Studies of Spontaneous Oxidative and Frameshift Mutagenesis in *Saccharomyces cerevisiae*

by

Sarah Victoria Mudrak

University Program in Genetics and Genomics
Duke University

Date:_____

Approved:

_____
Dr. Sue Jinks-Robertson, Supervisor

_____
Dr. Kenneth Kreuzer

_____
Dr. Daniel Lew

_____
Dr. John McCusker

_____
Dr. Thomas Petes

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the University Program in Genetics and Genomics
in the Graduate School of Duke University

2010

ABSTRACT

Studies of Spontaneous Oxidative and Frameshift Mutagenesis

in *Saccharomyces cerevisiae*

by

Sarah Victoria Mudrak

University Program in Genetics and Genomics
Duke University

Date:_____

Approved:

_____
Dr. Sue Jinks-Robertson, Supervisor

_____
Dr. Kenneth Kreuzer

_____
Dr. Daniel Lew

_____
Dr. John McCusker

_____
Dr. Thomas Petes

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the University Program in
Genetics and Genomics in the Graduate School
of Duke University

2010

Abstract

Preserving genome stability is critical to ensure the faithful transmission of intact genetic material through each cell division. One of the key components of this preservation is maintaining low levels of mutagenesis. Most mutations arise during replication of the genome, either as polymerase errors made when copying an undamaged DNA template or during the bypass of DNA lesions. Many different DNA repair proteins act both prior to and during replication to prevent the occurrence of these mutations. Although the mechanisms by which mutations occur and the various repair proteins that act to suppress mutagenesis are conserved throughout all species, they are best characterized in the yeast *Saccharomyces cerevisiae*. In this work, we have used this model system to study two types of spontaneous mutagenesis: oxidative mutagenesis and frameshift mutagenesis. In the first part of this work, we have examined mutagenesis that arises due to one of the most common oxidative lesions in the cell, 7,8-dihydro-8-oxoguanine or GO. When present during replication, these GO lesions generate characteristic transversion events that are accurately repaired by the mismatch repair pathway. We provide the first evidence that a second pathway involving the translesion synthesis polymerase Polη acts independently of the mismatch repair pathway to suppress GO-associated mutagenesis. We have also examined how differences in replication timing during S phase contribute to variations in the rate of these mutations across the genome. In the second part of this work, we have examined how spontaneous frameshift mutations are generated during replication. While most frameshift mutations occur in regions of repetitive DNA, we have designed a system to examine frameshifts that occur in very short repeats (< 4 nucleotides) and noniterated sequences. We have examined the

patterns of frameshifts at these sites and how the mismatch repair pathway acts to suppress these mutations. Together, the experiments presented here provide further insight into the different mechanisms that suppress and/or influence rates of oxidative mutagenesis and describe a system in which we have begun to characterize how frameshift mutations are generated at very short repeats and non-repetitive DNA.

Dedication

I would first and foremost like to thank Sue for introducing me to the world of genetics and taking a chance on a young undergraduate. Eight years later, I am still grateful every day for having such a wonderful teacher and mentor.

I would also like to thank my family for all of their support, both financially and mentally, over the years.

To my current and past labmates, especially Caroline, Nayun, Shannon, Kevin, Kat, and Stas, thank you so much for all your help, support, and laughter – you all made the work day much more enjoyable ☺.

Finally, to Ben, I will always think of grad school as not only the place where I earned my doctorate, but as the place that brought us together. Thank you, thank you, thank you, for everything.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

DNA represents the building blocks of every living organism, from the smallest bacteria to humans. Despite the vast phenotypic differences between organisms, most share the same basic DNA code and are very similar at the molecular level. All species share one common ancestor, and small changes in DNA over billions of years have yielded an enormous number of diverse species. These small changes in DNA are referred to as mutations, and collections of mutations define each species. Thus, the process of generating mutations, or mutagenesis, is necessary for evolution.

Despite the critical importance of mutagenesis to the development of new species, this process must be kept in a delicate balance. Mutagenesis is not guided; mutations do not occur at specific times or places or with a specific goal in mind. Thus, although some mutations are beneficial to a species, most have no effect or are even detrimental. Indeed, most human diseases, including cancer, have been linked to detrimental genetic mutations. To maintain a low level of mutagenesis, all species have evolved multiple mechanisms for preventing and removing mutations. Because these mechanisms are conserved across species, the genetic tractability of the yeast *Saccharomyces cerevisiae* has provided us with an invaluable model system in which to study these mechanisms. Much of the work in the mutagenesis field has been performed using *S. cerevisiae*, and the majority of mechanisms of mutation prevention and removal are best characterized in this species. Common types of mutations and mechanisms of mutation prevention and removal are discussed below.

1.1 Mutation Origins

Many different types of mutations can occur in DNA, including base substitutions, frameshifts, large deletions and duplications, translocations, and chromosome loss. The majority of these mutations occur within the context of DNA replication. Thus, studies of mutagenesis require an understanding of DNA replication.

For each cell cycle, the yeast cell must replicate its entire genome. DNA replication is a complicated process that requires several different protein complexes to ensure that the genome is copied with high fidelity. First, replication is initiated at distinct replication origins, referred to as autonomously replicating sequences (ARSs), across the genome. A complex of proteins including the hexameric Mcm2-7 complex, the origin recognition complex (ORC), Cdc6, Cdt1, and Mcm10 are localized to the origins and facilitate the unwinding and opening of the DNA duplex (shown in Figure 1.1a). Once the DNA duplex is opened by this complex, a DNA replication fork is formed on each side of the origin. Proliferating cell nuclear antigen (PCNA) is a homotrimer sliding clamp that is loaded onto the DNA by the RFC clamp loader complex and anchors DNA polymerases to the DNA (reviewed in Moldovan et al., 2007). The DNA polymerases can then begin copying each DNA strand, and the replication forks begin moving bidirectionally away from the origin. The anchoring of DNA polymerases by PCNA increases their processivity and facilitates DNA synthesis. Because DNA is double stranded, polymerases must copy each of the two strands. Replication of each strand is initiated by a short RNA primer, from which DNA polymerase continues synthesis using the parental DNA strand as a template. Because these strands are

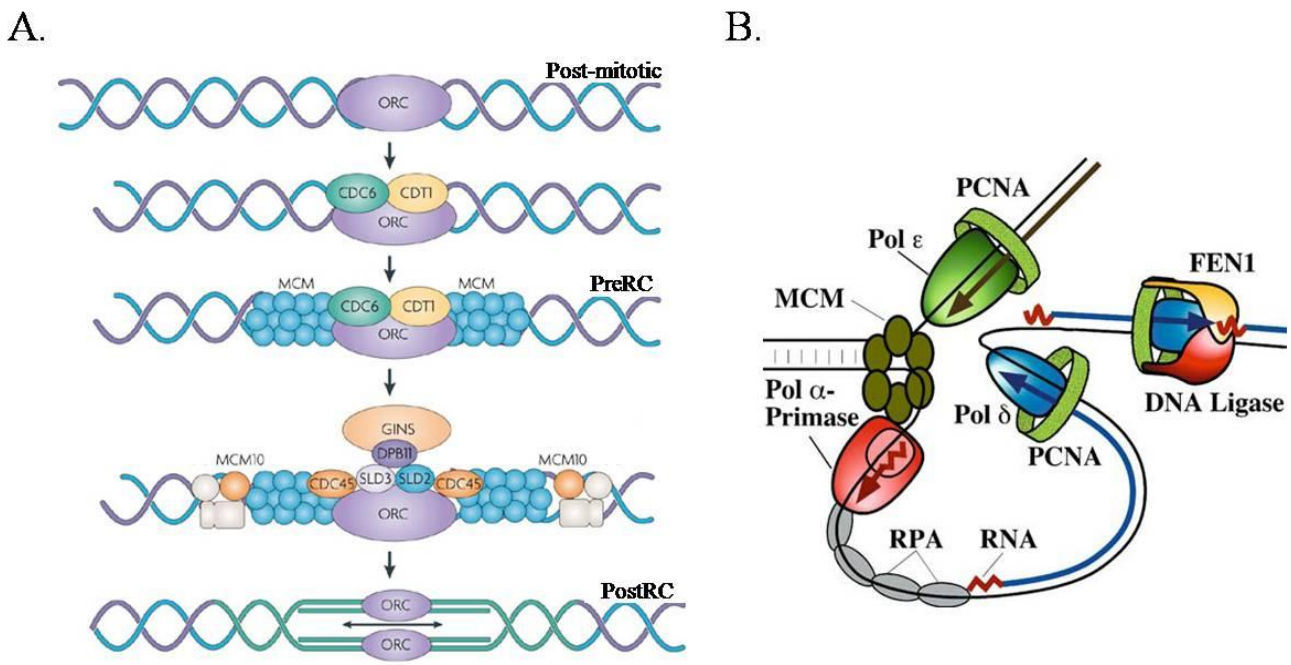

Figure 1.1 DNA Replication

(A) A complex of proteins including the Mcm2-7 complex (MCM), the origin recognition complex (ORC), Cdc6, Cdt1, and Mcm10 bind each origin, forming the pre-replication complex (PreRC). This complex opens the DNA duplex and permits access of PCNA and DNA polymerases. Image obtained from (Aladjem, 2007). (B) The "trombone" model of DNA synthesis proposes that the lagging strand loops out to enable synthesis of both strands in a concerted process. In leading-strand synthesis, Polε synthesizes DNA in a continuous manner. In lagging-strand synthesis, Polα initiates synthesis of individual Okazaki fragments. Polδ continues synthesis of the fragments, which are then ligated together by DNA ligase. Fen1 is required to remove the RNA primer of each Okazaki fragment, and RPA stabilizes the single-stranded regions of the lagging-strand template. Image obtained from (Garg and Burgers, 2005a).

antiparallel and DNA polymerization can occur in only the 5' to 3' direction, synthesis of the strands occurs in opposite directions, one away from the ARS and the other toward the ARS (Figure 1.1b). Synthesis away from the ARS is mostly continuous and is referred to as leading-strand synthesis. Leading-strand synthesis is very processive, generating one long strand of DNA that extends along with replication fork movement. Synthesis toward the ARS is discontinuous, however, and is referred to as lagging-strand synthesis. During this process, an RNA primer is used to initiate synthesis of individual 100-150-bp DNA fragments, referred to as Okazaki fragments. As the replication fork moves and the DNA duplex opens, new Okazaki fragments are initiated. Once the Okazaki fragments are synthesized, the RNA primer of each fragment is removed by the endo/exonuclease Fen1, and each of the fragments are subsequently ligated together by the DNA ligase encoded by the *CDC9* gene (Hubscher and Seo, 2001; MacNeill, 2001; Rossi et al., 2006). As shown in Figure 1.1b, the lagging-strand template is thought to loop out in such a way that the DNA polymerases are kept in close proximity and DNA replication of both the leading and lagging strand is coordinated. This model, referred to as the "trombone model" of DNA replication, was originally suggested by Bruce Alberts and colleagues in the 1970s (Hanawalt et al., 1975; Sinha et al., 1980), and electron microscopy has recently been used to visualize these loops in T7 bacteriophage (Park et al., 1998).

In *S. cerevisiae,* there are three DNA polymerases that are important for DNA replication: Polα, Polδ, and Polε. Polα is both a DNA polymerase and an RNA primase and is responsible for initiating DNA synthesis. *POL1* and *POL12* encode the catalytic and accessory components, respectively, of the DNA polymerase activity of Polα, and

PRI1 and *PRI2* encode the catalytic and accessory components, respectively, of the primase activity. For each strand, Polα synthesizes a small RNA primer (8-12 nucleotides) and then extends this primer by synthesizing ~20 DNA nucleotides (reviewed in Arezi and Kuchta, 2000). Polα is then replaced by either Polδ or Polε. Polδ is comprised of the catalytic subunit Pol3 and the accessory subunits Pol31 and Pol32. Although only Pol3 and Pol31 are essential for DNA replication, Pol32 increases the efficiency of Polδ DNA synthesis (Burgers and Gerik, 1998). Pol32 has been shown to interact with Polα, PCNA, and, interestingly, the translesion synthesis polymerase Polζ (discussed below; Gerik et al., 1998; Huang et al., 2000; Huang et al., 1999). Polδ has been shown to primarily replicate the lagging strand (Nick McElhinny et al., 2008). Because of the nature of lagging-strand synthesis, it is important to note that Polα and PCNA are required to begin synthesis of each Okazaki fragment. Polε is comprised of the catalytic subunit Pol2 and the accessory subunits Dpb2-4 and is thought to be responsible for synthesizing the leading strand (Pursell et al., 2007). Although both *POL2* and *DPB2* were originally characterized as being essential genes, later studies revealed that the catalytic domain of Pol2 was dispensable (Araki et al., 1991; Dua et al., 1999; Kesti et al., 1999; Morrison et al., 1990). Although it is important to note that *pol2* mutants are very sick, the ability of the cell to survive suggests that Polδ is able to partially compensate for the loss of Polε (Ohya et al., 2002).

These polymerases effectively read each base (adenine (A), thymine (T), guanine (G), or cytosine(C)) on the parental strand and select the complementary base (T, A, C, or G, respectively) to synthesize the new strand. Structural studies have shown that these polymerases have small active site pockets and that only correct base pairs fit snugly in

these pockets, resulting in a high selectivity for correct nucleotide insertion (reviewed in McCulloch and Kunkel, 2008). Aside from this high selectivity, Polδ and Polε also possess a proofreading function that serves to "double check" the new base pair. These polymerases are unable to efficiently extend from mismatched base pairs. If polymerase extension is delayed, the primer terminus is moved to the 3' exonuclease active site of the polymerase, which can remove newly inserted bases that are deemed incorrect. Consistent with Polα being less processive and lacking this 3' exonuclease activity, it seems to only be involved in initiating synthesis. It is then displaced by either Polδ or Polε, which are very processive and efficient. Interestingly, it was recently shown that Polδ efficiently proofreads errors that are generated by Polα, further supporting the assignment of Polδ as the lagging-strand polymerase (Pavlov et al., 2006).

Despite the enormous task of duplicating the entire genome (approximately 13 Mb in 30 minutes in yeast), this process is extremely efficient and occurs with high fidelity. It has been estimated that less than one error is generated for every 100,000-1,000,000 nucleotide insertions during DNA replication (reviewed in Kunkel and Burgers, 2008 and McCulloch and Kunkel, 2008). The most common types of replication errors are base substitions and frameshifts (Figure 1.2). Base substitions arise when a DNA polymerase inserts an incorrect base opposite the template DNA base. Frameshift mutations can be generated when either the template or newly synthesized strands misalign and DNA polymerase either misses a base or incorrectly inserts an extra base, respectively. This slippage is thought to occur primarily at runs of four or more identical bases, and generates either a deletion or an insertion, respectively.

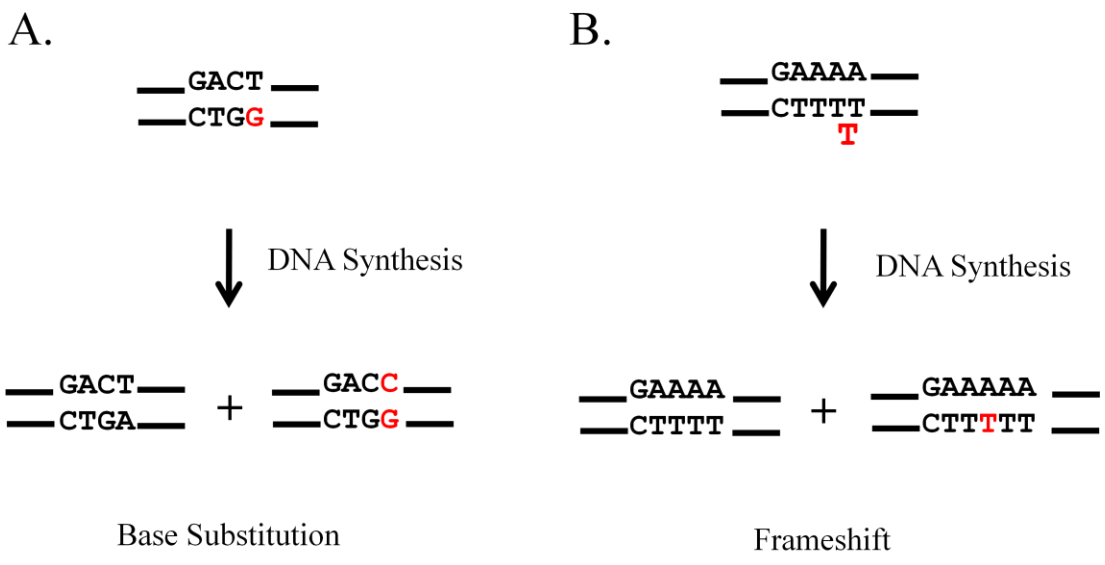

Figure 1.2 Replication Errors

(A) Misincorporation of guanine (G; in red) across thymine (T) results in a DNA duplex that contains a TA to CG base substitution after the next round of replication. (B) Misalignment or slippage of the primer strand (lower strand; red) results in an extra base being incorporated into one DNA duplex (frameshift; red) after the next round of replication.

While DNA replication alone can generate a low level of mutations, many replication errors are caused by underlying DNA damage. This damage may modify bases such that they are miscoding or actually block DNA polymerase during replication. At miscoding bases, DNA polymerase inserts the "correct" base (i.e., what it identifies as being complementary to the damaged base) and continues replication. Other types of DNA damage either do not provide coding information or are sufficiently bulky to block DNA polymerase. In either case, replication is stalled. Stalled replication is detected by checkpoint proteins (e.g., Mec1), which signal for cell cycle arrest to allow the stalled fork to be repaired. Without mechanisms to deal with stalled replication (discussed below), the cell will attempt to segregate an incompletely replicated genome, likely leading to genome instability and cell death.

There are two major types of DNA damage: induced and spontaneous. Induced DNA damage occurs when the cell is exposed to exogenous mutagens. These include harmful chemicals such as tobacco smoke, hydrocarbons, and other carcinogens, and physical agents such as UV irradiation from the sun and ionizing radiation from X-rays. Most mutagenesis research has focused on induced DNA damage because it is easy to control (i.e., precise dosage levels in a controlled environment) and the specific lesions that are caused by a given mutagen are often known (reviewed in Friedberg et al., 2006). For example, UV radiation generates dimers between adjacent pyrimidines in DNA, and ionizing radiation can generate protein- and DNA-DNA cross-links. These DNA lesions block DNA polymerases, preventing the completion of DNA replication.

Spontaneous DNA damage refers to any type of damage that is caused by normal cellular processes. As all cells contain water and oxygen, hydrolysis and reactive oxygen species (ROS) are the most common causes of spontaneous DNA damage. Hydrolysis can modify DNA directly, often resulting in the formation of abasic sites and the deamination of cytosine to form uracil. Abasic sites, also referred to as apurinic/apyrimidinic or AP sites, are places where the sugar-phosphate backbone of DNA remains intact but the base is missing. These are one of the most common types of spontaneous DNA damage, with an estimated 10,000 AP sites being generated every day per human cell (Lindahl, 1979). These sites provide no coding information for DNA polymerase, and replication is thus stalled at these sites. The deamination of cytosine into uracil generates a miscoding site in DNA; the DNA polymerase inserts an adenine opposite the uracil, generating a CG to TA transition.

ROS are generated when oxygen-containing compounds are broken down into highly reactive species, such as superoxide radical ($\bullet O_2^-$) and the hydroxyl radical (•OH) (reviewed in Friedberg et al., 2006). Although ROS can be generated by physical and chemical mutagens such as UV radiation and H_2O_2, they are also generated in the cell by normal metabolic processes such as aerobic respiration (Friedberg et al., 2006; Maynard et al., 2009). Cells contain multiple antioxidants and other proteins that protect the genome from oxidative damage, such as superoxide dismutase (Sod1) and other peroxiredoxins, but ROS are still implicated as causal agents of many diseases and of aging (D'Errico et al., 2008; Skinner and Turker, 2005). ROS can directly oxidize DNA, forming different types of modified bases. One common example of this is the oxidation of guanine, which forms 7,8-dihydro-8-oxoguanine or GO. GO lesions are often miscoding, with adenine often replacing cytosine opposite the GO lesion, thus generating GC to TA transversions. This specific lesion is discussed further in Chapters 2 and 3.

1.2 DNA Repair Pathways

While the previous section describes only a few of the common types of mutations that are observed in the cell, it is important to note that there are many others that frequently occur. Fortunately, the cell has developed several different mechanisms for dealing with different types of mutations and DNA damage. In most cases, damage is quickly and efficiently repaired or bypassed by one of the many highly conserved pathways. As shown in Figure 1.3, these pathways can be divided into three major groups based on when they are most active: those that act prior to replication, those that

act during replication (tolerance pathways), and those that act on newly synthesized DNA after replication.

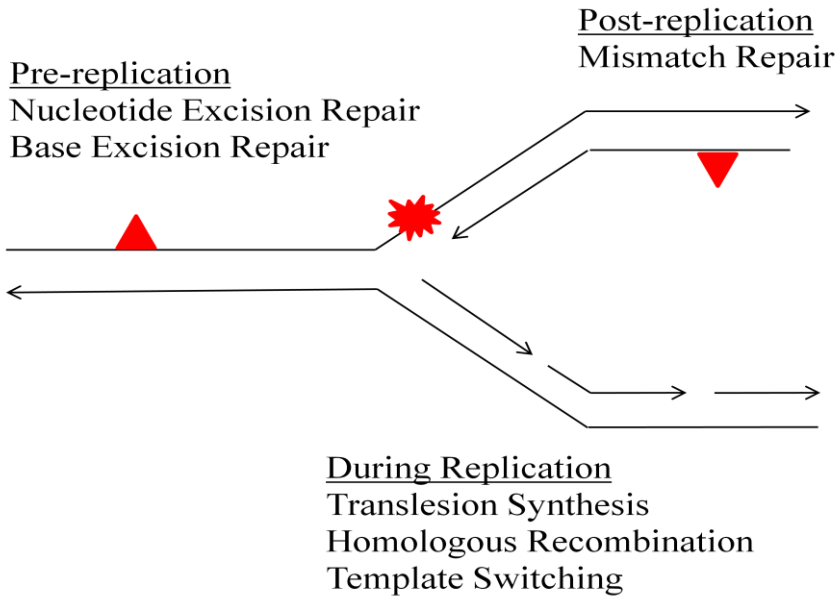

Figure 1.3 DNA Repair and Tolerance Pathways

1.2.1 Pre-replication Repair Pathways

The synthesis (S) phase of the cell cycle constitutes a small portion of the overall cell cycle. Thus, most DNA damage is thought to occur and be repaired outside of S phase. This enables the cell to remove damage before it is encountered by DNA polymerase, thereby preventing the generation of either a permanent mutation or a blocked replication fork. There are two major pre-replication repair pathways: nucleotide excision repair (NER) and base excision repair (BER; Figure 1.4).

NER is primarily responsible for removing bulky, helix-distorting DNA lesions. These include UV-induced dimers, ionizing radiation-induced cross-links, and chemical-induced DNA adducts. In yeast, the Rad4-Rad23 complex identifies the bulky lesion and is then joined by the Rad14, TFIIH, and RPA proteins. Rad14 is involved in damage

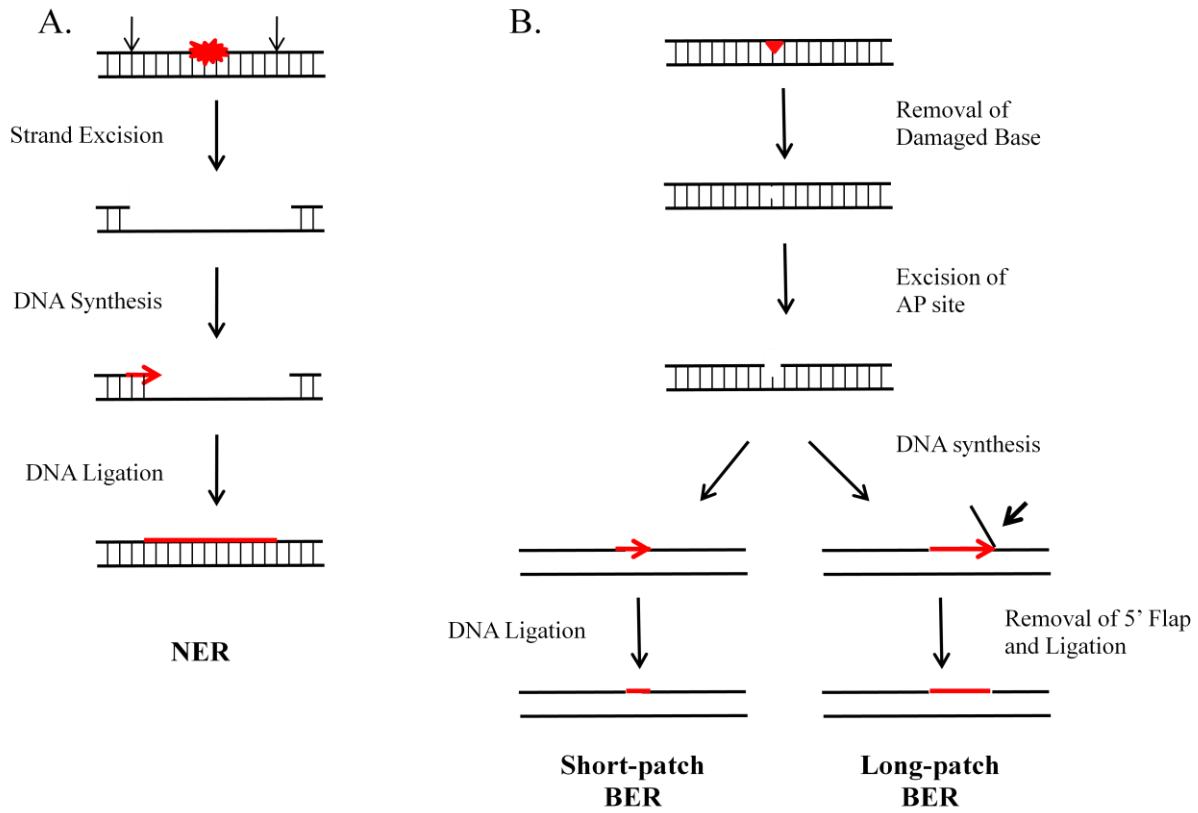

Figure 1.4 Nucleotide Excision Repair and Base Excision Repair

(A) In nucleotide excision repair (NER), bulky DNA lesions are removed by excising the region of the DNA strand (25-30 nucleotides) that contains the lesion. Once the lesion-containing region is removed, the remaining gap is filled in by DNA polymerase. (B) In base excision repair (BER), a DNA glycosylase or AP endonuclease will remove a damaged base or AP site, respectively. Once the damage is removed, DNA polymerase will fill in the remaining gap (short-patch repair) or will fill in the gap and continue synthesis for 2-10 nucleotides (long-patch). In long-patch repair, Fen1 is required to remove the displaced DNA strand.

recognition, TFIIH is a helicase that helps unwind the DNA, and RPA binds single-stranded DNA and removes secondary structure. Rad2 nicks the DNA on the 3' side of the lesion (approximately 2-8 nucleotides away from the lesion), and Rad1-Rad10 nicks the DNA on the 5' side (approximately 15-24 nucleotides away from the lesion; reviewed in Friedberg et al., 2006). The region that contains the lesion is thus excised, generating a 25-30-bp single-stranded gap. PCNA is loaded onto the DNA and then recruits either Polδ or Polε to synthesize a new DNA strand, and DNA ligase seals the remaining nick.

NER can occur either as a global damage recognition and removal pathway or in the context of DNA transcription, which is referred to as transcription-coupled repair. If DNA that is being transcribed contains a lesion, RNA polymerase can be blocked. This blockage signals the recruitment of the transcription-coupled repair machinery, which largely consists of NER proteins (reviewed in Mellon, 2005). The blocking lesion is removed as described above, and transcription can continue. Because of this repair mechanism, many studies have found higher mutation rates associated with the non-transcribed DNA strand (Mellon and Hanawalt, 1989; Mellon et al., 1987; Sweder and Hanawalt, 1992). It should be noted, however, that this bias may also be due to the increased susceptibility of the single-stranded, non-transcribed DNA strand to damage during transcription (Korzheva et al., 2000). Null mutations in yeast NER proteins are associated with increased sensitivity to damaging agents and increased rates of mutagenesis. In humans, mutant forms of NER proteins are associated with xeroderma pigmentosum, a condition defined by severely enhanced sensitivity to sunlight, and Cockayne syndrome, in which patients have enhanced sensitivity to sunlight and other developmental defects (Nouspikel, 2009).

Although BER is also involved in removing DNA lesions prior to replication, it is responsible for removing individual nucleotides rather than an oligonucleotide that contains a lesion. BER is initiated in response to spontaneous base loss (i.e., AP sites) or by specific DNA glycosylases that detect specific base lesions (e.g., GO lesions). The BER pathway in yeast contains five DNA glycosylases that each remove specific types of damaged bases and two AP endonucleases. DNA glycosylases catalyze the hydrolysis of the N-glycosidic bond between a base and the sugar-phosphate backbone of DNA. This activity effectively removes a base while keeping the DNA backbone intact, generating an AP site. Some DNA glycosylases are associated with an AP lyase activity (e.g., Ntg1, Ntg2, and Ogg1) that can nick the sugar-phosphate backbone on the 3' side of the AP site. Other DNA glycosylases do not have an associated AP lyase activity (e.g., Ung1 and Mag1) and require an additional AP endonuclease to nick the backbone on the 5' side of the AP site. After cleavage by either an AP endonuclease or AP lyase, the ends must be processed to generate either a 5' phosphate or a 3' hydroxyl end, respectively. A DNA polymerase will then fill in the 1-nucleotide gap (short-patch repair) or continue synthesis for a few bases (long-patch repair), displacing one of the DNA strands. In short-patch repair, an exonuclease or DNA-deoxyribophosphodiesterase is required to remove the 5' sugar-phosphate residue. In long-patch repair, Fen1 is required to remove the displaced DNA strand. In both types of BER, DNA ligase then seals the nick between the newly synthesized nucleotide(s) and the adjacent nucleotide, completing the repair process.

As mentioned above, yeast contain five DNA glycosylases: Ung1, Mag1, Ogg1, Ntg1, and Ntg2. Ung1 specifically removes uracil from DNA, which arises from either

deamination of cytosine or direct incorporation of uracil into DNA (Crosby et al., 1981), and recent work in our lab has shown that Ung1 plays an important role in removal of uracil that is selectively incorporated into highly transcribed regions of DNA (Kim and Jinks-Robertson, 2009). Mag1 specifically removes methylated DNA bases (Chen et al., 1989). Ung1 and Mag1 do not have AP lyase activity and thus generate AP sites that require the additional activity of an AP endonuclease. The Ogg1 glycosylase efficiently removes GO and FaPy oxidative lesions (van der Kemp et al., 1996). As discussed in Chapters 2 and 3, deletion of Ogg1 results in a specific increase in GO-associated mutations in the cell. Ntg1 and Ntg2 are redundantly involved in the removal of ring-saturated or fragmented pyrimidines (Senturker et al., 1998; Swanson et al., 1999). Ogg1, Ntg1, and Ntg2 possess an associated AP lyase activity and are thus efficient at removing their cognate lesions without AP endonuclease.

There are two AP endonucleases in yeast: Apn1 and Apn2. Apn1 is estimated to be responsible for 97% of all AP endonuclease activity in the cell, with Apn2 only serving a minor, redundant role (Popoff et al., 1990). Apn1 repairs all AP sites in the cell, both those generated spontaneously and those that are generated by the DNA glycosylases Ung1 and Mag1. Loss of Apn1 is associated with elevated rates of mutagenesis. Ntg1 and Ntg2 can also repair AP sites, as the combined removal of Apn1, Ntg1, and Ntg2 results in a synergistic increase in mutagenesis (Swanson et al., 1999).

The NER and BER pathways are often thought of as the cell's first line of defense againt DNA damage. These two pathways accurately remove DNA damage before it results in the generation of permanent mutations or blocked replication forks, which can result in cell death. Despite the efficiency of these pathways, however, some DNA

damage may remain as the cell enters S phase and begins replication of the genome. Once DNA polymerase encounters this damage, one of the tolerance pathways must be recruited to bypass the damage.

1.2.2 DNA Damage Tolerance Pathways

Because DNA replication is a streamlined, time-sensitive process, most repair of DNA damage is thought to occur outside of S phase. When DNA damage is encountered by the replication fork, it is instead tolerated or bypassed to enable replication to continue. Once replication is completed, the damage can be repaired by one of the two pathways discussed above, NER or BER. There are three major tolerance pathways that act during DNA replication: homologous recombination, template switching, and translesion synthesis (TLS; Figure 1.5).

Homologous recombination and template switching are both high-fidelity pathways that enable DNA polymerase to use complementary sequences to continue synthesis past a lesion. Some of the key proteins involved in homologous recombination include Rad51, Rad52, Rad55-Rad57, and Rad54 (reviewed in Krogh and Symington, 2004). During this process, the 3' end of a single strand of DNA is coated with Rad51 (Shinohara et al., 1992). Rad52 is bound to the tail end of the single strand and mediates both the interaction between Rad51 and the single-stranded DNA and the strand exchange that occurs when the single strand invades the homologous duplex (Benson et al., 1998; New et al., 1998; Shinohara and Ogawa, 1998). Rad55-Rad57 helps stabilize this process, and Rad54 induces topological changes in the DNA that help facilitate strand separation of the homologous duplex (Petukhova et al., 1999; Sung, 1997; Van Komen et al., 2000). Once the single strand invades the duplex, DNA polymerase uses

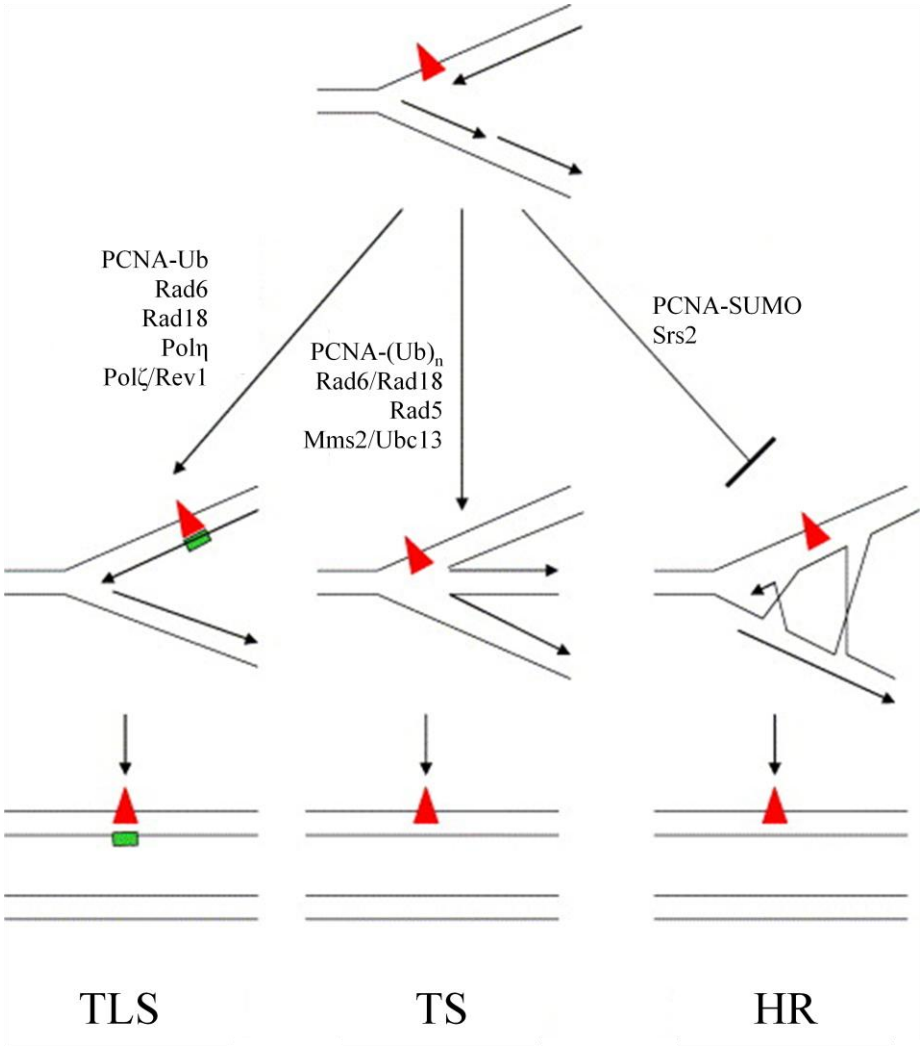

Figure 1.5 Tolerance Pathways

At a stalled or blocked replication fork, the cell can employ one of three pathways: translesion synthesis (TLS), template switching (TS), or homologous recombination (HR). Most TLS requires ubiquitination of PCNA (PCNA-Ub) by Rad6/Rad18 complex. TS requires polyubiquitination of PCNA (PCNA-$(Ub)_n$), Rad5, Mms2, and Ubc13. One common model of TS is that the replication fork regresses to allow DNA polymerase to use the newly synthesized sister DNA strand to template DNA synthesis and thereby bypass the lesion. In HR, the blocked DNA strand invades the complementary DNA duplex and uses the newly synthesized sister DNA to template DNA synthesis and thereby bypass the lesion. Image modified from (Watts, 2006).

the undamaged, complementary strand to template DNA synthesis and bypass the lesion

in the damaged DNA strand. It is important to note that homologous recombination is

not restricted to S phase and the tolerance of blocked replication forks; homologous recombination is used to repair single- and double-strand breaks that occur in any stage of the cell cycle. In contrast to homologous recombination, template switching is thought to involve regression of the replication fork to enable DNA polymerase to use the newly synthesized complementary strand as a template for continuing replication, thereby avoiding synthesis past the lesion. Although fork regression is just one model for template switching and this pathway is still being worked out mechanistically, it has been shown to require Rad6-Rad18, Rad5, Mms2, and Ubc13 (Ulrich and Jentsch, 2000). Rad6 is a ubiquitin-conjugating enzyme that acts with Rad18, an E3 ubiquitin ligase that binds single-stranded DNA, and ubiquitinates PCNA (Hoege et al., 2002). Mms2 and Ubc13 are ubiquitin-conjugating enzymes and can continue this ubiquitination, generating a polyubiquitin tail that is thought to signal template switching (Hoege et al., 2002). Rad5 is a helicase that interacts with Mms2 and Ubc13 and is required for fork regression (Blastyak et al., 2007; Ulrich and Jentsch, 2000). Although this process is in some ways similar to homologous recombination, it does not appear to require the same recombination proteins (i.e., Rad52) (Gangavarapu et al., 2007; Zhang and Lawrence, 2005).

The third tolerance pathway is TLS. Unlike the homologous recombination and template switching pathways, TLS is generally error prone. Instead of using a complementary DNA strand to bypass damage, TLS involves the recruitment of specialized DNA polymerases that can directly bypass the lesion by either inserting a nucleotide across from the lesion and/or extending from a lesion-base mispair (Prakash and Prakash, 2002; Woodgate, 2001). The TLS polymerases are able to bypass lesions

because of a relatively large catalytic active site pocket that can accommodate structurally deformed bases (Ling et al., 2001). As a result, TLS polymerases have high error rates when synthesizing across undamaged DNA (reviewed in McCulloch and Kunkel, 2008). These polymerases also have less processivity, and thus tend to fall off the DNA after incorporating only a few nucleotides. With one exception (Abdulovic et al., 2007), studies of TLS have found that PCNA is required for TLS polymerases to access DNA (reviewed in Prakash et al., 2005). In principle, when a replicative DNA polymerase is blocked by a lesion, a TLS polymerase is recruited to the site, replicates past the lesion, and is then replaced by the replicative polymerase, which continues replication. It is important to note, however, that evidence suggests that TLS may not always occur at the replication fork (Radman, 2005). Instead, DNA polymerase may stop replication, move past a lesion, and then continue replication once again. This generates single-stranded gaps in the newly synthesized DNA, and such gaps have been visualized by electron microscopy (Lopes et al., 2006). It has been suggested that TLS is recruited to these sites and fills in these gaps (reviewed in Sale et al., 2009). It is not yet clear whether DNA polymerase is more likely to stall and wait for a TLS-specific polymerase or reinitiate downstream of the lesion and continue replication. Although it seems likely that both scenarios are possible depending on the region of DNA and the type of blockage, further research is needed to better understand this process.

Because TLS polymerases are less accurate in replicating past DNA lesions, mutations are often generated. If the TLS polymerase continues to replicate past the DNA lesion before being replaced by a replicative polymerase, mutations can also be introduced into undamaged DNA. Although the low fidelity of the TLS polymerases

may make them seem an odd choice for a tolerance pathway, the alternative is often a blocked replication fork and therefore cell death.

There are three TLS polymerases in yeast: Polζ, Rev1, and Polη. Polζ is composed of the Rev3 catalytic subunit and the Rev7 accessory protein. The second polymerase, Rev1, was originally characterized by its biochemical activity as a deoxycytidyl transferase, specifically inserting a cytosine opposite a lesion (Nelson et al., 1996). However, subsequent *in vivo* studies revealed that this transferase activity does not appear to be the major function of Rev1 (reviewed in Prakash et al., 2005). Instead, Rev1 seems to play a structural role in TLS, and has been shown to be required for Polζ bypass in yeast. Rev1, in conjunction with Polζ, is thought to be responsible for at least 50% of all spontaneous mutations and for approximately 95% of all base-pair substitutions induced by UV irradiation (Lawrence, 2002; Quah et al., 1980). The third polymerase, Polη, is encoded by the *RAD30* gene. Though mutagenic in some cases, Polη is best known for its error-free bypass of UV-induced lesions and of oxidative GO lesions (Gibbs et al., 2005; Haracska et al., 2000; Johnson et al., 1999b; Maga et al., 2007; Yuan et al., 2000).

Despite the mutagenic potential of these polymerases, they play a critical role in the tolerance of DNA damage in all species. Mammalian cells contain homologs of Polζ, Rev1, and Polη as well as a number of other, mostly redundant TLS polymerases. The mouse homolog of Rev3 is known to be essential for early embryonic development (Bemark et al., 2000; Wittschieben et al., 2000), and the absence of Polη in humans has been shown to cause both fragile site instability (Rey et al., 2009) and a variant form of the cancer-prone disease xeroderma pigmentosum, which is characterized by highly

elevated susceptibility to UV-induced skin cancers (Gibbs et al., 2005; Johnson et al., 1999a; Masutani et al., 1999). As mentioned above, this condition is also caused by loss of NER proteins. Interestingly, the immune system of vertebrate species has made use of these low fidelity polymerases in a process called somatic hypermutation (reviewed in Diaz et al., 1999 and Simpson and Sale, 2003). In this process, TLS polymerases are recruited to elevate mutagenesis in the variable regions of immunoglobulin genes. The cell also uses cytosine deamination and recombination to diversify these regions. These variable regions will then be used as receptors on B cells to identify foreign antigens. In this scenario, larger collections of diverse (mutant) receptors are associated with a greater chance of identifying antigens and, thus, a better immune system.

When DNA damage is repaired prior to replication, the type of damage determines whether it will be repaired by NER or BER. This is not the case for the tolerance pathways, and it is still unclear how the cell makes the decision of which tolerance pathway to use at a given replication block. Although homologous recombination is a high-fidelity process, studies have shown that there at least two different mechanisms that prevent its activity during replication. The first mechanism involves modification of PCNA. During replication, Siz1 sumoylates (i.e., modifies with a small ubiquitin-related modifier or SUMO) PCNA at sites K127 and K164, which leads to the recruitment of Srs2 to replication forks. Srs2 blocks homologous recombination by removing Rad51 bound to single-stranded DNA (reviewed in Watts, 2006). Second, recent evidence in our lab has shown that the antirecombination activity of the mismatch repair pathway (MMR; discussed below) also suppresses homologous recombination,

thereby promoting TLS during replication (Lehner and Jinks-Robertson, 2009). These mechanisms prevent unwarranted recombination from disrupting replication.

Modifications of PCNA also help regulate template switching and TLS. As mentioned above, PCNA is sumoylated by Siz1 during S phase, and it appears that this sumoylation is required for both template switching and TLS (Branzei et al., 2008; Hoege et al., 2002). PCNA can be further modified by the Rad6-Rad18 complex, which ubiquitinates PCNA in response to DNA damage (Hoege et al., 2002). As shown in Figure 1.5, PCNA can then either remain monoubiquitinated or become polyubiquitinated by Mms2 and Ubc13. As described above, the polyubiquitinated form of PCNA is associated with template switching. In contrast, some TLS has been shown to require monoubiquitination of PCNA, while other TLS does not appear to require this modification (Chen et al., 2006; Garg and Burgers, 2005b; Haracska et al., 2006; Shen et al., 2006; van der Kemp et al., 2009a; Watanabe et al., 2004; Wood et al., 2007). TLS that does not require monoubiquitination of PCNA is dependent on Rad5 (Minesinger and Jinks-Robertson, 2005). It is currently unclear how this type of TLS occurs or why there are two different types of TLS.

Some TLS is also regulated, at least partially, at the protein level. Although there are low levels of Polη in the cell that function in the bypass of spontaneous DNA damage, studies in yeast have shown that Polη is transcriptionally upregulated in response to UV damage (McDonald et al., 1997). There is also evidence that Polη can be ubiquitinated and thereby targeted for proteolysis (McIntyre et al., 2006; Skoneczna et al., 2007). Rev1 has also been shown to be regulated by Mec1-dependent phosphorylation in a cell cycle-dependent manner (Sabbioneda et al., 2007; Waters and

Walker, 2006). Interestingly, levels of Rev1 were found to be highest in late S/G2 phase. This has led to the hypothesis that TLS is suppressed until the end of replication, where it can fill in single-stranded gaps that are left when DNA polymerase reinitiates synthesis downstream of a lesion (Waters and Walker, 2006).

Finally, recent evidence suggests that TLS may also be regulated by the alternative 9-1-1 checkpoint clamp (Sabbioneda et al., 2005). This clamp, composed of Rad17, Ddc1, and Mec1 in yeast, forms a heterotrimeric ring that physically resembles PCNA and functions in the DNA damage checkpoint response (Majka and Burgers, 2003). Studies in yeast have shown that this clamp physically interacts with Polζ and is at least partially required for Polζ-dependent TLS (Sabbioneda et al., 2005). Interestingly, recent evidence has shown that the 9-1-1 clamp is ubiquitinated by Rad18 in response to DNA damage (Fu et al., 2008). It is currently unclear if or how this ubiquitination helps regulate TLS in the cell.

1.2.3 Post-replication Repair Pathway

Although the cell has evolved a number of mechanisms to help prevent and remove DNA damage, mutations are still generated during DNA replication. As discussed above, these mutations can arise either as a response to DNA damage or as replication errors made by DNA polymerases. The cell has thus developed another repair mechanism that operates behind the replication fork, detecting and removing any mismatches that are present in newly synthesized DNA. This mechanism is referred to as mismatch repair (MMR; Figure 1.6; reviewed in Harfe and Jinks-Robertson, 2000b and Kunkel and Erie, 2005).

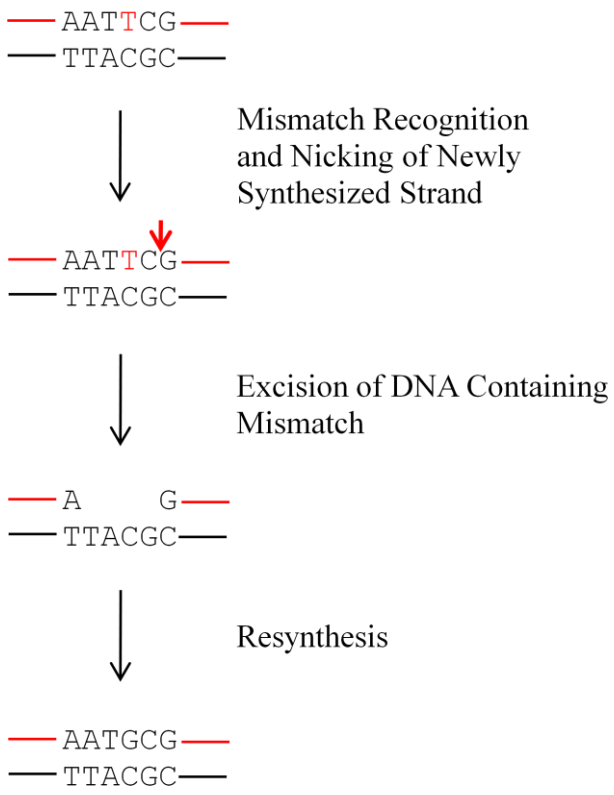

Figure 1.6 Mismatch Repair

Mismatches and frameshift intermediates generated during DNA replication are detected and excised by the mismatch repair (MMR) machinery. After the mismatch or frameshift intermediate is excised from the newly synthesized DNA strand (red), DNA polymerase fills in the remaining gap.

MMR is a complex of proteins responsible for the identification and removal of mismatches generated during DNA replication. Though originally discovered in *E. coli*, MMR is highly conserved in all species and has been well characterized in yeast and other eukaryotes. In yeast, there are two major classes of MMR proteins. First, there are six homologs of the *E. coli* MutS mismatch-recognition protein: Msh1-Msh6. With the exception of the mitochondria-specific Msh1 protein, these proteins combine into heterodimers. The two most relevant to nuclear mutagenesis are the MutSα dimer, composed of Msh2 and Msh6, and the MutSβ dimer, composed of Msh2 and Msh3.

MutSα recognizes base misinsertions and small frameshift intermediates, whereas MutSβ recognizes small and large frameshift intermediates. These two complexes are generally considered to be functionally redundant in the recognition and repair of small frameshift intermediates. Second, there are four homologs of the *E. coli* MutL protein: Mlh1-Mlh3 and Pms1. Mlh1 can form a heterodimer with any of the other three proteins, but MMR appears to use the Mlh1-Pms1 heterodimer most often. This second protein complex is thought to connect mismatch recognition with proteins involved in the downstream steps of mismatch removal. Once the mismatch has been identified, an exonuclease, most likely Exo1, degrades the newly synthesized strand containing the mismatch, allowing a DNA polymerase, most likely Pol δ, to resynthesize the DNA without a mismatch. MMR has been estimated to increase replication fidelity by at least a factor of 100, and cells without a fully functional MMR system display highly elevated rates of mutation (Harfe and Jinks-Robertson, 2000b; Kunkel and Erie, 2005). In humans, mutations in MMR genes have been implicated in hereditary non-polyposis colorectal cancer (HNPCC) (Buermeyer et al., 1999).

1.3 Factors Influencing Mutagenesis

Knowledge of how mutations arise and the mechanisms that act to prevent, remove, or tolerate them is not sufficient to understanding mutagenesis. Since scientists began studying mutagenesis, they noted that mutations do not occur with equal frequency throughout the genome – not all sites are created equal (Coulondre and Miller, 1977). We are only beginning to understand how many other factors in the cell affect

mutagenesis.

The depiction of DNA and its replication in Figure 1.1 is extremely oversimplified. DNA does not exist as a naked string of nucleotides but as a very complex structure referred to as chromatin. DNA is wound around histones, forming nucleosomes that resemble beads on a string. This structure is then further compacted to generate the chromosomes seen under a microscope. Areas of active gene transcription, or euchromatin, are less compacted than silenced regions of the genome, or heterochromatin.

This complex structure of DNA affects both where mutations occur and many of the DNA repair pathways. For example, clusters of mutations in immunoglobulin genes were found to correspond to nucleosome spacing (Shen et al., 2009). In these studies, deamination of cytosine by the AID protein was found to occur in DNA that was associated with nucleosomes, and the DNA between the nucleosomes was less likely to be mutated. In contrast, NER, BER, homologous recombination, and MMR have been shown to be less efficient at regions of DNA associated with nucleosomes (Chaudhuri et al., 2009; Li et al., 2009; Menoni et al., 2007; Osley et al., 2007). Efficient activity of these pathways requires that the histones associated with the nucleosomes be modified to allow access of repair proteins.

Aside from the complex structure of chromatin, other aspects of DNA affect where mutagenesis occurs. As described above, DNA replication is divided into leading-strand synthesis and lagging-strand synthesis. As the mechanisms involved in the synthesis of the leading and lagging strands during replication differ, the processes involved in handling DNA lesions on the two strands may also be expected to differ.

Indeed, differences in mutation rates on the leading and lagging strands were noted in *Escherichia coli* several years ago (reviewed in Radman, 1998). However, these studies have yielded conflicting results. Examinations of frameshift mutagenesis, for example, have found higher rates of mutation on either the leading or the lagging strand, depending on the experimental system used (Gawel et al., 2002; Iwaki et al., 1996; Trinh and Sinden, 1991; Yoshiyama et al., 2001). Studies of deletion mutagenesis have revealed no strand bias, whereas studies of base substitution mutagenesis have inferred a bias for mutations on the leading strand (Fijalkowska et al., 1998; Nagata et al., 2005). Studies in the yeast *S. cerevisiae* and in mammalian cells have resulted in more consistent conclusions. These studies have demonstrated differences in leading- and lagging-strand mutagenesis with a clear bias for mutations to occur on the leading strand (Dumstorf et al., 2006; McGregor et al., 1999; Pavlov et al., 2003; Touchon et al., 2005; Unniraman and Schatz, 2007).

It is possible that the leading-strand mutation bias is due to different fidelity of DNA synthesis on the two strands and/or to unequal repair of mutations that arise during DNA replication. Recent studies in yeast have demonstrated a higher efficiency of the MMR machinery on the lagging strand during replication (Kow et al., 2007; Pavlov et al., 2002). MMR may, therefore, play a major role in the differences seen in rates of leading- and lagging-strand mutagenesis. It is also possible that TLS bypass has an effect on the different mutations rates seen associated with either leading- or lagging-strand synthesis. Studies in *E. coli* have suggested that SOS-associated translesion bypass activity occurs differently on the two newly synthesized strands and previous work in our lab has shown that spontaneous Polη bypass in yeast may be occurring differently on the two strands as

well (Abdulovic et al., 2007; Maliszewska-Tkaczyk et al., 2000; Veaute and Fuchs, 1993).

Mutagenesis has also been shown to vary depending on the primary sequence and local sequence context. For example, previous work in our lab has shown that there are very specific hotspots of Polζ- and Polη-dependent spontaneous mutagenesis within a ~150-bp reversion window (Abdulovic et al., 2007; Harfe and Jinks-Robertson, 2000a). These hotspots appear to require their immediately surrounding sequence context, as identical sites with different surrounding sequence context are not hotspots. Interestingly, when some of these hotspots were individually moved within the reversion window, some retained their hotspot characteristics whereas others were still hotspots, but had a different spectrum of mutations (Abdulovic et al., 2008). The ability to recapitulate a hotspot at another location may suggest that local sequence context and not chromatin structure are critical for that specific site to be susceptible to mutagenesis.

Transcription can have both positive and negative effects on mutagenesis. As described above, lesions that are encountered by RNA polymerase during transcription can lead to the recruitment of NER proteins to remove the lesion. In this way, transcription serves as a monitor of DNA damage and can help remove damage before it is encountered by DNA replication. However, rates of transcription have also been shown to be directly proportional to rates of mutagenesis (Datta and Jinks-Robertson, 1995; Kim et al., 2007). This is likely due to the fact that during transcription, the transcribed strand forms a temporary DNA:RNA hybrid, while the non-transcribed strand is single stranded and thus more susceptible to DNA damage (Korzheva et al., 2000). Both cytosine deamination and the oxidation of guanine to form GO lesions has been

shown to occur more frequently on the non-transcribed strand in *E. coli* (Beletskii and Bhagwat, 1996; Klapacz and Bhagwat, 2005). Interestingly, recent evidence from our lab has shown that highly transcribed regions of DNA are associated with significantly increased incorporation of uracil during DNA replication (Kim and Jinks-Robertson, 2009). This suggests that transcription of DNA may occur in regions of the nucleus that contain high concentrations of dNTPs, including dUTP, which may lead to increased mutagenesis.

There are clearly many factors that are involved in mutagenesis in the cell. It is important to note that this introduction includes only some of these factors, as there is evidence that other cellular mechanisms, such as the maintenance of precise concentrations of dNTPs and accurate cell cycle and DNA damage checkpoints, influence mutagenesis. It is critical that these factors are taken into account in order to better understand all that is involved in the generation of mutations.

1.4 Summary

Despite our growing knowledge of how mutations occur, the mechanisms that are in place to prevent their occurrence and/or remove them, and other factors that may be contributing to these processes, there is still much that remains unknown. In the experiments presented in this thesis, I address some of the outstanding issues in the field. In Chapter 2, I have examined the role of Polη bypass in DNA replication of oxidative GO lesions. I have specifically addressed how this bypass is related to MMR of GO-associated mismatches. I have also used Polη-dependent bypass of GO lesions as a read out by which to characterize Polη activity and how it is regulated. In Chapter 3, I have

used GO-associated mutagenesis to examine the effect of replication timing on mutagenesis in yeast. Finally, in Chapter 4, I have examined spontaneous frameshift mutagenesis in regions of DNA that do not contain long mononucleotide runs. Although frameshifts are generally thought to occur in runs of four or more nucleotides, these experiments reveal that this is not the case; runs of two or three nucleotides can also be hotspots for frameshift mutagenesis. These experiments further examine the role of MMR proteins in preventing this mutagenesis. Together, these experiments provide further information about how mutations arise and the mechanisms that are in place to maintain low levels of mutagenesis in the cell.

Chapter 2: The Polη translesion synthesis DNA polymerase acts independently of the mismatch repair system to limit mutagenesis caused by 7,8-dihydro-8-oxoguanine

2.1 Summary

Reactive oxygen species are ubiquitous mutagens that have been linked to both disease and aging. The most studied oxidative lesion is 7,8-dihydro-8-oxoguanine (GO), which is often miscoded during DNA replication, resulting specifically in GC > TA transversions. In yeast, the mismatch repair (MMR) system repairs GO:A mismatches generated during DNA replication, and the Polη translesion synthesis DNA polymerase additionally promotes error-free bypass of GO lesions. It has been suggested that Polη limits GO-associated mutagenesis exclusively through its participation in the filling of MMR-generated gaps that contain GO lesions. In the experiments reported here, the *SUP4-o* forward mutation assay was used to monitor GC >TA mutation rates in strains defective in MMR (Msh2 or Msh6) and/or in Polη activity. Results clearly demonstrate that Polη can function independently of the MMR system to prevent GO-associated mutations, presumably through preferential insertion of cytosine opposite replication-blocking GO lesions. Furthermore, the Polη-dependent bypass of GO lesions is more efficient on the lagging strand of replication, requires an interaction with PCNA, and does not require ubiquitination of PCNA or the alternative 9-1-1 clamp. These studies establish a new paradigm for the prevention of GO-associated mutagenesis in eukaryotes.

2.2 Introduction

The most common oxidized DNA lesion is 7,8-dihydro-8-oxoguanine, which will

be referred to here as a GO lesion. The mutagenic potential of this lesion is due to miscoding during DNA synthesis, with replicative DNA polymerases usually misinserting adenine across from the lesion to generate GO:A mispairs and ultimately GC > TA transversions (Shibutani et al., 1991). Studies examining the crystal structure of T7 DNA polymerase complexed with a GO:C base pair or a GO:A mispair indicate the basis of this mutagenic specificity. Whereas the GO:C structure physically resembles that of a mismatch, the GO:A mispair structurally resembles a normal Watson-Crick base pair and, therefore, is likely to escape polymerase-associated proofreading activity (Brieba et al., 2004). A GO-containing nucleotide triphosphate (8-oxo-dGTP) can also be used by DNA polymerases during DNA synthesis, leading specifically to AT > CG transversion events (Cheng et al., 1992).

There are three major proteins in *Escherichia coli* that work independently to prevent GO-associated mutagenesis: MutM (Fpg), MutY, and MutT (Michaels and Miller, 1992). MutM is a DNA glycosylase that removes GO lesions in the GO:C base pairs created by oxidation of guanine in normal G:C base pairs, while MutY is an adenine-DNA glycosylase that removes adenines from the GO:A mispairs created by incorporation of adenine opposite a GO lesion. If DNA replication occurs before MutM can remove the GO lesion from a GO:C base pair, the lesion will likely generate a GO:A mispair, which is then subject to the A-specific activity of the MutY protein. Once MutY removes the adenine from the newly-synthesized strand, a cytosine can be inserted opposite the lesion, giving MutM another opportunity to excise the GO base. MutT is an 8-oxo-dGTPase that degrades 8-oxo-dGTP, thereby greatly reducing its incorporation into DNA. The postreplicative mismatch repair (MMR) pathway has also been

implicated in preventing GO-associated mutagenesis in *E. coli*, by functioning as an alternative to MutY or by helping MutY identify and remove mismatched adenines from GO:A mispairs (Bai and Lu, 2007; Wyrzykowski and Volkert, 2003).

In the yeast *Saccharomyces cerevisiae*, the Ogg1 protein is the functional homolog of MutM (van der Kemp et al., 1996) and thus removes GO lesions that are base paired with cytosine. The MMR machinery is functionally analogous to the MutY protein (Ni et al., 1999), excising adenines that are inserted opposite GO lesions during DNA replication. The mismatch-recognition MutSα complex (a heterodimer of the Msh2 and Msh6 proteins) specifically recognizes GO:A mispairs and initiates removal of the portion of the newly-synthesized strand containing the adenine (Ni et al., 1999). A homolog of MutT has yet to be identified in yeast, although one does exist in mammalian cells (Kakuma et al., 1995). It is possible either that the MutT homolog has eluded discovery because it is essential, because there is a redundant activity or because 8-oxo-dGTP is not a significant mutagen in yeast.

A third mechanism that limits GO-associated mutagenesis in yeast involves the translesion synthesis (TLS) polymerase Polη (eta), which is a member of the Y family of DNA polymerases and is encoded by the *RAD30* gene (Haracska et al., 2000; Yuan et al., 2000). Y family polymerases have a large active-site pocket that can accommodate structurally deformed bases, enabling them to insert a nucleotide opposite a lesion (Ling et al., 2001). Not only is such lesion bypass potentially error-prone, the larger active-site pocket of TLS polymerases imparts very low fidelity when copying undamaged DNA. Polη, for example, is error-prone when bypassing some lesions, such as abasic sites (Haracska et al., 2001b), but has relatively high fidelity when bypassing GO lesions,

usually inserting a cytosine across from the lesion (Haracska et al., 2000; Yuan et al., 2000). At GO lesions, Polη is 10-fold more accurate and efficient than Polδ (McCulloch et al., 2009). When given an undamaged DNA template, however, the base substitution error frequency of Polη *in vitro* is 3 orders of magnitude greater than that of a typical replicative polymerase (McCulloch et al., 2007). In addition to Polη, there are two other TLS polymerases in *S. cerevisiae* (Polζ and Rev1), but neither has been implicated in the bypass of GO lesions (de Padula et al., 2004; Sakamoto et al., 2007).

The most straightforward way for Polη to be involved in GO bypass would be for it to be recruited when a replicative polymerase stalls or leaves behind a gap. The replication-blocking potential of GO lesions, however, is unclear. Some *in vitro* studies have shown that replicative DNA polymerases stall or pause when encountering a template GO lesion (Haracska et al., 2000; Sabouri et al., 2008), but other studies have suggested that this is not the case (Shibutani et al., 1991). The currently-accepted model is that Polη is specifically recruited to fill the gaps generated by MMR when it initiates correction of GO:A mispairs (Haracska et al., 2000; van der Kemp et al., 2009b). This model of MMR-Polη cooperation in preventing GO-associated mutagenesis is based on epistasis analysis performed using the *CAN1* forward mutation assay (Haracska et al., 2000). Although the relationship between *msh2* and *rad30* was concluded to be epistatic, the data are also consistent with an additive relationship and, hence, potentially independent roles of Msh2 and Polη in limiting GO-associated mutagenesis. How and why the MMR pathway might specifically recruit a generally error-prone polymerase to fill the gaps in what is normally an extremely accurate repair process is not obvious.

In the present study, a *SUP4-o* forward mutation system was used to re-examine the relationship between MMR and Polη in preventing GO-associated mutagenesis in yeast. To enhance the accumulation of GO lesions, all experiments were conducted in mutants defective in removing GO from GO:C base pairs (an *ogg1* background). In addition, both *msh2* and *msh6* mutants were analyzed. In a *msh6* background, loss of the MutSα heterodimer eliminates the correction of GO:A mispairs, while retention of MutSβ (a heterodimer of Msh2 and Msh3) allows continued correction of most insertion-deletion loops. Finally, mutation spectra as well as mutation rates were considered in order to focus specifically on GC > TA mutations. The results reported here demonstrate that Polη can function independently of MMR to prevent GO-associated mutagenesis, presumably through its ability to bypass these lesions in an error-free manner. Data further indicate that the Polη-dependent bypass of GO lesions is more efficient on the lagging strand of replication and that it requires an interaction with proliferating cell nuclear antigen (PCNA). Interestingly, Polη bypass of GO lesions does not appear to require either ubiquitination of PCNA or the alternative checkpoint clamp.

2.3 Materials and Methods

Strain Constructions

All strains were derived from strain SJR576 (*MATa ura3ΔNco lys2-1$_{oc}$ can1-100$_{oc}$ ade2-1$_{oc}$ leu2-K*). To insert *SUP4-o* into the *HBN1* locus in the FORWARD orientation (*hbn1::SUP4-oF* allele), the following primers were used to amplify the allele from plasmid JF1754 (Pierce et al., 1987): forward primer 5'GGGAATGCAGCTGCGT ACGCTGGGAAGTCAGCCTTTAGCTTTTCAGTTACCTTGGGATCCGGGACCGG

A TAATT and reverse primer 5' GGCTATAGAAAGCCCTGCCGGTCAAAAGAGG CCTGC TTCAGCAAGGGATGAGGCCAATTCTTGAAAGAAATATTTC. The underlined portion of the sequence corresponds to sequence flanking *SUP4-o* and the non-underlined portion corresponds to the *HBN1* locus. *SUP4-o* was amplified and inserted in the REVERSE orientation at the *HBN1* locus (*hbn1::SUP4-oR* allele) using forward primer 5' GGGAATGCAGCTGCGTACGCTGGGAAGTCAGCCTTTAGCTT TTCAGTTACCTTGAATTCTTGAAAGAAATATTTC and reverse primer 5' GGCTA TAGAAAGCCCTGCCGGTCAAAAGAGGCCTGCTTCAGCAAGGGATGAGGCCG GATCCGGGACCGGATAATT. *SUP4-o* was inserted into the *AGP1* locus in the FORWARD orientation (*agp1::SUP4-oF* allele) using the forward primer 5' GCTTGATTAATTCTTCATCAAAGATTTGTCTATGAGAATCTAGGTCGAT CTTGTCGGATCCG GGACCGGATAATT and reverse primer 5' GGTCGGTAA CGGTACCGCGTTGGTTCATGCGGGTCCAGCTGGACTACTTATTAATTCT TGAAAGAAATATTTC. Finally, *SUP4-o* was inserted into the *AGP1* locus in the REVERSE orientation (*agp1::SUP4-oR* allele) using the forward primer 5' GCTTGATTAATTCTTCATCAAAGATTTGTCTATGAGAATCTAGGTCGATCT TGTCAATTCTTGAAAGAAATA TTTC and reverse primer 5' GGTCGGT AACGGTACCGCGTTGGTTCATGCGGGTCCAGCTGGACTACTTATT GGATCCGGGACCGGATAATT. Following transformation of SJR576 with the appropriate PCR product, Lys^+ colonies were selected, and the presence of *SUP4-o* was inferred by co-suppression of the *ade2-101* and *can1-100* ochre alleles. Insertion of *SUP4-o* at the correct location was confirmed by sequencing.

The *OGG1*, *RAD30*, *MSH2*, *MSH6*, *RAD17*, *RAD14*, and *MMS2* genes were deleted by transforming strains with a PCR-generated fragment containing a kanamycin resistance (Bahler et al., 1998), *URA3-Kl* (*URA3* gene from *Kluveromyces lactis*) (Gueldener et al., 2002) or hygromycin resistance marker (Goldstein and McCusker, 1999) with the appropriate flanking sequence of the target gene. *ogg1Δ::kan*, *msh2Δ::hyg*, *msh6Δ::hyg*, *rad17Δ::hyg*, and *rad14Δ::hyg* transformants were selected on YEPD medium (1% yeast extract, 2% Bacto-peptone, 2% dextrose, and 250 mg/L adenine) containing 200 μg/mL geneticin or 300 μg/mL hygromycin. *rad30Δ::URA3-Kl* and *mms2Δ::URA3-Kl* transformants were selected on synthetic complete medium containing 2% dextrose and lacking uracil (SCD-URA). Deletions were confirmed by PCR.

The *rad30(1-624)* allele (Haracska et al., 2001a), lacking the last eight amino acids of the Rad30 protein, was introduced using the *delitto perfetto* method (Storici et al., 2001) as previously described (Abdulovic et al., 2007). Transformants were confirmed by sequencing. A *POL30*-containing plasmid was obtained from Peter Burgers (Bauer and Burgers, 1990). The *POL30* allele was cut out and ligated adjacent to a *LEU2* selectable marker in a lab plasmid. The *pol30(K164R)* allele, which encodes an allele of PCNA that lacks the lysine required for ubiquitination, was generated by performing site-directed mutagenesis. The primers used for site directed mutagenesis were 5'GTGATTCTATTAATATCATGATCACCAGGGAAACAATAAAGTTTG TAGCTGACG 3' and 5'CGTCAGCTACAAACTTTATTGTTTCCCTGGTGATCA TGATATTAATAGAATCAC 3'. The underlined bases indicate where mutations were inserted, changing lysine to arginine. The *pol30(K164R)*-containing plasmid was

digested within the *POL30* allele and transformed into the appropriate yeast strains. This construct will recombine with the endogenous copy of *POL30,* and the adjacent *LEU2* selectable marker will be integrated into the genome at the same time. Transformants were selected on media lacking leucine (SCD-LEU) were sequenced to verify that the *pol30(K164R)* allele had replaced the endogenous copy of *POL30*.

Mutation Rate Analysis

To determine mutation rates, 4-5 individual colonies were used to inoculate a 5 mL starter culture. Following overnight growth at 30°C, the starter culture was used to inoculate independent 5 mL secondary cultures to a concentration of $2.5x10^5$ cells/mL. Two isolates were used for each strain, and at least 12 cultures were used for each isolate. These cultures were grown for 3 days at 30°C. Non-selective YEPGE medium (1% yeast extract, 2% Bacto-peptone, 2% glycerol, 2% ethanol, and 250 mg/L adenine) was used for both starter and secondary cultures. Appropriate dilutions of each culture were plated onto YEPGE medium to determine total cell number and on SCD-ARG plates supplemented with 60 μg/mL L-canavanine (Sigma) to select canavanine-resistant (Can^R) colonies. Colonies were designated as *SUP4-o* mutants if they were both resistant to canavanine and red (Ade^-), indicating loss of suppression of both the *can1-100* and *ade2-1* alleles. Mutation rates were determined using at least 24 cultures and the method of the median (Lea and Coulson, 1949), and 95% confidence intervals were calculated as previously described (Spell and Jinks-Robertson, 2004). Either comparison of the confidence intervals or the Mann-Whitney test (http://faculty.vassar.edu/lowry/utest.html) was used to determine whether two rates were significantly different. Mutation rates for specific mutation types were calculated by

multiplying the proportion of that event in the corresponding spectrum by the total mutation rate.

Mutation Spectra

To generate mutation spectra, DNA was extracted from purified Can^R, red colonies (http://jinks-robertsonlab.duhs.duke.edu/protocols/yeast_prep.html). The *SUP4-o* gene was amplified by PCR and sequenced using the *HBN1* sequencing primer (5' CCGCTTTCAACTCCCAGCC) or the *AGP1* sequencing primer (5' GGGTTATTGGTC GGTAACGG), as appropriate. Sequencing was performed by either the High-Throughput Genomics Unit (Seattle, Washington) or the Duke University DNA Analysis Facility (Durham, North Carolina). Proportions of leading- and lagging-strand mutations were analyzed using Chi Square analysis (http://faculty.vassar.edu/lowry/newcs.html). A p-value less than 0.05 was considered statistically significant.

2.4 Results

In the experiments reported here, the *SUP4-o* forward mutation system originally characterized by Pierce et al. (1987) was used to examine the roles of MMR and Polη in preventing GO-associated mutagenesis. *SUP4-o* is a mutant tRNA that suppresses ochre stop codons by inserting a tyrosine, and two ochre alleles were used to monitor *SUP4-o* function: *ade2-1* and *can1-100*. In the presence of *SUP4-o*, cells are phenotypically Ade^+ (white) and Can^S (canavanine sensitive). Mutations that inactivate *SUP4-o* can be identified by simultaneous loss of suppression of the *ade2-1* and *can1-100* alleles, resulting in Ade^- (red), canavanine-resistant (Can^R) colonies. *SUP4-o* is an ideal reporter to use for studying GO-associated mutagenesis because a mutation at virtually any site

disrupts *SUP4-o* function and allows phenotypic detection; the GC content is 51%, which is higher than the average GC content of the yeast genome and makes detection of GC > TA mutations more efficient; and the small size of the gene (89 bp) enables easy sequencing and determination of mutation patterns. It should be noted that the *SUP4-o* sequence reported throughout is that of the transcribed (non-coding) strand, which follows the convention established by Kunz and coworkers (Pierce et al., 1987).

Polη acts independently of MMR to limit GO-associated mutagenesis

The model that Polη and MMR cooperate to prevent GO-associated mutagenesis was based on epistasis analysis between *msh2* and *rad30* alleles in an *ogg1* background. In the *CAN1* forward mutation assay, Can^R rates were elevated 11-fold, 21-fold, and 27-fold in *ogg1 rad30*, *ogg1 msh2*, and *ogg1 rad30 msh2* mutants, respectively, relative to wildtype (WT) (Haracska et al., 2000). Although it was concluded that *msh2* and *rad30* alleles are epistatic, the use of mutation rate data to distinguish additive versus epistatic interactions becomes especially problematic when one rate is larger than the other. An additional issue with the analysis was that only total mutation rates, rather than the rates of GO-associated GC > TA mutations, were considered. The proportion of mutation spectra that are GC > TA transversions can vary dramatically in different genetic backgrounds (Ni et al., 1999), making total mutation rates a poor indicator of GO-specific GC >TA rates. Because of the inherent uncertainties associated with the earlier epistasis analysis, we re-examined the relationship between Polη and Msh2, focusing specifically on GO-associated GC >TA mutations in the *SUP4-o* reporter.

For initial analyses, we inserted *SUP4-o* into the nonessential *HBN1* locus (the *hbn1::SUP4-oF* allele), which is closely linked to the well-characterized *ARS306* origin

of replication on chromosome III. In an *ogg1* background, GO lesions persist in the genome, leading to GO:A mispairs during replication and increasing the relative contribution of GC > TA transversions to mutation spectra. Although the *hbn1::SUP4-oF* mutation rate was not significantly elevated in the *ogg1* strain, the proportion of GC > TA mutations increased from 27% to 67% (Table 2.1 and Figure 2.1). Taking the proportional increase in GC >TA transversions into account, the rate of these mutations was 3.3-fold higher in the *ogg1* mutant than in WT. Although Polη loss had no effect on mutagenesis when Ogg1 was present, the rate of GC > TA transversions in the double mutant was 10-fold higher than in WT and 3.1-fold higher than in the *ogg1* single mutant (Table 2.1). Thus, as reported in other studies (de Padula et al., 2004; Haracska et al., 2000), there was a synergistic effect of simultaneously removing Ogg1 and Polη, confirming that these proteins act in separate pathways to limit GO-associated mutations.

A similar synergistic interaction was evident when epistasis between *ogg1* and *msh2* was examined. Relative to WT, the total *SUP4-o* mutation rate was elevated 1.3-fold in the *ogg1* single mutant, 4.5-fold in the *msh2* single mutant, and 19-fold in the *ogg1 msh2* double mutant. The synergism was more striking when only the rates of GC >TA mutations were considered, with 3.3-fold, 4.7-fold, and 42-fold increases in the *ogg1* single, *msh2* single, and *ogg1 msh2* double mutants, respectively. As observed in the *CAN1* assay, the *SUP4-o* mutation rate in the *ogg1 rad30 msh2* triple mutant was slightly higher than in the *ogg1 msh2* mutant (5.5×10^{-6} and 3.0×10^{-6}, respectively), but not significantly so when confidence intervals were considered. When mutation spectra (Figure 2.1) were used to calculate rates of GO-associated GC >TA transversions, however, the effect of combining *msh2* and *rad30* was clearly more than additive, with

the GC >TA rate in the *ogg1 rad30 msh2* triple mutant being more than twice that in the *ogg1 msh2* double mutant (Table 2.1). The genetic data obtained here with the *hbn1::SUP4-oF* allele thus do not support a model in which Polη bypass occurs only downstream of Msh2-dependent GO:A mismatch removal. We suggest instead that Polη can directly bypass replication-blocking GO lesions in an error-free manner.

Table 2.1 Mutation Rates of *hbn1::SUP4-oF* Strains

Genotype	*SUP4-o* Mutation Rate[a]		GC > TA Mutation Rate		
	Total ($x10^{-7}$)	Relative to WT	Proportion	Rate ($x10^{-7}$)	Relative to WT
WT	1.6 (0.95-2.1)	1.0	43/160 (27%)	0.43	1.0
ogg1	2.1 (1.7-3.1)	1.3	175/262 (67%)	1.4	3.3
rad30	0.88 (0.52-2.0)	0.6	33/93 (35%)	0.31	0.72
msh2	7.2 (5.4-9.7)	4.5	31/111 (28%)	2.0	4.7
msh6	3.6 (2.3-8.2)	2.3	57/126 (45%)	1.6	3.7
ogg1 rad30	5.4 (3.6-11)	3.4	197/243 (81%)	4.4	10
ogg1 msh2	30 (22-41)	19	75/122 (61%)	18	42
ogg1 msh6	15 (11-19)	9.4	229/303 (76%)	11	26
ogg1 rad30 msh2	55 (33-87)	34	68/87 (78%)	43	100
ogg1 rad30 msh6	44 (30-64)	28	184/210 (88%)	38	88

[a] 95% confidence intervals are indicated in parentheses.

To further substantiate the MMR-independent role of Polη in preventing GO-associated mutagenesis, we also examined the epistatic interaction between *rad30* and *msh6* in an *ogg1* background. Given the greater proportion of base substitutions in *msh6*

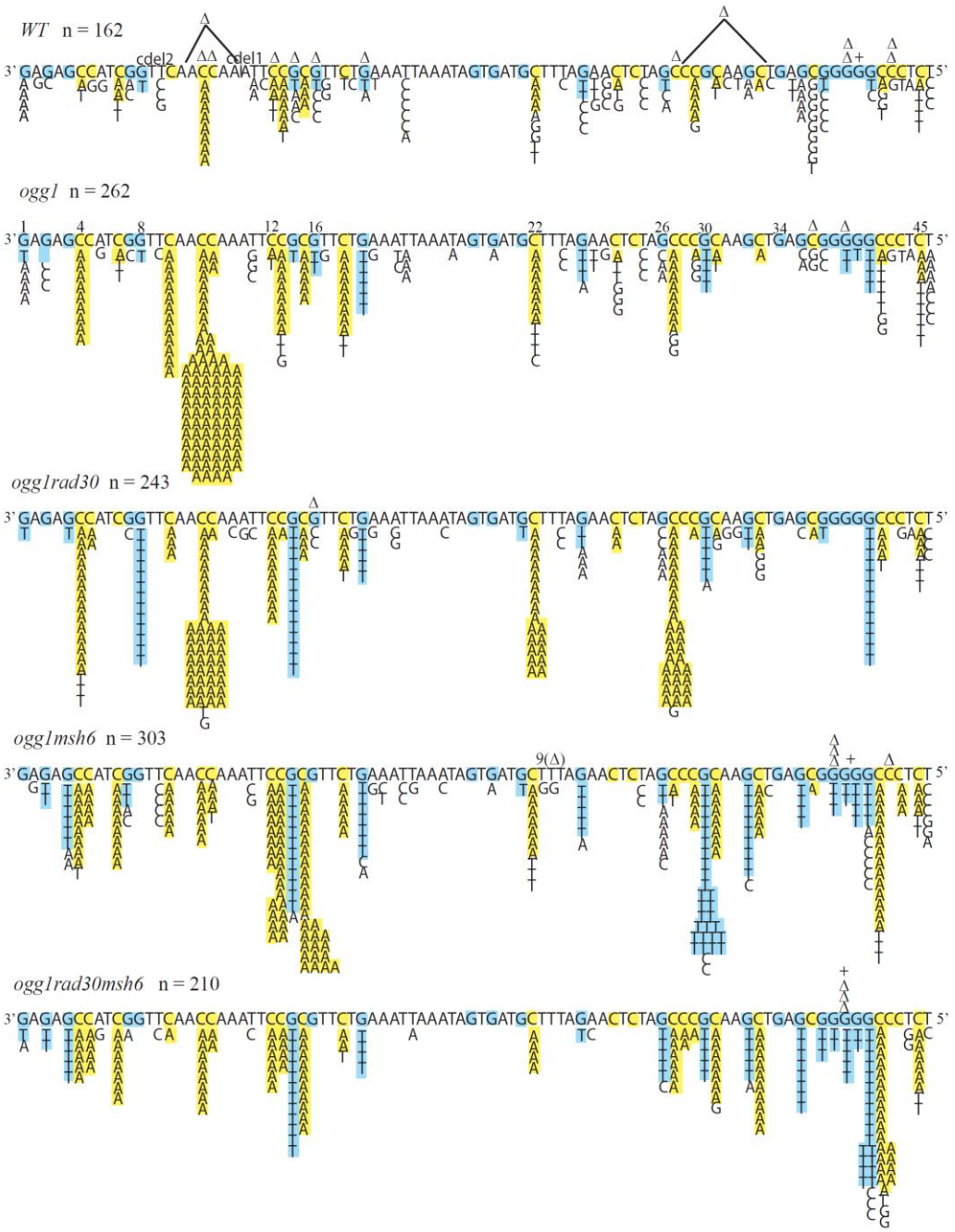

Figure 2.1 Mutation Spectra of *hbn1::SUP4-oF* Strains

The transcribed sequence of *SUP4-o* is shown in the 3’ to 5’ orientation from left to right. Base substitutions are indicated below the sequence of *SUP4-o*. Guanines and G > T transversions are shaded blue, and cytosines and C > A transversions are shaded yellow. Numbers above the sequence are given to indicate the relative position of the GC base pair, e.g., “4” indicates the fourth GC base pair in the sequence. + indicates an insertion event, Δ indicates a deletion event and cdel indicates a complex deletion, in which a deletion is associated with a base substitution event.

spectra than in *msh2* spectra (Ni et al., 1999), we speculated that the relationship between Polη and MMR might be more evident in *msh6* mutants. Loss of Msh6 led to 2.3-fold increase in the overall forward mutation rate of the *hbn1::SUP4-oF* allele, and simultaneous loss of Msh6 and Ogg1 resulted in a 9.4-fold increase in the overall mutation rate (Table 2.1). The *ogg1 rad30 msh6* triple mutant exhibited a 28-fold elevation in the overall *SUP4-o* mutation rate, which is clearly a synergistic increase relative to the 3.4- and 9.4-fold increases in the *ogg1 rad30* and *ogg1 msh6* double mutants, respectively (Table 2.1). The synergism was again much more striking when only GC > TA transversions were considered, with 10-, 26-, and 88-fold rate increases for the *ogg1 rad30* double, *ogg1 msh6* double, and *ogg1 msh6 rad30* triple mutant strains, respectively, relative to WT. In all subsequent experiments, *msh6* mutants were used as the MMR-defective background for analysis of GO-associated GC >TA transversions.

Roles of MMR and Polη in preventing GO-associated mutations at another genomic site

To exclude a site-specific anomaly in our data, we inserted *SUP4-o* into the *AGP1* locus, which positions the allele on the other side of *ARS306*, approximately 2.4 kb away from the *HBN1* locus. The analogous *ogg1*, *ogg1 rad30*, *ogg1 msh6*, and *ogg1 rad30 msh6* mutant strains containing the *agp1::SUP4-oF* allele were constructed and analyzed. Neither the overall mutation rates nor the proportions of GC > TA mutations were different from those in the equivalent strains with the *hbn1::SUP4-oF* allele (Table 2.2 and Figure 2.2). Because the WT mutation rate was slightly lower in the *agp1::SUP4-oF* strain, however, the rate increases in the mutant strains were even more dramatic than

those obtained in the analogous *hbn1::SUP4-oF* strains. Relative to the WT strain, the GC > TA mutation rate within the *agp1::SUP4-oF* allele increased 21-fold in the *ogg1 rad30* strain, 49-fold in the *ogg1 msh6* strain, and 241-fold in the *ogg1 rad30 msh6* strain.

Table 2.2 Mutation Rates of *agp1::SUP4-oF* Strains

	SUP4-o Mutation Rate[a]		GC > TA Mutation Rate		
Genotype	Total ($x10^{-7}$)	Relative to WT	Proportion	Rate ($x10^{-7}$)	Relative to WT
WT	0.76 (0.44-1.2)	1.0	36/161 (22%)	0.17	1.0
ogg1	1.4 (1.1-1.7)	1.8	96/152 (63%)	0.89	5.2
ogg1 rad30	4.1 (2.5-6.4)	5.4	189/219 (86%)	3.5	21
ogg1 msh6	11 (9.2-16)	14	167/226 (74%)	8.4	49
ogg1 rad30 msh6	45 (28-69)	59	148/164 (90%)	41	241

[a] 95% confidence intervals are in parentheses.

Requirements of Polη bypass of GO lesions

Polη contains a short, PCNA-interacting peptide (PIP) domain at its C-terminal end, and studies suggest that its bypass activity requires an interaction with PCNA (Haracska et al., 2001a; van der Kemp et al., 2009b). To determine if this interaction is required for the Polη-dependent bypass of GO lesions, we inserted the *rad30(1-624)* allele, which lacks the last eight amino acids of Rad30 and thereby removes the PIP domain, into the *ogg1* and the *ogg1 msh6* strains containing the *hbn1::SUP4-oF* allele. If an interaction with PCNA is required for Polη bypass of GO lesions, the *rad30(1-624)* allele should produce a phenotype similar to the *rad30* null allele. As shown in Table 2.3 and Figure 3.3, the GC > TA mutation rates were indistinguishable in the strains containing the *rad30Δ* versus the *rad30(1-624)* allele. Polη thus requires an interaction with PCNA for the efficient bypass of GO lesions in the *SUP4-o* system.

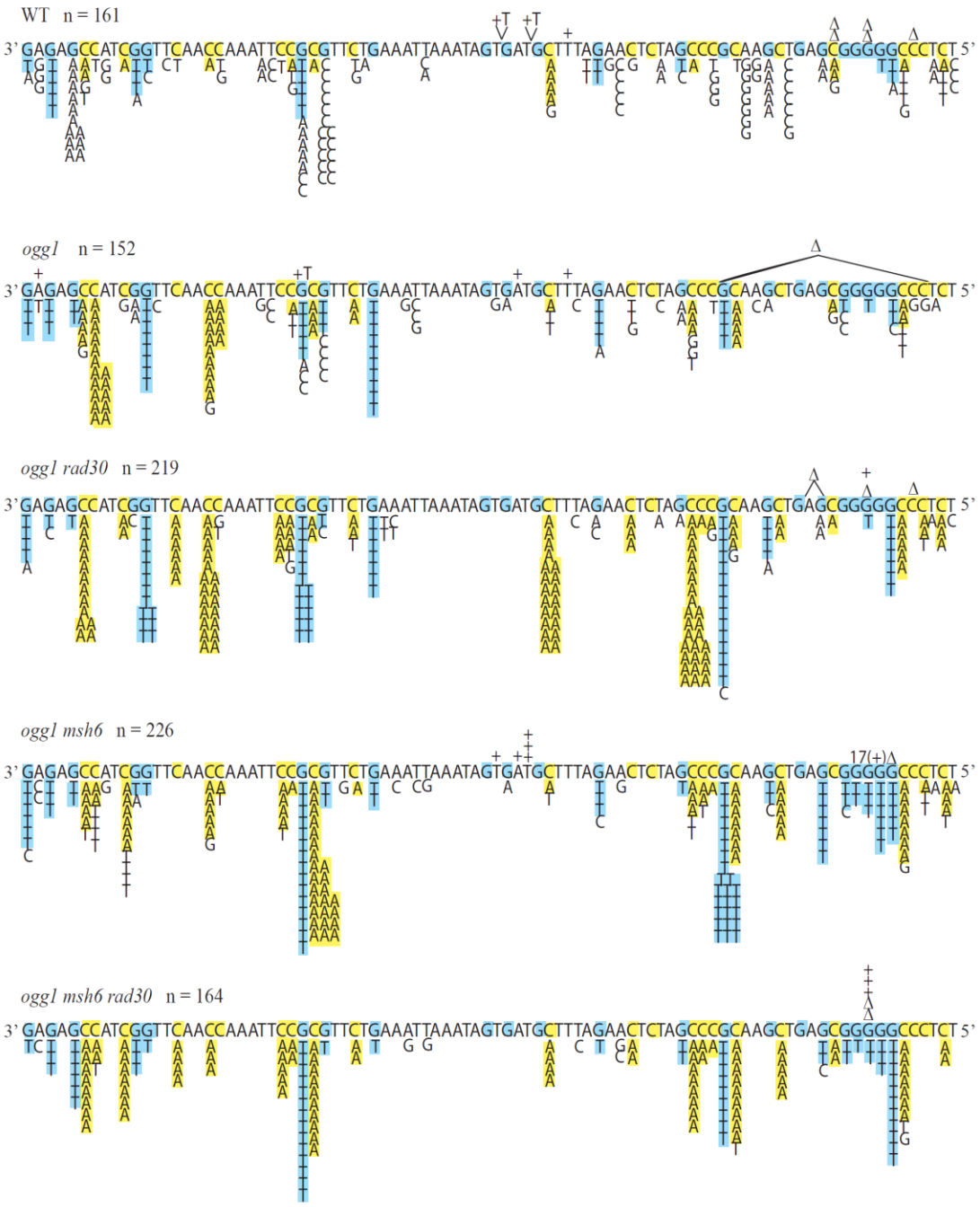

Figure 2.2 Mutation Spectra of *agp1::SUP4-oF* Strains

The transcribed sequence of *SUP4-o* is shown in the 3' to 5' orientation from left to right. Base substitutions are indicated below the sequence of *SUP4-o*. Guanines and G > T transversions are shaded blue, and cytosines and C > A transversions are shaded yellow. + indicates an insertion event, and Δ indicates a deletion event.

Table 2.3 Regulation of Polη Activity

Genotype[a]	*SUP4-o* Mutation Rate[b]		GC > TA Mutation Rate		
	Total ($x10^{-7}$)	Relative to WT	Proportion	Rate ($x10^{-7}$)	Relative to WT
WT	1.6 (0.95-2.1)	1.0	43/160 (27%)	0.43	1.0
ogg1	2.1 (1.7-3.1)	1.3	175/262 (67%)	1.4	3.3
ogg1 rad30	5.4 (3.6-11)	3.4	197/243 (81%)	4.4	10
ogg1 rad30(1-624)	6.1 (5.0-10)	3.8	210/258 (81%)	5.0	12
ogg1 rad30 msh6	44 (30-64)	28	184/210 (88%)	38	88
ogg1 rad30(1-624) msh6	42 (33-70)	26	136/165 (82%)	35	81
ogg1 pol30-K164R	3.5 (2.6-9.7)	2.2	128/202 (63%)	2.0	4.7
ogg1 rad17	3.4 (2.4-5.8)	2.1	74/156 (47%)	1.6	3.7

[a] All strains contain the *hbn1::SUP4-oF* allele.
[b] 95% confidence intervals are in parentheses.

In response to DNA damage, PCNA is ubiquitinated by the Rad6/Rad18 complex (Hoege et al., 2002). Studies have shown that Polη and Polζ bypass require ubiquitination of PCNA for some types of TLS (Garg and Burgers, 2005b; Haracska et al., 2004; Kannouche et al., 2004; Stelter and Ulrich, 2003; van der Kemp et al., 2009a; Zhuang et al., 2008). To determine if ubiquitination of PCNA is important for Polη bypass of GO lesions, we generated an allele of PCNA (encoded by the *POL30* gene) in which the lysine required for ubiquitination is mutated to an arginine. This allele, *pol30(K164R)*, was then transformed into the *hbn1::SUP4-oF ogg1* strain. If ubiquitination of PCNA is required for Polη bypass of GO lesions, strains carrying the *pol30(K164R)* allele should have the same phenotype as *rad30Δ* strains. Interestingly,

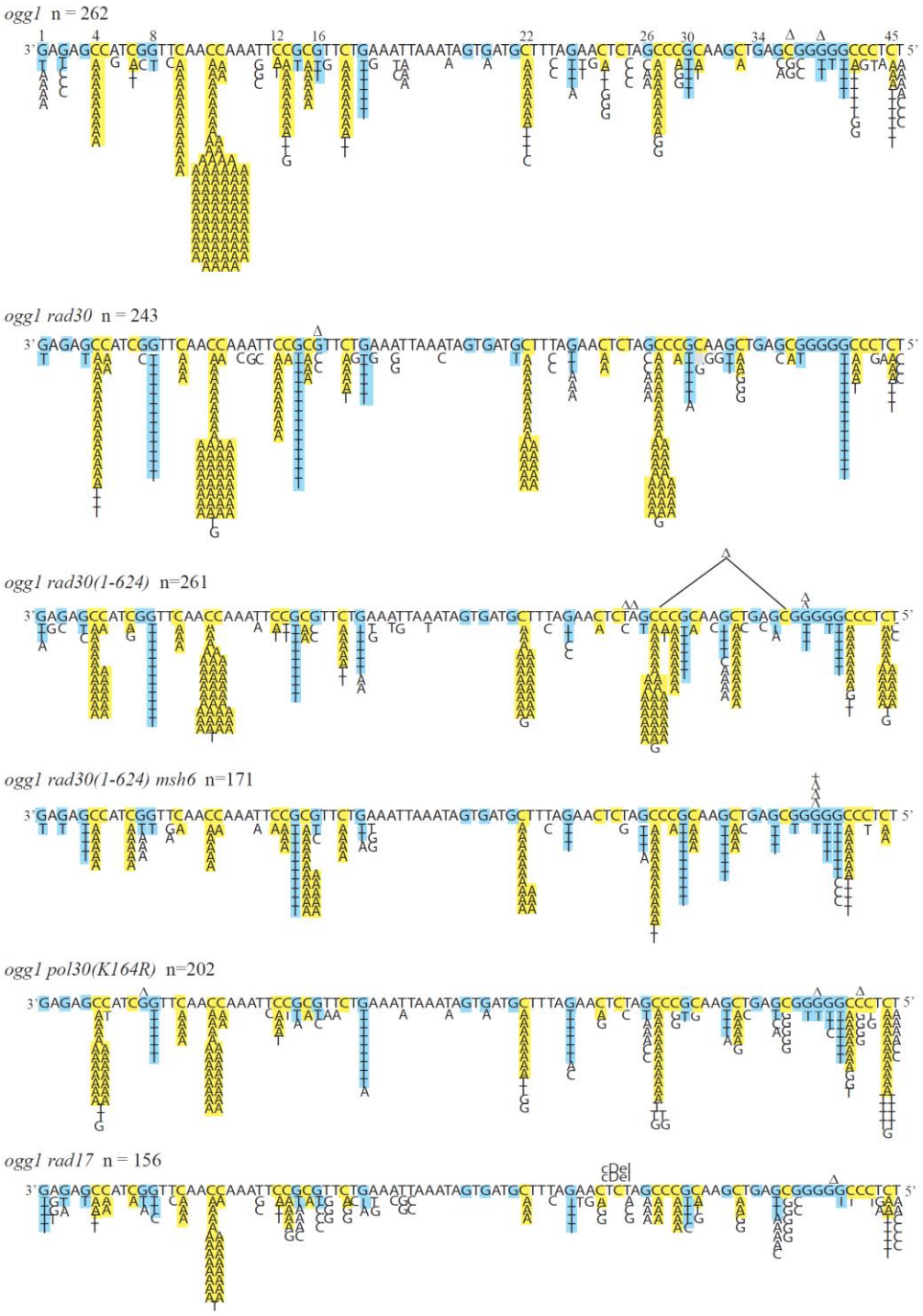

Figure 2.3 Mutation Spectra of *RAD30*, *POL30*, and *RAD17* Strains

The transcribed sequence of *SUP4-o* is shown in the 3' to 5' orientation from left to right. Base substitutions are indicated below the sequence of *SUP4-o*. Guanines and G > T transversions are shaded blue, and cytosines and C > A transversions are shaded yellow. + indicates an insertion event, and Δ indicates a deletion event.

the resulting strain exhibited the same rate of GO-associated mutations as the *ogg1* strain; loss of PCNA ubiquitination did not result in an increase in GO-associated mutagenesis (Table 2.3 and Figure 2.3). This indicates that Polη bypass of GO lesions does not require ubiquitination of PCNA.

Cells contain another, PCNA-like sliding clamp that is thought to help proteins access DNA and participate in the DNA damage checkpoint response. This clamp is composed of the Rad17, Ddc1, and Mec3 checkpoint proteins in yeast and is referred to as the 9-1-1 clamp (Majka and Burgers, 2003). Recent studies in our lab showed that this clamp physically interacts with Polζ and is partially required for Polζ-dependent spontaneous mutagenesis (Sabbioneda et al., 2005). To determine if this alternative clamp also participates in Polη bypass of GO lesions, the *RAD17* allele was disrupted in the *hbn1::SUP4-oF ogg1* strain. If the alternative clamp is involved in Polη bypass of GO lesions, *rad17Δ* strains should phenotypically resemble *rad30Δ* strains. As shown in Table 2.3 and Figure 2.3, deletion of *RAD17* did not have any effect on GO-associated mutagenesis. This suggests that the 9-1-1 clamp is not involved in Polη bypass of GO lesions.

Polη has a lagging-strand bias in GO bypass activity

Because the identity of the initiating lesion is known, GC > TA mutations in *ogg1* mutants can be attributed to GO:A mispairs rather than to C:T mispairs. This property of GO-associated mutagenesis can be used to assign the strand of origin of mutations and, as first described by Pavlov et al. (2002), to compare leading- and lagging-strand mutagenesis if the direction of replication is known. This approach was used previously to demonstrate that more GO-associated mutagenesis occurs during leading-strand

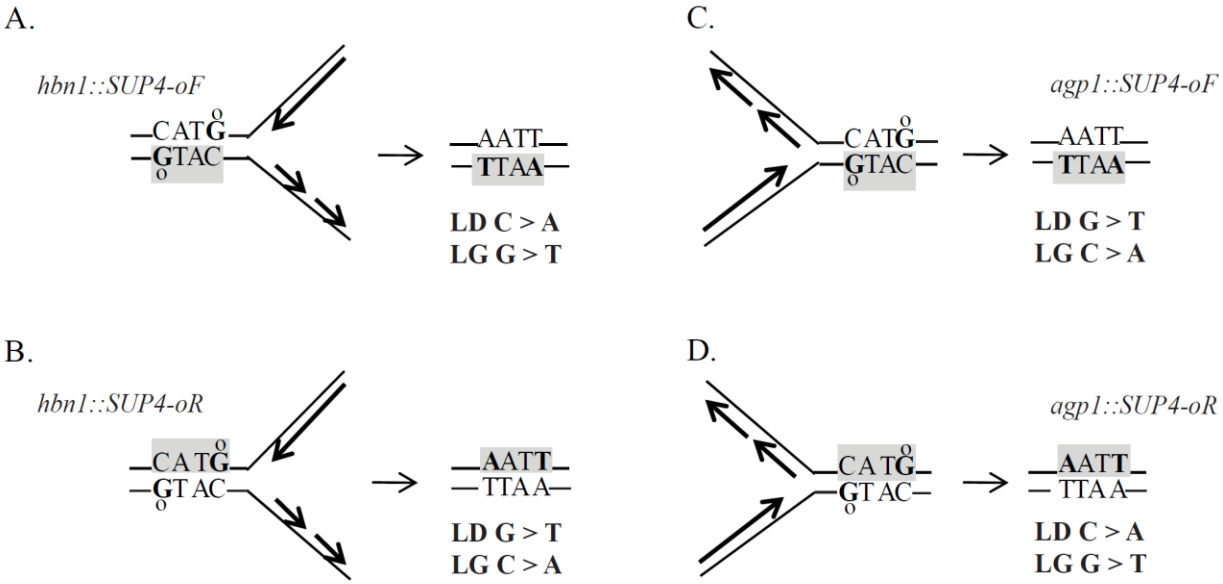

Figure 2.4 Orientations of the *SUP4-o* Alleles Relative to *ARS306*

(A and B) In the *hbn1::SUP4-o* strains, *SUP4-o* is on the left side of *ARS306*, with the replication fork moving from right to left. (A) In the *hbn1::SUP4-oF* allele, the transcribed strand is the lagging-strand template and is indicated by the gray box. Sequencing of the transcribed strand in *hbn1::SUP4-oF* mutants uncovers lagging-strand G > T mutations and leading-strand C > A mutations (indicated in bold). (B) In the *hbn1::SUP4-oR* allele, the transcribed strand is on the opposite DNA strand and is the leading-strand template. Sequencing of the transcribed strand in *hbn1:SUP4-oR* mutants uncovers leading-strand G > T mutations and lagging-strand C > A mutations. (C and D) In the *agp1::SUP4-o* strains, *SUP4-o* is on the right side of *ARS306*, with the replication fork moving from left to right. (C) In the *agp1::SUP4-oF* allele, the transcribed strand is the leading-strand template. Sequencing of the transcribed strand in *agp1::SUP4-oF* mutants uncovers leading-strand G > T mutations and lagging-strand C > A mutations. (D) In the *agp1:SUP4-oR* allele, the transcribed strand is on the opposite strand and is the lagging-strand template. Sequencing of the transcribed strand in *agp1::SUP4-oR* mutants uncovers lagging-strand G > T mutations and leading-strand C > A mutations. LD indicates leading-strand mutations and LG indicates lagging-strand mutations.

synthesis (Pavlov et al., 2002) and that most of this bias results from more efficient MMR during lagging-strand synthesis (Pavlov et al., 2003). As illustrated in Figure 2.4A, the transcribed strand is the lagging-strand template in the *hbn1::SUP4-oF* strain, meaning that G > T and C > A mutations arise from GO lesions on the lagging- and leading-strand templates, respectively. In the mutation spectrum of the *hbn1::SUP4-oF ogg1* strain, there were many more C > A than G > T mutations (57% and 9.9%, respectively; Figure 2.1), consistent with most GO-associated mutations being generated during leading-strand synthesis. Also in agreement with earlier studies (Pavlov et al., 2003), this bias largely disappeared in the *ogg1 msh6* double mutant, where the numbers of C > A and G > T mutations were more similar (46% and 30%, respectively).

Although an estimate of the ratio of leading- to lagging-strand errors can be obtained by comparing the numbers of G > T and C >A mutations in a given spectrum, a more accurate method is to examine exactly the same sequence when present on each of the two strands. The orientation of the *SUP4-o* reporter within the *HBN1* locus was thus reversed to generate the *hbn1::SUP4-oR* allele, in which the transcribed strand is switched from the lagging- to the leading-strand template during replication (Figure 2.4B). As done with the *hbn1::SUP4-oF* allele, mutation rates and spectra were generated for *ogg1*, *ogg1 rad30*, *ogg1 msh6*, and *ogg1 rad30 msh6* strains containing the *hbn1::SUP4-oR* allele (Table 2.4 and Figure 2.5). Because the rate of *SUP4-o* mutations in a given strain background was independent of gene orientation, the proportions of G > T (or C > A) mutations that arose during leading-strand synthesis were directly compared to those generated during lagging-strand synthesis.

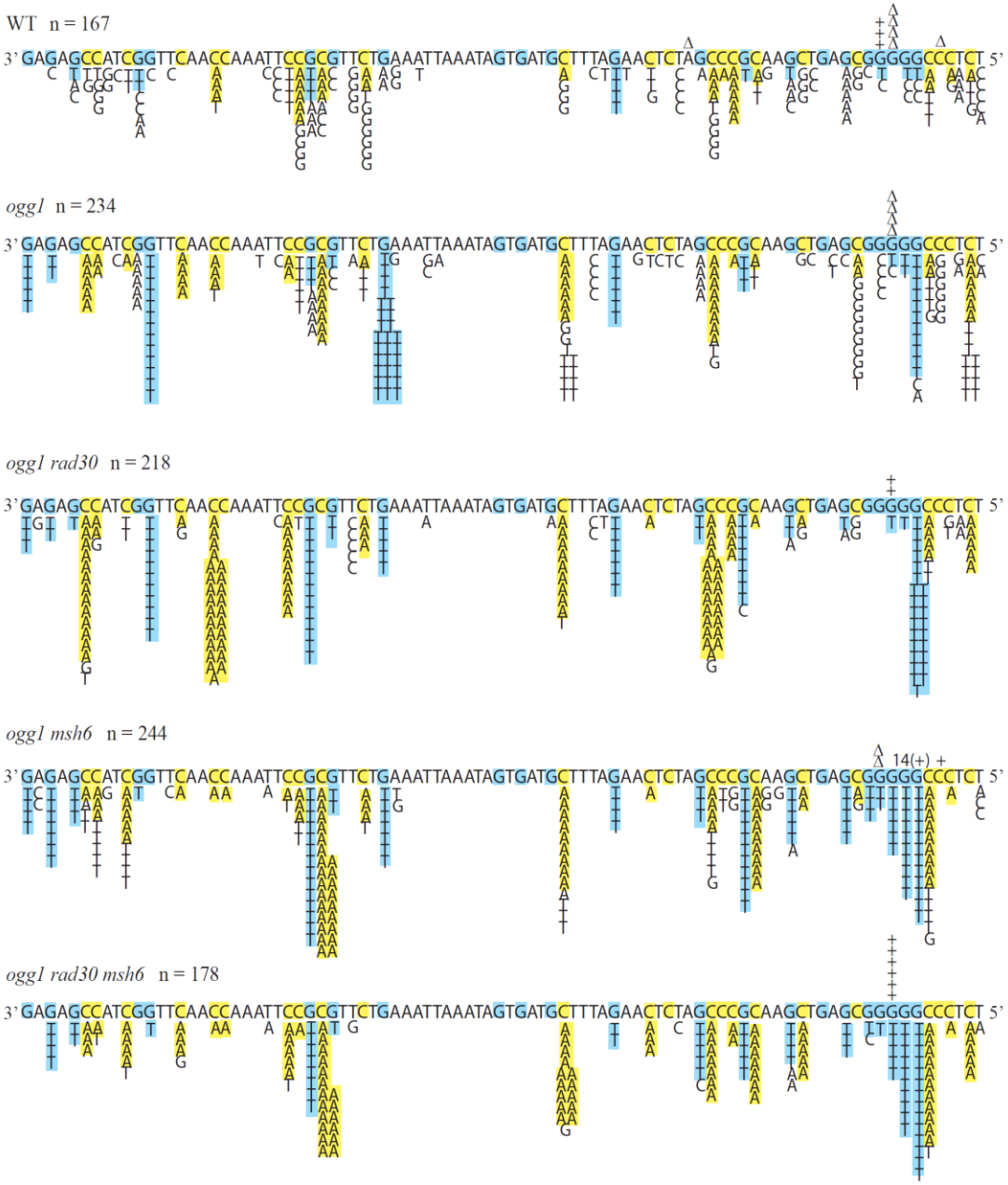

Figure 2.5 Mutation Spectra of *hbn1::SUP4-oR* Strains

The transcribed sequence of *SUP4-o* is shown in the 3' to 5' orientation from left to right. Base substitutions are indicated below the sequence of *SUP4-o*. Guanines and G > T transversions are shaded blue, and cytosines and C > A transversions are shaded yellow. + indicates an insertion event, and Δ indicates a deletion event.

Table 2.4 Mutation Rates of *hbn1:SUP4-oR* and *agp1::SUP4-oR* Strains

	hbn1::SUP4-oR	*agp1::SUP4-oR*
Genotype	Rate ($x10^{-7}$)[a]	Rate ($x10^{-7}$)[a]
ogg1	1.6 (1.2-2.2)	1.6 (1.3-3.2)
ogg1 rad30	4.4 (1.8-9.7)	3.0 (2.2-4.2)
ogg1 msh6	14 (9.2-24)	16 (8.9-20)
ogg1 rad30 msh6	31 (19-53)	44 (24-64)

[a] 95% confidence intervals are in parentheses.

As shown in Table 2.5, G > T leading-strand mutations accounted for 32% of the *ogg1* spectrum while G > T lagging-strand mutations comprised only 9.9% of the spectrum. There is thus an approximately 3-fold leading-strand bias for GO-associated mutations ($p < 0.0001$). In the *ogg1 msh6* strains, the proportion of both leading- and lagging-strand G > T mutations increased, but the proportion of lagging-strand mutations increased more than that of leading-strand mutations (9.9% to 30% for lagging-strand mutations and 32% to 41% for leading-strand mutations). Interestingly, a similar pattern was seen in the *ogg1 rad30* strains, with a 9.9% to 24% increase in lagging-strand mutations but only a 32% to 39% increase in leading-strand mutations. The leading-strand bias for G > T mutations was thus greatly reduced by elimination of either Msh6 or Polη, indicating that both processes are more efficient at reducing GO-associated mutagenesis during lagging-strand synthesis. Importantly, however, a significant bias still persisted in each corresponding double mutant. This bias was entirely eliminated in the *ogg1 rad30 msh6* triple mutant, with the proportion of G >T mutations being statistically the same on the two strands ($p = 1$). Similar results were obtained when the proportions of C > A mutations generated during leading- versus lagging-strand

Table 2.5 Leading- and Lagging-strand Mutagenesis in *hbn1::SUP4-o* Strains

	G > T Mutations			C > A Mutations		
	LD (*R*)	LG (*F*)	LD/LG	LD (*F*)	LG (*R*)	LD/LG
Genotype	Proportion (%)	Proportion (%)	Bias	Proportion (%)	Proportion (%)	Bias
ogg1	75/234 (32)	26/262 (9.9)	3.23 (p<0.0001)	149/262 (57)	52/234 (22)	2.59 (p<0.0001)
ogg1 rad30	84/218 (39)	58/243 (24)	1.63 (p=0.001)	139/243 (57)	99/218 (45)	1.27 (p=0.015)
ogg1 msh6	99/244 (41)	90/303 (30)	1.37 (p=0.01)	139/303 (46)	81/244 (33)	1.39 (p=0.004)
ogg1 rad30 msh6	66/178 (37)	79/210 (38)	0.97 (p=1)	105/210 (50)	92/178 (52)	0.96 (p=0.823)

LD indicates leading-strand mutations and LG indicates lagging-strand mutations. *F* indicates *hbn1::SUP4-oF* and *R* indicates *hbn1::SUP4-oR*. The proportions of mutations are given with the percentages (%) in parentheses. The LD/LG bias was calculated by dividing the proportion of LD mutations by the proportion of LG mutations. G > T mutations are on the transcribed strand and C > A mutations are on the nontranscribed strand. Significance was calculated by Chi Square.

replication were analyzed; the strong leading-strand bias in the *ogg1* mutant was reduced by loss of either Msh6 or Polη, but was completely eliminated only when both were absent. These results confirm previous observations that MMR is more efficient at repairing GO:A mismatches that arise during lagging-strand synthesis (Pavlov et al., 2003) and also demonstrate a clear lagging-strand bias for the error-free bypass of GO lesions by Polη. In addition, because some leading-strand bias for mutations persists when either Polη or Msh6 is present, the data indicate that the Polη strand-specific bias is at least partially independent of MMR and vice versa.

To examine whether genome position affects either the efficiency of GO:A mispair removal or error-free GO bypass, a similar analysis was done at the *AGP1* locus on the other side of *ARS306*. At this location, the transcribed strand of *SUP4-o* was the leading-strand template in the *agp1::SUP4-oF* allele and the lagging-strand template in the *agp1::SUP4-oR* allele (Figure 2.4C,D). Mutation rates and spectra for the *agp1::SUP4-oR* strains are presented in Table 2.4 and Figure 2.6, respectively. The comparisons of G > T and C > A mutations when *SUP4-o* was placed at *AGP1* are presented in Table 2.6. As in the *hbn1::SUP4-o ogg1* strains, there was a strong leading-strand bias for both G > T and C > A mutations in the *agp1::SUP4-o ogg1* strains ($p = 0.0001$ and $p = 0.0004$). The effects of Msh6 and Polη loss on this bias, however, were different at the *AGP1* location. While there was only a weak reduction in the leading- to lagging-strand bias when Msh6 was eliminated in the *ogg1* background (from 2.5 to 1.9 for G > T mutations and from 1.6 to 1.3 for C > A mutations), the bias for both G > T and C > A mutations was completely eliminated in the *ogg1 rad30* mutants ($p = 0.823$ and $p = 1$, respectively). Thus, at the *AGP1* position, the greater accumulation of GO-

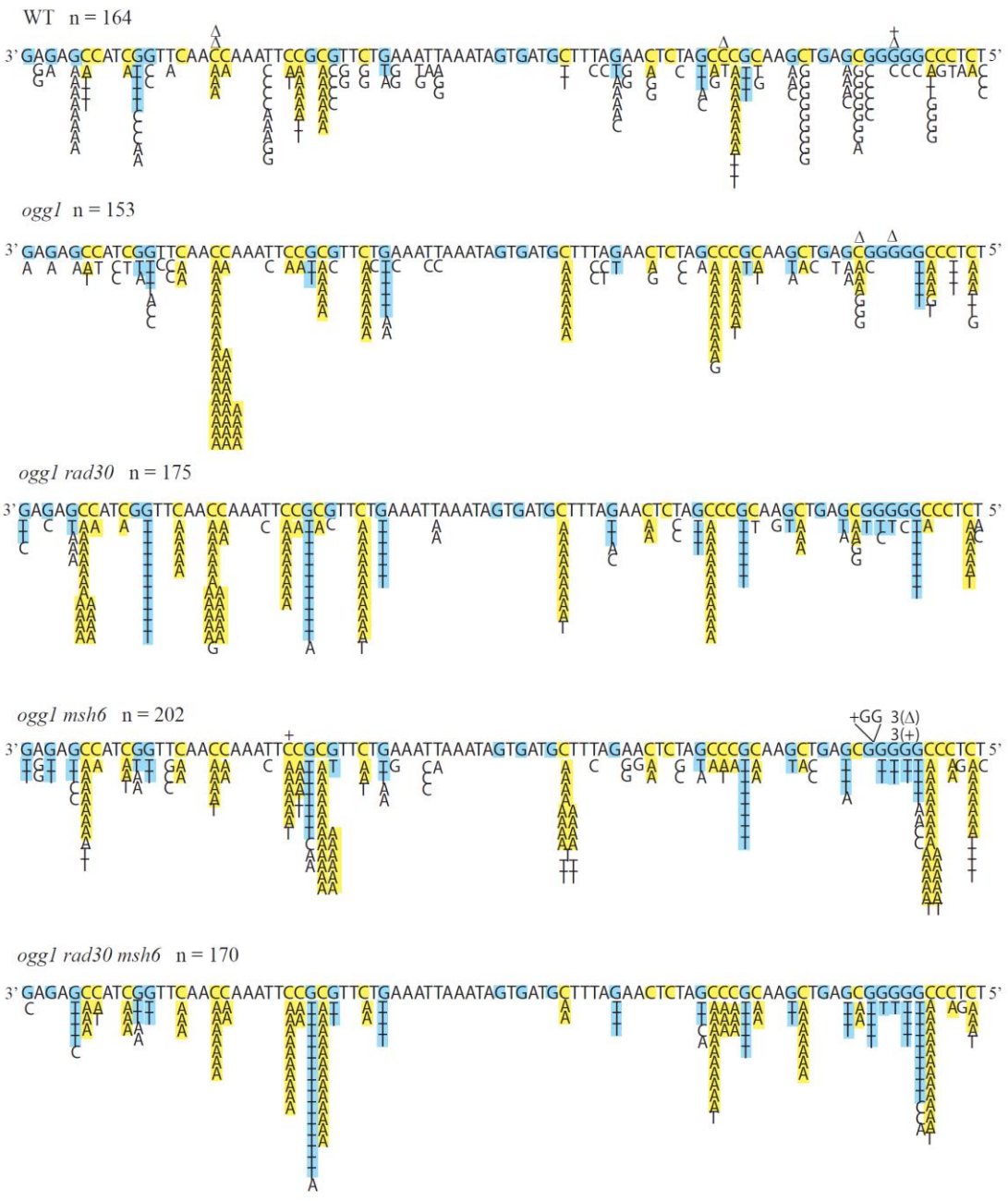

Figure 2.6 Mutation Spectra of *agp1::SUP4-oR* Strains

The transcribed sequence of *SUP4-o* is shown in the 3' to 5' orientation from left to right. Base substitutions are indicated below the sequence of *SUP4-o*. Guanines and G > T transversions are shaded blue, and cytosines and C > A transversions are shaded yellow. + indicates an insertion event, and Δ indicates a deletion event.

Table 2.6 Leading- and Lagging-strand Mutagenesis in *agp1::SUP4-o* Strains

	G > T Mutations			C > A Mutations		
	LD (*F*)	LG (*R*)	LD/LG	LD (*R*)	LG (*F*)	LD/LG
Genotype	Proportion (%)	Proportion (%)	Bias	Proportion (%)	Proportion (%)	Bias
ogg1	46/152 (30)	18/153 (12)	2.50 (p=0.0001)	82/153 (54)	50/152 (33)	1.64 (p=0.0004)
ogg1 rad30	71/219 (32)	54/175 (31)	1.03 (p=0.823)	94/175 (54)	118/219 (54)	1.00 (p=1)
ogg1 msh6	88/226 (39)	42/202 (21)	1.86 (p<0.0001)	95/202 (47)	79/226 (35)	1.34 (p=0.015)
ogg1 rad30 msh6	63/164 (38)	64/170 (38)	1.00 (p=1)	91/170 (54)	85/164 (52)	1.04 (p=0.84)

LD indicates leading-strand mutations and LG indicates lagging-strand mutations. *F* indicates *agp1::SUP4-oF* and *R* indicates *agp1::SUP4-oR*. The LD/LG bias was calculated by dividing the proportion of LD mutations by the proportion of LG mutations. G > T mutations are on the transcribed strand and C > A mutations are on the nontranscribed strand. Significance was calculated by Chi Square analysis.

associated mutations during leading-strand synthesis in the *ogg1* single mutant appears to be completely dependent on the more efficient bypass activity of Polη during lagging-strand synthesis.

It is important to note that the overall strand-related biases calculated in Tables 2.5 and 2.6 represent the cumulative effects at many sites across the *SUP4-o* sequence. When individual sites were analyzed in the *hbn1::SUP4-o ogg1* strains, the ratio of leading- to lagging-strand mutations ranged from 0.43 to 31 (Figure 2.7). Although the small numbers of mutations at most of the individual sites preclude accurate statistical analysis, the huge site-to-site variation in the leading- to lagging-strand bias demonstrates that mutagenesis is not equal at every site and that there is a wide range of variability in the GO-related strand bias for both MMR and Polη. When all sites were summed, the ratio of GC > TA leading-strand mutations to GC > TA lagging-strand mutations was 2.9. In the *ogg1 rad30*, *ogg1 msh6*, and *ogg1 rad30 msh6* strains, the range of the leading- to lagging-strand ratios at individual sites narrowed, and the ratios determined by summing all sites decreased to 1.4, 1.4, and 1.0, respectively.

Analysis of GO-associated mutagenesis at specific sites in the *hbn1::SUP4-o* alleles

Although mutation spectra alone provide information about the distributions of mutations, a more quantitative assessment of site-to-site variation in GC > TA transvervions can be obtained by calculating the mutation rates at individual sites. As described above, both Polη and MMR are more efficient on the lagging strand at the *HBN1* locus (Table 2.5). Therefore, GO lesions on the lagging-strand templates of the *hbn1::SUP4-o* alleles were specifically examined. This corresponds to G > T mutations in the *hbn1::SUP4-oF* allele and C > A mutations in the *hbn1::SUP4-oR* allele. A small

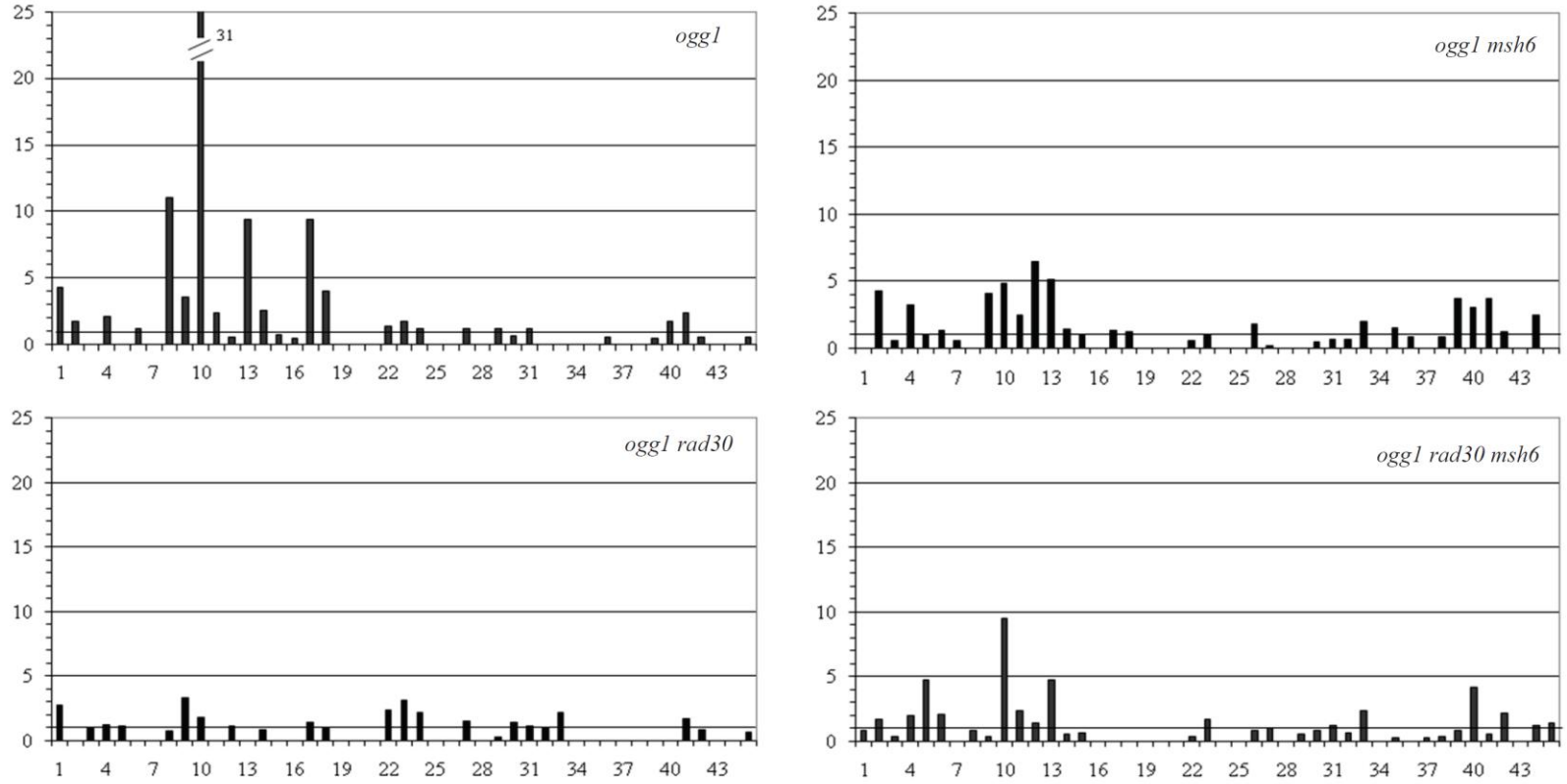

Figure 2.7 Leading- to Lagging-strand Bias at Specific Sites in *hbn1::SUP4-o*

Leading and lagging strand G > T mutations were obtained using the *hbn1::SUP4-oR* and *hbn1::SUP4-oF* alleles, respectively. Leading and lagging strand C > A mutations were obtained using the *hbn1::SUP4-oF* and *hbn1::SUP4-oR* alleles, respectively. At each site, the leading-strand mutation rate was divided by the lagging-strand mutation rate to obtain the bias. Numbers on the horizontal axis represent GC base pairs within *SUP4-o*, and numbers on the vertical axis indicate the ratio of leading- to lagging-strand bias. The dashed line indicates a ratio of one (i.e., no bias)

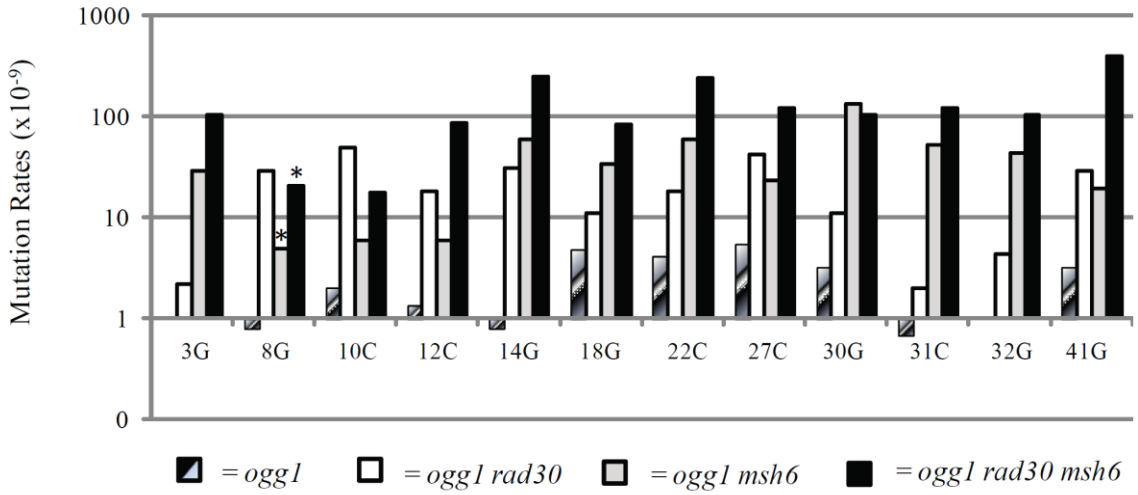

Figure 2.8 Site-specific Analysis of MMR and Pol η Activity in *hbn1::SUP4-o* Strains

Rates of GC > TA transversions on the lagging strand (G > T in the *hbn1::SUP4-oF* strain and C > A in the *hbn1::SUP4-oR* strain) are shown on a log scale. Each site is labeled according to its position in the *SUP4-o* gene, i.e., 3G is the third guanine in the gene. Only those sites where at least five G > T or five C > A mutations were observed in two different strain backgrounds are included. Striped bars: *ogg1*; white bars: *ogg1 rad30*; gray bars: *ogg1 msh6*; black bars: *ogg1 rad30 msh6*. An asterisk (*) indicates the mutation rate was calculated assuming one event, as none were detected at these specific sites.

subset of these sites is shown in Figure 2.8. Analysis of site-specific mutation rates reveals several important points. First, consistent with the overall rate measurements, there was a synergistic effect of removing Rad30 and Msh6 at the majority of sites. At some sites, however, the relative role of MMR or Polη was enhanced or diminished. At sites 8G and 10C, for example, Polη appears to be required for all error-free bypass of GO lesions, with MMR playing a relatively minor role. In contrast, at other sites, deletion of *RAD30* had little, if any, effect on mutation rate (e.g., site 30G), and only the MMR machinery seemed to be important for limiting GO-associated mutagenesis. Interestingly, in an *ogg1 rad30 msh6* background, rates of GO-associated mutations at

specific sites varied more than 20-fold. Even in the absence of repair, there clearly is significant variability in site-to-site mutagenesis.

More GO-associated mutagenesis occurs on the nontranscribed than the transcribed strand of *SUP4-o*

As presented here, G > T mutations always reflect GO lesions on the transcribed strand of the *SUP4-o* gene, and C > A mutations represent mutations generated in response to GO lesions on the nontranscribed strand. Interestingly, the proportion of mutations that were C > A transversions in the triple mutant *ogg1 rad30 msh6* background was greater than the proportion of mutations that were G > T transversions (Tables 2.5 and 2.6). At both the *HBN1* and *AGP1* locations, approximately 50% of mutations were C > A transversions, while only 40% were G > T transversions ($p = 0.0002$ for both sites). The greater abundance of C > A mutations suggests that there is more accumulation of GO lesions on the nontranscribed than on the transcribed strand of *SUP4-o*, regardless of whether the nontranscribed strand is the leading- or lagging-strand template during replication. This could be explained either by more efficient removal of lesions from the transcribed strand, or a greater accumulation of lesions on the nontranscribed strand. In the first scenario, GO lesions on the transcribed strand might block RNA polymerases, resulting in the activation of transcription-coupled nucleotide excision repair (TC-NER) and a concomitant reduction in potential mutagenesis. In the second scenario, the process of transcription might lead to the formation of a DNA:RNA hybrid between the DNA template and the newly-synthesized RNA, leaving the nontranscribed strand transiently single-stranded and more chemically reactive (Korzheva et al., 2000). To address this bias in our system, we generated strains that were deficient

in transcription-coupled nucleotide excision repair by deleting the *RAD14* gene in the *hbn1::SUP4-oF* strain. If the bias is due to TC-NER, the bias should be decreased when nucleotide excision repair pathway is deleted. Instead, we observed that the bias was not decreased in the *rad14Δ* strain (Table 2.7). This suggests that GO lesions are not subject to TC-NER, and likely accumulate more often on the nontranscribed strand (Table 2.7).

Table 2.7 Effect of Transcription-coupled Repair on Transcribed-strand Bias

Genotype	*SUP4-o* Mutation Rate[a]		Mutations on TS and NTS strands[b]				
	Total ($x10^{-7}$)	Relative to *ogg1*	TS ($x10^{-7}$)	Relative to *ogg1*	NTS ($x10^{-7}$)	Relative to *ogg1*	NTS/TS Bias
ogg1	2.1 (1.7-3.1)	1.0	0.21	1.0	1.2	1.0	5.7
ogg1rad14	5.6 (4.4-8.9)	2.7	0.14	0.67	2.1	1.8	15

[a] 95% confidence intervals are in parentheses.
[b] TS indicates transcribed strand and NTS indicates nontranscribed strand.

2.5 Discussion

The mutagenic potential of GO lesions and their relevance to cancer and aging have been well documented (D'Errico et al., 2008; Maynard et al., 2009; Skinner and Turker, 2005). Despite decades of study, however, the various mechanisms that act to prevent GO-associated mutagenesis in eukaryotes have yet to be fully described. In addition to Ogg1, which excises GO lesions from GO:C mispairs, both MMR and Polη reduce the mutagenesis that results from the insertion of adenine opposite a template GO lesion during DNA synthesis (Haracska et al., 2000; Ni et al., 1999; Yuan et al., 2000). Based on epistasis analysis between *msh2* and *rad30*, it was proposed that Polη works exclusively in the context of MMR to fill the gaps generated when the MMR system removes the adenine of GO:A mispairs (de Padula et al., 2004; Haracska et al., 2000; van

der Kemp et al., 2009b). The data presented here demonstrate that the full relationship between MMR and Polη in limiting GO-associated mutagenesis was obscured in earlier studies because *msh2* strains were used for epistasis analysis and because only total mutation rates were considered (Haracska et al., 2000). Using spectra to focus on GO-associated GC > TA mutations, we were able to clearly observe synergism between *msh2* and *rad30*. When *msh6* mutants were used instead of *msh2* mutants, the synergism with *rad30* was evident in total mutation rate measurements and was further exaggerated when GC >TA mutation rates were calculated. Finally, a simple comparison of the *ogg1 msh6* and *ogg1 msh6 rad30* spectra provide visual confirmation that *msh6* is not epistatic to

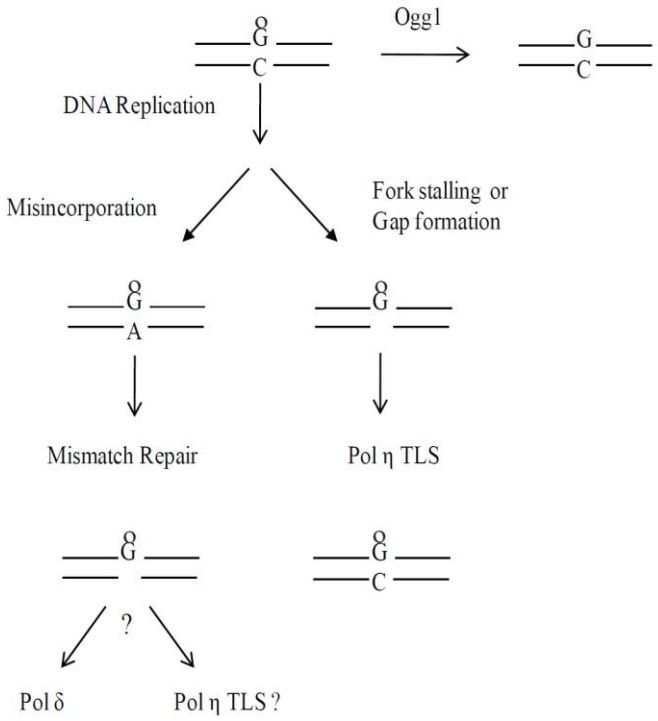

Figure 2.9 Roles of MMR and Polη Activities in Preventing GO-associated Mutagenesis

rad30; if it were, no changes would have been expected upon the additional deletion of *RAD30*. Our data thus clearly demonstrate that MMR and Polη can function

independently to limit GO-associated mutagenesis. As illustrated in Figure 2.9, we suggest that the MMR-independent role of Polη reflects its direct recruitment for error-free bypass of GO lesions that block the replicative DNA polymerases. Whether such recruitment involves a polymerase switch at the fork and/or occurs through the filling of gaps left behind the fork is not known. Interestingly, GO-associated mutagenesis is not elevated in strains in which the template switching pathway has been disabled (Table 2.8). This suggests that if GO lesions block replication, they do not trigger the template switching pathway and may instead only activate TLS.

Table 2.8 GO-associated Mutagenesis in *mms2Δ* Strains

	SUP4-o Mutation Rate[a]		GC > TA Mutation Rate		
Genotype	Total ($x10^{-7}$)	Relative to WT	Proportion	Rate ($x10^{-7}$)	Relative to WT
WT	1.6 (0.95-2.1)	1.0	43/160 (27%)	0.43	1.0
ogg1	2.1 (1.7-3.1)	1.3	175/262 (67%)	1.4	3.3
mms2	12 (7-16)	7.5	60/173 (35%)	4.2	9.8
ogg1mms2	7.8 (6.8-10)	4.9	70/161 (49%)	3.8	8.8

a 95% confidence intervals are in parentheses.

While our data clearly demonstrate that Polη can act independently of the MMR machinery to reduce GO-associated mutagenesis, they do not exclude the possibility that Polη may sometimes be recruited to fill GO-containing gaps created by MMR, as originally proposed (Haracska et al., 2000). We suggest, however, that Polη would only become involved in MMR if a replicative polymerase is blocked by a lesion during the gap-filling reaction. TLS in this context is likely no different from that triggered by any other replication-blocking lesion, and so it is not unique to the MMR process. Aside from this specific scenario, however, it is difficult to imagine a more global involvement of Polη in MMR; the relatively low fidelity of this enzyme on undamaged DNA would

lead to many nonspecific errors during MMR-associated gap filling (McCulloch et al., 2007). Indeed, the only unequivocal evidence of Polη functioning in MMR is in the specialized somatic hypermutation of B cells (Delbos et al., 2007).

This assay system provides us with a read out by which to study the requirements of Polη activity. Although it is not surprising that Polη appears to require an interaction with PCNA to bypass GO lesions, the finding that it does not require either ubiquitination of PCNA or the alternative 9-1-1 clamp is unexpected. Due to the error-prone nature of TLS polymerases, their activity is thought to be tightly regulated in the cell. Indeed, most studies have shown that activation of TLS requires PCNA ubiquitination by the Rad6/Rad18 complex in response to DNA damage (Garg and Burgers, 2005b; Haracska et al., 2004; Kannouche et al., 2004; Stelter and Ulrich, 2003; van der Kemp et al., 2009a; Watanabe et al., 2004; Zhuang et al., 2008). Moreover, Polζ-dependent bypass appears to be subject to further regulation by the 9-1-1 checkpoint clamp (Sabbioneda et al., 2005). Although there are a few examples of TLS activity without PCNA ubiquitination (Chen et al., 2006; Minesinger and Jinks-Robertson, 2005), it is not clear how this TLS is regulated or recruited to the site of damage. Our findings suggest that Polη activity is either not tightly regulated or is regulated by another, unknown mechanism. It is important to note, however, that the *pol30(K164R)* allele prevents not only ubiquitination of PCNA, but potentially most sumoylation of PCNA as well. It is possible that disruption of both ubiquitination and sumoylation results in other effects in the cell that directly or indirectly affect regulation of TLS.

One advantage of sequencing a small target is that mutation patterns can be easily discerned. In the case of the *SUP4-o* allele, there are 45 positions where GO-initiated GC

> TA tranversions can occur, and transversions at most of these sites were detected (Figure 1). As expected based on overall GC >TA rates, an analysis of mutation rates at individual sites within the *hbn1::SUP4-o* spectra confirmed that most sites were synergistically affected by the removal of both Polη and Msh6. There were sites, however, where loss of Polη had a greater-than-average effect and removal of Msh6 had little effect. We suggest that these sites correspond to locations where the replicative DNA polymerase is more efficiently blocked by a GO lesion. In contrast, other sites that were only affected by loss of Msh6 may be locations where the replicative polymerase rarely stalls, resulting in a high incidence of GO:A mismatches that are then subject to MMR. Because of this considerable site-to-site variation, our data underscore the great caution that needs to be exercised when using a single site (e.g., a reversion assay) in mutation analyses.

Even though mutations were much more evenly distributed in the *ogg1 rad30 msh6* triple mutant than in other genetic backgrounds, site-to-site differences persisted, indicating significant variability in the susceptibility of specific guanines to oxidative damage and/or in the propensity of the replicative DNA polymerases to insert A opposite GO lesions. This observed variability is not surprising, as previous studies have shown that mutation and repair rates vary across the genome (Hawk et al., 2005) and even within small stretches of DNA (Bebenek et al., 1999; Harfe and Jinks-Robertson, 2000a). Several studies, for example, have shown that GO lesions are less accessible to Ogg1 in areas of DNA that also contain AP sites, single-strand breaks, additional oxidative lesions and other types of damage (David-Cordonnier et al., 2001; Jiang and Wang, 2009; Pearson et al., 2004). Moreover, efficient nucleotide insertion opposite GO lesions and

primer extension from these insertions are also affected by nearby lesions and neighboring bases (Jiang and Wang, 2009; Yung et al., 2008). Based on previous studies and the work presented here using *SUP4-o*, it is clear that multiple factors determine the relative involvements of Ogg1, MMR, and Polη in limiting GO-associated mutagenesis at a given site.

Possible differences in the frequency or mechanism of mutagenesis during leading- versus lagging-strand synthesis have been examined using a variety of assays, mutagens, and strain backgrounds (Maliszewska-Tkaczyk et al., 2000; Pages et al., 2008; Pavlov et al., 2002; Thomas et al., 1993; Veaute and Fuchs, 1993; Watanabe et al., 2001). These different studies have inferred a leading-strand bias, a lagging-strand bias or the complete absence of a bias. A single, unifying explanation for these various findings is unlikely, as mutations result from a wide variety of initiating events and pathways. Our results with the *hbn1::SUP4-o* alleles demonstrate that, at this specific location, both MMR and Polη-dependent bypass are more efficient on the lagging strand of replication, leading to enhanced mutagenesis on the leading strand (Tables 2.5 and 2.6). We considered the possibility that the greater efficiency of Polη activity during lagging-strand synthesis at the *HBN1* location could simply reflect a role for Polη during MMR, and hence the strand bias of MMR. It should be noted, however, that some leading-strand mutation bias persisted in the absence of Msh6, and that this residual bias was completely eliminated upon additional removal of Polη. These results are consistent with earlier work demonstrating that MMR removes errors more efficiently during lagging- rather than during leading-strand synthesis (Pavlov et al., 2003) and provide the first evidence that the MMR-independent Polη bypass is also more efficient on the lagging

strand. Interestingly, all of the leading-strand bias in *SUP4-o* mutagenesis at the *AGP1* location required the presence of Polη, with the contribution of MMR being relatively minor. This suggests an additional layer of complexity, with effects of the immediate chromosomal environment extending into the mutational target.

It has been suggested that the lagging-strand bias of MMR is due to the increased accumulation either of nicks or of PCNA that accompanies discontinuous DNA synthesis (Pavlov et al., 2003). Just as PCNA would be left behind when the DNA polymerase is recycled to extend the next primer during lagging-strand synthesis, PCNA would presumably be left behind to mark a gap created when a lesion blocks a replicative polymerase. Because lagging-strand DNA replication is an inherently discontinuous process, we speculate that lesion-triggered gaps are more readily formed on the lagging strand than on the leading strand. This would account for the greater efficiency of Polη during lagging-strand synthesis as well as for the central role of PCNA in orchestrating the bypass reaction.

In addition to demonstrating replication-related strand differences in mutagenesis, the data presented here also indicate that there are more GO lesions present on the nontranscribed strand than on the transcribed strand of *SUP4-o* (Tables 2.5 and 2.6). Further experiments demonstrated that this bias was not due to transcription-coupled repair of the transcribed strand, suggesting that the nontranscribed strand is more susceptible to GO-associated mutagenesis (Table 2.7). It should be noted that both CG > TA mutations resulting from cytosine deamination and GO-dependent GC > TA transversions have been shown to preferentially accumulate on the nontranscribed of a

highly-transcribed gene in *E. coli* (Beletskii and Bhagwat, 1996; Klapacz and Bhagwat, 2005).

ROS are a constant threat to DNA integrity, and defining the mechanisms that limit their mutagenic effects is critical to understanding the regulation of eukaryotic genome stability. As in yeast, both MMR and Polη have been shown to prevent GO-associated mutagenesis in mammalian cells (Lee and Pfeifer, 2008; Mazurek et al., 2002; Russo et al., 2004; Russo et al., 2007). While purified yeast Polη is both more efficient and more accurate than Polδ in GO-lesion bypass, recent data suggest that human Polη has much lower fidelity (McCulloch et al., 2009). Whether this reflects a lack of relevant accessory proteins *in vitro* or a less prominent role of Polη in the error-free bypass of GO lesions in mammalian cells is unclear. The exact role of MMR in mammals is also unclear as the MutSβ complex does not bind GO mismatches at all and the MutSα complex does not bind or repair GO mismatches efficiently unless the mismatches are within repeats and are associated with slippage (Larson et al., 2003; Macpherson et al., 2005; Mazurek et al., 2002). Consistent with these *in vitro* data, oxidative lesions have been linked to frameshifts and microsatellite instability in many species (Gasche et al., 2001; Jackson et al., 1998; Vongsamphanh et al., 2006). Interestingly, human cells lacking Polη were recently shown to have increased rates of spontaneous fragile site instability (Rey et al., 2009). It is possible that replication is more likely to stall at GO lesions present in fragile sites and, therefore, that Polη is required for bypassing these lesions and maintaining genome stability. Future studies that further elucidate the mechanisms behind MMR-independent Polη bypass will provide additional insight into

the role of TLS in the prevention of oxidative mutations and specifically the role of human Polη in limiting mutagenesis and preventing human disease.

2.6 Acknowledgements

This work was supported by NIH grants (GM038464 and GM064769) awarded to S.J.-R. We would like to thank Jessica Franke and Amy Abdulovic for contributions made in the early phases of this work.

Chapter 3: Examination of the effect of replication timing on mutagenesis caused by 7,8-dihydro-8-oxoguanine

3.1 Summary

DNA replication is initiated at hundreds of individual replication origins across the genome. Each of these origins is temporally regulated to fire at a specific time during the S phase of the cell cycle, which results in early-replicating and late-replicating regions in the genome. Studies in various species have shown that increased rates of mutagenesis are associated with late-replicating regions of the genome. In this study, we examined whether this was also true of mutagenesis associated with the oxidative DNA lesion 7,8-dihydro-8-oxoguanine (GO) in the yeast *Saccharomyces cerevisiae*. We generated yeast strains in which the *SUP4-o* allele was inserted near an early-firing origin, *ARS306*, or a late-firing origin, *ARS501*. To specifically examine rates of GO-associated mutagenesis at these origins in the presence and absence of the translesion synthesis (TLS) polymerase Polη or mismatch repair (MMR), we generated *ogg1*, *ogg1 rad30*, and *ogg1 msh6* derivatives of each yeast strain. Although we did not observe significant variations in mutation rate in the *ogg1* and *ogg1 rad30* strains, the rate of GO-associated mutations was significantly increased in the *ARS501 ogg1 msh6* strain. To confirm the increased rate of mutagenesis near a late-firing origin, we then examined GO-associated mutagenesis at five other origins, two early-firing and three late-firing. While we observed variations in mutation rate at different origins, the variations do not appear to correlate with the replication times that have been obtained in previous studies.

We suggest that these variations are due to other differences in chromosomal context or that replication times are not consistent across various yeast strains.

3.2 Introduction

During the synthesis (S) phase of the cell cycle, the cell must replicate its entire genome quickly and efficiently. In *Saccharomyces cerevisiae,* for example, the 13 Mb genome is replicated in 25-30 minutes. The presence of hundreds of replication origins, also referred to as autonomous replicating sequences (ARSs), throughout the genome is critical for this process. Rather than starting DNA synthesis at one end of each chromosome and continuing until reaching the opposite end, several replication origins along each chromosome initiate replication of the DNA that lies both upstream and downstream. In this way, replication occurs all across the genome at the same time, significantly increasing the speed and efficiency of DNA synthesis.

Given their critical involvement in DNA replication, a large amount of research in the field has been aimed at identifying and characterizing these origins. The first origins in yeast were identified in screens designed to detect sequences that allowed propagation of an extrachromosomal plasmid (Hsiao and Carbon, 1981; Stinchcomb et al., 1980). Many of these sequences were then verified as origins with the use of two-dimensional gel electrophoresis (Brewer and Fangman, 1987, 1991). Mutagenesis assays revealed that each ARS contains a 100-150-bp variable region and an 11-17-bp consensus sequence, which is referred to as the ACS (Broach et al., 1983; Kearsey, 1984; Theis and Newlon, 1997). This consensus sequence alone is not sufficient for origin activity, as genome-wide searches of this sequence have uncovered over 10,000 matches and have

thus not been very helpful in the identification of true origins (Breier et al., 2004). The first attempts to identify all origins across a chromosome involved two-dimensional gel electrophoresis of many individual chromosome fragments (Friedman et al., 1997; Newlon et al., 1993; Reynolds et al., 1989). In recent years, microarray technology has been combined with various techniques to identify all origins across the yeast genome. For example, Raghuraman et al. generated replication profiles of each chromosome with the use of a density transfer technique similar to that used in the classic Meselson and Stahl experiment, which demonstrated the semi-conservative nature of DNA replication (Meselson and Stahl, 1958; Raghuraman et al., 2001). Other groups have mapped origins across the genome by identifying changes in copy number, by mapping regions of single-stranded DNA, and by identifying regions of DNA that bind various components of the pre-replication complex (pre-RC), such as the origin recognition complex (ORC) and the Mcm proteins (Feng et al., 2006; Wyrick et al., 2001; Xu et al., 2006; Yabuki et al., 2002). These studies have revealed that there are approximately 300-400 origins in the yeast genome, and information on each of these different origins can be found on the DNA Replication Origin Database (OriDB; Nieduszynski et al., 2007).

While some studies have focused on identifying origins, other studies have been aimed at further characterizing different origins and how they are regulated. Interestingly, each origin in yeast appears to be temporally regulated to initiate DNA replication at a specific time during S phase (e.g., early, mid-phase, or late; reviewed in Weinreich et al., 2004). Studies have shown that the pre-RC associates with each origin, thereby preparing it for initiation, early during the G1 phase of the cell cycle (Diffley, 2004; Raghuraman et al., 1997). While the essential components of the pre-RC appear to

be similar for all origins, various other components are involved in the temporal regulation of each origin. For example, although each origin contains a pre-RC at the beginning of S phase, the Cdc45 and RPA proteins, which are required for initiation of replication, only bind to an origin when it is time for it to initiate replication (Aparicio et al., 1999; Tanaka and Nasmyth, 1998; Zou and Stillman, 2000).

It appears that the variable flanking sequence of each origin contributes to its regulation. For example, the association of the late-firing origins such as *ARS501* with a telomere directly affects its replication timing, as moving the origin to a plasmid resulted in its firing earlier during S phase (Ferguson and Fangman, 1992). This is likely due to the "silenced" heterochromatic state of telomeres, which is regulated by the Sir proteins (Stevenson and Gottschling, 1999). The late-firing origin *ARS1413* is not associated with a telomere, however, and deletion studies have shown that multiple sequence elements near this origin contribute to its delayed firing (Friedman et al., 1996). It has been suggested that other types of chromatin modifications, such as histone acetylation, affect origin timing by silencing different regions until the origin is supposed to fire (Vogelauer et al., 2002). In support of this idea, the histone deacetylase Rpd3 has been shown to repress the firing of many late-firing origins in yeast (Knott et al., 2009). In contrast, the Clb5 protein has been shown to specifically activate late-firing origins later in S phase (Donaldson et al., 1998; McCune et al., 2008). Finally, there is evidence that heterochromatic regions of DNA are physically associated with the periphery of the nucleus during G1 phase, while euchromatic regions are more centrally localized (Gotta et al., 1996; Laroche et al., 1998; Li et al., 1998; Ma et al., 1998). This has been shown for both telomeres and various late-firing origins, regardless of whether they are

associated with telomeres (Heun et al., 2001). It is thus possible that the physical location of late-firing origins contributes to their regulation. While we are learning more and more about different types of origin regulation, it is still unclear how this regulation network is controlled and how different types of regulation affect each individual origin.

Studies of *Escherichia coli*, Drosophila, mice, primates, and humans have identified significantly increased mutation rates and increased proportions of certain types of nucleotides (e.g., GC-rich) and single nucleotide polymorphisms (SNPs) in late-replicating regions of the genome (Anderson et al., 2008; Deschavanne and Filipski, 1995; Diaz-Castillo and Golic, 2007; Pink and Hurst, 2009; Stamatoyannopoulos et al., 2009; Subramanian et al., 1996; Watanabe et al., 2002). These studies suggest that mutagenesis is affected by the timing of DNA replication. Mutagenesis studies in yeast, however, have produced conflicting results. Some studies have detected increased rates of mutagenesis in late-replicating DNA (Lang, 2007; Payen et al., 2009; Teytelman et al., 2008), while others have found variable mutation rates that were not associated with replication timing (Hawk et al., 2005; Ito-Harashima et al., 2002), and still others have found uniform mutation rates across the genome (Chin et al., 2005). The reasons for these discrepancies in yeast are unclear, but may be due to differences in assays and/or yeast strains.

Many factors likely contribute to increased mutagenesis in late-replicating areas of the genome. For example, late-replicating origins are often found near telomeres or other silenced or heterochromatic DNA regions. The compact nature of heterochromatin has been shown to prevent repair and recombination proteins from accessing DNA damage (Amouroux et al., 2010; Chaudhuri et al., 2009; Li et al., 2009; Menoni et al.,

2007; Osley et al., 2007). Several studies have also suggested that the error-prone translesion synthesis (TLS) polymerases are more active later in S phase. Experiments in *E. coli* have found that the SOS system, which is similar to the TLS system in eukaryotes, delays replication when DNA damage is encountered, allowing more efficient repair systems to correct the damage (Opperman et al., 1999). Only later in S phase do the proteins involved in this process proceed with lesion bypass. A recent study in yeast found that one TLS polymerase, Rev1, was upregulated in late S and in G2/M phase (Waters and Walker, 2006). These authors suggested that TLS polymerases could be more active later in S phase to fill in gaps that have been left behind the replication fork. Indeed, many studies have suggested that when a lesion is encountered during DNA replication, the replication machinery terminates DNA synthesis and then re-initiates synthesis downstream, leaving a lesion-containing gap (Lopes et al., 2006; Meneghini et al., 1981; Pages and Fuchs, 2003; Sale et al., 2009). As TLS polymerases are associated with elevated rates of spontaneous mutagenesis (Northam et al., 2010; Quah et al., 1980), rates of mutagenesis might be expected to be higher in late-replicating DNA, a time when TLS polymerases are most active.

In addition to possible cell cycle-dependent TLS activity, mismatch repair (MMR) has also been shown to have variable efficiencies at different regions of the genome. In one particular study, lower mutation rates were found in regions of the genome that are late-replicating (Hawk et al., 2005). Although these decreases were not statistically significant, it is possible that replication timing is at least partially involved in this difference in activity. Finally, it is possible that varying levels of dNTPs also contribute to increased rates of mutagenesis during late S phase. Increased levels of

dNTPs have been shown to result in increased DNA lesion bypass by replicative polymerases, which results in increased rates of mutagenesis (Sabouri et al., 2008). Moreover, studies in mammalian cells have shown that dNTP levels, and thereby the speed of replication, increase throughout S phase (Malinsky et al., 2001; Walters et al., 1973). It is therefore possible that dNTP levels are highest in late S phase, and that this might result in increased rates of mutagenesis in late-replicating regions of the genome.

We have previously used the *SUP4-o* forward mutation assay to study the pathways that are involved in preventing mutations that arise from the oxidative lesion 7,8-dihydro-8-oxoguanine or GO (see Chapter 2). GO lesions are one of the most common types of oxidative DNA damage and are relatively easy to study because they produce characteristic GC > TA transversion events. This assay enables us to sequence large numbers of mutants and thereby specifically calculate rates of GO-associated mutagenesis. In this study, we inserted the *SUP4-o* allele adjacent to seven different well characterized replication origins, three early-firing and four late-firing. We then increased the amount of GO-associated mutagenesis by deleting the *OGG1* gene, which encodes a DNA glycosylase that specifically removes GO lesions. The effect of replication timing on GO-associated mutagenesis in the presence and absence of MMR was examined.

3.3 Materials and Methods

Strain Construction

All strains were derived from strain SJR576 (*MATa ura3ΔNco lys2-1$_{oc}$ can1-100$_{oc}$ ade2-1$_{oc}$ leu2-K*). The *SUP4-o* allele-containing plasmid JF1754 was obtained

from Bernard A. Kunz (Pierce et al., 1987). *SUP4-o* was amplified from the plasmid and inserted into various regions of the genome using primers with homology to the *SUP4-o* plasmid and homology to the specific region of the genome where the allele was integrated. *SUP4-o* was integrated into the *HBN1* locus and the *AGP1* locus, which are near *ARS306*, as described in Chapter 2. *SUP4-o* was inserted ~250 bp from *ARS607* using the primers 5'GCATTTACGCACTCTAACTGGCATTTTAAAGAAAAAGGG ATAAATGCGGATCCGGGACCGGATAATT 3' and 5'CCATTTATCTTATGATAT CTAGCAAATCAAAGGATCCCTCAATTCTTGAAAGAAATATTTC 3'. *SUP4-o* was integrated ~200 bp from *ARS121* using the primers 5'GGGTACTAGATTCATT CATTTATTTCTATTCAAGGACAAGAACGGATCCGGGACCGGATAATT 3' and 5'CGTGATCTCTTTTAGAGAA AGGACTTAATCCGTACACAATGAATTCTTG AAAGAAATATTTC 3'. *SUP4-o* was integrated ~150 bp from *ARS1502* using the primers 5'CTATTCGTCAAGTC TTAAATCCATATTTTAATATTCATCAGGGAT CCGGGACCGGATAATT 3' and 5'CTGATAGGGTGAGCGCAAGAATTAGTAAT GTGTTGGTAACAATTCTTGAAAGAAATATTTC 3'. *SUP4-o* was integrated ~250 bp from *ARS1413* using the primers 5'GGTACTTTTGTCTGTTTTATAGTACTTGTA CATCAGACAAGG GGATCCGGGACCGGATAATT 3' and 5'CCTTTCACTGACCA ATCAACTGCTTTTCAATATCTTTATAAGCCAATTCTTGAAAGAAATATTTC 3'. *SUP4-o* was integrated ~210 bp from *ARS609* using the primers 5'GGGTAAAAGTC GAGCTTGTTTTCTGAAGCGGAAATTACAGCGGATCCGGGACCGGATAATT 3' and 5'GCGATTCCGTGCTCACCGCGGCTTCGCCTATATTCTACCTGAATTCTT GAAAGAAATATTTC 3'. *SUP4-o* was inserted ~1 kb from *ARS501* using the primers 5'GCTTCTTTGTGGCATCGCCCATGGGATCAAACCATACTGGTTTCTTTGTA

AAACAGGGGATCCGGGACCGGATAATT 3' and 5'GGTAGATCAGTCAAAC GGATGCGTCGCATAAATGGCTGATAAATTTTTCCACTACAATTCTTGAAAGAA ATATTTC 3'. For each primer set, the underlined portion of the sequence corresponds to sequence flanking *SUP4-o*, and the non-underlined portion corresponds to the region of the genome where *SUP4-o* was inserted. Following transformation of SJR576 with the appropriate PCR product, Lys^+ colonies were selected, and the presence of *SUP4-o* was inferred by co-suppression of the *ade2-101* and *can1-100* ochre alleles. Insertion of *SUP4-o* at the correct location was confirmed by sequencing.

The *OGG1*, *RAD30*, and *MSH6* genes were deleted by transforming strains with a PCR-generated fragment containing a kanamycin resistance (Bahler et al., 1998), *URA3-Kl* (*URA3* gene from *Kluveromyces lactis*) (Gueldener et al., 2002) or hygromycin resistance marker (Goldstein and McCusker, 1999) with the appropriate flanking sequence of the target gene. *ogg1Δ::kan* and *msh6Δ::hyg* transformants were selected on YEPD medium (1% yeast extract, 2% Bacto-peptone, 2% dextrose, and 250 mg/L adenine) containing 200 μg/mL geneticin or 300 μg/mL hygromycin, respectively. *rad30Δ::URA3-Kl* transformants were selected on synthetic complete medium containing 2% dextrose and lacking uracil (SCD-URA). Deletions were confirmed by PCR.

The *pGAL-RNR1* construct was obtained on a plasmid from Andrei Chabes (Chabes and Stillman, 2007). This plasmid also contains the selectable marker *URA3*. The plasmid was linearized with *Stu*I, which lies within *URA3*, and transformed into the appropriate yeast strains. The plasmid then recombined with the endogenous, mutant *ura3* allele, inserting the entire plasmid and generating a Ura^+ phenotype. Correct insertion of the plasmid was verified by PCR.

Mutation Rate and Spectra Analysis

To determine mutation rates, individual colonies were used to inoculate 1 mL cultures in non-selective YEPGE medium (1% yeast extract, 2% Bacto-peptone, 2% glycerol, 2% ethanol, and 250 mg/L adenine). These cultures were grown for 2 days at 30°C. Two isolates were used for each strain, and at least 12 cultures were used for each isolate. Appropriate dilutions of each culture were plated onto YEPGE medium to determine total cell number and on SCD-ARG plates supplemented with 60 μg/mL L-canavanine (Sigma) to select canavanine-resistant (Can^R) colonies. Colonies were designated as *SUP4-o* mutants if they were both resistant to canavanine and red (Ade^-), indicating loss of suppression of both the *can1-100* and *ade2-1* alleles. Mutation rates were determined using at least 24 cultures and the method of the median (Lea, 1949), and 95% confidence intervals were calculated as previously described (Spell and Jinks-Robertson, 2004). Either comparison of the confidence intervals or the Mann-Whitney test (http://faculty.vassar.edu/lowry/utest.html) was used to determine whether two rates were significantly different.

To generate mutation spectra, DNA was extracted from purified Can^R, red colonies (http://jinks-robertsonlab.duhs.duke.edu/protocols/yeast_prep.html). The *SUP4-o* gene was amplified by PCR and sequenced by the Duke University DNA Analysis Facility (Durham, North Carolina). Mutation rates for specific mutation types were calculated by multiplying the proportion of that event in the corresponding spectrum by the total mutation rate. Proportions of mutations were analyzed statistically using Chi Square analysis (http://faculty.vassar.edu/lowry/VassarStats.html). The efficiency of

MMR in suppressing GO-associated mutagenesis was calculated using the equation (*ogg1 msh6* rate – *ogg1* rate)/(*ogg1 msh6* rate), and the efficiency of Ogg1 in removing GO lesions was calculated using the equation (*ogg1* rate – WT rate)/(*ogg1* rate). The rates of GO-associated mutations were used for both calculations.

3.4 Results

Elevated rate of GO-associated mutagenesis near a late-firing origin

To examine the effect of replication timing on GO-associated mutagenesis, we used the *SUP4-o* forward mutation assay described in Chapter 2. This assay enables us to examine both the rate and mutation spectra of all forward mutations within a short, 89-bp window. As discussed in Chapter 2, GO-associated mutagenesis can be examined by deleting the *OGG1* gene, which encodes a DNA glycosylase that specifically removes GO lesions. This results in increased GO lesions during DNA replication and a concomitant increase in the proportion of GC > TA mutations, which result from an incorporation of adenine opposite a template GO lesion. The total number of GO-associated mutations is determined based on the number of G > T and C > A mutations in the mutation spectra. The reported mutation spectra are of the transcribed strand of *SUP4-o*.

The experiments discussed in Chapter 2 were performed using yeast strains in which the *SUP4-o* allele was inserted in either the *HBN1* gene or the *AGP1* gene, ~750 bp and ~1.5 kb, respectively, from the early-firing origin *ARS306* on chromosome III. With these strains, we showed that GO-associated mutations occur at a higher frequency on the leading strand relative to the lagging strand and on the nontranscribed strand

relative to the transcribed strand (see Chapter 2). It is therefore important to be aware of these strand differences when examining rates and spectra of GO-associated mutagenesis. In this system, G > T mutations reflect GO lesions that are on the transcribed strand, and C > A mutations reflect GO lesions that are on the nontranscribed strand. The orientation of *SUP4-o* relative to a replication origin determines which strand will be the leading-strand template and which will be the lagging-strand template. The positions of *SUP4-o* near *ARS306* are shown in Figure 3.1. In the *hbn1::SUP4-o* strain, the *SUP4-o* allele is in "Position 1" relative to the origin, and the transcribed strand is thus the lagging-strand template (LG). In the *agp1::SUP4-o* strain, the *SUP4-o* allele is in "Position 2" relative to the origin, and the transcribed strand is thus the leading-strand template (LD). For simplicity, we will refer to these and subsequent strains according to the origin that is adjacent to the *SUP4-o* allele and whether the transcribed strand is the lagging- or leading-strand template (i.e., *ARS306-LG* and *ARS306-LD*, respectively).

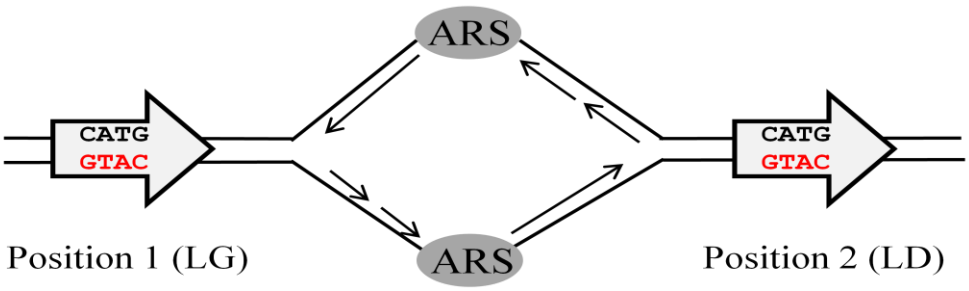

Figure 3.1 The orientation of *SUP4-o* alleles relative to a replication origin

The transcribed strand of *SUP4-o* is indicated in red. When *SUP4-o* is in "Position 1" relative to the origin, the transcribed strand is the lagging-strand template. In the text, these strains are denoted "*LG*." When *SUP4-o* is in "Position 2" relative to the origin, the transcribed strand is the leading-strand template. These strains are denoted "*LD*" in the text.

In this study, we generated a corresponding yeast strain in which the *SUP4-o* allele was inserted within an intergenic region ~1 kb from the late-firing origin *ARS501*

Table 3.1 Mutation Rates of *ARS306* and *ARS501* Strains

Origin[a]	Genotype	Overall Rate[b] ($x10^{-7}$)	GC > TA Mutation Rate		
			Total ($x10^{-7}$)	G > Ts	C > A
ARS306-LG (Early)	WT	1.6 (0.95-2.1)	0.43	0.09 (9/160)	0.34 (34/160)
	ogg1	2.1 (1.7-3.1)	1.4	0.21 (26/262)	1.2 (149/262)
	ogg1 rad30	5.4 (3.6-11)	4.4	1.3 (58/243)	3.1 (139/243)
	ogg1 msh6	15 (11-19)	11	4.5 (90/303)	6.9 (139/303)
ARS306-LD (Early)	WT	0.76 (0.44-1.2)	0.17	0.10 (21/161)	0.071 (15/161)
	ogg1	1.4 (1.1-1.7)	0.89	0.42 (46/152)	0.46 (50/152)
	ogg1 rad30	4.1 (2.5-6.4)	3.5	1.3 (71/219)	2.2 (118/219)
	ogg1 msh6	11 (9.2-16)	8.1	4.3 (88/226)	3.8 (79/226)
ARS501-LD (Late)	WT	1.6 (1.2-2.4)	0.43	0.22 (28/200)	0.21 (26/200)
	ogg1	3.8 (2.0-4.9)	2.2	0.99 (42/162)	1.2 (50/162)
	ogg1 rad30	6.4 (2.0-10)	5.8	2.7 (76/178)	3.1 (85/178)
	ogg1 msh6	36 (26-63)	24	11 (44/146)	13 (53/146)

[a] LG indicates that the transcribed strand of *SUP4-o* is on the lagging strand, and LD indicates that the transcribed strand is on the leading strand.
[b] 95% confidence intervals are in parentheses.

on chromosome V (Ferguson et al., 1991). In this strain, the *SUP4-o* allele is in "Position 2" relative to *ARS501*, and we will therefore refer to this strain as *ARS501-LD*. The effect of replication timing on GO-associated mutagenesis was initially examined by determining the rate of GC > TA mutations in both the *ARS306-LG ogg1* and *ARS306-LD* strains and the *ARS501-LD ogg1* strain. As shown in Table 3.1, the overall rate of GO-associated mutagenesis in the *ARS501-LD ogg1* strain was elevated slightly relative

to the *ARS306-LG ogg1* and *ARS501-LD ogg1* strains (2.5- and 1.6-fold, respectively). More striking differences between the strains were evident when rates of G > T and C > A mutations were considered separately. In the *ARS306-LG ogg1* strain, there were 5.7-fold more C > A mutations than G > T mutations (Figure 3.2 and Table 3.1). In this strain, C > A mutations are on the leading and nontranscribed strands, and G > T mutations are on the lagging and transcribed strands. This bias for C > A mutations thus reflects the increased proportions of mutations on the leading and nontranscribed strands, as discussed in Chapter 2. In contrast, the rates of C > A and G > T mutations in the *ARS306-LD ogg1* and *ARS501-LD ogg1* strains were similar ($0.42x10^{-7}$ and $0.46x10^{-7}$, and $0.99x10^{-7}$ and $1.2x10^{-7}$, respectively). In both of these strains, G > T mutations are on the leading and transcribed strands, and C > A mutations are on the lagging and nontranscribed strands. Although C > A mutations on the lagging strand should occur at a lower frequency than G > T mutations on the leading strand, we suggest that this lower frequency is masked by the increased frequency of mutations on the nontranscribed strand. This results in relatively equal rates of G > T and C > A mutations. In summary, the rates of GO-associated mutagenesis in an *ogg1* background vary slightly with replication timing, and differences in rates of G > T and C > A mutations can be attributed to differences in leading- and lagging-strand mutagenesis.

As discussed above, we previously showed that both MMR and the TLS polymerase Polη are involved in suppressing GO-associated mutagenesis (Table 2.1), and both MMR and TLS have been suggested to have variations in activity across the genome. To determine the effect of replication timing on the activity of MMR and Polη, we examined the rate of GO-associated mutagenesis in the absence of either MMR or

Polη in the *ARS306* and *ARS501* strains. To examine MMR, we deleted the *MSH6* gene, which is required for the MutSα mismatch recognition complex of MMR. To examine

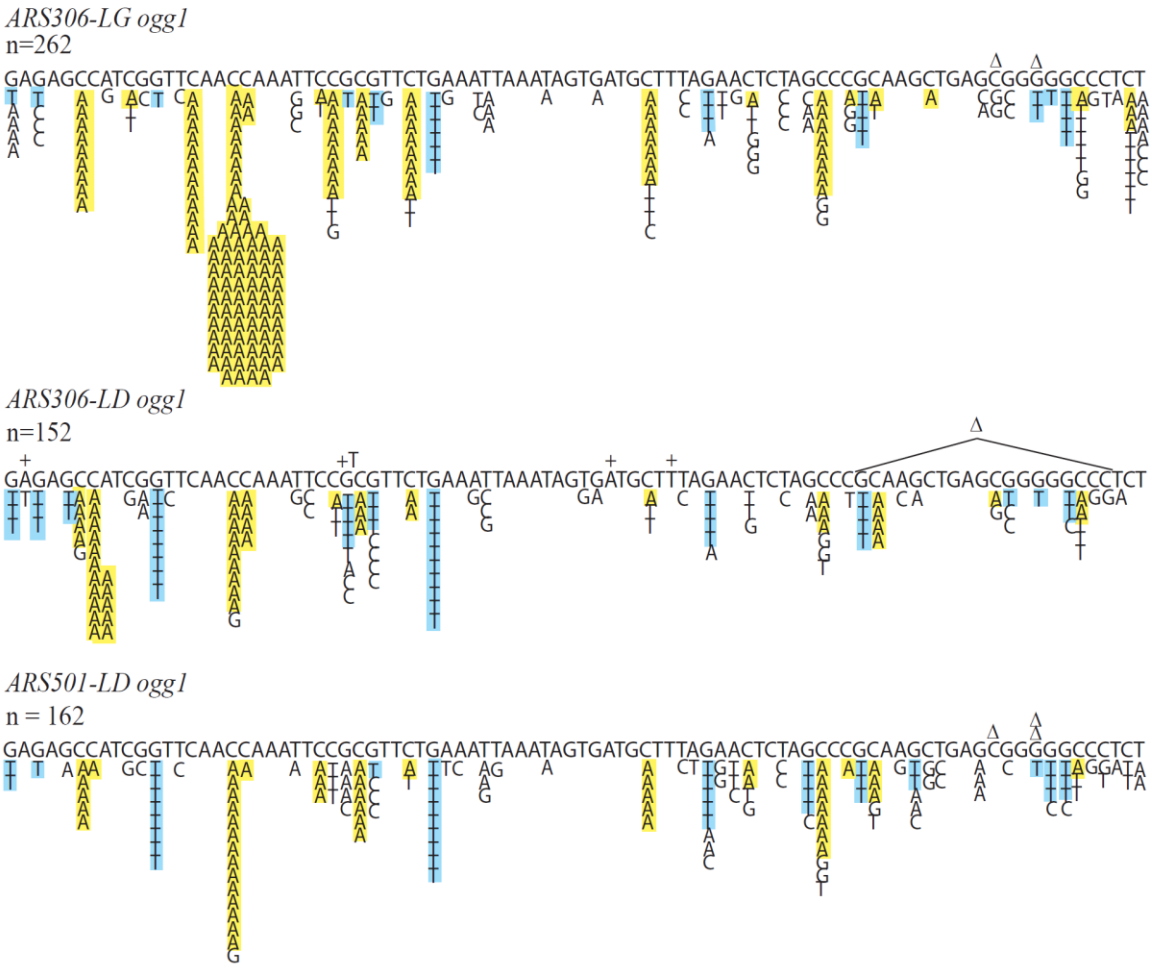

Figure 3.2 Mutation Spectra of *ARS306* and *ARS501 ogg1* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

Polη activity, we deleted the *RAD30* gene, which encodes Polη. Both of these deletions were made in an *ogg1* background.

As shown in Table 3.1, the rate of GO-associated mutagenesis was significantly elevated in the *ARS501-LD ogg1 msh6* strain relative to both the *ARS306-LG ogg1 msh6* and *ARS306-LD ogg1 msh6* strains. This suggests that in the absence of MMR, the rate of GO-associated mutagenesis is increased later in S phase. When we examined G > T and C > A mutations specifically in the three strains, we again observed that the

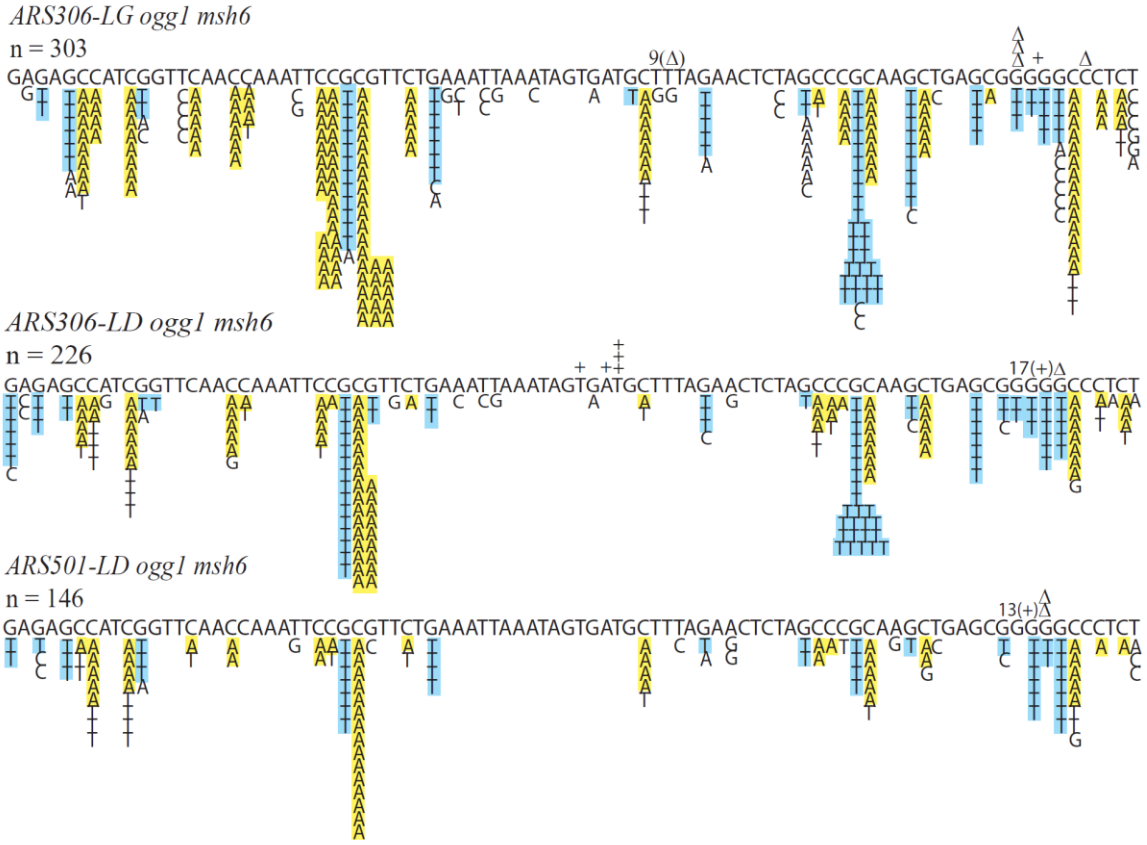

Figure 3.3 Mutation Spectra of *ARS306* and *ARS501 ogg1 msh6* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

proportions of G > T and C > A mutations were significantly different in the *ARS306-LG ogg1 msh6* strain (30% and 46%, respectively; $p<0.0001$; Figure 3.3). Interestingly, while the rate of C > A mutations was elevated 5.7-fold relative to the rate of G > T mutations in the *ARS306-LG ogg1* strain, the rate of C > A mutations was elevated only 1.5-fold relative to G > T mutations in the *ARS306-LG ogg1 msh6* strain. As MMR is more efficient on the lagging strand (see Chapter 2), this decreased bias is due to the greater increase in lagging-strand (G > T) mutations relative to leading-strand (C > A) mutations in the absence of MMR (21-fold and 5.8-fold, respectively). As expected, the proportions of G > T and C > A mutations were similar in both the *ARS306-LD ogg1 msh6* (39% and 35%, respectively; $p=0.4$) and *ARS501-LD ogg1 msh6* (30% and 36%, respectively; $p=0.3$) strains.

In contrast to the *ogg1 msh6* strains, the rate of GO-associated mutagenesis in the *ARS501-LD ogg1 rad30* strain was not elevated relative to the *ARS306-LG ogg1 rad30* or *ARS306-LD ogg1 rad30* strain (Table 3.1). Examination of the mutation spectra of these strains revealed the same differences observed with the *ogg1* and *ogg1 msh6* strains; while the proportions of C > A and G > T mutations were significantly different in the *ARS306-LG* strain, the proportions were similar in the *ARS306-LD* and *ARS501-LD* strains (Figure 3.4 and Table 3.1). Because the rates of GO-associated mutations in the three strains were similar, we concluded that the activity of Polη does not vary with replication timing and did not examine Polη further.

Mutation rates of *ogg1 msh6* strains at other early- and late-firing origins

To confirm that GO-associated mutagenesis is increased in areas of the genome that are late-replicating in the absence of MMR, we generated five more strains in which

the *SUP4-o* allele was localized near (within 300 bp) either an early- or late-firing origin. The two other early-firing origins chosen for this study were *ARS607* on chromosome VI

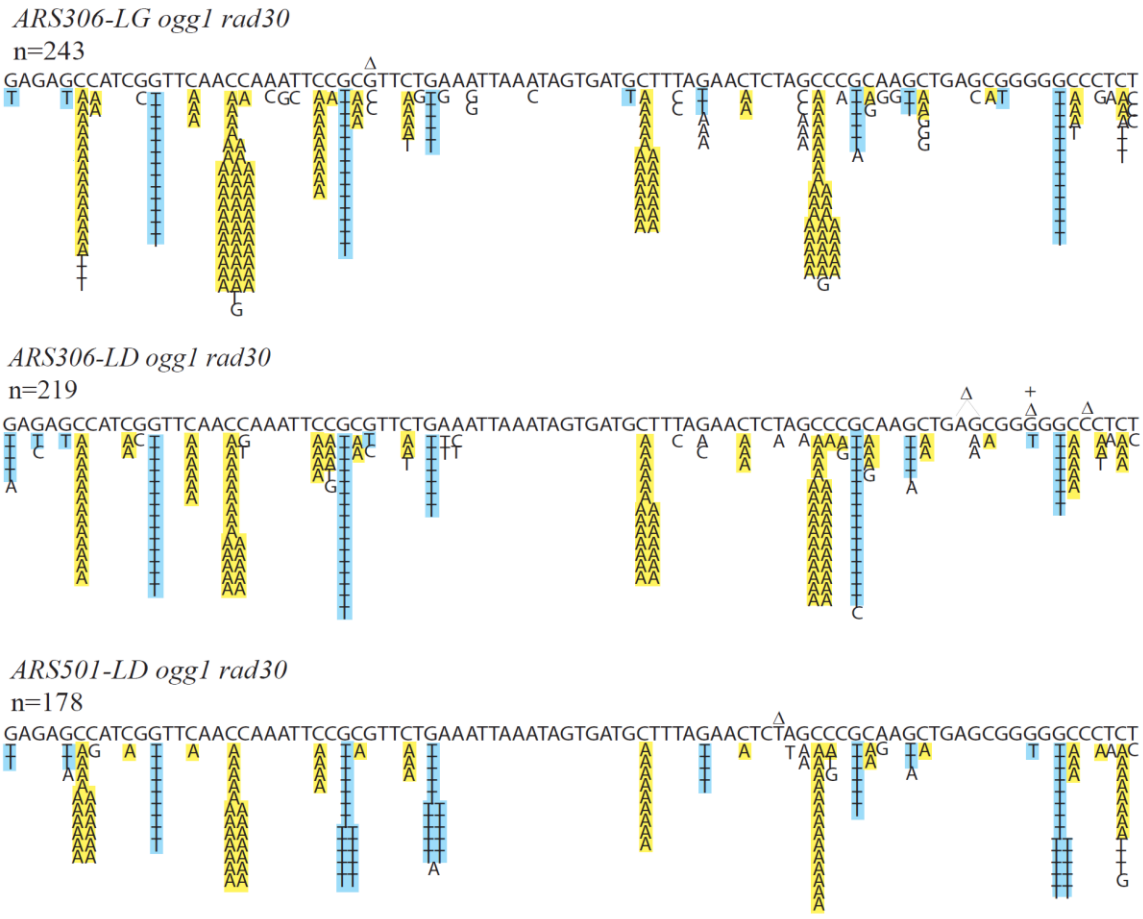

Figure 3.4 Mutation Spectra of *ARS306* and *ARS501 ogg1 rad30* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

and *ARS121* on chromosome X, which have been described as being efficient and early-firing in several studies (Feng et al., 2006; Friedman et al., 1997; Knott et al., 2009; McCune et al., 2008; Raghuraman et al., 2001; Yabuki et al., 2002; Yamashita et al.,

1997). The three other late-firing origins chosen for this study were *ARS1502* on chromosome XV, *ARS1413* on chromosome XIV, and *ARS609* on chromosome VI (Donaldson et al., 1998; Feng et al., 2006; Friedman et al., 1997; Knott et al., 2009; McCune et al., 2008; Raghuraman et al., 2001; Yabuki et al., 2002).

Table 3.2 Replication Times (minutes) of Selected Origins in Two Studies

		Raghuraman et al. (2001)	Yabuki et al. (2002)
Early	*ARS306*	11	18
↓	*ARS607*	13	20
↓	*ARS121*	16	18
↓	*ARS1502*	21	28
↓	*ARS1413*	29	29
↓	*ARS501*	33	28
Late	*ARS609*	42	32

It is important to note that origin firing does not occur at either a specific "early" or "late" time point, but occurs along a continuum throughout S phase; early-firing origins are characterized as firing before the midpoint of S phase, and late-firing origins are characterized as firing after the midpoint of S phase. Although the replication times of these origins do vary in some of these studies, they are almost always characterized as being either early- or late-firing. For example, the replication times determined by Raghuraman et al. (2001) and Yabjuki et al. (2002) for each of the origins in this study are shown in Table 3.2. The first study used a density transfer method to determine replication timing, and the second study used a copy number method. For both studies, *ARS306* was the earliest origin to fire. Both *ARS607* and *ARS121* were considered to be early-firing, although there is some discrepancy as to which origin fires first. Likewise, for both studies, *ARS1502*, *ARS1413*, *ARS609*, and *ARS501* are considered to be late-firing, but the order in which the origins fire is different. Other studies of these origins

present similar findings (Donaldson et al., 1998; Feng et al., 2006; Knott et al., 2009; McCune et al., 2008). Thus, for these studies, we attempted to identify correlations between mutation rates and origins that tend to fire either early or late during S phase.

We generated *ogg1* and *ogg1 msh6* derivatives of all of the new strains and determined rates of GO-associated mutagenesis in both the presence and absence of MMR. As shown in Table 3.3, the overall mutation rates of all of the *ogg1* strains varied between $1.4x10^{-7}$ and $3.8x10^{-7}$ (2.7-fold). Interestingly, the lowest mutation rate was observed for the *ARS306-LD* strain, and the highest mutation rate was observed for the *ARS501-LD* strain. The mutation rates of the remaining six strains varied only between $1.8x10^{-7}$ and $2.1x10^{-7}$ (1.2-fold), while the rates of GO-associated mutations varied between $0.82x10^{-7}$ and $2.2x10^{-7}$ (2.7-fold). As the lowest and highest rates were observed for the two latest origins, *ARS609-LG* and *ARS501-LD*, this variation did not appear to correlate with replication timing. Variations in the activity of Ogg1 are discussed below. The mutation spectra for the *ogg1* early- and late-firing strains are shown in Figures 3.5 and 3.6, respectively. As the orientation of the *SUP4-o* allele relative to the replication fork affects the proportions of G > T and C > A mutations in mutation spectra (discussed above), the strandedness of each strain is indicated; strains in which the transcribed strand of *SUP4-o* is the lagging-strand template are indicated as "*LG*", and strains in which the transcribed strand is the leading-strand template are indicated as "*LD*." The *ARS306-LG*, *ARS306-LD*, and *ARS501-LD* strains are included for comparison. As shown in Table 3.4, the proportions of G > T and C > A mutations are significantly different in all "*LG*" strains, while the proportions are similar for all "*LD*" strains. The one exception to this is

the *ARS609-LG* strain. This origin has been shown to be very inefficient, and it is likely that replication through this region also occurs via an origin lying on the other side of

Table 3.3 GO-associated Mutagenesis at Early- and Late-firing Origins

				GC > TA Mutations	
	Origin[a]	Genotype	Total Rate[b] ($x10^{-7}$)	Proportion	Rate ($x10^{-7}$)
Early	*ARS306-LG*	WT	1.6 (0.95-2.1)	43/160 (27%)	0.43
		ogg1	2.1 (1.7-3.1)	175/262 (67%)	1.4
		ogg1 msh6	15 (11-19)	229/303 (76%)	11
	ARS306-LD	WT	0.76 (0.44-1.2)	36/161 (22%)	0.17
		ogg1	1.4 (1.1-1.7)	96/152 (63%)	0.89
		ogg1 msh6	11 (9.2-16)	167/226 (74%)	8.1
	ARS607-LG	WT	1.2 (0.87-2.2)	23/65 (35%)	0.42
		ogg1	1.9 (1.6-2.7)	83/162 (51%)	0.97
		ogg1 msh6	23 (13-44)	131/166 (79%)	18
	ARS121-LD	WT	1.3 (0.75-2.3)	18/56 (32%)	0.42
		ogg1	1.9 (1.5-2.7)	76/157 (48%)	0.91
		ogg1 msh6	15 (11-23)	139/185 (75%)	11
	ARS1502-LG	WT	1.6 (1.3-2.7)	30/82 (37%)	0.59
		ogg1	1.8 (1.0-2.1)	95/148 (64%)	1.1
		ogg1 msh6	20 (15-32)	139/183 (76%)	15
	ARS1413-LG	WT	1.2 (0.84-1.9)	26/80 (33%)	0.40
		ogg1	1.8 (1.5-3.0)	101/171 (59%)	1.1
		ogg1 msh6	19 (14-24)	124/164 (76%)	14
	ARS501-LD	WT	1.6 (1.2-2.4)	54/200 (27%)	0.43
		ogg1	3.8 (2.0-4.9)	92/162 (57%)	2.2
		ogg1 msh6	36 (26-63)	97/146 (66%)	24
Late	*ARS609-LG*	WT	1.5 (1.1-1.9)	12/59 (20%)	0.3
		ogg1	1.9 (1.3-2.2)	70/161 (43%)	0.82
		ogg1 msh6	20 (17-31)	138/188 (73%)	15

[a] LG indicates that the transcribed strand of *SUP4-o* is on the lagging strand, and LD indicates that the transcribed strand is on the leading strand

[b] 95% confidence intervals are in parentheses.

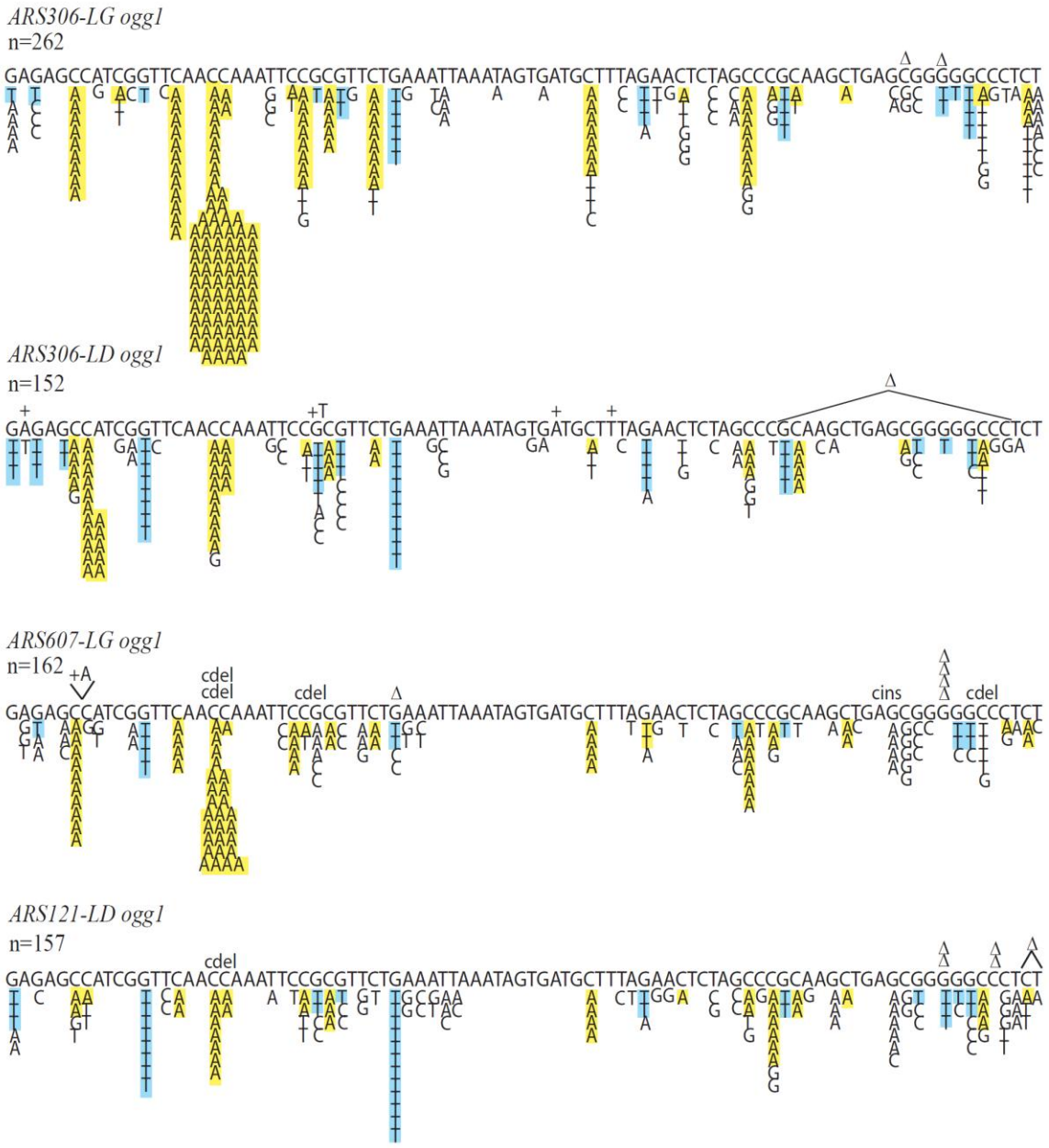

Figure 3.5 Mutation Spectra of Early-firing *ogg1* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

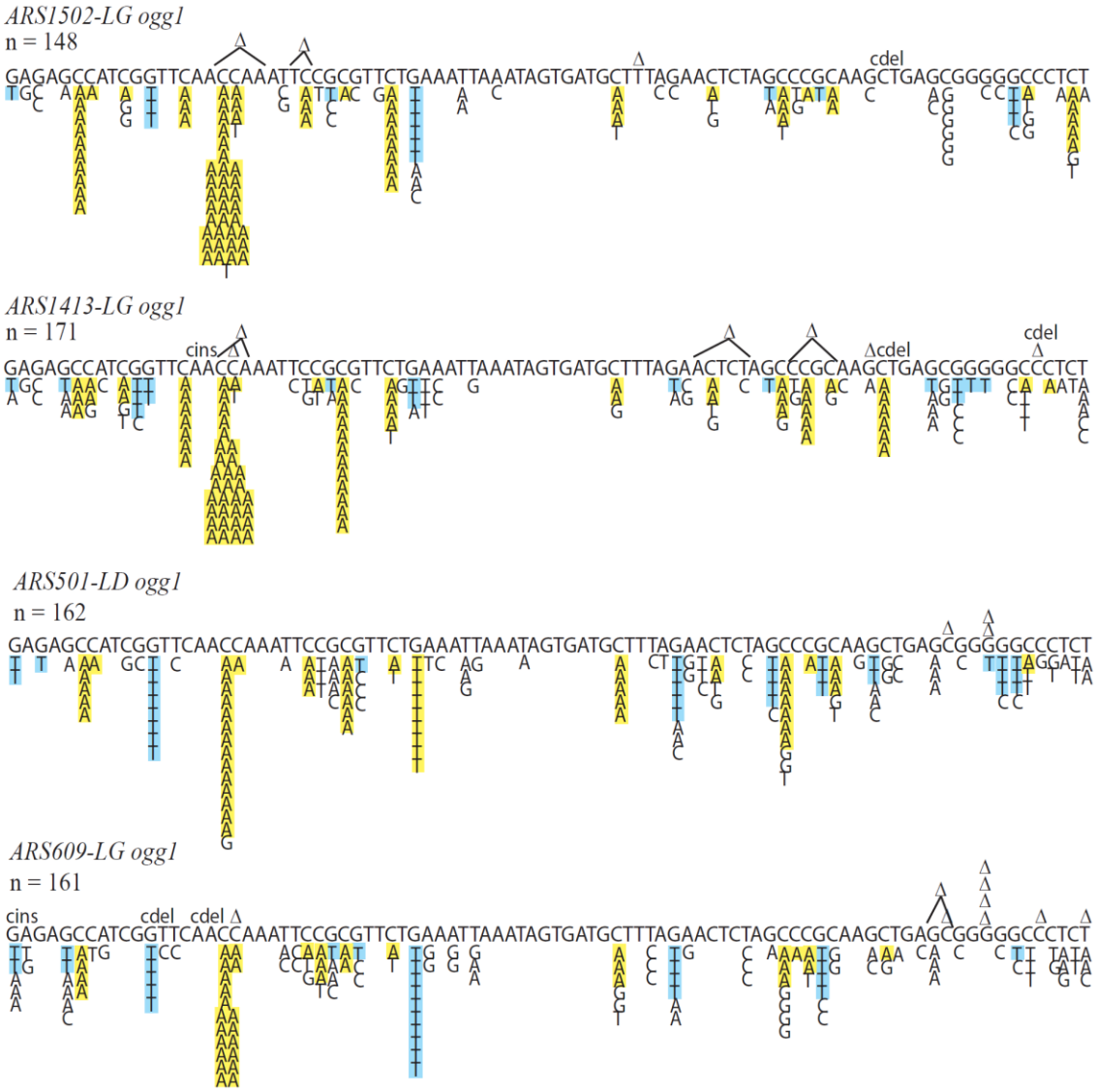

Figure 3.6 Mutation Spectra of Late-firing *ogg1* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

Table 3.4 Rates of G > T and C > A Mutations at Early- and Late-firing Origins

	Origin[a]	Genotype	Total Rate[b] ($x10^{-7}$)	G > T Mutations		C > A Mutations		G > T vs. C > A
				Rate (%)	Fold Increase	Rate (%)	Fold Increase	
Early	*ARS306-LG*	*ogg1*	2.1 (1.7-3.1)	0.21 (9.9%)		1.2 (57%)		p<0.0001
		ogg1 msh6	15 (11-19)	4.5 (30%)	21 (LG)	6.9 (46%)	5.8 (LD)	p<0.0001
	ARS306-LD	*ogg1*	1.4 (1.1-1.7)	0.42 (30%)		0.46 (33%)		p=0.7
		ogg1 msh6	11 (9.2-16)	4.3 (39%)	10 (LD)	3.9 (35%)	8.3 (LG)	p=0.4
	ARS607-LG	*ogg1*	1.9 (1.6-2.7)	0.18 (9.3%)		0.80 (42%)		p<0.0001
		ogg1 msh6	23 (13-44)	5.3 (23%)	30 (LG)	13 (56%)	16 (LD)	p<0.0001
	ARS121-LD	*ogg1*	1.9 (1.5-2.7)	0.45 (24%)		0.47 (25%)		p=0.9
		ogg1 msh6	15 (11-23)	6.3 (42%)	14 (LD)	5.0 (33%)	10 (LG)	p=0.09
	ARS1502-LG	*ogg1*	1.8 (1.0-2.1)	0.19 (11%)		0.96 (53%)		p<0.0001
		ogg1 msh6	20 (15-32)	5.5 (27%)	28 (LG)	9.7 (49%)	10 (LD)	p<0.0001
	ARS1413-LG	*ogg1*	1.8 (1.5-3.0)	0.18 (9.9%)		0.88 (49%)		p<0.0001
		ogg1 msh6	19 (14-24)	3.2 (17%)	18 (LG)	11 (59%)	13 (LD)	p<0.0001
	ARS501-LD	*ogg1*	3.8 (2.0-4.9)	0.99 (26%)		1.2 (31%)		p=0.4
		ogg1 msh6	36 (26-63)	11 (30%)	11 (LD)	13 (36%)	11 (LG)	p=0.3
Late	*ARS609-LG*	*ogg1*	1.9 (1.3-2.2)	0.35 (19%)		0.47 (25%)		p=0.2
		ogg1 msh6	20 (17-31)	6.3 (31%)	18 (LG)	8.4 (42%)	18 (LD)	p=0.04

[a] LG indicates that the transcribed strand of *SUP4-o* is the lagging-strand template, and LD indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

[b] 95% confidence intervals are in parentheses.

SUP4-o (Friedman et al., 1997). The absence of a bias for leading-strand C > A mutations in the *ARS609-LG* strain is consistent with the forks moving in both directions through this region.

As shown in Table 3.3, the mutation rates of the *ogg1 msh6* strains varied significantly. The mutation rate of the *ARS306-LD ogg1 msh6* strain was again the lowest ($11x10^{-7}$ overall and $8.1x10^{-7}$ for GO-associated mutations), and the mutation rate of the *ARS501-LD ogg1 msh6* strain was the highest ($36x10^{-7}$ overall and $24x10^{-7}$ for GO-associated mutations). In general, there appears to be a trend of lower mutation rates for the early-firing strains and higher mutation rates for the late-firing strains. An exception to this is the early-firing *ARS607* strain, which has the second highest mutation rate overall. Because the 95% confidence intervals for overall rates were large, it was difficult to determine the significance of the observed variations. For this reason, we also compared the *ogg1 msh6* mutation rates using the more sensitive Mann-Whitney statistical test, which compares the distributions of mutation rates of each individual culture for each strain. Specifically, the individual mutation rates of at least 12 cultures each for two strains are ranked together from lowest to highest. These rankings are analyzed to determine if one strain is significantly associated with either the lowest or highest rankings. In this way, although two strains may have overlapping outlier cultures, their overall mutation rate distributions may be significantly different. The results of these statistical analyses are shown in Table 3.5, with the strains ordered from earliest to latest origin based on the replication times determined by Raghuraman et al. (2001). For this analysis, *ARS306* indicates the *ARS306-LG* strain. Although significant differences in rates were evident, there did not appear to be a clear correlation between

Table 3.5 Mann-Whitney Analysis According to Replication Timing

		Early →						Late
		ARS306	*ARS607*	*ARS121*	*ARS1502*	*ARS1413*	*ARS501*	*ARS609*
Early	*ARS306*	X	p=0.02	p=0.4	p=0.5	p=0.04	p<0.0001	p=0.0004
↓	*ARS607*		X	p=0.1	p=0.08	p=0.4	p=0.006	p=0.6
	ARS121			X	p=0.9	p=0.3	p<0.0001	p=0.02
	ARS1502				X	p=0.3	p<0.0001	p=0.03
	ARS1413					X	p<0.0001	p=0.2
	ARS501						X	p=0.007
Late	*ARS609*							X

Table 3.6 Mann-Whitney Analysis According to Mutation Rate

		Low →						High
		ARS306	*ARS121*	*ARS1502*	*ARS1413*	*ARS607*	*ARS609*	*ARS501*
Low	*ARS306*	X	p=0.4	p=0.5	p=0.04	p=0.02	p=0.0004	p<0.0001
↓	*ARS121*		X	p=0.9	p=0.3	p=0.1	p=0.02	p<0.0001
	ARS1502			X	p=0.3	p=0.08	p=0.03	p<0.0001
	ARS1413				X	p=0.4	p=0.2	p<0.0001
	ARS607					X	p=0.6	p<0.0001
	ARS609						X	p<0.0001
High	*ARS501*							X

these differences and replication timing. The mutation rate of the *ARS306* strain was significantly lower than that of the *ARS607*, *ARS1413*, *ARS501*, and *ARS609* strains. The mutation rate of the *ARS501* strain was significantly higher than all other strains, and the mutation rate of the *ARS609* strain was significantly higher than the *ARS306*, *ARS121*, and *ARS1502* strains. From this analysis, it appears that the *ARS306*, *ARS121*, and *ARS1502* strains have the lowest mutation rate distributions, the *ARS501* and *ARS609* strains have the highest mutation rate distributions, and the *ARS607* and *ARS1413* strains fall somewhere in between. These groupings are more evident when the results of the analysis are rearranged from lowest to highest mutation rate distributions (Table 3.6).

The spectra for the *ogg1 msh6* early-firing and late-firing strains are presented in Figures 3.7 and 3.8, respectively. As for the *ogg1* strains, all of the "*LG*" strains, with the exception of *ARS609-LG*, have significantly different proportions of G > T and C > A mutations, while the proportions are similar in all of the "*LD*" strains (Table 3.4). Importantly, when we compared the *ogg1 msh6-LG* strains to the corresponding *ogg1-LG* strains, we always observed a larger increase in mutation rate for G > T lagging-strand mutations than for to C > A leading-strand mutations. This confirms our earlier findings that MMR is more efficient on the lagging stand.

Variations in the efficiency of Ogg1 and MMR

In the course of this study, we noticed that the overall proportions of GO-associated mutations varied greatly between the different strains. In the *ogg1* strains, for example, the proportions of GO-associated mutations in the spectra ranged from 43% to 67%. By comparing all of the *ogg1* strains to their corresponding wild-type strains, we

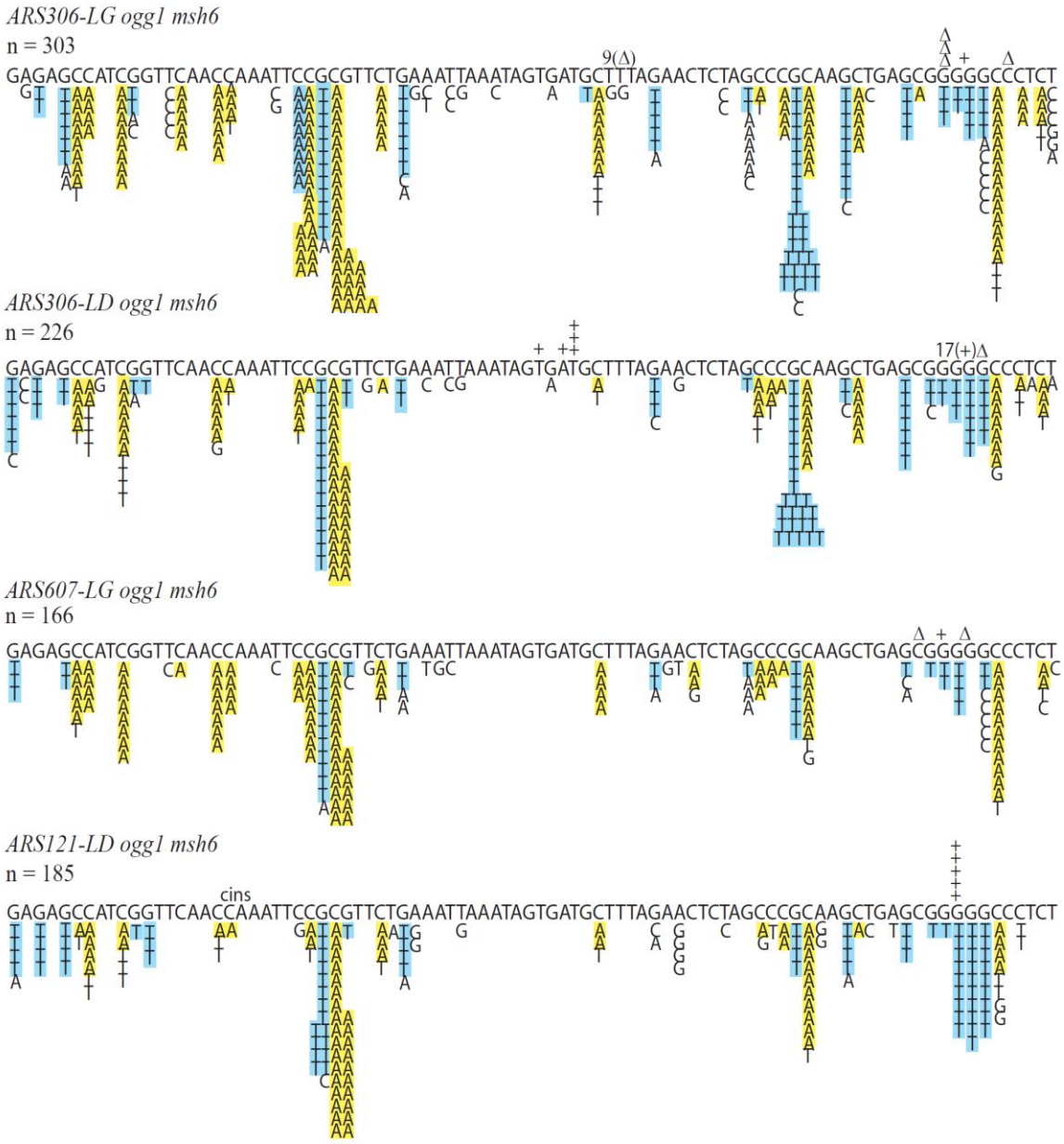

Figure 3.7 Mutation Spectra of Early-firing *ogg1 msh6* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

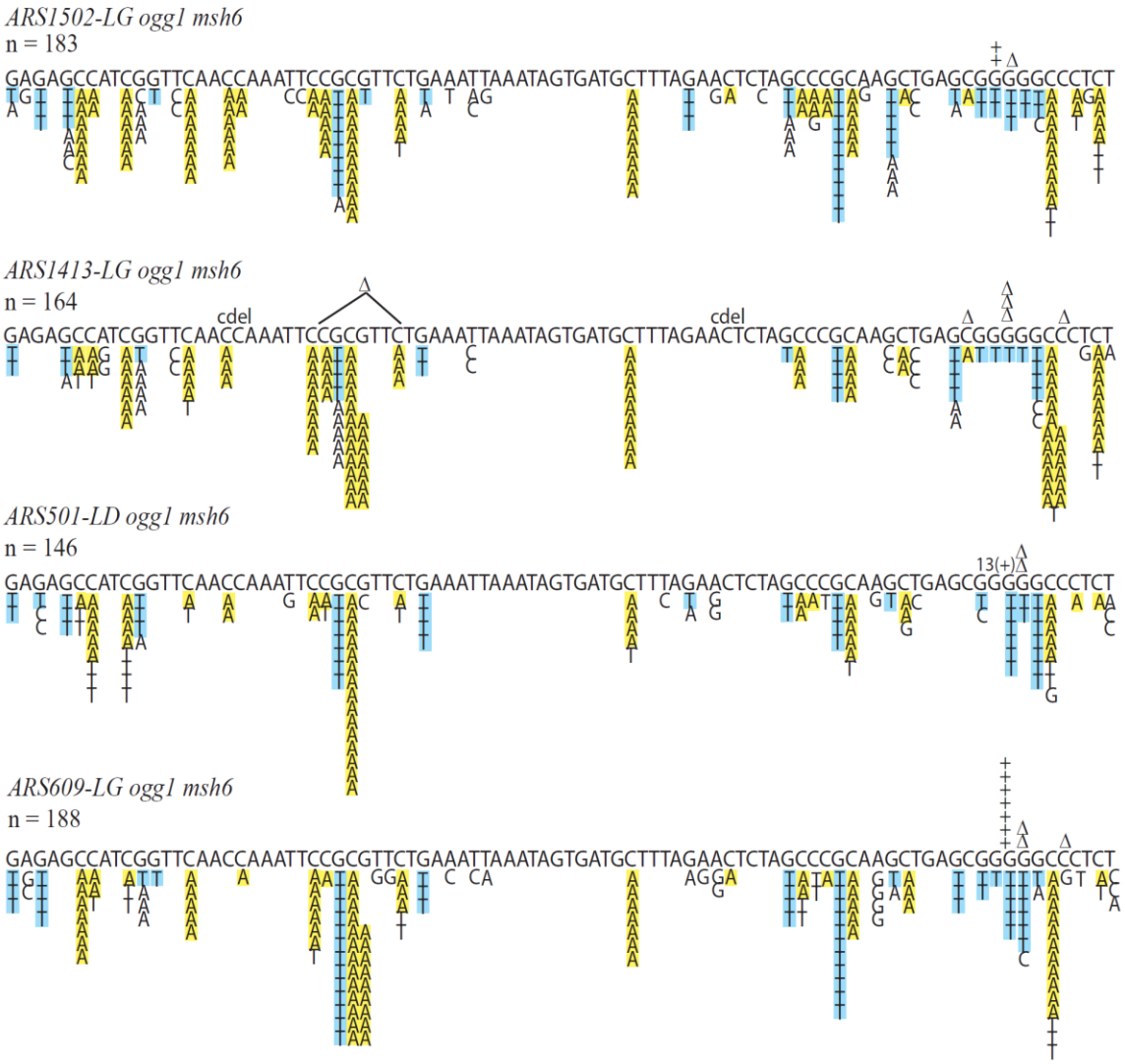

Figure 3.8 Mutation Spectra of Late-firing *ogg1 msh6* Strains

The transcribed sequence of *SUP4-o*, oriented from 3' to 5', is shown. All base substitutions are below the *SUP4-o* sequence, and all insertions and deletions are above the sequence. G > T mutations are highlighted in blue, and C > A mutations are highlighted in yellow. *LG* indicates that the transcribed strand of the *SUP4-o* allele is the lagging-strand template, and *LD* indicates that the transcribed strand of *SUP4-o* is the leading-strand template.

Table 3.7 Variations in Ogg1 and MMR Efficiency

		Ogg1 Efficiency (%)	MMR Efficiency (%)
Early	*ARS306*	69	87
↓	*ARS607*	57	95
↓	*ARS121*	54	92
↓	*ARS1502*	46	93
↓	*ARS1413*	64	92
↓	*ARS501*	80	91
Late	*ARS609*	63	95

were able to calculate the efficiency of Ogg1 at different chromosomal positions. As shown in Table 3.7, the efficiency of Ogg1 ranged from 46% to 80%. This variation does not appear to correlate with replication timing. We also compared the *ogg1 msh6* strains with their corresponding *ogg1* strains to calculate the efficiency of MMR at each origin. The efficiency of MMR varied less than that of Ogg1, from 87% to 95%, and also did not appear to correlate with replication timing.

3.5 Discussion

In the fields of DNA replication and mutagenesis, many groups have examined patterns of mutagenesis in an attempt to uncover how replication timing affects the generation and/or repair of mutations. While some studies have shown increased rates of mutagenesis in late-replicating regions of the genome, other studies have not observed this effect (Deschavanne and Filipski, 1995; Hawk et al., 2005; Ito-Harashima et al., 2002; Lang, 2007; Payen et al., 2009). Although the reason for this is not known, it may be due to the type of mutagenesis analyzed or differences in strain backgrounds. In this study, we have examined the effect of replication timing specifically on the rate of GO-

associated mutagenesis by moving the *SUP4-o* reporter allele near different early- and late-firing replication origins in the yeast genome.

We began this study by examining strains in which the *SUP4-o* allele was integrated near the early-firing origin *ARS306* or the late-firing origin *ARS501*. In *ogg1* strains, we observed slight (2-3 fold) variations in the rate of GO-associated mutations at the different origins. The rates of GO-associated mutagenesis in the *ogg1 rad30* strains were similar, indicating that Polη does not vary in activity across the genome. In contrast, we did observe some subtle, but significant, variations in mutation rate in the *ogg1 msh6* strains. To further examine this variable activity, we constructed and analyzed strains in which the *SUP4-o* allele was integrated near five other origins: early-firing *ARS607* and *ARS121*, and late-firing *ARS1502*, *ARS1413*, and *ARS609* (Feng et al., 2006; Ferguson et al., 1991; Friedman et al., 1997; Friedman et al., 1996; Knott et al., 2009; McCune et al., 2008; Newlon et al., 1993; Nieduszynski et al., 2005; Raghuraman et al., 2001; Yabuki et al., 2002; Yamashita et al., 1997). We again observed variations in mutation rate in the *ogg1 msh6* strains, suggesting that the activity of MMR and the rate of GO-associated mutagenesis in the absence of MMR vary across the genome.

To determine if these variations were correlated with replication timing, we compared these values with the replication times reported by Raghuraman et al. (2001). As shown in Table 3.2, the order of origins from earliest to latest is *ARS306*, *ARS607*, *ARS121*, *ARS1502*, *ARS1413*, *ARS501*, and *ARS609*. The mutation rates of the different strains are therefore presented in this order in Table 3.3. Based on the 95% confidence intervals of the different mutation rates, the mutation rate at the earliest origin, *ARS306*, was significantly lower than the mutation rate at the second latest origin, *ARS501*. It is

important to note that these intervals are for the overall mutation rate rather than the rate of GO-associated mutagenesis specifically. However, the relatively equal proportion of GO-associated mutations in all the *ogg1 msh6* spectra (66% to 79%) suggests that these rates and confidence intervals correspond well with rates of GO-associated mutagenesis. Although the mutation rates at these two origins are significantly different, the mutation rates at the other origins fall somewhere in between these two and are difficult to interpret statistically.

We then employed the Mann-Whitney statistical test, which can detect more subtle differences in mutation rates, to further analyze the rate variations. By Mann-Whitney (Tables 3.5 and 3.6), many of the mutation rate variations, though small, were significantly different. Interestingly, the different origins appear to fall into three groups; *ARS306*, *ARS121,* and *ARS1502* have the lowest mutation rates; *ARS501* and *ARS609* have the highest mutation rates; and *ARS1413* and *ARS607* fall somewhere in between. This suggests either that the mutation rates are not strictly correlated to replication timing or that the replication times of these origins in this strain background are not the same as those observed previously by Raghuraman et al. (2001). We favor the latter hypothesis because the times at which specific origins fire are not always consistent. As shown in Table 3.2, for example, early-firing *ARS607* was found to replicate before early-firing *ARS121* in one study, but was found to replicate after *ARS121* in another study, leading to the conclusion that it was late-firing (Yabuki et al., 2002). In yet another study, *ARS121* was concluded to be late-firing (Feng et al., 2006). In contrast to Raghuraman et al. (2001), two studies found that late-firing *ARS1413* replicates later, rather than before, late-firing *ARS501* (Donaldson et al., 1998; Yabuki et al., 2002). Finally, late-firing

ARS1502 was found to be regulated by one late origin-specific protein (Clb5) but not another (Rpd3), and late-firing *ARS609* does not appear to be regulated by either late origin-specific protein (Knott et al., 2009; McCune et al., 2008). Interestingly, some of these conflicting studies used the same parental yeast strains, indicating that discrepancies are likely due to differences in both assays and yeast strains.

Additional preliminary experiments also support the idea that the replication times of these origins may be different than those previously observed. As mentioned above, it has been suggested that dNTPs are elevated in late S phase and contribute to increased levels of mutagenesis. To test this idea, we inserted a copy of *RNR1* linked to an inducible *GAL1* promoter into the *ARS306-LG ogg1 msh6* strain. *RNR1* encodes one of the large subunits of ribonucleotide-diphosphate reductase, which catalyzes the rate-limiting step of dNTP synthesis. In the presence of galactose, expression of this *RNR1* allele, and thus the synthesis of dNTPs, is increased (Chabes and Stillman, 2007). The mutation rate of the *ARS306-LG ogg1 msh6 pGAL-RNR1* strain was significantly increased relative to the parental strain ($49x10^{-7}$, 95% confidence interval (CI): 44-55 vs. $26x10^{-7}$, 95% CI: 21-37). This small but significant difference in mutation rate is similar to the difference in mutation rate between the *ARS501* and *ARS306* strains.

Together, these results suggest that it is important that we determine the replication times of the *SUP4-o* alleles in our strains. This can be done in a variety of ways, including quantitative PCR, two-dimensional gel electrophoresis, and quantifying incorporation of BrdU at different sites. If the variations in mutation rate in our strains do reflect differences in replication timing, we expect the regions near the *ARS306*, *ARS121*, and *ARS1502* origins to replicate relatively early, the regions near the *ARS501*

and *ARS609* origins to replicate relatively late, and the regions near the *ARS1413* and *ARS607* origins to replicate sometime in between the two other groups. If we do not observe the expected replication times, this would suggest that the variations in mutation rate are not due to replication timing but to some other difference(s) in the origins or in their specific genomic context. This may include differences in replication fork dynamics, such as efficiency and rate of fork movement, the localization of the origins relative to telomeres, centromeres, genes, or other origins, or variations in nucleosome positioning and histone modifications.

Replication efficiency is unlikely to be a factor in this study, as each of these origins, except *ARS609* and *ARS1502*, have been shown to be very efficient (Ferguson et al., 1991; Friedman et al., 1997; Friedman et al., 1996; Nieduszynski et al., 2005; Poloumienko et al., 2001; Yamashita et al., 1997). Although it has been suggested that fork rates vary significantly across the genome (Raghuraman et al., 2001), the specific fork rates of different origins have not been yet characterized. As slowed replication has been shown to result in increased rates of mutagenesis and recombination (Lemoine et al., 2008; Palakodeti et al., 2010; Sabouri et al., 2008), it is possible that variations in fork rates play a role in the mutation rate variations observed in this study.

In an attempt to identify differences in the genomic locations of these origins, we have compared each location in terms of distance to telomeres and centromeres, the orientation of neighboring genes, and nearest origins. Interestingly, the two latest origins, which also had higher mutation rates, were closest to telomeres (15 kb for *ARS609* and 25 kb for *ARS501*). Aside from these two origins, however, we did not observe any correlations between mutation rate and distance to telomeres. Similarly, we did not

observe any correlations between mutation rate and distance to centromeres. For example, the two origins closest to centromeres, *ARS1413* (38 kb) and *ARS306* (39 kb), and the two origins furthest from centromeres, *ARS501* (400 kb) and *ARS1502* (275 kb), each had significantly different mutation rates. We also did not find any correlations between neighboring gene orientation and mutation rate; the neighboring genes of some origins were directed toward the origin (converged), some were directed away from the origin (diverged), and others were oriented in the same direction (one towards the origin and one away).

As it is formally possible that the insertion of *SUP4-o* ~150-300 bp away from an origin could impair or decrease the activity of that origin, we also identified the nearest neighboring origins of each location. However, almost all origins chosen in this study are in areas of similarly timed origins, i.e., the nearest origins to *ARS306* are also early-firing, and the nearest origins to *ARS501* are also late-firing. One exception to this is the early origin *ARS121*, as the nearest origins either upstream or downstream are mid- or late-firing. However, the lower mutation rate near *ARS121* may indicate that this is still an "earlier" origin even with the insertion of *SUP4-o*. Another possible exception to this is the early origin *ARS607*, as its nearest origin has been characterizing as being mid- to late-firing. However, as this neighboring origin has also been shown to be inactive in most cells, it is not clear if this origin would act in place of *ARS607*, and the next closest origin has been shown to fire at a similar time to *ARS607* (Friedman et al., 1997; Yamashita et al., 1997). Finally, even though these origins are neighbored by similarly timed origins, it is possible that impairment of these origins results in these regions replicating later than normal due to the increased distance between these regions and an

active origin. It is, therefore, important that we test the replication times of *SUP4-o* in each our strains.

As mentioned above, nucleosomes and some types of histone modifications have been shown to prevent different DNA repair proteins from accessing DNA, and may, therefore, have a role in the variations in mutation rate observed in this study. Although different groups are beginning to map nucleosome positions across the genome, there are clear differences between studies. While some nucleosomes are static, others are very dynamic and can change positions based on the cellular environment (Feng et al., 2010). Furthermore, recent studies suggest that origins tend to be in nucleosome-depleted regions and that increased amounts of nucleosomes result in decreased origin efficiency (Field et al., 2008; MacAlpine et al., 2010; Yin et al., 2009). It would thus be interesting to determine the nucleosome positioning around the origins studied here and whether or not the insertion of *SUP4-o* affects this positioning.

Nucleosomes are composed of four types of histones, which can each be modified by methyl, acetyl, phosphoryl, ubiquityl, and sumo groups at various locations. Increased and decreased levels of some types of acetylation and methylation are often associated with euchromatic and heterochromatic DNA, respectively. However, not all euchromatic and heterochromatic regions have the same patterns of histone modifications, and not all types of chromatin are either euchromatic or heterochromatic (reviewed in Millar and Grunstein, 2006). Although different groups have begun to map different types of histone modification across the genome, the complete genomic landscape of these variations in chromatin and how they contribute to mutagenesis are not yet clear. With

the information available to date, we have not found any correlations between known histone modification sites and the origins included in this study.

Aside from replication timing, differences in nucleosome positioning and histone modifications likely affect the variations in efficiency of Ogg1 and MMR observed in this study. These variations did not correlate with any of the published replication times of the origins or the variations in mutation rate. Ogg1 has been shown to be excluded from regions of heterochromatin in mammalian cells (Amouroux et al., 2010), and chromatin remodeling has been shown to facilitate Ogg1 activity in some chromatin contexts but not others (Menoni et al., 2007). Nucleosomes have also been shown to inhibit MMR (Li et al., 2009), yet another recent study showed that human MMR proteins were able to move nucleosomes (Javaid et al., 2009). It is thus likely that a combination of different chromatin factors affects the efficiency of both of these repair pathways. It is important to note that despite the overall variation in MMR activity, MMR was consistently more efficient on the lagging strand relative to the leading strand in strains in which we could detect differences between lagging- and leading-strand mutagenesis (i.e., *LG* strains).

The past decade has seen huge advances in the fields of DNA replication and mutagenesis, but it is clear that we are only beginning to understand how these two processes are interrelated. Although yeast has proven to be an excellent model system for studying both of these processes, the relationship between replication timing and mutagenesis is the least clear in this species. It has been suggested that overall mutation rate variations across the genome are much less pronounced in yeast relative to other systems (Fox et al., 2008), and this may explain why it has been difficult to obtain clear,

statistically significant results both here and in previous studies. Moreover, it is important to note that although some studies in other species have revealed a clear affect of replication timing on mutagenesis, these effects have been small overall (Stamatoyannopoulos et al., 2009). Additional experiments with the *SUP4-o* system may clarify if and how replication timing affects GO-associated mutagenesis in yeast.

3.6 Acknowledgements

This work was supported by NIH grants (GM038464 and GM064769) awarded to S.J.-R.

Chapter 4: Examination of frameshift mutagenesis in short runs and noniterated sequences

4.1 Summary

Small insertions or deletions that shift the reading frame of a gene are referred to as frameshifts. These events almost always result in the loss of function of the gene in which they are localized and occur most often in long homopolymer runs and other repeated sequences, such as dinucleotide repeats. Because longer nucleotide runs and repeat sequences are strong hotspots for frameshift mutagenesis, most of what we know about this type of mutagenesis has been focused on these sequences. Consequently, although frameshift mutations do occur at shorter nucleotide runs and noniterated sequences, little is known about how these mutations occur and which mechanisms act to prevent or repair them. By generating a modified version of a commonly used frameshift reversion allele in *Saccharomyces cerevisiae*, *lys2ΔBgl*, in which all runs ≥ 4N were deleted, we have successfully generated an in vivo assay in which to study frameshifts at short runs and noniterated sequences and the mechanisms that are involved in their generation and repair. In the *lys2ΔBgl,NR* (no run) strain, 72% of reversion events were simple -1 events at short, 3N and 2N runs and noniterated sequences. In the absence of mismatch repair (MMR), the overall rate of simple -1 events at short runs and noniterated sequences was increased 25-fold, and specific hotspots were increased up to 1000-fold. By comparing these results with those generated using a comparable +1 frameshift reversion assay (*lys2ΔA746,NR*), we find that -1 and +1 frameshifts at short runs and

noniterated sequences have different sequence hotspots and are not equivalent substrates for MMR.

4.2 Introduction

The two most common types of mutations are simple base substitutions and small insertions or deletions. A base substitution within a gene will often not have any significant effect, as it may not affect the specific amino acid where it is localized (i.e., a synonymous mutation) or the one mutant amino acid will not affect the protein as a whole. In contrast, insertions and deletions that are not a multiple of three nucleotides, which are referred to as frameshift mutations, are almost always detrimental because they shift the reading frame of the gene in which they are localized. These frameshifts cause all amino acids downstream to be read incorrectly and can create premature stop codons, resulting in either a significantly modified or truncated protein, respectively. Given the deleterious nature of frameshifts, it is critical that the cell recognizes and removes the corresponding mutational intermediates.

Frameshifts usually occur in repetitive sequences, such as homopolymer runs and dinucleotide repeats, and most are thought to occur as a result of polymerase slippage during replication, either spontaneously or as a result of DNA damage (Figure 4.1a and reviewed in Garcia-Diaz and Kunkel, 2006). When slippage occurs and the frameshift intermediate is on the primer strand, a +1 frameshift mutation will be generated in the next round of replication. Conversely, when the frameshift intermediate is on the template strand, a -1 mutation will be generated in the next round of replication. The proofreading activity of replicative DNA polymerases is often able to detect these

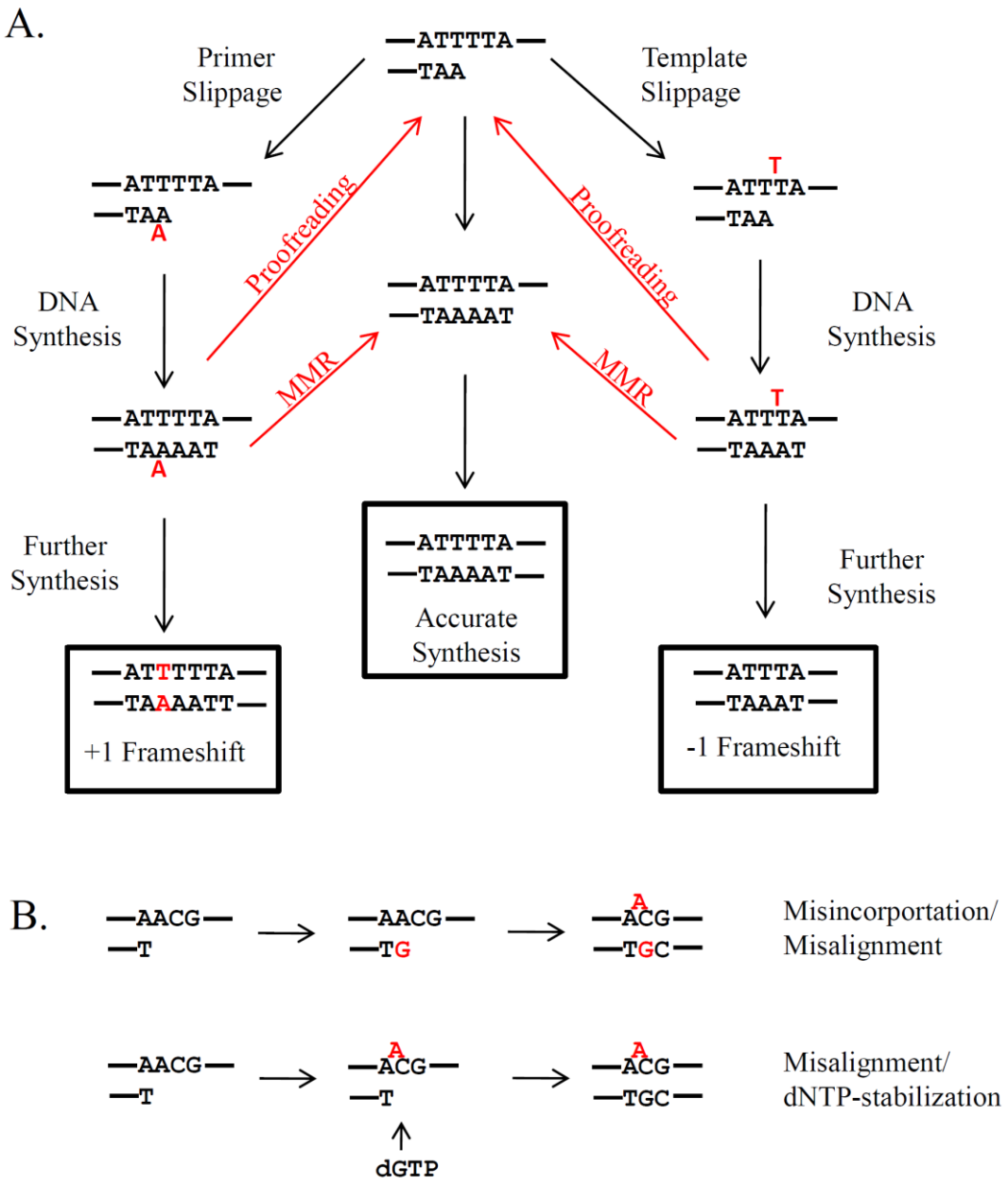

Figure 4.1 Mechanisms of Frameshift Mutagenesis

(A) Polymerase slippage of homopolymer runs ≥ 4N (modified from Garcia-Diaz and Kunkel, 2006). Slippage on the primer strand results in +1 frameshifts, and slippage on the template strand results in -1 frameshifts. Frameshift errors can be removed by the proofreading activity of DNA polymerase, and errors that escape proofreading can be removed after replication by MMR. (B) Mechanisms of frameshift mutagenesis at short runs and noniterated sequences. In the first model, misincorporation of an incorrect nucleotide results in strand misalignment to permit base pairing of the misincorporated 3' nucleotide with the template. In the second model, strand misalignment is stabilized by incorporation of a dNTP that base pairs with the next nucleotide.

frameshift intermediates as they arise and can remove them using their associated 3' to 5' exonuclease activity. This proofreading activity has been shown to be efficient in runs of three or less nucleotides, but is increasingly less efficient in longer runs (Kroutil et al., 1996). Furthermore, removal of the exonuclease activity of Polε and Polδ in yeast results in an increase in frameshift mutations at runs less than three nucleotides long (< 3N), with the majority of mutations occurring at noniterated (1N) sequences (Greene and Jinks-Robertson, 2001). Thus, frameshift mutations in runs ≥ 4N are more likely to escape proofreading, and the rate of mutations increases as the length of the homopolymer run increases (reviewed in Garcia-Diaz and Kunkel, 2006). Aside from the decreased efficiency of proofreading, misaligned intermediates in longer runs can also be stabilized by a correctly paired primer terminus, which enables efficient extension by DNA polymerase and continued replication.

Although frameshifts occur more frequently in runs ≥ 4N and other repetitive sequences, they also occur at shorter runs and noniterated sequences at low levels. In vitro experiments have shown that some of these frameshifts reflect replication errors caused by replicative DNA polymerases (Kunkel, 1985). Low fidelity polymerases such as human Polκ, yeast Polη, and *Sulfolobus solfataricus* Dpo4 have higher rates of frameshift mutagenesis at these sequences relative to replicative DNA polymerases (Gu and Wang, 2007; Kokoska et al., 2002; Ohashi et al., 2000). These polymerases generate frameshifts at even higher rates when they are replicating past certain types of DNA damage, such as abasic sites and the oxidative lesion 1,N^2-ethenoguanine (Choi et al., 2006; Kokoska et al., 2003). Crystallography studies have generated structures of polymerases with these different types of DNA damage and have shown that the presence

of a large active site in low fidelity polymerases facilitates misalignment of the DNA strands, which results in frameshift mutations (Garcia-Diaz et al., 2006; Ling et al., 2001).

Because polymerase slippage does not readily explain frameshift mutagenesis at short runs and noniterated sequences, two models have been proposed to explain the generation of these mutations (Figure 4.1b). In the first model, misincorporation of an incorrect nucleotide during DNA replication may trigger a subsequent misalignment of either the template or primer strand. The second model is similar to the first, but the order of events is switched; a misaligned base can be stabilized by incorporation of a dNTP that is complementary to the next base. In vitro studies have shown that polymerases do in fact generate frameshifts via the mechanisms described in these models, with different polymerases using different mechanisms in specific contexts (Tippin et al., 2004). Moreover, the occurrence of frameshifts at specific sites can be increased by altering the concentrations of specific dNTPs, which supports the idea that misincorporated dNTPs can stabilize frameshift intermediates (Bebenek et al., 1992). Although DNA polymerase proofreading is thought to be very efficient at detecting and removing these frameshift intermediates, a low level of frameshifts appears to escape this activity. It is currently unclear how this occurs.

As shown in Figure 4.1a, mismatch repair (MMR) also plays a role in preventing frameshift mutagenesis. Both the MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3) complexes have been shown to detect small, 1-2-bp frameshift intermediates, and MutSβ also detects larger (>2 bp) frameshift intermediates (Kunkel and Erie, 2005). These complexes interact with a MutL complex and signal for the removal of the intermediates

prior to the next round of replication. The Mlh1/Pms1 MutL complex appears to play the major role in suppressing frameshift mutagenesis in yeast, and mutations in either one of these proteins are phenotypically equivalent to mutations in Msh2 (Greene and Jinks-Robertson, 1997; Harfe and Jinks-Robertson, 1999). Mlh1 also interacts with Mlh2 and Mlh3, and these complexes have been shown to have minor roles in suppressing some types of frameshift mutagenesis (Flores-Rozas and Kolodner, 1998; Harfe et al., 2000). Mutations in MMR proteins have been shown to cause increased rates of frameshift mutagenesis in runs $\geq$ 4N, with mutation rates increasing further as the length of the run increases (Tran et al., 1997). Deficiencies in MMR are also associated with instability of microsatellites, which are regions of short, 1-6-bp repeats (Chen et al., 2005; Lipkin et al., 2000; Strand et al., 1995; Wierdl et al., 1997). In humans, mutations in MMR proteins are associated with Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC), and microsatellite instability is often used to diagnose this condition (reviewed in Shah et al., 2010).

Although MMR is clearly involved in the repair of frameshift intermediates in runs $\geq$ 4N, less is known about the activity of MMR at short runs and noniterated sequences. This is partially due to the low level, or even absence, of mutations at these sites, which hinders proper examination and analysis. However, the results of one study suggested that MMR was required to generate frameshifts at 2N runs and noniterated sequences, as the rate of these mutations was significantly decreased upon removal of MMR (Greene and Jinks-Robertson, 2001). Interestingly, recent work in our lab has shown that MMR may promote error-prone translesion synthesis (TLS) by suppressing

homologous recombination, thereby indirectly stimulating mutagenesis (Lehner and Jinks-Robertson, 2009).

As mentioned above, frameshifts at short runs and noniterated sites occur at low levels and are thus difficult to study. In vitro and crystallography experiments have examined frameshifts at short runs with the use of low fidelity polymerases and induced DNA damage, but little is known about frameshifts that occur spontaneously in the cell by normal, replicative DNA polymerases in the absence of exogenous DNA damage. In this study, we have generated a yeast frameshift reversion assay (*lys2ΔBgl,NR*) that can detect any compensatory -1 frameshift events within a ~150-bp window of the *LYS2* gene that does not contain any runs ≥ 4N. Using this assay, we have obtained mutation spectra in which the majority of mutations are frameshifts at short runs (< 3N) and noniterated sequences. Surprisingly, these mutation spectra also have a significant proportion of +2 frameshift events. We have been able to use this system to study the involvement of MMR in the repair of these frameshifts intermediates and the mechanisms that are involved in the generation of +2 events. We also created a frameshift reversion assay (*lys2ΔA746,NR*) that detects +1 events in the same ~150-bp window. By comparing the mutation spectra from these two assays, we have found that -1 and +1 frameshifts have different mutation hotspots and are repaired by MMR with varying efficiencies.

4.3 Materials and Methods

Strain Construction

All strains are isogenic derivatives of SJR195 (*MATα ade2-101 his3Δ200 ura3ΔNco*). The *lys2ΔBgl* allele was generated by digesting the *LYS2* allele with *Bgl*II and filling in the resulting ends with the Klenow fragment of *Escherichia coli* DNA polymerase I. This generates a direct duplication of GATC, which is equivalent to a +1 frameshift event (Steele and Jinks-Robertson, 1992). The *lys2ΔA746* allele was created by deleting the adenine at nucleotide 746 of the *LYS2* gene (nucleotides are numbered relative to the upstream *Xba*I site). Three other sites (T753, A767, and T781) were mutated to remove relevant stop codons in the *lys2ΔA746* allele (Harfe and Jinks-Robertson, 1999). Both the *lys2ΔBgl* and *lys2ΔA746* alleles contain coincident ~150-bp reversion windows that can be used to detect compensatory -1 and +1 frameshifts, respectively.

The *lys2ΔBgl,NR* (no run) and *lys2ΔA746,NR* alleles were created using site-directed mutagenesis of plasmids pSR699 (*lys2ΔBgl*; Steele and Jinks-Robertson, 1992) and pSR585 (*lys2ΔA746*; Harfe and Jinks-Robertson, 1999), respectively. The primers used to disrupt the 6A run were 5’ GCTAGCTG*AA****TC****AA*TTCAAAG and 5’ CTTTGA A*TT****GA****TT*CAGCTAGC. The primers used to disrupt the 5T and 4A runs were 5’ CGT TTGGCC*T****GTCT***GG*A****T****A****T***CCAAGATTTC and 5’ GAAATCTTGG***ATAT***CC*A****G****A****C****A* GGCCAAACG. The primers used to disrupt the 4C run were 5’ GGAAAGGA***GGCC***T CAGTTG and 5’ CAACTGA*GG****CC***TCCTTTCC. The italicized bases indicates the runs, and the bold bases indicate the sites mutated. The resulting plasmids containing the *lys2ΔBgl,NR* (pSR701) and *lys2ΔA726,NR* (pSR700) alleles were introduced into strain SJR195 using a two-step allele replacement technique (Rothstein, 1991), generating

strains SJR1468 and SJR1467, respectively. Integration of mutant alleles was confirmed by sequencing.

The *MSH2*, *MSH3*, and *MSH6* genes in SJR1468 and the *MSH2* gene in SJR1467 were disrupted using a *hisG-URA3-hisG* cassette as described previously (Greene and Jinks-Robertson, 1997). *rad14Δ*, *mlh2Δ*, and *mlh3Δ* strains were generated by transforming SJR1468 with a PCR-generated fragment containing a *URA3-Kl* marker (*URA3* gene from *Kluveromyces lactis*) (Gueldener et al., 2002) with the appropriate flanking sequence of the target gene. Similarly, *rev3Δ* strains were generated using a PCR-generated fragment containing a hygromycin resistance marker (Goldstein and McCusker, 1999) and *REV3* flanking sequence. *msh2Δ::hisG-URA3-hisG*, *msh3Δ::hisG-URA3-hisG*, *msh6Δ::hisG-URA3-hisG*, *rad14Δ::URA3-Kl*, *mlh2Δ::URA3-Kl*, and *mlh3Δ::URA3-Kl* transformants were selected on synthetic complete medium containing 2% dextrose and lacking uracil (SCD-URA). *rev3Δ::hyg* transformants were selected on YEPD medium (1% yeast extract, 2% Bacto-peptone, 2% dextrose, and 250 mg/L adenine) containing 300 μg/mL hygromycin. All deletions were confirmed by PCR.

Mutation Rate Analysis

To determine mutation rates, 4-5 individual colonies were used to inoculate a 5 mL starter culture. Following overnight growth at 30°C, the starter culture was used to inoculate independent 5 mL secondary cultures to a concentration of 2.5×10^5 cells/mL. Two isolates were used for each strain, and at least six cultures were used for each isolate. These cultures were grown for 3 days at 30°C. Non-selective YEPGE medium (1% yeast extract, 2% Bacto-peptone, 2% glycerol, 2% ethanol, and 250 mg/L adenine) was used for both starter and secondary cultures. Appropriate dilutions of each culture

were plated onto YEPGE medium to determine total cell number and onto SCD-LYS plates to select Lys^+ revertants. Mutation rates were determined using the method of the median (Lea and Coulson, 1949), and 95% confidence intervals were calculated as previously described (Spell and Jinks-Robertson, 2004). Mutation rates for specific mutation types were calculated by multiplying the proportion of that event in the corresponding spectrum by the total mutation rate.

Mutation Spectra

To generate mutation spectra, DNA was extracted from purified Lys^+ colonies (http://jinks-robertsonlab.duhs.duke.edu/protocols/yeast_prep.html). An appropriate portion of the *LYS2* gene was amplified by PCR and sequenced using the MO18 sequencing primer (5' GTAACCGGTGACGATGAT). Sequencing was performed by the Duke University DNA Analysis Facility (Durham, North Carolina). Proportions of mutations in different spectra were analyzed using the Fisher's exact test or contingency Chi Square analysis (http://faculty.vassar.edu/lowry/VassarStats.html). A p-value of less than 0.05 was considered statistically significant.

4.4 Results

Frameshift mutagenesis has been studied in vivo using various forward mutation and reversion assays. However, forward mutation spectra are composed of low levels of frameshifts, and many reversion assays utilize small reversion windows or specific sites (Giroux et al., 1988; Henderson and Petes, 1992; Kalinowski et al., 1995; Lee et al., 1988). In our lab, we have developed two frameshift reversion assays, *lys2ΔBgl* and *lys2ΔA746*, that can detect any compensatory -1 or +1 frameshift events, respectively,

within an approximately 150-bp window within the *LYS2* gene (Greene and Jinks-Robertson, 1997; Harfe and Jinks-Robertson, 1999; Kim and Jinks-Robertson, 2009; Minesinger and Jinks-Robertson, 2005). The *lys2ΔBgl* allele contains a 4-bp insertion, which is the equivalent of a +1 frameshift and results in a Lys^- phenotype (see Materials and Methods). A -1 (or equivalent) frameshift within the reversion window will restore the correct reading frame and result in a Lys^+ phenotype, which can be selected on medium lacking lysine. The *lys2ΔA746* allele contains a deletion of the adenine at position 746, resulting in a -1 frameshift and a Lys^- phenotype. A compensatory +1 frameshift will restore the reading frame and result in a Lys^+ phenotype.

These assays have been used to study patterns of frameshift mutagenesis and the mechanisms that act to prevent and repair frameshift intermediates. In these studies, frameshifts are primarily found in runs ≥ 4N; in the *lys2ΔBgl* assay, -1 frameshifts at these runs comprise 57% of the wild-type (WT) spectrum and 98% of the MMR-defective *msh2* spectrum (shown in Figure 4.2; Greene and Jinks-Robertson, 1997). Similarly, +1 frameshifts at these sites comprise 75% of the WT spectrum and 99% of the *msh2* spectrum in the *lys2ΔA746* assay (Harfe and Jinks-Robertson, 1999). As the proportions of frameshifts at 3N, 2N, and noniterated sequences were equivalent or less than expected based on their proportions of the reversion window, it was concluded from these studies that these sites are not hotspots for frameshift mutagenesis (Greene and Jinks-Robertson, 1997; Harfe and Jinks-Robertson, 1999). However, frameshifts do occur at low levels at these sites, and it is possible that the presence of runs ≥ 4N obscures these events. To test this hypothesis, we used site-directed mutagenesis to

remove all runs ≥ 4N in the *lys2ΔBgl* and *lys2ΔA746* alleles, generating the *lys2ΔBgl, "no run"* (*NR*) and *lys2ΔA746,NR* alleles.

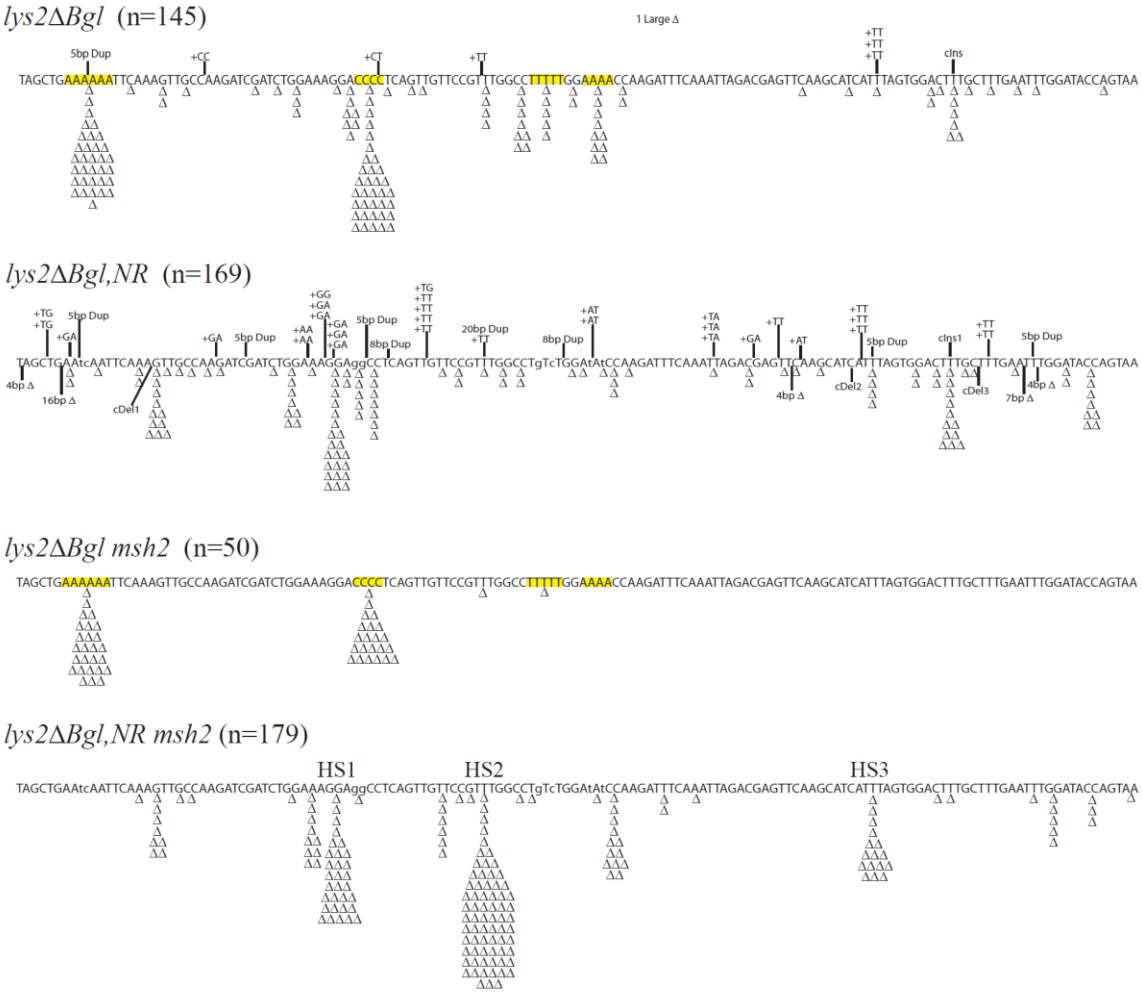

Figure 4.2 Mutation Spectra of *lys2ΔBgl* and *lys2ΔBgl,NR* WT and *msh2* Strains

All runs ≥ 4N in the *lys2ΔBgl* WT and *msh2* spectra are highlighted in yellow. Simple -1 frameshifts are indicated with a Δ symbol. "cIns" indicates a complex insertion, which is defined as an insertion associated with a base substitution within 5 bp, and "cDel" indicates a complex deletion, which is defined as a deletion associated with a base substitution within 5 bp. "Dup" indicates a duplication. HS1-3 indicates hotspots 1-3. The mutation spectra for the *lys2ΔBgl* WT and *msh2* strains were recreated with permission from Greene and Jinks-Robertson, 1997.

The *lys2ΔBgl,NR* allele reveals increased proportions of different classes of mutations and new hotspots for frameshift mutagenesis.

In the *lys2ΔBgl* strain, -1 events at the 6A and 4C runs dominated the mutation spectrum, and only 45% of -1 mutations occurred at short runs or noniterated sequences (Figure 4.2). In the *lys2ΔBgl,NR* strain, however, 72% ($p<0.0001$) of mutations were -1 events at short runs or noniterated sequences. Unexpectedly, the longest runs in the *lys2ΔBgl,NR* reversion window (3N runs) were not the only hotspots for frameshifts. Although there are nine 3N runs in the *lys2ΔBgl,NR* reversion window, the hottest position for frameshifts was localized at a 2N run of guanines (Figure 4.2). Frameshifts at this site accounted for 13% of the mutation spectrum, which corresponds to a 6.8-fold increase in mutation rate relative to the *lys2ΔBgl* strain. As these events were not present in the *lys2ΔBgl* strain, it is possible that by mutating the adjacent 4C run to create the *lys2ΔBgl,NR* allele, we indirectly created a new hotspot in this strain. Interestingly, of the mutations that occurred at short runs and noniterated sequences, the majority (60%) in the *lys2ΔBgl,NR* strain occurred at 2N runs (Table 4.1). Based on the expected proportions of 3N, 2N, and noniterated events, which were determined according to the proportions of these sequences within the reversion window, the number of -1 events at 2N runs was significantly increased relative to the expected number of events ($p<0.0002$; Table 4.2). This increased proportion can be accounted for by the hotspot at the 2N run of guanines mentioned above, as the proportion of observed events is not significantly different from the proportion of expected events when the events at this hotspot are excluded ($p=0.4$). The number of -1 events at 3N runs was not different from the

Table 4.1 Rates of Different Classes of Mutations in the *lys2ΔBgl* and *lys2ΔBgl,NR* WT and *msh2* Strains

		Rates of Different Classes of Mutations[b]								
		Simple Δ								
Genotype[a]	Total Rate ($x10^{-11}$)	Total	> 3N	3N	2N	1N	Simple +2	Large Δ	Large +	Complex
WT	280[c]	260 (135/145)	140 (74/145)	34 (18/145)	39 (20/145)	45 (23/145)	13 (7/145)	1.9 (1/145)	1.9 (1/145)	1.9 (1/145)
WT,NR	99 (76-160)	71 (121/169)	ND (0/145)	12 (20/169)	43 (73/169)	16 (28/169)	18 (31/169)	3.0 (5/169)	4.7 (8/169)	2.3 (4/169)
msh2	54,000[c]	54,000 (50/50)	53000 (49/50)	1100 (1/50)	ND (0/50)	ND (0/50)	ND (0/50)	ND (0/50)	ND (0/50)	ND (0/50)
msh2,NR	1900 (1600-3200)	1900 (172/179)	ND (0/179)	1100 (102/179)	680 (64/179)	140 (13/179)	ND (0/179)	ND (0/179)	ND (0/179)	ND (0/179)

[a] All strains contain the *lys2ΔBgl* allele.

[b] 95% confidence intervals of total mutation rates are indicated in parentheses. For different classes of mutations, numbers in parentheses indicate the proportion of events in the mutation spectra. ND indicates none detected.

[c] These mutation rates were obtained from Greene and Jinks-Robertson, 1997.

expected number (p=0.9), and the number of -1 events at noniterated sequences was significantly decreased (p<0.0003; Table 4.2).

Table 4.2 Proportions of Frameshifts at Different Sequences in the *lys2ΔBgl,NR* WT and *msh2* Strains

		lys2ΔBgl,NR		*lys2ΔBgl,NR msh2*	
Type of Run	Proportion of Window	No. of Expected Events	No. of Observed Events	No. of Expected Events	No. of Observed Events
3N (n=9)	27/146	22	20	33	102[a]
2N (n=26)	52/146	43	73[a]	64	64
Noniterated	67/146	56	28[a]	82	13[a]
	Total	121	121	179	179

[a] The difference between expected and observed in significant (p<0.0001).

We also observed a significant increase in the proportion of +2 mutations in the *lys2ΔBgl,NR* strain relative to the *lysΔBgl* strain (p=0.0005; Table 4.1), but note that the rates of these mutations were not different between the two strains. These events will be discussed further below. The proportions of large insertions and deletions and complex events, which are defined as frameshifts associated with a base substitution, were very small and did not appear to be different in the two WT strains. In summary, these spectra demonstrate that the *lys2ΔBgl,NR* allele will enable us to specifically study frameshifts at short runs and noniterated sequences.

Contributions of MMR proteins in the repair of frameshift intermediates.

To examine the relative involvement of different MMR proteins in the repair of frameshift intermediates at short runs and noniterated sequences, the mutation rates and

spectra of several different MMR-defective strains were analyzed. Deletion of *MSH2* effectively disables all MMR, as it is required for both the MutSα and MutSβ error recognition complexes. As shown in Figure 4.2 and Table 4.1, the *lys2ΔBgl msh2* strain displayed a highly elevated rate of frameshift mutagenesis, and all but one event occurred at runs ≥ 4N. As mentioned above, these data led to the conclusion that short runs and noniterated sequences are not hotspots for frameshift mutagenesis (Greene and Jinks-Robertson, 1997; Harfe and Jinks-Robertson, 1999). However, the mutation spectrum of the *lys2ΔBgl,NR msh2* strain demonstrates that this is not the case. As shown in Figure 4.2, the *lys2ΔBgl,NR msh2* spectrum was composed entirely of simple -1 events. The majority of frameshifts in the *lys2ΔBgl,NR msh2* strain occurred at three hotspots, which are referred to as HS1-3 and are shown in Figure 4.2. HS1 is located in the same 2N run of guanines that is a hotspot for mutations in the WT spectrum. In the *msh2* strain, the mutation rate at this location was elevated 25-fold relative to WT (Table 4.3). HS2 and HS3 are located in 3N runs of thymines. The relative mutation rate increases at these two spots in the *msh2* strain were 1000- and 74-fold, respectively (Table 4.3).

There were several other hotspots in the *lys2ΔBgl,NR msh2* spectrum that were localized at 3N, 2N, and noniterated sequences. The proportions of mutations at 3N and 2N runs were significantly greater than those in the *lys2ΔBgl,NR* strain ($p<0.0001$ for both; Table 4.1). Furthermore, the proportion of events at 3N runs was significantly more than expected based on their proportion of the window ($p<0.0001$; Table 4.2). The proportion of events at 2N runs was not significantly different from the expected proportion ($p=1$), and the proportion of events at noniterated sequences was significantly

Table 4.3 Rates of Different Classes of Mutations in MMR-deficient *lys2ΔBgl,NR* Strains

Genotype[a]	Total Rate ($x10^{-11}$)	Rates of Different Classes of Mutations[b]							
		Simple Δ	HS1	HS2	HS3	Simple +2	Large Δ	Large +	Complex
WT	99 (76-160)	71 (121/169)	13 (22/169)	0.58 (1/169)	2.3 (4/169)	18 (31/169)	3.0 (5/169)	4.7 (8/169)	2.3 (4/169)
msh2	1900 (1600-3200)	1900 (172/172)	320 (31/179)	740 (70/179)	170 (16/179)	ND (0/172)	ND (0/172)	ND (0/172)	ND (0/172)
msh3	150 (110-180)	120 (132/172)	29 (33/172)	5.3 (6/172)	14 (16/172)	11 (13/172)	11 (13/172)	4.4 (5/172)	7.8 (9/172)
msh6	120 (73-150)	100 (148/175)	28 (41/175)	12 (18/175)	1.3 (2/175)	5.5 (8/175)	6.8 (10/175)	3.5 (5/175)	2.7 (4/175)
mlh2	140 (110-170)	110 (133/172)	21 (25/172)	4.9 (6/172)	0.81 (1/172)	9.8 (12/172)	4.1 (5/172)	7.3 (9/172)	11 (13/172)
mlh3	130 (83-170)	95 (124/171)	23 (30/171)	4.6 (6/172)	3.0 (4/171)	9.9 (13/171)	8.3 (11/171)	9.9 (13/171)	7.5 (10/172)

[a] All strains contain the *lys2ΔBgl,NR* allele.

[b] 95% confidence intervals of total mutation rates are indicated in parentheses. For different classes of mutations, numbers in parentheses indicate the proportion of events in the mutation spectra. ND indicates none detected.

less than expected ($p<0.0001$). Relative to the WT strain, the mutation rates at 3N, 2N, and noniterated sequences were increased 92-, 16-, and 8.8-fold, respectively. It is thus clear that 3N runs are hotspots for frameshift mutagenesis and that a significant amount of frameshifts also occur at 2N runs and noniterated sequences.

To examine the relative roles of the MutSβ and MutSα complexes in the removal of frameshift intermediates at short runs and noniterated sequences, *msh3* and *msh6* strains were generated and analyzed. As shown in Figure 4.3, the *msh3* and *msh6* spectra looked more like the WT spectrum than the *msh2* spectrum. Also, the mutation rates of the *msh3* and *msh6* strains were not elevated relative to the WT strain (Table 4.3). This is not surprising, as Msh3 and Msh6 often act redundantly, especially at small insertion and deletion intermediates (Greene and Jinks-Robertson, 1997; Harfe and Jinks-Robertson, 1999). In the *msh3* strain, the mutation rates at HS1-3 were increased 2.2-, 9.1-, and 6.1-fold, respectively. This pattern is similar, but less pronounced, than that of the *msh2* strain (25-, 1000-, and 74-fold, respectively).

The proportions of simple -1 and +2 events and large insertions and deletions in the *msh6* were similar to those in the *msh3* strain (Table 4.3). Also similar to the *msh3* strain, the rate of mutations at HS1 in the *msh6* strain was increased 2.2-fold relative to WT. However, the *msh6* strain behaved differently from the *msh3* strain at HS2 and HS3. At HS2, the mutation rate in the *msh6* strain was elevated 21-fold over WT and 2.6-fold over *msh3* (Table 4.3). In contrast, the mutation rate at HS3 in the *msh6* strain was not significantly different from WT (1.3x10-11 and 2.3x10-11, respectively) and was 11-fold lower than that in the *msh3* strain. This suggests that Msh6 is more commonly

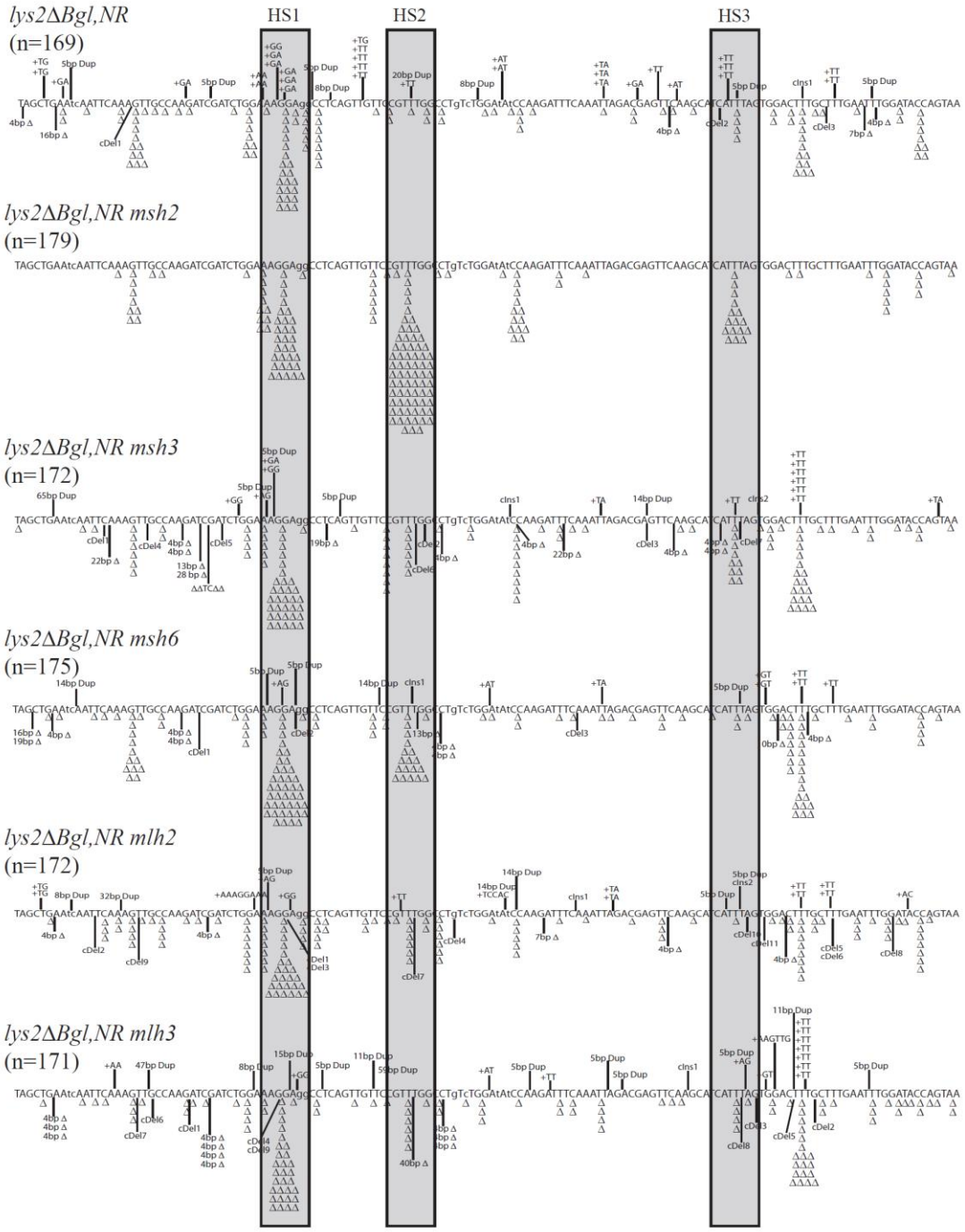

Figure 4.3 Mutation Spectra of MMR Mutant Strains

Simple -1 frameshifts are indicated with a Δ symbol. "cIns" indicates a complex insertion, which is defined as an insertion associated with a base substitution within 5 bp, and "cDel" indicates a complex deletion, which is defined as a deletion associated with a base substitution within 5 bp. "Dup" indicates a duplication. HS1-3 indicates hotspots 1-3.

used in the repair of frameshift intermediates at HS2, but that Msh3 can completely compensate for the loss of Msh6 at HS3.

The MutL homologs Mlh2 and Mlh3 have been shown to have minor roles in the repair of frameshift intermediates (Flores-Rozas and Kolodner, 1998; Harfe et al., 2000). To examine their involvement in the repair of frameshift intermediates at short runs and noniterated sequences, *mlh2* and *mlh3* strains were generated. Mutation rates and spectra reveal that the deletion of either of these genes had only a very minor effect on frameshift mutagenesis (Figure 4.3 and Table 4.3). Relative to WT, the *mlh2* and *mlh3* strains had the same proportion of large insertions and deletions and simple -1 events. The mutation rates at HS1-3 were elevated 1.6- and 1.7-fold, 8.4- and 7.9-fold, and 0.35- and 1.3-fold, respectively, relative to WT. Thus, Mlh2 and Mlh3 may play a small role in the removal of frameshift intermediates at HS2, but are either completely redundant or have no apparent role at HS1 or HS3. Interestingly, the rate of complex events in the *mlh2* strain was increased 4.8-fold relative to WT.

Examination of +2 frameshift events.

Although the rate of +2 events was not elevated in the *lys2ΔBgl,NR* strain relative to the *lys2ΔBgl* strain, the number of these events comprised a significantly increased proportion of the mutation spectrum (p=0.0002; Table 4.1). This increased proportion enables us to specifically examine these events. Except for one event, all of the +2 events (97%) were sequence duplications, and approximately half (46%) expanded 2N or 3N runs to 4N or 5N runs, respectively (Figure 4.2). Interestingly, the proportions of these events were significantly decreased in all of the MMR mutant strains examined, suggesting that the accumulation of +2 intermediates may be promoted rather than

prevented by the MMR system (Table 4.3). As mentioned above, previous work in our lab has shown that the suppression of recombination by the MMR system indirectly promotes the formation of TLS Polζ-dependent complex mutations (Lehner and Jinks-Robertson, 2009). It is also possible that these frameshift intermediates are generated in response to DNA damage outside the context of DNA replication and are therefore not subject to MMR. We thus hypothesized that these +2 events were due either to the activity of the error-prone TLS polymerase Polζ, which is known to be responsible for approximately 50% of all spontaneous mutagenesis in yeast (Northam et al., 2010; Quah et al., 1980), or to the repair or bypass of some type of DNA lesion. To test this hypothesis, we generated strains lacking Polζ by deleting the *REV3* gene and strains defective in nucleotide excision repair (NER) by deleting the *RAD14* gene.

If Polζ activity is responsible for generating the +2 events seen in the WT spectrum, these events should be decreased in the *rev3* spectrum. Surprisingly, the rate of +2 events in the *rev3* strain was slightly higher than the rate in the WT strain, and the proportion of +2 events was significantly increased ($p=0.007$; Figure 4.4 and Table 4.4). This indicates that the +2 events are not being generated by Polζ. Interestingly, the proportion of -1 events in the *rev3* strain was significantly decreased relative to WT ($p=0.0004$), suggesting that some of the -1 events may be due to Polζ activity.

If the +2 events are generated due to some type of bulky DNA lesion that is processed by NER, we would expect to see an increase in +2 events in the *rad14* strain. As shown in Figure 4.4 and Table 4.4, however, the rate of +2 events was not changed upon deletion of *RAD14*. This indicates that the +2 events were not generated by DNA damage that is subject to NER. In the *rad14* strain, we did see a significantly increased

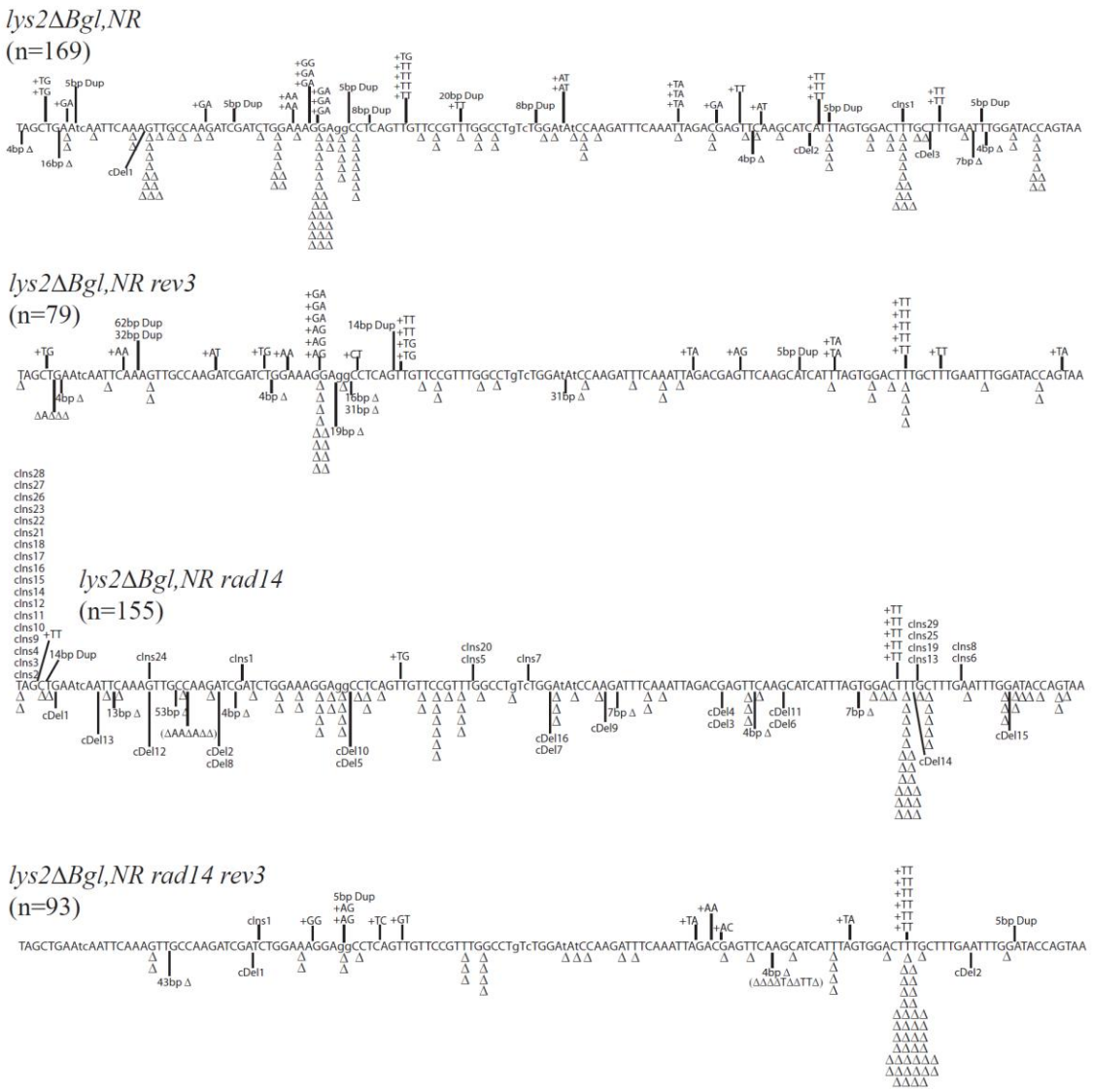

Figure 4.4 Mutation Spectra of *rev3* and *rad14* Strains

Simple -1 frameshifts are indicated with a Δ symbol. "cIns" indicates a complex insertion, which is defined as an insertion associated with a base substitution within 5 bp, and "cDel" indicates a complex deletion, which is defined as a deletion associated with a base substitution within 5 bp. "Dup" indicates a duplication.

rate of complex mutations (39-fold), which is consistent with previous studies (Harfe and Jinks-Robertson, 2000a; Minesinger and Jinks-Robertson, 2005). In earlier studies, the elevated rates of complex events in NER-deficient strains were dependent on Polζ activity. Deletion of *REV3* in the *rad14* strain indicated that this is also the case in this

Table 4.4 Rates of Difference Classes of Mutations in the *lys2ΔBgl,NR rev3* and *rad14* Strains

Genotype[a]	Total Rate ($x10^{-11}$)	Rates of Different Classes of Mutations[b]				
		Simple Δ	Simple +2	Large Δ	Large +	Complex
WT	99 (76-160)	71 (121/169)	18 (31/169)	3.0 (5/169)	4.7 (8/169)	2.3 (4/169)
rev3	86 (67-120)	41 (38/79)	29 (27/79)	7.7 (7/79)	4.4 (4/79)	NA (0/79)
rad14	310 (220-400)	190 (95/155)	14 (7/155)	14 (7/155)	2.0 (1/155)	90 (45/155)
rad14 rev3	90 (78-140)	68 (70/93)	15 (15/93)	2.9 (3/93)	1.9 (2/93)	2.9 (3/93)

[a] All strains contain the *lys2ΔBgl,NR* allele.

[b] 95% confidence intervals of total mutation rates are indicated in parentheses. For different classes of mutations, numbers in parentheses indicate the proportion of events in the mutation spectra.

system; the rate and proportion of complex mutations in the *rad14 rev3* strain was greatly reduced relative to the *rad14* strain (Figure 4.4 and Table 4.4).

Comparison of -1 and +1 frameshift assays.

All of the experiments described above were conducted using the *lys2ΔBgl,NR* allele, which specifically selects net -1 frameshift events. We also constructed the *lys2ΔA746,NR* allele, which specifically selects net +1 frameshift events. Although -1 and +1 frameshift intermediates are thought to be very similar, differing only in which strand contains the extrahelical base, comparison of the spectra generated from these two assays reveals striking differences. As shown in Figure 4.5 and Table 4.5, the majority (72%) of events in the *lys2ΔBgl,NR* spectrum were simple -1 events. Although the rate of +1 events in the *lys2ΔA746,NR* strain was similar to that of -1 events in the *lys2ΔBgl,NR* strain, these events only accounted for 23% of the *lys2ΔA746,NR* spectrum. There was also a significantly increased rate of large and complex insertions in the *lys2ΔA746*,*NR* strain relative to the *lys2ΔBgl,NR* strain (6.6-fold and 31-fold, respectively). We observed 35 large deletions in the *lys2ΔA746,NR* spectrum that were due to 10-bp direct repeats that lie just upstream and within the reversion window. As similar repeats that could generate large deletions are not present in the *lys2ΔBgl,NR* system, we did not include these events in the analysis.

To specifically examine the spectra of unrepaired -1 and +1 events in the two alleles, we generated the *lys2ΔA746,NR msh2* strain and compared it to the *lys2ΔBgl,NR msh2* strain (Figure 4.5). It is immediately apparent that the spectra look very different, as each strain has different hotspots for +1 and -1 events, respectively. In the *lys2ΔBgl,NR* allele, there are nine 3N runs. As shown in Figure 4.5 and Table 4.6, these

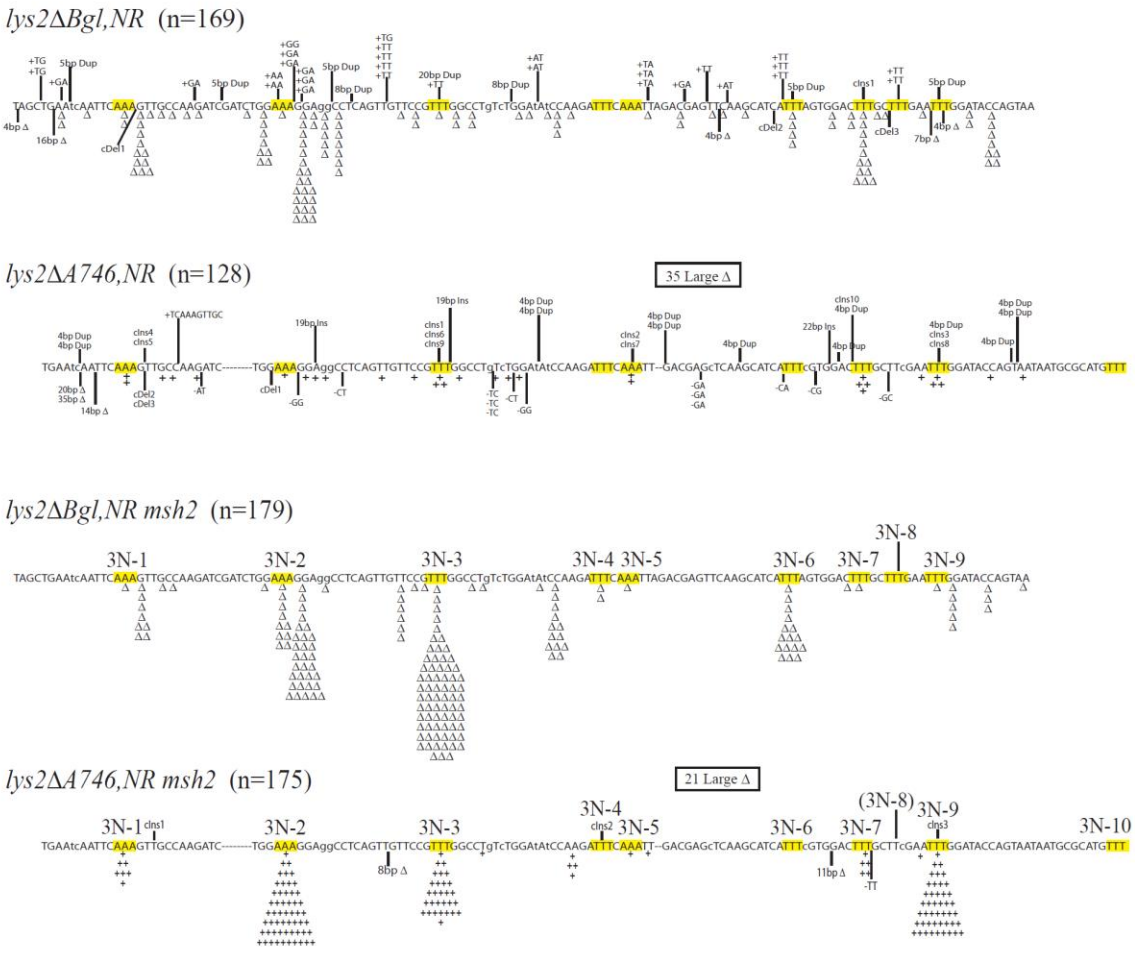

Figure 4.5 Mutation Spectra of *lys2ΔBgl,NR* and *lys2ΔA746,NR* WT and *msh2* Strains

Simple -1 frameshifts are indicated with a Δ symbol. "cIns" indicates a complex insertion, which is defined as an insertion associated with a base substitution within 5 bp, and "cDel" indicates a complex deletion, which is defined as a deletion associated with a base substitution within 5 bp. "Dup" indicates a duplication. 3N runs in *msh2* spectra are highlighted in yellow and labeled 3N-1 to 3N-10.

3N runs are not equal hotspots for frameshift mutagenesis. The proportion of events at these runs in the *msh2* strain ranged from none to 39%, with rate increases relative to WT ranging from undetectable to 1000-fold. In the *lys2ΔA746,NR* allele, there are also nine different 3N runs. The eighth 3N run in the *lys2ΔBgl,NR* allele (3N-8) was

Table 4.5 Rates of Different Classes of Mutations in the *lys2ΔBgl,NR* and *lys2ΔA746,NR* Strains

Genotype	Total Rate ($x10^{-11}$)	Rates of Different Classes of Mutations[a]							
		+1	-1	+2	-2	Large +	Large Δ	Complex Insertions	Complex Deletions
lys2ΔBgl,NR	99 (76-160)	ND (0/169)	71 (121/169)	18 (31/169)	ND (0/169)	4.7 (8/169)	2.9 (5/169)	0.59 (1/169)	1.8 (3/169)
lys2ΔA746,NR	230 (160-330)	54 (30/128)	ND (0/128)	ND (0/128)	25 (14/128)	31 (17/128)	5.4 (3/128)	18 (10/128)	5.4 (3/128)

[a] 95% confidence intervals of total mutation rates are indicated in parentheses. For different classes of mutations, numbers in parentheses indicate the proportion of events in the mutation spectra. ND indicates none detected.

Table 4.6 Rates of Events at 3N Runs in the *lys2ΔBgl,NR msh2* and *lys2ΔA746,NR msh2* Strains

Genotype	Total Rate ($x10^{-8}$)	Rates of Events at 3N Runs[a]									
		3N-1	3N-2	3N-3	3N-4	3N-5	3N-6	3N-7	3N-8	3N-9	3N-10
lys2ΔBgl,NR msh2 n=179	190 (16-32)	1.1 (1/179)	11 (10/179)	74 (70/179)	2.1 (2/179)	1.1 (1/179)	17 (16/179)	1.1 (1/179)	ND (0/179)	1.1 (1/179)	---
lys2ΔA746,NR msh2 n=175	170 (14-20)	6.8 (7/175)	53 (55/175)	29 (29/175)	ND (0/175)	0.97 (1/175)	ND (0/175)	4.9 (5/175)	---	44 (45/175)	ND (0/175)

[a] 95% confidence intervals of total rates are in parentheses. ND = none detected.

mutated in the construction of the *lys2ΔA746,NR* allele and is therefore absent in this strain, but there is an additional 3N run in the *lys2ΔA746,NR* allele (3N-10) that is not present in the *lys2ΔBgl,NR* allele (see Figure 4.5). The proportion of events at the 3N runs in the *lys2ΔA746,NR msh2* strain were also distributed unequally, with these sites containing none to 31% of events and rate increases relative to WT ranging from undetectable to 290-fold (Table 4.6). Importantly, sites that were hotspots in the *lys2ΔBgl,NR msh2* strain were not necessarily hotspots in the *lys2ΔA746,NR msh2* strain, and vice versa. For example, there were 5-fold and 40-fold more +1 than -1 events at sites 3N-2 and 3N-9, respectively. In contrast, there were 10-fold more -1 than +1 events at 3N-6. These findings suggest that specific sites can be hotspots for either -1 frameshifts or +1 frameshifts, or both.

In the *lys2ΔBgl,NR msh2* spectrum, 57% of -1 events were at 3N runs, 36% of -1 events were at 2N runs, and 7.3% of events were at noniterated sequences. In contrast, 81% of +1 events in the *lys2ΔA746,NR msh2* strain were at 3N runs, 4% of +1 events were at 2N runs, and there were no events at noniterated sequences. Thus, while -1 events were detected at all types of sequences, +1 events occurred almost exclusively at 3N runs (Figure 4.6). Interestingly, 38% of -1 events were located at GC base pairs, which comprise 40% of the reversion window, while only one +1 event (0.57%) was located at a GC base pair. This suggests that GC base pairs are not susceptible to +1 frameshift events in this system.

By comparing the *msh2* and WT strains, we were able to calculate the efficiency of MMR in repairing frameshift intermediates at different types of sequences using the equation (*msh2* rate – WT rate)/(*msh2* rate). As shown in Table 4.7, the overall rate

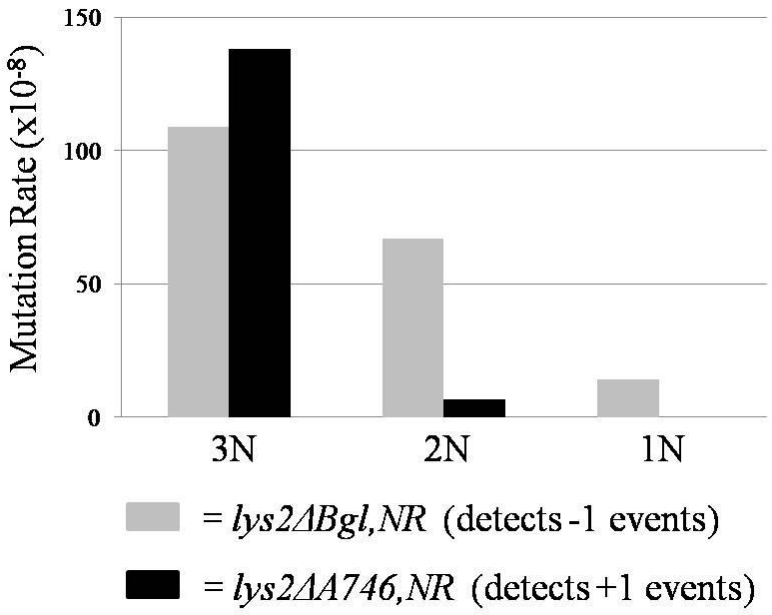

Figure 4.6 Distribution of Frameshifts in *lys2ΔBgl,NR msh2* and *lys2ΔA746,NR msh2* Spectra

Black bars represent *lys2ΔBgl,NR msh2* strain, and gray bars represent *lys2ΔA746,NR msh2* strain.

increase in -1 frameshifts in the *lys2ΔBgl,NR msh2* strain relative to WT was 27-fold, and the overall efficiency of MMR was 96%. In this assay, the efficiency of MMR at 3N and 2N runs and noniterated sequences was 99%, 94%, and 89%, respectively. Thus, MMR is more efficient at removing -1 frameshift intermediates at 3N runs than at 2N runs, and is more efficient at 2N runs than at noniterated sequences. The overall rate increase in +1 frameshifts in the *lys2ΔA726,NR msh2* strain relative to WT was 26-fold, and the overall efficiency of MMR was 96%. The efficiency of MMR at 3N and 2N runs was 98% and 74%, respectively. It should be noted that the number of events at 2N runs in the *lys2ΔA746,NR* strains was very low, and that estimation of MMR efficiency at these sites may therefore be inaccurate. The efficiency of MMR at noniterated sequences could not be calculated because no events were detected in the *msh2* strain. These results suggest

Table 4.7 MMR Efficiency in the *lys2ΔBgl,NR msh2* and *lys2ΔA746,NR msh2* Strains

	MMR Efficiency[a]			
Genotype	Overall	3N Runs	2N Runs	1Ns
lys2ΔBgl,NR msh2	96% (27x)	99% (92x)	94% (16x)	89% (8.8x)
lys2ΔA746,NR msh2	96% (26x)	98% (52x)	74% (3.8x)	ND

[a] Numbers in parentheses indicate fold rate increases relative to the corresponding WT strain. ND indicates not detected.

that MMR is much more efficient at repairing +1 frameshift intermediates at 3N runs than at 2N runs. Furthermore, MMR is more efficient at repairing -1 rather than +1 frameshift intermediates at both 3N and 2N runs.

The efficiency of MMR also varies widely at different runs. For example, in the *lys2ΔBgl,NR* strain, the efficiency of MMR is 99-100% at 3N-2, 3N-3, and 3N-6, 89% at 3N-1 and is only 36% at 3N-7. In the *lys2ΔA746,NR* strain, the efficiency of MMR was similar at 3N-2 and 3N-3, with values of 99.7% and 98%, respectively, but was different at other sites. For example, the efficiency of MMR at 3N-1 and 3N-7 was 95% and 85%, respectively. Together, these results demonstrate that there are large variations in the efficiency of MMR at different 3N runs within a ~150-bp window.

4.5 Discussion

The vast majority of studies of spontaneous frameshift mutagenesis have focused on frameshifts that occur in long homopolymer runs and repeated sequences. Although frameshifts have been observed in short runs and noniterated sequences, they occur at low levels and have therefore been difficult to study in vivo. Other groups have

overcome this issue by using in vitro systems and/or by using exogenous DNA damaging agents that increase the amount of frameshifts that occur (for example, Tippin et al., 2004). To study spontaneous frameshifts of this type in vivo, we have developed two reversion assays (*lys2ΔBgl,NR* and *lys2ΔA746,NR*) that can specifically detect spontaneous frameshifts that occur in short runs and noniterated sequences. Analysis of the mutation spectra generated in these systems has revealed novel information about frameshift mutagenesis.

In a WT background, the majority (72%) of *lys2ΔBgl,NR* reversion events were simple -1 frameshifts. Although we expected that most of these -1 events would be localized at 3N runs (the longest runs in the sequence), 60% were instead localized at 2N runs. The proportion of -1 events at 3N runs and noniterated sequences was 17% and 23%, respectively. This suggests that, at least in the window analyzed, 2N runs accumulate more susceptible frameshift mutations than 3N runs and noniterated sequences. As discussed below, subsequent analysis demonstrated that the decreased proportion of frameshifts at 3N runs reflects an increased efficiency of MMR at these runs relative to 2N runs. Interestingly, most (70%) of the -1 events occurred at GC base pairs. This proportion was significantly higher than expected ($p<0.0001$), as GC base pairs comprise only 40% of the reversion window. Previous studies have indicated that GC base pairs may be less accessible to DNA polymerase proofreading activity, and this may contribute to the increased proportion of events at GC base pairs (Bessman and Reha-Krantz, 1977; Goodman and Fygenson, 1998). It is also possible that MMR is less efficient at GC versus AT base pairs (Gragg et al., 2002).

In the MMR-defective *msh2* strain, all events were -1 events, and the majority (57%) of events occurred at 3N runs. This proportion was significantly elevated relative to the expected proportion based on the sequence of the reversion window, indicating that 3N runs are hotspots for DNA polymerase slippage in this strain. The proportion of events at 2N runs and noniterated sequences was 36% and 7.3%, respectively. These proportions correspond to 92-, 16-, and 8.8-fold increases in the rate of frameshifts at 3N, 2N, and noniterated sequences, respectively, relative to the WT strain. This indicates that the efficiency of MMR at repairing frameshift intermediates follows the order 3N>2N>1N, which is in agreement with previous studies that showed that MMR is more efficient as the length of a homopolymer runs increases (Tran et al., 1997). MMR efficiency is discussed further below.

The *lys2ΔBgl,NR* reversion spectra obtained from the *msh3* and *msh6* strains were very similar to each other and to that of the WT strain, suggesting that Msh3 and Msh6 have mostly redundant roles in the repair of frameshift intermediates in this system. However, some aspects of the *msh3* and *msh6* spectra were different. For example, HS2 was significantly hotter in the *msh6* strain than in the *msh3* strain, suggesting that MutSα plays the predominant role in removing -1 frameshift intermediates at this site. At HS3, however, the reverse specificity was observed. These differences between MutSα and MutSβ may be due to sequence context and/or differences in the underlying frameshift intermediate.

Mlh2 and Mlh3 have been shown to have minor roles in the repair of frameshift intermediates in other studies, with the deletion of either the *MLH2* or *MLH3* gene resulting in the increase of a specific type of event (Flores-Rozas and Kolodner, 1998;

Harfe et al., 2000). In the current study, the *mlh2* and *mlh3* spectra looked similar to that of WT, with three notable exceptions. First, both strains showed elevated rates of mutations at HS2 (8.4- and 7.9-fold, respectively). Interestingly, this is the hottest spot of frameshift mutagenesis in the *msh2* strain. It thus appears that multiple MutL complexes participate in the repair of frameshift intermediates that accumulate at this site. Second, similar to all other MMR mutants, the *mlh2* and *mlh3* strains had decreased proportions of +2 events relative to WT. These events are discussed further below. Finally, the rates of complex events in the *mlh2* and *mlh3* strains were increased 4.8- and 3.3-fold, respectively, relative to WT.

As mentioned above, we observed a significantly increased proportion of +2 events in the *lys2ΔBgl,NR* spectrum relative to the *lys2ΔBgl* spectrum. Although the rates of these events were not different between the two strains, the removal of runs $\geq$ 4N increased their proportion of the *lys2ΔBgl,NR* mutation spectrum. No +2 events were detected in the *msh2* background, and the rate of these events did not increase in any of the other MMR mutants. The data thus suggest that the corresponding frameshift intermediates are not subject to MMR, and we suggest that these events either escape detection by MMR or do not occur during DNA replication. For example, some types of DNA damage have been shown to cause frameshifts when bypassed by low fidelity TLS polymerases or when repaired (Efrati et al., 1997; Heidenreich et al., 2009; Kokoska et al., 2003; Zang et al., 2005). If the bypass or repair occurs outside of the context of DNA replication, it is possible that MMR does not detect the corresponding mutational intermediates. To examine these possibilities, we examined the accumulation of +2 events in strains deficient in the TLS polymerase Polζ or the NER pathway. Although

TLS has been assumed to primarily occur during replication, studies suggest that these polymerases also act outside of S phase (McHugh and Sarkar, 2006; Soria et al., 2009). NER is involved in the removal of bulky, helix-distorting DNA lesions. Surprisingly, neither removal of Polζ nor NER had any effect on the rate of +2 events.

If +2 mutations arise as a result of DNA damage, it is possible that the damage is bypassed by another translesion synthesis polymerase, Polη, or that Polη acts redundantly with Polζ. This could be determined by examining mutation spectra of Polη- and Polη/Polζ-deficient strains. It is also possible that the damage is not bulky and, therefore, not subject to NER. This could be tested by examining mutant strains defective in various components of the base excision repair pathway. Finally, it is possible that these events are, in fact, replication errors that are simply not detected by MMR. Although MMR has been shown to efficiently detect and remove 2-bp insertion and deletion intermediates that occur in dinucleotide repeats (Wierdl et al., 1997), it is possible that MMR is less efficient at detecting similar events in non-repetitive DNA. This could be addressed with the use of mutant DNA polymerases. As mentioned above, exonuclease-deficient Polδ and Polε cause increased rates of replication errors and may be useful in determining which mutations are generated during replication. There are also DNA polymerase mutants that specifically affect frameshift mutagenesis. For example, a mutant allele of Polδ, *pol3-447*, has been shown to cause a specific decrease in frameshift mutations (Hadjimarcou et al., 2001). If the +2 events are caused by Polδ replication errors, these events may be decreased in strains carrying the *pol3-447* mutant allele. There is also a mutant allele of Polε, *pol2-C1089Y*, which has been shown to cause a specific increase in frameshift mutagenesis (Kirchner et al., 2000). If the +2 events are

caused by Polε replication errors, we may see an increase in +2 events in strains carrying this mutant allele.

To further examine different types of frameshift intermediates, we generated an additional "no run" reversion allele, *lys2ΔA746,NR*, which specifically detects net +1 frameshifts in approximately the same window as that in the *lys2ΔBgl,NR* allele. Interestingly, although the majority (72%) of events in the WT -1 assay (*lys2ΔBgl,NR*) were simple -1 frameshifts, only 23% of events in the WT +1 assay (*lys2Δ736,NR*) were simple +1 frameshifts. We note, however, that the rates of +1 and -1 frameshift events were similar ($71x10^{-11}$ and $54x10^{-11}$, respectively). The rates of large and complex insertions were increased in the *lys2ΔA746,NR* relative to the *lys2ΔBgl,NR* strain (6.6-fold and 31-fold, respectively), while the rates of large and complex deletions were similar (Table 4.5). While the reason for this discrepancy is unknown, one possibility is that extrahelical DNA is more likely to accumulate on the primer rather than the template strand during DNA synthesis.

By deleting the *MSH2* gene in the *lys2ΔBgl,NR* and *lys2ΔA746,NR* strains, we were able to examine unrepaired -1 and +1 frameshift errors that escaped proofreading by DNA polymerase. Although the rates of frameshift errors were similar between the two strains, there were striking differences in their mutation spectra. First, although each allele contains nine 3N runs, only certain 3N runs were hotspots for frameshifts, and the distributions of events at each hotspot were very different in the two spectra. Specifically, sites that were hotspots for -1 events were not necessarily hotspots for +1 events, and vice versa. This suggests that flanking sequence context plays a big role in determining hotspots for frameshifts and that different sites can be more susceptible to

one or the other type of slippage intermediate. This may be due to increased susceptibility to some type of DNA damage or replication errors, or to variations in proofreading efficiency.

We also observed that GC base pairs appear to be hotspots for -1 frameshifts, but not +1 frameshifts. This may suggest that proofreading is less efficient at repairing an extrahelical G/C on the template strand than on the primer strand. It is also possible that, if these -1 frameshifts are the result of some underlying DNA damage, GC base pairs may be more susceptible to that type of damage. It will be interesting to examine the mutation spectra of various repair protein and polymerase mutants to determine why GC base pairs are only hotspots for certain types of frameshift mutagenesis.

By comparing the WT and *msh2* strains, we were able to calculate the efficiency of MMR at both -1 and +1 frameshift intermediates. While MMR has the same overall efficiency in the repair of the either type of frameshift intermediate (96%), MMR appears to be more efficient at repairing -1 frameshift intermediates than +1 frameshift intermediates at both 3N and 2N runs in this system. These results are in agreement with previous studies and suggest that MMR is more efficient at detecting extrahelical bases on the template strand (Gragg et al., 2002; Sia et al., 1997). We also observed a large amount of variability in the efficiency of MMR at different sites within the reversion windows of both strains, which demonstrates clear site-to-site variations in MMR activity. This may be due to an inability of MMR to identify frameshift intermediates in certain sequence contexts.

In summary, the experiments reported here have revealed several important features of frameshift mutagenesis. First, 3N runs are clearly hotspots for frameshift

mutagenesis, and a significant amount of frameshifts also occur at 2N runs and noniterated sequences. This suggests that DNA polymerase slippage occurs at 3N repeats. Second, GC base pairs appear to be hotspots for -1 frameshifts, but not +1 frameshifts. Third, the WT *lys2ΔBgl,NR* spectrum contains a significant proportion of +2 events that do not appear to be substrates for MMR. Finally, within the same ~150-bp window, different sites can be hotspots for either -1 or +1 frameshifts, or both. There are thus fundamental differences in the generation and repair of -1 and +1 frameshift intermediates, and it is likely that frameshifts occur via multiple mechanisms. The additional experiments suggested throughout the discussion will likely expand our understanding of these processes.

4.6 Acknowledgements

We thank Brenda Minesinger for constructing the *lys2ΔBgl,NR* and *lys2ΔA746,NR* alleles and Kevin Lehner for the experiments with the *lys2ΔA746,NR* strains. This work was supported by NIH grants GM038464 and GM064769 awarded to S. J.-R.

Chapter 5: Concluding Remarks

Mutagenesis is one of the most fascinating aspects of biology. It is the key mechanism by which all species evolve, but also by which many deleterious conditions, including numerous diseases and cancer, arise. Mutagenesis must therefore be kept at low levels to ensure a slow rate of evolution but a decreased chance of harmful DNA modifications. Although the term "mutagenesis" may evoke thoughts of carcinogens and/or irradiation, mutations frequently arise without exposure to any exogenous damaging agents. These spontaneous mutations are caused by endogenous damaging agents that are created as byproducts of normal cellular metabolism or by mistakes made by DNA polymerases. In the experiments described here, we have examined two types of spontaneous mutations: oxidative GO-associated transversions and frameshifts.

The oxidative GO lesion is one of the most common types of endogenously arising DNA damage. In humans, GO lesions are associated with aging and several diseases, including cancer and Huntington's disease (Kovtun et al., 2007; Skinner and Turker, 2005). At the time that this work began, the two major repair mechanisms known to be involved in suppressing GO-associated mutagenesis were the Ogg1 DNA glycosylase and MMR. There were hints that the TLS polymerase Polη was also involved in suppressing GO-associated mutagenesis, but the nature of TLS polymerases made it difficult to interpret how Polη could function in this capacity. Specifically, TLS polymerases are often error-prone and are known to contribute to a large proportion of all spontaneous mutations (Northam et al., 2010; Quah et al., 1980). However, Polη is actually very efficient at the error-free bypass of two types of lesions, UV-induced

pyrimidine dimers and GO lesions (Haracska et al., 2000; Johnson et al., 1999b; McDonald et al., 1997; Yoon et al., 2009; Yuan et al., 2000). As GO lesions are not thought to block DNA replication, which triggers TLS, it was proposed that Polη was somehow specifically recruited by the MMR machinery to bypass GO lesions (de Padula et al., 2004; Haracska et al., 2000). However, the experiments described in Chapter 2 clearly show that Polη acts independently of the MMR pathway to suppress GO-associated mutagenesis. Although we cannot exclude an additional role for Polη within the MMR pathway, it is difficult to imagine how a TLS polymerase could be specifically recruited to fill in gaps generated by MMR that contain GO lesions. The most plausible explanation is that GO lesions can at times stall DNA polymerase and thereby signal for TLS. This may happen during DNA replication or during the filling in of gaps generated by MMR.

When replication is stalled or blocked by a DNA lesion, the cell can use one of three different tolerance pathways (TLS, template switching, or homologous recombination) to bypass the damage and continue replication. Many lesions, including abasic sites and UV-induced pyrimidine dimers, are bypassed by more than one pathway (Lin et al., 2006; McDonald et al., 1997; Nakai and Mortimer, 1969; Swanson et al., 1999). It has been shown that homologous recombination or template switching is the preferred tolerance pathway for UV lesions (Zhang and Lawrence, 2005), and it is thought that this is likely the case for other replicating-blocking lesions as well. Surprisingly, our results and the results of others suggest that GO lesions do not trigger template switching or homologous recombination (Table 2.8; van der Kemp et al., 2009). It is thus possible that certain types of DNA lesions trigger specific bypass pathways

when encountered by DNA polymerase. It will be interesting to see if this is true of other types of damage and how the choice of tolerance pathway is regulated. It is currently unknown how or why one tolerance pathway is chosen over another.

On a similar note, it is surprising that Polη bypass of GO lesions does not appear to be regulated by ubiquitinated PCNA, as many studies have shown that this form of PCNA is required for TLS (Garg and Burgers, 2005b; Haracska et al., 2004; Kannouche et al., 2004; Stelter and Ulrich, 2003; van der Kemp et al., 2009a; Zhuang et al., 2008). Moreover, Polη activity was also unaffected by deletion of the alternative 9-1-1 clamp. Given the error-prone nature of TLS polymerases, TLS is thought to be tightly regulated. Although it is possible that the mutant allele of PCNA that cannot be ubiquitinated disrupts other aspects of damage tolerance that in turn obscure accurate analysis of Polη activity (e.g., this form of PCNA cannot be sumoylated), it is also possible that this form of PCNA is not required for this specific type of TLS. Indeed, it has recently been shown in vertebrates that TLS occurring at the replication fork requires Rev1, while TLS occurring during gap filling requires ubiquitinated PCNA (Sale et al., 2009). Similarly, Polη and Rev1 were shown to form a chromatin-bound complex in response to UV-induced replication fork arrest in human cells (Yuasa et al., 2006). To determine whether this is also true in yeast, experiments are currently under way to determine if Polη bypass of GO lesions requires Rev1. We hope that these experiments will provide further insight into how Polη activity is regulated in yeast.

As described in Chapter 3, we have attempted to examine the effect of replication timing on the activity of Polη and MMR at GO lesions. Surprisingly, the activity of MMR, but not Polη, varies throughout S phase. Unfortunately, it is unclear from our

studies whether this variation is directly correlated with replication timing. Although we see a significant difference between the mutation rate at the earliest replicating region in our study and one of the latest replication regions, variations in mutation rate at regions that replicate some time in between are subtle and difficult to interpret. It was previously noted that variations in mutation rate in yeast tend to be more subtle than in other species (Fox et al., 2008), and it is possible that this contributes to the difficulty we and others have had in examining the relationship between replication timing and mutation rate. Given the low level of variability, it is possible that clear differences in mutation rates will only be apparent for extremely early- and late-replicating regions.

It is important to note that if mutation rates do correlate with replication timing, our results suggest that the origins tested here do not fire at the times previously published (see, for example, Raghuraman et al., 2001 and Yabuki et al., 2002). For example, although *ARS607* has been characterized as an early-firing origin, the mutation rate associated with this origin suggests that adjacent regions replicate later in S phase. As the replication times of different origins have varied in different studies, it is possible that the origins in our yeast strain fire at different times than these the origins in other strains. It should also be noted that it has recently been suggested that origins in yeast do not have a defined firing time, but that origins appear early or late based on their probability of firing (reviewed in Rhind et al., 2009). It is possible that differences in strain backgrounds affect the probability of different origins firing. For these reasons, we are currently in the process of determining the replication times of the different *SUP4-o* alleles in our yeast strains. These experiments may reveal that origins vary in their firing time and efficiency across different yeast strains.

Aside from spontaneous oxidative mutagenesis, we have also examined how spontaneous frameshift mutations are generated in short runs and noniterated sequences (see Chapter 4). Because these mutations have been difficult to study in systems that contain large runs or repetitive DNA, we generated two reversion assays, *lys2ΔBgl,NR* and *lys2ΔA746,NR*, that specifically detect either -1 or +1 frameshift events, respectively, at short runs and noniterated sequences. We have used these assays to show that 3N runs are hotspots for frameshift mutagenesis due to polymerase slippage and that a significant number of frameshifts also occur at 2N runs and noniterated sequences. We have also shown that this system can be used to study spontaneous +2 frameshift events. It is currently unknown how these events are generated, but future experiments with this system will likely reveal the underlying mechanisms. Finally, we have shown that MMR appears to be more efficient at repairing frameshift intermediates at 3N runs than at 2N runs or noniterated sequences and is also more efficient at repairing -1 versus +1 frameshift intermediates. However, there is clear site-to-site variability in MMR efficiency. With this system, we hope to continue our characterization of how frameshift intermediates at these types of sequences are generated and repaired.

Studies of mutagenesis are critical for providing insight into how species have evolved and continue to evolve and for uncovering valuable information for the treatment of many diseases, including cancer. For example, because it is known that MMR has a critical role in the suppression of GO-associated mutagenesis, methotrexate, a drug that induces oxidative damage, has been used to selectively target MMR-defective tumor cells (Martin et al., 2009). As we have now shown that Polη also has a critical role in the suppression of this type of damage, similar treatments may be useful for Polη-deficient

tumor cells. Other groups have studied the ability of Polη to bypass lesions generated by other chemotherapeutic agents, such as cisplatin (Albertella et al., 2005). Lesions that cannot be bypassed by Polη will induce replication fork arrest and are therefore more likely to efficiently kill tumor cells. With this type of information, it may be possible to determine which chemotherapeutic agents will be most effective. In the case of UV-induced skin cancer, recent experiments have shown that overexpression of Polη provides increased error-free bypass of UV lesions, and thereby increased resistance to UV damage, without significantly increasing the overall rate of spontaneous mutations (Jung et al., 2010; King et al., 2005). Finally, chemotherapeutic treatments designed to delete trinucleotide repeats, which are prone to insertions and deletions, are also being developed (Hashem et al., 2004). The experiments presented here provide further insight into how spontaneous oxidative and frameshift mutations are generated and repaired and thus contribute to the field of mutagenesis.

References

Abdulovic, A.L., Minesinger, B.K., and Jinks-Robertson, S. (2007). Identification of a strand-related bias in the PCNA-mediated bypass of spontaneous lesions by yeast Polη. DNA Repair (Amst) *6*, 1307-1318.

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

Aladjem, M.I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat Rev Genet *8*, 588-600.

Albertella, M.R., Green, C.M., Lehmann, A.R., and O'Connor, M.J. (2005). A role for polymerase η in the cellular tolerance to cisplatin-induced damage. Cancer Res *65*, 9799-9806.

Amouroux, R., Campalans, A., Epe, B., and Radicella, J.P. (2010). Oxidative stress triggers the preferential assembly of base excision repair complexes on open chromatin regions. Nucleic Acids Res.

Anderson, J.A., Song, Y.S., and Langley, C.H. (2008). Molecular population genetics of Drosophila subtelomeric DNA. Genetics *178*, 477-487.

Aparicio, O.M., Stout, A.M., and Bell, S.P. (1999). Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A *96*, 9130-9135.

Araki, H., Hamatake, R.K., Johnston, L.H., and Sugino, A. (1991). *DPB2*, the gene encoding DNA polymerase II subunit B, is required for chromosome replication in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *88*, 4601-4605.

Arezi, B., and Kuchta, R.D. (2000). Eukaryotic DNA primase. Trends Biochem Sci *25*, 572-576.

Bahler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., 3rd, Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. Yeast *14*, 943-951.

Bai, H., and Lu, A.L. (2007). Physical and functional interactions between *Escherichia coli* MutY glycosylase and mismatch repair protein MutS. J Bacteriol *189*, 902-910.

Bauer, G.A., and Burgers, P.M. (1990). Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. Nucleic Acids Res *18*, 261-265.

Bebenek, K., Boyer, J.C., and Kunkel, T.A. (1999). The base substitution fidelity of HIV-1 reverse transcriptase on DNA and RNA templates probed with 8-oxo-deoxyguanosine triphosphate. Mutat Res *429*, 149-158.

Bebenek, K., Roberts, J.D., and Kunkel, T.A. (1992). The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. J Biol Chem *267*, 3589-3596.

Beletskii, A., and Bhagwat, A.S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. Proc Natl Acad Sci U S A *93*, 13919-13924.

Bemark, M., Khamlichi, A.A., Davies, S.L., and Neuberger, M.S. (2000). Disruption of mouse polymerase zeta (Rev3) leads to embryonic lethality and impairs blastocyst development in vitro. Curr Biol *10*, 1213-1216.

Benson, F.E., Baumann, P., and West, S.C. (1998). Synergistic actions of Rad51 and Rad52 in recombination and DNA repair. Nature *391*, 401-404.

Bessman, M.J., and Reha-Krantz, L.J. (1977). Studies on the biochemical basis of spontaneous mutation. V. Effect of temperature on mutation frequency. J Mol Biol *116*, 115-123.

Blastyak, A., Pinter, L., Unk, I., Prakash, L., Prakash, S., and Haracska, L. (2007). Yeast Rad5 protein required for postreplication repair has a DNA helicase activity specific for replication fork regression. Mol Cell *28*, 167-175.

Branzei, D., Vanoli, F., and Foiani, M. (2008). SUMOylation regulates Rad18-mediated template switch. Nature *456*, 915-920.

Breier, A.M., Chatterji, S., and Cozzarelli, N.R. (2004). Prediction of *Saccharomyces cerevisiae* replication origins. Genome Biol *5*, R22.

Brewer, B.J., and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell *51*, 463-471.

Brewer, B.J., and Fangman, W.L. (1991). Mapping replication origins in yeast chromosomes. Bioessays *13*, 317-322.

Brieba, L.G., Eichman, B.F., Kokoska, R.J., Doublie, S., Kunkel, T.A., and Ellenberger, T. (2004). Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. EMBO J *23*, 3452-3461.

Broach, J.R., Li, Y.Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A., and Hicks, J.B. (1983). Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harb Symp Quant Biol *47 Pt 2*, 1165-1173.

Buermeyer, A.B., Deschenes, S.M., Baker, S.M., and Liskay, R.M. (1999). Mammalian DNA mismatch repair. Annu Rev Genet *33*, 533-564.

Burgers, P.M., and Gerik, K.J. (1998). Structure and processivity of two forms of *Saccharomyces cerevisiae* DNA polymerase δ. J Biol Chem *273*, 19756-19762.

Chabes, A., and Stillman, B. (2007). Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *104*, 1183-1188.

Chaudhuri, S., Wyrick, J.J., and Smerdon, M.J. (2009). Histone H3 Lys79 methylation is required for efficient nucleotide excision repair in a silenced locus of *Saccharomyces cerevisiae*. Nucleic Acids Res *37*, 1690-1700.

Chen, C.C., Motegi, A., Hasegawa, Y., Myung, K., Kolodner, R., and D'Andrea, A. (2006). Genetic analysis of ionizing radiation-induced mutagenesis in *Saccharomyces cerevisiae* reveals TransLesion Synthesis (TLS) independent of PCNA K164 SUMOylation and ubiquitination. DNA Repair (Amst) *5*, 1475-1488.

Chen, J., Derfler, B., Maskati, A., and Samson, L. (1989). Cloning a eukaryotic DNA glycosylase repair gene by the suppression of a DNA repair defect in *Escherichia coli*. Proc Natl Acad Sci U S A *86*, 7961-7965.

Chen, P.C., Dudley, S., Hagen, W., Dizon, D., Paxton, L., Reichow, D., Yoon, S.R., Yang, K., Arnheim, N., Liskay, R.M., *et al.* (2005). Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. Cancer Res *65*, 8662-8670.

Cheng, K.C., Cahill, D.S., Kasai, H., Nishimura, S., and Loeb, L.A. (1992). 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G > T and A > C substitutions. J Biol Chem *267*, 166-172.

Chin, C.S., Chuang, J.H., and Li, H. (2005). Genome-wide regulatory complexity in yeast promoters: separation of functionally conserved and neutral sequence. Genome Res *15*, 205-213.

Choi, J.Y., Zang, H., Angel, K.C., Kozekov, I.D., Goodenough, A.K., Rizzo, C.J., and Guengerich, F.P. (2006). Translesion synthesis across 1,N2-ethenoguanine by human DNA polymerases. Chem Res Toxicol *19*, 879-886.

Coulondre, C., and Miller, J.H. (1977). Genetic studies of the lac repressor. IV. Mutagenic specificity in the lacI gene of *Escherichia coli*. J Mol Biol *117*, 577-606.

Crosby, B., Prakash, L., Davis, H., and Hinkle, D.C. (1981). Purification and characterization of a uracil-DNA glycosylase from the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res *9*, 5797-5809.

D'Errico, M., Parlanti, E., and Dogliotti, E. (2008). Mechanism of oxidative DNA damage repair and relevance to human pathology. Mutat Res *659*, 4-14.

Datta, A., and Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science *268*, 1616-1619.

David-Cordonnier, M.H., Boiteux, S., and O'Neill, P. (2001). Excision of 8-oxoguanine within clustered damage by the yeast Ogg1 protein. Nucleic Acids Res *29*, 1107-1113.

de Padula, M., Slezak, G., Auffret van Der Kemp, P., and Boiteux, S. (2004). The post-replication repair *RAD18* and *RAD6* genes are involved in the prevention of spontaneous mutations caused by 7,8-dihydro-8-oxoguanine in *Saccharomyces cerevisiae*. Nucleic Acids Res *32*, 5003-5010.

Delbos, F., Aoufouchi, S., Faili, A., Weill, J.C., and Reynaud, C.A. (2007). DNA polymerase η is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. J Exp Med *204*, 17-23.

Deschavanne, P., and Filipski, J. (1995). Correlation of GC content with replication timing and repair mechanisms in weakly expressed *E.coli* genes. Nucleic Acids Res *23*, 1350-1353.

Diaz-Castillo, C., and Golic, K.G. (2007). Evolution of gene sequence in response to chromosomal location. Genetics *177*, 359-374.

Diaz, M., Velez, J., Singh, M., Cerny, J., and Flajnik, M.F. (1999). Mutational pattern of the nurse shark antigen receptor gene (*NAR*) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic hypermutation. Int Immunol *11*, 825-833.

Diffley, J.F. (2004). Regulation of early events in chromosome replication. Curr Biol *14*, R778-786.

Donaldson, A.D., Raghuraman, M.K., Friedman, K.L., Cross, F.R., Brewer, B.J., and Fangman, W.L. (1998). *CLB5*-dependent activation of late replication origins in *S. cerevisiae*. Mol Cell *2*, 173-182.

Dua, R., Levy, D.L., and Campbell, J.L. (1999). Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* polε and its unexpected ability to support growth in the absence of the DNA polymerase domain. J Biol Chem *274*, 22283-22288.

Dumstorf, C.A., Clark, A.B., Lin, Q., Kissling, G.E., Yuan, T., Kucherlapati, R., McGregor, W.G., and Kunkel, T.A. (2006). Participation of mouse DNA polymerase ι in strand-biased mutagenic bypass of UV photoproducts and suppression of skin cancer. Proc Natl Acad Sci U S A *103*, 18083-18088.

Efrati, E., Tocco, G., Eritja, R., Wilson, S.H., and Goodman, M.F. (1997). Abasic translesion synthesis by DNA polymerase β violates the "A-rule". Novel types of nucleotide incorporation by human DNA polymerase β at an abasic lesion in different sequence contexts. J Biol Chem *272*, 2559-2569.

Feng, J., Dai, X., Xiang, Q., Dai, Z., Wang, J., Deng, Y., and He, C. (2010). New insights into two distinct nucleosome distributions: comparison of cross-platform positioning datasets in the yeast genome. BMC Genomics *11*, 33.

Feng, W., Collingwood, D., Boeck, M.E., Fox, L.A., Alvino, G.M., Fangman, W.L., Raghuraman, M.K., and Brewer, B.J. (2006). Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nat Cell Biol *8*, 148-155.

Ferguson, B.M., Brewer, B.J., Reynolds, A.E., and Fangman, W.L. (1991). A yeast origin of replication is activated late in S phase. Cell *65*, 507-515.

Ferguson, B.M., and Fangman, W.L. (1992). A position effect on the time of replication origin activation in yeast. Cell *68*, 333-339.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol *4*, e1000216.

Fijalkowska, I.J., Jonczyk, P., Tkaczyk, M.M., Bialoskorska, M., and Schaaper, R.M. (1998). Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. Proc Natl Acad Sci U S A *95*, 10020-10025.

Flores-Rozas, H., and Kolodner, R.D. (1998). The *Saccharomyces cerevisiae MLH3* gene functions in *MSH3*-dependent suppression of frameshift mutations. Proc Natl Acad Sci U S A *95*, 12404-12409.

Fox, A.K., Tuch, B.B., and Chuang, J.H. (2008). Measuring the prevalence of regional mutation rates: an analysis of silent substitutions in mammals, fungi, and insects. BMC Evol Biol *8*, 186.

Friedberg, E.C. , Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., and Ellenberger, T. (2006). DNA repair and mutagenesis, 2nd ed (Washington, D.C., ASM Press).

Friedman, K.L., Brewer, B.J., and Fangman, W.L. (1997). Replication profile of *Saccharomyces cerevisiae* chromosome VI. Genes Cells *2*, 667-678.

Friedman, K.L., Diller, J.D., Ferguson, B.M., Nyland, S.V., Brewer, B.J., and Fangman, W.L. (1996). Multiple determinants controlling activation of yeast replication origins late in S phase. Genes Dev *10*, 1595-1607.

Fu, Y., Zhu, Y., Zhang, K., Yeung, M., Durocher, D., and Xiao, W. (2008). Rad6-Rad18 mediates a eukaryotic SOS response by ubiquitinating the 9-1-1 checkpoint clamp. Cell *133*, 601-611.

Gangavarapu, V., Prakash, S., and Prakash, L. (2007). Requirement of *RAD52* group genes for postreplication repair of UV-damaged DNA in *Saccharomyces cerevisiae*. Mol Cell Biol *27*, 7758-7764.

Garcia-Diaz, M., Bebenek, K., Krahn, J.M., Pedersen, L.C., and Kunkel, T.A. (2006). Structural analysis of strand misalignment during DNA synthesis by a human DNA polymerase. Cell *124*, 331-342.

Garcia-Diaz, M., and Kunkel, T.A. (2006). Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem Sci *31*, 206-214.

Garg, P., and Burgers, P.M. (2005a). DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol *40*, 115-128.

Garg, P., and Burgers, P.M. (2005b). Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases η and Rev1. Proc Natl Acad Sci U S A *102*, 18361-18366.

Gasche, C., Chang, C.L., Rhees, J., Goel, A., and Boland, C.R. (2001). Oxidative stress increases frameshift mutations in human colorectal cancer cells. Cancer Res *61*, 7444-7448.

Gawel, D., Jonczyk, P., Bialoskorska, M., Schaaper, R.M., and Fijalkowska, I.J. (2002). Asymmetry of frameshift mutagenesis during leading and lagging-strand replication in *Escherichia coli*. Mutat Res *501*, 129-136.

Gerik, K.J., Li, X., Pautz, A., and Burgers, P.M. (1998). Characterization of the two small subunits of *Saccharomyces cerevisiae* DNA polymerase δ. J Biol Chem *273*, 19747-19755.

Gibbs, P.E., McDonald, J., Woodgate, R., and Lawrence, C.W. (2005). The relative roles in vivo of Saccharomyces cerevisiae Polη, Polζ, Rev1 protein and Pol32 in the bypass and mutation induction of an abasic site, T-T (6-4) photoadduct and T-T cis-syn cyclobutane dimer. Genetics *169*, 575-582.

Giroux, C.N., Mis, J.R., Pierce, M.K., Kohalmi, S.E., and Kunz, B.A. (1988). DNA sequence analysis of spontaneous mutations in the *SUP4-o* gene of *Saccharomyces cerevisiae*. Mol Cell Biol *8*, 978-981.

Goldstein, A.L., and McCusker, J.H. (1999). Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast *15*, 1541-1553.

Goodman, M.F., and Fygenson, K.D. (1998). DNA polymerase fidelity: from genetics toward a biochemical understanding. Genetics *148*, 1475-1482.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., and Gasser, S.M. (1996). The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. J Cell Biol *134*, 1349-1363.

Gragg, H., Harfe, B.D., and Jinks-Robertson, S. (2002). Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. Mol Cell Biol *22*, 8756-8762.

Greene, C.N., and Jinks-Robertson, S. (1997). Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. Mol Cell Biol *17*, 2844-2850.

Greene, C.N., and Jinks-Robertson, S. (2001). Spontaneous frameshift mutations in *Saccharomyces cerevisiae*: accumulation during DNA replication and removal by proofreading and mismatch repair activities. Genetics *159*, 65-75.

Gu, C., and Wang, Y. (2007). In vitro replication and thermodynamic studies of methylation and oxidation modifications of 6-thioguanine. Nucleic Acids Res *35*, 3693-3704.

Gueldener, U., Heinisch, J., Koehler, G.J., Voss, D., and Hegemann, J.H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res *30*, e23.

Hadjimarcou, M.I., Kokoska, R.J., Petes, T.D., and Reha-Krantz, L.J. (2001). Identification of a mutant DNA polymerase δ in *Saccharomyces cerevisiae* with an antimutator phenotype for frameshift mutations. Genetics *158*, 177-186.

Hanawalt, P.C., Goulian, M., University of California Los Angeles. Molecular Biology Institute., and ICN Pharmaceuticals inc. (1975). DNA synthesis and its regulation (Menlo Park, Calif., Benjamin).

Haracska, L., Kondratick, C.M., Unk, I., Prakash, S., and Prakash, L. (2001a). Interaction with PCNA is essential for yeast DNA polymerase η function. Mol Cell *8*, 407-415.

Haracska, L., Torres-Ramos, C.A., Johnson, R.E., Prakash, S., and Prakash, L. (2004). Opposing effects of ubiquitin conjugation and SUMO modification of PCNA on replicational bypass of DNA lesions in *Saccharomyces cerevisiae*. Mol Cell Biol *24*, 4267-4274.

Haracska, L., Unk, I., Prakash, L., and Prakash, S. (2006). Ubiquitylation of yeast proliferating cell nuclear antigen and its implications for translesion DNA synthesis. Proc Natl Acad Sci U S A *103*, 6477-6482.

Haracska, L., Washington, M.T., Prakash, S., and Prakash, L. (2001b). Inefficient bypass of an abasic site by DNA polymerase η. J Biol Chem *276*, 6861-6866.

Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L., and Prakash, S. (2000). Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase η. Nat Genet *25*, 458-461.

Harfe, B.D., and Jinks-Robertson, S. (1999). Removal of frameshift intermediates by mismatch repair proteins in *Saccharomyces cerevisiae*. Mol Cell Biol *19*, 4766-4773.

Harfe, B.D., and Jinks-Robertson, S. (2000a). DNA polymerase ζ introduces multiple mutations when bypassing spontaneous DNA damage in *Saccharomyces cerevisiae*. Mol Cell *6*, 1491-1499.

Harfe, B.D., and Jinks-Robertson, S. (2000b). Mismatch repair proteins and mitotic genome stability. Mutat Res *451*, 151-167.

Harfe, B.D., Minesinger, B.K., and Jinks-Robertson, S. (2000). Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast. Curr Biol *10*, 145-148.

Hashem, V.I., Pytlos, M.J., Klysik, E.A., Tsuji, K., Khajavi, M., Ashizawa, T., and Sinden, R.R. (2004). Chemotherapeutic deletion of CTG repeats in lymphoblast cells from DM1 patients. Nucleic Acids Res *32*, 6334-6346.

Hawk, J.D., Stefanovic, L., Boyer, J.C., Petes, T.D., and Farber, R.A. (2005). Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. Proc Natl Acad Sci U S A *102*, 8639-8643.

Heidenreich, E., Eisler, H., Lengheimer, T., Dorninger, P., and Steinboeck, F. (2010). A mutation-promotive role of nucleotide excision repair in cell cycle-arrested cell populations following UV irradiation. DNA Repair (Amst) *9*, 96-100.

Henderson, S.T., and Petes, T.D. (1992). Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol Cell Biol *12*, 2749-2757.

Heun, P., Laroche, T., Raghuraman, M.K., and Gasser, S.M. (2001). The positioning and dynamics of origins of replication in the budding yeast nucleus. J Cell Biol *152*, 385-400.

Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., and Jentsch, S. (2002). *RAD6*-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135-141.

Hsiao, C.L., and Carbon, J. (1981). Characterization of a yeast replication origin (ars2) and construction of stable minichromosomes containing cloned yeast centromere DNA (CEN3). Gene *15*, 157-166.

Huang, M.E., de Calignon, A., Nicolas, A., and Galibert, F. (2000). *POL32*, a subunit of the *Saccharomyces cerevisiae* DNA polymerase δ, defines a link between DNA replication and the mutagenic bypass repair pathway. Curr Genet *38*, 178-187.

Huang, M.E., Le Douarin, B., Henry, C., and Galibert, F. (1999). The *Saccharomyces cerevisiae* protein YJR043C (Pol32) interacts with the catalytic subunit of DNA polymerase α and is required for cell cycle progression in G2/M. Mol Gen Genet *260*, 541-550.

Hubscher, U., and Seo, Y.S. (2001). Replication of the lagging strand: a concert of at least 23 polypeptides. Mol Cells *12*, 149-157.

Ito-Harashima, S., Hartzog, P.E., Sinha, H., and McCusker, J.H. (2002). The tRNA-Tyr gene family of *Saccharomyces cerevisiae*: agents of phenotypic variation and position effects on mutation frequency. Genetics *161*, 1395-1410.

Iwaki, T., Kawamura, A., Ishino, Y., Kohno, K., Kano, Y., Goshima, N., Yara, M., Furusawa, M., Doi, H., and Imamoto, F. (1996). Preferential replication-dependent mutagenesis in the lagging DNA strand in *Escherichia coli*. Mol Gen Genet *251*, 657-664.

Jackson, A.L., Chen, R., and Loeb, L.A. (1998). Induction of microsatellite instability by oxidative DNA damage. Proc Natl Acad Sci U S A *95*, 12468-12473.

Javaid, S., Manohar, M., Punja, N., Mooney, A., Ottesen, J.J., Poirier, M.G., and Fishel, R. (2009). Nucleosome remodeling by hMSH2-hMSH6. Mol Cell *36*, 1086-1094.

Jiang, Y., Wang, Y., and Wang, Y. (2009). In Vitro Replication and Repair Studies of Tandem Lesions Containing Neighboring Thymidine Glycol and 8-Oxo-7,8-dihydro-2'-deoxyguanosine. Chem Res Toxicol *22*, 574-583.

Johnson, R.E., Kondratick, C.M., Prakash, S., and Prakash, L. (1999a). hRAD30 mutations in the variant form of xeroderma pigmentosum. Science *285*, 263-265.

Johnson, R.E., Prakash, S., and Prakash, L. (1999b). Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase Polη. Science *283*, 1001-1004.

Jung, Y.S., Liu, G., and Chen, X. (2010). Pirh2 E3 ubiquitin ligase targets DNA polymerase η for 20S proteasomal degradation. Mol Cell Biol *30*, 1041-1048.

Kakuma, T., Nishida, J., Tsuzuki, T., and Sekiguchi, M. (1995). Mouse MTH1 protein with 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-triphosphatase activity that prevents transversion mutation. cDNA cloning and tissue distribution. J Biol Chem *270*, 25942-25948.

Kalinowski, D.P., Larimer, F.W., and Plewa, M.J. (1995). Analysis of spontaneous frameshift mutations in *REV1* and *rev1-1* strains of *Saccharomyces cerevisiae*. Mutat Res *331*, 149-159.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004). Interaction of human DNA polymerase η with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Mol Cell *14*, 491-500.

Kearsey, S. (1984). Structural requirements for the function of a yeast chromosomal replicator. Cell *37*, 299-307.

Kesti, T., Flick, K., Keranen, S., Syvaoja, J.E., and Wittenberg, C. (1999). DNA polymerase ε catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. Mol Cell *3*, 679-685.

Kim, N., Abdulovic, A.L., Gealy, R., Lippert, M.J., and Jinks-Robertson, S. (2007). Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. DNA Repair (Amst) *6*, 1285-1296.

Kim, N., and Jinks-Robertson, S. (2009). dUTP incorporation into genomic DNA is linked to transcription in yeast. Nature *459*, 1150-1153.

King, N.M., Nikolaishvili-Feinberg, N., Bryant, M.F., Luche, D.D., Heffernan, T.P., Simpson, D.A., Hanaoka, F., Kaufmann, W.K., and Cordeiro-Stone, M. (2005). Overproduction of DNA polymerase η does not raise the spontaneous mutation rate in diploid human fibroblasts. DNA Repair (Amst) *4*, 714-724.

Kirchner, J.M., Tran, H., and Resnick, M.A. (2000). A DNA polymerase ε mutant that specifically causes +1 frameshift mutations within homonucleotide runs in yeast. Genetics *155*, 1623-1632.

Klapacz, J., and Bhagwat, A.S. (2005). Transcription promotes guanine to thymine mutations in the non-transcribed strand of an *Escherichia coli gene*. DNA Repair (Amst) *4*, 806-813.

Knott, S.R., Viggiani, C.J., Tavare, S., and Aparicio, O.M. (2009). Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in *Saccharomyces cerevisiae*. Genes Dev *23*, 1077-1090.

Kokoska, R.J., Bebenek, K., Boudsocq, F., Woodgate, R., and Kunkel, T.A. (2002). Low fidelity DNA synthesis by a y family DNA polymerase due to misalignment in the active site. J Biol Chem *277*, 19633-19638.

Kokoska, R.J., McCulloch, S.D., and Kunkel, T.A. (2003). The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase η and *Sulfolobus solfataricus* Dpo4. J Biol Chem *278*, 50537-50545.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S.A. (2000). A structural model of transcription elongation. Science *289*, 619-625.

Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H., and McMurray, C.T. (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature *447*, 447-452.

Kow, Y.W., Bao, G., Reeves, J.W., Jinks-Robertson, S., and Crouse, G.F. (2007). Oligonucleotide transformation of yeast reveals mismatch repair complexes to be differentially active on DNA replication strands. Proc Natl Acad Sci U S A *104*, 11352-11357.

Krogh, B.O., and Symington, L.S. (2004). Recombination proteins in yeast. Annu Rev Genet *38*, 233-271.

Kroutil, L.C., Register, K., Bebenek, K., and Kunkel, T.A. (1996). Exonucleolytic proofreading during replication of repetitive DNA. Biochemistry *35*, 1046-1053.

Kunkel, T.A. (1985). The mutational specificity of DNA polymerase β during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. J Biol Chem *260*, 5787-5796.

Kunkel, T.A., and Burgers, P.M. (2008). Dividing the workload at a eukaryotic replication fork. Trends Cell Biol *18*, 521-527.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annu Rev Biochem *74*, 681-710.

Lang, G. (2007). Mutation rate variation in the yeast *Saccharomyces cerevisiae* (Cambridge, MA, Harvard University).

Laroche, T., Martin, S.G., Gotta, M., Gorham, H.C., Pryde, F.E., Louis, E.J., and Gasser, S.M. (1998). Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. Curr Biol *8*, 653-656.

Larson, E.D., Iams, K., and Drummond, J.T. (2003). Strand-specific processing of 8-oxoguanine by the human mismatch repair pathway: inefficient removal of 8-oxoguanine paired with adenine or cytosine. DNA Repair (Amst) *2*, 1199-1210.

Lawrence, C.W. (2002). Cellular roles of DNA polymerase ζ and Rev1 protein. DNA Repair (Amst) *1*, 425-435.

Lea, D.E. and Coulson, C. A. (1949). The distribution of the number of mutants in bacterial populations. Journal of Genetics *49*, 264-285.

Lee, D.H., and Pfeifer, G.P. (2008). Translesion synthesis of 7,8-dihydro-8-oxo-2'-deoxyguanosine by DNA polymerase η in vivo. Mutat Res *641*, 19-26.

Lee, G.S., Savage, E.A., Ritzel, R.G., and von Borstel, R.C. (1988). The base-alteration spectrum of spontaneous and ultraviolet radiation-induced forward mutations in the *URA3* locus of *Saccharomyces cerevisiae*. Mol Gen Genet *214*, 396-404.

Lehner, K., and Jinks-Robertson, S. (2009). The mismatch repair system promotes DNA polymerase ζ-dependent translesion synthesis in yeast. Proc Natl Acad Sci U S A *106*, 5749-5754.

Lemoine, F.J., Degtyareva, N.P., Kokoska, R.J., and Petes, T.D. (2008). Reduced levels of DNA polymerase δ induce chromosome fragile site instability in yeast. Mol Cell Biol *28*, 5359-5368.

Li, F., Tian, L., Gu, L., and Li, G.M. (2009). Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells. J Biol Chem *284*, 33056-33061.

Li, G., Sudlow, G., and Belmont, A.S. (1998). Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise choreography of condensation/decondensation and nuclear positioning. J Cell Biol *140*, 975-989.

Lin, Q., Clark, A.B., McCulloch, S.D., Yuan, T., Bronson, R.T., Kunkel, T.A., and Kucherlapati, R. (2006). Increased susceptibility to UV-induced skin carcinogenesis in polymerase η-deficient mice. Cancer Res *66*, 87-94.

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. Prog Nucleic Acid Res Mol Biol *22*, 135-192.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. Cell *107*, 91-102.

Lipkin, S.M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A.D., Lynch, H.T., Elliott, R.M., and Collins, F.S. (2000). *MLH3*: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nat Genet *24*, 27-35.

Lopes, M., Foiani, M., and Sogo, J.M. (2006). Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. Mol Cell *21*, 15-27.

Ma, H., Samarabandu, J., Devdhar, R.S., Acharya, R., Cheng, P.C., Meng, C., and Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. J Cell Biol *143*, 1415-1425.

MacAlpine, H.K., Gordan, R., Powell, S.K., Hartemink, A.J., and Macalpine, D.M. (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res *20*, 201-211.

MacNeill, S.A. (2001). DNA replication: partners in the Okazaki two-step. Curr Biol *11*, R842-844.

Macpherson, P., Barone, F., Maga, G., Mazzei, F., Karran, P., and Bignami, M. (2005). 8-oxoguanine incorporation into DNA repeats in vitro and mismatch recognition by MutSα. Nucleic Acids Res *33*, 5094-5105.

Maga, G., Villani, G., Crespan, E., Wimmer, U., Ferrari, E., Bertocci, B., and Hubscher, U. (2007). 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. Nature *447*, 606-608.

Majka, J., and Burgers, P.M. (2003). Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. Proc Natl Acad Sci U S A *100*, 2249-2254.

Malinsky, J., Koberna, K., Stanek, D., Masata, M., Votruba, I., and Raska, I. (2001). The supply of exogenous deoxyribonucleotides accelerates the speed of the replication fork in early S-phase. J Cell Sci *114*, 747-750.

Maliszewska-Tkaczyk, M., Jonczyk, P., Bialoskorska, M., Schaaper, R.M., and Fijalkowska, I.J. (2000). SOS mutator activity: unequal mutagenesis on leading and lagging strands. Proc Natl Acad Sci U S A *97*, 12678-12683.

Martin, S.A., McCarthy, A., Barber, L.J., Burgess, D.J., Parry, S., Lord, C.J., and Ashworth, A. (2009). Methotrexate induces oxidative DNA damage and is selectively lethal to tumour cells with defects in the DNA mismatch repair gene *MSH2*. EMBO Mol Med *1*, 323-337.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. Nature *399*, 700-704.

Maynard, S., Schurman, S.H., Harboe, C., de Souza-Pinto, N.C., and Bohr, V.A. (2009). Base excision repair of oxidative DNA damage and association with cancer and aging. Carcinogenesis *30*, 2-10.

Mazurek, A., Berardini, M., and Fishel, R. (2002). Activation of human MutS homologs by 8-oxo-guanine DNA damage. J Biol Chem *277*, 8260-8266.

McCulloch, S.D., Kokoska, R.J., Garg, P., Burgers, P.M., and Kunkel, T.A. (2009). The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases δ and η. Nucleic Acids Res *37*, 2830-2840.

McCulloch, S.D., and Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res *18*, 148-161.

McCulloch, S.D., Wood, A., Garg, P., Burgers, P.M., and Kunkel, T.A. (2007). Effects of accessory proteins on the bypass of a cis-syn thymine-thymine dimer by *Saccharomyces cerevisiae* DNA polymerase η. Biochemistry *46*, 8888-8896.

McCune, H.J., Danielson, L.S., Alvino, G.M., Collingwood, D., Delrow, J.J., Fangman, W.L., Brewer, B.J., and Raghuraman, M.K. (2008). The temporal program of chromosome replication: genomewide replication in *clb5Δ Saccharomyces cerevisiae*. Genetics *180*, 1833-1847.

McDonald, J.P., Levine, A.S., and Woodgate, R. (1997). The *Saccharomyces cerevisiae RAD30* gene, a homologue of *Escherichia coli dinB* and *umuC*, is DNA damage inducible and functions in a novel error-free postreplication repair mechanism. Genetics *147*, 1557-1568.

McGregor, W.G., Wei, D., Maher, V.M., and McCormick, J.J. (1999). Abnormal, error-prone bypass of photoproducts by xeroderma pigmentosum variant cell extracts results in extreme strand bias for the kinds of mutations induced by UV light. Mol Cell Biol *19*, 147-154.

McHugh, P.J., and Sarkar, S. (2006). DNA interstrand cross-link repair in the cell cycle: a critical role for polymerase ζ in G1 phase. Cell Cycle *5*, 1044-1047.

McIntyre, J., Podlaska, A., Skoneczna, A., Halas, A., and Sledziewska-Gojska, E. (2006). Analysis of the spontaneous mutator phenotype associated with 20S proteasome deficiency in *S. cerevisiae*. Mutat Res *593*, 153-163.

Mellon, I. (2005). Transcription-coupled repair: a complex affair. Mutat Res *577*, 155-161.

Mellon, I., and Hanawalt, P.C. (1989). Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. Nature *342*, 95-98.

Mellon, I., Spivak, G., and Hanawalt, P.C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian *DHFR* gene. Cell *51*, 241-249.

Meneghini, R., Cordeiro-Stone, M., and Schumacher, R.I. (1981). Size and frequency of gaps in newly synthesized DNA of xeroderma pigmentosum human cells irradiated with ultraviolet light. Biophys J *33*, 81-92.

Menoni, H., Gasparutto, D., Hamiche, A., Cadet, J., Dimitrov, S., Bouvet, P., and Angelov, D. (2007). ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A.Bbd nucleosomes. Mol Cell Biol *27*, 5949-5956.

Meselson, M., and Stahl, F.W. (1958). The Replication of DNA in *Escherichia Coli*. Proc Natl Acad Sci U S A *44*, 671-682.

Michaels, M.L., and Miller, J.H. (1992). The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). J Bacteriol *174*, 6321-6325.

Millar, C.B., and Grunstein, M. (2006). Genome-wide patterns of histone modifications in yeast. Nat Rev Mol Cell Biol *7*, 657-666.

Minesinger, B.K., and Jinks-Robertson, S. (2005). Roles of *RAD6* epistasis group members in spontaneous polζ-dependent translesion synthesis in *Saccharomyces cerevisiae*. Genetics *169*, 1939-1955.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665-679.

Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K., and Sugino, A. (1990). A third essential DNA polymerase in *S. cerevisiae*. Cell *62*, 1143-1151.

Nagata, Y., Kawaguchi, G., Tago, Y., Imai, M., Watanabe, T., Sakurai, S., Ihara, M., Kawata, M., and Yamamoto, K. (2005). Absence of strand bias for deletion mutagenesis during chromosomal leading and lagging strand replication in *Escherichia coli*. Genes Genet Syst *80*, 1-8.

Nakai, S., and Mortimer, R.K. (1969). Studies on the genetic mechanism of radiation-induced mitotic segregation in yeast. Mol Gen Genet *103*, 329-338.

Nelson, J.R., Lawrence, C.W., and Hinkle, D.C. (1996). Deoxycytidyl transferase activity of yeast Rev1 protein. Nature *382*, 729-731.

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

Newlon, C.S., Collins, I., Dershowitz, A., Deshpande, A.M., Greenfeder, S.A., Ong, L.Y., and Theis, J.F. (1993). Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. Cold Spring Harb Symp Quant Biol *58*, 415-423.

Ni, T.T., Marsischky, G.T., and Kolodner, R.D. (1999). *MSH2* and *MSH6* are required for removal of adenine misincorporated opposite 8-oxo-guanine in *S. cerevisiae*. Mol Cell *4*, 439-444.

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. Mol Cell *30*, 137-144.

Nieduszynski, C.A., Blow, J.J., and Donaldson, A.D. (2005). The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. Nucleic Acids Res *33*, 2410-2420.

Nieduszynski, C.A., Hiraga, S., Ak, P., Benham, C.J., and Donaldson, A.D. (2007). OriDB: a DNA replication origin database. Nucleic Acids Res *35*, D40-46.

Northam, M.R., Robinson, H.A., Kochenova, O.V., and Shcherbakova, P.V. (2010). Participation of DNA polymerase ζ in replication of undamaged DNA in *Saccharomyces cerevisiae*. Genetics *184*, 27-42.

Nouspikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. Cell Mol Life Sci *66*, 994-1009.

Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W.J., Gerlach, V.L., Friedberg, E.C., Ohmori, H., and Kunkel, T.A. (2000). Fidelity and processivity of DNA synthesis by DNA polymerase κ, the product of the human *DINB1* gene. J Biol Chem *275*, 39678-39684.

Ohya, T., Kawasaki, Y., Hiraga, S., Kanbara, S., Nakajo, K., Nakashima, N., Suzuki, A., and Sugino, A. (2002). The DNA polymerase domain of polε is required for rapid, efficient, and highly accurate

chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. J Biol Chem *277*, 28099-28108.

Opperman, T., Murli, S., Smith, B.T., and Walker, G.C. (1999). A model for a umuDC-dependent prokaryotic DNA damage checkpoint. Proc Natl Acad Sci U S A *96*, 9218-9223.

Osley, M.A., Tsukuda, T., and Nickoloff, J.A. (2007). ATP-dependent chromatin remodeling factors and DNA damage repair. Mutat Res *618*, 65-80.

Pages, V., and Fuchs, R.P. (2003). Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo. Science *300*, 1300-1303.

Pages, V., Johnson, R.E., Prakash, L., and Prakash, S. (2008). Mutational specificity and genetic control of replicative bypass of an abasic site in yeast. Proc Natl Acad Sci U S A *105*, 1170-1175.

Palakodeti, A., Lucas, I., Jiang, Y., Young, D.J., Fernald, A.A., Karrison, T., and Le Beau, M.M. (2010). Impaired replication dynamics at the FRA3B common fragile site. Hum Mol Genet *19*, 99-110.

Park, K., Debyser, Z., Tabor, S., Richardson, C.C., and Griffith, J.D. (1998). Formation of a DNA loop at the replication fork generated by bacteriophage T7 replication proteins. J Biol Chem *273*, 5260-5270.

Pavlov, Y.I., Frahm, C., Nick McElhinny, S.A., Niimi, A., Suzuki, M., and Kunkel, T.A. (2006). Evidence that errors made by DNA polymerase α are corrected by DNA polymerase δ. Curr Biol *16*, 202-207.

Pavlov, Y.I., Mian, I.M., and Kunkel, T.A. (2003). Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. Curr Biol *13*, 744-748.

Pavlov, Y.I., Newlon, C.S., and Kunkel, T.A. (2002). Yeast origins establish a strand bias for replicational mutagenesis. Mol Cell *10*, 207-213.

Payen, C., Fischer, G., Marck, C., Proux, C., Sherman, D.J., Coppee, J.Y., Johnston, M., Dujon, B., and Neuveglise, C. (2009). Unusual composition of a yeast chromosome arm is associated with its delayed replication. Genome Res *19*, 1710-1721.

Pearson, C.G., Shikazono, N., Thacker, J., and O'Neill, P. (2004). Enhanced mutagenic potential of 8-oxo-7,8-dihydroguanine when present within a clustered DNA damage site. Nucleic Acids Res *32*, 263-270.

Petukhova, G., Van Komen, S., Vergano, S., Klein, H., and Sung, P. (1999). Yeast Rad54 promotes Rad51-dependent homologous DNA pairing via ATP hydrolysis-driven change in DNA double helix conformation. J Biol Chem *274*, 29453-29462.

Pierce, M.K., Giroux, C.N., and Kunz, B.A. (1987). Development of a yeast system to assay mutational specificity. Mutat Res *182*, 65-74.

Pink, C.J., and Hurst, L.D. (2009). Timing of replication is a determinant of neutral substitution rates but does not explain slow Y chromosome evolution in rodents. Mol Biol Evol.

Poloumienko, A., Dershowitz, A., De, J., and Newlon, C.S. (2001). Completion of replication map of *Saccharomyces cerevisiae* chromosome III. Mol Biol Cell *12*, 3317-3327.

Popoff, S.C., Spira, A.I., Johnson, A.W., and Demple, B. (1990). Yeast structural gene (*APN1*) for the major apurinic endonuclease: homology to *Escherichia coli* endonuclease IV. Proc Natl Acad Sci U S A *87*, 4193-4197.

Prakash, S., Johnson, R.E., and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu Rev Biochem *74*, 317-353.

Prakash, S., and Prakash, L. (2002). Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. Genes Dev *16*, 1872-1883.

Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA polymerase ε participates in leading-strand DNA replication. Science *317*, 127-130.

Quah, S.K., von Borstel, R.C., and Hastings, P.J. (1980). The origin of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics *96*, 819-839.

Radman, M. (1998). DNA replication: one strand may be more equal. Proc Natl Acad Sci U S A *95*, 9718-9719.

Radman, M. (2005). SOS replication: a distinct DNA replication mechanism which is induced by DNA-damaging treatments? 1970. DNA Repair (Amst) *4*, 732-738.

Raghuraman, M.K., Brewer, B.J., and Fangman, W.L. (1997). Cell cycle-dependent establishment of a late replication program. Science *276*, 806-809.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. Science *294*, 115-121.

Rey, L., Sidorova, J.M., Puget, N., Boudsocq, F., Biard, D.S., Monnat, R.J., Jr., Cazaux, C., and Hoffmann, J.S. (2009). Human DNA polymerase η is required for common fragile site stability during unperturbed DNA replication. Mol Cell Biol *29*, 3344-3354.

Reynolds, A.E., McCarroll, R.M., Newlon, C.S., and Fangman, W.L. (1989). Time of replication of ARS elements along yeast chromosome III. Mol Cell Biol *9*, 4488-4494.

Rhind, N., Yang, S.C., and Bechhoefer, J. (2009). Reconciling stochastic origin firing with defined replication timing. Chromosome Res.

Rossi, M.L., Purohit, V., Brandt, P.D., and Bambara, R.A. (2006). Lagging strand replication proteins in genome stability and DNA repair. Chem Rev *106*, 453-473.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol *194*, 281-301.

Russo, M.T., Blasi, M.F., Chiera, F., Fortini, P., Degan, P., Macpherson, P., Furuichi, M., Nakabeppu, Y., Karran, P., Aquilina, G., *et al.* (2004). The oxidized deoxynucleoside triphosphate pool is a significant contributor to genetic instability in mismatch repair-deficient cells. Mol Cell Biol *24*, 465-474.

Russo, M.T., De Luca, G., Degan, P., and Bignami, M. (2007). Different DNA repair strategies to combat the threat from 8-oxoguanine. Mutat Res *614*, 69-76.

Sabbioneda, S., Bortolomai, I., Giannattasio, M., Plevani, P., and Muzi-Falconi, M. (2007). Yeast Rev1 is cell cycle regulated, phosphorylated in response to DNA damage and its binding to chromosomes is dependent upon *MEC1*. DNA Repair (Amst) *6*, 121-127.

Sabbioneda, S., Minesinger, B.K., Giannattasio, M., Plevani, P., Muzi-Falconi, M., and Jinks-Robertson, S. (2005). The 9-1-1 checkpoint clamp physically interacts with polζ and is partially required for spontaneous polζ-dependent mutagenesis in *Saccharomyces cerevisiae*. J Biol Chem *280*, 38657-38665.

Sabouri, N., Viberg, J., Goyal, D.K., Johansson, E., and Chabes, A. (2008). Evidence for lesion bypass by yeast replicative DNA polymerases during DNA damage. Nucleic Acids Res *36*, 5660-5667.

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 ζ. DNA Repair (Amst) *6*, 1829-1838.

Sale, J.E., Batters, C., Edmunds, C.E., Phillips, L.G., Simpson, L.J., and Szuts, D. (2009). Timing matters: error-prone gap filling and translesion synthesis in immunoglobulin gene hypermutation. Philos Trans R Soc Lond B Biol Sci *364*, 595-603.

Senturker, S., Auffret van der Kemp, P., You, H.J., Doetsch, P.W., Dizdaroglu, M., and Boiteux, S. (1998). Substrate specificities of the Ntg1 and Ntg2 proteins of *Saccharomyces cerevisiae* for oxidized DNA bases are not identical. Nucleic Acids Res *26*, 5270-5276.

Shah, S.N., Hile, S.E., and Eckert, K.A. (2010). Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes. Cancer Res *70*, 431-435.

Shen, H.M., Poirier, M.G., Allen, M.J., North, J., Lal, R., Widom, J., and Storb, U. (2009). The activation-induced cytidine deaminase (AID) efficiently targets DNA in nucleosomes but only during transcription. J Exp Med *206*, 1057-1071.

Shen, X., Jun, S., O'Neal, L.E., Sonoda, E., Bemark, M., Sale, J.E., and Li, L. (2006). *REV3* and *REV1* play major roles in recombination-independent repair of DNA interstrand cross-links mediated by monoubiquitinated proliferating cell nuclear antigen (PCNA). J Biol Chem *281*, 13869-13872.

Shibutani, S., Takeshita, M., and Grollman, A.P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature *349*, 431-434.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. Cell *69*, 457-470.

Shinohara, A., and Ogawa, T. (1998). Stimulation by Rad52 of yeast Rad51-mediated recombination. Nature *391*, 404-407.

Sia, E.A., Kokoska, J.R., Dominska, M., Greenwell, P., and Petes, T.D. (1997). Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol *17*, 2851-2858.

Simpson, L.J., and Sale, J.E. (2003). Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. EMBO J *22*, 1654-1664.

Sinha, N.K., Morris, C.F., and Alberts, B.M. (1980). Efficient in vitro replication of double-stranded DNA templates by a purified T4 bacteriophage replication system. J Biol Chem *255*, 4290-4293.

Skinner, A.M., and Turker, M.S. (2005). Oxidative mutagenesis, mismatch repair, and aging. Sci Aging Knowledge Environ *2005*, re3.

Skoneczna, A., McIntyre, J., Skoneczny, M., Policinska, Z., and Sledziewska-Gojska, E. (2007). Polymerase η is a short-lived, proteasomally degraded protein that is temporarily stabilized following UV irradiation in *Saccharomyces cerevisiae*. J Mol Biol *366*, 1074-1086.

Soria, G., Belluscio, L., van Cappellen, W.A., Kanaar, R., Essers, J., and Gottifredi, V. (2009). DNA damage induced Pol η recruitment takes place independently of the cell cycle phase. Cell Cycle *8*, 3340-3348.

Spell, R.M., and Jinks-Robertson, S. (2004). Determination of mitotic recombination rates by fluctuation analysis in *Saccharomyces cerevisiae*. Methods Mol Biol *262*, 3-12.

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. Nat Genet *41*, 393-395.

Steele, D.F., and Jinks-Robertson, S. (1992). An examination of adaptive reversion i*n Saccharomyces cerevisiae*. Genetics *132*, 9-21.

Stelter, P., and Ulrich, H.D. (2003). Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. Nature *425*, 188-191.

Stevenson, J.B., and Gottschling, D.E. (1999). Telomeric chromatin modulates replication timing near chromosome ends. Genes Dev *13*, 146-151.

Stinchcomb, D.T., Thomas, M., Kelly, J., Selker, E., and Davis, R.W. (1980). Eukaryotic DNA segments capable of autonomous replication in yeast. Proc Natl Acad Sci U S A *77*, 4559-4563.

Storici, F., Lewis, L.K., and Resnick, M.A. (2001). In vivo site-directed mutagenesis using oligonucleotides. Nat Biotechnol *19*, 773-776.

Strand, M., Earley, M.C., Crouse, G.F., and Petes, T.D. (1995). Mutations in the *MSH3* gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *92*, 10418-10421.

Subramanian, P.S., Nelson, D.L., and Chinault, A.C. (1996). Large domains of apparent delayed replication timing associated with triplet repeat expansion at FRAXA and FRAXE. Am J Hum Genet *59*, 407-416.

Sung, P. (1997). Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. Genes Dev *11*, 1111-1121.

Swanson, R.L., Morey, N.J., Doetsch, P.W., and Jinks-Robertson, S. (1999). Overlapping specificities of base excision repair, nucleotide excision repair, recombination, and translesion synthesis pathways for DNA base damage in *Saccharomyces cerevisiae*. Mol Cell Biol *19*, 2929-2935.

Sweder, K.S., and Hanawalt, P.C. (1992). Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. Proc Natl Acad Sci U S A *89*, 10696-10700.

Tanaka, T., and Nasmyth, K. (1998). Association of RPA with chromosomal replication origins requires an Mcm protein, and is regulated by Rad53, and cyclin- and Dbf4-dependent kinases. EMBO J *17*, 5182-5191.

Teytelman, L., Eisen, M.B., and Rine, J. (2008). Silent but not static: accelerated base-pair substitution in silenced chromatin of budding yeasts. PLoS Genet *4*, e1000247.

Theis, J.F., and Newlon, C.S. (1997). The *ARS309* chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. Proc Natl Acad Sci U S A *94*, 10786-10791.

Thomas, D.C., Nguyen, D.C., Piegorsch, W.W., and Kunkel, T.A. (1993). Relative probability of mutagenic translesion synthesis on the leading and lagging strands during replication of UV-irradiated DNA in a human cell extract. Biochemistry *32*, 11476-11482.

Tippin, B., Kobayashi, S., Bertram, J.G., and Goodman, M.F. (2004). To slip or skip, visualizing frameshift mutation dynamics for error-prone DNA polymerases. J Biol Chem *279*, 45360-45368.

Touchon, M., Nicolay, S., Audit, B., Brodie of Brodie, E.B., d'Aubenton-Carafa, Y., Arneodo, A., and Thermes, C. (2005). Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. Proc Natl Acad Sci U S A *102*, 9836-9841.

Tran, H.T., Keen, J.D., Kricker, M., Resnick, M.A., and Gordenin, D.A. (1997). Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol Cell Biol *17*, 2859-2865.

Trinh, T.Q., and Sinden, R.R. (1991). Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. Nature *352*, 544-547.

Ulrich, H.D., and Jentsch, S. (2000). Two RING finger proteins mediate cooperation between ubiquitin-conjugating enzymes in DNA repair. EMBO J *19*, 3388-3397.

Unniraman, S., and Schatz, D.G. (2007). Strand-biased spreading of mutations during somatic hypermutation. Science *317*, 1227-1230.

van der Kemp, P.A., de Padula, M., Burguiere-Slezak, G., Ulrich, H.D., and Boiteux, S. (2009a). PCNA monoubiquitylation and DNA polymerase η ubiquitin-binding domain are required to prevent 8-oxoguanine-induced mutagenesis in *Saccharomyces cerevisiae*. Nucleic Acids Res *37*, 2549-2559.

van der Kemp, P.A., Thomas, D., Barbey, R., de Oliveira, R., and Boiteux, S. (1996). Cloning and expression in *Escherichia coli* of the *OGG1* gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. Proc Natl Acad Sci U S A *93*, 5197-5202.

Van Komen, S., Petukhova, G., Sigurdsson, S., Stratton, S., and Sung, P. (2000). Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. Mol Cell *6*, 563-572.

Veaute, X., and Fuchs, R.P. (1993). Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. Science *261*, 598-600.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., and Grunstein, M. (2002). Histone acetylation regulates the time of replication origin firing. Mol Cell *10*, 1223-1233.

Vongsamphanh, R., Wagner, J.R., and Ramotar, D. (2006). *Saccharomyces cerevisiae* Ogg1 prevents poly(GT) tract instability in the mitochondrial genome. DNA Repair (Amst) *5*, 235-242.

Walters, R.A., Tobey, R.A., and Ratliff, R.L. (1973). Cell-cycle-dependent variations of deoxyribonucleoside triphosphate pools in Chinese hamster cells. Biochim Biophys Acta *319*, 336-347.

Watanabe, K., Tateishi, S., Kawasuji, M., Tsurimoto, T., Inoue, H., and Yamaizumi, M. (2004). Rad18 guides polη to replication stalling sites through physical interaction and PCNA monoubiquitination. EMBO J *23*, 3886-3896.

Watanabe, T., van Geldorp, G., Najrana, T., Yamamura, E., Nunoshiba, T., and Yamamoto, K. (2001). Miscoding and misincorporation of 8-oxo-guanine during leading and lagging strand synthesis in *Escherichia coli*. Mol Gen Genet *264*, 836-841.

Watanabe, Y., Fujiyama, A., Ichiba, Y., Hattori, M., Yada, T., Sakaki, Y., and Ikemura, T. (2002). Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. Hum Mol Genet *11*, 13-21.

Waters, L.S., and Walker, G.C. (2006). The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G(2)/M phase rather than S phase. Proc Natl Acad Sci U S A *103*, 8971-8976.

Watts, F.Z. (2006). Sumoylation of PCNA: Wrestling with recombination at stalled replication forks. DNA Repair (Amst) *5*, 399-403.

Weinreich, M., Palacios DeBeer, M.A., and Fox, C.A. (2004). The activities of eukaryotic replication origins in chromatin. Biochim Biophys Acta *1677*, 142-157.

Wierdl, M., Dominska, M., and Petes, T.D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics *146*, 769-779.

Wittschieben, J., Shivji, M.K., Lalani, E., Jacobs, M.A., Marini, F., Gearhart, P.J., Rosewell, I., Stamp, G., and Wood, R.D. (2000). Disruption of the developmentally regulated *Rev3l* gene causes embryonic lethality. Curr Biol *10*, 1217-1220.

Wood, A., Garg, P., and Burgers, P.M. (2007). A ubiquitin-binding motif in the translesion DNA polymerase Rev1 mediates its essential functional interaction with ubiquitinated proliferating cell nuclear antigen in response to DNA damage. J Biol Chem *282*, 20256-20263.

Woodgate, R. (2001). Evolution of the two-step model for UV-mutagenesis. Mutat Res *485*, 83-92.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. Science *294*, 2357-2360.

Wyrzykowski, J., and Volkert, M.R. (2003). The *Escherichia coli* methyl-directed mismatch repair system repairs base pairs containing oxidative lesions. J Bacteriol *185*, 1701-1704.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavare, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. BMC Genomics *7*, 276.

Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. Genes Cells *7*, 781-789.

Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., and Shirahige, K. (1997). The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. Genes Cells *2*, 655-665.

Yin, S., Deng, W., Hu, L., and Kong, X. (2009). The impact of nucleosome positioning on the organization of replication origins in eukaryotes. Biochem Biophys Res Commun *385*, 363-368.

Yoon, J.H., Prakash, L., and Prakash, S. (2009). Highly error-free role of DNA polymerase η in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. Proc Natl Acad Sci U S A *106*, 18219-18224.

Yoshiyama, K., Higuchi, K., Matsumura, H., and Maki, H. (2001). Directionality of DNA replication fork movement strongly affects the generation of spontaneous mutations in *Escherichia coli*. J Mol Biol *307*, 1195-1206.

Yuan, F., Zhang, Y., Rajpal, D.K., Wu, X., Guo, D., Wang, M., Taylor, J.S., and Wang, Z. (2000). Specificity of DNA lesion bypass by the yeast DNA polymerase η. J Biol Chem *275*, 8233-8239.

Yuasa, M.S., Masutani, C., Hirano, A., Cohn, M.A., Yamaizumi, M., Nakatani, Y., and Hanaoka, F. (2006). A human DNA polymerase η complex containing Rad18, Rad6 and Rev1; proteomic analysis and targeting of the complex to the chromatin-bound fraction of cells undergoing replication fork arrest. Genes Cells *11*, 731-744.

Yung, C.W., Okugawa, Y., Otsuka, C., Okamoto, K., Arimoto, S., Loakes, D., Negishi, K., and Negishi, T. (2008). Influence of neighbouring base sequences on the mutagenesis induced by 7,8-dihydro-8-oxoguanine in yeast. Mutagenesis *23*, 509-513.

Zang, H., Goodenough, A.K., Choi, J.Y., Irimia, A., Loukachevitch, L.V., Kozekov, I.D., Angel, K.C., Rizzo, C.J., Egli, M., and Guengerich, F.P. (2005). DNA adduct bypass polymerization by *Sulfolobus solfataricus* DNA polymerase Dpo4: analysis and crystal structures of multiple base pair substitution and frameshift products with the adduct 1,N2-ethenoguanine. J Biol Chem *280*, 29750-29764.

Zhang, H., and Lawrence, C.W. (2005). The error-free component of the *RAD6/RAD18* DNA damage tolerance pathway of budding yeast employs sister-strand recombination. Proc Natl Acad Sci U S A *102*, 15954-15959.

Zhuang, Z., Johnson, R.E., Haracska, L., Prakash, L., Prakash, S., and Benkovic, S.J. (2008). Regulation of polymerase exchange between Polη and Polδ by monoubiquitination of PCNA and the movement of DNA polymerase holoenzyme. Proc Natl Acad Sci U S A *105*, 5361-5366.

Zou, L., and Stillman, B. (2000). Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. Mol Cell Biol *20*, 3086-3096.

Biography

Sarah V. Mudrak (née Wiley) was born in Tampa, FL, on October 5th, 1981. She stayed in that area until she went to college at Emory University in Atlanta, GA, in 2000. At Emory, she received a bachelor's of science in Biology, with a minor in French Studies. During her time at Emory, she completed an independent research project under the guidance of Sue Jinks-Robertson. The project was focused on translesion synthesis polymerases in yeast and enabled her to earn highest honors on her degree. After graduating from Emory in 2004, she started graduate school in the University Program in Genetics and Genomics at Duke University in Durham, NC. Soon after starting graduate school, Sue Jinks-Robertson moved her lab to Duke, and Sarah once again started working in the Jinks-Robertson lab studying mutagenesis in yeast. Most of the work presented in Chapter 2 has been published in the article "The Polη translesion synthesis DNA polymerase acts independently of the mismatch repair system to limit mutagenesis caused by 7,8-dihydro-8-oxoguanine in yeast" in the journal Molecular and Cellular Biology, 29(19):5316-26, in 2009. She also presented her work at the Genetic Toxicology Gordon Conference in Oxford, UK, in 2007. In May, 2009, she married fellow graduate student Benjamin Mudrak and plans to graduate with him in May, 2010.