Dynamics of the mismatch repair complexes during DNA replication

Joanna Elizabeth Haye

A Dissertation Presented to the Faculty of Princeton University
In Candidacy for the Degree of

Doctor of Philosophy

Recommended for Acceptance by the Department of Molecular Biology

Advisors: Alison Gammie & Mark Rose

September 2014
Acknowledgements

I am truly grateful and fortunate to have had the pleasure of working with Dr. Alison E. Gammie and Dr. Mark D. Rose. Alison has offered the necessary guidance I needed to grow and develop as an independent scientist. She has been an incredible thesis advisor, a mentor and a friend.

I would also like to thank Dr. Virginia Zakian, Dr. Paul Schedl and Dr. James Broach, for helpful discussion and valuable input into my work. Dr. Amy Caudy and Dr. Robyn Tanny for their initial contributions to earlier experiments.

My parents have been tremendously supportive and patient. I thank them for establishing very early in my childhood, the importance of setting goals and giving as much as it takes to accomplish them. Additionally, I am grateful to my friends who have shared with me in this journey. Finally, I would like to thank Victor A. Bertolozzi, for being my partner and friend who has been very supportive in the final stages of this process.
# Table of Contents

Acknowledgements ........................................................................................................ i
Table of Contents ........................................................................................................... ii
A Listing of Figures ......................................................................................................... iv
A Listing of Tables ........................................................................................................... v
Abstract .......................................................................................................................... vi

**Chapter 1: Introduction to DNA mismatch repair** ...................................................... 1
  DNA mismatch repair and strand discrimination in *Escherichia coli* ......................... 1
  MMR in eukaryotes and other prokaryotes .................................................................. 3
  Strand discrimination in eukaryotes and other prokaryotes: nicks and the sliding clamp .......................................................... 6
  Efficient mismatch scanning of the genome ................................................................. 12
  Mismatch repair specificity of MutSα and MutSβ ...................................................... 14

**Chapter 2: Materials & Methods** ........................................................................... 16
  Microbial manipulations and molecular techniques .................................................. 16
  Mutation Accumulation ............................................................................................. 18
  Strain constructions .................................................................................................... 19
  PIP box mutant ............................................................................................................ 19
  Biochemical assays ..................................................................................................... 20
  Synchronization of the Cell Cycle ............................................................................... 21
  Chromatin Immunoprecipitation and Hybridization of Custom Tiling Arrays .......... 21
  ChIP Data Analysis ..................................................................................................... 23
  Fork Progression Analysis ......................................................................................... 23
  Quantitative PCR ........................................................................................................ 24

**Chapter 3: Dynamics of MutS During Replication** .................................................. 25
Abstract .......................................................................................................................... 25
Introduction ...................................................................................................................... 26
Results ............................................................................................................................. 29
  Tracking the leading strand polymerase, Polɛ, during DNA synthesis....................... 29
  Polɛ binds at active origins after release from G1 .................................................... 38
  Polɛ binding is consistent with origin firing timing ................................................... 38
  Persistent Polɛ signal at certain origins may reflect an imprecision of firing times .......... 39
  Polɛ signal intensity is consistent with origin firing efficiency .................................. 39
  Polɛ signal advances bi-directionally away from origins with expected kinetics .......... 40
  Msh2 binds origins and spreads to adjacent regions consistent with the movement of leading strand polymerase during replication .................................................. 46
  MutS signal persists behind the advancing replisome ............................................... 48
  The leading strand DNA polymerase and MutS complex signals coincide at replication origins ........................................................................................................ 52
  MutS tracks with the replication machinery during DNA synthesis in replicate experiments ............................................................................................................. 56
  MutS shows an altered pattern of binding during replication in MMR defective PCNA/Pol30 variants ......................................................................................... 59
A Listing of Figures

Figure 1.1: Mismatch Repair in *E.coli* ................................................................. 3
Figure 1.2: Mismatch repair in eukaryotes.............................................................. 5
Figure 1.3: Crystal structures of Pol30$^{C22Y}$ and Pol30$^{C81R}$ .......................... 11
Figure 1.4: Leading and lagging strand synthesis during DNA replication......... 13
Figure 3.1: Experimental design and flow cytometry ........................................... 31
Figure 3.2: Confirmation of chromatin shearing.................................................. 32
Figure 3.3: Mcm binding at active and inactive origins of replication .................. 33
Figure 3.4: Reproducibility of Polε distribution..................................................... 37
Figure 3.5: Polε dynamics during DNA replication............................................... 41
Figure 3.6: MutS dynamics during DNA replication ............................................. 49
Figure 3.7: MutS and Polε dynamics during DNA replication .............................. 54
Figure 3.8: MutS persistence in the region behind the replisome......................... 55
Figure 3.9: MutS and Polε occupy the same regions during DNA replication..... 58
Figure 3.10: MutS and Polε co-localize in a MMR defective PCNA variant ......... 61
Figure 3.11: Cell synchrony of the PCNA, MutS double mutant strain ............... 63
Figure 3.12: MutS dynamics are altered in the double mutant strain................. 65
Figure 3.13: MutS Mismatch Recognition during DNA Replication................. 74
Figure 4.1: Msh2 is in a 2-fold excess of Msh3 and Msh6................................. 80
Figure 4.2: Msh3 and Msh6 are not reciprocally regulated............................... 82
Figure 4.3: MutSα distribution during DNA replication.................................... 84
A Listing of Tables

Table 1: Yeast strains used in this study ................................................................. 16
Table 2. Plasmids used in this study ......................................................................... 18
Table 3: Primers used for Q-PCR .............................................................................. 18
Table 4: ARSs on Tiling Array with Mcm Binding ....................................................... 34
Table 5. Example 1 - ARS coordinates and polymerase progression rates.............. 44
Table 6. Example 2 - ARS coordinates and polymerase progression rates.............. 44
Table 7: Replisome Progression Rates ....................................................................... 46
Table 8. Mutation accumulation in 210 generations .................................................. 78
DNA mismatch repair (MMR) functions mainly to correct mispaired bases that escape the proofreading activity of the DNA polymerase during replication. Defects in MMR genes have been linked to compromised genome stability and diseases including cancer. MMR is a highly conserved process and the yeast *Saccharomyces cerevisiae* is an ideal model organism to explore aspects of MMR because of the ease of manipulation and homology to the human system. MMR initiates when a mismatch in the DNA helix is recognized by MutS homologs. Subsequent events include excision of the error-containing strand followed by re-synthesis. A critical step in this process is directing repair to the newly synthesized strand, which requires a strand discrimination signal. Current data suggest that transient discontinuities in the DNA backbone, known as nicks, generated during replication serve as the strand discrimination signal. Additionally, histones have the capacity to block mismatch recognition and are known to rapidly assemble behind the replication fork. Thus, there must be a short window of opportunity for the MutS homologs to scan for mismatches and access the strand discrimination signals during replication. To address these unresolved issues, we hypothesize that the MMR machinery tracks with the replisome to allow for efficient scanning and access to the strand discrimination signal. We employed chromatin immunoprecipitation and DNA tiling microarrays (ChIP-chip) to determine the distribution of the eukaryotic MutS complexes during replication. The data indicate that during S-phase of the cell cycle MutS
binds origins of replication and shows bi-directional occupancy of regions flanking the origins over time with timing consistent with fork progression. Importantly, MutS displays the same origin binding and spreading pattern as the leading strand DNA polymerase over multiple experiments. In sum, our data supports the hypothesis of the MMR machinery tracking with the replisome.

There are two MutS complexes that occur in eukaryotes, MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3). Both complexes recognize the different types of mismatches that arise during DNA replication. Reporter constructs have traditionally been utilized to assay the types of mismatches targeted by each complex. With the availability of new techniques, we can now analyze the functions of MutSα and MutSβ on a genome wide scale. In this work, we used next generation sequencing to determine the mutation spectra in strains lacking MSH2, MSH3 or MSH6. Our studies confirm the findings of previous genetic experiments that MutSα and MutSβ are functionally redundant for repair at HPRs; however, each complex is essential for ~6-7% of the mismatches generated during replication. In this work, we provide evidence that both complexes should be in the vicinity of the replisome to ensure that majority of the mutations can be avoided during replication.
Chapter 1: Introduction to DNA mismatch repair

Accurate DNA replication preserves the integrity of the genome within a cell. In eukaryotes, the replicative DNA polymerases DNA Polymerase epsilon (Polɛ) and DNA polymerase delta (Polδ) perform leading and lagging strand synthesis respectively (Garg and Burgers, 2005; Pursell et al., 2007; reviewed in Kunkel and Burgers, 2008; McElhinny et al., 2008). Both polymerases have 3’-5’ exonuclease function which facilitate proofreading of mispaired nucleotides during DNA synthesis (reviewed in Garg and Burgers, 2005). Mismatch repair (MMR) complexes recognize errors that escape proofreading of the polymerases (Kunkel and Erie, 2005). The combined proofreading of the polymerases and recognition of single base mismatches or small insertions/deletion loops (indels) by the MMR complexes function to ensure faithful transmission of genetic information during each round of replication. Failure to transmit genetic information faithfully is fundamental, as defects in DNA MMR genes result in disease including hereditary and sporadic cancers (Peltomaki, 2005).

DNA mismatch repair and strand discrimination in *Escherichia coli*

During replication, the ability of the MMR machinery to recognize mismatches and target the newly synthesized strand for repair is critical. In *Escherichia coli* (*E. coli*), the parent strand is distinguished from the new strand containing the mismatch due to the difference in methylation state of the two strands (Welsh et al., 1987; Lahue and Modrich, 1988; Su et al., 1988). When an incorrect base is
incorporated during replication, it is recognized by the homodimer MutS. In *E. coli*, MutS recognizes single base-base mismatches (Su *et al.*, 1988) and short indels. Upon recognition of a mismatch, MutS recruits the MutL complex. The MutH endonuclease is then activated by MutL and cleaves the transiently unmethylated, newly synthesized strand at DNA adenine methylation sites (5’-GATC) (Langlerouault *et al.*, 1987; Lahue and Modrich, 1988; Su *et al.*, 1988). The helicase, helicase II (UvrD) is then loaded at the site of the cleaved DNA and unwinds the helix (Mechanic *et al.*, 2000). The region containing the mismatch is excised by one of the redundant exonucleases and the template strand in bound by single strand binding protein (Viswanathan *et al.*, 2001). The new strand is resynthesized by DNA pol III and ligated to form an error-free DNA molecule Figure 1.1 (Lahue *et al.*, 1989).
Figure 1.1: Mismatch Repair in *E.coli*

(a) MutS recognition of the mispaired base is followed by recruitment of MutL. The activation of MutH is orchestrated by MutL. This results in cleavage of the newly synthesized, unmethylated strand by MutH. (d-e) UvrD, unwinds the DNA, single strand binding protein binds the parent strand, and the error-containing strand is degraded by an exonuclease. (f) Finally, the degraded strand is re-synthesized by DNA polymerase III and ligation is performed by DNA ligase.

**MMR in eukaryotes and other prokaryotes**

The eukaryotic mismatch repair system is similar to the one found *E. coli* with a few differences. Two mismatch recognition complexes comprised of MutS homologs (Msh) function in eukaryotes. The recognition complexes include
MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3) (reviewed in Kunkel and Erie, 2005). MutSα, recognizes single base-base mismatches and single indel loops while MutSβ, recognizes single as well as larger indel loops (reviewed in Kunkel and Erie, 2005). When a mismatch is recognized by MutS, this results in recruitment and binding of MutLα (Mlh1/Pms1) (Prolla et al., 1994). The latent endonuclease of MutLα (Pms1 subunit) cleaves the newly replicated DNA strand 3’ or 5’ to the mismatch (Kadyrov et al., 2007). The cleavage is dependent upon the presence of a pre-existing nick and proliferating cell nuclear antigen (PCNA/Pol30) (Kadyrov et al., 2007; Pluciennik et al., 2010). The introduction of a nick 5’ of the mismatch provides a substrate for the 5’-3’ Exo1 exonuclease which degrades the new strand (Pluciennik et al., 2010) while single-stranded binding protein, replication protein A (RPA), protects the template strand (reviewed in Oakley and Patrick, 2010). Re-synthesis of the strands by DNA Polδ follows the strand excision step. Finally, ligation is performed by DNA ligase generating a newly synthesized mismatch free region (Johnston and Nasmyth, 1978) (Figure 1.2). It is of interest to note that the strand discrimination steps are equivalent to the eukaryotic system in many prokaryotic organisms that do not have methyl-directed MMR.
Figure 1.2: Mismatch repair in eukaryotes

(a) Mismatch recognition by MutSα

(b) MutLα recruitment

(c) Activation of latent endonuclease in Pms1 Cleavage 3' or 5' to the mismatch

(d) Excision of mismatch containing strand by Exo1. Binding of RPA

(e) Re-synthesis by DNA polymerase and ligation by DNA ligase

Figure 1.2: Mismatch repair in eukaryotes. (a) A mismatch is recognized by MutSα in the presence of PCNA. (b-c) MutLα (Mlh1/Pms1) is recruited and PCNA interactions lead to the activation of the latent endonuclease in the Pms1 subunit of MutLα, which cleaves the mismatch containing strand 3' or 5' of the mismatch. (d) The newly synthesized strand is degraded by the 5'-3' exonuclease Exo1 and RPA binds the parent DNA strand. (e) DNA polymerase re-synthesizes the new error free strand and ligation is performed by DNA ligase.
Strand discrimination in eukaryotes and other prokaryotes: nicks and the sliding clamp.

The method of strand discrimination observed in *E.coli* is distinct from most prokaryotes and all eukaryotes. As mentioned above, discontinuities in the DNA backbone (nicks) and the sliding clamp (β clamp in prokaryotes and PCNA in eukaryotes) function in the strand discrimination process. The evidence implicating nicks and PCNA in eukaryotic strand discrimination is provided below.

Earlier *in vitro* experiments using *Drosophila melanogaster* and human cell extracts demonstrated that a nick is sufficient to direct repair to the mismatch-containing strand (Holmes *et al.*, 1990; Thomas *et al.*, 1991). *In vivo*, nicks are also involved in directing repair to the newly synthesized strand. During DNA replication, DNA Polymerase α (Pol α) first synthesizes a ~10 base RNA primer (Kaufmann and Falk, 1982) and a short stretch of DNA consisting of 20 to 30 bases to initiate both leading and lagging strand synthesis (Nethanel and Kaufmann, 1990; Pursell *et al.*, 2007; McElhinny *et al.*, 2008). Polε and Polδ subsequently perform synthesis on leading and lagging strand, respectively (Pursell *et al.*, 2007; Nick McElhinny *et al.*, 2008). The lagging strand has nicks ~200 bp apart (reviewed in Garg and Burgers, 2005) (Figure 1.4). On the lagging strand the RNA primer is displaced by Polδ. Polδ conducts repeated cycles of strand opening and degradation (idling) (Garg *et al.*, 2004; reviewed in Garg and Burgers, 2005) generating a flap that is cleaved by Flap endonuclease 1 (*RAD27*) (Kao *et al.*, 2004). This process maintains nicks on the newly
synthesized strand that are then sealed by DNA ligase (Johnston and Nasmyth, 1978). The frequency of these nicks generated after primer removal differs on the leading and lagging strands. The leading strand, which is continuously synthesized, has nicks up to ~30 kb apart (Kunkel and Burgers, 2008). This poses a problem for strand discrimination in leading strand repair.

Recent work has provided evidence for how eukaryotes and most prokaryotes distinguish the new strand from the parent strand on the leading strand. Interestingly, there is an additional source of nicks generated during the replication process. On the leading strand incorporation of ribonucleotides (rNMP) occurs every ~1,250 dNMP (McElhinny et al., 2010c). The rNMPs are then cleaved by RNAase H2 (McElhinny et al., 2010a; McElhinny et al., 2010b), increasing the density of nicks during leading strand synthesis, thereby facilitating strand specificity (Ghodgaonkar et al., 2013; Lujan et al., 2013). Consistent with these data, the inactivation of RNase H2 results in a decrease in mismatch repair efficiency on the leading strand (Lujan et al., 2013). On the lagging strand, rNMP incorporation is reduced relative to the leading strand (1rNMP/5,000 dNMP) (McElhinny et al., 2010a; McElhinny et al., 2010b; McElhinny et al., 2010c). Therefore, the primary source of nicks on the lagging strand is not because of rNMP incorporation. The difference in the frequency of nicks results in a difference in efficiency of MMR on the leading strand (McElhinny et al., 2010a) (nicks every ~1,250 bp) relative to the lagging strand (nicks every ~200 bp).
Several lines of evidence support a role for PCNA as a strand discrimination candidate in MMR. First, PCNA functions as the processivity factor for both DNA Polɛ and Polδ and is a replisome component (reviewed in Johnson and O'Donnell, 2005). The PCNA sliding clamp loads at primer-template junctions during replication in an orientation-specific manner (reviewed in Majka and Burgers, 2004; reviewed in Stolimenov and Helleday, 2009). Additionally, PCNA remains bound to nicked DNA post-replication (reviewed in Stolimenov and Helleday, 2009). The retention of PCNA orientation as it diffuses along the DNA helix is thought to also ensure the maintenance of strand specificity (Pluciennik et al., 2010).

Previous experiments established a requirement for PCNA at earlier steps in the MMR process, prior to DNA re-synthesis (Umar et al., 1996). Additionally, PCNA was shown to interact with MutS complexes (Johnson et al., 1996). The sliding clamp was also shown to co-localize with MutS complexes during DNA replication (Kleczkowska et al., 2001). The orientation specific association of PCNA with the DNA helix presumably positions MutLα (Mlh1/Pms1) to cleave the newly synthesized nicked strand rather than the template strand (Pluciennik et al., 2010). The Mlh1 subunit of MutLα interacts with PCNA (Stolimenov and Helleday, 2009). As mentioned above, the latent endonuclease of MutLα can introduce a nick 3’ or 5’ to a mismatch. This function requires a pre-existing nick, the clamp loader - replication factor C (RFC) and the sliding clamp PCNA
(Kadyrov et al., 2007; Pluciennik et al., 2010). The orientation specificity of PCNA, co-localization and interaction with MutS and MutL complexes has established a key role for PCNA in strand discrimination.

Extensive studies have been performed to further elucidate the role of PCNA in MMR (Johnson et al., 1996; Umar et al., 1996; Flores-Rozas et al., 2000). PCNA interacts with numerous proteins that also include the MMR proteins Msh6, Msh3 and Mlh1 (Flores-Rozas et al., 2000; Stolimenov and Helleday, 2009). The interaction of Msh6 and Msh3 with PCNA occurs via a conserved region also found in other PCNA interacting proteins (PIP) (Stolimenov and Helleday, 2009). Mutations in the PIP box of Msh3 and Msh6 show partial MMR defects in vivo (Flores-Rozas et al., 2000). Separation of function variants of PCNA have been utilized in the analysis of MMR. More specifically two such separation of function alleles, identified in the yeast POL30 PCNA gene, have been particularly useful. The pol30-201 and pol30-204 mutations were determined to be replication proficient but MMR deficient (Lau et al., 2002). Both mutations result in modest structural alterations in PCNA (Dieckman et al., 2013). The pol30-201 results in a cysteine (C) to tyrosine (Y) change at amino acid 22 in the protein (Pol30(C22Y)). This change is on the inner region of the trimer that slides over DNA. This variant exhibits a MMR defect that is MutSα specific with no in vitro binding defect. The pol30-204 results in a C to arginine (R) change at amino acid 81 in the protein (Pol30(C81R)). This occurs at the monomer-monomer interface in Pol30 trimer and has a general defect in MMR. The locations of the amino acid changes in the
protein are illustrated in Figure 1.3. In double mutant strains, where

POL30/PCNA encodes a MMR defective allele in combination with mutation in

MSH6 disrupting the PIP binding site (Msh6PIP/Pol30C81R) causes the MutS-PCNA interaction to be lost in vitro (Lau et al., 2002). Further analysis of Msh6PIP/Pol30C81R has shown that the MutS/PCNA interaction is important for co-localization of MutSα with the leading strand polymerase (Hombauer et al., 2011a).

In summary, these experiments provided evidence for how the MutS complexes and other mismatch repair proteins might act in concert to direct repair specifically to the newly synthesized strand. The data are consistent with a model in which the in vitro and in vivo repair is faithfully directed to the new strand due to the presence of nicks in conjunction with PCNA.
Figure 1.3: Crystal structures of Pol30\textsuperscript{C22Y} and Pol30\textsuperscript{C81R}

The above structures depict the position of the amino acid changes in Pol30\textsuperscript{C22Y} and Pol30\textsuperscript{C81R} variants. The image was taken from Dieckman et al., 2013 (Dieckman et al., 2013). The C22Y mutation results in changes in the α-helices lining the central ring (left image). The C81R mutation causes a distortion in a loop at the monomer-monomer interface (right image). Both mutations result in structural changes in PCNA thought to specifically affect MutS interactions.
Efficient mismatch scanning of the genome

The discovery of the strand discrimination signal resolved a long-standing mystery in eukaryotic MMR systems. However another unresolved question is: how is the DNA efficiently scanned for mismatches in the presence of protein blockages? Chromatin efficiently packages eukaryotic DNA; however, the chromatin must be disrupted ahead of the replication fork and reassembled after fork passage. During replication, chromatin-remodeling complexes and other proteins are associated with the replisome, making DNA accessible (discussed in Azvolinsky et al., 2009). However, nucleosomes are known to rapidly reassemble immediately behind the advancing replisome (~250 bp) (Sogo et al., 1986). In the context of a dynamic replication process, the expectation is that MutS complexes should maintain close association with the traveling replisome throughout DNA replication. Interestingly, in human cell extracts MutSo, has been shown to interfere with nucleosome assembly in heteroduplex DNA and several lines of evidence have placed MutS in the vicinity of the replisome during DNA replication (Schoepf et al., 2012). Firstly, in yeast Msh6 was shown by live-cell imaging to co-localize with the leading strand polymerase during S-phase of the cell cycle and this association was PCNA dependent (Hombauer et al., 2011a). Consistent with yeast studies, mass spectrometry studies in human cells have identified MMR components at active replication forks (Sirbu et al., 2013).
Figure 1.4: Leading and lagging strand synthesis during DNA replication

One model for strand discrimination is that removed mis-incorporated ribonucleotides on the leading strand (every ~1250 bp) and nicks on the lagging strand (every ~200 bp) are the strand specificity signals. If the rNMP excision generated nicks on the leading strand were the main source of strand
discrimination, then deleting RNAase H2 should have had a larger effect on mismatch repair efficiency. However, RNAase H2 deletion has only a modest defect in MMR efficiency (Lujan et al., 2013). A third possible source for the signal could be the 3'-OH end of the newly synthesized strand. The 3'-OH of the newly synthesized strand could potentially serve as a proximal strand specificity signal during replication if MutS is closely associated with the replisome throughout S phase. A major objective of this thesis research was to employ high resolution techniques to determine whether MutS moves with the replisome during replication.

Mismatch repair specificity of MutSα and MutSβ

As mentioned previously, MutSα and MutSβ have different specificities for the range of mismatches generated during replication. The spectra of mismatches generated during DNA replication include single base-base mismatches, single nucleotide indel loops at homopolymeric runs (HPRs) and larger indel loops at microsatellites (MS). A simple model suggests that both MutS complexes are present at the replisome to efficiently recognize errors that escape the polymerase and initiate the repair process. The current literature provides conflicting data concerning the presence of both MutS complexes at the replisome. Fluorescence microscopy experiments in S. cerevisiae have demonstrated co-localization of MutSα, but not MutSβ, with the leading strand polymerase (Hombauer et al., 2011a). Whereas, mass spectrometry studies using human cells, have demonstrated that Msh3 (a component of MutSβ), is
present at the replisome (Sirbu et al., 2013). One major objective of this thesis research was to use high resolution techniques to determine whether both MutS complexes travel with the replisome to efficiently capture all mismatches as they emerge from the advancing polymerase.

In this study, we determined the distribution of the mismatch recognition complexes relative to the leading strand polymerase. Using chromatin immunoprecipitation combined with custom DNA tiling microarrays, we examined the dynamics of the both MutSα and MutSβ complexes during DNA replication. This thesis provides evidence that MutS travels with the replisome to capture the full spectrum of mismatches that escape the proof-reading function of the DNA polymerase.
Chapter 2: Materials & Methods

Microbial manipulations and molecular techniques

Yeast strains used in this work are listed in Table 1. Microbial manipulations were conducted according to previously published procedures (Ausubel et al., 1994; Burke et al., 2000). Molecular methods were carried out using standard protocols (Ausubel et al., 1994). Plasmid DNA extractions were performed using the Qiagen procedure (Qiagen Inc., Valencia, CA). Primers were synthesized by Integrated DNA Technologies Inc. (Coralville, IA). Restriction endonuclease digestions and polymerase chain reactions (PCR) were performed using the enzyme manufacturer recommended reaction conditions (New England Biolabs, Beverly, MA).

Table 1: Yeast strains used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGY1198</td>
<td>( \text{MAT} , \text{msh6}\Delta::\text{HphMX} , \text{ade2-1} , \text{trp1-1} , \text{ura3-1} , \text{leu2-3,112} )</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>AGY1196</td>
<td>( \text{MAT} , \text{msh3}\Delta::\text{HphMX} , \text{ade2-1} , \text{trp1-1} , \text{ura3-1} , \text{leu2-3,112} )</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>MY10733</td>
<td>( \text{MAT} , \text{MSH2-myc::KanMX} , \text{ura3-1} , \text{leu2-3,112} , \text{can1-100} , \text{his3-11,15} , \text{rad5-5} , \text{bar1 POL2-3xHA::LEU2} )</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>MY9704</td>
<td>( \text{MAT} , \text{msh6}\Delta::\text{KanMX} , \text{ade2-1} , \text{leu2-3,112} , \text{ura3-1} , \text{ade2-1} , \text{trp1-1} , \text{his3-11,15} , \text{LYS2 CAN1 RAD5} )</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>MY10567</td>
<td>( \text{MAT} , \text{MCM4-HA::TRP1} , \text{ade2-1} , \text{leu2-3,112} , \text{can1-100} , \text{his3-11,15} , \text{rad5-5} , \text{trp1-1} , \text{bar1::HISG} )</td>
<td>Stephen Bell Laboratory, MIT</td>
</tr>
<tr>
<td>MY11870</td>
<td>( \text{MAT} , \text{MSH2-myc:: NatMX POL2-3xHA::LEU2 pol30-201} , \text{bar1} , \text{ura3-1} , \text{leu2-3,112} , \text{can1-100} , \text{his3-11,15} , \text{rad5-5} )</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>MY13569</td>
<td>( \text{MAT} , \text{MSH6-myc::KanMX msh3}\Delta::\text{HphMX} , \text{ade2-1} , \text{trp1-1} , \text{ura3-1} , \text{leu2-3,112} , \text{can1-100} )</td>
<td>This study</td>
</tr>
</tbody>
</table>
### Chapter 2: Materials & Methods

<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY12073</td>
<td>MATα MSH6-myc::KanMX ura3-1 leu2-3,112 can1-100 his3-11,15</td>
<td>This study</td>
</tr>
<tr>
<td>MY12328</td>
<td>MATα MSH3-myc::NatMX ade2-1 trp1-1 ura3-1 his3-11,15 hom3-10</td>
<td>This study</td>
</tr>
<tr>
<td>MY14082</td>
<td>MATα MSH6-myc::KanMX MSH3-myc::KanMX ura3-1 leu2-3,112 his3-11,15</td>
<td>This study</td>
</tr>
<tr>
<td>MY14176</td>
<td>MATα MSH6-myc::KanMX msh2Δ::URA3 his3-11,15 trp1-1</td>
<td>This study</td>
</tr>
<tr>
<td>MY13798</td>
<td>MATα MSH6-myc::KanMX msh3Δ:: HphMX leu2-3,112 his3-11,15</td>
<td>This study</td>
</tr>
<tr>
<td>MY13797</td>
<td>MATα MSH3-myc::KanMX msh6Δ::HphMX ura3-1 leu2-3,112 his3-11,15 trp1-1 hom3-10 POL2-3xHA::LEU2</td>
<td>This study</td>
</tr>
<tr>
<td>MY14146</td>
<td>MATα MSH3-myc::KanMX msh2Δ:: URA3 leu2-3,112 his3-11,15 trp1-1</td>
<td>This study</td>
</tr>
<tr>
<td>MY11002</td>
<td>MATα MSH2-myc::KanMX msh6Δ::HphMX ade2-1 trp1-1 leu2-3,112</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>MY11003</td>
<td>MATα MSH2-myc::KanMX msh6Δ::HphMX msh3Δ::NatMX ade2-1 trp1-1 leu2-3,112</td>
<td>Gammie laboratory</td>
</tr>
<tr>
<td>MY14903</td>
<td>MATα MSH2-myc::KanMX msh3Δ::NatNX trp1-1 ura3-1 leu2-3,112</td>
<td>This study</td>
</tr>
<tr>
<td>MY14904</td>
<td>MATα MSH2-myc::KanMX MSH3-myc::KanMX MSH6-myc::KanMX ura3-1 his3-11,15 leu2-3,112</td>
<td>This study</td>
</tr>
<tr>
<td>MY13803</td>
<td>MATα MSH6-myc::KanMX POL2-3xHA::LEU2 can1-100 trp1-1 his3-11,15</td>
<td>This study</td>
</tr>
<tr>
<td>MY12327</td>
<td>MATα POL2-3xHA::LEU2 MSH2-myc::KanMX trp1-1 ura3-1 hom3-10 leu2-3,112 his3-11,15 bar1::HISG</td>
<td>This study</td>
</tr>
<tr>
<td>MY15095</td>
<td>MATα POL2-3xHA::LEU2 MSH2-myc::KanMX pol30-204 rad5-5 leu2-3,112 his3-11,15 trp1-1 ura3-1 bar1::HISG can1-100 msh6-F33AF34A</td>
<td>This study</td>
</tr>
<tr>
<td>MY14147</td>
<td>MATα MSH3-myc::KanMX MSH6-myc::KanMX msh2Δ:: URA3 leu2-3,112 his3-11,15</td>
<td>This study</td>
</tr>
</tbody>
</table>

All strains are derived from W303 and confirmed to be wild-type at the RAD5 locus by PCR (unless indicated) and at the CAN1 locus by canavanine resistance assays.
Table 2. Plasmids used in this study

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Relevant markers</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCORE</td>
<td><em>K. lactis URA3 KanMX amp</em>′</td>
<td>(Stoci and Resnick, 2006)</td>
</tr>
<tr>
<td>pRS415</td>
<td><em>CEN6 ARSH4 LEU2 amp</em>′</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pMSH6-MYC</td>
<td><em>MSH6-myc CEN6 ARSH4 LEU2 amp</em>′</td>
<td>Gammie Laboratory</td>
</tr>
<tr>
<td>pFA6a-13xmyc</td>
<td><em>KanMX amp</em>′</td>
<td>(Longtine <em>et al</em>., 1998)</td>
</tr>
<tr>
<td>pFA6a-3xHA</td>
<td><em>KanMX amp</em>′</td>
<td>(Longtine <em>et al</em>., 1998)</td>
</tr>
</tbody>
</table>

Table 3: Primers used for Q-PCR

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS305 origin fwd</td>
<td>GAT TGA GGC CAC AGC AAG AC</td>
</tr>
<tr>
<td>ARS305 origin rev</td>
<td>TCA CAC CGG ACA GTA CAT GA</td>
</tr>
<tr>
<td>ARS305 2.5 kb upstream fwd</td>
<td>AGA TCA AAG GCC AAA AGC AA</td>
</tr>
<tr>
<td>ARS305 2.5 kb upstream rev</td>
<td>TTG GAG ATT GCG AGT CAC AG</td>
</tr>
<tr>
<td>ARS305 2.5 kb downstream rev</td>
<td>AAC GCC ACT ACT CCA AAT GC</td>
</tr>
<tr>
<td>ARS305 2.5 kb downstream fwd</td>
<td>GTA GTC ACC GAA GCG GTA GG</td>
</tr>
<tr>
<td>ARS607 origin fwd</td>
<td>CCC AGC GTA AGG TAA ATA TTA TGG</td>
</tr>
<tr>
<td>ARS607 origin rev</td>
<td>CCT ACT GTG CCG AAT AAT GTG</td>
</tr>
<tr>
<td>ARS607 2.5 kb upstream fwd</td>
<td>TCA AAG GTC CTC CAA ACG AG</td>
</tr>
<tr>
<td>ARS607 2.5 kb upstream rev</td>
<td>TGA TCT GCA TGG GTT CGT TA</td>
</tr>
<tr>
<td>ARS607 2.5 kb downstream fwd</td>
<td>CGG ACT CTT CTC CTG CAA AG</td>
</tr>
<tr>
<td>ARS607 2.5 kb downstream rev</td>
<td>GCA GAT TCA CTG CTC CAA CA</td>
</tr>
</tbody>
</table>

Mutation Accumulation

For the mutation accumulation and whole genome sequencing, the knockout strains were propagated in rich medium (YEPD) for ~210 generations with bottlenecks every ~21 generations. Genomic DNA preparations, whole genome sequencing, and data analyses were as described previously (Lang *et al*., 2013).
Strain constructions

We identified a polymorphism present in *MSH6* in the strain MY10733 used in most of the MutS/Polε tracking experiments. This mutation resulted in a missense mutation changing a methionine to isoleucine amino acid position 896. To determine if the mutation affects MMR function, the full-length *MSH6* harboring the missense mutation (*msh6-M896I*) was amplified from MY10733 and cloned into pRS415 (*CEN/ARS LEU2*). The plasmid carrying the *msh6-M896I* gene was verified by sequence analysis and used to transform an *msh6* deletion strain (*msh6Δ*) MY9704. MMR function was tested by performing a canavanine resistance assays described previously (Gammie *et al.*, 2007). We determined that the *msh6-M896I* missense mutation did not alter MMR function.

To chromosomally tag *MSH6* and *MSH3* at the C-terminal coding regions, the myc or HA epitope tag and the kanamycin gene was amplified from the pFA6-x13myc or pFA6x3HA plasmids as described previously (Longtine *et al.*, 1998). PCR amplified products were transformed into wild-type W303. Integration was confirmed by PCR amplification of the epitope tag and sequencing. Western blot analysis was employed to confirm expression of the tag. Finally, the functionality of the fusions were confirmed by performing mismatch repair assays.

**PIP box mutant**

To produce the *msh6-F33AF34A*, we employed *in vivo* site directed mutagenesis (Storici and Resnick, 2006), with some modifications. Yeast strains were
transformed with the core cassette, from the pCore plasmid \((\text{KanMX-URA3})\) targeting the \(\text{MSH6}\) locus and selected on synthetic medium lacking uracil (SC–URA) plates. Integration of the pCORE cassette at the \(\text{MSH6}\) locus was confirmed by PCR. After confirmation, PCR reactions were performed to generate two fragments spanning full length \(\text{MSH6}\) and 500 bp up and downstream of \(\text{MSH6}\). The resulting two fragments overlapped in the region containing the desired nucleotide changes and were used to transform the strain with the pCORE cassette integrated to replace \(\text{MSH6}\) \((\text{msh6}\Delta::\text{KanMX-URA3})\).

Transformants were screened for replacement of the pCORE cassette using 5-fluro-orotic acid (FOA). Specifically, transformants were grown overnight at 30°C on YEPD and then replica-printed to medium containing FOA. Single FOA resistant colonies were selected for analysis. To confirm the mutation within the full length \(\text{MSH6}\) locus was amplified by PCR and sequenced. We confirmed a 5’-GCTGCC-3’ change from the WT sequence 5’-TTTTTC-3’. These specific nucleotide changes resulted in replacement of two conserved phenylalanines at positions 33 and 34 with alanines in the PIP box of Msh6.

**Biochemical assays**

Protein extractions and immunoblotting were conducted as described previously (Arlow \textit{et al.}, 2013). To confirm that each myc-tagged mismatch repair protein was immunoprecipitated during the time course experiments, a time point in S phase was collected in duplicate (for example, at 100 min). The sample was processed using identical conditions as other samples for ChIP-chip and the
crosslinks reversed. Next western blot analysis was performed. The proteins were detected using antibodies for the myc epitope, present in each protein.

**Synchronization of the Cell Cycle**

To achieve synchrony, cultures are grown to mid-exponential phase (~0.5 OD_{600}) in rich medium at 30°C. The cells were then shifted to 18°C to slow the growth rate and arrested in the G1 phase of the cell cycle with 10 μg/ml α-factor. The strains used lack the Bar1 protease that degrades α-factor and are thus more sensitive to the pheromone, allowing for near complete arrest in G1 (Chan and Otte, 1982). The cells were released from G1 arrest by washing the cells and resuspending in fresh medium. The washes and cell cycle arrest release were conducted at 18°C to maintain the slow growth rate, allowing for better resolution during S phase. Samples were taken initially at 6 or 10-minute intervals for ChIP-chip, followed by 30 minute intervals for continued analysis of DNA content. Samples were collected for each time-point; the cells were cross-linked with freshly made 4% para-formaldehyde (final concentration ~1%) and flash frozen in liquid nitrogen. Aliquots from each time point were processed and analyzed by flow cytometry to determine which samples correspond to the cells in S phase.

**Chromatin Immunoprecipitation and Hybridization of Custom Tiling Arrays**

An aliquot of the fixed samples were processed for flow cytometry as previously described with modifications (Haase and Reed, 2002). Briefly, cells were incubated with RNAase and SYTOX Green. The samples were analyzed using the Becton-Dickinson LSRII Multi laser analyzer. Samples corresponding to S-
phase of the cell cycle were then processed for ChIP-chip. The samples were processed for ChIP by mechanically disrupting the cell walls using a Fastprep®-24 instrument (MP Biomedicals LLC). Fastprep allows for rapid and thorough disruption of cells. Cell disruption is followed by sonication to generate DNA fragments averaging ~500 bp (Covaris S220 Focused-ultrasonicator). A portion of each sample was retained as the input DNA. The remaining sample was split into two equal fractions and the cross-linked protein/DNA complexes were immunoprecipitated with antibodies conjugated to agarose beads. The cross-links are reversed and the DNA was purified from the ChIPs and inputs. Routinely, a portion of the sample was quantified by Q-PCR. Both the input and ChIP DNA were amplified using ligation-mediated PCR (Carey et al., 2009) and labeled with fluorescent dyes Cy3 and Cy5 respectively. Differential labeling permits the assessment of sequences that are significantly enriched in the ChIP relative to the input. Finally, labeled samples were hybridized to custom DNA tiling arrays with 15,000 probes (Agilent technologies). The 24 regions represented include Autonomously Replicating Sequences (ARSs) or origins of replication and 20 kb of flanking DNA. Also included are telomere sequences, silenced loci, tRNA genes, highly transcribed genes and long terminal repeats, which are known replication pause sites (Aqvolsky et al., 2009). Additionally, mono-, di- and tri-nucleotide repeats were included because these regions are associated with insertion/deletion loops requiring mismatch repair (Streisin.G et al., 1966).
ChIP Data Analysis

An Agilent DNA microarray scanner was used to detect the fluorescence intensities for Cy3 and Cy5. The data was subsequently processed using Agilent Feature Extraction Software. Algorithms were used to correct for background and normalize the data and the log₂ ratios of ChIP/input were calculated for the adjusted data. The data were then uploaded into the Princeton University microarray database (PUMAdb). PUMAdb has features that facilitate data visualization and processing for a variety of programs. The data files processed in PUMAdb were converted to files compatible with Integrated Genome Browser (IGB) (Affymetrix). IGB was used to represent the data as log₂ ratios for individual experiments (Version 8.0.1)(Nicol et al., 2009).

For replicate experiments we used the software Chipper, described by Gibbons et al., (Gibbons et al., 2005) with minor modifications by Matthew Cahn (Princeton University). Chipper analysis generates the significance (p-values) of enrichment obtained from individual experiments by using variance stabilization and not log₂ ratios. The averaged data were visualized using IGB.

Fork Progression Analysis

To determine the rate of fork progression we measured the leading edge of the peak for each time point. The difference the time points (tp) were then taken and divided by the time interval (Rate of fork progression = (tp₁ - tp₂)/ time interval).

For two experiment sets, the average rate of fork progression of ~423 bp/min and
~ 438 kb/min respectively was determined for all ARSs by analyzing the bi-directional movement. In a few instances the leading edge was not discernable due to background signal therefore measurement was not possible.

**Quantitative PCR**

Quantification of DNA enrichment in the ChIP and the input was performed using Q-PCR (Power SYBR® Green PCR master mix, Applied biosystems). Three technical replicates were performed for each time point. Samples were amplified for and the threshold cycles (Ct) were determined using the Sequence Detection System, SDS version 2.3 software (Applied Biosystems). Primers used are listed in Table 3.
Chapter 3: Dynamics of MutS During Replication

Abstract

We examined the dynamics of the leading strand polymerase (Polε) and MutS during S-phase of the cell cycle using chromatin immunoprecipitation and DNA tiling arrays (ChIP-chip). By analyzing Polε dynamics, we determined the timing and efficiencies of origin firing as well as replisome progression rates are consistent with published data. Additionally, we established that MutS complexes have a distribution similar to Polε and remain associated with the replisome throughout S-phase of the cell cycle, presumably to retain access to the strand specificity signals. We show that MutS remains associated with the replisome in the presence of a PCNA variant (Pol30C22Y) that had previously been shown to have compromised mismatch repair function. Additionally, using another mismatch repair defective PCNA variant (Pol30C81R) in combination with the PCNA interacting region mutant of Msh6, we show that MutS is in the region of actively replicating DNA, but displays an abnormal distribution pattern. In summary, these experiments show that MutS complexes remain closely associated with the replisome during DNA replication and PCNA does not appear to be the sole determinant for localizing MutS to the replisome.
**Introduction**

Mismatch repair functions to increase the fidelity of DNA replication, recognizing errors that escape the proofreading of the DNA polymerases. A critical step during MMR is distinguishing the newly replicated DNA strand containing the error from the parental DNA strand. *E. coli* has a method to discriminate between the parental DNA and the newly replicated error-containing strand based on methylation differences between the two strands (Welsh *et al.*, 1987; Lahue and Modrich, 1988; Su *et al.*, 1988). It has become clear that methyl-directed MMR is not a widely used mechanism. For example, there is no DNA methylation in yeast and recognition of differential methylation patterns has been ruled out as a strand discrimination signal in higher eukaryotes (Petranovic *et al.*, 2000; Drummond and Bellacosa, 2001). As mentioned above, PCNA loading at nicks is important for distinguishing the newly synthesized, mismatch containing strand from the parent strand for both leading and lagging strand synthesis (Pluciennik *et al.*, 2010).

Accessibility to the nicks in the newly synthesized strand represents a major consideration for mismatch recognition repair. Taking into consideration the complicated nature of the *in vivo* DNA environment during replication, it is important to note that the newly replicated DNA is thought to quickly re-assemble into nucleosomes behind the replisome (Sogo *et al.*, 1986) after which, a mismatch is presumably less accessible to the MMR proteins. This potential for
diminished accessibility is based on the fact that it is known that nucleosomes without replication/repair associated histone modifications (Li et al., 2009) and other DNA bound proteins can block movement of MutS complexes along DNA (Mendillo et al., 2005; Pluciennik and Modrich, 2007). However, the ability of hMutSα to displace modified nucleosomes has been demonstrated (Javaid et al., 2009). Additionally, Chromatin assembly factor 1 interacts with MutS presumably to control chromatin assembly after fork progression until mismatch scanning is accomplished (Schoepf et al., 2012). Taken together, the most efficient mechanism for detecting mismatches and for accessing the strand specificity signal would involve a close association between MutS and the replisome within the region where chromatin has been cleared.

Current data are consistent with MutS localizing to the replisome. Mass spectrometry analyses of proteins at active replication forks, have identified hMSH2, hMSH6 and hMSH3 in cell lines (Sirbu et al., 2013). In yeast, live cell-imaging demonstrated co-localization of MutSα with Polε during S phase (Hombauer et al., 2011a). Additionally, yeast cells with MSH6 expressed from an S phase-cyclin promoter are MMR proficient, but are MMR deficient if expressed from a G2/M-cyclin promoter (Hombauer et al., 2011b). Specifically, these experiments demonstrated that early replicating genes did not accumulate mutations in the presence of the MSH6 S phase-cyclin promoter fusion; however, with the MSH6 G2/M-cyclin promoter fusion, of the genes analyzed, early
replicated genes accumulated mutations, but genes in late replication regions did not accumulate mutations.

To date, these experiments are consistent with a model that MMR proteins are recruited to the replisome at some point during DNA synthesis. They also indicate that the most effective MMR function is dependent on S phase expression and is especially important earlier in S phase. However, they fail to provide a clear picture of what occurs throughout S phase in vivo. Clearly, in vivo the accessibility of DNA is of critical importance. Therefore, to further address this issue we asked: how do the MMR proteins in eukaryotes access the strand specificity signal in vivo? We hypothesized that the MMR machinery remains in the vicinity of the replisome throughout S phase to allow for efficient scanning and access to the strand discrimination signal. The data presented in this work are consistent with the model that MutS tracks with the replisome during replication.
Results

Tracking the leading strand polymerase, Polε, during DNA synthesis

DNA replication initiates at origins of replication and the replisome moves bi-directionally away from the origins. In yeast, the origins are known as Autonomously Replicating Sequences (ARS). To determine if eukaryotic MutS is a replisome component, we first needed suitable controls to delineate the position of replication origins and to indicate the position of the advancing replisome during DNA replication. Mcm2-7 helicase (minichromosome maintenance, Mcm) has been established as a good predictor of potential ARSs (Wyrick et al., 2001), and is a component of pre-replication complexes that associate with origins at the G1 to S-phase transition (Aparicio et al., 1997). Our experiments employed a hemagglutinin (HA) tagged Mcm4 (Aparicio et al., 1997), a subunit of the helicase, to indicate replication origins. Additionally, in a separate strain, Polε, the leading strand polymerase, served as the control for replisome progression. Specifically, we used a HA tagged version of Pol2, the catalytic subunit of Polε (Bell Laboratory). All ChIP experiments included an untagged control for non-specific precipitation of certain DNA regions. This allowed for exclusion of regions of the genome that generate high background signal (e.g. highly transcribed regions were recently shown to have an increased tendency to ChIP) (Teytelman et al., 2013).

The S-phase time course experiments are similar to the approach taken in the whole genome analysis of the Go ichi ni san (GINS) complex progression, a
known component of the replisome (Sekedat et al., 2010). Experiments were performed at 18°C to slow the replication process, with the goal of improving resolution for ChIP-chip. The cells were first arrested in G1 with \( \alpha \)-factor. Next, the \( \alpha \)-factor was removed from the cultures, releasing the cells from arrest. Samples were collected at 6 or 10-minute intervals and fixed with paraformaldehyde. Initial arrest and progression through the cell cycle were monitored by flow cytometry. S phase cells were subsequently processed for ChIP-chip (Figure 3.1).
Figure 3.1: Experimental design and flow cytometry

Figure 3.1: A. Outline of the experimental design for cell synchrony. B. The flow cytometry data shown are representative of the cell cycle arrest and synchrony for the ChIP-chip experiments. Cells were arrested in G1 with α-factor at 18°C. The cells were washed twice to remove α-factor, resuspended in fresh medium and returned to 18°C. Samples were removed at the indicated time points and analyzed by Becton-Dickinson LSII Multi laser analyzer. The amount of SYTOX Green bound to DNA was measured by flow cytometry analysis. The data are shown in the graph where the x-axis represents DNA content per cell (haploid, 1N and diploid, 2N), the z-axis represents time points (min) after release from arrest and the y-axis denotes cell count. A total of 100,000 cells were collected for each time. The samples used in the ChIP-chip analysis are indicated, including 0 min and 108 - 144 min corresponding to complete G1 arrest and S-phase of the cell cycle respectively.

Cells fixed in S-phase were sonicated to generate DNA fragments averaging ~500 bp (Figure 3.2). After confirmation of fragment sizes, samples were labeled and hybridized to DNA tiling arrays.
Figure 3.2: Confirmation of chromatin shearing

Figure 3.2. Confirmation of chromatin Shearing. Formaldehyde fixed samples were sonicated to shear chromatin and crosslinked proteins were immunoprecipitated. After crosslink reversal, 5 µl of each IP and the input were run on a 1.5% agarose gel stained with SYBR® safe. The data shown is representative of the size fragments generated. Each lane is a single time point for each IP.

The DNA tiling arrays used for analysis included 24 regions of interest in the yeast genome, on 15,000 probes. There are 65 ARSs represented: 53 are confirmed ARSs, 3 have previously been identified as likely ARSs, 6 proposed ARSs and 3 as dubious ARSs (Siow et al., 2012). As mentioned above, the input and ChIP DNA were differentially labeled and hybridized to the arrays. The dye intensities were measured and the data were subsequently processed, corrected for background and normalized as described in the Materials and Methods. Next, log₂ ratios of the ChIP/Input data were visualized.

Previous experiments have examined Pol2 during S-phase using ChIP-chip (Hiraga et al., 2005; Lou et al., 2008). These experiments utilized oligonucleotides arrays of Chromosome VI at a single time point during S-phase or Chromosome III early in S-phase. Earlier studies have also examined Polc using ChIP-PCR (Aparicio et al., 1997). In this study, the approach taken was to
examine 65 ARSs throughout the yeast genome, over 6 time points (Table 4). By measuring the peaks corresponding to Mcm binding, we are able to mark the specific coordinates of the ARSs (Figure 3.3). We find that the maximum signal of each Mcm peak corresponds to the ~100-150 bp of each ARS element (Siow et al., 2012).

**Figure 3.3: Mcm binding at active and inactive origins of replication**

The number of times Mcm binding was observed in five independently performed experiments is indicated in Table 4. Previous microarray studies that have identified the ARSs included on the DNA tiling arrays are also indicated. These experiments are visualized using the Integrated Genome Browser, IGB, program (Affymetrix) and are depicted as peaks correspond to log2 ratios (ChIP/Input). The y-axis is set at 3 (or a ~8-fold maximum signal). Mcm signal is purple and the no tag control for non-specific binding is depicted in black. Black bars below the data denote position of ARSs in the genome database. Origins bound by Mcm helicase: active origins (ARS305 and ARS306) and adjacent inactive origins (ARS301, ARS303, ARS304 and ARS320).
results demonstrate the reliability of the Mcm4 ChIP method in defining origins of replication.

**Table 4: ARSs on Tiling Array with Mcm Binding**

<table>
<thead>
<tr>
<th>ARS</th>
<th>Number of experiments</th>
<th>Mcm loading observed</th>
<th>Other Microarray study identifying ARS *</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS104</td>
<td>5</td>
<td></td>
<td>a, c, d, e</td>
</tr>
<tr>
<td>ARS105</td>
<td>5</td>
<td></td>
<td>a, d</td>
</tr>
<tr>
<td>I-72μ</td>
<td>4</td>
<td></td>
<td>a, b</td>
</tr>
<tr>
<td>ARS216</td>
<td>5</td>
<td></td>
<td>a, d</td>
</tr>
<tr>
<td>ARS217</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARS300</td>
<td>4</td>
<td></td>
<td>c, d</td>
</tr>
<tr>
<td>ARS301</td>
<td>5</td>
<td></td>
<td>a, c, d</td>
</tr>
<tr>
<td>ARS302</td>
<td>3</td>
<td></td>
<td>a, c, d</td>
</tr>
<tr>
<td>ARS320</td>
<td>5</td>
<td></td>
<td>a, c, d</td>
</tr>
<tr>
<td>ARS303</td>
<td>5</td>
<td></td>
<td>a, c, d</td>
</tr>
<tr>
<td>ARS304</td>
<td>5</td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>ARS305</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS306</td>
<td>5</td>
<td></td>
<td>a, b, d, e</td>
</tr>
<tr>
<td>ARS308</td>
<td>5</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td>ARS309</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS315</td>
<td>5</td>
<td></td>
<td>a, b, d, e</td>
</tr>
<tr>
<td>ARS316</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS317</td>
<td>4</td>
<td></td>
<td>a, c, e</td>
</tr>
<tr>
<td>ARS318</td>
<td>5</td>
<td></td>
<td>a, c, e</td>
</tr>
<tr>
<td>ARS319</td>
<td>4</td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>ARS412</td>
<td>4</td>
<td></td>
<td>a, c, d</td>
</tr>
<tr>
<td>ARS413</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS414</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS415</td>
<td>4</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS416</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>IV-477μ</td>
<td>1</td>
<td></td>
<td>a, d, e</td>
</tr>
<tr>
<td>ARS433</td>
<td>4</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>ARS514</td>
<td>5</td>
<td></td>
<td>c, d, e</td>
</tr>
<tr>
<td>ARS522</td>
<td>5</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>ARS523</td>
<td>5</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>ARS603</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS607</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS608</td>
<td>5</td>
<td></td>
<td>a, d</td>
</tr>
<tr>
<td>ARS609</td>
<td>5</td>
<td></td>
<td>a, c, d, e</td>
</tr>
<tr>
<td>ARS610</td>
<td>5</td>
<td></td>
<td>c, d</td>
</tr>
<tr>
<td>VII-17μ</td>
<td>5</td>
<td></td>
<td>a, b, d</td>
</tr>
<tr>
<td>ARS</td>
<td>Number of experiments</td>
<td>Mcm loading observed</td>
<td>Other Microarray study identifying ARS</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>ARS702</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS919</td>
<td>5</td>
<td></td>
<td>a, c, d, e</td>
</tr>
<tr>
<td>ARS920</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1012</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1013</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1116</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1117</td>
<td>4</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1127</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1208</td>
<td>5</td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>ARS1209</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1210</td>
<td>3</td>
<td></td>
<td>b, c, d</td>
</tr>
<tr>
<td>ARS1211</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1211.5</td>
<td>3</td>
<td></td>
<td>a, b, d</td>
</tr>
<tr>
<td>ARS1213</td>
<td>4</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1407</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1412</td>
<td>5</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>XIV-180</td>
<td>1</td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>ARS1517</td>
<td>5</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>ARS1518</td>
<td>2</td>
<td></td>
<td>a, c, d</td>
</tr>
<tr>
<td>ARS1519</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1520</td>
<td>4</td>
<td></td>
<td>a, c-e</td>
</tr>
<tr>
<td>ARS1619</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1621</td>
<td>5</td>
<td></td>
<td>a-e</td>
</tr>
<tr>
<td>ARS1624</td>
<td>4</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>ARS1625</td>
<td>4</td>
<td></td>
<td>a-d</td>
</tr>
<tr>
<td>ARS1633</td>
<td>5</td>
<td></td>
<td>a, b, d, e</td>
</tr>
<tr>
<td>XVII-428</td>
<td>3</td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>XVII-553</td>
<td>4</td>
<td></td>
<td>a, b, d</td>
</tr>
<tr>
<td>ARS1634</td>
<td>5</td>
<td></td>
<td>a, b, d</td>
</tr>
</tbody>
</table>

* (a.) (Feng et al., 2006); (b.) (Raghuraman et al., 2001); (c.) (Wyrick et al., 2001); (d.) Xu et al. (2006)(Xu et al., 2006); (e.) (Yabuki et al., 2002). 
|<sup>*</sup>No ARS name has been assigned, chromosomal location obtained from OriDB (Siow et al., 2012).|

These experiments confirm Mcm4 signal functions as an excellent indicator for potential origins. Additionally, we determined that the Polɛ signal throughout S-phase functions as a good metric for replisome progression and that the method provides the precision required to determine whether MutS tracks with the replisome. Representative examples of the dynamics of Pol2 (Polɛ) progression
over multiple time points throughout S phase are shown below. Importantly, we established that the cell synchrony ChIP-chip experiments are able to reproducibly detect Pol€ binding to origins and movement bi-directionally to adjacent regions as the cells progress through S-phase of the cell cycle. Figure 3.4 illustrates the reproducibility of the ChIP experiments performed. The ~60 kb region of chromosome XVI, includes ARS1619 and ARS1633, an efficient and an inefficient ARS respectively.
Figure 3.4: Reproducibility of Polε distribution

Three independently performed experiments are depicted above. The tiling array data were visualized using IGB and are depicted as peaks correspond to log₂ ratios (ChIP/Input). For each experiment set the y-axis is set at 3 (or ~8-fold maximum). The region corresponds to chromosome XVI which includes ARS1619 and ARS1633.
Taken together, the data establish that chromatin immunoprecipitation in combination with DNA tiling arrays is an effective method for tracking the leading strand polymerase, Polε, at origins of replication during DNA synthesis.

**Polε binds at active origins after release from G1**

In G1 synchronized cells, all potential ARSs are detected by Mcm binding (Figure 3.3; Table 4), consistent with previous studies (Wyrick et al., 2001). In contrast, the Polε signal emerges only at active ARSs after release from G1. In these experiments, Polε is only detected at active origins, for example, ARS301, ARS303 and ARS320 are inactive origins (Vujic et al., 1999) and are bound by Mcm, but not by Polε (Figure 3.5 A). In contrast, ARS305 and ARS306 are known active origins (Fangman and Brewer, 1991; Raghuraman et al., 2001; Yabuki et al., 2002) and exhibit Mcm as well as Polε binding (Figure 3.5 A).

**Polε binding is consistent with origin firing timing**

We determined that in the ChIP-chip experiments, the Polε signal has the resolution to track origin firing times. Active origins of replication are known to fire at different times during S-phase (Raghuraman and Brewer, 2010). For this reason, we included origins that fire in early, middle and late S-phase on the DNA tiling array. A representative example of a late firing ARS is depicted in Figure 3.5 C where Polε shows a bi-directional migration from ARS609 at a much later time than earlier firing adjacent ARS607 and ARS608. ARS609, ARS607,
and ARS608 origins have previously been established as late inefficient, early efficient and early inefficient respectively (Siow et al., 2012).

**Persistent Polɛ signal at certain origins may reflect an imprecision of firing times**

The Polɛ signal typically diminishes at the origins after fork progression (Figure 3.5 A and Figure 3.5 E). However, in some cases signal is observed at the ARS at later time points (Figure 3.5 B and Figure 3.5 D). One explanation for this signal is lack of synchrony. Alternatively, this may be a consequence of some replication origins firing with less precision during the cell cycle (reviewed in Santocanale et al., 1999; Czajkowsky et al., 2008; Raghuraman and Brewer, 2010) (Figure 3.5 A-F). We favor the second explanation because there are examples from the same experiment where at certain origins the signal diminishes (e.g. at ARS305 and ARS306 in Figure 3.5 A), suggesting excellent synchrony, while at other ARSs the signal persists (e.g. ARS315, Figure 3.5 B), consistent with less precision of firing during the cell cycle.

**Polɛ signal intensity is consistent with origin firing efficiency**

The Polɛ signal observed at the ARSs is also in agreement with the known differences in firing efficiency of each ARS (Siow et al., 2012). Efficiency is an indication of the frequency of origin firing in a cell population during S-phase (Raghuraman and Brewer, 2010). This origin efficiency is observable in the distribution of detectable signal in the region flanking the origin as the cells
progress through S-phase. ARS315 is highly efficient and fires in ~90% of each S-phase of the cell cycle (Poloumienko et al., 2001). In this study, ARS315 exhibits robust Polε signal initially localized and then migrating away from the ARS over time (Figure 3.5 B). Additionally, efficiently firing ARS607 (Figure 3.5 C), which fires >85% of the cell cycles (Friedman et al., 1997) also displays a particularly robust Polε signal at the origin, representing a larger percentage of cells in the population with an activated origin. The adjacent ARS608 serves as an example previously determined to be less efficient, known to fire in <10% of the cell cycles (Friedman et al., 1997). In keeping with this finding, we observe a reproducibly reduced signal in the vicinity of ARS608. On chromosome XII (Figure 3.5 D), highly efficient ARSs such as ARS1207 and ARS1209, exhibit strong signal and Polε signal reproducibly migrates from these ARSs. Mcm signal is reduced at ARS1208 (positioned at CEN12) and it does not function as a site of initiation of fork progression, presumably because the adjacent ARS1207 and ARS1209 fire early and efficiently.

**Polε signal advances bi-directionally away from origins with expected kinetics**

We also observed that the Polε signal throughout S-phase is consistent with the advancing replisome kinetics. Specifically, Polε signal advances bi-directionally to adjacent regions as the cells progress through S-phase of the cell cycle (Figure 3.5 A-E). The ~100 kb region on chromosome IV represents a good example of Polε binding ARSs with the signal migrating to flanking regions up and downstream in subsequent time points (Figure 3.5 E).
Figure 3.5: Polε dynamics during DNA replication
Using data derived from two independently performed experiments, we determined the rate of fork progression. This was accomplished by measuring the leading edge of each peak at each time-point. For the origins of replication
included on the tiling arrays, distributed across 9 chromosomes, the average rate was calculated as ~430 base pairs per minute. Measurement of the leading edge both upstream and downstream of the origin was achieved only in cases where the leading edge was clearly defined (and not obscured by background signal) (Table 5 and Table 6).

### Table 5. Example 1 - ARS coordinates and polymerase progression rates

<table>
<thead>
<tr>
<th>ARS</th>
<th>Start coordinate</th>
<th>End coordinate</th>
<th>Upstream Rate bp/min</th>
<th>Downstream Rate bp /min</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS216</td>
<td>486273</td>
<td>487206</td>
<td>324</td>
<td>442</td>
</tr>
<tr>
<td>ARS305</td>
<td>39364</td>
<td>40030</td>
<td>459</td>
<td>572</td>
</tr>
<tr>
<td>ARS306</td>
<td>73714</td>
<td>75628</td>
<td>544</td>
<td>-</td>
</tr>
<tr>
<td>ARS309</td>
<td>131422</td>
<td>132972</td>
<td>390</td>
<td>-</td>
</tr>
<tr>
<td>ARS315</td>
<td>224155</td>
<td>225686</td>
<td>379</td>
<td>251</td>
</tr>
<tr>
<td>ARS413</td>
<td>329172</td>
<td>330356</td>
<td>473</td>
<td>359</td>
</tr>
<tr>
<td>ARS414</td>
<td>407607</td>
<td>408567</td>
<td>433</td>
<td>-</td>
</tr>
<tr>
<td>ARS1</td>
<td>461923</td>
<td>463443</td>
<td>271</td>
<td>494</td>
</tr>
<tr>
<td>ARS514</td>
<td>286912</td>
<td>288320</td>
<td>375</td>
<td>299</td>
</tr>
<tr>
<td>ARS607</td>
<td>198351</td>
<td>200345</td>
<td>466</td>
<td>499</td>
</tr>
<tr>
<td>ARS919</td>
<td>341049</td>
<td>342909</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ARS920</td>
<td>356021</td>
<td>358206</td>
<td>-</td>
<td>445</td>
</tr>
<tr>
<td>ARS1012</td>
<td>374194</td>
<td>375755</td>
<td>400</td>
<td>455</td>
</tr>
<tr>
<td>ARS1207</td>
<td>139167</td>
<td>140905</td>
<td>566</td>
<td>345</td>
</tr>
<tr>
<td>ARS1209</td>
<td>156018</td>
<td>157598</td>
<td>285</td>
<td>-</td>
</tr>
<tr>
<td>ARS1211</td>
<td>230313</td>
<td>231945</td>
<td>487</td>
<td>-</td>
</tr>
<tr>
<td>ARS1213</td>
<td>372681</td>
<td>373997</td>
<td>522</td>
<td>417</td>
</tr>
<tr>
<td>ARS1407</td>
<td>88901</td>
<td>90373</td>
<td>315</td>
<td>464</td>
</tr>
<tr>
<td>ARS1619</td>
<td>417440</td>
<td>419042</td>
<td>571</td>
<td>485</td>
</tr>
<tr>
<td>ARS1621</td>
<td>510933</td>
<td>512395</td>
<td>389</td>
<td>334</td>
</tr>
<tr>
<td>ARS1622</td>
<td>563042</td>
<td>564352</td>
<td>478</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 6. Example 2 - ARS coordinates and polymerase progression rates

<table>
<thead>
<tr>
<th>ARS</th>
<th>Start coordinate</th>
<th>End coordinate</th>
<th>Upstream Rate bp/min</th>
<th>Downstream Rate bp /min</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS216</td>
<td>486273</td>
<td>487206</td>
<td>353</td>
<td>416</td>
</tr>
<tr>
<td>ARS305</td>
<td>39364</td>
<td>40030</td>
<td>353</td>
<td>509</td>
</tr>
<tr>
<td>ARS306</td>
<td>73714</td>
<td>75628</td>
<td>551</td>
<td>-</td>
</tr>
<tr>
<td>ARS309</td>
<td>131422</td>
<td>132972</td>
<td>382</td>
<td>-</td>
</tr>
<tr>
<td>ARS</td>
<td>Start coordinate</td>
<td>End coordinate</td>
<td>Upstream Rate bp/min</td>
<td>Downstream Rate bp/min</td>
</tr>
<tr>
<td>---------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>ARS315</td>
<td>224155</td>
<td>225686</td>
<td>303</td>
<td>302</td>
</tr>
<tr>
<td>ARS413</td>
<td>329172</td>
<td>330356</td>
<td>470</td>
<td>-</td>
</tr>
<tr>
<td>ARS414</td>
<td>407607</td>
<td>408567</td>
<td>432</td>
<td>-</td>
</tr>
<tr>
<td>ARS415</td>
<td>434534</td>
<td>436004</td>
<td>442</td>
<td>-</td>
</tr>
<tr>
<td>ARS1</td>
<td>461923</td>
<td>463443</td>
<td>277</td>
<td>513</td>
</tr>
<tr>
<td>ARS607</td>
<td>198351</td>
<td>200345</td>
<td>475</td>
<td>512</td>
</tr>
<tr>
<td>ARS919</td>
<td>341049</td>
<td>342909</td>
<td>438</td>
<td>-</td>
</tr>
<tr>
<td>ARS920</td>
<td>356021</td>
<td>358206</td>
<td>498</td>
<td>448</td>
</tr>
<tr>
<td>ARS1012</td>
<td>374194</td>
<td>375755</td>
<td>404</td>
<td>443</td>
</tr>
<tr>
<td>ARS1207</td>
<td>139167</td>
<td>140905</td>
<td>559</td>
<td>324</td>
</tr>
<tr>
<td>ARS1209</td>
<td>156018</td>
<td>157598</td>
<td>273</td>
<td>448</td>
</tr>
<tr>
<td>ARS1211</td>
<td>230313</td>
<td>231945</td>
<td>551</td>
<td>491</td>
</tr>
<tr>
<td>ARS1213</td>
<td>372681</td>
<td>373997</td>
<td>570</td>
<td>508</td>
</tr>
<tr>
<td>ARS1407</td>
<td>88901</td>
<td>90373</td>
<td>310</td>
<td>414</td>
</tr>
<tr>
<td>ARS1619</td>
<td>417440</td>
<td>419042</td>
<td>565</td>
<td>544</td>
</tr>
</tbody>
</table>

Previous experiments have applied different methods, or examined different replisome components to monitor the rate of replication fork progression. These studies show that replication fork rate may vary with temperature, nutrient availability or drug treatment (Table 7). The rate of fork progression determined in this analysis is consistent with the doubling time at 18°C, the temperature at which the experiments were performed. Higher progressions rates were previously reported from experiments performed at higher temperatures. For example Sekedat et al., experiments were performed at room temperature and replication fork rate is 1.6 kb/min, consistent with a higher doubling time (Table 7).
### Table 7: Replisome Progression Rates

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Rate kb/min</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proline medium</td>
<td>0.56</td>
<td>(Rivin and Fangman, 1980)</td>
</tr>
<tr>
<td>New DNA synthesis</td>
<td>2.9</td>
<td>(Raghuraman et al., 2001)</td>
</tr>
<tr>
<td>New DNA synthesis</td>
<td>2.8</td>
<td>(Yabuki et al., 2002)</td>
</tr>
<tr>
<td>Psf2 subunit of GINS complex</td>
<td>1.6</td>
<td>(Sekeda et al., 2010)</td>
</tr>
<tr>
<td>New DNA synthesis in MMS</td>
<td>~0.45</td>
<td>(Zhong et al., 2013)</td>
</tr>
<tr>
<td><strong>Pol2 subunit of Polε at 18°C</strong></td>
<td><strong>~0.43</strong></td>
<td><strong>This study</strong></td>
</tr>
</tbody>
</table>

In summary, Polε signal is detected at active origins with expected timing (e.g. early, middle, or late firing during S phase) and firing efficiencies (frequency of firing per cell cycle). Additionally, the movement of the Polε signal is consistent with the advancing replisome during S phase. These experiments established the foundation for a comparative analysis of the eukaryotic MutS signal during S phase.

**Msh2 binds origins and spreads to adjacent regions consistent with the movement of leading strand polymerase during replication.**

With good controls for origin position and for replication fork migration established, we next aimed to address the position of MutS complexes during S phase. As mentioned above, MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3) are the two MutS complexes in eukaryotes that function in post-replicative mismatch repair (reviewed in Kunkel and Erie, 2005). Since Msh2 is the invariable
component of both complexes, we tagged Msh2 with the myc epitope, facilitating the detection of all MutS within the cell.

To determine the MutS dynamics during S phase we prepared samples as described above. Examples of the MutS signal distribution are shown in Figure 3.6. Like Polε, MutS signal is observed at origins in S-phase, but not G1, and the signal progresses away from the ARSs bi-directionally during S phase (Figure 3.6 A-E). Importantly, the distribution and movement of the MutS signal occurs is similar to the Polε signal (compare images in Figure 3.5 and Figure 3.6). For example in Figure 3.6 Panel A, MutS signal originates at ARS305 and ARS306 and migrates bi-directionally from each ARS. While Mcm signal is at ARS301, ARS303, ARS304 and ARS320 these inactive ARSs do not exhibit MutS signal (or Polε signal, Figure 3.5). The data show that MutS binds and moves bi-directionally away from active origins (Mcm and Polε both bound), but does not bind inactive origins (Mcm only).

Similar to Polε (Figure 3.5), the time of appearance and intensity of the MutS signal is consistent with the expected timing and firing efficiencies of the ARSs (Figure 3.6). For example, the intensity and distribution of MutS signal at ARS315 is consistent with the efficient, early firing of this origin (Figure 3.6 B). In addition, for the early-efficient ARS607 origin, the MutS signal is robust compared to the early-inefficient ARS608 origin (Figure 3.6 C). Correspondingly, the early-efficient ARS1207 and ARS1209 origins exhibit an early MutS signal that progressively
migrates bi-directionally away from both origins while the ARS1208 remains inactive (Figure 3.6 D) as was observed for Polε. After fork progresses bi-directionally from the ARSs, some signal can be observed in the position of the ARS. This is consistent with the imprecise nature of ARS firing during S-phase.

**MutS signal persists behind the advancing replisome**

There is one potential difference between the MutS (Figure 3.6) and Polε signal (Figure 3.5). The dissimilarity is the persistence of the MutS signal in regions behind the advancing replisome. While the Polε signal clears as the replisome advances, there is still some MutS signal in the region presumably behind the replication fork. The persistent MutS signal is significantly higher than is observed in the “no tag” control (ChIP data using cells that do not express the myc epitope, but are put through the same ChIP procedure used for MutS-myc).

The MutS and Polε distribution during S phase indicates that MutS loads at origins and migrates bi-directionally with a timing similar to DNA polymerase; however, some differences in the persistence behind the replisome might exist. In the next section, we explored the similarities and differences between the MutS and Polε in the same strain during S phase.
Figure 3.6: MutS dynamics during DNA replication
Chapter 3: Results

D

Time (min)

30
24
18
12
6
0
G1
G1
G1
CC

ARS1207  ARS1208  ARS1209

ARS607  ARS608  ARS609

CHIP Input

MutS
No tag
Mcm
Figure 3.6. MutS dynamics during DNA replication. Time course ChIP-chip experiment for Msh2-myc (MutS). Each row corresponds to ChIP-chip signal during G1 or to the time point series taken during S phase (0 to 30 minutes or 0 to 50 minutes). The tiling array data were visualized using the Integrated Genome Browser, IGB, program (Affymetrix) and are depicted as peaks correspond to Log2 ratios (ChIP/Input). The y-axis is set to 2.5 (or a ~6-fold maximum signal). Black bars below the data denote position of ARSs in the genome databases. The red bars represent ARSs not found in the genome database. CC-represents the chromosomal coordinates x10^3. Mcm4 (Mcm) signal, shown in purple, is visible at potential origins during G1 and non-specific signals shown in black are detected in the no tag control IP during G1. MutS signal (blue) is detected at active origins. Representative regions are shown including: (A) active origins (ARS305 and ARS306) and adjacent inactive origins (ARS301, ARS303, ARS304 and ARS320), (B) an early-efficient origin (ARS315), (C) adjacent early-efficient (ARS607), early-inefficient, (ARS608), late-inefficient, ARS609, (D) early-efficient origins (ARS1207 and ARS1209) flanking an inactive origin (ARS1208), (E) a ~100 kb region of chromosome IV where the advancing forks converge.

The leading strand DNA polymerase and MutS complex signals coincide at replication origins

The analyses of independent ChIP experiments show similarities in the general dynamics of the Polɛ and MutS complexes during S phase. We reasoned that if MutS is a component of the replisome or is consistently associated with the replisome, then these complexes should exhibit signal that overlaps with Polɛ at the same times during S-phase. It was therefore important to perform the time course experiment with both tagged proteins in the same strain under identical synchrony conditions to get the resolution needed to understand the similarities and differences in the signals.
To obtain signal corresponding to both MutS and Polɛ, fixation conditions were optimized. Fixation for 45 minutes with freshly prepared para-formaldehyde facilitated the immunoprecipitation of both proteins. Each sample was divided into equal portions and immunoprecipitated: one fraction with α-HA for Polɛ and the other with α-myc for MutS. By performing ChIP-chip for both proteins for each time point, we could determine the precise position corresponding to MutS relative to the leading strand DNA polymerase (Figure 3.7).

We examined the occupancy of both Polɛ and MutS at all of the 65 ARSs on the tiling arrays. Representative images are shown in Figure 3.7. The MutS signal can be observed in regions occupied by Polɛ at each corresponding time point (Figure 3.7). These data are consistent with MutS complexes loading at origins with a timing similar to the leading strand polymerase and associating with the replisome throughout DNA replication. Interestingly, there is a persistent MutS signal localized in the region behind the advancing replisome (Figure 3.8). The no tag control does not exhibit the signal that is present in MutS, suggesting that this signal is significant.
Figure 3.7: MutS and Polε dynamics during DNA replication

Figure 3.7. MutS and Polε dynamics during DNA replication. Cells were fixed for 45 minutes, the samples were divided and ChIP was performed with specified antibodies to detect Polε-HA (green) and MutS-myc (blue). The distribution was visualized using IGB, as log2 ratios (ChIP/Input) with the scale set at 2.5 (~ 6 fold increase) for all samples. Each row corresponds to ChIP-chip signal during G1 or to the time point series taken during S phase (0 to 30 minutes or 0 to 50 minutes). Black bars below the data denote position of ARSs in the genome databases. The red bars represent ARSs not found in the genome database. CC-represents the chromosomal coordinates x10⁷. Mcm4 (Mcm) signal, shown in purple, is visible at potential origins during G1 and non-specific signals shown in black are detected in the no tag control IP during G1. Representative regions are shown including: (A) ARS1407, where there is an initial unidirectional distribution of signal that is followed by bi-directional progression at later time point, and (B) the early-efficient ARS1012 and the early-inefficient ARS1013.
Figure 3.8: MutS persistence in the region behind the replisome. As described above, cells were arrested in G1 of the cell cycle using α-factor at 18°C to slow the process of DNA replication. The cells were then released from arrest, samples were collected at 10 minute intervals and crosslinked with paraformaldehyde. The samples were then immunoprecipitated with antibodies conjugated to agarose beads that were specific for the tagged protein. The crosslinks were reversed and immunoprecipitated DNA was labeled and hybridized to custom DNA tiling arrays. A sample from each time point was analyzed by flow cytometry, to monitor progression through the cell cycle (see methods). The y-axis is set at 2.5 (or ~6 fold maximum) for each row. CC represents the chromosome coordinates x 10^3. MutS (blue), Polɛ (green), no tag (black) and Mcm (purple). Region where signal persists is indicated with a red rectangle.
Chapter 3: Results

MutS tracks with the replication machinery during DNA synthesis in replicate experiments

As mentioned previously, the timing of origin binding and the movement of the replisome are qualitatively reproducible (Figure 3.4). To measure this quantitatively, we utilized a method that allows for the analysis of replicate experiments to determine the statistical significance of the signal. This was accomplished using the software Chipper (Gibbons et al., 2005) to assign $p$-values to the ChIP signals from the tiling arrays. We averaged three biological replicates for Mcm, Polε, MutS and the no tag control. The data are visualized as negative of the log$_{10}$ of the calculated $p$-values (Figure 3.9 A-C). Because the Mcm signal is so significant, the graphs are scaled to 27 (reflecting a $p$-value $\sim 10^{-27}$ for the most signal values); whereas for MutS and the “no tag” control the graphs are set to a maximum of 10 and the Polε graphs to 20.

We examined all potential ARSs represented on the tiling arrays in three independently performed experiments. Of the 65 putative ARSs on the arrays, 60 exhibited Mcm signal. These 60 potential ARSs were used to calculate the occupancy by Polε and MutS (see figure legend of Figure 3.9 for details). A total of 55 ARSs ($\sim 91\%$) showed Polε signal and 54 ARSs (90%) exhibited MutS signal. Importantly, the co-occupancy of Polε and MutS at origins and adjacent regions during replication is significant. For example, highly significant signal for both Polε and MutS is observed at ARS1207 and ARS1209 but not at ARS1208.
(Figure 3.9 A.). Additionally, ARS1213 and ARS416 also illustrate the overlapping, highly significant signal from Polε and MutS (Figure 3.9 B and C).

The origin that fails to show co-occupancy exhibits a weak Polε signal. Because the Polε signal is consistently more robust than MutS, it is not surprising that a region with diminished Polε signal would not show significant MutS signal. ARSs where no Polε or MutS signal is detected are origins that do not fire or have previously been established as inefficient and firing only in a small percentage of each round of replication (Friedman et al., 1997; Raghuraman et al., 2001; Yabuki et al., 2002).

In summary, the Mcm signal is highly significant and occurs precisely at each potential origin of replication. During S phase, MutS and Polε signal are coincident in majority of the active ARSs and flanking regions with high significance. These data further support the hypothesis that MutS remains closely associated with the replisome throughout S-phase of the cell cycle.
Figure 3.9: MutS and Polε occupy the same regions during DNA replication

Figure 3.9 MutS and Polε exhibit co-incident signal during S phase. Three independently performed experiments were used to calculate p-values. The samples were analyzed at the time of arrest (G1) and six additional time points (tp1- tp6) during S phase for Polε and MutS. The data are visualized as negative of the log_{10} of the calculated p-values using the Integrated Genome Browser for Mcm4 (Mcm, purple), no tag (black), Msh2 (MutS, blue) and Pol2 (Polε, green). Because the Mcm signal is so significant, the histogram is scaled to 27 (or reflecting a p-value \sim 10^{-27} for the most signal values). The MutS and the “no tag” control graphs are set to 10 and the Polε graphs to 20. Black bars below the data denote position of ARSs in the genome databases. The red bars represent ARSs not found in the genome database. CC-represents the chromosomal coordinates x10^3. Representative regions are shown including: (A) early-efficient origins (ARS1207 and ARS1209) flanking an inactive origin (ARS1208), (B) ARS1213, and (C) ARS1407
MutS shows an altered pattern of binding during replication in MMR defective PCNA/Pol30 variants

Having determined that MutS is closely associated with the replisome throughout S phase, we wanted to determine what factors might influence MutS loading at origins and scanning of the genome during replication. PCNA clearly has a critical role in MMR at multiple stages (Johnson et al., 1996; Umar et al., 1996; Flores-Rozas et al., 2000; Lau et al., 2002). To determine whether PCNA mutants implicated in MMR alter the binding and movement of MutS during S phase, we examined MutS and Polɛ dynamics in PCNA/Pol30 MMR defective strains. Two missense variants of PCNA/Pol30 (Pol30C22Y and Pol30C81R) were previously reported to disrupt MMR, but not to alter replication significantly (Lau et al., 2002). We reasoned that these “separation of function” variants would be ideal candidates for determining the role of PCNA in MutS dynamics during replication.

We first utilized the pol30-201 mutant coding for Pol30C22Y in ChIP-chip experiments as described previously. Pol30C22Y confers a partial MMR defect; however, the Pol30C22Y variant still interacts with MutS in vitro (Lau et al., 2002). It is important to note that, though not all time points were examined in this experiment, the Polɛ signal appears normally distributed (relative to strains expressing wild-type PCNA/Pol30), suggesting that there is no effect on the processivity of the polymerase (Figure 3.10 A and B). Cell cycle progression is also unaffected in this mutant (not shown), further supporting the absence of a
replication defect (Lau et al., 2002). Interestingly, although the Pol30<sup>C22Y</sup> variant is defective for MMR, the data show that MutS signal still coincides with Polε signal in the presence of Pol30<sup>C22Y</sup> (Figure 3.9 A and B). This is in agreement with <i>in vitro</i> studies that show interaction is not fully disrupted between MutS complexes and Pol30<sup>C22Y</sup> protein (Lau et al., 2002). However, the MutS signal does seem reduced in some regions adjacent to the ARSs (right side of ARS315 and ARS514 Figure 3.10), suggesting a potential defect in MutS association.
Figure 3.10: MutS and Polɛ co-localize in a MMR defective PCNA variant

Figure 3.10: MutS and Polɛ co-localize to origins during S phase in a strain expressing a PCNA/Pol30 MMR defective variant, Pol30<sup>C22Y</sup>. The samples were analyzed at the time of arrest (G1) and two additional time points in S phase, 10 minutes apart (20 min and 30 min) for Polɛ and MutS. The log<sub>2</sub> (ChIP/Input) were visualized as using the Integrated Genome Browser for Mcm4 (Mcm, purple), no tag (black), Msh2 (MutS, blue) and Pol2 (Polɛ, green). The graphs were set to 2.5 for all data (~6 fold maximum increase). Black bars below the data denote position of ARSs in the genome databases. CC-represents the chromosomal coordinates x10<sup>5</sup>. Representative regions are shown including: (A) ARS315 (B) ARS514
Previous work showed that the partial MMR defects caused by the Pol30\(^{C22Y}\) and Pol30\(^{C81R}\) missense variants are exacerbated by converting two conserved phenalynine residues to alanine in the PCNA interacting region (PIP box) of Msh6 (Lau et al., 2002). Additionally, strains expressing Pol30\(^{C81R}\) have a partial MMR defect that is more severe than is seen in strains expressing Pol30\(^{C22Y}\) (Lau et al., 2002). Additionally, in strains expressing Pol30\(^{C81R}\) and the Msh6 PIP box variant (Msh6\(^{PIP}\)), MutS\(\alpha\) associated replication foci are no longer visible (Hombauer et al., 2011a). With this in mind, we engineered strains containing \textit{pol30-204} (coding for Pol30\(^{C81R}\)) and the \textit{msh6-F33AF34A} PIP box mutation (expressing Msh6\(^{PIP}\)) to analyze MutS dynamics during replication. We hypothesized that by disrupting this interaction; signal corresponding to MutS binding should be abnormal, significantly reduced, or lost entirely. Consistent with the finding that Pol30\(^{C81R}\) does not affect replication, the engineered strain exhibited normal cell cycle progression (Figure 3.11).
Figure 3.11: Cell synchrony of the PCNA, MutS double mutant strain

The ChIP-chip data shows some MutS signal in the vicinity of the replisome in the strain expressing Pol30C81R and Msh6Pip (Figure 3.12); however the signal is not highly correlated with the Polc signal. Previous in vitro co-immunoprecipitation and in vivo microscopy studies show that the double mutant
exhibits loss of interaction and loss of co-localization respectively (Lau et al., 2002; Hombauer et al., 2011a). Our findings suggest that the MutS signal observed represents PCNA independent associations with the DNA during replication.
Figure 3.12: MutS dynamics are altered in the double mutant strain
Figure 3.12: MutS dynamics are altered in the double mutant strain. Double mutant cells (msh6-F33AF34A pol30-204) expressing the Msh6$^{\text{PIP}}$ variant and Pol30$^{\text{C81R}}$ were analyzed at the time of arrest (G1) and additional time points in S phase, 10 minutes during S phase for Polε and MutS. The log$_2$ (ChIP/Input) were visualized as using the Integrated Genome Browser for Mcm (purple), no tag (black), MutS (blue) and Polε (green). The graphs were set to 2.5 for all data (~6 fold maximum increase). Black bars below the data denote position of ARSs in the genome databases. CC-represents the chromosomal coordinates x10$^3$. Representative regions are shown including: (A) ARS315 (B) ARS919 and ARS920.
Discussion

In this study we show that MutS complexes track with the replisome throughout DNA replication. Specifically, we examined the dynamics of the mismatch recognition complexes relative to the leading strand DNA polymerase. The movement of the leading strand polymerase served as the control for the advancing replisome and Mcm helicase functioned as the metric for ARS demarcation. The MutS complex signal was distributed in the region occupied by the leading strand DNA polymerase and appeared to persist after fork passage. These findings support the model that the mismatch repair recognition complex remains in close proximity to the errors as they emerge from the replisome as well as to the replication-specific nicked DNA that serve as strand specificity signals.

Polε binds active origins after release from G1

Prior to this work it was known that Mcm loads at potential origins during G1 (Aparicio et al., 1997). However, until now the dynamics of the DNA polymerase during DNA replication was not examined with this detail using ChIP-chip. Using high resolution methods with synchronized cells we were able to show that Polε loads in S phase in a manner consistent with established origin firing times. Using ChIP-PCR, Polε has been shown to assemble at origins during G1 of the cell cycle (Aparicio et al., 1997). The data presented in this work does not show significant Polε signal above background at the origins at the time of α-factor
arrest (G1). This absence of Polɛ binding may reflect a weak association with the DNA during early S-phase as Polα is initially loaded onto the DNA for the priming step. Alternatively, the higher resolution methods used in this work may allow for more precision of timing to be delineated. Accordingly, the timing, precision and efficiency of firing at active origins is highly tractable using this method and is likely to be a more reliable way to detect the kinetics of Polɛ than ChIP-PCR.

**Potential Role for PCNA in MutS Loading during Replication**

Importantly, in this work we showed that MutS loads at origins of replication with a timing similar to the DNA polymerase during S phase. How MutS loads onto DNA is not known. One hypothesis we tested was that PCNA was responsible for the loading and potentially for aiding in the scanning efficiency of MutS. Interestingly, we found that MutS signal is still observed in the presence of the PCNA variants that partially disrupt MMR function; however, the signal is aberrant in the mutant strains.

It is of interest that we are able to detect MutS signal in a strain in which the interaction between PCNA and MutSα should be abolished. Two explanations could account for the signal. First, it is possible that the signal observed may be from MutSβ (Msh2/Msh3), which is known to be partially redundant with MutSα (Msh2/Msh6) (discussed in detail in Chapter 4). This explanation contradicts the studies showing that MutSβ does not co-localize with the leading strand.
polymerase (Hombauer et al., 2011a); however it is of importance to note that studies using human cell lines have also identified MutSβ at sites of active replication (Sirbu et al., 2013). Additionally, using mass spectrometry, MutSβ was shown to interact with the replisome in Schizosaccharomyces pombe (McDonald, Cristea, and Zakian, personal communication). The second explanation for the MutS signal in the strain where the PCNA/MutSα interaction is disrupted is that there is a PCNA independent association of MutS complexes with the replisome that may account for the observed signal. We favor a model where PCNA plays an important role in MutS associations with replicating DNA, but it is not the sole determinant controlling MutS loading at origins.

Potential Role for Histone Modification in MutS Loading at Origins during Replication

If PCNA is not the sole determinant recruiting MutS to the replisome, other origin specific components must be involved. Recently, human Msh6 of the MutSα complex has been shown to interact with modified nucleosomes during G1 and S-phase of the cell cycle, establishing a requirement for Histone 3 trimethylated Lys36 (H3K36me3) and the methyl transferase SETD2 for recruitment of human MutSα during DNA replication (Li et al., 2013). Yeast Msh6 does not contain a Pro-Trp-Trp-Pro (PWWP) domain found in hMsh6, thought to be required for H3K36me3 interaction; however, it is plausible that there exist a similar mechanism in yeast or other eukaryotes exploiting a different modification. In
yeast, histones at active replication origins are acetylated by Gcn5 (Vogelauer et al., 2002; Unnikrishnan et al., 2010). It is tempting to speculate that nucleosomes containing acetylated histones marking the origins may be important for MutS loading. Our laboratory has preliminary data showing that Msh2 levels are reduced in a strain lacking the Gcn5, the protein responsible for acetylating histones at active origins (Kim and Gammie, unpublished). We speculate that if MutS recruitment to origins is less efficient if the histones are not acetylated, then the unbound MutS complexes may be targeted for degradation.

Mismatch Scanning in the Context of Chromatin

Once MutS is loaded at origins, the complex must have the ability to efficiently scan for mismatches. Several lines of evidence show that in vitro, chromatin and histone modifications may influence the accessibility of DNA by MutS complexes. Firstly, in vitro experiments show that unmodified nucleosomes block mismatch recognition (Li et al., 2009). It is also known that MutS translocation on DNA can be impeded by proteins bound to the same DNA molecule (Mendillo et al., 2005; Pluciennik and Modrich, 2007). Additionally, single molecule studies show that movement of one MutS complex along the DNA occurs via a sliding mechanism and can be inhibited when a nucleosome is encountered (Gorman et al., 2007; Gorman and Greene, 2008). There is evidence suggesting that modifications of histones might influence MutS accessibility to DNA. Histone 3 acetylated at lysine 56 (H3K56ac), is known to be associated with repair (Chen et al., 2008). In vitro, H3K56ac has been shown to facilitate nucleosome disassembly by hMutSα.
(Javaid et al., 2009) (Forties et al., 2011). In this case, if nucleosomes with specific modifications are already assembled on the DNA, MutSα can then displace such nucleosomes in order to gain access to the mismatched DNA.

Finally, MutSα is also capable of interfering with nucleosome assembly (Schoepf et al., 2012). This is relevant to replication because it is possible that MutS blocks assembly behind the replisome until the entire genome has been efficiently scanned. However it has yet to be determined in yeast if MutS complexes may be capable of displacing modified histones in vivo.

A Model for Efficient Mismatch Scanning of Newly Replicated DNA

Using chromatin immunoprecipitation and DNA tiling arrays, we are able to visualize the scanning dynamics of MutS binding during S phase; however, two models for movement along the DNA are consistent with the data: (1) MutS loads at origins and scans immediately behind the advancing replisome facilitated by direct interactions with replisome components, or (2) MutS loads at origins, but scans independently of the replisome. Because the two models involve loading of MutS at active origins where the chromatin have been cleared, they both address the protein blockage problems discussed earlier.

The first model is dependent upon a physical connection between MutS and the replisome. Live-cell imaging during S-phase of S. cerevisiae cells show that Msh6 co-localizes with Pol2 (Hombauer et al., 2011a). Additionally, MutS has
been shown to interact with PCNA (Johnson et al., 1996) and PCNA is associated with the replisome (Johnson and O'Donnell, 2005). Finally, in *Bacillus subtilis* MutS and MutL have been shown to interact with the catalytic subunit of the DNA polymerase III (DnaE) *in vitro* (Klocko et al., 2011). *In vivo* experiments in *B. subtilis* using GFP-tagged DnaE showed that mismatch detection causes the polymerase to disengage the DNA during replication (Klocko et al., 2011). These experiments support the model that MutS is directly associated with the replisome. Thus, we favor the first model based on the previous studies and our observations that the distribution of MutS signal is very similar to the distribution of the leading strand DNA polymerase as the replisome advances during S phase. The first model is also appealing because tracking directly behind the replisome ensures that the MutS complexes are always in close proximity to a strand specificity signal, the 3’-OH of the newly synthesized strand.

A further refinement of the model includes the following: MutS loads at origins and scans immediately behind the advancing replisome as well as in the regions behind the replication fork. The addition to the model is based on the fact that the MutS signal persists in the newly replicated region even when the leading strand polymerase appears to have cleared the region. The persistence of signal could be explained by the interaction of MutS and PCNA. In eukaryotes, PCNA is known to accumulate behind the replisome (Shibahara and Stillman, 1999) and in *B. subtilis*, DnaN (PCNA) clamp zones have been shown to remain behind the replication zone (Su'etsugu and Errington, 2011). This DnaN-mediated
recruitment of MutS is responsible for 90% of repair in *B. subtilis* with the remaining MMR being DnaN-independent (Lenhart et al., 2013). Taken together, we favor a model in which the persistent MutS signal after fork passage is explained, in part, by interactions with PCNA molecules that remain behind the replisome.

Figure 3.12, illustrates a model for MutS signal distribution during S phase. In the model, MutS complexes bind to activated origins during S phase with a timing similar to DNA polymerase. The MutS loading may be facilitated by direct PCNA interactions or modified histones may function to recruit MutS to active origins. Once MutS is loaded, a close association with the advancing replisome ensures that mismatches are rapidly detected and that the MMR machinery always has a proximal replication-specific nick to direct repair to the newly synthesizes strand. In the model, upon detection of a mismatch, the most proximal signal is the 3’-OH of the newly synthesized strand. The MutS signal persisting behind the advancing replisome may be a consequence of PCNA interactions. PCNA is bound to nicks behind the replisome created during lagging strand synthesis and caused by rNMP excision. The nick/PCNA/MutS interactions may account for the persistent signal behind the advancing replication fork. Finally, in this model, chromatin formation in inhibited by MutS until all of the mismatches and nicks have been repaired.
Figure 3.13: MutS Mismatch Recognition during DNA Replication

Figure 3.13: Model for MutS mismatch recognition during replication. The model described in the text is depicted above with schematics of MutS complexes (green and red) with a flexible tether (purple); DNA polymerase (multisubunit complexes shown in light blue and grey); PCNA (dark blue circles), unmodified histones (green circles); modified histones (green circles with blue tag); single stranded binding proteins (orange circles); DNA polymerase alpha (multisubunit complex shown in purple circles); and the Mcm helicase (red). The template DNA is shown with black lines and the newly synthesized DNA with green lines. The direction of polymerization of the DNA is shown with arrowheads.
Chapter 4: "MutSβ and MutSα Dynamics During Replication"

Abstract

Using next generation sequencing we determined the spectrum of mutations generated in the absence of individual components of the mismatch recognition complexes relative to the ancestral strain. Our data show that MutS complexes are redundant for the repair of single base insertion deletion loops at homopolymeric runs, but that each MutS heterodimer is uniquely required for ~6% of mismatches generated during replication. We examined the relative abundance of the components of each MutS complex and found that Msh2 levels are in 2-fold excess of Msh3 and Msh6, while Msh3 and Msh6 have equivalent levels in the cell. These findings are consistent with equal amounts of MutSα and MutSβ in the cell.
Introduction

MutS complexes function in post-replicative repair by recognizing errors that escape the proofreading of the DNA polymerases. The errors include single base mismatches, single nucleotide indel loops at homopolymeric runs (HPRs) and larger nucleotide indel loops at microsatellites (MS). Microsatellites are repeat regions of 1-10 bp repeat units, which frequently undergo expansion and contraction due to slippage of the polymerases during replication (Bhargava and Fuentes, 2010). In prokaryotes, homodimeric MutS binds the full range of mismatches (reviewed in Kunkel and Erie, 2005). In eukaryotes, MutS complexes are heterodimers with differing mismatch recognition capabilities. MutSα (Msh2/Msh6) recognizes single base mismatches and single nucleotide insertion/deletion (indel) loops, and MutSβ (Msh2/Msh3) complex recognizes single nucleotide and larger indel loops (reviewed in Kunkel and Erie, 2005).

In *S. cerevisiae* MutSα is thought to be the major MutS complex in the cell because MutSα, but not MutSβ, co-localize with the leading strand polymerase during S-phase of the cell cycle (Hombauer et al., 2011a). We provide evidence to support an alternative hypothesis in which both complexes MutSα and MutSβ are in the vicinity of the replisome to capture all of the mismatches that escape polymerase proofreading. In the following sections we show that MutSα and MutSβ are required to efficiently repair the entire spectrum of mutations during replication and that MutSα and MutSβ are found in a stoichiometry suggesting equivalent levels in the cell.
Results & Discussion

MutSβ and MutSα are required for the full spectrum of mismatches generated during replication

To confirm on a genome-wide level that both MutS complexes are required for the full spectrum of mutations generated during replication, we performed mutation accumulation experiments followed by whole genome sequencing in strains lacking one of the components of the two mismatch recognition complexes (Table 8). We have included previously published msh2Δ and WT data (Lang et al., 2013) for comparison. We determined that mutation rate for DNA mismatch repair null strains was ~1 mutation per genome per generation, 225-fold higher than the wild-type rate. The mutation spectra for mismatch repair defective cells included insertions/deletions at homopolymeric runs (~87%) and at larger microsatellites (~6%), as well as transitions (~5%) and transversions (~2%) (Lang et al., 2013).

Mutation accumulation analysis of the msh3Δ strain revealed an increase in indels at larger microsatellites comparable to msh2Δ. Interestingly, no single base substitutions and only a few single base indels were observed (Table 8). These data are consistent with having a fully functional MutSα (Msh2/Msh6). MutSα is capable of repairing single base substitutions and single nucleotide indel loops in the absence of Msh3p.
The msh6Δ strain, acquired 14 single base substitutions. This observed number is also comparable to the single base substitutions observed in the msh2Δ strain. Of the 14 mutations observed 12 were transitions while 2 were tranversions, consistent with the ratio of transitions to transversions observed previously (Lang et al., 2013). The msh6Δ strain accumulated 3 indels at homopolymeric runs. Larger indels are not observed at MS in this strain, consistent with the repair of larger indels being MutSβ specific.

After ~210 generations, the msh2Δ strain accumulated a large number of indels at HPRs (177) relative to the single deletion of the binding partners. We observe only 5 and 3 indels at HPRs in msh3Δ and msh6Δ respectively (Table 8). This underscores the functional redundancy of MutSα and MutSβ for repair at HPRs and is in agreement with previous genetic analyses showing that MutSα and MutSβ are redundant for the repair of a homopolymeric run (Marsischky et al., 1996). Additionally, analyses in human cell lines of MutS complexes also demonstrate redundancy in repair of indels (Genschel et al., 1998).

Table 8. Mutation accumulation in 210 generations

<table>
<thead>
<tr>
<th>Relevant genotype</th>
<th>Single nucleotide polymorphisms</th>
<th>HPR insertion or deletion</th>
<th>MS insertion or deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>msh2Δ</td>
<td>15</td>
<td>177</td>
<td>8</td>
</tr>
<tr>
<td>msh3Δ</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>msh6Δ</td>
<td>14</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

In summary, using mutation accumulation assays we confirmed that Msh6 and Msh3 are fully redundant for repair of single-base indels at homopolymers across
the entire genome; however, each MutS complex is needed to repair the entire spectrum of mismatches generated during replication.

**The levels of the individual MutS subunits are such that MutSα and MutSβ should be found at equivalent levels**

Although both MutS complexes are needed for the full spectrum of mismatches generated during replication, it is possible that MutSα was found to be the major replisome associated mismatch recognition complex because MutSα is more abundant. Previous studies examined the levels of the MutS complexes in human and yeast cells. In human cells, the complexes were determined to occur at a MutSα:MutSβ ratio of 6:1 or 10:1, depending on the analysis (Drummond et al., 1997; Genschel et al., 1998). In yeast, high throughput abundance studies reported ~1,000 copies of Msh2, ~5,000 of Msh6 and ~700 of Msh3 per cell (Ghaemmaghami et al., 2003). Taken together, the data suggest that MutSα accounts for a greater percentage of the MutS complexes present in the cell. However, the high throughput experiments in yeast required validation.

We examined the relative abundance of the individual components of both mismatch recognition complexes, using western blot analysis. We engineered a strain in which all three proteins were tagged with an identical epitope (myc). The fusions were engineered at the endogenous chromosomal positions using the native promoters. The tagged proteins were shown to be functional for mismatch repair in vivo (not shown).
Chapter 4: Results and Discussion

The data reproducibly show that the relative abundance of Msh2: Msh6: Msh3 is 2:1:1 (Figure 4.1). This ratio is visualized by examining the protein extract from the strain in which all three proteins are identically tagged (lane 2, Figure 4.1) and therefore a direct comparison of the abundance is possible. In summary, we find that Msh2 is in a 2-fold excess of the Msh6 and Msh3 binding partners such that there should be equal levels of MutSα and MutSβ in the cell. The discrepancy with our results and the high throughput method may be due to differences in the method of visualizing protein levels or the epitope tag used.

**Figure 4.1: Msh2 is in a 2-fold excess of Msh3 and Msh6**

![Image of protein extract](image.png)

**Figure 4.1: Msh2 protein levels are in excess of Msh3 and Msh6.** Cultures were grown to ~0.4 O.D$_{600}$ and proteins were extracted and detected by immunoblotting (see Materials and Methods). The proteins were detected using antibodies for the myc epitope. Lane 1: contains Msh2-myc tagged only. Lane 2: all three components of the MutS complexes are myc-tagged. Lane 3: both Msh3 and Msh6 are myc-tagged. The loading control was detected using α-Kar2 antibody. The bands were quantified using image J software.
Chapter 4: Results and Discussion

The levels of Msh3 or Msh6 do not increase to compensate for the absence of the redundant partner.

The results regarding stoichiometry suggest that the levels of the MutS subunits are in balance such that there should be equal MutSα and MutSβ in the cell (Figure 4.1). Our laboratory previously determined that monomeric Msh2 and Msh6 are targeted for degradation (Arlow et al., 2013). It is possible that in the absence of one binding partner, the levels of the other partner may be stabilized upon binding to Msh2. To determine if Msh3 or Msh6 levels are altered in the absence of the redundant partner, we performed western blot analysis in strains where one component is myc-tagged, while the other is absent in the cell. The results are illustrated in (Figure 4.2). The absence of Msh3 does not alter the levels of Msh6. Similarly, the absence of Msh6 in the cell does not change Msh3 levels.
Figure 4.2: Msh3 and Msh6 are not reciprocally regulated.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msh6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Msh3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Msh2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kar2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.2: Msh3 and Msh6 are not reciprocally regulated. Cultures were grown to mid-exponential phase and proteins were processed for immunoblotting. Lanes 1-3 show the levels of Msh6, Msh3, or Msh2 expressed from stains were the proteins are individually tagged with the myc epitope. Lane 4 is an extract from a strain expressing both Msh3-myc and Msh6-myc. Lane 5: Msh3 levels in the absence of Msh6 (msh6Δ). Lane 6: Msh6 levels in the absence of Msh3 (msh3Δ). Kar2 is the loading control.

**Dynamics of the MutSα (Msh2/Msh6) during DNA replication**

MutSα and MutSβ are needed to cover the full spectrum of mutations and the levels of the protein subunits suggest that the stoichiometry of the MutS complexes are in balance; however it is still possible that MutSα is the major complex found at the replisome and that MutSβ only binds when larger indels form at the advancing fork. We therefore aimed to determine whether both complexes track with the replisome during replication.

The MutS tracking expreriments presented above, used Msh2-myc and therefore do not allow us to determine if one or both mismatch recognition complexes are
co-incident with the polymerase throughout S-phase. To determine if MutSα (Msh6/Msh2) and MutSβ (Msh3/Msh2) are both in the vicinity of the replisome throughout S-phase, we performed ChIP-chip time course experiments using strains with Pol2-HA tagged (Polε) and either Msh6-myc tagged (MutSα) or Msh3-myc tagged (MutSβ). As described in the previous Chapter, the control for binding at origins of replication was Mcm4 (HA tagged), a subunit of the Mcm2-7 (Mcm) replicative helicase that binds replication origins during G1 of the cell cycle (Aparicio et al., 1997).

Having previously established the distribution of the polymerase throughout S phase (Chapter 3), a simplified time course time course experiment was performed to examine MutSα or MutSβ binding during S-phase. Briefly, samples were taken at the time of arrest and two additional times during S phase of the cell cycle and processed for ChIP-chip.

Figure 4.3 shows an example of the data for MutSα during replication. As for Msh2, we observed binding of Msh6 in regions corresponding to Polε binding. This is consistent with previous studies that show co-localization of Msh6 and the leading strand polymerase (Hombauer et al., 2011a); however, it does not exclude the possibility that MutSβ is also at the replisome.
Figure 4.3: MutSα distribution during DNA replication. As described above, cells were arrested in G1 of the cell cycle using α-factor at 18°C to slow the process of DNA replication. The cells were then released from arrest, samples were collected at 10 minute intervals and crosslinked with para-formaldehyde. The samples were then immunoprecipitated with antibodies conjugated to agarose beads that were specific for the tagged protein. The crosslinks were reversed and immunoprecipitated DNA was labeled and hybridized to custom DNA tiling arrays. A sample from each time point was analyzed by flow cytometry, to monitor progression through the cell cycle (see methods). The y-axis is set at 3 (or ~8 fold maximum) for each row. CC represents the chromosome coordinates x 10^3. Msh6 (brown), Polɛ (green), no tag (black) and Mcm (purple).

We performed a time course experiment with MutSβ tracking, however because of time constraints the data are not yet available.
Chapter 4: Results and Discussion

Summary

In this chapter, we provide data consistent with a hypothesis positing that both MutSα and MutSβ track with the replisome to capture the entire spectrum of mismatches that escape DNA polymerase proofreading. We determined that on a genome-wide level that MutSα and MutSβ are required to efficiently repair single base pair substitutions (MutSα), single base indels at homopolymers (MutSα and MutSβ) and larger indels at microsatellites (MutSβ). Additionally, we showed that MutSα and MutSβ are found in a stoichiometry suggesting equivalent levels of MutSα and MutSβ in the cell. Finally we determined that MutSα is detected at the replisome.

The hypothesis that both MutS complexes are at the replisome contradicts earlier findings that failed to show co-localization of MutSβ with the leading strand polymerase (Hombauer et al., 2011a). However, the hypothesis is supported by other findings. For example, experiments in human cell lines that show hMSH2, hMSH3 and hMSH6 are found in regions of active replication (Sirbu et al., 2013). Additionally, using mass spectrometry, MutSβ was shown to interact with the replisome in Schizosaccharomyces pombe (McDonald, Cristea, and Zakian, personal communication). Taken together, we propose a model in which both MutS complexes are found at the replisome.
References


Chapter 5: References


