

**UNDERSTANDING THE MECHANISMS UNDERLYING DSB
REPAIR-INDUCED MUTAGENESIS AT DISTANT LOCI IN YEAST**

A Dissertation
Presented to
The Academic Faculty

by

Natalie Saini

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
in the
School of Biology

Georgia Institute of Technology
May, 2014

Copyright © 2014 by Natalie Saini

**UNDERSTANDING THE MECHANISMS UNDERLYING DSB
REPAIR-INDUCED MUTAGENESIS AT DISTANT LOCI**

Approved by:

Dr. Kirill S. Lobachev, Advisor
School of Biology
Georgia Institute of Technology

Dr. Yuhong Fan
School of Biology
Georgia Institute of Technology

Dr. Yury Chernoff
School of Biology
Georgia Institute of Technology

Dr. Milo Fasken
School of Medicine
Emory University

Dr. Francesca Storici
School of Biology
Georgia Institute of Technology

Date Approved: 03/06/2014

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude towards my advisor Dr. Kirill Lobachev. I have very high regard for him as he introduced me to the stimulating field of DNA repair. My time in Georgia Institute of Technology would not have been as successful and memorable without his ability to motivate and encourage his students to do their best. He has and will continue to be a positive influence in my career.

I am also highly thankful to all my committee members, Dr. Yuri Chernoff, Dr. Francesca Storici, Dr. Yuhong Fan and Dr. Milo Fasken for their valuable suggestions during the course of Ph.D.

I would like to thank all the members of my lab. Their constant help and support were extremely valuable in both my research and my life. I hold their friendship in high esteem and will be forever grateful to them.

I would like to extend a special thanks to my husband for his unflinching love and support, without which I would not be able to be as successful as I am today. Lastly I would like to thank my parents and my family for always being there for me.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF SYMBOLS AND ABBREVIATIONS	ix
SUMMARY	xi
 <u>CHAPTER</u>	
1 Introduction and literature review	1
1.1 Double-strand breaks cause mutagenesis of the flanking regions	1
1.2 Secondary structure-forming DNA sequences are an intrinsic source of DSBs	4
1.2.1 Hairpin and cruciform-forming long inverted repeats	4
1.2.2 Triplex forming GAA/TTC repeats	7
1.3 Overview of the dissertation	9
1.4 References	10
2 Fragile DNA motifs trigger mutagenesis at distant chromosomal loci in <i>Saccharomyces cerevisiae</i>	19
2.1 Abstract	19
2.2 Introduction	20
2.3 Results	24
2.3.1 Experimental system	24
2.3.2 A defect in DNA replication leads to increased <i>Alu</i> -quasipalindrome-induced breakage and mutagenesis at the <i>CAN1</i> locus	25
2.3.3 A 1.3 kb perfect palindrome induces mutagenesis at the <i>CAN1</i> locus even in the wild-type strains	25

2.3.4	Mutagenesis by inverted repeats depends on the distance of the reporter from the DSB site and on the activity of the Sae2 protein	31
2.3.5	Increase in mutagenesis observed in replication-deficient and -proficient strains is mostly attributed to the activity of Pol ζ translesion polymerase	34
2.3.6	GAA/TTC fragile motif also induces mutagenesis at distant chromosomal loci that is partly Pol ζ dependent	37
2.3.6	GAA/TTC fragile motif also induces mutagenesis at distant chromosomal loci that is partly Pol ζ dependent	37
2.4	Discussion	38
2.5	Materials and Methods	44
2.5.1	Yeast Strains	44
2.5.2	GCR and mutation rates estimations	45
2.5.3	DSB detection and quantification	45
2.5.4	Sequence analysis of mutations at <i>CANI</i>	46
2.5.5	Estimation of Pol3 and Rfa2 expression	47
2.6	Acknowledgements	47
2.7	References	47
3	Migrating bubble during break-induced replication drives conservative DNA synthesis	53
3.1	Introduction, Results and Discussion	53
3.1	Materials and Methods	65
3.2.1	Media and strains	65
3.2.2	Analysis of BIR efficiency	67
3.2.3	2D analysis of molecular intermediates of BIR	67
3.2.4	DNA combing and fluorescent <i>in situ</i> hybridization	70
3.2.5	Mutagenesis associated with DSB repair	72
3.2.6	Analysis of BIR-induced Lys ⁺ mutants	74

3.2.7 Analysis of mutation spectra of <i>ura3-29</i> Ura ⁺ reversions	74
3.3 References	74
4 Conclusions and future directions	78
4.1 Conclusions	78
4.2 Future directions	82
4.2.1 Do other translesion polymerases function during DSB repair -associated mutagenesis?	82
4.2.2 Determining the proteins responsible for BIR progression.	83
4.2.3 Can BIR occur in the S-phase?	83
4.3 References	84
APPENDIX A: Supplementary information for chapter 2	87
APPENDIX B: Supplementary information for chapter 3	99
PUBLICATIONS	110

LIST OF TABLES

	Page
Table 2.1: Pol ζ - and Sae2-dependent mutagenesis by <i>Alu</i> -quasi-palindrome in replication mutants	28
Table 2.2: Pol ζ - and Sae2-dependent mutagenesis by <i>IS50</i> -perfect palindrome	31
Table 2.3: Mutagenesis by fragile <i>Alu</i> -IRs depends on the distance of reporter from the DSB site	33
Table 2.4: Mutation spectra in <i>CANI</i> reporter	36
Table 2.5: Mutagenesis in <i>CANI</i> reporter stimulated by GAA/TTC repeats	38
Table A.1: Sequences of mutations analyzed in <i>CANI</i> in wild-type strain containing inverted repeats	87
Table A.2: Sequences of mutations analyzed in <i>CANI</i> in <i>pol3-P664L</i> mutant strain carrying inverted repeats	89
Table A.3: Sequences of mutations analyzed in <i>CANI</i> in <i>pol3-P664L</i> $\Delta rev3$ mutant strain carrying inverted repeats	91
Table A.4: Sequences of mutations analyzed in <i>CANI</i> in <i>pol3-P664L</i> mutant strain carrying no inverted repeats	93
Table A.5: Sequences of mutations analyzed in <i>CANI</i> in wild-type strain carrying <i>IS50</i> -perfect palindrome	96
Table B.1: The rate of spontaneous and DSB-associated Ura ⁺ mutations	99
Table B.2: The rate of DSB-associated Lys ⁺ mutations (top), and the rate of spontaneous Lys ⁺ mutations (bottom)	100
Table B.3: Strains used in this study	101

LIST OF FIGURES

	Page
Figure 2.1: Experimental system to study fragile motif-induced mutagenesis	25
Figure 2.2: Analysis of protein levels of Pol3 and Rfa2 in the wild-type and tetracycline downregulatable strains	27
Figure 2.3: Inverted repeat and GAA/TTC-induced DSB detection in wild-type and mutant strains	30
Figure 2.4: Model for mutagenesis induced by <i>Alu</i> -IRs and GAA/TTC repeats	43
Figure 3.1: The mode of DNA synthesis during BIR	55
Figure 3.2: BIR-induced mutations	57
Figure 3.3: DNA synthesis during BIR is conservative	59
Figure 3.4: Molecular intermediates of BIR	62
Figure A.1: Sensitivities of the replication-deficient strains to DNA-damaging agents	98
Figure B.1: BIR efficiency during molecular combing analysis of molecular intermediates of BIR	102
Figure B.2: Analysis of molecular mechanism and mutagenesis associated with BIR	103
Figure B.3: Molecular outcomes of BIR	104
Figure B.4: Conservative DNA synthesis associated with BIR	106
Figure B.5: BIR kinetics during 2D analysis of molecular intermediates of BIR	107
Figure B.6: The structure of molecular intermediates of BIR	108
Figure B.7: Molecular intermediates of BIR	109

LIST OF SYMBOLS AND ABBREVIATIONS

<i>Alu</i> -IRs	Inverted <i>Alu</i> Repeats
BIR	Break induced replication
BrdU	Bromodeoxyuridine
CHEF	Contour-clamped Homogeneous Electric Field
ChIP	Chromatin Immunoprecipitation
CST	Cdc13-Stn1-Ten1
DSB	Double Strand Break
GC	Gene Conversion
GCR	Gross Chromosomal Rearrangement
<i>IS50</i> -PAL	Palindromic <i>IS50</i> repeats
MMS	Methyl-methane Sulfonate
MRX	Mre11-Rad50-Xrs2
PCR	Polymerase Chain Reaction
SDSA	Synthesis Dependent Strand Annealing
SSA	Single-Strand Annealing
ssDNA	Single-stranded DNA

TD	Telomere-distal
TLS	Translesion Synthesis
TP	Telomere-proximal
2D gel	Two-dimensional Gel Electrophoresis
5-FOA	5-Fluoroorotic Acid

SUMMARY

Increased mutagenesis is a hallmark of cancers. On the other hand, this can trigger the generation of polymorphisms and lead to evolution. Lately, it has become clear that one of the major sources of increased mutation rates in the genome is chromosomal break formation and repair.

A variety of factors can contribute to the generation of breaks in the genome. A paradoxical source of breaks is the sequence composition of the genomic DNA itself. Eukaryotic and prokaryotic genomes contain sequence motifs capable of adopting secondary structures often found to be potent inducers of double strand breaks culminating into rearrangements. These regions are therefore termed fragile sequence motifs. Here, we demonstrate that in addition to being responsible for triggering chromosomal rearrangements, inverted repeats and GAA/TTC repeats are also potent sources of mutagenesis. Repeat-induced mutagenesis extends up to 8 kb on either side of the break point. Remarkably, error-prone repair of the break by Pol ζ reconstitutes the repeats making them a long term source of mutagenesis.

Despite its negative connotations for genome stability, the mechanisms underlying the unstable nature of double strand break repair pathways are not known. Previous studies have demonstrated that break induced replication (BIR), a mechanism employed to repair broken chromosomes with only one repairable end, is highly mutagenic, undergoes frequent template switching and often yields half-crossovers. In the work presented here, we show that the instabilities inherent to BIR can be attributed to its unusual mode of synthesis. We determined that BIR proceeds via a migrating

bubble with long stretches of single-stranded DNA and culminates with conservative inheritance of the newly synthesized DNA.

We propose that the mechanisms described here might be important for generation of repair-associated mutagenesis in higher organisms. Secondary structure forming repeats like inverted repeats have been found to be enriched in cancer cells. These motifs often constitute chromosomal rearrangement hot-spots and demonstrate the phenomenon of kataegis. This study provides a mechanistic insight into how such breakage-prone motifs contribute to hypermutability of cancer genomes.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Double-strand breaks cause mutagenesis of the flanking regions

Mutagenesis is a hallmark of cancers and most human tumors have augmented mutation rates. These mutations include “driver” mutations that are directly related to the carcinogenesis or “passenger” mutations in genes not shown to cause cancer development. Tumors contain, on an average, ~10,000 mutations and up to 200,000 mutations have been identified in glioblastomas and myelomas (Hahn et al., 1999; Mahale et al., 2008). Based on this, the average rate of accumulation of point mutations in most cancers is estimated to be $\sim 10^{-8}$ /bp/cell (Jones et al., 2008). However, estimates of mutation rates in cultured fibroblasts are $\sim 10^{-10}$ /bp/cell (Fox et al., 2013) and the mutator frequencies of normal stem cells is 100X lower (Cervantes et al., 2002; Fox et al., 2013). Taking these data into account it is unclear how tumor cells can accumulate so many mutations within short generation spans.

One source of mutagenesis in cancer cells is double-strand break repair. With the advent of whole-genome sequence analysis, it is becoming more and more evident that the mutational landscape of cancer genomes is not uniform. Cancer genomes often demonstrate clusters of mutations termed “kataegis” (Nik-Zainal et al., 2012). Further analysis of the clusters demonstrates a 100- to 300-fold increase in mutation rates around chromosomal breakpoints across many cancer types (Drier et al., 2013). The predominant mutation around the breakpoints was found to be C-G transversions which

may be a product of damage accumulated in the single-stranded region created around the break due to resection during DSB repair (Degtyareva et al., 2013; Jansen et al., 2006). C-G transversions were also seen within a nucleotide context of TpC surrounding break sites, suggestive of the action of the apolipoprotein B mRNA-editing enzymes (APOBEC1 and APOBEC3s) that act on single-stranded DNA. APOBEC mutation patterns have been found in bladder, cervical, breast, head and neck and lung cancers and can comprise up to 68% of the mutations in certain samples and have been shown to be strand-coordinated (Roberts et al., 2013).

Studies in the model organism *Saccharomyces cerevisiae* showed long and persistent single-stranded DNA arising from resection at a double-strand break (DSB) accrues similar strand-coordinated mutation clusters (Roberts et al., 2012). These studies strongly implicate single-stranded DNA generated during double-strand break repair in kataegis in both humans and model organisms. Elevated mutation rates have been seen in *Salmonella* undergoing transduction with phages identical to the hosts' (Demerec, 1962, 1963). Also, studies from *E. coli* determined that adaptive and DSB induced mutagenesis in the LacZ gene is dependent on RecA, RecBCD and RuvABC (Harris et al., 1994; Harris et al., 1996; He et al., 2006; Ponder et al., 2005). The mutagenesis was surmised to have arisen due to a switch in the key polymerases during repair synthesis from the error-free PolII and PolIII to the erroneous translesion polymerases PolIV and PolV (Hastings et al., 2010). Furthermore, the fidelity of PolII at D-loop substrates was shown to be markedly lower than at normal replication intermediates (Pomerantz et al., 2013). Breaks generated during meiosis have also been shown to contribute to

mutagenesis in both the budding yeast and the fungus *Ascobolus immerses* (Magni, 1963; Paszewski and Surzycki, 1964). In mitotically growing yeast, chromosomal breakage and repair via the presumably error-free pathway of gene conversion was shown to dramatically increase mutations in the regions flanking the break (Esposito and Bruschi, 1993; Hicks et al., 2010; Rattray et al., 2002). Depending on the system, the mutations were found to be dependent on either erroneous activities of the replicative polymerases Pol δ and Pol ϵ or on the action of the translesion polymerase Pol ζ (Hicks et al., 2010; Rattray et al., 2002). Increased mutation rates were also seen during filling in of single-stranded DNA during DSB repair and break-induced replication (BIR) pathways (Burch et al., 2011; Deem et al., 2011; Strathern et al., 1995). The mutagenesis in these systems was predominantly confined to the regions flanking the break site, and very low levels of mutations could be detected elsewhere in the genomes (Burch et al., 2011).

The above clearly demonstrate that chromosomal double-strand breaks and their repair are capable of providing cells, at least in part; a way to bypass the low error rates of normal DNA replication and accumulate clusters of mutations within a few cellular divisions. Considering that tumor cells are frequently associated with replication stalling, chromosomal rearrangements and often even with chromothripsis (shattering and repair of a single chromosome), it is highly likely that error-prone DSB repair is the key mechanism to generate kataegis.

1.2 Secondary structure-forming DNA sequences are an intrinsic source of DSBs

One of the most prevalent sources of single-or double-stranded breaks is DNA replication itself. On an average, cells experience almost 10 breaks with each replication cycle with certain regions of the genome (fragile sites) having a higher predisposition to breakage (Haber, 1999). DNA fragile sites, many of which encompass DNA motifs capable of adopting a variety of non-canonical secondary structures including hairpin, cruciform, G-quadruplex, triplex and Z-DNA, are hot-spots of chromosomal breakage and rearrangements [reviewed in (Mirkin, 2008; Zhao et al., 2010)]. Due to their potential to alter the genome, secondary structure-forming DNA sequences have also been called “At-risk motifs” (Gordenin and Resnick, 1998). In the following sections, two at-risk motifs pertinent to this study will be discussed.

1.2.1 Hairpin and cruciform-forming long inverted repeats

Long inverted repeats (~100bp or more in length each) capable of forming either a hairpin or cruciform structure are a prominent source of genomic instability in both prokaryotes and eukaryotes. Long inverted repeats are often found to be separated by a unique spacer. Perfect inverted repeats that do not have a spacer between them are also referred to as DNA palindromes. Long palindromic sequences are frequently deleted in *E. coli* and closely spaced inverted repeats (<2bp spacer) are extremely rare in the human genome pointing towards their extremely unstable nature (Leach, 1994; Lobachev et al., 2000). Imperfect inverted repeats separated by a spacer, although more stable, are highly recombinagenic and can cause gross chromosomal rearrangements in both eukaryotes and prokaryotes (Darmon et al., 2010; Gordenin et al., 1993; Lemoine et al., 2005; Lobachev et al., 2007; Narayanan et al., 2006). Rearrangements initiated at inverted repeats have

been implicated in a variety of human diseases. Notable examples include recurrent and non-recurrent translocations at palindromic AT-rich repeats in Emanuel syndrome (Gotter et al., 2007; Gotter et al., 2004; Kurahashi et al., 2003; Nimmakayalu et al., 2003; Sheridan et al., 2010); palindrome mediated deletions and insertions in types of $\epsilon\gamma\delta\beta$ thalassemia (Rooks et al., 2012) and formation of isodicentric chromosomes at inverted repeats on the Y chromosome leading to infertility (Carvalho et al., 2011). Inverted repeats have also been found to be responsible for initiating gene amplification events in colon and breast cancers, medulloblastoma and lymphoma (Ford and Fried, 1986; Guenthoer et al., 2012; Mangano et al., 1998; Neiman et al., 2006; Neiman et al., 2008; Tanaka et al., 2007).

The unstable nature of inverted repeats has been attributed to their predisposition to DSB formation. In bacteria, inverted repeats on the chromosome were shown to be attacked by the SbcCD nuclease on the lagging strand (Connelly and Leach, 1996; Cromie et al., 2000; Darmon et al., 2010; Eykelenboom et al., 2008). Similarly, in *Schizosaccharomyces pombe*, the Mre11-Rad50-Xrs2 (MRX) complex (homolog of the SbcCD nuclease) was shown to be accountable for generating DSBs at perfect palindromes (Farah et al., 2005). On the other hand, Mus81/Mms4 and Gen1 were implicated in breakage at inverted repeats on plasmids in the budding yeast and human cells, respectively (Cote and Lewis, 2008; Inagaki et al., 2013). Interestingly, breakage at long inverted repeats present on yeast chromosomes was demonstrated to be independent of the activities of the above mentioned nucleases (Casper et al., 2009). Further analysis of the breakage intermediates revealed that the breaks were symmetrical and hairpin capped. These broken molecules were found to be substrates for Sae2 and the MRX

complex, which open the hairpin and initiate DNA resection and repair (Lobachev et al., 2002). These data point towards cruciform resolution as the most likely mode of breakage at inverted repeats. However, identity of the nuclease responsible for the breakage is currently unknown.

Inverted repeats on plasmids and in the chromosomal context can also block replication progression in *E. coli*, yeast and mammalian cell lines (Voineagu et al., 2008). Fork stall, breakage and consequently rearrangements at long inverted repeats are further augmented when the replication machinery is compromised (Casper et al., 2008; Voineagu et al., 2008; Zhang et al., 2013). Formation of hairpin structure on the lagging strand was deemed the culprit as the intensity of the blockage directly correlated with the propensity of the repeats to adopt non-B forms. In replication-deficient yeast strains, replication block, DSBs and the associated GCRs were also dependent on the activity of homologous recombination. It was suggested that bypass of the replication barrier imposed by hairpins formed on the lagging strand via template switching and reannealing of the newly synthesized DNA to the lagging strand might culminate in the formation of a cruciform structure. Such a structure could further be attacked by a putative nuclease leading to the formation of hairpin-capped breaks (Zhang et al., 2013).

Replication of the hairpin-capped breaks at inverted repeats was shown to lead to the formation of dicentric chromosomes (Narayanan et al., 2006). Another mechanism to yield these intermediates could be template switching at the repeats due to a replication fork stall (Mizuno et al., 2009). These structures are inherently unstable and can lead to

DSB formation. The major mechanism employed to stabilize the broken molecules is resection and one-ended invasion into regions of microhomology or delta and Ty elements, followed by break induced replication (BIR) (Narayanan et al., 2006; VanHulle et al., 2007). Since BIR has been shown to be a highly mutagenic process (Deem et al., 2011), it is highly probable that breakage at inverted repeats and subsequent repair can be a potent source of mutagenesis.

1.2.2 Triplex forming GAA/TTC repeats

Homopurine homopyrimidine sequences with mirror symmetry can form triplex structures wherein a third DNA strand can pair with the duplex DNA via Hoogsteen hydrogen bonds. Long GAA/TTC tracts are exemplar of triplex forming DNA motifs and are abundant in both prokaryotic and eukaryotic genomes (Clark et al., 2006; Kassai-Jager et al., 2008). ~2700 loci have been mapped to contain GAA/TTC repeats of varying lengths in the human genome (Kozlowski et al., 2010). These repeats have the potential to both expand and contract and expansion of the repeats has been associated with the neurodegenerative disease Friedreich's ataxia where expanded repeats (66-1700 copies from <65 repeats) in the first intron of the FXN gene can inactivate the gene (Campuzano et al., 1996).

Long GAA/TTC repeats were also shown to stimulate mitotic crossing over and gross chromosomal rearrangements in yeast (Kim et al., 2008; Tang et al., 2011; Zhang et al., 2012). Repeat instability correlates with the size of the repeats and their orientation relative to the replication origin. Strains containing long repeat tracts with GAA on the

lagging strand exhibit augmented chromosomal fragility as compared to strains with GAA on the leading strand. Increased fragility was associated with an increase in replication fork stalling at the repeats. Furthermore, long GAA/TTC tracts on plasmids and on chromosome can block replication fork progression in yeast and mammalian cells (Follonier et al., 2013; Kim et al., 2008; Krasilnikova and Mirkin, 2004). The extent of the replication block was also dependent on the length of the repeats and was found to be elevated when the GAA containing tract was on the lagging strand (Kim et al., 2008). These observations indicate that triplex formation at GAA/TTC repeats on the lagging strand is the main intermediate capable of triggering chromosomal breakage. Interestingly, augmented rates of GCRs were observed in yeast strains with defective transcription machinery. Furthermore, increased breakage intermediates could be detected in non-dividing wild-type and transcription-deficient yeast strains held in the stationary-phase (Zhang et al., 2012). These data indicate a replication-independent mechanism of DSB formation at GAA/TTC tracts.

Similar to inverted repeats, in yeast, the major mechanism to repair chromosomal breaks generated at GAA/TTC repeats was surmised to be BIR. The breakage intermediates could be stabilized via invasion into GAA/TTC rich regions found in the MNN4 gene on chromosome XI or the FPR3 gene on chromosome XIII and BIR resulting in terminal deletions and concomitant duplication of the donor chromosomal region (Kim et al., 2008).

1.3 Overview of the dissertation

Repair of DNA DSBs is crucial for the maintenance of genome integrity as imprecise repair can lead to genome rearrangements and mutagenesis, the hallmarks of cancers and hereditary diseases. Secondary structure forming DNA repeats are a potent source of DSBs and have been shown to be the driving force behind gross-chromosomal rearrangements (GCRs) due to their inherent fragility in both humans and model-organisms. In order to safeguard the genome against the detrimental effects of DSBs, the cell has evolved a variety of repair mechanisms. Recent studies involving chromosomal breaks generated using site-specific endonucleases indicate that repair-associated DNA synthesis is extremely error prone. However, unlike natural fragile sites, site-specific endonucleases are extremely efficient resulting in up to 100% breaks. It is unclear if the amount of breaks generated by fragile sites is sufficient to induce mutagenesis in the flanking regions. This dissertation focuses on determining the mutagenic potential of two secondary structure forming motifs – inverted repeats capable of adopting hairpin or cruciform structures and the triplex-forming long GAA/TTC repeats as inherent sources of DSB-associated mutagenesis. In addition, the major mechanism of repair of secondary structure-induced DSBs surmised from the previous studies from this lab is BIR. This mode of repair synthesis has been shown to be highly unstable and mutagenic. This study will further elucidate the intermediates formed during BIR that are responsible for its mutagenic nature.

Chapter 2 presents analysis of the mutagenic potential of secondary structure forming DNA motifs. We demonstrated that in yeast, fragility at both inverted repeats

and long GAA/TTC repeats is associated with augmented mutation levels in a reporter located ~8kb from the repeats. The increase in mutation frequency is dependent on the presence of repeats and the nuclease Sae2 and the translesion polymerase Pol ζ . These results indicate that secondary structure-forming DNA repeats pose a dual threat to the genome by inducing chromosomal aberrations and mutations in the flanking regions.

The study presented in Chapter 3 demonstrates that BIR proceeds via an unusual bubble-like replication fork driven by the Pif1 helicase, leading to a conservative mode of inheritance for the newly synthesized strands. Additionally, this non-canonical replication fork generates large single-stranded intermediates. This work enabled us to determine that the structure of the migrating bubble itself makes BIR prone to instabilities including high levels of mutagenesis.

1.4 References

Burch, L.H., Yang, Y., Sterling, J.F., Roberts, S.A., Chao, F.G., Xu, H., Zhang, L., Walsh, J., Resnick, M.A., Mieczkowski, P.A., et al. (2011). Damage-induced localized hypermutability. *Cell Cycle* 10, 1073-1085.

Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423-1427.

Carvalho, C.M., Zhang, F., and Lupski, J.R. (2011). Structural variation of the human genome: mechanisms, assays, and role in male infertility. *Systems biology in reproductive medicine* 57, 3-16.

Casper, A.M., Greenwell, P.W., Tang, W., and Petes, T.D. (2009). Chromosome aberrations resulting from double-strand DNA breaks at a naturally occurring

yeast fragile site composed of inverted ty elements are independent of Mre11p and Sae2p. *Genetics* 183, 423-439, 421SI-426SI.

Casper, A.M., Mieczkowski, P.A., Gawel, M., and Petes, T.D. (2008). Low levels of DNA polymerase alpha induce mitotic and meiotic instability in the ribosomal DNA gene cluster of *Saccharomyces cerevisiae*. *PLoS genetics* 4, e1000105.

Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A., and Stambrook, P.J. (2002). Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3586-3590.

Clark, R.M., Bhaskar, S.S., Miyahara, M., Dalglish, G.L., and Bidichandani, S.I. (2006). Expansion of GAA trinucleotide repeats in mammals. *Genomics* 87, 57-67.

Connelly, J.C., and Leach, D.R. (1996). The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. *Genes to cells : devoted to molecular & cellular mechanisms* 1, 285-291.

Cote, A.G., and Lewis, S.M. (2008). Mus81-dependent double-strand DNA breaks at in vivo-generated cruciform structures in *S. cerevisiae*. *Molecular cell* 31, 800-812.

Cromie, G.A., Millar, C.B., Schmidt, K.H., and Leach, D.R. (2000). Palindromes as substrates for multiple pathways of recombination in *Escherichia coli*. *Genetics* 154, 513-522.

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

Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS biology* 9, e1000594.

Degtyareva, N.P., Heyburn, L., Sterling, J., Resnick, M.A., Gordenin, D.A., and Doetsch, P.W. (2013). Oxidative stress-induced mutagenesis in single-strand DNA occurs primarily at cytosines and is DNA polymerase zeta-dependent only for adenines and guanines. *Nucleic acids research* 41, 8995-9005.

- Demerec, M. (1962). "Selfers"-Attributed to Unequal Crossovers in Salmonella. Proceedings of the National Academy of Sciences of the United States of America 48, 1696-1704.
- Demerec, M. (1963). Selfer Mutants of Salmonella Typhimurium. Genetics 48, 1519-1531.
- Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M., Beroukhim, R., and Getz, G. (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. Genome research 23, 228-235.
- Esposito, M.S., and Bruschi, C.V. (1993). Diploid yeast cells yield homozygous spontaneous mutations. Current genetics 23, 430-434.
- Eykelenboom, J.K., Blackwood, J.K., Okely, E., and Leach, D.R. (2008). SbcCD causes a double-strand break at a DNA palindrome in the Escherichia coli chromosome. Molecular cell 29, 644-651.
- Farah, J.A., Cromie, G., Steiner, W.W., and Smith, G.R. (2005). A novel recombination pathway initiated by the Mre11/Rad50/Nbs1 complex eliminates palindromes during meiosis in Schizosaccharomyces pombe. Genetics 169, 1261-1274.
- Follonier, C., Oehler, J., Herrador, R., and Lopes, M. (2013). Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. Nature structural & molecular biology 20, 486-494.
- Ford, M., and Fried, M. (1986). Large inverted duplications are associated with gene amplification. Cell 45, 425-430.
- Fox, E.J., Prindle, M.J., and Loeb, L.A. (2013). Do mutator mutations fuel tumorigenesis? Cancer metastasis reviews 32, 353-361.
- Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E., and Resnick, M.A. (1993). Inverted DNA repeats: a source of eukaryotic genomic instability. Molecular and cellular biology 13, 5315-5322.
- Gordenin, D.A., and Resnick, M.A. (1998). Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. Mutation research 400, 45-58.

- Gotter, A.L., Nimmakayalu, M.A., Jalali, G.R., Hacker, A.M., Vorstman, J., Conforto Duffy, D., Medne, L., and Emanuel, B.S. (2007). A palindrome-driven complex rearrangement of 22q11.2 and 8q24.1 elucidated using novel technologies. *Genome research* 17, 470-481.
- Gotter, A.L., Shaikh, T.H., Budarf, M.L., Rhodes, C.H., and Emanuel, B.S. (2004). A palindrome-mediated mechanism distinguishes translocations involving LCR-B of chromosome 22q11.2. *Human molecular genetics* 13, 103-115.
- Guenthoer, J., Diede, S.J., Tanaka, H., Chai, X., Hsu, L., Tapscott, S.J., and Porter, P.L. (2012). Assessment of palindromes as platforms for DNA amplification in breast cancer. *Genome research* 22, 232-245.
- Haber, J.E. (1999). DNA recombination: the replication connection. *Trends in biochemical sciences* 24, 271-275.
- Hahn, W.C., Counter, C.M., Lundberg, A.S., Beijersbergen, R.L., Brooks, M.W., and Weinberg, R.A. (1999). Creation of human tumour cells with defined genetic elements. *Nature* 400, 464-468.
- Harris, R.S., Longrich, S., and Rosenberg, S.M. (1994). Recombination in adaptive mutation. *Science* 264, 258-260.
- Harris, R.S., Ross, K.J., and Rosenberg, S.M. (1996). Opposing roles of the holliday junction processing systems of *Escherichia coli* in recombination-dependent adaptive mutation. *Genetics* 142, 681-691.
- Hastings, P.J., Hersh, M.N., Thornton, P.C., Fonville, N.C., Slack, A., Frisch, R.L., Ray, M.P., Harris, R.S., Leal, S.M., and Rosenberg, S.M. (2010). Competition of *Escherichia coli* DNA polymerases I, II and III with DNA Pol IV in stressed cells. *PloS one* 5, e10862.
- He, A.S., Rohatgi, P.R., Hersh, M.N., and Rosenberg, S.M. (2006). Roles of *E. coli* double-strand-break-repair proteins in stress-induced mutation. *DNA repair* 5, 258-273.
- Hicks, W.M., Kim, M., and Haber, J.E. (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329, 82-85.

- Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M., Emanuel, B.S., and Kurahashi, H. (2013). Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nature communications* 4, 1592.
- Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., and de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *The Journal of experimental medicine* 203, 319-323.
- Jones, S., Chen, W.D., Parmigiani, G., Diehl, F., Beerewinkel, N., Antal, T., Traulsen, A., Nowak, M.A., Siegel, C., Velculescu, V.E., et al. (2008). Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences of the United States of America* 105, 4283-4288.
- Kassai-Jager, E., Ortutay, C., Toth, G., Vellai, T., and Gaspari, Z. (2008). Distribution and evolution of short tandem repeats in closely related bacterial genomes. *Gene* 410, 18-25.
- Kim, H.M., Narayanan, V., Mieczkowski, P.A., Petes, T.D., Krasilnikova, M.M., Mirkin, S.M., and Lobachev, K.S. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *The EMBO journal* 27, 2896-2906.
- Kozlowski, P., de Mezer, M., and Krzyzosiak, W.J. (2010). Trinucleotide repeats in human genome and exome. *Nucleic acids research* 38, 4027-4039.
- Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)_n repeats in vivo. *Molecular and cellular biology* 24, 2286-2295.
- Kurahashi, H., Shaikh, T., Takata, M., Toda, T., and Emanuel, B.S. (2003). The constitutional t(17;22): another translocation mediated by palindromic AT-rich repeats. *American journal of human genetics* 72, 733-738.
- Leach, D.R. (1994). Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays : news and reviews in molecular, cellular and developmental biology* 16, 893-900.

- Lemoine, F.J., Degtyareva, N.P., Lobachev, K., and Petes, T.D. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* 120, 587-598.
- Lobachev, K.S., Gordenin, D.A., and Resnick, M.A. (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* 108, 183-193.
- Lobachev, K.S., Rattray, A., and Narayanan, V. (2007). Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Frontiers in bioscience : a journal and virtual library* 12, 4208-4220.
- Lobachev, K.S., Stenger, J.E., Kozyreva, O.G., Jurka, J., Gordenin, D.A., and Resnick, M.A. (2000). Inverted Alu repeats unstable in yeast are excluded from the human genome. *The EMBO journal* 19, 3822-3830.
- Magni, G.E. (1963). The Origin of Spontaneous Mutations during Meiosis. *Proceedings of the National Academy of Sciences of the United States of America* 50, 975-980.
- Mahale, A.M., Khan, Z.A., Igarashi, M., Nanjangud, G.J., Qiao, R.F., Yao, S., Lee, S.W., and Aaronson, S.A. (2008). Clonal selection in malignant transformation of human fibroblasts transduced with defined cellular oncogenes. *Cancer research* 68, 1417-1426.
- Mangano, R., Piddini, E., Carramusa, L., Duhig, T., Feo, S., and Fried, M. (1998). Chimeric amplicons containing the c-myc gene in HL60 cells. *Oncogene* 17, 2771-2777.
- Mirkin, S.M. (2008). Discovery of alternative DNA structures: a heroic decade (1979-1989). *Frontiers in bioscience : a journal and virtual library* 13, 1064-1071.
- Mizuno, K., Lambert, S., Baldacci, G., Murray, J.M., and Carr, A.M. (2009). Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes & development* 23, 2876-2886.
- Narayanan, V., Mieczkowski, P.A., Kim, H.M., Petes, T.D., and Lobachev, K.S. (2006). The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* 125, 1283-1296.

- Neiman, P.E., Burnside, J., Elsaesser, K., Hwang, H., Clurman, B.E., Kimmel, R., and Delrow, J. (2006). Analysis of gene expression, copy number and palindrome formation with a Dt40 enriched cDNA microarray. *Sub-cellular biochemistry* 40, 245-256.
- Neiman, P.E., Elsaesser, K., Loring, G., and Kimmel, R. (2008). Myc oncogene-induced genomic instability: DNA palindromes in bursal lymphomagenesis. *PLoS genetics* 4, e1000132.
- Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979-993.
- Nimmakayalu, M.A., Gotter, A.L., Shaikh, T.H., and Emanuel, B.S. (2003). A novel sequence-based approach to localize translocation breakpoints identifies the molecular basis of a t(4;22). *Human molecular genetics* 12, 2817-2825.
- Paszewski, A., and Surzycki, S. (1964). 'Selfers' and High Mutation Rate during Meiosis in *Ascobolus Immersus*. *Nature* 204, 809.
- Pomerantz, R.T., Goodman, M.F., and O'Donnell, M.E. (2013). DNA polymerases are error-prone at RecA-mediated recombination intermediates. *Cell Cycle* 12, 2558-2563.
- Ponder, R.G., Fonville, N.C., and Rosenberg, S.M. (2005). A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Molecular cell* 19, 791-804.
- Rattray, A.J., Shafer, B.K., McGill, C.B., and Strathern, J.N. (2002). The roles of REV3 and RAD57 in double-strand-break-repair-induced mutagenesis of *Saccharomyces cerevisiae*. *Genetics* 162, 1063-1077.
- Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics* 45, 970-976.
- Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., et al. (2012). Clustered mutations in

yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular cell* 46, 424-435.

- Rooks, H., Clark, B., Best, S., Rushton, P., Oakley, M., Thein, O.S., Cuthbert, A.C., Britland, A., Ruf, A., and Thein, S.L. (2012). A novel 506kb deletion causing epsilon-gamma-thalassemia. *Blood cells, molecules & diseases* 49, 121-127.
- Sheridan, M.B., Kato, T., Haldeman-Englert, C., Jalali, G.R., Milunsky, J.M., Zou, Y., Klaes, R., Gimelli, G., Gimelli, S., Gemmill, R.M., et al. (2010). A palindrome-mediated recurrent translocation with 3:1 meiotic nondisjunction: the t(8;22)(q24.13;q11.21). *American journal of human genetics* 87, 209-218.
- Strathern, J.N., Shafer, B.K., and McGill, C.B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics* 140, 965-972.
- Tanaka, H., Cao, Y., Bergstrom, D.A., Kooperberg, C., Tapscott, S.J., and Yao, M.C. (2007). Intrastrand annealing leads to the formation of a large DNA palindrome and determines the boundaries of genomic amplification in human cancer. *Molecular and cellular biology* 27, 1993-2002.
- Tang, W., Dominska, M., Greenwell, P.W., Harvanek, Z., Lobachev, K.S., Kim, H.M., Narayanan, V., Mirkin, S.M., and Petes, T.D. (2011). Friedreich's ataxia (GAA)_n*(TTC)_n repeats strongly stimulate mitotic crossovers in *Saccharomyces cerevisiae*. *PLoS genetics* 7, e1001270.
- VanHulle, K., Lemoine, F.J., Narayanan, V., Downing, B., Hull, K., McCullough, C., Bellinger, M., Lobachev, K., Petes, T.D., and Malkova, A. (2007). Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. *Molecular and cellular biology* 27, 2601-2614.
- Voineagu, I., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2008). Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proceedings of the National Academy of Sciences of the United States of America* 105, 9936-9941.
- Zhang, Y., Saini, N., Sheng, Z., and Lobachev, K.S. (2013). Genome-wide screen reveals replication pathway for quasi-palindrome fragility dependent on homologous recombination. *PLoS genetics* 9, e1003979.

Zhang, Y., Shishkin, A.A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M., and Lobachev, K.S. (2012). Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. *Molecular cell* 48, 254-265.

Zhao, J., Bacolla, A., Wang, G., and Vasquez, K.M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and molecular life sciences : CMLS* 67, 43-62.

CHAPTER 2

FRAGILE DNA MOTIFS TRIGGER MUTAGENESIS AT DISTANT CHROMOSOMAL LOCI IN *Saccharomyces cerevisiae*

2.1 Abstract

DNA sequences capable of adopting non-canonical secondary structures have been associated with gross-chromosomal rearrangements in humans and model organisms. Previously we have shown that long inverted repeats that form hairpin and cruciform structures and triplex-forming GAA/TTC repeats induce the formation of double-strand breaks which trigger genome instability in yeast. In this study, we demonstrate that breakage at both inverted repeats and GAA/TTC repeats is augmented by defects in DNA replication. Increased fragility is associated with increased mutation levels in the reporter genes located as far as 8 kb from both sides of the repeats. The increase in mutations was dependent on the presence of inverted or GAA/TTC repeats and activity of the translesion polymerase Pol ζ . Mutagenesis induced by inverted repeats also required Sae2 which opens hairpin-capped breaks and initiates end resection. The amount of breakage at the repeats is an important determinant of mutations as a perfect palindromic sequence with inherently increased fragility was also found to elevate mutation rates even in replication-proficient strains. We hypothesize that the underlying mechanism for mutagenesis induced by fragile motifs involves the formation of long single-stranded regions in the broken chromosome, invasion of the undamaged sister chromatid for repair, and faulty DNA synthesis employing Pol ζ . These data demonstrate that repeat-mediated breaks pose a dual threat to eukaryotic genome integrity by inducing chromosomal aberrations as well as mutations in flanking genes.

2.2 Introduction

Chromosomal instability and mutagenesis are two fundamental processes that alter prokaryotic and eukaryotic genomes. The deleterious consequences of excessive DNA perturbations are hereditary diseases and cancer in humans (reviewed in (Charames and Bapat, 2003; Harper and Elledge, 2007; Jackson and Bartek, 2009)). At the same time, a fine balance between acquiring genetic changes and restoring original DNA content is paramount for organismal development, adaptation, polymorphism and evolution (for example (Pollard *et al.*, 2006; Stamatoyannopoulos *et al.*, 2009; Zhang *et al.*, 2009)).

Double-strand breaks (DSBs) in DNA are a driving force for both chromosomal instability and accumulation of mutations. DSBs are a well-established source of a variety of chromosomal aberrations including translocations and copy number variations (Hastings *et al.*, 2010; Wyman and Kanaar, 2006). It has also become evident from studies in bacteria and yeast that DSB formation and repair are associated with an increased level of mutations, even during homologous recombination which was considered to be an error-free process. In *E. coli*, the role of DSB formation in the induction of mutagenesis was first inferred based on the requirement of RecA and RecBCD for the occurrence of adaptive mutations in the *LacZ* gene (Harris *et al.*, 1994) and was later directly demonstrated by using I-*SceI* endonuclease-induced breaks (Ponder *et al.*, 2005). In yeast, elevated levels of base substitutions and frame shift mutations were shown to be due to DSB repair in meiosis (Magni, 1963) and as a result of induction of DSBs in mitotically-dividing cells as shown in gene conversion (GC) (Ratray *et al.*, 2002; Strathern *et al.*, 1995), break-induced replication (BIR) (Deem *et al.*, 2011) and single-strand annealing (SSA) assays (Yang *et al.*, 2008). The proposed mechanism for

break-induced mutagenesis, surmised from these studies, involves the formation of long regions of single-stranded DNA (ssDNA) as a result of DSB end resection. Mutations arise during error-prone synthesis either across the damaged ssDNA template or during synthesis following invasion into the undamaged donor strand. There are two lines of evidence supporting this mechanism. First, Yang *et al.*, 2008 have shown in yeast, that single-stranded DNA is drastically more prone to the accumulation of mutations with and without treatment with DNA damaging agents than double-stranded DNA. Second, in several studies, mutagenesis was shown to be fully or partially dependent on highly inaccurate translesion polymerases (TLS). The bacterial TLS polymerase, DinB is responsible for 85% of the mutations triggered by DSB repair during adaptive mutagenesis (McKenzie *et al.*, 2001). In yeast, depending on the assay and nature of mutations, DSB-induced mutagenesis is either completely (SSA (Yang *et al.*, 2008)), partially (GC next to the DSB site and BIR (Deem *et al.*, 2011; Rattray *et al.*, 2002; Strathern *et al.*, 1995)) or not (classical GC assay (Hicks *et al.*, 2010)) attributed to the activity of the error-prone TLS polymerase, Pol ζ .

Problems encountered by DNA replication machinery are a major source of spontaneous chromosomal breakage in eukaryotes, estimated to be approximately 10 DSBs per cell cycle in human cells (reviewed in Haber, 1999). Certain chromosomal regions, the fragile sites, often containing secondary structure-forming repeats, are susceptible to breakage especially under conditions of replication stress (Schwartz *et al.*, 2006). The mutagenic potential of replication-associated breaks has not been studied in detail. It is also unknown what the level of breaks during replication should be for mutagenesis to be manifested. The latter is important considering the fact that in

previous studies mutagenesis was detected under conditions of extremely high levels of DSBs, reaching up to 100% as seen in the case of site-specific endonucleases. Whether fragile motifs on their own or under conditions of replication stress could be a potent endogenous source of mutations remains to be established.

In this study, we investigate the mutagenic potential of two sequence motifs, inverted repeats and GAA/TTC tracts, which are natural chromosomal fragile sites (Kim *et al.*, 2008; Lobachev *et al.*, 2002) under conditions of unperturbed and compromised replication. Long inverted repeats can adopt non-B DNA secondary structures such as hairpins and cruciforms owing to their internal symmetry. They are a potent source of genome rearrangements in both prokaryotes and eukaryotes including humans (Darmon *et al.*, 2010; Edelman *et al.*, 2001; Kurahashi and Emanuel, 2001; Narayanan *et al.*, 2006). We have previously demonstrated that in yeast a 320 bp *Alu*-quasi-palindrome triggers gross chromosomal rearrangements by inducing special type of DSBs that have hairpin-capped termini (Lobachev *et al.*, 2002; Narayanan *et al.*, 2006). The hairpin ends are a substrate for opening and processing by Sae2 and the Mre11/Rad50/Xrs2 (MRX) complex. In $\Delta mre11$, $\Delta rad50$, $\Delta xrs2$, or $\Delta sae2$ mutants, the resection of broken ends is completely blocked, giving rise to inverted dimers. GAA/TTC tracts adopt another kind of non-canonical DNA structure, namely, H-DNA or triplex DNA (reviewed in (Frank-Kamenetskii and Mirkin, 1995)). The triplex secondary structure is a driving force for the expansions of GAA tracts, a phenomenon responsible for Friedreich's ataxia in humans (Campuzano *et al.*, 1996). Triplex-adopting sequences, including GAA/TTC repeats, are also responsible for breakage and induction of recombination and rearrangements in bacteria, yeast and humans (Blaszak *et al.*, 1999; Kim *et al.*, 2008;

Napierala *et al.*, 2004; Patel *et al.*, 2004; Raghavan *et al.*, 2005a; Raghavan *et al.*, 2005b; Wang and Vasquez, 2004). Using yeast as an experimental system, we previously demonstrated that triplex structure-imposed replication problems can contribute to breakage at long GAA/TTC tracts (Kim *et al.*, 2008). At the same time, GAA-mediated breaks can occur in non-dividing cells where transcription is an important determinant of DSBs (Tang *et al.*, 2011; Zhang *et al.*, 2012). H-DNA forming sequences are mutagenic in yeast and mammalian systems (Shah *et al.*, 2012; Shishkin *et al.*, 2009; Tang *et al.*, 2013; Wang and Vasquez, 2004), albeit, direct evidence that repeat-induced fragility is the reason for mutagenesis in the vicinity of the repeats remains to be found.

In this work, we demonstrate that increased break formation at the location of inverted repeats causes mutagenesis at distances up to 8 kb away from the DSB site. The accumulation of mutations requires the Sae2 protein, indicating that resection and generation of long ssDNA is a critical parameter for this phenomenon. We have found that error-prone synthesis involving the translesion polymerase Pol ζ during repair is primarily responsible for the observed mutagenesis. We also show that in replication-deficient strains the triplex-adopting GAA/TTC repeats are associated with hypermutability at distant loci, suggesting that a similar mechanism of mutagenesis can operate at repeat-associated chromosomal break sites under conditions of replication stress. These data demonstrate that secondary structure-mediated breaks pose a dual threat to eukaryotic genome integrity by inducing chromosomal aberrations and mutations extending to distant chromosomal sites. It is conceivable that the mechanisms of DSB-induced mutagenesis uncovered in this study are also relevant to human evolution, polymorphism and tumorigenesis.

2.3 Results

2.3.1 Experimental system

The experimental system used to assess the mutagenic potential of fragile inverted and GAA/TTC repeats in this study is based on the GCR assay described in (Kim *et al.*, 2008; Narayanan *et al.*, 2006). Briefly, the *LYS2* gene containing the fragile motifs was inserted 43 kb from the telomere on the left arm of chromosome V in haploid yeast strains (Figure 2.1). There are no essential genes between the left telomere and the *LYS2* gene. *CANI* is located 8 kb telomere-proximal to the *LYS2* gene. The insertion of *ADE2* between *CANI* and *LYS2* allows for the differentiation between two types of events on media containing canavanine and low amounts of adenine. Breakage at the location of structure-forming repeats leads to the loss of the terminal 43 kb of the chromosomal arm containing both *CANI* and *ADE2*, resulting in canavanine-resistant red-colored colonies (Can^RAde⁻). On the other hand, mutations in *CANI* are manifested as white-colored canavanine-resistant colonies (Can^RAde⁺) (Figure 2.1). The correlation between colony color and the requirement of adenine for growth was verified by replica plating the Can^R colonies to media lacking adenine. The three fragile motifs inserted into *LYS2* were 100% homologous inverted *Alu* repeats, 320 bp each with a 12 bp spacer (*Alu*-IRs); 100% homologous *IS50* palindromic repeats (*IS50*-PAL), 1.3 kb each; and 230 repeats of GAA/TTC in the orientation wherein the GAA sequence is the template for the lagging strand synthesis.

To estimate how far mutagenesis can extend from the break site, *URA3* was inserted into chromosome V telomere-proximal (TP) 0.6 kb and 30 kb away from the repeats, and telomere-distal (TD) 0.4 kb, 8 kb and 30 kb away from the repeats (Figure 2.1). Mutations in *URA3* were measured on 5-fluoroorotic acid-containing media lacking

adenine (5-FOA^R Ade⁺), allowing us to preferably select these events in contrast to GCRs that give rise to 5-FOA^R Ade⁻ colonies.

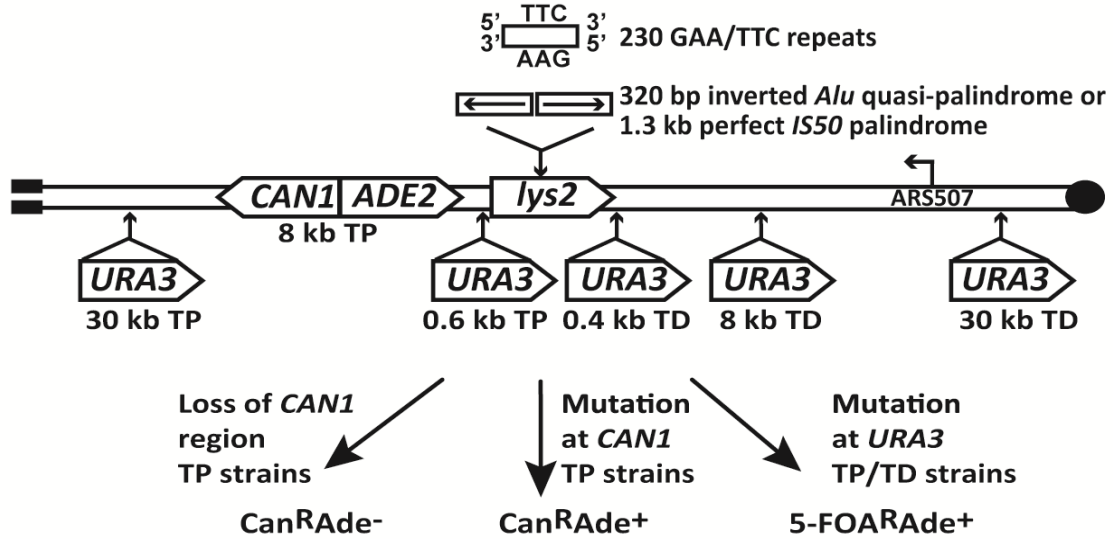


Figure 2.1 Experimental system to study fragile motif-induced mutagenesis. *Alu*-quasi-palindrome, *IS50*-palindrome or GAA/TTC repeats were inserted into *LYS2* gene on the left arm of chromosome V. Positions of *CAN1* and *URA3* reporters located telomere-proximal (TP) or telomere-distal (TD) to the repeat insertion are shown. The position of the *ARS507* and the direction of replication fork migrating through the repeat region are indicated. Breakage at the location of secondary-structure-adopting repeats can lead to loss of 43 kb telomere-proximal deletion resulting in red-colored Can^RAde⁻ clones. Mutations in *CAN1* reporter will yield white-colored Can^RAde⁺ isolates. Mutations in *URA3* gene will give rise to colonies resistant to medium containing 5-fluorototic acid (5-FOA^R).

2.3.2 A defect in DNA replication leads to increased *Alu*-quasipalindrome-induced breakage and mutagenesis at the *CAN1* locus

Mutation levels in the *CAN1* locus in wild-type strains with inverted *Alu*-quasipalindrome are not different from strains that lack the sequence motif. We wanted to determine whether addition of replication stress will enhance the fragility potential of these repeats and increase mutagenesis. In a screen for mutants that exhibit an increased level of

hairpin-capped DSBs we identified the *pol3-P664L* allele that affects the functions of replicative polymerase δ responsible for synthesis of the lagging strand (Nick McElhinny *et al.*, 2008). The *P664L* mutation is located in the polymerase domain of Pol δ (Pavlov *et al.*, 2006) and the yeast strains carrying this mutant allele exhibit temperature-sensitive growth at 37°C (data not shown). The rate of *CAN1* region loss in strains containing *Alu*-quasipalindrome was 40-fold higher in *pol3-P664L* mutants than in wild-type (Table 2.1). Moreover, *pol3-P664L* strains with *Alu*-IRs exhibited elevated levels of mutagenesis in *CAN1* loci located 8 kb away from the DSB site. Notably, the mutagenesis was completely dependent on the presence of fragile motifs, suggesting that the mutator phenotype is not a feature of the *pol3* allele but rather is a consequence of increased breakage. It is important to note that in *pol3-P664L* strains without the *Alu*-quasipalindrome, the relative rate of arm loss was nearly 3-fold higher than in wild-type strains. However, the fragility due to deficiency in Pol δ is not high enough to induce mutagenesis.

A similar increase in *Alu*-IR-dependent fragility and mutagenesis in *CAN1* gene was observed in strains where the *POL3* expression was under the control of a tetracycline-repressible promoter (*tetO7*) (Belli *et al.*, 1998). Belli *et al.*, 1998 showed that *tetO7*-driven expression of genes in the presence of the antibiotic leads to a reduction in protein levels in comparison to conditions when genes were expressed from their native promoters. Western blotting analysis of c-Myc-tagged Pol3 revealed that upon treatment of cells with doxycycline the protein level was indeed ~10 fold decreased in comparison with the wild-type level (Figure 2.2). Hence, we refer to TET-*POL3* as a

mutant allele and all further tests were carried out in the presence of doxycycline (see Material and Methods).

We also replaced the native promoter of another replication gene, *RFA2*, which encodes one of the subunits of the single-stranded DNA-binding protein participating in DNA replication and repair, with the *tetO7* promoter. Upon downregulation with doxycycline, the expression of Rfa2 was ~ 4-fold lower than the wild-type level (Figure 2.2). Similar to the TET-*POL3* strain, the TET-*RFA2* strain exhibited increased levels of arm loss and mutagenesis (Table 2.1).

It is important to note that neither the *pol3-P664L* strain nor the TET-*POL3* and TET-*RFA2* strains grown in the presence of doxycycline at chosen concentrations showed sensitivity to DNA damaging agents such as MMS and camptotechin, indicating that they are proficient in DNA repair (Figure A.1).

Overall, these data show that mutations at distant loci require the presence of fragile motifs and are dependent on the amount of replication-associated breaks.

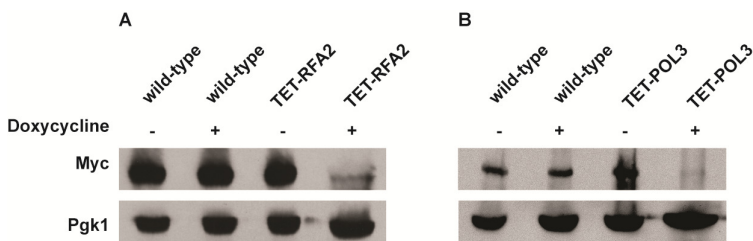


Figure 2.2 Analysis of protein levels of Pol3 and Rfa2 in the wild-type and tetracycline downregulatable strains. (A) Effect of downregulation of Pol3. (B) Effect of downregulation of Rfa2. Pol3 and Rfa2 were c-Myc tagged at the C-terminus in the wild-type, TET-*POL3* and TET-*RFA2* strains respectively. Proteins were extracted with (+) or without (-) treatment with doxycycline. Pol3 and Rfa2 were detected by Western blot with anti-c-MYC antibody. The protein levels were compared against Pgk1 levels (detected with anti-Pgk1 specific antibody) which acted as the loading control. Upon treatment with doxycycline, Pol3 expression was lowered 10 fold (average of 9, 10 and 12) and Rfa2 expression was lowered 4-fold (average of 4.1, 4.4 and 3.7).

Table 2.1 Pol ζ - and Sae2-dependent mutagenesis by *Alu*-quasi-palindrome in replication mutants

Genetic background	Rate of arm loss x 10 ⁻⁷ (Can ^R Ade ⁻)		<i>CAN1</i> mutation rate x 10 ⁻⁷ (Can ^R Ade ⁺)	
	No repeats	Inverted repeats	No repeats	Inverted repeats
wild-type	0.05 (0.03-0.06) ^a	520 (450-630)	3 (1-6)	5* (3-7)
<i>pol3-P664L</i>	0.14 (0.05-0.2)	21000 (14000-27000)	5 (3-9)	60* (50-80)
TET- <i>POL3</i>	0.6 (0.5-0.8)	12000 (10000-14000)	8 (7-13)	140* (90-200)
TET- <i>RFA2</i>	20 (11-30)	8600 (6600-15000)	13 (8-23)	92* (35-190)
$\Delta rev3$	0.05 (0.03-0.1)	643 (280-2200)	2 (1-4)	2 (1-2)
$\Delta sae2$	0.4 (0.3-0.6)	1600 (1100-2200)	2 (1-3)	3 (2-4)
<i>pol3-P664L</i> $\Delta rev3$	0.3 (0.2-0.5)	24000 (21000-31000)	3 (2-5)	8 [#] (5-10)
<i>pol3-P664L</i> $\Delta sae2$	8 (5-10)	15000 (11000-21000)	5 (4-8)	7 [#] (6-9)
TET- <i>POL3</i> $\Delta rev3$	0.7 (0.5-1)	8700 (7000-9600)	8 (7-11)	18 [#] (12-24)
TET- <i>POL3</i> $\Delta sae2$	4 (2-5)	12500 (12000-15000)	4 (3-6)	6 [#] (5-10)
TET- <i>RFA2</i> $\Delta rev3$	20 (14-27)	7900 (6000-9000)	21 (17-40)	23 [#] (13-30)
TET- <i>RFA2</i> $\Delta sae2$	13 (10-15)	14000 (12000-19000)	20 (14-30)	18 [#] (16-26)

^a Numbers in parentheses correspond to the 95% confidence interval

* Depicts mutation rates significantly higher than the wild-type strain (P<0.05)

Depicts mutation rates in $\Delta rev3$ and $\Delta sae2$ strains significantly lower than corresponding replication-deficient strains (P<0.01)

2.3.3 A 1.3 kb perfect palindrome induces mutagenesis at the *CAN1* locus even in the wild-type strains

We addressed directly whether a fragile site can induce mutations at distant loci in replication-proficient strains. This experiment also helps to distinguish which is the key factor in mutagenesis, the level of breakage or repair of broken molecules by faulty replication proteins. The 1.3 kb long *IS50* palindrome was found to induce mutagenesis in *CAN1* when replication was unimpaired (3-fold). The increased length of the interacting arms and the lack of a spacer between them likely create a problem even for intact replication machinery and render this motif highly fragile with a 14-fold increase in GCR rates as compared to the *Alu*-IR strain (Table 2.2). Consistently, using Southern hybridization, we estimated the level of breakage at this palindrome to be 4.8% (average of 4.6%, 5% and 4.9%) which is ~3-times higher than in strains carrying the *Alu*-quasi-palindrome (1.6%, average of 1.4%, 1.5% and 1.9%) (Figure 2.3). Taking into account that a deficiency in Pol3 causes a 7-fold increase in *Alu*-IR-mediated breakage (11%, average of 11%, 10% and 11%) and a 12-fold increase in mutagenesis, it is evident that the levels of DSB formation and not DSB repair by defective replication proteins are the important determinant of mutagenesis.

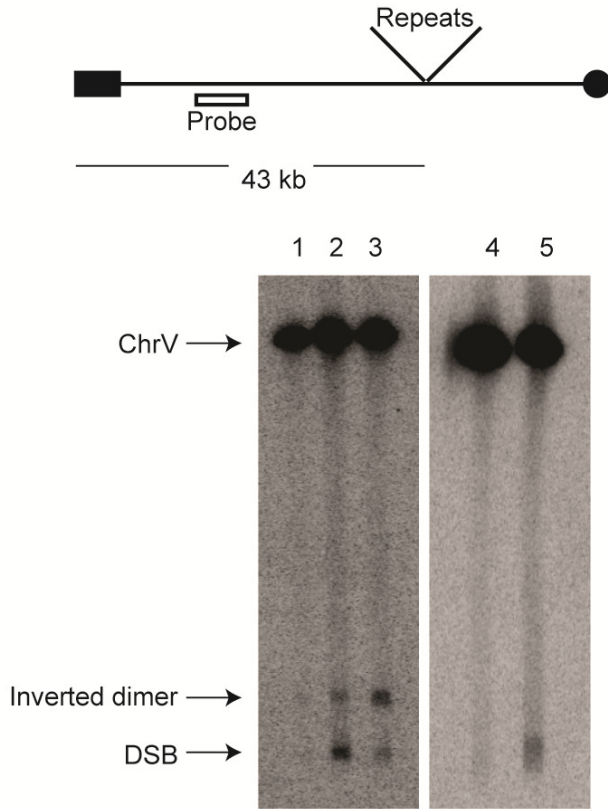


Figure 2.3 Inverted repeat and GAA/TTC-induced DSB detection in wild-type and mutant strains. Upper panel depicts the relative positions of the inverted repeats and the probe (open rectangle) used. For the detection of inverted repeat-mediated breaks $\Delta sae2$ strains were used as in these mutants the hairpin-capped breaks are not opened and resection is abolished [21]. As a consequence, inverted dimer molecules accumulate in $\Delta sae2$ mutants as previously demonstrated. Contour-clamped homogeneous electric field gel electrophoresis and Southern hybridization were used to highlight the intact chromosome V and the broken fragment. Lanes 1, 2 and 3 depict the *Alu-IR*, *pol3-P664L Alu-IR*, and *IS50-IR* strains respectively. Lanes 4 and 5 depict *GAA/TTC₍₂₃₀₎* and *TET-POL3 GAA/TTC₍₂₃₀₎* strains respectively. Intact chromosome V, DSB fragments and inverted dimers (in the case of inverted repeats) are indicated.

Table 2.2 Polζ- and Sae2-dependent mutagenesis by *IS50*-perfect palindrome

Genetic background	Rate of arm loss x 10 ⁻⁷ (Can ^R Ade ⁻)	<i>CAN1</i> mutation rate x 10 ⁻⁷ (Can ^R Ade ⁺)
wild-type	7000 (6000-9000) ^a	13* (10-19)
<i>Δrev3</i>	9000 (7000-11000)	4 [#] (3-6)
<i>Δsae2</i>	6000 (5000 – 7000)	3 [#] (2-5)

^a Numbers in parentheses correspond to the 95% confidence interval

* Depicts mutation rates significantly higher than the wild-type strain with *Alu*-IR (see Table 1) (P<0.05)

Depicts mutation rates in *Δrev3* and *Δsae2* strains significantly lower than the wild-type strain (P<0.01)

2.3.4 Mutagenesis by inverted repeats depends on the distance of the reporter from the DSB site and on the activity of the Sae2 protein

Previously, we have shown that *Alu*-IRs induce DSBs that have hairpin-capped termini (Lobachev *et al.*, 2002). The resection and repair of these DSBs requires the hairpin-opening activity of Sae2 and the Mre11 nuclease (Lengsfeld *et al.*, 2007). To test if ssDNA generated as a result of 5'-3' DSB end resection is a critical requirement for repeat-induced mutagenesis, the *SAE2* gene was disrupted in *pol3-P664L*, TET-*POL3* and TET-*RFA2* *Alu*-IR strains. The level of *CAN1* mutagenesis in *pol3-P664LΔsae2*, TET-*POL3Δsae2* and TET-*RFA2Δsae2* mutants was reduced to levels observed in strains without inverted *Alus* (Table 2.1), indicating that mutations are indeed a consequence of DSB resection. Similarly, in the strains carrying *IS50* repeats, mutation rates declined upon deletion of *SAE2* (Table 2.2).

To determine to what distance the DSB-associated mutagenesis can spread on either side of the fragile site, we inserted the *URA3* reporter 0.4, 8 and 30 kb telomere-distal (TD) and 0.6 and 30 kb telomere-proximal (TP) to *Alu*-IRs in *pol3-P664L* strains (Figure 1). The average length of ssDNA generated via DSB end resection in yeast varies from 2 kb to 10 kb (Zhu *et al.*, 2008). This predicts that mutations in *URA3* situated past 10 kb should diminish. Consistently, although mutation rates at 0.4 kb, 0.6 kb and 8 kb were approximately the same (10-15-fold higher than in wild-type strain), at 30 kb the rate of *ura3* mutations significantly decreased in TP and TD constructs (Table 2.3).

The dependence of the efficiency of mutagenesis on the activity of Sae2 and the distance of the reporter from DSB site demonstrates that ssDNA is an intermediate for the occurrence of mutations.

Table 2.3 Mutagenesis by fragile *Alu*-IRs depends on the distance of reporter from the DSB site

Location of the reporter from <i>Alu</i> -IRs		<i>URA3</i> Mutation rate x 10 ⁻⁷ (5-FOA ^R Ade ⁺)					
		No repeats		Inverted repeats			
		wild-type	<i>pol3-P664L</i>	wild-type	<i>pol3-P664L</i>	$\Delta rev3$	<i>pol3-P664L</i> $\Delta rev3$
TP ^a constr ucts	0.6kb	1.5 (1-1.9) ^c	1 (0.5-2)	1.3 (0.8-2)	15* (8-20)	0.7 (0.5-0.9)	3# (2-5)
	30kb	0.6 (0.5-1.3)	1.7 (1.3-2.2)	0.9 (0.6-1.1)	2.9 (2.7-4.7)	ND	ND
TD ^b constr ucts	0.4kb	0.7 (0.4-1.4)	1.1 (0.4-1.9)	1 (0.5-2)	14* (9-17)	1 (0.7-1)	2# (0.9-4)
	8kb	0.7 (0.6-1)	0.8 (0.7-1.5)	0.6 (0.4-0.8)	9* (6-17)	0.6 (0.3-0.7)	1.4# (0.8-2)
	30kb	0.6 (0.4-0.9)	1.7 (0.8-3.8)	0.7 (0.2-1.9)	1 (0.8-1.4)	ND	ND

^a TP denotes telomere-proximal location of *URA3* with respect to the *Alu*-IRs

^b TD denotes telomere-distal location of *URA3* with respect to the *Alu*-IRs

^c Numbers in parentheses indicate 95% confidence intervals

* Depicts mutation rates significantly higher than the wild-type strain at the respective loci (P<0.01)

Depicts mutation rates in $\Delta rev3$ and $\Delta sae2$ strains significantly lower than the *pol3-P664L* strain at the respective loci (P<0.01)

ND - not determined

2.3.5 Increase in mutagenesis observed in replication-deficient and –proficient strains is mostly attributed to the activity of Polζ translesion polymerase

Holbeck and Strathern, 1997 and Rattray et al., 2002 showed that Polζ translesion synthesis activity is required for the generation of base substitutions in a reporter located 0.3 kb from the site of an HO-endonuclease-induced break. To assess if mutagenesis induced by fragile motifs depends on translesion synthesis, we disrupted the *REV3* gene encoding the catalytic subunit of Polζ (Prakash *et al.*, 2005) in wild-type, *pol3-P664L*, TET-*POL3* and TET-*RFA2* strains with *Alu-IRs* (Table 2.1). *REV3* disruption in replication-defective strains brought the mutation level in the *CANI* reporter to almost the level observed in the wild-type strain. In replication-proficient strains carrying *Δrev3*, only a modest 2-fold decrease in *CANI* mutation rate was observed. Augmented mutation rates in the strains with *IS50* repeats were also dependent on the activity of Polζ (Table 2.2).

To gain further insight into the spectrum of mutations generated at distant loci as a result of DSB formation by inverted repeats, we sequenced 22-31 independent *Can^RAde⁺* isolates from wild-type, *pol3-P664L* and *pol3-P664LΔrev3* strains, respectively. In the wild-type strain with *Alu-IRs*, 85% of the mutations were base substitutions and 15% were single base deletions (Table 2.4 and Table A.1). A similar mutation spectrum was also observed in other studies (Lang and Murray, 2008; Sakamoto *et al.*, 2007). This correlates with the lack of increase of *CANI* mutagenesis in the wild-type *Alu-IR* strain (Table 2.1), indicating that the observed mutations in replication-proficient strains were a result of spontaneous mutagenesis rather than secondary structure-induced DSBs. In the *pol3-P664L* strain the mutation spectrum was changed.

There was a significant increase in the frequencies of base substitutions, particularly G:C→T:A and G:C→C:G transversions characteristic of Polζ errors during spontaneous mutagenesis (Endo *et al.*, 2007) (Table 2.4 and Table A.2). Increases in deletions ranging from 1 to 5 bp and complex mutations (two or more changes in a run of 10 bp) were also observed. These types of changes were also previously attributed to the TLS activity of Polζ (Abdulovic *et al.*, 2008; Sakamoto *et al.*, 2007). A similar mutation spectrum was also seen for Can^RAde⁺ clones from strains containing the *IS50*-perfect palindrome (Table 2.4 and Table A.5). Since mutagenesis observed in these strains requires the activity of Polζ, it is likely that error-prone synthesis by the TLS polymerase during DSB repair causes base substitutions as well as deletions and complex mutations. Consistently, errors that could be assigned to the activity of Polζ were suppressed in *pol3-P664LΔrev3* strains (Table 2.4 and Table A.3). We also uncovered large deletions (up to 39 bp) and a duplication of 27 bp flanked by short direct repeats in *pol3* mutants with or without *Alu*-IRs. This is most probably attributed to the defective Polδ. Notably, *pol3-P664L* strains that lack fragile motifs also exhibited complex mutations (Table 2.4 and Table A.4). Taking into account that fragility in *pol3-P664L* without *Alu*-IRs is low, it can be inferred that these changes reflect mutations arising during DNA replication carried out by a faulty DNA polymerase (a process that also might require TLS polymerases (Northam *et al.*, 2010)) rather than a consequence of error-prone synthesis during DSB repair.

Overall, analysis of mutation spectra in wild-type and replication-deficient strains is in agreement with genetic analysis and supports the conclusion that repeat-mediated

mutations are generated by error-prone Pol ζ and do not occur due to faulty synthesis by replicative polymerases.

Table 2.4 Mutation spectra in *CAN1* reporter

Class of mutation ^a	Mutation rate x 10 ⁻⁷				
	wild-type + <i>Alu</i> -IR	<i>pol3-P664L</i> + <i>Alu</i> -IR	<i>pol3-P664L</i> Δ <i>rev3</i> + <i>Alu</i> -IR	<i>pol3-P664L</i> - <i>Alu</i> -IR	wild-type + <i>IS50</i> -PAL
Base substitutions	4.2 (23 ^b , 85% ^c)	27 (14, 45%)	3.2 (9, 41%)	2.4 (14, 48%)	9.4(18, 72%)
- G:C→T:A, G:C→C:G transversions	2.0 (11, 41%)	19* (10, 32%)	1.8 (5, 22%)	1.0 (6, 20%)	4.7* (9, 36%)
- other substitutions	2.2 (12, 44%)	8 (4, 13%)	1.5 (4, 18%)	1.3 (8, 27%)	4.7 (9, 36%)
Frameshifts^d	0.7 (4, 15%)	17* (9, 29%)	0.4 (1, 5%)	0.7 (4, 14%)	2.1* (4, 16%)
Complex mutations^e	< 0.1 ^g (0, 0%)	12* (6, 19%)	< 0.4 (0, 0%)	0.5 (3, 10%)	1.6* (3, 12%)
Slippage ^f	< 0.1 (0, 0%)	4 (2, 6%)	4.3 (12, 54%)	1.2 (7, 24%)	< 0.5 (0, 0%)
Total	5 (27, 100%)	60 (31, 100%)	8 (22, 100%)	5 (29, 100%)	13 (25, 100%)

^a Classes of mutations characteristic for Pol ζ are indicated in bold.

^b Parentheses indicate the number of sequenced isolates of each subclass.

^c Relative fraction of the isolates of each subclass among the total isolates sequenced.

^d Frameshift mutations are insertions and deletions of 1-5 nucleotides

^e Complex mutations are more than one nucleotide change in a span of 10 nucleotides.

^f Slippage subclass includes deletions from -15 to -39 nucleotides and duplication of 27 nucleotides flanked by short direct repeats.

^g < value reflects less than measurable rate of mutations.

* indicates the classes demonstrating increased rates of mutagenesis due to the activity of Pol ζ .

2.3.6 GAA/TTC fragile motif also induces mutagenesis at distant chromosomal loci that is partly Pol ζ dependent

To determine if DSB-induced mutagenesis can be observed at another fragile motif, we assessed *CAN1* mutation rate in strains carrying 230 repeats of the triplex-adopting GAA/TTC (Figure 2.1). Although the rate of Can^R mutations was unaltered in wild-type strains, a 4-fold increase in mutagenesis was detected in *pol3-P664L* and TET-*POL3* strains (Table 2.5). The level of DSB formation at GAA/TTC repeats in the TET-*POL3* strain was estimated to be 3.3% (average of 3.1%, 3.3% and 3.6%, Figure 3). This is a minimal estimation of GAA/TTC-mediated DSBs since, unlike the situation with palindromic sequences, resection of the broken fragments cannot be prevented by *SAE2* disruption and a proportion of degraded DSBs are excluded from detection.

Similar to *Alu*-IR-mediated mutagenesis, Pol ζ plays a role in the induction of mutations by GAA/TTC repeats. There was a mild but statistically significant reduction (2-fold) in mutagenesis in *pol3-P664L Δ rev3* versus *pol3-P664L* ($p < 0.05$) and TET-*POL3 Δ rev3* versus TET-*POL3* ($p < 0.05$) strains as determined using an unpaired t-test. Although it is difficult to evaluate the contribution of resection and long ssDNA to GAA/TTC-associated mutagenesis, the involvement of *REV3* suggests that the mechanism underlying mutagenesis in the case of inverted repeats and GAA/TTC fragile sites can be similar.

Table 2.5 Mutagenesis in *CAN1* reporter stimulated by GAA/TTC repeats

Genetic background	Rate of arm loss x 10 ⁻⁷ (Can ^R Ade ⁻)		<i>CAN1</i> mutation rate x 10 ⁻⁷ (Can ^R Ade ⁺)	
	No repeats	GAA/TTC ₍₂₃₀₎	No repeats	GAA/TTC ₍₂₃₀₎
wild-type	0.03 (0.01-0.04) ^a	20 (10-30)	3 (3-4)	5 (3-9)
<i>pol3-P664L</i>	0.2 (0.1-0.3)	240 (190-260)	5 (3-6)	19* (15-24)
TET- <i>POL3</i>	0.6 (0.5-0.8)	180 (130-240)	9 (8-10)	30* (20-50)
$\Delta rev3$	0.08 (0.05-0.1)	20 (10-30)	2 (1-2)	3 (2-4)
<i>pol3-P664L</i> $\Delta rev3$	0.4 (0.3 – 0.6)	214 (176-266)	4 (3-6)	12 [#] (10-16)
TET- <i>POL3</i> $\Delta rev3$	0.7 (0.6-0.9)	160 (140-200)	9 (7-10)	14 [#] (9-16)

^a Numbers in parentheses correspond to the 95% confidence interval

* Depicts mutation rates significantly higher than the wild-type strain (P<0.05)

Depicts mutation rates in $\Delta rev3$ and $\Delta sae2$ strains significantly lower than corresponding replication-deficient strains (P<0.05)

2.4 Discussion

The induction of DSBs using site-specific endonucleases has been shown to drive mutagenesis (Deem *et al.*, 2011; Rattray *et al.*, 2002; Strathern *et al.*, 1995; Yang *et al.*, 2008). This study demonstrates that natural chromosomal fragile sites comprising of sequence motifs that can adopt non-B DNA structures are also mutagenic. Under conditions of replication stress, the mutagenesis can reach up to the levels caused by deficiency in the mismatch repair system (Chen *et al.*, 1999). We also show that the mutations are a consequence of error-prone repair of repeat-induced DSBs. Overall, we establish secondary structure-forming motifs as a potent source of endogenous mutagenesis and reveal the mechanism underlying this phenomenon.

In this study we found that when replication is compromised, *Alu*-quasi-palindrome promotes chromosomal fragility and mutagenesis at *CAN1* and *URA3* reporters located 8 kb from the break site. Mutations were also increased in strains with a perfect *IS50*-palindrome with inherently higher fragility even in replication-proficient strains. We have previously shown that inverted repeats induce hairpin-capped DSBs in replication-proficient strains (Lobachev *et al.*, 2002). We have found that in replication-defective mutants the DSBs mediated by the *Alu*-quasi-palindrome also have hairpin-capped termini (Zhang *et al.*, 2013). The opening of the hairpins necessitates the nuclease activity of the MRX complex and Sae2. The requirement of Sae2 for mutagenesis at distant loci unequivocally demonstrates that mutations are a consequence of DSB formation (Table 2.1 and Table 2.2). Moreover, these data also implicate the formation of long ssDNA upon resection of DSB ends as the second step in repeat-mediated mutagenesis. ssDNA has been shown to be prone to accumulation of mutations during SSA or in a situation where the telomeres become uncapped (Yang *et al.*, 2008). Therefore, it is possible that hairpin-processing generates damaged ssDNA that can serve as a faulty template for synthesis during SSA or GC. Alternatively, the undamaged ssDNA can be involved in strand invasion and mutations could arise due to error-prone synthesis during homologous recombination as suggested in other studies (Deem *et al.*, 2011; Harris *et al.*, 1994; Strathern *et al.*, 1995). Error-prone synthesis of the undamaged DNA template in replication deficient strains by Pol ζ was observed by Northam *et al.*, 2010.

Although we cannot determine whether mutagenesis is due to accumulation of damage in resected DNA or error-prone synthesis on undamaged template, our data point

towards synthesis-dependent strand annealing (SDSA) as the underlying mechanism for mutagenesis (Figure 2.4). None of the analyzed Can^R clones contained interstitial deletions and all of the clones retained intact *Alu*-IRs or *IS50*-palindrome (data not shown). This suggests that SSA is unlikely to operate during *Alu*-IR-mediated mutagenesis and alludes to a template-dependent repair process that involves the undamaged sister chromatid. Thus, we favor a scenario wherein hairpin-capped DSBs are induced in late S or G2 stage of the cell cycle. Upon hairpin opening by Sae2 and MRX, the 3' end of the resected DSB invades the intact sister chromatid template. The requirement for invasion in mutagenesis is a likely step but ultimately cannot be proven by using *rad51* or *rad52* mutants for two reasons: these strains exhibit a mutator phenotype on their own (Endo *et al.*, 2007) and Rad51 and Rad52 proteins are required for DSB formation at the *Alu*-IRs in replication-defective strains (Zhang *et al.*, 2013). It is conceivable that the invasion event can proceed either as a BIR or as an SDSA event. SDSA is the most probable mechanism owing to the fact that mutations were observed in both TP and TD reporters and that reduced mutation rates were measured at reporters 30 kb from the break site (Table 2.3). It is important to note that SDSA preserves the original inverted repeats that can trigger additional rounds of breakage and associated mutagenesis. If extrapolated to humans, these observations identify secondary structure-forming repeats as a potent source of mutagenesis that can change the expression of flanking genes during the lifetime of healthy individuals even in the absence of exogenous damage.

Mutations generated in the reporter 8 kb away from DSB site were also strongly dependent on the activity of Pol ζ (Table 2.1 and Table 2.2) indicating that the error-prone

translesion synthesis operates during SDSA. This is consistent with the Hirano and Sugimoto, 2006 study that showed that Mec1 kinase is needed to recruit the Pol ζ -Rev1 complex to the DSB site (Hirano and Sugimoto, 2006) and other studies where DSB-induced mutagenesis required Pol ζ (Deem *et al.*, 2011; Rattray *et al.*, 2002; Strathern *et al.*, 1995; Yang *et al.*, 2008). Analysis of mutations in the *CAN1* locus of the hyperfragile strains revealed an increase in G:C→T:A and G:C→C:G transversions, frameshift and complex mutations that are signatures of Pol ζ (Table 2.4) (Abdulovic *et al.*, 2008; Endo *et al.*, 2007; Sakamoto *et al.*, 2007).

In this study we also show that DSB-triggering long GAA/TTC repeats induce mutagenesis at distant loci, indicating that a similar underlying mechanism of mutagenesis described above for inverted repeats can operate for triplex-forming motifs. The requirement of Rev3 for mutagenesis is more evident for inverted repeats than for GAA/TTC repeats. It would be interesting to see if other TLS polymerases, Rev1 and Pol η , besides Pol ζ operate in GAA/TTC-associated mutagenesis and to determine if the mutation spectra in GAA/TTC- and inverted repeat-containing strains differ. It is also important to note that in our experimental system, we observe mutagenesis by GAA/TTC tracts only under conditions of compromised replication wherein the repeat-mediated fragility is further increased. In other studies, mutagenesis is induced by GAA/TTC repeats in replication-proficient strains (Shishkin *et al.*, 2009; Tang *et al.*, 2013). These discrepancies might reflect the distance of the used reporter from the fragile motif. It is possible that at closer distances, SSA might be the predominant pathway for mutagenesis where mutations introduced by Pol ζ can be scored above the spontaneous level of mutagenesis, while SDSA requires higher frequencies of breakage and longer ssDNA.

This can be checked experimentally in future studies. Our data are also in agreement with the recent study by Shah *et al.*, 2012 wherein GAA/TTC-induced mutations in a closely juxtaposed reporter in Pol δ mutants were dependent on Pol ζ .

Overall, this study demonstrates that fragile sequence motifs that are found in eukaryotic genomes, including humans, can be potent inducers of mutagenesis. Thus, secondary structure-adopting repeats can represent a dual threat to DNA stability by changing the structural organization of the genome and causing mutations. Recent studies linking the occurrence of mutations near chromosomal rearrangement break-points in primates and humans suggest that error-prone repair of DSBs can operate during speciation, evolution and tumorigenesis (Berger *et al.*, 2011; De and Babu, 2010; Drier *et al.*, 2013; Roberts *et al.*, 2012). Thus it is likely that fragile and mutagenic non B-DNA-forming motifs are contributing factors to these processes.

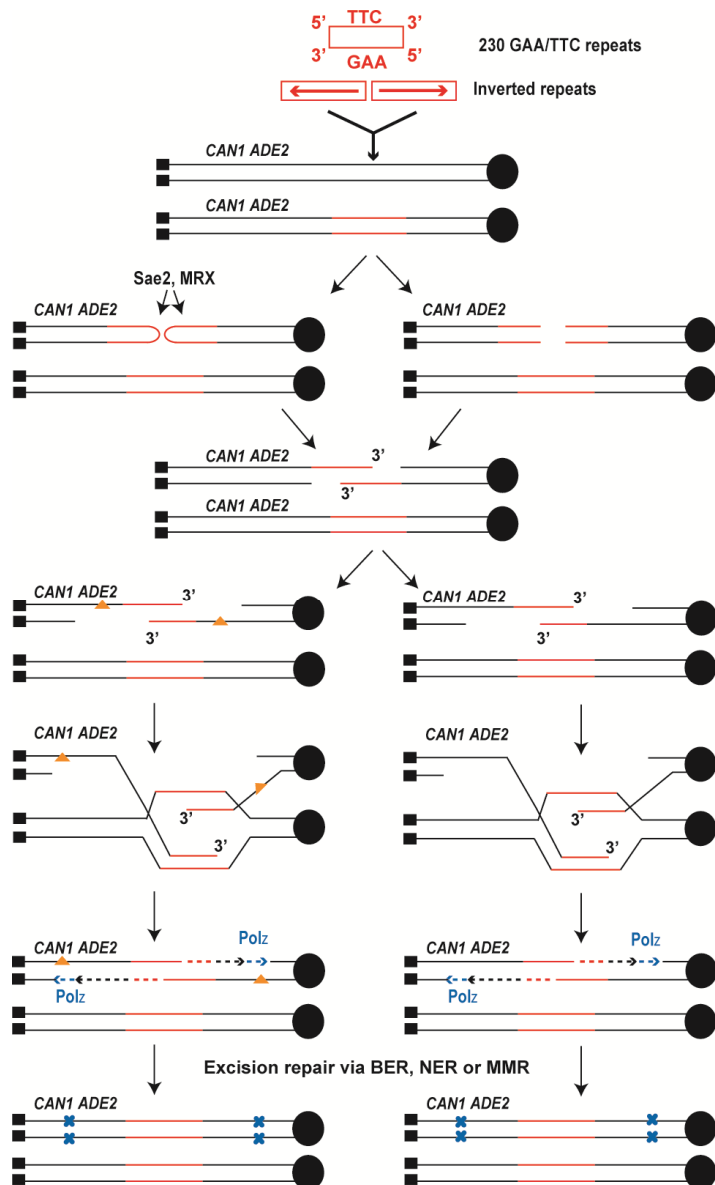


Figure 2.4 Model for mutagenesis induced by *Alu*-IRs and GAA/TTC repeats. The inverted repeats and 230 repeats of GAA/TTC inserted into *LYS2* are shown in red (not drawn to scale). Centromere (filled black circle) and telomeres (filled black squares) are also shown. Inverted repeats and GAA/TTC repeats trigger DSBs in late S or in G2 wherein the intact sister chromatid is present. The inverted repeats-induced hairpin-capped DSB are processed by Sae2 and the MRX complex (shown on the left). GAA/TTC tracts induce DSBs that have exposed 5' and 3' termini (shown on the right). Two scenarios for the accumulation of mutations are shown. On the left, ssDNA generated as a result of extensive resection can accumulate damages (orange triangles). Error-prone synthesis during the fill in reaction will lead to mutations (shown as blue x). On the right, errors can be made by Polz during synthesis across the undamaged template. Mutations will be incorporated either due to the action of mismatch repair or in next round of DNA replication (not shown).

2.5 Materials and Methods

2.5.1 Yeast Strains

The yeast strains used for the analysis of the inverted repeat-induced mutagenesis were derivatives of the KT19 strain (MATa, *bar1*- Δ , *his7-2*, *trp1*- Δ , *ura3*- Δ , *leu2-3,112*, *ade2*- Δ , *lys2*- Δ , *cup1*- Δ , *yhro54c*- Δ , *cup2*- Δ , *V34205::ADE2lys2::Alu-IRs*, *V29616::CUP1*). GAA/TTC-mediated mutagenesis was measured in strains that were derivative of YKL36 (MATa, *bar1*- Δ , *his3*- Δ , *trp1*- Δ , *ura3*- Δ , *leu2*- Δ , *ade2*- Δ , *lys2*- Δ , *V34205::ADE2lys2::(GAA)₂₃₀*). *Alu*-IRs and GAA repeats were inserted into the *Bam*HI site and the *IS50* palindrome was inserted into *Hpa*I site in *LYS2*. The strains without repeats had an intact *LYS2* gene. For measuring the distance dependence of repeat-induced mutagenesis, *URA3* was amplified from pRS306 with flanking regions for the points of insertion into chromosome V. *URA3* was inserted close to the repeats in *lys2* (586 bp TP and 352 bp TD), ~8 kb TD of the repeat locus between SGD coordinates 42096 and 42097 and ~30 kb TP between SGD coordinates 11910 and 11911 and TD between coordinates 64686 and 64687 (Figure 1, Table S6). The *pol3-P664L* allele was created via site-directed mutagenesis using p170 (Kokoska *et.al*, 1998). The mutation *P664L* results in the appearance of the *Ase*I site. The plasmid was digested with *Hpa*I and the mutation was obtained using pop-in pop-out methodology. The mutant shows mild temperature sensitivity at 37°C. The tetracycline promoter construct was obtained from Euroscarf (pCM225). PCR was performed with primers carrying overhangs for *RFA2* and *POL3* promoter regions and one-step integration was used to replace the promoters for *RFA2* and *POL3* (Table S6). *REV3* was replaced with the *kanMX* cassette in wild-type and *pol3-P664L* strain and with the *hphMX* cassette amplified from pAG32

in the TET-*POL3* and TET-*RFA2* strains {Goldstein, 1999 #235}. *SAE2* was disrupted with the *kanMX* cassette in the wild-type strains and with *TRP1* in *pol3-P664L*, TET-*POL3* and TET-*RFA2* strains.

2.5.2 GCR and mutation rates estimations

Fluctuation tests were carried out to estimate mutation and GCR rates. The strains were allowed to grow on YPD agar for 3 days at 30°C. The TET-*POL3* and TET-*RFA2* strains were grown on YPD containing 2 µg/ml and 0.1 µg/ml doxycycline, respectively. At these chosen concentrations of doxycycline the proteins are downregulated leading to an increase in fragility without significantly affecting viability of the strains. 14 individual colonies were diluted in 0.25 ml water each and serial dilutions were made to approximately 1:10000. The cultures were plated on YPD and on L-canavanine (60 mg/L) low adenine (5 mg/L) containing synthetic media in order to obtain approximately several hundred colonies per plate after incubating for 3 days at 30°C. White colonies on canavanine-containing media are indicative of mutations in *CAN1* while red colonies depict GCR events. For mutation rate estimation at *URA3*, the cultures were appropriately diluted and plated on 5-FOA (1 g/L) containing synthetic media lacking adenine. Mutation rates and 95% confidence intervals were calculated as previously described (Lobachev *et al.*, 2002).

2.5.3 DSB detection and quantification

Yeast cells were embedded into agarose plugs at a concentration of 2×10^9 cells/ml for detection of inverted-repeat-mediated DSBs and at a concentration of 8×10^9

cells/ml for detection of GAA/TTC-induced DSBs. The chromosomes were separated using contour-clamped homogeneous electric field gel electrophoresis as described previously (Zhang *et al.*, 2012) and transferred onto a nylon membrane. Southern hybridization was carried out using a probe specific to *HPA3* that is telomere-proximal to the repeats. Densitometry analysis was performed using ImageJ (NIH) and the intensity of the broken product was normalized against the intact chromosome V.

2.5.4 Sequence analysis of mutations at *CANI*

CANI mutants were obtained by plating approximately 30 individual colonies from two independent isolates for each strain on L-canavanine low adenine containing synthetic media. The Can^RAde⁺ isolates were then streaked out on YPD to obtain single colonies from which DNA was extracted. PCR was carried out using primers 60 bp upstream and 158 bp downstream of *CANI*. The PCR product was sequenced using 4 internal primers for *CANI* such that the entire gene would be covered at least twice during sequencing. The primers used for sequencing *CANI* are
can1-o1: 5'CATCTACTGGTGGTGACAAAG3';
can1-s1: 5'GCCACGGTATTTCAAAGCTTGC3';
can1-s2: 5'GGCTCTTGGAACGGATTTTC3';
can1-s3: 5'TGTAGCCATTTCAACCAAGG3'. The sequencing results are depicted in Table 4 and Tables S1-S5.

2.5.5 Estimation of Pol3 and Rfa2 expression

To quantify the changes in protein expression *POL3* was tagged with 13 copies of c-Myc-epitope tag and *RFA2* was tagged with 9 copies of the c-Myc-epitope tag at the C-terminus in both wild-type and tetracycline downregulatable strains. TET-*POL3* was grown in the presence of 2 µg/ml doxycycline overnight, while TET-*RFA2* was grown with 0.1 µg/ml doxycycline overnight. Total protein was extracted as previously described and separated using SDS-polyacrylamide gel electrophoresis on an 8% gel. After electrophoresis the gel was blotted on a nitrocellulose membrane and probed with an antibody specific to c-Myc (Genscript) and an antibody for Pgk1 (Invitrogen). The membrane was further treated with anti-mouse secondary antibody (Genscript) and chemiluminescent detection was carried out using the protocol described by GE Healthcare. Densitometry analysis was performed using ImageJ (NIH) and the intensity of Pol3 and Rfa2 were normalized against the loading control Pgk1.

2.6 Acknowledgements

We thank Dr. Katie Rudd, Sriram Vijayraghavan, Alexandra Wagner, and Anastasiya Lobacheva for critical reading of manuscript and useful suggestions and Clara Moon for initial characterization of *pol3-P664L* mutant. We are also grateful to Dr. Grzegorz Ira for providing us with the Pol3-13XMyC allele.

2.7 References

Abdulovic, A.L., Minesinger, B.K., and Jinks-Robertson, S. (2008). The effect of sequence context on spontaneous Polzeta-dependent mutagenesis in *Saccharomyces cerevisiae*. *Nucleic acids research* 36, 2082-2093.

- Belli, G., Gari, E., Aldea, M., and Herrero, E. (1998). Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. *Yeast* 14, 1127-1138.
- Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., *et al.* (2011). The genomic complexity of primary human prostate cancer. *Nature* 470, 214-220.
- Blaszak, R.T., Potaman, V., Sinden, R.R., and Bissler, J.J. (1999). DNA structural transitions within the PKD1 gene. *Nucleic acids research* 27, 2610-2617.
- Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., *et al.* (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423-1427.
- Charames, G.S., and Bapat, B. (2003). Genomic instability and cancer. *Current molecular medicine* 3, 589-596.
- Chen, C., Merrill, B.J., Lau, P.J., Holm, C., and Kolodner, R.D. (1999). *Saccharomyces cerevisiae* pol30 (proliferating cell nuclear antigen) mutations impair replication fidelity and mismatch repair. *Molecular and cellular biology* 19, 7801-7815.
- Darmon, E., Eykelenboom, J.K., Lincker, F., Jones, L.H., White, M., Okely, E., Blackwood, J.K., and Leach, D.R. (2010). *E. coli* SbcCD and RecA control chromosomal rearrangement induced by an interrupted palindrome. *Molecular cell* 39, 59-70.
- De, S., and Babu, M.M. (2010). A time-invariant principle of genome evolution. *Proceedings of the National Academy of Sciences of the United States of America* 107, 13004-13009.
- Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS biology* 9, e1000594.
- Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M., Beroukhim, R., and Getz, G. (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome research* 23, 228-235.
- Edelmann, L., Spiteri, E., Koren, K., Pulijaal, V., Bialer, M.G., Shanske, A., Goldberg, R., and Morrow, B.E. (2001). AT-rich palindromes mediate the constitutional t(11;22) translocation. *American journal of human genetics* 68, 1-13.

- Endo, K., Tago, Y., Daigaku, Y., and Yamamoto, K. (2007). Error-free RAD52 pathway and error-prone REV3 pathway determines spontaneous mutagenesis in *Saccharomyces cerevisiae*. *Genes & genetic systems* 82, 35-42.
- Frank-Kamenetskii, M.D., and Mirkin, S.M. (1995). Triplex DNA structures. *Annual review of biochemistry* 64, 65-95.
- Haber, J.E. (1999). DNA recombination: the replication connection. *Trends in biochemical sciences* 24, 271-275.
- Harper, J.W., and Elledge, S.J. (2007). The DNA damage response: ten years after. *Molecular cell* 28, 739-745.
- Harris, R.S., Longerich, S., and Rosenberg, S.M. (1994). Recombination in adaptive mutation. *Science* 264, 258-260.
- Hastings, P.J., Hersh, M.N., Thornton, P.C., Fonville, N.C., Slack, A., Frisch, R.L., Ray, M.P., Harris, R.S., Leal, S.M., and Rosenberg, S.M. (2010). Competition of *Escherichia coli* DNA polymerases I, II and III with DNA Pol IV in stressed cells. *PloS one* 5, e10862.
- Hicks, W.M., Kim, M., and Haber, J.E. (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329, 82-85.
- Hirano, Y., and Sugimoto, K. (2006). ATR homolog Mec1 controls association of DNA polymerase zeta-Rev1 complex with regions near a double-strand break. *Current biology : CB* 16, 586-590.
- Holbeck, S.L., and Strathern, J.N. (1997). A role for REV3 in mutagenesis during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* 147, 1017-1024.
- Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* 461, 1071-1078.
- Kim, H.M., Narayanan, V., Mieczkowski, P.A., Petes, T.D., Krasilnikova, M.M., Mirkin, S.M., and Lobachev, K.S. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *The EMBO journal* 27, 2896-2906.
- Kurahashi, H., and Emanuel, B.S. (2001). Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Human molecular genetics* 10, 2605-2617.
- Lang, G.I., and Murray, A.W. (2008). Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178, 67-82.

- Lengsfeld, B.M., Rattray, A.J., Bhaskara, V., Ghirlando, R., and Paull, T.T. (2007). Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Molecular cell* 28, 638-651.
- Lobachev, K.S., Gordenin, D.A., and Resnick, M.A. (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* 108, 183-193.
- Magni, G.E. (1963). The Origin of Spontaneous Mutations during Meiosis. *Proceedings of the National Academy of Sciences of the United States of America* 50, 975-980.
- McKenzie, G.J., Lee, P.L., Lombardo, M.J., Hastings, P.J., and Rosenberg, S.M. (2001). SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Molecular cell* 7, 571-579.
- Napierala, M., Dere, R., Vetcher, A., and Wells, R.D. (2004). Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. *The Journal of biological chemistry* 279, 6444-6454.
- Narayanan, V., Mieczkowski, P.A., Kim, H.M., Petes, T.D., and Lobachev, K.S. (2006). The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* 125, 1283-1296.
- Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M., and Kunkel, T.A. (2008). Division of labor at the eukaryotic replication fork. *Molecular cell* 30, 137-144.
- Northam, M.R., Robinson, H.A., Kochenova, O.V., and Shcherbakova, P.V. (2010). Participation of DNA polymerase zeta in replication of undamaged DNA in *Saccharomyces cerevisiae*. *Genetics* 184, 27-42.
- Patel, H.P., Lu, L., Blaszkak, R.T., and Bissler, J.J. (2004). PKD1 intron 21: triplex DNA formation and effect on replication. *Nucleic acids research* 32, 1460-1468.
- Pavlov, Y.I., Shcherbakova, P.V., and Rogozin, I.B. (2006). Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. *International review of cytology* 255, 41-132.
- Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., *et al.* (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS genetics* 2, e168.
- Ponder, R.G., Fonville, N.C., and Rosenberg, S.M. (2005). A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Molecular cell* 19, 791-804.

- Prakash, S., Johnson, R.E., and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annual review of biochemistry* 74, 317-353.
- Raghavan, S.C., Chastain, P., Lee, J.S., Hegde, B.G., Houston, S., Langen, R., Hsieh, C.L., Haworth, I.S., and Lieber, M.R. (2005a). Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *The Journal of biological chemistry* 280, 22749-22760.
- Raghavan, S.C., Swanson, P.C., Ma, Y., and Lieber, M.R. (2005b). Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. *Molecular and cellular biology* 25, 5904-5919.
- Rattray, A.J., Shafer, B.K., McGill, C.B., and Strathern, J.N. (2002). The roles of REV3 and RAD57 in double-strand-break-repair-induced mutagenesis of *Saccharomyces cerevisiae*. *Genetics* 162, 1063-1077.
- Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., *et al.* (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular cell* 46, 424-435.
- Sakamoto, A.N., Stone, J.E., Kissling, G.E., McCulloch, S.D., Pavlov, Y.I., and Kunkel, T.A. (2007). Mutator alleles of yeast DNA polymerase zeta. *DNA repair* 6, 1829-1838.
- Schwartz, M., Zlotorynski, E., and Kerem, B. (2006). The molecular basis of common and rare fragile sites. *Cancer letters* 232, 13-26.
- Shah, K.A., Shishkin, A.A., Voineagu, I., Pavlov, Y.I., Shcherbakova, P.V., and Mirkin, S.M. (2012). Role of DNA polymerases in repeat-mediated genome instability. *Cell reports* 2, 1088-1095.
- Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Molecular cell* 35, 82-92.
- Stamatoyannopoulos, J.A., Adzhubei, I., Thurman, R.E., Kryukov, G.V., Mirkin, S.M., and Sunyaev, S.R. (2009). Human mutation rate associated with DNA replication timing. *Nature genetics* 41, 393-395.
- Strathern, J.N., Shafer, B.K., and McGill, C.B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics* 140, 965-972.
- Tang, W., Dominska, M., Gawel, M., Greenwell, P.W., and Petes, T.D. (2013). Genomic deletions and point mutations induced in *Saccharomyces cerevisiae* by the

trinucleotide repeats (GAA.TTC) associated with Friedreich's ataxia. *DNA repair* 12, 10-17.

- Tang, W., Dominska, M., Greenwell, P.W., Harvanek, Z., Lobachev, K.S., Kim, H.M., Narayanan, V., Mirkin, S.M., and Petes, T.D. (2011). Friedreich's ataxia (GAA)_n*(TTC)_n repeats strongly stimulate mitotic crossovers in *Saccharomyces cerevisiae*. *PLoS genetics* 7, e1001270.
- Wang, G., and Vasquez, K.M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13448-13453.
- Wyman, C., and Kanaar, R. (2006). DNA double-strand break repair: all's well that ends well. *Annual review of genetics* 40, 363-383.
- Yang, Y., Sterling, J., Storici, F., Resnick, M.A., and Gordenin, D.A. (2008). Hypermutable of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLoS genetics* 4, e1000264.
- Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics* 10, 451-481.
- Zhang, Y., Saini, N., Sheng, Z., and Lobachev, K.S. (2013). Genome-wide screen reveals replication pathway for quasi-palindrome fragility dependent on homologous recombination. *PLoS genetics* 9, e1003979.
- Zhang, Y., Shishkin, A.A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M., and Lobachev, K.S. (2012). Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. *Molecular cell* 48, 254-265.
- Zhu, Z., Chung, W.H., Shim, E.Y., Lee, S.E., and Ira, G. (2008). Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* 134, 981-994.

Disclaimer

This chapter was published under the same name in the journal *PLoS Genetics*

Saini N., Zhang Y., Nishida Y., Sheng Z., Choudhury S., Mieczkowski P. and Lobachev

K.S. (2013) "Fragile DNA motifs trigger mutagenesis at distant chromosomal loci."

PLOS Genetics, doi: 10.1371/journal.pgen.1003551

CHAPTER 3

MIGRATING BUBBLE DURING BREAK-INDUCED REPLICATION DRIVES CONSERVATIVE DNA SYNTHESIS

3.1 Introduction, Results and Discussion

The repair of chromosomal double strand breaks (DSBs) is crucial for the maintenance of genomic integrity. However, the repair of DSBs can also destabilize the genome by causing mutations and chromosomal rearrangements, the driving forces for carcinogenesis and hereditary diseases. Break-induced replication (BIR) is one of the DSB repair pathways that is highly prone to genetic instability (Deem *et al.*, 2008; Llorente *et al.*, 2008; Malkova and Haber, 2012). BIR proceeds by invasion of one broken end into a homologous DNA sequence followed by replication that can copy hundreds of kilobases of DNA from a donor molecule all the way through its telomere (Davis and Symington, 2004; Malkova and Haber, 2012). The resulting repaired chromosome comes at a great cost to the cell, as BIR promotes mutagenesis, loss of heterozygosity, translocations, and copy number variations, all hallmarks of carcinogenesis (Bosco and Haber, 1998; Davis and Symington, 2004; Deem *et al.*, 2011; Hastings *et al.*, 2009; Malkova *et al.*, 2005; Payen *et al.*, 2008). BIR uses most known replication proteins to copy large portions of DNA, similar to S-phase replication (Lydeard *et al.*, 2007; Lydeard *et al.*, 2010). It has thus been suggested that BIR proceeds by semiconservative replication; however, the model of a bona fide, stable replication fork contradicts the known instabilities associated with BIR such as a 1,000-

fold increase in mutation rate compared to normal replication (Deem *et al.*, 2011). Here we demonstrate that in budding yeast the mechanism of replication during BIR is significantly different from S-phase replication, as it proceeds via an unusual bubble-like replication fork that results in conservative inheritance of the new genetic material. We provide evidence that this atypical mode of DNA replication, dependent on Pif1 helicase, is responsible for the marked increase in BIR-associated mutations. We propose that the BIR mode of synthesis presents a powerful mechanism that can initiate bursts of genetic instability in eukaryotes, including humans.

Theoretically, BIR might constitute a unidirectional, bona fide replication fork producing two semiconservatively replicated molecules (Lydeard *et al.*, 2010; Malkova *et al.*, 2005) (Figure 3.1 a, i). Alternatively, a D-loop formed by invasion of the broken chromosome may persist throughout BIR, migrating down the length of the chromosome, creating an unusual condition of conservative inheritance of newly synthesized DNA (Llorente *et al.*, 2008; Malkova and Ira, 2013; Smith *et al.*, 2007) (Figure. 3.1 a, ii-iv).

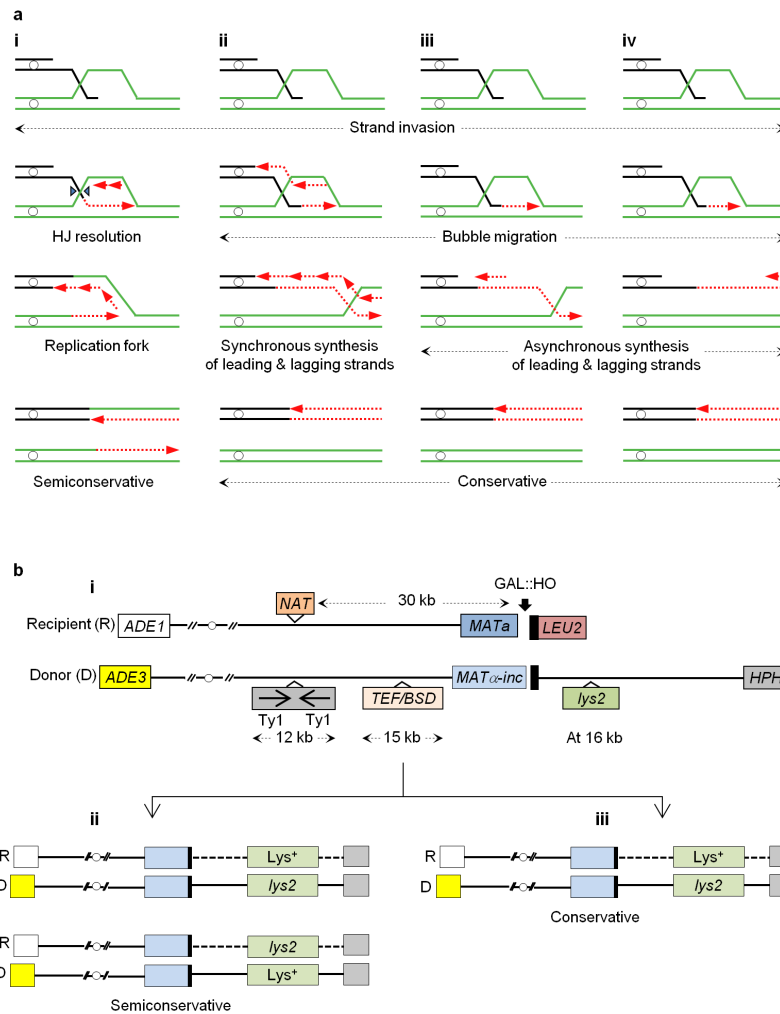


Figure 3.1 The mode of DNA synthesis during BIR. a, The models of BIR. **i**, Replication fork proceeds semiconservatively. **ii-iv**, Migrating bubble leads to conservative inheritance of new DNA. Synchronous (**ii**) and asynchronous (**iii,iv**) synthesis of leading and lagging DNA strands. **b**, **i**, The BIR frameshift mutation assay. A DSB is induced at *MAT α* of the recipient chromosome III. *lys2* reporter is inserted in the donor chromosome 16 or 36 kb telomere-proximal from *MAT α -inc*. *Lys*⁺ mutations would be inherited equally by the donor (D) or recipient (R) if BIR is semiconservative (**ii**), but only by recipient if BIR is conservative (**iii**).

To distinguish between these models, we used a disomic yeast system (Fig. 3.1 b, i) containing a second, truncated copy of chromosome III, cleaved by HO endonuclease under control of a galactose-inducible promoter (Deem *et al.*, 2008). The HO-induced DSB possesses only one efficiently repairable end that invades the second copy of

chromosome III, and initiates BIR that copies over 100 kilobases (kb) of the distal part of the chromosome. Using this system, we recently demonstrated that BIR stimulates mutagenesis along the path of DNA synthesis at a series of *lys2* frameshift reporters (Deem *et al.*, 2011). Here we examined these Lys⁺ mutations to determine whether errors during BIR were acquired semiconservatively (inherited by either the donor or recipient molecule; Figure 3.1 b, ii) or conservatively (inherited only by the recipient molecule; Figure 3.1 b, iii). Pulse-field gel electrophoresis (PFGE) was used to separate donor and recipient molecules from Lys⁺ BIR outcomes resulting from mutations in a *lys2* reporter located 16 or 36 kb distal to the site of BIR initiation (Figure. 3.2 a, b). Sequencing of the polymerase chain reaction (PCR) products derived from the separated chromosomes revealed that the great majority of heterozygous frameshift mutations (58 of 58 and 68 of 77 from strains with reporters at 16 and 36 kb, respectively) were inherited by the recipient molecule, whereas the donor sequence remained unchanged. Overall, the mutation pattern supports a conservative replication mechanism for BIR. However, because this conclusion was based on analysis of selected mutation events, we developed a non-selective test to analyse BIR microscopically by DNA combing.

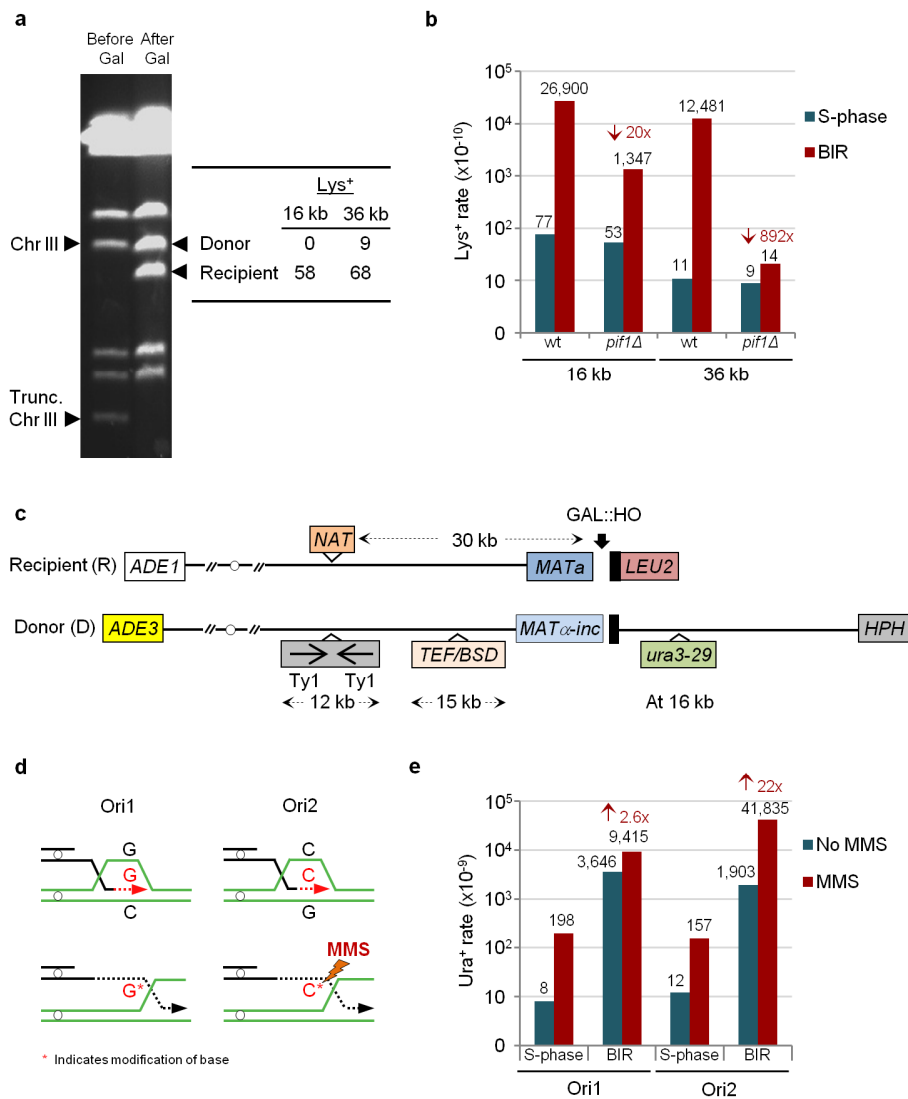


Figure 3.2 BIR-induced mutations. **a**, The sequencing of the separated donor and recipient chromosomes of heterozygous Lys^+ mutants. **b**, The effect of $\Delta pif1$ on BIR-induced frameshifts. Medians of mutation rates are shown. The arrows represent a reduction as compared to wild type (WT). **c**, The assay to study BIR-induced base substitutions in *ura3-29* reporter. **d**, Depending on orientation, the selectable position of *ura3-29* leading strand includes cytosine (C) or guanine (G). **e**, MMS amplifies BIR-induced base substitutions in an orientation-dependent way. The arrows indicate an increase as compared to no-MMS control. Statistical analysis and for the ranges of medians shown in **e** and **b**.

The experiments were conducted in nocodazole-arrested cells of disomic BIR strain bearing a cassette facilitating BrdU incorporation in yeast (Viggiani and Aparicio,

2006) (Figure. 3.3 a, b). BrdU was added 3.5 h after DSB induction. After completion of BIR, PFGE-separated donor and recipient molecules (Figure. 3.3 c and Figure. B.1) were analysed by molecular combing and fluorescent *in situ* hybridization (FISH). We used an anti-BrdU antibody, the P1 probe specific to the tandem repeat of *TEF1/BSD* inserted 14 kb centromere-proximal to *MAT* in the donor chromosome, the P2 probe specific to the 20-kb region of chromosome III where invasion occurs, and the P3 probe specific to the 15-kb region near the telomere (Figure. 3.3 a) to characterize BIR. We observed BrdU tracts approximately 100 kb in length in 70 of the 98 repaired recipient molecules analysed (Figure. 3.3 d, e and Figure B.2 a). These tracts include the entire chromosome region marked by P2 and P3 and, therefore, represent BIR that copied the donor chromosome through to its telomere. Additionally, 14% of recipient molecules contained long (>30 kb) BrdU tracts that overlapped with P2 but not with P3 (Figure. 3.3 d and Figure B.3 a). These molecules probably represent repair events where BIR was interrupted, resulting in half-crossover formation (Deem *et al.*, 2008; Smith *et al.*, 2009).

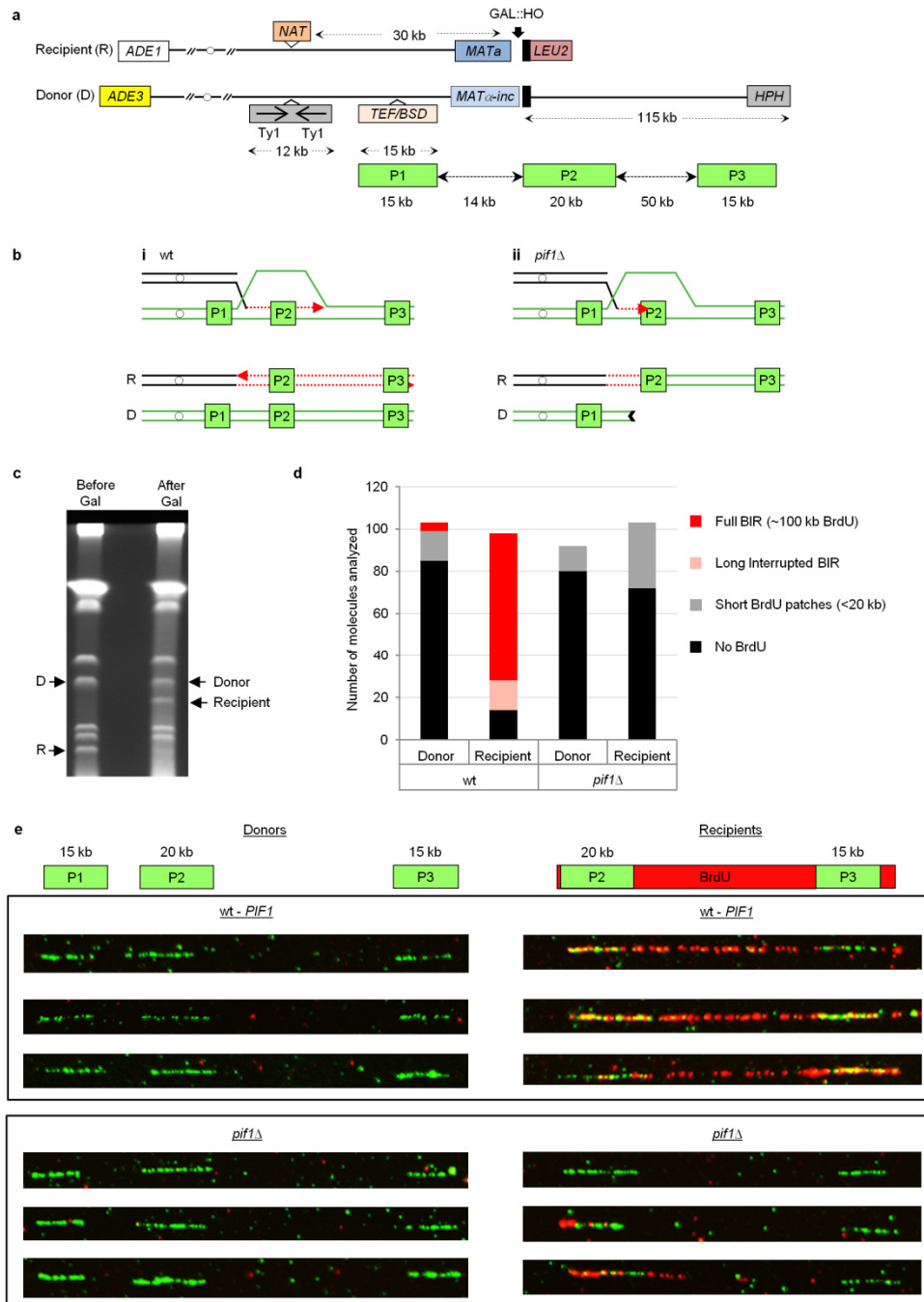


Figure 3.3 DNA synthesis during BIR is conservative. **a**, Experimental system to assay BIR using dynamic molecular combing including the position of hybridization probes P1, P2 and P3. **b**, Left: BrdU incorporation in the recipient is expected from conservative BIR (red). Right: formation of half-crossovers in $\Delta pif1$ leads to short patches of BrdU in the recipient. **c**, Donor and recipient chromosomes separated using PFGE. **d**, The summary of molecular combing analysis. **e**, The donors and recipients of wild-type (*PIF1*) and $\Delta pif1$. Each molecule was hybridized with P1, P2 and P3 probes (green tracts) and treated with anti-BrdU antibodies (red tracts).

Our analysis of donor molecules supports a conservative mode of DNA replication during BIR, as only 4 out of 103 donor molecules were illuminated by >30 kb BrdU tracts (Figure. B.3 d, e and Figure B.2 a). These data confirm a strong bias ($P < 0.0001$) towards BrdU tracts present only in the recipient chromosome. The four cases of BrdU incorporation in the donor could result from rare semiconservative synthesis or from BIR initiated >30 kb proximal to the DSB site, which would result in a donor-like size and hybridization pattern due to copying of regions unique to the donor molecule (VanHulle *et al.*, 2007). On the basis of these data, we estimate that, even if semiconservative synthesis occurs, it can account for no more than 8% of the BIR events that we analysed (see Figure.B.4 for the results of another series of experiments supporting this conclusion).

The unusual mode of replication prompted us to characterize the structure of BIR molecular intermediates at *LYS2* inserted ~16 kb from the point of strand invasion. Genomic DNA extracted from nocodazole-arrested cells undergoing BIR was digested with PstI and analysed by two-dimensional (2D) gel electrophoresis using a *LYS2*-specific probe (Figure. 3.4 a, top panel). We detected bubble-like structures between 3 and 7 h after DSB induction (Figure. 3.4 b–d), but not at 13 h, consistent with the timing of BIR progression (Deem *et al.*, 2011) (Figure B.5). All bubble-like intermediates were markedly different from the Y structures indicative of S-phase replication forks observed before addition of nocodazole and induction of the break (Figure. 3.4 c, 0 h). Furthermore, no bubble-like structures were observed in control strains in which HO endonuclease cannot initiate a DSB (Fig. 3.4 d, no-cut), thus linking these structures to BIR exclusively. The bubble-like structures observed in BIR were reminiscent of

bubbles routinely detected at replication origins (Fangman and Brewer, 1991), with one important difference: the BIR bubbles included a long, high-molecular-mass tail that extended well beyond the size expected for complete replication (arrows in Figure. 3.4 c, d). We proposed that initiation of BIR lagging-strand synthesis is often delayed compared to leading strand, resulting in accumulation of single-stranded DNA (ssDNA) behind the BIR bubble, which makes the region around *LYS2* refractory to PstI digestion. Indeed, pre-incubation of genomic DNA with oligonucleotides (PstO3 and PstO4; Figure. 3.4 a, middle and bottom panels) complimentary to the Watson strand of two PstI sites flanking the *LYS2* gene eliminated the tail and resulted in a second arc that probably corresponds to molecular intermediates with bubbles consisting of one double-stranded branch (leading-strand synthesis) and one single-stranded branch (lagging-strand synthesis) (Figure 3.4 a, b, d and Figure B.6). Similar results were also obtained using BglII digestion (Figure B.7). Notably, whereas simultaneous addition of oligonucleotides BglO3 and BglO4, complimentary to the Watson strand of two BglII sites, eliminated the ssDNA tail, the addition of each of these oligonucleotides individually failed to eliminate the tail. This confirms that two types of DNA intermediates contribute to the observed ssDNA tail: those containing ssDNA centromere proximal to *LYS2* and those with ssDNA distal to *LYS2* (Figure 3.4 a and Figure B.7 a, panels ii, iii). Addition of oligonucleotides complimentary to the Crick strand did not have any effect (data not shown). Bubble migration intermediates were also detected with an *HPH*-specific probe that hybridizes to the end of the donor chromosome (Figure. 3.4 a, e). These data strongly support a migrating D-loop type of DNA replication (Ferguson and Holloman, 1996; Formosa and Alberts, 1986).

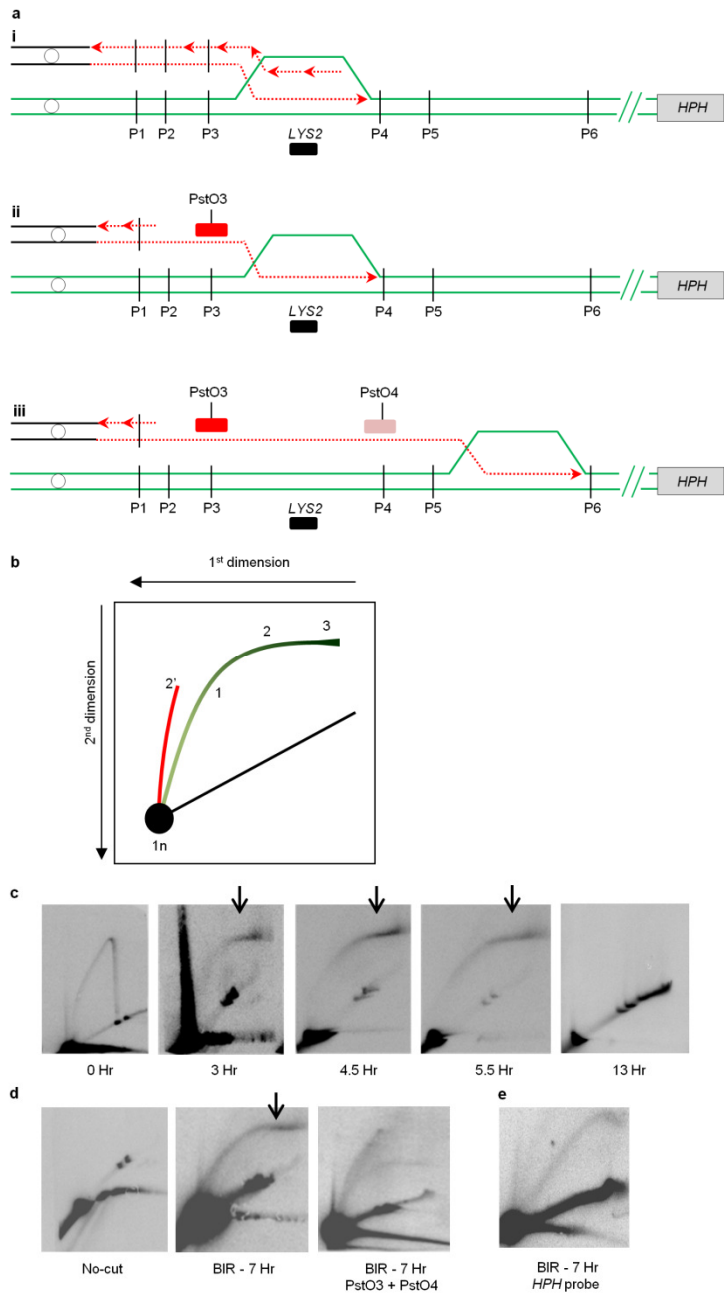


Figure 3.4 Molecular intermediates of BIR. **a**, D-loop migration during coordinated (top) and uncoordinated (bottom and middle) leading- and lagging-strand synthesis. **b**. Schematic of 2D gel with BIR bubbles forming an arc (1, 2) with an extension (3) representing ssDNA tail. Annealing with PstO3 and PstO4 allows PstI digestion changing the mobility of the intermediate (red, 2'). **c**, 2D analysis of Y-arc during normal replication (0 h) and bubble-like structures at time points after BIR induction hybridized to *LYS2*-specific probe. Similar bubble structures were observed in other nine independent experiments. **d**, High-molecular-mass tails (arrows) disappear after annealing with PstO3 and PstO4. The arc is absent in no-cut controls. **e**, BIR intermediates highlighted with *HPH*-specific probe.

We proposed that ssDNA accumulated behind the migrating BIR bubble is the cause of BIR-associated mutagenesis because of the propensity of ssDNA to accumulate unrepaired DNA lesions (Yang *et al.*, 2008). This was tested by using methyl methanesulphonate (MMS), a DNA damaging agent that predominantly creates mutagenic lesions in cytosines of ssDNA (Roberts *et al.*, 2012; Yang *et al.*, 2010). In addition, a *ura3-29* reporter (Shcherbakova and Pavlov, 1996), which can revert to Ura⁺ via three different base substitutions at one C•G pair (Figure. 3.2 c), was inserted in the donor chromosome in two different orientations (Ori1 and Ori2). We expected that MMS will specifically elevate the level of BIR-associated mutagenesis in Ori2, where cytosine is located in the mutant position of the leading (ssDNA) strand, but not in Ori1, which contains guanine instead (Figure. 3.2 d). Indeed, we observed that even though BIR markedly stimulated base substitutions in *ura3-29* irrespective of its orientations, the effect of MMS was orientation dependent (Figure. 3.2 e and Table B.1). Specifically, MMS highly amplified BIR-induced mutagenesis in cells containing *ura3-29* in Ori2, whereas its effect on BIR mutagenesis in Ori1 was relatively modest. This observation supports the conjecture that ssDNA accumulated behind the BIR bubble is the cause of BIR-associated mutagenesis. Additionally, the spectrum of BIR-induced mutations was also orientation dependent, supporting our conclusion (Figure B.2 b).

Because the Pif1 helicase is a key component of the BIR machinery (Chung *et al.*, 2010), we proposed that Pif1 is essential for long-range BIR. We observed that even though BIR-sized products were formed in $\Delta pif1$ mutant (Figure B.1 a, b), no extended BrdU tracts were observed in either the donor or recipient chromosomes (Figure. 3.3 d,

e). In addition, approximately 22% of recipient molecules contained short (<20 kb) BrdU patches that co-localized with probe P2 (Figure. 3.3 d, e and Figure B.2 a) and probably represented DNA synthesis that was prematurely terminated. Therefore, it is likely that most outcomes in *Δpif1* mutants formed during the time frame of these experiments were half-crossovers (Figure B.1 c), supporting our hypothesis that Pif1 is required for BIR-associated DNA synthesis. The low amount of BIR precluded 2D analysis of BIR intermediates in *Δpif1*. We investigated whether Pif1 may be necessary for BIR-induced frameshift mutations. Notably, we observed that all BIR-induced frameshift mutations were eliminated in the *Δpif1* mutant at the 36 kb position, and there was a 20-fold reduction in frameshift mutations at the 16 kb position (Figure. 3.2 b and Supplemental Table B.2). Thus, whereas BIR may initiate in the absence of Pif1, these data support that Pif1 is required for long-range synthesis during BIR. Therefore, Pif1 can be added to the list of other previously identified proteins, including Pol δ , Pol ζ , Msh2, Mlh1, Dun1 and others that are involved in BIR and associated mutagenesis (Deem *et al.*, 2008; Deem *et al.*, 2011; Lydeard *et al.*, 2007; Smith *et al.*, 2009).

Overall, the results of this study demonstrate that BIR is carried out by a migrating replication bubble driven by Pif1 with asynchronous synthesis of leading and lagging strands resulting in a mutation-prone accumulation of ssDNA, and leads to conservative inheritance of the new genetic material. The bubble migration mechanism and associated mutagenesis may be relevant to cellular processes where BIR has been implicated, such as alternative telomere lengthening and mitochondrial maintenance (Hashimoto and Costanzo, 2011; Kreuzer *et al.*, 1995; Kuzminov, 1995; Le *et al.*, 1999; Lydeard *et al.*, 2007; Pohjoismaki and Goffart, 2011), where Pif1 has an important role.

An intriguing possibility is that the burst of mutations recently linked to replication stress/fork collapse in pre-cancerous cells (Halazonetis *et al.*, 2008) may be linked to conservative synthesis initiated by BIR.

3.2 Materials and Methods

3.2.1 Media and strains

All yeast strains (Table B.3) were isogenic to AM1003 (Deem *et al.*, 2008) which is a chromosome III disome with the following genotype: *hmlΔ::ADE1/hmlΔ::ADE3 MATα-LEU2-tel/MATΔ-inc hmrΔ::HPH FS2Δ::NAT/FS2 leu2/leu2-3,112 thr4 ura3-52 ade3::GAL::HO ade1 met13*.

AM1291 and AM1482 are derivatives of AM1003 and were created by deleting *LYS2* from its native location, and inserting *lys2-Ins A₄(A₄)* at different positions of chromosome III (Deem *et al.*, 2011). AM2191 and AM 2198 were constructed from AM1291 and AM1482 by replacement of *PIF1* with *KANMX* module (Wach *et al.*, 1994). Control strains AM1449, AM1649, AM2247 and AM2257, which contained no HO cut site in the recipient chromosome III, were obtained from AM1291, AM1482, AM2191 and AM2198 as previously described (Deem *et al.*, 2011). AM2439 and AM2438 were created by integrating three and two copies of *TEF1/BSD-snt1* into *SNT1* of AM1291 and AM1482, respectively. The *TEF1/BSD-snt1* plasmid was constructed by cloning of a PCR-amplified 1-kb region of *SNT1* (from 185626 to 186589 positions of chromosome III) into the BamHI/HindIII fragment of *TEF1/BSD* (Invitrogen). The resulting plasmid was linearized by SnaBI and integrated at *SNT1* to introduce a donor-specific region into the *MATα-inc* containing copy of chromosome III. The selection of

transformants with integration of multiple copies of the plasmid was achieved by PFGE followed by Southern hybridization with *TEF1/BSD* used as a probe. AM2118 was isogenic to AM1247, but contained *KANMX* module at chromosome II between *PTC4* and *TPS1*.

AM2110 is a derivative of AM1003, and was created by deleting *URA3* (using delitto perfetto approach) and replacing *hmr::HPH* with *hmr::KANMX*. In addition, it contains *lys2-InsA₄* inserted at *SED4* (36 kb centromere distal to *MAT α -inc*). AM2161 and AM2820 were derivatives of AM2110 where *ura3-29-HPH* fragments (Ori1 and Ori2 respectively) were inserted 16 kb centromere distal to *MAT α -inc* between *RSC6* and *THR4*. The *ura3-29-HPH* cassettes containing *ura3-29* allele (Shcherbakova and Pavlov, 1996) in two orientations were a gift from Y. Pavlov. The insertion of *ura3-29-HPH* 16 kb centromere distal to *MAT α -inc* was achieved by transformation of AM2110 with DNA fragments generated by PCR amplification of *ura3-29-HPH* using the following primers with targeting tails (uppercase) and *ura3-29-HPH* amplification sequence (lower case): 5'-

TCTTTCTGCAATTATTGCACGCCTCCTCGTGAGTAGTGACCGTGCGAACAAAA

GAGTCATTACAACGAGGAAATAGAAGAagtcagtgagcgaggaagc-3' and 5'-

ATATTTGCTGCTATACTACCAAATGGAAAAATATAAGATACACAATATAGAT

AGTATTAAAAAACGTGTATACGTTATTattgtactgagagtcacc-3'. Control strains

AM2442, AM2259 and AM2842, which contained no HO cut site in the recipient chromosome III, were obtained from AM2118, AM2161 and AM2820 as previously described (Deem *et al.*, 2011).

AM2406 is a derivative of AM1003 that was constructed by inserting BrdU cassette (with the human equilibrative nucleoside transporter (hENT1) and the herpes simplex virus thymidine kinase) (Viggiani and Aparicio, 2006) into *URA3* to facilitate efficient BrdU incorporation in yeast. In particular, the p306-BrdU plasmid was linearized with *StuI* and inserted by transformation into the *URA3* gene (chromosome V). In addition, AM2406 contained insertion of three tandem arrays of the *TEF1/BSD-snt1* at *SNT1*, and replacement of *TPS1* with a *KANMX* module. *TPS1* was deleted to reduce accumulation of trehalose, which interfered with DNA purification.

YEP-raffinose, YEP-lactate and YEP-galactose were made as described (Chung *et al.*, 2010; Deem *et al.*, 2011). Cultures were grown at 30°C.

3.2.2 Analysis of BIR efficiency

DSBs were initiated by HO induction by addition of galactose (Deem *et al.*, 2008). BIR efficiency was determined genetically and by physical analysis in time-course experiments using PFGE as previously described (Deem *et al.*, 2008). The average efficiency of BIR at each time point was calculated based on results of four independent experiments.

3.2.3 2D analysis of molecular intermediates of BIR

Cells were grown overnight in synthetic leucine drop-out media, transferred to YEP-raffinose, and incubated for ~16 h, until cell density reached $\sim 1 \times 10^7$ cells ml⁻¹. An aliquot was taken for analysis of the S-phase replication fork, and 2% galactose was added to induce HO endonuclease in the remainder of the culture. In these experiments, the efficiency of BIR was $80 \pm 15\%$, as determined by PFGE analysis 10 h after DSB (Figure B.5 c). DSB induction led to G2/M arrest ~3 h after galactose addition as cells

were in the process of completing BIR repair (Figure B.5 d). At this point, nocodazole was added to the culture to a final concentration of 0.015 mg ml^{-1} to maintain the arrest. Cells were collected at different intervals following the break and subjected to psoralen crosslinking that allowed one to constrain branch migration during DNA purification as previously described (Oh *et al.*, 2009). Chromosomal DNA was extracted and neutral/neutral 2D analysis was carried out (Friedman and Brewer, 1995). PstI-digested DNA was separated in the first dimension on a 0.4% gel without ethidium bromide in $1\times$ TBE buffer at 1 V cm^{-1} for 22 h. The second dimension was run at 6 V cm^{-1} in $1\times$ TBE buffer containing $0.3 \text{ }\mu\text{g ml}^{-1}$ ethidium bromide for 12 h.

Alternatively, to guarantee that the observed intermediates do not result from mechanical stress during genomic DNA preparation, we conducted 2D-gel electrophoresis using chromosomal DNA embedded in agarose plugs. In particular, cells collected at different intervals after induction of BIR were treated with psoralen as described previously (Oh *et al.*, 2009). The cells were then re-suspended in $750 \text{ }\mu\text{l}$ solution of 1 M sorbitol, 0.5 M EDTA (pH 8) and treated with 0.2 mg ml^{-1} lyticase for 1 h at $37 \text{ }^\circ\text{C}$. The spheroplasts were washed in a solution of 50 mM Tris, 50 mM EDTA and 100 mM NaCl. The spheroplasts were then embedded in 0.8% low melt agarose at a concentration of $1.5 \times 10^{10} \text{ cells ml}^{-1}$. The chromosomal DNA embedded in agarose was digested with BglII, and 2D-gel electrophoresis was carried out as described for the 2D analysis of PstI-digested chromosomal DNA.

To identify regions of single-stranded DNA, a PstI or BglII digest was preceded by pre-incubation of genomic DNA with oligonucleotides that were complimentary to the PstI or BglII sites flanking the *LYS2* gene and had the following sequences: 5'-

GGTCGCCCTGCAGCACAAGC-3' (PstO3), 5'-
GTCCTTTCCAGATCTTGGCAACTTT-3' (BglO3), 5'-
GCTTGTGCTGCAGGGCGACC-3' (PstO5), 5'-
AAAGTTGCCAAGATCTGGAAAGGAC-3' (BglO5), where 'O3' and 'O5' indicate
oligonucleotides that are complimentary to the Watson and Crick strands at the
centromere-proximal site, respectively; and 5'-TAGATGGCTGCAGAACCAGT-3'
(PstO4), 5'-TGGATCTGGTAGATCTGTAAACTTGG-3' (BglO4), 5'-
ACTGGTTCTGCAGCCATCTA-3' (PstO6), 5'-
CCAAGTTTACAGATCTACCAGATCCA-3' (BglO6), where 'O4' and 'O6' indicate
oligonucleotides that are complimentary to the Watson and Crick strands at the telomere-
proximal site, respectively.

Southern hybridization was performed using *LYS2* fragment obtained by PCR amplification of a 0.6 kb region of *LYS2* (from 471835 to 472443 kb positions of chromosome II) or using *HPH*-hybridizing fragment obtained by PCR amplification of *HPH* from the pAG32 plasmid.

Along with analysis of BIR intermediates, cell cycle distribution was analysed by flow cytometry and BIR kinetics were analysed by PFGE. For PFGE, chromosome plugs were prepared with genomic DNA embedded in plugs of 1% low-melting agarose and separated at 6 V cm^{-1} for 40 h using the CHEF DRII apparatus. PFGE was followed by Southern analysis with an *ADE1*-specific probe labelled with P^{32} . Images were analysed using a Molecular Dynamics PhosphorImager.

3.2.4 DNA combing and fluorescent *in situ* hybridization

Cells were grown overnight in synthetic leucine drop-out media, transferred to YEP-lactate, and incubated for ~20 h, until cell density reached $\sim 1 \times 10^7$ cells ml⁻¹. Cells were arrested by nocodazole added to 0.015 mg ml⁻¹, and DSBs were induced 2.5 h later by addition of galactose to the final concentration of 2%. When experiments were performed according to this protocol, the efficiency of BIR was $54.0 \pm 9.8\%$, as determined by PFGE analysis (Deem *et al.*, 2008) 11 h after DSB induction (Supplemental Figure B.1a, b). BrdU was added to the culture 3.5 h after DSB induction by galactose to the final concentration of 0.4 mg ml⁻¹ after all normal DNA replication was completed but before the beginning of BIR. Aliquots were removed to embed cells into agarose plugs before and 11 h after induction of DSBs with galactose. In experiments involving *Δpif1* strains, the analysis was performed 13 h after DSB induction due to slower kinetics of DSB repair in *Δpif1* (data not shown). The uniform arrest of cells at G2/M was confirmed by the absence of BrdU incorporation in any chromosomes other than chromosome III, which was assayed by PFGE analysis of yeast chromosomes extracted from samples taken before the addition of BrdU and 11 or 13 h after DSB induction and probing with anti-BrdU antibodies.

Genomic DNA preparation and molecular combing were performed as described (Conti *et al.*, 2001). Colour hybridization of chromosome III molecules was performed using three fluorescent probes. P1 probe was prepared using the *TEF1/BSD* plasmid (Invitrogen) and hybridized to the 15-kb region containing three tandem repeats of the *TEF1/BSD-snt1* plasmid inserted into the donor copy of chromosome III at position 186535. P2 probe marked the position close to strand invasion during BIR and was

comprised of a set of four 5-kb fragments that corresponded to the following positions on the donor chromosome III: 200205 to 205140, 205117 to 210385, 210361 to 215385, and 215361 to 220337. The P3 probe highlights the region close to the telomeric end of chromosome III and is made up of three 5-kb fragments corresponding to the following positions on the donor chromosome: 274778 to 279801, 279778 to 284814 and 284791 to 289782. The probes were made by PCR amplification of genomic DNA from AM2406. Nucleotide sequences of the primers used to generate fragments for labelling are available upon request. Probes were labelled with biotin-dUTP. Hybridization and fluorescent detection of combed DNA molecules were achieved according to protocols described (Conti *et al.*, 2001) with a few modifications. Successive layers of fluorophore-conjugated antibodies diluted in 1× PBST (1× PBS + 0.05% Tween) were used. For the biotin-conjugated probes, the following series was used at a dilution of 1:4,000: (1) Alexa-488-Streptavidin (Molecular Probes; Life Technologies, catalogue no. 32354); (2) biotinylated antistreptavidin (from Vector Lab, catalogue no. BA-0500); (3) Alexa-488-streptavidin; (4) biotinylated anti-streptavidin; and (5) Alexa-488-Streptavidin. To detect BrdU incorporation, the following series were used at the indicated dilutions: (1) 1:20 dilution of mouse anti-BrdU (BD Biosciences, catalogue no. 347580); (2) 1:50 dilution of Cy3-coupled rat anti-mouse (Jackson ImmunoResearch Lab, catalogue no. 415-165-166); and (3) 1:50 dilution of Cy3-mouse anti-rat (Jackson ImmunoResearch Lab, catalogue no. 212-165-168). All images were acquired using the Zeiss LSM 510 Confocal Microscope with 100× objective. The lengths of the fluorescent stretches were calculated by comparison with the length of P1, P2 and P3 hybridization signals.

The statistical comparison between donor and recipient chromosomes in respect to BrdU incorporation was performed using the Chi-square test. For each experiment, the frequency of semiconservative BIR (F) was calculated as follows: $F = A/N \times f \times b$, where A represents the number of donor molecules with long BrdU tracts; N represents the total number of analysed donor molecules; f represents the efficiency of BIR in the experiment (calculated by physical analysis as a percentage of the truncated chromosome III converted in the BIR product); and b represents the fraction of recipient molecules containing full and long interrupted BIR tracts.

3.2.5 Mutagenesis associated with DSB repair

To determine mutation frequency associated with BIR, yeast strains were grown from individual colonies with agitation in liquid synthetic media lacking leucine for approximately 20 h, diluted 20-fold with fresh YEP-Lac, and grown to logarithmic phase for approximately 16 h. Next, 20% galactose was added to the culture to a final concentration of 2%, and cells were incubated with agitation for 7 h. Samples from each culture were plated at appropriate concentrations on adenine drop-out media and on media omitting lysine and adenine before (0 h) and 7 h after the addition of galactose (7 h) to measure the frequency of Lys^+ cells. To measure the frequency of Ura^+ cells, samples were plated at appropriate concentrations on adenine drop-out media and on media omitting uracil and adenine before (0 h) and 7 h after the addition of galactose (7 h). To determine spontaneous mutation frequencies, no-DSB strains were grown similarly to the DSB-containing strains. Because spontaneous mutation frequencies were calculated based on the number of mutations accumulated during many cell generations,

the rate of spontaneous mutagenesis in no-DSB control strains was calculated using the following modification of Drake equation: $\mu = 0.4343 f / \log(N\mu)$, where μ is the rate of spontaneous mutagenesis, f is mutation frequency, and N is the number of cells in yeast culture. The rate of mutations after galactose treatment (μ_7) was determined using a simplified version of the Drake equation: $\mu_7 = (f_7 - f_0)$, where f_7 and f_0 are the mutation frequencies among Ade⁺ cells at times 7 h and 0 h, respectively. This modification was necessary because experimental strains did not divide or underwent ≤ 1 division between 0 h and 7 h.

MMS was added at 1.5 mM 30 min after galactose addition. Cells were incubated with agitation for 7 h, treated with 10% sodium thiosulphate to inactivate MMS, diluted and plated. The loss of viability after MMS treatment was barely detectable and never exceeded 40% independently of *ura3-29* orientation. The rate of mutations following MMS treatment was determined using a simplified version of the Drake equation: $\mu_7 = (f_7 - f_0)$, where f_7 and f_0 are the mutation frequencies among Ade⁺ cells at times 7 h (following MMS treatment) and 0 h, respectively. This modification was necessary because experimental strains did not divide or underwent ≤ 1 division between 0 h and 7 h in the presence of MMS.

Rates are reported as the median value and the 95% confidence limits for the median are calculated for the strains with a minimum of six individual experiments. For strains with four–five individual experiments, the range of the median was calculated. Statistical comparisons between median mutation rates were performed using the Mann–Whitney *U*-test.

3.2.6 Analysis of BIR-induced Lys⁺ mutants

Lys⁺ revertants were obtained in BIR mutagenesis experiments. After phenotypic examination, cultures were grown from mutants for chromosome analysis by PFGE using 1% low-melting agarose at 6 V cm⁻¹ for 48 h. DNA bands corresponding to the donor and repaired recipient chromosome III were excised, equilibrated in β -agarase buffer (NEB), melted at 65 °C, and subjected to β -agarase treatment for 1 h at 40 °C. The obtained DNA was PCR amplified using *LYS2*-specific DNA primers, followed by sequencing analysis.

3.2.7 Analysis of mutation spectra of *ura3-29* Ura⁺ reversions

To determine the spectrum of Ura⁺ in individual experiments, a portion of the *URA3* gene from independent Ura⁺ was PCR-amplified using *URA3*-specific primers: 5'-GTGTGCTTCATTGGATGTTCGTAC-3' and 5'-AAAAGGCCTCTAGGTTCCCTTTGTT-3' followed by sequencing analysis using 5'-CTGGAGTTAGTTGAAGCATTAGG-3' as a primer.

For experimental strains undergoing BIR repair, 7 h Ura⁺ BIR events (confirmed as Ade⁺Leu⁻ on selective media) were sequenced. Because these cells underwent ≤ 1 division between the 0 h and 7 h time points and the Ura⁺ frequency at 7 h significantly exceeded that at 0 h, all Ura⁺ events resulting from DSB repair were considered independent.

3.3 References

Bosco, G., and Haber, J.E. (1998). Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* 150, 1037-1047.

- Chung, W.H., Zhu, Z., Papusha, A., Malkova, A., and Ira, G. (2010). Defective resection at DNA double-strand breaks leads to de novo telomere formation and enhances gene targeting. *PLoS genetics* 6, e1000948.
- Conti, C., Caburet, S., Schurra, C., and Bensimon, A. (2001). Molecular combing. *Current protocols in cytometry / editorial board, J. Paul Robinson, managing editor ... [et al.] Chapter 8, Unit 8 10.*
- Davis, A.P., and Symington, L.S. (2004). RAD51-dependent break-induced replication in yeast. *Molecular and cellular biology* 24, 2344-2351.
- Deem, A., Barker, K., Vanhulle, K., Downing, B., Vayl, A., and Malkova, A. (2008). Defective break-induced replication leads to half-crossovers in *Saccharomyces cerevisiae*. *Genetics* 179, 1845-1860.
- Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS biology* 9, e1000594.
- Fangman, W.L., and Brewer, B.J. (1991). Activation of replication origins within yeast chromosomes. *Annual review of cell biology* 7, 375-402.
- Ferguson, D.O., and Holloman, W.K. (1996). Recombinational repair of gaps in DNA is asymmetric in *Ustilago maydis* and can be explained by a migrating D-loop model. *Proceedings of the National Academy of Sciences of the United States of America* 93, 5419-5424.
- Formosa, T., and Alberts, B.M. (1986). DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* 47, 793-806.
- Friedman, K.L., and Brewer, B.J. (1995). Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. *Methods in enzymology* 262, 613-627.
- Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science* 319, 1352-1355.
- Hashimoto, Y., and Costanzo, V. (2011). Studying DNA replication fork stability in *Xenopus* egg extract. *Methods Mol Biol* 745, 437-445.
- Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS genetics* 5, e1000327.

- Kreuzer, K.N., Saunders, M., Weislo, L.J., and Kreuzer, H.W. (1995). Recombination-dependent DNA replication stimulated by double-strand breaks in bacteriophage T4. *Journal of bacteriology* *177*, 6844-6853.
- Kuzminov, A. (1995). Collapse and repair of replication forks in *Escherichia coli*. *Molecular microbiology* *16*, 373-384.
- Le, S., Moore, J.K., Haber, J.E., and Greider, C.W. (1999). RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics* *152*, 143-152.
- Llorente, B., Smith, C.E., and Symington, L.S. (2008). Break-induced replication: what is it and what is it for? *Cell Cycle* *7*, 859-864.
- Lydeard, J.R., Jain, S., Yamaguchi, M., and Haber, J.E. (2007). Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* *448*, 820-823.
- Lydeard, J.R., Lipkin-Moore, Z., Sheu, Y.J., Stillman, B., Burgers, P.M., and Haber, J.E. (2010). Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes & development* *24*, 1133-1144.
- Malkova, A., and Haber, J.E. (2012). Mutations arising during repair of chromosome breaks. *Annual review of genetics* *46*, 455-473.
- Malkova, A., and Ira, G. (2013). Break-induced replication: functions and molecular mechanism. *Current opinion in genetics & development* *23*, 271-279.
- Malkova, A., Naylor, M.L., Yamaguchi, M., Ira, G., and Haber, J.E. (2005). RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Molecular and cellular biology* *25*, 933-944.
- Oh, S.D., Jessop, L., Lao, J.P., Allers, T., Lichten, M., and Hunter, N. (2009). Stabilization and electrophoretic analysis of meiotic recombination intermediates in *Saccharomyces cerevisiae*. *Methods Mol Biol* *557*, 209-234.
- Payen, C., Koszul, R., Dujon, B., and Fischer, G. (2008). Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS genetics* *4*, e1000175.
- Pohjoismaki, J.L., and Goffart, S. (2011). Of circles, forks and humanity: Topological organisation and replication of mammalian mitochondrial DNA. *BioEssays : news and reviews in molecular, cellular and developmental biology* *33*, 290-299.

- Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., *et al.* (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular cell* 46, 424-435.
- Shcherbakova, P.V., and Pavlov, Y.I. (1996). 3'→5' exonucleases of DNA polymerases epsilon and delta correct base analog induced DNA replication errors on opposite DNA strands in *Saccharomyces cerevisiae*. *Genetics* 142, 717-726.
- Smith, C.E., Lam, A.F., and Symington, L.S. (2009). Aberrant double-strand break repair resulting in half crossovers in mutants defective for Rad51 or the DNA polymerase delta complex. *Molecular and cellular biology* 29, 1432-1441.
- Smith, C.E., Llorente, B., and Symington, L.S. (2007). Template switching during break-induced replication. *Nature* 447, 102-105.
- VanHulle, K., Lemoine, F.J., Narayanan, V., Downing, B., Hull, K., McCullough, C., Bellinger, M., Lobachev, K., Petes, T.D., and Malkova, A. (2007). Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. *Molecular and cellular biology* 27, 2601-2614.
- Viggiani, C.J., and Aparicio, O.M. (2006). New vectors for simplified construction of BrdU-Incorporating strains of *Saccharomyces cerevisiae*. *Yeast* 23, 1045-1051.
- Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 10, 1793-1808.
- Yang, Y., Gordenin, D.A., and Resnick, M.A. (2010). A single-strand specific lesion drives MMS-induced hyper-mutability at a double-strand break in yeast. *DNA repair* 9, 914-921.
- Yang, Y., Sterling, J., Storici, F., Resnick, M.A., and Gordenin, D.A. (2008). Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLoS genetics* 4, e1000264.

Disclaimer

This chapter was published under the same name in the journal *Nature*

Saini N., Ramakrishnan S., Ayyar S., Zhang Y., Elango R., Deem A., Haber J.E.,

Lobachev K.S., Malkova A. (2013) “Migrating bubble during break-induced replication drives conservative DNA synthesis.” *Nature*, doi: 10.1038/nature12584

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusions

Increased mutation rates are a hallmark of cancers (Hanahan and Weinberg, 2011). On the other hand, mutagenesis also provides the genome with the plasticity that enables an accumulation of polymorphisms and subsequently promotes evolution as described in (Pollard *et al.*, 2006; Stamatoyannopoulos *et al.*, 2009). The classical view prevailing is that the majority of mutations arise due to replicative errors (Fox *et al.*, 2013). However, the low frequencies of polymerase inaccuracies and the efficacy of DNA repair pathways discredit this assumption. Several studies using site-specific endonucleases have revealed that DSBs are potent inducers of mutations in the regions flanking the break site (Deem *et al.*, 2011; Hicks *et al.*, 2010; Rattray *et al.*, 2002; Strathern *et al.*, 1995; Yang *et al.*, 2008). Localized hypermutability of single-stranded regions created during DSB repair, the utilization of error-prone polymerases during synthesis and the inability of the mismatch-repair pathway to correct repair-associated mutations have been proposed as the mechanisms leading to increased mutagenesis (Deem *et al.*, 2011; Hicks *et al.*, 2010; Yang *et al.*, 2008).

Secondary structure-forming DNA repeats have been shown to be the driving force behind gross-chromosomal rearrangements due to their inherent fragility in both humans and model-organisms. Previously we have shown that long inverted repeats that form hairpin and cruciform structures; and triplex forming GAA/TTC repeats induce the formation of a double strand break which triggers genome instability in yeast (Kim *et al.*,

2008; Lobachev *et al.*, 2002; Narayanan *et al.*, 2006). This study demonstrates that in addition to GCRs, fragile DNA motifs are an endogenous source of DSB-associated mutagenesis in the genome. Augmented mutation rates were found to extend up to 8kb from the repeats and required the activity of the error-prone translesion polymerase Pol ζ . Inverted repeat-induced mutagenesis was also dependent on the Sae2 protein which along with the MRX complex was shown to open the hairpin-capped breaks and initiate resection (Lobachev *et al.*, 2002). These observations indicate that breakage and resection at fragile DNA motifs leads to extensive single-strandedness in the flanking DNA regions. The bypass of the damage accumulated in the single-stranded DNA during repair synthesis might necessitate a switch from the error-free polymerases to the translesion polymerase Pol ζ . Pol ζ has also been shown to be recruited to sites of DSBs by the Mec1 kinase and might be capable of replacing the replicative polymerases during repair synthesis even in the absence of a damaged template (Hirano and Sugimoto, 2006). These data bring forth secondary structure-forming DNA motifs as a dual threat to the genome capable of inducing both chromosomal rearrangements and mutagenesis at distant loci from the repeats.

Breakage at both inverted repeats and long GAA/TTC repeats has been shown to result in deletions and non-reciprocal translocations. The primary repair mechanism implicated in generating these gross-chromosomal rearrangements is break induced replication (Kim *et al.*, 2008; Narayanan *et al.*, 2006). BIR is a mode of DSB repair that is employed to repair DSBs with only one ended homology such as those generated upon the collapse of replication forks and at eroded telomeres (Llorente *et al.*, 2008). BIR

proceeds via invasion of one of the DSB ends into a homologous sequence followed by synthesis of the chromosome arm culminating in a large duplication with a concomitant deletion (Davis and Symington, 2004; Malkova *et al.*, 2005). Due to the requirement of large scale DNA synthesis and the utilization of almost all DNA replicative proteins, BIR has been considered analogous to the S-phase replication fork (Malkova *et al.*, 2005; Wang *et al.*, 2004). However, the *bona fide* semiconservative mode of synthesis fails to explain the instabilities inherent to BIR including 1000X higher mutation rates and frequent template switches (Deem *et al.*, 2011; Smith *et al.*, 2007; Stafa *et al.*, 2014). This work establishes that the innate instability of BIR is explained by its atypical mode of synthesis. Unlike the Y-shaped canonical replication intermediate, BIR was found to proceed via an unusual bubble-like replication fork driven by the Pif1 helicase, leading to conservative mode of inheritance for the newly synthesized strands. The migrating bubble creates large regions of single-strandedness in its wake suggesting that during BIR the leading and lagging strand synthesis is not synchronized. This work enabled us to demonstrate that the structure of the migrating bubble itself during BIR makes it prone to instabilities including high levels of mutagenesis. The presence of single-stranded DNA might provide a substrate to accumulate damage and enhance mutations in the recipient molecule. Taking into account the conservative nature of inheritance of the newly synthesized DNA and that the migrating bubble might not allow a stable heteroduplex to be formed upon mis-incorporation of nucleotides during synthesis, it is also unlikely that the mismatch repair machinery would be capable of correcting the errors, leading to increased mutagenesis. Additionally, the branched bubble-like structure might be a substrate for various structure-specific nucleases in the cell explaining the high

frequencies of half-crossovers generated during BIR. Finally, the unstable nature of the migrating bubble itself might be causative of the increased incidences of template switches seen in the path of BIR progression. These observations might also explain the mutagenic nature of long range synthesis associated with other DSB repair pathways like SDSA. Furthermore, this study might provide insights into the mechanisms involved in alternative telomere lengthening and mitochondrial DNA synthesis which proceed via a bubble-like replication intermediate (Lydeard *et al.*, 2007; Pohjoismaki and Goffart, 2011).

The conclusions from this study have been summarized below.

1. DSB formation at secondary-structure forming repeats is increased in replication-deficient strains.
2. Increase in fragility is followed with mutagenesis up to 8kb from the break and is diminishes at 30kb.
3. Augmented mutagenesis is dependent on the presence of the repeats and the translesion polymerase Pol ζ .
4. In strains with inverted repeats, mutagenesis depends on the Sae2 protein, indicating a requirement for DSB formation and initiation of resection.
5. Mutagenic DSB repair reconstitutes the fragile motif making secondary structure forming repeats are long term source of mutations.
6. DNA synthesis during break induced replication proceeds via a migrating bubble driven by the Pif1 helicase.

7. Asynchronous synthesis of the leading and lagging strands during BIR yields large stretches of single stranded DNA.

8. The newly synthesized strands in BIR are conservatively inherited.

4.2 Future directions

This study has provided an insight into the mechanisms leading to DSB repair-associated mutagenesis. Below are presented research directions that might enable us to get an in-depth understanding of the protein players in this phenomenon.

4.2.1 Do other translesion polymerases function during DSB repair-associated mutagenesis?

Despite the correlation of mutagenesis induced by fragile DNA motifs with the presence of the translesion polymerase Pol ζ , the strains with GAA/TTC repeats only depict a partial reduction in mutations upon deletion of *REV3*. It is therefore of interest to determine if other known translesion polymerases encoded by *RAD30* or *REVI* might be responsible for the remaining mutagenesis. In addition, increased mutation rates during gene conversion assays demonstrate a strong dependence of the mutations on the erroneous activities of the replicative polymerases. Hence, testing proofreading deficient polymerase δ and ϵ mutants in inverted repeat-mediated mutagenesis might elucidate the roles that these proteins play during DSB repair.

4.2.2 Determining the proteins responsible for BIR progression.

Previous studies using ChIP analysis at the site of BIR initiation in the genome demonstrated that almost all replication proteins including Mcm2-7, Cdc45, GINS, Cdc7, Sld3, Dpb11 are associated with these sites. However, this study clearly demonstrates that the main helicase required for robust BIR progression is Pif1 indicating that BIR initiation and progression likely operate through slightly different mechanisms. Using regulatable alleles of the various replication proteins it is possible to deplete them after the induction of break in the BIR assay described above. Further analysis using 2D gel electrophoresis and Southern hybridization using probes specific for either the point of invasion or downstream in the path of BIR would enable us to determine the involvement of these proteins in initiation or progression of BIR, respectively.

4.2.3 Can BIR occur in the S-phase?

In this study, BIR intermediates were analyzed in cells arrested in the G2-phase of the cell cycle in order to prevent interference of the output by DNA replication. However, it is intriguing to understand if BIR can function to repair one-ended breaks formed during DNA replication in S-phase. It is possible to convert single-stranded breaks generated by site-specific nickases into DSBs upon replication which can be repaired via BIR over the sister chromatid. This system can be utilized to determine the mutagenic potential of single-stranded breaks in the genome and to shed light upon the protein players that function in the repair of replication-associated DSBs. Furthermore, using 2D gels it is possible to determine the structural intermediates of the repair pathway.

These studies using yeast as a model organism provide a base that can be extended to the understanding of the pathways leading to hypermutagenesis in cancer cells and would potentially contribute to strategizing drug design and therapy.

4.3 References

- Davis, A.P., and Symington, L.S. (2004). RAD51-dependent break-induced replication in yeast. *Molecular and cellular biology* 24, 2344-2351.
- Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS biology* 9, e1000594.
- Fox, E.J., Prindle, M.J., and Loeb, L.A. (2013). Do mutator mutations fuel tumorigenesis? *Cancer metastasis reviews* 32, 353-361.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646-674.
- Hicks, W.M., Kim, M., and Haber, J.E. (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329, 82-85.
- Hirano, Y., and Sugimoto, K. (2006). ATR homolog Mec1 controls association of DNA polymerase zeta-Rev1 complex with regions near a double-strand break. *Current biology : CB* 16, 586-590.
- Kim, H.M., Narayanan, V., Mieczkowski, P.A., Petes, T.D., Krasilnikova, M.M., Mirkin, S.M., and Lobachev, K.S. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *The EMBO journal* 27, 2896-2906.
- Llorente, B., Smith, C.E., and Symington, L.S. (2008). Break-induced replication: what is it and what is it for? *Cell Cycle* 7, 859-864.
- Lobachev, K.S., Gordenin, D.A., and Resnick, M.A. (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* 108, 183-193.

- Lydeard, J.R., Jain, S., Yamaguchi, M., and Haber, J.E. (2007). Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448, 820-823.
- Malkova, A., Naylor, M.L., Yamaguchi, M., Ira, G., and Haber, J.E. (2005). RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Molecular and cellular biology* 25, 933-944.
- Narayanan, V., Mieczkowski, P.A., Kim, H.M., Petes, T.D., and Lobachev, K.S. (2006). The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* 125, 1283-1296.
- Pohjoismaki, J.L., and Goffart, S. (2011). Of circles, forks and humanity: Topological organisation and replication of mammalian mitochondrial DNA. *BioEssays : news and reviews in molecular, cellular and developmental biology* 33, 290-299.
- Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., *et al.* (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS genetics* 2, e168.
- Rattray, A.J., Shafer, B.K., McGill, C.B., and Strathern, J.N. (2002). The roles of REV3 and RAD57 in double-strand-break-repair-induced mutagenesis of *Saccharomyces cerevisiae*. *Genetics* 162, 1063-1077.
- Smith, C.E., Llorente, B., and Symington, L.S. (2007). Template switching during break-induced replication. *Nature* 447, 102-105.
- Stafa, A., Donnianni, R.A., Timashev, L.A., Lam, A.F., and Symington, L.S. (2014). Template Switching During Break-Induced Replication Is Promoted by the Mph1 Helicase in *Saccharomyces cerevisiae*. *Genetics*.
- Stamatoyannopoulos, J.A., Adzhubei, I., Thurman, R.E., Kryukov, G.V., Mirkin, S.M., and Sunyaev, S.R. (2009). Human mutation rate associated with DNA replication timing. *Nature genetics* 41, 393-395.
- Strathern, J.N., Shafer, B.K., and McGill, C.B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics* 140, 965-972.
- Wang, X., Ira, G., Tercero, J.A., Holmes, A.M., Diffley, J.F., and Haber, J.E. (2004). Role of DNA replication proteins in double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 24, 6891-6899.
- Yang, Y., Sterling, J., Storici, F., Resnick, M.A., and Gordenin, D.A. (2008). Hypermutability of damaged single-strand DNA formed at double-strand breaks

and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLoS genetics* 4, e1000264.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A.1. Sequences of mutations analyzed in *CAN1* in wild-type strain containing inverted repeats

Isolate	Coordinate in <i>CAN1</i> (coding strand) ^a	Wild-type base	Mutant base	Insertion/deletion (\pm #bases)	Wild-type sequence context	Type of mutation ^b
1	571	g	a		AGTTGGCCAAgT CATTCAATT	sub
2	1623	g	a		TAGCTGTTTGgA TCTTATTTC	sub
3	973	g	t		CAAAAAAGTTgT TTCCGTAT	sub
4	789	g	a		TCCGTTATTGgA GAAACCCAG	sub
5	670	g	c		CAAATATTACgG TGAATTCTGA	sub
6	807	g	a		CAGGTGCCTGgG GTCCAGGTA	sub
7	470	c	a		ACAGTTTTCTcA CAAAGATTC	sub
8	449	c	a		ATCCCTGTTAcA TCCTCTTTC	sub
9	804	c	-	-1	ACCCAGGTGCcT GGGGTCCAG	indel
10	1314	c	a		CTTTGGCTTAcA TGGAGACAT	sub
11	895	g	c		CCTTCACATTTC AAgGTAAGTAA	sub
12	910	g	a		TGAACTAGTTgG TATCACTGC	sub
13	580	-	t	+1	AGTCATTCAA- TTTTGGACGT	indel
14	887	c	-	-1	GCTGCCTTCAcA TTTCAAGGT	indel
15	937	c	a		AGCTGCAAACcC AGAAAATCC	sub
16	1214	c	g		TTTGGTCTATcA AAGAACAAG	sub

17	1178	a	t		GCAAATTCAAaT ATTACGTT	sub
18	284	t	g		CATATTGGTAtG ATTGCCCTT	sub
19	1253	c	t		TCAAGGACCAcC AAAGGTGGT	sub
20	356	c	g		AACGCCGGCCc AGTGGGCGCT	sub
21	290	c	a		GGTATGATTGcC CTTGGTGGT	sub
22	541	t	g		GGCAATCACTtT TGCCCTGGA	sub
23	663	-	a	+1	CCCTGTCAAA- TATTACGGTG	indel
24	1392	g	a		TTTTTGCATGgT TATTTATCT	sub
25	895	g	a		CACATTTCAAaG TACTGAACT	sub
26	670	g	t		CAAATATTACgG TGAATTCGA	sub
27	857	g	a		TTCTTAGGTTgG GTTTCCTCT	sub

^a Coordinates of the first nucleotide in the mutated sequence are indicated based on the *CAN1* coding strand sequence.

^b sub - base substitutions, indel - insertions or deletions

Table A.2 Sequences of mutations analyzed in *CAN1* in *pol3-P664L* mutant strain carrying inverted repeats

Isolate	Coordinate in <i>CAN1</i> (coding strand) ^a	Wild-type base	Mutant base	Insertion/deletion (\pm #bases)	Wild-type sequence context	Type of mutation ^b
1	527	c	t		TATTGGTTTTcT TGGGCAATC	sub
2	530	g	a		TGGTTTTCTTgG GCAATCACT	sub
3	581	tttgg	-	-5	GTCATTCAATtttg gACGTACAAAG	indel
4	625	t	-	-1	TAGTATTTTTtG GGTAATTAT	indel
5	629	ta	--	-2	ATTTTTTGGGtaA TTATCACAA	indel
6	655	c	a		GAACTTGTTcC TGTCAAATA	sub
7	679	g	a		CGGTGAATTCgA GTTCTGGGT	sub
8	686	gGGTCGC t	aGGT CGCa		CGAGTTCTgGGT CGctTCCATCA	complex
9	770	c	-	-1	GTTACCGGCCcA GTTGGATTC	indel
10	803	cc	-a		AACCCAGGTGcc TGGGGTCCAG	complex
11	892	c	a		CTTCACATTTcA AGGTA CTGA	sub
12	920	c	a		GGTATCACTGcT GGTGAAGCT	sub
13	926	a	t		ACTGCTGGTGaA GCTGCAAAC	sub
14	937	cCc	aC-		AGCTGCAAACcC cAGAAAATCCG	complex
15	977	t	-	-1	AAAGTTGTTtC CGTATCTTA	indel
16	979	c	g		AGTTGTTTTcCg TATCTTAAC	sub
17	979	c	a		AGTTGTTTTcCg TATCTTAAC	sub
18	1001	gct	c--		TTCTACATTGgct CTCTATTATT	complex
19	1002	ctctctattattc attgg	-	-19	CTACATTGGctctc tattattcattggACTT	slippage

20	1129	at	--	-2	TTTGCCACATatC TTCAACGCT	indel
21	1214	- AAAgAA C	tAAAA AAC		GGTCTATCAAAa AACAAGTTGGC	complex
22	1267	cc	aa		AGGTGGTGTtcc ATACATTGCA	complex
23	1268	c	a		GGTGGTGTTCcA TACATTGCA	sub
24	1272	c	g		GTGTTCCATAcA TTGCAGTTT	sub
25	1324	tctactggtg tgaca	-	-16	GGAGACA <u>tctactg</u> <u>gtggtgaca</u> AAG	slippage
26	1351	tg	--	-2	AGTTTTCGAAtg GCTATTAAAT	indel
27	1386	t	-	-1	CAGGCTTTTTtG CATGGTTAT	indel
28	1429	g	c		ATTTATGCAAgC TTTGAAATA	sub
29	1537	c	a		TATCATTATTcA AGGTTTCAC	sub
30	1351	tg	--	-2	AGTTTTCGAAtg GCTATTAAAT	indel
31	892	c	a		CTTCACATTTcA AGGTAAGTGA	sub

^a Coordinates of the first nucleotide in the mutated sequence are indicated based on the *CAN1* coding strand sequence.

^b sub - base substitutions, indel - insertions or deletions, complex - complex mutations, slippage- slippage events between short direct repeats that are indicated by underlined sequences.

Table A.3 Sequences of mutations analyzed in *CAN1* in *pol3-P664L Δrev3* mutant strain carrying inverted repeats

Isolate	Coordinate in <i>CAN1</i> (coding strand) ^a	Wild-type base	Mutant base	Insertion/deletion (±#bases)	Wild-type sequence context	Type of mutation ^b
1	276	-	tattggta tgattgcc cttggtg gtac	+27	ACATATTGGT <u>ATGATTGCC</u> TTGGTGGTAC <u>TATTGGTACA</u>	slippage
2	284	tgattgcct tggtgtac tattggta	-	-27	<u>CATATTGGTA</u> tgattgccttggtgt actattggtaCA	slippage
3	754	gctgggggt	-	-8	GGTTTGTGGT gctgggggtTACCG GCCCA	slippage
4	274	c	g		TAAGCAAAG AcATATTGGT AT	sub
5	1069	gtttctac	-	-8	TACTTCCTAC gtttctacTTCTCC CTTT	slippage
6	299	g	a		GCCCTTGGTG gTACTATTGG T	sub
7	530	g	a		TGGTTTTCTTg GGCAATCACT	sub
8	1002	ctctctattat tcattgg	-	-19	TCTAC <u>ATTGG</u> ctctctattattcattgg ACTTT	slippage
9	1195	cgtattttatt tggtctatca aagaacaa gttggtcc	-	-38	GTTCCcgtattttat ttggtctatcaaagaac aagttggtccTAA AT	slippage
10	1324	tctactggtg gtgaca	-	-16	CATGGAGAC <u>Atctactggtggtgac</u> aAAGTTTTTCG	slippage
11	836	a	-	-1	AAGGATAAA AaCGAAGGGA GG	indel
12	724	ctaatatact gttt	-	-14	TATCGGGTTT ctaatatactgtttTT GTATGGTT	slippage
13	687	g	a		TCGAGTTCTG gGTCGCTTCC A	sub

14	432	tacattcatc cctgt	-	-15	GTGAAATGGC <u>tacattcatc</u> cctgtT <u>ACATCCTCT</u>	slippage
15	311	c	a		ACTATTGGTA cAGGTCTTTT C	sub
16	669	c	a		TCAAATATTA cGGTGAATTC G	sub
17	284	tgattgcct tgggtgtac tattgga	-	-27	<u>CATATTGGTA</u> tgattgccttgggtgt actattggaCA	slippage
18	1324	tctactggtg gtgaca	-	-16	<u>CATGGAGAC</u> <u>Atctactggtggtgac</u> aAAGTTTTCG	slippage
19	1195	cgtattttatt tggctatca aagaacaa gttggctcc	-	-38	TGGTTCCcgtatt ttatttggctatcaaag aacaagttggctccT AAAT	slippage
20	1068	c	g		CTACTTCCTA cGTTTCTACTT	sub
21	980	g	a		GTTGTTTTCCg TATCTTAACC	sub
22	374	c	a		GCTCTTATAT cATATTTATTT	sub

^a Coordinates of the first nucleotide in the mutated sequence are indicated based on the *CAN1* coding strand sequence.

^b sub - base substitutions, indel - insertions or deletions, complex - complex mutations, slippage- slippage events between short direct repeats that are indicated by underlined sequences.

Table A.4 Sequences of mutations analyzed in *CAN1* in *pol3-P664L* mutant strain carrying no inverted repeats

Isolate	Coordinate in <i>CAN1</i> (coding strand) ^a	Wild-type base	Mutant base	Insertion/deletion (\pm #bases)	Wild-type sequence context	Type of mutation ^b
1	1195	cgtattttatt ggtctatcaa agaacaagtt ggctcc	-	-38	GGT <u>TCC</u> cgtattt tatttggtctatcaaag aacaagttggtccT AA	slippage
2	502	g	-	-1	ATTTGGTGC GgCCAATGG TTA	indel
3	447	t	c		TCATCCCTGT tACATCCTCT T	sub
4	1002	ctctctattatt cattgg	-	-18	TCTACATTG Gctctctattattcatt ggACTTTTA	slippage
5	804	ctggggtc	-tgggggtt		ACCCAGGTG CctggggtcCAG GTATAA	complex
6	1536	tc	-t		TTATCATTAT tcAAGGTTTC AC	complex
7	658	g	c		CTTGTTCCCT gTCAAATATT A	sub
8	395	t	a		ATGGGTTCT TtGGCATATT CT	sub
9	1054	c	t		TAAACTAAC AcAATCTACT TC	sub
10	979	c	-	-1	AGTTGTTTTTC cGTATCTTAA C	indel
11	1099	g	t		TATTGCTATT gAGAACTCT GG	sub
12	804	c	-	-1	ACCCAGGTG CcTGGGGTCC AG	indel
13	434	c	a		GAAATGGCT AcATTCATCC CT	sub

14	1403	c	g		TTATTTATCT cAATCTCGCA C	sub
15	1625	tcttattcaat gcatattcag atgcagattt atttga	-	-39	CTGTTTGGAt cttatttcaatgcatatt cagatgcagatttatt ggaAGAT	slippage
16	922	g	a		TATCACTGC TgGTGAAGC TGC	sub
18	1039	gaccctaaa ctaacacaat	-	-19	ATACAATgac cctaaactaacacaa tCTAC	slippage
19	687	g	a		TCGAGTTCT GgGTCGCTTC CA	sub
20	1473	a	-	-1	TACCATTTA AaGCTAAATT AA	indel
21	974	tt	ga		AAAAAAGTT GttTTCCGTAT CT	complex
22	673	g	t		ATATTACGG TgAATTCGA GTT	sub
23	1195	c	t		CGTTGGTTC CcGTATTTTA TT	sub
24	668	a	g		GTCAAATAT TaCGGTGAAT TC	sub
25	352	g	c		GACCAACGC CgGCCCAGT GGG	sub
26	1195	cgtattttatt ggtctatcaa agaacaagtt ggctcc	-	-38	GGTTCcgtattt tatttggctatcaaag aacaagttggctccT AAAT	slippage
27	284	tgattgccctt ggtgtacta ttgta	-	-27	CATATTGGT Atgattgcccttgg gtactattggtaCA GG	slippage
28	284	tgattgccctt ggtgtacta ttgta	-	-27	CATATTGGT Atgattgcccttgg gtactattggtaCA GG	slippage

29	238	c	t		AGGAGAAGT AcAGAACGC TGA	sub
----	-----	---	---	--	-------------------------------	-----

^a Coordinates of the first nucleotide in the mutated sequence are indicated based on the *CAN1* coding strand sequence.

^b sub - base substitutions, indel - insertions or deletions, complex - complex mutations, slippage- slippage events between short direct repeats that are indicated by underlined sequences.

Table A.5 Sequences of mutations analyzed in *CAN1* in wild-type strain carrying *IS50*-perfect palindrome

Isolate	Coordinate in <i>CAN1</i> (coding strand) ^a	Wild-type base	Mutant base	Insertion/deletion (\pm #bases)	Wild-type sequence context	Type of mutation ^b
1	901	g	c		TCAAGGTAC TgAACTAGTT GG	sub
2	446	cTg	-Tt	-1	ACATTCATC CctgTTACATC CTC	complex
3	1195	cG-	aGt	1	CGTTGGTTC CcgTATTTTA TTT	complex
4	1394	t	a		CATGGAGAC AtCTACTGGT GG	sub
5	1754	t	g		TGGGACAAA TtTTGGAATG TT	sub
6	1796	aTCACA TTt	cTCAC ATTc		ATTACCTTTG aTCACATTtC CACGCCATT	complex
7	969	a	g		CCATCAAAA AaGTTGTTTT CC	sub
8	1065	-	c	1	AATCTACTT C- CTACGTTTCT	indel
9	1163	c	a		ACCATTATTT cTGCCGCAA AT	sub
10	522	g	a		ACATGTATT GgTTTTCTTG GG	sub
11	1123	-	c	1	AAAGGTTTT G- CCACATATC T	indel
12	244	g	t		AGTACAGAA CgCTGAAGT GAA	sub
13	591	c	a		TTTGGACGT AcAAAGTTC CAC	sub

14	890	-	t		GCCTTCACA T- TTCAAGGTA C	indel
15	317	c	-		TGGTACAGG TcTTTTTCATT GG	indel
16	351	g	a		GACCAACGC CgGCCCAGT GGG	sub
17	464	t	g		TCTTTCACA GtTTTCTCAC AA	sub
18	112	c	a		AGACGTGGG TcAATACCAT TG	sub
19	146	t	a		AGTAAAGAA TtGTATCCAT TG	sub
20	896	g	a		ACATTTCAA GgTACTGAA CTA	sub
21	48	c	g		AGCATATGT AcAATGAGC CGG	sub
22	1314	c	g		CTTTGGCTTA cATGGAGAC AT	sub
23	275	a	g		AAGCAAAGA CaTATTGGTA TG	sub
24	1314	c	g		CTTTGGCTTA cATGGAGAC AT	sub
25	1600	a	g		CTATATCTCT aTTTTCTGT T	sub

^a Coordinates of the first nucleotide in the mutated sequence are indicated based on the *CAN1* coding strand sequence.

^b sub - base substitutions, indel - insertions or deletions, complex - complex mutations.

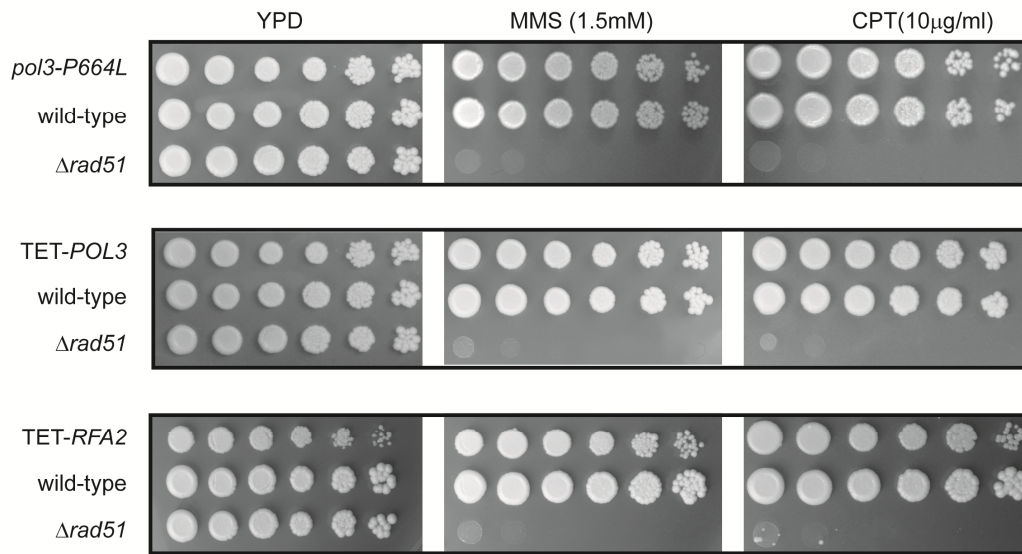


Figure A.1 Sensitivities of the replication-deficient strains to DNA-damaging agents. Four-fold serial dilutions of wild-type strains and the mutant alleles were plated on YPD and YPD containing 1.5mM MMS and 10µg/ml camptothecin (CPT). *Δrad51* was used as a control since it exhibits extreme sensitivity to the drugs used. The middle and bottom panels depict strains grown in the presence of 2µg/ml and 0.1µg/ml doxycycline, respectively.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Table B.1 The rate of spontaneous and DSB-associated Ura⁺ mutations

Orientation of <i>ura3-29</i>	Rate of Ura ⁺ (x10 ⁹) *											
	Before galactose(0h)			After galactose (frequency (7h-0h))				Fold-above no damage [‡] (P-value)	BIR efficiency (%) [§]			
	HO site	Median	CI [†]	No MMS		1.5mM MMS			No MMS		1.5mM MMS	
				Median	CI [†]	Median	CI [†]		Ade ⁺	Ade ⁺ Ura ⁺	Ade ⁺	Ade ⁺ Ura ⁺
Ori1	DSB	28	10-47 [7]	3,646	2,305- 4,159	9,415	4,911- 11,061	2.6 (0.0006)	80± 8	94± 8	84± 10	98± 0.4
Ori2	DSB	6	5-17 [7]	1,903	979- 2,941	41,835	34,830- 79,488	22 (0.0006)	74± 9	94± 2	76± 14	92 ±2
Ori1	No	8	5-10 [17]	0	0-11	198	49-358	24.8	N/A	N/A	N/A	N/A
Ori2	No	12	5-16 [13]	0	0-7	157	101- 245	13.1	N/A	N/A	N/A	N/A

* Rates calculated at 0 h based on 0 h frequencies using the Drake equation (see Methods for details). At 7 h, rates were calculated as (7 h frequency – 0 h frequency); differences <0 are reported as '0'.

[†] For strains with ≥6 experiments, the 95% CI of the median is given.

[‡] Statistically significant elevation of 7 h mutation rate in strains in the presence of MMS over 7 h mutation rate in the absence of MMS.

[§] Percent of BIR (average ± s.d.) calculated based on 3-6 experiments among DSB repair outcomes collected at 7 h on either adenine dropout media (Ade⁺) or on adenine/uracil dropout media (Ade⁺Ura⁺).

Table B.2 The rate of DSB-associated Lys⁺ mutations (top), and the rate of spontaneous Lys⁺ mutations (bottom)

Rate of Lys ⁺ (x10 ⁹) *										
Position	Construct	HO site	Genotype	Before galactose (0h)		After galactose (frequency (7h-0h))		Fold-below WT [‡] (P-value)	BIR efficiency (%) [§]	
				Median	CI [†]	Median	CI [†]		Ade ⁺	Ade ⁺ Lys ⁺
16 kb	A ₄	DSB	wt	40	12.7-64.3	2,690	1,703-4,361	N/A	77±12	99.7±0.5
16 kb	A ₄	DSB	<i>Δpif1</i>	6	40-10.4	134.7	104-1,580	20 (0.0001)	73±11	00±3
36 kb	A ₄	DSB	wt	5.3	2.7-15.2	1,248	860-1,552	N/A	80±1	99±1
36 kb	A ₄	DSB	<i>Δpif1</i>	1	0.5-12	1.4	0-4.7	892 (0.0003)	91±4	100 [#]

Rate of Lys ⁺ (x10 ⁹) *					
Position	Construct	HO site	Genotype	Median	CI [†]
16 kb	A ₄	No	wt	7.7	3.3-34
16 kb	A ₄	No	<i>Δpif1</i>	5.3	3.3-7.5
36 kb	A ₄	No	wt	1.1	0.7-5.4
36 kb	A ₄	No	<i>Δpif1</i>	0.9	0.6-3.9

* Rates calculated at 0 h based on 0 h frequencies using the Drake equation (see Methods for details). At 7 h, rates were calculated as (7 h frequency – 0 h frequency); differences <0 are reported as "0".

[†] For strains with ≥6 experiments, the 95% CI of the median is given. For the strains with <6 experiments, the median range is given.

[‡] Statistically significant decrease of median rate at 7 h in *Δpif1* compared to wild type.

[§] Percent of BIR (average ± s.d.) calculated based on 4-8 experiments among DSB repair outcomes collected at 7 h on either adenine dropout media (Ade⁺) or on adenine/lysine dropout media (Ade⁺ Lys⁺).

[#] No s.d. could be calculated because of a very low number of Lys⁺ (between 1 and 5) in each experiment.

Table B.3 Strains used in this study

Strain name	Genotype
AM1003	<i>MATa-LEU2-tel/MATΔ-inc ade1 met13 ura3 leu2/leu2-3,112 thr4 lys5 hmlΔ::ADE1/hmlΔ::ADE3 hmrΔ::HPH ade3::GAL::HO. FS2Δ::NAT/FS2</i>
AM1291	AM1003 but <i>lys2Δ thr4::lys2-Ins(A4)</i>
AM1449	AM1291 but <i>MATα-inc-LEU2-tel</i>
AM1482	AM1003 but <i>lys2 Δ sed::lys2- Ins(A4)</i>
AM1649	AM1482 but <i>MATα-inc-LEU2-tel</i>
AM2191	AM1291 but <i>pif1::KANMX</i>
AM2247	AM1291 but <i>MATα-inc-LEU2-tel</i>
AM2198	AM1482 but <i>pif1::KANMX</i>
AM2257	AM2198 but <i>MATα-inc-LEU2-tel</i>
AM1247	AM1003 but <i>lys2Δ thr4::LYS2</i>
AM2439	AM1291 but <i>snt1:(TEF1/BSD)3</i>
AM2438	AM1482 but <i>snt1:(TEF1/BSD)2</i>
AM2118	AM1247, ChrII::KANMX
AM2442	AM2118 but <i>MATα-inc-LEU2-tel</i>
AM2406	AM1003 but <i>ura3::p306-BrdU tps1::KANMX snt1:(TEF1/BSD)3</i>
AM2846	AM2406 but <i>tps1::BLEO pif1::KANMX</i>
AM2110	AM1003 but <i>lys2Δ ura3Δ sed4::lys2- Ins(A4) hmr::KANMX</i>
A2161	AM2110 but <i>thr4::ura3-29 (Ori1)</i>
AM2259	AM2161 but <i>MATα-inc-LEU2-tel</i>
AM2820	AM2110 but <i>thr4::ura3-29 (Ori2)</i>
AM2842	AM2820 but <i>MATα-inc-LEU2-tel</i>

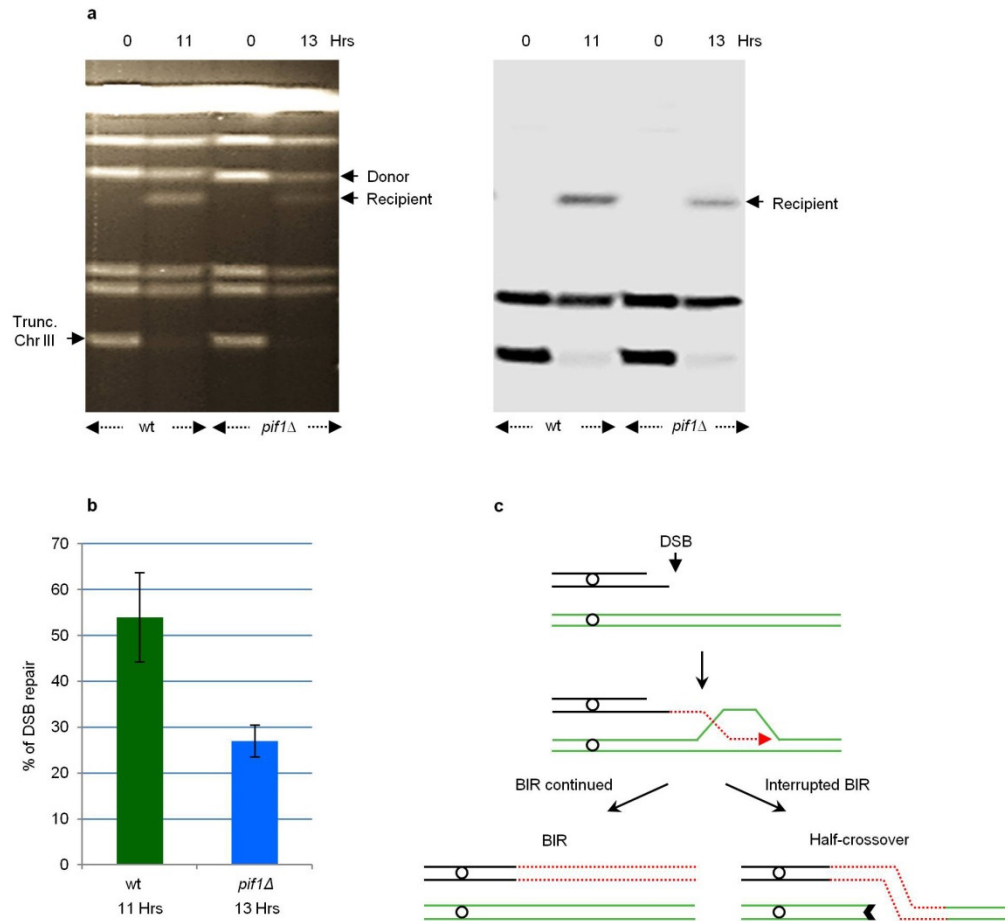


Figure B.1 BIR efficiency during molecular combing analysis of molecular intermediates of BIR. **a**, BIR efficiency was analysed by PFGE from samples used for dynamic molecular combing analysis. DNA was prepared from cells containing truncated chromosome III (Trunc. Chr III) before DSB induction and 11 h or 13 h after DSB induction from wild-type (*PIF1*) and $\Delta pif1$ cells, respectively. In $\Delta pif1$, a later time point (13 h) was analysed owing to slower kinetics of DSB repair in $\Delta pif1$ as compared to *PIF1*. Chromosomes were separated by PFGE followed by Southern hybridization with an *ADE1*-specific probe. **b**, Quantification of DSB repair efficiency (BIR, or other recombination pathways) based on the results of 3–5 individual experiments and presented as average \pm s.d. **c**, Schematic of the BIR assay. Interruption of BIR leads to the resolution of BIR intermediates resulting in half-crossover formation.

a

Relevant Genotype	Molecule analyzed	Full BIR (~100 kb BrdU)	Long BIR (>30 kb BrdU)	Short BrdU* patches (<20 kb BrdU)	No BrdU	Total
wt (<i>PIF1</i>)	Donor	4	---	14 [4] (10)	85	103
	Recipient	70	14	---	14	98
<i>pif1</i> Δ	Donor	0	0	12 [5] (7)	80	92
	Recipient	0	0	31 [23] (8)	72	103

* - BrdU patches between P2 and P3 are included. [] - overlaps with P2; () - overlaps with P3

b

Orientation of <i>ura3-29</i>	MMS ^a	BIR <i>Ura</i> ⁺ mutations		
		C→A	C→T	C→G
Ori1	No	79 (80%)	18 (18%)	2 (2%)
	Yes	31 (66%)	12 (25%)	4 (9%)
Ori2	No	13 (28%)	24* (51%)	10 (21%)
	Yes	21 (36%)	26** (45%)	11 (19%)

^a - Cells exposed to 1.5 mM MMS; * and ** - Statistically different from Ori1 (P<0.0001 and P=0.04 respectively)

FigureB.2 Analysis of molecular mechanism and mutagenesis associated with BIR.

a. The summary of molecular combing analysis presented in Figure 3 and Figure B.3 is shown. A strong bias towards BrdU tracts present only in the recipient chromosome was also observed in three additional independent experiments. **b,** Mutation spectra of BIR-induced base substitutions in *ura3-29* in the presence or absence of 1.5 mM MMS is shown.

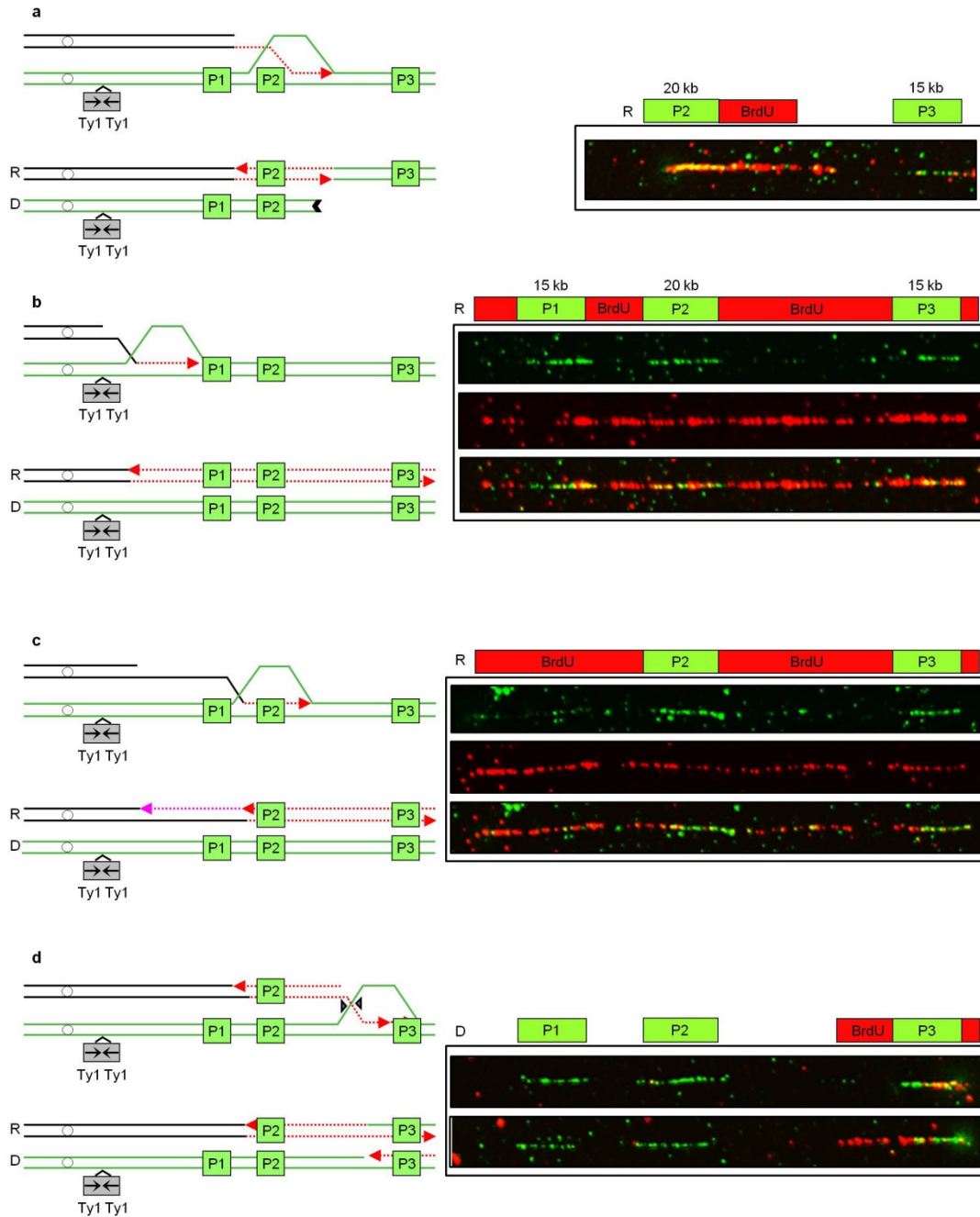


Figure B.3 Molecular outcomes of BIR. **a**, Left: interrupted BrdU tract in recipient may result from half-crossover. Right: an example of wild-type (*PIF1*) recipient with interrupted BrdU tract hybridized to P1, P2, P3 probes (green) and treated with anti-BrdU antibody (red). **b**, Left: BIR initiated by strand invasion between FS2 (inverted repeat of Ty1 located 30 kb centromere proximal to *MAT*) and P1 results in formation of recipients hybridizing to P1, P2, P3 and BrdU. Right: an example of wild-type (*PIF1*) recipient. Top: hybridization to P1, P2, P3. Middle: treatment with anti-BrdU antibody. Bottom: merge. **c**, Left: BrdU incorporation in the recipient resulting from BIR (red) and from

filling-in synthesis (pink) following extensive resection. Right: an example of wild-type (*PIF1*) recipient. Top: hybridization to P1, P2, P3. Middle: treatment with anti-BrdU antibody. Bottom: merge. **d**, Left: HJ resolution at the end of BIR progression leads to switch from conservative to semiconservative BIR resulting in a short patch of BrdU overlapping with P3 in the donor. Right: an example of BrdU incorporation in the donor from wild-type (*PIF1*) strain hybridized to P1, P2, P3 and treated with anti-BrdU antibody.

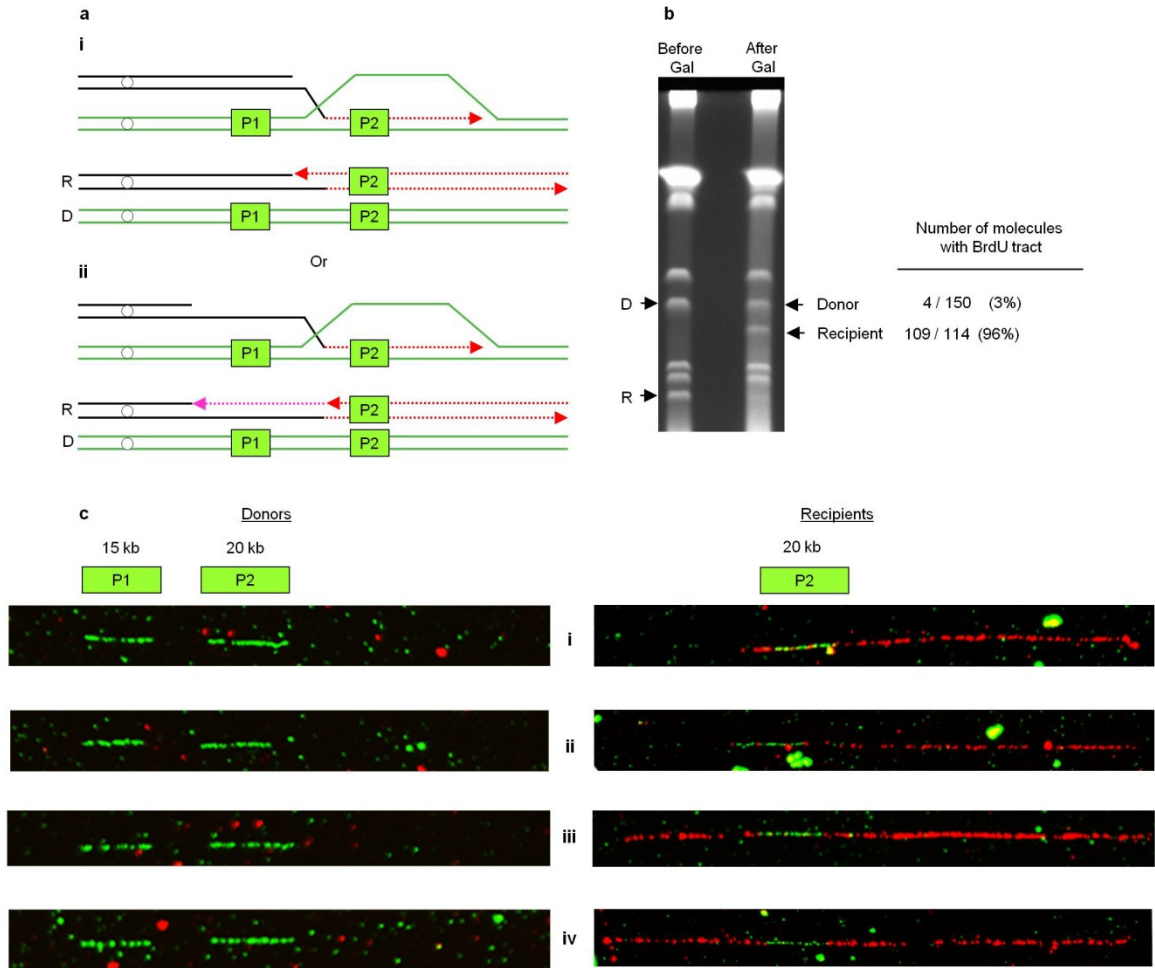


Figure B.4 Conservative DNA synthesis associated with BIR. Results from a series of 3 experiments where only P1, P2 and anti-BrdU antibody were used. **a**, BrdU incorporation in the recipient is expected from conservative BIR (i; red) and from filling-in synthesis (pink) following extensive resection (ii). **b**, **c**, Analysis of the donor (D) and repaired recipient (R) chromosomes extracted after PFGE (**b**) and hybridization with probes (green tract) and treatment with anti-BrdU antibodies (red tract) (**c**). No BrdU tracts are visible in more than 97% of donors. The repaired recipient contains long stretches of BrdU overlapping with the P2 region.

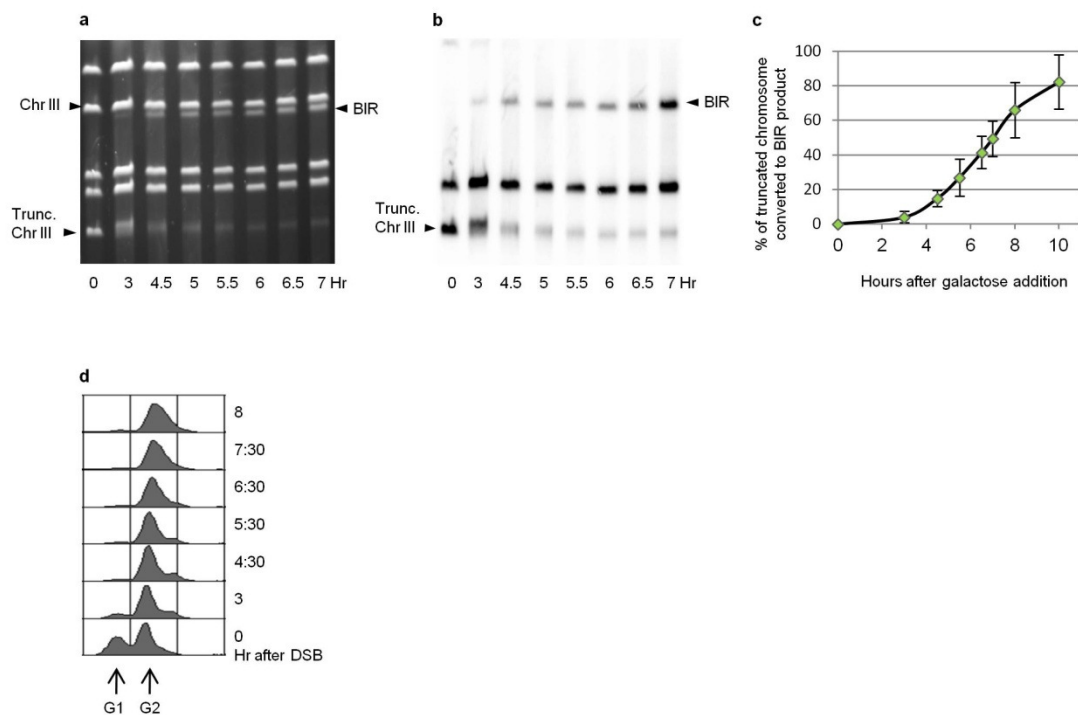


Figure B.5 BIR kinetics during 2D analysis of molecular intermediates of BIR. **a**, BIR kinetics was analysed by PFGE from samples used to determine the structure of BIR intermediates by 2D electrophoresis (Figure.3 4 c, d). DNA was prepared for PFGE at intervals after induction of DSBs at *MATa* and separated by PFGE (**a**) followed by Southern hybridization with an *ADE1*-specific probe (**b**). **c**, BIR efficiency quantified based on the results of four individual experiments. Presented as average \pm s.d. **d**, Flow cytometry of DNA analysis of cells undergoing BIR repair.

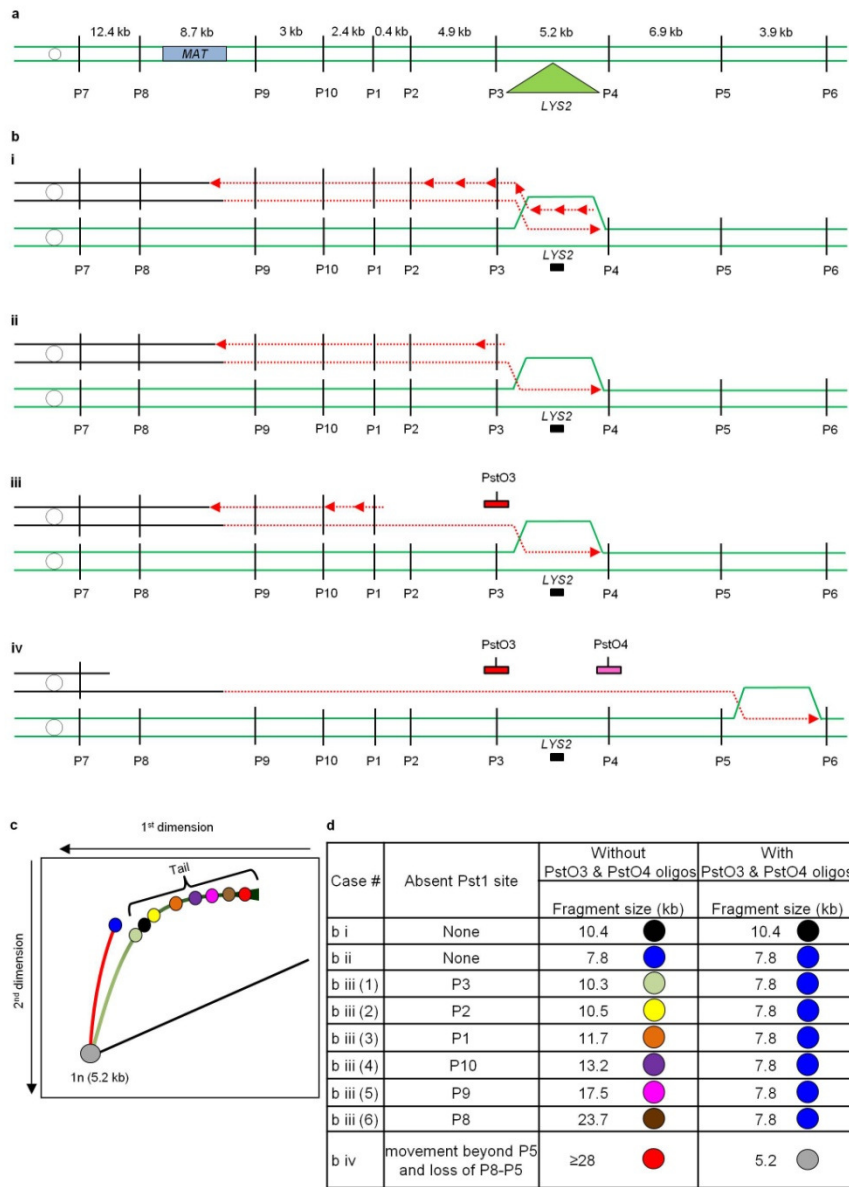


Figure B.6 The structure of molecular intermediates of BIR. **a**, The structure of the chromosome III region with *LYS2* inserted 16 kb centromere distal to *MAT α -inc*. P1, P2, P3, and so on designate positions of PstI sites flanking *LYS2*. **b**, The structure of replication bubbles migrating through *LYS2* (with black rectangle designating *LYS2*-specific probe). i, Replication bubble with synchronous leading and lagging strands (double-stranded). ii, Replication bubble with delayed initiation of the lagging strand with respect to the leading strand (partially single-stranded bubble). iii, A partially single-stranded bubble with one or several PstI sites behind the bubble inactivated due to accumulation of single-stranded DNA. Red and pink rectangles represent oligonucleotides PstO3 and PstO4, respectively. iv, A single-stranded bubble that has passed beyond the P3–P4 region. **c**, Theoretical bubble-migration curves for the intermediates shown in **b**. **d**, Calculation of parameters of the bubble-like structures for the intermediates shown in **b**.

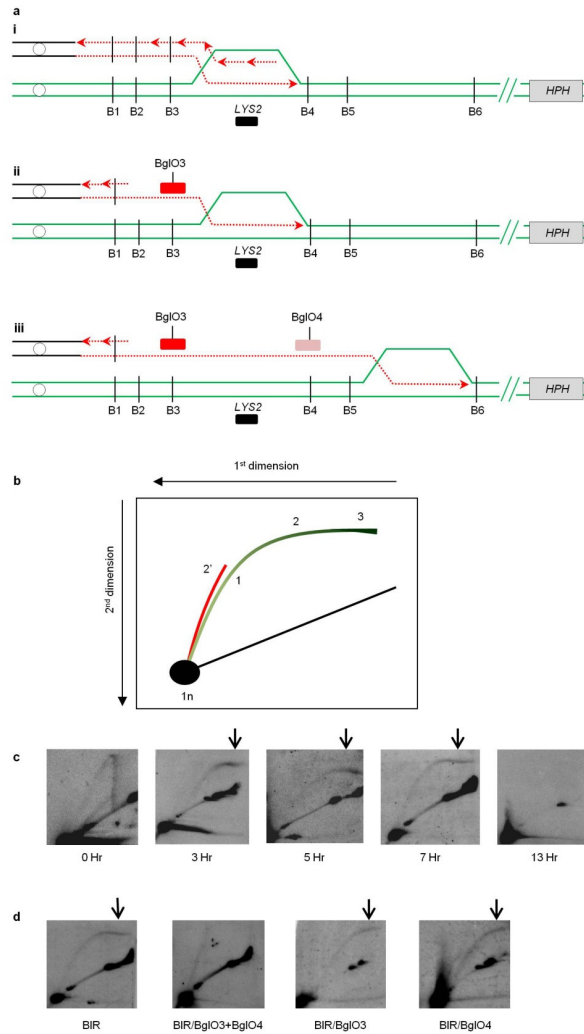


Figure B.7 Molecular intermediates of BIR. BIR intermediates were analysed by 2D gel electrophoresis of BglII-digested intact chromosomal DNA embedded in agarose plugs. **a**, D-loop migration in 2D gels (hybridized to *LYS2*, black rectangle) during coordinated (i) and uncoordinated (ii, iii) leading- and lagging-strand synthesis. **b**, Schematic of 2D gel separation of replication and BIR intermediates. Annealing to oligonucleotides (BglO3 and BglO4) restores BglII sites (B) in ssDNA (see **a**, ii) and changes migration of the intermediate as shown by 2' (red). **c**, 2D analysis of Y-arc during normal replication (0 Hr) and bubble-like structures at time points after BIR induction. Similar bubble structures were observed in nine additional independent experiments (see the legend to Figure 3.4). **d**, High-molecular-mass tails (arrows) disappear after simultaneous addition of BglO3 and BglO4 (BIR/BglO3+BglO4). The addition of each of these oligonucleotides individually (BIR/BglO3 or BIR/BglO4) failed to eliminate the tail.

PUBLICATIONS

Zhang Y., Saini N., Sheng Z. and Lobachev K.S. (2013) “Genome-wide screen reveals replication pathway for quasi-palindrome fragility dependent on homologous recombination.” PlosGenetics (accepted and in press)

Saini N., Ramakrishnan S., Ayyar S., Zhang Y., Elango R., Deem A., Haber J.E., Lobachev K.S., Malkova A. (2013) “Migrating bubble during break-induced replication drives conservative DNA synthesis.” Nature, doi: 10.1038/nature12584

Saini N., Zhang Y., Nishida Y., Sheng Z., Choudhury S., Mieczkowski P. and Lobachev K.S. (2013) “Fragile DNA motifs trigger mutagenesis at distant chromosomal loci.” PLOS Genetics, doi: 10.1371/journal.pgen.1003551

Saini N., Zhang Y., Usdin K., Lobachev K.S. (2013) “When secondary comes first – the importance of non-canonical DNA structures.” Biochimie, 95(2):117-23, PMID: 23084930

Zhang Y., Shishkin A. A., Nishida Y., Marcinkowski-Desmond D., Saini N., Volkov K., Mirkin S. and Lobachev K.S. (2012) “Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansion in dividing and non-dividing yeast cells.” Molecular Cell, 48(2):254-65, PMID: 22959270