UNIVERSITY OF CALIFORNIA, SAN DIEGO

Mechanisms of Mispair Recognition by Msh2-Msh3

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Jill Mae Dowen

Committee in charge:

Professor Richard D. Kolodner, Chair
Professor John Carethers
Professor Karen Oegema
Professor Amy Pasquinelli
Professor Bing Ren

2009

The dissertation of Jill Mae Dowen is approved, and it is

acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2009

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would first like to pay tribute to my thesis advisor Richard Kolodner. He has always provided guidance and mentoring when needed but given me the freedom to ask my own questions and also learn from my own mistakes. Richard's lab has provided the resources and scientific interactions necessary for me to grow into an independent researcher. He is also an entertaining guy whose perspective on many topics, including the way science is done, has definitely influenced my way of thinking for the better.

I would like to thank the lab as a whole for being a great place to do science. Many former members were vital to my early training in yeast genetics and biochemical techniques. In particular, I am thankful to Dr. Dan Mazur, Dr. Marc Mendillo, Dr. Scarlet Shell, Dr. Chris Putnam and Jason Chan for the time and expertise they leant to thinking about my projects.

I would also like to thank my family whose love and support has been unconditional. Without fully understanding why I moved so far away or exactly what I do on a daily basis, they are nonetheless infinitely proud of me.

Finally, I thank my husband Robbie Dowen. It is difficult to describe the impact he has had on my time in graduate school. Not only did he provide a wonderful friendship in the beginning that has grown into a marriage, but he has always been available to discuss science of any kind. Our conversations on many topics, from ideas about my project to new findings from another field to technical problems with an experiment, have been invaluable. I respect his

ability to simultaneously be my biggest critic and my biggest supporter. My time in graduate school would not have been as fun or fulfilling had I not met him here in San Diego.

The text of Chapter 2 is a full reprint of the material as it appears in *Mol Cell Biol*, 2007(18):6546-64, Harrington, JM, Kolodner RD. The dissertation author was the primary researcher and author of this paper.

The text of Chapter 3, in full, is in preparation for submission to *Nat. Struct. Mol. Biol*, 2009, Dowen, JM, Putnam CD, Kolodner, RD. The dissertation author was the primary researcher and author of this paper.

Vita

2002 BS, Microbiology, University of Iowa, Iowa City

2009 PhD, Biomedical Sciences, University of California, San Diego

Publications

Dowen, JM, Putnam, CD, Kolodner, RD. *Msh2-Msh3 Mispair Recognition Involves DNA Bending and Strand Separation.* In preparation. 2009.

Harrington, JM, Kolodner, RD. *Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs*. Molecular and Cellular Biology, 2007 (18);6546-54.

Vieira AR, Avila JR, Daack-Hirsch S, Dragan E, Félix TM, Rahimov F, Harrington J, Schultz RR, Watanabe Y, Johnson M, Fang J, O'Brien SE, Orioli IM, Castilla EE, Fitzpatrick DR, Jiang R, Marazita ML, Murray JC. *Medical sequencing of candidate genes for nonsyndromic cleft lip and palate*. PLoS Genetics, 2005 (6):e64.

ABSTRACT OF THE DISERTATION

Mechanisms of Mispair Recognition by Msh2-Msh3

by

Jill Mae Dowen

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2009

Professor Richard D. Kolodner, Chair

DNA mismatch repair is the process of fixing errors that arise in the genome during DNA replication and recombination. This is accomplished by the recognition of DNA lesions and the subsequent recruitment of factors to remove the damaged DNA and then incorporate the correct DNA sequence. The faithful maintenance of the genomic sequence is important since the

accumulation of somatic mutations can disrupt basic processes within a cell and the accumulation of germline mutations can lead to inherited mutations or even inviable daughter cells.

Errors in basic repair processes have severe consequences for stability of the genome and the development of cancer. Disruption of mismatch repair genes themselves leads to an overall defect in DNA mismatch repair causing the genome to accumulate mutations at a high rate. This hypermutation phenotype will eventually cause mutations at gene loci important for regulating the cell cycle and therefore lead to tumor development. A variety of studies presented here, demonstrate the importance of a functional mismatch repair system in the cell and provide a detailed analysis of mismatch recognition by the Msh2-Msh3 complex.

Chapter 1: Overview of Mismatch Repair

1.1 Mismatch repair and cancer

1.1.1 Mismatch repair maintains genome fidelity

DNA mispairs are rare, spontaneously occurring events that a cell must repair. Mispairs can arise during DNA replication due to polymerase errors, aberrant recombinational events, and the incorporation of damaged DNA or DNA precursors (Kolodner and Marsischky, 1999). If mispairs are not recognized and repaired, and a cycle of DNA replication ensues, they become fixed in the genome as a mutation. The mismatch repair (MMR) proteins are a surveillance system dedicated to recognizing and repairing these lesions in DNA, thus preserving the fidelity of the genome (Kolodner, 1996; Kolodner and Marsischky, 1999).

If the mismatch repair system becomes inactivated, the mutation rate of a cell increases 50-1,000-fold (Iyer et al., 2006). The events that accumulate in a mismatch repair-defective cell include base substitutions, frameshifts, and insertions and deletions of larger sequences, often mediated by homology or secondary structures. Faithfully maintaining the sequence of the genome is such an important task for a cell that the mismatch repair pathway is preserved through evolution from bacteria to humans.

1.1.2 Links between mismatch repair defects and cancer

In humans, defects in the mismatch repair pathway primarily lead to Lynch Syndrome (also called hereditary non-polyposis colorectal cancer (HNPCC)). Patients that inherit mutations in the *MSH2* or *MLH1* genes are predisposed to primarily develop early onset gastrointestinal and endometrial cancers as well as other cancers at lower frequencies (Peltomaki, 2003). The mismatch repair-defective tumor cells exhibit a characteristic instability of microsatellite sequences in the genome. This is due to the fact that these highly repetitive sequence elements are prone to polymerase slippage events and mismatch repair is required to preserve the original tract length.

1.2 The prokaryotic MutS homodimer

The mechanism of DNA mismatch repair was first described in prokaryotes. After 2 decades of research, this pathway is well understood and the reactions can be reconstituted *in vitro* with purified components (Kunkel and Erie, 2005; Lahue et al., 1989). MMR in *Escherichia coli* initiates when a homodimer of MutS specifically recognizes and binds to mispaired DNA. The mispairs recognized are base-base mismatches and insertion/deletions of 1 to several nucleotides caused by polymerase errors (Modrich, 1991).

Recently, much has been learned about how the various domains of the MutS protein contribute to the overall mechanism of mismatch recognition. Importantly, while MutS is a homodimer, it binds to mispaired DNA

asymmetrically. The mispair binding domain (MBD) from one subunit makes direct contact with the mispair while both subunits provide loose contacts with the DNA backbone (Lamers et al., 2000). Upon binding by MutS, the DNA becomes kinked by 60 degrees at the mispair site, thus opening up the minor groove for access to the mispaired nucleotide (Lamers et al., 2000). Other domains of MutS have been shown to play important roles in dimerization, binding and hydrolysis of ATP, transmitting conformation changes, and providing interfaces for interaction with other proteins like MutL and the processivity factor PCNA.

After recognition of the mismatch by MutS, there is a subsequent recruitment of a homodimer of MutL. This mismatch-MutS-MutL complex signals for the recruitment of the endonuclease MutH (Acharya et al., 2003; Galio et al., 1999; Grilley et al., 1989; Hall and Matson, 1999; Selmane et al., 2003). This endonuclease places a nick in the newly replicated daughter strand, which is transiently unmethylated at a GATC site after DNA replication (Au et al., 1992; Welsh et al., 1987). From this nick, there are four exonucleases capable of excising the damage-containing DNA strand back through the mispair site (Burdett et al., 2001; Matson and Robertson, 2006). The gap left behind can be filled in by DNA polymerase III and the action of DNA ligase to seal the strand (Lahue et al., 1989; Modrich, 2006).

1.3 The eukaryotic Msh2-Msh6 and Msh2-Msh3 heterodimers

The mismatch repair pathway in eukaryotes is more complex than prokaryotes (Iyer et al., 2006). Instead of a single homodimer of MutS there are two MutS Homologoue complexes that perform mismatch recognition: Msh2-Msh6 and Msh2-Msh3. Msh2 is the equivalent of the non-mispair contacting subunit while Msh3 and Msh6 make direct contact with the mispaired DNA, and therefore determine mispair specificity.

1.3.1 The roles of the eukaryotic heterodimers

The two eukaryotic heterodimers recognize different DNA mispairs. Msh2-Msh6 (MutSα) recognizes single base-base mismatches and small insertion/deletion mispairs of 1 or 2 nucleotides, although it can recognize larger insertion/deletions to a lesser extent. Msh2-Msh3 (MutSβ) recognizes loops of DNA from 1 to 14 nucleotides in size, and various flaps and hairpins of DNA thought to form during genetic recombination (Marsischky et al., 1996). Therefore, the two heterodimers are partially redundant in their ability to recognize 1 and 2 nucleotide insertions. Due to this partial redundancy the loss of a functional Msh6 or Msh3 protein leads to a partial defect in MMR and a moderate increase in mutation rate (Marsischky et al., 1996). Loss of Msh2 causes a complete defect in MMR and high rate of mutation, and therefore mutations in *MSH2* are commonly found in HNPCC.

Much of the work on eukaryotic MMR has focused on the Msh2-Msh6 complex and has shown striking similarities to the prokaryotic MutS mechanism of mismatch recognition. While Msh2-Msh6 only acts in traditional MMR processes, Msh2-Msh3 has additionally been shown to play a role in repairing intermediate structures that form during recombinational events. For example, during the repair of a double strand break, exonuclease processing leaves 3' ssDNA tails that are recognized and bound by Msh2-Msh3. Together with Rad1-Rad10, Msh2-Msh3 is essential for removing regions of nonhomology in the tails, and thus, preventing aberrant recombination (Selva et al., 1997; Surtees and Alani, 2006).

1.3.2 Biochemical characterization of Msh2-Msh6 and Msh2-Msh3

It is known that there are several important properties of the mismatch recognition proteins that are essential for mismatch repair to occur *in vivo.* Previous studies have measured the affinity of the two eukaryotic heterodimers for various mispaired DNA substrates. From other work it is known that these proteins are able to hydrolyze ATP causing a conformational change that is essential for signaling for downstream events in the repair process. Additionally, the relative protein levels have been estimated in HeLa cell extracts where there is a 10-fold excess of the Msh2-Msh6 heterodimer to Msh2-Msh3 (Genschel et al., 1998).

1.3.2.1 Substrate specificities

In vitro binding studies have been performed with the purified human and yeast Msh2-Msh6 heterodimer on defined mispaired substrates. Msh2-Msh6 shows high affinity for most base:base mispaired substrates but can also bind to substrates with +1 to +10 insertions (Kolodner and Marsischky, 1999). The specificity of mispair recognition seems to be influenced by the affinity for a specific mismatch and the flexibility of the local DNA sequence. It is important to note that during *in vitro* studies these proteins do exhibit a low level of binding to completely basepaired DNA and the ends of DNA substrates.

While very few studies have examined the mispair specificity of the Msh2-Msh3 heterodimer, it has been shown to bind to loops of DNA from 1 to 14nt in size, hairpins such as a $(CAG)_{15}$ trinucleotide repeat, DNA flap substrates with 3' or 5' single stranded DNA overhangs and a splayed Y structure that is partially double stranded and partially single stranded (Palombo et al., 1996; Surtees and Alani, 2006). Msh2-Msh3 appears to have the highest affinity for 8 nucleotide loops of DNA and 3' flaps of DNA. It is unclear how Msh2-Msh3 can recognize this diversity of substrates that display such a wide variety of secondary structures.

1.3.2.2 Mechanism of mismatch recognition

One of the critical questions in the field of MMR is how mismatches are recognized. The mechanism of mismatch recognition for MutS and Msh2-Msh6 has been well characterized biochemically as well as by crystallographic studies. The MBD of MutS and Msh6 share a conserved phenylalanine residue (Phe36 and Phe337, in *E. coli* and *S. cerevisiae*, respectively) that when mutated to alanine causes severe defects in MMR (Alani, 1996; Das Gupta and Kolodner, 2000; Drotschmann et al., 2001; Dufner et al., 2000; Yamamoto et al., 2000). Mutation of the equivalent phenylalanine or tyrosine residue present in Msh2 does not cause a MMR defect, suggesting that Msh2 does not make direct contact with the mispaired base (Bowers et al., 1999; Dufner et al., 2000).

A crystal structure of MutS bound to a G:T mispair revealed several key features of mismatch recognition (Lamers et al., 2000). First of all, the DNA is bent by nearly 60 degrees when the mismatch is bound by MutS. Importantly, this bending causes the mispaired T to be flipped out of the DNA helix and available for interaction directly with the phenylalanine residue that mediates mismatch recognition. The specific interaction is via π-stacking of the phenylalanine ring with the mispaired T or +T insertion. Other contacts were identified between the surrounding DNA backbone and both subunits of the MutS homodimer. Many of these contacts are conserved between MutS and the eukaryotic Msh2-Msh6 heterodimer.

More recently, a crystal structure of human Msh2-Msh6 on a G:T mismatch and +1 insertion have been solved and confirmed the important features in common with MutS-mismatch recognition (Warren et al., 2007). Genetic studies in *S. cerevisiae* have confirmed that these conserved amino acids in Msh2 and Msh6 are indeed playing critical roles in mismatch recognition *in vivo* (Das Gupta and Kolodner, 2000; Drotschmann et al., 2001; Holmes et al., 2007).

There is no structural information available for any Msh2-Msh3 heterodimer. Msh3 does not have the conserved phenylalanine and glutamate residues found in Msh6 and MutS, that are essential for interaction with the a mismatched base or 1 nucleotide insertion. Furthermore, the wide variety of secondary structures recognized by Msh2-Msh3 indicates that there are indeed other amino acid residues within Msh3 that are directly interacting with DNA substrates.

Two studies have examined the footprint of the Msh2-Msh3 heterodimer bound to mispaired DNA substrates *in vitro*. The pattern of hydroxy-radical protection of a (CAG) hairpin substrate bound by Msh2-Msh3 is quite different from the pattern on a $(CA)_4$ loop (Owen et al., 2005). On the ideal loop substrate there is protection of the loop and also the 30 basepairs of duplex DNA at the base of the loop. In contrast, Msh2-Msh3 bound to a hairpin substrate shows minimal contact with the duplex DNA at the base of a

hairpin and reduced catalytic activity, suggesting an altered function for the Msh2-Msh3 complex.

1.3.2.3 Domain features

Each subunit of a mismatch recognition complex is organized into five domains and one floppy N-terminal region. Domains I and IV encircle the DNA strands making important contacts with the mispair and the DNA backbone. Domain V contains the ATP binding site and dimerization interface. The N-terminal region of Msh6 has been examined structurally and shown to act as a floppy unstructured tether to PCNA (Shell et al., 2007b). PCNA forms a trimeric ring that encircles DNA, controls processivity of DNA polymerase δ and ε and interacts with many repair proteins including Rad27, DNA helicases, DNA ligase, base-excision proteins Ung1 and Apn, nucleotide excision protein XPG, chromatin assembly factor Cac1 and mismatch repair proteins Msh6 and Msh3. It is thought that PCNA recruits these proteins, to the site of newly replicated DNA and enhances the overall fidelity of duplicating the genome, but that remains to be shown for Msh3.

In order to better understand the shared and unique features of the two eukaryotic heterodimers, previous studies have swapped domains between the Msh3 and Msh6 proteins. This work revealed that a portion of the Msh3 MBD could be placed into Msh6 creating a functional chimeric protein with the mispair specificity of Msh3 (Shell et al., 2007a). Many other domain swap

alleles lead to nonfunctional proteins presumably due to a disruption of the precise domain structure.

1.3.3 Transmission of the mismatch recognition signal

Binding to mispaired DNA causes a conformational change in the MutS and Msh2-Msh6 complexes that is important for downstream steps of MMR. It has been reported that the interaction with mispaired DNA stabilizes the MBD during crystallization studies (Obmolova et al., 2000). Also, mispair binding changes the affinity of MutS homolgoues for adenine nucleotide binding and hydrolysis (Antony and Hingorani, 2003; Bjornson et al., 2000; Gradia et al., 2000). This is quite interesting because the MBD and the ATP-binding domain are on opposite ends of a MutS homolgoue protein (Obmolova et al., 2000). Therefore, nucleotide binding on one end of the heterodimer induces a signal that is transmitted to the ATPase domain on the other end by a conformational change.

The transmission of the mismatch recognition signal is critical for the MMR process as evidenced by mutations, found in HNPCC patients as well as other studies, that disrupt the ATP pocket or transmission of conformational change, leading to complete lack of MMR activity. As previously mentioned, HNPCC families have mutations in the *MSH2* and *MLH1* gene, the non-redundant components of the eukaryotic MutSα and MutLα homologues (Fishel et al., 1993; Huang et al., 2001b; Peltomaki, 2003). Additionally,

mutations have been identified in *MSH6* that appear to cause milder cancer-predisposition syndromes but no mutations have been found in the *MSH3* gene that are associated with cancer susceptibility (Edelmann et al., 2000; Huang et al., 2001a; Kolodner and Marsischky, 1999; Wagner et al., 2001).

It is important to note the complexity of the ATP binding domains. First of all, each ATPase domain can bind to ATP, ADP or remain empty. Secondly, each ATPase domain is comprised of residues from both subunits and the conformational change brought about by mispair binding is transmitted through each subunit (Lamers et al., 2000). Thirdly, the ATP-binding site comprised mostly of Msh6 residues has a higher affinity for ATP than the site comprised of mostly Msh2 residues (Antony and Hingorani, 2004; Lamers et al., 2003). Finally, the two ATP-binding sites can communicate with each other since ATP binding by the Msh6 site causes the Msh2 site to lose affinity for ADP (Mazur et al., 2006).

Recent studies have led to a model for the relationship between mispair binding, nucleotide occupancy of the ATPase domains and movement of the MutS homologue away from the mispair site. Specific binding to a mispair occurs when both ATPase domains are empty or Msh2 is bound to ADP. This causes Msh6 to bind ATP stably and induces Msh2 to also bind ATP. The heterodimer next converts into a clamp that slides off of the mispair but remains on the DNA and is competent to interact with a MutL homologue

(Acharya et al., 2003; Gradia et al., 1999; Mazur et al., 2006; Mendillo et al., 2005).

1.4 MutL homologues

The MutL proteins play an important role in coordinating subsequent steps in MMR. Like MutS proteins, they form dimers, bind and hydrolyze ATP and interact with other proteins. They also interact with DNA in a mispair-independent manner. In eukaryotes there are two MutL complexes: Mlh1-Pms1 (MutLα) and Mlh1-Mlh3 (MutLβ). Mlh1-Pms1 appears to play a more important role in MMR and interacts primarily with Msh2-Msh6 but also with Msh2-Msh3 (Li and Modrich, 1995; Prolla et al., 1994b). Mlh1-Mlh3 is thought to only interact with Msh2-Msh3 (Flores-Rozas and Kolodner, 1998; Wang et al., 1999).

These general interactions were initially discovered by genetic experiments where a *pms1* mutant strain shows a high mutation rate indistinguishable from *msh2* (Li and Modrich, 1995; Prolla et al., 1994a; Prolla et al., 1994b), and an *mlh3* strain accumulates a small but significant increase in frameshift events, like those seen in an *msh3* strain. Also, the *mlh3 msh6* double mutant strain shows an increase in frameshifts indicating that Mlh3 is involved in Msh3-based repair, distinct from Msh6 (Flores-Rozas and Kolodner, 1998). In subsequent biochemical studies, the relative protein levels in HeLa cell extracts were measured to be a 10–fold excess human

MutLα complex (hMlh1-hPms2) to human MutLβ (hMlh1-hPms1) (Cannavo et al., 2005; Raschle et al., 1999).

A MutL complex is capable of binding to a MutS complex on mispaired DNA, in the presence of ATP, suggesting the formation of a large stable ternary complex (Acharya et al., 2003; Mendillo et al., 2005). This has been observed by gel shift analysis, surface plasmon resonance and DNA footprinting studies where the region of DnaseI protection increases from 20 to 100 basepairs with the addition of MutL (Mendillo et al., 2005; Schofield et al., 2001). The human MutLα complex shows MMR activity *in vitro* when lysates from mismatch repair-deficient cell lines are complemented with hMlh1 and hPms2 proteins, but the MutLβ complex did not show MMR activity (Raschle et al., 1999). One study has suggested that yeast MutLα can form a ternary complex with yMsh2-Msh3 *in vitro* and enhance mismatch binding (Habraken et al., 1997).

Until recently, no activity has been assigned to MutL other than its ability to recruit MutH, an endonuclease, to the site of a MutS bound mispair. Recent studies have identified an endonuclease activity of the human Mlh1-Pms2, yeast Mlh1-Pms1 and yeast Mlh1-Mlh3 complexes (Erdeniz et al., 2007; Kadyrov et al., 2006; Kadyrov et al., 2007; Nishant et al., 2008). While the active site residues of this endonuclease motif are not found in MutL sequences from MutH-containing prokaryotes, this activity is required for eukaryotic MMR (Kadyrov et al., 2006). This endonuclease activity from MutL

homologues in eukaryotes may act in the initiation of the excision step of MMR.

As mentioned previously, in prokaryotes, MutH is responsible for placing nicks on the newly replicated strand of DNA, which is discriminated by its incomplete methylation by Dam methylase immediately following replication. In eukaryotes, it now seems likely that MutL homologues are able to place nicks in DNA *in vitro*, but it remains to be shown whether this provides effective strand discrimination *in vivo*. The possibility remains that the transient nicks present in the lagging strand immediately following replication may direct MMR to the appropriate strand (Kadyrov et al., 2006).

Defects in the MutL homologues are also associated with the cancer predisposition syndrome HNPCC. HNPCC patients have been identified with mutations in the *hMLH1* gene while mutations in the *hPMS2* and *hPMS1* genes are extremely rare (Kolodner and Marsischky, 1999). The partial redundancy of the Mutsα and Mutsβ complexes as well as the redundancy of the MutLα and MutLβ complexes could explain why cancer causing mutations are mostly associated with the non-redundant component of each heterodimer. It is also clear that an imbalance in the ratio of subunit proteins, can drive the formation of homodimers rather than heterodimers, and cause a mutator phenotype, as observed in case of overexpression of Mlh1 (Shcherbakova et al., 2001).

Many questions about the mechanism of MMR remain to be answered including the mechanism of mispair recognition by Msh2-Msh3. Does the type of DNA lesion recognized by Msh2-Msh3 determine whether the Mlh1-Pms1 or Mlh1-Mlh3 heterodimer is subsequently recruited? How does the recruitment of the Mlh1-Pms1 versus Mlh1-Mlh3 complex effect downstream repair events?

Chapter 2: *Saccharomyces cerevisiae* Msh2-Msh3 acts in repair of single base-base mispairs

2.1 Introduction

For a cell to survive and grow normally, it must maintain the fidelity of its genome. To do this the cell utilizes multiple mechanisms to minimize the rate at which mutations occur. DNA mismatch repair is one such highly conserved mechanism that recognizes and repairs mispaired bases in DNA caused by replication errors, recombination or chemical damage to DNA and DNA precursors (Kolodner and Marsischky, 1999; Modrich, 1991). The importance of mismatch repair is evidenced by the fact that inherited mutations in two human mismatch repair genes, *MSH2* and *MLH1,* are responsible for most cases of Hereditary Non-Polyposis Colorectal Cancer (HNPCC) and that epigenetic silencing of *MLH1* underlies most cases of sporadic mismatch repair defective cancer (Lynch and de la Chapelle, 2003; Peltomaki, 2003).

The mechanism of mismatch repair is best understood in *E. coli* where mismatch repair has been reconstituted *in vitro* with purified proteins and defined DNA substrates (Harfe and Jinks-Robertson, 2000; Modrich, 1991; Modrich and Lahue, 1996). In this reaction, the MutS protein homodimer recognizes the abnormal DNA structure of base:base or insertion/deletion mispairs (Joshi et al., 2000; Schofield et al., 2001; Su et al., 1988). The MutL

homodimer binds to the MutS-DNA complex and activates the MutH endonuclease which nicks the unmethylated DNA strand at a hemi-methylated GATC sequence, targeting repair to newly synthesized DNA strands (Acharya et al., 2003; Au et al., 1992; Galio et al., 1999; Grilley et al., 1989; Hall and Matson, 1999; Selmane et al., 2003; Welsh et al., 1987). The nicked DNA strand is unwound by the UvrD helicase and degraded by a number of exonucleases, resulting in excision of the mispaired base; repair is completed by resynthesis of the excised strand (Burdett et al., 2001; Lahue et al., 1989; Matson and Robertson, 2006). The detailed molecular mechanisms of many aspects of this reaction are still under investigation (Acharya et al., 2003; Allen et al., 1997; Junop et al., 2001; Selmane et al., 2003).

The mismatch repair system in eukaryotes, while conserved with that of bacteria, is more complex. Nonetheless, many of the proteins involved have been identified, their general biochemical properties determined and at least partial repair reactions resembling those of bacteria have been reconstituted *in vitro* with purified proteins (Constantin et al., 2005; Zhang et al., 2005). In eukaryotes, the dimeric MutS mismatch recognition protein has been replaced by two different heterodimers of MutS homologue proteins, the Msh2-Msh6 and Msh2-Msh3 complexes (Kolodner and Marsischky, 1999). Similarly, the MutL dimer has been replaced by two different heterodimers of MutL homologue proteins, the Mlh1-Pms1 (Pms2 in humans) and Mlh1-Mlh3 complexes (Cannavo et al., 2005; Flores-Rozas and Kolodner, 1998; Prolla et

al., 1994b; Wang et al., 1999). In addition, it has been suggested that a third MutL homologue complex, Mlh1-Mlh2 (Pms1 in humans), may play a minor role in mismatch repair although biochemical studies do not support this (Harfe et al., 2000; Raschle et al., 1999). DNA polymerase δ, RPA, PCNA, RFC and Exo1 have been shown to act in eukaryotic mismatch repair, although evidence suggests that additional proteins may be involved (Modrich, 2006).

Current models of eukaryotic mismatch repair suggest that the Msh2-Msh6 complex is the major mismatch recognition complex and functions in repair of base:base and insertion/deletion mispairs (Harfe and Jinks-Robertson, 2000; Kolodner and Marsischky, 1999; Modrich, 1991; Modrich, 2006; Modrich and Lahue, 1996). The Msh2-Msh3 complex is redundant with the Msh2-Msh6 complex with respect to the repair of small insertion/deletion mispairs and is also able to recognize larger insertion/deletion mispairs (Marsischky et al., 1996; Marsischky and Kolodner, 1999; Sia et al., 2001). A number of genetic results are consistent with this scenario: null mutations in *MSH2* result in a strong mutator phenotype characterized by the accumulation of base substitution and frameshift mutations; *MSH6* defects result in a strong mutator phenotype with respect to base substitutions but only a small increase in frameshift mutations; *MSH3* defects cause weak mutator phenotypes characterized by the accumulation of frameshift mutations, however in assays where larger frameshift mutations are analyzed, stronger mutator phenotypes are observed; lastly an *msh3 msh6* double mutant recapitulates the mutator

phenotype of an *msh2* single mutant (Marsischky et al., 1996; Sia et al., 2001).

Similar studies have led to the view that the Mlh1-Pms1 complex is the major MutL homologue complex that functions in eukaryotic mismatch repair whereas the Mlh1-Mlh3 complex plays a minor role in mismatch repair and is partially redundant with the Mlh1-Pms1 complex (Flores-Rozas and Kolodner, 1998; Prolla et al., 1994b; Wang et al., 1999). Genetic results supporting this view are as follows: null mutations in *MLH1 and PMS1* result in a strong mutator phenotype characterized by the accumulation of base substitution and frameshift mutations; *MLH3* defects result in a weak mutator phenotype primarily characterized by the accumulation of frameshift mutations; deletion of both *MLH3* and *PMS1* (*PMS2* in human and mouse) is required to recapitulate the mutator phenotypes (and cancer prone phenotype in mice) caused by a defect in *MLH1* (Chen et al., 2005; Heyer et al., 1999). Genetic analysis has also suggested that the Mlh1-Mlh3 complex primarily functions in conjunction with the Msh2-Msh3 complex (Cannavo et al., 2005; Flores-Rozas and Kolodner, 1998; Prolla et al., 1994b; Raschle et al., 1999; Wang et al., 1999). Biochemical studies are consistent with the Mlh1-Pms1 complex playing the major role in mismatch repair, whereas the Mlh1-Mlh3 complex, which has been much less studied, only has weak *in vitro* mismatch repair activity (Cannavo et al., 2005; Constantin et al., 2005).

While the studies establishing the roles of the eukaryotic MutS and MutL homologue complexes in mismatch repair seem quite definitive, it is important to note that they have some limitations. First, the genetic results are based on a few types of assays. Reversion assays can only detect a limited number of types of mutations. Forward mutation assays are less biased but prior mutation spectrum analysis was performed at a time when it was not feasible to sequence large numbers of mutations in large unbiased forward mutation targets like the *CAN1* gene. Even with analysis of small forward mutation targets, where large numbers of mutations can be analyzed, it is difficult to control for biological variation within mutation spectrum analysis experiments. Second, the mutations observed in a given mutant background are the result of a complex process involving misincorporation errors at individual sites combined with how efficiently other competing pathways, including editing exonucleases, bypass DNA polymerases and the different mismatch repair pathways act on mispairs and mispair producing errors. Third, because of the low mutation rates caused by defects in *MSH3* and *MLH3*, it has been difficult to genetically characterize the roles of Msh2-Msh3 and Mlh1-Mlh3 *in vivo*, which is further complicated by the fact that these defects are masked by the activity of the Msh2-Msh6 and Mlh1-Pms1 complexes, respectively (Flores-Rozas and Kolodner, 1998; Marsischky et al., 1996; Sia et al., 2001). Lastly, biochemical studies have used a limited diversity of substrates and mispairs predicted to occur *in vivo* have generally

not been used as substrates *in vitro*. Here we have used a genetic approach to identify mutations that arose in the absence of the *S. cerevisiae* proteins Msh6 or Msh3 *in vivo* and then used DNA substrates derived from the mutated sequences to analyze Msh2-Msh3 and Msh2-Msh6 binding affinities *in vitro*. Our results indicate that Msh2-Msh3 plays a previously unrecognized role in the repair of specific base:base mispairs and implies that the Mlh1-Mlh3 complex may also function in similar repair reactions. Additionally, we demonstrate that Msh2-Msh3 and Mlh1-Mlh3 play a previously unrecognized role in the suppression of homology-mediated duplication and deletion mutations.

2.2 Mutation spectra analysis of mismatch repair deficient strains

The roles of Msh2, Msh3 and Msh6 in mismatch repair were initially inferred by determining the rates and spectra of reversion of specific frameshift mutations as well as forward mutation of the *CAN1* gene in *S. cerevisiae* strains lacking different combinations of Msh2, Msh3 and Msh6 (Harfe and Jinks-Robertson, 2000; Marsischky et al., 1996; Sia et al., 2001). However, these studies only analyzed small numbers of forward mutations, potentially limiting the conclusions of these early experiments. To re-investigate the role of Msh3 in mismatch repair, we analyzed the spectrum of mutations that accumulate in an *msh3* strain (Table 2.1) in a larger number of independent mutants than previously analyzed. The *msh3* strain displayed an increase in

the proportion of frameshift and previously un-described homology-mediated duplication and deletions mutations (Supplementary Table 2.1) and a decrease in base substitution mutations when compared to the wild-type strain (p= 0.0001, chi squared “goodness of fit” test). The duplicated and deleted sequences were flanked by direct repeats that varied in length from 4 to 8 nucleotides; these duplication and deletion mutations were of the same type as previously seen in *rad27* mutants (Tishkoff et al., 1997) and many of the duplication and deletion mutations found in the *msh3* mutant were identical to those found in a *rad27* mutant (data not shown). In contrast, the mutation spectrum of the *msh6* strain consisted almost exclusively of base substitutions as previously reported (Marsischky et al., 1996) and showed no homology-mediated duplication and deletion mutations. The mutation spectrum of the *mlh3* strain was intermediate between that of the *msh3* and wild-type strains. For purposes of reference, data for a *mlh1* mutant strain, expected to be null for mismatch repair, are presented in Tables 2.1-2.3.

To further characterize the mutator phenotype of the mismatch repair defective strains the overall mutation rate of each strain was determined by fluctuation analysis and the rate of each class of mutation was then calculated (Table 2.2). The overall mutation rates of the *msh3* and *mlh3* strains were significantly higher than the wild-type strain (p= 0.0001, p= 0.0117, respectively, two-tailed Mann-Whitney test), and were indistinguishable from each other (p= 0.3735). The rate of base substitutions in the *msh3* strain was

not different from that in the wild-type strain (p= 0.7263) whereas the *msh3* strain had significantly increased rates of both frameshift mutations (p= 0.0001) and homology-mediated duplication and deletion mutations (p= 0.0001). Compared to the wild-type strain the *mlh3* mutant showed a slight increase in base substitutions (p= 0.0308) and an increase in both frameshift (p= 0.0002) and homology-mediated deletion and duplication mutations (p= 0.0002), albeit not as large as seen in the *msh3* mutant. This is consistent with previous models that Mlh1-Mlh3 is a MutL homologue protein complex that acts in a subset of Msh2-Msh3 dependent mismatch repair events (Flores-Rozas and Kolodner, 1998; Harfe and Jinks-Robertson, 2000). These mutation spectra were different from those of the *msh6* and *mlh1* strains, which showed an increase in only frameshift and base substitution mutations. It should be noted that an *mlh1* mutant is completely mismatch repair defective resulting in sufficiently high rates of base substitution and frameshift mutations that, given our sample size, homology-mediated duplication and deletion mutations would not be detected had they occurred at the same rate as in *msh3* or *mlh3* mutants.

To further study mispair recognition by the Msh2-Msh6 and Msh2-Msh3 heterodimers we analyzed the spectra and classes of base substitutions that accumulate in mismatch repair defective strains in detail (Table 2.3 and Supplementary Figures 2.1-2.4). We found that although the overall rate of base substitutions in the *msh3* mutant was not significantly different than that

of the wild-type strain, the spectrum of base substitutions observed was different and showed very little overlap between the positions of either single or recurrent mutations between the two spectra. Of the 68 and 61 base substitutions observed in wild-type and *msh3* strains respectively, only 8 of the mutations in the wild-type spectrum were in common with the *msh3* spectrum and 5 of the mutations in the *msh3* spectrum were in common with the wild-type spectrum; in addition, none of the six hotspots from the *msh3* spectrum overlaped with the wild-type mutation spectrum. A particularly interesting example was seen at the mutation hotspot *CAN1* codon 399 that was mutated from CGT (Arg) to CCT (Pro) in *msh3* two times and mutated to CAT (His) three times in wild-type but not mutated in the other strains analyzed. Similarly, there was also little or no overlap between the *mlh3* or *msh6* mutation spectra and the wild-type mutation spectrum (see Supplementary Figures 2.1-2.4). These data raise the possibility that *msh3* and *mlh3* mutations, like *msh6* mutations, alter the repair of base-base mispairs, although there are limitations to this analysis because it is difficult to ensure that the different mutation spectra have reached saturation. We did not make direct comparisons using the *mlh1* mutation spectrum because the *mlh1* mutator phenotype is dominated by frameshift mutations and hence the base substitution mutation sample size was small.

Although saturation of the mutation target did not occur, we could analyze the overlap of the mutation spectra for the various strains. By fitting

the base substitution mutation spectra data (specifically the number of unique and recurrent mutations) to Poisson distributions, we estimated that the wild-type strain contains 159 sites available for mutation, and the *msh3* strain contains 259 sites; the larger number of mutable sites in the *msh3* mutant is consistent with a defect in repair leading to base substitution mutations. Next we examined the relationship between the 159 wild-type and 259 *msh3* mutation sites. We found that: 1) it is unlikely that the wild-type and *msh3* strains explicitly share the same mutation sites (p=0.0215 assuming 259 sites available to both strains; $p=7.992 \times 10^{-5}$ assuming 159 sites available to both strains) and that 2) it is unlikely that the deletion of *MSH3* simply added additional mutation sites to a wild-type strain (p=0.032). Thus, our data suggests that while the spectra of base substitution mutations in both the wild-type and *msh3* strains overlap to some degree, they are likely to be different from each other.

To further analyze the spectrum of base substitution mutations seen in the *msh3*, *msh6* and *mlh3* mutants, the classes of observed base substitution mutations were compared (Table 2.3). The *msh3*, *msh6* and *mlh3* strains showed statistically different spectra of mutation classes compared to wild-type (p= 0.0001, p= 0.0003 and p= 0.0147, respectively, two-tailed chi squared “goodness of fit” test). Compared to wild-type, the *msh3* strain showed increases in GC basepair (bp) to CG bp and AT bp to TA bp mutations and decreases in GC bp to AT bp and AT bp to CG bp mutations.

Consistent with the known role of Msh6 in the repair of base-base mispairs, the *msh6* strain showed distinct differences from the wild-type strain including a striking absence of GC bp to CG bp base substitutions. One hypothesis that could explain these data is that the Msh2-Msh3 complex is able to specifically recognize one of the base:base mispairs involved in these two classes of base substitutions seen to increase: either a GG or CC mispair involved in a GC bp to CG bp base substitution and either an AA or TT mispair involved in an AT bp to TA bp base substitution.

2.3 Expression of Msh3

We initially attempted to overproduce the Msh2-Msh3 complex by fusing the *GAL10* promoter and an optimal Kozak consensus sequence to codon 1 of *MSH3* and co-express it with Msh2 but were unable to detect significant amounts of Msh3-FLAG by Western blotting. Similarly, a reconstructed version of a previously published Msh3 expression vector containing the *GAL10* promoter upstream of codon 1 (Habraken et al., 1996) only resulted in low level Msh3 expression. This led us to consider whether Met codon 1 was in fact the correct translational start codon. By aligning the conserved PCNA binding motif present in various *Saccharomyces* Msh3 proteins, we found that the equivalent of *S. cerevisiae* Met codon 30 was conserved among all of the Msh3 proteins analyzed whereas only *S. cerevisiae* and *S. paradoxus MSH3* contained 29 upstream codons including Met codon 1, suggesting that the start site for translation might be at codon 30

(Figure 2.1A). Plasmids were then constructed that contained the native *MSH3* promoter and gene to analyze Ala substitution mutations at each position, M1A and M30A. The *msh3-M1A* allele was able to complement the mutator phenotype of the *msh3 msh6* strain to the same level as wild-type *MSH3* (Figure 2.1B) but neither the *msh3-M30A* allele or vector control were able to complement. These results indicate that the Met codon 30 is the initiation codon for *in vivo* translation of the *S. cerevisiae MSH3* gene. Consistent with this, fusion of the *GAL10* promoter and an optimal Kozak consensus sequence to codon 30 of *MSH3*, resulted in approximately 10 times higher levels of Msh3 expression when Msh2 was co-expressed compared to that observed with the longer *MSH3* allele (Habraken et al., 1996).

2.4 Biochemical characterization of the Msh2-Msh3 complex

The *S. cerevisiae* Msh2-Msh3 complex was overexpressed and purified to near homogeneity as described under “Materials and Methods”. The mispair recognition properties of the Msh2-Msh3 heterodimer were investigated using electrophoretic mobility shift assays and compared to Msh2-Msh6 (Figure 2.2). As a positive control for Msh2-Msh3 mispair binding, we used a 38 basepair oligonucleotide duplex based on a previously described backbone and containing a 6 nucleotide insertion (Habraken et al., 1996; Hess et al., 2002). In addition, we created a series of 27 different control and mispair-containing 38 basepair oligonucleotide duplexes derived

from the sequence of the *CAN1* gene (Supplementary Table 2.2). The mispaired DNA substrates generated were from sites found to be mutated in the *msh3, msh6* or wild-type mutation spectra.

We initially examined binding of Msh2-Msh3 and Msh2-Msh6 to a GC basepair and GG, CC, AC and GT mispairs at *CAN1* coding nucleotide 1196 (Figure 2.2A) which is the position in codon 399 where a GC bp to CG bp mutation was found two times exclusively in the *msh3* mutation spectrum. As predicted, Msh2-Msh3 bound most robustly to the +6 insertion mispair and high level binding to the CC mispair was also found. Msh2-Msh3 also bound to the GG, AC and GT mispairs at a level which was weak but above the GC binding background. The Msh2-Msh6 complex bound the GG, AC, GT and +6 insertion substrates weakly and did not appear to significantly bind the CC mispair. Furthermore, the addition of ATP caused a decrease in binding of Msh2-Msh3 to CC and +6 insertion mispairs, consistent with rapid release via sliding (Figure 2.2B, 2.2C) (Mendillo et al., 2005; Wilson et al., 1999). This mispair binding specificity is consistent with the formation of a CC mispair at nucleotide 1196 of *CAN1* that subsequently escapes repair in an *msh3* mutant.

To extend the above results, binding of Msh2-Msh3 to a greater diversity of CAN1 derived base-base mispairs was analyzed (Table 2.4); 3 were at sites found to be mutated in the msh3 spectrum, and 2 each were at sites found to be mutated in the msh6 or wild-type spectra. The substrates

analyzed included sites that were and were not mutated in msh3 mutants and also included mispairs that were and were not predicted to underlie the classes of mutations that were preferentially found in the msh3 mutation spectrum. In these experiments, very strong binding (greater than 4-fold over the GC control) was observed for four mispairs (CC 1196, CT 1196, AA 1193 and AC 1193) and binding that was at least 2-fold above the control binding was observed for an additional seven mispairs (CC 413, AA 1196, AC 1196, GG 1196, AA 1628, AC 807 and AG 1196). Weak or no binding was observed for an additional fourteen mispairs. ATP promoted dissociation from the mispair in each case (data not shown). Of the eleven mispairs showing the strongest binding, six were of the classes suggested by mutation spectra analysis to undergo Msh2-Msh3 dependent repair: GG or CC and AA or TT. Of the seven total sites analyzed, Msh2-Msh3 did not show high-level mispair binding at two sites: 955, found to be mutated in an msh6 mutant and 314 found to be mutated in an msh3 mutant. However, Msh2-Msh3 did show binding at five other sites: 1193 and 1196 found to be mutated in an msh3 mutant, 413 and 1628 found to be mutated in the wild-type strain, and 807 found to be mutated in an msh6 mutant.

2.5 Discussion

In the present study we have used a combined genetic and biochemical approach to investigate the role of the Msh2-Msh3 complex in mismatch repair. Mutation spectrum analysis showed that *msh3* mutants, while having

low overall rates of base substitution mutations nonetheless accumulated a spectrum of such mutations that appeared distinct from those that accumulated in either *msh6* mutants or the wild-type strain, suggesting a role for the Msh2-Msh3 complex in the repair of some base:base mispairs. Mutation spectra analysis similarly suggested that the Mlh1-Mlh3 heterodimer might also function in the repair of base:base mispairs. In addition, *msh3* and *mlh3* but not *msh6* or *mlh1* mutants accumulated homology-mediated duplication and deletion mutations of the type only previously seen in *rad27* mutants (Tishkoff et al., 1997). The parallels between the mutation spectra of *mlh3* and *msh3* mutants is consistent with previous observations showing that the Mlh1-Mlh3 complex functions in conjunction with the Msh2-Msh3 complex (Flores-Rozas and Kolodner, 1998). Mispair binding analysis with DNA substrates derived from *CAN1* sequences found to be mutated *in vivo* demonstrated that Msh2-Msh3 had robust binding to specific base:base mispairs that was reduced upon the addition of ATP. Overall, the results presented here are consistent with an unexpected role of the Msh2-Msh3 complex and the Mlh1-Mlh3 complex in the repair of base:base mispairs as well as in the suppression of homology-mediated duplication and deletion mutations. This result could explain how, in one genetic study, Msh3 could partially suppress homeologous recombination between substrates containing four single base differences, although note that the effect of individual single base differences were not examined in the study (Nicholson et al., 2000).

Additionally, another study examining the repair of defined mismatches on transformed plasmids detected possible repair defects of some individual base:base mismatches caused by an *msh3* mutant in one strain background but not in another strain background (Luhr et al., 1998). The observation presented here that Msh2-Msh3 is able to bind to specific mispairs *in vitro* that were initially identified as potential mutation intermediates *in vivo*, demonstrates a specific mechanism that explains some of these earlier studies.

Previous studies have analyzed the effect of different mismatch repair defects in a number of mutator assays sometimes combined with sequencing of mutation spectra to infer the role of different proteins in mismatch repair. As noted in the "Introduction", these types of studies have a number of limitations. In the current study we sequenced larger numbers of independent mutations in a large relatively unbiased forward mutation target than in prior studies and found that a *msh3* strain appeared to accumulate a spectrum of base substitutions that differed from that of wild-type or *msh6* strains; differences in the spectrum of frameshift mutations were also observed, although we did not further analyze these mutations as it is well accepted that Msh2-Msh3 and Msh2-Msh6 both function in the repair of insertion/deletion mispairs (Harfe and Jinks-Robertson, 2000; Kolodner and Marsischky, 1999; Modrich, 1991; Modrich, 2006; Modrich and Lahue, 1996). While it is probably difficult to completely saturate the *CAN1* mutation spectrum in the strains tested,

nonetheless there was very little overlap in the mutation spectra and there were significant differences between the overall mutation spectra observed as well as in the classes of base substitutions seen. These data support the hypothesis that the Msh2-Msh3 complex functions in the repair of base:base mispairs. However, it is difficult to determine how efficiently the Msh2-Msh3 complex can act in such repair events *in vivo* because competition by Msh2-Msh6 dependent repair clearly obscures the *msh3* mutator phenotype. In mammalian cells, the Msh6 mismatch repair pathway dominates mismatch repair because the Msh2-Msh6 complex is found at 6- to 10-fold higher levels than the Msh2-Msh3 complex (Drummond et al., 1997; Genschel et al., 1998). The ratio of the two complexes is not known in *S. cerevisiae* due to the lack of good Msh3 antibodies. However, the observations that *msh3* mutants have detectable mutator phenotypes (Marsischky et al., 1996; Sia et al., 2001) and that a single copy *MSH3* plasmid can suppress dominant *msh6* mutations (Das Gupta and Kolodner, 2000) suggests that the Msh3 pathway plays a significant role in mismatch repair in wild-type *S. cerevisiae*.

By performing mispair binding studies with oligonucleotide duplexes based on the sequence of the *CAN1* gene that contained mispairs that were or were not found at sites mutated in the *msh3* mutation spectrum and were or were not the mispairs predicted to underlie the mutation seen, we were able to demonstrate that the Msh2-Msh3 complex could robustly bind specific base:base mispairs including ones that were not well recognized by Msh2-

Msh6; base:base mispairs that were not bound by Msh2-Msh3 were also found. The Msh2-Msh3 base:base mispair binding was sensitive to ATP addition, upon which Msh2-Msh3 quickly dissociated from the DNA substrate, as predicted for bona-fide mispair binding (Mendillo et al., 2005; Wilson et al., 1999). Significant binding of base:base mispairs by Msh2-Msh3 has not previously been observed; our use of mispairs based on *in vivo* mutation sites is likely what made it possible to observe binding of base:base mispairs by Msh2-Msh3. These results support the hypothesis that Msh2-Msh3 can function in the repair of base:base mispairs and suggest that such repair augments Msh2-Msh6 dependent repair of base:base mispairs. A considerable amount of structural information is available on how MutS recognizes mispairs, and Msh6 shares the key MutS mispair recognition structural determinants including the Phe residue that stacks on the mispaired base and other residues that contact the DNA backbone (Drotschmann et al., 2001; Lamers et al., 2000; Natrajan et al., 2003). Msh3 lacks this key Phe residue but retains six residues that contact the DNA backbone in MutS and are present in Msh6 as well (Lee et al., 2007). However, four of these residues have been mutated in Msh3 and only one was found to be important for Msh3 dependent mismatch repair (Drotschmann et al., 2001; Lee et al., 2007). Given our analysis indicating that the Msh2-Msh6 and Msh2-Msh3 complexes can recognize the same classes of mispairs (i.e., base:base and insertion/deletion mispairs), the lack of the mispair-contacting Phe residue and

the lack of a requirement for many of the other predicted critical DNA contacting residues suggests that Msh3 may utilize a distinct structural mechanism for mispair recognition.

Additionally, we found that *msh3* and *mlh3* mutants exhibit a significant increase in the rate of accumulation of homology-mediated duplication and deletion mutations. This type of mutation has been suggested to arise in *rad27* mutants due to errors in processing the ends of Okazaki fragments, leading to double strand breaks and aberrant repair of the double strand breaks (Tishkoff et al., 1997) possibly by single strand annealing recombination (SSA). However, the deletion and duplication mutations seen in an *msh3* mutant (and probably in an *mlh3*) probably cannot be mediated by SSA because Msh3 is required for SSA and in particular those events that occur by SSA between short homologous DNA sequences like those implicated in the deletion and duplication mutations seen here. It seems unlikely that loss of Msh3 and Mlh3 cause the same type of defects as loss of Rad27 since an *msh3* mutation did not cause an increase in the rate of gross chromosomal rearrangements as seen for *rad27* mutants (data not shown) (Chen and Kolodner, 1999). It is also unlikely that all errors arising in the absence of Rad27, which result in duplication and deletion events, are normally repaired by mismatch repair since mismatch repair defects do not result in a synergistic increase in mutation rate when combined with a *rad27* mutation (unpublished results) (Tishkoff et al., 1997). More likely possibilities

are that either Rad27, Msh2-Msh3 and Mlh1-Mlh3 function together in a subclass of repair events or only a proportion of the errors induced by the absence of Rad27 are repaired by Msh2-Msh3 and Mlh1-Mlh3 dependent mismatch repair. For example, Msh2-Msh3 and possibly Rad27, could interact with aberrant branched structures that form at stalled or damaged replication forks (Surtees and Alani, 2006). Additional studies will be required to elucidate the exact mechanisms involved.

In summary, the results presented here indicate a need for modification of the current models of mismatch repair such that in the early step of mismatch repair, both Msh2-Msh6 and Msh2-Msh3 recognize base:base and insertion/deletion mispairs; this redundancy likely increases the overall efficiency of mismatch repair. In addition, our results have implicated the Msh2-Msh3 and Mlh1-Mlh3 complexes in the suppression of homology-mediated duplication and deletion mutations like those that occur in *rad27* mutants thus expanding current views of the role of mismatch repair in suppressing mutations.

Table 2.1 Mutation spectra analysis of mismatch repair-deficient strains

Genotype	Strain	No. (%) of indicated mutation types				n^d
		Base substitution	Frameshift (±1 or 2)	Homologous duplication/ deletion[b]	Other[c]	
WT[a]	RDKY3686	68 (81)	12 (14)	2 (2)	2 (2)	84
msh3	RDKY4149	61 (46)	45 (34)	23 (18)	2 (2)	132
msh6	RDKY4151	90 (89)	10 (10)	0	1 (1)	101
mlh3	RDKY5295	50 (69)	17 (24)	4 (5)	1 (2)	72
mlh1	RDKY4237	23 (41)	32 (57)	0	1 (1)	56

[a] WT, wild type.

[b] Homologous duplication/deletions are defined as the addition or removal of a sequence flanked by blocks of 4 to 8 nucleotides of homology (51); also see Table S1 in the supplemental material.

[c] Other mutation types include single isolates that contain more than one mutation.

[d] *n*, number of mutations.

Table 2.2 Mutation rate analysis of mismatch repair-deficient strains

Genotype	Strain	Rate (fold increase) for indicated mutation[a]			
		Overall Can[r]	Base substitution	Frameshift (±1 or 2)	Homologous duplication/deletion
WT[b]	RDKY3686	$7.5\ [3.8–8.9] \times 10^{-8}$ (1)	6.1×10^{-8} (1)	1.1×10^{-8} (1)	1.5×10^{-9} (1)
msh3	RDKY4149	$1.3\ [1.1–1.4] \times 10^{-7}$ (1.6)	5.8×10^{-8} (0.94)	4.4×10^{-8} (4)	2.3×10^{-8} (15)
msh6	RDKY4151	$6.7\ [5.9–10.1] \times 10^{-7}$ (8.9)	6.0×10^{-7} (9)	6.7×10^{-8} (6)	0
mlh3	RDKY5295	$1.1\ [0.7–1.6] \times 10^{-7}$ (1.4)	7.6×10^{-8} (1.3)	2.6×10^{-8} (2.5)	5.5×10^{-9} (3.7)
mlh1	RDKY4237	$3.2\ [2.3–4.1] \times 10^{-6}$ (43)	1.3×10^{-6} (21)	1.8×10^{-6} (164)	0

[a] The increase relative to the value for the wild type is shown in parentheses. Ninety-five percent confidence intervals are shown in brackets. The overall mutation rate was used to calculate the rate of each class of mutations.
[b] WT, wild type.

Table 2.3 Classes of base substitutions found in mismatch repair-deficient strains[a]

Base pair mutation class	% Substitutions for indicated genotype[b]				
	WT	*msh6*	*msh3*	*mlh3*	*mlh1*
GC to CG	12 (8)	0	18 (11)	10 (5)	9 (2)
GC to AT	38 (26)	42 (38)	27 (17)	30 (15)	39 (9)
GC to TA	26 (18)	38 (35)	30 (18)	32 (16)	26 (6)
AT to TA	4 (3)	2 (2)	13 (8)	10 (5)	13 (3)
AT to GC	6 (4)	10 (9)	7 (4)	8 (4)	4 (1)
AT to CG	13 (9)	8 (7)	5 (3)	10 (5)	9 (2)

[a] *n* was 68 for the wild type (WT), 91 for *msh6*, 61 for *msh3*, 50 for *mlh3*, and 23 for *mlh1*.

[b] The observed number of mutations of each type is shown in parentheses. For the GC to CG base pair class, *msh6* significantly differed from the wild type (the Fisher exact test, $P = 6.8 \times 10^{-5}$). For the AT bp to TA bp class, *msh3* significantly differed from the wild type (the Fisher exact test, $P = 0.04$). Both of these analyses used wild-type mutation spectrum data combined with that of Tishkoff et al. (51). Two other comparisons were of borderline significance: the GC to CG base pair class in *msh3* compared to that in the wild type and the GC bp to TA bp class in *msh6* compared to that in the wild type. For these latter comparisons to reach statistical significance, three-times- and two-times-larger sample sizes would be required, respectively.

Table 2.4 Binding of Msh2-Msh3 to mispairs in different sequence contexts[a]

Origin	Mispair	Position in *CAN1*	% Binding[b]	Score[c]	Expected[d]
msh3	GC	1196	7.1	−	No
	GG	1196	18.5	+	Yes
	CC	1196	45.4	++	Yes
	AA	1196	20.2	+	Yes
	TT	1196	9.5	−	No
	GT	1196	13.5	−	No
	AC	1196	18.6	+	No
	CT	1196	37	++	No
	AG	1196	25.5	+	No
	+6	NA	75.7	++	NA
msh3	GC	1193	8.4	−	No
	GG	1193	8.7	−	No
	AA	1193	33.3	++	Yes
	TT	1193	13.4	−	No
	GT	1193	11	−	No
	AC	1193	47.9	++	No
	CT	1193	9.6	−	No
	AG	1193	12	−	No
	+6	NA	85.4	++	NA
WT	GG	413	8.4	−	No
	CC	413	17.3	+	Yes
msh6	AA	955	9.9	−	No
	TT	955	8.7	−	No
WT	AA	1628	16.2	+	Yes
	TT	1628	12.6	−	No
msh3	GT	314	11.4	−	No
	AC	314	9.7	−	No
msh6	GT	807	11.6	−	No
	AC	807	22.3	+	No
	+6	NA	44	++	NA

[a] NA, not applicable; WT, wild type.

[b] The percentage of substrate shifted from the total signal per lane was calculated by densitometry and is shown as "% Binding" for the base pairs/mispairs at the indicated *CAN1* positions.

[c] The three degrees of binding are noted under "Score." ++, increased binding of more than fourfold compared to GC; +, increased binding of two- to fourfold compared to GC; −, increased binding of less than twofold compared to GC.

[d] "Expected" indicates whether a mispair was predicted to be recognized based on the mutations seen in the *CAN1* mutation spectra.

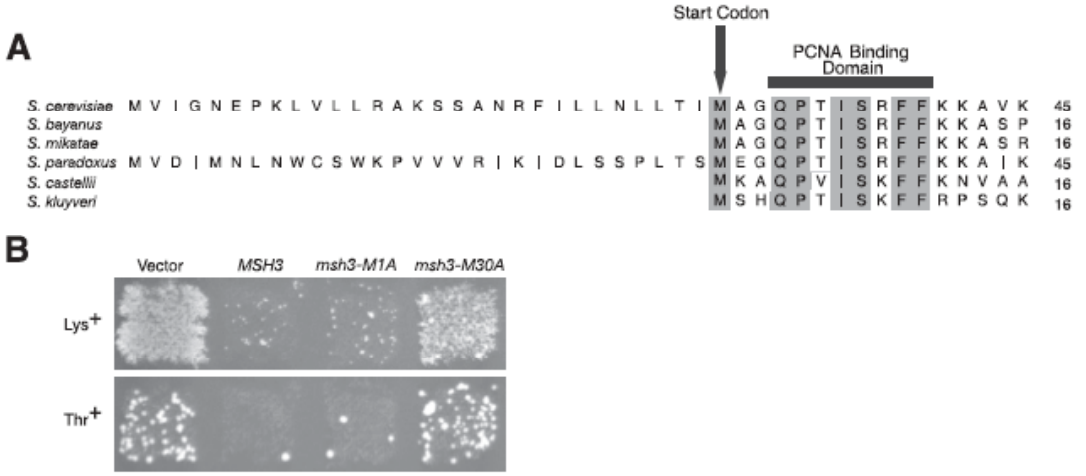


Figure 2.1 Identification of *MSH3* translation start site
(A) Alignment of predicted fungal Msh3 protein sequences. The protein sequence, according to the *Saccharomyces* genome database, of Msh3 from various organisms is aligned based on the conserved PCNA-binding motif. (B) *MSH3* complementation of a *msh3 msh6* strain. The *msh3* alleles were expressed on a low-copy-number plasmid bearing a marker allowing growth on media lacking Leu. Plasmids were transformed into the *msh3 msh6* strain, and isolates were patched onto plates lacking leucine and then replica plated onto plates lacking lysine and threonine as shown.

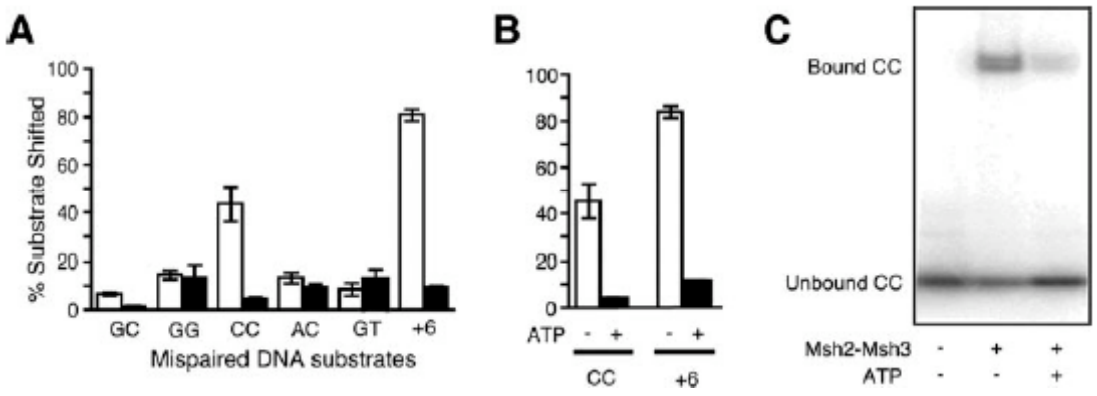


Figure 2.2 Binding of Msh2-Msh3 to mispaired DNA substrates *in vitro*

(A) Sixteen nanomolar of purified heterodimer (white bars, Msh2-Msh3; black bars, Msh2-Msh6) was incubated with 14nM ^{32}P-labeled substrates as described in Materials and Methods and analyzed by gel shift assay. The percentage of substrate shifted from the total signal per lane was calculated by densitrometry. Error bars indicate one standard deviation. (B) The addition of ATP promotes dissociation of Msh2-Msh3 from mispaired DNA. Binding reactions were performed, and then 500 υM ATP was added on ice for 15 min (black bars). -, absence of; +, presence of. (C) Gel shift analysis of Msh2-Msh3 specifically bound to a CC mispair.

Supplementary Table 2.1 Insertion and deletion mutations found in *CAN1*

Strain	Mutation	Occurrence
RDKY4149 *msh3*	I_{46-88} (43,4)	1
	$I_{257-273}$ (17,5)	1
	$I_{286-303}$ (18,4)	1
	$I_{310-326}$ (18,6)	1
	$D_{382-386}$ (5,5)	1
	$I_{382-386}$ (5,5)	1
	$D_{383-386}$ (4,4)	2
	$I_{512-524}$ (13,5)	1
	$D_{753-760}$ (8,3)	1
	$I_{756-803}$ (48,4)	1
	$I_{761-809}$ (49,5)	1
	$D_{970-975}$ (3,3)	1
	$I_{1073-1086}$ (14,3)	1
	$I_{1265-1336}$ (72,7)	1
	$D_{1303-1310}$ (8,6)	2
	$D_{1323-1338}$ (16,4)	1
	$D_{1386-1398}$ (13,3)	1
	$I_{1401-1424}$ (24,6)	1
	$D_{1529-1534}$ (6,4)	1
	$I_{1529-1534}$ (6,4)	1
	$I_{1580-1617}$ (38,6)	1
RDKY5295 *mlh3*	$I_{219-250}$ (32,5)	1
	$D_{226-251}$ (26,4)	1
	$I_{399-432}$ (33,4)	1
	$D_{1305-1311}$ (7,3)	1
RDKY5724 *rad27*	$I_{207-231}$ (25,4)	1
	$I_{289-306}$ (18,3)	1
	$I_{290-307}$ (18,4)	2
	$I_{310-327}$ (18,6)	2
	$I_{419-454}$ (36,4)	1
	$I_{761-809}$ (49,5)	3
	$I_{1003-1020}$ (18,6)	1
	$I_{1032-1065}$ (34,4)	1
	$I_{1073-1086}$ (14,3)	1
	$I_{1190-1227}$ (38,4)	1
	$D_{1354-1392}$ (39,4)	1
	$I_{1387-1398}$ (12,3)	1
	$I_{1401-1425}$ (24,6)	1
	$I_{1486-1518}$ (33,3)	1
	$I_{1580-1618}$ (38,6)	1

Supplementary Table 2.1 (continued) Insertion and deletion mutations found in *CAN1*

The nomenclature for insertion and deletion mutations has been described previously (Tishkoff et al., 1997). I, indicating insertion, and D, indicating deletion, is followed by the coordinates of the sequence affected by insertion or deletion. The first number in parenthesis is the length of the inserted or deleted sequence and the second number is the length of the flanking repeated sequence

Supplementary Table 2.2 DNA substrates

CAN1 Nt Position	Basepair/ Mispair	Oligo	Sequence
1193	CG	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1193	TG	3-Top 1193T	5′-AAATATTTACGTTGGTTTCCGTATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1193	AG	4-Top 1193A	5′-AAATATTTACGTTGGTTACCGTATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1193	CA	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		5-Bot 1193A	3′-TTTATAAATGCAACCAAAGGCATAAAATAAACCAGATA–5′
1193	CT	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		6-Bot 1193T	3′-TTTATAAATGCAACCAATGGCATAAAATAAACCAGATA–5′
1193	GG	11-Top 1193G	5′-AAATATTTACGTTGGTTGCCGTATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1196	GC	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1196	AC	7-Top 1196A	5′-AAATATTTACGTTGGTTCCCATATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1196	CC	8-Top 1196C	5′-AAATATTTACGTTGGTTCCCCTATTTTATTTGGTCTAT-3′
		2-Bot 1193G	3′-TTTATAAATGCAACCAAGGGCATAAAATAAACCAGATA-5′
1196	GT	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		9-Bot 1196T	3′-TTTATAAATGCAACCAAGGGTATAAAATAAACCAGATA–5′
1196	GA	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		10-Bot 1196A	3′-TTTATAAATGCAACCAAGGGAATAAAATAAACCAGATA–5′
1196	TT	12-Top 1196T	5′-AAATATTTACGTTGGTTCCCTTATTTTATTTGGTCTAT-3′
		9-Bot 1196T	3′-TTTATAAATGCAACCAAGGGTATAAAATAAACCAGATA–5′
1196	GG	1-Top 1193C	5′-AAATATTTACGTTGGTTCCCGTATTTTATTTGGTCTAT-3′
		13-Bot 1196G	3′-ATAGACCAAATAAAATAGGGGAACCAACGTAAATATTT-5′
1196	AA	7-Top 1196A	5′-AAATATTTACGTTGGTTCCCATATTTTATTTGGTCTAT-3′
		10-Bot 1196A	3′-TTTATAAATGCAACCAAGGGAATAAAATAAACCAGATA–5′

Supplementary Table 2.2 (continued) DNA substrates

1196	CT	8-Top 1196C	5′-AAATATTTACGTTGGTTCCC**C**TATTTTATTTGGTCTAT-3′
		9-Bot 1196T	3′-TTTATAAATGCAACCAAGGG**T**ATAAAATAAACCAGATA–5′
106	GT	13-Top 106G	5′-GTGGTACTATTGGTACAG**G**TCTTTTCATTGGTTTATC-3′
		14-Bot 106T	3′-CACCATGATAACCATGTC**T**AGAAAAGTAACCAAATAG-5′
106	AC	15-Top 106A	5′-GTGGTACTATTGGTACAG**A**TCTTTTCATTGGTTTATC-3′
		16-Bot 106C	3′-CACCATGATAACCATGTC**C**AGAAAAGTAACCAAATAG-5′
139	AG	17-Top 139A	5′-TGGCATATTCTGTCACGC**A**GTCCTTGGGTGAAATGGC-3′
		18-Bot 139G	3′-ACCGTATAAGACAGTGCG**G**CAGGAACCCACTTTACCG-5′
139	CT	19-Top 139C	5′-TGGCATATTCTGTCACGC**C**GTCCTTGGGTGAAATGGC-3′
		20-Bot 139T	3′-ACCGTATAAGACAGTGCG**T**CAGGAACCCACTTTACCG-5′
232	GG	21-Top 232G	5′-GAATTCGAGTTCTGGGTC**G**CTTCCATCAAAGTTTTAG-3′
		22-Bot 232G	3′-CTTAAGCTCAAGACCCAG**G**GAAGGTAGTTTCAAAATC-5′
232	CC	23-Top 232C	5′-GAATTCGAGTTCTGGGTC**C**CTTCCATCAAAGTTTTAG-3′
		24-Bot 232C	3′-CTTAAGCTCAAGACCCAG**C**GAAGGTAGTTTCAAAATC-5′
270	GT	25-Top 270G	5′-GAGAAACCCAGGTGCCTG**G**GGTCCAGGTATAATATCT-3′
		26-Bot 270T	3′-CTCTTTGGGTCCACGGAC**T**CCAGGTCCATATTATAGA-5′
270	AC	27-Top 270A	5′-GAGAAACCCAGGTGCCTG**A**GGTCCAGGTATAATATCT-3′
		28-Bot 270C	3′-CTCTTTGGGTCCACGGAC**C**CCAGGTCCATATTATAGA-5′
320	AA	29-Top 320A	5′-CCCAGAAAATCCGTTCC**A**AGAGCCATCAAAAAAGTTG-3′
		30-Bot 320A	3′-GGGTCTTTTAGGCAAGGT**A**CTCGGTAGTTTTTTCAAC-5′
320	TT	31-Top 320T	5′-CCCAGAAAATCCGTTCCAT**G**AGCCATCAAAAAAGTTG-3′
		32-Bot 320T	3′-GGGTCTTTTAGGCAAGGT**T**CTCGGTAGTTTTTTCAAC-5′
544	TT	33-Top 544T	5′-TCTTAGCTGTTTGGATCTTATTTCAATGCATATTCAG-3′
		34-Bot 544T	3′-AGAATCGACAAACCTAGA**T**TAAAGTTACGTATAAGTC-5′
544	AA	35-Top 544A	5′-TCTTAGCTGTTTGGATCT**A**ATTTCAATGCATATTCAG-3′
		36-Bot 544A	3′-AGAATCGACAAACCTAGA**A**TAAAGTTACGTATAAGTC-5′
+6	+6	Top+ATGCTA 43mer	**-ATGCTA-** 5′-ATGTGAATCAGTATG GTTCCTATCTGCTGAAGGAAAT-3′
		Bot-biot 37mer	3′-Bio-TACACTTAGTCATAC CAAGGATAGACGACTTCCTTTA-5′

Supplementary Figure 2.1 *CAN1* mutation spectra wild-type versus *msh3*

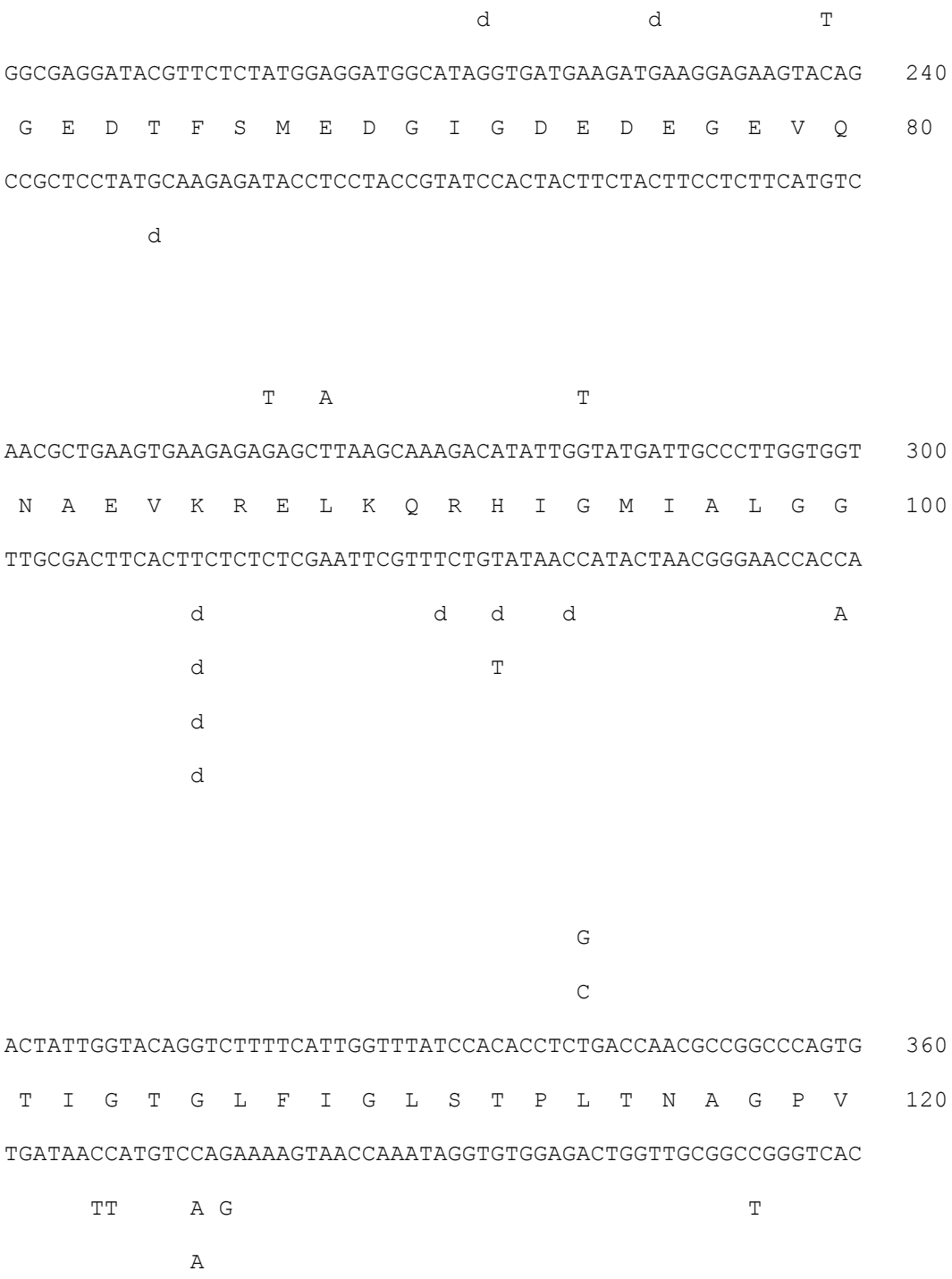
Top= RDKY3686 wild-type

Bottom= RDKY4149 *msh3*

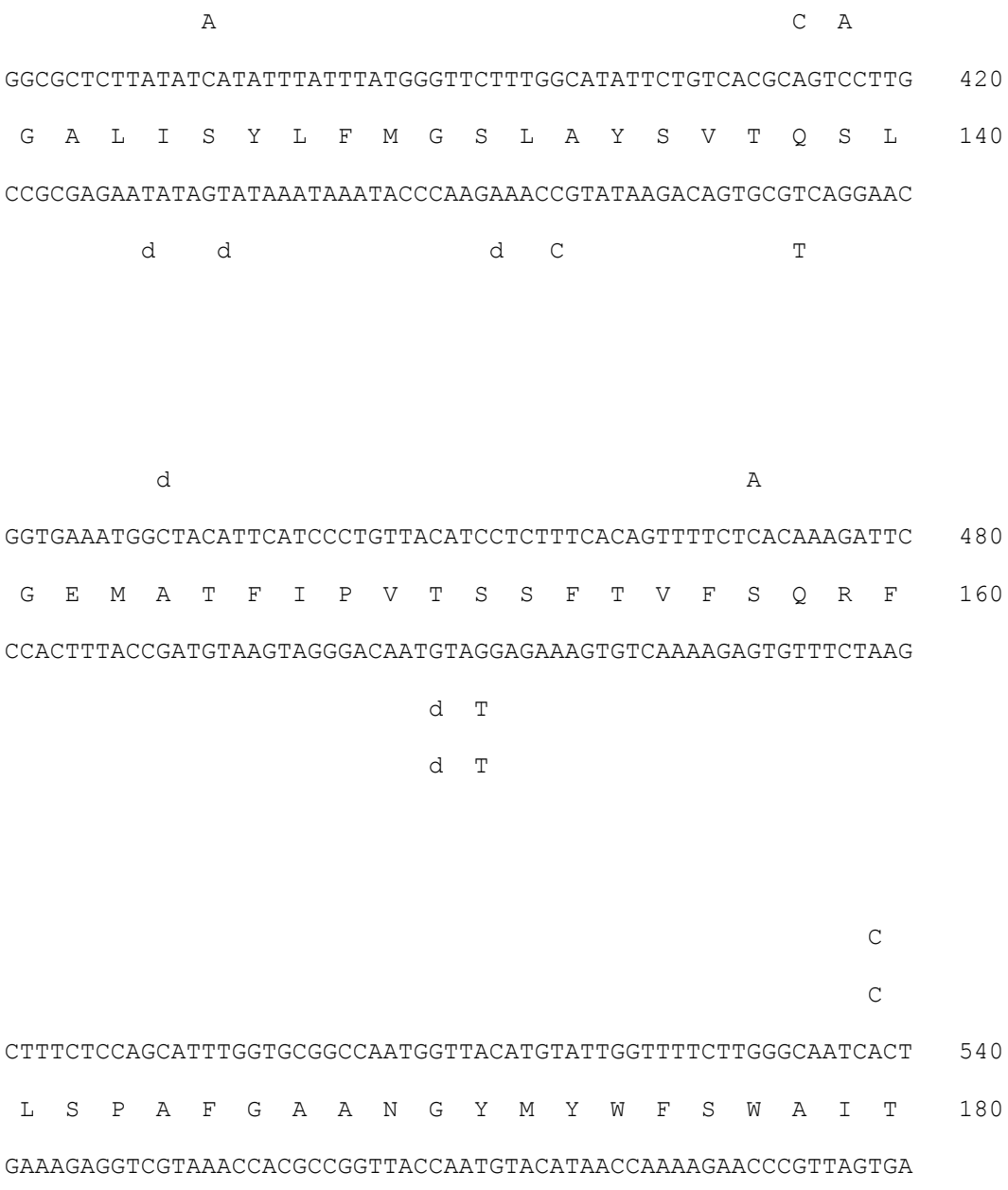
Base substitutions shown as a different nucleotide at the corresponding position. Frameshift deletions of -1 or -2 shown as d. Frameshift insertions of +1 or +2 as i.

```
ATGACAAATTCAAAAGAAGACGCCGACATAGAGGAGAAGCATATGTACAATGAGCCGGTC   60
 M  T  N  S  K  E  D  A  D  I  E  E  K  H  M  Y  N  E  P  V    20
TACTGTTTAAGTTTTCTTCTGCGGCTGTATCTCCTCTTCGTATACATGTTACTCGGCCAG
          d
          d

                                                      G
ACAACCCTCTTTCACGACGTTGAAGCTTCACAAACACACCACAGACGTGGGTCAATACCA   120
 T  T  L  F  H  D  V  E  A  S  Q  T  H  H  R  R  G  S  I  P    40
TGTTGGGAGAAAGTGCTGCAACTTCGAAGTGTTTGTGTGGTGTCTGCACCCAGTTATGGT

TTGAAAGATGAGAAAAGTAAAGAATTGTATCCATTGCGCTCTTTCCCGACGAGAGTAAAT   180
 L  K  D  E  K  S  K  E  L  Y  P  L  R  S  F  P  T  R  V  N    60
AACTTTCTACTCTTTTCATTTCTTAACATAGGTAACGCGAGAAAGGGCTGCTCTCATTTA
                                        d
```

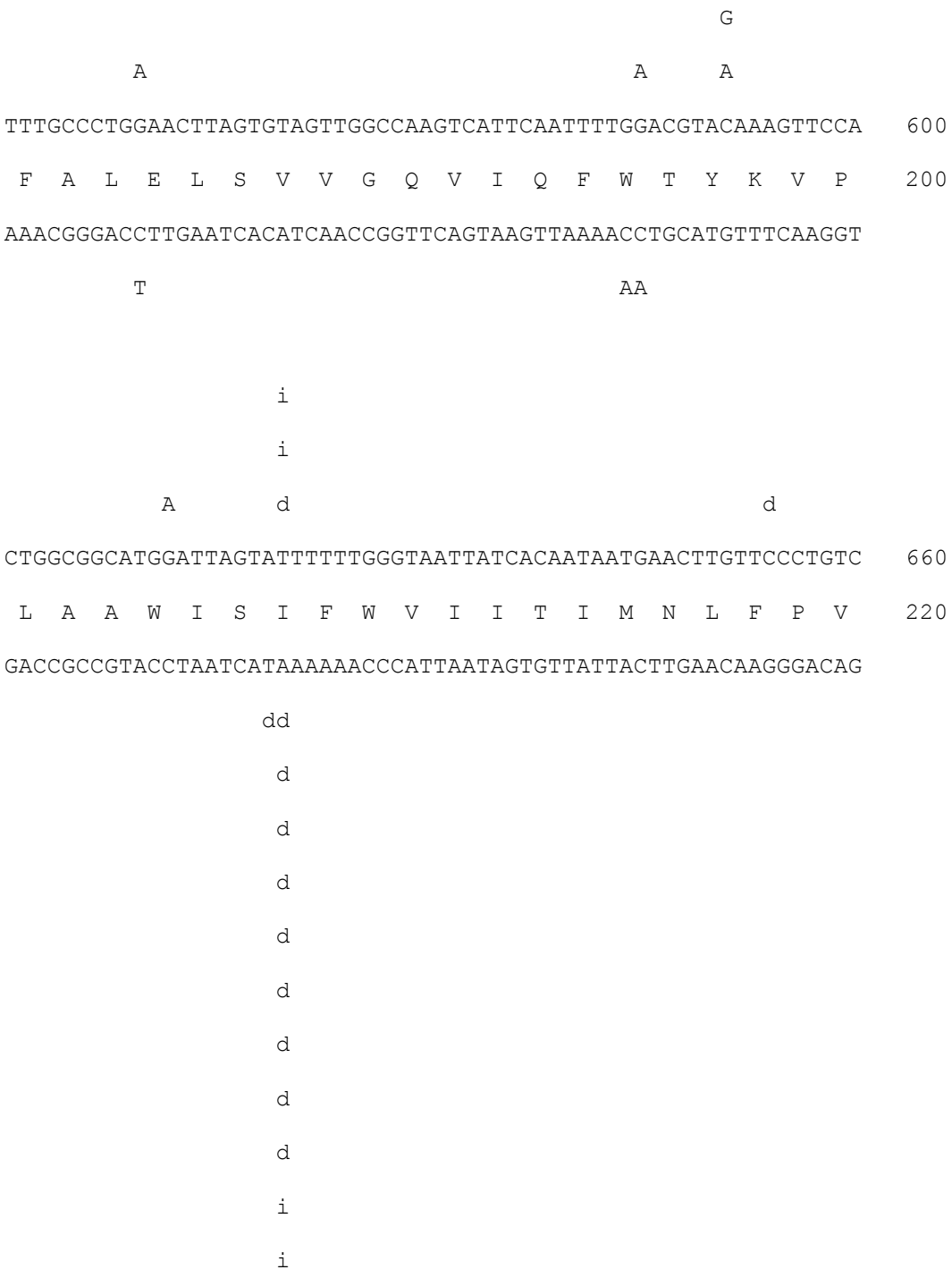
Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



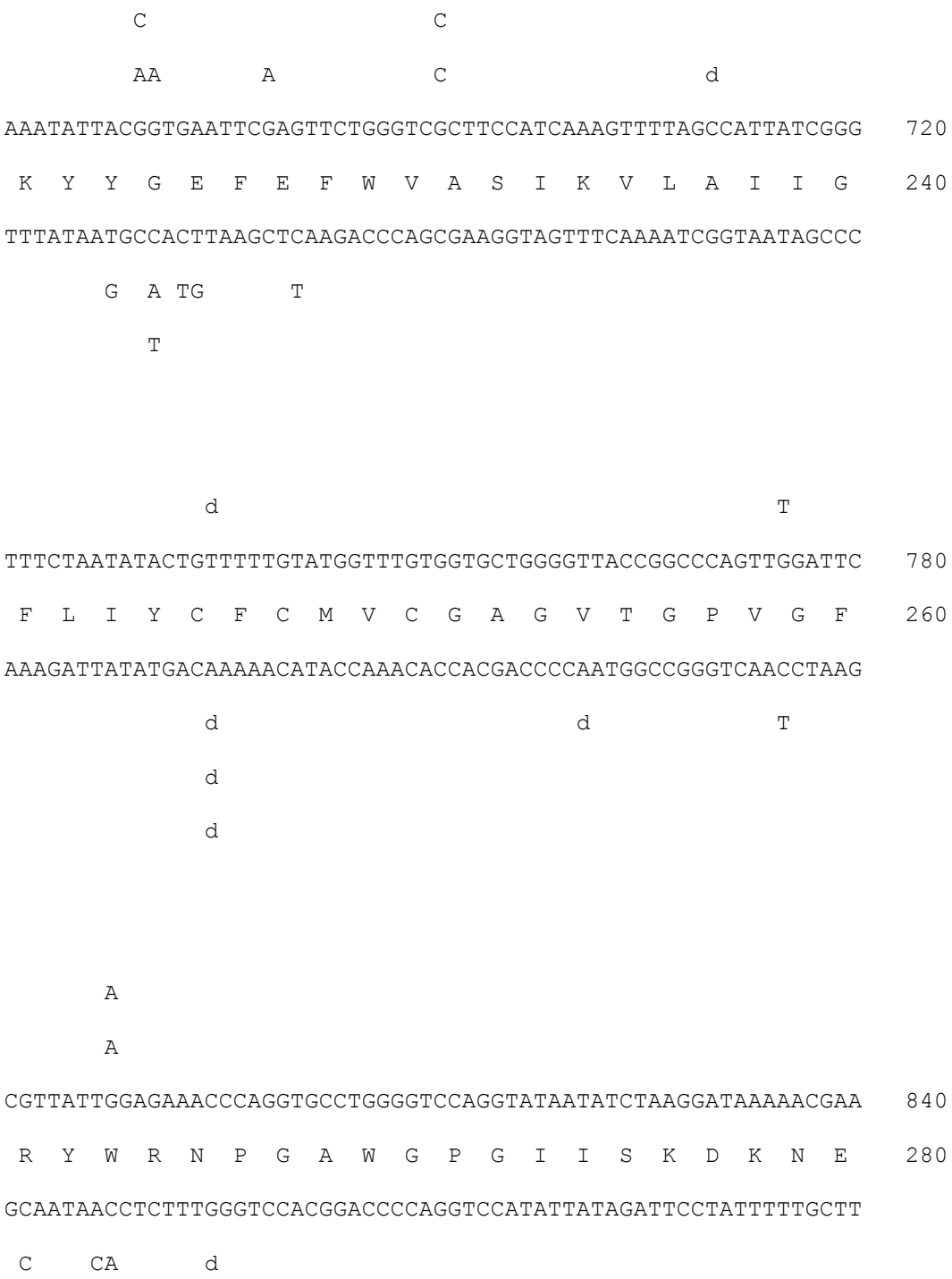
Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



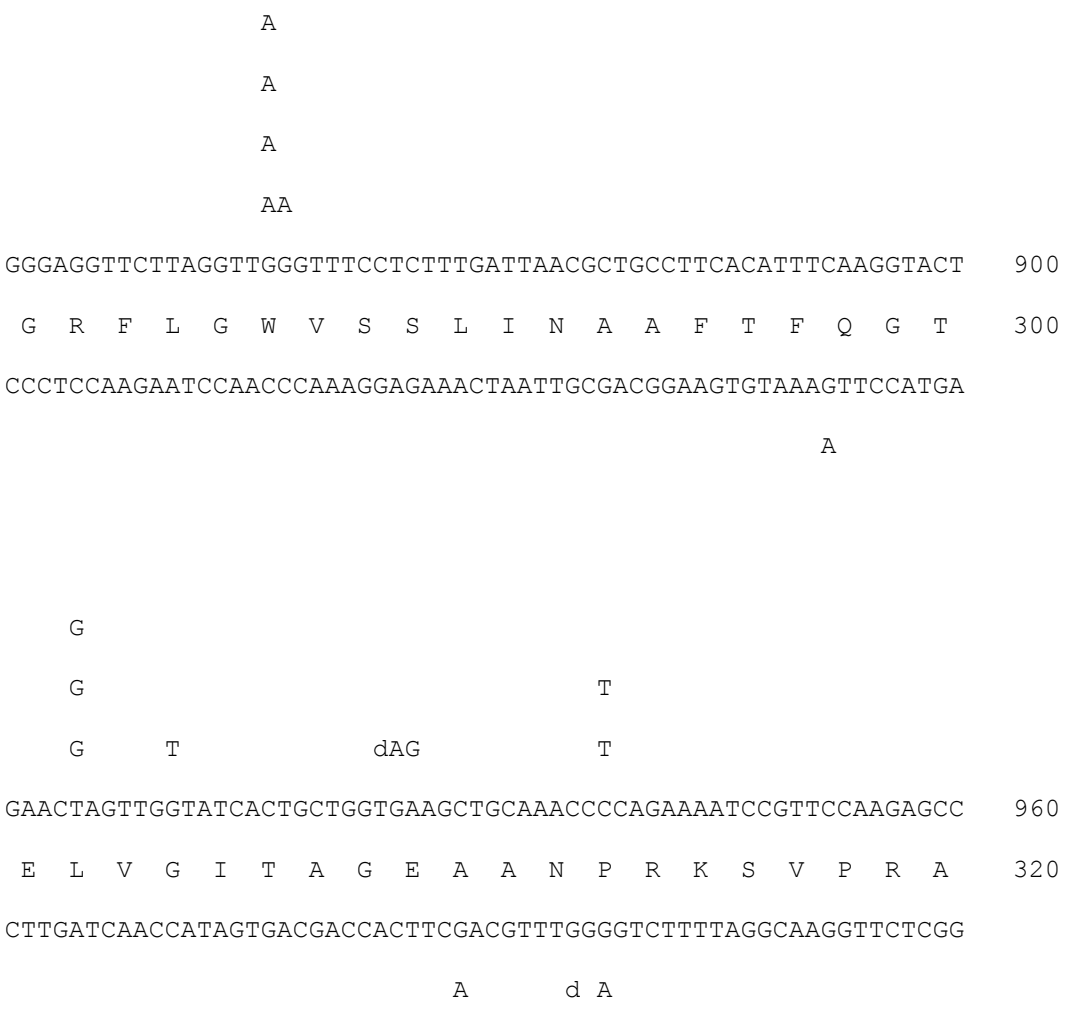
Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



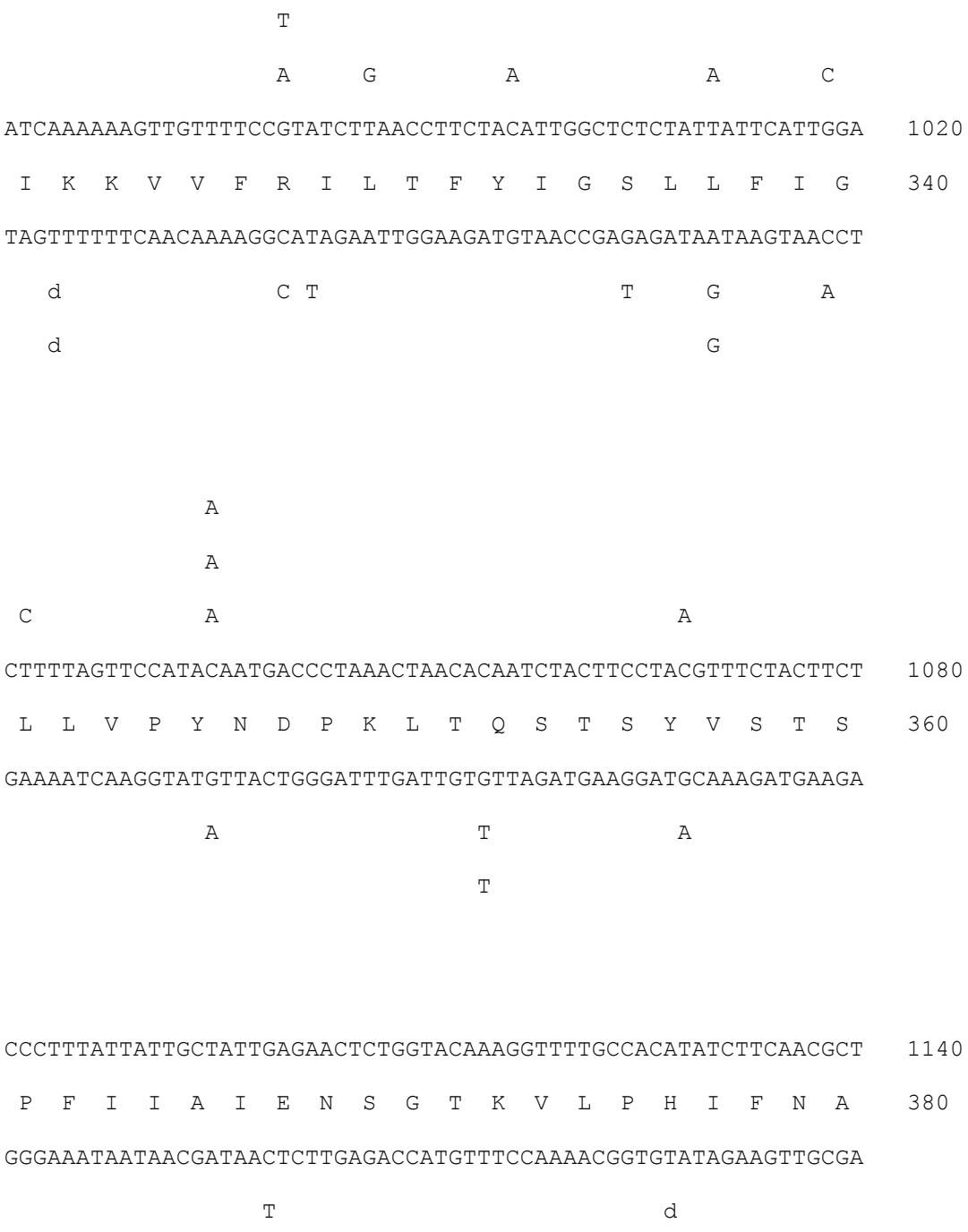
Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



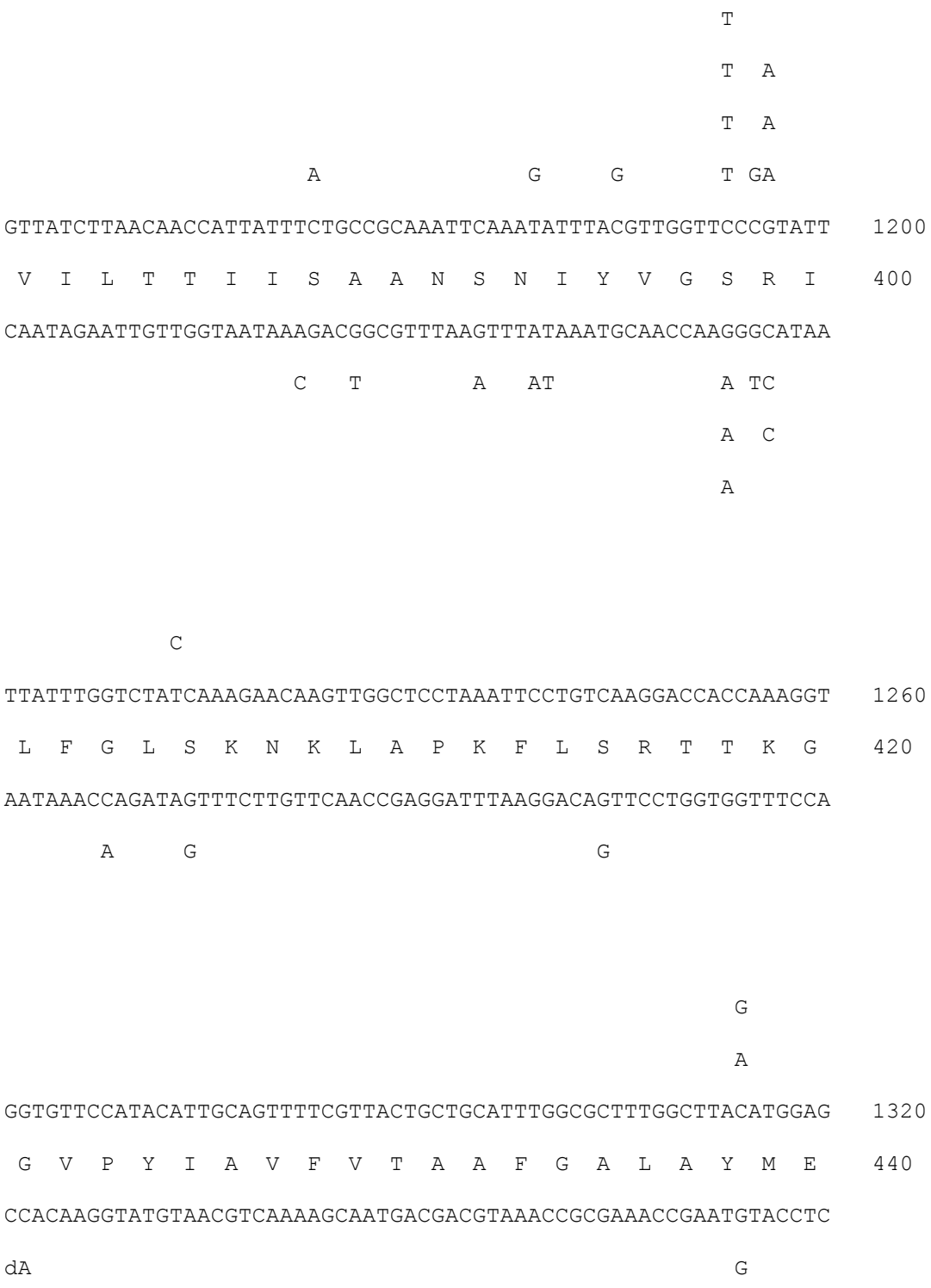
Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



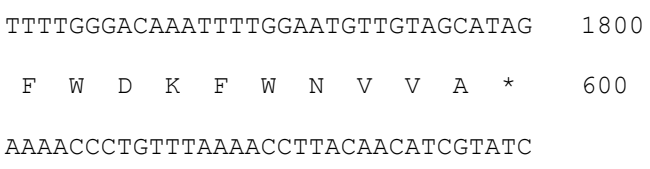
Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*

```
ACATCTACTGGTGGTGACAAAGTTTTCGAATGGCTATTAAATATCACTGGTGTTGCAGGC   1380
 T  S  T  G  G  D  K  V  F  E  W  L  L  N  I  T  G  V  A  G    460
TGTAGATGACCACCACTGTTTCAAAAGCTTACCGATAATTTATAGTGACCACAACGTCCG

d        A
TTTTTTGCATGGTTATTTATCTCAATCTCGCACATCAGATTTATGCAAGCTTTGAAATAC   1440
 F  F  A  W  L  F  I  S  I  S  H  I  R  F  M  Q  A  L  K  Y    480
AAAAAACGTACCAATAAATAGAGTTAGAGCGTGTAGTCTAAATACGTTCGAAACTTTATG
d                    G           T                    G
d

                                          A
CGTGGCATCTCTCGTGACGAGTTACCATTTAAAGCTAAATTAATGCCCGGCTTGGCTTAT   1500
 R  G  I  S  R  D  E  L  P  F  K  A  K  L  M  P  G  L  A  Y    500
GCACCGTAGAGAGCACTGCTCAATGGTAAATTTCGATTTAATTACGGGCCGAACCGAATA
                                       G   d           A
                                           d
```

Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*

```
TATGCGGCCACATTTATGACGATCATTATCATTATTCAAGGTTTCACGGCTTTTGCACCA   1560
 Y  A  A  T  F  M  T  I  I  I  I  I  Q  G  F  T  A  F  A  P    520
ATACGCCGGTGTAAATACTGCTAGTAATAGTAATAAGTTCCAAAGTGCCGAAAACGTGGT
                               A  d                    i


AAATTCAATGGTGTTAGCTTTGCTGCCGCCTATATCTCTATTTTCCTGTTCTTAGCTGTT   1620
 K  F  N  G  V  S  F  A  A  A  Y  I  S  I  F  L  F  L  A  V    540
TTTAAGTTACCACAATCGAAACGACGGCGGATATAGAGATAAAAGGACAAGAATCGACAA
d


   d  A
TGGATCTTATTTCAATGCATATTCAGATGCAGATTTATTTGGAAGATTGGAGATGTCGAC   1680
 W  I  L  F  Q  C  I  F  R  C  R  F  I  W  K  I  G  D  V  D    560
ACCTAGAATAAAGTTACGTATAAGTCTACGTCTAAATAAACCTTCTAACCTCTACAGCTG


ATCGATTCCGATAGAAGAGACATTGAGGCAATTGTATGGGAAGATCATGAACCAAAGACT   1740
 I  D  S  D  R  R  D  I  E  A  I  V  W  E  D  H  E  P  K  T    580
TAGCTAAGGCTATCTTCTCTGTAACTCCGTTAACATACCCTTCTAGTACTTGGTTTCTGA
```

Supplementary Figure 2.1 (continued) *CAN1* mutation spectra wild-type versus *msh3*



Supplementary Figure 2.2 *CAN1* mutation spectra wild-type versus *msh6*

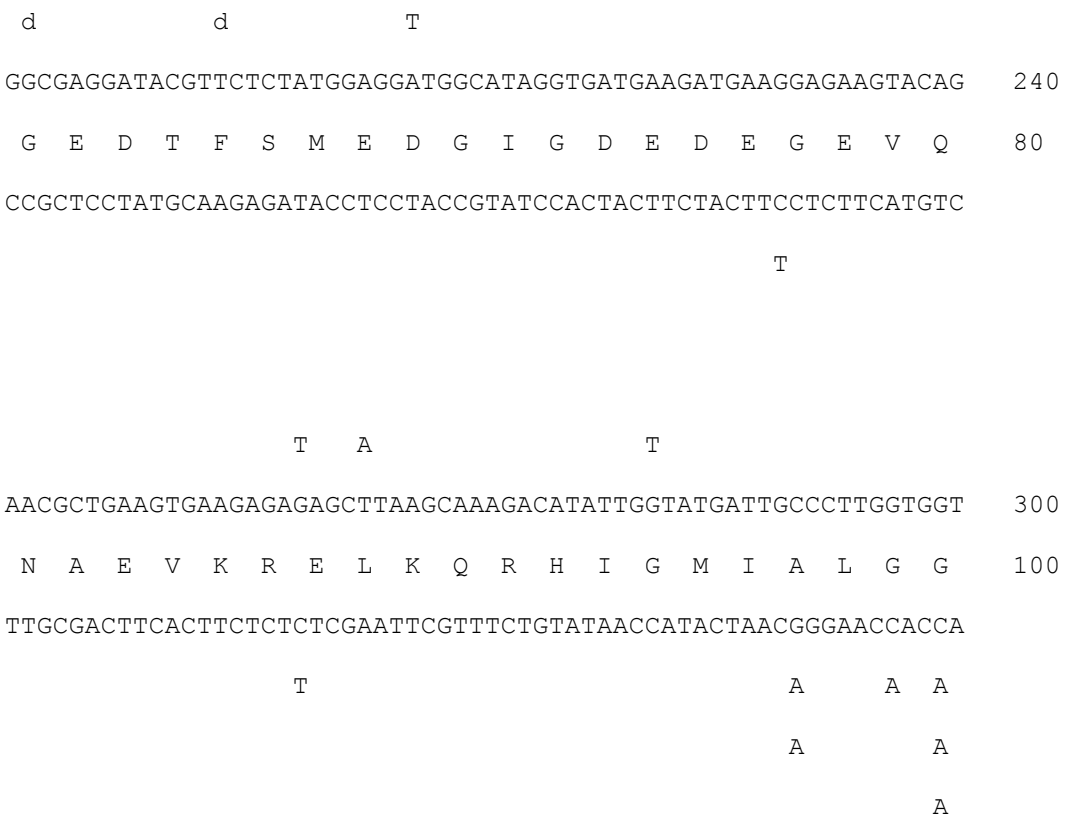
Top= RDKY3686 wild-type

Bottom= RDKY4151 *msh6*

Base substitutions shown as a different nucleotide at the corresponding position. Frameshift deletions of -1 or -2 shown as d. Frameshift insertions of +1 or +2 as i.

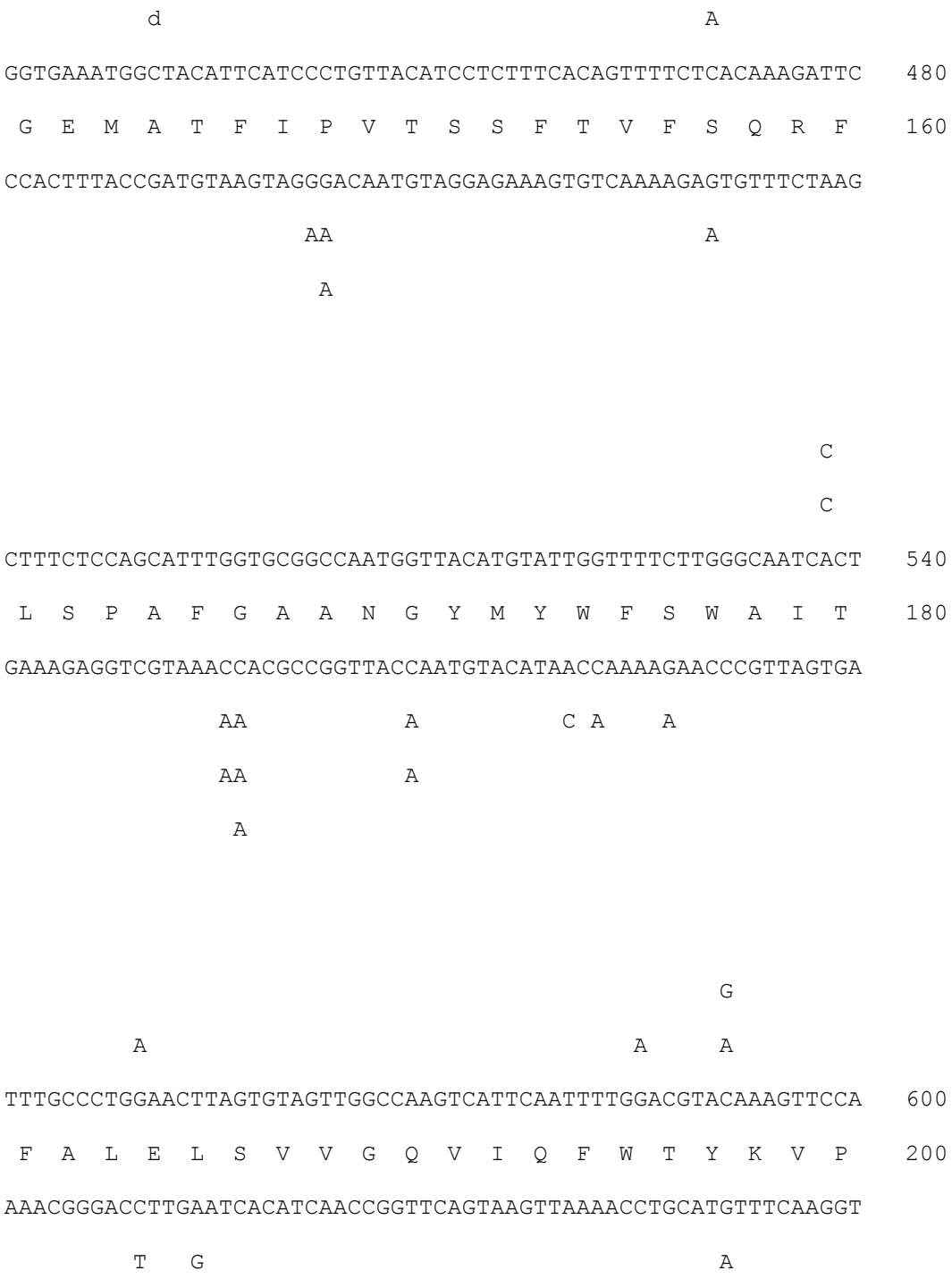
```
ATGACAAATTCAAAAGAAGACGCCGACATAGAGGAGAAGCATATGTACAATGAGCCGGTC   60
 M  T  N  S  K  E  D  A  D  I  E  E  K  H  M  Y  N  E  P  V    20
TACTGTTTAAGTTTTCTTCTGCGGCTGTATCTCCTCTTCGTATACATGTTACTCGGCCAG
                               T

                                                      G
ACAACCCTCTTTCACGACGTTGAAGCTTCACAAACACACCACAGACGTGGGTCAATACCA   120
 T  T  L  F  H  D  V  E  A  S  Q  T  H  H  R  R  G  S  I  P    40
TGTTGGGAGAAAGTGCTGCAACTTCGAAGTGTTTGTGTGGTGTCTGCACCCAGTTATGGT
                               T
                               T

TTGAAAGATGAGAAAAGTAAAGAATTGTATCCATTGCGCTCTTTCCCGACGAGAGTAAAT   180
 L  K  D  E  K  S  K  E  L  Y  P  L  R  S  F  P  T  R  V  N    60
AACTTTCTACTCTTTTCATTTCTTAACATAGGTAACGCGAGAAAGGGCTGCTCTCATTTA
                     T
```

Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*

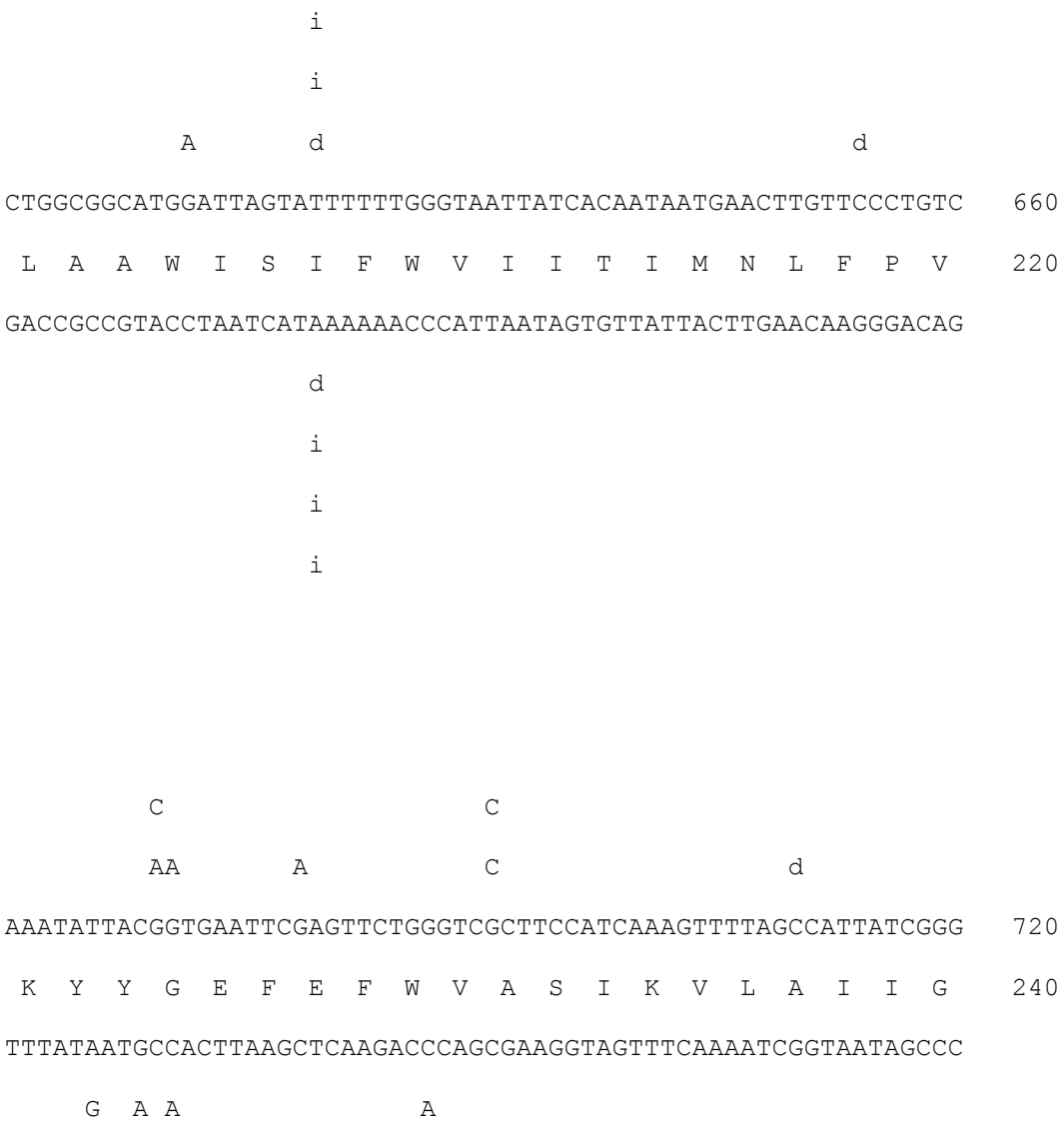


Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*

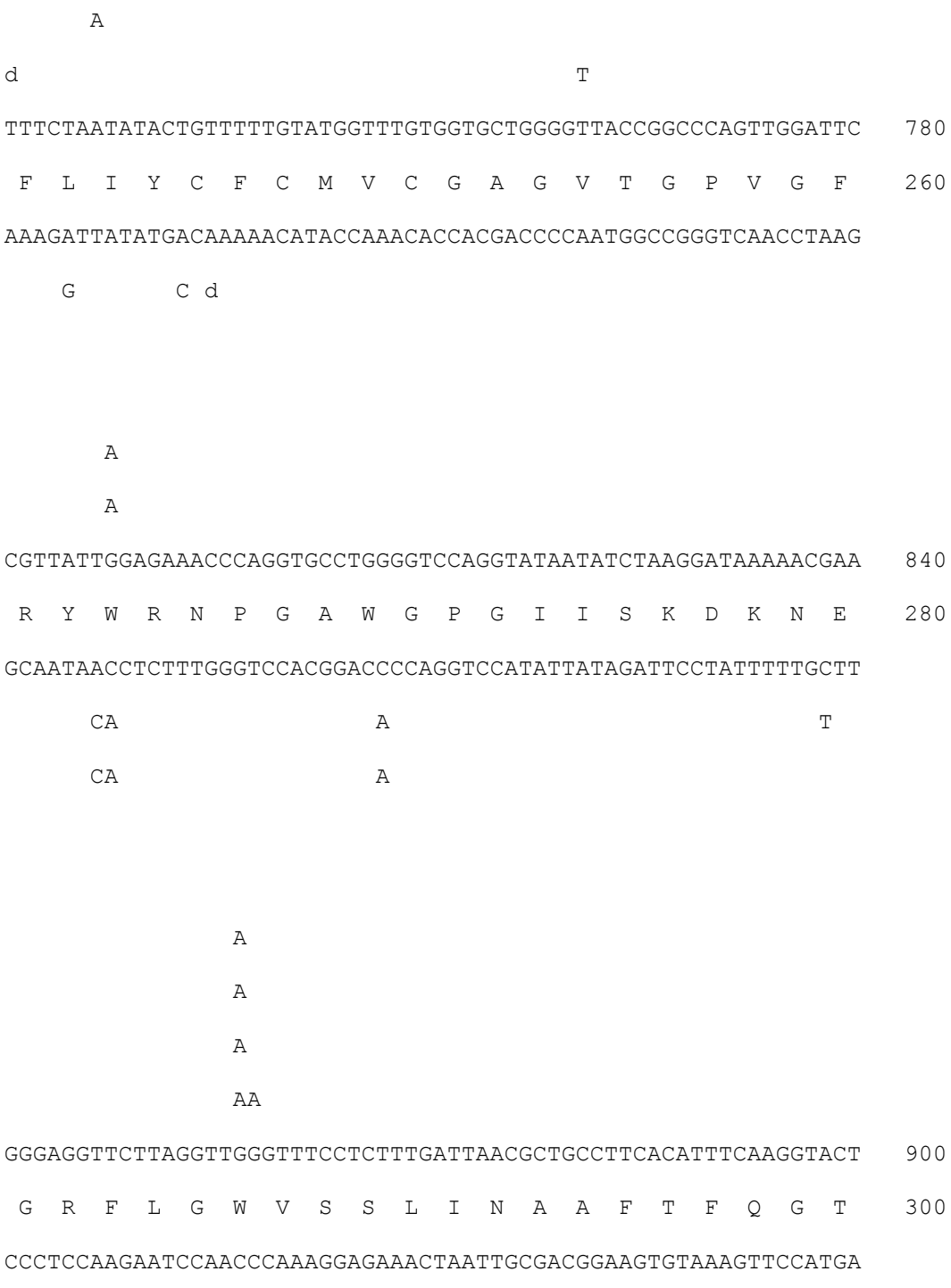
Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



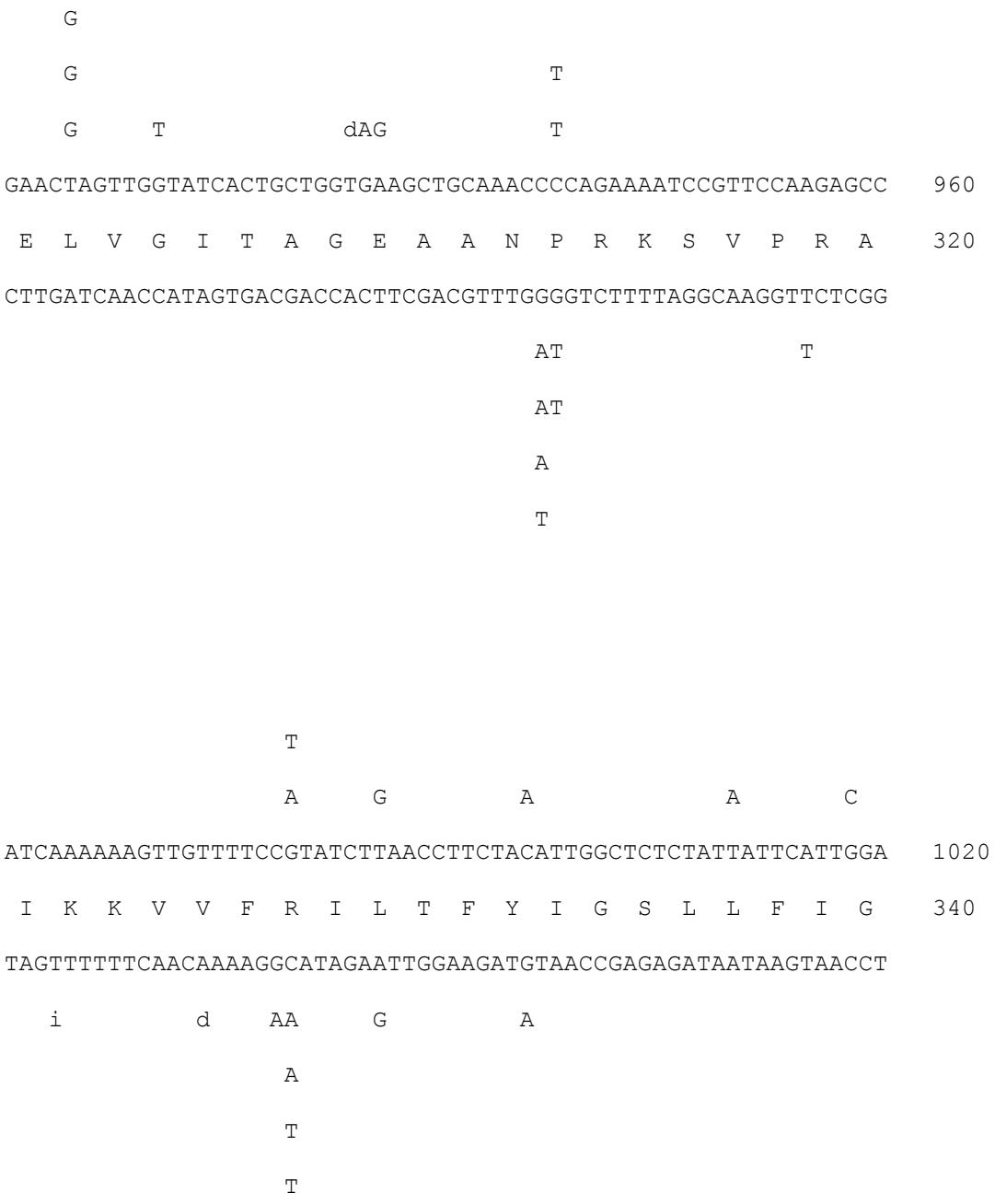
Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



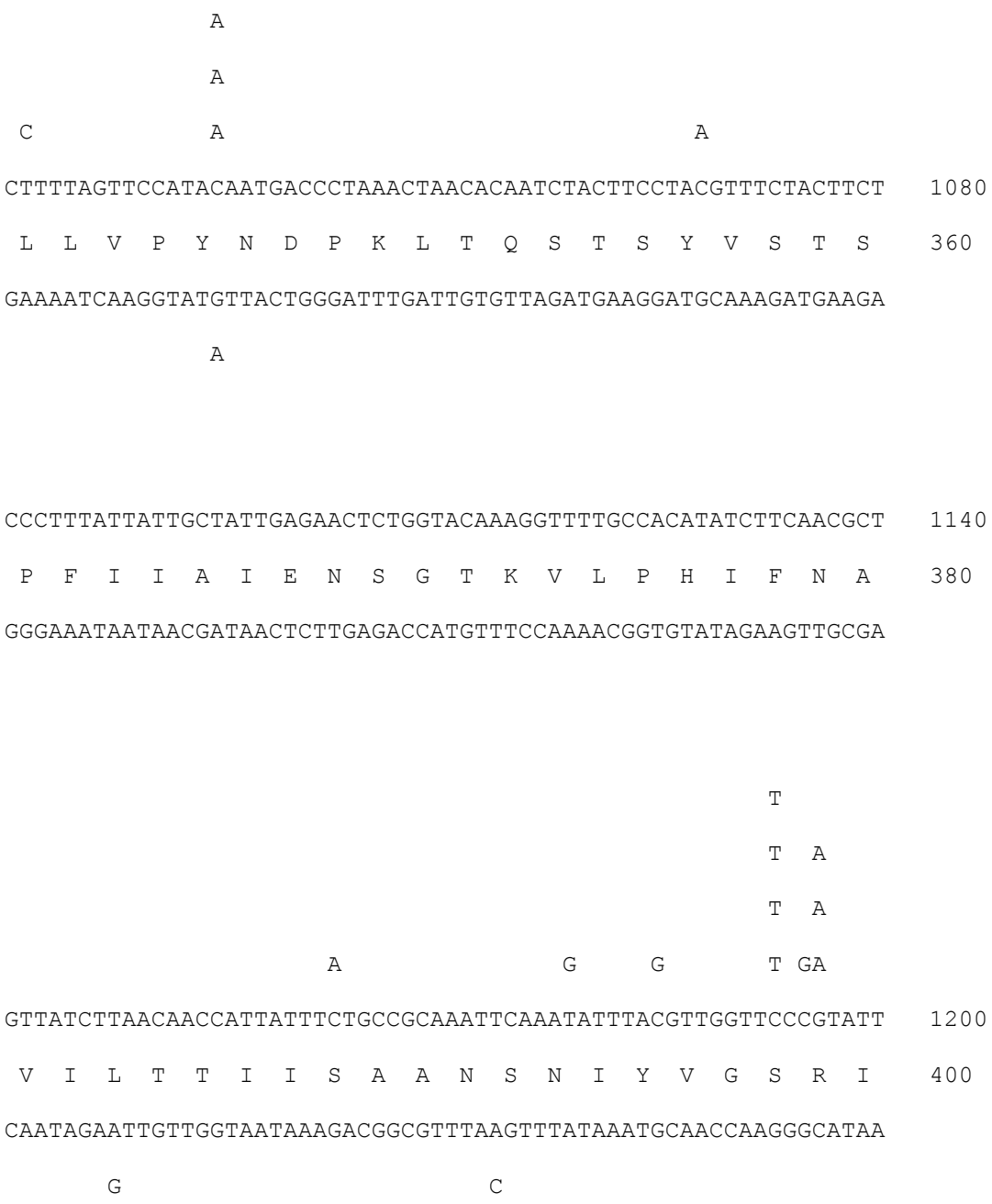
Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



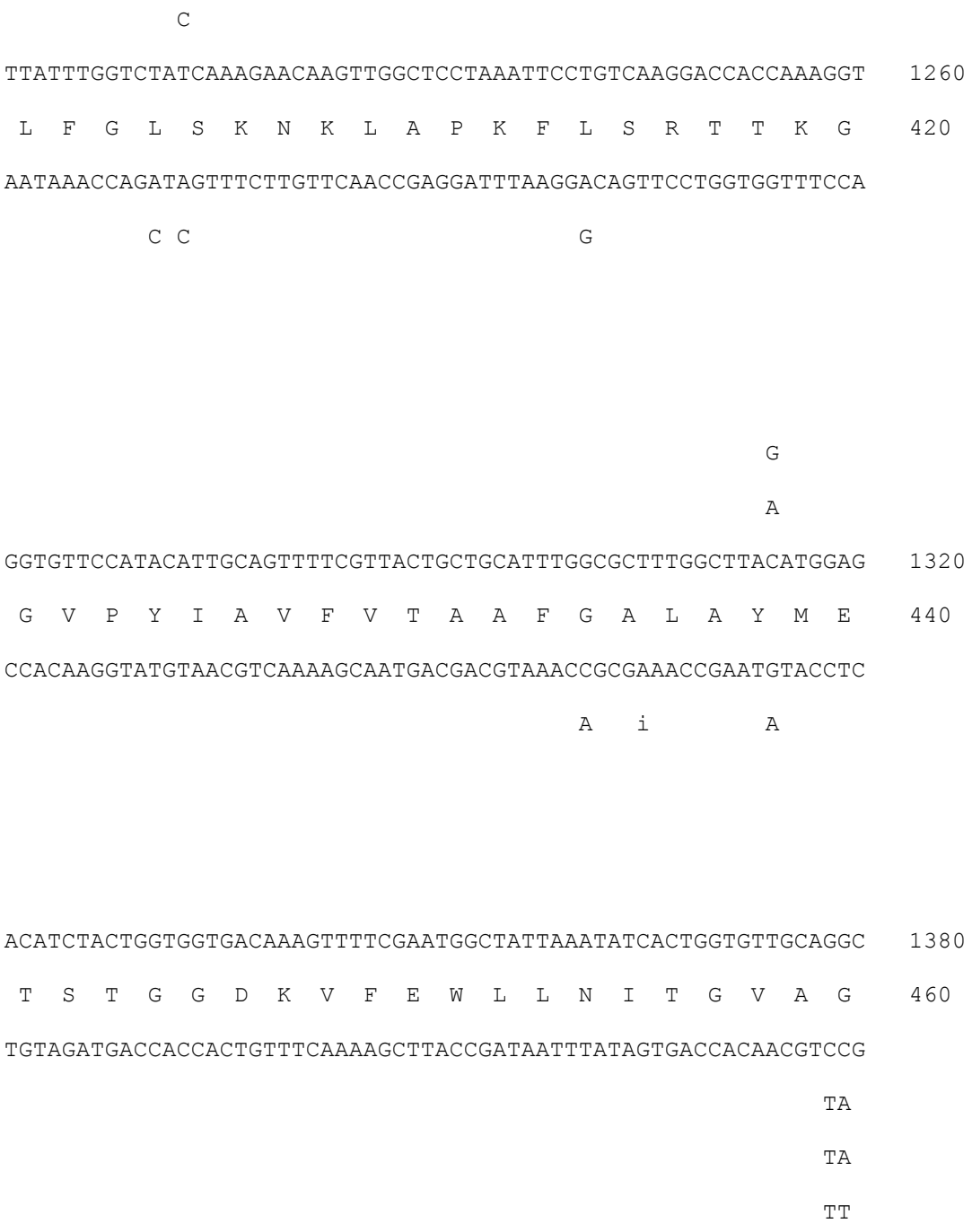
Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



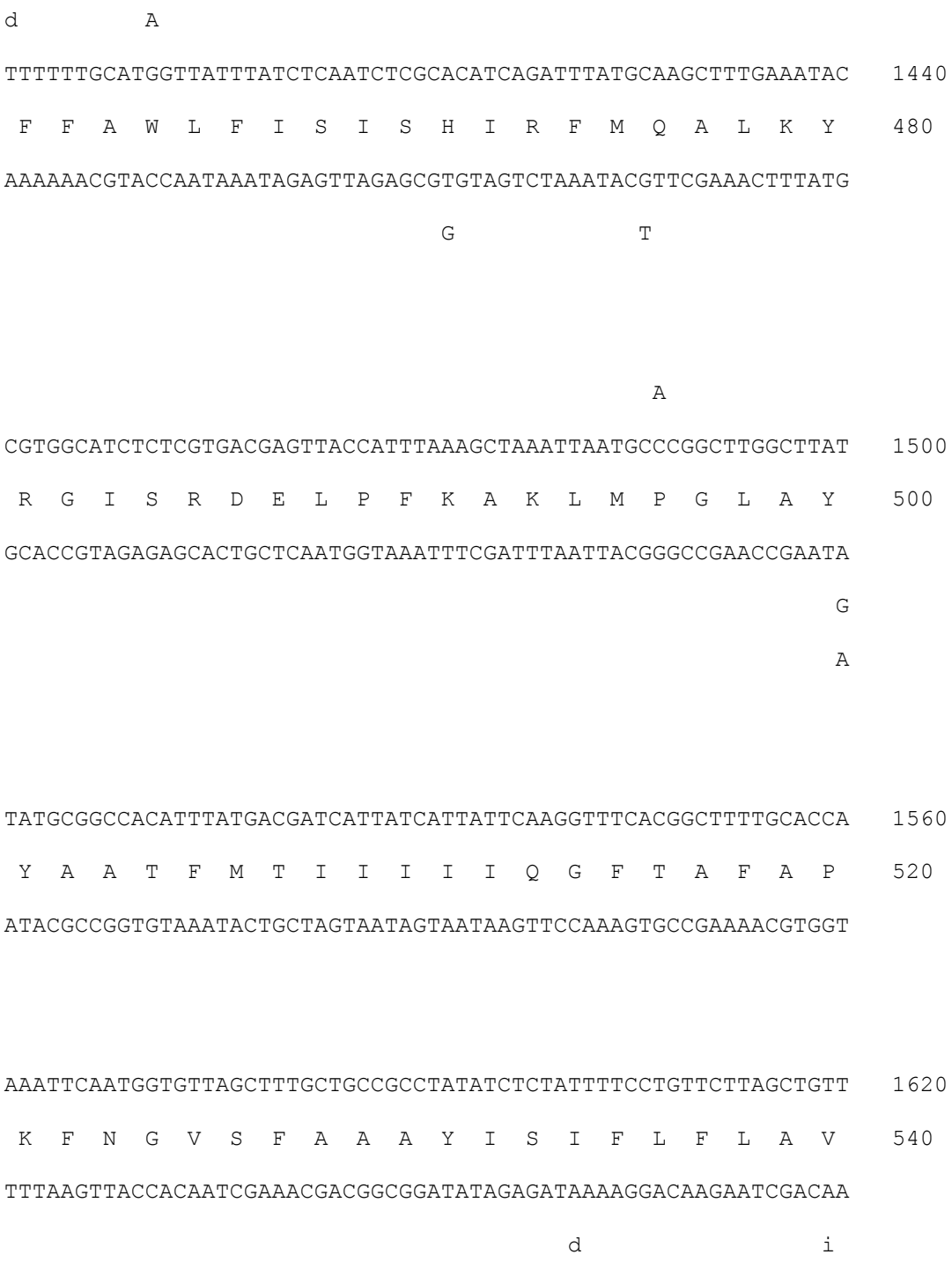
Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*



Supplementary Figure 2.2 (continued) *CAN1* mutation spectra wild-type versus *msh6*

Supplementary Figure 2.3 *CAN1* mutation spectra wild-type versus *mlh3*

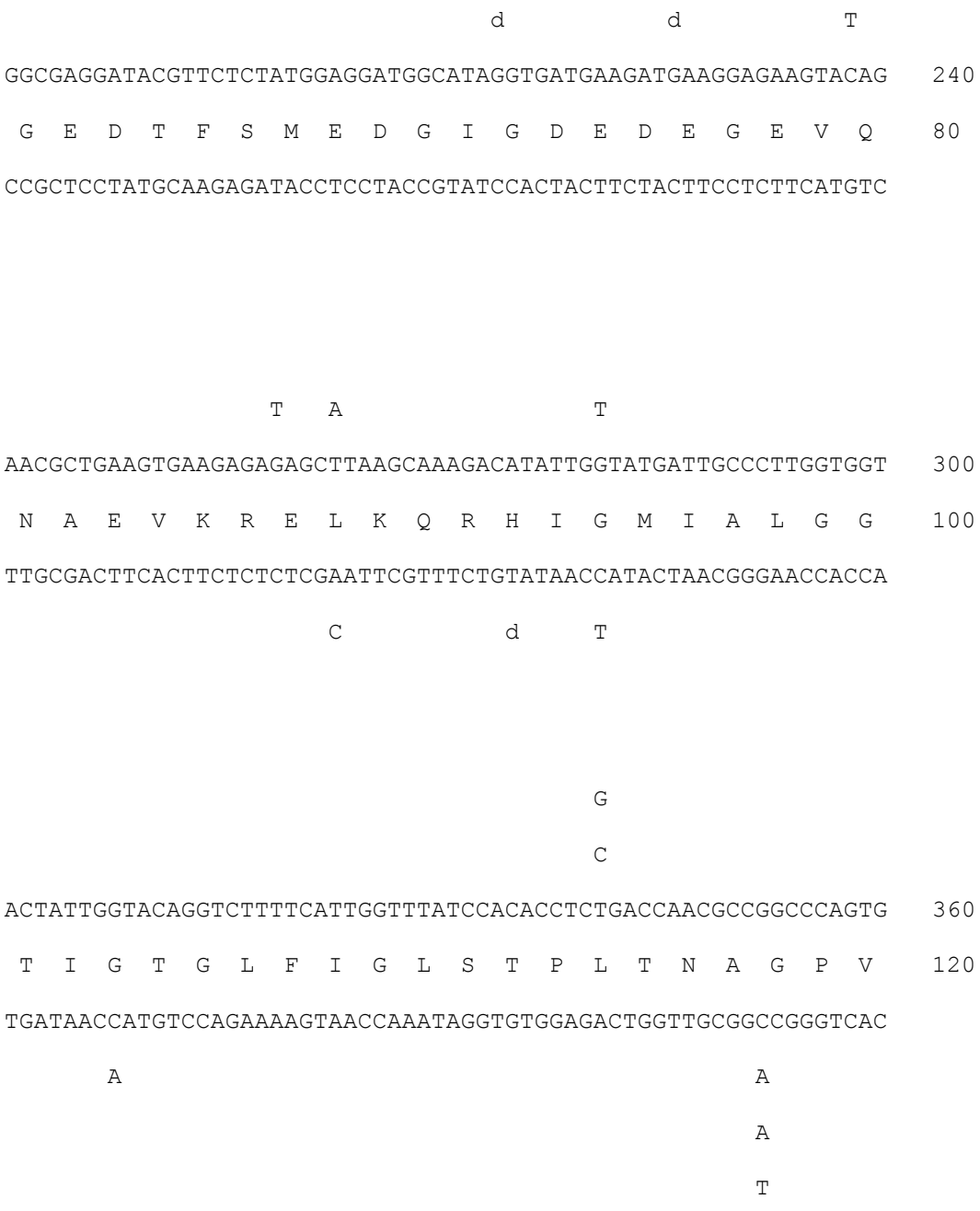
Top= RDKY3686 wild-type

Bottom= RDKY5295 *mlh3*

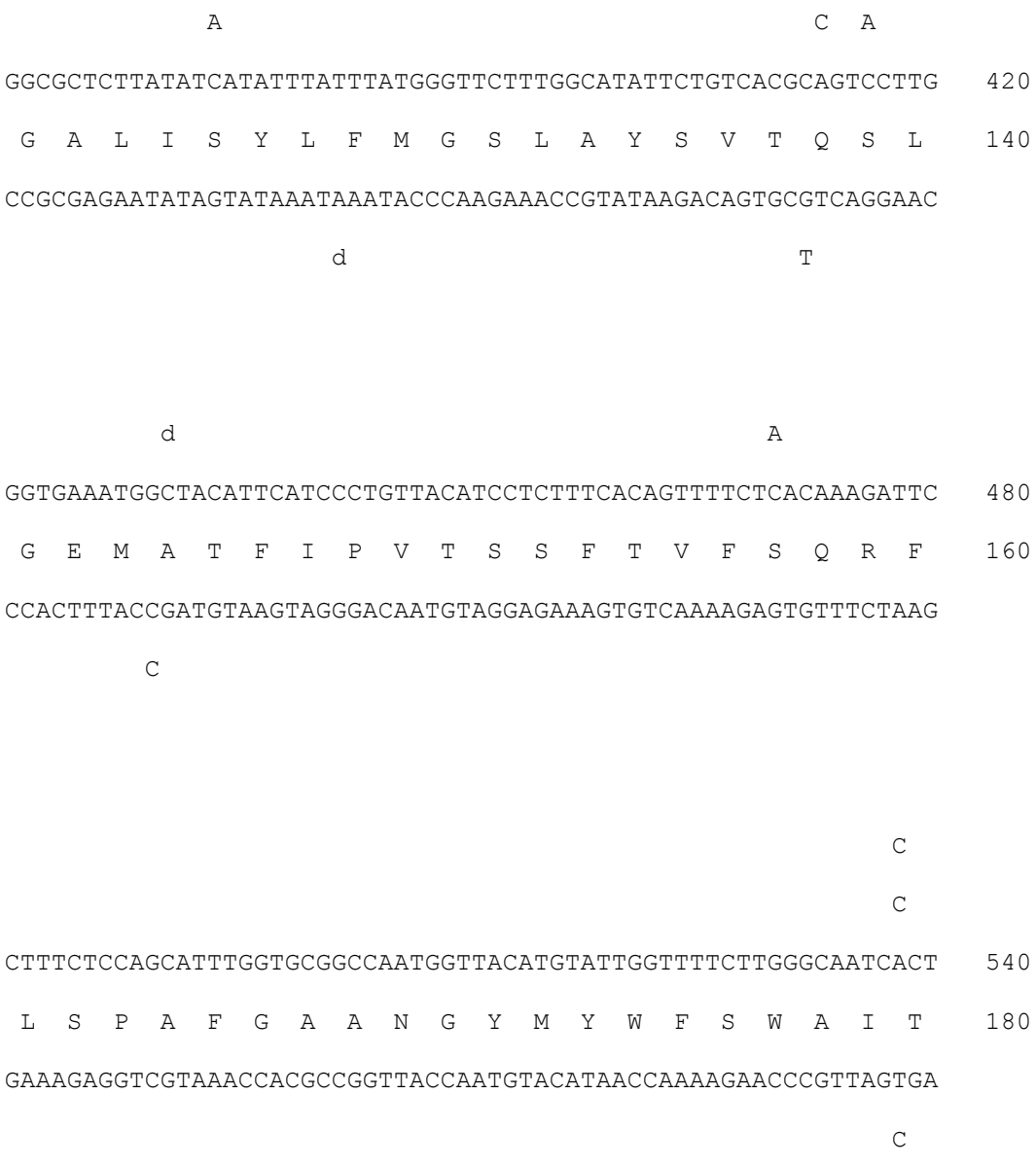
Base substitutions shown as a different nucleotide at the corresponding position. Frameshift deletions of -1 or -2 shown as d. Frameshift insertions of +1 or +2 as i.

```
ATGACAAATTCAAAAGAAGACGCCGACATAGAGGAGAAGCATATGTACAATGAGCCGGTC   60
 M  T  N  S  K  E  D  A  D  I  E  E  K  H  M  Y  N  E  P  V    20
TACTGTTTAAGTTTTCTTCTGCGGCTGTATCTCCTCTTCGTATACATGTTACTCGGCCAG

                                                    G
ACAACCCTCTTTCACGACGTTGAAGCTTCACAAACACACCACAGACGTGGGTCAATACCA   120
 T  T  L  F  H  D  V  E  A  S  Q  T  H  H  R  R  G  S  I  P    40
TGTTGGGAGAAAGTGCTGCAACTTCGAAGTGTTTGTGTGGTGTCTGCACCCAGTTATGGT
                                                    A

TTGAAAGATGAGAAAAGTAAAGAATTGTATCCATTGCGCTCTTTCCCGACGAGAGTAAAT   180
 L  K  D  E  K  S  K  E  L  Y  P  L  R  S  F  P  T  R  V  N    60
AACTTTCTACTCTTTTCATTTCTTAACATAGGTAACGCGAGAAAGGGCTGCTCTCATTTA
                 T                    d
```

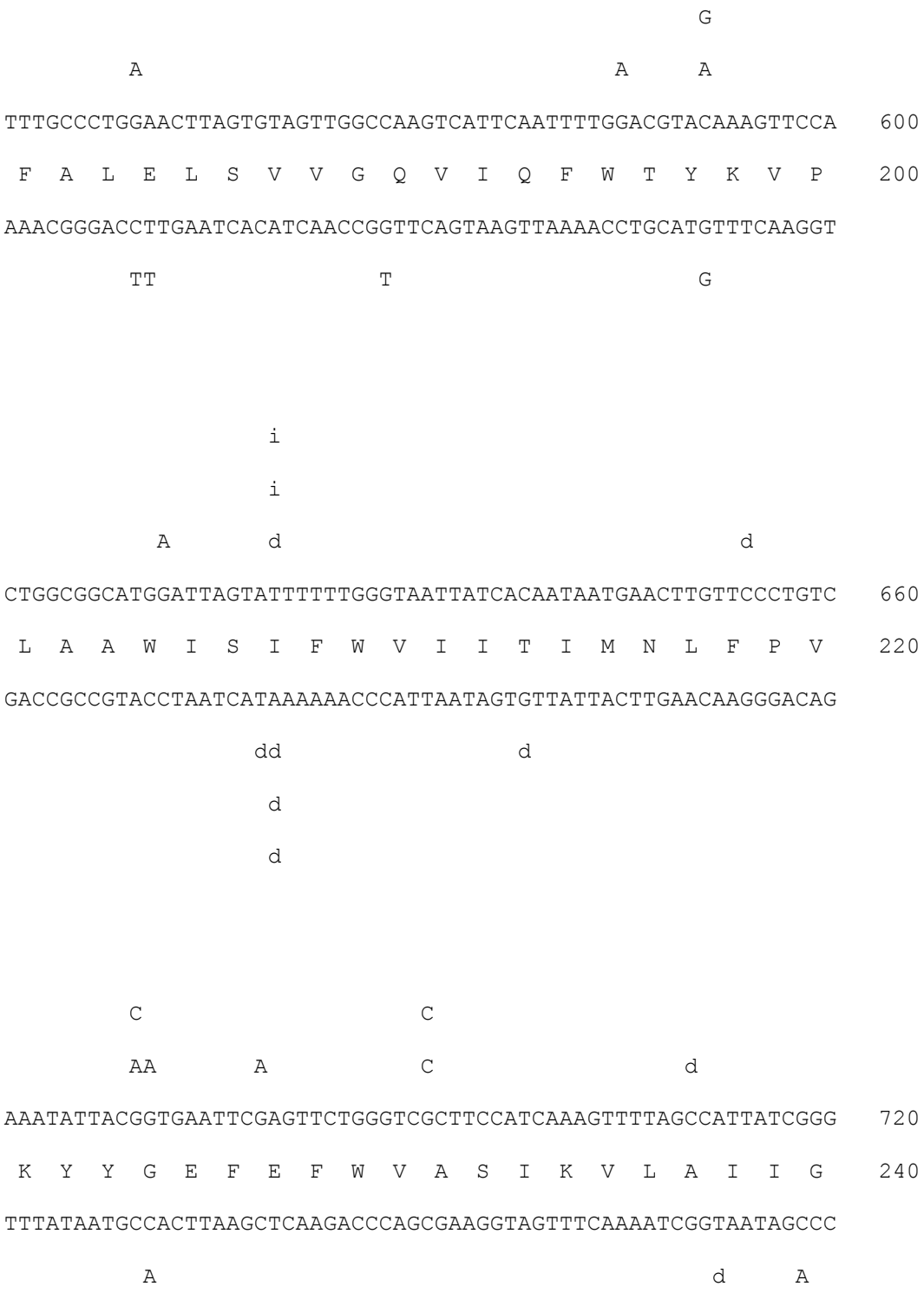
Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



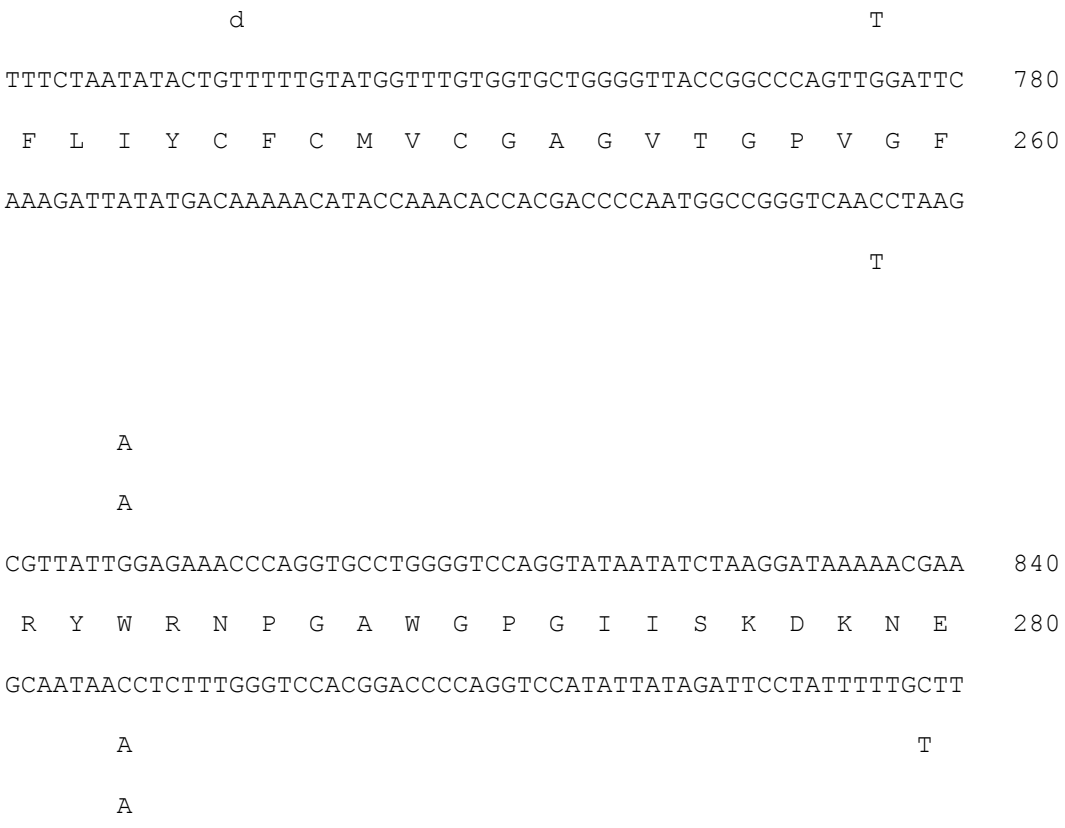
Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



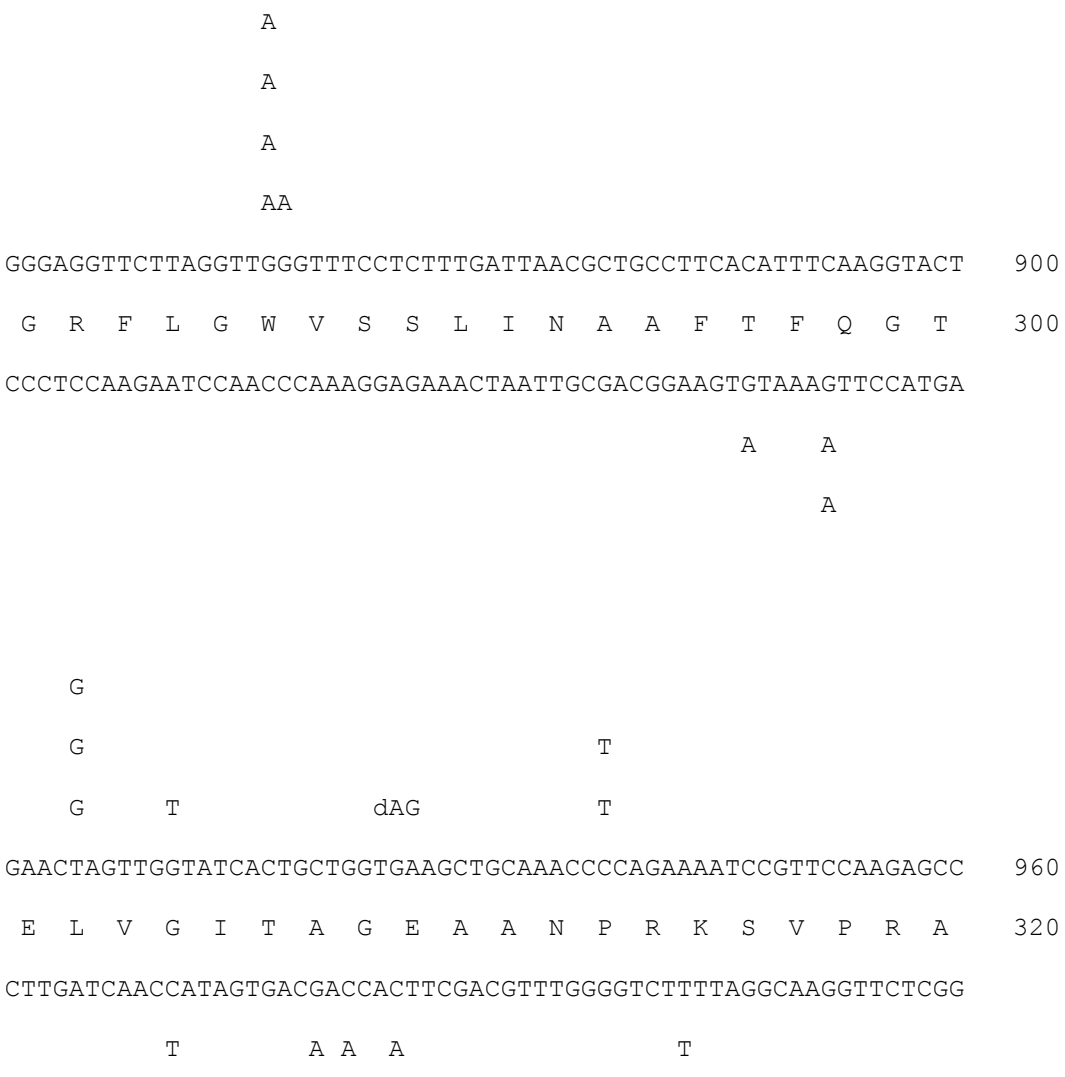
Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



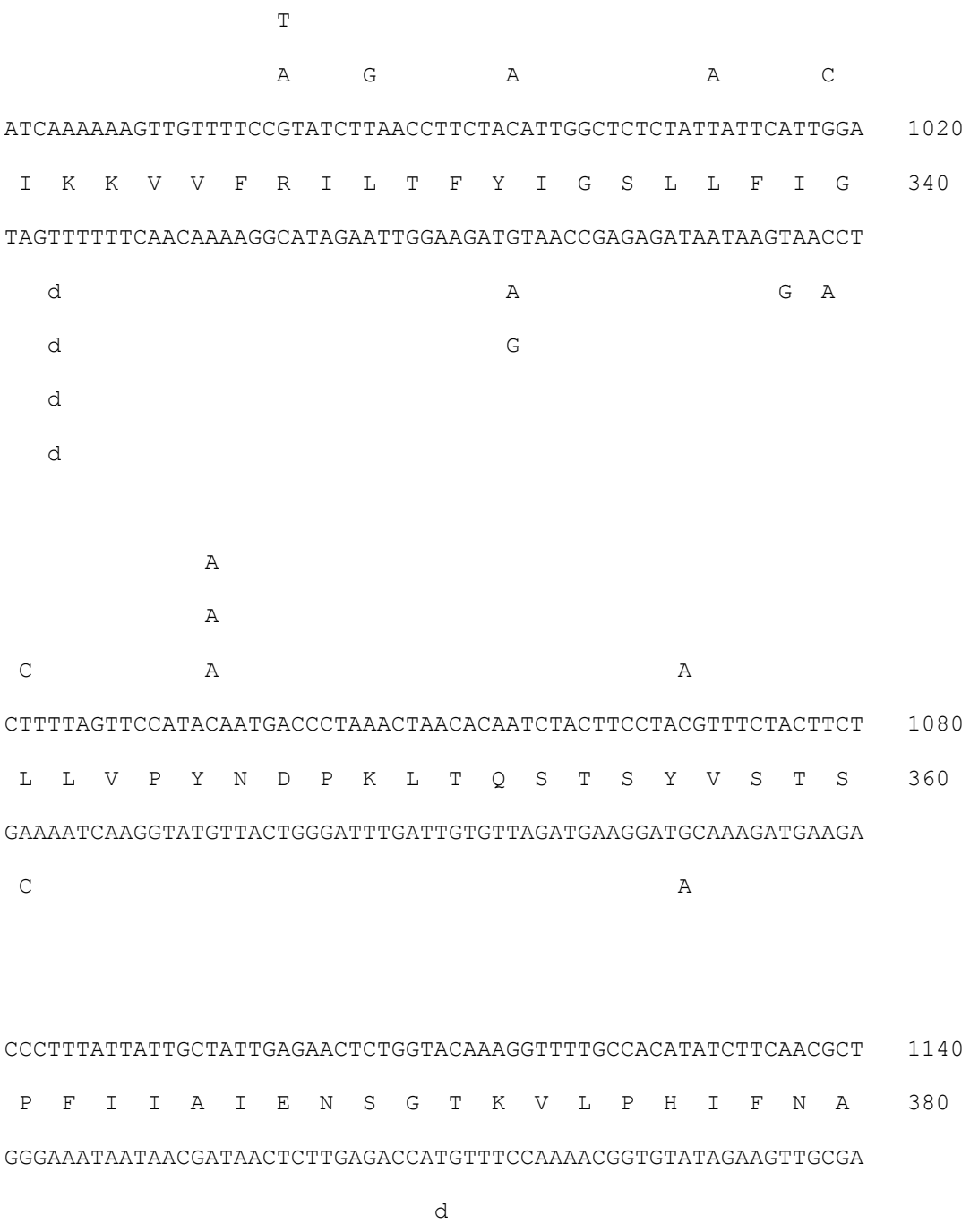
Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



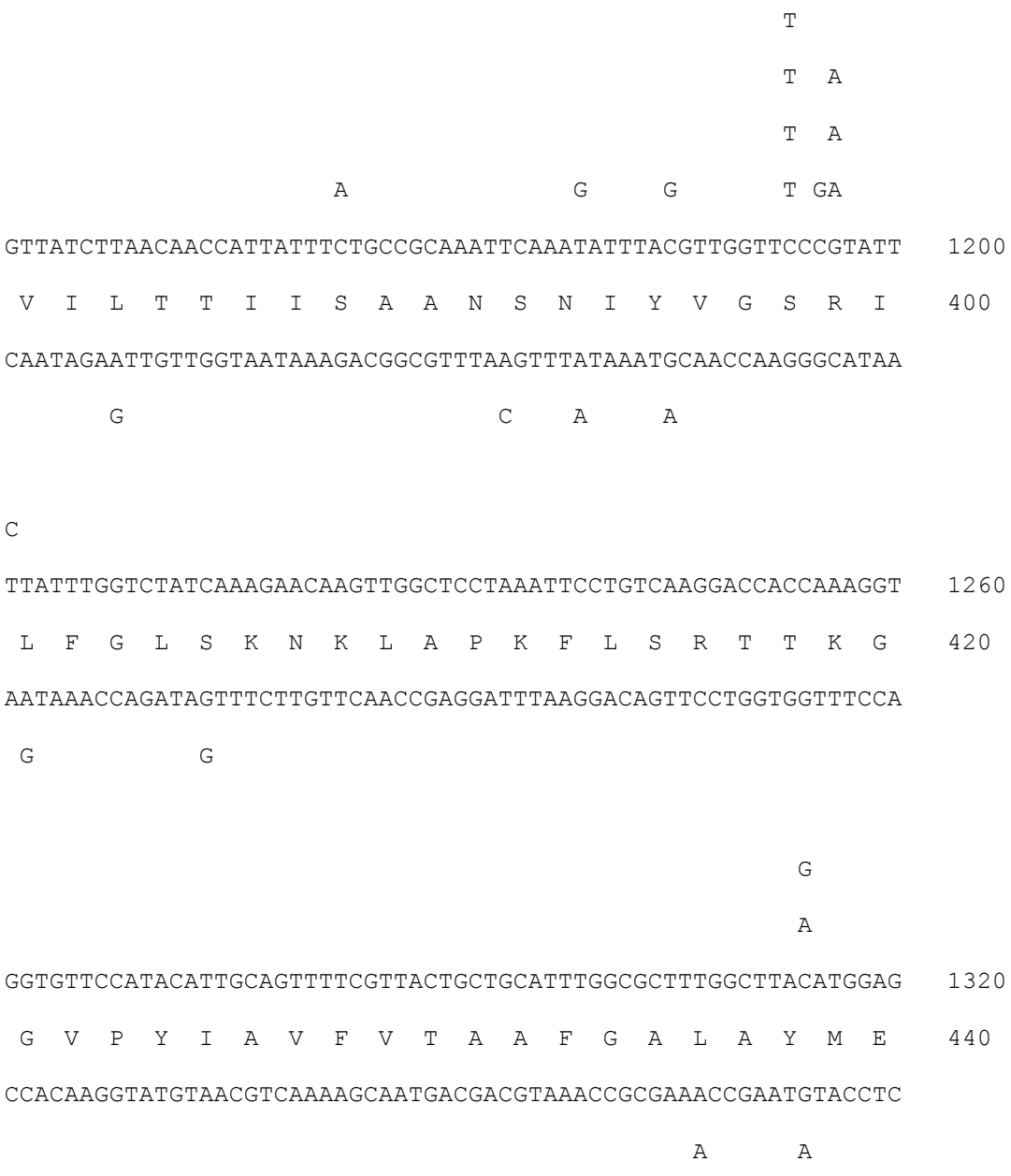
Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*



Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*

```
ACATCTACTGGTGGTGACAAAGTTTTCGAATGGCTATTAAATATCACTGGTGTTGCAGGC   1380
 T  S  T  G  G  D  K  V  F  E  W  L  L  N  I  T  G  V  A  G    460
TGTAGATGACCACCACTGTTTCAAAAGCTTACCGATAATTTATAGTGACCACAACGTCCG
                                                    T


d         A
TTTTTTGCATGGTTATTTATCTCAATCTCGCACATCAGATTTATGCAAGCTTTGAAATAC   1440
 F  F  A  W  L  F  I  S  I  S  H  I  R  F  M  Q  A  L  K  Y    480
AAAAAACGTACCAATAAATAGAGTTAGAGCGTGTAGTCTAAATACGTTCGAAACTTTATG
         CA d                                       C


                                           A
CGTGGCATCTCTCGTGACGAGTTACCATTTAAAGCTAAATTAATGCCCGGCTTGGCTTAT   1500
 R  G  I  S  R  D  E  L  P  F  K  A  K  L  M  P  G  L  A  Y    500
GCACCGTAGAGAGCACTGCTCAATGGTAAATTTCGATTTAATTACGGGCCGAACCGAATA


TATGCGGCCACATTTATGACGATCATTATCATTATTCAAGGTTTCACGGCTTTTGCACCA   1560
 Y  A  A  T  F  M  T  I  I  I  I  I  Q  G  F  T  A  F  A  P    520
ATACGCCGGTGTAAATACTGCTAGTAATAGTAATAAGTTCCAAAGTGCCGAAAACGTGGT
G      d                                                   A
```

Supplementary Figure 2.3 (continued) *CAN1* mutation spectra wild-type versus *mlh3*

```
AAATTCAATGGTGTTAGCTTTGCTGCCGCCTATATCTCTATTTTCCTGTTCTTAGCTGTT   1620
 K  F  N  G  V  S  F  A  A  A  Y  I  S  I  F  L  F  L  A  V    540
TTTAAGTTACCACAATCGAAACGACGGCGGATATAGAGATAAAAGGACAAGAATCGACAA
                                       d


   d   A
TGGATCTTATTTCAATGCATATTCAGATGCAGATTTATTTGGAAGATTGGAGATGTCGAC   1680
 W  I  L  F  Q  C  I  F  R  C  R  F  I  W  K  I  G  D  V  D    560
ACCTAGAATAAAGTTACGTATAAGTCTACGTCTAAATAAACCTTCTAACCTCTACAGCTG
  A

ATCGATTCCGATAGAAGAGACATTGAGGCAATTGTATGGGAAGATCATGAACCAAAGACT   1740
 I  D  S  D  R  R  D  I  E  A  I  V  W  E  D  H  E  P  K  T    580
TAGCTAAGGCTATCTTCTCTGTAACTCCGTTAACATACCCTTCTAGTACTTGGTTTCTGA


TTTTGGGACAAATTTTGGAATGTTGTAGCATAG   1800
 F  W  D  K  F  W  N  V  V  A  *    600
AAAACCCTGTTTAAAACCTTACAACATCGTATC
```

Supplementary Figure 2.4 *CAN1* mutation spectra wild-type versus *mlh1*

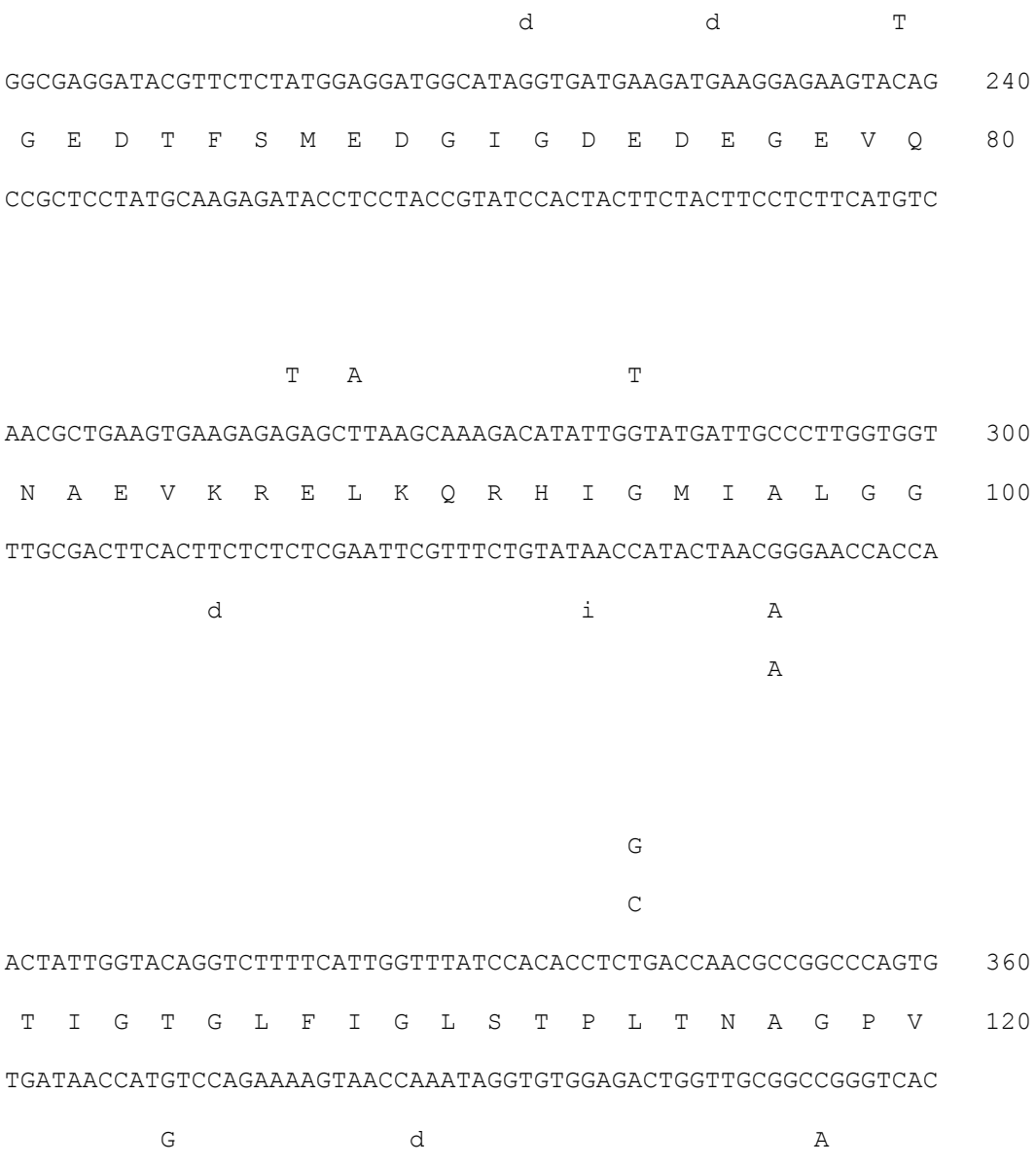
Top = 3686 wild-type

Bottom = 4237 *mlh1*

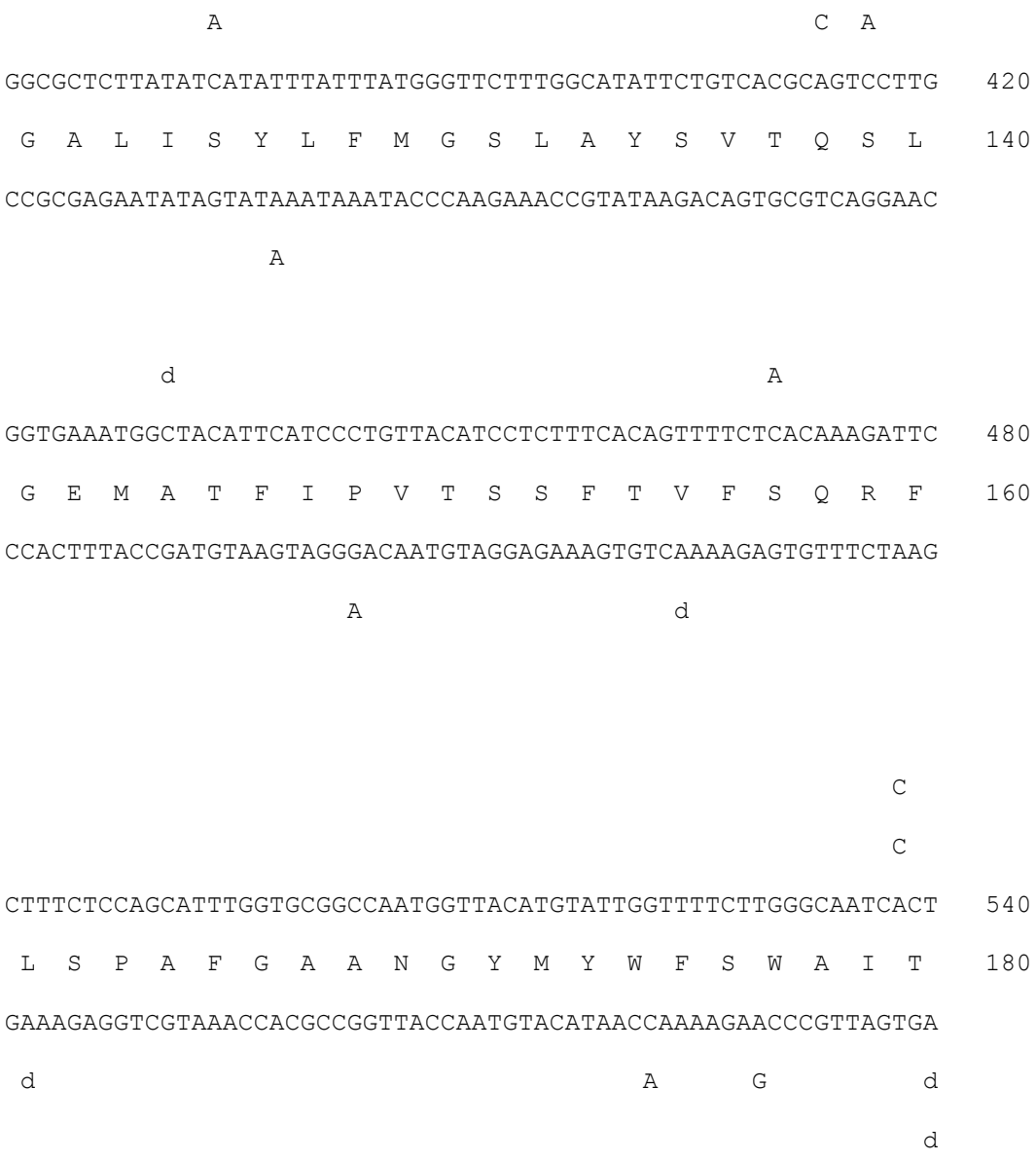
Base substitutions shown as a different nucleotide at the corresponding position. Frameshift deletions of -1 or -2 shown as d. Frameshift insertions of +1 or +2 as i.

```
ATGACAAATTCAAAAGAAGACGCCGACATAGAGGAGAAGCATATGTACAATGAGCCGGTC   60
 M  T  N  S  K  E  D  A  D  I  E  E  K  H  M  Y  N  E  P  V    20
TACTGTTTAAGTTTTCTTCTGCGGCTGTATCTCCTCTTCGTATACATGTTACTCGGCCAG

                                                    G
ACAACCCTCTTTCACGACGTTGAAGCTTCACAAACACACCACAGACGTGGGTCAATACCA   120
 T  T  L  F  H  D  V  E  A  S  Q  T  H  H  R  R  G  S  I  P    40
TGTTGGGAGAAAGTGCTGCAACTTCGAAGTGTTTGTGTGGTGTCTGCACCCAGTTATGGT

TTGAAAGATGAGAAAAGTAAAGAATTGTATCCATTGCGCTCTTTCCCGACGAGAGTAAAT   180
 L  K  D  E  K  S  K  E  L  Y  P  L  R  S  F  P  T  R  V  N    60
AACTTTCTACTCTTTTCATTTCTTAACATAGGTAACGCGAGAAAGGGCTGCTCTCATTTA
```

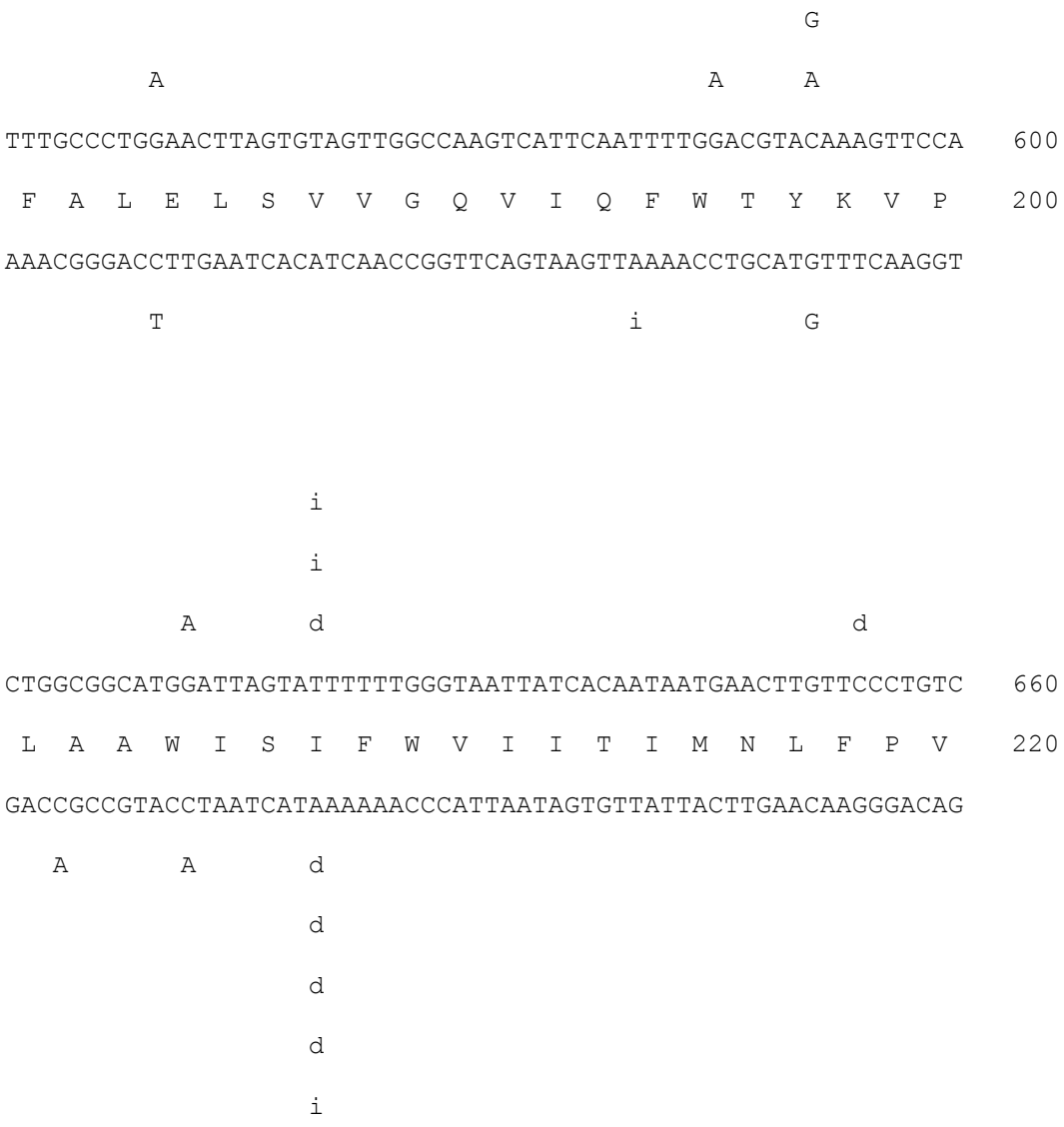
Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



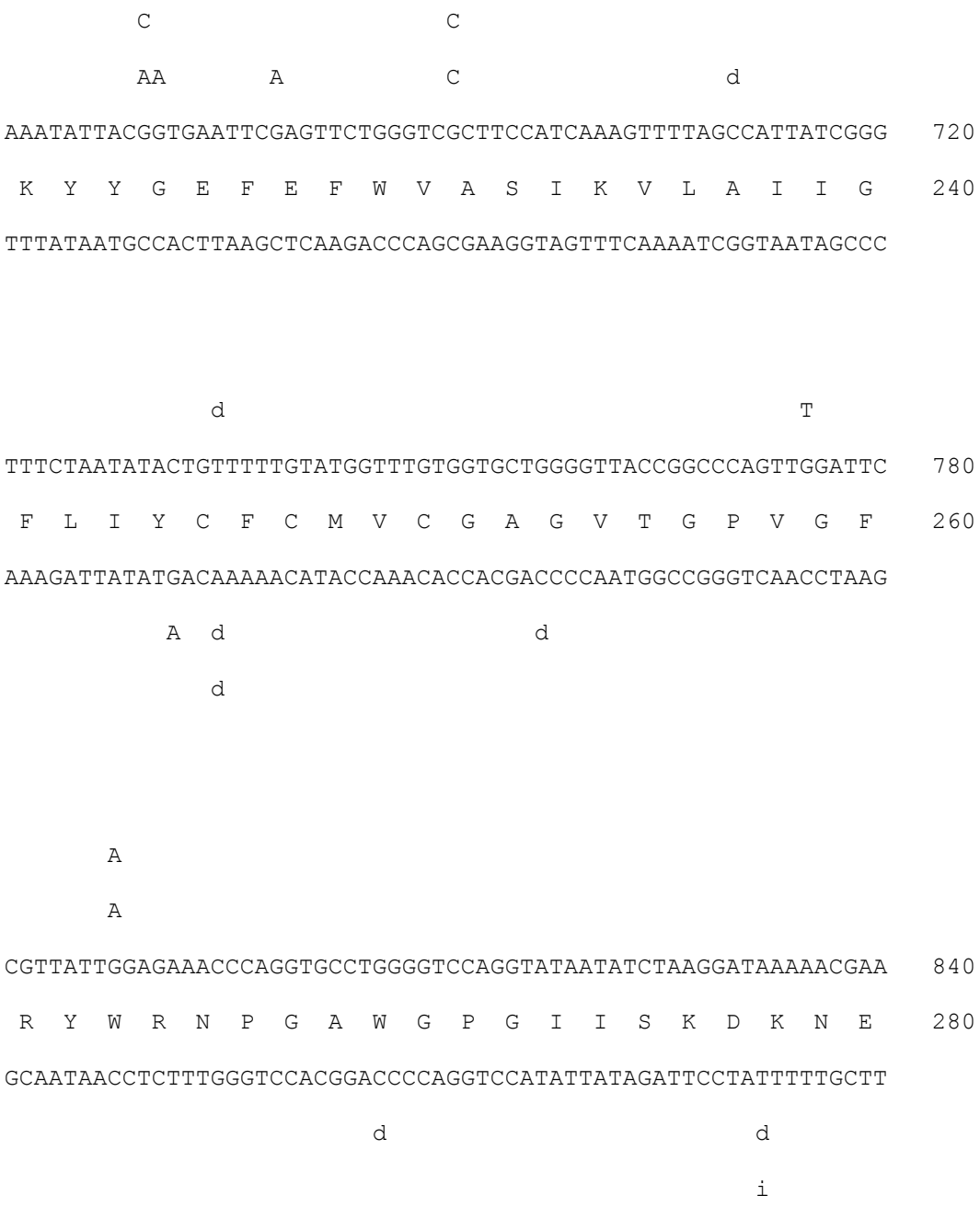
Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



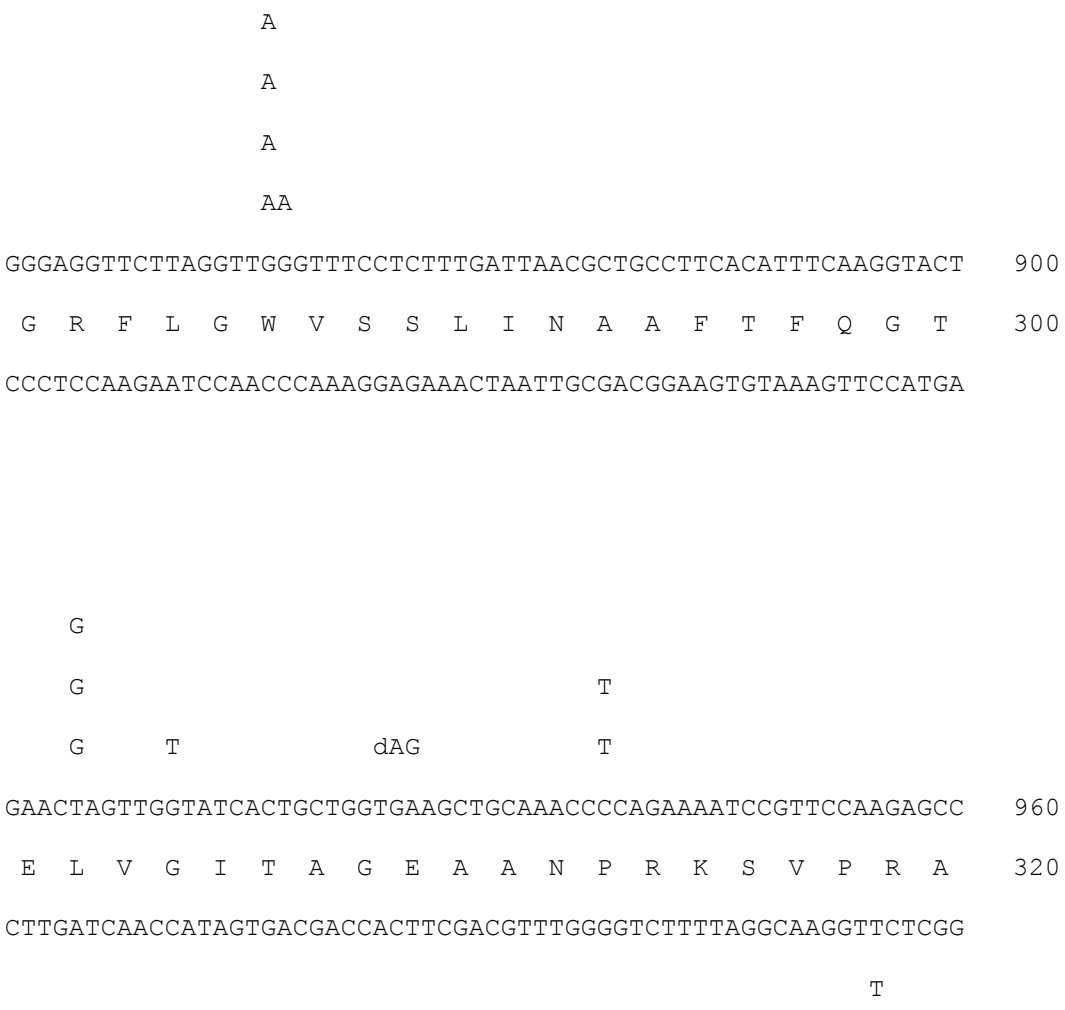
Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



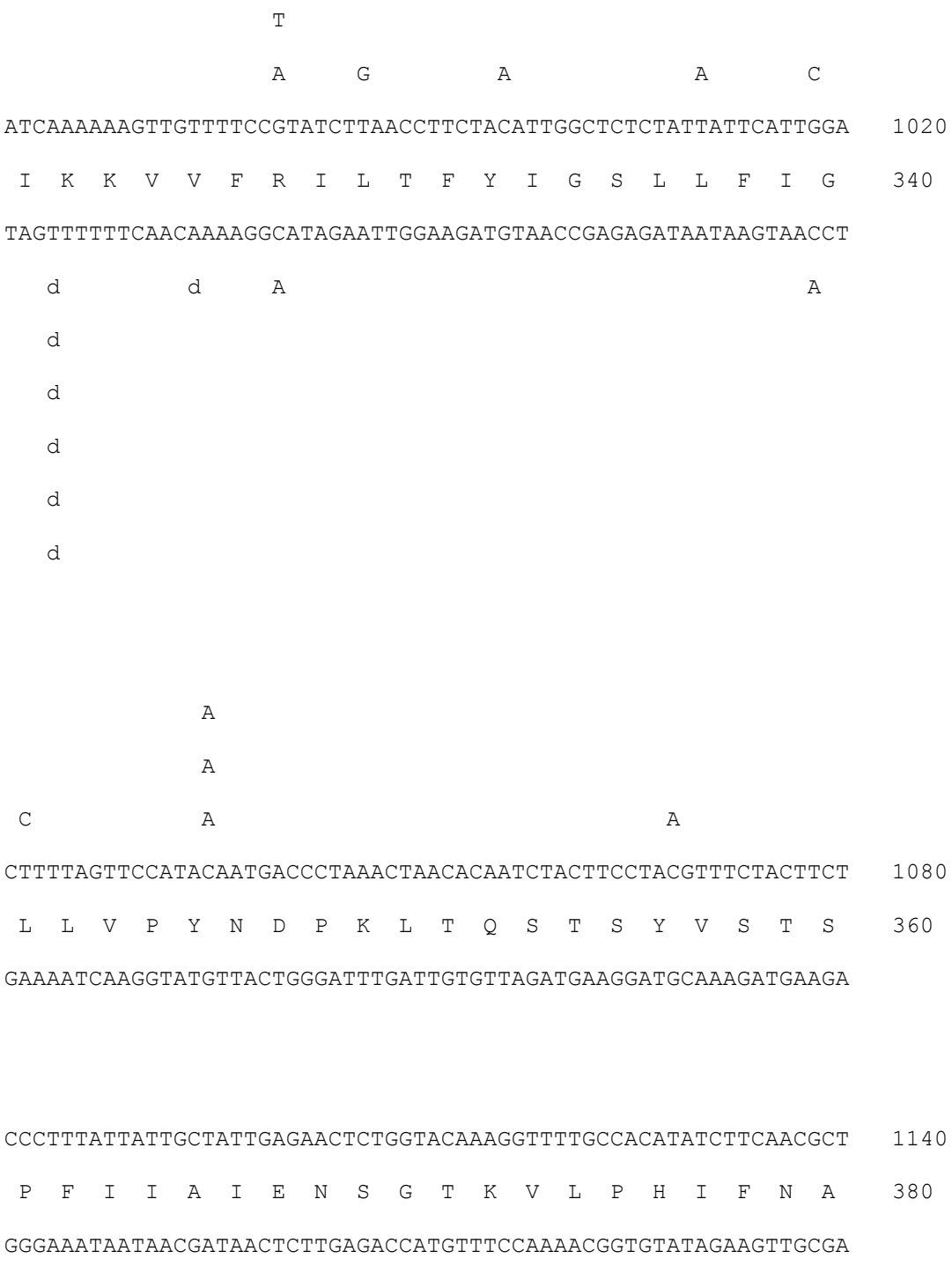
Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



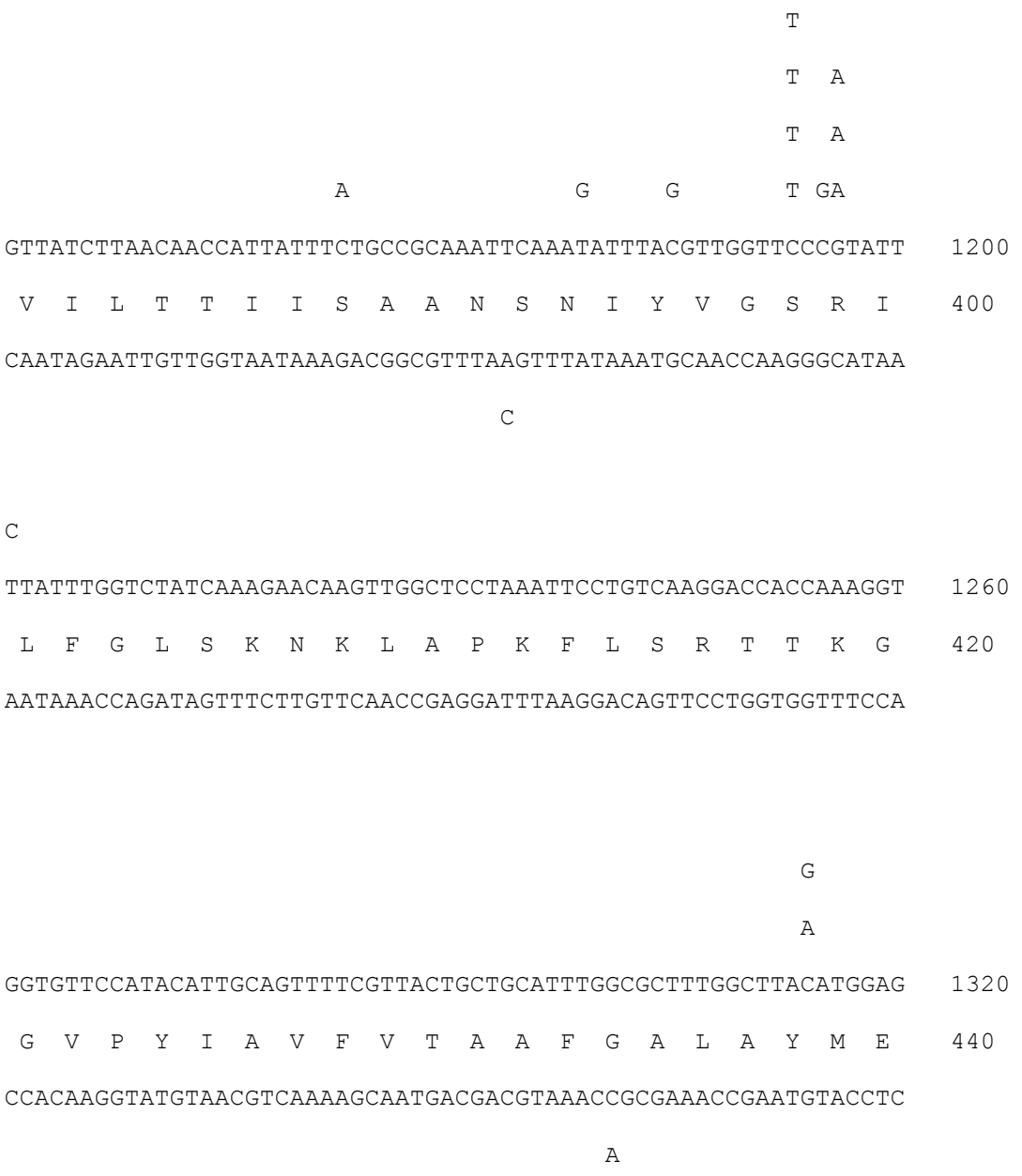
Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



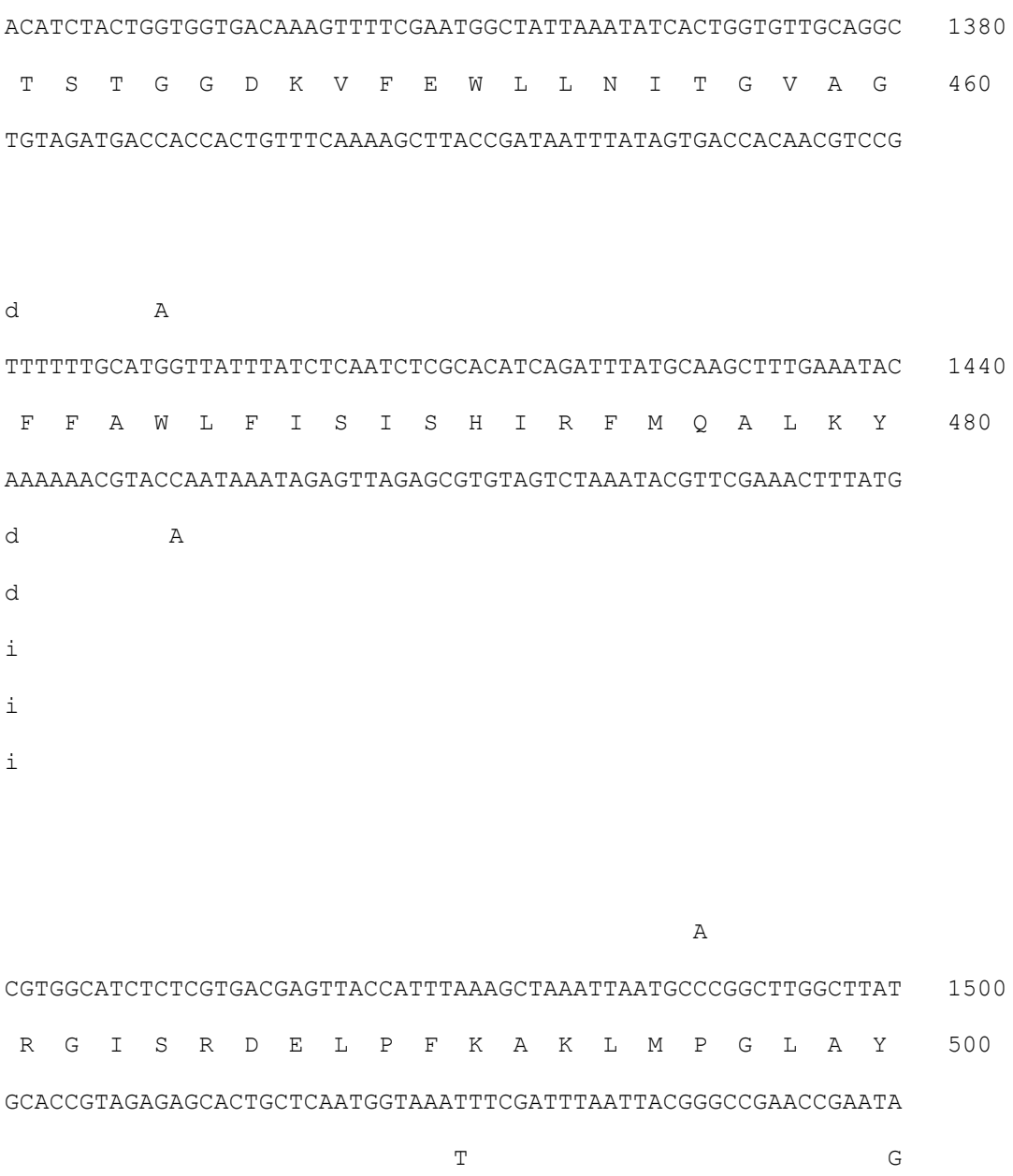
Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*



Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*

```
TATGCGGCCACATTTATGACGATCATTATCATTATTCAAGGTTTCACGGCTTTTGCACCA   1560
 Y  A  A  T  F  M  T  I  I  I  I  I  Q  G  F  T  A  F  A  P    520
ATACGCCGGTGTAAATACTGCTAGTAATAGTAATAAGTTCCAAAGTGCCGAAAACGTGGT

AAATTCAATGGTGTTAGCTTTGCTGCCGCCTATATCTCTATTTTCCTGTTCTTAGCTGTT   1620
 K  F  N  G  V  S  F  A  A  A  Y  I  S  I  F  L  F  L  A  V    540
TTTAAGTTACCACAATCGAAACGACGGCGGATATAGAGATAAAAGGACAAGAATCGACAA
                                       d

   d  A
TGGATCTTATTTCAATGCATATTCAGATGCAGATTTATTTGGAAGATTGGAGATGTCGAC   1680
 W  I  L  F  Q  C  I  F  R  C  R  F  I  W  K  I  G  D  V  D    560
ACCTAGAATAAAGTTACGTATAAGTCTACGTCTAAATAAACCTTCTAACCTCTACAGCTG
  A             A

ATCGATTCCGATAGAAGAGACATTGAGGCAATTGTATGGGAAGATCATGAACCAAAGACT   1740
 I  D  S  D  R  R  D  I  E  A  I  V  W  E  D  H  E  P  K  T    580
TAGCTAAGGCTATCTTCTCTGTAACTCCGTTAACATACCCTTCTAGTACTTGGTTTCTGA
```

Supplementary Figure 2.4 (continued) *CAN1* mutation spectra wild-type versus *mlh1*

```
TTTTGGGACAAATTTTGGAATGTTGTAGCATAG   1800
 F  W  D  K  F  W  N  V  V  A  *    600
AAAACCCTGTTTAAAACCTTACAACATCGTATC
```

Chapter 2, in full, is a reprint of the material as it appears in *Mol Cell Biol*, 2007(18):6546-64, Harrington, JM, Kolodner RD. The dissertation author was the primary researcher and author of this paper.

Chapter 3: Msh2-Msh3 mispair recognition involves DNA bending and strand separation

3.1 Introduction

The DNA mismatch repair (MMR) pathway recognizes and repairs mispaired and damaged bases in DNA, which primarily result from replication errors but also result from recombination and chemical damage to DNA and DNA precursors (Kolodner and Marsischky, 1999; Modrich, 1991). Repairing mispairs improves the overall fidelity of DNA replication and is important for genome stability (Modrich and Lahue, 1996). Inherited defects in MMR are responsible for most cases of Lynch Syndrome [Hereditary Non-Polyposis Colorectal Cancer (HNPCC)] and, furthermore, the epigenetic silencing of one of the genes involved in MMR, *MLH1*, underlies most cases of sporadic MMR-defective cancer (Lynch and de la Chapelle, 2003; Peltomaki, 2003).

MMR is initiated by the recognition of base:base mismatches or insertion/deletion mispairs. In bacteria, the homodimeric MutS complex directly binds mispairs, bending the mispair-containing DNA by almost 60 degrees, and shifting one of the mispaired bases, such as the thymidine base from G:T or +T mispairs, out of the DNA base stack (Lamers et al., 2000). The mispaired base is stabilized by π-stacking with a conserved phenylalanine (Constantin et al., 2005; Warren et al., 2007; Zhang et al., 2005). DNA binding induces a functional asymmetry to the MutS complex; one subunit directly

recognizes the mispair via a mispair-binding domain (MBD), whereas the MBD of the second subunit primarily is involved in non-specific backbone interactions (Lamers et al., 2000).

In eukaryotes, mitotic MMR utilizes two heterodimeric complexes of MutS Homologs: Msh2-Msh6 and Msh2-Msh3 (Kolodner and Marsischky, 1999). In these asymmetric heterodimers, Msh6 and Msh3 directly recognize the mispair via their MBD, whereas the Msh2 subunit appears to be functionally equivalent to the MutS subunit that non-specifically binds the DNA backbone. The Msh2-Msh6 heterodimer primarily recognizes base:base mispairs and small 1 or 2 nucleotide insertion/deletions (Harfe and Jinks-Robertson, 2000; Kolodner and Marsischky, 1999; Modrich, 1991; Modrich, 2006; Modrich and Lahue, 1996). The crystal structure of human Msh2-Msh6 revealed that mispair recognition by Msh6 shares many details with *E. coli* MutS, including the π-stacking phenylalanine (Drotschmann et al., 2001; Holmes et al., 2007; Warren et al., 2007). In contrast, the Msh2-Msh3 heterodimer primarily recognizes insertions and deletions from 1 to 14 nucleotides in size (Habraken et al., 1996; Marsischky et al., 1996; Marsischky and Kolodner, 1999; Palombo et al., 1996; Sia et al., 2001; Surtees and Alani, 2006; Wilson et al., 1999), although we have previously shown that Msh2-Msh3 also recognizes some base:base mispairs with a preference for those that have weak hydrogen bonding (Harrington and Kolodner, 2007).

While no structural information is available for any Msh3 homolog, several lines of evidence suggest that mispairs are recognized by Msh2-Msh3 in a substantially different way than mispairs are recognized by MutS and Msh2-Msh6. First, Msh3 lacks the conserved π-stacking phenylalanine present in both MutS and Msh6, which is required for MMR by these proteins *in vivo* (Drotschmann et al., 2001; Lee et al., 2007). In contrast, mutagenesis of the *Saccharomyces cerevisiae* Msh3 residue located in the equivalent position to the phenylalanine conserved in MutS and Msh6 (K158, called K187 prior to the identification of the correct start codon (Harrington and Kolodner, 2007)) only caused a modest MMR defect (Lee et al., 2007). Second, when other conserved residues and predicted DNA-backbone contacting residues in *S. cerevisiae* Msh3 were mutated to alanine, only R247A (previously called R276A) showed a significant defect in the repair of 1, 2, and 4 nucleotide-long insertion/deletion mispairs (Lee et al., 2007).

Despite these differences, the Msh3 MBD is likely related to the MBD of Msh6 and MutS. Replacement of the Msh6 MBD with the Msh3 MBD generated a functional chimera possessing Msh3 substrate specificity (Shell et al., 2007a). Moreover, combining the Msh3 K158A mutation with K160A gave rise to a *msh3* mutant with a greater MMR defect than either single mutant alone. This double mutant caused a loss of specificity for mispaired DNA (Lee et al., 2007). Together these data indicate not only that mispair specificity is determined by the Msh3 MBD, but also that the critical region of the Msh3

MBD mediating mispair recognition likely overlaps the same region as the MBDs of MutS and Msh6, even if the nature of the recognition is different. We have therefore used homology modeling and site-directed mutagenesis to gain insight into how Msh3 recognizes a diverse array of mispairs.

3.2 Homology model of the Msh3 MBD

The extensive conservation between the MBD of MutS from bacteria, Msh6 from yeast and humans, and Msh3 from yeast and humans (Fig. 3.1a), the similar patterns of predicted secondary structure (data not shown), and the ability to form a functional Msh6 chimera with a Msh3 MBD (Shell et al., 2007a) all argue that the overall fold of the Msh3 MBD is conserved with other MutS homologs. We therefore generated a homology model of the *S. cerevisiae* Msh3 MBD (Fig. 3.1b) using the structure of the human Msh6 MBD (Warren et al., 2007). Superimposition of this model on human Msh2-Msh6 complexed with a G:T mispair revealed a number of clues to differences between DNA binding features of the Msh6 and Msh3 MBDs. Both K158, which is conserved in Msh3 and aligns with the π-stacking phenylalanine in MutS and Msh6 (Fig. 3.1a), and S201, which is also conserved in Msh3 and aligns with a conserved glycine in MutS and Msh6 that packs against the displaced nucleotide (Fig. 3.1a), sterically clash with the displaced thymidine in the Msh2-Msh6 complex (Fig. 3.1c). This model suggests that the displacement and stabilization of a single nucleotide from the base stack by MutS and Msh6 either does not occur or occurs in a different fashion in Msh3.

3.3 *msh3* mutants differentially repair different DNA lesions

To experimentally probe the interactions between Msh3 and mispaired DNA, we designed a series of *msh3* point mutant alleles in the MBD focusing on residues predicted to be at the MBD-DNA interface, but also including residues from other regions of the MBD. These *msh3* alleles were tested by expression from the native *MSH3* promoter on a low copy number plasmid in a *msh3Δ msh6Δ* yeast strain and evaluated for their effect on MMR proficiency using the -1 nucleotide *hom3-10* frameshift reversion assay (Marsischky et al., 1996; Wang et al., 1990) (Fig. 3.2). Four *msh3* alleles had wild-type phenotypes and were not further studied including E164A, R171A, H174A and H194E (Suppl. Table 3.1). Of these mutations, only H174A and H194E affected residues with side chains predicted to be within 6 Å of the DNA. However, alleles predicted to effect amino acid residues at the MBD-DNA interface as well as some slightly removed from the interface had a defect in the *hom3-10* reversion assay including Y157S, K158D, K160D, F162A, R195D, F197A, Y199A, S201G, R206A, and H210A.

When alleles defective in the *hom3-10* frameshift assay were tested for their effects in the repair of 2 and 4 nucleotide microsatellite stability assays (Sia et al., 1997), the alleles fell into two distinct classes (Fig. 3.2; Suppl. Table 3.1). One class also had defects for the repair of 2 nucleotide and 4 nucleotide loops and included Y157S, K158D, K160D, F162A, F197A, and H210A. This class also included the ERN allele that replaced the insertion

between β3 and β4 (G180 to Q196; Fig. 3.1a) in the MBD with the sequence ERN from Msh3 from the fungus *Ustilago maydis*. The other class had no defect or nearly no defect in microsatellite stability and included R195D, Y199A, S201G and R206A.

Two mutations that caused specific defects in frameshift repair when changed to the equivalent Msh6 or MutS residues, S201G and R206A, were used to design *msh6* alleles encoding the Msh3 residue, G368S and S373R, to analyze their effect on Msh6-mediated frameshift repair. Neither *msh6* allele enhanced frameshift repair in the *hom3-10* reversion assay; the *msh6-G368S* allele was completely defective whereas the *msh6-S373R* allele did not cause any defect (Suppl. Fig. 3.1).

3.4 Additional mutations in the Msh3 MBD-DNA interface fall into two classes

To further investigate the *msh3* Y157S, K158D, F162A, F197A, Y199A, and S201G alleles, we generated additional mutations that resulted in different amino acid substitutions at each position, and tested them using the *hom3-10* frameshift reversion assay and 4 nucleotide microsatellite stability assay.

Msh3 Y157 is positioned in our Msh3 MBD model to stack on bases in the strand that does not contain the mispair in the Msh6 structure (Fig. 3.1d). This role would be predicted to be eliminated by the Y157S substitution. Consistent with this role, Y157F and Y157A were less defective for frameshift repair than Y157D and Y157L (Fig. 3.3a, b); however, both showed substantial

defects relative to wild-type. In contrast to Y157S and Y157D, alleles Y157F, Y157A, Y157L, and Y157A K158A were much more defective for frameshift repair than microsatellite stability (Fig. 3.3b; Suppl. Table 3.2).

Msh3 K158, which aligns the π-stacking phenylalanine in MutS and Msh6, was inactive when mutated to aspartate or glutamate. In contrast, K158R was indistinguishable from wild-type by 95% confidence intervals in the frameshift and microsatellite stability assays (Fig. 3.3a, b; Suppl. Table 3.2). Both K158M and K158A caused a slight defect primarily in the frameshift repair assay.

Msh3 F162Y caused a 18-fold defect in frameshift repair, but was indistinguishable from the wild-type rate for microsatellite stability. In contrast, F162A and F162S caused complete defects in both assays (Fig. 3.3a, b; Suppl. Table 3.2). Importantly, the relative defect in the frameshift assay was similar to the relative defect observed in the microsatellite stability assay for each of the F162 alleles (Fig. 3.3b).

Msh3 F197H caused a 114-fold defect in frameshift repair, but a much more modest defect in microsatellite stability. In contrast, F197A was indistinguishable from the empty vector control for both frameshift and microsatellite stability assays (Fig. 3.3a, b).

Msh3 Y199A was also changed to leucine, aspartate and lysine. When qualitatively tested for MMR proficiency using patch tests, the Y199D allele was completely defective in both assays, and the Y199K allele was partially

defective in both assays, similar to the original Y199A allele. The Y199L caused a greater defect in frameshift repair than 4 nucleotide microsatellite stability assays (Suppl. Figure 3.2).

Msh3 S201G was changed to leucine, aspartate and arginine residues. The S201L allele was partially defective in both frameshift and microsatellite assays. The S201D and S201R alleles caused null phenotypes in both frameshift repair and microsatellite stability assays (Suppl. Figure 3.2).

Mapping alleles causing MMR defects onto the Msh3 MBD model (Fig. 3.4a, b) revealed that a central region, likely directly involved in mispair recognition, contains positions that when mutagenized only cause equivalent defects in all MMR assays, other positions that only cause greater defects in frameshift repair than microsatellite stability assays, and yet other positions that cause either type of defect depending on the specific amino acid substitution. Remarkably, most of the central positions can be mutated to alleles that either equally affect frameshift repair and microsatellite stability or primarily affect frameshift repair. Those sites only associated with defects that primarily affect frameshift repair tend to be on the periphery of the core recognition region.

3.5 Discussion

Here we have demonstrated by theoretical modeling and analysis of point mutations that mismatch recognition by Msh3 differs from MutS and Msh6. Unlike MutS and Msh6, there is no clear equivalent in Msh3 to the π-

stacking phenylalanine involved in stabilizing bases in the mismatch as at least some alternative amino acids could be tolerated at each of the positions tested. Additionally, swapping individual amino acid residues or short stretches between the Msh3 and Msh6 MBD has not successfully altered mispair specificity as demonstrated here and previously (Lee et al., 2007)(Shell et al., 2007a). We have also shown that mutations affecting the Msh3 MBD fall into two classes. One class, including Y157D, Y157S, K158D, K158R, K158E, F162A, F162S, F197A, Y199D, Y199K, S201D, S201L, S201R, and H210A, had similar effects on all Msh3-based repair. The second class, including Y157F, Y157A, Y157L, Y157A K158A, K158A, K158M, F162Y, R195D, F197H, Y199A, Y199L, S201G, and R206A, selectively disrupted 1 nucleotide frameshift repair but not 2 and 4 base loop repair; we would also anticipate that these mutations would prevent repair of the A:A, A:T, C:C and C:T base-base mismatches that are recognized and repaired by Msh3 but we did not specifically test this (Harrington and Kolodner, 2007). Importantly, we have not identified any mutations that specifically cause defects in 2 and 4 base loop repair but were still proficient for 1 base frameshift repair, which indicates that the larger loops are more readily repaired by Msh2-Msh3 than frameshifts.

Why should repair of DNA loops present in larger insertion/deletion mispairs be less sensitive to mutation of the Msh3 MBD than repair of smaller frameshift mispairs? Structures of DNAs containing insertions of several

nucleotides (+ 5A insertion) demonstrate that these insertions form loops that cause the DNA helix to bend and forces the inserted nucleotides to separate from the other strand (Fig. 3.4d) (Dornberger et al., 1999). The overall orientation and bend of the DNA strands are highly reminiscent of the G:T mispaired DNA bound by Msh2-Msh6 (Fig. 3.4c) (Warren et al., 2007), which is stabilized by Msh2-Msh6 binding (Warren et al., 2007). Smaller frameshift mispairs, on the other hand, are substantially less bent and the loop-containing strand is not as separated as seen with DNAs containing large loops (Fig. 3.4e) (Natrajan et al., 2003; Warren et al., 2007). Thus, we propose that frameshift mutations require additional stabilization relative to large loops to be bent and recognized by the Msh3 MBD. This hypothesis would explain why we observe a class of mutations that are specifically defective in the repair of 1 base frameshift insertions and why we do not observe mutations that are specifically defective in the repair of larger loops. This hypothesis also is consistent with the fact that positions that only affect frameshift repair when mutated are on the outside of the central recognition region (Fig. 3.4a, b). The fact that the central region typically contains positions that when mutated can affect both frameshift and microsatellite repair or primarily frameshift repair suggests that mispair recognition features of Msh2-Msh3 are frequently the same features that stabilize induced conformations in small insertion/deletion mispairs and that these sites cannot be cleanly separated.

Analysis of individual mutants in the context of the homology model also suggests that strand separation is important for mispair recognition by Msh2-Msh3, which is distinct from how Msh2-Msh6 and MutS recognize mispairs. Msh3 Y157 is well positioned to stack with bases of the non-loop containing strand (Fig. 3.1d), whereas Msh3 K158, K160, and S201 could be part of either a steric wedge separating the two strands and hydrogen bonding to bases at the insertion/deletion site or a specific surface that interacts with and stabilizes phosphates of a displaced and nucleotide-flipped loop-containing strand (Figs. 3.1c, 3.4e). Charge and size seem to be critical for the role of K158: K158R was mostly functional; K158A and K158M had increased defects primarily in frameshift repair; and the negatively charged K158D or K158E caused a substantial MMR defect as did the negatively charged K160D. If Msh3 binds to and stabilizes a strand-separated substrate, then residues like F197 might π-stack with bases in the loop. We note that the more conservative F197H allele that could retain some π-stacking ability was less defective for Msh3 repair than F197A.

Recognition of a bent and strand-separated substrate could easily allow recognition of a range of different loop sizes, consistent with the wide range of sizes recognized by Msh2-Msh3 (from 1 to 14 nucleotides) (Habraken et al., 1996; Lee et al., 2007). This model is also consistent with the fact that Msh2-Msh3 has been observed to bind and distort some DNA substrates containing secondary structures, including substrates with 3' ssDNA overhangs and a

splayed Y structure (Surtees and Alani, 2006). The large loop-containing strand would also be positioned close to Msh2 domain I (*S. cerevisiae* Msh2 residues 2-133), which is equivalent to the Msh3 and Msh6 MBD. Intriguingly, Msh3, but not Msh6, requires Msh2 Domain I for repair (Lee et al., 2007), although this is not a fundamental requirement of the Msh3 MBD as an Msh6 chimera containing the Msh3 MBD was independent of Msh2 Domain I (Shell et al., 2007a). If the Msh3 scaffold has evolved to require loop binding by Msh2 Domain I, this could explain the failure of the reverse chimera constructs with the Msh6 MBD placed into the Msh3 scaffold to support MMR (Shell et al., 2007a). The model presented here provides an explanation of the flexibility of Msh3 recognition of substrates from weakly hydrogen bonded base:base mispairs to large insertion/deletion loops; however, analysis of the precise details of the interface await structure determination of Msh2-Msh3 complexes with insertion/deletion mispairs at atomic resolution

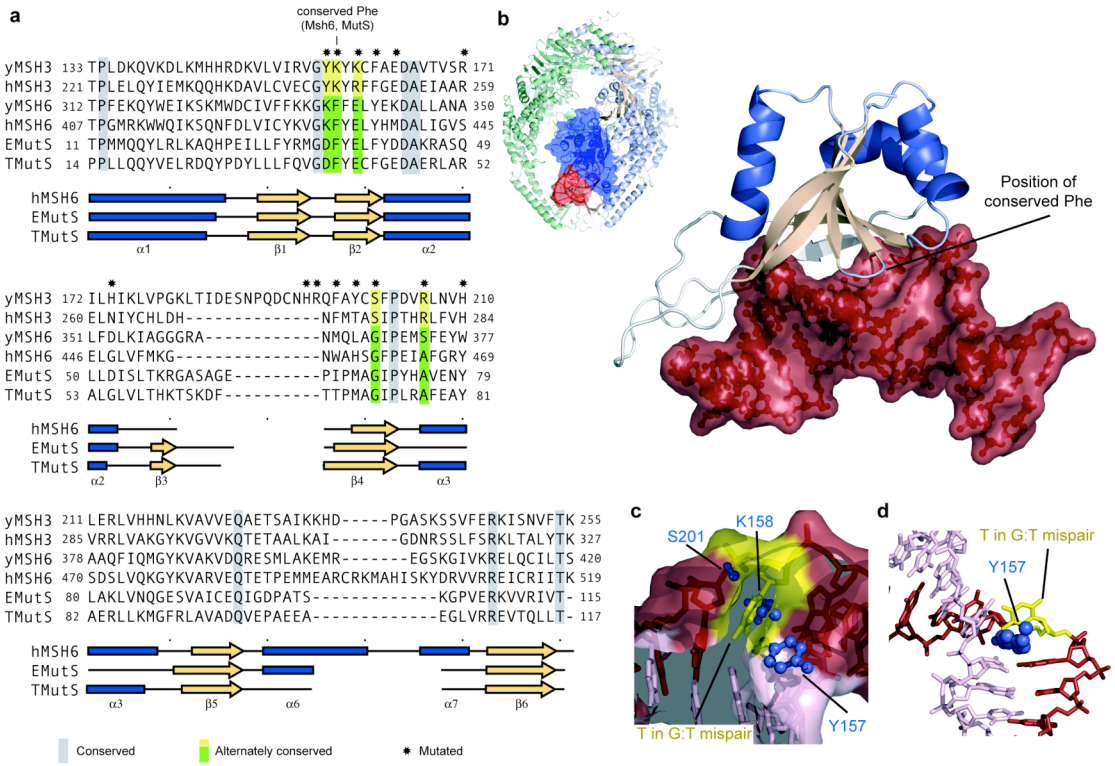


Figure 3.1 Modeling of Msh3 MBD

(a) Alignment of the MutS homologue protein sequences. Msh3 from *Saccharomyces cerevisiae* (y), *Homo sapiens* (h), Msh6 from *Saccharomyces cerevisiae* (y), *Homo sapiens* (h), and MutS from *Escherichia coli* (E), *Thermus aquaticus* (T). Grey boxes indicate conserved amino acid residues, green and yellow boxes indicate amino acid residues differentially conserved between Msh3, Msh6 and MutS. Asterisks indicate residues that were mutated in this study. Secondary structure for the *E. coli* MutS (PDB id 1e3m)(Lamers et al., 2000), *Thermus aquaticus* MutS (PDB id 1fw6)(Junop et al., 2001) and human Msh6 (PDB id 2o8b)(Warren et al., 2007) are shown below the amino acid sequence. Blue bars are a-helices, and peach arrows are b-sheets. (b) Model of Msh3 MBD on a G:T mispaired DNA (red) from the Msh2-Msh6 crystal structure (PDB id 2o8b)(Warren et al., 2007). Regions of low-confidence (see Methods) are shown in white. Inset shows the Msh2-Msh6 heterodimer on G:T mispaired DNA, with the Msh6 MBD in dark blue. (c) Model of Msh3 MBD residues on a G:T mispair reveals steric clash of K158, S201, and possibly Y157 (blue) with the mispaired T (yellow). (d) Possible stacking of Y157 with the bases of the non-T containing strand (pink). Molecular images generated with PyMOL (DeLano, 2002).

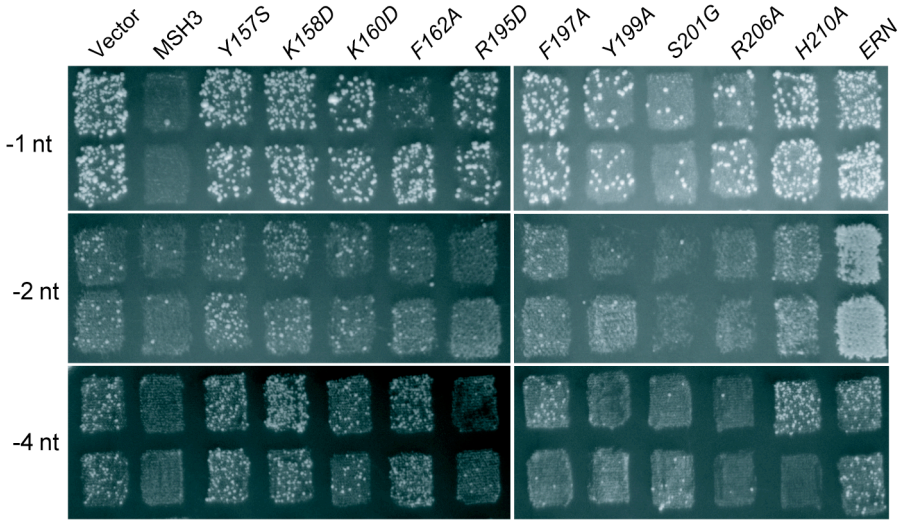


Figure 3.2 Supression of the *msh3Δ* phenotype by plasmid-borne *msh3* mutant alleles in MMR assays

Patches of *msh3Δ msh6Δ* strains expressing plasmid-borne *msh3* alleles were replica plated onto –threonine plates for the -1 nucleotide *hom3-10* reversion assay. Patches of *msh3Δ msh6Δ* strains expressing *msh3* alleles and containing a microsatellite plasmid with an in frame 2 or 4 nucleotide repeat sequence upstream of the *URA3* gene were replica plated onto –leucine –tryptophan +uracil +5-fluoroorotic acid plates as shown.

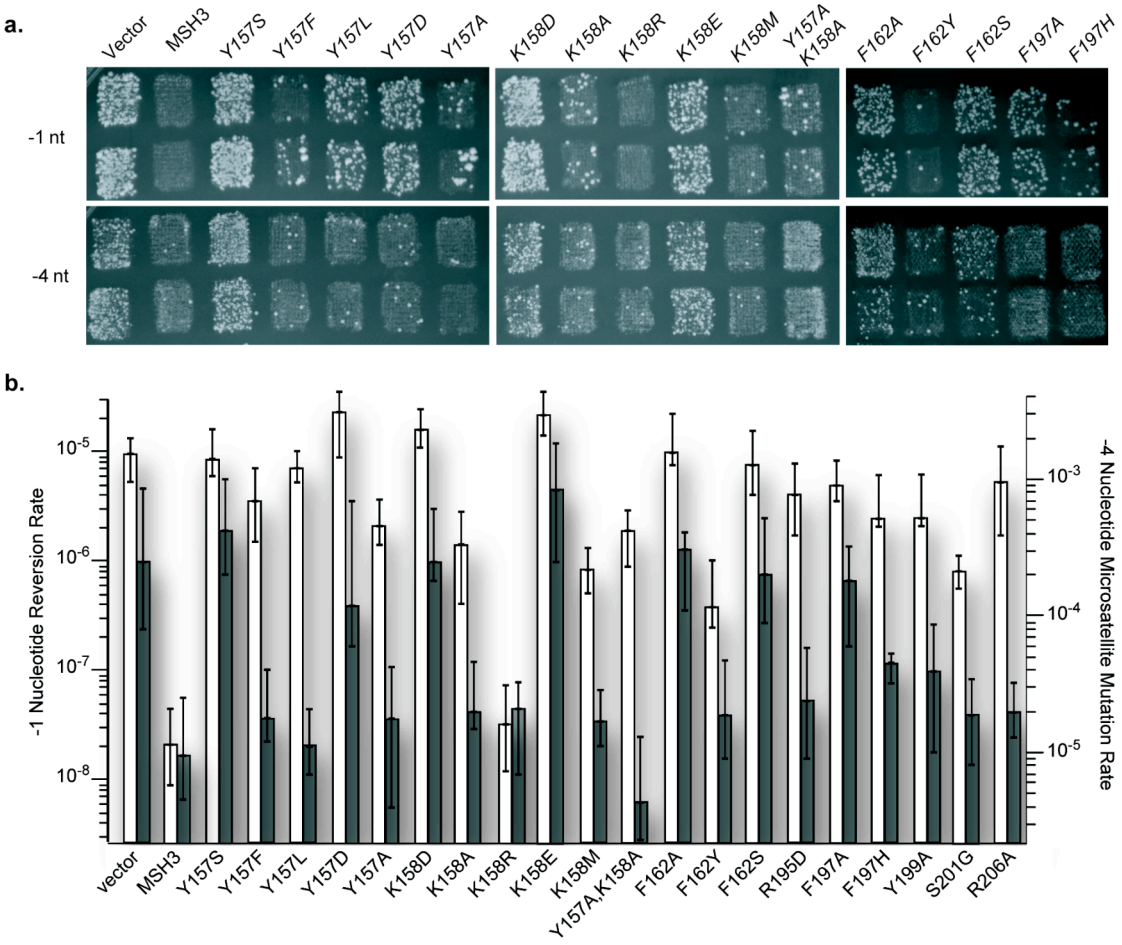
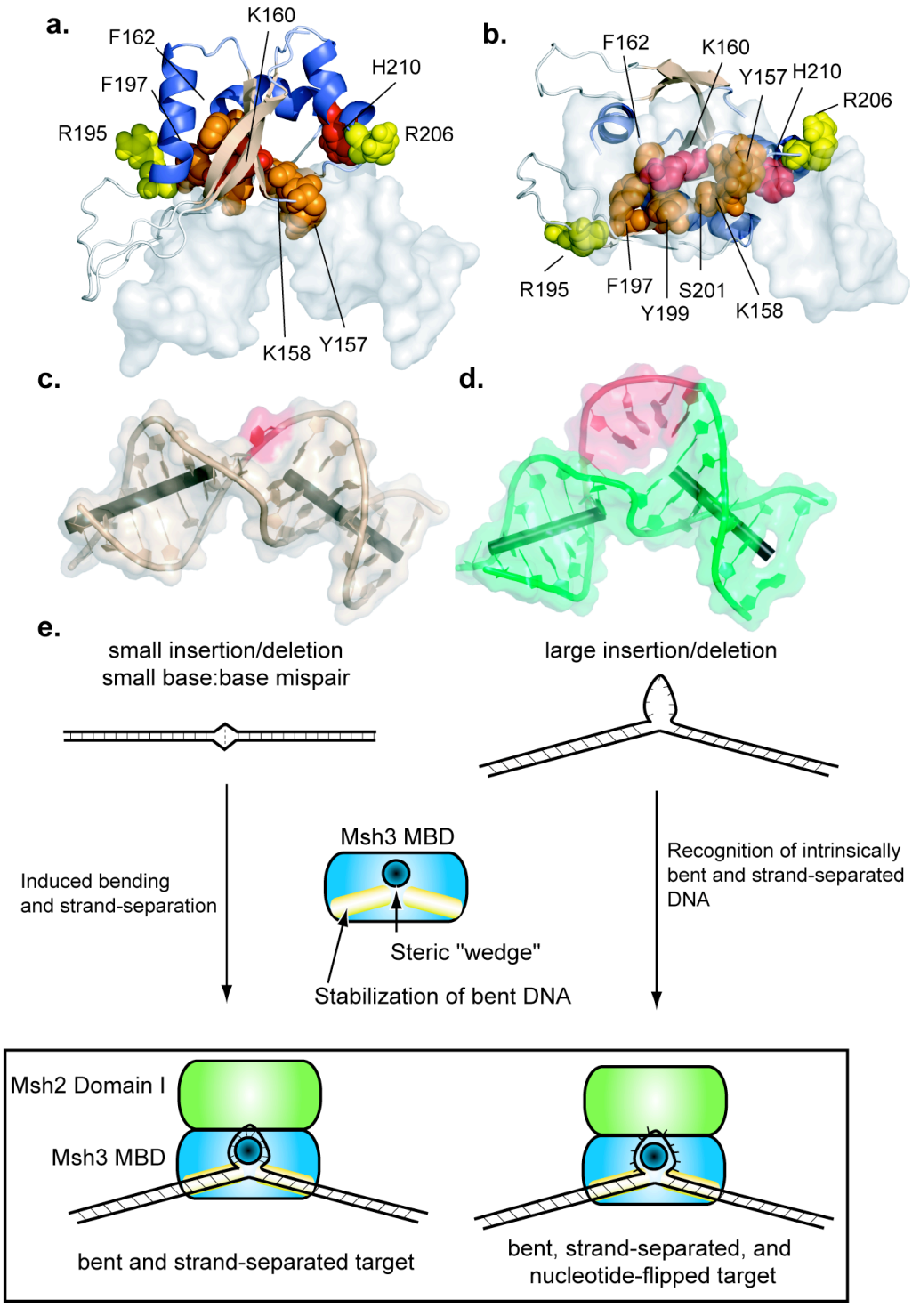


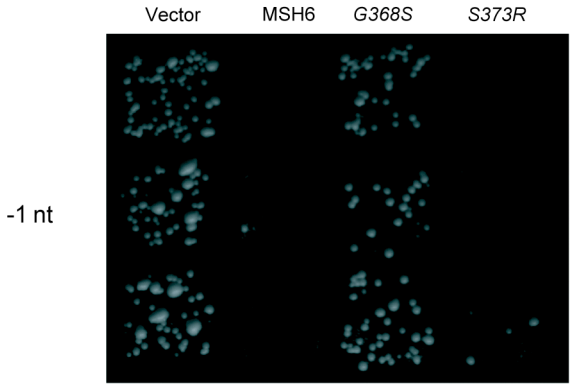
Figure 3.3 Supression of the *msh3Δ* phenotype by alternate amino acid substitutions in *msh3* mutant alleles in MMR assays

(a) Patches of *msh3Δ msh6Δ* strains expressing *msh3* alleles were replica plated onto –threonine plates for the -1 nucleotide *hom3-10* reversion assay. Patches of *msh3Δ msh6Δ* strains expressing *msh3* alleles and containing a microsatellite plasmid with an in frame 4 nucleotide repeat sequence upstream of the *URA3* gene were replica plated onto –leucine –tryptophan +uracil +5-fluoroorotic acid plates as shown. (b) Mutation rates caused by *msh3* mutant alleles in the frameshift (open bars) and 4 nucleotide microsatellite assays (closed bars).

Figure 3.4 Differential effect of *msh3* mutant alleles in frameshift repair versus microsatellite stability assays

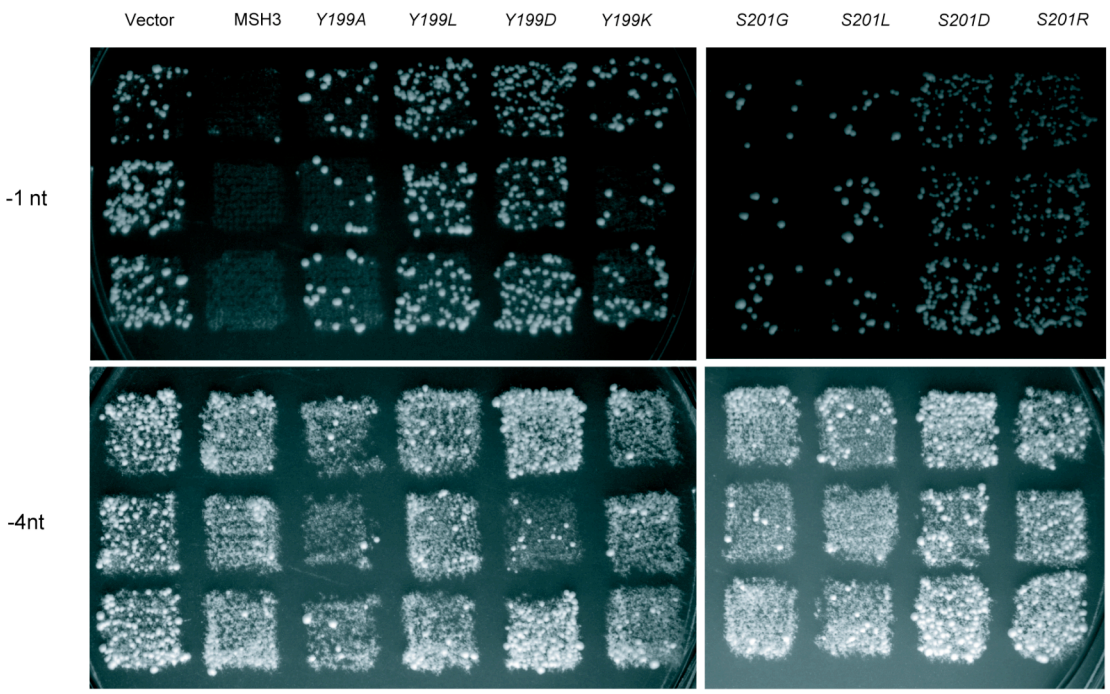
(a,b) Mutations mapped onto the model of Msh3 MBD placed on G:T-containing DNA from the Msh2-Msh6 crystal structure (white) (PDB id 2o8b)(Warren et al., 2007). Red residues correspond to positions that when mutated cause relative defects that are similar in all MMR assays. Yellow residues are positions that cause more severe defects in the frameshift reversion assay than the microsatellite stability assay. Orange residues are positions that, depending on the specific amino acid substitution, can cause equivalent defects in all MMR assays or greater defects in the frameshift assay than the microsatellite stability assay. (c) Structure of a DNA containing a G:T mismatch whose bend is induced by Msh2-Msh6 binding (PDB id 2o8b)(Warren et al., 2007). *red,* T mispair. (d) Structure of an intrinsically bent DNA containing a +5A insertion (red) (PDB id 1qsk)(Dornberger et al., 1999). Molecular images generated by PyMOL (DeLano, 2002). (e) Model of the Msh3 MBD binding to intrinisically bent DNA containing large insertions or inducing and stabilizing non-bent DNA containing small DNA insertions. Recognition likely involves a steric wedge inserting between the DNA strands and stabilization of the DNA bend.


a.
F162
K160
F197
H210
R195
R206
K158
Y157
b.
F162
K160
Y157
H210
R206
R195
F197
S201
Y199
K158
c.
d.
e.
small insertion/deletion
small base:base mispair
large insertion/deletion
Induced bending
and strand-separation
Msh3 MBD
Steric "wedge"
Stabilization of bent DNA
Recognition of intrinsically
bent and strand-separated
DNA
Msh2 Domain I
Msh3 MBD
bent and strand-separated target
bent, strand-separated, and
nucleotide-flipped target



Supplementary Figure 3.1 Phenotype caused by *msh6* mutant alleles in the *hom3-10* reversion

The *msh6* alleles were expressed on a low copy-number plasmid bearing the original promoter sequence and a marker allowing growth on –leucine media. Plasmids were transformed into the *msh3Δ msh6Δ* strain and isolates were patched onto –leucine plates, then replica plated onto –threonine plates as shown for the -1 nucleotide *hom3-10* reversion assay.



Supplementary Figure 3.2 Phenotype caused by *msh3* mutant alleles in the MMR assays

The *msh3* alleles were expressed on a low copy-number plasmid bearing the original promoter sequence and a marker allowing growth on –leucine media. Plasmids were transformed into the *msh3Δ msh6Δ* strain and isolates were patched onto –leucine plates, then replica plated onto –threonine plates as shown for the -1 nucleotide *hom3-10* reversion assay. For the microsatellite stability assay, a plasmid containing a 4 nucleotide repeat sequence inserted in frame and prior to the *URA3* gene was transformed into the *msh3Δ msh6Δ* strain containing a *msh3* allele on a low-copy number *LEU2* plasmid and patched onto –leucine -tryptophan, then replica plated onto –leucine –tryptophan +uracil +5-fluoroorotic acid plates as shown.

Supplementary Table 3.1 Mutator phenotype caused by *msh3* alleles as observed by patch test in MMR assays

Genotype	frameshift	2 nt loop	4 nt loop
MSH3	wt	wt	wt
vector	null	null	null
msh3-Y157S	null	null	null
msh3-K158D	null	null	null
msh3-K160D	null	null	null
msh3-F162A	null	null	null
msh3-E164A	wt	wt	wt
msh3-R171A	wt	wt	wt
msh3-H174A	wt	wt	wt
msh3-H194E	wt	wt	wt
msh3-R195D	null/partial	wt	partial
msh3-F197A	null/partial	partial	null/partial
msh3-Y199A	partial	wt	wt
msh3-S201G	partial	wt	wt
msh3-R206A	partial	wt	wt
msh3-H210A	null	null	partial
msh3-ERN	null	null	null

wt, indicates a rate similar to wild-type, null, indicates a rate similar to the vector alone. Partial indicates an intermediate phenotype.

Supplementary Table 3.2 Mutation rate caused by *msh3* alleles in the *hom3-10* frameshift reversion and 4 nucleotide microsatellite staility assays

Genotype	Frameshift		4 bp unit	
	Mutation Rate	relative rate	Mutaton Rate	relative rate
MSH3	2.1 [0.9-4.5] x 10^{-8}	1	9.5 [4.5-24.8] x 10^{-6}	1
vector	9.6 [5.2-13.3] x 10^{-6}	457	2.5 [0.8-8.5] x 10^{-4}	26
msh3-Y157S	8.7 [6.1-16.1] x 10^{-6}	414	4.2 [2.0-10.0] x 10^{-4}	44
msh3-Y157F	3.5 [1.5-7.0] x 10^{-6}	166	1.8 [1.2-4.0] x 10^{-5}	1.9
msh3-Y157L	7.1 [5.3-10.0] x 10^{-6}	338	1.1 [0.7-2.1] x 10^{-5}	1.2
msh3-Y157D	2.3 [0.9-3.5] x 10^{-5}	1095	1.2 [0.6-7.0] x 10^{-4}	13
msh3-Y157A	2.1 [1.4-3.7] x 10^{-6}	100	1.8 [0.4-4.3] x 10^{-5}	1.9
msh3-K158D	1.6 [1.1-2.5] x 10^{-5}	761	2.5 [1.8-6.0] x 10^{-4}	26
msh3-K158A	1.4 [0.4-2.8] x 10^{-6}	66	2.0 [1.5-4.6] x 10^{-5}	2.1
msh3-K158R	3.2 [1.2-7.3] x 10^{-8}	1.5	2.1 [0.7-3.3] x 10^{-5}	2.2
msh3-K158E	2.2 [1.4-3.6] x 10^{-5}	1047	8.4 [2.5-18.1] x 10^{-4}	88
msh3-K158M	8.3 [5.1-13.1] x 10^{-7}	39	1.7 [1.1-2.9] x 10^{-5}	1.8
msh3-Y157A K158A	1.9 [0.9-2.9] x 10^{-6}	90	4.3 [2.3-13.0] x 10^{-6}	0.5
msh3-K160D	null		null	
msh3-F162A	9.9 [7.5-22.2] x 10^{-6}	471	3.1 [1.1-4.1] x 10^{-4}	33
msh3-F162Y	3.8 [2.5-10.1] x 10^{-7}	18	1.9 [0.9-4.7] x 10^{-5}	2
msh3-F162S	7.6 [4.0-15.7] x 10^{-6}	362	2.0 [0.9-5.2] x 10^{-4}	21
msh3-E164A	wt		wt	
msh3-R171A	wt		wt	
msh3-H174A	wt		wt	
msh3-H194E	wt		wt	
msh3-R195D	4.1 [1.7-7.9] x 10^{-6}	195	2.4 [0.9-5.9] x 10^{-5}	2.5
msh3-F197A	5.0 [3.6-8.4] x 10^{-6}	238	1.8 [0.6-3.2] x 10^{-4}	19
msh3-F197H	2.4 [1.5-6.6] x 10^{-6}	114	4.5 [3.2-5.3] x 10^{-5}	4.7
msh3-Y199A	2.5 [2.1-6.3] x 10^{-6}	119	3.9 [1.0-8.7] x 10^{-5}	4.1
msh3-Y199L	null		partial	
msh3-Y199D	null		null	
msh3-Y199K	partial		partial	
msh3-S201G	8.2 [5.6-11.4] x 10^{-7}	39	1.9 [0.8-3.4] x 10^{-5}	2
msh3-S201L	partial		partial	
msh3-S201D	null		null	
msh3-S201R	null		null	
msh3-R206A	5.2 [1.7-11.3] x 10^{-6}	247	2.0 [1.3-3.3] x 10^{-5}	2.1
msh3-H210A	null		partial	
msh3-ERN	null		null	
Msh6	wt		null	
vector	null		null	
msh6-G368S	null		null	
msh6-S373R	wt		null	

Rate is shown quantitatively if measured, or qualitatively if observed by patch test. The increase relative to the value of the wild-type is shown. Ninety-five percent confidence intervals are shown in brackets. Fourteen independent isolates were used to calculate each value.

Supplementary Table 3.3 Oligonucleotides used to create *msh3* and *msh6* mutant alleles

Msh3 codon Position	Oligo Name	Sequence
Y157S	JH145	AGAGATAAAGTGCTTGTTATTAGAGTAGGCagcAAGTACAAATGTTTTGCAGAGGATGCAGT
	JH146	ACTGCATCCTCTGCAAAACATTTGTACTTgctGCCTACTCTAATAACAAGCACTTTATCTCT
Y157F	JH165	AGAGATAAAGTGCTTGTTATTAGAGTAGGCtttAAGTACAAATGTTTTGCAGAGGATGCAGT
	JH166	ACTGCATCCTCTGCAAAACATTTGTACTTaaaGCCTACTCTAATAACAAGCACTTTATCTCT
Y157L	JH167	AGAGATAAAGTGCTTGTTATTAGAGTAGGCcttAAGTACAAATGTTTTGCAGAGGATGCAGT
	JH168	ACTGCATCCTCTGCAAAACATTTGTACTTaagGCCTACTCTAATAACAAGCACTTTATCTCT
Y157D	JH169	AGAGATAAAGTGCTTGTTATTAGAGTAGGCgatAAGTACAAATGTTTTGCAGAGGATGCAGT
	JH170	ACTGCATCCTCTGCAAAACATTTGTACTTatcGCCTACTCTAATAACAAGCACTTTATCTCT
Y157A	JH171	AGAGATAAAGTGCTTGTTATTAGAGTAGGCgcaAAGTACAAATGTTTTGCAGAGGATGCAGT
	JH172	ACTGCATCCTCTGCAAAACATTTGTACTTtgcGCCTACTCTAATAACAAGCACTTTATCTCT
K158D	JH147	GATAAAGTGCTTGTTATTAGAGTAGGCTACgatTACAAATGTTTTGCAGAGGATGCAGTAAC
	JH148	GTTACTGCATCCTCTGCAAAACATTTGTAatcGTAGCCTACTCTAATAACAAGCACTTTATC
K158A	JH173	GATAAAGTGCTTGTTATTAGAGTAGGCTACgcaTACAAATGTTTTGCAGAGGATGCAGTAAC
	JH174	GTTACTGCATCCTCTGCAAAACATTTGTAtgcGTAGCCTACTCTAATAACAAGCACTTTATC
K158R	JH175	GATAAAGTGCTTGTTATTAGAGTAGGCTACagaTACAAATGTTTTGCAGAGGATGCAGTAAC
	JH176	GTTACTGCATCCTCTGCAAAACATTTGTAtctGTAGCCTACTCTAATAACAAGCACTTTATC
K158E	JH177	GATAAAGTGCTTGTTATTAGAGTAGGCTACgagTACAAATGTTTTGCAGAGGATGCAGTAAC
	JH178	GTTACTGCATCCTCTGCAAAACATTTGTActcGTAGCCTACTCTAATAACAAGCACTTTATC
K158M	JH179	GATAAAGTGCTTGTTATTAGAGTAGGCTACatgTACAAATGTTTTGCAGAGGATGCAGTAAC
	JH180	GTTACTGCATCCTCTGCAAAACATTTGTAcatGTAGCCTACTCTAATAACAAGCACTTTATC
Y157A K158A	JH200	AGAGATAAAGTGCTTGTTATTAGAGTAGGCgcagcaTACAAATGTTTTGCAGAGGATGCAGTAAC
	JH201	GTTACTGCATCCTCTGCAAAACATTTGTAtgctgcGCCTACTCTAATAACAAGCACTTTATCTCT
K150D	JH149	GTGCTTGTTATTAGAGTAGGCTACAAGTACgatTGTTTTGCAGAGGATGCAGTAACGGTTAGC
	JH150	GCTAACCGTTACTGCATCCTCTGCAAAACAatcGTACTTGTAGCCTACTCTAATAACAAGCAC
F162A	JH129	GTTATTAGAGTAGGCTACAAGTACAAATGTgcaGCAGAGGATGCAGTAACGGTTAGCAGAATA
	JH130	TATTCTGCTAACCGTTACTGCATCCTCTGCtgcACATTTGTACTTGTAGCCTACTCTAATAAC

Supplementary Table 3.3 (continued) Oligonucleotides used to create *msh3* and *msh6* mutant alleles

F162Y	JH181	GTTATTAGAGTAGGCTACAAGTACAAATGTtacGCAGAGGATGCAGTAACGGTTAGCAGAATA
	JH182	TATTCTGCTAACCGTTACTGCATCCTCTGCgtaACATTTGTACTTGTAGCCTACTCTAATAAC
F162S	JH183	GTTATTAGAGTAGGCTACAAGTACAAATGTagcGCAGAGGATGCAGTAACGGTTAGCAGAATA
	JH184	TATTCTGCTAACCGTTACTGCATCCTCTGCgctACATTTGTACTTGTAGCCTACTCTAATAAC
E164A	JH131	AGAGTAGGCTACAAGTACAAATGTTTTGCAgcaGATGCAGTAACGGTTAGCAGAATACTTCAC
	JH132	GTGAAGTATTCTGCTAACCGTTACTGCATCtgcTGCAAAACATTTGTACTTGTAGCCTACTCT
R171A	JH133	TGTTTTGCAGAGGATGCAGTAACGGTTAGCgcaATACTTCACATCAAACTTGTGCCTGGAAAA
	JH134	TTTTCCAGGCACAAGTTTGATGTGAAGTATtgcGCTAACCGTTACTGCATCCTCTGCAAAACA
H174A	JH141	GAGGATGCAGTAACGGTTAGCAGAATACTTgcaATCAAACTTGTGCCTGGAAAATTGACTATC
	JH142	GATAGTCAATTTTCCAGGCACAAGTTTGATtgcAAGTATTCTGCTAACCGTTACTGCATCCTC
H194E	JH155	ATCGATGAGTCTAATCCTCAAGATTGCAATgagAGGCAGTTTGCGTACTGTTCTTTCCCGGAT
	JH156	ATCCGGGAAAGAACAGTACGCAAACTGCCTctcATTGCAATCTTGAGGATTAGACTCATCGAT
R195D	JH157	GATGAGTCTAATCCTCAAGATTGCAATCATgatCAGTTTGCGTACTGTTCTTTCCCGGATGTC
	JH158	GACATCCGGGAAAGAACAGTACGCAAACTGatcATGATTGCAATCTTGAGGATTAGACTCATC
F197A	JH137	TCTAATCCTCAAGATTGCAATCATAGGCAGgcaGCGTACTGTTCTTTCCCGGATGTCAGATTA
	JH138	TAATCTGACATCCGGGAAAGAACAGTACGCtgcCTGCCTATGATTGCAATCTTGAGGATTAGA
F197H	JH185	TCTAATCCTCAAGATTGCAATCATAGGCAGcatGCGTACTGTTCTTTCCCGGATGTCAGATTA
	JH186	TAATCTGACATCCGGGAAAGAACAGTACGCatgCTGCCTATGATTGCAATCTTGAGGATTAGA
Y199A	JH143	CCTCAAGATTGCAATCATAGGCAGTTTGCGgcaTGTTCTTTCCCGGATGTCAGATTAAACGTT
	JH144	AACGTTTAATCTGACATCCGGGAAAGAACAtgcCGCAAACTGCCTATGATTGCAATCTTGAGG
Y199L	JH219	CCTCAAGATTGCAATCATAGGCAGTTTGCGttaTGTTCTTTCCCGGATGTCAGATTAAACGTT
	JH220	AACGTTTAATCTGACATCCGGGAAAGAACAtaaCGCAAACTGCCTATGATTGCAATCTTGAGG
Y199D	JH221	CCTCAAGATTGCAATCATAGGCAGTTTGCGgatTGTTCTTTCCCGGATGTCAGATTAAACGTT
	JH222	AACGTTTAATCTGACATCCGGGAAAGAACAatcCGCAAACTGCCTATGATTGCAATCTTGAGG
Y199K	JH223	CCTCAAGATTGCAATCATAGGCAGTTTGCGaaaTGTTCTTTCCCGGATGTCAGATTAAACGTT
	JH224	AACGTTTAATCTGACATCCGGGAAAGAACAtttCGCAAACTGCCTATGATTGCAATCTTGAGG
S201G	JH135	GATTGCAATCATAGGCAGTTTGCGTACTGTggtTTCCCGGATGTCAGATTAAACGTTCACCTA
	JH136	TAGGTGAACGTTTAATCTGACATCCGGGAAaccACAGTACGCAAACTGCCTATGATTGCAATC
S201L	JH229	GATTGCAATCATAGGCAGTTTGCGTACTGTttaTTCCCGGATGTCAGATTAAACGTTCACCTA
	JH230	TAGGTGAACGTTTAATCTGACATCCGGGAAtaaACAGTACGCAAACTGCCTATGATTGCAATC

Supplementary Table 3.3 (continued) Oligonucleotides used to create *msh3* and *msh6* mutant alleles

S201D	JH225	GATTGCAATCATAGGCAGTTTGCGTACTGTgatTTCCCGGATGTCAGATTAAACGTTCACCTA
	JH226	TAGGTGAACGTTTAATCTGACATCCGGGAAatcACAGTACGCAAACTGCCTATGATTGCAATC
S201R	JH227	GATTGCAATCATAGGCAGTTTGCGTACTGTaggTTCCCGGATGTCAGATTAAACGTTCACCTA
	JH228	TAGGTGAACGTTTAATCTGACATCCGGGAAcctACAGTACGCAAACTGCCTATGATTGCAATC
R206A	JH151	CAGTTTGCGTACTGTTCTTTCCCGGATGTCgcgTTAAACGTTCACCTAGAGAGACTTGTGCAT
	JH152	ATGCACAAGTCTCTCTAGGTGAACGTTTAAcgcGACATCCGGGAAAGAACAGTACGCAAACTG
H210A	JH153	TGTTCTTTCCCGGATGTCAGATTAAACGTTgcgCTAGAGAGACTTGTGCATCATAATTTAAAG
	JH154	CTTTAAATTATGATGCACAAGTCTCTCTAGcgcAACGTTTAATCTGACATCCGGGAAAGAACA
ERN	JH163	AGCAGAATACTTCACATCAAACTTGTGCCTgagagaaatTTTGCGTACTGTTCTTTCCCGGATGTCAGA
	JH164	TCTGACATCCGGGAAAGAACAGTACGCAAAatttctctcAGGCACAAGTTTGATGTGAAGTATTCTGCT
msh6-G368S	JH189	GGTGGAGGACGCGCTAATATGCAACTAGCTtcaATTCCAGAGATGTCATTTGAATATTGGGCC
	JH190	GGCCCAATATTCAAATGACATCTCTGGAATtgaAGCTAGTTGCATATTAGCGCGTCCTCCACC
msh6-S373R	JH191	AATATGCAACTAGCTGGGATTCCAGAGATGcgcTTTGAATATTGGGCCGCTCAGTTTATCCAA
	JH192	TTGGATAAACTGAGCGGCCCAATATTCAAAgcgCATCTCTGGAATCCCAGCTAGTTGCATATT

Sequences are shown in the 5’ to 3’ direction. The codon substitution is shown in lower case letters.

The text of Chapter 3, in full, is in preparation for submission to *Nat. Struct. Mol. Biol*, 2009, Dowen, JM, Putnam CD, Kolodner, RD. The dissertation author was the primary researcher and author of this paper. The co-authors listed in this publication directed and supervised the research that forms the basis of this chapter.

Chapter 4: Conclusions and future directions

The work presented here has lead to a new understanding of the role of Msh2-Msh3 in MMR. Genetic and biochemical approaches have shown that the eukaryotic Msh2-Msh3 heterodimer is able to recognize and repair base:base mismatches in DNA, a previously unknown function. Our subsequent studies have revealed the mechanism of mismatch recognition for this new class of base:base mismatch substrates and the previously identified class of insertion/deletion substrates of Msh2-Msh3.

In the case of large insertion/deletions in DNA, Msh2-Msh3 recognizes the unpaired bases in the loop and the intrinsically bent structure of the surrounding DNA which is due to the extra sequence on one strand. In the case of small insertion/deletions and base:base mispairs, Msh2-Msh3 must make additional contacts with the DNA that force the DNA sequence into a bent conformation for recognition. By studying the recognition process of many kinds of DNA mispairs, rather than the recognition of only an ideal substrate, this work has shed light on the flexible and dynamic nature of these kinds of DNA-binding proteins. Therefore, this work has lead to a better understanding of the complete scope of what MMR proteins are capable of doing in the cell, a previously underappreciated aspect of MMR.

In the future, it will be of great interest to solve the crystal structure of Msh2-Msh3 bound to various DNA substrates including a base:base mispair, a small insertion/deletion of 1 or 2 nucleotides, a larger insertion/deletion of 8 nt

a hairpin and a 3'-flap containing substrate. We have shown, by functional analysis using *in vivo* MMR assays, that we have identified many residues of the Msh3-MBD that mediate interactions with these various DNA substrates. The critical test of our hypothesis of mismatch recognition by Msh2-Msh3, will be to determine the atomic resolution interactions between the Msh3 MBD and the mispair containing DNA as well as measure the biophysical forces of DNA bending applied to these various DNA mispairs.

In addition to studying the lesion specificity of Msh2-Msh3 in mismatch repair, the approaches we have used here could also be applied to the study of double strand break repair (DSBR). During the repair of a double strand break the DNA on each side of the break undergoes resection in the 5' to 3' direction, creating long 3' single stranded tails (Paques and Haber, 1999). In gene conversion, these tails invade a homologous sequence and serve as primers for new DNA synthesis. In single strand annealing, these tails anneal to each other through regions of homology resulting in deletion of the intervening sequence. Msh2-Msh3 has been shown to act with Rad1-Rad10 to trim off nonhomologous portions of these tails leaving a homologous portion capable of repair (Sugawara et al., 1997).

It has been suggested that Msh2-Msh3 recognizes the branched structure of these recombination intermediates and, upon binding, is able to stabilize them for Rad1-Rad10 cleavage (Paques and Haber, 1999; Sugawara et al., 1997). It is also possible that Msh2-Msh3 may recruit Rad1-Rad10 to

these structures through a direct interaction (Bertrand et al., 1998). It would be interesting to test the *msh3* mutants identified in this study for their ability to perform DSBR. This work could determine whether branched DNA structures are recognized by Msh2-Msh3 in the same way as large insertion/deletion mispairs.

A broader and more challenging question raised by this work is the relationship of Msh2-Msh3 with the two MutL complexes, Mlh1-Pms1 and Mlh1-Mlh3. It is of great interest to determine whether the lesion specificity determines which MutL complex interacts with Msh2-Msh3. It is possible that Msh2-Msh3 when complexed with a C:C mispair is bound by Mlh1-Pms1 and Msh2-Msh3 when complexed with a +8 nucleotide loop is bound by Mlh1-Mlh3. The differential recruitment of one MutL heterodimer over the other could be dependent on a subtle conformational difference in Msh2-Msh3 due to the lesion size.

What is the nature of the two possible ternary complexes: Msh2-Msh3-Mlh1-Pms1 and Msh2-Msh3-Mlh1-Mlh3? There are many unanswered questions regarding the overlapping and unique roles of these two ternary complexes in the cell including recruitment to specific sites in the genome, function in IgG class switch recombination and somatic hypermutation, capacity for movement from the mispair site, tissue and temporal expression patterns, and ultimate repair outcomes. Further studies are needed to fill these gaps in our understanding of the basic mechanisms of MMR

Methods

General methods and strains

All media including dropout media and canavanine-containing dropout media have been previously described (Alani et al., 1994; Amin et al., 2001; Reenan and Kolodner, 1992). All strains used in this study were derivatives of the S288c strain RDKY3686 *MATα, ura3-52, leu2-1, trp1-63, hom3-10, his3-200, lys2-10A* (Amin et al., 2001). The relevant genotypes of these strains are as follows: RDKY4149 *msh3::hisG*, RDKY4151 *msh6::hisG*, RDKY5295 *mlh3::HIS3 and* RDKY4237 *mlh1::hisG*. The protease deficient strain RDKY2418 *MATα, ura3-52, leu2-1, his3-200, pep4::HIS3, prb1-1.6R, can1, msh2::hisG, msh6::hisG* was used to overexpress proteins for purification (Hess et al., 2002). Genetic complementation of *MSH3* derivatives was measured in yeast strain RDKY4234 *MATα, ura3-52, leu2-1, trp1-63, hom3-10, his3-200, lys2-10A, msh3::hisG, msh6::hisG*. Mutation rates were determined by fluctuation analysis using at least 14 independent colonies from each strain as previously described (Alani et al., 1994; Amin et al., 2001; Das Gupta and Kolodner, 2000; Reenan and Kolodner, 1992).

Plasmid construction

Site-directed mutagenesis of a wild-type *MSH3* low copy-number, *LEU2* plasmid (Shell et al., 2007a) was performed to generate mutations affecting the Msh3 MBD using primers listed in Supplemental Figure 3.3. The *msh3*

mutant plasmids were sequenced to confirm that only the desired mutation was present. All DNA sequencing was performed by using an Applied Biosystems 3730XL DNA sequencer and standard chemistry. Sequence analysis was performed using Sequencher 4.2.2 (Gene Codes, Ann Arbor, MI).

Genetic complementation

Site-directed mutagenesis of a wild-type *MSH3* low copy-number, LEU2 plasmid was performed to mutate the Met codon at position 1 to Ala (M1A) or the Met codon at position 30 to Ala (M30A). Primers to create the *msh3-M1A* allele were: JH67 5'-AATTTTGACAAAGCCAATTTGAACTCCAAAGCTGCCCCAGCTACCCCTAAACTTCTAAGACT with JH68 5'-AGTCTTAGAAGTTTTAGGGGTAGCTGGGGCAGCTTTGGAGTTCAAATTGGCTTTGTCAAAATT. Primers to create the *msh3-M30A* allele were: JH69 5'-GAAAATGGCTCCACATCTTCTCAAAAGAAAGCTAAGCAATCGAGTTTGTTATCTTTTTTCTCA with JH70 5'-TGAGAAAAAAGATAACAAACTCGATTGCTTAGCTTTCTTTTGAGAAGATGTGGAGCCATTTTC. The *msh3* mutant plasmids were sequenced to confirm that only the desired mutation was present. Plasmids were then transformed into the strain RDKY4234 and transformants were patched onto –leucine media to maintain plasmid selection. Patches were then replica plated onto –lysine and –threonine plates and grown at 30^{o}C for 2 days to select for *lys2-*

10A and *hom3-10* revertants so as to visualize the mutator phenotype of the different plasmid containing strains.

Canavanine mutation analysis

Strains of interest were first streaked for single colonies on YPD plates and then individual colonies were patched onto YPD plates. The patches were replica plated onto –arginine +canavanine selective media and canavanine resistant mutants were allowed to grow at 30 ^{o}C for 2 days. Mutation spectra were analyzed by isolating chromosomal DNA from one Can^{r} mutant per patch, amplifying the *CAN1* gene by PCR and sequencing to determine the inactivating mutation in the *CAN1* gene (Das Gupta and Kolodner, 2000; Flores-Rozas and Kolodner, 1998; Marsischky et al., 1996). The PCR primer pair used for amplification of *CAN1* was CAN1FX 5'-GTTGGATCCAGTTTTTAATCTGTCGTC and CAN1RX 5'-TTCGGTGTATGACTTATGAGGGTG. The three primers used for sequencing *CAN1* were CAN1G 5'-CAGTGGAACTTTGTACGTCC, CANSEQ3 5'-TTCTGTCACGCAGTCCTTGG and CANSEQ5 5'-AACTAGTTGGTATCACTGCT**.** All DNA sequencing was performed by using an Applied Biosystems 3730XL DNA sequencer and standard chemistry. Sequence analysis was performed using Sequencher 4.2.2 (Gene Codes, Ann Arbor, MI).

Frameshift and microsatellite stability assays

Patches grown on –leucine plates from RDKY4234 containing various plasmid-borne *msh3* alleles were replica plated onto –threonine plates and grown at 30^{o}C for 2 days to select for *hom3-10* revertants. The microsatellite instability assay was performed by transforming a microsatellite containing plasmid into the RDKY4234 strain containing a plasmid-borne *msh3* allele. The microsatellite plasmid had a *TRP1* selectable marker and contained the microsatellite repeats sequences $(GT)_{16.5}$ or $(CAGT)_{16}$ for 2 and 4 nucleotide repeats, respectively, in frame and prior to the *URA3* gene (Sia et al., 1997). Strains with plasmids for both the *msh3* allele and the microsatellite assay were grown in patches on –leucine -tryptophan plates and then replica plated onto –leucine –tryptophan +uracil +5-fluoroorotic acid plates and grown at 30^{o}C for 2 to 3 days. Quantitative mutation rates were determined by fluctuation analysis using at least 14 independent colonies from each strain as previously described (Alani et al., 1994; Amin et al., 2001; Das Gupta and Kolodner, 2000; Reenan and Kolodner, 1992).

Statistical analysis

The significance of the observed overlap between the *CAN1* base substitution mutation spectra in different strains was calculated with a Monte Carlo technique. Since the observed base substitution mutations were unlikely to be saturating, the total number of readily mutable *CAN1*-inactivating mutation sites was estimated by fitting the observed distribution of singly and

multiply observed base substitution mutations to a theoretical Poisson distribution. We minimized the root-mean-square error between the expected and observed number of singly and multiply observed mutations using the equation

$$\varepsilon = \sqrt{\frac{\sum_{i=1}^{p} \left(Npois(i,\lambda) - a_i\right)^2}{p}}$$

where p is the maximum number of events for any single mutation, a_i is the number of mutations observed i times, and $pois(i,\lambda)$ is the probability with the parameter $\lambda = C/N$, with C being the number of observed events and N being the total number of possible mutation sites, defined by both the position and the base substitution at that position. Minimization of ε by varying N in the range [1,600] gives the total number of mutations, including those not observed in experimental sampling. By using the Poisson distribution, we assumed that all observed base substitution mutations within a strain occur with equal efficiency and that mutations in multiple isolates are independent of each other. For the wild-type and *msh3* strains, the best fitted Poisson curves used values of N=159 and N=259, respectively.

Using the total number of inactivating mutations, the results for two different models were calculated. In model 1, the readily mutated *CAN1*-inactivating mutations are identical in both strains, and each strain was

allowed to accumulate mutations at any of the 259 mutation sites or 159 mutation sites (the predicted number of mutations in the *msh3* and wild-type strains, respectively). In model 2, the mutational spectra were treated as overlapping, but distinct in that the wild-type strain was only allowed to accumulate mutations at 159 mutation sites of the 259 mutation sites for the *msh3* strain. Mutations were then randomly selected for both strains using each of the two models. The total number of randomly chosen mutations was equal to the number of observed mutations in each strain. The overlap of these theoretical distributions of mutation sites was then calculated. We repeated this process 50,000 times and used a Z-score test to calculate the significance of the observed overlaps using the null hypothesis that differences in overlap in base substitution mutation spectra between wild-type and *msh3* strains was due to sampling and not due to differences in the specificity of mutation accumulation.

The two-tailed Mann-Whitney test, chi squared "goodness of fit" and the Fisher Exact probability test were performed on the VassarStats website: http://faculty.vassar.edu/lowry/VassarStats.html

Overexpression and purification of Msh2-Msh3 complex

The *S. cerevisiae* Msh2-Msh3 heterodimer was coexpressed from *GAL10* promoter plasmids in the protease deficient yeast strain RDKY2418. The Msh2 expression vector contains a *GAL10* promoter fused to the *MSH2* gene on a 2μ *URA3*, Amp^r plasmid. The Msh3 expression vector contains a

GAL10 promoter fused to the *MSH3* FLAG tagged gene on a 2μm *LEU2*, Amp^r plasmid. The Msh3 expression vector fuses the *GAL10* promoter to the methionine at amino acid position 30 according to the *Saccharomyces* Genome Database coding sequence (http://www.yeastgenome.org/) using the leader sequence AAGGAGATATACATatg and contains a C-terminal FLAG-tag sequence cacGACTACAAGGACGACGATGACAAGtga where the last codon of *MSH3* (cac 1047) is shown in lower case and the FLAG codons are shown in upper case followed by the stop codon; genetic complementation studies showed that this FLAG tag did not affect the biological function of *MSH3* (data not shown). A fermentor was used to grow 10 L of cells in synthetic dropout media lacking uracil and leucine and containing 2% raffinose to an OD of 0.8 at 30°C. The expression of Msh2 and Msh3 was then induced by adding galactose to a final concentration of 2% for 8 hours. The cells were harvested by centrifugation and the resulting cell pellet was resuspended in lysis buffer (500 mM NaCl, 50 mM Tris-HCl, ph 8.0, 1 mM EDTA, 5 mM DTT, 10% glycerol, 1 mM phenyl-methyl-sulfonyl fluoride, leupeptin, benzamidine, pepstatin) and lysed with glass beads (Sigma) in a beadbeater (Biospec Products, Inc., Bartlesville, OK). Msh2-Msh3 heterodimer was purified by sequential chromatography on a 30 ml Polybuffer Exchanger 94 resin, 10 ml High Trap Q, 5 ml Heparin, 10 ml DNA cellulose, 1 ml SP-sepherose, and 1 ml DEAE columns. Fractions were either frozen directly in liquid N_2 and stored at -80°C or concentrated by centrifugation in a Centricon YM30 (Millipore,

Billerica, MA) and then frozen. Protein concentrations were determined by comparison to known protein concentrations on a coomassie stained gel. The yield from 10 L of cells was 15 μg of Msh2-Msh3 protein. Additionally, the purified protein was digested with trypsin and subjected to mass spectrometry to confirm its identity. Msh2-Msh6 was provided by Dr. Dan Mazur (Mazur et al., 2006).

DNA substrates

Oligonucleotides were synthesized by Midland Certified Reagent Company (Midland, TX). Double stranded DNA substrates were constructed by annealing 38 basepair complementary oligonucleotides at 95 ^{o}C for 5 min in annealing buffer (0.5M NaCl, 10mM Tris-Hcl, ph7.5, 1mM EDTA) followed by slowly cooling over 2 hours. DNA duplexes were purified by high pressure liquid chromatography using a Waters GEN-PAK FAX column (Marsischky and Kolodner, 1999). The sequences of the different oligonucleotides and double stranded DNA substrates are presented in Supplementary Table 2.

***In vitro* DNA binding experiments**

Purified DNA substrates were 5'-end-labeled using [γ-32]ATP and T4 polynucleotide kinase and purified by centrifugation through mini Quick Spin Oligo Columns (Roche, Indianapolis, IN). DNA binding assays were performed by combining 16 nM protein (Msh2-Msh6 or Msh2-Msh3 heterodimer) with 14 nM ^{32}P-labeled substrate in a final volume of 10 μl

Binding Buffer (20 mM Tris-HCl, ph 8.0, 110 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10 μM ADP, 5% glycerol, 70 nM unlabeled GC homoduplex, 100 mM bovine serum albumin). Reactions were incubated on ice for 15 min and then 500 μM ATP was added as indicated in individual experiments for 15 min before loading dye was added. Gel electrophoresis of the samples was performed on a 4-20% gradient TBE Criterion gel (BioRad, Hercules, CA) run in 0.5X TBE (45 mM Tris borate, 1 mM EDTA, ph 8.0), 5% glycerol for 3 hours at 150 V at 4°C. Gels were then soaked for 1 hour in 40% MeOH, 10% Acetic Acid, 5% glycerol before being dried and analyzed using a PhosphorImager and ImageQuant software (Molecular Dynamics, Sunnyvale, CA).

Molecular modeling

An initial homology model for the *S. cerevisiae* Msh3 MBD (residues 133-255) using the human Msh6 MBD (PDB id 2o8b)(Warren et al., 2007) using SWISS-MODEL (Schwede et al., 2003). Two regions of the resulting model were treated as "low-confidence" regions. These regions were residues S230-V244 (corresponding to a three-fold crystal contact between Msh6 MBD domains in the Msh2-Msh6 crystal structure) and residues I175-N193 (corresponding to a 14 amino acid insertion not present in Msh6). Both low-confidence regions were outside of the core recognition region of interest here and were rebuilt manually and refined with CNS (Brunger et al., 1998) to resolve steric problems in the original model built by SWISS-MODEL

References

Acharya, S., Foster, P. L., Brooks, P., and Fishel, R. (2003). The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. Mol Cell *12*, 233-246.

Alani, E. (1996). The Saccharomyces cerevisiae Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. Mol Cell Biol *16*, 5604-5615.

Alani, E., Reenan, R. A., and Kolodner, R. D. (1994). Interaction between mismatch repair and genetic recombination in Saccharomyces cerevisiae. Genetics *137*, 19-39.

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., and Griffith, J. D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. Embo J *16*, 4467-4476.

Amin, N. S., Nguyen, M. N., Oh, S., and Kolodner, R. D. (2001). exo1-Dependent mutator mutations: model system for studying functional interactions in mismatch repair. Mol Cell Biol *21*, 5142-5155.

Antony, E., and Hingorani, M. M. (2003). Mismatch recognition-coupled stabilization of Msh2-Msh6 in an ATP-bound state at the initiation of DNA repair. Biochemistry *42*, 7682-7693.

Antony, E., and Hingorani, M. M. (2004). Asymmetric ATP binding and hydrolysis activity of the Thermus aquaticus MutS dimer is key to modulation of its interactions with mismatched DNA. Biochemistry *43*, 13115-13128.

Au, K. G., Welsh, K., and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. J Biol Chem *267*, 12142-12148.

Bertrand, P., Tishkoff, D. X., Filosi, N., Dasgupta, R., and Kolodner, R. D. (1998). Physical interaction between components of DNA mismatch repair and nucleotide excision repair. Proc Natl Acad Sci U S A *95*, 14278-14283.

Bjornson, K. P., Allen, D. J., and Modrich, P. (2000). Modulation of MutS ATP hydrolysis by DNA cofactors. Biochemistry *39*, 3176-3183.

Bowers, J., Sokolsky, T., Quach, T., and Alani, E. (1999). A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2-MSH6 complex disrupts mismatch recognition. J Biol Chem *274*, 16115-16125.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T., and Modrich, P. (2001). In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. Proc Natl Acad Sci U S A *98*, 6765-6770.

Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S. M., Fischer, F., Cejka, P., and Jiricny, J. (2005). Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. Cancer Res *65*, 10759-10766.

Chen, C., and Kolodner, R. D. (1999). Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. Nat Genet *23*, 81-85.

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

Constantin, N., Dzantiev, L., Kadyrov, F. A., and Modrich, P. (2005). Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. J Biol Chem *280*, 39752-39761.

Das Gupta, R., and Kolodner, R. D. (2000). Novel dominant mutations in Saccharomyces cerevisiae MSH6. Nat Genet *24*, 53-56.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA, USA, DeLano Scientific).

Dornberger, U., Hillisch, A., Gollmick, F. A., Fritzsche, H., and Diekmann, S. (1999). Solution structure of a five-adenine bulge loop within a DNA duplex. Biochemistry *38*, 12860-12868.

Drotschmann, K., Yang, W., Brownewell, F. E., Kool, E. T., and Kunkel, T. A. (2001). Asymmetric recognition of DNA local distortion. Structure-based functional studies of eukaryotic Msh2-Msh6. J Biol Chem *276*, 46225-46229.

Drummond, J. T., Genschel, J., Wolf, E., and Modrich, P. (1997). DHFR/MSH3 amplification in methotrexate-resistant cells alters the hMutSalpha/hMutSbeta ratio and reduces the efficiency of base-base mismatch repair. Proc Natl Acad Sci U S A *94*, 10144-10149.

Dufner, P., Marra, G., Raschle, M., and Jiricny, J. (2000). Mismatch recognition and DNA-dependent stimulation of the ATPase activity of hMutSalpha is abolished by a single mutation in the hMSH6 subunit. J Biol Chem *275*, 36550-36555.

Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., Kneitz, B., Avdievich, E., Fan, K., Wong, E.*, et al.* (2000). The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res *60*, 803-807.

Erdeniz, N., Nguyen, M., Deschenes, S. M., and Liskay, R. M. (2007). Mutations affecting a putative MutLalpha endonuclease motif impact multiple mismatch repair functions. DNA Repair (Amst) *6*, 1463-1470.

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell *75*, 1027-1038.

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

Galio, L., Bouquet, C., and Brooks, P. (1999). ATP hydrolysis-dependent formation of a dynamic ternary nucleoprotein complex with MutS and MutL. Nucleic Acids Res *27*, 2325-2331.

Genschel, J., Littman, S. J., Drummond, J. T., and Modrich, P. (1998). Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. J Biol Chem *273*, 19895-19901.

Gradia, S., Acharya, S., and Fishel, R. (2000). The role of mismatched nucleotides in activating the hMSH2-hMSH6 molecular switch. J Biol Chem *275*, 3922-3930.

Gradia, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J., and Fishel, R. (1999). hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. Mol Cell *3*, 255-261.

Grilley, M., Welsh, K. M., Su, S. S., and Modrich, P. (1989). Isolation and characterization of the Escherichia coli mutL gene product. J Biol Chem *264*, 1000-1004.

Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1996). Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. Curr Biol *6*, 1185-1187.

Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1997). Enhancement of MSH2-MSH3-mediated mismatch recognition by the yeast MLH1-PMS1 complex. Curr Biol *7*, 790-793.

Hall, M. C., and Matson, S. W. (1999). The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. J Biol Chem *274*, 1306-1312.

Harfe, B. D., and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. Annu Rev Genet *34*, 359-399.

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

Harrington, J. M., and Kolodner, R. D. (2007). Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs. Mol Cell Biol *27*, 6546-6554.

Hess, M. T., Gupta, R. D., and Kolodner, R. D. (2002). Dominant Saccharomyces cerevisiae msh6 mutations cause increased mispair binding and decreased dissociation from mispairs by Msh2-Msh6 in the presence of ATP. J Biol Chem *277*, 25545-25553.

Heyer, J., Yang, K., Lipkin, M., Edelmann, W., and Kucherlapati, R. (1999). Mouse models for colorectal cancer. Oncogene *18*, 5325-5333.

Holmes, S. F., Scarpinato, K. D., McCulloch, S. D., Schaaper, R. M., and Kunkel, T. A. (2007). Specialized mismatch repair function of Glu339 in the Phe-X-Glu motif of yeast Msh6. DNA Repair (Amst) *6*, 293-303.

Huang, J., Kuismanen, S. A., Liu, T., Chadwick, R. B., Johnson, C. K., Stevens, M. W., Richards, S. K., Meek, J. E., Gao, X., Wright, F. A.*, et al.* (2001a). MSH6 and MSH3 are rarely involved in genetic predisposition to nonpolypotic colon cancer. Cancer Res *61*, 1619-1623.

Huang, S. C., Lavine, J. E., Boland, P. S., Newbury, R. O., Kolodner, R., Pham, T. T., Arnold, C. N., Boland, C. R., and Carethers, J. M. (2001b). Germline characterization of early-aged onset of hereditary non-polyposis colorectal cancer. J Pediatr *138*, 629-635.

Iyer, R. R., Pluciennik, A., Burdett, V., and Modrich, P. L. (2006). DNA mismatch repair: functions and mechanisms. Chem Rev *106*, 302-323.

Joshi, A., Sen, S., and Rao, B. J. (2000). ATP-hydrolysis-dependent conformational switch modulates the stability of MutS-mismatch complexes. Nucleic Acids Res *28*, 853-861.

Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P., and Yang, W. (2001). Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. Mol Cell *7*, 1-12.

Kadyrov, F. A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. Cell *126*, 297-308.

Kadyrov, F. A., Holmes, S. F., Arana, M. E., Lukianova, O. A., O'Donnell, M., Kunkel, T. A., and Modrich, P. (2007). Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease. J Biol Chem *282*, 37181-37190.

Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev *10*, 1433-1442.

Kolodner, R. D., and Marsischky, G. T. (1999). Eukaryotic DNA mismatch repair. Curr Opin Genet Dev *9*, 89-96.

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

Lahue, R. S., Au, K. G., and Modrich, P. (1989). DNA mismatch correction in a defined system. Science *245*, 160-164.

Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N., and Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature *407*, 711-717.

Lamers, M. H., Winterwerp, H. H., and Sixma, T. K. (2003). The alternating ATPase domains of MutS control DNA mismatch repair. Embo J *22*, 746-756.

Lee, S. D., Surtees, J. A., and Alani, E. (2007). Saccharomyces cerevisiae MSH2-MSH3 and MSH2-MSH6 Complexes Display Distinct Requirements for DNA Binding Domain I in Mismatch Recognition. J Mol Biol *366*, 53-66.

Li, G. M., and Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. Proc Natl Acad Sci U S A *92*, 1950-1954.

Luhr, B., Scheller, J., Meyer, P., and Kramer, W. (1998). Analysis of in vivo correction of defined mismatches in the DNA mismatch repair mutants msh2, msh3 and msh6 of Saccharomyces cerevisiae. Mol Gen Genet *257*, 362-367.

Lynch, H. T., and de la Chapelle, A. (2003). Hereditary colorectal cancer. N Engl J Med *348*, 919-932.

Marsischky, G. T., Filosi, N., Kane, M. F., and Kolodner, R. (1996). Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev *10*, 407-420.

Marsischky, G. T., and Kolodner, R. D. (1999). Biochemical characterization of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 complex and mispaired bases in DNA. J Biol Chem *274*, 26668-26682.

Matson, S. W., and Robertson, A. B. (2006). The UvrD helicase and its modulation by the mismatch repair protein MutL. Nucleic Acids Res *34*, 4089-4097.

Mazur, D. J., Mendillo, M. L., and Kolodner, R. D. (2006). Inhibition of Msh6 ATPase activity by mispaired DNA induces a Msh2(ATP)-Msh6(ATP) state capable of hydrolysis-independent movement along DNA. Mol Cell *22*, 39-49.

Mendillo, M. L., Mazur, D. J., and Kolodner, R. D. (2005). Analysis of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 and MLH1-PMS1 complexes with DNA using a reversible DNA end-blocking system. J Biol Chem *280*, 22245-22257.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. Annu Rev Genet *25*, 229-253.

Modrich, P. (2006). Mechanisms in eukaryotic mismatch repair. J Biol Chem *281*, 30305-30309.

Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem *65*, 101-133.

Natrajan, G., Lamers, M. H., Enzlin, J. H., Winterwerp, H. H., Perrakis, A., and Sixma, T. K. (2003). Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. Nucleic Acids Res *31*, 4814-4821.

Nicholson, A., Hendrix, M., Jinks-Robertson, S., and Crouse, G. F. (2000). Regulation of mitotic homeologous recombination in yeast. Functions of mismatch repair and nucleotide excision repair genes. Genetics *154*, 133-146.

Nishant, K. T., Plys, A. J., and Alani, E. (2008). A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in Saccharomyces cerevisiae. Genetics *179*, 747-755.

Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature *407*, 703-710.

Owen, B. A., Yang, Z., Lai, M., Gajek, M., Badger, J. D., 2nd, Hayes, J. J., Edelmann, W., Kucherlapati, R., Wilson, T. M., and McMurray, C. T. (2005). (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. Nat Struct Mol Biol *12*, 663-670.

Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996). hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. Curr Biol *6*, 1181-1184.

Paques, F., and Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev *63*, 349-404.

Peltomaki, P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. J Clin Oncol *21*, 1174-1179.

Prolla, T. A., Christie, D. M., and Liskay, R. M. (1994a). Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. Mol Cell Biol *14*, 407-415.

Prolla, T. A., Pang, Q., Alani, E., Kolodner, R. D., and Liskay, R. M. (1994b). MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science *265*, 1091-1093.

Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P., and Jiricny, J. (1999). Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. J Biol Chem *274*, 32368-32375.

Reenan, R. A., and Kolodner, R. D. (1992). Characterization of insertion mutations in the Saccharomyces cerevisiae MSH1 and MSH2 genes: evidence for separate mitochondrial and nuclear functions. Genetics *132*, 975-985.

Schofield, M. J., Nayak, S., Scott, T. H., Du, C., and Hsieh, P. (2001). Interaction of Escherichia coli MutS and MutL at a DNA mismatch. J Biol Chem *276*, 28291-28299.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385.

Selmane, T., Schofield, M. J., Nayak, S., Du, C., and Hsieh, P. (2003). Formation of a DNA mismatch repair complex mediated by ATP. J Mol Biol *334*, 949-965.

Selva, E. M., Maderazo, A. B., and Lahue, R. S. (1997). Differential effects of the mismatch repair genes MSH2 and MSH3 on homeologous recombination in Saccharomyces cerevisiae. Mol Gen Genet *257*, 71-82.

Shcherbakova, P. V., Hall, M. C., Lewis, M. S., Bennett, S. E., Martin, K. J., Bushel, P. R., Afshari, C. A., and Kunkel, T. A. (2001). Inactivation of DNA mismatch repair by increased expression of yeast MLH1. Mol Cell Biol *21*, 940-951.

Shell, S. S., Putnam, C. D., and Kolodner, R. D. (2007a). Chimeric Saccharomyces cerevisiae Msh6 protein with an Msh3 mispair-binding domain combines properties of both proteins. Proc Natl Acad Sci U S A *104*, 10956-10961.

Shell, S. S., Putnam, C. D., and Kolodner, R. D. (2007b). The N terminus of Saccharomyces cerevisiae Msh6 is an unstructured tether to PCNA. Mol Cell *26*, 565-578.

Sia, E. A., Dominska, M., Stefanovic, L., and Petes, T. D. (2001). Isolation and characterization of point mutations in mismatch repair genes that destabilize microsatellites in yeast. Mol Cell Biol *21*, 8157-8167.

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

Su, S. S., Lahue, R. S., Au, K. G., and Modrich, P. (1988). Mispair specificity of methyl-directed DNA mismatch correction in vitro. J Biol Chem *263*, 6829-6835.

Sugawara, N., Paques, F., Colaiacovo, M., and Haber, J. E. (1997). Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination. Proc Natl Acad Sci U S A *94*, 9214-9219.

Surtees, J. A., and Alani, E. (2006). Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination. J Mol Biol *360*, 523-536.

Tishkoff, D. X., Filosi, N., Gaida, G. M., and Kolodner, R. D. (1997). A novel mutation avoidance mechanism dependent on S. cerevisiae RAD27 is distinct from DNA mismatch repair. Cell *88*, 253-263.

Wagner, A., Hendriks, Y., Meijers-Heijboer, E. J., de Leeuw, W. J., Morreau, H., Hofstra, R., Tops, C., Bik, E., Brocker-Vriends, A. H., van Der Meer, C.*, et al.* (2001). Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. J Med Genet *38*, 318-322.

Wang, Q., Hennig, U. G. G., Ritzel, R. G., Savage, E. A., and von Borstel, R. C. (1990). Double-stranded base sequencing confirms the genetic evidence that the hom3-10 allele of Saccharomyces cerevisiae is a frameshift mutation (abstract). Yeast *6*, S76.

Wang, T. F., Kleckner, N., and Hunter, N. (1999). Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. Proc Natl Acad Sci U S A *96*, 13914-13919.

Warren, J. J., Pohlhaus, T. J., Changela, A., Iyer, R. R., Modrich, P. L., and Beese, L. S. (2007). Structure of the human MutSalpha DNA lesion recognition complex. Mol Cell *26*, 579-592.

Welsh, K. M., Lu, A. L., Clark, S., and Modrich, P. (1987). Isolation and characterization of the Escherichia coli mutH gene product. J Biol Chem *262*, 15624-15629.

Wilson, T., Guerrette, S., and Fishel, R. (1999). Dissociation of mismatch recognition and ATPase activity by hMSH2-hMSH3. J Biol Chem *274*, 21659-21664.

Yamamoto, A., Schofield, M. J., Biswas, I., and Hsieh, P. (2000). Requirement for Phe36 for DNA binding and mismatch repair by Escherichia coli MutS protein. Nucleic Acids Res *28*, 3564-3569.

Zhang, Y., Yuan, F., Presnell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L., and Li, G. M. (2005). Reconstitution of 5'-directed human mismatch repair in a purified system. Cell *122*, 693-705.