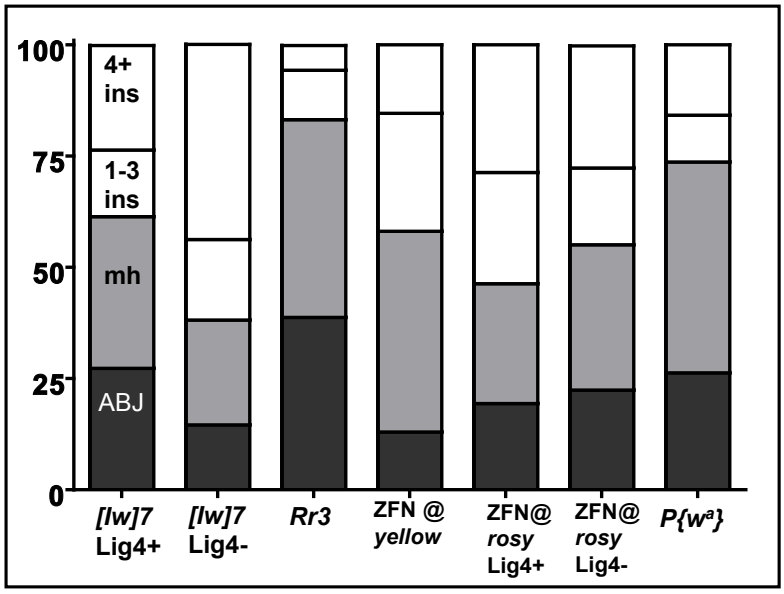


Figure 4-5. Relative proportions of different junction types by DSB substrate

Total proportions of apparent blunt joins (ABJs; dark gray), microhomology junctions (mh; light gray), 1-3 bp insertions (white), and 4+ bp insertions (white) observed at different DSBs. Descriptions of the different DSBs are given in the text. Total n and genotypes of data sets are as follows: *[Iw]7* (Lig4+/+):n=293. *[Iw]7* (Lig4-/-): n=144. Rr3 (WT): n=90. ZFN@*yellow* (WT): n=162. ZFN@*rosy* (Lig4+/+):n=108 . ZFN@*rosy*(Lig4-/-): n=58. *P{w^a}*(WT and Blm-/-) n=38. Sequences are as follows:

[Iw]7 **(I-*Sce*I)**	AGCAGCAGCCGAATTCGGTACATTACCCTG***TTAT*** CCCTAGCGGCCGCATAGGCCACTAGT TCGTCGTCGGCTTAAGCCATGTAATGGGAC ***AATA***GGGATCGCCGGCGTATCCGGTGATCA
Rr3 **(I-*Sce*I)**	ACGGTGTTGTGGCGGCCGTATATTACCCTG***TTAT*** CCCTAGCCGGGGAAGGACAGCTTCTT TGCCACAACACCGCCGGCATATAATGGGAC ***AATA***GGGATCGGCCCCTTCCTGTCGAAGAA
yellow **(ZFN)**	GCGCCAACAGTATTACCACTGCCTACCGCA ***TTAA***AGTGGATGAGTGTGGTCGGCTGTGGG CGCGGTTGTCATAATGGTGACGGATGGCGT***AATT*** TCACCTACTCACACCAGCCGACACCC
rosy **(ZFN)**	ATCCCAATGCCCGCACCTATAGCTACTACA ***CGAA***TGGCGTGGGAGTCACTGTGGTAGAGA TAGGGTTACGGGCGTGGATATCGATGATGT***GCTT*** ACCGCACCCTCAGTGACACCATCTCT

Even without considering the issue of SD-MMEJ, there is substantial variability in the relative proportions of different junction structures produced by end-joining repair at different DSBs. Comparison of the proportions of apparent blunt joins, microhomology junctions, and insertion junctions (Figure 4-5) between different systems suggests that not only are DSBs with

identical ends repaired differently, but that the effect of genetic background on repair outcome may depend heavily on idiosyncratic characteristics of the particular break being repaired.

As predicted by the SD-MMEJ model, although the structure and sequence at the ends of the *[Iw]7* and Rr3 DSBs are identical, different repair product spectra are observed, with substantially fewer insertion junctions in the Rr3 system, even when genetic background is taken into account.

Even more interesting, however, is the comparison between the effect of Lig4 status on repair junction spectra in the contexts of the *[Iw]7* and *rosy* DSBs. Although insertions at the *[Iw]7* DSB are substantially increased in *lig4* mutants, no substantial change in insertion frequency is observed at the *rosy* DSB in *lig4* mutants. This is consistent with the hypothesis suggested by the study in Chapter 3 that a complex interaction exists between genetic background and sequence context during end-joining repair.

Although the absolute percentage of apparent blunt joins and microhomology junctions varies from system to system, the overall general trend appears to be that there are proportionately more microhomology junctions than apparent blunt joins recovered. To ask if this might be due to random chance, we ran a simulation in which randomly generated deletions at the given DSB resulted in an apparent blunt join or a microhomology junction. The results are given in Table 4-7. Contrary to the observed trend, the simulation predicts that in all systems assayed if microhomology is not influencing repair, proportionately more apparent blunt joins should be observed than microhomology junctions. It also suggests that if microhomology length is not a factor (as was observed in the study in Chapter 3), opportunities for MMEJ repair are approximately equivalent in the systems tested.

System	%blunt	%mh
Rr3 (I-*Sce*I)	65	35
***FokI* ZFN (*yellow* locus)**	70	30
***FokI* ZFN (*rosy* locus)**	61	39

Table 4-7. Estimated proportions of junction types by random chance for DSB substrates

Relative percentage of blunt joins and microhomology junctions generated by generation of 10,000 random deletions at the specified DSB. To avoid confounding results due to mechanistic considerations (see chapter 1), the simulation was run as a blunt-ended deletion but tallied only the presence or absence of microhomology at a junction, not the length of the microhomology. Microhomologies of all lengths were considered.

The SD-MMEJ model suggests that most short insertions in most systems are templated from surrounding sequence

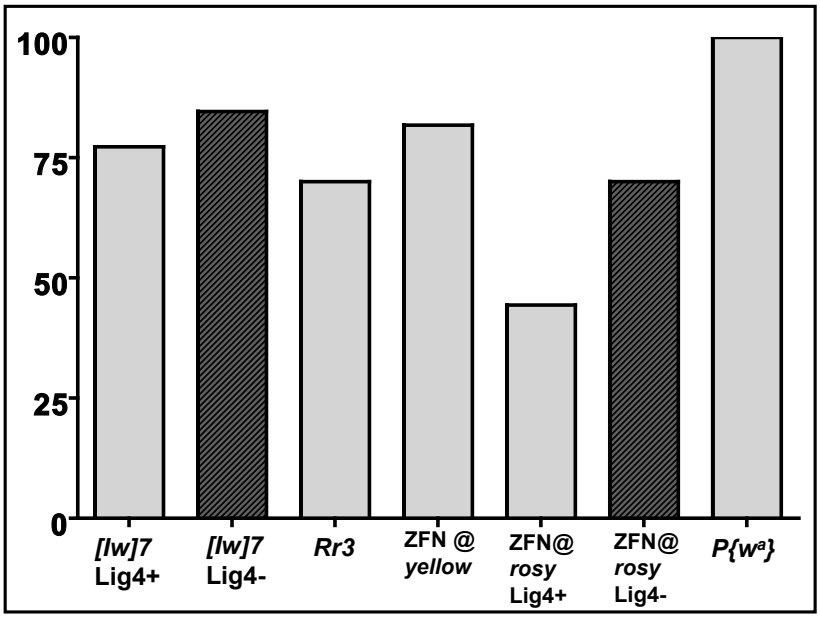


Figure 4-6. Proportion of SD-MMEJ consistent 1-3 bp insertions by DSB and Lig4 status

[lw]7 (Lig4+/+): n=44; *[lw]7* (Lig4-/-): n=26; Rr3 (WT): n=10; *yellow* (WT), n=44; *rosy* (Lig4+/+): n=27, *rosy* (Lig4-/-) n=10; *P{w^a}* (WT and Blm-/-): n=4. Data for Lig4 competent backgrounds is light gray; Lig4 deficient backgrounds are indicated by dark shaded gray.
Sequences are as follows:

[lw]7 **(I-*Sce*I)**	AGCAGCAGCCGAATTCGGTACATTACCCTG***TTAT*** CCCTAGCGGCCGCATAGGCCACTAGT TCGTCGTCGGCTTAAGCCATGTAATGGGAC ***AATA***GGGATCGCCGGCGTATCCGGTGATCA
Rr3 **(I-*Sce*I)**	ACGGTGTTGTGGCGGCCGTATATTACCCTG***TTAT*** CCCTAGCCGGGGAAGGACAGCTTCTT TGCCACAACACCGCCGGCATATAATGGGAC ***AATA***GGGATCGGCCCCTTCCTGTCGAAGAA
yellow **(ZFN)**	GCGCCAACAGTATTACCACTGCCTACCGCA ***TTAA***AGTGGATGAGTGTGGTCGGCTGTGGG CGCGGTTGTCATAATGGTGACGGATGGCGT***AATT*** TCACCTACTCACACCAGCCGACACCC
rosy **(ZFN)**	ATCCCAATGCCCGCACCTATAGCTACTACA ***CGAA***TGGCGTGGGAGTCACTGTGGTAGAGA TAGGGTTACGGGCGTGGATATCGATGATGT***GCTT*** ACCGCACCCTCAGTGACACCATCTCT

It has generally not been considered possible to determine whether short insertions of 1-3 bp at end-joining repair junctions are templated from surrounding sequence. However, at most DSBs observed, the SD-MMEJ model accounts for the origins of most insertions (Figure 4-6). It also extends the concept of a complex interaction between sequence context and genotype. In

the context of the *[Iw]7* DSB, the overall proportion of insertions increased in *lig4* backgrounds, but here we see that the proportion of consistent junctions is unchanged. Conversely, at the *rosy* DSB, a trend towards increased SD-MMEJ consistency of short insertions is observed, although it does not reach statistical significance (P=0.27, Fisher's Exact Test), likely due to the small number of short insertions at the *rosy* DSB (ten) recovered in the *lig4* background.

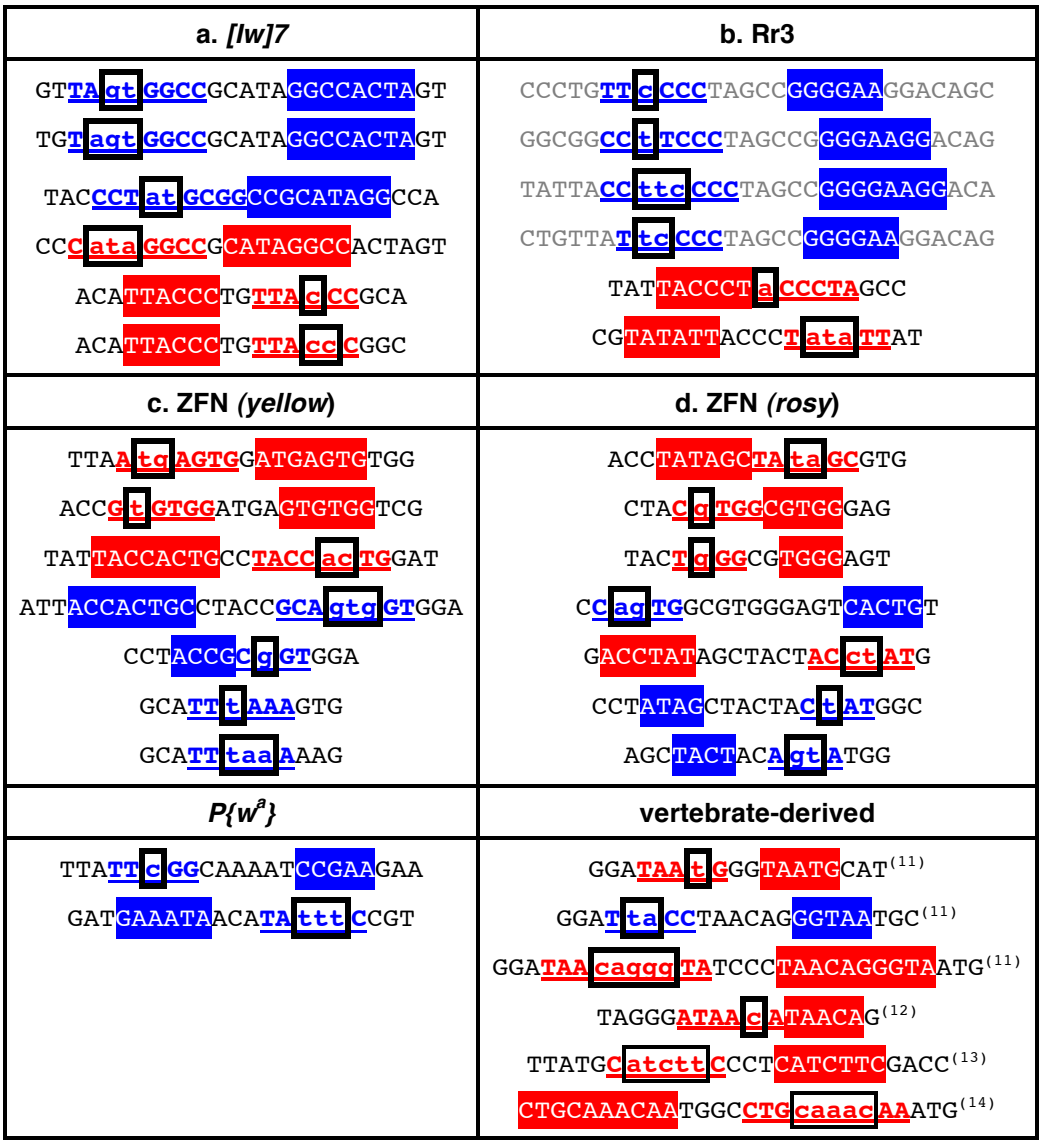


Table 4-8. Examples of representative patterns of SD-MMEJ consistent repeats at repair junctions with 1-3 bp insertions.

Superscripts associated with vertebrate-derived sequences reference the study in which the sequence was observed.

Short Insertions At [Iw]7 *and Rr3:*

Comparing spectra of representative SD-MMEJ consistent insertions observed at the *[Iw]7* and Rr3 DSBs (compare Table 4-8 a and b) supports the prediction that the sequences of the short insertions observed at each DSB should be consistent with synthesis primed from repeats near the DSB, again confirming that sequence near, but not adjacent to, the junction can exert a strong effect on repair outcome. Examination of non SD-MMEJ consistent junctions at the Rr3 DSB (Table 4-9) suggests that some amount of templated synthesis took place before end joining either by blunt ligation or by a synthesis event creating a <4bp repeat.

```
TGGCGGCCGTATATTACCCTactTCCCTAGCCGGGGAAGGACA
----------------CCCTA---CCCTA--------------
--------------------CTTCCC------GGGAAG----
```

Table 4-9. Example of a non SD-MMEJ consistent short insertion recovered from the Rr3 DSB.

Note the two repeats intersecting the insertion junctions the bottom sequence, suggesting that templated synthesis took place in the course of repair.

Short insertions at yellow:

Patterns of priming at longer repeats near the DSB are also evident at the *yellow* DSB. A second class of SD-MMEJ consistent insertions at yellow appears to derive from very short additions templated from the AT-rich single stranded ends, resulting in overlapping SD-MMEJ consistent repeats. Interestingly, the latter pattern is also apparent at nearly all the rare non SD-MMEJ consistent insertion junctions at the *yellow* DSB (Table 4-10).

i	CCACTGCCTACCGCATTAAA**a**GTGGATGAGTGTGGTCGGCT (x2)
ii	TTACCACTGCCTACCGCATT**aa**AAAGTGGATGAGTGTGGTCG
iii	TTACCACTGCCTACCGCATT**aat**AAAGTGGATGAGTGTGGTCG (x3)
iv	TATTACCACTGCCTACCGCA**tc**GTGGATGAGTGTGGTCGGCT
v	CAACAGTATTACCACTGCCT**tt**TGGATGAGTGTGGTCGGCTG

Table 4-10. Non SD-MMEJ consistent short insertions recovered from the yellow DSB.

Note that most junctions suggest production of short overhangs from repeated slipped mispairing events between the recessed 3' end and single stranded 5' overhang (see Table 4-6).

Short insertions at rosy:

The spectrum of short insertions observed at the *rosy* DSB is unusual in having a large proportion of non SD-MMEJ consistent junctions, and the structures of the observed SD-MMEJ consistent repair junctions are unusually diverse, even though 3 bp repeats that appear as if they ought to be strong primer repeats are present near both recessed 3' ends. Conversely, the sequences of the non SD-MMEJ consistent 1-3 bp repeats recovered from the *rosy* DSB show some striking similarities (Table 4-11). This sequence is an excellent candidate for further analysis to refine the SD-MMEJ model.

i	CGCACCTATAGCTACTACAC**c**GAATGGCGTGGGAGTCACTG (x6)
ii	CACCTATAGCTACTACACGA**aa**ATGGCGTGGGAGTCACTGTG
iii	ACCTATAGCTACTACACGAA**a**TGGCGTGGGAGTCACTGTGG (x4)
iv	ACCTATAGCTACTACACGAA**a**TGGCGTGGGAGTCACTGTGG (x2)
v	CCCGACCTATAGCTACTACA**ta**CGAATGGCGTGGGAGTCACT
vi	GCCCGCACCTATAGCTACTA**tac**GGCGTGGGAGTCACTGTGGT
vii	GCACCTATAGCTACTACACG**cg**AATGGCGTGGGAGTCACTGT
viii	GCACCTATAGCTACTACACG**cga**AATGGCGTGGGAGTCACTGT (x2)

Table 4-11. Non SD-MMEJ consistent short insertions recovered from the *rosy* DSB

Structural similarities between sequences (compare v and vi, vii and viii) suggest that repair took place by a nonrandom as yet unelucidated mechanism.

Short insertions at P{w^a}:

Sequences of *P{w^a}* junctions with short insertions tend to show deletion boundaries correlating with relatively repetitive sequences, which could support SD-MMEJ via slipped mispairing.

Microhomology junctions

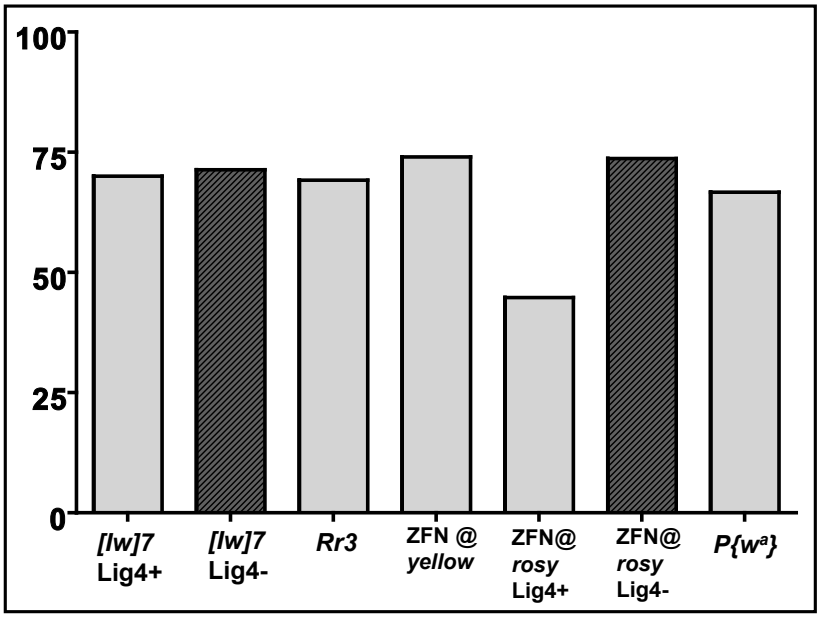


Figure 4-7. Proportions of SD-MMEJ consistent microhomology junctions

[Iw]7 (Lig4+/+): n=99; *[Iw]7* (Lig4-/-): n=35. Percent total SD-MMEJ consistent *[Iw]7* microhomology junctions is significantly greater than predicted by random chance (P<0.05, Fisher's Exact Test). Rr3 (WT): n=39. Percent total SD-MMEJ consistent microhomology junctions at Rr3 is significantly greater than predicated by random chance (P<0.01, Fisher's Exact Test). *yellow* (WT): n=73; percent total SD-MMEJ consistent microhomology junctions for the *yellow* DSB significantly greater than predicted by random chance (P<0.01, Fisher's Exact Test). *rosy* (Lig4+/+): n=29; *rosy* (Lig4-/-): n=19. *P{w^a}* (WT and Blm-/-): n=18. Sequences are as follows:

[Iw]7 **(I-*Sce*I)**	AGCAGCAGCCGAATTCGGTACATTACCCTG***TTAT*** CCCTAGCGGCCGCATAGGCCACTAGT TCGTCGTCGGCTTAAGCCATGTAATGGGAC ***AATA***GGGATCGCCGGCGTATCCGGTGATCA
Rr3 (I-*Sce*I)	ACGGTGTTGTGGCGGCCGTATATTACCCTG***TTAT*** CCCTAGCCGGGGAAGGACAGCTTCTT TGCCACAACACCGCCGGCATATAATGGGAC ***AATA***GGGATCGGCCCCTTCCTGTCGAAGAA
yellow **(ZFN)**	GCGCCAACAGTATTACCACTGCCTACCGCA ***TTAA***AGTGGATGAGTGTGGTCGGCTGTGGG CGCGGTTGTCATAATGGTGACGGATGGCGT***AATT*** TCACCTACTCACACCAGCCGACACCC
rosy **(ZFN)**	ATCCCAATGCCCGCACCTATAGCTACTACA ***CGAA***TGGCGTGGGAGTCACTGTGGTAGAGA TAGGGTTACGGGCGTGGATATCGATGATGT***GCTT*** ACCGCACCCTCAGTGACACCATCTCT

Given the generally high probability that a microhomology junction will be SD-MMEJ consistent by random chance, it is especially striking that in three of the systems investigated (*[Iw]*7, Rr3, and the *yellow* ZFN DSB), the observed proportion of SD-MMEJ consistent

microhomology junctions is significantly greater than would be expected by random chance. This renders it tempting to speculate that in at least some contexts, observed increases in microhomology junctions in C-NHEJ deficient backgrounds could also correlate with an increase in SD-MMEJ consistent repair products.

The contrasting effect of the *lig4* genetic background observed at the *[Iw]7* and *rosy* DSBs is observed again for microhomology junctions: the proportion of consistent microhomology junctions is unchanged in *[Iw]7*, but increased at *rosy*.

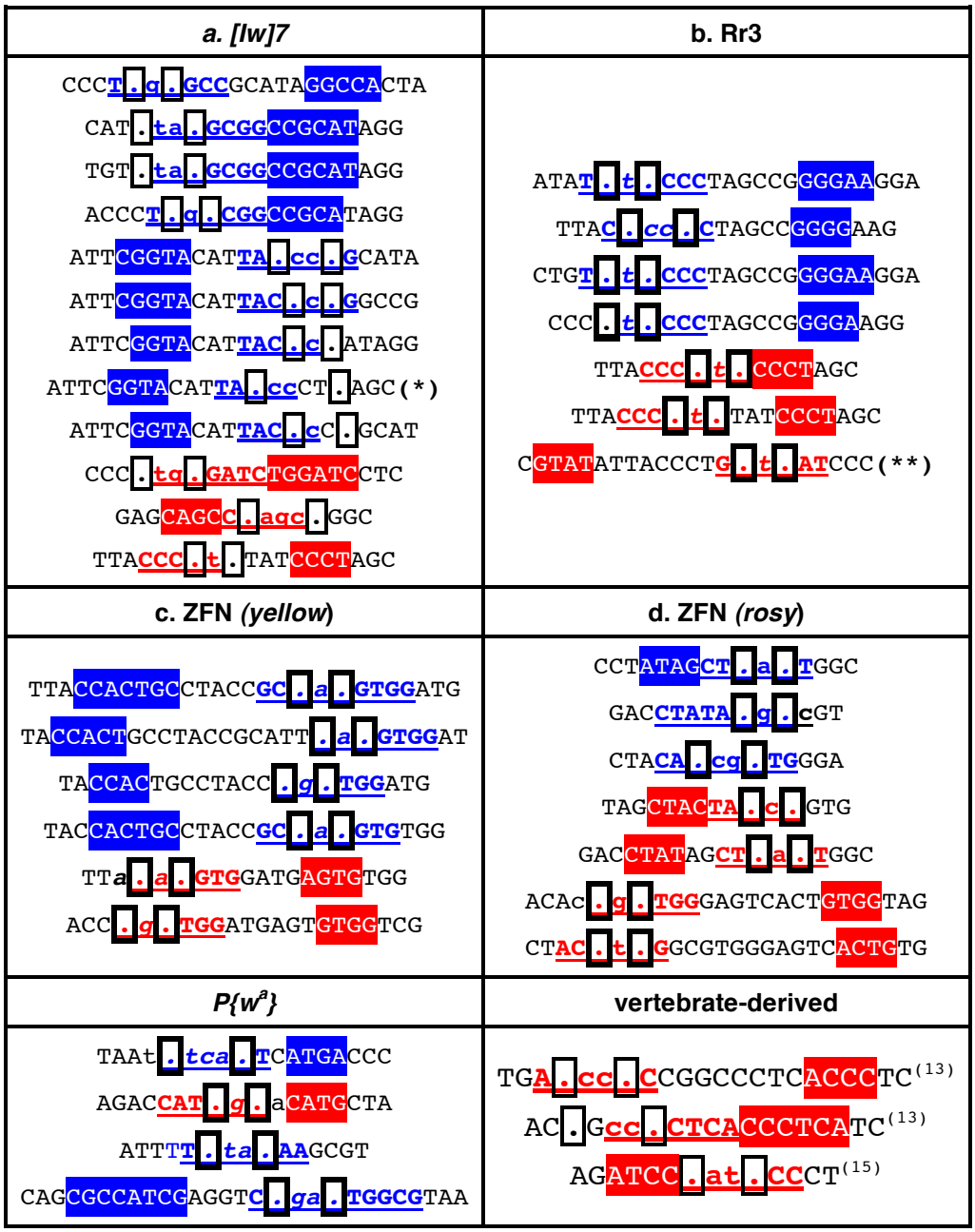


Table 4-12. Representative patterns of SD-MMEJ consistent microhomology junctions

The repair junction consistent with repair by MMEJ at the 4 bp CCCT microhomology (*) is SD-MMEJ consistent in the *[lw]7* system, but not in the Rr3 system. Note the high degree of similarity among the SD-MMEJ consistent repair junctions at the *yellow* DSB.

Microhomology junctions at [Iw]7 *and Rr3:*

In the *[Iw]7* system, the end-joining repair junction consistent with MMEJ at the 4 bp CCCT microhomology contained within the I-*Sce*I site is SD-MMEJ consistent (Table 4-12, asterisked), rendering the origin of those junctions ambigious and raising the question of whether the high proportion of SD-MMEJ consistent microhomology junctions at the *[Iw]7* DSB could be fortuitous. In the Rr3 system, the CCCT microhomology is not SD-MMEJ consistent. Nevertheless, the proportions of SD-MMEJ consistent microhomology junctions in both I-*Sce*I based systems are essentially equivalent, suggesting that the SD-MMEJ consistency of the CCCT repeat did not artificially "inflate" the frequency of SD-MMEJ consistent junctions in the *[Iw]7* system.

Interestingly, one of the most frequently observed SD-MMEJ consistent microhomology junctions from the Rr3 DSB (marked with (**) in Table 4-12; recovered 11 independent times) is an end mispairing that is SD-MMEJ consistent in the Rr3 system, but not in the *[Iw]7* system. These observations are consistent with the idea that the frequency at which specific short junctional microhomologies are observed correlates with whether or not they are SD-MMEJ consistent. In fact, in the Rr3 system, all of the 1 and 2 bp microhomology junctions recovered are SD-MMEJ consistent.

Microhomology junctions at yellow:

There are few substantial pre-existing microhomologies near the *yellow* DSB, which could help explain the high proportion of SD-MMEJ consistent repair junctions.

The diversity of SD-MMEJ consistent microhomology junctions is especially low at the *yellow* DSB; identical repair junctions were independently recovered up to 18 times, and the different SD-MMEJ consistent junctions resemble each other so closely on the sequence level that they

are difficult to distinguish (Table 4-13). Interestingly, the only junction consistent with a single SD-MMEJ consistent repeat was recovered only once. This could reflect a requirement for synthesis from both break ends for efficient SD-MMEJ in this context, or, perhaps, the proximity of the direct repeat to the break renders it more efficient in driving repair.

i	TTACCACTGCCTACCGCATTa.a.GTGGATGAGTGTGGTCGGCT TTACCACTGCCTACCGCATTa.a.GTGGATGAGTGTGGTCGGCT x16
ii	TTACCACTGCCTACCGCATT.a.GTGGATGAGTGTGGTCGGCT TTACCACTGCCTACCGCATT.a.GTGGATGAGTGTGGTCGGCT x9
iii	GTATTACCACTGCCTACCGC.a.GTGGATGAGTGTGGTCGGCT GTATTACCACTGCCTACCGC.a.GTGGATGAGTGTGGTCGGCT x18
iv	TATTACCACTGCCTACCGC.a.GTGGATGAGTGTGGTCGGCT TATTACCACTGCCTACCGC.a.GTGGATGAGTGTGGTCGGCT x2
v	CAGTATTACCACTGCCTACC.g.TGGATGAGTGTGGTCGGCTG CAGTATTACCACTGCCTACC.g.TGGATGAGTGTGGTCGGCTG x8
vi	GTATTACCACTGCCTACCGC.a.GTGTGGTCGGCTGTGGGTTT x1

Table 4-13. SD-MMEJ consistent microhomology junctions recovered from the *yellow* DSB.

Number of independent recoveries indicated to the right of the sequence. Repair junctions i - v are consistent with either of two SD-MMEJ mechanisms (indicated by blue and red highlighted repeats). Sequence vi was recovered only once, suggesting that the GTG direct repeat rather than the TCG/GCA inverted repeat more heavily influences repair.

Microhomology junctions at rosy*:*

The proportion of SD-MMEJ consistent microhomology joins shows what appears to be a substantial increase in the *lig4* mutant background (Figure 4-6). Although the proportion of consistent junctions is not statistically significant from random chance in either genetic background, perhaps due to the relatively high likelihood of fortuitously consistent microhomology junctions at this DSB (estimated to be approximately 63%), the difference in the frequency of SD-MMEJ consistent microhomology junctions between the Lig4 competent and

deficient backgrounds at the *rosy* DSB does approach statistical significance (*P*=0.075, Fisher's Exact Test).

Microhomology junctions at P{w^a}:

Annealing at relatively long microhomologies appears to be a prominent mode of repair at this DSB, and, accordingly, the observed junctional microhomologies do not show convincing evidence of being influenced by SD-MMEJ consistency.

Apparent blunt joins

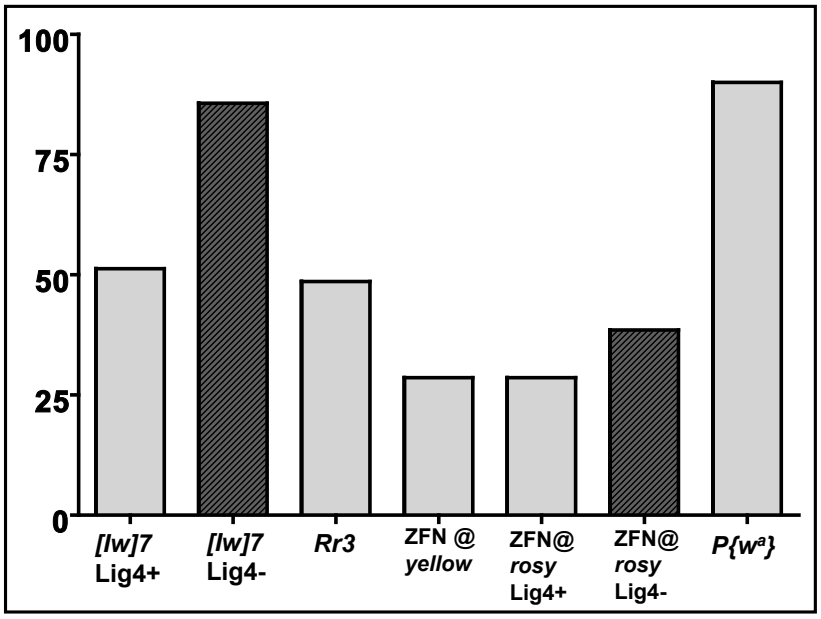


Figure 4-8. Proportion of SD-MMEJ consistent apparent blunt joins by DSB and Lig4 status

[lw]7 (Lig4+/+): n=80; *[lw]7* (Lig4-/-): n=21; Rr3 (WT): n=35. *yellow* (WT): n=21. *rosy* (Lig4+/+): n=21. *rosy* (Lig4-/-): n=13; *P{w^a}* (WT and Blm-/-): n=4. Data for Lig4 competent backgrounds is light gray; Lig4 deficient backgrounds are indicated by dark shaded gray.
Proportion of total *[lw]7* SD-MMEJ consistent apparent blunt joins is greater than predicted by random chance (*P*<0.05, Fisher's exact test). Despite small sample size (n=10), *P{w^a}* apparent blunt joins show a definite trend towards greater frequency than predicted by random simulation (*P*=0.14, Fisher's Exact Test). Sequences are as follows:

[lw]7 **(I-*Sce*I)**	AGCAGCAGCCGAATTCGGTACATTACCCTG***TTAT*** CCCTAGCGGCCGCATAGGCCACTAGT TCGTCGTCGGCTTAAGCCATGTAATGGGAC ***AATA***GGGATCGCCGGCGTATCCGGTGATCA
Rr3 (I-*Sce*I)	ACGGTGTTGTGGCGGCCGTATATTACCCTG***TTAT*** CCCTAGCCGGGGAAGGACAGCTTCTT TGCCACAACACCGCCGGCATATAATGGGAC ***AATA***GGGATCGGCCCCTTCCTGTCGAAGAA
yellow **(ZFN)**	GCGCCAACAGTATTACCACTGCCTACCGCA ***TTAA***AGTGGATGAGTGTGGTCGGCTGTGGG CGCGGTTGTCATAATGGTGACGGATGGCGT***AATT*** TCACCTACTCACACCAGCCGACACCC
rosy **(ZFN)**	ATCCCAATGCCCGCACCTATAGCTACTACA ***CGAA***TGGCGTGGGAGTCACTGTGGTAGAGA TAGGGTTACGGGCGTGGATATCGATGATGT***GCTT*** ACCGCACCCTCAGTGACACCATCTCT

The proportions by DSB context and genotype of SD-MMEJ apparent blunt joins again differ, but this time with the opposite trend observed for the other junction types: for apparent blunt joins, the proportion of SD-MMEJ consistent apparent junctions is markedly increased in

the *lig4* background at the *[Iw]7* DSB, but not at the *rosy* DSB, underscoring the importance of sequence context in determining repair outcome.

a. *[Iw]7*	b. Rr3
ACATTACCCTGTTA-CCCTAGC TGTTAT-GCGGCCGCATAGGC TGTTA-GGCCGCATAGGCCACT CCTGT-GGCCGCATAGGCCACTAG TGTT-GCGGCCGCATAG TACATTACCCTGTTAT-TAGCG	TTACCCT-TCCCTAGCCGGGAAGGACA TTACC-TTATCCCTAGCCGGGGAAGG TACCCT-TTATCCCTAGCCGGGGAAGG GTATATTACCCTGTTAT-TAGC ATATTACCCTGTTA-CCCTAGC
c. ZFN *(yellow)*	**d. ZFN *(rosy)***
CATT-GTGGATGAGTGTGGTCG CGCAT-GATGAGTG TTACCACTGCCTAC-AGTGGATG CTACCGC-GGATG	CTATAGCTA-GCGTG CGAA-GTGGGAGTCACTGTG ACACG-TGGCGTGGGAG CGACCTATAGCTACTA-GGCGT ACTACA-GGAGTCACTGTGGT
P{w^a}	**vertebrate-derived**
AAAACATTAAC-atgtt CTTTACCTATGCTA-ccgac	CGGGCCAAATGGCCGGAGTTGTCAGATCCATT-TGGCAG[15]

Table 4-14. Representative sequences showing patterns of SD-MMEJ consistent apparent blunt joins
Superscript references the study in which the DSB was observed.

Apparent blunt joins at [Iw]7 *and Rr3*

Considering only Lig4 proficient genetic backgrounds, the frequency of SD-MMEJ consistent apparent blunt joins was not different from random chance in either system. However, there is a dramatic shift towards SD-MMEJ consistent apparent blunt joins at the *[Iw]7* DSB in a *lig4* background. It would be interesting to determine whether this would also be observed at the Rr3 DSB.

Although the numerical proportion of SD-MMEJ consistent blunt joins observed at the Rr3 DSB was not different from random chance, observation of the specific SD-MMEJ

consistent repair junctions recovered suggests a pattern of repair mediated by the longer primer repeats available near the DSB.

Interestingly, distinct spectra of non SD-MMEJ consistent apparent blunt joins are observed in the two systems. For example, the 4 bp deletion corresponding to trimming of the 4 bp 3' overhangs and blunt ligation is not SD-MMEJ consistent in either system. This was a major repair product in the Rr3 system, comprising nearly a third of apparent blunt joins, but did not occur with any notable frequency in the *[Iw]7* apparent blunt joins.

Apparent blunt joins at yellow:

In striking contrast to the high proportion of SD-MMEJ consistent short insertions and microhomology junctions recovered from the yellow DSB, relatively few of the infrequent apparent blunt joins (21 out of 162 sequences) recovered from the *yellow* DSB were SD-MMEJ consistent (6 of 21). However, as at the Rr3 DSB, the patterns of which particular SD-MMEJ consistent repeats were observed (compare with patterns observed at microhomology junctions) again suggest a nonrandom process.

Examination of the sequences of the non SD-MMEJ consistent junctions from the yellow DSB implies that the infrequency of SD-MMEJ consistent apparent blunt joins could be in part due to the readily available opportunities to create complementary ends via addition of one or two base pairs templating from the 5' overhangs. The most frequently recovered apparent blunt join (independently recovered 10 times) has a structure consistent with such an event.

Apparent blunt joins at rosy*:*

The overall frequency of SD-MMEJ consistent apparent blunt joins at the *rosy* locus is very nearly statistically significantly *lower* than would be predicted by random chance (32 versus 51 percent; P=0.074, Fisher's Exact Test), suggesting that some other mode of repair is the predominant generator of apparent blunt joins at this DSB. Interestingly, the observed spectrum of non SD-MMEJ consistent apparent blunt joins at *rosy* is quite diverse, lacking a clear relationship between sequence content and repair outcome.

Discussion

The results of this meta-analysis indicate that the high proportions of SD-MMEJ consistent repeats observed in the *[Iw]*7 system was likely not an artifact of the sequence context of that particular DSB. SD-MMEJ is likely a prominent mode of end-joining repair in Drosophila. Even in cases where the majority of junction sequences were not consistent with SD-MMEJ repair, consistency in the specific patterns of observed SD-MMEJ consistent repeats suggests that even if the sequence was not necessarily an efficient substrate for SD-MMEJ, SD-MMEJ was likely contributing to some repair.

The results of this meta-analysis firmly establish the generality of the observation made in study in Chapter 3 that a complex and poorly understood interrelationship exists between the sequence surrounding a DSB, genetic background, and the outcome of end-joining repair. The especially dramatic contrast between the effect of the *lig4* mutant background on repair of the *[Iw]7* and *rosy* DSBs emphasizes that it may be perilous to draw general conclusions about the effect of C-NHEJ deficiency on end-joining repair from a study that examines only a single molecularly defined DSB.

Contrasting the proportions of SD-MMEJ consistent apparent blunt joins, microhomology junctions, and insertion junctions obtained within the context of a single break in some cases to renders an apparently paradoxical result: specifically, the proportion of SD-MMEJ consistent junctions for different junction types resulting from the same break can differ substantially. For instance, at the *yellow* DSB, we observe high proportions of SD-MMEJ consistent insertion and microhomology junctions, but few SD-MMEJ consistent apparent blunt joins. On the surface, this observation would seem to contradict the assertion that SD-MMEJ is a unifying model for all three classes of repair junction structures. However, closer inspection of

the overhangs likely to be synthesized at available primer repeats and the opportunities for base-pairing on the other side of the DSB (Table 4-15) suggests a different interpretation: SD-MMEJ apparent blunt joins are probably infrequent at the *yellow* DSB because few of the available combinations of primer and microhomology repeats can mediate an SD-MMEJ repair event that does not produce a net insertion.

	Potential overlaps with sequence to right (bold)			
left 3' end	**ABJ overlap**	**overlap with 1bp insert**	**overlap with 2bp insert**	**overlap with 3bp insert**
CCAACAGTATTACCACTGCCTACCGCA **GGTTGTCATAATGGTGACGGATGGCGTAATT**				
CCAACAGTAT**TACC**ACTGCC**TACC**(actg) GGTTGTCATAATGGTGACGGATGG	*TT***AAA**GTGG**A**TG**A**GTG aa**t**t**T**CA**C**CT**A**C**T**CAC	TTAAAGTGGATGAGTG aatttCACCTACTCAC	***TT***AAAG**TG**GA**TG**AG**TG** aatttCA**C**CT**A**C**T**CAC	TTAAA**G**T**GG**AT**G**AG**TG** aatttC**A**CCTA**C**T**C**AC
CCAA**CAGT**ATTACC**ACTG**(ttgg) GGTTGTCATAATGGTGAC	***TT***AAAG**T**GGA**T**GAG**T**G aa**t**t**T**CA**C**CT**A**C**T**CAC	TTAAAG**TGGATG**AG**TG** aatttCA**CC**TA**C**TCAC	TTAAA**G**T**GG**AT**G**AG**TG** aatttCA**C**CTA**C**T**C**AC	TTAAA**G**T**GG**AT**G**AG**TG** aatttC**A**CCTA**C**T**C**AC
CCAACAGTATTACCAC**TGC**CTACC**GCA**(gtgg) GGTTGTCATAATGGTGACGGATGGCGT	TTAAA**GTGG**ATGA**GTG** aattT**CACC**TACT**CAC**	TTAAAGTGGA**TG**AG**TG** aatttCACCT**AC**TC**AC**	TTAAAGTGGATGAGTG aatttCACCTA**C**T**C**AC	TTAAA**G**T**GG**AT**G**AG**TG** aatttC**A**CCTA**C**T**C**AC

Table 4-15. The number of possible SD-MMEJ events is large.

Example illustrating that due to the large numbers of possible SD-MMEJ events, obtaining maximum predictive power from the SD-MMEJ model will require computerized analysis. Candidate strong primer repeats near the left 3' end of the *yellow* DSB (highlighted), overhang sequence that would result from synthesis using the indicated primer repeats (lowercase in parentheses), and potential base pairings with sequence on the opposite side of the DSB that would result in an apparent blunt join or a 1-3 bp insertion as indicated (white on black). 3' end sequence that would result from filling in the single stranded 5' overhangs to the right of the DSB is indicated by lowercase.

The example in Table 4-15 illustrates that there are so many possible SD-MMEJ events at any given DSB it is not feasible to carry out the sequence analysis necessary to predict the outcome of SD-MMEJ at a given DSB by hand. It will be necessary to automate this analysis to design SD-MMEJ reporters with combinations of primer and microhomology repeats able to efficiently promote creation of multiple junction types by SD-MMEJ. However, there are few, if any, sequence analysis programs currently available that are optimized for identification and characterization of short exact repeats, and, to our knowledge, none that are able to effectively model the outcome of iterative rounds of synthesis at a DSB. As will be described in Chapter 5, we are actively developing new software to address this need.

Given the complexity of the relationship between sequence context and end-joining repair, design of appropriate SD-MMEJ reporters will be an interesting, though ultimately tractable, challenge. Rather than attempting to design the sequences outright, it will likely be most efficient to generate large numbers of random sequences with base composition or positioning rendering the probability of fortuitous SD-MMEJ consistency low *a priori*, and then analyzing the randomly generated sequences to identify the likely primer repeats. Alternatively, computational analysis could focus on generating slight variations on the DSB substrates analyzed in this chapter. For example, the *[Iw]*7 DSB appears to already be a surprisingly good substrate for SD-MMEJ, but could be improved as a reporter by small base pair alterations, for example to make the products of snap-back versus loop-out synthesis at the GGCC repeat more readily distinguishable. It would also be interesting to generate slight variations on the *rosy* DSB to investigate which, if any, improved its ability to support SD-MMEJ.

The difficulties of designing a good SD-MMEJ reporter became clearly apparent during the process of selecting data sets to include in the meta-analysis. Many data sets appeared, upon casual inspection, to show convincing evidence of SD-MMEJ repair. However, when subjected to simulations designed to predict the expected frequency of fortuitous SD-MMEJ consistent apparent blunt joins and microhomology junctions, the expected frequency often turned out to be unacceptably high. For example, non-complementary ends produced by double I-*Sce*I cuts, which are widely used in reporters in vertebrate-derived systems, have a repetitive structure that results in nearly all possible microhomology junctions near the DSB being SD-MMEJ consistent, rendering it impossible to discern whether or not patterns of junctional microhomology are due to use of SD-MMEJ repair.

Fortunately, the results of the data sets that were appropriate for inclusion in this meta-analysis demonstrate that it is indeed possible to specify a molecularly defined DSB at which the probability of fortuitous SD-MMEJ consistent apparent blunt joins and/or microhomology junctions under maximally inclusive consistency criteria is acceptably low. This result is especially welcome from the perspective of designing SD-MMEJ reporters for use in vertebrate-derived systems, in which apparent blunt joins and microhomology junctions tend to dominate repair junction spectra.

It is quite possible that studies in non-insect systems may show that SD-MMEJ is not an equally important mechanism for end joining in all organisms. If this is the case, it would make the SD-MMEJ model especially interesting from the perspective of molecular evolution, as differential use of SD-MMEJ repair across organisms could help account for interspecies differences in the evolution of genome architecture (16).

The most important conclusions to be drawn from this study, however, are not specific to the SD-MMEJ model. It has been shown many times over that when the most reliable result across studies is the degree to which the results appear inconsistent, it is highly probable that fundamental approaches to data analysis need to be reconsidered before further progress can be made. This study demonstrates that reevaluation of the ways in which sequence data are interpreted can lead to fresh insights regarding the relationship between DNA sequence and the mechanisms of mutagenesis.

Materials And Methods

End-joining sequence data:

The Rr3 sequence data were generated in the course of the study described in (3). The complete data set with additional flanking sequence was provided by Dr. William Engels (personal communication). The zinc finger nuclease DSB junction data were generated in the course of the studies described in (17, 18). Junction sequences not explicitly published, along with additional flanking sequence data, were provided by Dr. Dana Carroll (personal communication). Additional flanking sequence at the *P{w^a}* junctions published in Adams et al. were reconstructed by the author of this study based on the complete *P{w^a}* sequence (Mitch McVey, personal communication). All junction sequences were checked by alignment with the original sequence. Sequences that could not be unambiguously reconstructed from the available information (for example, sequences with large insertions of unspecified base composition, or deletions identified as having junctional microhomology for which two copies of the repeated motif could not be located in the original sequence) were discarded, as were the few sequences from the I-*Sce*I and ZFN data sets with deletions so large as to preclude meaningful comparison with the other junction sequences (typically >40 bp). Some junctions were reclassified in this study based on the author's alignment; for example, for the I-*Sce*I DSB, the 3 bp insertion consistent with mispairing between the 3' overhangs was reclassified as a microhomology junction rather than an insertion. For the purposes of this study, an "insertion" was defined to be all base pairs intervening between the left and right boundaries demarcating the left and rightmost position at which the original and repair product sequence did not match, even if some of the bases in the interval between the left and the rightmost mismatched positions could be aligned with the original sequence. This criterion for defining an insertion was made on the

consideration that for a base pair to be changed or removed, the base pairs intervening between the changed/removed base and the end of the DNA would also need to be removed.

Computational analysis:

All computer programs used to carry out the simulations and analyses in this chapter were written by the author in the Python programming language.

The program that generated and analyzed the random apparent blunt joins and microhomology junctions summarized in Figure 4-1, Table 4-1, and Table 4-2 takes as input the type of junction to be generated (microhomology or apparent blunt join), relative proportions of bases to the left and right of the DSB (and in the microhomology if specified), the minimum match length required to classify a sequence as SD-MMEJ consistent or not, the length of the random sequence to be generated, the length of the junctional microhomology (if applicable), and the number of times the simulation should be run. For each run, the simulation randomly generates independent DNA sequences with the specified base compositions as the right and left sides of the repair junction (and the microhomology, if appropriate) and joins them into a single sequence. The program then uses the randomly generated sequence to determine the base composition of all motifs spanning the point(s) at which the sequences were joined that, if repeated, would result in the randomly generated junction being SD-MMEJ consistent. Once the breakpoint-spanning sequences are determined, the program uses a sliding window to systematically check the sequence surrounding the breakpoint for matches to the breakpoint-spanning motifs. As soon as a match is found, the program increments a counter keeping track of the number of matches, aborts the current run, and begins the next run with a new randomly generated repair junction. If no match is found, a corresponding counter is incremented and the next run begins, again with a new repair junction. The error-free operation of the program was

confirmed via manual examination of the program output, an example of which is given in Table 4-16 below:

Table 4-16. Sample output from SD-MMEJ analysis software

Example output from a computer program that generates random blunt joins of a given sequence composition, tests them for SD-MMEJ consistency, and tabulates the results of multiple runs.

```
========================================
Running HazConsistentRepeat, wordLen=4
AGCCATGACGCCCATGCCCG (len(seq)=20, bkpt=10)
refMotifIndices are [7, 8, 9]
========================================
```

Example of the report output at the beginning of each run, showing the sequence that has been randomly generated for that run.

```
Running IsRep(refIndex=9, testIndex=12, wordLen=4, debug=True)
AGCCATGACGCCCATGCCCG seq
---------GCCC------- [9, 13] refSeq (rc: GGGC)
------------catg---- [12, 16] testSeq
does not match!

Running IsRep(refIndex=9, testIndex=13, wordLen=4, debug=True)
AGCCATGACGCCCATGCCCG seq
---------GCCC------- [9, 13] refSeq (rc: GGGC)
-------------atgc--- [13, 17] testSeq
does not match!

Running IsRep(refIndex=9, testIndex=14, wordLen=4, debug=True)
AGCCATGACGCCCATGCCCG seq
---------GCCC------- [9, 13] refSeq (rc: GGGC)
--------------tgcc-- [14, 18] testSeq
does not match!

Running IsRep(refIndex=9, testIndex=15, wordLen=4, debug=True)
AGCCATGACGCCCATGCCCG seq
---------GCCC------- [9, 13] refSeq (rc: GGGC)
---------------gccc- [15, 19] testSeq
match found!
```

The four last iterations of the sliding window matching algorithm for the run in the first panel.

Table 4-16 (continued)

```
========================================
ConsistentRandomBluntSim Final Results
----------------------------------------
trials=10, seqLen=20, wordLen=4
LpctA=25, LpctT=20, LpctG=30, LpctC=25
RpctA=10, RpctT=30, RpctG=10, RpctC=50

2 consistent; 8 not consistent
----------------------------------------
========================================
```

Report generated upon completion of the program, including the number of trials (here, 10), the length of the random sequences generated (20), the minimum length for an SD-MME J consistent repeat (4), the base pair compositions to the left and right of the DSB, and the total numbers of SD-MMEJ consistent/not consistent sequences tabulated.

Two programs were used to determine the expected frequency of SD-MMEJ consistent repeats in the context of the specific DSB substrates given in Table 4-3. The first program, which was also used to generate the expected proportions of apparent blunt joins and microhomology junctions in Table 4-7, took as input the original DNA sequence to the left and right of the top strand nick, two lists of integers from which deletion lengths to the left and right of the nick were randomly selected at each run ("deletion pools"), and the number of trials to run. The content of the deletion pools were specified based on the observed range of deletion lengths in the data sets. At each run, the program randomly selected deletions to the left and right of the top strand nick, applied the deletions to the original sequence, and compared the resulting internally deleted sequence to the original sequence to determine whether or not the deletion had junctional microhomology. Internally deleted sequences were enumerated, aligned with the original sequence, and written to an output file sorted by the junction type. The error-free operation of the program was verified via examination of output files. An excerpt from a typical output file is shown in the Table 4-17 below:

Table 4-17. Sample random deletions output from SD-MMEJ analysis software

```
rosy_ABJ 6175
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTACA CGAATGGCGTGGGAGTCACTGTGGTAGAGATCGATTGCCT
CCGGAAACGAATCCCAATGCCCGCACCTATAG-------- -GAATGGCGTGGGAGTCACTGTGGTAGAGATCGATTGCCT

rosy_ABJ 6176
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTACA CGAATGGCGTGGGAGTCACTGTGGTAGAGATCGATTGCCT
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTACA -----GGCGTGGGAGTCACTGTGGTAGAGATCGATTGCCT

rosy_MH 1
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTACA CGAATGGCGTGGGAGTCACTGTGGTAGAGATCGATTGCCT
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTAC- --------GTGGGAGTCACTGTGGTAGAGATCGATTGCCT

rosy_MH 2
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTACA CGAATGGCGTGGGAGTCACTGTGGTAGAGATCGATTGCCT
CCGGAAACGAATCCCAATGCCCGCACCTATAGCTACTAC- --------GTGGGAGTCACTGTGGTAGAGATCGATTGCCT
```

Sample output from a computer program that generates internally deleted sequences with deletions to the left and right of a breakpoint randomly selected from pools of user-specified deletion lengths. Internally deleted sequences are output as alignments with the original sequence. The sample output above was generated from the original sequence at the *rosy* zinc finger DSB.

To determine the proportions of the randomly generated sequences that were SD-MMEJ consistent, sets of randomly generated apparent blunt joins or microhomology junctions were manually edited into a format compatible with a variant of the program that produced the output in Table 4-16, which had been modified to take user-specified sequence batch files as input, instead of generating random sequences.

For the sequences in the actual data sets, we wanted to identify the complete set of SD-MMEJ consistent repeats rather than simply asking whether or not an SD-MMEJ consistent repeat was present in the sequence. This was accomplished by a program that also used a sliding-window matching algorithm, but rather than terminating the run when a match was found and incrementing a counter, this program iterated through all possible comparisons of subsequences that would be SD-MMEJ consistent repeats if those subsequences were direct or reverse complement repeats. The program compiled a list of the locations of all SD-MMEJ

consistent repeats in the sequence and generated as output an alignment of the original sequence and the complete set of SD-MMEJ consistent repeats. The error-free operation of this program was verified by inspection of the output, examples of which are presented in Figure 4-18 below. SD-MMEJ consistency of insertion junctions was tabulated manually from the generated alignments.

Table 4-18. Sample output from SD-MMEJ analysis software.

Sample output from the computer program used to identify SD-MMEJ consistent repeats in junction sequences from data sets. Here, points of apparent ligation that would be represented by box boundaries in formatted output are indicated by upper/lower case transitions in the original sequence. SD-MMEJ consistent repeats are given as alignments below the junction sequence.

a.

```
seqID: @RYS1CAR4(X2/WT1/OK1)
seqType: ABJ
GCACCTATAGCTACTACACGtggcgtgggagtcactgtgg
----------------CACG--------------------
------------------CGtg------------------

------------------CGtg------------------
------------------------cgtg------------

--------------------Gtgg----------------
------------------------gtgg------------

--------------------Gtgg----------------
-----------------------------------gtgg
```

b.

```
seqID: @RYS1CAR3(X1/OK1)
seqType: ABJ
CCCGCACCTATAGCTACTACcgaatggcgtgggagtcact
```

a,b: Sample output showing SD-MMEJ consistent (a) and not consistent (b) apparent blunt joins at the *rosy* DSB.

Table 4-18 (continued)

c.

```
seqID: ENGwt-25x1
seqType: MH
GCGGCCGTATATTACCCTGTtCCCTAGCCGGGGAAGGACAG
-------------------TtCC------------------
-------------------------------GGAA------

--------------------tCCC-----------------
-----------------------------GGGA-------
```

d.

```
seqID: ENGwt-4x8
seqType: MH
GTTGTGGCGGCCGTATATTAccctAGCCGGGGAAGGACAGCTTC
```

c,d: Sample output from the Rr3 data set showing SD-MMEJ consistent (c) and not consistent (d) microhomology junctions Note that the non-consistent junction shown in panel (d) is the CCCT 4 bp microhomology within the I-*Sce*I target sequence that is SD-MMEJ consistent in the context of the *[Iw]7* DSB but not the Rr3 DSB.

Table 4-18 (continued)

e.

```
seqID: @YCARWT65X1
seqType: INST
ACAGTATTACCACTGCCTACacTGGATGAGTGTGGTCGGCTG
------------------ACac--------------------
----------------------------GTGT---------

----------CACT----------------------------
-------------------CacT-------------------

-------------------CacT-------------------
----------------------------AGTG----------

-CAGT-------------------------------------
--------------------acTG------------------

-----------ACTG--------------------------
--------------------acTG------------------
```

f.

```
seqID: @YCARWT32X3
seqType: INST
TTACCACTGCCTACCGCATTaatAAAGTGGATGAGTGTGGTCG
-----------------ATTa----------------------
-------------------Taat--------------------
```

e,f: Sample output from the *yellow* DSB data set showing SD-MMEJ consistent (e) and non-consistent (f) short insertions. Note that although the insertion in panel (f) has one repeat intersecting a point of ligation (upper/lowercase transition), it is nevertheless not SD-MMEJ consistent because there is no repeat intersecting the other point of ligation.

Chapter 5

Future Directions

Development of software for sequence analysis and modeling of end-joining repair events is ongoing. The software used in the meta-analysis presented in Chapter 4 is highly reliable and enables rapid, accurate studies of data sets far larger than could be analyzed by hand. However, it has certain limitations that will ultimately need to be addressed on a fundamental design level. This will require implementation of more sophisticated, and in some cases completely novel, data structures.

For example, in Table 4-18, SD-MMEJ consistent repeats longer than the minimum consistency threshold of 4 bp are reported as overlapping 4 bp-long substrings. It would be desirable modify the program to report only full-length SD-MMEJ consistent repeats. However, the sliding-window algorithm renders implementation of this enhancement surprisingly difficult because the rules for identifying which substrings comprise a single SD-MMEJ consistent repeat turn out to be unexpectedly complex.

Another enhancement which is highly desirable, but essentially impossible to implement with a sliding-window algorithm, is automated evaluation of longer insertion junctions for SD-MMEJ consistency. As described in Chapter 3, irrespective of junction type, the task of determining whether or not a junction is SD-MMEJ consistent is in essence determination of whether or not that junction has sequence patterns that could have been produced by a specific biological process. This means that to write software that evaluates junction sequences for SD-MMEJ consistency, a fundamentally biological concept must be translated into a set of pattern-matching rules for sorting repair junction sequences. For apparent blunt joins and microhomology junctions, this conceptual translation is straightforward: the sequence is consistent if it contains a repeated motif intersecting an apparent point of ligation. This can easily be implemented as a sliding-window pattern-matching algorithm.

For insertion junctions, however, the additional requirement that existence of the synthesis template be verified at each step means SD-MMEJ consistency cannot be evaluated by simple substring pattern matching in the final repair product. Instead, to evaluate the SD-MMEJ consistency of an insertion junction, it is necessary to directly model the repair process. An algorithm evaluating an insertion junction must first identify the sequence that was present at the beginning of the repair process, and then iteratively determine all possible combinations of net synthesis events (if any) that could culminate in production of the observed repair product, updating the available template at each step. The sliding-window algorithm is clearly not adequate to demands of this problem.

To address the fundamental limitations of the SD-MMEJ consistency evaluation software used in this study, the version of the SD-MMEJ consistency evaluation software currently under active development uses a fundamentally different approach to identifying repeated sequences. Instead of representing a DNA sequence as a linear string of bases, it stores the DNA sequence in a data structure called an affix tree (1, 2). This strategy for representing DNA sequences inside the computer enables very efficient and flexible search algorithms, which can readily be tuned to identify only full-length SD-MMEJ consistent repeats. Additionally, the affix tree representation of the DNA sequence can be efficiently updated in a way that facilitates implementation of the modeling necessary to evaluate SD-MMEJ consistency of sequences with longer insertions. The same basic strategy can also be used to generate detailed predictions about likely outcomes for SD-MMEJ repair at any specific sequence. This will greatly aid development of novel SD-MMEJ reporter sequences. A working prototype of this software has been achieved, and is under active testing.

A second software development project currently underway is automation of the modeling approach used in Chapter 1 to demonstrate that increased nuclease activity, as well as annealing at longer microhomologies, could account for observed increases in junctional microhomology. Currently, many DNA sequence analysis programs, depending on application, either treat DNA sequences as if they were text or take detailed thermodynamic considerations into account. In contrast to this, the SD-MMEJ model was developed by modeling DNA at an intermediate level of complexity. Base pairing of both strands was taken into account, but the details of the thermodynamics were abstracted away. Computational tools for automating the pencil-and-paper thought experiments that led to the SD-MMEJ model are not to our knowledge available in the toolkits provided by bioinformatics programming resources such as BioPython (3). It is our goal to remedy that deficiency. This software should find general application among researchers interested in modeling the consequences of hypotheses concerning enzymatic action during DNA repair.

Appendix A

Rapid Screening Of I-*Sce*I DSB End-Joining Repair Junctions In Drosophila Embryos

Introduction

To further refine the SD-MMEJ model, we wish to compare end-joining repair junction sequences from several different DSB substrates in a variety of different genetic backgrounds. Because the time and work involved in creating the many transgenic Drosophila lines necessary to carry out these experiments in a chromosomal context is prohibitive, we decided to develop an alternative higher-throughput screening approach.

Previous work in our lab and publications in the literature (1) suggested that such screens could be done efficiently via injection of plasmid DNA into fly embryos. In these studies, Drosophila embryos were injected with linearized plasmid DNA, and recircularized repair products recovered via transformation into *E. coli*. Characterization of repair product sequences demonstrated the presence of insertions at repair junctions, suggesting that repair of linearized plasmids proceeds via a mechanism similar to the mechanism producing insertions at chromosomal DSBs. However, unlike the majority of chromosomal repair products we observed in the *[Iw]7* assay, substantial deletions were frequently observed in recircularization products from linear injected DNA. Since testing the SD-MMEJ model requires comparison of end-joining repair junctions created in a specifically defined sequence context, a system biased towards production of large deletions of variable extent is not ideal.

We speculated that the large deletions incurred during repair of linear injected DNA could be due to nuclease activity unrelated to the actual repair event, and that the frequency of large deletions could be reduced if the DNA were injected in circularized form and the DSB induced inside the nucleus, as in the chromosomal assay. Efficient cutting of chromosomal substrates by maternal effect I-*Sce*I in Drosophila embryos has been previously demonstrated (2).

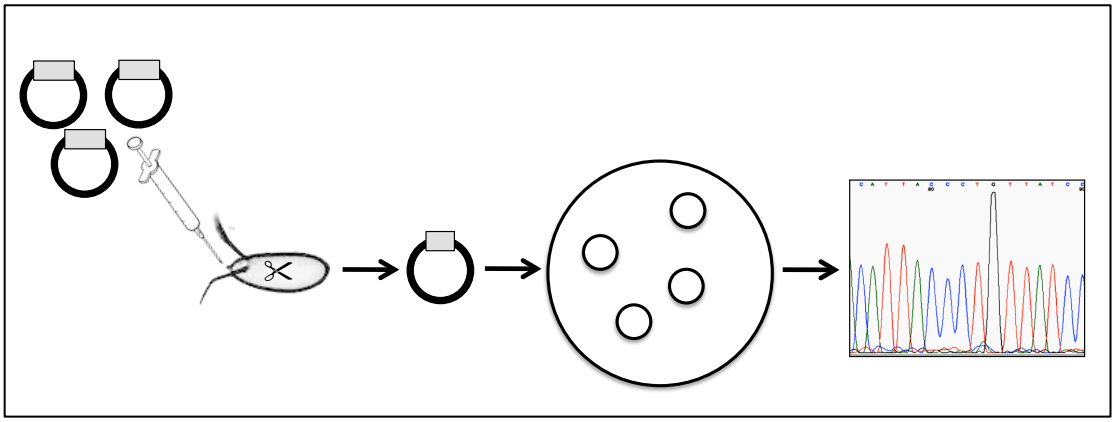


Figure A1-1. Experimental strategy for rapid screening of end-joining repair junctions in Drosophila embryos.

Plasmids containing an I-*Sce*I cut site (gray box) are injected into embryos containing maternal effect I-*Sce*I, which results in successive rounds of cutting and repair. I-*Sce*I resistant plasmids (shortened gray box) are recovered from embryos by alkaline lysis, transformed, and sequenced.

Materials And Methods

Plasmids: A 1.3 kb *[Iw]7* PCR product was TA cloned into the pGEM-T Easy vector (Invitrogen) according to the manufacturer's recommendations. Plasmids were prepared for injection with a midiprep kit (Qiagen), resuspended in water, and stored in aliquots at -20° until use.

Fly stocks: Embryos were obtained from w^{1118}; UIE[I-*Sce*I], Sp / CyO[I-*Sce*I]2R (2) females, which harbor two chromosomally integrated I-*Sce*I expression constructs under a consistitutive ubiquitin promoter.

Embryo injections:

Pre-cellularized embryos were collected at intervals of 30 minutes, dechorionated, and prepared for injection via standard methods. Plasmid DNA was injected at a concentration of approximately 0.1μM. Only pre-cellularized embryos were injected. After injection, embryos were held at room temperature to allow cycles of cutting and repair. In initial experiments,

embryos were held overnight. However, repair products obtained from these experiments contained a substantial proportion of what appeared to be homologous recombination mediated plasmid concatamers and rearrangements. A previous study (3) suggested that homologous recombination mediated plasmid concatamerization occurs primarily in later embryogenesis. Therefore, to maximize the proportion of end-joining repair products, subsequent studies held injected embryos at room temperature for up to 4 hours, after which embryos were frozen at -20° until recovery of plasmid DNA.

Recovery of repaired plasmids:

Injected embryos were ground and low molecular weight DNA recovered by alkaline lysis miniprep and ethanol precipitation. Recovered plasmids were subjected to in vitro I-*Sce*I digestion to eliminate any uncut or perfectly repaired plasmids. Digested plasmids were transformed into chemically competent XL-1 Blue *E. Coli*. To ensure that only independent repair products were recovered, cells were allowed only 20 minutes recovery time before plating.

Screening of transformants:

Individual bacterial colonies were screened for the presence of inaccurate repair junctions via colony PCR with the PE5' and jn3B (See Chapter 3) primers. The PCR products were directly digested with I-*Sce*I and run on a 1% agarose gel. Any samples showing evidence of digestion products were not analyzed further. Colonies yielding I-*Sce*I resistant PCR products were miniprepped and test digested with *XmnI* and *SphI* to exclude repair products with deletions substantially larger than those observed in the chromosomal assay. Repair products that did not yield appropriately sized fragments (approximately 2757 and 1607 bp for *XmnI*; 3912 and 450 for *SphI*) were not analyzed further. I-*Sce*I resistant plasmids that yielded the expected restriction fragments upon *XmnI* and *SphI* digestion were sequenced.

Results

Sequencing of *[Iw]7* repair junctions demonstrated that the spectrum of end-joining repair junctions recovered from the embryo plasmid injection assay closely resemble the junctions observed in the chromosomal assay.

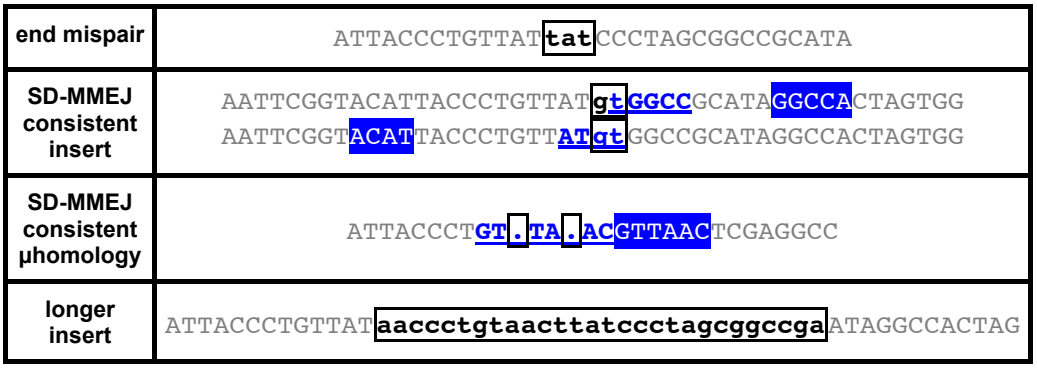


Figure A1-2. End-joining repair junctions obtained from initial studies in the plasmid embryo injection assay.

Note that the SD-MMEJ consistent short insertion obtained resembles SD-MMEJ consistent junctions recovered from the chromosomal assay in that the repair event appears to have been mediated by the GGCC repeat. Additionally, the longer insertion appears to have been synthesized from templates in the surrounding sequence.

Discussion

We have demonstrated that plasmids containing I-*Sce*I target sites injected into embryos obtained from female flies expressing I-*Sce*I are cut and subjected to inaccurate end-joining repair, and that the sequences of the end-joining repair junctions so obtained resemble the junction sequences obtained from repair products of the same DSB induced in a chromosomal context. This is, to the authors' knowledge, a novel approach in Drosophila, and suggests that the end-joining repair mechanisms that operate on plasmid and chromosomal DSBs in this model organism may be in some important respects largely similar. This opens new possibilities for rapid screening approaches to investigate the relationship between sequence context and end-joining repair.

Works Cited

References For Chapter 1

1. Pardo,B., Gomez-Gonzalez,B. and Aguilera,A. (2009) DNA repair in mammalian cells: DNA double-strand break repair: How to fix a broken relationship. *Cell Mol. Life Sci.,* **66,** 1039-1056.

2. Mahaney,B.L., Meek,K. and Lees-Miller,S.P. (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem. J.,* **417,** 639-650.

3. DeFazio,L.G., Stansel,R.M., Griffith,J.D. and Chu,G. (2002) Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.,* **21,** 3192-3200.

4. Spagnolo,L., Rivera-Calzada,A., Pearl,L.H. and Llorca,O. (2006) Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol. Cell,* **22,** 511-519.

5. Lee,J.W., Yannone,S.M., Chen,D.J. and Povirk,L.F. (2003) Requirement for XRCC4 and DNA ligase IV in alignment-based gap filling for nonhomologous DNA end joining in vitro. *Cancer Res.,* **63,** 22-24.

6. Gu,J., Lu,H., Tippin,B., Shimazaki,N., Goodman,M.F. and Lieber,M.R. (2007) XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.,* **26,** 1010-1023.

7. Chen,L., Trujillo,K., Sung,P. and Tomkinson,A.E. (2000) Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J. Biol. Chem.,* **275,** 26196-26205.

8. Moore,J.K. and Haber,J.E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in saccharomyces cerevisiae. *Mol. Cell. Biol.,* **16,** 2164-2173.

9. Taccioli,G.E., Rathbun,G., Oltz,E., Stamato,T., Jeggo,P.A. and Alt,F.W. (1993) Impairment of V(D)J recombination in double-strand break repair mutants. *Science,* **260,** 207-210.

10. Bogue,M.A., Wang,C., Zhu,C. and Roth,D.B. (1997) V(D)J recombination in Ku86-deficient mice: Distinct effects on coding, signal, and hybrid joint formation. *Immunity,* **7,** 37-47.

11. Liang,F., Romanienko,P.J., Weaver,D.T., Jeggo,P.A. and Jasin,M. (1996) Chromosomal double-strand break repair in Ku80-deficient cells. *Proc. Natl. Acad. Sci. U. S. A.,* **93,** 8929-8933.

12. Nick McElhinny,S.A., Snowden,C.M., McCarville,J. and Ramsden,D.A. (2000) Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol. Cell. Biol.,* **20,** 2996-3003.

13. Teo,S.H. and Jackson,S.P. (2000) Lif1p targets the DNA ligase Lig4p to sites of DNA double-strand breaks. *Curr. Biol.,* **10,** 165-168.

14. Guirouilh-Barbat,J., Rass,E., Plo,I., Bertrand,P. and Lopez,B.S. (2007) Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proc. Natl. Acad. Sci. U. S. A.,* **104,** 20902-20907.

15. Iwabuchi,K., Hashimoto,M., Matsui,T., Kurihara,T., Shimizu,H., Adachi,N., Ishiai,M., Yamamoto,K., Tauchi,H., Takata,M., et al. (2006) 53BP1 contributes to survival of cells irradiated with X-ray during G1 without Ku70 or artemis. *Genes Cells,* **11,** 935-948.

16. Schulte-Uentrop,L., El-Awady,R.A., Schliecker,L., Willers,H. and Dahm-Daphi,J. (2008) Distinct roles of XRCC4 and Ku80 in non-homologous end-joining of endonuclease- and ionizing radiation-induced DNA double-strand breaks. *Nucleic Acids Res.,* **36,** 2561-2569.

17. Frank,K.M., Sekiguchi,J.M., Seidl,K.J., Swat,W., Rathbun,G.A., Cheng,H.L., Davidson,L., Kangaloo,L. and Alt,F.W. (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature,* **396,** 173-177.

18. Lee,Y. and McKinnon,P.J. (2002) DNA ligase IV suppresses medulloblastoma formation. *Cancer Res.,* **62,** 6395-6399.

19. Losada,R., Rivero,M.T., Slijepcevic,P., Goyanes,V. and Fernandez,J.L. (2005) Effect of wortmannin on the repair profiles of DNA double-strand breaks in the whole genome and in interstitial telomeric sequences of chinese hamster cells. *Mutat. Res.,* **570,** 119-128.

20. Nevaldine,B., Longo,J.A. and Hahn,P.J. (1997) The scid defect results in much slower repair of DNA double-strand breaks but not high levels of residual breaks. *Radiat. Res.,* **147,** 535-540.

21. Perrault,R., Wang,H., Wang,M., Rosidi,B. and Iliakis,G. (2004) Backup pathways of NHEJ are suppressed by DNA-PK. *J. Cell. Biochem.,* **92,** 781-794.

22. Corneo,B., Wendland,R.L., Deriano,L., Cui,X., Klein,I.A., Wong,S.Y., Arnal,S., Holub,A.J., Weller,G.R., Pancake,B.A., et al. (2007) Rag mutations reveal robust alternative end joining. *Nature,* **449,** 483-486.

23. Yan,C.T., Boboila,C., Souza,E.K., Franco,S., Hickernell,T.R., Murphy,M., Gumaste,S., Geyer,M., Zarrin,A.A., Manis,J.P., et al. (2007) IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature,* **449,** 478-482.

24. Boboila,C., Jankovic,M., Yan,C.T., Wang,J.H., Wesemann,D.R., Zhang,T., Fazeli,A., Feldman,L., Nussenzweig,A., Nussenzweig,M., et al. (2010) Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. *Proc. Natl. Acad. Sci. U. S. A.,* **107,** 3034-3039.

25. Boboila,C., Yan,C., Wesemann,D.R., Jankovic,M., Wang,J.H., Manis,J., Nussenzweig,A., Nussenzweig,M. and Alt,F.W. (2010) Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. *J. Exp. Med.,* **207,** 417-427.

26. Ma,J.L., Kim,E.M., Haber,J.E. and Lee,S.E. (2003) Yeast Mre11 and Rad1 proteins define a ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol. Cell. Biol.,* **23,** 8820-8828.

27. McVey,M., Radut,D. and Sekelsky,J.J. (2004) End-joining repair of double-strand breaks in drosophila melanogaster is largely DNA ligase IV independent. *Genetics,* **168,** 2067-2076.

28. Brown,M.L., Franco,D., Burkle,A. and Chang,Y. (2002) Role of poly(ADP-ribosyl)ation in DNA-PKcs- independent V(D)J recombination. *Proc. Natl. Acad. Sci. U. S. A.,* **99,** 4532-4537.

29. Ahmad,A., Robinson,A.R., Duensing,A., van Drunen,E., Beverloo,H.B., Weisberg,D.B., Hasty,P., Hoeijmakers,J.H. and Niedernhofer,L.J. (2008) ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Mol. Cell. Biol.,* **28,** 5082-5092.

30. Mansour,W.Y., Rhein,T. and Dahm-Daphi,J. (2010) The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. *Nucleic Acids Res.,* **38,** 6065-6077.

31. Williams,R.S., Moncalian,G., Williams,J.S., Yamada,Y., Limbo,O., Shin,D.S., Groocock,L.M., Cahill,D., Hitomi,C., Guenther,G., et al. (2008) Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell,* **135,** 97-109.

32. Sandoval,A. and Labhart,P. (2004) High G/C content of cohesive overhangs renders DNA end joining ku-independent. *DNA Repair (Amst),* **3,** 13-21.

33. Daley,J.M. and Wilson,T.E. (2005) Rejoining of DNA double-strand breaks as a function of overhang length. *Mol. Cell. Biol.,* **25,** 896-906.

34. McVey,M. and Lee,S.E. (2008) MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends Genet.,* **24,** 529-538.

35. Letsolo,B.T., Rowson,J. and Baird,D.M. (2010) Fusion of short telomeres in human cells is characterized by extensive deletion and microhomology, and can result in complex rearrangements. *Nucleic Acids Res.,* **38,** 1841-1852.

36. Decottignies,A. (2005) Capture of extranuclear DNA at fission yeast double-strand breaks. *Genetics,* **171,** 1535-1548.

37. Babushok,D.V., Ostertag,E.M., Courtney,C.E., Choi,J.M. and Kazazian,H.H.,Jr. (2006) L1 integration in a transgenic mouse model. *Genome Res.,* **16,** 240-250.

38. Miller,D.G., Petek,L.M. and Russell,D.W. (2003) Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol. Cell. Biol.,* **23,** 3550-3557.

39. De Buck,S., Jacobs,A., Van Montagu,M. and Depicker,A. (1999) The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. *Plant J.,* **20,** 295-304.

40. Povirk,L.F. (2006) Biochemical mechanisms of chromosomal translocations resulting from DNA double-strand breaks. *DNA Repair (Amst),* **5,** 1199-1212.

41. Simsek,D. and Jasin,M. (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat. Struct. Mol. Biol.,* **17,** 410-416.

42. Lee,J.A., Carvalho,C.M. and Lupski,J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell,* **131,** 1235-1247.

43. Hecht,M.M., Nitz,N., Araujo,P.F., Sousa,A.O., Rosa Ade,C., Gomes,D.A., Leonardecz,E. and Teixeira,A.R. (2010) Inheritance of DNA transferred from american trypanosomes to human hosts. *PLoS One,* **5,** e9181.

44. Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.,* **5,** 2599-2607.

45. Pfeiffer,P., Thode,S., Hancke,J. and Vielmetter,W. (1994) Mechanisms of overlap formation in nonhomologous DNA end joining. *Mol. Cell. Biol.,* **14,** 888-895.

46. Lee,K. and Lee,S.E. (2007) Saccharomyces cerevisiae Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics,* **176,** 2003-2014.

47. Adams,M.D., McVey,M. and Sekelsky,J.J. (2003) Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science,* **299,** 265-267.

48. Perrin,A., Buckle,M. and Dujon,B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.,* **12,** 2939-2947.

49. Decottignies,A. (2007) Microhomology-mediated end joining in fission yeast is repressed by pku70 and relies on genes involved in homologous recombination. *Genetics,* **176,** 1403-1415.

50. Chan,C.Y. and Schiestl,R.H. (2009) Rad1, rad10 and rad52 mutations reduce the increase of microhomology length during radiation-induced microhomology-mediated illegitimate recombination in saccharomyces cerevisiae. *Radiat. Res.,* **172,** 141-151.

51. Lee-Theilen,M., Matthews,A.J., Kelly,D., Zheng,S. and Chaudhuri,J. (2011) CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat. Struct. Mol. Biol.,* **18,** 75-79.

52. Yun,M.H. and Hiom,K. (2009) CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature,* **459,** 460-463.

53. van der Burg,M., Verkaik,N.S., den Dekker,A.T., Barendregt,B.H., Pico-Knijnenburg,I., Tezcan,I., vanDongen,J.J. and van Gent,D.C. (2007) Defective artemis nuclease is characterized by coding joints with microhomology in long palindromic-nucleotide stretches. *Eur. J. Immunol.,* **37,** 3522-3528.

54. Liang,L., Deng,L., Chen,Y., Li,G.C., Shao,C. and Tischfield,J.A. (2005) Modulation of DNA end joining by nuclear proteins. *J. Biol. Chem.,* **280,** 31442-31449.

55. Tseng,H.M. and Tomkinson,A.E. (2004) Processing and joining of DNA ends coordinated by interactions among Dnl4/Lif1, Pol4, and FEN-1. *J. Biol. Chem.,* **279,** 47580-47588.

56. Akyuz,N., Boehden,G.S., Susse,S., Rimek,A., Preuss,U., Scheidtmann,K.H. and Wiesmuller,L. (2002) DNA substrate dependence of p53-mediated regulation of double-strand break repair. *Mol. Cell. Biol.,* **22,** 6306-6317.

57. Pan-Hammarstrom,Q., Lahdesmaki,A., Zhao,Y., Du,L., Zhao,Z., Wen,S., Ruiz-Perez,V.L., Dunn-Walters,D.K., Goodship,J.A. and Hammarstrom,L. (2006) Disparate roles of ATR and ATM in immunoglobulin class switch recombination and somatic hypermutation. *J. Exp. Med.,* **203,** 99-110.

58. Rahal,E.A., Henricksen,L.A., Li,Y., Williams,R.S., Tainer,J.A. and Dixon,K. (2010) ATM regulates Mre11-dependent DNA end-degradation and microhomology-mediated end joining. *Cell. Cycle,* **9,** 2866-2877.

59. Bothmer,A., Robbiani,D.F., Feldhahn,N., Gazumyan,A., Nussenzweig,A. and Nussenzweig,M.C. (2010) 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med.,* **207,** 855-865.

60. Pawelczak,K.S. and Turchi,J.J. (2008) A mechanism for DNA-PK activation requiring unique contributions from each strand of a DNA terminus and implications for microhomology-mediated nonhomologous DNA end joining. *Nucleic Acids Res.,* **36,** 4022-4031.

61. Willers,H., Husson,J., Lee,L.W., Hubbe,P., Gazemeier,F., Powell,S.N. and Dahm-Daphi,J. (2006) Distinct mechanisms of nonhomologous end joining in the repair of site-directed chromosomal breaks with noncomplementary and complementary ends. *Radiat. Res.,* **166,** 567-574.

62. Robert,I., Dantzer,F. and Reina-San-Martin,B. (2009) Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *J. Exp. Med.,* **206,** 1047-1056.

63. Audebert,M., Salles,B. and Calsou,P. (2008) Effect of double-strand break DNA sequence on the PARP-1 NHEJ pathway. *Biochem. Biophys. Res. Commun.,* **369,** 982-988.

64. Zhong,Q., Boyer,T.G., Chen,P.L. and Lee,W.H. (2002) Deficient nonhomologous end-joining activity in cell-free extracts from Brca1-null fibroblasts. *Cancer Res.,* **62,** 3966-3970.

65. Zhuang,J., Zhang,J., Willers,H., Wang,H., Chung,J.H., van Gent,D.C., Hallahan,D.E., Powell,S.N. and Xia,F. (2006) Checkpoint kinase 2-mediated phosphorylation of BRCA1 regulates the fidelity of nonhomologous end-joining. *Cancer Res.,* **66,** 1401-1408.

66. Katsura,Y., Sasaki,S., Sato,M., Yamaoka,K., Suzukawa,K., Nagasawa,T., Yokota,J. and Kohno,T. (2007) Involvement of Ku80 in microhomology-mediated end joining for DNA double-strand breaks in vivo. *DNA Repair (Amst),* **6,** 639-648.

67. Bentley,J., Diggle,C.P., Harnden,P., Knowles,M.A. and Kiltie,A.E. (2004) DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. *Nucleic Acids Res.,* **32,** 5249-5259.

68. Fan,J., Li,L., Small,D. and Rassool,F. (2010) Cells expressing FLT3/ITD mutations exhibit elevated repair errors generated through alternative NHEJ pathways: Implications for genomic instability and therapy. *Blood,* **116,** 5298-5305.

69. Babbe,H., McMenamin,J., Hobeika,E., Wang,J., Rodig,S.J., Reth,M. and Leder,P. (2009) Genomic instability resulting from blm deficiency compromises development, maintenance, and function of the B cell lineage. *J. Immunol.,* **182,** 347-360.

70. Bannister,L.A., Waldman,B.C. and Waldman,A.S. (2004) Modulation of error-prone double-strand break repair in mammalian chromosomes by DNA mismatch repair protein Mlh1. *DNA Repair (Amst),* **3,** 465-474.

71. Garcia,P.B., Robledo,N.L. and Islas,A.L. (2004) Analysis of non-template-directed nucleotide addition and template switching by DNA polymerase. *Biochemistry,* **43,** 16515-16524.

72. Brissett,N.C., Pitcher,R.S., Juarez,R., Picher,A.J., Green,A.J., Dafforn,T.R., Fox,G.C., Blanco,L. and Doherty,A.J. (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science,* **318,** 456-459.

73. Sekelsky,J.J., Brodsky,M.H. and Burtis,K.C. (2000) DNA repair in drosophila: Insights from the drosophila genome sequence. *J. Cell Biol.,* **150,** F31-6.

74. Zhang,Y., Wu,X., Yuan,F., Xie,Z. and Wang,Z. (2001) Highly frequent frameshift DNA synthesis by human DNA polymerase mu. *Mol. Cell. Biol.,* **21,** 7995-8006.

75. Capp,J.P., Boudsocq,F., Bertrand,P., Laroche-Clary,A., Pourquier,P., Lopez,B.S., Cazaux,C., Hoffmann,J.S. and Canitrot,Y. (2006) The DNA polymerase lambda is required for the repair of non-compatible DNA double strand breaks by NHEJ in mammalian cells. *Nucleic Acids Res.,* **34,** 2998-3007.

76. Yu,A.M. and McVey,M. (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res.,* **38,** 5706-5717.

77. Chan,S.H., Yu,A.M. and McVey,M. (2010) Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in drosophila. *PLoS Genet.,* **6,** e1001005.

78. Liang,L., Deng,L., Nguyen,S.C., Zhao,X., Maulion,C.D., Shao,C. and Tischfield,J.A. (2008) Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Res.,* **36,** 3297-3310.

79. Mukherjee,S., LaFave,M.C. and Sekelsky,J. (2009) DNA damage responses in drosophila nbs mutants with reduced or altered NBS function. *DNA Repair (Amst),* **8,** 803-812.

80. Guirouilh-Barbat,J., Huck,S., Bertrand,P., Pirzio,L., Desmaze,C., Sabatier,L. and Lopez,B.S. (2004) Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol. Cell,* **14,** 611-623.

81. Guirouilh-Barbat,J., Huck,S. and Lopez,B.S. (2008) S-phase progression stimulates both the mutagenic KU-independent pathway and mutagenic processing of KU-dependent intermediates, for nonhomologous end joining. *Oncogene,* **27,** 1726-1736.

82. Glover,L., Jun,J. and Horn,D. (2011) Microhomology-mediated deletion and gene conversion in african trypanosomes. *Nucleic Acids Res.,* **39,** 1372-1380.

83. Bentley,J., L'Hote,C., Platt,F., Hurst,C.D., Lowery,J., Taylor,C., Sak,S.C., Harnden,P., Knowles,M.A. and Kiltie,A.E. (2009) Papillary and muscle invasive bladder tumors with distinct genomic stability profiles have different DNA repair fidelity and KU DNA-binding activities. *Genes Chromosomes Cancer,* **48,** 310-321.

84. Rockwood,L.D., Nussenzweig,A. and Janz,S. (2003) Paradoxical decrease in mutant frequencies and chromosomal rearrangements in a transgenic lacZ reporter gene in Ku80 null mice deficient in DNA double strand break repair. *Mutat. Res.,* **529,** 51-58.

85. Kabotyanski,E.B., Gomelsky,L., Han,J.O., Stamato,T.D. and Roth,D.B. (1998) Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic Acids Res.,* **26,** 5333-5342.

86. Pan-Hammarstrom,Q., Jones,A.M., Lahdesmaki,A., Zhou,W., Gatti,R.A., Hammarstrom,L., Gennery,A.R. and Ehrenstein,M.R. (2005) Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. *J. Exp. Med.,* **201,** 189-194.

87. Bogue,M.A., Wang,C., Zhu,C. and Roth,D.B. (1997) V(D)J recombination in Ku86-deficient mice: Distinct effects on coding, signal, and hybrid joint formation. *Immunity,* **7,** 37-47.

88. Paull,T.T. and Lee,J.H. (2005) The Mre11/Rad50/Nbs1 complex and its role as a DNA double-strand break sensor for ATM. *Cell. Cycle,* **4,** 737-740.

89. Kuhfittig-Kulle,S., Feldmann,E., Odersky,A., Kuliczkowska,A., Goedecke,W., Eggert,A. and Pfeiffer,P. (2007) The mutagenic potential of non-homologous end joining in the absence of the NHEJ core factors Ku70/80, DNA-PKcs and XRCC4-LigIV. *Mutagenesis,* **22,** 217-233.

90. Scuric,Z., Chan,C.Y., Hafer,K. and Schiestl,R.H. (2009) Ionizing radiation induces microhomology-mediated end joining in trans in yeast and mammalian cells. *Radiat. Res.,* **171,** 454-463.

91. Chan,C.Y., Kiechle,M., Manivasakam,P. and Schiestl,R.H. (2007) Ionizing radiation and restriction enzymes induce microhomology-mediated illegitimate recombination in saccharomyces cerevisiae. *Nucleic Acids Res.,* **35,** 5051-5059.

92. Feeney,A.J. (1992) Predominance of VH-D-JH junctions occurring at sites of short sequence homology results in limited junctional diversity in neonatal antibodies. *J. Immunol.,* **149,** 222-229.

93. Seluanov,A., Mittelman,D., Pereira-Smith,O.M., Wilson,J.H. and Gorbunova,V. (2004) DNA end joining becomes less efficient and more error-prone during cellular senescence. *Proc. Natl. Acad. Sci. U. S. A.,* **101,** 7624-7629.

94. Barjesteh van Waalwijk van Doorn-Khosrovani,S., Janssen,J., Maas,L.M., Godschalk,R.W., Nijhuis,J.G. and van Schooten,F.J. (2007) Dietary flavonoids induce MLL translocations in primary human CD34+ cells. *Carcinogenesis,* **28,** 1703-1709.

References For Chapter 2

1. Lehoczky P, McHugh PJ, Chovanec M (2007) DNA interstrand cross-link repair in Saccharomyces cerevisiae. FEMS Microbiol Rev 31: 109-133.

2. Li X, Heyer WD (2008) Homologous recombination in DNA repair and DNA damage tolerance. Cell Res 18: 99-113.

3. Pardo B, Gomez-Gonzalez B, Aguilera A (2009) DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci 66: 1039-1056.

4. Nohmi T, Masumura K (2005) Molecular nature of intrachromosomal deletions and base substitutions induced by environmental mutagens. Environ Mol Mutagen 45: 150-161.

5. Raschle M, Knipscheer P, Enoiu M, Angelov T, Sun J, et al. (2008) Mechanism of replication-coupled DNA interstrand crosslink repair. Cell 134: 969-980.

6. Weterings E, Chen DJ (2008) The endless tale of non-homologous end-joining. Cell Res 18: 114-124.

7. Lieber MR, Lu H, Gu J, Schwarz K (2008) Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system. Cell Res 18: 125-133.

8. McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24: 529-538.

9. Simsek D, Jasin M (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat Struct Mol Biol 17: 410-416.

10. Rass E, Grabarz A, Plo I, Gautier J, Bertrand P, et al. (2009) Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat Struct Mol Biol 16: 819-824.

11. Ma JL, Kim EM, Haber JE, Lee SE (2003) Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol Cell Biol 23: 8820-8828.

12. Weinstock DM, Brunet E, Jasin M (2007) Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. Nat Cell Biol 9: 978-981.

13. Zhu C, Mills KD, Ferguson DO, Lee C, Manis J, et al. (2002) Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. Cell 109: 811-821.

14. Bentley J, Diggle CP, Harnden P, Knowles MA, Kiltie AE (2004) DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. Nucleic Acids Res 32: 5249-5259.

15. Mattarucchi E, Guerini V, Rambaldi A, Campiotti L, Venco A, et al. (2008) Microhomologies and interspersed repeat elements at genomic breakpoints in chronic myeloid leukemia. Genes Chromosomes Cancer 47: 625-632.

16. Corneo B, Wendland RL, Deriano L, Cui X, Klein IA, et al. (2007) Rag mutations reveal robust alternative end joining. Nature 449: 483-486.

17. McVey M, Radut D, Sekelsky JJ (2004) End-joining repair of double-strand breaks in Drosophila melanogaster is largely DNA ligase IV independent. Genetics 168: 2067-2076.

18. Sekelsky JJ, Brodsky MH, Burtis KC (2000) DNA repair in Drosophila: insights from the Drosophila genome sequence. J Cell Biol 150: F31-36.

19. Takasu-Ishikawa E, Yoshihara M, Hotta Y (1992) Extra sequences found at P element excision sites in Drosophila melanogaster. Mol Gen Genet 232: 17-23.

20. Ducau J, Bregliano JC, de La Roche Saint-Andre C (2000) Gamma-irradiation stimulates homology-directed DNA double-strand break repair in Drosophila embryo. Mutat Res 460: 69-80.

21. Welzel N, Le T, Marculescu R, Mitterbauer G, Chott A, et al. (2001) Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. Cancer Res 61: 1629-1636.

22. Jager U, Bocskor S, Le T, Mitterbauer G, Bolz I, et al. (2000) Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. Blood 95: 3520-3529.

23. Denny CT, Hollis GF, Magrath IT, Kirsch IR (1985) Burkitt lymphoma cell line carrying a variant translocation creates new DNA at the breakpoint and violates the hierarchy of immunoglobulin gene rearrangement. Mol Cell Biol 5: 3199-3207.

24. Verkaik NS, Esveldt-van Lange RE, van Heemst D, Bruggenwirth HT, Hoeijmakers JH, et al. (2002) Different types of V(D)J recombination and end-joining defects in DNA double-strand break repair mutant mammalian cells. Eur J Immunol 32: 701-709.

25. Lundberg R, Mavinakere M, Campbell C (2001) Deficient DNA end joining activity in extracts from fanconi anemia fibroblasts. J Biol Chem 276: 9543-9549.

26. Boyd JB, Sakaguchi K, Harris PV (1990) mus308 mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease. Genetics 125: 813-819.

27. Harris PV, Mazina OM, Leonhardt EA, Case RB, Boyd JB, et al. (1996) Molecular cloning of *Drosophila mus308*, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes. MCB 16: 5764-5771.

28. Muzzini DM, Plevani P, Boulton SJ, Cassata G, Marini F (2008) Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. DNA Repair (Amst) 7: 941-950.

29. Inagaki S, Suzuki T, Ohto MA, Urawa H, Horiuchi T, et al. (2006) Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. Plant Cell 18: 879-892.

30. Seki M, Marini F, Wood RD (2003) POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. Nucleic Acids Res 31: 6117-6126.

31. Shima N, Munroe RJ, Schimenti JC (2004) The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. Mol Cell Biol 24: 10381-10389.

32. Zan H, Shima N, Xu Z, Al-Qahtani A, Evinger Iii AJ, et al. (2005) The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. Embo J 24: 3757-3769.

33. Masuda K, Ouchida R, Hikida M, Nakayama M, Ohara O, et al. (2006) Absence of DNA polymerase theta results in decreased somatic hypermutation frequency and altered mutation patterns in Ig genes. DNA Repair (Amst) 5: 1384-1391.

34. Seki M, Masutani C, Yang LW, Schuffert A, Iwai S, et al. (2004) High-efficiency bypass of DNA damage by human DNA polymerase Q. Embo J 23: 4484-4494.

35. Yoshimura M, Kohzaki M, Nakamura J, Asagoshi K, Sonoda E, et al. (2006) Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. Mol Cell 24: 115-125.

36. Prasad R, Longley MJ, Sharief FS, Hou EW, Copeland WC, et al. (2009) Human DNA polymerase {theta} possesses 5'-dRP lyase activity and functions in single-nucleotide base excision repair in vitro. Nucleic Acids Res. 37: 1868-1877.

37. Kawamura K, Bahar R, Seimiya M, Chiyo M, Wada A, et al. (2004) DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. Int J Cancer 109: 9-16.

38. Pang M, McConnell M, Fisher PA (2005) The Drosophila mus 308 gene product, implicated in tolerance of DNA interstrand crosslinks, is a nuclear protein found in both ovaries and embryos. DNA Repair (Amst) 4: 971-982.

39. Arana ME, Seki M, Wood RD, Rogozin IB, Kunkel TA (2008) Low-fidelity DNA synthesis by human DNA polymerase theta. Nucleic Acids Res 36: 3847-3856.

40. Boyd JB, Golino MD, Shaw KES, Osgood CJ, Green MM (1981) Third-chromosome mutagen-sensitive mutants of *Drosophila melanogaster*. Genetics 97: 607-623.

41. Laurencon A, Orme CM, Peters HK, Boulton CL, Vladar EK, et al. (2004) A large-scale screen for mutagen-sensitive loci in Drosophila. Genetics 167: 217-231.

42. Staeva-Vieira E, Yoo S, Lehmann R (2003) An essential role of DmRad51/SpnA in DNA repair and meiotic checkpoint control. Embo J 22: 5863-5874.

43. Adams MD, McVey M, Sekelsky JJ (2003) Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. Science 299: 265-267.

44. Beall EL, Rio DC (1997) Drosophila P-element transposase is a novel site-specific endonuclease. Genes Dev 11: 2137-2151.

45. McVey M, Adams M, Staeva-Vieira E, Sekelsky JJ (2004) Evidence for multiple cycles of strand invasion during repair of double-strand gaps in Drosophila. Genetics 167: 699-705.

46. Johnson-Schlitz DM, Flores C, Engels WR (2007) Multiple-pathway analysis of double-strand break repair mutations in Drosophila. PLoS Genet 3: e50.

47. Wei DS, Rong YS (2007) A genetic screen for DNA double-strand break repair mutations in Drosophila. Genetics 177: 63-77.

48. McVey M, Larocque JR, Adams MD, Sekelsky JJ (2004) Formation of deletions during double-strand break repair in Drosophila DmBlm mutants occurs after strand invasion. Proc Natl Acad Sci U S A 101: 15694-15699.

49. Srivastava A, Simmonds AJ, Garg A, Fossheim L, Campbell SD, et al. (2004) Molecular and functional analysis of scalloped recessive lethal alleles in Drosophila melanogaster. Genetics 166: 1833-1843.

50. Preston CR, Flores CC, Engels WR (2006) Differential usage of alternative pathways of double-strand break repair in Drosophila. Genetics 172: 1055-1068.

51. Weinstock DM, Nakanishi K, Helgadottir HR, Jasin M (2006) Assaying double-strand break repair pathway choice in mammalian cells using a targeted endonuclease or the RAG recombinase. Methods Enzymol 409: 524-540.

52. Rong YS, Golic KG (2003) The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in Drosophila. Genetics 165: 1831-1842.

53. Yu AM, McVey M (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res. 2010 May 11. [Epub ahead of print].

54. Lee K, Lee SE (2007) Saccharomyces cerevisiae Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. Genetics 176: 2003-2014.

55. Decottignies A (2007) Microhomology-mediated end joining in fission yeast is repressed by pku70 and relies on genes involved in homologous recombination. Genetics 176: 1403-1415.

56. Taylor EM, Cecillon SM, Bonis A, Chapman JR, Povirk LF, et al. (2010) The Mre11/Rad50/Nbs1 complex functions in resection-based DNA end joining in Xenopus laevis. Nucleic Acids Res 38: 441-454.

57. Dinkelmann M, Spehalski E, Stoneham T, Buis J, Wu Y, et al. (2009) Multiple functions of MRN in end-joining pathways during isotype class switching. Nat Struct Mol Biol 16: 808-813.

58. Xie A, Kwok A, Scully R (2009) Role of mammalian Mre11 in classical and alternative nonhomologous end joining. Nat Struct Mol Biol 16: 814-818.

59. Deriano L, Stracker TH, Baker A, Petrini JH, Roth DB (2009) Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. Mol Cell 34: 13-25.

60. Nassif N, Penney J, Pal S, Engels WR, Gloor GB (1994) Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol 14: 1613-1625.

61. Yusufzai T, Kadonaga JT (2008) HARP is an ATP-driven annealing helicase. Science 322: 748-750.

62. Brissett NC, Pitcher RS, Juarez R, Picher AJ, Green AJ, et al. (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. Science 318: 456-459.

63. Audebert M, Salles B, Calsou P (2004) Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J Biol Chem 279: 55117-55126.

64. Wang H, Rosidi B, Perrault R, Wang M, Zhang L, et al. (2005) DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. Cancer Res 65: 4020-4030.

65. Grossmann KF, Ward AM, Matkovic ME, Folias AE, Moses RE (2001) S. cerevisiae has three pathways for DNA interstrand crosslink repair. Mutat Res 487: 73-83.

66. Sarkar S, Davies AA, Ulrich HD, McHugh PJ (2006) DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. Embo J 25: 1285-1294.

67. Jonnalagadda VS, Matsuguchi T, Engelward BP (2005) Interstrand crosslink-induced homologous recombination carries an increased risk of deletions and insertions. DNA Repair (Amst) 4: 594-605.

68. Liu Y, Nairn RS, Vasquez KM (2008) Processing of triplex-directed psoralen DNA interstrand crosslinks by recombination mechanisms. Nucleic Acids Res 36: 4680-4688.

69. Zheng H, Wang X, Legerski RJ, Glazer PM, Li L (2006) Repair of DNA interstrand cross-links: interactions between homology-dependent and homology-independent pathways. DNA Repair (Amst) 5: 566-574.

70. Scaringe WA, Li K, Gu D, Gonzalez KD, Chen Z, et al. (2008) Somatic microindels in human cancer: the insertions are highly error-prone and derive from nearby but not adjacent sense and antisense templates. Hum Mol Genet 17: 2910-2918.

71. Kusano K, Johnson-Schlitz DM, Engels WR (2001) Sterility of Drosophila with mutations in the Bloom syndrome gene--complementation by Ku70. Science 291: 2600-2602.

72. Windhofer F, Krause S, Hader C, Schulz WA, Florl AR (2008) Distinctive differences in DNA double-strand break repair between normal urothelial and urothelial carcinoma cells. Mutat Res 638: 56-65.

73. Marculescu R, Vanura K, Montpellier B, Roulland S, Le T, et al. (2006) Recombinase, chromosomal translocations and lymphoid neoplasia: Targeting mistakes and repair failures. DNA Repair (Amst) 5: 1246-1258.

74. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, et al. (2008) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803-806.

75. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, et al. (2008) An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798-802.

76. Koundakjian EJ, Cowan DM, Hardy RW, Becker AH (2004) The Zuker Collection: A Resource for the Analysis of Autosomal Gene Function in Drosophila melanogaster. Genetics 167: 203-206.

77. McVey M, Andersen SL, Broze Y, Sekelsky J (2007) Multiple functions of Drosophila BLM helicase in maintenance of genome stability. Genetics 176: 1979-1992.

78. Kurkulos M, Weinberg JM, Roy D, Mount SM (1994) *P* element-mediated *in vivo* deletion analysis of *white-apricot*: deletions between direct repeats are strongly favored. Genetics 136: 1001-1011.

79. Zider A, Paumard-Rigal S, Frouin I, Silber J (1998) The vestigial gene of Drosophila melanogaster is involved in the formation of the peripheral nervous system: genetic interactions with the scute gene. J Neurogenet 12: 87-99.

80. Gloor GB, Preston CR, Johnson-Schlitz DM, Nassif NA, Phillis RW, et al. (1993) Type I repressors of P element mobility. Genetics 135: 81-95.

References For Chapter 3

1. McVey,M., Radut,D. and Sekelsky,J.J. (2004) End-joining repair of double-strand breaks in drosophila melanogaster is largely DNA ligase IV independent. Genetics, 168, 2067-2076.

2. Johnson-Schlitz,D.M., Flores,C. and Engels,W.R. (2007) Multiple-pathway analysis of double-strand break repair mutations in drosophila. PLoS Genet., 3, e50.

3. Wei,D.S. and Rong,Y.S. (2007) A genetic screen for DNA double-strand break repair mutations in drosophila. Genetics, 177, 63-77.

4. Rong,Y.S. and Golic,K.G. (2003) The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in drosophila. Genetics, 165, 1831-1842.

5. Bozas,A., Beumer,K.J., Trautman,J.K. and Carroll,D. (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in drosophila. Genetics, 182, 641-651.

6. Chan,S.H., Yu,A.M. and McVey,M. (2010) Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in drosophila. PLoS Genet., 6, e1001005.

7. Sakaguchi,K., Harris,P.V., van Kuyk,R., Singson,A. and Boyd,J.B. (1990) A mitochondrial nuclease is modified in drosophila mutants (mus308) that are hypersensitive to DNA crosslinking agents. Mol. Gen. Genet., 224, 333-340.

8. Boyd,J.B., Sakaguchi,K. and Harris,P.V. (1990) mus308 mutants of drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease. Genetics, 125, 813-819.

9. McVey,M., Adams,M., Staeva-Vieira,E. and Sekelsky,J.J. (2004) Evidence for multiple cycles of strand invasion during repair of double-strand gaps in drosophila. Genetics, 167, 699-705.

10. Ducau,J., Bregliano,J.C. and de La Roche Saint-Andre,C. (2000) Gamma-irradiation stimulates homology-directed DNA double-strand break repair in drosophila embryo. Mutat. Res., 460, 69-80.

11. Howlett,N.G., Scuric,Z., D'Andrea,A.D. and Schiestl,R.H. (2006) Impaired DNA double strand break repair in cells from nijmegen breakage syndrome patients. DNA Repair (Amst), 5, 251-257.

12. Rahal,E.A., Henricksen,L.A., Li,Y., Williams,R.S., Tainer,J.A. and Dixon,K. (2010) ATM regulates Mre11-dependent DNA end-degradation and microhomology-mediated end joining. Cell. Cycle, 9, 2866-2877.

13. Xie,A., Kwok,A. and Scully,R. (2009) Role of mammalian Mre11 in classical and alternative nonhomologous end joining. Nat. Struct. Mol. Biol., 16, 814-818.

14. Mukherjee,S., LaFave,M.C. and Sekelsky,J. (2009) DNA damage responses in drosophila nbs mutants with reduced or altered NBS function. DNA Repair (Amst), 8, 803-812.

15. Williams,R.S., Moncalian,G., Williams,J.S., Yamada,Y., Limbo,O., Shin,D.S., Groocock,L.M., Cahill,D., Hitomi,C., Guenther,G., et al. (2008) Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. Cell, 135, 97-109.

16. Kotnis,A., Du,L., Liu,C., Popov,S.W. and Pan-Hammarstrom,Q. (2009) Non-homologous end joining in class switch recombination: The beginning of the end. Philos. Trans. R. Soc. Lond. B. Biol. Sci., 364, 653-665.

17. Zhang,Y. and Jasin,M. (2011) An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat. Struct. Mol. Biol., 18, 80-84.

18. Bibikova,M., Golic,M., Golic,K.G. and Carroll,D. (2002) Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. Genetics, 161, 1169-1175.

19. Sekelsky,J.J., Brodsky,M.H. and Burtis,K.C. (2000) DNA repair in drosophila: Insights from the drosophila genome sequence. J. Cell Biol., 150, F31-6.

20. Roth,D.B., Chang,X.B. and Wilson,J.H. (1989) Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. Mol. Cell. Biol., 9, 3049-3057.

21. Varga,T. and Aplan,P.D. (2005) Chromosomal aberrations induced by double strand DNA breaks. DNA Repair (Amst), 4, 1038-1046.

22. Scaringe,W.A., Li,K., Gu,D., Gonzalez,K.D., Chen,Z., Hill,K.A. and Sommer,S.S. (2008) Somatic microindels in human cancer: The insertions are highly error-prone and derive from nearby but not adjacent sense and antisense templates. Hum. Mol. Genet., 17, 2910-2918.

23. Pan-Hammarstrom,Q., Jones,A.M., Lahdesmaki,A., Zhou,W., Gatti,R.A., Hammarstrom,L., Gennery,A.R. and Ehrenstein,M.R. (2005) Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. J. Exp. Med., 201, 189-194.

24. Secretan,M.B., Scuric,Z., Oshima,J., Bishop,A.J., Howlett,N.G., Yau,D. and Schiestl,R.H. (2004) Effect of Ku86 and DNA-PKcs deficiency on non-homologous end-joining and homologous recombination using a transient transfection assay. Mutat. Res., 554, 351-364.

25. Kawamura,K., Bahar,R., Seimiya,M., Chiyo,M., Wada,A., Okada,S., Hatano,M., Tokuhisa,T., Kimura,H., Watanabe,S., et al. (2004) DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. Int. J. Cancer, 109, 9-16.

26. Jager,U., Bocskor,S., Le,T., Mitterbauer,G., Bolz,I., Chott,A., Kneba,M., Mannhalter,C. and Nadel,B. (2000) Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: Novel insights into the mechanism of t(14;18) translocation. Blood, 95, 3520-3529.

27. Welzel,N., Le,T., Marculescu,R., Mitterbauer,G., Chott,A., Pott,C., Kneba,M., Du,M.Q., Kusec,R., Drach,J., et al. (2001) Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: Implications for the mechanism of translocation and the origin of mantle cell lymphoma. Cancer Res., 61, 1629-1636.

28. Murga Penas,E.M., Callet-Bauchu,E., Ye,H., Gazzo,S., Berger,F., Schilling,G., Albert-Konetzny,N., Vettorazzi,E., Salles,G., Wlodarska,I., et al. (2009) The t(14;18)(q32;q21)/IGH-MALT1 translocation in MALT lymphomas contains templated nucleotide insertions and a major breakpoint region similar to follicular and mantle cell lymphoma. Blood, .

29. Hogg,M., Seki,M., Wood,R.D., Doublie,S. and Wallace,S.S. (2011) Lesion bypass activity of DNA polymerase theta (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. J. Mol. Biol., 405, 642-652.

30. Marini,F., Kim,N., Schuffert,A. and Wood,R.D. (2003) POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. J. Biol. Chem., 278, 32014-32019.

31. Marini,F. and Wood,R.D. (2002) A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. J. Biol. Chem., 277, 8716-8723.

32. McCaffrey,R., St Johnston,D. and Gonzalez-Reyes,A. (2006) Drosophila mus301/spindle-C encodes a helicase with an essential role in double-strand DNA break repair and meiotic progression. Genetics, 174, 1273-1285.

33. Moldovan,G.L., Madhavan,M.V., Mirchandani,K.D., McCaffrey,R.M., Vinciguerra,P. and D'Andrea,A.D. (2010) DNA polymerase POLN participates in cross-link repair and homologous recombination. Mol. Cell. Biol., 30, 1088-1096.

34. Seki,M., Marini,F. and Wood,R.D. (2003) POLQ (pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. Nucleic Acids Res., 31, 6117-6126.

35. Ripley,L.S. (1982) Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. Proc. Natl. Acad. Sci. U. S. A., 79, 4128-4132.

36. Greenblatt,M.S., Grollman,A.P. and Harris,C.C. (1996) Deletions and insertions in the p53 tumor suppressor gene in human cancers: Confirmation of the DNA polymerase slippage/misalignment model. Cancer Res., 56, 2130-2136.

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

38. Brissett,N.C., Pitcher,R.S., Juarez,R., Picher,A.J., Green,A.J., Dafforn,T.R., Fox,G.C., Blanco,L. and Doherty,A.J. (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. Science, 318, 456-459.

39. Wessler,S., Tarpley,A., Purugganan,M., Spell,M. and Okagaki,R. (1990) Filler DNA is associated with spontaneous deletions in maize. Proc. Natl. Acad. Sci. U. S. A., 87, 8731-8735.

40. Lee,J.A., Carvalho,C.M. and Lupski,J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell, 131, 1235-1247.

41. Feschenko,V.V. and Lovett,S.T. (1998) Slipped misalignment mechanisms of deletion formation: Analysis of deletion endpoints. J. Mol. Biol., 276, 559-569.

42. Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol., 5, 2599-2607.

43. Reichel,M., Gillert,E., Breitenlohner,I., Angermuller,S., Fey,G.H., Marschalek,R., Repp,R., Greil,J. and Beck,J.D. (2001) Rapid isolation of chromosomal breakpoints from patients with t(4;11) acute lymphoblastic leukemia: Implications for basic and clinical research. Leukemia, 15, 286-288.

44. Daley,J.M. and Wilson,T.E. (2008) Evidence that base stacking potential in annealed 3' overhangs determines polymerase utilization in yeast nonhomologous end joining. DNA Repair (Amst), 7, 67-76.

45. Iwabuchi,K., Hashimoto,M., Matsui,T., Kurihara,T., Shimizu,H., Adachi,N., Ishiai,M., Yamamoto,K., Tauchi,H., Takata,M., et al. (2006) 53BP1 contributes to survival of cells irradiated with X-ray during G1 without Ku70 or artemis. Genes Cells, 11, 935-948.

46. Guirouilh-Barbat,J., Rass,E., Plo,I., Bertrand,P. and Lopez,B.S. (2007) Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. Proc. Natl. Acad. Sci. U. S. A., 104, 20902-20907.

47. Schulte-Uentrop,L., El-Awady,R.A., Schliecker,L., Willers,H. and Dahm-Daphi,J. (2008) Distinct roles of XRCC4 and Ku80 in non-homologous end-joining of endonuclease- and ionizing radiation-induced DNA double-strand breaks. Nucleic Acids Res., 36, 2561-2569.

48. Gu,J. and Lieber,M.R. (2008) Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining. Genes Dev., 22, 411-415.

49. Lieber,M.R., Lu,H., Gu,J. and Schwarz,K. (2008) Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: Relevance to cancer, aging, and the immune system. Cell Res., 18, 125-133.

50. Preston,C.R., Flores,C.C. and Engels,W.R. (2006) Differential usage of alternative pathways of double-strand break repair in drosophila. Genetics, 172, 1055-1068.

51. Staeva-Vieira,E., Yoo,S. and Lehmann,R. (2003) An essential role of DmRad51/SpnA in DNA repair and meiotic checkpoint control. EMBO J., 22, 5863-5874.

52. Koundakjian,E.J., Cowan,D.M., Hardy,R.W. and Becker,A.H. (2004) The zuker collection: A resource for the analysis of autosomal gene function in drosophila melanogaster. Genetics, 167, 203-206.

53. Oshige,M., Aoyagi,N., Harris,P.V., Burtis,K.C. and Sakaguchi,K. (1999) A new DNA polymerase species from drosophila melanogaster: A probable mus308 gene product. Mutat. Res., 433, 183-192.

54. Gloor,G.B., Preston,C.R., Johnson-Schlitz,D.M., Nassif,N.A., Phillis,R.W., Benz,W.K., Robertson,H.M. and Engels,W.R. (1993) Type I repressors of P element mobility. Genetics, 135, 81-95.

55. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: The european molecular biology open software suite. Trends Genet., 16, 276-277.

References For Chapter 4

1. Windels,P., De Buck,S., Van Bockstaele,E., De Loose,M. and Depicker,A. (2003) T-DNA integration in arabidopsis chromosomes. presence and origin of filler DNA sequences. *Plant Physiol.,* **133,** 2061-2068.

2. Jager,U., Bocskor,S., Le,T., Mitterbauer,G., Bolz,I., Chott,A., Kneba,M., Mannhalter,C. and Nadel,B. (2000) Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: Novel insights into the mechanism of t(14;18) translocation. *Blood,* **95,** 3520-3529.

3. Preston,C.R., Flores,C.C. and Engels,W.R. (2006) Differential usage of alternative pathways of double-strand break repair in drosophila. *Genetics,* **172,** 1055-1068.

4. Beumer,K., Bhattacharyya,G., Bibikova,M., Trautman,J.K. and Carroll,D. (2006) Efficient gene targeting in drosophila with zinc-finger nucleases. *Genetics,* **172,** 2391-2403.

5. Bibikova,M., Golic,M., Golic,K.G. and Carroll,D. (2002) Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics,* **161,** 1169-1175.

6. Bozas,A., Beumer,K.J., Trautman,J.K. and Carroll,D. (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in drosophila. *Genetics,* **182,** 641-651.

7. Adams,M.D., McVey,M. and Sekelsky,J.J. (2003) Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science,* **299,** 265-267.

8. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.,* **11,** 636-646.

9. Perrin,A., Buckle,M. and Dujon,B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.,* **12,** 2939-2947.

10. McVey,M., Larocque,J.R., Adams,M.D. and Sekelsky,J.J. (2004) Formation of deletions during double-strand break repair in drosophila DmBlm mutants occurs after strand invasion. *Proc. Natl. Acad. Sci. U. S. A.,* **101,** 15694-15699.

11. Simsek,D. and Jasin,M. (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat. Struct. Mol. Biol.,* **17,** 410-416.

12. Weinstock,D.M., Elliott,B. and Jasin,M. (2006) A model of oncogenic rearrangements: Differences between chromosomal translocation mechanisms and simple double-strand break repair. *Blood,* **107,** 777-780.

13. Bannister,L.A., Waldman,B.C. and Waldman,A.S. (2004) Modulation of error-prone double-strand break repair in mammalian chromosomes by DNA mismatch repair protein Mlh1. *DNA Repair (Amst),* **3,** 465-474.

14. Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.,* **5,** 2599-2607.

15. Honma,M., Sakuraba,M., Koizumi,T., Takashima,Y., Sakamoto,H. and Hayashi,M. (2007) Non-homologous end-joining for repairing I-SceI-induced DNA double strand breaks in human cells. *DNA Repair (Amst),* **6,** 781-788.

16. Ragg,H. (2011) Intron creation and DNA repair. *Cell Mol. Life Sci.,* **68,** 235-242.

17. Bibikova,M., Golic,M., Golic,K.G. and Carroll,D. (2002) Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics,* **161,** 1169-1175.

18. Bozas,A., Beumer,K.J., Trautman,J.K. and Carroll,D. (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in drosophila. *Genetics,* **182,** 641-651.

References for Chapter 5

1. Huo,H., Wang,X. and Stojkovic,V. (2009) Repeats identification using improved suffix trees. *Int. J. Comput. Biol. Drug Des.,* **2,** 264-277.

2. Gusfield,D. (1997) Algorithms on Strings, Trees, and Sequences : Computer Science and Computational Biology. Cambridge University Press, Cambridge England ; New York.

3. Cock,P.J., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B., et al. (2009) Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics,* **25,** 1422-1423.

References for Appendix A

1. Ducau,J., Bregliano,J.C. and de La Roche Saint-Andre,C. (2000) Gamma-irradiation stimulates homology-directed DNA double-strand break repair in drosophila embryo. *Mutat. Res.,* **460,** 69-80.

2. Preston,C.R., Flores,C.C. and Engels,W.R. (2006) Differential usage of alternative pathways of double-strand break repair in drosophila. *Genetics,* **172,** 1055-1068.

3. Steller,H. and Pirrotta,V. (1985) Fate of DNA injected into early drosophila embryos. *Dev. Biol.,* **109,** 54-62.