ROLES FOR THE INTRINSICALLY DISORDERED LINKER ARMS OF MLH PROTEINS IN DNA MISMATCH REPAIR

A Dissertation

Presented to the Faculty of the Graduate School

Of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Christopher Michael Furman

May 2021

ROLES FOR THE INTRINSICALLY DISORDERED LINKER ARMS OF MLH PROTEINS IN DNA MISMATCH REPAIR

Christopher Michael Furman, Ph. D.

Cornell University 2021

Eukaryotic mismatch repair (MMR) is initiated when a misincorporation event occurring during DNA replication is recognized by MutS homolog (MSH) heterodimeric proteins. MSH complexes then recruit MutL homolog (MLH) heterodimeric complexes. Once recruited to the MSH-mismatch site, MLH complexes undergo ATP-dependent conformational changes that results in the nicking of the newly replicated strand. This nicking step provides entry for downstream repair factors that subsequently excise and repair the error. MLH proteins contain long and flexible intrinsically disorder regions (IDRs) that connect structured N- and C- terminal domains. I characterized how IDRs regulate the functions of the yeast Mlh1-Pms1 mismatch repair (MMR) complex. Shortening or scrambling the IDRs in both subunits ablated MMR *in vivo*. Mlh1-Pms1 complexes with shorter IDRs that disrupt MMR retained wild-type DNA binding affinity but were impaired for diffusion on both naked and nucleosome-coated DNA. Moreover, the IDRs also regulated the ATPase and nuclease activities encoded in the structured N- and C-terminal domains of the complex. This combination of phenotypes underlies the catastrophic MMR defect seen with the mutant Mlh1-Pms1 *in vivo*. More broadly, this work highlighted an unanticipated multi-functional role for IDRs in regulating both facilitated diffusion on chromatin and nucleolytic processing of a DNA substrate.

These studies encouraged me to perform a set of experiments where I inserted FRB and FKBP dimerization domains at various positions within the Mlh1 and Pms1 IDRs that did not disrupt MMR. I then induced rapamycin-dependent FRB-FKBP interactions both *in vivo* and *in*

vitro. Through this approach, I created an *mlh1* allele that can be reversibly disrupted for MMR upon rapamycin treatment, providing a new tool to disrupt MMR on demand. I then showed that restraining the MLH linkers disrupted the coordinated movement of the N-terminal MLH ATP binding domains and caused defects in MMR. In contrast, restraints predicted to maintain free movement of the N-terminal ATP binding domains had weak effects on MMR. Restriction of coordinated movements disrupted DNA and PCNA dependent regulation of Mlh1-Pms1 ATPase activity and Mlh1-Pms1 DNA binding affinity. Together, this work provides support for MLH linker domains mediating distinct conformational steps in DNA MMR and is consistent with a two-step clamp model for MLH proteins acting in eukaryotic MMR.

Additionally, I performed a separate set of studies aimed at understanding how Mlh1-Mlh3 acts in its major role to resolve double Holliday junctions into crossovers that facilitate the Meiosis I division. Through a phylogenetic approach, I identified residues in baker's yeast Mlh3 critical for its meiotic functions. Sites in Mlh3 were changed to the Pms1 equivalent in conserved positions located outside of motifs found in all MLH family members. I also made changes in sites that are conserved in one family but missing from the other, and constructed Mlh3/Pms1 chimeras. The resulting strains showed phenotypes similar to *mlh3* hypomorph or null alleles, and in some cases showed phenotypes stronger than the *mlh3* null, providing evidence that all three domains in the MLH protein family are critical for conferring pathway specificity. Importantly, *mlh3* ATP binding and endonuclease domain alleles improved MMR functions in *pms1Δ* strains without disrupting meiotic functions, indicating an expansion of Mlh3 functions, and suggesting that MLH proteins have both mismatch repair and crossover functions. This strategy provides an approach to understand how paralogs have evolved to support distinct cellular processes.

BIOGRAPHICAL SKETCH

The author was born in Fredericksburg, Virginia and grew up in the DC metropolitan area and lived there until he finished both his undergraduate and Master's degree in Biology at The Catholic University of America (CUA). While at CUA, the author worked and performed research in the lab of Dr. John Golin. In the Golin lab, he studied the protein Pdr5 and how it contributes to the Xenobiotic resistance of baker's yeast. The author's mother and father both have successful jobs in the food marketing and distribution fields respectively. Both parents are loving, kind and very hard workers and have passed down these traits to the author. His sister is also extremely talented and is pursuing a PhD in Organic Chemistry at the University of Tennessee in the lab of David Jenkins.

The author's interest in biological research started in college at CUA, where he started as a summer undergraduate researcher and ended up changing his degree to a five year BS/MS program in order to pursue research full time. This led the author to seek a job in science where he first worked at the company DuPont, where he was as an Associate Investigator in the Central Research and Development group studying how to metabolically engineer organisms to make products. At DuPont the author decided to further pursue his scientific career, thus deciding to go back to school to obtain a PhD. In 2015 the author joined Cornell University in the lab of Dr. Eric Alani where he continued to study yeast and began exploring the topic of DNA repair. In the Alani Lab the author worked to further develop his skills as a biochemist and yeast geneticist. In his time at Cornell, the author took up several opportunities to better impact the broader graduate community by starting a graduate student association for his department and became a social fellow of the Big Red Barn, the Graduate and Professional Student Center. When not in lab the author also enjoys spending time with his wife and his two young children.

I would like to dedicate my thesis to my grandfathers, Patrick Beuchert and Reed Furman, two men who have greatly shaped who I am today. Not a day goes by that I don't think or miss either of you.

ACKNOWLEDGMENTS

I want to take this opportunity to express my deepest gratitude towards my PhD advisor Dr. Eric Alani. He is one of the smartest and most brilliant scientist that I know and has been the best mentor any one could ask for. He is by far one of the kindest and most patient human on this planet. I continue to learn from him every day and I greatly appreciate that he has constantly encouraged me and has been extremely supportive during my entire time at Cornell and in Ithaca. I only hope that I can be as good of a leader and mentor as him in the future.

I also want to thank my thesis committee members, Dr. Marcus Smolka and Dr. Joe Peters. I am lucky to have these two brilliant scientists on my side both as mentors and as people who truly care about my science and career. Not only do they provide crucial critiques of my work, but they are extremely encouraging.

I also want to thank my fellow members of Alani Lab, both past and present. They have been one of the best group of people to work with. I especially want to thank my early mentors, Carol Manhart and Vandana Ragahaven, who taught me all that I know about biochemistry and protein purification. I want to thank Najla-Al-Sweel, Ujani Chakraborty, Beata Mackenroth, Gianno Pannifino, Michael Gioia, Lisette Payero, Michelle Scotland and Jun Jie Chen, with whom I have enjoyed many discussions about science and more importantly life. I also have to thank my collaborators with whom I have been lucky to have learnt a great deal from. I also have to give a special thank you to my undergraduate mentee, Ryan Elbashir, who taught me how to be a better mentor and an even better person.

I want to thank all my friends in the staff at the Big Red Barn including my two supervisors, Kris Corda and Jen Forbes, who gave me an escape from science and many friendships that I hold dear.

I also have to thank all my friends at Cornell and in BMCB with whom I have shared many great times over drinks while discussing science or on the golf course, these are the friendships that I will carry with me forever. I also would like to thank the GFAs of BMCB-Casey, Vic and Ginger for all the amazing work that they do and for being so friendly and helpful.

More importantly, I want to especially thank all of my family who have been with me through this entire process and who put up with all my craziness and eccentricities. To my wife Caroline, who has been a pillar of support throughout my time at Cornell and the whole time we have known each other. You are always there to make me laugh and look on the positive side of things. I would not have made it without you and cannot thank you enough for being there for me. To my children, Patrick and Mary, for providing continual joy in my life even when my experiments failed to work. To my sister, for being an awesome sibling who pushes me to always be better. Finally to my parents, who have always taken pride in everything that I do and have given me constant support and instilled in me the work ethics and values that shape who I am as a person and a scientist.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CREDITS

Chapter 1 was published in Yeast: Furman CM, Elbashir R, Alani E. (2020). Expanded roles for the MutL family of DNA mismatch repair proteins. Yeast doi: 10.1002/yea.3512. C.F. wrote two thirds of the text sections while R.E. wrote one third of the text sections. C.F. assembled Figures 1.2, 1.3, 1.4, 1.5., R.E. and E.A. assembles Figures 1.1 and 1.6.

Chapter 2 C.F., R.E. and E.A. co-wrote all sections of the text. C.F. assemble data for Tables 2.1, 2.2 and helped assembled all figures. R.E. contributed data to Table 2.1 and 2.2 and Figures 2.6, 2.7 and 2.10. G.P. contributed data to Table 2.1 and helped assembled Figure 2.2.

Chapter 3 was published in Nucleic Acids Research: Kim, Y., Furman, C. M., Manhart, C. M., Alani, E., & Finkelstein, I. J. (2019). Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex. *Nucleic Acids Research, 47,* 1823-1835. Y.K collected data and assembled Figures 3.3C-D, 3.4, 3.5, 3.6, 3.9 and Tables 3.3, 3.4 C.F collected data and assembled Figures 3.1, 3.2 3.3A-B, 3.9 and Table 3.1 and 3.2 C.M. collected data and assembled Figure 3.7, 3.8.

Chapter 4 C.F. wrote all sections of the text and collected data for all figures except 4.4 which was performed by T.W. and Q.Z.

CHAPTER 1

Expanded roles for the MutL family of DNA mismatch repair proteins

This chapter was published in the journal Yeast: Furman CM, Elbashir R, Alani E. (2020). Expanded roles for the MutL family of DNA mismatch repair proteins. Yeast doi: 10.1002/yea.3512.

ABSTRACT

The MutL family of DNA mismatch repair proteins plays a critical role in excising and repairing misincorporation errors during DNA replication. In many eukaryotes, members of this family have evolved to modulate and resolve recombination intermediates into crossovers during meiosis. In these organisms, such functions promote the accurate segregation of chromosomes during the Meiosis I division. What alterations occurred in MutL homolog (MLH) family members that enabled them to acquire these new roles? In this review we present evidence that the yeast Mlh1-Mlh3 and Mlh1-Mlh2 complexes have evolved novel enzymatic and non-enzymatic activities and protein-protein interactions that are critical for their meiotic functions. Curiously, even with these changes, these complexes retain backup and accessory roles in DNA mismatch repair during vegetative growth.

INTRODUCTION

During meiosis a diploid cell undergoes successive reductional and equational divisions to form haploid gametes. In most organisms, proper chromosome disjunction during the reductional division requires crossing over between homolog pairs (Figure 1.1). Meiotic crossing over is thought to have evolved from DNA repair mechanisms present in vegetative (somatic) growth to allow for increased genetic recombination and a shift from sister chromatid to interhomolog repair (Barton & Charlesworth, 1998; Cavalier-Smith, 2002; Hunter, 2006; Kleckner, 1996; Marcon & Moens, 2005; Villeneuve & Hillers, 2001). Meiosis-specific functions needed to accomplish these tasks are thought to have emerged from the duplication of genes involved in cellular growth and maintenance, with such events and adaptive mechanisms permitting the acquisition of meiosis-specific gene expression and function. As described below, the MutL homolog (MLH) family of DNA mismatch repair (MMR) proteins provides a clear example of how genes that act in vegetative growth have acquired meiotic functions (Marcon & Moens, 2005).

Yeast provides an ideal model to understand how gene duplication result in the expansion of functional roles for highly conserved proteins. The adaptation of such functions results from changes in gene expression and patterns of post-translational modifications, as well as the gain or loss of interacting protein partners and enzymatic functions. One such example in baker's yeast involves the Rad51 and Dmc1 proteins, which promote early steps in homologous recombination by catalyzing the invasion of a 3' single stranded end originating from a DNA double-strand break into a homologous template (Figure 1.1; Chan, Zhang, Weissman, & Bishop, 2019; Cloud, Chan, Grubb, Budke, & Bishop, 2012; Crickard, Kaniecki, Kwon, Sung, & Greene, 2018;

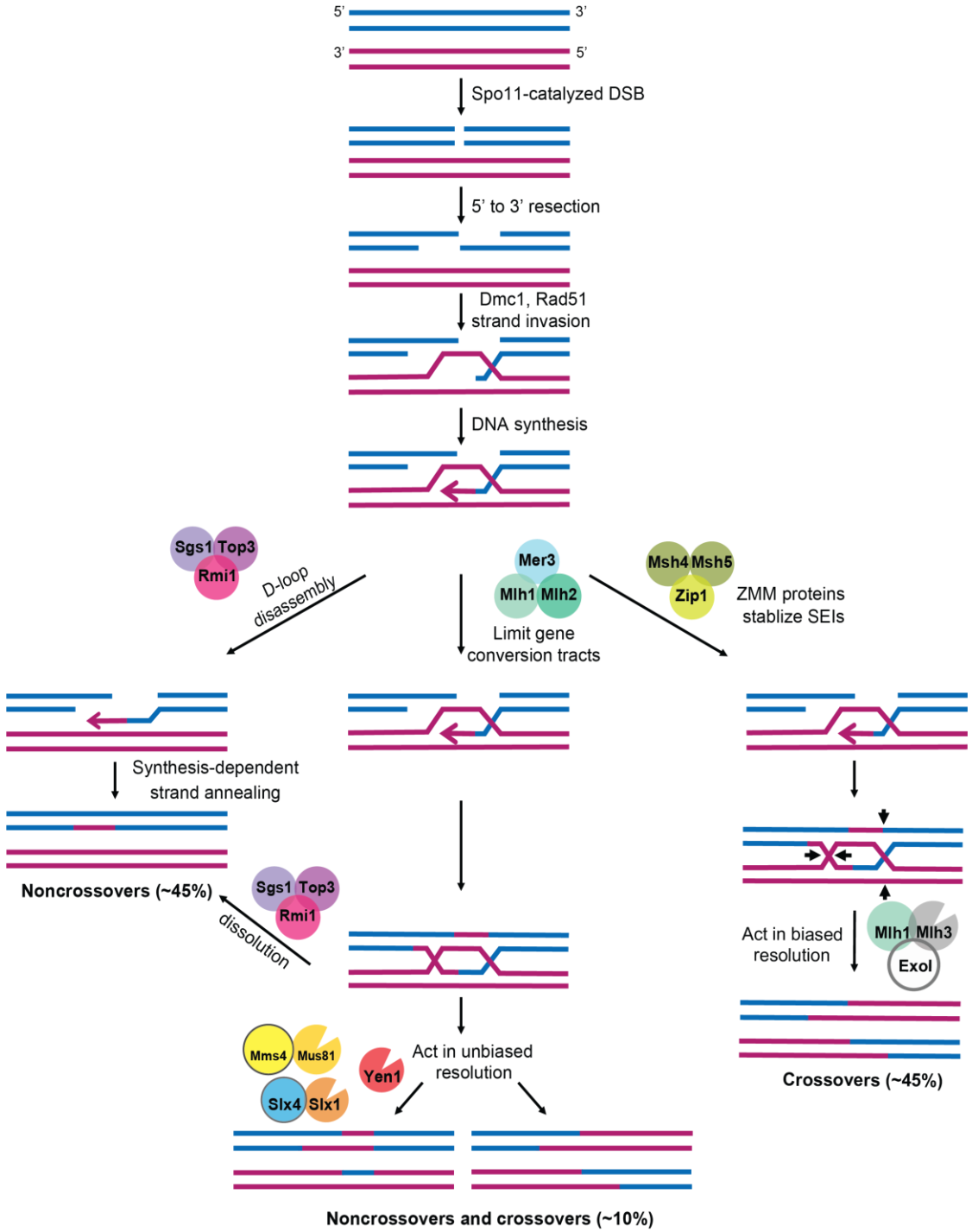


Figure 1.1. A model indicating roles for MLH proteins in meiotic recombination. In diploid yeast induced to undergo meiosis, Spo11 catalyzes double-strand breaks (DSBs) throughout the genome, which undergo 5' to 3' resection. The resulting 3' single stranded tail can invade the homologous chromosome to form a D-loop intermediate. In the ZMM-stabilized (Class I) pathway, the single end invasion intermediate (SEI) is stabilized by the actions of Msh4-Msh5 and Zip3 to promote D-loop extension through DNA repair synthesis. A double Holliday junction (dHJ) is then formed by second end capture that is asymmetrically cleaved in an Mlh1-Mlh3-dependent step to yield primarily crossover products. Note that in this model

Figure 1.1. (con't)
branch migration of Holliday junctions can occur but is not shown. Noncrossovers and a minority of crossovers (class II) are thought to occur through the other pathways shown (reviewed in Manhart & Alani, 2016; Wild & Matos, 2016; Zakharyevich et al., 2012). Roles for the Mlh1-Mlh2 complex in limiting gene conversion tract length, and Mlh1-Mlh3 in biased resolution of double-Holliday junctions to form crossovers, are shown. Mlh1-Pms1 acts to repair DNA mismatches that form in heteroduplex DNA in all pathways. The distribution of the types of meiotic recombination events in baker's yeast is approximate and was calculated based on studies showing that Spo11 catalyzed DSBs produce equal numbers of crossovers and noncrossovers (Mancera, Bourgon, Brozzi, Huber, & Steinmetz, 2008; Martini, Diaz, Hunter, & Keeney, 2006; Marsolier-Kergoat et al., 2018). Of the crossovers, 85% are estimated to be class I, and 15% class II (Cooper et al., 2018; Jessop & Lichten, 2008; Oh, Lao, Taylor, Smith, & Hunter, 2008). Of the noncrossovers, 15% are estimated to result from the resolution of dHJs in the class II pathway (DeMuyt et al., 2012), with the remainder occurring through other mechanisms such as synthesis-dependent strand annealing. See text for details and Marsolier-Kergoat et al. (2018) and Pyatnitskaya, Borde, & De Muyt (2019) for more detailed models.

Steinfeld et al., 2019). These proteins are thought to have evolved early in the evolution of eukaryotes from a single gene duplication of an ancestral archaea RadA recombinase. The duplication of an ancestral gene may have coincided with, or potentially enabled, meiosis (Lin, Kong, Nei, & Ma, 2006; Ramesh, Malik, & Logsdon, 2005). In support of these ideas, Rad51 is the only strand exchange protein present in vegetative growth, whereas both Rad51 and Dmc1 act in meiosis (Bishop, 2012; Bishop, Park, Xu, & Kleckner, 1992; Neale & Keeney, 2006). Rad51 and Dmc1 have biochemical and structural differences as well as unique subsets of protein interactors/regulators (Brown & Bishop, 2014; Chan et al., 2019; Lee et al., 2015; Steinfeld et al., 2019). In meiosis, Rad51 and Dmc1 are both needed to promote recombination between homologs, known as interhomolog bias (Figure 1.1). Interestingly, in a mutant background that lacks Dmc1 but is activated for Rad51 strand exchange activity, the bias towards interhomolog recombination is disrupted, and obligate crossovers are missing that are required for accurate chromosome segregation (Bishop 2012; Callender et al., 2016; Lao et al., 2013; Cloud et al., 2012). Additional studies showed that Rad51-mediated strand invasion activity is not required for meiotic recombination, but such an activity is critical for Dmc1 function (Cloud et al., 2012). Together, these studies identified a critical role for Dmc1 in strand exchange steps in meiosis. However, Rad51 meiotic functions are still essential in this process; Rad51 acts with the Mei5-Sae3 complex as a Dmc1 accessory factor to assist in the formation of the Dmc1 strand exchange filament (Chan et al., 2019; Cloud et al., 2012).

MutS provides another example of a single bacterial gene that duplicated and evolved early in eukaryogenesis to yield gene families with distinct vegetative growth and meiotic functions (Eisen, 1998; Lin, Nei, & Ma, 2007). In prokaryotic and eukaryotic MMR, MutS homolog (MSH) proteins recognize base-base and insertion/deletion mismatches that escape

DNA polymerase proofreading, and transmit the recognition signal to downstream repair proteins such as the MLH family, which coordinate excision of the replication error and DNA resynthesis using the parental DNA strand as a repair template (Figures 1.2 and 1.3). Inactivation of MMR results in an increase in mutation rates by up to several orders of magnitude. Such an increase reduces fitness due to the accumulations of deleterious mutations yet can provide a source of beneficial mutations for adaptation to stressful environments (Kunkel & Erie, 2015; Raghavan, Aquadro & Alani, 2019). MutS is thought to have entered archaea and eukaryotes through horizontal gene transfer, followed by a series of gene duplications (Eisen, 1998; Lin et al., 2006). In baker's yeast these events gave rise to MutS homologs (MSH) that function in the following pathways: 1. Msh2-Msh6 and Msh2-Msh3, which act to repair misincorporation errors during DNA replication and mismatches that form in heteroduplex DNA during genetic recombination; 2. Msh1, which functions in mitochondrial DNA metabolism; and 3. Msh4-Msh5, which promotes crossover formation in meiosis (Figures 1.1 and 1.3; Hollingsworth, Ponte, & Halsey, 1995; Kunkel & Erie, 2015; Reenan & Kolodner, 1992; Ross-Macdonald & Roeder, 1994; Snowden, Acharya, Butz, Berardini, & Fishel, 2004). Interestingly, Msh2-Msh6 (base-base and single nucleotide insertion/deletion mismatches) and Msh2-Msh3 (preference for insertion/deletion mismatches of up to 17 nucleotides in size) have evolved different mismatch repair specificities, and the Msh4-Msh5 complex lacks the amino-terminal mismatch recognition domain found in the other yeast MSH proteins, but recognizes Holliday junctions, a critical intermediate in meiotic recombination (Jensen, Jauert, & Kirkpatrick, 2005; Kunkel & Erie, 2015; Lahiri, Li, Hingorani & Mukerji, 2018; Pochart, Woltering, & Hollingsworth, 1997; Snowden et al., 2004).

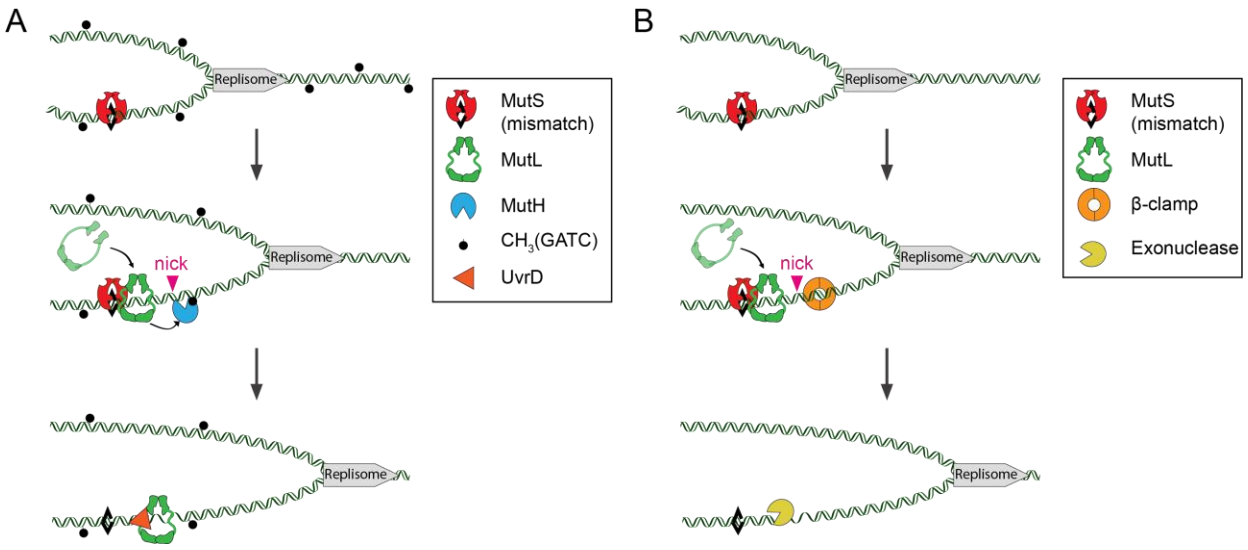


Figure 1.2. Model for prokaryotic MMR. A model for mismatch repair in bacteria. A mismatch in DNA formed due to DNA polymerase misincorporation/DNA slippage is shown as a black diamond. MutS binds to the mismatch and undergoes an ATP-dependent conformational change to act as a sliding clamp and recruit MutL. (A) In *E. coli,* the MutS-MutL complex then acts as a sliding clamp complex that recruits MutH to nick the unmethylated (newly replicated) strand at hemi-methylated d(GATC) sites. These nicks allow for the helicase UvrD to unwind the DNA which is then excised, by a variety of nucleases, depending on the polarity of the nick relative to the mismatch. Alternatively, UvrD and MutL processively unwind the DNA between two d(GATC) sites (depicted). The resulting gap is then filled in by DNA polymerase. (B) In most other bacteria, MutL contains an intrinsic endonuclease activity that is stimulated by the DNA replication processivity factor β-clamp. These nicks act as entry sites for exonucleases to excise the mismatch, followed by re-synthesis of the gapped DNA by DNA polymerase. It is not clear if a UvrD helicase acts during MMR in bacteria that do not use MutH and Dam methylase. In such bacteria a recent model suggests that analogous to eukaryotic MMR, an exonuclease acts on double-stranded DNA to excise the mismatch, thus not requiring a UvrD helicase type activity (as shown, see Lenhart, Pillon, Guarne, Biteen, & Simmons, 2016).

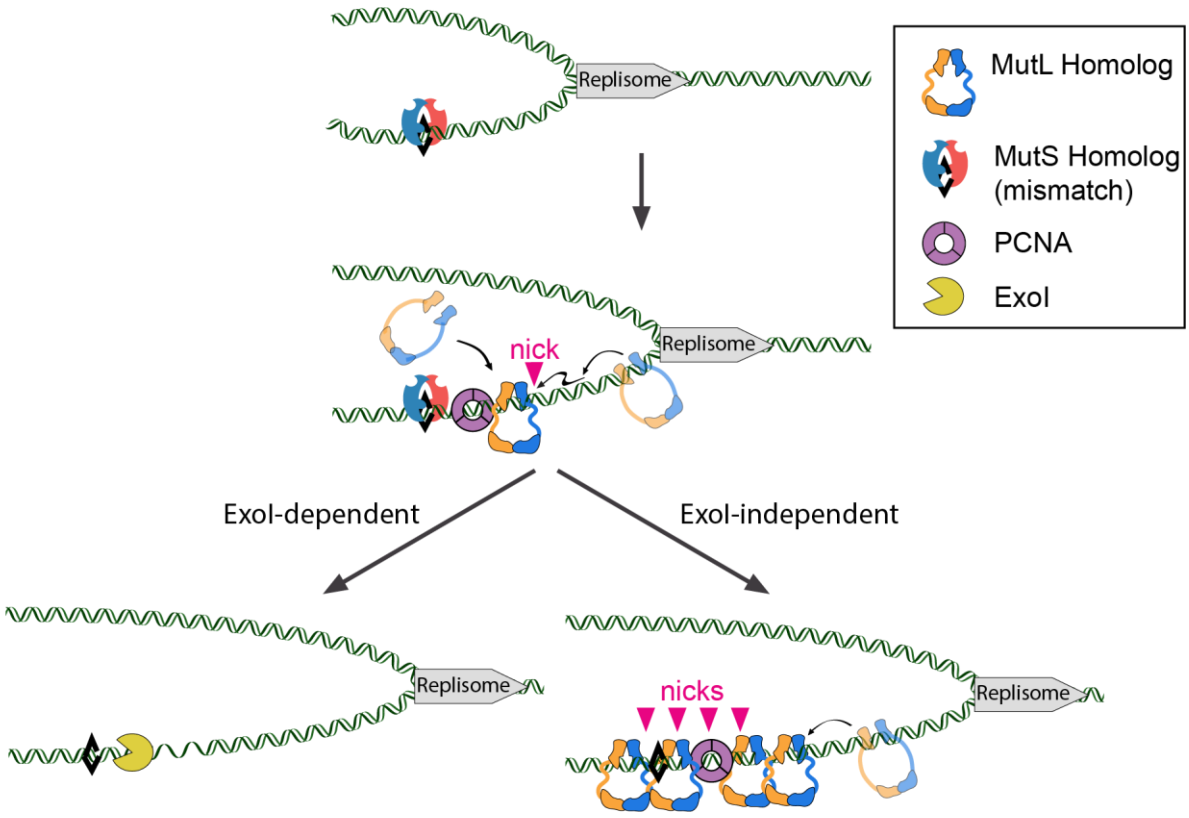


Figure 1.3. Model for MMR in baker's yeast. A mismatch in DNA formed due to DNA polymerase misincorporation is shown as black diamond. MSH complexes (Msh2-Msh3 or Msh2-Msh6) bind to the mismatch and undergo an ATP-dependent conformational change that allows them to act as a sliding clamp and recruit MLH complexes. Mlh1-Pms1 or Mlh1-Mlh3 undergo ATP-dependent conformational changes and orientation-specific interactions with PCNA (see text for a review of hypotheses for how these interactions with PCNA could be initiated), resulting in incision of the newly replicated DNA strand. This is followed by two potential pathways, ExoI-independent and -dependent, that can recruit downstream repair factors to excise and resynthesize the DNA and ligate the repaired strand. In the ExoI-dependent pathway, ExoI uses the nicks made by MLH complexes as an entry site to excise the mismatch containing DNA in a 5'-3' direction. In the ExoI-independent pathway, multiple nicks made by MLH complexes in the vicinity of the mismatch promote DNA strand displacement and synthesis by DNA polymerase to repair the mismatch. Adapted from Kim et al., (2019).

This review focuses on how the MutL homolog (MLH) family of eukaryotic DNA MMR proteins has evolved distinct vegetative and meiotic life cycle functions; the causes of this specificity in function are not as well studied compared to those identified for the Rad51 and Dmc1 proteins. Most eukaryotes encode at least two MLH heterodimeric complexes that play critical roles in vegetative MMR and/or meiotic recombination (Kadyrov, Dzantiev, Constantin, & Modrich, 2006; Kaydrov et al., 2007; Räschle, Dufner, Marra, & Jiricny, 2002). In baker's yeast three MLH complexes have been identified: Mlh1-Pms1, Mlh1-Mlh2, and Mlh1-Mlh3 (Wang, Kleckner, & Hunter, 1999). All three function during vegetative MMR, with Mlh1-Mlh3 and Mlh1-Mlh2 displaying minor and more specialized roles. (Romanova & Crouse, 2013; Harfe, Minesinger, & Jinks-Robertson, 2000). In meiosis, Mlh1-Mlh3 acts in the biased cleavage of double Holliday junctions to yield crossovers that are critical for the formation of gametes, Mlh1-Mlh2 regulates gene conversion tract lengths, and Mlh1-Pms1, analogous to its role in MMR, repairs mismatches that form in heteroduplex DNA during genetic recombination (Figure 1.1; Manhart & Alani, 2016; Zakharyevich, Tang, Ma, & Hunter., 2012; Abdullah, Hoffmann, Cotton, & Borts, 2004; Campbell et al., 2014; Duroc et al., 2017; Harfe et al., 2000; Hunter & Borts, 1997). These observations suggest that MLH complexes have evolved different roles during the vegetative and meiotic life cycles of a yeast cell. Such distinct roles have been conserved in higher eukaryotes, and in humans, defects in these factors have been linked to distinct diseases such as hereditary forms of colon cancer and infertility (Gray & Cohen, 2016; Lynch, Snyder, Shaw, Heinen, & Hitchens, 2015).

In addition to being functionally conserved, the MLH family of proteins are structurally conserved (Figure 1.4). Each subunit in the dimeric MLH complexes contain globular amino (N) and carboxy (C-) terminal domains connected by linker arms. Structures of some of the N- and

C-terminal domains have been solved for prokaryotic and eukaryotic proteins (Figure 1.4; Arana et al., 2010; Ban & Yang, 1998; Guarne et al., 2004; Gueneau et al., 2013). The N-terminal domain of these proteins contains an ATP-binding site similar in structure to the GHKL family of ATPases (DNA **G**yrase, **H**sp90, histidine **k**inases, Mut**L)** (Dutta & Inouye, 2000). Upon nucleotide binding, the two ATP binding domains in the dimer physically interact, with the linker domains becoming more ordered, and the MLH complex primed for activation (Ban & Yang, 1998; Sacho, Kadryov, Modrich, Kunkel & Erie, 2008). The main dimerization interface for the two subunits is located at the C-terminal domain, and for most MLH complexes, a metal binding domain is also found at the C-terminus that is critical for endonuclease activity (Guarne et al., 2004; Kadyrov et al., 2006).

What changes occurred in MLH family members that enabled them to acquire distinct vegetative and meiotic roles? As described below, there is strong conservation in the N- and C-terminal domains of the MLH proteins, but there are also amino acid sequences that are uniquely conserved for each MLH protein that likely confer different enzymatic functions and protein-protein interactions for the different MLH complexes. Also, the MLH proteins contain highly variable and non-conserved intrinsically disordered linker arms that range in size from about 100 amino acids (aa) in bacteria to 300 aa in yeast (Figure 1.5; Guarne et al., 2004; Sacho et al., 2008). In higher eukaryotes these linkers can approach 650 aa in length, as seen for the mouse and human Mlh3 linkers (Lipkin et al., 2000). The properties of these linkers have been proposed to facilitate conformational changes needed to activate latent MutL activities (Sacho et al., 2008), and recent studies in bacteria and yeast suggest that the linker arms are also important for MLH proteins to overcome barriers in the DNA landscape that would need to be traversed to locate the replication machinery and MSH proteins (Kim, Furman, Manhart, Alani, & Finkelstein, 2019;

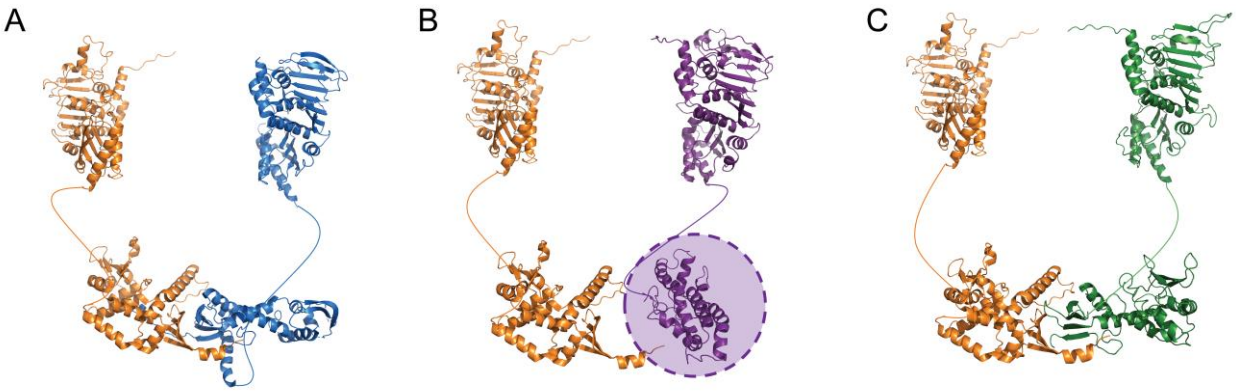


Figure 1.4. Structural homology of yeast Mlh1-Pms1, Mlh1-Mlh2, and Mlh1-Mlh3. (A) For yeast Mlh1-Pms1, the N-terminal domain of Pms1 was obtained from PDB 3h4l, and the N-terminal domain of Mlh1 was modeled by a Phyre2 (amino acids 1-367 of Mlh1; http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). The structure of the Mlh1-Pms1 C-terminus (PDB 4e4w) was solved by Gueneau et al. (2013). (B) For yeast Mlh1-Mlh2, the N-terminal domains of Mlh1 and Mlh2 were created using a Phyre2 homology model (amino acids 1-371 of Mlh2). The C-terminal domain of Mlh1 was obtained from PDB 4e4w. The C-terminal domain of Mlh2 (amino acids 503-695), created using a Phyre2 homology model, did not align well with the C-terminal domains of the other MLH proteins; for this reason, it is indicated by a shaded oval. (C) For yeast Mlh1-Mlh3, the N-terminal domains of Mlh1 and Mlh3 were created by a homology model presented in Al-Sweel et al. (2017). The C-terminal domain of Mlh1 was obtained from PDB 4e4w, and the C-terminal domain of Mlh3 (amino acids 491-715 of Mlh3) was created by a homology model presented in Al-Sweel et al. (2017). In all panels the unstructured linker domains are shown as curved lines.

Mardenborough et al., 2019; Plys, Rogacheva, Greene, & Alani, 2012). Why linker lengths show such variability is not known, but they too could have evolved for different life cycle functions.

OVERVIEW OF PROKARYOTIC MUTL PROTEINS AND THEIR RELATIONSHIP TO EUKARYOTIC MLH PROTEINS

Through genetic, *in vitro* reconstitution, and single-molecule analyses, scientists have obtained a thorough understanding of MMR in the bacteria *E. coli*. This work has provided detailed mechanistic information with respect to mismatch recognition, strand discrimination, and repair DNA synthesis steps (Figure 1.2; Jiricny 2013; Lahue, Au, & Modrich, 1989; Lamers et al., 2000; Liu et al., 2016; Modrich & Lahue, 1996; Obmolova, Ban, Hsieh, & Yang, 2000). In the model for *E. coli* shown in Figure 1.2, MMR is initiated by the MutS homodimer binding to a mismatch. This binding causes an ATP-dependent conformational change in MutS to form a sliding clamp that recruits the MutL homodimer (Acharya et al., 2003; Liu et al., 2016). The MutS-MutL complex then activates MutH, a restriction endonuclease-like protein, which nicks the unmethylated strand of hemimethylated DNA at d(GATC) sites. Importantly, *E. coli* uses the time interval needed for DNA adenine methyltransferase (Dam) to methylate newly replicated DNA at d(GATC) sites as a strand discrimination signal for MutH incision. After activating MutH incision, MutL recruits the UvrD helicase to unwind the newly replicated strand containing the mismatch, which in turns allows for single stranded exonucleases with different polarities (RecJ, ExoI, ExoVII, ExoX) to excise mismatches 5' or 3' to the nicked site (Burdett et al., 2001, but see Liu et al. (2019) for an exonuclease-independent UvrD unwinding model). MMR is completed upon DNA resynthesis by DNA Polymerase III, followed by sealing of nicks by DNA ligase.

A

Length in amino acids and position of linker in MLH proteins

Species	Mlh1 (MutL for *E. coli* and *B. subtilis*	Pms1 (Pms2)	Mlh3	Mlh2 (Pms1)
E. coli	97 (334-431)	-	-	-
B. subtilis	108 (327-435)	-	-	-
S. cerevisiae	164 (335-499)	295 (364-659)	113 (375-488)	245 (372-617)
S. pombe	92 (359-451)	234 (326-560)	-	-
M. musculus	171 (331-502)	284 (383-667)	612 (419-1031)	290 (416-706)
H. sapien	170 (331-501)	283 (385-668)	642 (396-1038)	303 (419-722)
D. melanogaster	87 (354-441)	307 (399-695)	-	-
C. elegans	145 (347-496)	249 (367-616)	-	-
A. thaliana	100 (357-457)	364 (344-698)	511 (357-868)	-
D. rerio	160 (322-482)	288 (374-662)	303 (358-661)	334 (342-672)

B

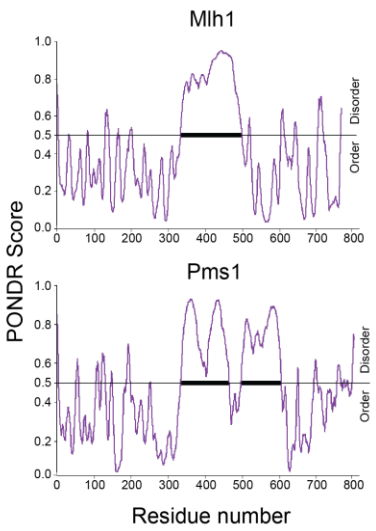


Figure 1.5. Composition of linker variation among MutL and MLH proteins. (A) Proposed linker lengths of various MutL and MLH proteins from bacteria to higher eukaryotes. Amino acid sequences were obtained from UniProt (https://www.uniprot.org), and linker lengths were determined using the Predictor of Natural Disorder (PONDR) VSL2 disorder prediction algorithm. (B) Shown is a PONDR analysis of *S. cerevisiae* Mlh1 and Pms1 (Kim et al., 2019; Obradovic et al., 2003).

In addition to roles in MMR, MutS and MutL act to prevent recombination between divergent homologous sequences, a process that is also thought to promote speciation. This heteroduplex rejection mechanism, which involves interactions with non-MMR proteins, acts to prevent deleterious recombination events, and is also seen in eukaryotes (Chakraborty & Alani, 2016; Rayssiguier, Thaler, & Radman, 1989; Worth, Clark, Radman, & Modrich, 1994).

E. coli is an outlier among prokaryotes in having acquired DNA methylation to distinguish template from nascent strands during replication fork passage. The acquisition of Dam and MutH nuclease functions is thought to be accompanied by the loss of MutL endonuclease activity (reviewed in Putnam 2016). In fact, most prokaryotes and all eukaryotes examined lack a MutH-Dam mechanism (Lenhart, Pillon, Guarné, Biteen, & Simmons, 2016), and in these organisms, MutL family proteins display an intrinsic endonuclease activity that is thought to be directed to nick the nascent strand through orientation-specific interactions with the DNA polymerase processivity clamp (β-clamp in prokaryotes, PCNA in eukaryotes; Kadyrov et al., 2006; Kadyrov et al., 2007; Pillon et al., 2010). While it is not known how this interaction occurs, recent studies have suggested that the discontinuity of Okazaki fragments on the lagging strand, processivity clamps remaining on DNA after replication, replication restart on the leading strand, and excision of ribonucleotides could provide the correct loading of the processivity clamp for strand-specific nicking by MLH endonucleases (Heller & Marians, 2006; Kawasoe, Tsurimoto, Nakagawa, Masukata, & Takahashi, 2016; Kunkel & Erie, 2015; Pluciennik et al., 2010).

BAKER'S YEAST MLH PROTEINS

MLH1-PMS1 PLAYS A MAJOR ROLE IN MMR

In current models for eukaryotic MMR (Figure 1.3; Kunkel & Erie, 2015), once an MSH complex binds to a mismatch, it undergoes, like MutS, an ADP-ATP exchange that alters the conformation of the complex to form a sliding clamp, allowing it to recruit the major MLH MMR complex, Mlh1-Pms1 (Gradia et al., 1999; Gorman et al., 2012). During this process, MSH heterodimers can interact with the replication processivity clamp PCNA that is loaded onto DNA by clamp loader (Clark, Valle, Drotschmann, Gary, & Kunkel, 2000; Johnson et al., 1996; Kleczkowska, Marra, Lettieri, & Jiricny, 2001). The MSH-PCNA interaction has been hypothesized to connect at least a portion of the MSH proteins to the replication fork and inhibit the unloading of PCNA during MMR, thus increasing the time interval for MMR (Hombauer, Campbell, Smith, Desai, & Kolodner, 2011; Kawasoe, Tsurimoto, Nakagawa, Masukata, & Takahashi, 2016; Kleczkowska, Marra, Lettieri, & Jiricny, 2001). PCNA then activates the endonuclease activity of Mlh1-Pms1 to nick the newly replicated daughter strand (Erdeniz, Dudley, Gealy, Jinks-Robertson, & Liskay, 2005; Kadyrov et al., 2006; Kawasoe, Tsurimoto, Nakagawa, Masukata, & Takahashi, 2016; Pluciennik et al., 2010). These nicks can facilitate two excision pathways (Figure 1.3). One is Exo1-dependent, where a nick 5′ to the mispair serves as an entry site for ExoI, a 5' to 3' exonuclease, to enter the DNA and excise DNA containing the mismatch. The resulting single-stranded DNA gap is coated by Replication Protein A (RPA), which facilitates DNA resynthesis steps. The second is Exo1-independent; Mlh1-Pms1 makes multiple nicks in the vicinity of the mismatch, allowing for strand displacement by Polymerase δ or ε. In both Exo1-dependent and independent mechanisms, nicks are sealed by DNA ligase,

thus repairing the DNA and maintaining the original genetic information (reviewed in Goellner, Putnam, & Kolodner, 2015).

MLH1-MLH3 PLAYS A MAJOR ROLE IN MEIOTIC CROSSING OVER

Mlh1-Mlh3 plays a minor role in MMR; genetic studies have suggested that it is recruited by Msh2-Msh3 to repair primarily insertion/deletion mismatches (Flores-Rozas & Kolodner, 1998; reviewed in Manhart & Alani, 2016; Romanova & Crouse, 2013). However, Mlh1-Mlh3 plays a critical role in resolving recombination intermediates during meiotic crossing over, with *mlh1* and *mlh3* null mutants showing meiotic crossover defects (Hunter & Borts, 1997; Manhart & Alani, 2016).

Meiotic recombination is initiated by Spo11, which catalyzes 150-200 DNA double strand breaks genome-wide (Figure 1.1; Keeney, Giroux, & Kleckner, 1997; Pan et al., 2011; Pyatnitskaya, Borde, & De Muyt, 2019; Robine et al., 2007). The DSBs are resected in a 5' to 3' direction in steps involving Exo1 to form 3' single stranded tails at each side of the double-strand break. The RecA family proteins Dmc1 and Rad51 form a filament on the tails and invade a homologous template, creating a displacement loop (D-loop; Brown, Grubb, Zhang, Rust, & Bishop, 2015). The D-loop intermediates serve as a key substrate for interconnected recombination pathways. In one pathway, the Sgs1-Top3-Rmi1 helicase/topoisomerase complex (STR), mediates the disassembly of the D-loop intermediate to form noncrossovers in a process known as synthesis-dependent strand annealing. In a second pathway the D-loop is stabilized by a class of ZMM (yeast Zip1/Zip2/Zip3/Zip4, Msh4/Msh5, Mer3) proteins to form a single-end invasion intermediate (SEI; Hunter and Kleckner, 2001; Pyatnitskaya, Borde, & De Muyt, 2019). Second-end capture of this intermediate, followed by branch migration, DNA synthesis, and

ligation, results in the formation of a double-Holliday junction (dHJ) which is resolved in a biased manner to form crossovers (class I) that display interference (Allers & Lichten, 2001; Hunter, 2006; Zakharyevich et al., 2012). In a third pathway the D-loop progresses to form dHJs that are resolved by structure selective nucleases such as Mus81-Mms4 to form noncrossovers and crossovers (class II) that lack interference and are distributed randomly (rather than the tendency for crossovers to be evenly spaced; de los Santos et al., 2003; De Muyt et al., 2012; Zakharyevich, Tang, Ma, & Hunter, 2012).

Work in yeast demonstrated that the dHJs formed in the ZMM pathway are resolved through the actions of the Msh4-Msh5, Exo1, and the Mlh1-Mlh3 endonuclease. Specifically, Msh4-Msh5 acts to stabilize single-end invasion and Holliday junctions, after which Mlh1-Mlh3 and Exo1 promote biased resolution of double-Holliday junctions to form crossovers (Borner, Kleckner, & Hunter, 2004; Manhart et al., 2017; Ranjha, Anand, & Cejka, 2014; Rogacheva et al., 2014; Snowden et al., 2004; Zakharyevich et al., 2012). This model is supported by genetic and physical studies demonstrating that Mlh1-Mlh3 acts downstream of Msh4-Msh5 in meiosis (Kolas et al., 2005; Moens et al., 2002). Relevant to these studies is the work of Marsolier-Kergoat, Khan, Schott, Zhu, & Llorente (2018), who used a high-throughput DNA sequencing approach to analyze the meiotic progency of a hybrid yeast strain. In their study they analyzed heteroduplex DNA tracts associated with noncrossover and crossover events. Interestingly, they observed a biased pattern of crossover resolution events consistent with the resolution of double-Holliday junctions being directed towards strands containing newly replicated DNA near the junctions. These observations suggest that cleavage of double-Holliday junctions through the actions of the Mlh1-Mlh3 endonuclease might be similar to that seen for the Mlh1-Pms1 nuclease during MMR, rather than Mlh1-Mlh3 acting as a structure-specific nuclease that binds

and symmetrically cleaves Holliday junctions (see Mlh1-Mlh3 polymerization models presented in Manhart et al., 2017).

It is important to note that the Mlh1-Mlh3 complex is likely to play similar roles in higher eukaryotes. For example, *Mlh3*$^{-/-}$ mice are viable, but sterile, and display dramatic decreases in meiotic crossovers. *Mlh3*$^{-/-}$ spermatocytes undergo apoptosis after metaphase, whereas oocytes fail to complete Meiosis I (Lipkin et al., 2002). These findings suggest that like in Baker's yeast, Mlh3 plays a critical role in higher eukaryotes to direct the maturation of recombination intermediates into crossovers, promote accurate chromosome segregation, and generate viable gametes. In humans, there are few clinical studies linking *hMLH3* polymorphisms to infertility; however, a detailed analysis of population variants in *hMLH3* could provide mechanistic insights (Markandona et al., 2015).

Many questions remain for how Mlh1-Mlh3 resolves double-Holliday junctions into crossover products. It is not fully understood what kinds of chromosomal architecture or protein-protein interactions are required to stabilize Mlh1-Mlh3 or direct its function (Manhart et al., 2017). There is clear evidence for post-translational modifications regulating meiotic crossover formation (He et al., 2020; Hollingsworth & Gaglione, 2019; Sanchez et al., 2020; Sourirajan & Lichten, 2008; Wild & Matos, 2016; Wild et al., 2019). For example, the structure-specific endonucleases Mus81-Mms4 and Yen1 that act in the interference-independent crossover pathway are activated and inhibited, respectively, by phosphorylation (Figure 1.1; Matos, Blanco, Maslen, Skehel, & West, 2011; Wild & Matos, 2016). At present there is no evidence for phosphorylation directly regulating Mlh1-Mlh3 endonuclease activity, but it is likely that factors that interact with Mlh1-Mlh3 are regulated by phosphorylation and/or other modifications such as SUMOylation (Cheng et al., 2006**;** He et al., 2020; Manhart & Alani, 2016). It is also of

interest to understand how the chromatin landscape is regulated to allow access of Mlh1-Mlh3 to recombination substrates (Wild et al., 2019).

Recent studies have also suggested roles for Mlh1-Mlh3 in chromosome disjunction and crossover interference (Chakraborty et al., 2017; Claeys-Bouuaert & Keeney, 2017). Crossover interference, seen in many organisms such humans, mice, *Drosophila*, and *C. elegans*, regulates the spatial distribution of crossovers; the formation of a crossover lowers the probability of a second crossover nearby (Wang et al., 2019). Widely spaced crossovers, exhibiting interference, help to ensure the fidelity of chromosome segregation. How Mlh1-Mlh3 participates in these roles is unclear.

Lastly, recent studies in yeast and higher eukaryotes have suggested a role for Mlh1-Mlh3 in mediating somatic CAG trinucleotide repeat instability (Pinto et al., 2013; Su & Freudenreich, 2017). In humans, CAG repeat expansions in exon 1 of the *HTT* gene results in Huntington's disease, an inherited neurodegenerative disease, where age of disease onset depends on the length of the CAG repeat tract. It is unclear whether Mlh1-Mlh3 acts alone, or if this process also requires other MLH proteins (see Gomes-Pereira, Fortune, Ingram, McAbney, & Monckton, 2004). It could be that, analogous to MMR, there is some partial redundancy to Pms1 and Mlh3's role in generating expansions. Curiously, Mlh1-Mlh3 has been shown to bind specifically to, but not cleave, DNA branch structures such as Holliday junctions (Ranjha, Anand, & Cejka, 2014; Rogacheva et al., 2014), substrates that could resemble intermediates that form during trinucleotide repeat expansion.

MLH1-MLH2 MODULATES MEIOTIC RECOMBINATION TRACT LENGTH

Unlike Mlh3 and Pms1, Mlh2 lacks an endonuclease motif, and although *mlh2* mutants display sensitivity to DNA damaging agents, deletion of *MLH2* confers only a mild mutator phenotype. Harfe et al. (2000) demonstrated that although not required for vegetative MMR, Mlh2 has a minor role in the repair of frameshift mutations. Additionally, Campbell et al. (2014) showed that Mlh1-Mlh2 plays a non-essential role as an accessory factor in DNA MMR, but whose function becomes more significant when essential MMR factors become limiting. More specifically, they showed through live cell imaging that Mlh2 forms foci in the presence of mispaired DNA, and that focus formation was dependent on *MSH2*, *MSH6*, and *MLH1*. Biochemical assays revealed that Mlh1-Mlh2 is recruited to DNA mismatches by both Msh2-Msh6 and Msh2-Msh3, and *in vivo* studies showed that deletion of *MLH2* in backgrounds where essential MMR proteins were missing or reduced caused synergistic increases in mutation rate (Campbell et al., 2014). Taken together, these findings suggest that Mlh2 plays a minor role in MMR, potentially overlapping with Pms1 functions.

In contrast to minor roles in MMR, Mlh2 has been implicated through genetic analyses to interact with the Mer3 helicase to regulate meiotic gene conversion tract lengths (Figure 1.1; Abdullah et al., 2004; Duroc et al., 2017; Marsolier-Kergoat et al., 2018). Furthermore, Abdullah et al. (2004) suggested, in their analysis of meiotic recombination in *mlh2 msh4* and *mlh2 msh5* double mutants, that Mlh2 channels meiotic recombination intermediates into the Msh4-Msh5 interference-dependent meiotic crossover pathway. Consistent with regulating meiotic gene conversion tracts, Duroc et al. (2017) showed that the Mer3 helicase interacted with Mlh1-Mlh2 *in vitro* and *in vivo*, Mer3 and Mlh1-Mlh2 each preferentially bound to D-loop intermediates, Mlh2 was recruited to meiotic recombination hotspots through its interaction with

Mer3, and Mlh2 recruitment did not require the presence of DNA mismatches or interactions with mismatch recognition factors. These studies suggested Mer3, initially in a structural role, binds to D-loop intermediates, and then recruits Mlh1-Mlh2 to prevent excessive D-loop extension and DNA synthesis. Such a mechanism was hypothesized to limit the exposure to DNA sequences that could participate in inappropriate meiotic recombination events that result in genome rearrangements, as well as to reduce the fraction of the genome converted at each meiosis (Duroc et al., 2017).

It is unclear whether Mlh2's role in repairing frameshift mutations is mechanistically related to its meiotic recombination role, especially because its meiotic role appears distinct from MMR. A phylogenetic analysis of for Mlh2 in the unikont taxa revealed that *S. cerevisiae MLH2* is the homolog to metazoan *PMS1* (*S. cerevisiae PMS1's* homolog is metazoan *PMS2*), which also lacks an endonuclease motif and exhibits a weak mutator phenotype (Campbell et al., 2014). *MLH2* homologs were identified in species that diverged from *S. cerevisiae* prior to the whole genome duplication event, suggesting that the origin of *MLH2* predated the yeast whole genome duplication event. Interestingly, a subset of unikont genomes contain a yeast *MLH3* homolog but lack the yeast *MLH2* homolog (the human homolog is *PMS1*), and other unikonts lack both yeast *MLH3* and *MLH2* homologs (see Vakirlis et al., 2016). Also, in a small number of fungal species, *MLH2* homologs contained stop codons and frameshifts that could reflect inactivation of the *MLH2* gene, providing support for accessory roles for Mlh2 (Campbell et al., 2014).

HOW DIFFERENT ARE THE ACTIVITIES OF MLH FAMILY PROTEINS?

There are many structural and functional similarities and differences between the Pms1, Mlh2, and Mlh3 proteins (Figure 1.4). Pms1 and Mlh3 share highly conserved functional domains: an

ATP binding site in the N-terminus that drives the conformational changes observed in MLH proteins, an intrinsically disordered linker arm, and a C-terminus containing the endonuclease motif and Mlh1 dimerization site. All three proteins compete for the same dimerization/binding site on the Mlh1 protein, and this competition could regulate the levels of each heterodimer present at each stage of the yeast life cycle (Kondo, 2001). One interesting feature of these proteins is the variability in the lengths of their intrinsically disordered linkers (Figure1.5). As mentioned earlier, the linker domains are poorly conserved, and in yeast the length of the Pms1 linker is twice that of Mlh1 and three-times that of Mlh3. Curiously, the Mlh3 linker has expanded significantly in higher eukaryotes. The mouse and human Mlh3 linkers are roughly six-times larger than the yeast linker. While the functions of these linkers are thought to primarily regulate ATP-driven conformational changes, a difference in length and sequence could facilitate physical interactions with other proteins or DNA, or serve as substrates for post-translational modifications (Claeys Bouuaert & Keeney, 2017; Sacho, Kadryov, Modrich, Kunkel & Erie, 2008; Kim et al., 2019).

MLH ENDONUCLEASE DOMAINS SHOW SUBTLE DIFFERENCES

Mlh1-Pms1 and Mlh1-Mlh3 display endonuclease activities dependent on highly conserved metal binding motifs present in both Pms1 and Mlh3 (Figure1.6; Gueneau et al., 2013; Kadyrov et al., 2006; Rogacheva et al., 2014). Five conserved residues predicted to form the endonuclease active site in Pms1 were found in homologous positions in Mlh3. These proteins maintain the ancestral endonuclease motif, but also contain distinct residues that are likely required for their function and pathway specificity. The outlier is Mlh2, which has very limited conservation in its C-terminal domain, has lost the conserved endonuclease motif, and has no characterized

enzymatic function. However, Mlh2 has adopted a novel function in regulating the length of gene conversion tract lengths, as well as playing a role as an accessory factor in MMR, but this was only seen when Pms1 levels were reduced (Abdullah et al., 2004; Campbell et al., 2014; Duroc et al., 2017). The endonuclease motif in Mlh3's C-terminus also overlaps with the Mlh1 dimerization domain. Al Sweel et al. (2017) identified alleles in the endonuclease motif that disrupt Mlh1-Mlh3 interaction, but interestingly, these alleles are functional in meiotic crossing over and defective in MMR. These observations suggest the presence of other protein-protein interactions during meiosis that stabilize Mlh1-Mlh3 and promote crossover resolution. The identification and characterization of such interactions would also allow us to better understand how protein-protein interactions provide pathway specificity.

MLH PROTEINS INTERACT WITH SPECIFIC SETS OF PROTEINS

The MLH proteins do not act alone. The kinds of selective forces that drove the evolutionary adaptation of these genes also likely acted on other genes in the same pathway or molecular network (Clark, Alani, & Aquadro, 2013). For example, in MMR, the endonuclease activity of Mlh1-Pms1 is activated through its interaction with the DNA processivity factor PCNA. PCNA is a ring-shaped structure that acts to enhance the processivity of DNA polymerase during replication by creating a topological link with the DNA template and enabling sliding during chain elongation (Kelman, 1997; Pillon, Miller, & Guarné, 2011). Eukaryotic Msh2-Msh6 and Mlh1-Pms1 both interact with PCNA through the PCNA-interacting peptide motif (PIP Box), which is defined as a six amino acid residue consisting of Qxφ[L/I]xP, where φ is a hydrophobic residue and x is any amino acid (Kelman, 1997**;** Pillon, Miller, & Guarné, 2011).

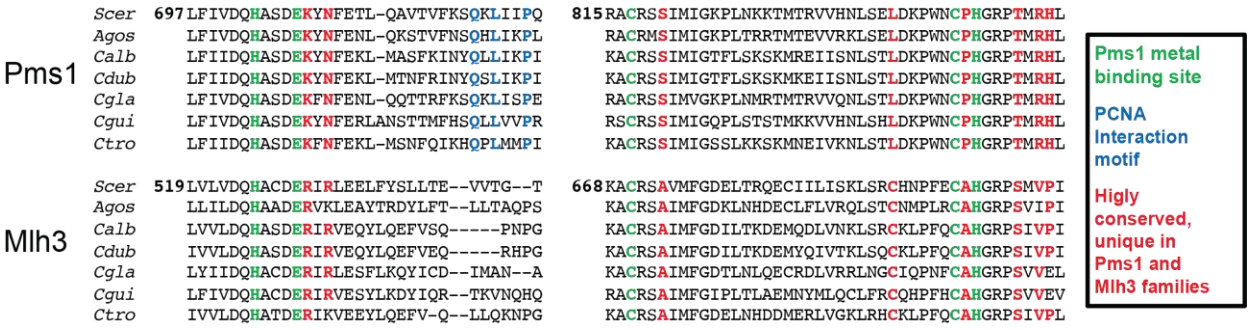


Figure 1.6. Alignment of the endonuclease domain of *S. cerevisiae* Pms1 (Gueneau et al., 2013) with Pms1 and Mlh3 homologs from 18 budding yeast (family *Saccharomycetaceae*) species (Clark, Alani, & Aquadro, 2012). *Saccharomyces cerevisiae (Scer)* Pms1 and Mlh3 amino acid sequences are shown, followed by sequences from homologs from six of the 18 species. The metal binding site of Pms1 (green font), which forms the endonuclease active site, contains five residues (H703, E707, C817, C848, H850) that are highly conserved (100% identity in 18 Pms1 and 18 Mlh3 sequences; Clark, Alani, & Aquadro, 2012). The QXLXXP motif (blue font), important for interactions with PCNA (PIP), is highly conserved in the Pms1 sequences (>94% identity; Genschel et al., 2017), but is absent in Mlh3 sequences. Also shown are residues in red that are uniquely conserved (61 to 100% identity) within each Pms1 and Mlh3 homolog family. Homologs were obtained from *Saccharomyces cerevisiae (Scer), Ashbya gossypii (Agos), Candida albicans (Calb), Candida dubliniensis (Cdub), Candida glabrata (Cgla), Candida guilliermondii (Cgui), Candida tropicalis (Ctro), Saccharomyces paradoxus, Saccharomyces mikatae, Saccharomyces bayanus, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces waltii, Kluyveromyces lactis, Kluyveromyces polysporus, Candida lusitaniae, Lodderomyces elongisporus, Debaryomyces hansenii,* and *Pichia stipitis.* Because of incomplete sequence information, the Pms1 homolog for *Debaryomyces hansenii*, and the Mlh3 homolog of *Candida lusitaniae* were not included.

The endonuclease activity of yeast Pms1 is highly stimulated by PCNA through the PIP-motif located in close proximity to the endonuclease motif (Genschel et al., 2017; Pluciennik et al., 2010). Mlh3, however, lacks any sort of motif at this specific site (Figure1.6). Historically, *in vitro* studies have shown no requirement for PCNA to stimulate the endonuclease activity of Mlh1-Mlh3 (Ranjha, Anand, & Cejka, 2014; Rogacheva et al., 2014). However, studies (Cannavo et al., 2020; Kulkarni et al., 2020), have showed that PCNA stimulates the Mlh1-Mlh3 endonuclease activity in conjunction with Msh4-Msh5 and Exo1, and one group (Cannavo et al., 2020) characterize a potential PIP-motif in Mlh1, but whether this is a bonafide PCNA interaction site is not clear (Genschel et al., 2017; Lee & Alani, 2006). One explanation for these properties is that the MLH proteins have undergone changes in their ancestral catalytic motif to bind and resolve different substrates, and/or have shifted interactions with protein partners during MMR and meiosis.

DO MLH PROTEINS SHARE FUNCTIONS?

MLH proteins are likely to share many functions, with some of their specificities driven by recruitment to different substrates or different MSH proteins. For example, Mlh1-Mlh2 meiotic functions in limiting gene conversion tracts could be explained by its interactions with Mer3, and Mlh1-Mlh3's functions actions in meiotic crossing over could be explained through its recruitment by Msh4-Msh5 to recombination intermediates (Duroc et al., 2017; Kolas et al., 2005; Moens et al., 2002). In support of this idea, work by Marsolier-Kergoat et al. (2018) showed that cleavage of double Holliday junctions, presumably through the actions of the Mlh1-Mlh3 endonuclease, was directed towards strands containing newly replicated DNA; this mechanism is analogous to Msh2-Msh6 or Msh2-Msh3 and PCNA directing the endonuclease

activity of Mlh1-Pms1 to the newly replicated strand during MMR. Further support for this idea comes from studies which showed that Msh2-Msh3 recruits Mlh1-Mlh3 primarily for the repair of deletion mispairs (Flores-Rozas & Kolodner, 1998; Romanova & Crouse, 2013). Thus, the recruitment of MLH complexes to specific substrates by specialized MutS proteins could contribute to our understanding of how protein-protein interactions influence the difference in functionality. Further studies on Mlh1-Mlh3's protein-protein interactions in meiosis would allow us to understand how DNA binding proteins and protein-protein interactions of MutL proteins evolved to adapt to meiotic processes. For example, if the residues that specify different MSH interactions were identified, and then introduced into the different MLH proteins, one could test if such interactions are sufficient for specificity, or if additional changes have occurred in the MLH proteins that specify their functions such as acquiring novel DNA binding activities.

BIOINFORMATIC AND STRUCTURAL APPROACHES

In addition to analyzing protein-protein interactions as descried above, bioinformatic strategies can be used to study the specificity of MLH complexes. Steinfeld et al. (2019) identified and analyzed amino acid residues that are well conserved within the Rad51 or Dmc1 lineage but differ between the two recombinases. This same methodology can be applied through computational approaches. For example, Multi-Harmony uses multiple sequence alignments between subfamilies of proteins, homology models, and multi-Relief and sequence-Harmony algorithms to identify amino acids that may suggest functional specificity (Brandt, Feenstra, & Heringa, 2010). By aligning Mlh3 and Pms1 amino acid sequences from 19 different budding yeast species, multi-Harmony alignments can identify residues that are well conserved within

each subfamily but differ between Mlh3 and Pms1 (Figure 1.6). Having a three-dimensional structure of Mlh3 would allow for these residues to be mapped precisely.

REGULATION BY POST-TRANSLATIONAL MODIFICATIONS

At present there is little evidence in yeast to suggest that any of the MutL proteins are functionally regulated by phosphorylation. However, in mammalian cells, post translational modifications of MLH proteins have been proposed to regulate their functions. For example, recent studies have suggested that human Mlh1-Pms2 is phosphorylated by Casein Kinase II (CK2) at Mlh1's S477 *in vitro*. Phosphorylated Mlh1 loses MMR activity and levels of phosphorylated Mlh1 vary during the cell cycle. In tumors, CK2 is overexpressed and appears to inactivate MMR. This could be analogous to mechanisms of tumorigenesis where hyper-methylation of Mlh1 drives microsatellite instability (Webbecher & Brieger, 2018). The modification of MLH proteins in yeast is not well studied but could be an interesting avenue to explore with respect to understanding unique MLH functions.

CONCLUSIONS AND FUTURE DIRECTIONS

This review is focused on understanding how and why eukaryotes acquired multiple MLH complexes that act in the vegetative and meiotic stages of the yeast life cycle. In addition to the MLH, MSH, and Rad51/Dmc1 examples presented, many other components of the homologous recombination machinery have been adapted to function in meiosis (Marcon & Moens, 2005; Steinfeld et al., 2019). Thus, studies performed with MLH proteins will likely provide a model to understand how large classes of proteins have adapted novel roles. Such adaptations could involve new partner interactions (see Wild et al., 2019) or yield novel

biochemical functions. We believe that a combination of structural (Figure 1.4) and computational evolutionary (e.g. using Multi-Harmony), analyses of MLH family proteins will encourage the creation of chimeric proteins that show how MLH proteins have evolved meiotic roles. We are optimistic that such an approach will be useful because aspects of this strategy have been successfully used by Steinfeld et al. (2019) to identify functional differences between Rad51 and Dmc1.

Molecular evolution provides another approach to study how proteins have adapted to new roles in the yeast life cycle. In work performed by Hsieh, Makrantoni, Robertson, Marston, & Murray (2020), the vegetative cohesin Scc1 was replaced with vegetative growth expression of the meiotic cohesin paralog Rec8. After passaging this strain through 1,750 vegetative growth generations, Hsieh et al. (2020) identified evolved fit populations that acquired mutations primarily in known and novel partner proteins of cohesins. Thus, this work provided a clear example of how one could "turn back the clock" to understand how proteins evolved to take on new roles. For example, a strain lacking Pms1 could be evolved to become fully functional in MMR. Such an experiment could be used to learn how the Mlh2, Mlh3, and Pms1 proteins diverged, with the goal of identifying novel protein domains and/or interacting partners.

ACKNOWLEDGMENTS

We thank Beata Mackenroth, Sara Liang, Michael Lichten, Nathan Clark, and anonymous reviewers for helpful comments on the manuscript, and Gianno Pannafino for help on the protein structure analysis of MLH proteins. C. M. F., R. E., and E.A. are supported by the National Institute of General Medical Sciences of the National Institutes of Health: R35GM134872. R. E. was also funded by the Cornell McNair Scholars Program, and the Frank L. and Lynnet Douglas

Fellowship in Chemistry, Chemical Biology, and Biochemistry. The content of this review is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AN OVERVIEW OF THIS THESIS

In Chapter 2 I present an approach to understand how Mlh3 and Pms1 have evolved to support distinct functions in genetic recombination and MMR. Using a combination of genetic, phylogenetic and computational modeling analyses, I showed that all three domains of both proteins play important roles in defining their protein functions. Furthermore, I observed that distinct amino acid residues that are conserved in Pms1 family proteins and are either absent or display differential consensus motifs in Mlh3 family proteins can be changed in Mlh3 to partially complement the loss of Pms1 function. Importantly, my work provides a framework to understand how paralogs have evolved to support distinct cellular processes.

In Chapter 3, I present work aimed at understanding the role that the intrinsically disordered linker domains of Mlh1-Pms1 play in MMR. I showed that the linker domains contain specific sequence and position requirements that cannot be generically substituted. I also showed that the loss of the linker arms contributed to the disruption of processes that require multiple rounds or processive action of the Mlh1-Pms1 conformation change and did not appear critical for single round actions. This indicates that the linker arms play a critical role in allowing Mlh1-Pms1 to undergo multiple rounds of conformational change required for MMR.

In Chapter 4, I continued to investigate the role that the linker arms play in Mlh1-Pms1 protein function. First, I created a drug-inducible system to disrupt MMR using FRB-FKBP protein domains. Second, my work showed that constricting movement of the Mlh1-Pms1 linker arms disrupted MMR functions as measured genetically and in DNA binding, ATPase, and endonuclease assays. These observations provide new insights into the choreography of conformational changes in the MLH proteins that facilitate MMR.

Chapter 5 outlines an expanded future directions section for Chapter 2, 3, and 4.

REFERENCES

Abdullah, M. F., Hoffmann, E. R., Cotton, V. E., & Borts, R. H. (2004). A role for the MutL homolog MLH2 in controlling heteroduplex formation and in regulating between two different crossover pathways in budding yeast. *Cytogenetic Genome Research, 107*, 180-190.

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

Al-Sweel, N., Raghavan, V., Dutta, A., Ajith, V. P., Di Vietro, L., Khondakar, N., Manhart, C. M., Surtees, J. A., Nishant, K. T., & Alani, E. (2017). *mlh3* mutations in baker's yeast alter meiotic recombination outcomes by increasing noncrossover events genome-wide. *PLoS Genetics*, *13*, e1006974.

Allers, T., & Lichten, M. (2001). Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell, 106,* 47-57.

Arana, M. E., Holmes, S. F., Fortune, J. M., Moon, A. F., Pedersen, L. C., & Kunkel, T. A. (2010). Functional residues on the surface of the N-terminal domain of yeast Pms1. *DNA Repair*, *9*, 448–457.

Ban, C., & Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell, 95,* 541–552.

Barton, N. H. & Charlesworth, B. (1998). Why sex and recombination? *Science, 281*, 1986-1990.

Bishop D. K. (2012). Rad51, the lead in mitotic recombinational DNA repair, plays a supporting role in budding yeast meiosis. *Cell Cycle*, *11*, 4105–4106.

Bishop, D. K., Park, D., Xu, L., & Kleckner, N. (1992). *DMC1*: a meiosis-specific yeast homolog of *E. coli recA* required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell, 69,* 439-456.

Börner, G. V., Kleckner, N., & Hunter, N. (2004). Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell, 117*, 29-45.

Brandt, B. W., Feenstra, K.A., & Heringa, J. (2010). Multi-Harmony: Detecting functional specificity from sequence alignment. *Nucleic Acids Research*, *38*, W35-W40.

Brown, M. S., & Bishop, D. K. (2014). DNA strand exchange and RecA homologs in meiosis. *Cold Spring Harbor Perspectives in Biology*, *7*, a016659.

Brown, M. S., Grubb, J., Zhang, A., Rust, M. J., & Bishop, D. K. (2015). Small Rad51 and Dmc1 complexes often co-occupy both ends of a meiotic DNA double strand break. *PLoS Genetics, 11,* e1005653.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T., & Modrich, P. (2001). *In vivo* requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences, USA*, *98*, 6765–6770.

Callender, T. L., Laureau, R., Wan, L., Chen, X., Sandhu, R., Laljee, S., Zhou, S., Suhandynata, R. T., Prugar, E., Gaines, W. A,, Kwon, Y., Börner, G. V., Nicolas, A., Neiman, A. M., & Hollingsworth, N. M. (2016). Mek1 Down Regulates Rad51 Activity during Yeast Meiosis by Phosphorylation of Hed1. *PLoS Genetics, 12,* e1006226.

Campbell, C. S., Hombauer, H., Srivatsan, A., Bowen, N., Gries, K., Desai, A., Putnam, C. D., & Kolodner, R. D. (2014). Mlh2 is an accessory factor for DNA mismatch repair in *Saccharomyces cerevisiae*. *PLoS Genetics, 10*, e1004327.

Cannavo, E., Sanchez, A., Anand, R., Ranjha, L., Hugener, J., Adam, C., Acharya, A., Weyland, N., Aran-Guiu, X., Charbonnier, J.-B., Hoffmann, E. R., Borde, V., Matos, J., & Cejka, P. (2020). Regulation of the MLH1-MLH3 endonuclease in meiosis. *BioRxiv*, 2020.02.12.946293. https://doi.org/10.1101/2020.02.12.946293

Cavalier-Smith, T. (2002). Origins of the machinery of recombination and sex. *Heredity, 88*, 125–141.

Chakraborty, U., & Alani, E. (2016). Understanding how mismatch repair proteins participate in the repair/anti-recombination decision. *FEMS Yeast Research, 16*, pii:fow071.

Chakraborty, P., Pankajam, A. V., Lin, G., Dutta, A., Krishnaprasad, G. N., Tekkedil, M. M., Shinohara, A., Steinmetz, L.M., & Nishant, K. T. (2017). Modulating crossover frequency and interference for obligate crossovers in *Saccharomyces cerevisiae* meiosis. *G3 (Bethesda), 7*, 1511-1524.

Chan, Y. L., Zhang, A., Weissman, B. P., & Bishop, D. K. (2019). RPA resolves conflicting activities of accessory proteins during reconstitution of Dmc1-mediated meiotic recombination. *Nucleic Acids Research*, *47*, 747–761.

Cheng, C. H., Lo, Y. H., Liang, S. S., Ti, S. C., Lin, F. M., Yeh, C. H., Huang, H. Y., & Wang, T. F. (2006). SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of *Saccharomyces cerevisiae*. *Genes and Development*, *20*, 2067–2081.

Claeys Bouuaert, C., & Keeney, S. (2017). Distinct DNA-binding surfaces in the ATPase and linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. *PLoS Genetics, 13,* e1006722.

Clark, N. L., Alani, E., & Aquadro, C. F. (2012). Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Research, 22,* 714-720.

Clark, N. L., Alani, E., & Aquadro, C. F. (2013). Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. *Genetics, 193*, 529–538.

Clark, A. B., Valle, F., Drotschmann, K., Gary, R. K., & Kunkel, T. A. (2000). Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *Journal of Biological Chemistry, 275,* 36498-36501.

Cloud, V., Chan, Y. L., Grubb, J., Budke, B., & Bishop, D. K. (2012). Rad51 is an accessory factor for Dmc1-mediated joint molecule formation during meiosis. *Science*, *337*, 1222–1225.

Cooper, T. J., Crawford, M. R., Hunt, L. J., Marsolier-Kergoat, M. C., Llorente, B., & Neale, M. J. (2018). Mismatch repair impedes meiotic crossover interference. *BioRxiv,* 480418. https://doi.org/10.1101/480418.

Crickard, J. B., Kaniecki, K., Kwon, Y., Sung, P., & Greene, E. C. (2018). Spontaneous self-segregation of Rad51 and Dmc1 DNA recombinases within mixed recombinase filaments. *Journal of Biological Chemistry*, *293*, 4191–4200.

de los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., & Hollingsworth, N. M. (2003). The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics, 164*, 81-94.

De Muyt, A., Jessop, L., Kolar, E., Sourirajan, A., Chen, J., Dayani, Y., & Lichten, M. (2012). BLM helicase ortholog Sgs1 is a central regulator of meiotic recombination intermediate metabolism. *Molecular Cell, 46,* 43-53.

Duroc, Y., Kumar, R., Ranjha, L., Adam, C., Guérois, R., Md Muntaz, K., Marsolier-Kergoat, M. C., Dingli, F., Laureau, R., Loew, D., Llorente, B., Charbonnier, J. B., Cejka, P., & Borde, V. (2017). Concerted action of the MutLβ heterodimer and Mer3 helicase regulates the global extent of meiotic gene conversion. *eLife, 6*, e21900.

Dutta, R., & Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences, 25*, 24–28.

Eisen, J. A. (1998). A phylogenomic study of the MutS family of proteins. *Nucleic Acids Research, 26*, 4291-4300.

Erdeniz, N., Dudley, S., Gealy, R., Jinks-Robertson, S., & Liskay, R. M. (2005). Novel *PMS1* alleles preferentially affect the repair of primer strand loops during DNA replication. *Molecular Cell Biology, 25,* 9221-9231.

Flores-Rozas, H., & Kolodner, R. D. (1998). The *Saccharomyces cerevisiae MLH3* gene functions in MSH3-dependent suppression of frameshift mutations. *Proceedings of the National Academy of Sciences, USA, 95,* 12404-12409.

Genschel, J., Kadyrova, L. Y., Iyer, R. R., Dahal, B. K., Kadyrov, F. A. & Modrich, P. (2017). Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair. *Proceedings of the National Academy of Sciences, USA, 114,* 4930–4935.

Goellner, E.M., Putman, C. D., & Kolodner, R. D. (2015). Exonuclease 1-dependent and independent mismatch repair. *DNA Repair, 32,* 24–32.

Gomes-Pereira, M., Fortune, M. T., Ingram, L., McAbney, J. P., & Monckton, D. G. (2004). Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Human Molecular Genetics, 13*, 1815-1825.

Gorman, J., Wang, F., Redding, S., Plys, A. J., Fazio, T., Wind, S., Alani, E. E., & Greene, E. C. (2012). Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proceedings of the National Academy of Sciences, USA, 109*, E3074-E3083.

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

Gray, S., Cohen, P. E. (2016). Control of meiotic crossovers: from double-strand break formation to designation. *Annual Review of Genetics, 50*, 175-210.

Guarné, A., Ramon-Maiques, S., Wolff, E.M., Ghirlando, R., Hu, X., Miller, J. H. & Yang, W. (2004). Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO Journal, 23,*4134–4145.

Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., Londino, F., Quemener, C., Le Du, M. H., Márquez, J. A., Moutiez, M., Gondry, M., Boiteux, S., & Charbonnier, J. B. (2013). Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nature Structural and Molecular Biology, 20,* 461–468.

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

He, W., Rao, H. B. D. P., Tang, S., Bhagwat, N., Kulkarni, D.S., Ma, Y., Chang, M. A.W., Hall, C., Bragg, J. W., Manasca, H.S., Baker, C., Verhees, G. F., Ranjha, L., Chen, X., Hollingsworth, N. M., Cejka, P., & Hunter, N. (2020). Regulated proteolysis of MutSγ controls meiotic crossing over. *Molecular Cell, 78,* 168-183.

Heller, R. C., & Marian, K. J. (2006). Replication fork reactivation downstream of a blocked nascent leading strand. *Nature, 439*, 557-562.

Hollingsworth, N. M., & Gaglione, R. (2019). The meiotic-specific Mek1 kinase in budding yeast regulates interhomolog recombination and coordinates meiotic progression with double-strand break repair. *Current Genetics, 65,* 631-641.

Hollingsworth, N. M., Ponte, L., & Halsey, C. (1995). MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes and Development, 9,*1728–1739.

Hombauer, H., Campbell, C. S., Smith, C.E., Desai, A. & Kolodner, R. D. (2011). Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell, 147,* 1040–1053.

Hsieh, Y.-Y. P., Makrantoni, V., Robertson, D., Marston, A. L., & Murray, A. W. (2020). Evolutionary repair: changes in multiple functional modules allow meiotic cohesin to support mitosis. *PLoS Biology, 18*, e3000635.

Hunter, N. (2006). Meiotic recombination. *Topics in Current Genetics, 17,* 381–442.

Hunter, N., & Borts, R. H. (1997). Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes and Development, 11,* 1573-1582.

Hunter, N., & Kleckner, N. (2001). The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell, 106,* 59-70.

Jensen, L. E., Jauert, P. A., & Kirkpatrick, D. T. (2005). The large loop repair and mismatch repair pathways of *Saccharomyces cerevisiae* act on distinct substrates during meiosis. *Genetics*, *170*, 1033–1043.

Jessop, L., & Lichten, M. (2008). Mus81/Mms4 endonuclease and Sgs1 helicase collaborate to ensure proper recombination intermediate metabolism during meiosis. *Molecular Cell 31*, 313-323.

Jiricny, J. (2013). Postreplicative mismatch repair. *Cold Spring Harbor Perspectives in Biology, 5*, 1–23.

Johnson, R. E., Kovvali, G.K., Guzder, S. N., Amin, N. S., Holm, C., Habraken, Y., Sung, P., Prakash, L., & Prakash, S. (1996). Evidence for involvement of yeast proliferating cell nuclear antigen in DNA mismatch repair. *Journal of Biological Chemistry, 271,* 27987-27990.

Kadyrov, F. A., Dzantiev, L., Constantin, N., & 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. & Modrich, P. (2007). *Saccharomyces cerevisiae* MutLalpha is a mismatch repair endonuclease. *Journal of Biological Chemistry, 282,* 37181–37190.

Kawasoe, Y., Tsurimoto, T., Nakagawa, T., Masukata, H. & Takahashi, T.S. (2016). MutSα maintains the mismatch repair capability by inhibiting PCNA unloading. *eLife, 5,* e15155.

Keeney, S., Giroux, C. N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell, 88*, 375-84.

Kelman, Z. (1997). PCNA: Structure, functions and interactions. *Oncogene, 14*, 629–640.

Kim, Y., Furman, C. M., Manhart, C. M., Alani, E., & Finkelstein, I. J. (2019). Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex. *Nucleic Acids Research, 47,* 1823-1835.

Kleckner, N. (1996). Meiosis: How could it work? *Proceedings of the National Academy of Sciences, USA, 93*, 8167–8174.

Kleczkowska, H. E., Marra, G., Lettieri, T., & Jiricny, J. (2001). hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes and Development*, *15*, 724–736.

Kolas, N. K., Svetlanov, A., Lenzi, M. L., Macaluso, F. P., Lipkin, S. M., Liskay, R. M., Greally, J., Edelmann, W., & Cohen, P. E. (2005). Localization of MMR proteins on meiotic chromosomes in mice indicates distinct functions during prophase I. *Journal of Cell Biology*, *171*, 447–458.

Kondo, E. (2001). The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Research*, *29*, 1695–1702.

Kulkarni, D. S., Owens, S., Honda, M., Ito, M., Yang, Y., Corrigan, M. W., Chen, L., Quan, A. L., & Hunter, N. (2020). PCNA activates the MutLγ endonuclease to promote meiotic crossing over. *BioRxiv*, 2020.02.12.946020. https://doi.org/10.1101/2020.02.12.946020

Kunkel, T. A., & Erie, D. A. (2015). Eukaryotic mismatch repair in relation to DNA replication. *Annual Reviews of Genetics, 49,* 291-313.

Lahiri, S., Li, Y., Hingorani, M. M., & Mukerji, I. (2018). MutSγ-induced DNA conformational changes provide insights into its role in meiotic recombination. *Biophysical Journal, 115*, 2087-2101.

Lahue, R. S., Au, K. G., & 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., & Sixma, T. X. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature, 407,* 711–717.

Lao, J. P., Cloud, V., Huang, C.-C., Grubb, J., Thacker, D., Lee, C.-Y., Dresser, M. E., Hunter, N., & Bishop, D. K. (2013). Meiotic crossover control by concerted action of Rad51-Dmc1 in homolog template bias and robust homeostatic regulation. *PLoS Genetics*, *9*, e1003978.

Lee, S. D., & Alani, E. (2006). Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA. *Journal of Molecular Biology*, *355*, 175–184.

Lee, J.Y., Terakawa, T., Qi, Z., Steinfeld, J. B., Redding, S., Kwon, Y., Gaines, W. A., Zhao, W., Sung, P., & Greene, E. C. (2015). Base triplet stepping by the Rad51/RecA family of recombinases. *Science, 349,* 977-981.

Lenhart, J. S., Pillon, M. C., Guarné, A., Biteen, J. S., & Simmons, L. A. (2016). Mismatch repair in Gram-positive bacteria. *Research in Microbiology,* 167, 4–12.

Lin, Z., Kong, H., Nei, M., & Ma, H. (2006). Origins and evolution of the *recA/RAD51* gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences, USA*, *103*, 10328–10333.

Lin, Z., Nei, M., & Ma, H. (2007). The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution, *Nucleic Acids Research*, *35,* 7591–7603.

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

Lipkin, S. M., Moens, P. B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., Thomas, J., Cheng, J., Touchman, J. W., Green, E. D., Schwartzberg, P., Collins, F. S., & Cohen, P. E. (2002). Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nature Genetics, 31,* 385–90.

Liu, J., Hanne, J., Britton, B. M., Bennett, J., Kim, D., Lee, J. B. & Fishel, R. (2016). Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature, 539,* 583–587.

Liu, J., Lee, R., Britton, B. M., London, J. A., Yang, K., Hanne, J., Lee, J. B., & Fishel, R. (2019). MutL sliding clamps coordinate exonuclease-independent *Escherichia coli* mismatch repair. *Nature Communications*, *10*, 5294.

Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D., & Hitchins, M. P.(2015). Milestones of Lynch syndrome: 1895-2015. *Nature Reviews Cancer, 15*, 181-194.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., & Steinmetz, L. M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature, 454*, 479-485.

Manhart, C. M., Alani, E. (2016). Roles for mismatch repair family proteins in promoting meiotic crossing over. *DNA Repair 38,* 84-93.

Manhart, C. M., Ni, X., White, M.A., Ortega, J., Surtees, J.A., & Alani, E. (2017). The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLoS Biology, 15,* e2001164.

Marcon, E., & Moens, P. B. (2005). The evolution of meiosis: Recruitment and modification of somatic DNA-repair proteins. *BioEssays*, *27*, 795–808.

Mardenborough, Y., Nitsenko, K., Laffeber, C., Duboc, C., Sahin, E., Quessada-Vial, A., Winterwerp, H., Sixma, T. K., Kanaar, R., Friedhoff, P., Strick, T. R., & Lebbink, J. (2019). The unstructured linker arms of MutL enable GATC site incision beyond roadblocks during initiation of DNA mismatch repair. *Nucleic Acids Research*, *47*, 11667–11680.

Markandona, O., Dafopoulos, K., Anifandis, G., Messini, C.I., Dimitraki, M., Tsezou, A., Georgoulias, P., & Messinis, I. E. (2015). Single-nucleotide polymorphism rs 175080 in the *MLH3* gene and its relation to male infertility. *Journal of Assisted Reproduction and Genetics, 32,* 1795-1799.

Marsolier-Kergoat, M. C., Khan, M. M., Schott, J., Zhu, X., & Llorente, B. (2018). Mechanistic view and genetic control of DNA recombination during meiosis. *Molecular Cell, 70,* 9-20.

Martini, E., Diaz, R. L., Hunter, N., & Keeney, S. (2006). Crossover homeostasis in yeast meiosis. *Cell*, *126*, 285-295.

Matos, J., Blanco, M. G., Maslen, S., Skehel, J. M., & West, S. C. (2011). Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. *Cell, 147,* 158-172.

Modrich, P., & Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Reviews of Biochemistry,* 65, 101–133.

Moens, P. B., Kolas, N. K., Tarsounas, M., Marcon, E., Cohen, P. E., & Spyropoulos, B. (2002). The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination. *Journal of Cell Science, 115*, 1611-22.

Neale, M. J., & Keeney, S. (2006). Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature, *442*, 153–158.

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

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., & Dunker, A. K. (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins*, *53*, 566–572.

Oh, S. D., Lao, J. P., Taylor, A.F., Smith, G. R., & Hunter, N. (2008). RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. *Molecular Cell, 31,* 324-336.

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield, S. E., Zhu, X., Neale, M. J., Jasin, M., Socci, N. D., Hochwagen, A., & Keeney, S. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell, 144*, 719-731.

Pillon, M. C., Lorenowicz, J.J., Uckelmann, M., Klocko, A.D., Mitchell, R. R., Chung, Y. S., Modrich, P., Walker, G. C., Simmons, L. A., Friedhoff, P., & Guarné, A. (2010). Structure of the endonuclease domain of MutL: unlicensed to cut. *Molecular Cell, 39,* 145–151.

Pillon, M. C., Miller, J. H., & Guarné, A. (2011). The endonuclease domain of MutL interacts with the β sliding clamp. *DNA Repair, 10*, 87–93.

Pinto, R. M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., St Claire, J., Panigrahi, G. B., Hou, C., Holloway, K., Gillis, T., Guide, J. R., Cohen, P. E., Li, G. M., Pearson, C. E., Daly, M. J., & Wheeler, V. C. (2013). Mismatch repair genes *Mlh1* and *Mlh3* modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genetics, 9,* e1003930.

Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., & Modrich, P. (2010). PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences, USA, 107,* 16066–16071.

Plys, A. J., Rogacheva, M. V., Greene, E. C., Alani, E. (2012). The unstructured linker arms of Mlh1-Pms1 are important for interactions with DNA during mismatch repair. *Journal of Molecular Biology, 422,* 192–203.

Pochart, P., Woltering, D., & Hollingsworth, N. M (1997). Conserved properties between functionally distinct MutS homologs in yeast. *Journal of Biological Chemistry, 272,* 30345-30349.

Putnam, C. D. (2016). Evolution of the methyl directed mismatch repair system in *Escherichia coli*. *DNA Repair 38*, 32–41.

Pyatnitskaya, A., Borde, V., & De Muyt, A. (2019). Crossing and zipping: molecular duties of the ZMM proteins in meiosis. *Chromosoma, 128*, 181-198.

Raghavan, V., Aquadro, C.F., & Alani, E. (2019). Baker's yeast clinical isolates provide a model for how pathogenic yeasts adapt to stress. *Trends in Genetics, 35,* 804-817.

Ramesh, M. A., Malik, S. B., & Logsdon, J. M. (2005). A phylogenomic inventory of meiotic genes: Evidence for sex in Giardia and an early eukaryotic origin of meiosis. *Current Biology*, *15*, 185–191.

Ranjha, L., Anand, R., & Cejka, P. (2014). The *Saccharomyces cerevisiae* Mlh1-Mlh3 heterodimer is an endonuclease that preferentially binds to Holliday Junctions. *Journal of Biological Chemistry 289,* 5674–5686.

Räschle, M., Dufner, P., Marra, G. & Jiricny, J. (2002). Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *Journal of Biological Chemistry, 277,* 21810–21820.

Rayssiguier, C., Thaler, D. S., & Radman, M. (1989). The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature*, *342*, 396–401.

Reenan, R. A., & 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.

Robine, N., Uematsu, N., Amiot, F., Gidrol, X., Barillot, E., Nicolas, A., & Borde, V. (2007). Genome-wide redistribution of meiotic double-strand breaks in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, *27*, 1868–1880.

Rogacheva, M.V., Manhart, C.M., Chen, C., Guarne, A., Surtees, J., & Alani, E. (2014). Mlh1-Mlh3, A meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *Journal of Biological Chemistry, 289,* 5664–5673.

Romanova, N.V., & Crouse, G.F. (2013). Different roles of eukaryotic MutS and MutL complexes in repair of small insertion and deletion loops in yeast. *PLoS Genetics, 9,* e1003920.

Ross-Macdonald, P., & Roeder, G. S. (1994). Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell, 79,* 1069-1080.

Sacho, E.J., Kadyrov, F.A., Modrich, P., Kunkel, T.A., & Erie, D.A. (2008). Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha. *Molecular Cell, 29,* 112-121.

Sanchez, A., Adam, C., Rauh, F., Duroc, Y., Ranjha, L., Lombard, B., Mu, X., Loew, D., Keeney, S., Cejka, P., Guérois, R., Klein, F., Charbonnier, J.-B., & Borde, V. (2020). Mechanism *of in vivo* activation of the MutLγ-Exo1 complex for meiotic crossover formation. *BioRxiv*, 2019.12.16.876623. https://doi.org/10.1101/2019.12.16.876623.

Snowden, T., Acharya, S., Butz, C., Berardini, M., & Fishel, R. (2004). hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Molecular Cell, 15,* 437-451.

Sourirajan, A., & Lichten, M. (2008). Polo-like kinase Cdc5 drives exit from pachytene during budding yeast meiosis. *Genes and Development, 22*, 2627-2632.

Steinfeld, J. B., Beláň, O., Kwon, Y., Terakawa, T., Al-Zain, A., Smith, M. J., Crickard, J. B., Qi, Z., Zhao, W., Rothstein, R., Symington, L. S., Sung, P., Boulton, S. J., & Greene, E. C. (2019). Defining the influence of Rad51 and Dmc1 lineage-specific amino acids on genetic recombination. *Genes and Development*, *33*, 1191–1207.

Su, X. A., & Freudenreich, C. H. (2017). Cytosine deamination and base excision repair cause R-loop–induced CAG repeat fragility and instability in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences, USA*, *114*, E8392–E8401.

Vakirlis, N., Sarilar, V., Drillon, G., Fleiss, A., Agier, N., Meyniel, J.P., Blanpain, L., Carbone, A., Devillers, H., Dubois, K., Gillet-Markowska, A., Graziani, S., Huu-Vang, N., Poirel, M., Reisser, C., Schott, J., Schacherer, J., Lafontaine, I., Llorente, B., Neuvéglise, C., & Fischer, G. (2016). Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Research, 26,* 918-932.

Villeneuve, A. M., & Hillers, K. J. (2001). Whence meiosis? *Cell*, *106*, 647-650.

Wang, T.F., Kleckner, N., & 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. *Proceedings of the National Academy of Sciences, USA, 96,* 13914-13919.

Wang, S., Liu, Y., Shang, Y., Zhai, B., Yang, X., Kleckner, N., & Zhang, L. (2019). Crossover interference, crossover maturation, and human aneuploidy. *Bioessays, 41,* e1800221.

Weßbecher, I.M., & Brieger, A., (2018). Phosphorylation meets DNA mismatch repair. *DNA Repair, 72*,107-114.

Wild, P., & Matos, J. (2016). Cell cycle control of DNA joint molecule resolution. *Current Opinion in Cell Biology*, *40*, 74–80.

Wild, P., Susperregui, A., Piazza, I., Dörig, C., Oke, A., Arter, M., Yamaguchi, M., Hilditch, A. T., Vuina, K., Chan, K.C., Gromova, T., Haber, J. E., Fung, J. C., Picotti, P., & Matos, J. (2019) Network rewiring of homologous recombination enzymes during mitotic proliferation and meiosis. *Molecular Cell, 75,* 859-874.

Worth, L., Clark, S., Radman, M., & Modrich, P. (1994). Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proceedings of the National Academy of Sciences, USA*, *91*, 3238–3241.

Zakharyevich, K., Tang, S., Ma, Y., & Hunter, N. (2012). Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. *Cell, 149,* 334-347.

CHAPTER 2

Evolution of pathway specificity for the highly divergent Mlh3 and Pms1 mismatch repair factors.

Abstract

Gene duplication events provide an important path to evolve novel protein functions. In baker's yeast there are many examples of duplicated genes, with the MLH family of mismatch repair proteins showing a particularly dramatic expansion of functions; Mlh1-Pms1 acts to excise DNA misincorporation errors whereas Mlh1-Mlh3 acts in its major role to resolve double Holliday junctions into crossovers that facilitate the Meiosis I division. Mlh3 and Pms1 share only 21% amino acid identity, with much of the conservation in the ATP binding and endonuclease domains, thus providing a difficult challenge to reconstruct an ancient divergence. We set out to understand how paralogs derived from an ancestral family acquire new functions, focusing on Pms1 and Mlh3 because phylogenetic analyses indicated that Mlh3 and Pms1 had split into sister groups. Through this approach we identified residues in baker's yeast Mlh3 critical for its meiotic functions. Sites in Mlh3 were changed to the Pms1 family member equivalent in conserved positions located outside of motifs found in all MLH family members. We also made changes in sites that are conserved in one family but missing from the other, and constructed Mlh3/Pms1 chimeras. The resulting strains showed phenotypes similar to *mlh3* hypomorph or null alleles, and in some cases showed phenotypes stronger than the *mlh3* null, providing evidence that all three domains in the MLH protein family are critical for conferring pathway specificity. Importantly, *mlh3* ATP binding and endonuclease domain alleles improved MMR functions in *pms1Δ* strains without disrupting meiotic functions, indicating an expansion of Mlh3 functions, and suggesting that MLH proteins are capable of having both mismatch repair and crossover functions. This strategy provides an approach to understand how paralogs have evolved to support distinct cellular processes.

Introduction

Gene duplications, occurring through events such as polyploidization or unequal crossing over, can create new gene families that play important roles in adaptive evolution [1,2]. In baker's yeast there are many examples of duplicated genes, some that function redundantly, whereas others have specialized roles [3]. Various models have been proposed to explain how novel protein functions have evolved after gene duplication, with an attractive model proposing the presence of an ancestral gene capable of performing two distinct functions, with the duplicated genes being subjected to purifying selection, resulting in gene specialization [2]. Such a model can be difficult to test because the duplication event leading to specialization is typically ancient and the resulting genes display significant sequence divergence. As described below, we studied the outcome of gene duplication events that diversified the functions of the baker's yeast MLH family of mismatch repair genes. These events predated the whole genome duplication event that took place ~ 100 million years ago [3,4]. First, we describe the MLH family proteins and their roles in MMR and meiotic crossing over, and then present our approach to understand how the specialization process unfolded.

Mismatch repair (MMR) is a highly conserved mechanism that reduces the genome mutation rate through the action of MutS homolog (MSH) proteins that bind to base-base and insertion/deletion mismatches that form as the result of DNA replication errors, and MutL homolog (MLH) proteins that transmit this recognition signal to downstream repair events (Fig2. 1A). Genetic and biochemical analyses in eukaryotic systems have identified MSH and MLH factors that have evolved new functions [5-7]. For example, in baker's yeast, the MSH family members Msh2-Msh6 and Msh2-Msh3 act in the repair of different subsets of mismatches, Msh1 acts in mitochondrial DNA metabolism, and Msh4-Msh5 promotes crossover formation in

meiosis [8-10]. For the MLH family, Mlh1-Pms1 plays a major role in MMR, whereas Mlh1-Mlh2 and Mlh1-Mlh3 display minor and more specialized MMR roles. In meiosis, Mlh1-Mlh2 acts to regulate gene conversion tract length and Mlh1-Mlh3 acts in the biased cleavage of double-Holiday junctions to form crossovers [9,11-15]. While studies have been performed to understand how protein functionalization is accomplished for mismatch recognition in MSH family complexes, few, if any have been performed for the MLH family [13,16,17].

During MMR in *S. cerevisiae*, MSH recognition of mismatches results primarily in the recruitment of the latent endonuclease Mlh1-Pms1 (MutLα), which is activated through interactions with MSH proteins and the DNA replication processivity factor, PCNA. The Mlh1-Pms1 endonuclease is activated to nick the newly replicated daughter strand [18-21], with these nicks serving as substrates for two excision pathways, one of which is dependent on the 5' to 3' Exo1-nuclease to excise DNA containing the mismatch. The resulting single-stranded DNA is repaired through DNA synthesis steps. The second pathway is Exo1-independent and is thought to occur through Mlh1-Pms1 making multiple nicks near the mismatch, allowing for strand displacement and repair synthesis by Polymerase δ or ε. In both pathways, nicks are sealed by DNA ligase to repair DNA and maintain the original template information [22-24]. It is important to note that while Mlh1-Pms1, Mlh1-Mlh2, and Mlh1-Mlh3 have all been implicated in MMR, the Mlh1-Mlh3 endonuclease is recruited primarily by Msh2-Msh3 to repair deletion mispairs, primarily in a backup role [17,25], and Mlh1-Mlh2, which lacks endonuclease activity, acts in an accessory role [13]. Curiously Mlh1-Mlh3 plays a critical role in baker's yeast to resolve recombination intermediates into crossovers. Crossing over (CO) of parental homologs during meiotic prophase facilitates their segregation into viable gametes (Fig 2.1B; [9]). In the absence of at least one crossover event per homolog pair, nondisjunction events

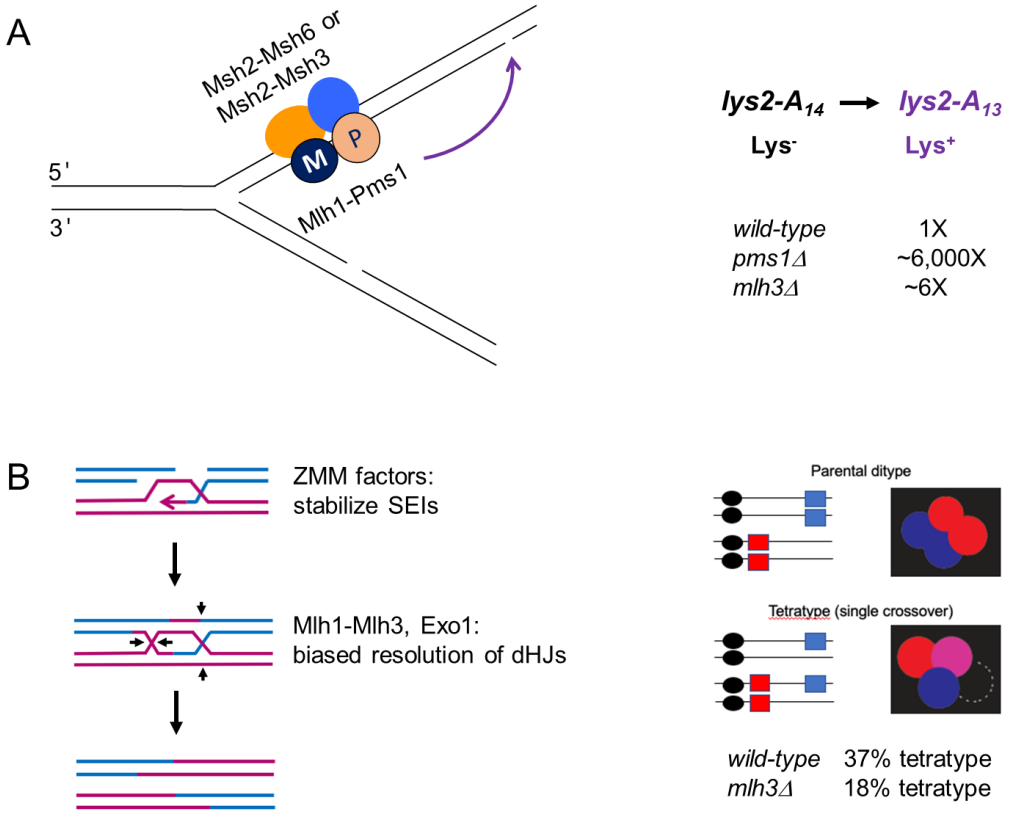


Figure 2.1. Models for Mlh1-Pms1 functioning in MMR and Mlh1-Mlh3 in meiotic crossing over. A. In MMR, MSH complexes bind to DNA mismatches (Msh2-Msh6 primarily base-base and 1 nt insertions/deletions for Msh2-Msh2 and primarily larger insertions/deletions for Msh2-Msh3), and then undergo an ATP-dependent conformational change to form sliding clamps that recruit MLH complexes. Mlh1-Pms1 then undergoes ATP-dependent conformational changes and interacts with PCNA, resulting in incision of the newly replicated DNA strand. This is followed by recruitment of downstream repair factors to excise and resynthesize the DNA and ligate the repaired strand. The Mlh1-Mlh3 complex plays a minor role in MMR and is recruited by Msh2-Msh3 to repair insertion/deletion mismatches. The right panel shows the *lys2-A*$_{14}$ reversion assay to measure defects in MMR (Materials and Methods). Also shown is the effect of *pms1Δ* and *mlh3Δ* mutations on mutation rate using this assay. B. In wild-type yeast induced to undergo meiosis, the Spo11 complex catalyzes double-strand breaks (DSBs) which then undergo 5' to 3' resection. 3' single stranded tails invade the homologous chromosome to form a D-loop intermediate which, in the major pathway for crossover formation, can be further stabilized by ZMM proteins (Zip3, Msh4-Msh5) to enable DNA synthesis and branch migration. A double Holliday junction is then formed that is asymmetrically cleaved in an Mlh1-Mlh3-dependent step to yield primarily crossover products [9,11]. The right panel shows the spore autonomous fluorescence assay used to measure meiotic crossing over in the *CEN8-THR1* interval of chromosome VIII [44]. Also shown is the effect of the *mlh3Δ* mutation on crossing over using this assay.

occur, leading to the formation of aneuploid gametes. In baker's yeast the meiotic crossover pathway is initiated by the Spo11 complex, which catalyzes 150-200 DNA double strand breaks genome wide [26,27]. These DSBs are resected in a 5' to 3' direction to form 3' single stranded tails. These 3' tails, through actions requiring single-strand DNA binding and RecA family strand exchange proteins, invade the homologous template to create a D-loop intermediate. In the major crossover pathway, the D-loop is further stabilized by ZMM proteins (Zip1-4, Msh4-Msh5, Mer3, Spo16) to form a single-end invasion intermediate and is extended through DNA repair synthesis. Extension is followed by branch migration and second-end capture, ultimately resulting in the formation of a double Holliday junction (dHJ) intermediate that is asymmetrically cleaved in an Mlh1-Mlh3-dependent step to yield primarily crossover products [9,11]. Recent work suggested that Mlh1-Mlh3 endonuclease functions are directed towards newly replicated DNA formed to create the dHJ intermediate in a mechanism analogous to that proposed for the Mlh1-Pms1 nuclease during MMR [28-31]. In this model Mlh1-Mlh3 does not act as a structure-specific nuclease that binds and cleaves Holliday junctions; instead its endonuclease activity is recruited to meiotic recombination intermediates in a mechanism that requires interactions with specific meiotic factors (e.g. the MSH family complex Msh4-Msh5) and the DNA polymerase processivity factor PCNA which presumably aids in the imposition of DNA strand specificity [14,32,33].

Mlh1-Mlh3 plays a critical role in the asymmetric cleavage of double Holliday junctions formed to yield crossovers that are critical for the formation of gametes, indicating that Mlh3 has acquired a distinct functional identity from Pms1 (reviewed in [9]). Such distinct roles are critical in both yeast and humans to maintain genomic stability and prevent chromosome aneuploidy, yet both Pms1 and Mlh3 interact with Mlh1 to form functional endonucleases [9,34].

How did the specialization process that led to the Pms1 and Mlh3 proteins unfold? Can we develop a strategy to identify the critical changes that occurred during the specialization of a protein, reverse them and/or expand the functions of that protein? We employed evolutionary and structure-function analyses to make Mlh3/Pms1 chimeras and *mlh3* alleles that contained blocks of amino acid substitutions in residues that were conserved in Mlh3 but were absent in Pms1 or conserved with a different set of amino acids in Pms1. This work is based on an evolutionary analysis indicating an initial divergence of Mlh1 homologs, followed by another split into Pms1 and Mlh3 sister groups. As described below, these efforts provided evidence that all three functional domains of the MLH proteins have evolved for gene specialization. Importantly, a small number of mutations were created in *MLH3* that expanded its MMR specificity without disrupting its role in meiotic crossing over. This combination of approaches provides a strategy to understand how organisms have evolved paralog complexes with distinct cellular functions.

Results

Experimental rationale

MLH family proteins contain three structurally conserved regions; an N-terminal ATP binding domain that facilitates conformational changes, an intrinsically disordered linker arm, and a C-terminal domain containing a region required for dimerization with other MLH proteins. Most MLH proteins also contain a functional endonuclease domain within the C-terminal region (Fig 2.2A; 2.2C; [35-38]. While showing an overall organizational similarity, Mlh1, Mlh3, and Pms1 proteins display significant amino acid divergence. For example, baker's yeast Mlh3 (715 amino acids in length) and Pms1 (873 amino acids in length) display limited (21.2%) amino acid identity over a 966 amino acid alignment (35.6% gaps, EMBOSS Needle;

https://www.ebi.ac.uk/Tools/psa/emboss_needle/), with the greatest divergence seen within the linker domain. While the Mlh3 linker can be weakly aligned to the Pms1 linker (15.9% identity), this alignment contains many gaps that are difficult to align because the Pms1 linker is significantly longer (~2.5-fold; Fig 2.2B).

Phylogenetic analysis supports the divergence of Mlh1 (outgroup) first, followed by another split into Pms1 and Mlh3 as sister groups

To better understand the evolutionary organization of Mlh1, Mlh3 and Pms1, we performed a phylogenetic tree analysis of homologs of these proteins from 34 fungal species, using *E. coli* MutS and *B. subtilis* MutL to root the tree (Fig 2.3; Materials and Methods; [39]). This analysis includes species both before and after the whole-genome duplication event (~100 million years ago) which took place after the split of the yeast species into the *Kluyveromyces* and *Saccharomyces* clades, indicating that Mlh1, Mlh3 and Pms1 predated the whole genome duplication event [3]. Our analysis supports the divergence of Mlh1 (outgroup) first, followed by another split into Pms1 and Mlh3 as sister groups. While the trees are consistent with Mlh3 and Pms1 sharing a more recent common ancestor and function than with Mlh1, it is difficult to definitely conclude directionality because the events are ancient. To further scrutinize the inferred divergence order, we analyzed a select subset of more conserved and confidently aligned amino acid sites. Specifically, we used the GBlocks program to select more conserved blocks of the alignment lacking insertion and deletion events. The phylogenetic tree inferred with those 212 sites also showed Mlh1 diverging first, followed by Mlh3 and Pms1, and those relationships were supported by high approximate likelihood ratio test branch support values (Fig 2.4).

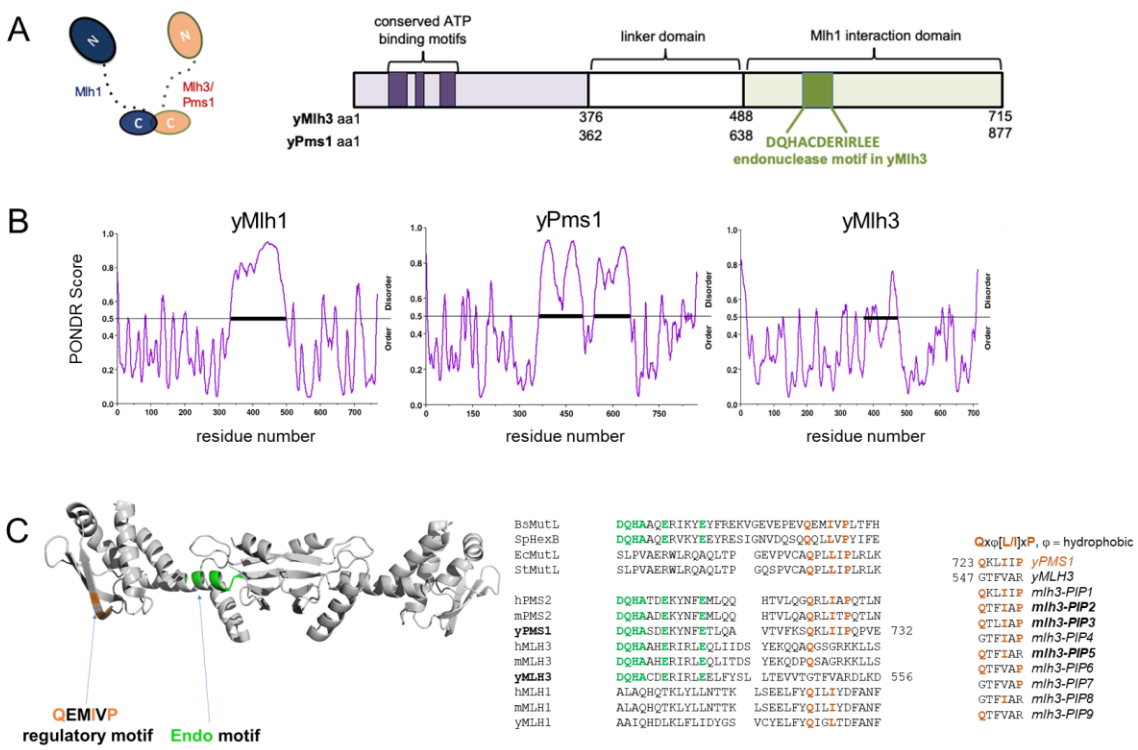


Figure 2.2 Functional domains of yeast Mlh3. A. Cartoon depictions of the Mlh1-Mlh3 and Mlh1-Pms1 complexes, with the N-terminal ATP binding and C-terminal endonuclease/Mlh1 interaction domains separated by intrinsically disordered linker domains (left panel). The amino acid locations of the domains in yeast Mlh3 and Pms1 are shown (right panel). The organization of Mlh3/Pms1 chimeras was based on sequence homology with Mlh1-Pms1, and crystallographic analysis of the ATP binding and C-terminal domains of Mlh3 proteins (see text). B. PONDR analysis of yeast Mlh1, Pms1, and Mlh3 illustrates the locations of the intrinsically disordered linker domains, which are highlighted by the black bars (see Materials and Methods). C. Ribbon diagram (left panel) of the C-terminal domain of *B. subtilis* MutL, with the regulatory (QEMIVP) and endonuclease motifs highlighted ([43]; PDB 3KDK, RCSB Protein Data Bank, modeled using PyMOL). Sequence alignments (right panel) of the endonuclease (green) and PCNA binding motifs (brown) found in a subset of MLH proteins. Sequences for the yeast *mlh3-PIP* mutations analyzed in this study are also shown. Bacterial MutL protein sequences are indicated as follows: Bs, *Bacillus subtilis*; Sp, *Streptococcus pneumoniae*; Ec, *E. coli*; St, *Salmonella typhimurium*. The eukaryotic MLH1, PMS1 (PMS2 in humans) and MLH3 protein sequences are indicated with an h for human, m for mouse, and y for baker's yeast.

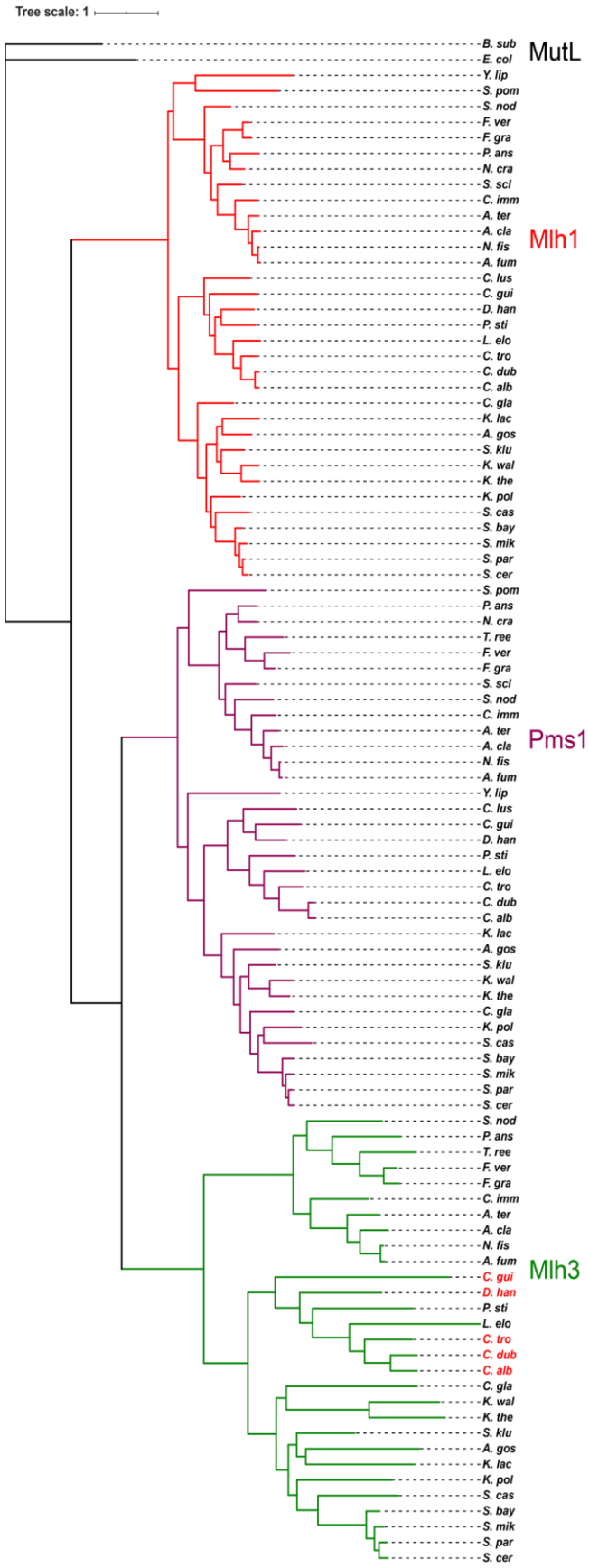


Figure 2.3 Evolutionary analysis of Mlh1, Pms1 and Mlh3 homologs from 34 fungal species. Mlh1 (red branches), Pms1 (orange branches), and Mlh3 (green branches) homologs from 34 fungal species were exported for phylogenic analysis (Materials and Method). *E. coli* MutS and *B. subtilis* MutL were used to root the tree. The scale bar indicates the number of changes per amino acid site. Abbreviations for the fungal species that the Mlh1, Mlh3 and Pms1 homologs were derived from are as follows: *Saccharomyces cerevisiae (Scer), Aphis gossypii (Agos), Candida albicans (Calb), Candida dubliniensis (Cdub), Candida glabrata (Cgla), Candida guilliermondii (Cgui), Candida lusitaniae (Clus), Candida tropoicalis (Ctro), Debaryomyces hansenii (Dhan), Kluyveromyces lactis (Klac), Kluyveromyces polysporus (Kpol), Kluyveromyces thermotolerans (Kthe), Kluyveromyces waltii (Kwal), Saccharomyces bayanus (Sbay), Saccharomyces castellii (Scas), Saccharomyces kluyveri (Sklu), Saccharomyces mikatae (Smik), Saccharomyces paradoxus (Spar), Lodderomyces elongisporus (Lelo), Pichia stipitis (Psti), Yarrowia lipomyces (Ylip), Aspergillus fumigatus (Afum), Neosartorya fischeri (Nfis), Aspergillus clavatus (Acla), Aspergillus terreus (Ater), Coccidioides immitis (Cimm), Neurospora crassa (Ncra), Podospora anserina (Pans), Fusarium graminearum (Fgra), Fusarium verticillioides (Fver), Trichoderma reesei (Tree), Sclerotinia sclerotiorum (Sscl), Stagonospora nodorom (Snod),* and *Schizosaccharomyces pombe (Spom).* Species with non-canonical (or lacking) meiosis (*C. alb, C. dub, C. tro, C. gui and D. han*) are highlighted in red in the Mlh3 clade [39].

Interestingly, the data in Fig 2.3 and Fig 2.4 suggest an acceleration of changes in the Mlh3 clade; one explanation for such an acceleration is the neofunctionalization of Mlh3 to act in meiotic recombination. However, it is difficult to make a definitive conclusion on this point because the events are ancient. For example, an alternative explanation is that the apparent acceleration could be due to a portion of Mlh3 deteriorating under less constraint. Estimating selective pressure during that ancient divergence is challenging (e.g. measuring d_N/d_S, the ratio of the number of nonsynonymous to synonymous substitutions per site), because outside of the genus *Saccharomyces* there are so many synonymous site substituions that these are therefore saturated. The data also suggest that Mlh1 remains under stronger constraint relative to Mlh3, which is curious because Mlh1 and Mlh3 are hypothesized to be coevolving as measured by ERC analysis. It is important to note, however, that proteins with different average evolutionary rates can still have correlated changes in their rates over time.

ERC signals are seen between sets of domains for three MLH proteins

Evolutionary rate covariation (ERC), which identifies protein pairs with correlated changes in evolutionary rate, has been used to make functional inferences [39,40]. In general, ERC values between unrelated proteins do not show correlated rate changes, whereas protein pairs in shared Pathways, complexes, and functions show positively correlated rates [40]. ERC is calculated as the correlation coefficient between the phylogenetic branch-specific rates of one protein vs. another. A value of one indicates perfect rate covariation and a value near zero represents little or no covariation. Previous studies had shown an elevated ERC for Mlh1 and Mlh3 with each other and with meiotic crossover proteins [39], but ERC values for Mlh1 and Mlh3 with the mismatch repair specific components Msh2 and Msh6 were not elevated, suggesting that

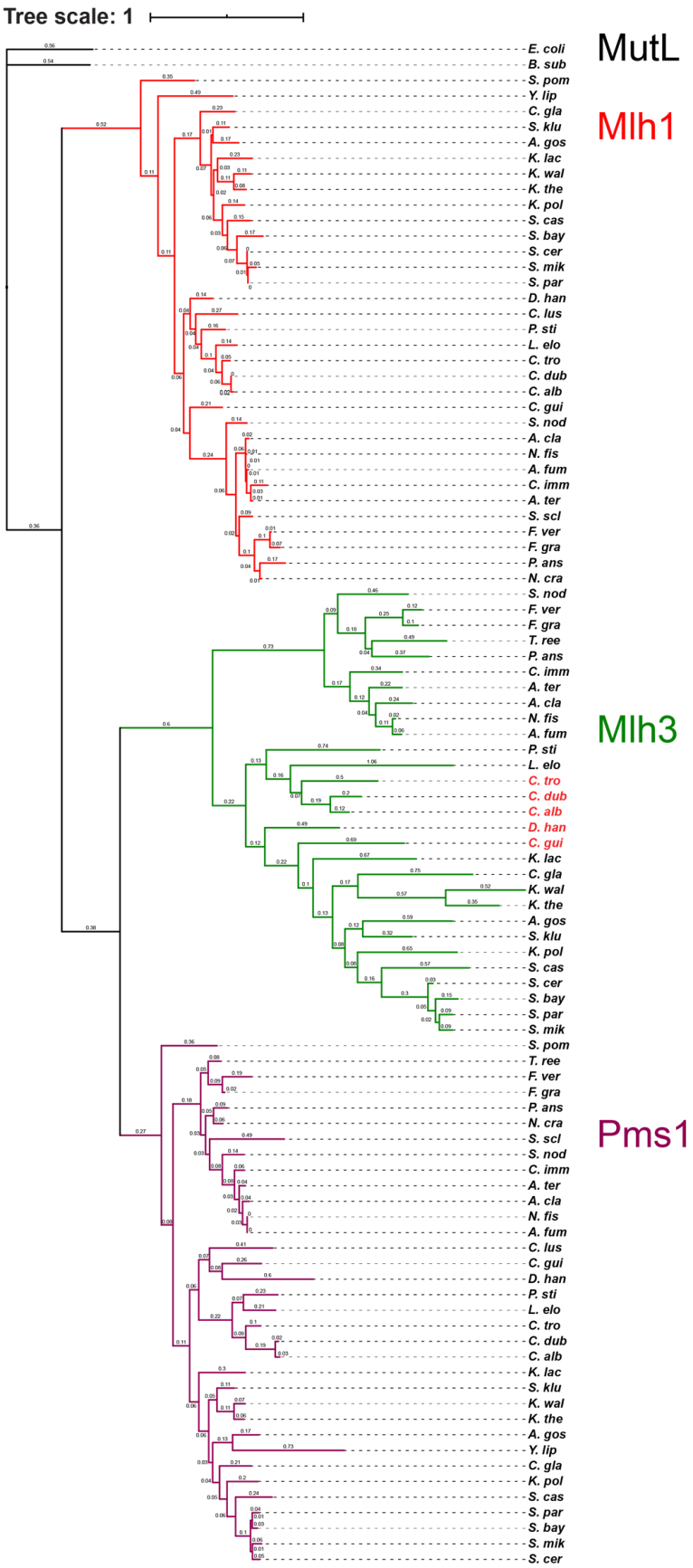


Figure 2.4. Mlh3, Pms1, Mlh1 phylogenetic analysis using GBlocks. A select subset of more conserved and confidently aligned amino acid sites. Specifically, GBlocks [78] was used to select more conserved blocks of the alignment used to create the phylogenetic analysis presented in Fig 2.3. The conserved blocks that were used lacked insertion and deletion events, with the phylogenetic tree inferred from 212 sites. The scale bar indicates the number of changes per amino acid site.

the evolutionary forces relating to meiotic crossing over had a greater effect on Mlh1 and Mlh3 than mismatch repair.

ERC has been used previously to compare rates between full-length protein sequences. In this study, we performed ERC analysis on the whole length proteins as well as the ATP binding (N-terminal), linker, and endonuclease/MLH interaction (C-terminal) domains of Mlh1, Mlh3, and Pms1 using 18 yeast species (including *Saccharomyces cerevisiae*; Materials and Methods; [39]). The purpose of this analysis was to determine if any one specific domain of the MLH proteins was showing ERC with another, or if multiple domains displayed such co-variation as compared to the whole protein. This analysis involved the 12 X 12 matrix presented in Fig 2.5. It is important to note that our new analysis showed slightly lower levels of ERC between Mlh1 and Mlh3 than previously reported [39]. This difference is due to normalizing branch lengths more carefully, resulting in an improved way to compute relative evolutionary rates (RER, Materials and Methods). With this new method we see similarly significant ERC signals between Mlh1 and Mlh3, and Mlh1 and Pms1 ($p < 0.05$). For the sub domains of the MLH proteins, we observed stronger ERC signals between sets of domains of all three proteins. The strongest ERC signals were between the Mlh1-N terminal and both Mlh3 N- and C-terminal domains and between the Mlh3 N-terminal and Mlh3 C-terminal domains. In addition, the Pms1 linker showed strong signals with the Mlh1 N-terminal domain and all three domains of Mlh3. These observations do not show a specific pattern of signals between specific domains of MLH proteins, but are consistent with structural studies indicating that ATP-dependent conformational rearrangements involving the linker regions of the MLH proteins are important for the positioning of the two N-terminal MLH domains and bound DNA near the endonuclease active site [41-43].

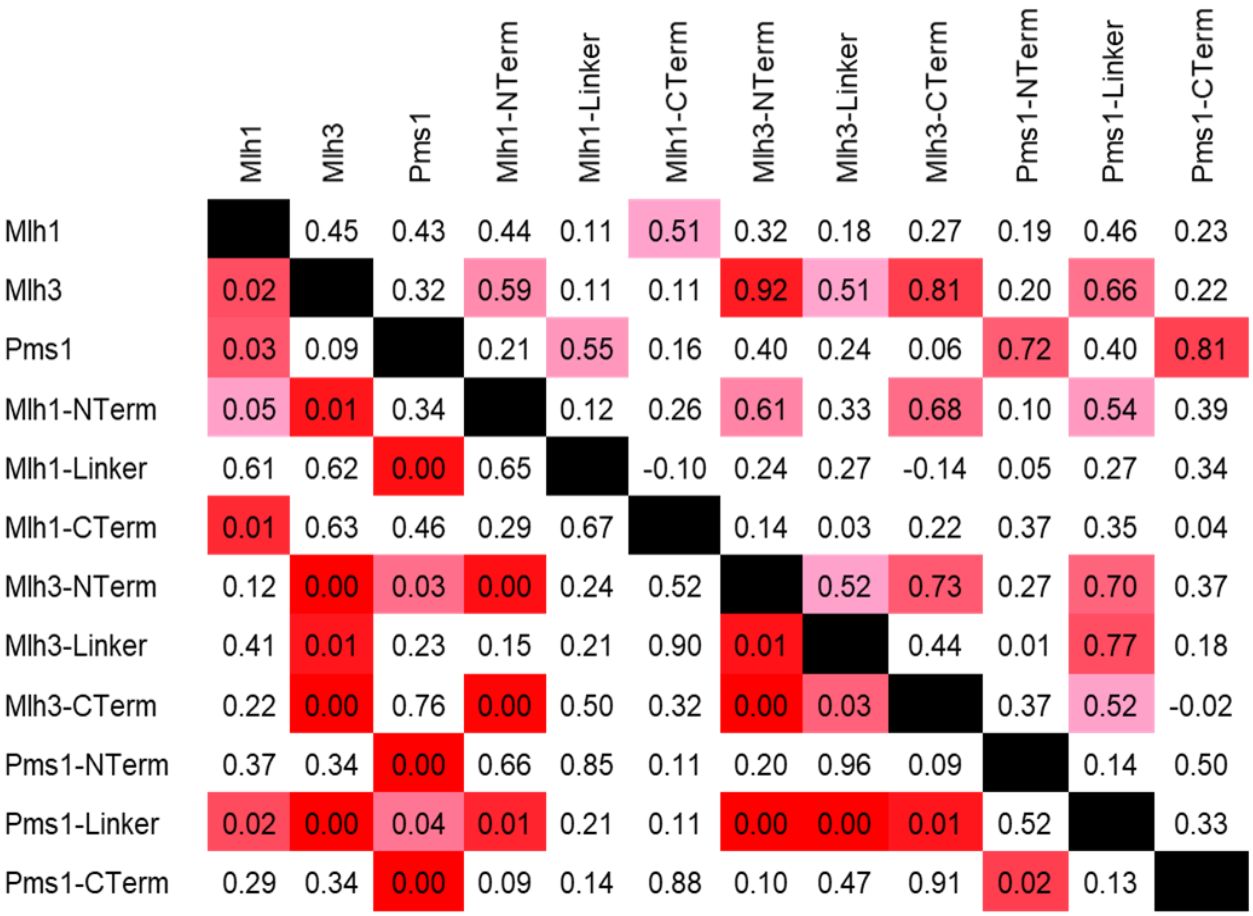


Figure 2.5. ERC is elevated between domains of MLH proteins. This pairwise matrix shows all comparisons between the full length Mlh1, Mlh3 and Pms1 proteins and their ATP binding (abbreviated as Nterm), linker, and endonuclease/MLH interaction (abbreviated as Cterm) domains. ERC values are above the diagonal and empirical P-values are below. The colors of ERC cells range from pink at values of 0.5 to red at 1.0. P-value cells are pink at 0.05 and become red as they approach zero.

Table 2.1. Complementation of *MLH3/PMS1* chimera and *mlh3* alleles in mismatch repair and meiotic crossing over (CO).

	Mismatch Repair			Meiotic CO	Phenotype	
Genotype	Rate x 10^{-6} (n)	95% CI. X 10^{-6}	Relative to WT	% tetratype	MMR	CO
MLH3	1.03 (42)	0.81-1.39	1	37.1 (1023)	+	+
mlh3Δ	6.24 (39)	4.53-8.51	6.05	18.3 (1239)	-	-
***MLH3-PMS1* chimeras**						
PPP (PMS1)	8.29 (15)	2.85-22.2	8.1	17.8 (549)	-	-
PMM	15.6 (15)	3.29-29.0	15.1	19.2 (530)	-	-
MMP	19.1 (15)	12.9-25.4	18.6*	21.4 (524)	- -	-
PPM	15.9 (15)	11.8-128	15.4*	19.8 (824)	- -	-
MPP	15.4 (15)	12.4-32.3	14.9*	22.3 (837)	- -	-
PMP	17.2 (15)	13.1-18.2	16.7*	20.7 (1,287)	- -	-
MPM	8.75 (15)	5.11-11.4	8.49	21.4 (1,196)	-	-
MPM-19	not tested			18.8 (449)		-
MPM-40	not tested			22.4 (540)		-
MPM-60	not tested			19.3 (507)		-
MPM-80	not tested			21.1 (508)		-
MPM-100	not tested			23.0 (521)		-
MPM-N-full	not tested			21.4 (500)		-
MPM-middle-full	not tested			18.6 (560)		-
MPM-C-full	not tested			20.9 (521)		-
***mlh3* Block mutations**						
Block 1, ATP binding						
mlh3-K17T,A20Q,S24D, R30K,Q34D	2.26 (15)	1.18-3.78	2.2	35.6 (513)	+/-	+
Block 2, Mlh1 interaction						
mlh3-Y493M,N497G,V499F, D500N,K502G	3.71 (15)	2.04-4.98	3.6	29.1 (769)	+/-	+/-
mlh3-D500N	1.73 (15)	1.33-6.98	1.68	33.1 (801)	+/-	+
mlh3-K502G	2.84 (15)	2.13-4.76	2.76	19.1 (761)	+/-	-
Block 3, Endonuclease motif						
mlh3-R530K	2.38 (15)	1.61-3.28	2.31	25.3 (771)	+/-	+/-
mlh3-R532N	1.25 (15)	0.91-2.72	1.21	29.5 (774)	+	+/-
mlh3-R530K,R532N	4.49 (15)	3.64-7.24	4.36	20.4 (509)	-	-
Block 4, PCNA interaction motif						
mlh3-PIP1	3.49 (15)	1.96-4.35	3.39	30.9 (742)	-	+/-
mlh3-PIP2	0.567 (15)	0.41-0.75	0.55	37.1 (792)	+	+

Table 2.1 Cont'd						
Block 5, Helix 2 *mlh3-V660K,N666A,F676I,D678K*	4.44 (15)	2.45-12.8	4.32	18.2 (760)	-	-
mlh3-D678K	0.831 (15)	0.57-1.18	0.81	32.1 (772)	+	+/-
Block 6, Helix1 *mlh3-C695L,F699W,A702P, S707T,V709R,P710H*	3.38 (15)	2.37-4.76	3.28	19.5 (527)	+/-	-
Mutant combinations Block 1 (*ATP binding*), Block 4 (*PIP2*)	2.56 (15)	1.95-4.05	2.48	31.6 (509)	+/-	+/-
Block 5 (*Helix 2*), Block 6 (*Helix 1*)	4.11 (15)	2.40-5.28	3.99	17.9 (514)	+/-	-
S288C background						
wild-type + empty vector	0.28 (15)	0.19-0.47	**1**		+	
mlh3Δ+ empty vector	1.61 (15)	1.37-2.08	5.9		-	
mlh3Δ + pMLH3$_{SK1}$ ARS-CEN	0.27 (15)	0.18-0.39	0.98		+	
mlh3Δ + pMLH3$_{SK1}$-2μ	223 (15)	209-291	81.8*		- -	

The indicated chimeras and *mlh3* alleles (Figs. 2.2C; 2.6; 2.7; Table 2.3) were integrated into the *MLH3* locus in the SK1 strain background and tested for DNA mismatch repair functions using the *lys2-A$_{14}$* reversion assay with 95% CI (confidence interval) presented, and for meiotic crossover functions using a spore autonomous assay that measures genetic map distances in the *CEN8-THR1* interval on Chromosome VIII (Materials and Methods; Table 2.4). For the S288c background experiments, EAY1269 (*wild-type*, S288c background) was transformed with pRS415 (empty vector). and EAY4595 (*mlh3Δ*) was transformed with pRS415 (empty vector), *MLH3$_{SK1}$-ARS-CEN* (pEAA566), and *MLH3$_{SK1}$-2 μ* (pEAM65; Table 2.3). n represents the number of independent measurements from at least two transformants. WT, wild type. +, indistinguishable from *MLH3* as measured by 95% CI or $\chi 2$ ($p<0.0001$ for % tetratype). -, indistinguishable from *mlh3Δ* as measured by 95% CI or or $\chi 2$ ($p<0.0001$). +/- distinguishable from both *MLH3* and *mlh3Δ* as measured by 95% CI or $\chi 2$ ($p<0.05$). *mutation rate higher than *mlh3Δ*; illustrated as a - - phenotype.

Chimeric Mlh3/Pms1 proteins can interfere with Mlh3-dependent MMR

The above evolutionary analyses encouraged us to ask if we could disrupt the specificity of the Mlh1-Pms1 and Mlh1-Mlh3 complexes by performing domain swaps or swaps of residues between Mlh3 and Pms1 that are specifically conserved in Pms1 or Mlh3 family proteins, but not in both. We tested for altered specificities in MMR in *mlh3Δ* and *pms1Δ* strains using a highly sensitive *lys2-A*$_{14}$ reporter containing a centrally located homopolymeric run of 14 deoxyadenosine residues that result in a +1-frameshift mutation (Fig 2.1A; [22]). DNA slippage events that restore the reading frame, primarily the result of -1-frameshift mutations, confer reversion to Lys^{+} (Materials and Methods). We measured meiotic crossing over using a spore-autonomous fluorescence assay to measure percent tetratype at the *CEN8-THR1* interval (Fig 2.1B; [44]).

Six Mlh3/Pms1 chimeric proteins were constructed to assess the role of specific MLH domains in MMR and meiotic crossing over (abbreviated in Fig 2.6 as PMM, MMP, PPM, PMP, MPP, MPM). Specifically, Mlh3/Pms1 chimeras were assessed for their ability to complement Mlh3's major function in crossing over and minor function in MMR (Table 2.1), and their ability to complement Pms1's major function in MMR (Table 2.2). The end points for the domain swaps are presented in Fig 2.2B. We found that none of the chimeras complemented either MMR or meiotic crossing over functions (Fig 2.6; Tables 2.1; 2.2), suggesting that critical specificity-determining residues in all three domains are required for functions in MMR and meiotic crossing over. These observations are consistent with previous work showing that the highly conserved ATP binding, linker and endonuclease/Mlh1 interaction domains of both Mlh1 and Mlh3 are required for MMR and meiotic crossover functions [45-48].

Al-Sweel et al. [46] identified *mlh3-32* as a separation of function allele in Mlh3's linker that conferred a defect in crossing over, but not MMR. Claeys Bouuaert and Keeney [49] identified mutations that overlapped with those mutated in *mlh3-32* allele which affected the DNA binding activity of Mlh1-Mlh3, suggesting that DNA binding within the linker region may be important for meiotic functions. To test whether sequences surrounding the *mlh3-32* allele formed a meiosis-specific domain, we substituted increasingly larger regions of wild-type Mlh3 linker centered at the site of the *mlh3-32* allele into the Pms1 linker region present in the MPM chimera (Fig 2.6; Table 2.1). The resulting mutants were examined for their ability to complement meiotic crossing over defects seen in *mlh3Δ* strains using the spore autonomous fluorescence assay. None of the constructs conferred meiotic crossover function, indicating that the spatial organization of the Mlh3 linker with respect to the N- and C-terminal domains was critical for function. These observations are also consistent with the work of Kim et al. [24] who showed that the linker arms of Mlh1-Pms1 require specific amino acid sequence and organizational contexts.

Mlh3/Pms1 chimeric proteins whose N- or C-termini were changed (MMP, PPM, PMP, MPP; Fig 2.6), conferred mutator phenotypes that were higher than seen for *mlh3Δ*. It is likely that these chimeras interfere with other MLH MMR pathways; for example, the chimeras could sequester Mlh1 from interacting with Pms1 in MMR; examples of MLH proteins interfering with MMR functions have been reported previously [50,51]. In summary, the higher mutator phenotypes seen in chimeric proteins suggested that they were expressed, but that no single domain could confer functional specificity.

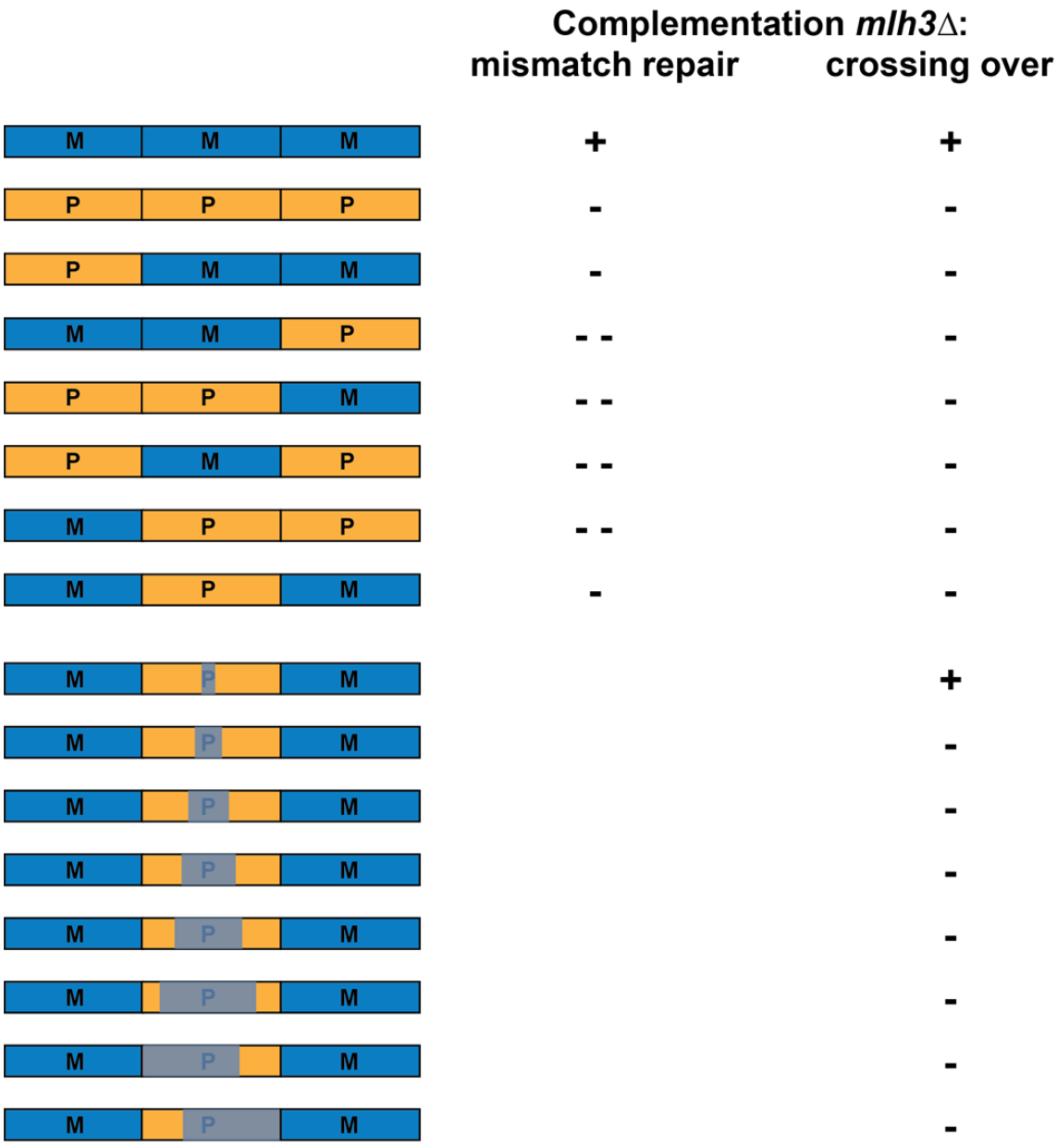


Figure 2.6 Mlh3/Pms1 chimeras analyzed in this study, with MLH3 phenotypes conferred in MMR and meiotic crossing over assays. Chimeras (PMM, MMP, PPM, MPP, MPM, PMP; P= Pms1 domain, M = Mlh3 domain) were constructed using the domain boundary definitions shown in Fig 2.2A (Materials and Methods). In addition, increasing amounts of the Mlh3 linker domain (shown in light blue shading and in detail in Fig 2.10) were inserted into the MPM construct as illustrated. Chimeras were expressed from the native *MLH3* promoter and were tested for *MLH3* functions in mismatch repair (MMR) using a *lys2-A*$_{14}$ reversion assay, and meiotic crossing in an interval in the *CEN8-THR1* interval using a spore autonomous fluorescence microscopy assay (Fig 2.1B; [44]). Quantitative data are presented in Tables 2.1 and 2.2. +/- phenotype, distinct from wild-type (+) or null (-); - - phenotype, higher mutation rate compared to *mlh3Δ*.

Table 2.2. *mlh3-ATP binding and mlh3-PIP* mutations modestly rescue *pms1* null mismatch repair defects.

pms1Δ strain with indicated plasmids	rate x 10^{-7} (n)	95% CI x 10^{-7}	relative to wild type
PMS1, ARS-CEN	2.99 (30)	2.11-3.85	**1**
PMS1, 2μ	22.4 (18)	15.1-63.5	7.5
ARS-CEN (LEU2)	20,300 (29)	12,900-28,700	6,800
2 μ	18,900 (18)	15,900-21,300	6,320
MLH3-PMS1* chimeras, *2 μ			
MMM	13,200 (28)	8,280-20,600	4,430
PPP	42.9 (18)	24.1-91.4	14.4
PMM	19,600 (18)	8,530-31,200	6,550
MMP	14,600 (18)	11,600-21,100	4,870
PPM	16,100 (18)	12,900-18,500	5,380
MPP	14,800 (18)	11,800-18,300	4,960
PMP	15,300 (18)	12,800-21,300	5,130
MPM	20,400 (18)	18,100-35,100	6,830
mlh3* Block mutations, *2 μ			
Block 1, ATP binding			
mlh3-K17T, A20Q, S24D, R30K, Q34D	9,720 (18)	6,360-11,300	3,250*
Block 2, Mlh1 interaction			
mlh3-Y493M, N497G, V499F, D500N, K502G	15,400 (18)	10,700-23,700	5,160
mlh3-D500N	10,900 (18)	9,540-16,100	3,630
mlh3-K502G	16,500 (18)	10,200-19,100	5,530
Block 3, Endonuclease motif			
mlh3-R530K	20,500 (18)	16,900-28,600	6,840
mlh3-R532N	15,700 (18)	10,600-21,600	5,240
mlh3-R530K, R532N	13,300 (18)	10,100-17,200	4,440
Block 4, PCNA interaction motif			
MLH3 (GTFVAR)	13,200 (28)	8,280-20,600	4,430
mlh3-PIP1 (QKLIIP)	13,700 (25)	7,940-17,200	4,590
mlh3-PIP2 (QTFIAP)	6,300 (44)	4,120-7,470	2,110*
mlh3-PIP3 (QTLIAP)	8,520 (18)	6,950-17,600	2,850
mlh3-PIP4 (GTFIAP)	13,800 (18)	7,390-22,400	4,610
mlh3-PIP5 (QTFIAR)	6,890 (18)	4,250-9,890	2,300*
mlh3-PIP6 (QTFVAP)	15,700 (18)	13,100-25,800	5,240
mlh3-PIP7 (GTFVAP)	15,500 (18)	10,500-22,500	5,190
mlh3-PIP8 (GTFIAR)	19,500 (18)	15,100-25,900	6,510
mlh3-PIP9 (QTFVAR)	8,520 (18)	4,890-10,800	3,060*
Block 5, Helix 2			
mlh3- V660K, N666A, F676I, D678K	14,900 (18)	10,700-19,700	4,970
mlh3-D678K	14,700 (18)	11,900-22,200	4,920

Table 2.2 cont'd			
Block 6, Helix 1			
mlh3-C695L, F699W, A702P, S707T, V709R, P710H	14,100 (18)	12,100-21,100	4,710
Double Mutants			
Block 1, Block 4 (*PIP2*)	5,480 (18)	4,340-6,710	1,830*
Block 2 (*D500N),* Block 4 *(PIP2*)	12,200 (18)	6,340-15,100	4,078
Block 2 (*K502G),* Block 4 *(PIP2*)	11,900 (18)	6,830-19,100	3,970
Block 3 (*R530K),* Block 4 *(PIP2*)	7,350 (18)	2,490-12,700	2,460*
Block 3 (*R532N),* Block 4 *(PIP2*)	7,110 (18)	3,240-9,650	2,380*
Block 3 (*R530K, R532N),* Block 4 *(PIP2*)	17,600 (18)	13,400-19,800	5,880
Block 4 *(PIP2*), Block 5 (*D678K*)	11,000 (18)	5,750-12,900	3,690*
Block 4 (*PIP2*), Block 5 (*Helix 2*)	18,700 (18)	12,200-30,900	6,240
Block 4 (*PIP2*), Block 6 (*Helix 1*)	20,600 (18)	14,400-27,900	6,904
Block 5 (*Helix 2*), Block 6 (*Helix 1*)	20,200 (18)	17,500-38,800	6,740

EAY3097 (*pms1Δ,S288c background*) was transformed with pJH481 ($PMS1_{S288c}$, *ARS-CEN*), pEAM50 ($PMS1_{S288c}$, *2μ*), and pRS415 (dummy vector; Table 2.3) to analyze *PMS1*, *PMS1-2μ* and *pms1Δ* genotypes, respectively. *MLH3-PMS1* chimera constructs (Fig. 2.6; Table 2.3) *mlh3* substitution alleles (Blocks 1-6, Figures 2.2C; 2.7; Table 2.3) and were also transformed into EAY3097. *mlh3* alleles and chimeras were expressed from an *MLH3* promoter in *2μ* and *ARS-CEN* vectors as indicated. All strains were analyzed for mutation rate using the $lys2\text{-}A_{14}$ reversion assay as described in the Materials and Methods with the 95% confidence interval (CI) presented. n represents the number of independent measurements obtained from at least two transformants. *indicates complementation of *pms1Δ* as measured by non-overlap in 95% CI.

Identifying residues critical for Mlh3 MMR and crossover functions

The phylogenetic analysis indicating Mlh3 and Pms1 as sister groups encouraged us to perform a more targeted approach to identify residues in the MLH proteins required for their specific functions. Because Mlh3 plays a minor role in MMR, and Pms1 does not play a direct role in dHJ resolution, we hypothesized that our best opportunity to study the specificity of the MLH proteins would be to make substitution mutations in *MLH3* that affected its MMR and meiotic crossing over functions. Such substitutions might also allow Mlh3 to substitute at least partially for Pms1 MMR functions (see below). We performed this analysis recognizing that the divergence of Mlh3 and Pms1 is an ancient event, and any changes in specificity would be informative as refinement of protein function would likely require large numbers of changes over a long evolutionary time scale.

We used multi-Harmony analysis as a targeted approach to identify residues in Mlh3 and Pms1 required for MMR function. Multi-Harmony uses multiple sequence alignments between subfamilies of proteins, homology models, and multi-Relief and sequence-Harmony algorithms to identify amino acids that may suggest functional specificity (Materials and Methods; [52]. By aligning Mlh3 and Pms1 amino acid sequences from 34 fungal species (Fig 2.3; Fig 2.7; Fig 2.8A-D), residues were identified that are well conserved within each subfamily but differed between Mlh3 and Pms1. When mapped onto the three-dimensional structure of Mlh1-Pms1 on PyMOL, five clusters of amino acids were identified and labeled as “blocks.” The substitutions in each block are shown in the Mlh3 sequence with the Mlh3 residue shown first, the amino acid position in Mlh3, and the equivalent position in Pms1.

Block 1 maps to the ATP binding domain; Block 2 to the Mlh1 interaction motif, and Blocks 3, 5 and 6, to the MLH endonuclease motif, (Fig 2.7). This analysis also encouraged us

to test single amino acid substitutions with two (D500N, K502G) mapping to Block 2, two (R530K, R532N) to Block 3, and one (D678K) to Block 5. In addition, a Block 4 was created based on previous studies showing that eukaryotic MutSα and MutLα complexes each interact with the DNA replication processivity clamp PCNA through a PIP Box defined as a six amino acid sequence consisting of Qxφ[L/I]xP, where φ is a hydrophobic residue and x is any amino acid ([53-55]**;** Fig 2.2C). For MutSα, this interaction helps to tether at least a subset of MutSα complexes to the replication fork [56]. For MutLα, interaction with PCNA through a PIP-motif present in Pms1 is critical to activate MutLα's endonuclease activity [21, 55]. Mlh1-Mlh3 does not contain this motif, and its endonuclease activity is not stimulated by PCNA under conditions where the Mlh1-Pms1 endonuclease is activated [29,57,58]. The *mlh3* Block 1-6 substitutions were tested for complementing Mlh3 functions in MMR (Table 2.1) and meiotic crossing over, and for their ability to complement Pms1 functions in MMR (Table 2.2). A brief summary of the phenotypes conferred by these phenotypes is shown below.

The *mlh3* Block 1 (ATP binding) allele conferred an intermediate mutator phenotype but did not affect meiotic crossing over.

The *mlh3* Block 2 (Mlh1 interaction) allele conferred an intermediate phenotype in both MMR and meiotic crossing over, but with one allele (*mlh3-K502G*) conferring a more severe defect in the meiotic crossover assay.

On the whole, Block 3, 5 and 6 (Mlh3 endonuclease motifs) alleles conferred intermediate to severe defects in both MMR and meiotic crossover assays, with some alleles conferring a more severe defect in one assay vs. the other. It was interesting to see a more severe defect for the Block 3 *mlh3-R530K,R532N* mutation compared to the single mutations, indicating important roles for the two residues in Mlh3 function.

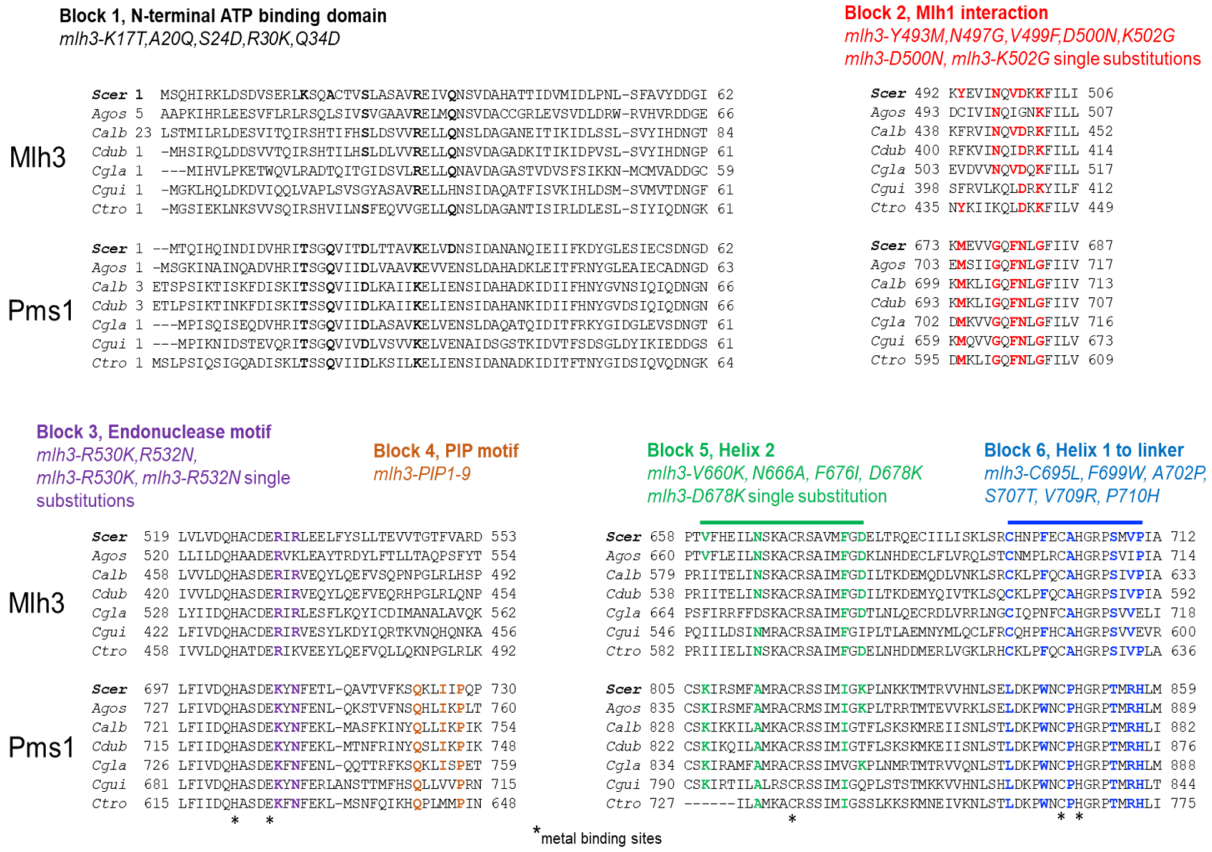


Figure 2.7. Mutations made in *MLH3* to revert back to conserved *PMS1* sequences. Six blocks of mutations were made in Mlh3; Block 1-ATP binding, Block 2- Mlh1 interaction, and Blocks 3 to 6-Endonuclease/PCNA interaction. The multiple amino acid substitutions are shown for each block as well as single substitutions that were made in each region. For Blocks 1, 2, 5, and 6. Multi-Relief and Sequence Harmony algorithms were used to identify functionally specific residues in Pms1 (Materials and Methods). Briefly, Mlh3 and Pms1 amino acid sequences from 34 different fungal species were aligned and presented in multi-Harmony. Shown are alignments of the regions showing 7 fungal species (*S. cerevisiae-Scer, Ashbya gossypii-Agos, Candida albicans-Calb, Candida dubliniensis-Cdub, Candida glabrata-Cgla, Candida guilliermondii-Cgui, Candida tropicalis-Ctro*). The more complete lists of species alignments are shown in Fig 2.3 legend. The N and C-terminal domains of Mlh3 were mapped onto the three-dimensional structure of Mlh1-Pms1, and four amino acid clusters were identified for substitution analysis [52,82]. Block 3 spans the endonuclease motif found in Pms1 and Mlh3 (Fig 2.2C). Block 4 contains the QXLXXP motif important for interactions with PCNA (PIP) which is highly conserved in the Pms1 sequences (>94% identity; [55]) but is absent in Mlh3 sequences. Nine PIP mutations were made as shown in Fig 2.2C and Tables 2.1 and 2.2. The asterisks indicate highly conserved metal binding residues (H703, E707, C817, C848 and H850) in yeast Pms1, which form the endonuclease active site [38].

Figure 2.8 A & B

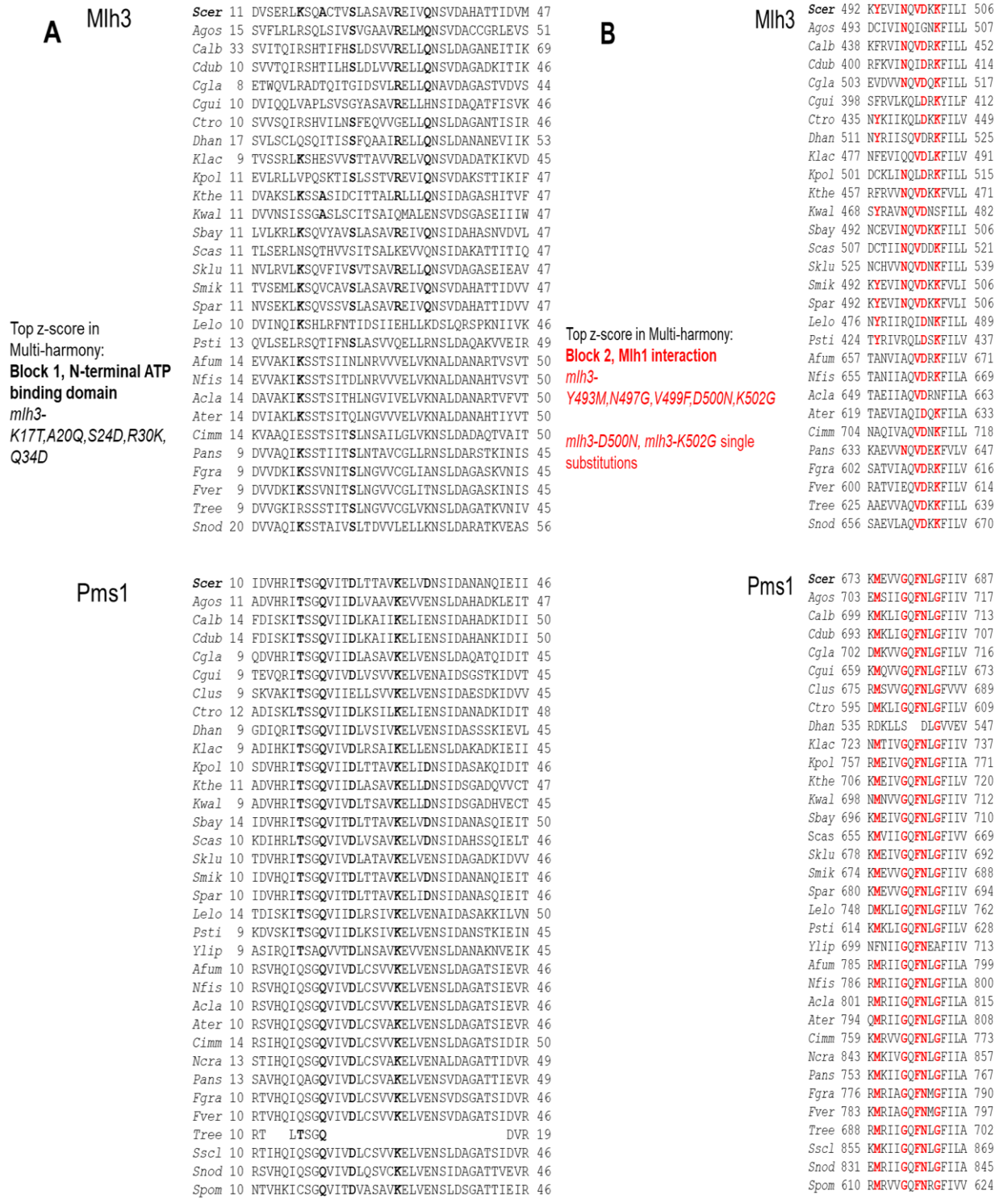
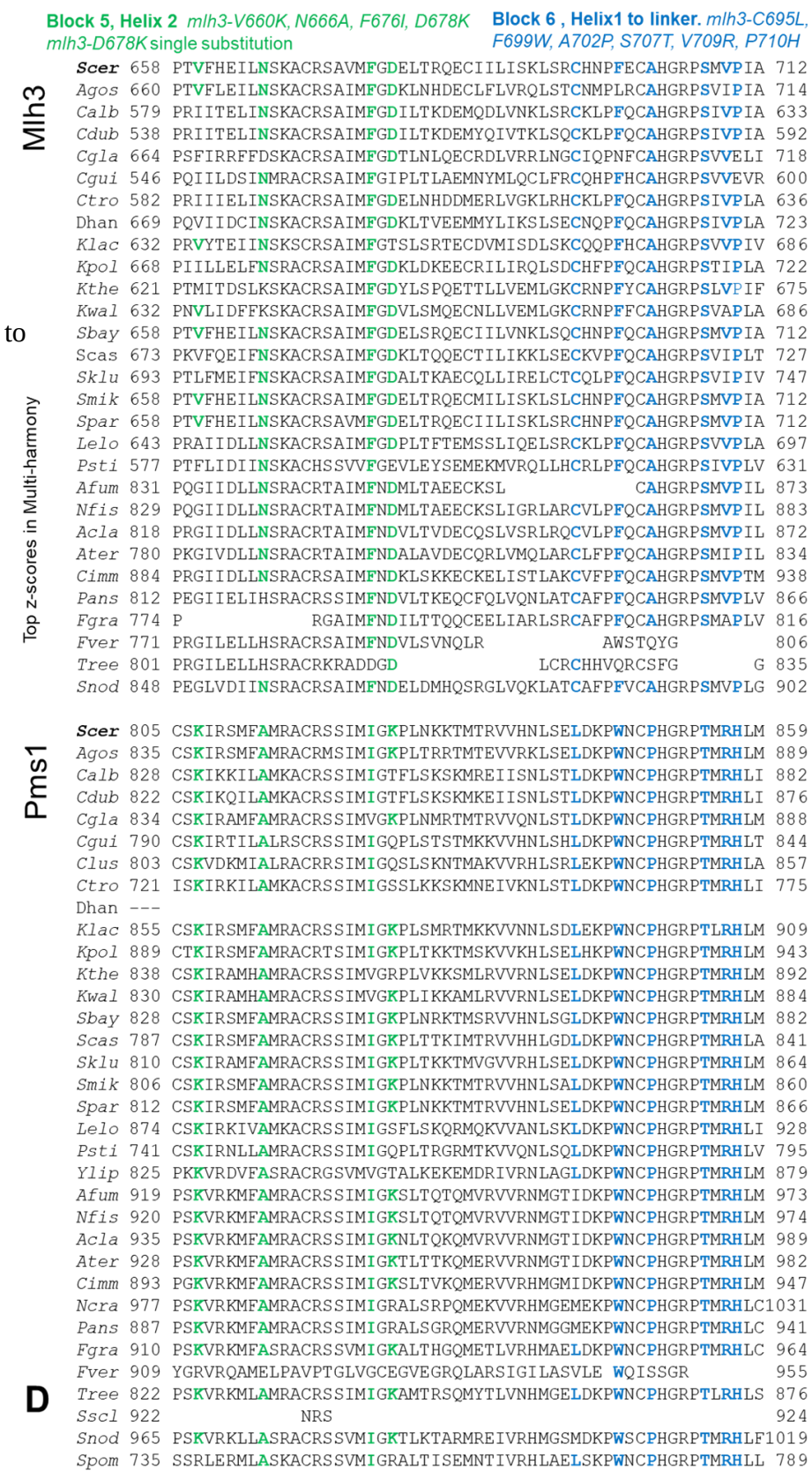

A
Mlh3
Top z-score in Multi-harmony:
Block 1, N-terminal ATP binding domain
mlh3-K17T,A20Q,S24D,R30K,Q34D
Scer 11 DVSERLKSQACTVSLASAVREIVQNSVDAHATTIDVM 47
Agos 15 SVFLRLRSQLSIVSVGAAVRELMQNSVDACCGRLEVS 51
Calb 33 SVITQIRSHTIFHSLDSVVRELLQNSLDAGANEITIK 69
Cdub 10 SVVTQIRSHTILHSLDLVVRELLQNSVDAGADKITIK 46
Cgla 8 ETWQVLRADTQITGIDSVLRELLQNAVDAGASTVDVS 44
Cgui 10 DVIQQLVAPLSVSGYASAVRELLHNSIDAQATFISVK 46
Ctro 10 SVVSQIRSHVILNSFEQVVGELLQNSLDAGANTISIR 46
Dhan 17 SVLSCLQSQITISSFQAAIRELLQNSLDANANEVIIK 53
Klac 9 TVSSRLKSHESVVSTTAVVRELVQNSVDADATKIKVD 45
Kpol 11 EVLRLLVPQSKTISLSSTVREVIQNSVDAKSTTIKIF 47
Kthe 11 DVAKSLKSSASIDCITTALRLLLQNSIDAGASHITVF 47
Kwal 11 DVVNSISSGASLSCITSAIQMALENSVDSGASEIIIW 47
Sbay 11 LVLKRLKSQVYAVSLASAVREIVQNSIDAHASNVDVL 47
Scas 11 TLSERLNSQTHVVSITSALKEVVQNSIDAKATTITIQ 47
Sklu 11 NVLRVLKSQVFIVSVTSAVRELLQNSVDAGASEIEAV 47
Smik 11 TVSEMLKSQVCAVSLASAVREIVQNSVDAHATTIDVV 47
Spar 11 NVSEKLKSQVSSVSLASAVREIVQNSVDAHATTIDVV 47
Lelo 10 DVINQIKSHLRFNTIDSIIEHLLKDSLQRSPKNIIVK 46
Psti 13 QVLSELRSQTIFNSLASVVQELLRNSLDAQAKVVEIR 49
Afum 14 EVVAKIKSSTSIINLNRVVVELVKNALDANARTVSVT 50
Nfis 14 EVVAKIKSSTSITDLNRVVVELVKNALDANAHTVSVT 50
Acla 14 DAVAKIKSSTSITHLNGVIVELVKNALDANARTVFVT 50
Ater 14 DVIAKLKSSTSITQLNGVVVELVKNALDANAHTIYVT 50
Cimm 14 KVAAQIESSTSITSLNSAILGLVKNSLDADAQTVAIT 50
Pans 9 DVVAQIKSSTIITSLNTAVCGLLRNSLDARSTKINIS 45
Fgra 9 DVVDKIKSSVNITSLNGVVCGLIANSLDAGASKVNIS 45
Fver 9 DVVDKIKSSVNITSLNGVVCGLITNSLDAGASKINIS 45
Tree 9 DVVGKIRSSSTITSLNGVVCGLVKNSLDAGATKVNIV 45
Snod 20 DVVAQIKSSTAIVSLTDVVLELLKNSLDARATKVEAS 56
B
Mlh3
Top z-score in Multi-harmony:
Block 2, Mlh1 interaction
mlh3-Y493M,N497G,V499F,D500N,K502G
mlh3-D500N, mlh3-K502G single substitutions
Scer 492 KYEVINQVDKKFILI 506
Agos 493 DCIVINQIGNKFILL 507
Calb 438 KFRVINQVDRKFILL 452
Cdub 400 RFKVINQIDRKFILL 414
Cgla 503 EVDVVNQVDQKFILL 517
Cgui 398 SFRVLKQLDRKYILF 412
Ctro 435 NYKIIKQLDKKFILV 449
Dhan 511 NYRIISQVDRKFILL 525
Klac 477 NFEVIQQVDLKFILV 491
Kpol 501 DCKLINQLDRKFILL 515
Kthe 457 RFRVVNQVDKKFVLL 471
Kwal 468 SYRAVNQVDNSFILL 482
Sbay 492 NCEVINQVDKKFILI 506
Scas 507 DCTIINQVDDKFILL 521
Sklu 525 NCHVVNQVDNKFILL 539
Smik 492 KYEVINQVDKKFVLI 506
Spar 492 KYEVINQVDKKFVLI 506
Lelo 476 NYRIIRQIDNKFILL 489
Psti 424 TYRIVRQLDSKFILV 437
Afum 657 TANVIAQVDRKFILV 671
Nfis 655 TANIIAQVDRKFILA 669
Acla 649 TAEIIAQVDRNFILA 663
Ater 619 TAEVIAQIDQKFILA 633
Cimm 704 NAQIVAQVDNKFILL 718
Pans 633 KAEVVNQVDEKFVLV 647
Fgra 602 SATVIAQVDRKFILV 616
Fver 600 RATVIEQVDRKFILV 614
Tree 625 AAEVVAQVDKKFILL 639
Snod 656 SAEVLAQVDKKFILV 670
Pms1
Scer 10 IDVHRITSGQVITDLTTAVKELVDNSIDANANQIEII 46
Agos 11 ADVHRITSGQVIIDLVAAVKEVVENSLDAHADKLEIT 47
Calb 14 FDISKITSSQVIIDLKAIIKELIENSIDAHADKIDII 50
Cdub 14 FDISKITSSQVIIDLKAIIKELIENSIDAHANKIDII 50
Cgla 9 QDVHRITSGQVIIDLASAVKELVENSLDAQATQIDIT 45
Cgui 9 TEVQRITSGQVIVDLVSVVKELVENAIDSGSTKIDVT 45
Clus 9 SKVAKITSGQVIIELLSVVKELVENSIDAESDKIDVV 45
Ctro 12 ADISKLTSSQVIIDLKSILKELIENSIDANADKIDIT 48
Dhan 9 GDIQRITSGQVIIDLVSIVKELVENSIDASSSKIEVL 45
Klac 9 ADIHKITSGQVIVDLRSAIKELLENSLDAKADKIEII 45
Kpol 10 SDVHRITSGQVIIDLTTAVKELIDNSIDASAKQIDIT 46
Kthe 11 ADVHRITSGQVIIDLASAVKELLDNSIDSGADQVVCT 47
Kwal 9 ADVHRITSGQVIVDLTSAVKELLDNSIDSGADHVECT 45
Sbay 14 IDVHRITSGQVITDLTTAVKELVDNSIDANASQIEIT 50
Scas 10 KDIHRLTSGQVIVDLVSAVKELVDNSIDAHSSQIELT 46
Sklu 10 TDVHRITSGQVIVDLATAVKELVENSIDAGADKIDVV 46
Smik 10 IDVHQITSGQVITDLTTAVKELVDNSIDANANQIEIT 46
Spar 10 IDVHRITSGQVITDLTTAVKELIDNSIDANASQIEIT 46
Lelo 14 TDISKITSGQVIIDLRSIVKELVENAIDASAKKILVN 50
Psti 9 KDVSKITSGQVIIDLKSIVKELVENSIDANSTKIEIN 45
Ylip 9 ASIRQITSAQVVTDLNSAVKEVVENSLDANAKNVEIK 45
Afum 10 RSVHQIQSGQVIVDLCSVVKELVENSLDAGATSIEVR 46
Nfis 10 RSVHQIQSGQVIVDLCSVVKELVENSLDAGATSIEVR 46
Acla 10 RSVHQIQSGQVIVDLCSVVKELVENSLDAGATSIEVR 46
Ater 10 RSVHQIQSGQVIVDLCSVAKELVENSLDAGATSIEVR 46
Cimm 14 RSIHQIQSGQVIVDLCSVVKELVENSLDAGATSIDIR 50
Ncra 13 STIHQIQSGQVIVDLCSVAKELVENALDAGATTIDVR 49
Pans 13 SAVHQIQAGQVIVDLCSVAKELVENSVDAGATTIEVR 49
Fgra 10 RTVHQIQSGQVIVDLCSVVKELVENSVDSGATSIDVR 46
Fver 10 RTVHQIQSGQVIVDLCSVVKELVENSVDAGATSIDVR 46
Tree 10 RT LTSGQ DVR 19
Sscl 10 RTIHQIQSGQVIVDLCSVVKELVENSLDAGATSIDVR 46
Snod 10 RSVHQIQSGQVIVDLQSVCKELVENSIDAGATTVEVR 46
Spom 10 NTVHKICSGQVITDVASAVKELVENSLDSGATTIEIR 46
Pms1
Scer 673 KMEVVGQFNLGFIIV 687
Agos 703 EMSIIGQFNLGFIIV 717
Calb 699 KMKLIGQFNLGFIIV 713
Cdub 693 KMKLIGQFNLGFIIV 707
Cgla 702 DMKVVGQFNLGFILV 716
Cgui 659 KMQVVGQFNLGFILV 673
Clus 675 RMSVVGQFNLGFVVV 689
Ctro 595 DMKLIGQFNLGFILV 609
Dhan 535 RDKLLS DLGVVEV 547
Klac 723 NMTIVGQFNLGFIIV 737
Kpol 757 RMEIVGQFNLGFIIA 771
Kthe 706 KMEIVGQFNLGFILV 720
Kwal 698 NMNVVGQFNLGFIIV 712
Sbay 696 KMEIVGQFNLGFIIV 710
Scas 655 KMVIIGQFNLGFIVV 669
Sklu 678 KMEIVGQFNLGFIIV 692
Smik 674 KMEVVGQFNLGFIIV 688
Spar 680 KMEVVGQFNLGFIIV 694
Lelo 748 DMKLIGQFNLGFILV 762
Psti 614 KMKLIGQFNLGFILV 628
Ylip 699 NFNIIGQFNEAFIIV 713
Afum 785 RMRIIGQFNLGFILA 799
Nfis 786 RMRIIGQFNLGFILA 800
Acla 801 RMRIIGQFNLGFILA 815
Ater 794 QMRIIGQFNLGFILA 808
Cimm 759 KMRVVGQFNLGFILA 773
Ncra 843 KMKIVGQFNLGFIIA 857
Pans 753 KMKIIGQFNLGFILA 767
Fgra 776 RMRIAGQFNMGFIIA 790
Fver 783 KMRIAGQFNMGFIIA 797
Tree 688 RMRIIGQFNLGFIIA 702
Sscl 855 KMKIIGQFNLGFILA 869
Snod 831 EMRIIGQFNLGFIIA 845
Spom 610 RMRVVGQFNRGFIVV 624

Figure 2.8C


C
Mlh3
Block 3, endonuclease motif
mlh3-R530K, mlh3-R532N, mlh3-R530K,R532N
Block 4, PIP motif
mlh3-PIP1-9
Scer 519 LVLVDQHACDERIRLEELFYSLLTEVVTGTFVARD 553
Agos 520 LLILDQHAADERVKLEAYTRDYLFTLLTAQPSFYT 554
Calb 458 LVVLDQHASDERIRVEQYLQEFVSQPNPGLRLHSP 492
Cdub 420 IVVLDQHASDERIRVEQYLQEFVEQRHPGLRLQNP 454
Cgla 528 LYIIDQHACDERIRLESFLKQYICDIMANALAVQK 562
Cgui 422 LFIVDQHACDERIRVESYLKDYIQRTKVNQHQNKA 456
Ctro 458 IVVLDQHATDERIKVEEYLQEFVQLLQKNPGLRLK 492
Dhan 541 ILVVDQHACDERIRVEALFKDFILMLLDKTLGIEL 575
Klac 497 LLMLDQHACHERILVENMLKETIIKCMNKCFNYVK 531
Kpol 528 LLLIDQHACDERIKLEEYLKEYLTSVRSKTVTTRK 562
Kthe 483 LVLLDQHAADERIKYESLLNGFIWGMLTSPHLHIR 517
Kwal 494 LLLVDQHAADERMRFESLMSDFLYEFITAPYLQTL 528
Sbay 519 LILVDQHACDERIRLEDLLHSLLSEVLTETFVTQD 553
Scas 533 LLIVDQHACDERIKLESYLKDFFQDVINGTISSQP 567
Sklu 552 LLIVDQHACDERIRLESYIKEFVFNALSFRGLNTV 586
Smik 519 LILVDQHACDERIRLEGLLHNLLTEVISGTFVAQD 553
Spar 519 LVLVDQHACDERIRLEDLFHNLLTEIITGTFVAQD 553
Lelo 517 LVVLDQHASDERVKIEKLIKEFVDEMSANPGLRLC 451
Psti 448 LLVIDQHACDERIKVEALFKDFIFLVLDAHTNLLL 482
Afum 683 LVLIDQHAADERCRIESLFGAMFADGHRQVQSIRI 717
Nfis 681 LVLIDQHAADERCRIESLFGDMFADGHRQVQTIRI 715
Acla 675 LILVDQHAADERCRVEHLFGGLFADNDNSLCPIHT 709
Ater 645 LVLIDQHAADERCRVERLFGEFFADDRSGRVQTIT 679
Cimm 734 LVLVDQHAADERIRVERLFDELCGSSPSHTVDTTP 768
Pans 667 LVIIDQHAADERYRVEALLKEYFVPNPDDGRSLVA 701
Fgra 637 LVMLDQHAVDERCQLEELMLEYFTTDPLTNQVLPQ 671
Fver 635 LVMLDQHAADERCQLEDLMASYFAHDPSTNQTSSV 669
Tree 658 LVMLDQHAADERCRLEDLMAGYFTHDSSIGAIRAV 692
Snod 689 LVLIDQHAADERIRVEALLRELCSPMDTHCSGYQS 723
Pms1
Scer 697 LFIVDQHASDEKYNFETL QAVTVFKSQKLIIPQP 730
Agos 727 LFIVDQHASDEKYNFENL QKSTVFNSQHLIKPLT 760
Calb 721 LFIIDQHASDEKYNFEKL MASFKINYQLLIKPIK 754
Cdub 715 LFIIDQHASDEKYNFEKL MTNFRINYQSLIKPIK 748
Cgla 726 LFIVDQHASDEKFNFENL QQTTRFKSQKLISPET 759
Cgui 681 LFIVDQHASDEKYNFERLANSTTMFHSQLLVVPRN 715
Clus 695 LFIVDQHALDEIFNYERL MQSLVLRAQPLVIPRL 728
Ctro 615 LFIIDQHASDEKFNFEKL MSNFQIKHQPLMMPIN 648
Dhan 555 RRDRDNNIEDEKY 567
Klac 747 LFIVDQHASDEKYNFEML QKNTVFKSQSLLSLKT 780
Kpol 781 LFIIDQHASDEKYNFETL QKSTVFKSQKLIVPQP 814
Kthe 730 LFIIDQHASDEKYNFEKL QKNTVFKSQKLLAPQI 763
Kwal 722 LFIVDQHASDEKYNFETL QRTTVFRSQRLIAPQV 755
Sbay 720 LFIVDQHASDEKYNFETL QAVTVFKSQKLIVPQP 753
Scas 679 LFIVDQHASDEKYNFEML QKETVFNSQRLIAPQP 712
Sklu 702 MFIVDQHASDEKYNFETL QKTTVFKSQRLLSPQI 735
Smik 698 LFIVDQHASDEKYNFETM QAVTVFKSQKLIIPQP 731
Spar 704 LFIVDQHASDEKYNFETL QAVTVFKSQKLITPQP 737
Lelo 768 LFIIDQHASDEKFNFERL LETFAVNYQPLITPLF 801
Psti 635 LFIIDQHASDEKYNFERL NQELSIKIQRLIVPQT 668
Ylip 719 LFIIDQHASDEKYNFERL QRDTKITPQPFVNPLT 752
Afum 809 LFIIDQHASDEKFNFERL QAETVVQNQRLVQPKR 842
Nfis 810 LFIIDQHASDEKFNFERL QAETVVQNQRLVQPKR 843
Acla 825 LFIIDQHASDEKFNFERL QAETVVQNQRLVQSKR 858
Ater 818 LFIIDQHASDEKFNFERL QAETVVQNQRLVQPKR 851
Cimm 783 LFIIDQHASDEKYNFERL QAETVVQNQRLVRPKT 816
Ncra 867 LFIIDQHASDEKYNFERL QSTTTVQSQRLVQPKP 900
Pans 777 LFIIDQHASDEKYNFERL QSTTTVQSQRLVQPKP 810
Fgra 800 LFIIDQHATDEKYNFERL QEVQTVQSQRLVHPKR 833
Fver 807 LFIIDQHATDEKFNFERL QEIQTVQSQRLVHPKR 840
Tree 712 LFIIDQHASDEKYNYERL QRTTEIQSQRLVHPKR 745
Sscl 879 VFIIDQHSSDEKYNFERL QATTIVQSQRLVYPQN 912
Snod 855 LFIIDQHASDEKYNFERL SATTTLVSQRLVHPHP 888
Spom 630 LFIIDQHASDEKFNYEHL KSNLVINSQDLVLPKR 663

Figure 2.8D



to

Fungal alignments of Block 1 (A), Blocks 2 (B), Blocks 3-4 (C) and Blocks 5-6 (D) mutations analyzed in this study. Amino acid sequences of Mlh3 and Pms1 Homologs were obtained from the list presented in the legend Fig 2.3.

Finally, two *mlh3-PIP* mutant alleles (Block 4, addition of PCNA interaction (PIP) motif) were created. *mlh3-PIP1* contains the yeast Pms1 PIP motif (**QKLIIP**) and *mlh3-PIP2* contains just the three critical consensus residues (**Q**TF**I**A**P**), with the other residues derived from Mlh3 (Fig 2.2C). The *mlh3-PIP2* mutation complemented Mlh3 functions in both MMR and meiotic crossing over, whereas *mlh3-PIP1* conferred an intermediate defect in both assays. These observations encouraged us to perform the detailed analysis of the PIP motif described below.

These observations confirmed that regions of high conservation specific to the Mlh3 family were critical for Mlh3 function; every Block substitution group conferred defects in MMR and meiotic crossing over, though with different phenotypic strengths. Also, the finding that some Block mutations differentially affected Mlh3 MMR and meiotic functions is consistent with the previous identification of Mlh3 separation of function mutations that may disrupt interactions with specific pathway components (Table 2.1; [46]).

Identifying Block *mlh3* mutations with expanded functions in DNA mismatch repair

The complementation studies presented in Table 2.1 encouraged us to test if the Block 1-6 *mlh3* alleles could complement Pms1 MMR functions (Table 2.2). We performed these studies by expressing *mlh3* alleles through their native promoter on a *2μ LEU2* pRS derived vector. This was done because an analysis of multiple protein abundance data sets [59] indicated that Mlh3 protein levels (median of 220 molecules per cell) are lower than those seen for Pms1 (median of 648 molecules per cell). Karim et al. [60] determined in an S288c strain that such *LEU2 2μ* vectors were present at 14 copies/cell, and so the expression of *MLH3* in a *2μ LEU2* vector is predicted to yield Mlh3 molecule numbers per cell that are only a few-fold higher than the native Pms1 levels.

We tested if the *MLH3*$_{SK1}$ gene used to make all of our multi-Harmony substitutions could complement, when expressed on an *ARS-CEN* vector, the MMR functions of an S288c *mlh3Δ* strain (EAY4595). As shown in Table 2.1, this plasmid displayed full complementation, indicating that it was an appropriate gene construct for our analysis. Interestingly, mild overexpression of *MLH3* on a *2μ* vector conferred a mutator phenotype significantly higher than seen for in *mlh3Δ*, but significantly lower than seen in the *pms1Δ* strains. This was not a surprise as we had shown previously that robust overexpression of *MLH3* through a *GAL10* promoter conferred mutator phenotypes in a *wild-type* strain similar to that seen in *pms1Δ*, most likely by sequestering Mlh1 from interacting with Pms1 [61]. We reasoned that any improved function for an *mlh3* allele in *PMS1*-dependent MMR should overcome the mutator phenotype associated with *MLH3* overexpression.

As shown In Table 2.2, *mlh3* Block 1 (~2-fold) and 4 (~3.5 fold) alleles conferred modest suppression of the *pms1Δ* mutator phenotype whereas *mlh3* Blocks 2, 3, 5, 6 did not. Curiously, this suppression was not seen for the *mlh3-PIP1* allele, or when *mlh3-PIP2* was expressed on an *ARS-CEN* vector, indicating that *MLH3* expression on a 2*μ* vector was important to detect the suppression phenotype. The failure of the *mlh3-PIP1* allele to suppress the *pms1Δ* mutator phenotype suggested that the suppression seen with *mlh3-PIP1* was not due to the Mlh3 being able to interact with PCNA in MMR. To test this, we introduced into *MLH3* seven other mutations that deviated from the PCNA motif in Pms1 (Fig 2.2C). Of these, two conferred phenotypes similar to *mlh3-PIP2*, *mlh3-PIP3* (QTLIAP) and *mlh3-PIP5* (QTFIAR). The remainder, several of which contained PIP consensus sequences, conferred mutation rates similar to *mlh3-PIP1*. Together, these findings suggest that the addition of a functional PIP motif was unlikely to explain the improved MMR phenotype seen in *mlh3-PIP2, -PIP3*, and *-PIP5*. In

addition, they do not address whether the mlh3-PIP proteins are directly interacting with PCNA, or if a structural change induced by the insertion of this motif alters its endonuclease activity *in vivo* to be more similar to that seen with Mlh1-Pms1.

The above observations encouraged us to determine if the modest complementation seen for the *mlh3-Block 1* and *mlh3-PIP2 alleles* could be enhanced or maintained in combination with other *mlh3* alleles. As shown in Table 2.2 and Fig 2.9, the *mlh3-Block1, Block 4(PIP2)* allele complemented the *pms1Δ* MMR defect as well as, if not slightly better than the single mutants. Other combinations involving the addition of silent mutations (*D678K*) or weak *mlh3* hypomorphs (*R530K, R532N*) did not affect the ability of *mlh3-Block 4 (PIP2)* to complement the *pms1Δ* MMR defect, whereas others did (*D500N, K502G*). It was interesting to see that the more severe *mlh3-R530K,R532N* and *mlh3-Block 6* mutations disrupted *mlh3-Block 4 (PIP2)* complementation, consistent with the idea that Block mutations that could not complement *mlh3Δ* and *pms1Δ* functions were deleterious for all MLH functions. Together, these observations indicate that multi-Harmony is a valuable approach to identify residues outside of consensus motifs that are critical for specificity of subsets of protein family members. The finding that alleles were found in both the ATP-binding and endonuclease domains of Mlh3 and Pms1 could alter Mlh3 specificity supports our evolutionary analysis suggesting that multiple domains across the MLH proteins are co-evolving to provide substrate specificity.

Discussion

We used the MLH family of proteins as a model to better understand the steps that lead to gene specialization following a gene duplication event. We focused on two highly conserved yeast MLH proteins, Mlh3 and Pms1, which share limited DNA mismatch repair functions, with only

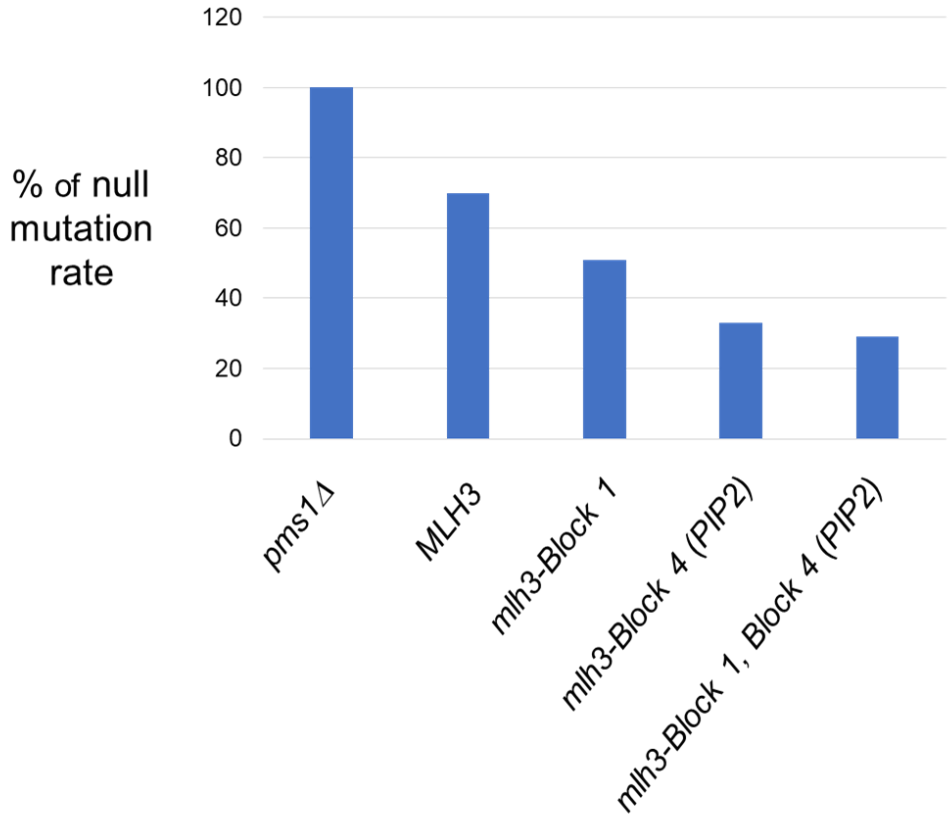


Figure 2.9. Block 1 and Block 4 *mlh3* mutations can alleviate the *pms1Δ* mutator phenotype. Mutation rates of *mlh3-Block 1* and *mlh3-Block 4* single- and double-mutant strains were determined in the *lys2-A*$_{14}$ assay as described in Materials and Methods. Rates are shown as a percentage of the corresponding *pms1Δ* (Table 2.2).

Mlh3 having a role in meiotic crossing over (reviewed in [6]). An advantage of working with yeast MLH proteins is the availability of sensitive genetic assays to assess MMR and meiotic crossing over, as well as the availability of fungal genomes to perform phylogenetic and ERC analyses. The baker's yeast Mlh3 and Pms1 proteins show significant divergence (21% identity), suggesting that an extensive number of evolutionary steps, some major, others involving subtle refinements, likely took place to diversify the MLH proteins, and that it would not be trivial to reconstruct paths leading to their diversification.

By leveraging phylogenetic, multi-Harmony, and molecular genetic analyses we identified critical residues in Mlh3 that specifies its roles in MMR and meiotic crossing over, and importantly, we were able to alter the specificity of Mlh3 to modestly rescue MMR functions in the absence of Pms1. We were surprised by this result because the functional advantage conferred by making a limited number of mutations in proteins that differ by hundreds of amino acids would likely be difficult to detect. The altered specificity seen in *mlh3* alleles was further confirmed by combining mutant combinations. Overall, combining beneficial alleles maintained or slightly improved altered specificity, whereas combining a neutral allele with a beneficial one did not affect outcome, and combining a deleterious allele with a beneficial one disrupted altered specificity. These data, combined with chimera and ERC analyses, are consistent with the three domains of the MLH proteins being interdependent, having co-evolved for millions of years, and contributing to the distinct specificities in MMR and meiotic crossing over seen for the Pms1 and Mlh3 proteins, respectively.

As described in the Introduction, Hughes [2] proposed an elegant model to explain how novel protein functions evolved after gene duplication. In this model, gene specialization evolved in two steps; first, an ancestral gene is capable of performing two distinct functions. Second,

gene duplication followed by purifying selection enabled specialization of the duplicated genes. We find it intriguing that two *mlh3* alleles (*Block 1* and *Block 4*) displayed partial complementation of Pms1 MMR functions yet maintained Mlh3 meiotic crossover functions, suggesting that the specialization of the MLH functions could be reversed, with perhaps an ancestral MLH protein displaying both MMR and meiotic functions that were specialized following a gene duplication event. Consistent with this model is work showing that the *E. coli* MutS protein can repair base-base mismatches and small insertion/deletion mismatches up to 3 nt in size [62]. However, this repair became more specialized in the eukaryotic MSH protein complexes; Msh2-Msh6 repairs primarily base-base mismatches and 1 nt insertion/deletion mismatches whereas Msh2-Msh3 primarily repairs insertion/deletion mismatches up to 17 nt in size [8, 63].

Our work complements two recently published studies which examined the evolution of factors that act in vegetative and meiotic functions in baker's yeast. In baker's yeast the Rad51 and Dmc1 strand exchange proteins, which share ~50% amino acid identity and are thought to evolve from a single gene duplication of an ancestral recombination protein, play critical roles in promoting homologous recombination (reviewed in [6]). Both proteins are able to catalyze the invasion of a 3' single stranded end into a homologous duplex template to form postsynaptic complexes in three nucleotide steps [64-67]. Most eukaryotes contain a Rad51 protein that is expressed in both mitotic and meiotic cell cycles, and a Dmc1 protein which is meiosis-specific and unlike Rad51, can stabilize heteroduplex DNA with mismatch containing base triplets. In meiosis, Dmc1 and Rad51 play somewhat complementary roles, with Dmc1 catalyzing interhomolog recombination, and Rad51 promoting the assembly of the Dmc1 presynaptic filament [64,68,69]. Using structural and phylogenetic approaches, Steinfeld et al. [67] identified

residues that are conserved within the Rad51 or Dmc1 lineages but differ between them. They then made Rad51/Dmc1 chimeras and identified a set of three amino acid substitutions in the L1 DNA binding loop of these proteins that conferred differences in how the two proteins stabilized recombination intermediates bearing base triplets with mismatches; only Dmc1 can stabilize such intermediates. While such changes in specificity were observed biochemically, they did not result in the formation of functional proteins *in vivo*; the author speculated that swapping this motif "may hinder some downstream step in the HR pathway." In addition, Steinfeld et al. [67] identified 19 lineage specific amino acids in other regions of the two recombinases that they speculated were important for Rad51 and Dmc1-specific protein-protein contacts.

In the second study, Hsieh et al. [70] analyzed kleisin paralogs that act in sister chromosome cohesion in vegetative growth (Scc1) and meiosis (Rec8). Specifically, they performed an experimental evolution experiment to ask how yeast evolve when forced to use the Rec8 kleisin in the vegetative life cycle. Their findings demonstrated that the vast majority of the evolved populations acquired mutations not in Rec8, but in known and novel partners of kleisins. The observations did not address the molecular changes that took place between the vegetative and meiotic kleisins; the authors argue that "the selective advantage conferred by individual mutations is likely to be modest," which is supported by our analysis of *mlh3* altered specificity mutants.

Below is a brief overview of our strategy to identify residues critical for specificity in divergent paralogs. We feel that such an approach will also be valuable to identify protein-protein interactions that are specific for each paralog.

I. Identify protein families with highly conserved organization and motifs. We took advantage of previous genetic and biochemical studies which showed that Mlh3 and Pms1

displayed highly conserved (from bacteria to mammals) protein organizations that included an ATP binding motif in their N-termini, a linker region, and a metal binding motif ($DQHA(X)_2E(X)_4E$) in their C-termini. These motifs and domains have been shown to be critical for the function of MLH paralogs (reviewed in [8]). We also took advantage of previous structural studies that provided domain boundaries for the N-terminal ATP-binding and C-terminal Metal binding/MLH interaction domains. These boundaries provided the inputs for the chimera and ERC analyses that were performed.

II. Perform a phylogenetic approach with relatively closely related species. We included a group of 34 fungal species in this analysis because Mlh3 and Pms1 homologs could be identified in almost all of the species. This analysis allowed us to determine if the Mlh3 and Pms1 orthologs split into phylogenetic sister groups, which provided a justification for the mutagenesis, or one of the orthologs split into an outgroup which would have discouraged us from implementing the multi-Harmony strategy.

III. Perform ERC on functional domains to test the idea that these domains tend to change in parallel over evolutionary time. The results of such an analysis would encourage the construction of chimeras that are substituted in parallel or separately. While our findings showed that none of the chimeric Mlh3/Pms1 proteins constructed complemented MMR or meiotic crossing over functions, most of them interfered with MMR functions, providing evidence that functions required for MLH-specificity were compromised. The chimera studies also provided evidence that the linker regions of the MLH proteins are critical for paralog specificity, as we were unable to recreate functional MLH proteins despite making a large number of linker substitutions.

IV. Using multi-Harmony to identify paralog-specific amino acid substitutions. The above observations encouraged us to use the multi-Harmony computational approach to identify blocks of amino substitutions between Mlh3 and Pms1 that would improve the functions of Mlh3 in MMR. This analysis allowed us to identify residues in both the ATP binding and DNA binding domains that appeared important for specificity. Interestingly, multi-Harmony analysis revealed that every block of substitutions that we made affected the functions of Mlh3; it also showed that changes in multiple domains were required for the altered specificity, which is consistent with our ERC analysis.

An approach to study human disease

Germline mutations in MMR genes in humans, specifically *MSH2, MLH1,* and *PMS2*, are associated with microsatellite instability, and an increase in predisposition of a hereditary form of colorectal cancer (HNPCC; [8,71]). While it is known that alterations in *MLH* genes are associated with colorectal cancer susceptibility, it still not completely determined which gene variants promote tumorigenesis. Thus, identifying specificity-determining residues in Pms1 and Mlh3 in baker's yeast can allow one to map critical residues required for Pms1's (PMS2 in Humans) MMR role, and thus provide insights in the human homologs, with the hopes of identifying deleterious alleles associated with cancer susceptibility.

Materials and Methods

Media

S. cerevisiae SK1 and S288c strains were grown at 30^{o}C in either yeast extract-peptone (YPD) media or minimal selective media (SC; [72]). When required, geneticin (Invitrogen, San Diego) was added at 200 μg/ml [73], and sporulation plates were prepared as described [74].

Plasmids

Plasmid and strain background derivation for the relevant MMR genes are listed in Table 2.3. The DNA sequence of the open reading frame (including 300 bp upstream and 150 bp downstream) of constructs was confirmed by Sanger DNA sequencing (Cornell BioResource Center). MMR gene constructs built by Gibson cloning were resub-cloned into the backbones of expression vectors. Oligonucleotides and sequences of plasmids are available upon request.

Mlh3/Pms1 chimera (Fig 2.6) integration vectors containing $MLH3_{SK1}$ and $PMS1_{SK1}$ sequences were derived from pEAI254, a 7.8 kb $MLH3_{SK1}$::*KanMX* integrating vector [46], and pEAA238, a 9,1 kb $PMS1_{SK1}$ *ARS-CEN HIS3* vector, respectively. The Block mutation (Fig 2.7; Fig 2.8A-D) integration vectors were derived from pEAM168, a 10.7 kb $MLH3_{SK1}$::*KanMX* *2μ* vector, fragments of which could be isolate to integrate *mlh3* alleles. The *MLH3* and *PMS1* boundaries in the chimera plasmids was chosen based structural and homology model analyses of MLH proteins that yielded three distinct regions; an ATP binding domain, a linker region, and C-terminal endonuclease/MLH interaction domain (Fig 2.2A; [35.38]; Mlh3 homology model in [46]). The chimera plasmids were constructed by linking the different domains of PCR-amplified *MLH3* and *PMS1* DNA sequences using NEB HiFi DNA Assembly cloning (New England Biolabs, Ipswich, MA). For Mlh3 these domains were: ATP binding, aa 1 to 375; linker, aa 376

to 488; C-terminal endonuclease, 489 to 715. For Pms1 the domains were: ATP binding, aa 1 to 361; linker, aa 362 to 638; C-terminal endonuclease, 639 to 877 (SK1).

The MPM chimeric construct pEAI403, (Table 2.3; Fig 2.6; Fig 2.10) was modified in plasmids pEAI428 to pEAI435 (Table 2.3) to contain Mlh3 linker insertions of different sizes (19, 40, 60, 80, 100, 113-full length amino acids). The amino acids mutated in the *mlh3-32* allele, which confers a more severe defect in meiotic crossing over, compared to MMR [46], were centered within the Mlh3 linker insertions. These plasmids were constructed by inserting the different linker sizes into pEAI403 construct using NEB HiFi DNA Assembly cloning. Pms1 linker sequences were deleted to accommodate linker sizes similar to those seen for the Mlh3 and Pms1 linker. The exact junction points for these constructs are presented in Fig 2.10. *mlh3* Block mutations were constructed using NEB HiFi DNA Assembly cloning and/or Q5 site-directed mutagenesis kits (New England Biolabs, Ipswich, MA). The integration plasmids were digested with *Bam*HI and *Sal*I prior to transformation into EAY3255 using methods described by Gietz et al. [75]. At least three independent transformants for each genotype were made and genotyped by PCR (presence of specific alleles also confirmed by DNA sequencing) using primers that map outside of the restriction sites used for integration.

MLH3/PMS1 chimera and Block mutation alleles (SK1 background) were also expressed from the native *MLH3* promoter on *2μ LEU2* plasmids (pEAM). pEAI254 ($MLH3_{SK1}$*::KanMX*), pEAM168 ($MLH3_{SK1}$*::KanMX, 2μ*) and pEAM65 ($MLH3_{SK1}$*::LEU2, 2μ*) were the parental plasmids for these constructs. These plasmids were tested for complementation of MMR defects in the S288c strain EAY3097 (relevant genotype *pms1Δ, lys2::insE-*A_{14}).

Table 2.3. Plasmids used in this study.

Plasmid	Promoter	Relevant Genotype	Source
pRS415		*ARS-CEN, LEU2*	Christianson et al., [83]
pRS425		*2m, LEU2*	Christianson et al., [83]
pJH481	*PMS1*	*PMS1$_{S288c}$, ARS-CEN, LEU2*	James Haber
pEAA238	*PMS1*	*PMS1$_{S288c}$, ARS-CEN, HIS3*	
pEAM50	*PMS1*	*PMS1$_{S288c}$, 2m, LEU2*	
pEAM65	*MLH3*	*MLH3, 2m, LEU2*	
pEAM168	*MLH3*	*MLH3::KanMX, 2m, URA3*	
pEAI254	*MLH3*	*MLH3::KanMX*	
pEAA566	*MLH3*	*MLH3, ARS-CEN, LEU2*	
Chimera constructs containing *MLH3(M)* and *PMS1(P)* domains			
pEAI398	*MLH3*	*PMM::KanMX*	
pEAI399	*MLH3*	*PPP::KanMX*	
pEAI400	*MLH3*	*MMP::KanMX*	
pEAI401	*MLH3*	*PPM::KanMX*	
pEAI402	*MLH3*	*MPP::KanMX*	
pEAI404	*MLH3*	*PMP::KanMX*	
pEAI403	*MLH3*	*MPM::KanMX*	
pEAI428	*MLH3*	*MPM-19::KanMX*	
pEAI429	*MLH3*	*MPM-40::KanMX*	
pEAI430	*MLH3*	*MPM-60::KanMX*	
pEAI431	*MLH3*	*MPM-80::KanMX*	
pEAI432	*MLH3*	*MPM-100::KanMX*	
pEAI433	*MLH3*	*MPM-N-full::KanMX*	
pEAI434	*MLH3*	*MPM-middle-full::KanMX*	
pEAI435	*MLH3*	*MPM-C-full::KanMX*	
pEAM296	*MLH3*	*PMM::KanMX, 2m, LEU2*	
pEAM297	*MLH3*	*PPP::KanMX, 2m, LEU2*	
pEAM298	*MLH3*	*MMP::KanMX, 2m, LEU2*	
pEAM299	*MLH3*	*PPM::KanMX, 2m, LEU2*	
pEAM300	*MLH3*	*MPP::KanMX, 2m, LEU2*	
pEAM301	*MLH3*	*PMP::KanMX, 2m, LEU2*	
pEAM302	*MLH3*	*MPM::KanMX, 2m, LEU2*	
Block mutations in *MLH3:* used to make integrations			
pEAM313	*MLH3*	Block 1, ATP binding-*mlh3-K17T,A20Q,S24D,R30K,Q34D::KanMX*	
pEAM314	*MLH3*	Block 2, Mlh1 interaction-*mlh3-Y493M,N497G,V499F, D500N,K502G::KanMX*	
pEAM320	*MLH3*	Block 2-*mlh3-D500N*::*KanMX*	
pEAM321	*MLH3*	Block 2-*mlh3-K502G::KanMX*	
pEAM324	*MLH3*	Block 3, Endonuclease motif-*mlh3-R530K::KanMX*	
pEAM322	*MLH3*	Block 3-*mlh3-R532N::KanMX*	
pEAM325	*MLH3*	Block 3-*mlh3-R530K,R532N::KanMX*	
pEAI439	*MLH3*	Block 4, PCNA interaction motif-*mlh3-PIP1::KanMX*	
pEAI440	*MLH3*	Block 4-*mlh3-PIP2::KanMX*	

Table 2.3. (con't)

pEAM315	*MLH3*	Block 5, Helix 2-*mlh3-V660K,N666A,F676I,D678K::KanMX*
pEAM323	*MLH3*	Block 5-*mlh3-D678K::KanMX*
pEAI479	*MLH3*	Block 6, Helix 1-*mlh3-C695L,F699W,A702P,S707T,V709R,P710H::KanMX*
pEAM316	*MLH3*	Block 1 (*ATP binding*), Block 4 *(PIP2*)*::KanMX* combination
pEAM326	*MLH3*	Block 5 *(Helix 2),* Block 6 (*Helix 1*)*::KanMX* combination
Block mutations in *MLH3: 2m, LEU2* vectors used for expression		
pEAM303	*MLH3*	Block 1-*mlh3-K17T,A20Q,S24D,R30K,Q34D*
pEAM304	*MLH3*	Block 2-*mlh3-Y493M,N497G,V499F,D500N,K502G*
pEAM307	*MLH3*	Block 2-*mlh3-D500N*
pEAM308	*MLH3*	Block 2-*mlh3-K502G*
pEAM311	*MLH3*	Block 3-*mlh3-R530K*
pEAM309	*MLH3*	Block 3-*mlh3-R532N*
pEAM312	*MLH3*	Block 3-*mlh3-R530K,R532N*
pEAM278	*MLH3*	Block 4-*mlh3-PIP1::KanMX*
pEAM279	*MLH3*	Block 4-*mlh3-PIP2::KanMX*
pEAM291	*MLH3*	Block 4-*mlh3-PIP3::KanMX*
pEAM285	*MLH3*	Block 4-*mlh3-PIP4::KanMX*
pEAM287	*MLH3*	Block 4-*mlh3-PIP5::KanMX*
pEAM286	*MLH3*	Block 4-*mlh3-PIP6::KanMX*
pEAM288	*MLH3*	Block 4-*mlh3-PIP7::KanMX*
pEAM289	*MLH3*	Block 4-*mlh3-PIP8::KanMX*
pEAM290	*MLH3*	Block 4-*mlh3-PIP9::KanMX*
pEAM305	*MLH3*	Block 5-*mlh3-V660K,N666A,F676I,D678K*

Unless indicated, *MLH3* and *PMS1* genes were derived from the SK1 strain background and all plasmids with the pEA plasmid designation were built in the Alani laboratory (Materials and Methods). The designation of the *M* (*MLH3*-derived) and *P* (*PMS1*-derived) abbreviations, and for the *MLH3-PMS1* chimeras and the block mutations are presented in the Materials and Methods and Figs. 2.2, 2.6 and 2.7

Strains

The SK1 strains EAY3252, EAY3255, and EAY3486 and indicated derivatives in Table 2.4 were used to measure Mlh3-dependent meiotic crossing over and MMR functions (Table 2.3). EAY3255 and derivatives contain the *lys2::insE-A*$_{14}$ allele to measure mutation rate [22]. EAY3252/EAY3486 (*wild-type*), EAY3255/EAY3486 (*mlh3Δ*) and EAY3255::*mlh3* alleles/EAY3486 diploids contain spore autonomous fluorescence markers to measure meiotic crossing over in the *CEN8-THR1* interval [44]. The S288c strain EAY3097 (relevant genotype *pms1Δ*, *lys2::insE-A*$_{14}$) was used to determine if *mlh3* alleles (SK1 background) expressed on *2μ* vectors could restore MMR functions in *pms1Δ* strains (Table 2.2). The S288c strain EAY4595 (relevant genotype *mlh3Δ* was used to determine if the SK1 derived *MLH3* gene could complement the MMR defects seen in the strain.

***lys2-A*$_{14}$ reversion assay**

The haploid strains described in Tables 2.1, 2.2, and 2.4 were analyzed for MMR functions using the *lys2-A*$_{14}$ reversion assay ([22]; Fig 2.1A). *ARS-CEN* and *2μ* vectors were maintained by growing strains in minimal leucine dropout media. Rates of *lys2::insE-A*$_{14}$ reversion were calculated as $\mu = f/\ln(N \cdot \mu)$ where f is the reversion frequency and N is the total number of revertant in the culture [22]. For each strain, 15-44 independent cultures, obtained from two to four independent transformants, were assayed on at least two different days to prevent batch effects, and 95% confidence intervals were determined as described by Dixon and Massey [76]. The Mann Whitney U test was used to calculate median reversion rates [77].

Figure 2.10. MPM chimera substitution constructs. pEAI403 containing the MPM chimeric construct was modified in plasmids pEAI428 to pEAI435 (Table 2.3) to contain Mlh3 linker regions of the indicated sizes (red font) inserted into the Pms1 linker (black font). The amino acids mutated in the *mlh3-32* allele (shown as KMK) are shown. The exact junction points for the Mlh3 and Pms1 linker sequences are shown with the surrounding Mlh3 amino acid sequences in the MPM chimera highlighted in purple font.

Spore autonomous fluorescence assay to measure percent tetratype

Diploids in the EAY3252/EAY3486 background (Table 2.4) were used for analysis of meiotic crossing over phenotypes. These diploids contain a spore autonomous fluorescent protein marker (CFP) linked to *THR1* of Chromosome VIII and RFP marker linked to *CEN8* in the second copy (Fig 2.1B; [44]). Diploids were selected by mating parental and derived EAY3252 and EAY3486 strains on media lacking tryptophan and leucine and maintained as stable strains. Fluorescence microscopy was used to quantify parental ditypes and tetratypes resulting from single crossover events. Sporulation plates were prepared as described by Detloff et al. [74] and incubations were performed at 30ºC. Spores were treated with 0.5% NP40 and sonicated for 5-10 seconds before analysis using the Zeiss Axioimager.M2. 250-1,000 tetrads for each *mlh3* allele were counted to determine % tetratype (# tetratypes/(tetratypes + parental ditypes)) two to three independent transformants were measured per allele on at least two different days to prevent batch effects. $\chi 2$ analysis was used to classify each allele as exhibiting a wild-type, intermediate, or null phenotype (http://vassarstats.net/). In this assay, wild-type SK1 *S. cerevisiae* strains gave single crossover events at 37.1% frequency, whereas *mlh3* null strains gave single crossover events at 18.3% frequency (Table 2.1). A Pearson Chi-Squared contingency test (http://vassarstats.net/) was used to test statistical significance. We applied a Benjamini-Hochberg correction at a 5% false discovery rate to minimize α inflation due to multiple comparisons (30 comparisons, with a $p<0.0183$ cut off for significance; 2.5 Table).

Table 2.4 Strains used in this study.

Parental S288c strains (Tables 2.1 and 2.2)	
EAY1269	*MATa, ura3-52, leu2D1, trp1D63, lys2::insE-A$_{14}$*
EAY4595	EAY1269, *mlh3D::KanMX*
EAY3097	*MATa, ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys::insE-A$_{14}$, pms1Δ::KanMX4*
Parental SK1 strains (Table 2.1)	
EAY3252	*MATα ho::hisG, ura3, leu2::hisG, trp1::hisG, ADE2, HIS4, CEN8Tomato::LEU2, MLH3, lys2::insE-A$_{14}$*
EAY3255	*MATα ho::hisG, ura3, leu2::hisG, trp1::hisG, ADE2, his4xB, CEN8Tomato::LEU2, mlh3Δ::NATMX, lys2::insE-A14*
EAY3486	*MATa ho::LYS2, lys2, ura3, leu2::hisG, trp1::hisG, THR1::m-Cerulean-TRP1, mlh3Δ::NatMX*
Chimera integrations in the *MLH3* locus in EAY3255 (Table 2.1)	
EAY3937	*PMM::KanMX*
EAY3938	*PPP::KanMX*
EAY3939	*MMP::KanMX*
EAY3940	*PPM::KanMX*
EAY3941	*MPP::KanMX*
EAY3943	*PMP::KanMX*
EAY3942	*MPM::KanMX*
EAY4097-4099	*MPM-19::KanMX*
EAY4100-4102	*MPM-40::KanMX*
EAY4103-4105	*MPM-60::KanMX*
EAY4106-4108	*MPM-80::KanMX*
EAY4109-4111	*MPM-100::KanMX*
EAY4112-4113	*MPM-N-full::KanMX*
EAY4116-17	*MPM-middle-full::KanMX*
EAY4118-4120	*MPM-C-full::KanMX*
Block mutation integrations at the *MLH3* locus in EAY3255 (Table 2.1)	
EAY4563-4564	Block 1 (ATP binding)-*mlh3-K17T,A20Q,S24D,R30K,Q34D::KanMX*
EAY4565-4567	Block 2 (Mlh1 interaction)-*mlh3-Y493M,N497G,V499F,D500N,K502G::KanMX*
EAY4575-4577	Block 2-*mlh3-D500N::KanMX*
EAY4578-4580	Block 2-*mlh3-K502G::KanMX*
EAY4581-4583	Block 3 (Endonuclease motif)-*mlh3-R532N::KanMX*
EAY4587-4589	Block 3-*mlh3-R530K::KanMX*
EAY4590-4591	Block 3-*mlh3-R530K,R532::KanMX*
EAY4121-4123	Block 4 (PCNA interaction motif)-*mlh3-PIP1::KanMX*
EAY4124-4126	Block 4 -*mlh3-PIP2::KanMX*
EAY4568-4570	Block 5 (Helix 2)-*mlh3-V660K,N666A,F676I,D678K::KanMX*
EAY4584-4586	Block 5-*mlh3-D678K::KanMX*
EAY4571-4572	Block 6 (Helix 1)-*mlh3-C695L,F699W,A702P,S707T,V709R,P710H::KanMX*
EAY4573-4574	Block 1 (*ATP binding),* Block 4 *(PIP2)::KanMX* combination
EAY4592-4594	Block 5 (*Helix 2*), Block 6 (*Helix 1*)*::KanMX* combination

lys2-A$_{14}$ reversion assays were performed on the SK1 strains EAY3252 (*wild-type*), EAY3255 (*mlh3Δ*), and EAY3255 derivatives containing integrated *mlh3$_{SK1}$::KanMX* alleles (Table 2.1). Meiotic crossover assays were performed on the following diploids: EAY3252/EAY3486 (*wild-type*), EAY3255/EAY3486 (*mlh3Δ*), and EAY3255::*mlh3* alleles/EAY3486 (Table 2.1). *lys2-A$_{14}$* reversion assays were performed on the S288c strains EAY1269, EAY4595::*mlh3Δ*, and EAY3097 transformed with plasmids as indicated (Tables 2.1 and 2.2).

Evolutionary analysis of fungal Mlh1, Mlh3 and Pms1 proteins

The inferred amino acid sequences of Mlh1, Mlh3 and Pms1 were taken from the complete genome sequences of 34 diverse fungal species (GenBank, https://www.ncbi.nlm.nih.gov/genbank/; Fig 2.3). *E. coli* MutS and *B. subtilis* MutL were used to potentially root the tree. 29 Mlh3, 34 Pms1 and 33 Mlh1 proteins were analyzed in the tree analysis. A gene missing from a particular species does not necessarily indicate gene loss in that species but could also reflect incomplete genome sequencing or other bioinformatic difficulty in locating the orthologous sequence. To decipher relationships between MutL homologs, amino acid sequences from the entire open reading frames of the Mlh1, Pms1 and Mlh3 proteins were aligned in SeaView (http://doua.prabi.fr/software/seaview) using the Clustal Omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/). Phylogenetic analyses were performed with PhyML (http://www.atgc-montpellier.fr/phyml/) in SeaView using model-provided amino acid equilibrium frequencies, optimized across site rate variations, and bootstrapping with 100 replicates. A second phylogenetic analysis was performed on a subset of more conserved alignment columns as selected by the Gblocks program (Fig 2.4; [78]). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. This reduced the alignment from 1,509 to 212 highly conserved and more confidently aligned positions. A phylogenetic tree was then inferred using PhyML under the same parameters and model as before. Phylogenetic tree images were created and annotated using interactive Tree of Life v.4 online tool [79].

Table 2.5. Statistics for meiotic crossing over data presented in Table 2.1

mlh3 allele	PD	TT	PD+TT	TT fraction	P-value to *MLH3*	P-value to null	Phenotype
MLH3	641	382	1023	0.37	**1**	<0.0001	Plus
mlh3Δ	1002	227	1229	0.18	<0.0001	**1**	minus
PPP (PMS1)	451	98	549	0.18	<0.0001	0.7548	minus
PMM	431	102	533	0.19	<0.0001	0.7415	minus
MMP	412	112	524	0.21	<0.0001	0.1588	minus
PPM	661	163	824	0.20	<0.0001	0.4578	minus
MPP	650	187	837	0.22	<0.0001	0.0309	minus
PMP	1020	267	1287	0.21	<0.0001	0.1509	minus
MPM	940	256	1196	0.21	<0.0001	0.0705	minus
MPM-19	365	84	449	0.19	<0.0001	0.9116	minus
MPM-40	419	121	540	0.22	<0.0001	0.0551	minus
MPM-60	409	98	507	0.19	<0.0001	0.6765	minus
MPM-80	401	107	508	0.21	<0.0001	0.2123	minus
MPM-100	401	120	521	0.23	<0.0001	0.0286	minus
MPM-N-full	393	107	500	0.21	<0.0001	0.1618	minus
MPM-middle-full	456	104	560	0.19	<0.0001	0.9593	minus
MPM-C-full	412	109	521	0.21	<0.0001	0.2339	minus
mlh3-K17T, A20Q, S24D, R30K, Q34D	330	183	513	0.36	0.5224	<0.0001	Plus
mlh3-Y493M, N497G, V499F, D500N, K502G	545	224	769	0.29	0.0003	<0.0001	Plus/minus
mlh3-D500N	536	265	801	0.33	0.0593	<0.0001	Plus
mlh3-K502G	616	145	761	0.19	<0.0001	0.7456	minus
mlh3-R530K	576	195	771	0.25	<0.0001	0.0003	Plus/minus
mlh3-R532N	546	228	774	0.29	0.0005	<0.0001	Plus/minus
mlh3-R530K,R532N	405	104	509	0.20	<0.0001	0.3432	minus
mlh3-PIP1	513	229	742	0.31	0.0047	<0.0001	Plus/minus
mlh3-PIP2	498	294	792	0.37	0.9234	<0.0001	Plus
mlh3-V660K,N666A,F676I,D678K	622	138	760	0.18	<0.0001	0.8612	minus
mlh3-D678K	524	248	772	0.32	0.0219	<0.0001	Plus/minus
mlh3-C695L,F699W,A702P, S707T,V709R,P710H	434	103	527	0.20	<0.0001	0.7246	minus
Block 1 (*ATP binding*), Block 4 (*PIP2*)	348	161	509	0.32	0.0277	<0.0001	Plus/minus
Block 5 (*Helix 2*), Block 6 (*Helix 1*)	422	92	514	0.18	<0.0001	0.7784	minus

P-values were determined using a Pearson Chi-Squared contingency test, Vassar Stats, with a Benjamini-Hochberg for 32 comparisons. A $p<0.002$ cut off was used for statistical significance.

Evolutionary rate covariation (ERC) analysis

The 18 species included for the ERC analysis were: *S. cerevisiae, S. paradoxus, S. mikatae, S. bayanus, Naumovozyma castellii, Candida glabrata, Vanderwaltozyma polysporus, Lachancea kluyveri, L. thermotolerans, L. waltii, Kluyveromyces lactis, Eremothecium gossypii, C. tropicalis, C. albicans, C. dubliniensis, C. lusitaniae, C. guilliermondii, and Debaryomyces hansenii*. Multiple species alignments for Mlh1, Mlh3, and Pms1 were subdivided into N-terminal, Linker, and C-terminal domains so that they could be analyzed separately. ERC values between a given pair of proteins or domains were calculated as initially described in Clark et al. [39,40] with modifications made to improve the relative rate normalization (Fig 2.5) [80]. Briefly, the branch lengths for each domain or protein tree were first normalized into relative evolutionary rates (RERs) using the RERconverge package [81]. Normalization vectors were estimated from the branch lengths of 4,458 proteins plus the three MLH domains in Mlh1, Pms1 and Mlh3 (9 in total). The resulting RERs were used to calculate the ERC values between all pairwise comparisons of the three domains in Mlh1, Mlh3, and Pms1. ERC values were calculated as the Pearson correlation coefficients between each pair of domains with extreme values controlled by limiting the 2 most extreme positive and negative branch RERs to the third most extreme value (*i.e.*, Winsorization of the 2 most extreme outliers).

Using multi-Harmony to identify sites for mutagenesis in *MLH3*

To identify specificity determining residues in Mlh3 and Pms1, alignments of Pms1 and Mlh3 from 34, for Pms1, and 29, for Mlh3, fungal species were used for multi-Harmony analysis (Fig 2.3; Fig 2.7; Fig 2.8A-D; [52]). Using multi-Harmony, amino acid positions significantly different between the two groups of sequences were identified. An identity cutoff of a

combination of a score for multi-relief greater than 0.8 and a score less than 0.5 for sequence harmony was used to reduce the number of interesting amino acids resulting in a set of groups-specific positions. Once identified, the amino acids were mapped onto a homolog model of Mlh3 from Al-Sweel et al. [46] using PyMOL to verify and group amino acids in close physical proximity. This analysis led to the identification of Blocks 1, 2, 5, and 6 (Fig 2.7; Fig 2.8A-D). Blocks 3 and 4 were identified for analysis based on the identification of endonuclease and PCNA binding motifs found in *B. subtilis* MutL and a subset of MLH homologs ([54]; Fig 2.2C).

REFERENCES

1. Ohno S (1970) Evolution by gene duplication. New York: Springer-Verlag.

2. Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc London B 256: 119-124.

3. Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.

4. Marcet-Houben M, Gabaldón T (2015) Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLoS Biol. 13: e1002220. doi: 10.1371/journal.pbio.1002220.

5. Culligan KM (2000) Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. Nucleic Acids Res 28: 463–471.

6. Furman CM, Elbashir R, Alani E (2020) Expanded roles for the MutL family of DNA mismatch repair proteins. Yeast doi: 10.1002/yea.3512.

7. Pyatnitskaya A, Borde V, DeMuyt A (2019) Crossing and zipping: molecular duties of the ZMM proteins in meiosis. Chromosoma 128: 181-198.

8. Kunkel TA, Erie DA (2015) Eukaryotic mismatch repair in relation to DNA replication. Ann Rev Genet 49: 291-313.

9. Manhart CM, Alani E (2016) Roles for mismatch repair family proteins in promoting meiotic crossing over. DNA Repair 38: 84-93.

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

11. Zakharyevich K, Tang S, Ma Y, Hunter N (2012) Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. Cell 149: 334-347.

12. Abdullah MF, Hoffmann ER, Cotton VE, Borts RH (2004) A role for the MutL homolog MLH2 in controlling heteroduplex formation and in regulating between two different crossover pathways in budding yeast. Cytogenet Genome Res 107: 180-190.

13. Campbell CS, Hombauer H, Srivastan A, Bowen N, Gries K, et al. (2014) Mlh2 is an accessory factor for DNA mismatch repair in *Saccharomyces cerevisiae*. PLoS Genetics 10: e1004327

14. Duroc Y, Kumar R, Ranjha L, Adam C, Guérois R et al. (2017) Concerted action of the MutLβ heterodimer and Mer3 helicase regulates the global extent of meiotic gene conversion. eLife 6: e21900.

15. Harfe BD, Minesinger BK, 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.

16. Shell SS, Putnam CD, Kolodner RD (2007) Chimeric *Saccharomyces cerevisiae* Msh6 protein with an Msh3 mispair-binding domain combines properties of both proteins. Proc Natl Acad Sci USA 104: 10956-10961.

17. Romanova NV, Crouse GF (2013) Different roles of eukaryotic MutS and MutL complexes in repair of small insertion and deletion loops in yeast. PLoS Genet 9: e1003920.

18. Erdeniz N, Dudley S, Gealy R, Jinks-Robertson S, Liskay RM (2005) Novel *PMS1* alleles preferentially affect the repair of primer strand loops during DNA replication. Mol Cell Biol 25: 9221-9231.

19. Kadyrov FA, Dzantiev L, Constantin N, Modrich P (2006) Endonucleolytic function of MutLalpha in human mismatch repair. Cell 126: 297–308.

20. Kawasoe Y, Tsurimoto T, Nakagawa T, Masukata H, Takahashi TS (2016) MutSα maintains the mismatch repair capability by inhibiting PCNA unloading. eLife 5: e15155.

21. Pluciennik A, Dzantiev L, Iyer RR, Constantin N, Kadyrov FA, Modrich P (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc Natl Acad Sci USA 107: 16066–16071.

22. Tran HT, Keen JD, Kricker M, Resnick MA, Gordenin DA (1997) Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol Cell Biol 17: 2859–2865.

23. Goellner EM, Putman CD, Kolodner RD (2015) Exonuclease 1-dependent and independent mismatch repair. DNA Repair 32: 24–32.

24. Kim Y, Furman C, Manhart C, Alani E, Finkelstein I (2019) Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex. Nucleic Acids Res 47: 1823-1835.

25. Flores-Rozas H, Kolodner RD (1998) The *Saccharomyces cerevisiae MLH3* gene functions in MSH3-dependent suppression of frameshift mutations. Proc Natl Acad Sci USA 95: 12404-12409.

26. Robine N, Uematsu N, Amiot F, Gidrol X, Barillot E, Nicolas A, Borde V (2007) Genome-wide redistribution of meiotic double-strand breaks in *Saccharomyces cerevisiae*. Mol Cell Biol 27: 1868–1880.

27. Keeney S, Giroux CN, Kleckner N (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88: 375-384.

28. Marsolier-Kergoat MC, Khan MM, Schott J, Zhu X, Llorente B (2018) Mechanistic view and genetic control of DNA recombination during meiosis. Mol Cell 70: 9–20. https://doi.org/10.1016/j.molcel.2018.02.032

29. Manhart CM, Ni X, White MA, Ortega J, Surtees JA, Alani E (2017) The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. PLoS Biol 15: e2001164.

30. Cannavo E, Sanchez A, Anand R, Ranjha L, Hugener J, et al. (2020) Regulation of the MLH1-MLH3 endonuclease in meiosis. Nature 586: 618-622. doi: 10.1038/s41586-020-2592-2.

31. Kulkarni DS, Owens SN, Honda M, Ito M, Yang Y et al. (2020) PCNA activates the MutLγ endonuclease to promote meiotic crossing over. Nature 586: 623-627. doi: 10.1038/s41586-020-2645-6.

32. Kolas NK, Svetlanov A, Lenzi ML, Macaluso FP, Lipkin SM et al. (2005) Localization of MMR proteins on meiotic chromosomes in mice indicates distinct functions during prophase I. J Cell Biol 171: 447–458. https://doi.org/10.1083/jcb.200506170

33. Moens PB, Kolas NK, Tarsounas M, Marcon E, Cohen PE, Spyropoulos B (2002) The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA–DNA interactions without reciprocal recombination. J Cell Sci 115: 1611–1622.

34. Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, et al. (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet 9: e1003930.

35. Arana, ME, Holmes, SF, Fortune, JM, Moon, AF, Pederson, LC, Kunkel, TA (2010) Functional residues on the surface of the N-terminal domain of yeast Pms1. DNA Repair 9: 448-457.

36. Ban C, Yang W (1998) Crystal structure and ATPase activity of MutL: Implications for DNA repair and mutagenesis. Cell 95: 541–552.

37. Guarné A, Ramon-Maiques S, Wolff EM, Ghirlando R, Hu X, Miller JH, Yang W (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. EMBO J 23: 4134–4145.

38. Gueneau E, Dherin C, Legrand P, Tellier-Lebegue C, Gilquin B et al. (2013) Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. Nat Struct Mol Biol 20: 461–468.

39. Clark NL, Alani E, Aquadro CF (2013) Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. Genetics 193: 529–538.

40. Clark NL, Alani E, Aquadro CF (2012) Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res *22:* 714-720.

41. Sacho EJ, Kadryrov FA, Modrich P, Kunkel TA, Erie DA (2008) Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha. Mol Cell: 29: 112–121.

42. Ban C, Junop M, Yang W (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell 97: 85–97.

43. Pillon MC. Lorenowicz JJ, Uckelmann M, Klocko AD, Mitchell RR, et al. (2010) Structure of the endonuclease domain of MutL: unlicensed to cut. Mol Cell 39: 145–151.

44. Thacker D, Lam I, Knop M, Keeney S (2011) Exploiting Spore-Autonomous Fluorescent protein expression to quantify meiotic chromosome behaviors in *Saccharomyces cerevisiae*. Genetics 189: 423–439.

45. Argueso, J. L., Kijas, A. W., Sarin, S., Heck, J., Waase, M., Alani, E (2003) Systematic mutagenesis of the *Saccharomyces cerevisiae MLH1* gene reveals distinct roles for Mlh1p in meiotic crossing over and in vegetative and meiotic mismatch repair. Mol Cell Biol 23: 873-886.

46. Al-Sweel N, Raghavan V, Dutta A, Ajith VP, DiVietro L et al. (2017) *mlh3* mutations in baker's yeast alter meiotic recombination outcomes by increasing noncrossover events genome-wide. PLoS Genetics 13: e1006974.

47. Cotton VE, Hoffmann ER, Borts RH (2010) Distinct Regulation of Mlh1p Heterodimers in Meiosis and Mitosis in *Saccharomyces cerevisiae*. Genetics 185: 459-467.

48. Sonntag Brown M, Lim E, Chen C, Nishant KT, Alani E (2013) Genetic analysis of *mlh3* mutations reveals interactions between crossover promoting factors during meiosis in baker's yeast. G3 3: 9-22.

49. Claeys Bouuaert C, Keeney S (2017) Distinct DNA-binding surfaces in the ATPase and linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. PLoS Genet 13: e1006722. https://doi.org/10.1371/journal.pgen.1006722

50. Shcherbakova PV, Kunkel TA (1999) Mutator phenotypes conferred by MLH1 overexpression and by heterozygosity for *mlh1* mutations. Mol Cell Biol 19: 3177-3183.

51. Smith CE, Mendillo ML, Bowen N, Hombauer H, Campbell CS, Desai A, Putnam CD, Kolodner RD (2013) Dominant mutations in S. cerevisiae PMS1 identify the Mlh1-Pms1 endonuclease active site and an exonuclease 1-independent mismatch repair pathway. PLoS Genet 9: e1003869.

52. Brandt BW, Feenstra KA, Heringa J (2010) Multi-Harmony: Detecting functional specificity from sequence alignment. Nucleic Acids Res 38: W35-W40.

53. Kelman Z (1997) PCNA: Structure, functions and interactions. Oncogene 14: 629–640.

54. Pillon MC, Miller JH, Guarné A (2011) The endonuclease domain of MutL interacts with the β sliding clamp. DNA Repair 10: 87–93.

55. Genschel J, Kadryrova LY, Iyer RR, Dahal BK, Kadyrov FA, Modrich P (2017) Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair. Proc Natl Acad Sci USA 114: 4930–4935.

56. Hombauer H, Campbell CS, Smith CE, Desai A, Kolodner RD (2011) Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. Cell 147: 1040–1053.

57. Rogacheva MV, Manhart CM, Chen C, Guarne A, Surtees J, Alani E (2014) Mlh1-Mlh3, A meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. J Biol Chem 289: 5664–5673.

58. Ranjha L, Anand R, Cejka P (2014) The *Saccharomyces cerevisiae* Mlh1-Mlh3 heterodimer is an endonuclease that preferentially binds to Holliday Junctions. J Biol Chem 289: 5674–5686.

59. Ho B, Baryshnikova A, Brown GW (2018) Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. Cell Syst 6: 192-205.

60. Karim AS, Curran KA, Alper HS (2013) Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering operations. FEMS Yeast Res 13: 107-116. doi: 10.1111/1567-1364.12016.

61. Nishant KT, Plys AJ, 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. doi: 10.1534/genetics.108.086645.

62, Parker BO, Marinus MG (1992) Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. Proc Natl Acad Sci USA 89: 1730-1734. doi: 10.1073/pnas.89.5.1730.

63. Jensen LE, Jauert PA, Kirkpatrick DT (2005) The large loop repair and mismatch repair pathways of *Saccharomyces cerevisiae* act on distinct substrates during meiosis. Genetics 170: 1033-1043. doi: 10.1534/genetics.104.033670.

64. Brown MS, Bishop DK (2014) DNA strand exchange and RecA homologs in meiosis. Cold Spring Harb Perspect Biol 7: a016659.

65. Chan YL, Zhang A, Weissman BP, Bishop DK (2019) RPA resolves conflicting activities of accessory proteins during reconstitution of Dmc1-mediated meiotic recombination. Nucleic Acids Res 47: 747–761. https://doi.org/10.1093/nar/gky1160

66. Lee JY, Terakawa T, Qi Z, Steinfeld JB, Redding S, Kwon Y, Greene EC (2015) Base triplet stepping by the Rad51/RecA family of recombinases. Science 349: 977–981. https://doi.org/10.1126/science.aab2666

67. Steinfeld JB, Beláň O, Kwon Y, Terakawa T, Al-Zain A, Smith MJ, Greene EC (2019) Defining the influence of Rad51 and Dmc1 lineage-specific amino acids on genetic recombination. Genes Dev 33: 1191–1207. https://doi.org/10.1101/gad.328062.119

68. Neale MJ, Keeney S (2006) Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature 442: 153–158. doi:10.1038/nature04885

69. Lao JP, Cloud V, Huang CC, Grubb J, Thacker D et al. (2013) Meiotic crossover control by concerted action of Rad51–Dmc1 in homolog template bias and robust homeostatic regulation. PLoS Genet 9: e1003978. doi:10.1371/journal.pgen.1003978

70. Hsieh YP, Makrantoni V, Robertson D, Marston AL, Murray AW (2020) Evolutionary repair: changes in multiple functional modules allow meiotic cohesin to support mitosis. PLoS Biol 18: e3000635.

71. Peltomaki P (2016) Update on lynch syndrome genomics. Fam Cancer 15: 385–393. doi: https://doi.org/10. 1007/s10689-016-9882-8, PMID: 26873718

72. Rose MD, Winston F, Hieter P (1990) Methods in yeast genetics: A Cold Spring Harbor Laboratory course manual, Cold Spring Harbor Laboratory Press, NY.

73. Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast 15: 1541-1553.

74. Detloff P, Sieber J, Petes TD (1991) Repair of specific base pair mismatches formed during meiotic recombination in the yeast *Saccharomyces cerevisiae*. Mol Cell Biol 11: 737–745.

75. Gietz RD, Schiestl RH, Willems AR, Woods RA (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast 11: 355–360.

76. Dixon WJ, Massey FJ (1969) Introduction to statistical analysis. McGraw-Hill, New York.

77. Drake JW (1991) Spontaneous mutation. Ann Rev Genet, 25: 125–146.

78. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Bio Evol 17: 540-552. DOI: 10.1093/oxfordjournals.molbev.a026334

79. Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47: W256–W259, https://doi.org/10.1093/nar/gkz239

80. Partha R, Kowalczyk A, Clark NL, Chikina M (2019) Robust method for detecting convergent shifts in evolutionary rates. Mol Biol Evol 36: 1817-1830. doi: 10.1093/molbev/msz107.

81. Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, Chikina M (2019) RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35: 4815-4817. doi: 10.1093/bioinformatics/btz468.

82. Manning JR, Jefferson ER, Barton GJ (2008) The contrasting properties of conservation and correlated phylogeny in protein functional residue prediction. BMC Bioinformatics 9: 51. doi: 10.1186/1471-2105-9-51.

83. Christianson TW, Sikorksi RS, Dante M, Shero JH, Hieter P (1992) Multifunctional yeast high-copy-number shuttle vectors. Gene 110: 119–122.

CHAPTER 3

Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex

This chapter was published in the journal Nucleic Acids Research: Kim, Y., Furman, C. M., Manhart, C. M., Alani, E., & Finkelstein, I. J. (2019). Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex. *Nucleic Acids Research, 47,* 1823-1835.

ABSTRACT

Intrinsically disordered regions (IDRs) are present in at least 30% of the eukaryotic proteome and are enriched in chromatin-associated proteins. Here, we use a combination of genetics, biochemistry, and single-molecule biophysics to characterize how IDRs regulate the functions of yeast MutLα (Mlh1-Pms1), a critical DNA mismatch repair (MMR) complex. Shortening or scrambling the IDRs in both subunits ablates MMR in vivo. Mlh1-Pms1 complexes with shorter IDRs that disrupt MMR retain wild-type DNA binding affinity but are impaired for diffusion on both naked and nucleosome-coated DNA. Moreover, the IDRs also regulate the ATP hydrolysis and nuclease activities that are encoded in the structured N- and C-terminal domains of the complex. This combination of phenotypes underlies the catastrophic MMR defect seen with the mutant Mlh1-Pms1 in vivo. More broadly, this work highlights an unanticipated multi-functional role for IDRs in regulating both facilitated diffusion on chromatin and nucleolytic processing of a DNA substrate.

INTRODUCTION

Intrinsically disordered regions (IDRs) are structurally heterogeneous protein domains that encode diverse functions. IDRs are conformationally flexible, facilitating interactions with multiple partners through intramolecular and intermolecular mechanisms (1, 2). IDRs are often found as linkers connecting functional domains where they can regulate protein stability (1). IDRs are prevalent in chromatin-binding proteins, and the IDRs in these proteins have been implicated in bridging DNA strands, chromatin remodeling, and interacting with other key proteins in DNA metabolic pathways (3, 4). Moreover, IDRs in transcription factors and single-strand DNA binding (SSB) proteins have been reported to tune the DNA binding affinities of these proteins (5–10). Whether these IDRs also regulate scanning on chromatin and other catalytic processes is an open question. This is partly because mutations in such regions often do not confer a specific phenotype, and in some cases, the amino acid sequences contained within IDRs, which are typically poorly conserved among family members, can be critical for the function of a specific IDR-containing protein. Using the mismatch repair protein Mlh1-Pms1 as a case study, we explore the role of IDRs in regulating the DNA scanning and enzymatic activities of a critical eukaryotic DNA repair factor.

The MutL homolog family protein MutLα (MLH; Mlh1-Pms1 in baker's yeast) is essential for eukaryotic DNA mismatch repair (MMR). Mlh1-Pms1 organizes into a ring-like structure that links the ordered N- and C-terminal domains via 160-290 amino acid-long IDRs (11–16) (**Figure 3.1A**; amino acids 335-499 in Mlh1, 364-659 in Pms1). Mlh1-Pms1 searches for MutS homologs (MSH) bound to DNA mismatches (16–18). A latent MLH endonuclease activity then nicks the newly-synthesized DNA strand resulting in excision of the mismatch (19).

This activity requires PCNA, and multiple nicks are thought to enhance the excision step of MMR (20–27).

All MLH-family proteins encode an IDR between the structured N- and C-termini. However, the functional role(s) of the IDRs in Mlh1-Pms1 is enigmatic. The composition and length of the MLH IDRs are critical for efficient MMR in yeast, and missense/deletion mutations within these linkers are found in human cancers (11, 28, 29). We previously proposed that the Mlh1-Pms1 IDRs are sufficiently long to accommodate a nucleosome within the complex, possibly allowing Mlh1-Pms1 to navigate on chromatin *in vivo* (17, 30). In support of this model Mlh1-Pms1 foci that were visualized in live yeast were short-lived (~1.5 min on average), and displayed rapid movements in the nucleus (31). In addition, the Mlh1-Pms1 IDRs display nucleotide-dependent conformational transitions, with ATP binding bringing the N- and C-terminal subunits close together (32–34). This ATPase activity is required for MMR *in vivo* and can stimulate the endonuclease activity *in vitro* (20, 35–38). ATP-dependent conformational rearrangements involving the IDRs are hypothesized to position bound DNA near the endonuclease active site and presumably change Mlh1-Pms1 affinity for DNA (32, 34, 35). Together, these studies suggest that MLH proteins may use conformational changes mediated by the ATP cycle to modulate affinity for DNA, navigate on chromatin, and introduce nicks on a DNA substrate for efficient MMR. However, these possible functions of the IDRs have not been tested directly.

Here, we use a combination of genetics, ensemble biochemistry, and single-molecule biophysics to investigate how the Mlh1-Pms1 IDRs promote both DNA scanning and nuclease activities. We show that both the sequence composition and the precise length of the IDRs are required for optimal MMR *in vivo*. Having mapped genetic requirements for MMR, we next

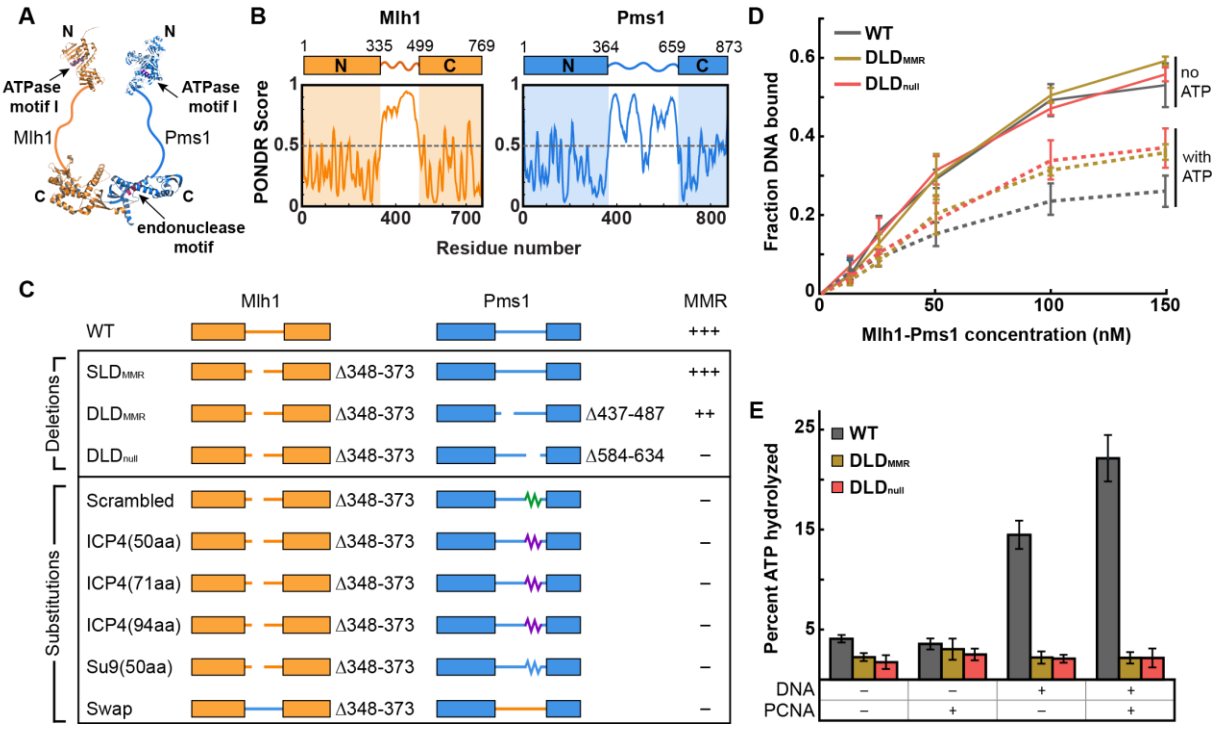


Figure 3.1. The IDR of Mlh1-Pms1 is critical for MMR *in vivo* and ATP hydrolysis *in vitro*. (A) Illustration of Mlh1-Pms1 highlighting the structured N- and C-terminal domains separated by IDRs (solid lines). (B) Bioinformatic prediction of long IDRs in both Mlh1 (amino acids 335-499) and Pms1 (amino acids 364-659) using the PONDR VSL2 predictor (73). Any value above 0.5 is considered disordered. (C) Schematic of IDR sequence changes made in Mlh1-Pms1, followed by the mutator phenotype conferred by the indicated alleles. +++ wild-type mutation rate, ++ hypomorph, – null. See text for a description of the specific sequences. (D) DNA binding activities for each complex analyzed by filter binding in the presence (dashed line) and absence (solid line) of 1 mM ATP. Mlh1-Pms1 variants were included at final concentrations of 12.5 nM, 25 nM, 50 nM, 100 nM, and 150 nM in buffer containing 25 mM NaCl. DNA binding of a 49 bp oligonucleotide was quantified by scintillation counting. Three replicates were averaged; error bars indicate ± one SD. (E) ATP hydrolysis activities of WT and mutant Mlh1-Pms1 complexes (0.40 μM) were determined alone, and in the presence of PCNA (0.5 μM), or 49-bp homoduplex DNA (0.75 μM), and both PCNA (0.5 μM) and 49-bp homoduplex DNA (0.75 μM). Error bars indicate ± one SD of three replicates.

biochemically characterized a double linker deletion (DLD) mutant that was almost completely defective in MMR (DLD_{null}), and another linker deletion mutant that retains partial *in vivo* MMR function (DLD_{MMR}). Interestingly, both mutants can bind and diffuse on DNA, retain single-round endonuclease activities, but show reduced DNA-dependent ATPase and nucleosome bypass activities. Furthermore, DLD_{null} is unable to navigate dense nucleosome arrays and is defective in multiple rounds of DNA nicking. These results establish that the IDRs license Mlh1-Pms1 to navigate chromatin and nick DNA at multiple sites to promote efficient MMR *in vivo*. They suggest that the IDRs play a critical role in regulating how a DNA repair enzyme scans chromatin for a specific target and how the enzyme activates its endonuclease activity. More broadly, these results expand the functions of IDRs in regulating the DNA scanning and enzymatic activities of chromatin-associated complexes.

MATERIAL AND METHODS

Bulk biochemical assays

DNA substrates for bulk biochemical assays: pUC18 (2.7 kb, Invitrogen) was used as the closed circular substrate for endonuclease assays presented in **Figure 3.7A** and Figure 3.8**A–B**. A 49-mer homoduplex DNA substrate was used in the DNA binding and ATPase experiments presented in **Figure 3.1D, E, and Figure 3.3B**. This substrate was made as follows. AO3142-5'-GGGTCAACGTGGGCAAAGATGTCCTAGCAAGTCAGAATTCGGTAGCGTG-3' was labeled on the 5' end with ^{32}P labeled phosphate using T4 polynucleotide kinase (New England Biolabs). Unincorporated nucleotide was removed using a P30 spin column (BioRad). The two oligonucleotides were annealed by combining end-labeled AO3142 with a 2-fold molar excess of unlabeled AO3144-5'-

CACGCTACCGAATTCTGACTTGCTAGGACATCTTTGCCCACGTTGACCC-3' in buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM EDTA. Annealing was accomplished by incubating the DNA substrates at 95 °C for 5 min, followed by cooling to 25 °C at a rate of 1 °C/min. Following annealing, excess single-stranded DNA was removed using an S300 spin column (GE). 2.7 kb pUC18 for endonuclease assays on circular DNA was purchased from Thermo. For **Figure 3.7B and Figure 3.8C–D**, the pBR322 plasmid (4.4 kb, Thermo) was linearized using *Hind*III (NEB) by incubation at 37 ºC for 60 min, followed by enzyme inactivation at 80 ºC for 20 min. Linearized fragments were isolated using a PCR clean-up kit (Zymo Research).

Protein purification: Yeast WT, DLD_{MMR} and DLD_{null} Mlh1-Pms1 variants (**Figure 3.3A**) were purified from galactose-induced *S. cerevisiae* BJ2168 (*MATa, ura3–52, leu2–3, 112, trp1–289, prb1–1122, prc1–407, pep4–3*) containing expression vectors as previously described (11, 39). Mlh1 contains a FLAG tag at position 499 in wild-type at the equivalent position in Mlh1 truncation mutants. Yeast RFC and PCNA were expressed and purified from *E. coli* (40, 41). RPA-RFP was expressed and purified from Rosetta(DE3)/pLysS cells as described previously (42).

Endonuclease assay: Endonuclease reactions were performed in a buffer containing: 20 mM HEPES- KOH (pH 7.5), 20 mM KCl, 2.5 mM $MnSO_4$, 0.2 mg/mL BSA, and 1 % glycerol(43). Reactions were stopped by the addition of 0.1 % SDS, 14 mM EDTA, and 0.1 mg/mL Proteinase K (NEB). For reactions on a circular DNA substrate, products were resolved by 1.2 % agarose gel containing 0.1 μg mL^{-1} ethidium bromide, which causes covalently closed circular DNA isoforms to separate from nicked DNA product. Gels were run in 1x TAE (Tris-acetate-EDTA) at 100 V for 45 min. Negative control lanes were used as background and were subtracted out of

reported quantifications. Endonuclease assays on linear substrates were carried out and stopped as described for circular DNA substrates. Denaturing agarose gels consist of 1 % (w/v) agarose, 30 mM NaCl, 2 mM EDTA pH 7.5 run in a buffer containing 30 mM NaOH and 2 mM EDTA(44). Immediately prior to sample loading, reactions were supplemented with 30 mM NaOH, 1 mM EDTA, 3 % glycerol, and 0.02 % bromophenol blue (final concentrations), heated for 5 min at 70 ºC, then cooled for 3 min on ice. Gels were run at 50 V for ~3 h. After running, alkaline agarose gels were neutralized in 0.5 M Tris base (pH 7.5) for 30 min and stained with 0.5 μg mL^{-1} ethidium bromide for ~2 h. GelEval (FrogDance Software, v1.37) was used to quantify gels.

Filter binding assay: DNA binding assays were performed as described previously (45). Briefly, 20 μL reactions containing 4 nM ^{32}P-labeled homoduplex substrate and 11 nM unlabeled homoduplex substrate were combined with increasing amounts of protein in a reaction buffer containing 20 mM Tris-HCl (pH 7.5), 20mM NaCl, 0.01 mM EDTA, 2 mM $MgCl_2$, 40 μg mL^{-1} BSA, and 0.1 mM DTT. Assays with nucleotide contain 1 mM ATP. Reactions were incubated for 10 min at 30 ºC after addition of WT, DLD_{MMR} and DLD_{null} Mlh1-Pms1. Reactions were then filtered through KOH-treated nitrocellulose filters using a Hoefer FH225V filtration device for approximately 1 min. Filters were analyzed by scintillation counting to determine DNA binding efficiency.

ATPase assay: ATPase activity was determined using the Norit A absorption method as described previously (43). Briefly, 30 μL reactions contained 0.4 μM of Mlh1-Pms1 (WT, DLD_{MMR} and DLD_{null}), 100 μM [γ-32P]-ATP, 20 mM Tris, pH 7.5, 2.0 mM $MgCl_2$, 0.1 mM DTT, 1 mM $MnSO_4$, 75 mM NaCl, 1 % glycerol, 40 μg/ml BSA. Reactions were incubated for

40 min at 37 °C. When specified, DNA (49-mer homoduplex DNA substrate as described above) and PCNA were included at 0.75 μM and 0.5 μM, respectively.

Strains and plasmids: Yeast strains were grown in yeast extract/ peptone/dextrose, minimal complete, or minimal selective media (46). Plasmids used in this study are listed in **Table 3.1**. Full details of plasmid and strain constructions are available upon request. Expression vectors were derived from pMH1 (*GAL1-MLH1-VMA-CBD,2μ, TRP1*) and pMH8 (*GAL10-PMS1,2μ, LEU2*) (39).

Linker arm replacement series: A series of *ARS-CEN* vectors were created to test if the 50 amino acid deletion made in the Pms1 linker arm (*pms1Δ584–634*) could be replaced by other sequences (**Table 3.1**). These vectors were derived from pEAA238, which expresses *PMS1* from its native promoter (47). Vectors used to overexpress and purify Mlh1-Pms1 were derived from pMH1 (*GAL1-MLH1-VMA-CBD,2μ, TRP1*) and pMH8 (*GAL10-PMS1,2μ, LEU2*)(39). Insertion plasmids were constructed using NEB HiFi DNA Assembly cloning (pEAA644-656) and Q5 mutagenesis (pEAA659-665). The desired DNA sequence (PCR amplified from specific plasmid or constructed as gBlocks, IDT) was inserted into the deleted region (amino acids 584 to 634) of the Pms1 linker (**Table 3.1**). The DNA sequence of vectors constructed using PCR amplified vector backbones and linker inserts were confirmed by DNA sequencing (Cornell BioResource Center).

lys2::insE-A_{14} reversion assay (**Table 3.2***):* Assays were performed as described previously (11). Briefly, pEAA238 (*PMS1*), pEAA548 (*pms1Δ584-634*) and derivative linker insertion plasmids of pEAA548 were transformed into EAY3097 (*MATa, ura3–52, leu2Δ1, trp1Δ63, his3Δ200, lys2::insE-A_{14}, pms1Δ::KanMX4*) using standard methods (46, 48). Plasmids were maintained by growing strains in minimal selective histidine dropout media. When tested in combination,

pEAA238, pEAA548 (*pms1Δ584-634*) or derivative linker insertion plasmids were co-transformed with pEAA213 (*MLH1*) or pEAA526 (*mlh1Δ348–373* (FLAG499)) into EAY1365 (*MATa, ura3–52, leu2Δ1, trp1Δ63, his3Δ200, lys2::insE-A$_{14}$, mlh1Δ::KanMX4, pms1Δ::KanMX4*). Plasmids were maintained by growing strains in minimal selective histidine and leucine dropout media. Null controls were transformed with pRS413 and pRS415 dummy vectors (49). Rates of *lys2::insE-A$_{14}$* reversion were calculated as $\mu=f/\ln(N\cdot\mu)$, where f is reversion frequency and N is the total number of revertants in the culture (50). For each strain, 15–45 independent cultures, obtained from two to three independent transformants bearing a unique allele, were assayed to determine the mutation rate; 95% confidence intervals and all computer-aided rate calculations were performed as previously described (11).

Single-molecule experiments and analysis

Data collection on TIRF microscopy: All single-molecule images were collected with Nikon Ti-E microscope equipped with a customized prism-TIRF configuration. The fluorescent samples were illuminated by a 488 nm laser (Coherent) or 532 nm laser (Coherent) through a quartz prism (Tower Optical Co.) depending on the fluorescent dye used. The laser light was adjusted to deliver 40 mW or 15 mW of power at the front face of the prism for 488 nm or 532 nm laser, respectively. Fluorescence was collected by two EM-CCD cameras (Andor iXon DU897, -80°C) using a 638 nm dichroic beam splitter (Chroma), and NIS-Elements software (Nikon) was used to collect the single-molecule data at 50 - 100 ms frame rates. All images were saved as TIFF files without compression for further image analysis in ImageJ (NIH). Experiments were conducted on a floating TMC optical table to avoid spatial drift.

Table 3.1. Plasmids used in this study

Plasmid	Relevant genotype	Vector type	Source
pRS413		*ARS-CEN, HIS3*	(Christianson et al., 1992)
pRS415		*ARS-CEN, LEU2*	(Christianson et al., 1992)
pEAA213	*MLH1*	*ARS-CEN, LEU2*	(Plys et al., 2012)
pEAA526	*mlh1*$\Delta_{348–373}$ *(FLAG*$_{499}$*)*	*ARS-CEN, LEU2*	(Plys et al., 2012)
pEAA238	*PMS1*	*ARS-CEN, HIS3*	(Plys et al., 2012)
pEAA544	*pms1*$\Delta_{437-487}$ *(HA*$_{565}$*)*	*ARS-CEN, HIS3*	(Plys et al., 2012)
pEAA548	*pms1*$\Delta_{584–634}$ *(HA*$_{565}$*)*	*ARS-CEN, HIS3*	(Plys et al., 2012)
pEAA644	*pms1*$_{50\text{-}scramble1}$	*ARS-CEN, HIS3*	This study
pEAA645	*pms1*$_{50\text{-}scramble2}$	*ARS-CEN, HIS3*	This study
pEAA646	*pms1*$_{50\text{-}scramble3}$	*ARS-CEN, HIS3*	This study
pEAA647	*pms1*$_{50\text{-}scramble4}$	*ARS-CEN, HIS3*	This study
pEAA648	*pms1*$_{50\text{-}scramble5}$	*ARS-CEN, HIS3*	This study
pEAA649	*pms1*$_{40\text{-}alpha\text{-}helix1}$	*ARS-CEN, HIS3*	This study
pEAA650	*pms1*$_{40\text{-}alpha\text{-}helix2}$	*ARS-CEN, HIS3*	This study
pEAA651	*pms1*$_{50\text{-}SRR\text{-}1}$	*ARS-CEN, HIS3*	This study
pEAA652	*pms1*$_{52\text{-}SRR}$	*ARS-CEN, HIS3*	This study
pEAA653	*pms1*$_{71\text{-}SRR}$	*ARS-CEN, HIS3*	This study
pEAA654	*pms1*$_{92\text{-}SRR}$	*ARS-CEN, HIS3*	This study
pEAA655	*pms1*$_{94\text{-}SRR}$	*ARS-CEN, HIS3*	This study
pEAA656	*pms1*$_{50\text{-}SRR\text{-}2}$	*ARS-CEN, HIS3*	This study
pEAA657	*pms1*$_{full\ mlh1\ Linker}$	*ARS-CEN, HIS3*	This study
pEAA658	*mlh1*$_{full\ pms1\ Linker}$	*ARS-CEN, LEU2*	This study
pEAA659	*pms1*$_{Y613A}$	*ARS-CEN, HIS3*	This study
pEAA660	*pms1*$_{Y594A}$	*ARS-CEN, HIS3*	This study
pEAA661	*pms1*$_{scramble584\text{-}593}$	*ARS-CEN, HIS3*	This study
pEAA662	*pms1*$_{scramble594\text{-}603}$	*ARS-CEN, HIS3*	This study
pEAA663	*pms1*$_{scramble604\text{-}613}$	*ARS-CEN, HIS3*	This study
pEAA664	*pms1*$_{scramble614\text{-}623}$	*ARS-CEN, HIS3*	This study
pEAA665	*pms1*$_{scramble624\text{-}633}$	*ARS-CEN, HIS3*	This study
pEAE269	*GAL1-MLH1(FLAG*$_{499}$*)-VMA1-CBD*	*2μ, TRP1*	(Plys et al., 2012)
pEAE308	*GAL1-mlh1*$\Delta_{348–373}$*(FLAG*$_{499}$*)-VMA1-CBD*	*2μ, TRP1*	(Plys et al., 2012)
pMH8	*GAL10-PMS1*	*2μ, LEU2*	(Hall and Kunkel, 2001)
pEAE388	*GAL10-pms1*$\Delta_{437-487}$	*2μ, LEU2*	This study
pEAE419	*GAL10-pms1*$\Delta_{584–634}$	*2μ, LEU2*	This study

Plasmid constructs expressing pms1 proteins in which the specified amino acid sequences replace those deleted in *pms1*$\Delta_{584–634}$.

Table 3.1 con't

Plasmid	Insert	*amino acid sequence replacement*
pEAA644	*pms1*$_{50\text{-}scramble1}$	*SSKSNKFGINSNKSLIDGKRNERFLMLDK LKNSEQIISTRESDSKYYETHI*
pEAA645	*pms1*$_{50\text{-}scramble2}$	*ISKMSKILKKNIESSSGEYNKSFLNSEQRN SYSGTLINLKDDTEDIKFRHR*
pEAA646	*pms1*$_{50\text{-}scramble3}$	*DTIERNGESIKLDFSESFKKSISTKSKLKN MESSIRYKIRGLLNDNYQSNH*
pEAA647	*pms1*$_{50\text{-}scramble4}$	*NKSSQSKIGLRSRLYIGTMTENSSKSFEDI KKHSDSNILLYKNFRDEKINE*
pEAA648	*pms1*$_{50\text{-}scramble5}$	*GGRNIKEINFKLIQIKKEDSLLSKSIDMR HTKSNFTNYESYSDKESRSSNL*
pEAA649	*pms1*$_{40\text{-}alpha\text{-}helix1}$	*NSRKSEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKQMSSII*
pEAA650	*pms1*$_{40\text{-}alpha\text{-}helix2}$	*SMISQEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKNKISSR*
pEAA651	*pms1*$_{50\text{-}SRR\text{-}1}$	*SSSTSSDSGSSSSSSASSSSGSSSTSSDSGSS SSSSASSSSGSGTMKHGT*
pEAA652	*pms1*$_{52\text{-}SRR}$	*RRSSSTSSDSGSSSSSSASSSSGSSSTSSDSG SSSSSSASSSSGSGTMKHGT*
pEAA653	*pms1*$_{71\text{-}SRR}$	*SSSTSSDSGSSSSSSASSSSGSSSTSSDSGSS SSSSASSSSGSSSTSSDSGSSSSSSASSSSGS GTMKHGT*
pEAA654	*pms1*$_{92\text{-}SRR}$	*SSSTSSDSGSSSSSSASSSSGSSSTSSDSGSS SSSSASSSSGSSSTSSDSGSSSSSSASSSSGSS STSSDSGSSSSSSASSSSGSGTMKHGT*
pEAA655	*pms1*$_{94\text{-}SRR}$	*RRSSSTSSDSGSSSSSSASSSSGSSSTSSDSG SSSSSSASSSSGSSSTSSDSGSSSSSSASSSSG SSSTSSDSGSSSSSSASSSSGSGTMKHGT*
pEAA656	*pms1*$_{50\text{-}SRR\text{-}2}$	*MASTRVLASRLASQMAASAKVARPAVRV AXVSKRTIQTGSPLQTRAYSS*
Plasmid constructs expressing pms1 proteins in which the specified amino acid sequences replace those deleted in the designated region		
pEAA661	*pms1*$_{scramble584\text{-}593}$	*ISKMSKILKK*
pEAA662	*pms1*$_{scramble594\text{-}603}$	*NIESSSGEYN*
pEAA663	*pms1*$_{scramble604\text{-}613}$	*KSFLNSEQRN*
pEAA664	*pms1*$_{scramble614\text{-}623}$	*SYSGTLINLK*
pEAA665	*pms1*$_{scramble624\text{-}633}$	*DDTEDIKFRH*

Full plasmid descriptions can be found in the Materials and Methods.

Preparation of single-molecule DNA substrates: DNA substrates for single-molecule imaging were prepared by modifying the cohesive ends of λ-DNA (New England Biolabs; NEB). Briefly, 125 μg λ-DNA was mixed with 2 μM IF003 and IF004 in T4 DNA ligase reaction buffer (NEB) and heated to 70°C for 15 minutes followed by gradual cooling to 15°C for 2 hours. After the oligomer hybridization, T4 DNA ligase (2000 units; NEB) was added to the mixture and incubated overnight at room temperature to seal nicks on DNA. The ligase was inactivated with 2 M NaCl, and the reaction was passed over an S-1000 gel filtration column (GE) to remove excess oligonucleotides and proteins. Typically, ~ 10 mL fractions from the first peak were collected and stored at 4˚C.

Nucleosomes were deposited on the DNA substrate as described previously with minor modifications (51). The DNA substrate was ligated to the oligo handles, mixed with sodium acetate (pH 5.5) to 0.3 M and isopropanol to 1:1 (v/v), and then precipitated by centrifugation at 15,000 g for 30 minutes. The invisible DNA precipitate was washed with 70% ethanol and dissolved in 2 M TE buffer (10 mM Tris–HCl pH 8.0, 1mM EDTA, 2 M NaCl) to obtain concentrated DNA at ~ 150 ng μL^{-1}. For reconstitution, 0.8 nM of the DNA was prepared in 2 M TE buffer with 1 mM DTT for a total volume of 100 μL. Human histone octamers (3xHA H2A with wild-type H2B, H3, H4; Histone Source) were added to the DNA, and the mixture was dialyzed using a mini dialysis button (10 kDa molecular weight cutoff, BioRad) against 400 mL dialysis buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA, 1 mM DTT, and gradually decreasing concentration of NaCl). The salt gradient dialysis was performed in a cold room and started with 1.5 M NaCl dialysis buffer for 1 hour. The buffer was exchanged every 2 hours to decrease salt in the order of 1 M, 0.8 M, 0.6 M, 0.4 M, and 0.2 M. The last 0.2 M NaCl buffer was used for

overnight dialysis. The ratio of DNA to octamer was adjusted to have 3 to 10 nucleosomes per DNA for single nucleosome bypass experiments.

Imaging Mlh1-Pms1 on DNA curtains: The Mlh1-Pms1 complexes used in this study contain a FLAG epitope tag at residue 499 on the Mlh1 subunit. We have previously confirmed that placing a FLAG epitope at this position supports full MMR activity *in vivo*, does not disrupt Mlh1-Pms1 biochemical activities (e.g., ATPase, nuclease, MSH2-6 interactions), and is suitable for single-molecule imaging (11, 15, 17, 30). 25 nM of FLAG-tagged proteins were conjugated with 30 nM biotinylated anti-FLAG antibody (Sigma-Aldrich, F9291-2MG) and 25 nM streptavidin QDs (Life Tech, Q10163MP) in a total volume of 60 μL on ice for 7 minutes. The mixture was supplemented with 100 μL biotin and diluted to a total volume of 150 μL in BSA buffer (40 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.2 mg mL^{-1} BSA, 50 mM NaCl, 1 mM DTT). The fluorescently labeled proteins were injected into the flowcell immediately after the conjugation at a 200 μL min^{-1} flow rate.

Mlh1-Pms1 loading on DNA is sensitive to the salt concentration in the loading buffer. Therefore, we developed a protocol to efficiently load the fluorescently-labeled protein onto DNA curtains. Mlh1-Pms1 was initially injected into the flowcell containing double-tethered DNA curtains with BSA buffer and 50 mM NaCl to assist its DNA binding. Next, the buffer was switched to imaging buffer (40 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.2 mg mL^{-1} BSA, 150 mM NaCl, 1 mM DTT, 1 mM nucleotides as indicated). After the flowcell was completely washed with the imaging buffer, flow was terminated to observe 1D diffusion on doubly-tethered DNA substrates. We note that Mlh1-Pms1 diffusion trajectories were indistinguishable between this protocol and complexes that were both loaded and imaged at 150 mM NaCl concentration (**Figure 3.3C**).

Fluorescent labeling of nucleosomes: Nucleosomes were fluorescently labelled *in situ* after Mlh1-Pms1 diffusion trajectories were recorded on the DNA substrates. An anti-HA antibody targeting (Immunology Consultants Laboratory, RHGT-45A-Z) was diluted 100-fold in BSA buffer and injected into the flowcell at 10 nM final concentration for 5 minutes. Next, 10 nM secondary antibody was injected and incubated for 7 minutes, then buffer flow was stopped to visualize nucleosomes on double-tethered DNA molecules. We have used anti-rabbit Alexa488 (Life Tech, A-11008) or anti-rabbit ATTO647N (Sigma-Aldrich, 40839-1mL) for the secondary antibody.

Particle tracking: Fluorescently-labeled proteins were tracked in ImageJ with a custom-written particle tracking script (available upon request) and the resulting trajectories further analyzed in MATLAB (R2015a, Mathworks). The positions of labeled proteins were determined by fitting every single fluorescent particle to a two-dimensional Gaussian distribution, and the series of time-dependent sub-pixel positions generated each trajectory.

Diffusion coefficients are a measure of a molecule's movement over an entire trajectory whereas nucleosome bypass is a measure of the local stepping through a nucleosome barrier. Thus, two different approaches were used to calculate these two experimental observables. Mlh1-Pms1 diffusion coefficients were determined by using the trajectories of individual moving molecules on double-tethered DNA curtains in the absence of buffer flow. The one-dimensional (1D) mean squared displacement (MSD) of each particle was determined as a function of the time interval, Δt using following equation:

$$MSD(n\Delta t) = \frac{1}{N\text{-}n}\sum_{i=1}^{N\text{-}n}(y_{i+n}\text{-}y_i)^2$$

where N is the total number of frames in the trajectory, n is the number of frames for a given time interval and ranges from 1 to N, Δt is the frame rate, and y_i is the Mlh1-Pms1 position at frame i. The MSD was calculated using the first ten time intervals (e.g. Δt = 0.05 s to 0.5 s when the frame rate was 0.05 s) and plotted as a function of Δt. Plots were fit to a line and the slope was used to calculate diffusion coefficients of individual Mlh1-Pms1 molecules. Diffusion coefficients were calculated for ≥ 30 molecules in all experiments and are reported as a mean ± standard error of the mean (S.E.M).

Measuring single nucleosome bypass frequencies: Fluorescently-labeled Mlh1-Pms1 was loaded onto double-tethered nucleosomal DNA curtains as described above. All nucleosome bypass experiments were done in imaging buffer containing 150 mM NaCl and either no nucleotide or with 1 mM ATP. We determined each collision and bypass event from individual Mlh1-Pms1 trajectories. First, a 'collision zone' was defined around each nucleosome position as described in **Figure 3.5C**. Next, the positions of diffusing Mlh1-Pms1 were plotted relative to the center of the nucleosome collision zone. The number of collisions was determined by counting the number of times that Mlh1-Pms1 entered the nucleosome collision zone. Bypass events were defined as collisions that had Mlh1-Pms1 cross from the first to the second side of the nucleosome collision zone. Non-bypass events had Mlh1-Pms1 start and end the collision on the same side relative to the nucleosome. The bypass activity measures how frequently Mlh1-Pms1 passes each nucleosome barrier. To compare the probability of bypassing single roadblock between different conditions with a statistical test, we coded each bypass event as '1' and no bypass as '0' and fit the data to a binary distribution using MATLAB.

Statistical methods: We conducted the two-sample Kolmogorov-Smirnov (K-S) test to determine whether average diffusion coefficient differ based on nucleotide types using the PAST3 software package (52). Error bars on the quantified single nucleosome bypass and percentage of moving molecules were calculated in MATLAB using bootstrap analysis with replacement (53). P-values between conditions on single nucleosome bypass experiments were determined in MATLAB using a binary regression model. The significance threshold was set at 0.05 in all tests.

Single-molecule nicking assay: 5 nM PCNA was loaded by 1.5 nM RFC on double-tethered DNA curtain in Mlh1-Pms1 endonuclease buffer (40 mM Tris-HCl pH 8.0, 0.2 mg mL^{-1} BSA, 50 mM NaCl, 2 mM $MnCl_2$, 1 mM DTT, 1 mM ATP) (54). $MgCl_2$ was used instead of $MnCl_2$ for manganese negative control. RFC was washed out by injecting endonuclease buffer with 300 mM NaCl for 2 minutes. 20 nM Mlh1-Pms1 complexes were loaded on the PCNA-containing DNA and incubated for 20 min at 30˚ C followed by washing with 1 M NaCl for 2 min. 50 nM RPA-RFP was then injected to label any gaps larger than 10 nucleotides. For a photobleaching experiment, RPA-RFP was imaged by a 532 nm laser (100 mW at the prism face) with 250 ms exposure time (**Figure 3.7D–E**). To assess RPA foci, data were collected every 5 seconds with a shutter to reduce photobleaching (**Figure 3.7F–G**).

RESULTS

The IDRs of Mlh1-Pms1 are critical for mismatch repair.

We first examined whether the IDRs of Mlh1 (~160 amino acids) and Pms1 (~290 amino acids) contain functionally important amino acids (**Figure 3.1A–B**). Our previous study established that MMR was ablated in yeast cells that lacked the Mlh1 IDR residues 348-373 and Pms1 residues 548-634 (MMR-null double-linker deletion, DLD_{null}; mlh1Δ348-373-pms1Δ584-

634). This result was surprising because deleting the same residues in the individual subunits conferred very mild MMR defects (**Table 3.2**) (11). Here, we expand on this early study by defining whether the composition and/or the lengths of the IDRs are critical for supporting MMR.

We first tested whether restoring the IDR of *pms1Δ584-634* to its full length rescued MMR (**Figure 3.1C, Figure 3.2**). *PMS1* was chosen because truncating its IDR at different positions showed only minor MMR defects and thus may be more likely to restore function with a synthetic linker (11). The substitutions included random scrambling of the 50 critical amino acids (584-634) in Pms1, as well as two biophysically characterized serine-rich regions that were equal or longer than 50 amino acids (obtained from the Herpes Virus ICP4 and *Neurospora crassa* Su9 proteins) (55, 56). All substitutions were initially examined in the wild-type (WT) *MLH1* background, where they did not restore function. The MMR defects conferred by these *pms1* mutants were similar to the *pms1Δ584-634* allele, indicating that the insertions are unlikely to disrupt the stability of the Mlh1-Pms1 complex (**Figure 3.1C, Figure 3.2, and Table 3.2**). If complex stability was compromised, these *pms1* mutants would have shown a MMR defect similar to *pms1Δ* (~8000-fold higher mutation rate compared to *PMS1* in the *lys2::insE-A*$_{14}$ reversion assay, **Table 3.2**). In the *mlh1Δ348-373* background, *PMS1* linker substitutions all conferred a nearly-null MMR phenotype that was reminiscent of the DLD_{null} MMR defect. We also performed a full-length linker swap between the IDRs in *MLH1* and *PMS1* (**Figure 3.1C, Table 3.2**); these alleles, as swaps or single substitutions, were unable to confer MMR function. Lastly, fine-scale mapping of the

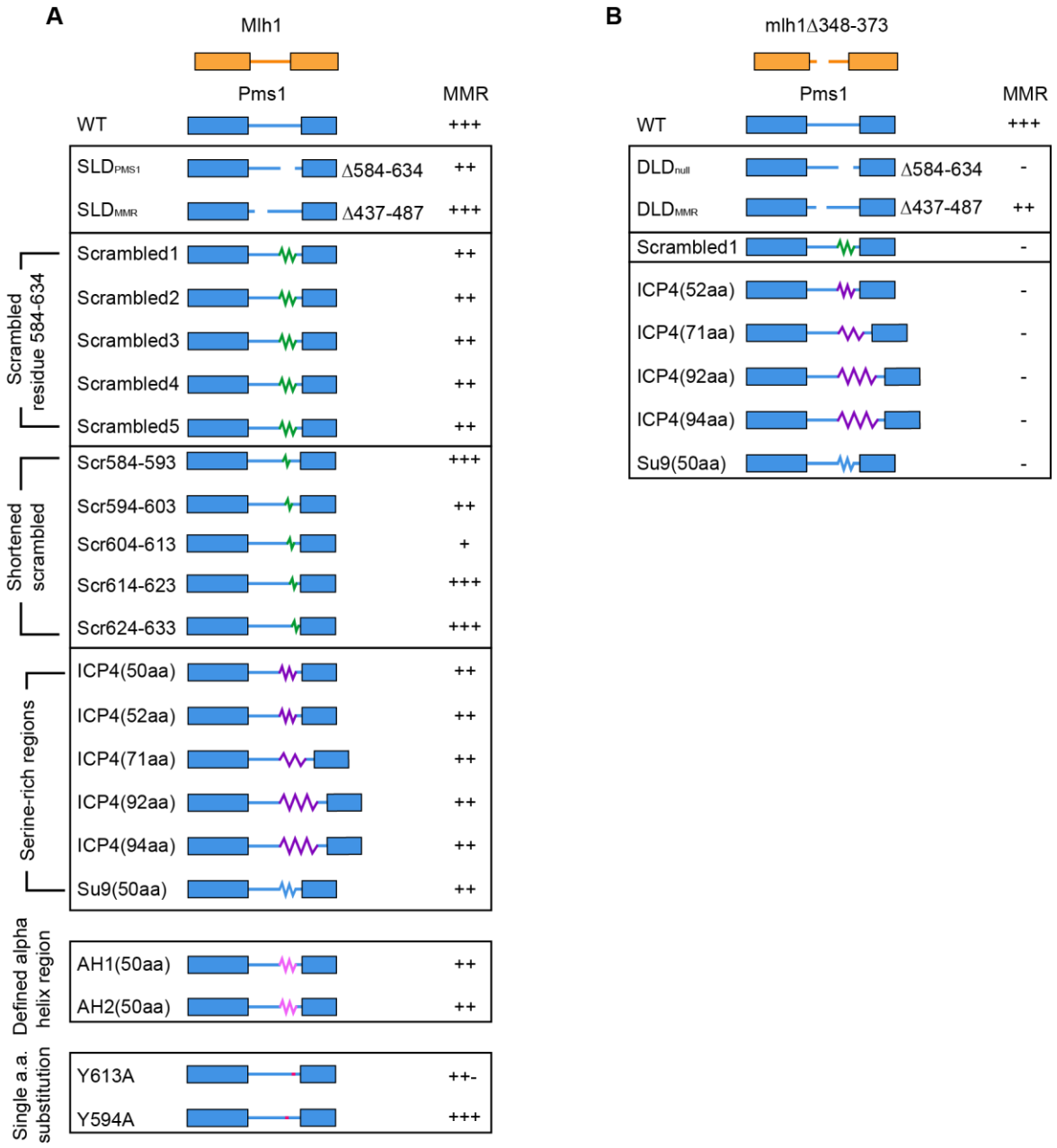


Figure 3.2. Schematic of IDR replacements (A) Schematic of specific sequences that replace the IDRs in Pms1, followed by the mutator phenotype conferred by the indicated alleles in the *MLH1* strain background. +++ indicates a wild-type mutation rate, ++, ++- and + indicate hypomorph phenotypes, and – indicates a null phenotype. See text and ***Table 3.2*** and ***Table 3.1*** for quantitative data and detailed description of the specific sequences. (B) Analysis of linker alleles presented in panel A in the *mlh1-Δ348-373* background.

Table 3.2. *pms1* linker arm insertions are unable to rescue *pms1Δ584-634* MMR defects

relevant genotype	mutation rate (x 10^{-7}) (95% CI)	*n*	relative to wild type
*pms1*Δ (EAY3097)	25,700 (15,400-36,700)	45	7,930
PMS1	3.24 (2.75-4.32)	40	1.0
$pms1\Delta_{584\text{-}634}$	308 (230-452)	24	95
$pms1_{50\text{-}scramble1}$	332 (213-414)	20	103
$pms1_{50\text{-}scramble2}$	238 (184-338)	20	74
$pms1_{50\text{-}scramble3}$	335 (303-405)	20	104
$pms1_{50\text{-}scramble4}$	413 (339-637)	20	127
$pms1_{50\text{-}scramble5}$	520 (357-615)	20	161
$pms1_{40\text{-}alpha\text{-}helix1}$	378 (214-505)	20	117
$pms1_{40\text{-}alpha\text{-}helix2}$	465 (335-645)	20	144
$pms1_{50\text{-}SRR\text{-}1}$	436 (391-575)	20	135
$pms1_{52\text{-}SRR}$	315 (210-444)	20	97
$pms1_{71\text{-}SRR}$	413 (318-592)	20	128
$pms1_{92\text{-}SRR}$	421 (314-639)	20	130
$pms1_{94\text{-}SRR}$	514 (414-745)	20	159
$pms1_{50\text{-}SRR\text{-}2}$	229 (120-340)	20	71
$pms1_{scramble584\text{-}593}$	33.4(12.1-58.3)	15	10
$pms1_{scramble594\text{-}603}$	286(157-428)	15	88
$pms1_{scramble604\text{-}613}$	1,180(870-1,360)	15	365
$pms1_{scramble614\text{-}623}$	28.6(13.8-42.6)	15	9
$pms1_{scramble624\text{-}633}$	26.3(16.7-30.8)	15	8
$pms1_{Y613A}$	89(55-128)	20	27
$pms1_{Y594A}$	22.7(8.8-32.8)	15	7
*mlh1*Δ, *pms1*Δ (EAY1365)	13,400 (9,330-15,400)	30	4,570
MLH1, PMS1	2.94 (2.2-4.1)	20	1.0
$mlh1\Delta_{348\text{-}373}$, *PMS1*	206 (118-256)	20	70
$mlh1\Delta_{348\text{-}373}$, $pms1\Delta_{584\text{-}634}$	8,420 (4,460-10,300)	20	2,860
$mlh1\Delta_{348\text{-}373}$, $pms1\Delta_{437\text{-}487}$	429 (393-857)	20	146
$mlh1\Delta_{348\text{-}373}$, $pms1_{50\text{-}scramble1}$	13,100 (9,280-17,400)	20	4,450
$mlh1\Delta_{348\text{-}373}$, $pms1_{52\text{-}SRR}$	3,000 (2,150-7,470)	20	1,020
$mlh1\Delta_{348\text{-}373}$, $pms1_{71\text{-}SRR}$	3,858 (3,260-7,150)	20	1,320
$mlh1\Delta_{348\text{-}373}$, $pms1_{92\text{-}SRR}$	5,870 (5,100-7,330)	20	2,000
$mlh1\Delta_{348\text{-}373}$, $pms1_{50\text{-}SRR\text{-}2}$	3,430 (2,890-6,180)	20	1,170
Mlh1, $pms1_{Mlh1\ linker}$	13,700 (9,580-19,700)	15	4,670
$mlh1_{Pms1\ linker}$, *Pms1*	7,680 (4,170-11,700)	15	2,610
$mlh1_{Pms1\ linker}$, $pms1_{Mlh1\ linker}$	8,330 (3,130-20,400)	15	2.830

The indicated *mlh1* and *pms1* alleles were tested in the *lys2::insE-A*$_{14}$ reversion assay, and Lys^+ reversion rates (CI, confidence interval) were calculated as described in the *Materials and Methods*. n, number of independent measurements. The *PMS1* and *pms1* alleles were expressed from the native *PMS1* promoter in pRS413 derived *ARS-CEN HIS3* plasmids. The *MLH1* and *mlh1* alleles were expressed from the native *MLH1* promoter in pRS415 derived *ARS-CEN LEU2* plasmids. These plasmids (***Table 3.1***) were transformed into EAY3097 (*pms1*Δ) and EAY1365 (*mlh1*Δ, *pms1*Δ) strains.

PMS1 584-634 region using scrambled and single amino acid substitution analyses identified a 20 amino acid region, 594-613, that plays a critical role for the function of the linker. A single substitution in this region, *pms1-Y613A*, conferred a mutator phenotype, (p-value <.00001 to WT; p-value<.00001 to *pms1Δ584-634* compared by Mann-Whitney U test) (**Figure 3.2, Table 3.2**). The *pms1-Y613A* substitution maps to a region of PMS1 that is disordered (**Figure 3.1B**) and does not encode any known PCNA-interaction motifs, as identified in yeast PMS1 (721QRLIAP), human PMS1 (723QKLIIP), and *B. subtilis* MutL (QEMIVP) (27). Consistent with this, the endonuclease activity of MLH complexes containing the *pms1Δ584-634* mutation is stimulated by PCNA, indicating that this region is not required for PCNA interactions (see below). Together these experiments establish that the specific sequence of the IDR, but not the flexibility, length or disorder is important for efficient MMR.

IDRs regulate Mlh1-Pms1 ATPase activity in the presence of DNA and PCNA. Mlh1-Pms1 is a DNA-stimulated ATPase and PCNA-activated endonuclease. Nucleolytic cleavage of the newly-synthesized DNA strand by Mlh1-Pms1 is proposed to be a critical strand discrimination signal during MMR (20, 57, 58). We sought to understand the role(s) of the IDRs in promoting the enzymatic activities of Mlh1-Pms1. We compared WT Mlh1-Pms1 to two additional mutant complexes: one mostly functional in MMR (DLD_{MMR}; mlh1Δ348-373-pms1Δ437-487), and a second defective (DLD_{null}; mlh1Δ348-373-pms1Δ584-634) (**Figure 3.1**).

All Mlh1-Pms1 variants bound similarly to a 49 bp duplex oligonucleotide in the absence of ATP (**Figure 3.1D and Figure 3.3B**). In the presence of ATP, Mlh1-Pms1 displayed reduced binding to DNA, but both DLD_{MMR} and DLD_{null} displayed DNA binding levels that were higher than WT. These results show that the two DLD complexes are impaired in ATP-dependent interactions with DNA (**Figure 3.1D**). The ATPase activities of the WT complex are stimulated

by DNA and PCNA (27, 34). However, neither DLD complex exhibited such stimulation (**Figure 3.1E**). We conclude that the IDRs facilitate interactions between Mlh1-Pms1 and DNA, and either directly or indirectly affect the DNA-dependent stimulation of ATP hydrolysis. Remarkably, both DLD_{MMR} and DLD_{null} showed similar defects in DNA binding and ATPase activities but had very different MMR phenotypes (**Table 3.2**). This puzzle encouraged us to further explore the role of the IDRs in MMR.

The IDR linker promotes facilitated 1D diffusion of Mlh1-Pms1 on DNA.

DNA-binding proteins, including Mlh1-Pms1, locate their targets using facilitated 1-dimensional (1D) diffusion along the genome (17, 18, 30, 59). Based on the biochemical results presented above, we hypothesized that the IDRs of Mlh1-Pms1 are essential for efficient 1D diffusion on chromatin. We examined Mlh1-Pms1 diffusion on double-tethered DNA curtains (**Figure 3.4A–B**). In this assay, a 48.5 kb-long DNA substrate is extended over a fluid lipid bilayer between two microfabricated chromium barriers (51, 60, 61). The lipid bilayer provides a biomimetic surface that passivates the flowcell surface from non-specific adsorption by DNA-binding proteins. A single FLAG epitope was inserted at amino acid 499 of Mlh1 for downstream fluorescent labeling. The FLAG epitope does not impact Mlh1-Pms1 activities in vitro and in vivo (17, 30). For fluorescent labeling, Mlh1 was conjugated with an anti-FLAG antibody harboring a fluorescent quantum dot (QD) (17, 30). Using this assay, we characterized WT Mlh1-Pms1, as well as DLD_{MMR} and DLD_{null} variants. All three Mlh1-Pms1 complexes readily bound DNA and >90% of the molecules rapidly diffused along the entire length of the DNA substrate (**Figure 3.4C and Figure 3.3C**; WT: 97%, N=62/64; DLD_{MMR}: 97%, N=79/81; DLD_{null}: 90%, N=60/67). Analysis of the movement showed linear mean-squared displacement (MSD) plots, verifying that all three Mlh1-Pms1 complexes freely diffuse on DNA.

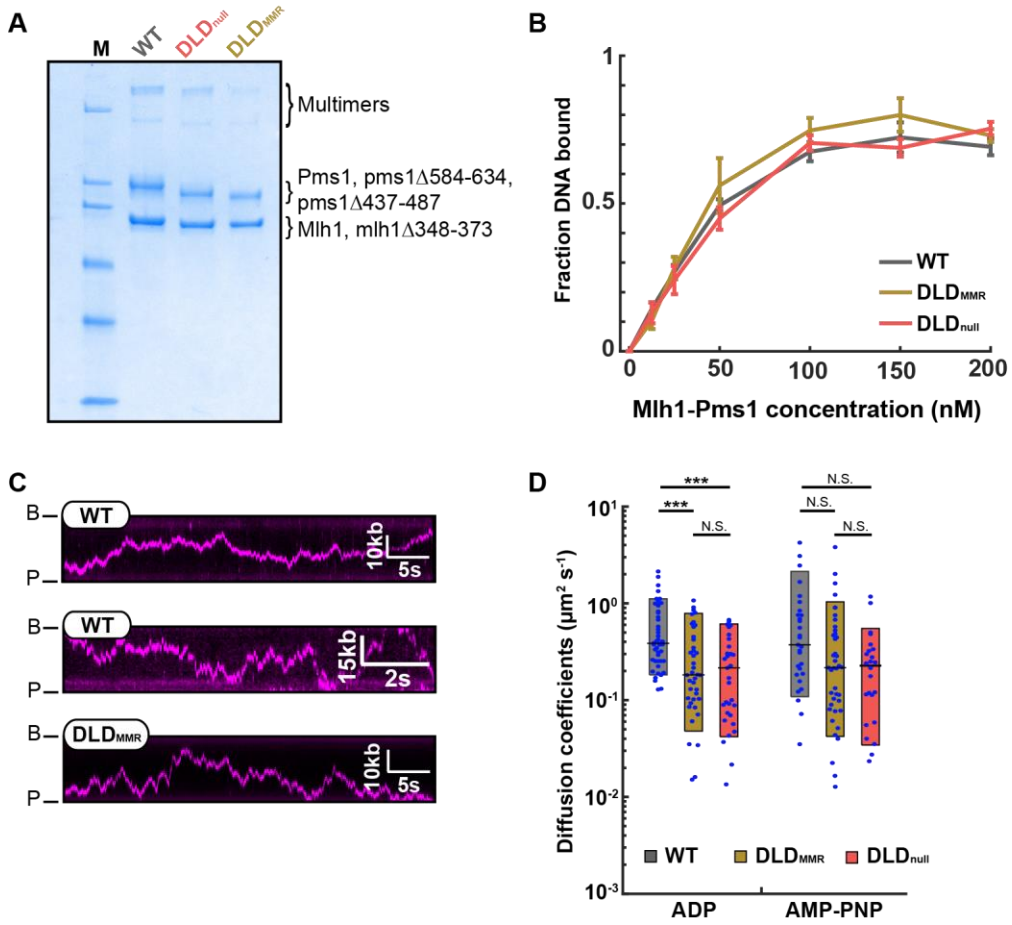


Figure 3.3. (A) SDS-PAGE analysis (8% Coomassie blue R250 stained gel) of purified wild-type (WT) Mlh1-FLAG-Pms1, DLD_{null} (mlh1Δ348-373-FLAG-pms1Δ584-634), and DLD_{MMR} (mlh1Δ348-373-FLAG-pms1Δ438-487). (B) Analysis of DNA binding in the absence of nucleotide. For each complex, DNA binding was analyzed by filter binding. Mlh1-Pms1 variants were included at final concentrations of 12.5 nM, 25 nM, 50 nM, 100 nM, 150 nM and 200 nM in buffer containing 25 mM NaCl. DNA binding was quantified by scintillation counting. Four replicates were averaged; error bars indicate the SD. (C) Representative kymographs of WT and DLD_{MMR} complexes loaded at 50 mM NaCl and imaged at 150 mM NaCl (top and bottom), and WT Mlh1-Pms1 loaded and imaged at 150 mM NaCl (middle). These images show that the proteins are freely diffusing on DNA. Complexes that were loaded at low or high NaCl concentration were indistinguishable in the single-molecule assays. B and P indicate barrier and pedestals, respectively. (D) Diffusion coefficients of the four Mlh1-Pms1 complexes with ADP or AMP-PNP. The black bar in the box plot represents the median of the distribution. * *P*-values <0.05, ** *P*-value < 0.01, and *** *P* value < 0.005. N.S. indicates p > 0.05.

ATP binding to Mlh1-Pms1 results in dimerization of the N-terminal domains, compaction of the IDRs, and the formation of a ring-like sliding clamp on DNA (32, 34, 36, 62). To probe the functional significance of this conformational change, we measured the diffusion coefficients of the Mlh1-Pms1 variants as a function of ATP. Diffusion coefficients in the ATP-bound state were significantly increased compared to the apo (no nucleotide) condition for all complexes. These results are consistent with a prior single-molecule report of ATP-dependent diffusion of *E. coli* MutL homodimer (59). However, compared to WT and DLD_{MMR}, the mean DLD_{null} diffusion coefficient is ~six-fold lower on DNA in the presence and absence of ATP (**Figure 3.4D and Table 3.3**). While DLD_{null} displayed the lowest diffusion coefficient of all the complexes in the absence or presence of ATP (**Figure 3.4D**), DLD_{null} and DLD_{MMR} displayed similar diffusion coefficients in the presence of ADP or AMP-PNP (**Figure 3.3D**; See Discussion). We conclude that the IDRs of Mlh1-Pms1 are critical for efficient facilitated diffusion on DNA.

IDRs promote facilitated diffusion on nucleosome-coated DNA.

Mlh1-Pms1 must efficiently traverse chromatin to locate mismatch-bound MSH complexes. To investigate how the IDRs regulate movement on chromatin, we imaged Mlh1-Pms1 on nucleosome-coated DNA substrates. Nucleosomes were assembled using salt gradient dialysis with increasing concentrations of histone octamers to DNA molecules to recapitulate both sparse and dense nucleosome arrays (17, 63). Single nucleosomes were visualized via a fluorescent antibody directed against a triple HA epitope on the N-terminus of H2A. Nucleosomes were distributed over the entire length of the DNA molecule, with a weak preference for GC-rich segments, as described previously (**Figure 3.5A**) (64).

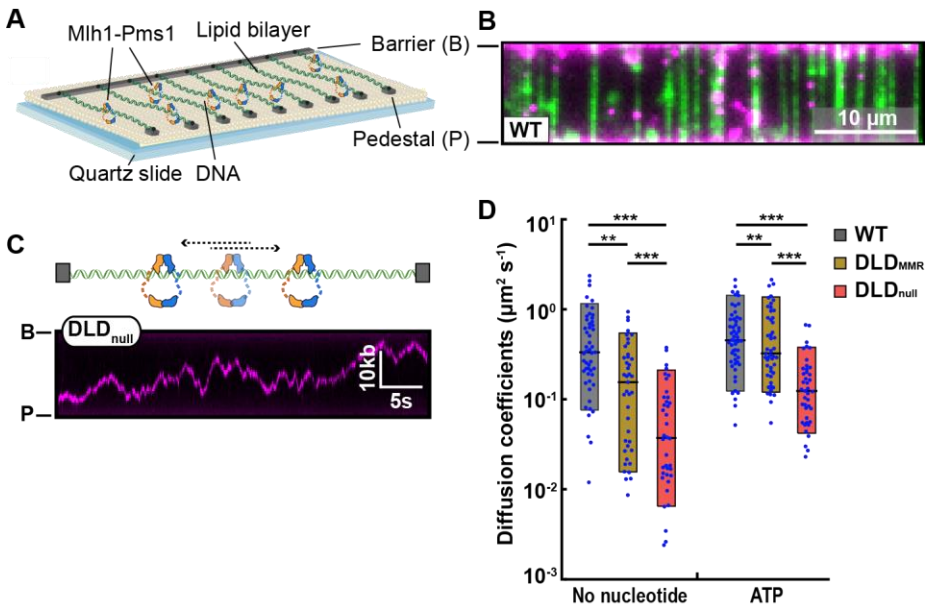


Figure 3.4. **The IDRs promote facilitated Mlh1-Pms1 diffusion on DNA**. (A) Schematic of the DNA curtains assay. Fluorescently-labeled Mlh1-Pms1 is injected into the flowcell and visualized on double-tethered DNA substrates in the absence of buffer flow. (B) An image of Mlh1-Pms1 (magenta puncta) on double-tethered DNA molecules (green). To avoid interference from the DNA-intercalating dye, the DNA is not fluorescently stained during analysis of Mlh1-Pms1 movement on DNA. (C) A schematic (top) and representative kymograph of a DLD_{null} Mlh1-Pms1 diffusing on DNA (bottom). (D) Diffusion coefficients of Mlh1-Pms1 complexes in the absence and presence of ATP. Boxplots indicate the median, 10th, and 90th percentiles of the distribution. P-values are obtained from K-S test: * *P*-values <0.05, ** *P*-value < 0.01, and *** *P*-value < 0.005.

Table 3.3. Nucleotide-dependent Mlh1-Pms1 diffusion coefficients

Nucleotide type	Protein type	1D diffusion ± S.E.M. ($\mu m^2 s^{-1}$)	Number of molecules	p-value (relative to WT)	
				t-test	K-S test
None	WT	0.507 ± 0.07	56	N/A	
	DLD_{MMR}	0.223 ± 0.04	41	0.1	0.01
	DLD_{null}	0.078 ± 0.01	39	$1.5x10^{-6}$	$6.7x10^{-9}$
ADP	WT	0.584 ± 0.07	46	N/A	
	DLD_{MMR}	0.303 ± 0.04	40	0.001	0.0007
	DLD_{null}	0.376 ± 0.14	34	0.15	0.001
AMP-PNP	WT	0.753 ± 0.18	30	N/A	
	DLD_{MMR}	0.451 ± 0.11	41	0.13	0.99
	DLD_{null}	0.275 ± 0.05	26	0.02	0.07
ATP	WT	0.918 ± 0.10	59	N/A	
	DLD_{MMR}	0.646 ± 0.10	50	0.069	0.007
	DLD_{null}	0.171 ± 0.02	42	$8.2x10^{-8}$	$9.3x10^{-13}$

All data points were acquired in imaging buffer containing 150 mM NaCl.

We first determined whether the Mlh1-Pms1 IDRs regulate diffusion past a single nucleosome. DNA substrates with one to seven nucleosomes were assembled into double-tethered DNA curtains (**Figure 3.5B**). Mlh1-Pms1 was added to the flowcell prior to fluorescently labeling the nucleosomes. Keeping the nucleosomes unlabeled guaranteed that Mlh1-Pms1 was not blocked by the H2A-targeting antibody. After recording 10-15 minutes of Mlh1-Pms1 diffusion, a fluorescently labeled anti-HA antibody visualized the nucleosome positions. Diffusing Mlh1-Pms1 complexes encountered and occasionally bypassed individual nucleosomes (**Figure 3.6A**). To quantitatively determine the probability of bypassing a single nucleosome, we defined a 'collision zone' for each nucleosome which encompasses three standard deviations of the spatial resolution of our single-molecule assay (0.08 μm; ~ 300 bp) (**Figure 3.5C** and *Materials and Methods*). Diffusing Mlh1-Pms1 that entered this collision zone from one side of the nucleosome and emerged from the other side was counted as a bypass event. Events where Mlh1-Pms1 entered and emerged from the same side of the nucleosome collision zone were scored as non-bypass encounters. This quantification likely underestimates the frequency of microscopic Mlh1-Pms1 nucleosome bypass events that are below our spatial resolution but does not change any of the underlying conclusions comparing the different complexes.

WT Mlh1-Pms1 bypassed nucleosomes 30 ± 0.3% of the time (**Table 3.4**). A molecule that travels via a 1D random walk involving facilitated diffusion has a 50% probability of stepping forward or backward on DNA. This 50% probability value is the maximum theoretical bypass probability in the absence of any nucleosome obstacles. Thus, Mlh1-Pms1 is capable of efficiently bypassing a nucleosome obstacle. In contrast, both DLD_{MMR}

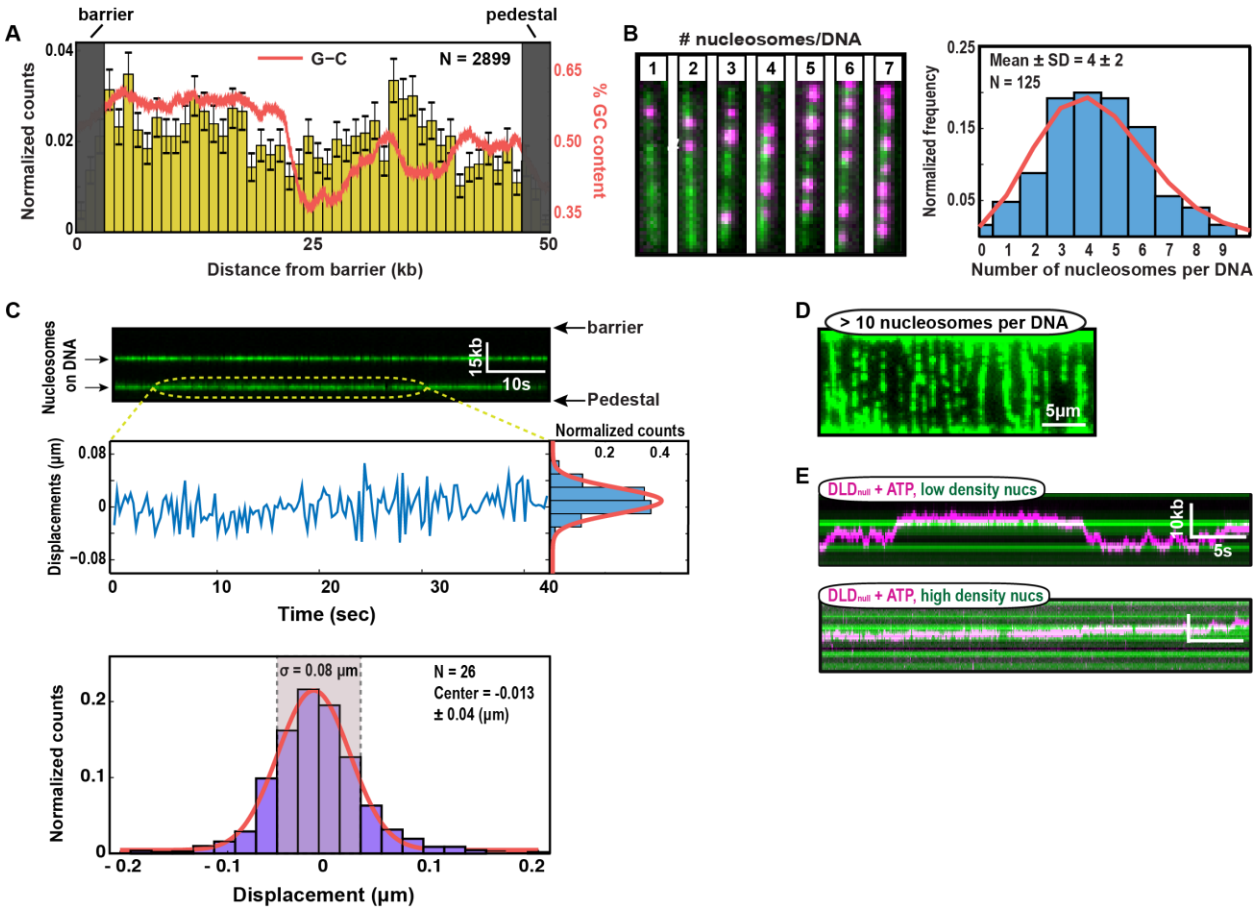


Figure 3.5. Analysis of nucleosome position in DNA curtain assays (A) Distribution of human nucleosomes on DNA curtains indicates a weak preference for GC-rich sequences. Error bars were generated by bootstrap analysis. The % GC content for a 2 kbp sliding window is shown in the red line. (B) Representative images of various numbers of nucleosomes per DNA (left), and a histogram of frequency of nucleosome deposition on DNA fitted to a Poisson distribution (red line) with the mean of the data (right). (C) Definition of 'nucleosome zone' for bypass analysis. To determine the spatial resolution, a distribution of the net displacement of single nucleosomes was fit to a Gaussian distribution (red line). For analyzing single nucleosome bypass frequencies, the nucleosome zone was defined as a three-sigma region surrounding the mean nucleosome position (bottom). (D) A fluorescent image of double-tethered DNA curtain with > 10 nucleosomes per DNA. Nucleosomes were labeled with anti-HA antibody conjugated QDs (green). The position of each nucleosome cannot be determined due to overlapping fluorescent nucleosome signals. (E) Representative kymograph of DLD_{null} on different nucleosome density substrate in the presence of ATP.

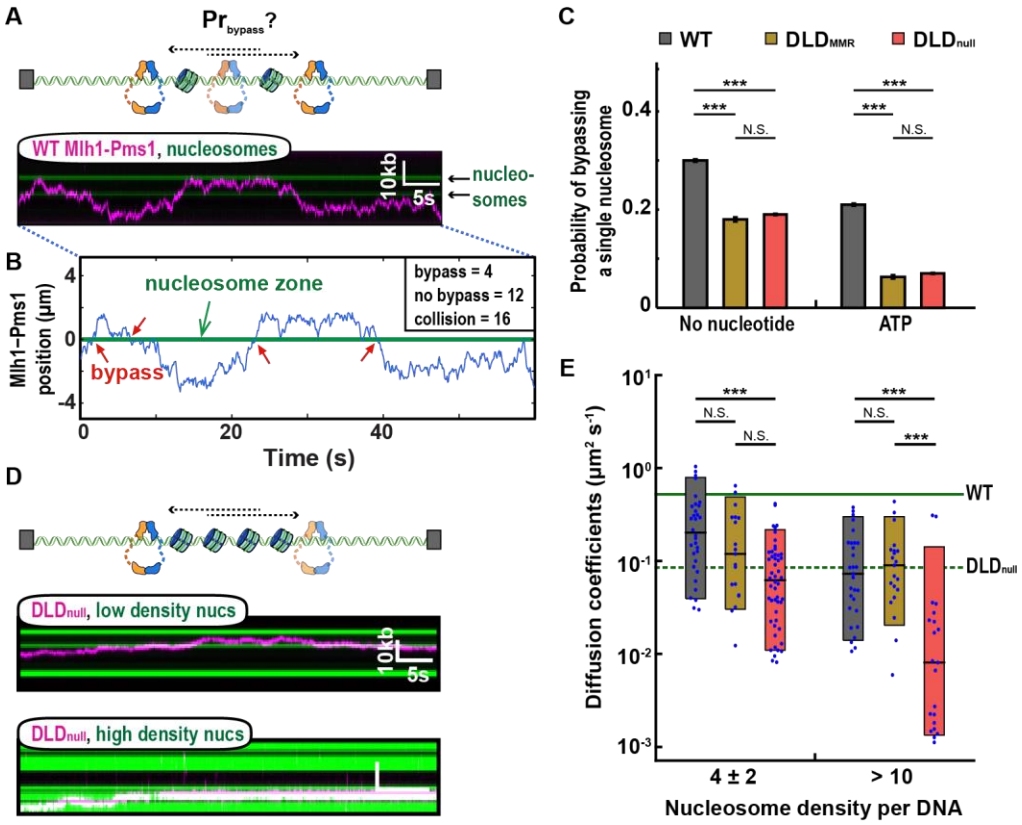


Figure 3.6 The IDRs increase Mlh1-Pms1 movement on nucleosome-coated DNA. (A) An illustration (top) and a representative kymograph of WT Mlh1-Pms1 diffusing past a nucleosome (bottom). (B) Trajectory analysis of single nucleosome bypass events. The nucleosome collision zone (green) is defined as three standard deviations of the experimental resolution of the nucleosome position (see *Materials and Methods*). (C) The values obtained from the analysis shown in (B) are fit to binary logistic regression to obtain predicted probability of nucleosome bypass. (D) A cartoon (top) and representative kymographs of DLD_{null} (magenta) on DNA containing 4 ± 2 nucleosomes (middle, green) or > 10 nucleosomes (bottom, green) in the absence of ATP. Nucleosomes are labeled with fluorescent anti-HA antibodies after Mlh1-Pms1 trajectories are recorded. (E) Diffusion coefficients of the three Mlh1-Pms1 complexes on nucleosome-coated DNA. The solid and dashed green lines indicate the mean of the diffusion coefficients of WT and DLD_{null} on naked DNA, respectively. P-values are obtained from K-S test: * *P*-values <0.05, ** *P* value < 0.01, and *** *P*-value < 0.005. N.S. indicates p > 0.05.

Table 3.4. Probability of single nucleosome bypass by Mlh1-Pms1 complexes

Nucleotide type	Protein-Condition	Number of trajectories	Number of collisions	Probability ± std. dev.	P-value (relative to WT)	P-value (relative to DLD_{MMR})	Average number of collisions per trajectory ± std. dev.
Minus	WT	31	1361	0.30 ± 0.003	N/A		44 ± 44
	DLD_{MMR}	29	1166	0.18 ± 0.005	$2.8x10^{-11}$		40 ± 29
	DLD_{null}	27	1033	0.19 ± 0.002	$6.1x10^{-9}$	0.54	38 ± 24
ATP	WT	34	1223	0.21 ± 0.007	N/A		35 ± 35
	DLD_{MMR}	30	1572	0.063 ± 0.004	$4.8x10^{-28}$		56 ± 60
	DLD_{null}	30	1111	0.070 ± 0.002	$1.9x10^{-19}$	0.08	37 ± 30

P-values are determined from fitting binary logistic regression relative to WT with same nucleotide condition.

and DLD_{null} complexes had a 2-fold reduced nucleosome bypass frequency (18 ± 0.5%; N=29 for DLD_{MMR}; 19 ± 0.2%; N=27 for DLD_{null}).

Next, we explored how ATP-induced conformational changes affect nucleosome bypass by Mlh1-Pms1 (**Figure 3.5C**). In the presence of ATP, all Mlh1-Pms1 variants exhibited a reduced bypass probability, with a significantly larger, ~2 to 3-fold, decrease in nucleosome bypass probability for both DLD variants. The decrease in bypass probabilities for DLD_{MMR} and DLD_{null} mirrors their ATPase activities (**Figure 3.1E**). Taken together, these data suggest that ATP-dependent dimerization of the N-terminal domains accompanied by conformational compaction of the IDRs reduces dynamic movement on nucleosome-coated DNA.

We reasoned that the combination of a reduced diffusion coefficient and less efficient nucleosome bypass observed with DLD_{null} may compromise its ability to navigate on dense nucleosome arrays. To test this, we increased the histone octamer to DNA ratio during salt dialysis to deposit >10 nucleosomes per DNA substrate (**Figure 3.5D**). At this high density, each nucleosome is optically indistinguishable due to the diffraction limit of light. Nonetheless, by using two-color fluorescent imaging we can still track individual diffusing Mlh1-Pms1 complexes on this nucleosome-coated DNA substrate (**Figure 3.6D and Figure 3.5E**). The 1D diffusion of all Mlh1-Pms1 complexes was restricted on this high nucleosome density substrate compared to naked DNA. Notably, while 1D diffusion coefficients of WT and DLD_{MMR} decreased by 3-fold compared to naked DNA, the DLD_{null} diffusion coefficient decreased 12-fold on this chromatinized DNA substrate (**Figure 3.6E**). Thus, the IDRs are important for promoting rapid facilitated diffusion on naked DNA but are especially critical for navigating on chromatin.

The IDRs are required for multiple rounds of endonucleolytic cleavage.

After MSH recognition, Mlh1-Pms1 nicks the mismatch-containing DNA strand for efficient MMR (24, 65). Motivated by the importance of the IDRs in promoting diffusion on both naked and nucleosome-coated DNA, we tested how these domains regulate Mlh1-Pms1 endonuclease activity. We first assayed the ability of Mlh1-Pms1 variants to nick supercoiled DNA in a well-established mismatch- and MSH-independent endonuclease reaction (20, 27, 35, 57). This assay requires the ATP-dependent clamp loader RFC to load PCNA on the closed circle DNA substrate (**Figure 3.7A**) (66, 67). The endonuclease activity of WT Mlh1-Pms1 was indistinguishable from the DLD_{null} and DLD_{MMR} variants (**Figure 3.7A and Figure 3.8A–B**). However, this assay cannot distinguish between singly- and multiply-nicked DNA substrates. This assay also cannot report the ATP dependence of the Mlh1-Pms1 endonuclease activity because ATP is required for RFC-dependent PCNA loading. To resolve these limitations, we established the alkaline gel-based and single-molecule endonucleolytic assays described below. We directly tested the role(s) of ATP in Mlh1-Pms1 endonuclease activation on linear DNA substrates analyzed by denaturing gel electrophoresis. PCNA can thread onto the ends of linear DNA, abrogating the need for RFC and ATP (**Figure 3.8C–D**) (27, 57, 68). We observed that Mlh1-Pms1 endonuclease requires PCNA and is further enhanced by ATP binding (**Figure 3.8C–D**). ATP hydrolysis was not required because ATPγS could support the reaction to the same extent or better than ATP, as suggested for the *E. coli* and *Bacillus* MutL (34, 35, 62). Although the DLD_{MMR} variant hydrolyzed linear DNA to approximately the same extent as wild-type Mlh1-Pms1, DNA degradation was attenuated with the DLD_{null} variant (**Figure 3.7B and Figure 3.8D**). This was seen in the presence of ATP, but less so in the presence of ATPγS (**Figure 3.8C**, *see Discussion*). In the assay in **Figure 3.7B**, extensive nicking on each DNA

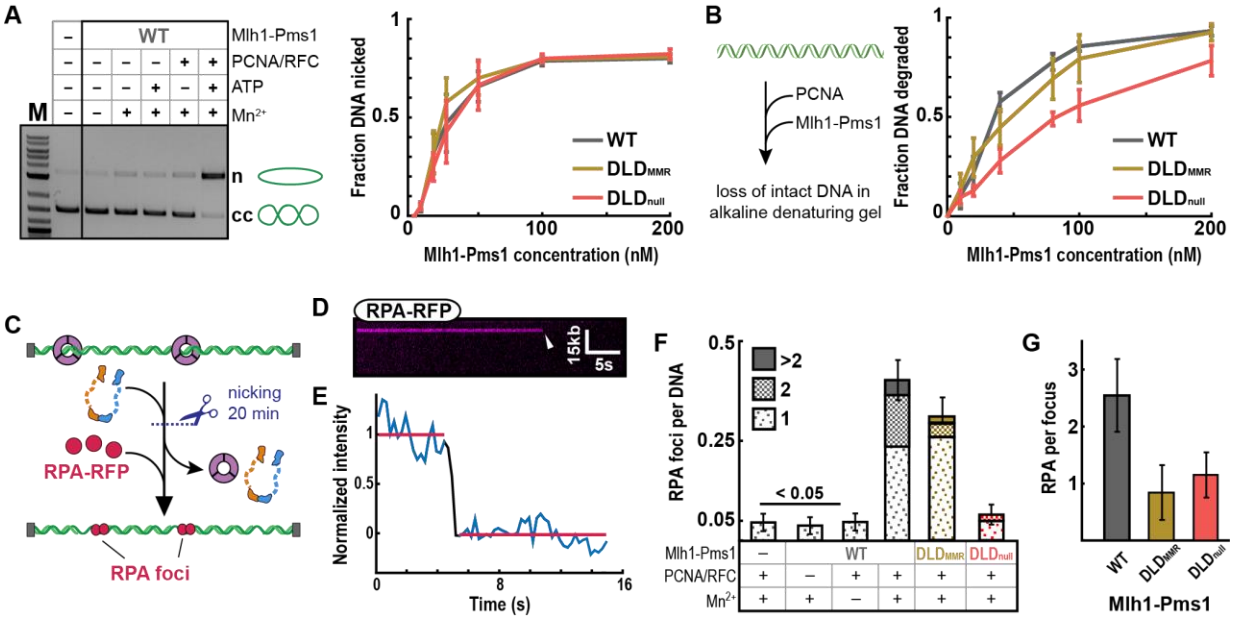


Figure 3.7. **The IDRs regulates extensive DNA nicking.** (A) Endonuclease activity on closed circular DNA in the presence (+) or absence (-) of $MnSO_4$, ATP, and yeast PCNA/RFC (left panel). Where + is indicated, the concentration of $MnSO_4$ was 2.5 mM, ATP was 0.5 mM, RFC and PCNA were each 500 nM. The final concentration of WT Mlh1-Pms1 was 100 nM. In the presence of $MnSO_4$, ATP, RFC, and PCNA at the above concentrations, Mlh1-Pms1 variants were titrated from 0-200 nM (right panel). Error-bars: SD of three replicates. (B) Illustration (left panel) and quantification (right panel) of endonuclease activity of wild-type and mutant Mlh1-Pms1 complexes (titrated from 0-200 nM) on linear DNA (also see **Figure 3.8C–D** for controls). All reactions contain 500 nM PCNA, 0.5 mM ATP, and 5 mM $MnSO_4$. Error-bars: SD of four replicates. (C) Schematic of the single-molecule endonuclease assay. Formation of ssDNA gaps via PCNA-activated Mlh1-Pms1 nuclease activity was visualized by injecting RPA-RFP into the flowcell. (D) Kymograph and (E) fluorescent intensity profile of an RPA-RFP punctum with a single-step photobleaching event (arrow), indicating a single RPA-RFP molecule on the ssDNA. (F) The number of RPA foci per DNA molecule for the indicated Mlh1-Pms1 variants. (G) The number of RPA molecule per punctum for the three Mlh1-Pms1 complexes. To estimate the number of RPA molecules per ssDNA segment, the fluorescent intensity for each punctum was measured and normalized to that of a single RPA-RFP (see *Methods* for details).

molecule accounted for the observed substrate loss, and the reduced nicking by the DLD_{null} complex suggested another *in vivo* MMR defect.

Next, we developed a single-molecule assay to probe the limited nicking that likely occurs for DLD_{null} *in vivo*. This reaction was carried out in two steps. First, PCNA was loaded by RFC on double-tethered DNA curtains in the presence of ATP, as described previously (54). After flushing out RFC, Mlh1-Pms1 was incubated in the flowcell for 20 minutes (**Figure 3.7D**). PCNA and Mlh1-Pms1 were washed out by 1 M NaCl followed by injecting 50 nM RPA-RFP to visualize the ssDNA gaps made by multiple rounds of Mlh1-Pms1 endonuclease activity (**Figure 3.7E**). These ssDNA gaps arise from loss of short oligos formed by multiple nicks that are deposited in close proximity by multiple Mlh1-Pms1 molecules. Closely-spaced nicks allow fraying of ssDNAs that are subsequently bound and displaced by RPA (69). Note that we would not be able to detect RPA foci if the nicks on the same strand created by Mlh1-Pms1 were far apart. We quantified the number of RPA foci per DNA and the number of RPA per focus via single-molecule photobleaching. RPA preferentially binds ~30 nt of ssDNA, but individual RPA molecules can bind ssDNA as short as 10 nucleotides (70, 71). Thus, we estimate that puncta with one RPA contain approximately 10-30 nt of ssDNA, whereas puncta with three or more RPA expose > 60 nt of ssDNA. Interestingly, DLD_{null} generated 6-fold fewer RPA foci (0.07 ± 0.02 RPA/DNA; N=307) than WT Mlh1-Pms1 (0.40 ± 0.02 RPA/DNA; N=382). In contrast, DLD_{MMR} was only mildly compromised (0.28 ± 0.02 RPA foci/DNA; N=420) compared to WT complex (**Figure 3.7F**). We also estimated the length of the exposed ssDNA by counting the number of RPA molecules bound on DNA. The number of RPA per focus was comparable for DLD_{null} (1.1 ± 0.56 RPA; N=20) and DLD_{MMR} (0.9 ± 0.58 RPA; N=68) but was substantially

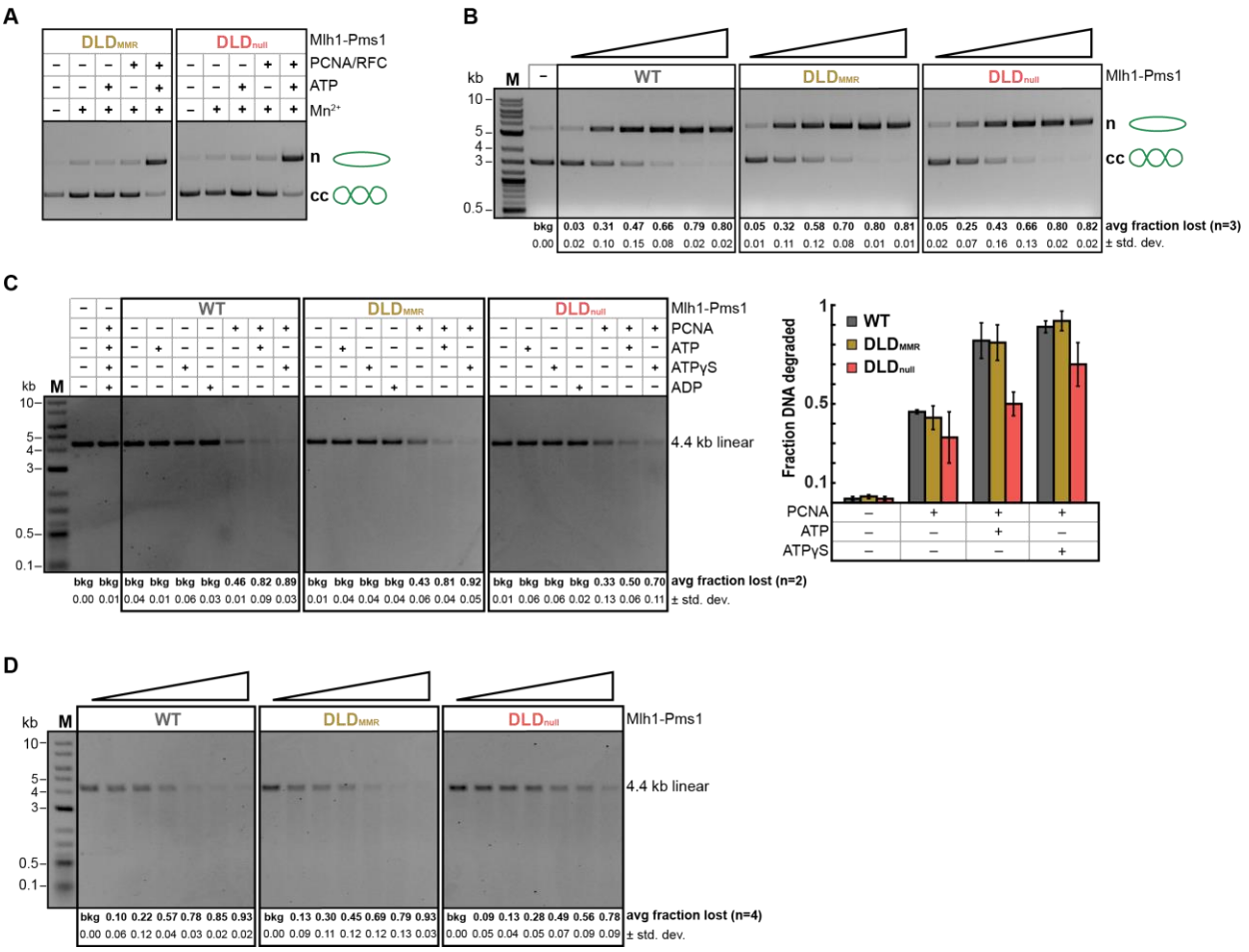


Figure 3.8. Endonuclease activity of mutant protein complexes on closed circle and linear DNA (A) Endonuclease activity on a closed circular DNA in the presence (+) or absence (-) of $MnSO_4$, ATP, and yeast RFC/PCNA. Where + is indicated, the concentration of $MnSO_4$ was 2.5 mM, ATP was 0.5 mM, RFC and PCNA were each 500 nM. The final concentration of Mlh1-Pms1 variants was 100 nM. All variants were analyzed side-by-side with WT as a control (***Figure 3.7A***). (B) Gel for data in ***Figure 3.7A*** right panel. Mlh1-Pms1 variants were titrated from 0-200 nM in the presence of 2.5 mM $MnSO_4$, 0.5 mM, ATP, 500 nM RFC, and 500 nM PCNA. All variants were analyzed side-by-side with WT as a control. Final image is arranged for simplicity. Quantification from three separate experiments is shown below the gel. (C) Left panel, where indicated, the final concentration of nucleotide is 0.5 mM. The amount of substrate lost was quantified and expressed as a fraction by comparing to the average of lanes 2-3. The average fraction lost was calculated from duplicate experiments. The standard deviation between two independent experiments is expressed below the gel. All variants were analyzed side-by-side with WT as a control. Final image is arranged for simplicity. Right panel, analysis of endonuclease activity of wild-type and mutant Mlh1-Pms1 complexes (100 nM) on linear DNA. Reactions contain 500 nM PCNA and 0.5 mM ATP or ATPγS as indicated. 5 mM $MnSO_4$ is included in all reactions. (D) Gel for data reported in ***Figure 3.7B***. The amount of substrate lost was quantified and expressed as a fraction by comparing to lane 2, 9, or 16 for each variant. The average fraction lost and SD were calculated from four replicates (SD expressed below the gel). All variants were analyzed in a single gel. Final image is arranged for simplicity.

lower than WT Mlh1-Pms1 (2.6 ± 1.2 RPA; N=79) (**Figure 3.7G**). These data indicate that IDRs are crucial for multiple rounds of DNA nicking during strand excision.

DISCUSSION

All MLH proteins—from the *E. coli* MutL to the human Mlh1-Pms1—contain IDRs that link the structured N- and C-terminal domains. The importance of these IDRs have been recognized in both bacterial and eukaryotic MMR, but the functions of this domain have remained elusive (11, 16). Here, we show that shortening, scrambling, lengthening, or swapping the IDRs caused mild to severe MMR defects, and even a single amino acid substitution in the IDR of Pms1, Y613A, caused an MMR defect *in vivo* (**Figure 3.1 and Figure 3.2**). We therefore used three representative Mlh1-Pms1 complexes (WT, DLD_{MMR}, DLD_{null}) to further probe the mechanistic implications of altered IDRs.

The Mlh1-Pms1 IDRs undergo conformational changes throughout the ATP hydrolysis cycle (32–34, 62). Upon ATP binding, Mlh1-Pms1 adopts a ring-like, scrunched conformation (32). ATP hydrolysis reverts the complex back to the extended open state where it is likely to dissociate from DNA (32, 34, 36, 62). Here, we show that the ATPase activity is disrupted when the IDRs are shortened (**Figure 3.1E**), indicating that disrupting this conformational cycle feeds back on the ATPase activity encoded in the structured N-terminus of both subunits. These data motivated us to assay the roles of the IDRs in both facilitated diffusion and nucleolytically processing of the DNA.

Mispair recognition by an MSH complex catalytically loads MLH proteins onto DNA. Evidence for Mlh1-Pms1 loading includes an accumulation of Pms1 foci under conditions requiring Msh2-Msh6 and mispaired bases and the identification of *msh6* dominant mutations

that prevent Mlh1-Pms1 recruitment *in vitro* and Pms1 foci formation *in vivo* (31, 72). Therefore, Mlh1-Pms1 complexes must scan the genome for mismatch-bound MSH as nucleosomes are being assembled onto the newly synthesized DNA. Strikingly, the DLD_{null} complex is significantly impaired in 1D diffusion on naked DNA and this defect is further exacerbated on dense nucleosome arrays, where the diffusion coefficient of DLD_{null} is decreased by 12-fold compared to that of WT Mlh1-Pms1 (**Figure 3.4D, 3E**). The different activities of DLD_{null} and DLD_{MMR} suggest that the residues spanning the 584-634 aa region in Pms1 are especially critical for MMR. These residues likely contribute to the conformational rearrangement of the entire complex. A second possibility is that the IDR reorganizes how DNA is channeled through the Mlh1-Pms1 complex. Further structural and biophysical studies will be required to probe the conformational transitions of these IDR variants on DNA. Taken together, our data establish that the IDRs regulate facilitated diffusion of Mlh1-Pms1 on both naked and nucleosome-coated DNA substrates.

Recent studies suggest an alternative EXO1-independent MMR pathway that requires iterative nicking involving multiple Mlh1-Pms1 molecules that are activated via interactions with MSH complexes and PCNA. When Exo1 is absent, multiple nicks may promote strand removal via displacement and/or exonucleolytic activities of Polymerase δ (21, 22, 24, 26, 65). The IDRs may control this activity by ATP-dependent conformational rearrangements that bring the DNA strand close to the nuclease active site. Indeed, ATP-dependent structural rearrangement stimulates the nuclease activity in the bacterial MutL system (35, 62). Consistent with this idea, DLD_{null} was defective in carrying out multiple rounds of DNA cleavage, as seen in both ensemble and single-molecule nuclease assays (**Figure 3.7**).

Figure 3.9 summarizes a working model for how MLH IDRs promote mismatch repair. Mlh1-Pms1 rapidly diffuses on nucleosome-coated DNA in search of lesion-bound MSH complexes. The IDRs play a critical role in promoting facilitated diffusion on chromatin to accelerate the search for MSH-bound lesions. Mlh1-Pms1 is activated by PCNA to nick DNA proximal to an MSH-bound mismatch. This activity may be further regulated by conformational changes in the IDRs that are coupled to ATP hydrolysis. The degree of Mlh1-Pms1 nicking *in vivo* may depend on the concentration of complexes in the vicinity of the mismatch, as well as the availability of the Exo1 nuclease. When Exo1 is unavailable, extensive Mlh1-Pms1-induced nicking provides an alternative strand excision pathway. The loss of MMR observed for the DLD_{null} mutant stems from the combination of defects in ATPase, facilitated diffusion on chromatin, and endonuclease activities. A subset of these phenotypes explains the partial MMR defects of the other IDR variants that we assayed genetically (**Figure 3.1C and Figure 3.2**). Additional studies with the fully-reconstituted mismatch-provoked repair system will provide additional insights into how nicking by Mlh1-Pms1 is regulated at the repair site. More broadly, our results highlight that conformational changes in intrinsically disordered linkers can profoundly alter DNA interactions and enzymatic activities of neighboring structured domains. This work adds additional details to the emerging disorder-function paradigm emerging from biophysical studies of intrinsically disordered proteins.

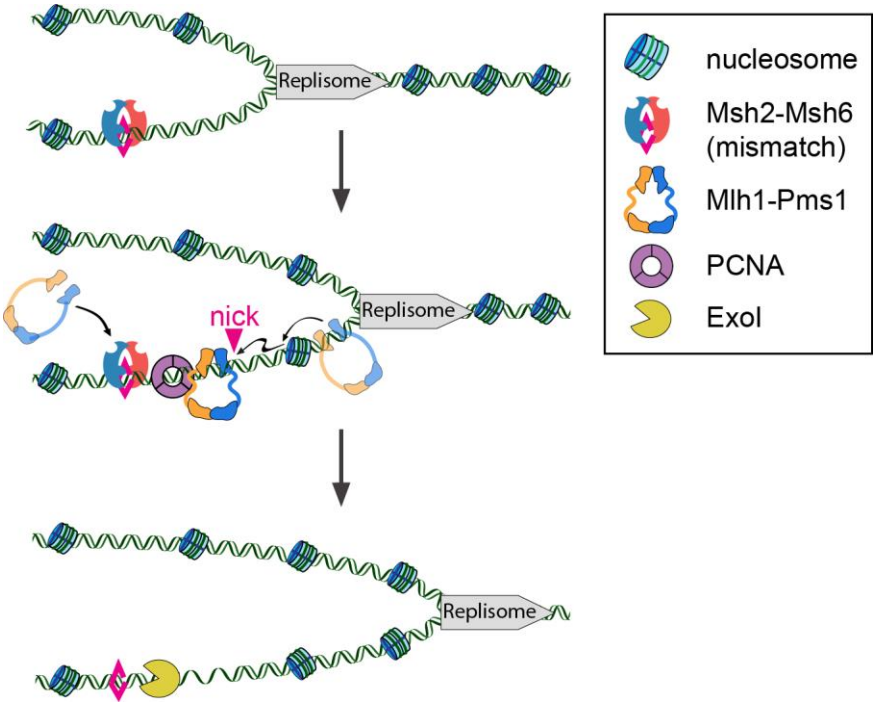


Figure 3.9. **A model of replication coupled mismatch repair.** Msh2-Msh6 locates mismatches in DNA that is nucleosome-free during replication. Mlh1-Pms1 sliding clamps efficiently bypass nucleosomes that are deposited on the newly-replicated DNA. The ability of Mlh1-Pms1 to traverse nucleosomes is important to locate mismatch-bound Msh2-Msh6 but not during endonucleolytic DNA cleavage. Exo1 binds one or more Mlh1-Pms1-generated nicks and degrades the nascent daughter DNA strand.

ACKNOWLEDGEMENT

We thank Dr. Alba Guarne for sharing information on IDRs and the MLH nicking assay, Dr. Titia Sixma for sharing the MutL over-expression vector, Dr. Andreas Matouschek for sharing plasmids encoding the flexible protein linkers used in the genetic assays, Jeff Schaub for RPA-RFP, Dr. Eric Greene, and members of the Finkelstein and Alani labs for helpful discussions and comments on the manuscript. Funding:Howard Hughes Medical Institute (to Y.K.); Cancer Prevention & Research Institute of Texas (to I.J.F.); National Institutes of Health [GM112435 to C.M.M.]; National Institute of General Medical Sciences [GM120554 to I.J.F. and GM53085 to E.A.]; The Welch Foundation [F-1808 to I.J.F.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

REFERENCES

1. Oldfield,C.J. and Dunker,A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.*, **83**, 553–584.

2. van der Lee,R., Buljan,M., Lang,B., Weatheritt,R.J., Daughdrill,G.W., Dunker,A.K., Fuxreiter,M., Gough,J., Gsponer,J., Jones,D.T., *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, **114**, 6589–6631.

3. Andres,S.N. and Williams,R.S. (2017) CtIP/Ctp1/Sae2, molecular form fit for function. *DNA Repair*, **56**, 109–117.

4. Warren,C. and Shechter,D. (2017) Fly Fishing for Histones: Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. *J. Mol. Biol.*, **429**, 2401–2426.

5. van Leeuwen,H.C., Strating,M.J., Rensen,M., de Laat,W. and van der Vliet,P.C. (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding. *EMBO J.*, **16**, 2043–2053.

6. Kozlov,A.G., Weiland,E., Mittal,A., Waldman,V., Antony,E., Fazio,N., Pappu,R.V. and Lohman,T.M. (2015) Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA. *J. Mol. Biol.*, **427**, 763–774.

7. Liu,J., Perumal,N.B., Oldfield,C.J., Su,E.W., Uversky,V.N. and Dunker,A.K. (2006) Intrinsic disorder in transcription factors. *Biochemistry*, **45**, 6873–6888.

8. Currie,S.L., Lau,D.K.W., Doane,J.J., Whitby,F.G., Okon,M., McIntosh,L.P. and Graves,B.J. (2017) Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Res.*, **45**, 2223–2241.

9. Kozlov,A.G., Shinn,M.K., Weiland,E.A. and Lohman,T.M. (2017) Glutamate promotes SSB protein-protein Interactions via intrinsically disordered regions. *J. Mol. Biol.*, **429**, 2790–2801.

10. Vuzman,D. and Levy,Y. (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. Biosyst.*, **8**, 47–57.

11. Plys,A.J., Rogacheva,M.V., Greene,E.C. and Alani,E. (2012) The unstructured linker arms of Mlh1-Pms1 are important for interactions with DNA during mismatch repair. *J. Mol. Biol.*, **422**, 192–203.

12. Bronner,C.E., Baker,S.M., Morrison,P.T., Warren,G., Smith,L.G., Lescoe,M.K., Kane,M., Earabino,C., Lipford,J., Lindblom,A., *et al.* (1994) Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. *Nature*, **368**, 258–261.

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

14. Hall,M.C., Shcherbakova,P.V., Fortune,J.M., Borchers,C.H., Dial,J.M., Tomer,K.B. and Kunkel,T.A. (2003) DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair. *Nucleic Acids Res.*, **31**, 2025–2034.

15. Argueso,J.L., Kijas,A.W., Sarin,S., Heck,J., Waase,M. and Alani,E. (2003) Systematic mutagenesis of the Saccharomyces cerevisiae MLH1 gene reveals distinct roles for Mlh1p in meiotic crossing over and in vegetative and meiotic mismatch repair. *Mol. Cell. Biol.*, **23**, 873–886.

16. Guarné,A., Ramon-Maiques,S., Wolff,E.M., Ghirlando,R., Hu,X., Miller,J.H. and Yang,W. (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO J.*, **23**, 4134–4145.

17. Gorman,J., Plys,A.J., Visnapuu,M.-L., Alani,E. and Greene,E.C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, **17**, 932–938.

18. Groothuizen,F.S., Winkler,I., Cristóvão,M., Fish,A., Winterwerp,H.H.K., Reumer,A., Marx,A.D., Hermans,N., Nicholls,R.A., Murshudov,G.N., *et al.* (2015) MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *elife*, **4**, e06744.

19. Kunkel,T.A. and Erie,D.A. (2015) Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu. Rev. Genet.*, **49**, 291–313.

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

21. Kadyrov,F.A., Genschel,J., Fang,Y., Penland,E., Edelmann,W. and Modrich,P. (2009) A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 8495–8500.

22. Goellner,E.M., Smith,C.E., Campbell,C.S., Hombauer,H., Desai,A., Putnam,C.D. and Kolodner,R.D. (2014) PCNA and Msh2-Msh6 activate an Mlh1-Pms1 endonuclease pathway required for Exo1-independent mismatch repair. *Mol. Cell*, **55**, 291–304.

23. Lu,A.L. (1987) Influence of GATC sequences on Escherichia coli DNA mismatch repair in vitro. *J. Bacteriol.*, **169**, 1254–1259.

24. Hermans,N., Laffeber,C., Cristovão,M., Artola-Borán,M., Mardenborough,Y., Ikpa,P., Jaddoe,A., Winterwerp,H.H.K., Wyman,C., Jiricny,J., *et al.* (2016) Dual daughter strand incision is processive and increases the efficiency of DNA mismatch repair. *Nucleic Acids Res.*, **44**, 6770–6786.

25. Gueneau,E., Dherin,C., Legrand,P., Tellier-Lebegue,C., Gilquin,B., Bonnesoeur,P., Londino,F., Quemener,C., Le Du,M.-H., Márquez,J.A., *et al.* (2013) Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nat. Struct. Mol. Biol.*, **20**, 461–468.

26. Goellner,E.M., Putnam,C.D. and Kolodner,R.D. (2015) Exonuclease 1-dependent and independent mismatch repair. *DNA Repair*, **32**, 24–32.

27. Genschel,J., Kadyrova,L.Y., Iyer,R.R., Dahal,B.K., Kadyrov,F.A. and Modrich,P. (2017) Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 4930–4935.

28. Wanat,J.J., Singh,N. and Alani,E. (2007) The effect of genetic background on the function of Saccharomyces cerevisiae mlh1 alleles that correspond to HNPCC missense mutations. *Hum. Mol. Genet.*, **16**, 445–452.

29. Ou,J., Niessen,R.C., Vonk,J., Westers,H., Hofstra,R.M.W. and Sijmons,R.H. (2008) A database to support the interpretation of human mismatch repair gene variants. *Hum. Mutat.*, **29**, 1337–1341.

30. Gorman,J., Wang,F., Redding,S., Plys,A.J., Fazio,T., Wind,S., Alani,E.E. and Greene,E.C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3074-3083.

31. Hombauer,H., Campbell,C.S., Smith,C.E., Desai,A. and Kolodner,R.D. (2011) Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell*, **147**, 1040–1053.

32. Sacho,E.J., Kadyrov,F.A., Modrich,P., Kunkel,T.A. and Erie,D.A. (2008) Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha. *Mol. Cell*, **29**, 112–121.

33. Claeys Bouuaert,C. and Keeney,S. (2017) Distinct DNA-binding surfaces in the ATPase and linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. *PLoS Genet.*, **13**, e1006722.

34. Ban,C., Junop,M. and Yang,W. (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, **97**, 85–97.

35. Pillon,M.C., Lorenowicz,J.J., Uckelmann,M., Klocko,A.D., Mitchell,R.R., Chung,Y.S., Modrich,P., Walker,G.C., Simmons,L.A., Friedhoff,P., *et al.* (2010) Structure of the endonuclease domain of MutL: unlicensed to cut. *Mol. Cell*, **39**, 145–151.

36. Tran,P.T. and Liskay,R.M. (2000) Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha. *Mol. Cell. Biol.*, **20**, 6390–6398.

37. Räschle,M., Dufner,P., Marra,G. and Jiricny,J. (2002) Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *J. Biol. Chem.*, **277**, 21810–21820.

38. 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.

39. Hall,M.C. and Kunkel,T.A. (2001) Purification of eukaryotic MutL homologs from Saccharomyces cerevisiae using self-affinity technology. *Protein Expr. Purif.*, **21**, 333–342.

40. Cooper,M.P., Balajee,A.S. and Bohr,V.A. (1999) The C-terminal domain of p21 inhibits nucleotide excision repair In vitro and In vivo. *Mol. Biol. Cell*, **10**, 2119–2129.

41. Finkelstein,J., Antony,E., Hingorani,M.M. and O'Donnell,M. (2003) Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy. *Anal. Biochem.*, **319**, 78–87.

42. Modesti,M. (2011) Fluorescent labeling of proteins. *Methods Mol. Biol.*, **783**, 101–120.

43. Rogacheva,M.V., Manhart,C.M., Chen,C., Guarne,A., Surtees,J. and Alani,E. (2014) Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *J. Biol. Chem.*, **289**, 5664–5673.

44. Manhart,C.M., Ni,X., White,M.A., Ortega,J., Surtees,J.A. and Alani,E. (2017) The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLoS Biol.*, **15**, e2001164.

45. Alani,E., Chi,N.W. and Kolodner,R. (1995) The Saccharomyces cerevisiae Msh2 protein specifically binds to duplex oligonucleotides containing mismatched DNA base pairs and insertions. *Genes Dev.*, **9**, 234–247.

46. Rose MD, Winston F and Hieter P (1990) Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual Cold Spring Harbor Laboratory Press.

47. Heck,J.A., Argueso,J.L., Gemici,Z., Reeves,R.G., Bernard,A., Aquadro,C.F. and Alani,E. (2006) Negative epistasis between natural variants of the Saccharomyces cerevisiae MLH1 and PMS1 genes results in a defect in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 3256–3261.

48. Gietz,R.D., Schiestl,R.H., Willems,A.R. and Woods,R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355–360.

49. Christianson,T.W., Sikorski,R.S., Dante,M., Shero,J.H. and Hieter,P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.

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

51. Brown,M.W., Kim,Y., Williams,G.M., Huck,J.D., Surtees,J.A. and Finkelstein,I.J. (2016) Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. *Nat. Commun.*, **7**, 10607.

52. Hammer,Ø., Harper,D.A. and Ryan,P.D. (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, **4**, 1–9.

53. Efron,B. and Tibshirani,R.J. (1993) An Introduction to the Bootstrap Softcover reprint of the original 1st ed. 1993 edition. Chapman and Hall/CRC, New York.

54. Kim,Y., de la Torre,A., Leal,A.A. and Finkelstein,I.J. (2017) Efficient modification of λ-DNA substrates for single-molecule studies. *Sci. Rep.*, **7**, 2071.

55. Yu,H., Singh Gautam,A.K., Wilmington,S.R., Wylie,D., Martinez-Fonts,K., Kago,G., Warburton,M., Chavali,S., Inobe,T., Finkelstein,I.J., *et al.* (2016) Conserved sequence preferences contribute to substrate recognition by the proteasome. *J. Biol. Chem.*, **291**, 14526–14539.

56. Fishbain,S., Inobe,T., Israeli,E., Chavali,S., Yu,H., Kago,G., Babu,M.M. and Matouschek,A. (2015) Sequence composition of disordered regions fine-tunes protein half-life. *Nat. Struct. Mol. Biol.*, **22**, 214–221.

57. Pluciennik,A., Dzantiev,L., Iyer,R.R., Constantin,N., Kadyrov,F.A. and Modrich,P. (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 16066–16071.

58. Kawasoe,Y., Tsurimoto,T., Nakagawa,T., Masukata,H. and Takahashi,T.S. (2016) MutSα maintains the mismatch repair capability by inhibiting PCNA unloading. *eLife*, **5**, e15155.

59. Liu,J., Hanne,J., Britton,B.M., Bennett,J., Kim,D., Lee,J.-B. and Fishel,R. (2016) Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, **539**, 583–587.

60. Gallardo,I.F., Pasupathy,P., Brown,M., Manhart,C.M., Neikirk,D.P., Alani,E. and Finkelstein,I.J. (2015) High-throughput universal DNA curtain arrays for single-molecule fluorescence imaging. *Langmuir*, **31**, 10310–10317.

61. Soniat,M.M., Myler,L.R., Schaub,J.M., Kim,Y., Gallardo,I.F. and Finkelstein,I.J. (2017) Next-Generation DNA Curtains for Single-Molecule Studies of Homologous Recombination. *Meth. Enzymol.*, **592**, 259–281.

62. Ban,C. and Yang,W. (1998) Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell*, **95**, 541–552.

63. Dyer,P.N., Edayathumangalam,R.S., White,C.L., Bao,Y., Chakravarthy,S., Muthurajan,U.M. and Luger,K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth. Enzymol.*, **375**, 23–44.

64. Visnapuu,M.-L. and Greene,E.C. (2009) Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat. Struct. Mol. Biol.*, **16**, 1056–1062.

65. Smith,C.E., Bowen,N., Graham,W.J., Goellner,E.M., Srivatsan,A. and Kolodner,R.D. (2015) Activation of Saccharomyces cerevisiae Mlh1-Pms1 Endonuclease in a Reconstituted Mismatch Repair System. *J. Biol. Chem.*, **290**, 21580–21590.

66. Sakato,M., Zhou,Y. and Hingorani,M.M. (2012) ATP binding and hydrolysis-driven rate-determining events in the RFC-catalyzed PCNA clamp loading reaction. *J. Mol. Biol.*, **416**, 176–191.

67. Dionne,I., Brown,N.J., Woodgate,R. and Bell,S.D. (2008) On the mechanism of loading the PCNA sliding clamp by RFC. *Mol. Microbiol.*, **68**, 216–222.

68. Pillon,M.C., Babu,V.M.P., Randall,J.R., Cai,J., Simmons,L.A., Sutton,M.D. and Guarné,A. (2015) The sliding clamp tethers the endonuclease domain of MutL to DNA. *Nucleic Acids Res.*, **43**, 10746–10759.

69. Lao,Y., Lee,C.G. and Wold,M.S. (1999) Replication protein A interactions with DNA. 2. Characterization of double-stranded DNA-binding/helix-destabilization activities and the role of the zinc-finger domain in DNA interactions. *Biochemistry*, **38**, 3974–3984.

70. Kim,C., Snyder,R.O. and Wold,M.S. (1992) Binding properties of replication protein A from human and yeast cells. *Mol. Cell. Biol.*, **12**, 3050–3059.

71. Blackwell,L.J. and Borowiec,J.A. (1994) Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol. Cell. Biol.*, **14**, 3993–4001.

72. Hess,M.T., Mendillo,M.L., Mazur,D.J. and Kolodner,R.D. (2006) Biochemical basis for dominant mutations in the Saccharomyces cerevisiae MSH6 gene. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 558–563.

73. Obradovic,Z., Peng,K., Vucetic,S., Radivojac,P., Brown,C.J. and Dunker,A.K. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins*, **53 Suppl 6**, 566–572.

74. Reyes,G.X., Schmidt,T.T., Kolodner,R.D. and Hombauer,H. (2015) New insights into the mechanism of DNA mismatch repair. *Chromosoma*, 10.1007/s00412-015-0514-0.

CHAPTER 4

Evidence for the coordinated assembly of MLH clamps in eukaryotic mismatch repair

Abstract

Eukaryotic mismatch repair (MMR) is initiated when a misincorporation event occurring during DNA replication is recognized by MutS homolog (MSH) heterodimeric proteins. MSH complexes then recruit MutL homolog (MLH) heterodimeric complexes. Once recruited to the MSH-mismatch site, MLH complexes undergo ATP-dependent conformational changes that result in the nicking of the newly replication strand. This nicking step provides entry for downstream repair factors that subsequently excise and repair the error. MLH proteins contain long and flexible intrinsically disordered regions (IDRs) that connect structured N- and C-terminal domains. Recent studies have led to models in which Mlh1 triggers, through its IDR, ATP-dependent conformational steps that enable MLH proteins to form a clamp on DNA, interact with MSH proteins, and activate DNA nicking. We tested such models by inserting FRB and FKBP dimerization domains at various positions within the Mlh1 and Pms1 IDRs. Upon confirmation that these recombinant MLH proteins were proficient for MMR functionality, we then induced rapamycin-dependent FRB-FKBP interactions both *in vivo* and *in vitro*. Through this approach we created an *mlh1* allele that could be reversibly disrupted for MMR upon rapamycin treatment, providing a new tool to disrupt MMR on demand. We further showed that restraining the MLH linkers near the N-terminal ATP binding domains caused defects in MMR. In contrast, restraints predicted to maintain free movement of the N-terminal ATP binding domains had very weak effects on MMR. Restriction predicted to disrupt movement altered DNA and PCNA dependent regulation of Mlh1-Pms1 ATPase activity, DNA binding, and endonuclease activities. Together, our work provides support for MLH linker domains mediating distinct conformational steps in DNA mismatch repair and is consistent with a two-step clamp model for MLH proteins acting in eukaryotic MMR. Our works also highlights roles for intrinsically disordered regions mediating step-wise changes critical for protein function.

Introduction

Intrinsically disordered regions (IDRs), characterized as being conformationally flexible, are present in roughly 30% of eukaryotic proteins. While many IDRs have unknown roles, some facilitate interactions with other proteins or between regions of the same protein. IDRs are often critical for overall protein function, but insights into how they act are often lacking because they lack conservation, and amino acid substitutions in IDRs are often silent (Oldfield and Dunker, 2014; van der Lee et. al., 2014; Uversky, 2019).

Our study is focused on the IDRs present in the highly conserved MutL family of DNA mismatch repair (MMR) proteins. MMR acts to correct DNA misincorporation errors that arise during replication (reviewed in Kunkel & Erie, 2015). In baker's yeast, MMR is initiated upon mismatch recognition by the Msh2-Msh6 (MutSα) and Msh2-Msh3 (MutSβ) MutS homolog (MSH) proteins. These proteins act as ATP-modulated sliding clamps to recruit MLH proteins, principally Mlh1-Pms1, which nick the newly replicated strand of DNA in the vicinity of the mismatch through steps that require interactions with the DNA polymerase processivity clamp PCNA (Gradia et al., 1999; Gorman et al., 2012; Pluciennik et al., 2010; Kawasoe et al., 2016). Nicks located 5' to the mismatch act as entry sites for Exo1, a 5' to 3' exonuclease, to excise the DNA through the mismatch site. The single stranded binding protein RPA coats the ssDNA gap, after which a DNA polymerase (δ or ε) resynthesizes the DNA, thus removing the misincorporation error. An Exo1-independent mechanism is also thought to act where Mlh1-Pms1 makes multiple nicks on the newly replicated strand in the vicinity of the mismatch, enabling Polymerases δ and/or ε to remove the mismatch through DNA synthesis and strand displacement steps. In both mechanisms, remaining nicks are sealed by DNA ligase (Kadyrov et al., 2009; Goellner et al., 2014; Goellner et al., 2015; Hermans et al., 2016).

In baker's yeast, Mlh1-Pms1 (MutLα) acts as a central player to transduce mismatch recognition signals produced by MSH mismatch binding to downstream excision steps. It displays a latent ATPase activity that is required for MMR *in vivo,* and *in vitro* studies have shown that ATP binding by Mlh1-Pms1 stimulates its endonuclease activity on double-stranded DNA. This stimulation is significantly greater on DNA substrates loaded with PCNA (Kadyrov et al., 2006; Tran et al., 2000; Räschle et al., 2002; Kim et al., 2019; Genschel et al., 2017). Mlh1 (164 amino acid long IDR) and Pms1 (295 amino acid long IDR) contain IDRs separated by globular N- and C- terminal domains (Figure 4.1A; Plys et al., 2012; Hall et al., 2003; Prolla et al., 1994; Argueso et al., 2003; Guarné et al., 2004). Interestingly, atomic force microscopy and ATP hydrolysis analyses provided evidence that ATP binding by Mlh1-Pms1 is accompanied by conformational rearrangements involving IDRs in both subunits (Sacho et al., 2008; Hall et al., 2001; Hall et al., 2003).

Sacho et al. (2008) and Kunkel and Erie (2015) proposed that ATP binding to the Mlh1 subunit of the Mlh1-Pms1 heterodimer promotes the formation of a one-armed complex in which the N-terminal ATP binding domain of Mlh1 and its IDR fold into or near the C-terminal Mlh1-Pms1 domain (Figure 4.1B). ATP binding to Pms1 then condenses its linker arm to yield a condensed complex that positions the N-terminus of Mlh1 and Pms1 in proximity to the C-terminal domains. This conformational change, modulated through the IDR domains of both proteins, is hypothesized to change Mlh1–Pms1 affinity's for DNA and activate its endonuclease activity facilitated by the C-terminus of Pms1 (Sacho et al., 2008; Ban et al., 1999; Pillon et al., 2010; Gueneau et al., 2013; Tran and Liskay, 2000). Supporting this idea, Mlh1 displays a higher

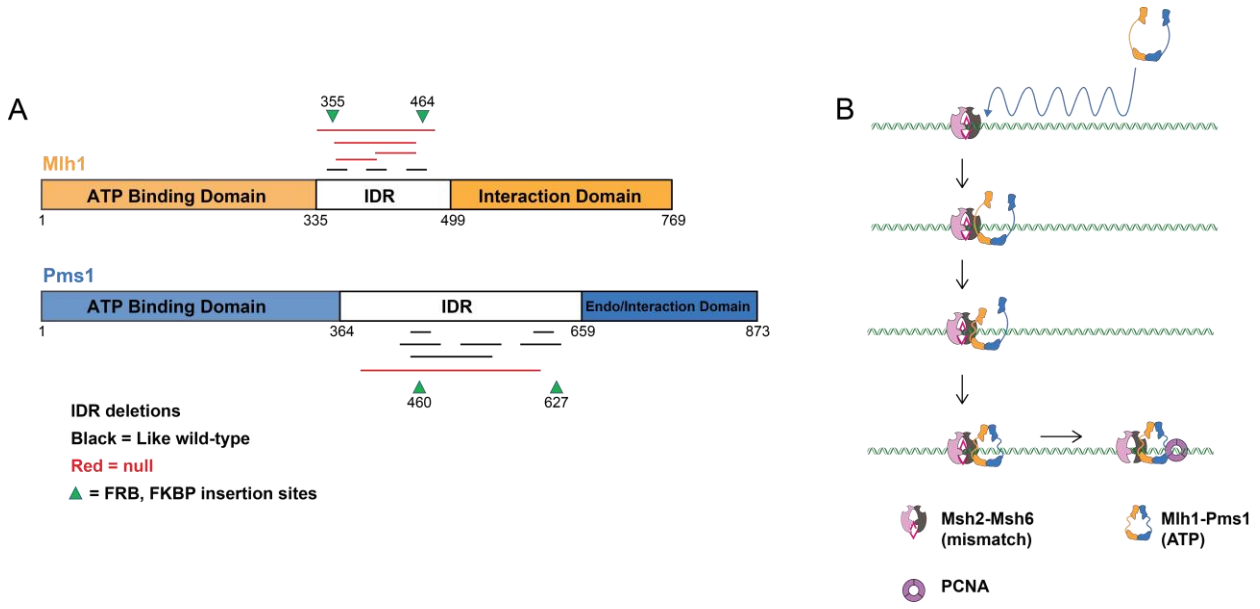


Figure 4.1. A mechanistic model for Mlh1-Pms1 acting in MMR. (A) Functional domains of yeast Mlh1 and Pms1. Cartoon representations of the Mlh1 and Pms1 subunits, with the N-terminal ATP binding and C-terminal endonuclease/MLH interaction domains separated by IDRs. Amino acid locations of the domains in yeast Mlh1 and Pms1 are shown. IDR deletions that do not disrupt or only weakly disrupt Mlh1 or Pms1 functions (black lines) or confer *mlh1* or *pms1* null phenotypes (red lines), are shown (Plys et al., 2012). The green triangles indicate insertion sites for FRB and FKBP into the Mlh1 and PMs1 IDRs. FRB and FKBP were inserted immediately after the indicated amino acid position of Mlh1 and Pms1. (B) A model for how Mlh1-Pms1 locates and interacts with MSH proteins during MMR. See Introduction for details.

affinity for ATP compared to Pms1 (Hall et al. 2003), mutations in the *MLH1* ATP binding domain confer significantly higher mutation rates compared to analogous positions in *PMS1* (Tran and Liskay, 2000) and proteolytic cleavage of one or both IDRs in Mlh1-Pms1 abolished the DNA-binding activity of the complex *in vitro* and disrupted MMR *in vivo* (Plys et al., 2012; Gorman et al., 2010). In addition, deletion analysis of the IDRs showed that both were required for MMR but acted differently. Plys et al. (2012) showed that the Mlh1 IDR was more sensitive to deletion compared to Pms1 (Figure 4.1A) and that deleting the Pms1 IDR resulted in a Mlh1-pms1 complex that could not bind DNA or form a ternary complex with Msh2-Msh6 and a DNA mismatch substrate *in vitro,* which was not the case when the Mlh1 IDR was deleted.

Single molecule studies have provided a mechanistic view of how Mlh1-Pms1 interacts with MSH proteins and mismatched DNA (Figure 4.1B). Studies by Gorman et al. (2010, 2012) showed that Mlh1-Pms1 is targeted to lesion-bound Msh2-Msh6 by one-dimensional hopping and three-dimensional diffusion mechanisms, and introducing ATP provoked the release of MutSα/MutLα from mismatched DNA. Interestingly, Kim et al. (2019) showed that Mlh1-Pms1 complexes containing shorter IDRs that caused defects in MMR retained wild-type DNA binding affinity but showed diffusion defects on both naked and nucleosome-coated DNA. They also showed that the IDRs regulated the ATP hydrolysis and nuclease activities encoded by the N- and C-terminal domains of the complex, respectively. These studies illustrated the multi-functional roles for IDRs in regulating Mlh1-Pms1 diffusion on DNA and nucleolytic processing. More recently Liu et al. (2016) showed that *E. coli* MutS sliding clamps recruited MutL on mismatched DNA to form a MutS-MutL search complex. In their model they hypothesized that “an open conformation of EcMutL is required to interact with EcMutS sliding clamps, which then binds ATP to form a second, exceedingly stable, ring-like clamp” that acts as

a search complex which leaves the mismatch to find other strand discrimination factors needed for proper MMR. (Liu et al., 2016; Groothuizen et al., 2015).

The models outlined above would be greatly strengthened by direct evidence implicating the Mlh1 and Pms1 IDRs in facilitating coordinated steps in MMR. In this study we sought to restrict the motion of the IDRs by introducing FRB and FKBP, two small protein domains that dimerize in the presence of rapamycin (Spencer et al., 1993), into the IDR domains of Mlh1 and Pms1. First, we introduced both FRB and FKBP into the Mlh1 IDR and showed that the presence of rapamycin conferred a null MMR phenotype, consistent with previous deletion analysis of the Mlh1 IDR (Plys et al. 2012), and thus providing an inducible system to disrupt MMR. Next, we asked if locking the linker regions of Mlh1 and Pms1 proteins through their respective IDRs would also disrupt MMR. These studies showed that restraining the MLH IDRs near the N-terminal domain of Mlh1 or Pms1 caused defects in MMR *in vivo*, and disrupted DNA and PCNA dependent regulation of Mlh1-Pms1 ATPase activity and Mlh1-Pms1 DNA binding activity *in vitro*. Remarkably, restraining the MLH linkers in a manner predicted to maintain free movement of the N-terminal ATP binding domains had very weak effects on MMR. These observations are consistent with a two-step clamp model for MLH proteins acting in eukaryotic MMR and highlight how IDRs mediate coordinated conformational changes that are critical for protein function.

Results

Rationale

MLH proteins are hypothesized to interact with MSH proteins bound to mismatched DNA through the MSH sliding clamp interacting with an open MLH complex, followed by MLH ring

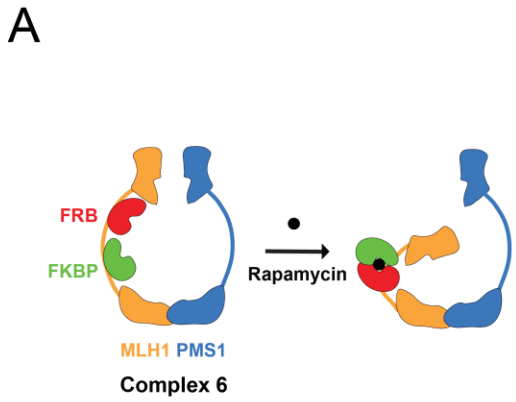
closure (Figure 4.1B; Gorman et al., 2010; 2012; Lui et al., 2016). The MLH IDRs are thought to play key roles in facilitating these steps (Sacho et al., 2008; Kim et al., 2019; Mardenborough et al., 2019). A prediction from this model is that restricting the coordinated movements of the IDRs in the MLH proteins would disrupt the formation of a stable MLH clamp on DNA. We tested this using a chemically inducible dimerization system in which one can control the dimerization of a pair of proteins with a small molecule that acts as a "dimerizer" (Fegan et al., 2008; Spencer et al., 1993). Dimerizers can bring regions of proteins together that normally lack interaction and have been used to study mechanistic aspects of protein localization, targeted protein degradation, transcriptional control and signal transduction (Xu et al., 2010; Voß et al., 2015; Inobe and Nukina, 2016; Zhu et al., 2017; Geda et al., 2008). Here we targeted the Mlh1 and Pms1 IDRs through chemically inducible dimerization experiments involving the FKBP–rapamycin binding (FRB) domain of the mammalian target of rapamycin protein (mTOR) protein and the 12-kDa FK506 binding protein (FKBP). The FRB (93 amino acids) and FKBP (107 amino acids) domains are relatively small and each bind to the small molecule macrolide rapamycin. Rapamycin binding to each domain facilitates a stable and tight FRB-FKBP heterodimer (Kd in the fmol range; Banaszynski et al., 2005).

Using our previous structure-function analysis of the Mlh1 and Pms1 IDRs, we inserted FRB and FKBP domains into two locations within the Mlh1 and Pms1 IDRs that did not affect or minimally affected Mlh1-Pms1 function (Figures 4.1A, 4; Plys et al., 2012). With this system in place, we tested if we could first reversibly inhibit MMR by restricting the motion of the linker arm of Mlh1 in the presence of rapamycin. We chose Mlh1 as our initial target because previous work suggested that Mlh1 acts as the initiating partner to condense the Mlh1-Pms1 complex in a stepwise fashion that permits interactions with MSH proteins. This is then followed

by the formation of an MLH clamp through a condensation step involving Pms1 (Sacho et al., 2008; Kunkel and Erie, 2015; Liu et al., 2016). Consistent with this idea was the finding that mlh1-Pms1 complexes lacking the Mlh1 IDR can still bind to DNA and interact with MSH proteins on mismatched DNA, whereas Mlh1-pms1 complexes lacking the Pms1 IDR fail to bind to DNA and interact with the Msh2-Msh6-mismatch complex (Plys et al., 2012). We then tested if locking the IDRs together using the FRB-FKBP system would prevent their coordinated movement as predicted by the two-step model and thus disrupt MMR. We tested the effects of disrupting IDR movement genetically, using a sensitive mutator assay, and biochemically, by measuring Mlh1-Pms1 ATPase, DNA binding, and endonuclease activities. As shown below, our data are consistent with the two-step clamp model for MLH proteins acting in eukaryotic MMR.

Rapamycin-induced disruption of MMR *in vivo.*

We inserted the FRB and FKBP domains into two positions in the Mlh1 IDR, immediately after amino acids 355 and 464. The effects of these insertions were assessed using a highly sensitive *lys2-A$_{14}$* reversion assay, where an *mlh1Δ* strain shows a three-orders of magnitude higher mutation rate compared to *wild-type* (Tran et al., 1997; Kim et al., 2019). As shown in Figure 4.2B and Table 4.1, single insertions of FRB or FKBP in Mlh1 conferred mutation rates that were marginally higher than *wild-type* (1.8 to 4.9-fold). The double insertion allele, *mlh1-FRB$_{355}$-FKBP$_{464,}$* (referred to as Complex 6 in Figure 4.2) also conferred a very weak mutator phenotype (46-fold). In the presence of rapamycin, the single domain insertion *mlh1* alleles conferred mutator phenotypes that were stronger than in the absence of rapamycin, (68 to 114-fold) but much lower than the *mlh1Δ* (5,820-fold). One explanation for why the single domain



B

Genotype	relative to *wild-type*	relative to *wild-type* with rapamycin
wild-type	1.0	1.0
mlh1Δ	5,110	5,820
mlh1-FRB$_{355}$-PMS1	1.8[a,b]	112[a,b]
mlh1-FKBP$_{464}$-PMS1	4.9[a,b]	114[a,b]
mlh1-FRB$_{464}$-PMS1	2.6[a,b]	68.4[a,b]
mlh1-FRB$_{355}$,FKBP$_{464}$-PMS1	46.4[a,b]	4,770[a]

lys2-A$_{14}$ reversion

6

C

Genotype	relative to *wild-type*	with rapamycin	post rapamycin
wild-type	1.0	1.2	1.0
mlh1-FRB$_{355}$,FKBP$_{464}$-PMS1	47	4,800	53

lys2-A$_{14}$ reversion

6

Figure 4.2. Rapamycin-induced disruption of MMR in a strain containing FRB and FKBP domain insertions in the Mlh1 IDR. (A) Illustration of FRB-FKBP domain insertions into the Mlh1 IDR, in the presence or absence of rapamycin. (B) Mutation rate analysis of *mlh1* alleles containing the indicated FRB and FKBP insertions in the Mlh1 IDR. Mutation rate was determined using the *lys2-A$_{14}$* reversion assay (Materials and Methods). Mutation rates are presented relative to *wild-type*, with the more complete dataset presented in Table 4.1. [a,b]Significantly different from wild-type and *mlh1Δ* with respect non-overlapping 95% confidence intervals (CIs). [a]Significantly different from *wild-type* as determined by non-overlapping 95% CIs. (C) Wild-type and *mlh1-FRB$_{355}$,FKBP$_{464}$-PMS1* strains were grown in the presence and absence of rapamycin, after which with the mutation rate was determined. Independent colonies from strains grown in the presence of rapamycin were then grown in the absence of rapamycin, after which mutation rate was determined in the *lys2-A$_{14}$* reversion assay (Materials and Methods).

Table 4.1. *FRB* and *FKBP* domain insertions into the linker arms of *mlh1 and pms1* confer MMR defects in the presence of rapamycin.

relevant genotype in *tor1-1, fpr1Δ* background	mutation rate ($\times 10^{-7}$) (95% CI)	*n*	relative to *wild-type*	mutation rate ($\times 10^{-7}$) (95% CI)	*n*	relative to *wild-type*
				with 2 ug/ml rapamycin		
wild-type	2.44 (1.96-2.91)	64	**1.0**	2.25 (2.04-3.32)	38	**1.0**
*mlh1*Δ	12,500 (10,900-17,100)	15	5,112	13,100 (7,890-18,200)	15	5,820
MLH1 inserts						
mlh1-FRB$_{355}$	4.45 (3.18-15.6)	15	1.82	253 (125-327)	15	112
mlh1-FKB$_{P464}$	12.1 (5.79-26.5)	15	4.94	256 (183-305)	15	114
mlh1-FRB$_{464}$	6.39 (3.75-7.59)	15	2.62	154 (118-294)	15	68.4
mlh1-FRB$_{355,}$ FKBP$_{464}$ (#6)	113 (69.1-254)	15	46.4	11,700 (7,720-16,100)	15	4,770
PMS1 inserts						
pms1-FRB$_{460}$	4.73 (2.21-6.47)	15	1.93	24.5 (18.1-31.5)	15	10.9
pms1-FKBP$_{460}$	4.03 (1.84-4.35)	15	1.65	54.9 (45.1-82.4)	15	24.4
pms1-FRB$_{627}$	8.16 (5.72-11.7)	15	3.34	64.9 (53.9-131)	15	28.8
pms1-FKBP$_{627}$	11.4 (8.47-16.4)	15	4.69	242 (220-286)	15	108
MLH1, PMS1 inserts						
mlh1-FKBP$_{464}$, pms1-FRB$_{627}$ (#1)	16.4 (12.2-19.2)	15	6.72	194 (142-232)	15	86.1
mlh1-FRB$_{464}$, pms1-FKBP$_{460}$ (#2)	4.03 (3.37-6.35)	15	1.65	2,400 (1,740-3,490)	15	1,070
mlh1-FKBP$_{464}$, pms1-FRB$_{460}$ (#3)	5.49 (3.47-6.56)	15	2.25	2,830 (2,230-4,530)	15	1,260
mlh1-FRB$_{355}$, pms1-FKBP$_{460}$ (#4)	12.5 (4.58-15.8)	15	5.12	3,520 (2,690-4,160)	15	1,560
mlh1-FRB$_{355}$, pms1-FKBP$_{627}$ (#5)	184 (165-220)	15	75.5	2,630 (2,340-4,200)	15	1,170

The indicated alleles were integrated into the *MLH1* or *PMS1* or both locus in the S288C strain background and tested for DNA mismatch repair functions both in the presence and absence of 2ug/ml of rapamycin using the *lys2-A$_{14}$* reversion assay with 95% CI (confidence interval) presented (Materials and Methods; Table 4.5). For these experiments a *wild-type* strain, EAY4450 was generated by deleting the *FPR1* locus and introducing the *tor1-1 (S1972A)* allele into the EAY1269 S288c background. n represents the number of independent measurements from at least two transformants.

insertion alleles display higher mutation rates in the presence of rapamycin is that FRB and FKBP are each competent to bind rapamycin, and such binding bind could impact IDR functions by altering their flexibility. Alternatively, rapamycin binding proteins in the yeast cell other than the Fpr1 protein (yeast homolog of human FKBP, deleted in our strain background) may be able to weakly interact with FRB or FKBP, sequestering Mlh1 function. Consistent with the latter idea is that the single domain insertion alleles were recessive (Figure 4.3; Table 4.2); if rapamycin impacted IDR function, one might have expected to still see a mutator phenotype when the FRB and FKBP insertion alleles were expressed in the presence of wild-type *MLH1*, though at lower levels (see below). In contrast to the single *FRB* and *FKBP mlh1* insertion alleles, a complete null phenotype was observed for the *mlh1-FRB*$_{355}$*-FKBP*$_{464}$ double insertion allele (Complex 6) grown in the presence of rapamycin, a phenotype identical to that seen when the Mlh1 IDR was completely deleted (Plys et al., 2012).

We then asked if we could reverse the MMR null phenotype observed for *mlh1-FRB*$_{355}$*-FKBP*$_{464}$ strains grown in the presence of rapamycin, by removing rapamycin, and after a period of outgrowth, repeating the mutation rate assay. As shown in Figure 4.2C and Table 4.3, the mutation rate of the *mlh1-FRB*$_{355}$*-FKBP*$_{464}$ strain (Complex 6) was reduced to levels seen in the absence of rapamycin. To our knowledge, this is the first reported system in which MMR functions could be modulated using a small molecule; applications using this approach are presented in the Discussion.

Cross-linking mass spectroscopy analysis of Mlh1-Pms1 reveals inter-subunit interactions involving the N-terminal, IDR and C-terminal domains of both proteins.

Genotype	relative to *wild-type* with rapamycin	relative to *wild-type* with rapamycin and pEAA213 (*MLH1*)
wild-type	1.0	1.0
mlh1Δ	5,820	1.2
mlh1-FRB$_{355}$-PMS1	112*	5.1
mlh1-FKBP$_{464}$-PMS1	114*	6.8*
mlh1-FRB$_{464}$-PMS1	68.4*	2.5

lys2-A$_{14}$ reversion

Figure 4.3. Wild-type Mlh1 rescues the rapamycin-induced minor MMR defect seen in strains containing FRB and FKBP insertions in the Mlh1 IDR. Mutation rate analysis of strains containing the indicated FRB and FKBP insertions in the Mlh1 IDR. Strains were grown in the presence or absence of an *ARS-CEN* vector containing *MLH1* (pEAA213). The assay was performed in the presence of 2 μg/ml rapamycin, with the mutation rate presented relative to *wild-type*. A detailed analysis is presented in Table 4.2. *Significantly different from *wild-type* as determined by non-overlapping 95% CIs.

Table 4.2. *FRB* and *FKBP* domain insertions into the linker arms of *MLH1 and PMS1* confer a dominant phenotype for MMR in the presence of rapamycin.

relevant genotype in *tor1-1, fpr1Δ* background	mutation rate ($\times 10^{-7}$) (95% CI)	*n*	relative to *wild-type*	mutation rate ($\times 10^{-7}$) (95% CI)	*n*	relative to *wild-type*
				with 2 ug/ml rapamycin		
wild-type	2.44 (1.96-2.91)	64	**1.0**	2.25 (2.04-3.32)	38	**1.0**
*mlh1*Δ	12,500 (10,900-17,100)	15	5,112	13,100 (7,890-18,200)	15	5,820
mlh1* strains containing *pMLH1						
wild-type, pEAA213	2.35 (1.93-3.97)	15	**1.0**	1.77 (1.11-3.45)	15	**1.0**
*mlh1*Δ, pEAA213	2.69 (1.22-4.78)	15	1.14	2.11 (1.85-4.78)	15	1.2
mlh1-FRB$_{355}$, pEAA213	3.54 (2.68-5.08)	15	1.27	9.45 (2.47-13.8)	15	5.1
mlh1-FKB$_{P464}$, pEAA213	4.98 (2.58-6.22)	15	1.79	12.6 (5.54-20.3)	15	6.82
mlh1-FRB$_{464}$, pEAA213	3.22 (1.59-5.56)	15	1.16	4.62 (2.51-7.07)	15	2.49
mlh1, pms1* strains containing *pMLH1, PMS1						
wild-type, pEAA67	2.75 (1.94-3.57)	15	**1.0**	2.80 (2.06-4.45)	15	**1.0**
*mlh1*Δ, pEAA67	4.16 (2.52-5.93)	15	1.51	4.32 (2.56-6.67)	15	1.5
mlh1-FKBP$_{464}$, pms1-FRB$_{627}$, pEAA67	3.01 (1.74-5.83)	15	1.1	68.7 (55.8-86.9)	15	24.5
mlh1-FRB$_{464}$, pms1-FKBP$_{460}$, pEAA67	4.37 (3.03-6.64)	15	1.59	314 (186-446)	15	112
mlh1-FKBP$_{464}$, pms1-FRB$_{460}$, pEAA67	4.32 (2.45-8.83)	15	1.57	687 (496-1,327)	15	245
mlh1-FRB$_{355}$, pms1-FKBP$_{460}$, pEAA67	6.15 (5.24-7.57)	15	2.24	898 (875-1,078)	15	320
mlh1-FRB$_{355}$, pms1-FKBP$_{627}$, pEAA67	71.7 (47.6-93.5)	15	26.11	788 (529-1131)	15	281
***mlh1, pms1* strains containing empty vector**						
wild-type, pRS416	3.22 (1.79-4.06)	15	**1.0**	2.11 (1.54-3.61)	15	**1.0**
*mlh1*Δ, pRS416	23,300 (11,500-24,200)	15	7,230	13,100 (11,100-18,100)	15	6,190
mlh1-FKBP$_{464}$, pms1-FRB$_{627}$, pRS416	26.3 (17.5-31.1)	15	8.15	303 (288-337)	15	143
mlh1-FRB$_{464}$, pms1-FKBP$_{460}$, pRS416	8.23 (6.02-13.2)	15	2.55	2,520 (2,340-3,840)	15	1,190
mlh1-FKBP$_{464}$, pms1-FRB$_{460}$, pRS416	8.95 (4.95-11.5)	15	2.78	2,950 (1,550-5,040)	15	1,390
mlh1-FRB$_{355}$, pms1-FKBP$_{460}$, pRS416	25.8 (17.7-34.8)	15	8.02	3,480 (2,450-4,080)	15	1,650
mlh1-FRB$_{355}$, pms1-FKBP$_{627}$, pRS416	432 (284-526)	15	135.6	3,170 (2,450-3,980)	15	1,500

The indicated alleles were transformed with either a plasmid expressing a *wild-type* copy of *MLH1* (pEAA213), a *wild type* copy of *MLH1*, *PMS1*, *MSH2* (pEAA67), or an empty vector (pRS416), and tested for DNA mismatch repair functions both in the presence and absence of 2 μg/ml of rapamycin using the *lys2-A$_{14}$* reversion assay. 95% CIs (confidence intervals) are presented (Materials and Methods; Table 4.6). n represents the number of independent measurements from at least two transformants.

Table 4.3. *FRB* and *FKBP* domain insertions into the linker arm of *MLH1 confers* MMR defects in the presence of rapamycin that can be reversed.

relevant genotype in *tor1-1, fpr1Δ* background	mutation rate ($\times 10^{-7}$) (95% CI)	*n*	relative to *wild-type*	mutation rate ($\times 10^{-7}$) (95% CI)	*n*	relative to *wild-type*
				with 2 ug/ml rapamycin		
wild-type	3.04 (2.3-4.4)	15	**1.0**	3.7 (1.95-10.7)	15	1.23
mlh1-FRB$_{355,}$FKBP$_{464}$ (Complex 6)	143 (113-186)	15	47	14,600 (10,400-21,200)	15	4,806
Post growth in rapamycin						
wild-type	2.36 (1.67-3.47)	15	0.78	3.11 (1.80-6.78)	15	1.02
mlh1-FRB$_{355,}$FKBP$_{464}$ (Complex 6)	99.8 (68.3-171)	15	32.9	162 (77.5-190)	15	53.4

Wild-type and *mlh1-FRB$_{355}$,FKBP$_{464}$-PMS1* (Complex 6) strains were grown in the presence and absence of rapamycin, after which the mutation rate was determined. Independent colonies from strains grown in the presence of rapamycin were then grown in the absence of rapamycin, after which the mutation rate was determined. The *lys2-A$_{14}$* reversion assay was used to measure mutation rate, with 95% CIs presented (Materials and Methods).

The above studies suggesting that the restriction of Mlh1 IDR movements disrupts MMR encouraged us to perform a cross-linking mass spectrometry (XL-MS) analysis of the Mlh1-Pms1 complex. Such an analysis could help us determine if the IDRs interact with each other as predicted by the two-step clamp model presented in Figure 4.1B. We used reversible cross-linking coupled with mass spectrometry (XL-MS) to identify regions of Mlh1-Pms1 that are in close proximity to each other during cross-linking with the reversible crosslinker disuccinimidyl sulfoxide (DSSO; Materials and Methods). DSSO is used to cross-link exposed lysine residues that are in close proximity (10.1 angstrom spacer length; Yu and Huang, 2018; Yugandhar et al., 2020). While this method cannot yield detailed structural information, it can identify regions of proteins that potentially interact and thus give insights into the movements of Mlh1-Pms1 during MMR.

As shown below, XL-MS of Mlh1-Pms1 (Figure 4.4 and Table 4.7) provided several key pieces of information that informed our studies of the Mlh1 and Pms1 IDRs. First, we observed a large number of crosslinks that mapped within an IDR or between IDRs, but not to regions where we inserted the FRB and FKBP domains. These observations are consistent with the IDRs being highly flexible, solvent exposed, and capable of interacting with nearby partners. It also confirmed our decision to insert the FRB and FKBP domains into specific regions of Mlh1 and Pms1. Second, many of the inter-subunit crosslinks map to regions of Mlh1 and Pms1 that were previously shown by genetic analysis to be critical for their function (see asterisks in Figure 4.4, Table 4.4; Argueso et al., 2003; Tran and Liskay, 2000; Smith et al., 2013; Plys et al., 2012). Third, we observed cross-links between the ATP binding (N-terminal) domains of Mlh1 and Pms1, consistent with previous structural and biochemical studies that show these domains interact when both domains are ATP bound (Ban et al., 1999; Tran and Liskay, 2000). Fourth,

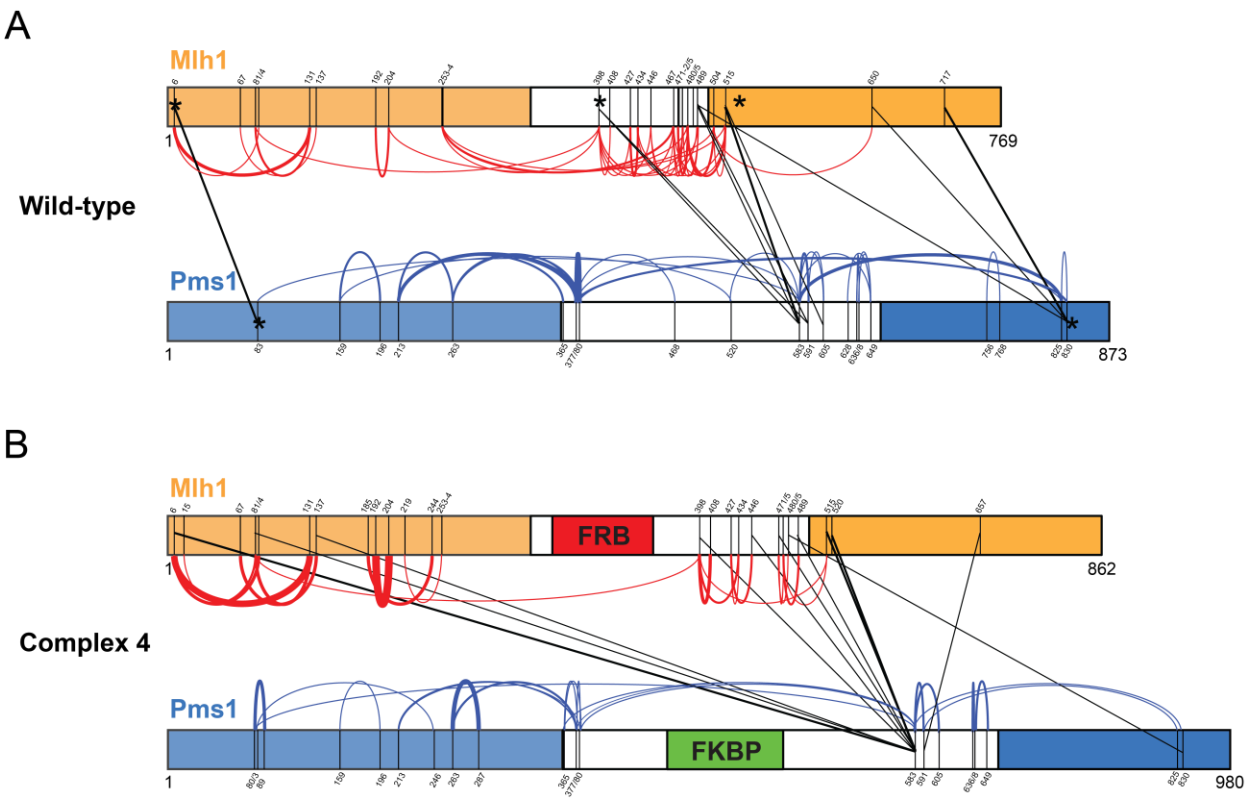


Figure 4.4. Crosslinking mass spectrometry of Mlh1-Pms1 demonstrate linkers arms interact with the C-terminal region of Pms1. Mlh1-Pms1 (A) and *mlh1-FRB*$_{355}$, *pms1-FKBP*$_{460}$ (B) were crosslinked with DSSO in the absence of rapamycin and then subjected to cross-linking mass spectrometry (XL-MS; Materials and Methods). Shown are the inter- and intra-subunit lysine residue crosslinks with respect to the ATP binding, IDR and endonuclease/interaction domains of Mlh1 and Pms1. * represent positions in Mlh1 and Pms1 which were shown previously to disrupt MMR.

Table 4.4. Phenotypic analysis of *mlh1* and *pms1* alleles that map within or near Mlh1-Pms1 crosslinking sites as determined by XL-MS.

Crosslinking sites in Mlh1	*mlh1* allele	MMR Phenotype	Reference
6	*R4A, K6A*	+/-	Argueso et al., 2003
131	*E129A, K131A*	+	Argueso et al., 2003
253,254	*K253A, K254A*	-	Argueso et al., 2003
398	*E396A, K398A*	+, but – nearby	Argueso et al., 2003
467,471,472	*K467A, D468A, K471A, K472A, K473A*	+	Argueso et al., 2003
490	*D486A, D487A, E488A, K489A*	+	Argueso et al., 2003
504	*K504A, E505A, R506A*	+/-	Argueso et al., 2003
515	*K515A, K516A*	+/-	Argueso et al., 2003
717	*E714A, D715A, E716A, K717A*	+	Argueso et al., 2003
Crosslinking sites in Pms1			
83	*G97A*	-	Tran and Liskay, 2000
583-606	*IDR deletion (584-634)*	+/-	Plys et al., 2012
825-830	*C817R, C848S*	-	Smith et al., 2013

+ indicates phenotype similar to wild-type, - indicates similar to MMR null, +/- intermediate phenotype.

we observed cross-links between the Mlh1 and Pms1 C-terminal domains, which consistent with this domain being required for Mlh1-Pms1 interactions (Gueneau et al., 2013). Lastly, we observed crosslinks between the Mlh1 IDR and the Pms1 endonuclease domain, consistent with models in which the condensation of the N- and C-termini of MLH proteins is important to activate MLH functions (Sacho et al., 2008; Ban et al., 1999; Pillon et al., 2010; Gueneau et al., 2013).

We then analysed an Mlh1-Pms1 complex containing FRB in Mlh1 and FKBP in Pms1 in the absence of rapaymicn (Complex 4; Figures 4.4, 4.5, Table 4.1). A similar number of crosslinks were observed for Complex 4 and Mlh1-Pms1 (Table 4.7). In general, the intra and inter-subunit IDR interactions also appeared; however, novel cross-links were seen for Complex 4 between the Mlh1 N-terminal domain and the Pms1 IDR, reminiscent of the formation of the single-arm condensed complex involving the N-terminal domain of Mlh1 observed by Sacho et al. (2008). One explanation for this observation is that the FRB domain in Complex 4 is poised to perform the condensation steps that are normally seen for Mlh1-Pms1 in the presence of ATP (see below). Taken together, the XL-MS analysis both confirmed previous structural analyses of the MLH proteins and also identified regions for IDR interactions that would be very difficult to obtain by conventional structural approaches.

It is important to note that we were unable to perform XL-MS in the presence of ATP or the non-hydrolysable ATP analog AMP-PMP because the primary amines present in these molecules interfered with our detection of DSSO-induced cross-linking by mass spectrometry. We did not attempt XL-MS with an ATP analog that lacks the primary amine, 6-methyl ATP, because this analog was unable to activate Mlh1-Pms1's latent endonuclease activity. Lastly, we

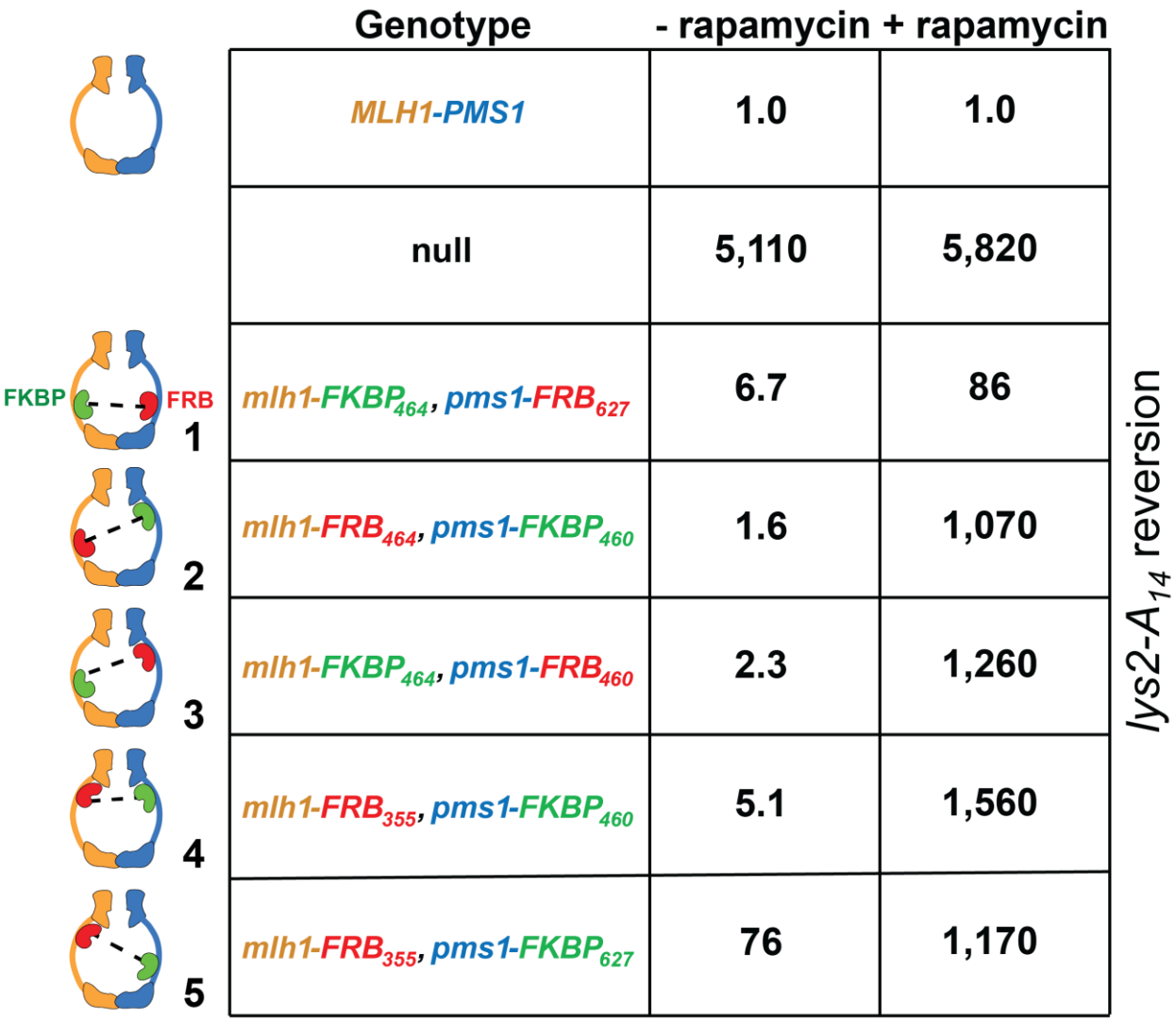


Figure 4.5. Rapamycin-induced disruption of MMR in strains containing FRB and FKBP domain insertions in the Mlh1 and Pms1 IDRs. *Wild-type* and strains containing the indicated FRB and FKBP Mlh1 and Pms1 IDR insertions were grown in the absence and presence of rapamycin (2 μg/ml) and then analyzed for mutation rate using the *lys2-A*$_{14}$ reversion assay (Materials and Methods). Mutation rate is presented relative to *wild-type*, with a detailed analysis presented in Table 4.1. All complexes displayed mutation rates that were significantly different from *wild-type* and *mlh1Δ* as determined by non-overlapping 95% CIs.

were unable to perform XL-MS in the presence of rapamycin because its addition interfered with sample preparation.

Rapamycin-induced interactions between Mlh1 and Pms1 IDRs disrupt MMR.

To further test the two-step clamp model, we hypothesized that the prevention of MLH clamp formation through FRB-FKBP interactions would inhibit MMR, and that such inhibition could be detected in biochemical assays. We inserted FRB and FKBP into several positions within the Mlh1 and Pms1 IDRs and obtained five Mlh1-Pms1 FRB/FKBP insertion complexes that maintained roughly wild-type levels of MMR in the absence of rapamycin (Figure 4.5). This wild-type function was also seen when one subunit contained the FRB/FKBP insertion and the other a wild-type partner, indicating that the FRB/FKBP insertions were well tolerated (Figures 4.2, 4.6).

We then tested the MMR functions of strains bearing the five complexes in the presence of rapamycin. Strains bearing a complex predicted to restrain the IDR of Mlh1 or Pms1 near the N-terminal region displayed a strong MMR defect (Complexes 2-5; Figure 4.5; Table 4.1), whereas Complex 1, predicted to restrain the IDRs of Mlh1 and Pms1 near their C-terminal regions, conferred relatively weak MMR defects. Interestingly, the small increase in mutation rate for Complex 1 seen in the presence of rapamycin was very similar to the mutation rate seen when the FRB and FKBP insertion alleles in Complex 1 were expressed in the presence of a wild-type partner, indicating that the effect of rapamycin on Complex 1 function was likely to be minimal. One explanation for these phenotypes is that interactions predicted to restrain the coordinated movements of the N-terminal ATP binding domains (Complexes 2-5) disrupt MMR whereas those predicted to maintain free movement of the N-terminal ATP binding domains (Complex 1) do not.

Genotype	relative to *wild-type*	relative to *wild-type* with rapamycin
wild-type	1.0	1.0
MLH1-pms1-FRB$_{460}$	1.9	11
MLH1-pms1-FKBP$_{460}$	1.7	24
MLH1-pms1-FRB$_{627}$	3.3	29
MLH1-pms1-FKBP$_{627}$	4.7	108

lys2-A$_{14}$ reversion

Figure 4.6. Single FRB or FKBP insertions into the Pms1 IDR confer minor defects in MMR in the presence of rapamycin. Mutation rate analysis of strains containing the indicated FRB and FKBP insertions in the Pms1 IDR. The mutation rate was determined using the *lys2-A*$_{14}$ reversion assay (Materials and Methods) and is presented relative to *wild-type*, with a detailed analysis presented in Table 4.1. *Significantly different from *wild-type* as determined by non-overlapping 95% CIs.

A challenge using the rapamycin system with proteins that already form a complex is that it is not easy to directly show that the FRB and FKBP domains are interacting in the presence of rapamycin. Because of this issue and the fact that we were unable to use XL-MS to identify FRB-FKBP interactions in Complexes 1-5 in the presence of rapamycin, we performed the following genetic and biochemical analyses that are consistent with rapamycin-induced FRB-FKBP interactions occurring efficiently within Complexes 1-5. First, we could purify Complexes 1-4, indicating that they form stable heterodimers and in the presence of rapamycin the complexes remained in solution and active with respect to ATPase and endonuclease activities (Figures 4.7, 4.9. 4.10 and 4.11). These properties were expected because the local concentration of the FRB and FKBP domains within a single heterodimer is expected to be much higher than predicted between Mlh1-Pms1 heterodimers, which are present at roughly ~600 copies in a single nucleus (Karim et al., 2013). Second, we observed similarly strong mutator phenotypes for Complexes 2 to 5 in the presence of rapamycin, suggesting that FRB and FKBP insertion at both positions in the Mlh1 and Pms1 IDRs are capable of conferring a mutator phenotype upon rapamycin addition. Third, our XL-MS analysis (Figure 4.4) shows that the C-terminal regions of the IDR are capable of forming a large number of crosslinks and are therefore in close proximity, providing evidence that the FRB and FKBP domains in Complex 1 can dimerize in the presence of rapamycin, consistent with the idea that Complexes 2-5 show MMR defects in the presence of rapamycin due to IDR restriction whereas rapamycin induced FRB-FKBP interactions only mildly affect Complex 1. It is also not surprising that Complex 1 is functional in MMR since deletions in the regions where FRB and FKBP were inserted conferred very weak MMR defects *in vivo* (Plys et al., 2012).

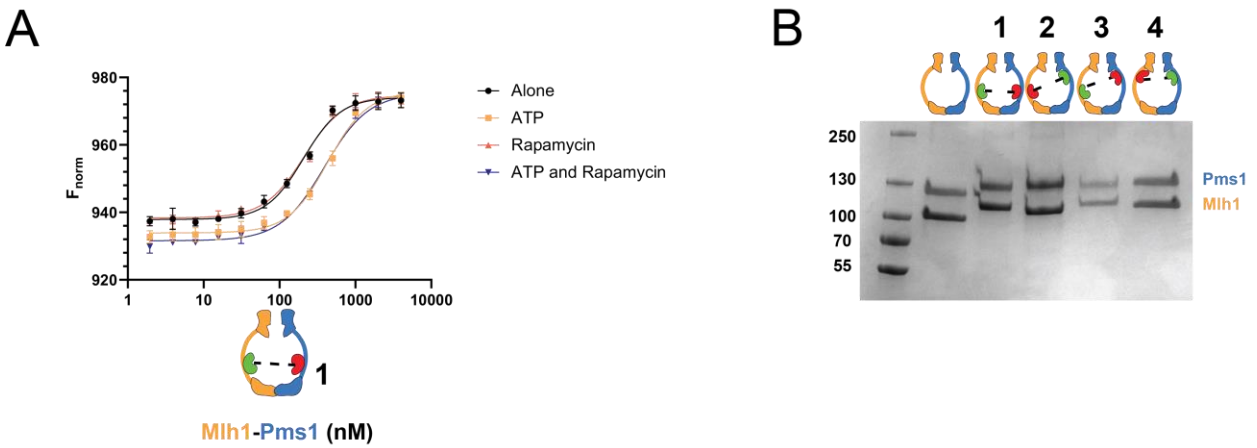


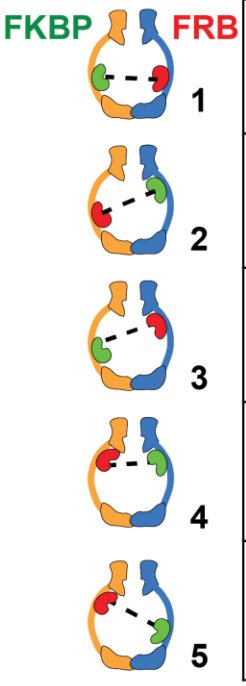
Figure 4.7. MST analysis of Complex 1 and purification of other complexes. (A) MST analysis of Complex 1. See Figure 4.11B and Materials and Methods for details. (B) SDS-PAGE analysis (8% Coomassie blue R250 stained gel) of wild-type, mlh1-FKBP$_{464}$, pms1-FRB$_{627}$ (Complex 1), mlh1-FRB$_{464}$, pms1-FKBP$_{460}$ (Complex 2), and mlh1-FKBP$_{464}$, pms1-FRB$_{460}$ (Complex 3), and mlh1-FRB$_{355}$-pms1-FKBP$_{460}$ (Complex 4).

Finally, we asked if Complexes 1-5 expressed in the presence of similar levels of wild-type complex would display evidence of subunit mixing. If this was the case, then ¼ of the complexes in a cell would be expected to contain FRB and FKBP insertions in the same heterodimer, and the mutation rate for cells expressing both the complex and wild-type Mlh1-Pms1 in the presence of rapamycin would be expected to be ¼ of the rate seen when only the complex was expressed. If intermolecular interactions were prevalent then we would have expected to see a two-fold reduction in mutation rate. As shown in Figure 4.8 and Table 4.2, the mutation rate for all five complexes expressed in the presence of wild-type Mlh1-Pms1 and rapamycin was roughly ¼ of the rate seen when only the complex was expressed, consistent with rapamycin inducing interactions between FRB and FKBP insertions within single heterodimers.

Rapamycin induced N-terminal dimerization pre-activates Mlh1-Pms1 *in vitro*.

To further understand why Complexes 2-5 but not Complex 1 displayed significantly increased mutation rates in the presence of rapamycin, we purified Complexes 1 and 4 and assayed their activity in ATPase and endonuclease assays (Figure 4.9; Kadyrov et al., 2006; Pluciennik et al., 2010; Kim et al., 2019). In the absence of rapamycin both complexes displayed normal levels of PCNA stimulation of endonuclease activity compared to Mlh1-Pms1 (Figure 4.9B). We then performed these experiments in the presence of rapamycin and found that rapamycin did not impact Mlh1-Pms1 or Complex 1 endonuclease activity; however, Complex 4 displayed an endonuclease activity that was higher than seen for Mlh1-Pms1 (Figure 4.9B, C).

To better understand the enhanced endonuclease activity of Complex 4 in the presence of rapamycin we measured the ATPase activity of Mlh1-Pms1 and Complexes 1 and 4. Previous studies showed that Mlh1-Pms1 ATPase activity is stimulated in the presence of DNA duplex



Relevant genotype	+ rapamycin, empty vector	+ rapamycin, MLH1, PMS1
wild-type	1.0	1.0
mlh1Δ	6,190	1.57
mlh1-FKBP$_{464}$, *pms1-FRB*$_{627}$	143	24.5
mlh1-FRB$_{464}$, *pms1-FKBP*$_{460}$	1,190	112
mlh1-FKBP$_{464}$, *pms1-FRB*$_{460}$	1,390	245
mlh1-FRB$_{355}$, *pms1-FKBP*$_{460}$	1,650	320
mlh1-FRB$_{355}$, *pms1-FKBP*$_{627}$	1,500	281

lys2-A$_{14}$ reversion

Figure 4.8. Rapamycin induced dimerization causes a dominant MMR defect *in vivo*
Wild-type and strains containing the indicated FRB and FKBP Mlh1 and Pms1 IDR insertions and either an empty (pRS416) or *MLH1* and *PMS1 ARS-CEN* vector were grown in the absence and presence of rapamycin (2 μg/ml) and then analyzed for mutation rate in the *lys2-A*$_{14}$ reversion assay (Materials and Methods). Mutation rate is presented relative to *wild-type*, with a detailed analysis presented in Table 4.2. All complexes were significantly different from *wild-type* and *mlh1Δ* as determined by non-overlapping 95% CIs.

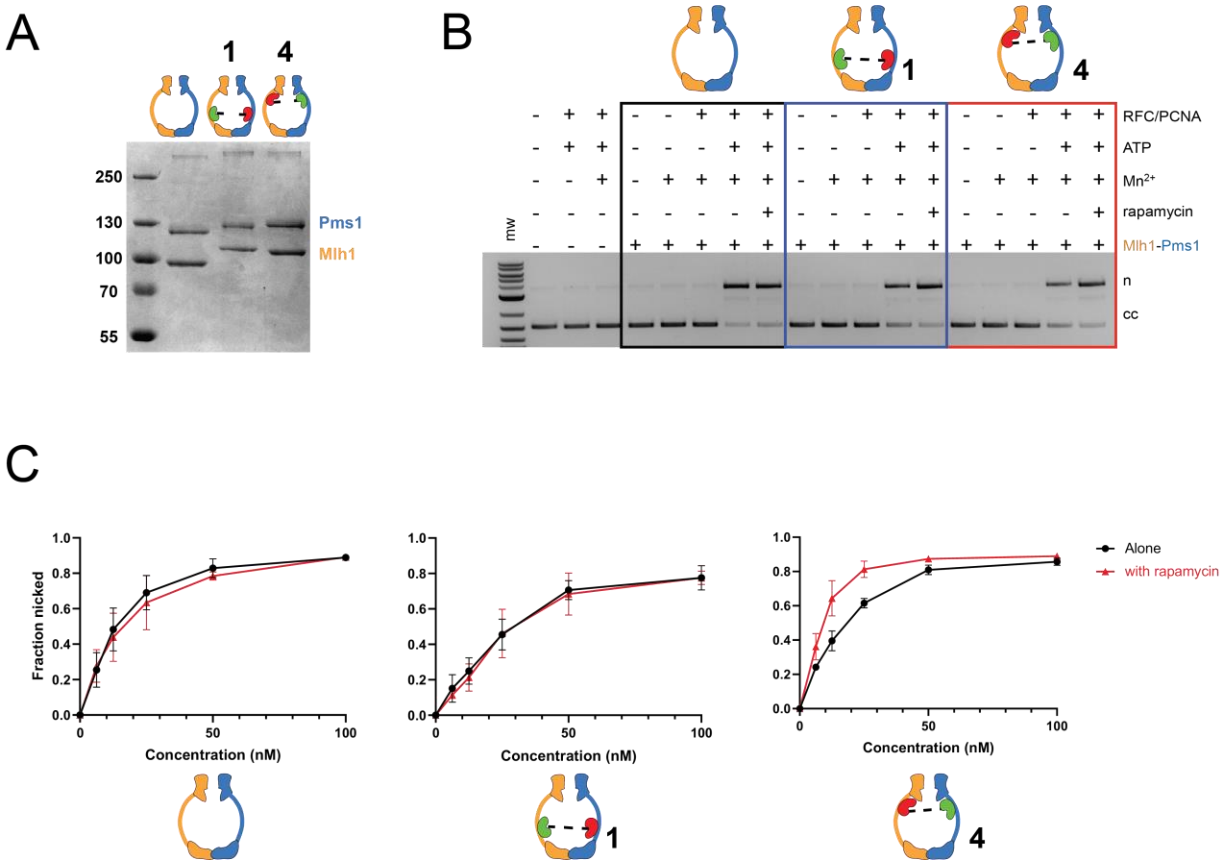


Figure 4.9. **Enhanced nicking observed for mlh1-FRB$_{355}$, pms1-FKBP$_{460}$ in the presence of rapamycin.** (A) SDS-PAGE analysis (8% Coomassie blue R250 stained gel) of wild-type, mlh1-FKBP$_{464}$, pms1-FRB$_{627}$ (Complex 1) and mlh1-FRB$_{355}$, pms1-FKBP$_{460}$ (Complex 4). (B) Endonuclease activity of Mlh1-Pms1 and mutant derivatives (50 nM each) was determined on a closed circular DNA substrate in the presence (+) or absence (-) of $MnSO_4$, ATP, rapamycin, and yeast PCNA/RFC (Materials and Methods). $MnSO_4$, ATP, rapamycin, RFC and PCNA were included at 5 mM, 0.5 mM, 1 μM, 125 nM and 250 nM, respectively. (C) Endonuclease activity was determined at the indicated concentrations of wild-type, Complex 1 and Complex 4. Assays were performed in the presence of $MnSO_4$, ATP, RFC, PCNA. Rapamycin was included as indicated. Error-bars indicate the standard deviation of three replicates (Figure 4.10).

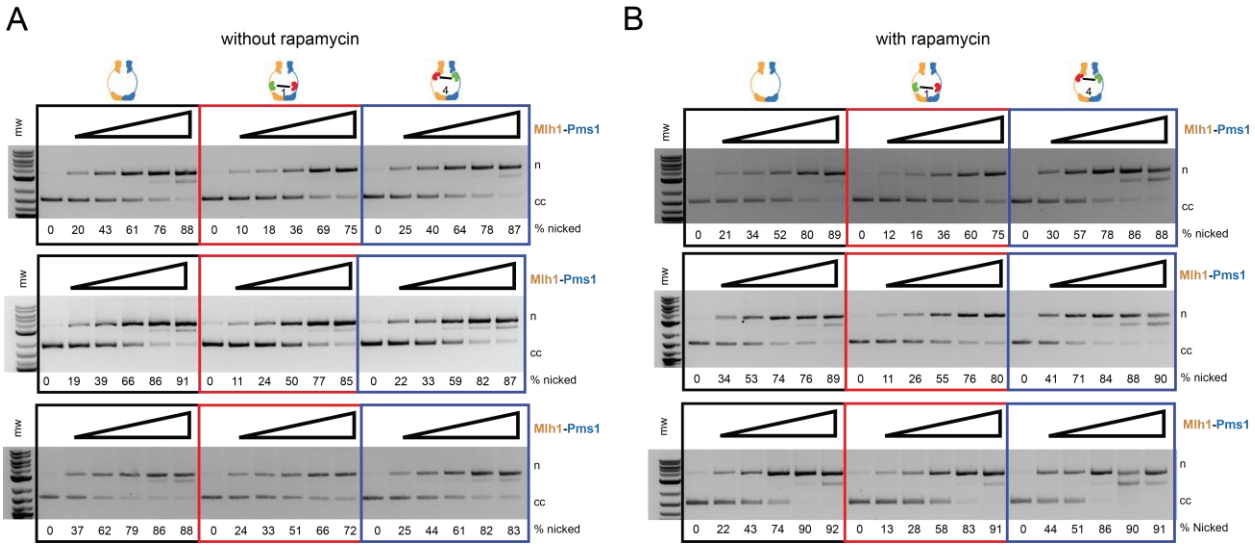


Figure 4.10. Representative gel images of protein titration endonuclease activity assays. Individual titrations are shown which were used to make the graphs presented in Figure 4.9C. Endonuclease assays (Materials and Methods) were performed on closed circular plasmid DNA in the presence of $MnSO_4$, ATP, RFC, PCNA without (A) or with (B) rapamycin. Mlh1-Pms1 complexes were titrated from 0-100 nM.

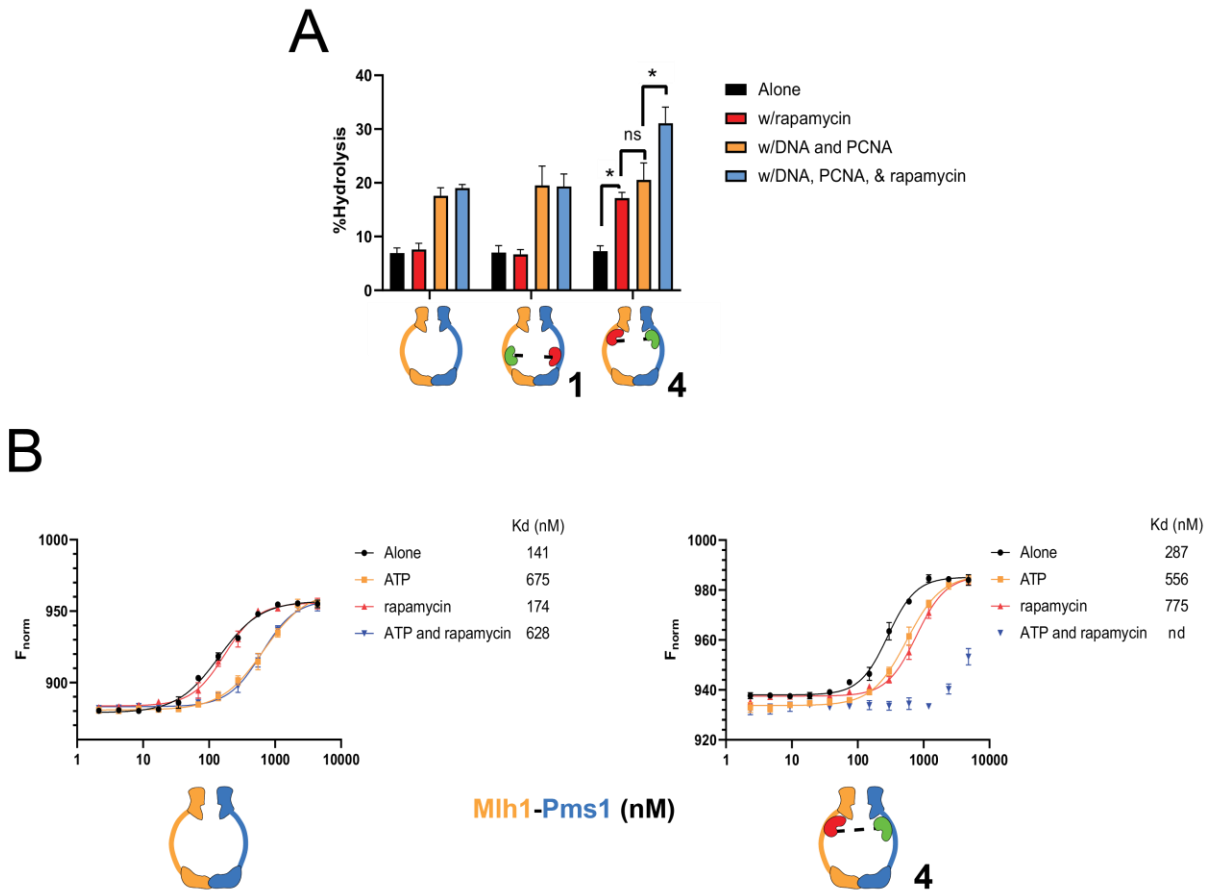


Figure 4.11. mlh1-FRB$_{355}$, pms1-FKBP$_{460}$ (Complex 4) displays enhanced ATP hydrolysis and defective DNA binding in the presence of rapamycin. (A) The ATP hydrolysis activities of Mlh1-Pms1, mlh1-FKBP$_{464}$, pms1-FRB$_{627}$ (Complex 1) and mlh1-FRB$_{355}$, pms1-FKBP$_{460}$ (Complex 4; 0.40 μM each) were determined in the presence and absence of PCNA (0.5 μM), 49-bp homoduplex DNA (0.75 μM), PCNA (0.250 μM), and rapamycin (1 mM). Error bars indicate ± one standard deviation for three replicates. * denotes statistical significance in a student's T-test. (B) MST analysis of Mlh1-Pms1 and mlh1-FRB$_{355}$, pms1-FKBP$_{460}$ (Complex 4) in the presence and absence of 49 bp homoduplex DNA (20 nM), ATP (1 mM) and rapamycin (1 mM). Experiments represent the mean ± standard deviation of three independent experiments and were performed using at least two independently purified batches of each protein. F_{norm} was calculated by dividing F_{hot} (average fluorescence value in the heated state) by F_{cold} (average fluorescence value measured in the cold state before the IR laser is turned on) and plotted as parts per thousand (%). Binding constants were determined by nonlinear curve fitting using GraphPad Prism. Calculated K_d values are shown. See the Materials and Methods for details.

oligonucleotide and PCNA (Kadyrov et al., 2007; Tran et al., 2000; Räschle et al., 2002; Kim et al., 2019; Genschel et al., 2017). As shown in Figure 4.11A, we found that Mlh1-Pms1 and Complex 1 displayed similar levels of ATPase activity in the presence and absence of rapamycin; this similar activity was also seen in the presence of DNA and PCNA. Like Mlh1-Pms1 and Complex 1, Complex 4 displayed an ATPase activity that was similar to that seen for Mlh1-Pms1 and Complex 4 in the absence of rapamycin; however, a significant increase in ATPase activity was observed in the presence of rapamycin both with and without DNA and PCNA. One explanation for the increased endonuclease and ATPase activities of Complex 4 in the presence of rapamycin is that it has been inappropriately switched to an activated state due to the N-terminal ATP binding domains of Mlh1 and Pms1 being placed in close proximity. Such a hypothesis could account for the increased mutation rate for Complex 4, but not for Complex 1, which would be predicted to maintain flexible N-terminal domains. Complex 1 appears to have a slightly reduced overall endonuclease activity, which could be due to the presence of FRB and FKBP domains in the vicinity of the endonuclease region. Taken together, these observations support the idea that in Complex 4 we are chemically inducing the localization of the Mlh1 and Pms1 N-terminal domains in an inappropriate manner that disrupts MMR.

Rapamycin induced N-terminal dimerization prevents Complex 4 from binding to DNA.

As indicated above, one explanation for the mutator phenotypes seen for strains bearing Complex 4 grown in the presence of rapamycin is that Complex 4 is inappropriately switched to an activated state. A prediction of this hypothesis and the model presented in Figure 4.1B is that Complex 4 would show defects in forming a stable complex on DNA. To test this, we assayed DNA binding activity of Mlh1-Pms1, and Complexes 1 and 4 to a 49 bp homoduplex oligo using

Microscale Thermophoresis (MST; Figure 4.7A, 4.11B). In these experiments we observed binding of Mlh1-Pms1 that was reduced in the presence of ATP. This binding was unaffected by the presence of rapamycin. The reduced binding of Mlh1-Pms1 in the presence of ATP is consistent with previous studies indicating that Mlh1-Pms1 ring opening and closing is linked to interactions with DNA (Kim et al., 2019; Gorman et al., 2010; Hall et al., 2003). For Complex 4 analysed in the absence of rapamycin, we observed DNA binding affinities similar to Mlh1-Pms1 both in the presence and absence of ATP. In contrast, in the presence of rapamycin, we observed a significantly reduced DNA binding affinity for Complex 4 both in the presence and absence of ATP. In fact, the Kd for binding of Complex 4 in the presence of ATP and rapamycin could not be determined due to our inability to saturate binding of Complex 4 to DNA due to the aggregation of Mlh1-Pms1 at high concentrations. In contrast, Complex 1, which showed a mild mutator phenotype in the presence of rapamycin, displayed DNA binding activity in the presence of rapamycin that appeared similar to wild-type (Figure 4.7A). Together these experiments are consistent with the locking of the IDRs in Complex 4 blocking the ability of Mlh1-Pms1 to bind in a stepwise fashion to DNA and interact with MSH complexes, thus accounting for the MMR defect seen *in vivo*. This is also consistent with our previous observation that a deletion of the Pms1 IDR disrupted DNA binding and prevented ternary complex formation of MSH and MLH proteins on a mismatched substrate (Plys et al., 2012).

Discussion

Single molecule studies involving the bacterial MutS, MutL, and MutH MMR proteins have led to a model where MutS interacts with an open complex of MutL at a mismatch site, followed by, in steps requiring ATP hydrolysis, the closing of the MutL clamp (Figure 4.1B). The resulting

MutS-MutL sliding clamp then acts as a search complex to interact with the MutH endonuclease to initiate downstream steps in MMR (Liu et al. 2016; Groothuizen et al., 2015). In yeast, single molecule analysis showed that Mlh1-Pms1 is targeted to a mismatch bound Msh2-Msh6 by a one-dimensional hopping and three-dimensional diffusion mechanism, followed by release of MutSα/MutLα from mismatched DNA upon binding ATP (Gorman et al., 2010, 2012).

We tested the above models by examining roles for the MLH IDRs in coordinating the recruitment of MLH proteins during MMR. Our work suggests that the two-step clamp model presented by Liu et al. (2016) is likely to be relevant for the eukaryotic MMR system. We found that restricting the mobility of the IDRs of Mlh1 and Pms1 through the use of the rapamycin-induced FRB-FKBP dimerization pre-activated the Mlh1-Pms1 complex as measured in ATPase and DNA nicking assays. These biochemical phenotypes were accompanied by significant MMR defects *in vivo*. As predicted by models presented in Figure 4.1B, the introduction of rapamycin to Complex 4 prevented it from forming a stable complex on DNA (Figure 4.12). This result is consistent with the locking of the IDRs, preventing Mlh1-Pms1 from binding in a stepwise fashion to DNA. It also supports previous work showing that Mlh1-Pms1 ring opening and closing is linked to interactions with DNA (Kim et al., 2019; Gorman et al., 2010; Hall et al., 2003), and is consistent with work showing that deletion of the Pms1 IDR disrupted both binding of the Mlh1-pms1 complex to DNA and the formation of an MSH-MLH ternary complex on mismatched DNA (Plys et al., 2012).

As outlined in the introduction, ATP binding to the Mlh1 subunit of the Mlh1-Pms1 heterodimer is thought to initially promote the formation of a one-armed complex in which the N-terminal ATP binding domain of Mlh1 is in proximity to the C-terminal Mlh1-Pms1 domain.

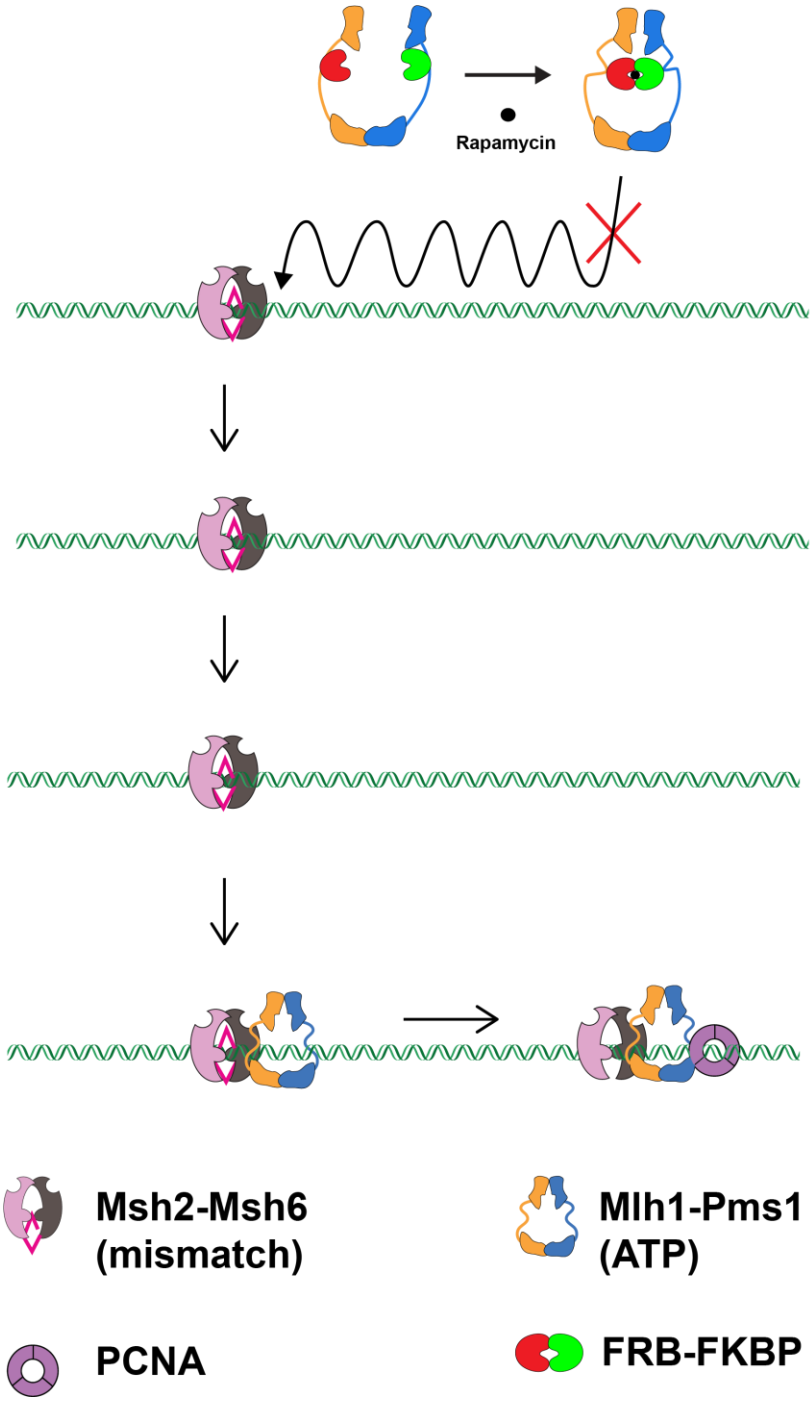


Figure 4.12. A model explaining how rapamycin induced dimerization of FRB-FKBP IDR insertions disrupts MMR. Summary of results indicating that constriction of the Mlh1 and Pms1 IDRs disrupts MMR by preventing the ability of the complex to undergo its 1D hoping/stepping diffusion on DNA to locate the mistmatch bound MSH proteins. This in tern limits the complex from interacting with MSH proteins and forming a stable clamp on DNA. See text for details.

This conformational change is then thought to be followed by ring closure through the Pms1 N-terminal domain, resulting in a closed Mlh1-Pms1 conformation competent for endonuclease activation (Sacho et al., 2008; Ban et al., 1999; Pillon et al., 2010; Gueneau et al., 2013; Kunkel and Erie, 2015). Our work supports the importance of Mlh1 as a "trigger" to initiate protein-protein and enzymatic functions in MMR. More specifically, we saw novel cross-links for Complex 4 between the Mlh1 N-terminal domain and the Pms1 IDR that are reminiscent of the formation of the single-arm condensed complex involving the N-terminal domain of Mlh1 observed by Sacho et al. (2008). Consistent with this observation is the finding that Complex 4 displayed an increased ATPase and nicking activity in the presence of rapamycin, suggesting that it had been preactivated. The idea of Mlh1 serving as a trigger is intriguing because Mlh1 serves as a common subunit for three MutL complexes in baker's yeast, Mlh1-Pms1, Mlh1-Mlh3 and Mlh1-Mlh2, where Mlh1-Mlh3 plays a minor role in MMR and a major role in the resolution of double-Holliday junctions in meiosis to form crossovers, and Mlh1-Mlh2 regulates gene conversion tract lengths in meiosis (reviewed in Furman et al., 2020). In such a model, recruitment of various MLHs is accomplished through the initial interactions of Mlh1 with the particular repair pathway, followed by clamp formation by the specificity subunit (Mlh2, Mlh3, or Pms1) that activates the complex for its specific role. Thus, our studies provide mechanistic support for the idea that IDRs license Mlh1-Pms1 to interact with MSH proteins during MMR and introduce nicks on the newly replicated DNA strand. It complements previous studies showing that IDRs regulate how a DNA repair enzyme scans chromatin for DNA lesions and how repair functions (for example, nicking of the newly replicated strand during MMR) are

activated (Gorman et al., 2010, 2012; Liu et al., 2016; Kim et al., 2019; Mardenborough et al., 2019).

Modulation of MMR using a small molecule

We present in this study a novel and reversible mechanism to disrupt MMR. More specifically, we were able to modulate MMR functions through the addition and removal of the small molecule rapamycin. Such a system can provide a valuable way to understand how an increased mutation rate can provide beneficial mutations to an organism adapting to changing environments (reviewed in Raghavan et al., 2018; 2019). Elevated mutation rates, while providing a source of beneficial mutations for adaptation, are associated with long-term fitness costs due to the accumulation of deleterious mutations (Lynch et al., 2016; Bui et al., 2015; 2017). Previously we developed a model of incompatibility between naturally segregating *MLH1* and *PMS1* alleles in yeast populations that would generate mutators that could facilitate adaptation to changing environments. We provided evidence that such incompatibilities could be generated through mating, followed by entry into meiosis to yield gametes with incompatible genotypes (Bui et al., 2015; 2017; Raghavan et al., 2018). In this model, incompatible populations that have adapted to an environmental stress can mate back to available strains to regain MMR functions and thus avoid fitness costs. However this model was not directly tested.

In the present study, we constructed Mlh1-Pms1 complexes that can confer a range of mutation rates in the presence of rapamycin. In the *lys2-A$_{14}$* reversion assay where the difference between *wild-type* and MMR defective is roughly 5,000-fold, we identified complexes that showed *wild-type* or near *wild-type* phenotypes in the absence of rapamycin, but a wide-range of elevated rates in its presence (Figures 4.2, 4.5). For example, in the presence of rapamycin,

Complex 1 conferred an 86-fold elevated rate compared to *wild-type*, Complex 2, a 1,100-fold higher rate, Complex 4, a 1,600-fold higher rate, and Complex 6, a rate indistinguishable from *mlh1* null. Most informative was that the null phenotype in Complex 6 could be reversed to the level observed in the absence of rapamycin. This resource can thus allow one to test the incompatibility models presented above. In such experiments, cells grown in stressed environments (such as high salt) could be pulsed with rapamycin for different time intervals under a variety of mutation supply conditions and compared to non-stressed conditions. Such studies would allow one to determine the impact of cellular fitness as a function of mutational load and would be relevant to understand the progression of disease states accelerated by increased mutation rate.

Materials and Methods

Media. *S. cerevisiae* strains were grown at 30^{o}C in either yeast extract-peptone (YPD) media or minimal selective media (SC; Rose et al., 1990). When required, geneticin (Invitrogen, San Diego) was added at 200 μg/ml (Goldstein and McCusker, 1999).

Strains. S288c background derived yeast strains are listed in Table 4.5, with details regarding their construction available upon request. Briefly, the rapamycin resistant strain EAY4450, derived from EAY1269, was constructed as described by Zhu et al. (2017) for the yeast strain JJY70. Plasmids bearing *MLH1* and *PMS1* derivatives (Table 4.6) were digested with the appropriate restriction enzymes and introduced into EAY4450 using methods described in Rose et al. (1990) and Gietz et al. (1995). *mlh1Δ* derivatives of EAY4450, EAY4488-4490, were

constructed by digesting pEAI160 (*mlh1Δ::KanMX*) with the *Sph*I and *Kpn*I prior to transformation. All integrants were genotyped by PCR using primers that map outside of the restriction sites used for integration, and the presence of specific alleles was confirmed by DNA sequencing.

Plasmids. Plasmids used in this study are listed in Table 4.6. Full details of plasmid constructions and maps are available upon request. Briefly, integration vectors (pEAA672, 674, 675, 713 derived from pEAA213=*MLH1::KanMX*; pEAI453, 454, 455, 468, derived from pEAA238-*PMS1*) containing *FRB* or *FKBP* insertions were constructed using HiFi Gibson cloning (New England Biolabs, Ipswich, MA), with PCR fragments generated from pEAA213 or pEAA238, and gBlocks (Integrated DNA Technologies, Coralville, IA) encoding FRB and FKBP protein domains (Plys et al., 2012; Xu et al., 2010; Geda et al., 2008). The FRB and FKBP domains were inserted immediately after the indicated amino acid position of Mlh1 and Pms1 (Figure 4.1A). *PMS1* integration vectors were constructed through HiFi Gibson cloning by inserting the *LEU2* gene from pRS415 (Christianson et al., 1992) downstream of *PMS1* in the *ARS-CEN* vectors pEAA671, 676, 677, and 678. *MLH1* (pEAE 269, 446, 447, 448) and *PMS1* (pEAE431, 433, 435) protein expression vectors were derived from pMH1 (*GAL1-MLH1-VMA-CBD,2μ, TRP1*) and pMH8 (*GAL10-PMS1,2μ, LEU2*), respectively (Hall and Kunkel, 2001). The DNA sequence of the open reading frames (including 300 bp upstream and 150 bp downstream) of constructs was confirmed by Sanger DNA sequencing (Cornell BioResource Center). The amino acid sequence of the FRB insertions, with glycine and serine linkers shown in bold is:

GSILWHEMWHEGLEEASRLYFGERNVKGMFEVLEPLHAMMERGPQTLKETSFNQAYGRDLMEAQEWCR KYMKSGNVKDLLQAWDLYYHVFRRISK**GS**

Table 4.5. Strains used in this study.

EAY1269	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$*
EAY4209-4211	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A*
EAY4212-4215, 4449	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::KANMX*
EAY4450	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX*
EAY4488-4490	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1Δ::KANMX*
JJY70	*MATa, fpr1Δ::NAT, tor1-1, leu2-3,112, ura3-52, his3-Δ200, trp1-Δ901, suc2-Δ9 lys2-801; GAL*
BJ2168	*MATa, ura3–52, leu2–3, 112, trp1–289, prb1–1122, prc1–407, pep4–3*
***mlh1* integrations**	
EAY4470-4471	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FKBP$_{464}$::KANMX*
EAY4472-4473	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FRB$_{355}$::KANMX*
EAY4474-4475,4491	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FRB$_{464}$::KANMX*
EAY4598-4600	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FRB$_{355}$,FKBP$_{464}$::KANMX*
***pms1* integrations**	
EAY4492-4494, 4504	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, pms1-FRB$_{460}$::LEU2*
EAY4495-4496, 4505	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, pms1-FKBP$_{460}$::LEU2*
EAY4497-4498, 4506	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, pms1-FRB$_{627}$::LEU2*
EAY4499-4500, 4507	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, pms1-FKBP$_{627}$::LEU2*
***mlh1, pms1* integrations**	
EAY4451-4452, 4486	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FKBP$_{464}$::KANMX, pms1-FRB$_{627}$::LEU2*
EAY4453-4454, 4508	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FRB$_{464}$::KANMX, pms1-FKBP$_{460}$::LEU2*
EAY4455-4456, 4509	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FKBP$_{464}$::KANMX, pms1-FRB$_{460}$::LEU2*
EAY4457-4458, 4487	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FRB$_{355}$::KANMX, pms1-FKBP$_{460}$::LEU2*
EAY4501-4503	*MATa, ura3-52, leu2Δ1, trp1Δ63, lys2::insE-A$_{14}$, tor1-1-S1972A, fpr1Δ::NATMX, mlh1-FRB$_{355}$::KANMX, pms1-FKBP$_{627}$::LEU2*

Strains used in this study were derived from the S288c background.

Table 4.6. Plasmids used in this study.

Plasmid	Relevant genotype	Vector type	Source
pRS415	empty vector	*ARS-CEN, LEU2*	Christianson et al., 1992
pRS416	empty vector	*ARS-CEN, URA3*	Christianson et al., 1992
pEAA213	*MLH1::KANMX*	*ARS-CEN, LEU2*	Alani lab
pEAA238	*PMS1*	*ARS-CEN, HIS3*	Alani lab
pEAA67	*MLH1, PMS1, MSH2*	*ARS-CEN, URA3*	Alani lab
pEAA672	*mlh1-FRB$_{355}$::KANMX*	*ARS-CEN, LEU2*	Alani lab; this study
pEAA674	*mlh1-FKBP$_{464}$::KANMX*	*ARS-CEN, LEU2*	Alani lab; this study
pEAA675	*mlh1-FRB$_{464}$::KANMX*	*ARS-CEN, LEU2*	Alani lab; this study
pEAA713	*mlh1-FRB$_{355}$,FKBP$_{464}$::KANMX*	*ARS-CEN, LEU2*	Alani lab; this study
pEAA676	*pms1-FRB$_{460}$*	*ARS-CEN, HIS3*	Alani lab; this study
pEAA677	*pms1-FKBP$_{460}$*	*ARS-CEN, HIS3*	Alani lab; this study
pEAA678	*pms1-FRB$_{627}$*	*ARS-CEN, HIS3*	Alani lab; this study
pEAA671	*pms1-FKBP$_{627}$*	*ARS-CEN, HIS3*	Alani lab; this study
pEAE448	*GAL1-mlh1-FRB$_{355}$(FLAG$_{499}$)-VMA1-CBD*	*2μ, TRP1*	Alani lab; this study
pEAE447	*GAL1-mlh1-FKBP$_{464}$(FLAG$_{499}$)-VMA1-CBD*	*2μ, TRP1*	Alani lab; this study
pEAE446	*GAL1-mlh1-FRB$_{464}$(FLAG$_{499}$)-VMA1-CBD*	*2μ, TRP1*	Alani lab; this study
pEAE435	*GAL10-pms1-FRB$_{460}$*	*2μ, LEU2*	Alani lab; this study
pEAE433	*GAL10-pms1-FKBP$_{460}$*	*2μ, LEU2*	Alani lab; this study
pEAE431	*GAL10-pms1-FRB$_{627}$*	*2μ, LEU2*	Alani lab; this study
pEAI453	*pms1-FRB$_{460}$::LEU2*	*ARS-CEN, LEU2, HIS3*	Alani lab; this study
pEAI454	*pms1-FKBP$_{460}$::LEU2*	*ARS-CEN, LEU2, HIS3*	Alani lab; this study
pEAI455	*pms1-FRB$_{627}$::LEU2*	*ARS-CEN, LEU2, HIS3*	Alani lab; this study
pEAI468	*pms1-FKBP$_{627}$::LEU2*	*ARS-CEN, LEU2, HIS3*	Alani lab; this study
pEAI160	*mlh1Δ::KANMX*	*Integration*	Alani lab
pEAE269	*GAL1-MLH1(FLAG$_{499}$)-VMA1-CBD*	*2μ, TRP1*	Alani lab
pMH8	*GAL10-PMS1*	*2μ, LEU2*	Hall and Kunkel, 2001

Full plasmid descriptions can be found in the Materials and Methods. *mlh1-FRB$_{XXX}$*, *mlh1-FKBP$_{XXX}$*, *pms1-FRB$_{XXX}$*, and *pms1-FKBP$_{XXX}$* refer to the amino acids (XXX) in Mlh1 or Pms1 immediately after which the FRB or FKBP domains were inserted (Figure 4.1; Materials and Methods).

The amino acid sequence of the FKBP insertions, with glycine and serine linkers shown in bold is:

GSGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEVI RGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE**GS** (Spencer et al., 1993; Geda et al., 2008).

***lys2::insE-A₁₄* reversion assay** (**Tables 4.1, 4.2, and 4.3; Figures 4.2, 4.3, and 4.6).** Assays were performed as described previously (Plys et al., 2012; Kim et al., 2019). Briefly, strains listed in Table 4.5 were freshly struck from frozen stocks and grown in synthetic complete media in the presence or absence of rapamycin (2 μg/ml) and then inoculated in liquid complete media maintaining either the presence or absence of rapamycin prior to plating onto complete and lysine dropout plates. Strains containing plasmids were grown in minimal minimal selective leucine dropout, to maintain pEAA213, or uracil dropout, to maintain pRS416 and pEAA67. Rapamycin was included in growth media until just before cell cultures were plated onto complete and lysine dropout plates to measure *lys2::insE-A₁₄* reversion. Rates of *lys2::insE-A₁₄* reversion were calculated as $\mu = f/\ln(N \cdot \mu)$, where f is reversion frequency and N is the total number of revertants in the culture (Tran et al., 1997). For each strain, 15–64 independent cultures, obtained from two to three independent transformants bearing a unique allele, were assayed on at least two different days to prevent batch effects, and 95% confidence intervals and all computer-aided rate calculations were performed as described previously (Dixon and Massey, 1969; Kim et al., 2019). For assays in Table 4.3, the *wild-type* and complex 6 strains were grown both in the presence and absence of rapamycin (2 μg/ml) and then inoculated in liquid complete media maintaining either the presence or absence of rapamycin prior to plating onto

complete and lysine dropout plates. Upon colony counting after 3 days of incubation, a single colony from the complete plate was inoculated in liquid media lacking rapamycin and then assayed again for mutator phenotype.

Cross-linking Mass Spectrometry (XL-MS). Mlh1-Pms1 and Complex 1 and 4 derivatives were crosslinked with disuccinimidyl sulfoxide (DSSO; Thermo Fisher Scientific). A 50 mM stock solution of DSSO was freshly prepared by dissolving DSSO into anhydrous dimethyl sulfoxide (DMSO). 5 μg of Mlh1-Pms1 and Complex 1 and 4 derivatives were incubated in 50 μl of buffer containing 25mM mM HEPES pH 8.0, 180 mM NaCl, 10% glycerol). DSSO was included at a final concentration of 1.25 mM and the reaction was then incubated for 30 min at room temperature. Reactions were quenched by the addition of Tris-HCl pH 8.0 to a final concentration of 10 mM. Samples were digested and processed for MS as described in Yugandhar et al. (2020). In brief, the cross-linked samples were denatured with 1% sodium dodecyl sulfate (SDS) at 65 °C for 15 min, reduced by 1 mM dithiothreitol (DTT) at room temperature for 15 min, and then alkylated with 25 mM iodoacetamide at room temperature for 15 min. Proteins were precipitated by adding 3X volumes of cold acetone/ethanol/acetic acid solution (50:49.9:0.1, v/v/v). The precipitates were resuspended in 8 M Urea, 50 mM Tris-Cl and 150 mM NaCl, pH 8.0. After dilution to 2 M Urea, trypsin (Trypsin Gold, Promega, Madison, WI)) digestion was performed at 37 °C overnight. Trifluoroacetic acid-formic acid (TFA-FA) solution was applied to terminating digestion. The digested peptides were desalted using Sep-Pak C18 cartridge (Waters, Dublin, Ireland), dried using SpeedVac^{TM} Concentrator (Thermo Fisher Scientific, Pittsburgh, PA) and stored in -80 °C for further analysis. Crosslink maps shown are a composite of results from two independent crosslink trials show in Table 4.7.

Wild-type								
Score	#CSMs	z	Peptide A	Protein A id	Location	Peptide B	Protein B id	Location
59.99	1	4	VPkER	Mlh1	C-Term-504	VNVNLTSIkK	Mlh1	C-Term-515
60	1	4	VPkER	Mlh1	C-Term-504	LkSLPLLLK	Mlh1	C-Term-650
60	1	5	VPkER	Mlh1	C-Term-504	LGDYkVPSIADDEKNALPISK	Mlh1	Linker-480
59.89	1	4	VPkER	Mlh1	C-Term-504	LGDYkVPSIADDEK	Mlh1	Linker-480
80	1	4	VNVNLTSIkK	Mlh1	C-Term-515	DANTINDNDLkDQPK	Mlh1	Linker -467
80	1	5	VNVNLTSIkK	Mlh1	C-Term-515	QLSEPkVTNVSHSQEAEK	Mlh1	Linker-434
81	2	4	VNVNLTSIkK	Mlh1	C-Term-515	LGDYkVPSIADDEK	Mlh1	Linker-480
60	1	4	VNVNLTSIkK	Mlh1	C-Term-515	kSISPIFFINNR	Mlh1	N-Term-254
60	1	4	VNVNLTSIkK	Mlh1	C-Term-515	kFEDEILEYNLSTK	Pms1	Linker-605
80	1	4	QENkLVR	Mlh1	Linker-398	IDASQAkITSFLSSSQQFNFEGSSTK	Mlh1	Linker-408
79.95	1	5	QENkLVR	Mlh1	Linker-398	VTNVSHSQEAEkLTLNESEQPR	Mlh1	Linker-446
60	1	4	QENkLVR	Mlh1	Linker-398	FTTSkLQK	Mlh1	N-Term-81
40	1	4	QENkLVR	Mlh1	Linker-398	SISkDNYR	Pms1	Linker-591
60	1	4	ITSFLSSSQQFNFEGSSTkR	Mlh1	Linker-427	QENkLVR	Mlh1	Linker-398
81	2	5	ITSFLSSSQQFNFEGSSTkR	Mlh1	Linker-427	QLSEPkVTNVSHSQEAEK	Mlh1	Linker-434
82	3	4	VTNVSHSQEAEkLTLNESEQPR	Mlh1	Linker-446	DANTINDNDLkDQPK	Mlh1	Linker -467
60	1	4	VTNVSHSQEAEkLTLNESEQPR	Mlh1	Linker-446	QkLGDYK	Mlh1	Linker-475
80	1	4	DANTINDNDLKDQPkKK	Mlh1	Linker-471	LGDYkVPSIADDEK	Mlh1	Linker-480
80	1	4	DANTINDNDLKDQPkK	Mlh1	Linker-471	VPSIADDEkNALPISK	Mlh1	Linker-489
60	1	5	DANTINDNDLKDQPKkK	Mlh1	Linker-472	LGDYkVPSIADDEK	Mlh1	Linker-480
80	1	5	QkLGDYKVPSIADDEK	Mlh1	Linker-475	DANTINDNDLkDQPK	Mlh1	Linker -467
81	2	4	LGDYkVPSIADDEK	Mlh1	Linker-480	DANTINDNDLkDQPK	Mlh1	Linker -467
81	2	5	LGDYkVPSIADDEKNALPISK	Mlh1	Linker-480	DANTINDNDLkDQPK	Mlh1	Linker -467
59.99	1	4	LGDYkVPSIADDEK	Mlh1	Linker-480	QENkLVR	Mlh1	Linker-398
80	1	4	LGDYkVPSIADDEK	Mlh1	Linker-480	VcNLNFISkK	Mlh1	N-Term-253
79.78	1	4	LGDYKVPSIADDEkNALPISK	Mlh1	Linker-489	VPkER	Mlh1	C-Term-504
80	1	4	VPSIADDEkNALPISK	Mlh1	Linker-489	VNVNLTSIkK	Mlh1	C-Term-515
60	1	4	VPSIADDEkNALPISK	Mlh1	Linker-489	QENkLVR	Mlh1	Linker-398
60	1	5	VPSIADDEkNALPISK	Mlh1	Linker-489	QLSEPkVTNVSHSQEAEK	Mlh1	Linker-434
60	1	4	VPSIADDEkNALPISK	Mlh1	Linker-489	QkLGDYK	Mlh1	Linker-475
59.95	1	4	VPSIADDEkNALPISK	Mlh1	Linker-489	kTMTR	Pms1	C-Term-830
60	1	4	VPSIADDEkNALPISK	Mlh1	Linker-489	SISkDNYR	Pms1	Linker-591
80	1	4	VcNLNFISkK	Mlh1	N-Term 253	QENkLVR	Mlh1	Linker-398
80.97	2	4	kSISPIFFINNR	Mlh1	N-Term 254	DANTINDNDLkDQPK	Mlh1	Linker -467
80	1	4	mLESPkPVAGK	Mlh1	N-Term-137	VLQITDNGSGINkADLPILcER	Mlh1	N-Term-67
79.95	1	4	FGDSNYSLSVkPSYTVQDR	Mlh1	N-Term-204	VPkER	Mlh1	C-Term-504
60.95	2	4	FGDSNYSLSVkPSYTVQDR	Mlh1	N-Term-204	DIGFSckK	Mlh1	N-Term-192
83	4	5	IkALDASVVNK	Mlh1	N-Term-6	VSYAEGkMLESPKPVAGK	Mlh1	N-Term 131
81	2	4	IkALDASVVNK	Mlh1	N-Term-6	VSYAEGkmLESPKPVAGK	Mlh1	N-Term 131
60	1	4	IkALDASVVNK	Mlh1	N-Term-6	LQkFEDLSQIQTYGFR	Mlh1	N-Term-84
79.99	1	4	VLQITDNGSGINkADLPILcER	Mlh1	N-Term-67	MLESPkPVAGK	Mlh1	N-Term-137
81	2	5	FTTSkLQK	Mlh1	N-Term-81	VSYAEGkMLESPKPVAGK	Mlh1	N-Term 131
80	1	4	VkLLSLPTSK	Pms1	C-Term-768	LkIDEEEEFGSR	Pms1	C-Term-756
79.86	1	4	SSIMIGkPLNKK	Pms1	C-Term-825	LkTEVFDDR	Pms1	Linker-380
60.94	2	4	kTMTR	Pms1	C-Term-830	VDTSDASLSEDEkAQFINR	Mlh1	C-Term-717
39.29	1	4	kTMTR	Pms1	C-Term-830	SSIMIGkPLNK	Pms1	C-Term-825
60.88	2	4	kTMTR	Pms1	C-Term-830	McSQSEQQAQkR	Pms1	Linker-377
79.95	1	4	QELALPkR	Pms1	Linker-365	kNISSVFGAGGMR	Pms1	N-Term-213
79.67	1	4	QELALPkR	Pms1	Linker-365	kNISSVFGAGGmR	Pms1	N-Term-213
81	2	4	McSQSEQQAQkR	Pms1	Linker-377	kNISSVFGAGGMR	Pms1	N-Term-213
79.95	1	4	mcSQSEQQAQkR	Pms1	Linker-377	kNISSVFGAGGMR	Pms1	N-Term-213
40	1	4	mcSQSEQQAQkR	Pms1	Linker-377	kNISSVFGAGGmR	Pms1	N-Term-213
80	1	4	McSQSEQQAQkR	Pms1	Linker-377	VkGYISQNSFGcGR	Pms1	N-Term-263
84	5	4	LkTEVFDDR	Pms1	Linker-380	McSQSEQQAQkR	Pms1	Linker-377
81	2	4	LkTEVFDDR	Pms1	Linker-380	mcSQSEQQAQkR	Pms1	Linker-377
81	2	4	LkTEVFDDR	Pms1	Linker-380	kNISSVFGAGGmR	Pms1	N-Term-213
80	1	4	LkTEVFDDR	Pms1	Linker-380	kNISSVFGAGGMR	Pms1	N-Term-213
81	2	4	LkTEVFDDR	Pms1	Linker-380	VkGYISQNSFGcGR	Pms1	N-Term-263
60.94	2	4	kLPSIK	Pms1	Linker-468	McSQSEQQAQkR	Pms1	Linker-377
59.86	1	4	VVEEPVYFDIDGEkFQEK	Pms1	Linker-520	QkEFSK	Pms1	N-Term-159
80.99	2	4	TPLkNSR	Pms1	Linker-583	VNVNLTSIkK	Mlh1	C-Term-515
80	1	4	TPLkNSR	Pms1	Linker-583	QENkLVR	Mlh1	Linker-398
79.92	1	4	TPLkNSR	Pms1	Linker-583	VPSIADDEkNALPISK	Mlh1	Linker-489
81	2	4	TPLkNSR	Pms1	Linker-583	SSImIGkPLNK	Pms1	C-Term-825
59.91	1	4	TPLkNSR	Pms1	Linker-583	SSIMIGkPLNK	Pms1	C-Term-825
79.9	1	4	TPLkNSR	Pms1	Linker-583	McSQSEQQAQkR	Pms1	Linker-377
79.88	1	4	TPLkNSR	Pms1	Linker-583	VVEEPVYFDIDGEkFQEK	Pms1	Linker-520
80.99	2	4	TPLkNSR	Pms1	Linker-583	SISkDNYR	Pms1	Linker-591
59.97	1	4	TPLkNSR	Pms1	Linker-583	kFEDEILEYNLSTK	Pms1	Linker-605
59.95	1	4	TPLkNSR	Pms1	Linker-583	NkDELEDFEQGEK	Pms1	Linker-649
79.95	1	4	TPLkNSR	Pms1	Linker-583	IAkFQDVAK	Pms1	N-Term-83
40	1	4	SISkDNYR	Pms1	Linker-591	kFEDEILEYNLSTK	Pms1	Linker-605
39.88	1	4	NGkQMSSIISK	Pms1	Linker-628	NkDELEDFEQGEK	Pms1	Linker-649
79.9	1	4	kSEAQENIIK	Pms1	Linker-638	QmSSIISkR	Pms1	Linker-636
80	1	4	kSEAQENIIK	Pms1	Linker-638	NkDELEDFEQGEK	Pms1	Linker-649
80	1	4	NkDELEDFEQGEK	Pms1	Linker-649	QmSSIISkR	Pms1	Linker-636
60.84	2	4	QkEFSK	Pms1	N-Term-159	FSVWNITPkGK	Pms1	N-Term-196
80.95	2	4	VkGYISQNSFGcGR	Pms1	N-Term-263	kNISSVFGAGGMR	Pms1	N-Term-213
81	2	4	IAkFQDVAK	Pms1	N-Term-83	IkALDASVVNK	Mlh1	N-Term-6

Table 4.7. XL-MS lysine crosslinks. List of crosslinks from MS3 analysis. Lowercase k indicates the lysine residue that was crosslinked. Lowercase C and M mark a modified residue, either carbamidomethylation or oxidation modifications for cysteine and methionine, respectively.

Complex4								
Score	#CSMs	z	Peptide A	Protein_A id	Location	Peptide_B	Protein_B id	Location
79.87	1	4	QENkLVR	Mlh1	Linker-398	VNVNLTSIkK	Mlh1	C-Term-515
79.7	1	4	QENkLVR	Mlh1	Linker-398	FTTSkLQK	Mlh1	N-Term-81
59.87	1	5	QENkLVR	Mlh1	Linker-398	QLSEPkVTNVSHSQEAEK	Mlh1	Linker-434
19.65	1	5	RQENkLVR	Mlh1	Linker-398	QLSEPkVTNVSHSQEAEK	Mlh1	Linker-434
61.86	3	4	IDASQAkITSFLSSSQQFNFEGSSTK	Mlh1	Linker-408	QENkLVR	Mlh1	Linker-398
81	2	5	ITSFLSSSQQFNFEGSSTkR	Mlh1	Linker-427	VTNVSHSQEAEkLTLNESEQPR	Mlh1	Linker-446
29.72	1	5	ITSFLSSSQQFNFEGSSTkR	Mlh1	Linker-427	QLSEPkVTNVSHSQEAEK	Mlh1	Linker-434
60	1	5	DANTINDNDLKDQPkKK	Mlh1	Linker-471	LGDYkVPSIADDEK	Mlh1	Linker-480
59.95	1	4	QkLGDYK	Mlh1	Linker-475	VNVNLTSIkK	Mlh1	C-Term-515
19.69	1	5	QkLGDYK	Mlh1	Linker-475	DANTINDNDLKDQPkKK	Mlh1	Linker-471
60.95	2	4	VPSIADDEkNALPISK	Mlh1	Linker-489	QkLGDYK	Mlh1	Linker-475
87	8	5	VSYAEGkMLESPKPVAGK	Mlh1	N-Term 131	IkALDASVVNK	Mlh1	N-Term-6
82	3	4	VSYAEGkMLESPKPVAGK	Mlh1	N-Term 131	FTTSkLQK	Mlh1	N-Term-81
79.85	1	5	VSYAEGkmLESPKPVAGK	Mlh1	N-Term 131	IkALDASVVNK	Mlh1	N-Term-6
31.63	3	5	YAIHSkDIGFScK	Mlh1	N-Term-185	VEDLNLESVDGkVcNLNFISK	Mlh1	N-Term244
65.94	7	4	FGDSNYSLSVkPSYTVQDR	Mlh1	N-Term-204	DIGFSckK	Mlh1	N-Term-192
39.95	1	4	TVFNkSVASNLITFHISK	Mlh1	N-Term-219	VcNLNFISkK	Mlh1	N-Term 253
81.96	3	4	IkALDASVVNK	Mlh1	N-Term-6	LQkFEDLSQIQTYGFR	Mlh1	N-Term-84
80.86	2	4	IkALDASVVNK	Mlh1	N-Term-6	FTTSkLQK	Mlh1	N-Term-81
82	3	4	VLQITDNGSGINkADLPILcER	Mlh1	N-Term-67	MLESPkPVAGK	Mlh1	N-Term-137
79.87	1	4	FTTSkLQK	Mlh1	N-Term-81	ALDASVVNkIAAGEIIISPVNALK	Mlh1	N-Term-15
19.48	1	4	kTMTR	Pms1	C-Term-830	LGDYkVPSIADDEK	Mlh1	Linker-480
81	2	4	McSQSEQQAQkR	Pms1	Linker-377	VkGYISQNSFGcGR	Pms1	N-Term-263
81	2	4	mcSQSEQQAQkR	Pms1	Linker-377	kNISSVFGAGGMR	Pms1	N-Term-213
80.98	2	4	McSQSEQQAQkR	Pms1	Linker-377	kNISSVFGAGGMR	Pms1	N-Term-213
80	1	4	mcSQSEQQAQkR	Pms1	Linker-377	LkTEVFDDR	Pms1	Linker-380
60.96	2	4	McSQSEQQAQkR	Pms1	Linker-377	kNISSVFGAGGmR	Pms1	N-Term-213
60	1	4	McSQSEQQAQkR	Pms1	Linker-377	QELALPkR	Pms1	Linker-365
81.93	3	4	LkTEVFDDR	Pms1	Linker-380	kNISSVFGAGGMR	Pms1	N-Term-213
79.98	1	4	LkTEVFDDR	Pms1	Linker-380	McSQSEQQAQkR	Pms1	Linker-377
60	1	4	LkTEVFDDR	Pms1	Linker-380	QELALPkR	Pms1	Linker-365
80.9	2	4	TPLkNSR	Pms1	Linker-583	IkALDASVVNK	Mlh1	N-Term-6
79.9	1	5	TPLkNSR	Pms1	Linker-583	VTNVSHSQEAEkLTLNESEQPR	Mlh1	Linker-446
79.79	1	4	TPLkNSR	Pms1	Linker-583	MLESPkPVAGK	Mlh1	N-Term-137
79.77	1	4	TPLkNSR	Pms1	Linker-583	HYTSkIAK	Pms1	N-Term80
79.77	1	4	TPLkNSR	Pms1	Linker-583	VPSIADDEkNALPISK	Mlh1	Linker-489
60.96	2	4	TPLkNSR	Pms1	Linker-583	VNVNLTSIkK	Mlh1	C-Term-515
60.96	2	4	TPLkNSR	Pms1	Linker-583	kFEDEILEYNLSTK	Pms1	Linker-605
60.88	2	4	TPLkNSR	Pms1	Linker-583	EkVDDSIHR	Mlh1	C-Term-520
59.92	1	4	TPLkNSR	Pms1	Linker-583	KSISkDNYR	Pms1	Linker-591
59.9	1	4	TPLkNSR	Pms1	Linker-583	QELALPkR	Pms1	Linker-365
59.89	1	4	TPLkNSR	Pms1	Linker-583	SISkDNYR	Pms1	Linker-591
59.87	1	4	TPLkNSR	Pms1	Linker-583	McSQSEQQAQkR	Pms1	Linker-377
59.85	1	5	TPLkNSR	Pms1	Linker-583	DANTINDNDLKDQPkKK	Mlh1	Linker-471
59.85	1	4	TPLkNSR	Pms1	Linker-583	FTTSkLQK	Mlh1	N-Term-81
59.79	1	4	TPLkNSR	Pms1	Linker-583	RQENkLVR	Mlh1	Linker-398
39.95	1	4	TPLkNSR	Pms1	Linker-583	LkTEVFDDR	Pms1	Linker-380
39.05	1	4	TPLkNSR	Pms1	Linker-583	kTMTR	Pms1	C-Term-830
79.85	1	4	SISkDNYR	Pms1	Linker-591	SSIMIGkPLNK	Pms1	C-Term-825
59.88	1	4	SISkDNYR	Pms1	Linker-591	SLPLLLkGYIPSLVK	Mlh1	C-Term-657
80.89	2	4	kSEAQENIIK	Pms1	Linker-638	NkDELEDFEQGEK	Pms1	Linker-649
79.92	1	4	kSEAQENIIK	Pms1	Linker-638	QMSSIISkR	Pms1	Linker-636
59.95	1	4	kSEAQENIIK	Pms1	Linker-638	QmSSIISkR	Pms1	Linker-636
84	5	4	QFIYVNkRPVEYSTLLK	Pms1	N-Term 287	VkGYISQNSFGcGR	Pms1	N-Term-263
60.82	2	4	QkEFSK	Pms1	N-Term-159	FSVWNITPkGK	Pms1	N-Term-196
40	1	4	MLGkYTDDPDFLDLDYK	Pms1	N-Term-246	IAkFQDVAK	Pms1	N-Term-83
81.95	3	4	HYTSkIAK	Pms1	N-Term-80	FQDVAkVQTLGFR	Pms1	N-Term-89

Protein purification. Mlh1-Pms1 and Complex 1 and 4 variants (Figure 4.9A) were purified as described from galactose-induced cultures of BJ2168 (*MATa, ura3–52, leu2–3,112, trp1–289, prb1–1122, prc1–407, pep4–3*) bearing pEAE expression vectors (Tables 4.5, 4.6; Plys et al., 2012; Hall and Kunkel, 2001). The *MLH1* pEAE672, 674, 675 and 713 expression constructs contain a FLAG tag at position 499 (with respect to the wild-type sequence). RFC and PCNA were expressed and purified from *E. coli* (Cooper et al., 1999; Finkelstein et al., 2003).

DNA substrates for biochemical assays. Closed circular pUC18 (2.7 kb, Thermo Fisher Scientific, Waltham, MA) was used as the DNA substrate in the endonuclease assays presented in Figures 4.9 and 4.10. A 49-mer homoduplex DNA substrate was included in the ATPase experiments presented in Figure 4.11A. This substrate was made by annealing AO3142 (5'GGGTCAACGTGGGCAAAGATGTCCTAGCAAGTCAGAATTCGGTAGCGTG) and AO3144 (5'CACGCTACCGAATTCTGACTTGCTAGGACATCTTTGCCCACGTTGACCC). AO3142 and AO3144 were added at an equal molar ratio in buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM EDTA. These oligonucleotides were annealed through an incubation at 95 °C for 5 min, followed by cooling to 25 °C at a rate of 1 °C/min. Following annealing, excess single-stranded DNA was removed using an S300 spin column (GE Healthcare). For the microscale thermophoresis (MST) analysis presented in Figure 4.11B, a 48-mer homoduplex DNA substrate was made by annealing AO4549 (5'-6-FAM-CTGGACGGGTTAAGACCGAACGTGGCTCCAGAAACGGGTGCAACTGGG; synthesized by Integrated DNA technologies) and AO4548 (5'CCCAGTTGCACCCGTTTCTGGAGCCACGTTCGGTCTTAACCCGTCCAG) as described above.

Endonuclease assays. Reactions were performed in 1X endonuclease buffer (20 mM HEPES-KOH (pH 7.5), 20 mM KCl, 2.5 mM $MnSO_4$, 0.2 mg/mL BSA, and 1 % glycerol; Rogacheva et al., 2014). When indicated 1 μM rapamycin was also included. Rapamycin was dissolved into DMSO at a concentration of 10 mM then serially diluted in DMSO to a final concentration of 100 μM. The 100 μM solution was then diluted in 1x endonuclease reaction buffer to a concentration of 20 μM before being added to individual reaction tubes at a final concentration of 1 μM. Rapamycin was added to the reaction before the addition of the DNA substrate and the reaction was incubated at 37ºC for 5 mins prior to DNA addition. Reactions (37ºC, 40 min) were started following the addition of pUC18 and stopped by the addition (final concentrations shown) of 0.1 % SDS, 14 mM EDTA, and 0.1 mg/ml Proteinase K (New England Biolabs). DNA was electrophoresed in 1.2% agarose gels in 1xTAE (Tris-acetate-EDTA) containing 0.1μg/mL^{-1} ethidium bromide, which causes covalently closed circular DNA isoforms to separate from nicked DNA product. Gels were run in 1x TAE at 100 V for 45 min. Negative control lanes were used as background and were subtracted out of reported quantifications. BioRad Image Lab Software, v5.2.1 was used to quantify gels.

ATPase assays. ATPase activity was determined using the Norit A absorption method described previously (Rogacheva et al., 2014; Kim et al., 2019). Briefly, 30 μl reactions contained 0.4 μM of Mlh1-Pms1, Complex 1, or Complex 4, 100 μM [γ-32P]-ATP, 20 mM Tris, pH 7.5, 2.0 mM $MgCl_2$, 0.1 mM DTT, 1 mM $MnSO_4$, 75 mM NaCl, 1% glycerol, 40 μg/ml BSA. Reactions were incubated for 40 min at 37 °C. When specified, DNA (49-mer homoduplex DNA substrate), PCNA, and rapamycin were included at 0.75 μM, 0.5 μM, and 1 mM, respectively. Just prior to

addition, rapamycin was diluted and included in reactions as described for the endonuclease assays.

Microscale Thermophoresis (MST) assay. MST was performed using a Monolith NT.115 instrument (NanoTemper Technologies) equipped with red and blue filters using methods described in Hosford et al. (2020). Mlh1-Pms1, Complex 1, and Complex 4 were serially diluted 16 times in two-fold steps, and then mixed with a stock of 6-FAM-labeled duplex oligonucleotide (40 nM, sequence described above). Reaction buffer contained 20 mM Tris-HCl (pH 7.5), 20mM NaCl, 0.01 mM EDTA, 2 mM $MgCl_2$, 40 μg mL^{-1} BSA, 0.1 mM DTT and 0.05% TWEEN-20. Assays with nucleotide, rapamycin, or both contained 1 mM ATP and 1 mM rapamycin. ATP and rapamycin were added to reactions containing Mlh1-Pms1 prior to DNA addition. The reactions were then incubated for 5 min at room temperature. Following DNA addition, reactions were incubated at room temperature at 30 ºC for 15 min. They were then loaded into standard capillaries (NanoTemper Technologies). Reactions were then tested with 40% excitation power, medium MST power, and measured using M.O. Control software (NanoTemper Technologies). M.O. Affinity Analysis software (version 2.3, Nanotemper Technologies) was use to analyze data and determine the normalized fluorescence (F_{nom}) for each concentration. F_{norm} is calculated by dividing F_{hot} (average fluorescence value in the heated state) by F_{cold} (average fluorescence value measured in the cold state before the infrared (IR) laser is turned on) and plotted as parts per thousand (%). Three independent reactions were measured to obtain F_{norm} values, which were then averaged (mean ± standard deviation) and plotted against the respective concentration of Mlh1-Pms1 (Figure 4.11B). Binding constants (K_d) were

determined by nonlinear curve fitting using GraphPad Prism 9. All experiments were performed using at least two independently purified proteins.

Acknowledgments

We are grateful to Eric Greene for suggesting the use of the FRB-FKBP dimerization to probe Mlh1-Pms1 functions, and Carol Manhart, Joshua Chappie, Philip Cole, Ilya Finkelstein, Myfanwy Adams, and Jeff Jorgensen for providing their expertise, advice, and use of equipment. C.M.F. and E.A. were supported by the National Institute of General Medical Sciences of the National Institutes of Health (https://www.nih.gov/): R35GM134872, and T.-Y.W., Q.Z. and H.Y. by R01GM124559. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Argueso, J.L., Kijas, A.W., Sarin, S., Heck, J., Waase, M. and Alani, E., 2003. Systematic mutagenesis of the Saccharomyces cerevisiae MLH1 gene reveals distinct roles for Mlh1p in meiotic crossing over and in vegetative and meiotic mismatch repair. *Molecular and Cellular Biology*, *23*(3), pp.873-886.

Ban, C., Junop, M. and Yang, W., 1999. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, *97*(1), pp.85-97.

Banaszynski, L.A., Liu, C.W. and Wandless, T.J., 2005. Characterization of the FKBP⊙ Rapamycin⊙ FRB Ternary Complex. *Journal of the American Chemical Society*, *127*(13), pp.4715-4721.

Bui, D.T., Dine, E., Anderson, J.B., Aquadro, C.F. and Alani, E.E., 2015. A genetic incompatibility accelerates adaptation in yeast. *PLoS Genet*, *11*(7), p.e1005407.

Bui, D.T., Friedrich, A., Al-Sweel, N., Liti, G., Schacherer, J., Aquadro, C.F. and Alani, E., 2017. Mismatch repair incompatibilities in diverse yeast populations. *Genetics*, *205*(4), pp.1459-1471.

Cooper, M.P., Balajee, A.S. and Bohr, V.A., 1999. The C-terminal domain of p21 inhibits nucleotide excision repair in vitro and in vivo. *Molecular biology of the cell*, *10*(7), pp.2119-2129.

Christianson, T.W., Sikorski, R.S., Dante, M., Shero, J.H. and Hieter, P., 1992. Multifunctional yeast high-copy-number shuttle vectors. *Gene*, *110*(1), pp.119-122.

Fegan, A., White, B., Carlson, J.C. and Wagner, C.R., 2010. Chemically controlled protein assembly: techniques and applications. *Chemical reviews*, *110*(6), pp.3315-3336.

Furman, C.M., Elbashir, R. and Alani, E., 2020. Expanded roles for the MutL family of DNA mismatch repair proteins. *Yeast*.

Finkelstein, J., Antony, E., Hingorani, M.M. and O'Donnell, M., 2003. Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy. *Analytical biochemistry*, *319*(1), pp.78-87.

Geda, P., Patury, S., Ma, J., Bharucha, N., Dobry, C.J., Lawson, S.K., Gestwicki, J.E. and Kumar, A., 2008. A small molecule-directed approach to control protein localization and function. *Yeast*, *25*(8), pp.577-594.

Gietz, R.D., Schiestl, R.H., Willems, A.R. and Woods, R.A., 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, *11*(4), pp.355-360.

Goellner, E.M., Smith, C.E., Campbell, C.S., Hombauer, H., Desai, A., Putnam, C.D. and Kolodner, R.D., 2014. PCNA and Msh2-Msh6 activate an Mlh1-Pms1 endonuclease pathway required for Exo1-independent mismatch repair. *Molecular cell*, *55*(2), pp.291-304.

Goellner, E.M., Putnam, C.D. and Kolodner, R.D., 2015. Exonuclease 1-dependent and independent mismatch repair. *DNA repair*, *32*, pp.24-32.

Goldstein, A.L. and McCusker, J.H., 1999. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast*, *15*(14), pp.1541-1553.

Gorman, J., Plys, A.J., Visnapuu, M.L., Alani, E. and Greene, E.C., 2010. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nature structural & molecular biology*, *17*(8), p.932.

Gorman, J., Wang, F., Redding, S., Plys, A.J., Fazio, T., Wind, S., Alani, E.E. and Greene, E.C., 2012. Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proceedings of the National Academy of Sciences*, *109*(45), pp.E3074-E3083.

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. *Molecular cell*, *3*(2), pp.255-261.

Groothuizen, F.S., Winkler, I., Cristóvão, M., Fish, A., Winterwerp, H.H., Reumer, A., Marx, A.D., Hermans, N., Nicholls, R.A., Murshudov, G.N. and Lebbink, J.H., 2015. MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *Elife*, *4*, p.e06744.

Guarné, A., Ramon-Maiques, S., Wolff, E.M., Ghirlando, R., Hu, X., Miller, J.H. and Yang, W., 2004. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *The EMBO journal*, *23*(21), pp.4134-4145.

Hall, M.C. and Kunkel, T.A., 2001. Purification of eukaryotic MutL homologs from Saccharomyces cerevisiae using self-cleaving affinity technology. *Protein expression and purification*, *21*(2), pp.333-342.

Hall, M.C., Wang, H., Erie, D.A. and Kunkel, T.A., 2001. High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer. *Journal of molecular biology*, *312*(4), pp.637-647.

Hall, M.C., Shcherbakova, P.V., Fortune, J.M., Borchers, C.H., Dial, J.M., Tomer, K.B. and Kunkel, T.A., 2003. DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair. *Nucleic acids research*, *31*(8), pp.2025-2034.

Hermans, N., Laffeber, C., Cristovao, M., Artola-Borán, M., Mardenborough, Y., Ikpa, P., Jaddoe, A., Winterwerp, H.H., Wyman, C., Jiricny, J. and Kanaar, R., 2016. Dual daughter strand incision is processive and increases the efficiency of DNA mismatch repair. *Nucleic acids research*, *44*(14), pp.6770-6786.

Hosford, C.J., Adams, M.C., Niu, Y. and Chappie, J.S., 2020. The N-terminal domain of Staphylothermus marinus McrB shares structural homology with PUA-like RNA binding proteins. *Journal of structural biology*, *211*(3), p.107572.

Inobe, T. and Nukina, N., 2016. Rapamycin-induced oligomer formation system of FRB–FKBP fusion proteins. *Journal of bioscience and bioengineering*, *122*(1), pp.40-46.

Kadyrov, F.A., Dzantiev, L., Constantin, N. and Modrich, P., 2006. Endonucleolytic function of MutLα in human mismatch repair. *cell*, *126*(2), pp.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 MutLα is a mismatch repair endonuclease. *Journal of Biological Chemistry*, *282*(51), pp.37181-37190.

Kadyrov, F.A., Genschel, J., Fang, Y., Penland, E., Edelmann, W. and Modrich, P., 2009. A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proceedings of the National Academy of Sciences*, *106*(21), pp.8495-8500.

Karim, A.S., Curran, K.A. and Alper, H.S., 2013. Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *FEMS yeast research*, *13*(1), pp.107-116.

Kawasoe, Y., Tsurimoto, T., Nakagawa, T., Masukata, H. and Takahashi, T.S., 2016. MutSα maintains the mismatch repair capability by inhibiting PCNA unloading. *elife*, *5*, p.e15155.

Kim, Y., Furman, C.M., Manhart, C.M., Alani, E. and Finkelstein, I.J., 2019. Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex. *Nucleic acids research*, *47*(4), pp.1823-1835.

Kunkel, T.A. and Erie, D.A., 2015. Eukaryotic mismatch repair in relation to DNA replication. *Annual review of genetics*, *49*, pp.291-313.

Liu, J., Hanne, J., Britton, B.M., Bennett, J., Kim, D., Lee, J.B. and Fishel, R., 2016. Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, *539*(7630), pp.583-587.

Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K. and Foster, P.L., 2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, *17*(11), p.704.

Mardenborough, Y.S., Nitsenko, K., Laffeber, C., Duboc, C., Sahin, E., Quessada-Vial, A., Winterwerp, H.H., Sixma, T.K., Kanaar, R., Friedhoff, P. and Strick, T.R., 2019. The unstructured linker arms of MutL enable GATC site incision beyond roadblocks during initiation of DNA mismatch repair. *Nucleic acids research*, *47*(22), pp.11667-11680.

Oldfield, C.J. and Dunker, A.K., 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual review of biochemistry*, *83*, pp.553-584.

Pillon, M.C., Lorenowicz, J.J., Uckelmann, M., Klocko, A.D., Mitchell, R.R., Chung, Y.S., Modrich, P., Walker, G.C., Simmons, L.A., Friedhoff, P. and Guarné, A., 2010. Structure of the endonuclease domain of MutL: unlicensed to cut. *Molecular Cell*, *39*(1), pp.145-151.

Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A. and Modrich, P., 2010. PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences*, *107*(37), pp.16066-16071.

Plys, A.J., Rogacheva, M.V., Greene, E.C. and Alani, E., 2012. The unstructured linker arms of Mlh1–Pms1 are important for interactions with DNA during mismatch repair. *Journal of molecular biology*, *422*(2), pp.192-203.

Prolla, T.A., Christie, D.M. and Liskay, R.M., 1994. Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. *Molecular and cellular biology*, *14*(1), pp.407-415.

Raghavan, V., Aquadro, C.F. and Alani, E., 2019. Baker's Yeast Clinical Isolates Provide a Model for How Pathogenic Yeasts Adapt to Stress. *Trends in Genetics*, *35*(11), pp.804-817.

Raghavan, V., Bui, D.T., Al-Sweel, N., Friedrich, A., Schacherer, J., Aquadro, C.F. and Alani, E., 2018. Incompatibilities in mismatch repair genes MLH1-PMS1 contribute to a wide range of mutation rates in human isolates of baker's yeast. *Genetics*, *210*(4), pp.1253-1266.

Räschle, M., Dufner, P., Marra, G. and Jiricny, J., 2002. Mutations within the hMLH1 and hPMS2 subunits of the human MutLα mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSα. *Journal of Biological Chemistry*, *277*(24), pp.21810-21820.

Rogacheva, M.V., Manhart, C.M., Chen, C., Guarne, A., Surtees, J. and Alani, E., 2014. Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *Journal of Biological Chemistry*, *289*(9), pp.5664-5673.

Rose, M.D., Winston, F. and Hieter, P., 1990. Laboratory course manual for methods in yeast genetics. *Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. Open Biology*, *6*(150178), p.14.

Sacho, E.J., Kadyrov, F.A., Modrich, P., Kunkel, T.A. and Erie, D.A., 2008. Direct visualization of asymmetric adenine nucleotide-induced conformational changes in MutLα. *Molecular Cell*, *29*(1), pp.112-121.

Spencer, D.M., Wandless, T.J., Schreiber, S.L. and Crabtree, G.R., 1993. Controlling signal transduction with synthetic ligands. *Science*, *262*(5136), pp.1019-1024.

Smith, C.E., Mendillo, M.L., Bowen, N., Hombauer, H., Campbell, C.S., Desai, A., Putnam, C.D. and Kolodner, R.D., 2013. Dominant mutations in S. cerevisiae PMS1 identify the Mlh1-Pms1 endonuclease active site and an exonuclease 1-independent mismatch repair pathway. *PLoS Genetics*, *9*(10), p.e1003869.

Tran, H.T., Keen, J.D., Kricker, M., Resnick, M.A. and Gordenin, D.A., 1997. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Molecular and cellular biology*, *17*(5), pp.2859-2865.

Tran, P.T. and Liskay, R.M., 2000. Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLα. *Molecular and Cellular Biology*, *20*(17), pp.6390-6398.

Uversky, V.N., 2019. Intrinsically disordered proteins and their "mysterious"(meta) physics. *Frontiers in Physics*, *7*, p.10.

Xu, T., Johnson, C.A., Gestwicki, J.E. and Kumar, A., 2010. Conditionally controlling nuclear trafficking in yeast by chemical-induced protein dimerization. *Nature protocols*, *5*(11), pp.1831-1843.

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T. and Kim, P.M., 2014. Classification of intrinsically disordered regions and proteins. *Chemical reviews*, *114*(13), pp.6589-6631.

Voß, S., Klewer, L. and Wu, Y.W., 2015. Chemically induced dimerization: reversible and spatiotemporal control of protein function in cells. *Current opinion in chemical biology*, *28*, pp.194-201.

Yu, C. and Huang, L., 2018. Cross-linking mass spectrometry: an emerging technology for interactomics and structural biology. *Analytical chemistry*, *90*(1), pp.144-165.

Yugandhar, K., Wang, T.Y., Leung, A.K.Y., Lanz, M.C., Motorykin, I., Liang, J., Shayhidin, E.E., Smolka, M.B., Zhang, S. and Yu, H., 2020. MaXLinker: proteome-wide cross-link identifications with high specificity and sensitivity. *Molecular & Cellular Proteomics*, *19*(3), pp.554-568.

Zhu, L., Jorgensen, J.R., Li, M., Chuang, Y.S. and Emr, S.D., 2017. ESCRTs function directly on the lysosome membrane to downregulate ubiquitinated lysosomal membrane proteins. *elife*, *6*, p.e26403.

CHAPTER 5

Conclusions and Future Directions:

Better understanding the roles of MutL homolog proteins in MMR and meiotic crossing over

Overview

Much has been learned about the roles of the MutL homolog family of proteins. Through the research of many outstanding scientists, we have identified three distinct MutL homolog complexes, Mlh1-Pms1, Mlh1-Mlh2, and Mlh1-Mlh3. In MMR, Mlh1-Pms1 is the major endonuclease responsible for nicking the newly replicated strand in order to allow for efficient repair. Mlh1-Mlh3 plays a minor role, helping to recognize and repair a distinct subset of errors, and Mlh1-Mlh2 plays a minor accessory role. In meiosis, Mlh1-Mlh3 plays a major role as the main Holliday junction resolvase during crossing over, Mlh1-Mlh2 has a minor role in meiosis regulating gene tracts lengths, and Mlh1-Pms1 acts to repair mismatches in heteroduplex DNA that form during genetic recombination (reviewed in Furman et al., 2020). While the field has made great strides genetically and biochemically to understand these processes, the basic understandings of the mechanism of how these MutL homologs proteins are involved in their respective actions is not fully understood.

The work outlined in Chapters 2, 3 and 4 of my thesis present new insights into roles for MLH proteins in mismatch repair and genetic recombination. In Chapter 2, I outlined how we were able to use a combination of computational biology and genetics to understand what amino acids convey the defined specialization of the Mlh3 and Pms1 proteins. In Chapter 3, I showed that the intrinsically disorder linker regions (IDRs) of the Mlh1 and Pms1 proteins are essential for the proteins to function during MMR and that they play a role in how the proteins undergo multiple rounds of a defined conformational change. In Chapter 4, I showed that if the IDRs are constricted using a drug inducible system, one can severely inhibit MMR and that by doing this one can block the ability of Mlh1-Pms1 to interact properly with DNA and possibly interact with

other MMR proteins. In this chapter, I outline future questions and experimental ideas for each of these three projects as well as some for the MMR and meiosis fields as a whole.

Expanded roles in the divergence of MutL homologs

Using a computational Multi-Harmony approach, we were able to identify specific residues that are conserved in the Pms1 family and are either present as a different conserved amino acid in the Mlh3 family or are non-conserved (Chapter 2). I found that a small number of amino acid changes in Mlh3 were able to modestly improve a MMR deficient strain that lacks Pms1. However, I was unable to achieve a full complementation in a *pms1Δ* strain with a novel *mlh3* allele. I hypothesized that these proteins are highly specialized and have significantly diverged from the same ancestral protein. It is likely that many small refinements of both proteins over millions of years have contributed to their specialization; it would be difficult if not impossible to identify these steps using our method which involved analyzing a limited number of fungal MLH proteins. Regardless, are there any ways in which we can improve our approach?

One idea is to see if we can identify an ancestral state/protein that is capable of performing both MMR and meiotic crossing. I would first try using more advanced computation modeling, such as ancestral protein reconstruction or ancestral gene resurrection studies (Thronton, 2004; Merkl & Sterner, 2016; Gumulya & Gillam, 2017) to find the gene sequence that can fully complement a *pms1Δ* strain in a sensitive mutator assay and complement an *mlh3Δ* strain in a meiotic crossing over assay (outlined in Chapter 2). There is some promise for finding such a gene since we were able to show some improvements by changing only a few amino acids in Mlh3 to rescue a *pms1Δ* strain. However, it may prove to be unrealistic in that both Pms1 and Mlh3 do not work alone in their respective systems and instead rely on a multitude of other proteins to help them function properly (Furman et al., 2020).

Another approach is to perform large scale mutational challenge trialls with either protein, similar to how Hsieh et al. (2020) analyzed the kleisin paralogs that act in sister chromosome cohesion in vegetative growth and meiosis. In order to do this, I would conduct a random mutagenesis of *MLH3* expressed on a plasmid vector, and then transform this library into a *pms1Δ pol3-01* strain containing *PMS1* on a *URA3 ARS-CEN* vector. The *pol3-01* allele contains a mutation in the 3' to 5' polymerase delta editing function; this mutation in combination with a *pms1* mismatch repair defect confers lethality due to a catastrophic mutation rate, but the strain is viable because it contains the wild type *PMS1* gene on a plasmid (Morrison et al. 1993; Argueso et al. 2002). By performing a plasmid swap, replacing the *PMS1, ARS-CEN, URA3* vector with a *MLH3* mutagenized library vector, I can then screen for viable strains and identify *MLH3* alleles that can replace some aspect of *PMS1* function. A problem with this approach is that any mutagenesis that we perform might not be sufficient to confer such a radical change in *MLH3* function. Alternatively one could try to identify *PMS1* alleles that could complement *MLH3* meiotic functions; however, we would need to identify a strong selection for meiotic inviability in an *mlh3* background that could be overcome by the *PMS1* allele.

While we were able to use our pipeline to see how Mlh3 could be changed to restore Pms1 function, this pipeline could easily be used to study the reverse. One could simply make the opposite of the mutations that we made in *MLH3*, replacing them in *PMS1*, and then looking via the spore autonomous assay to see if these mutant *PMS1* genes can complement a strain lacking *MLH3* or containing a hypo-morphic *mlh3* allele. Again, this pipeline, which I used in Chapter 2, can easily be done for the remaining MutL homolog proteins Mlh1 and Mlh2. Using the multiple sequence alignments obtained from Nathan Clark's lab, one could look at residues in each of these protein families that seem to be characteristically conserved only in their family.

For Mlh1, this may be challenging since it is lacking an endonuclease domain but since we have the structural information of the endonuclease motif in Pms1 (Gueneau et al., 2013), this could be added in the corresponding position in Mlh1 and favor a homodimer formation, which has been hypothesized to occur (Shcherbakova et al., 2001), thus resulting in an Mlh1 protein that can act by itself as a nuclease in MMR or meiosis. While it would be easy to see residues that are only conserved in the N-terminal ATPase domain of Mlh2 because of the high conservation, the rest of the protein lacks any conservation with the rest of the MutL homologs and thus this method would fail to shed light on interesting amino acids in the C-terminus. Mlh2 would also prove challenging since it has no known enzymatic function and a feasible output for mutational analysis is nonexistent.

This brings me to the final question that can be looked at using our pipeline, which is examining further the N-terminal ATPase region of all four MutL homolog proteins. It has been hypothesized that the N-terminus contains critical residues that are important for each proteins ability to bind a wide variety of DNA substrates (Guarné et al, 2001; Bolz et al., 2012). The question of how each of these complexes bind to DNA and what amino acids are involved is relatively unknown and any information here would be highly important. By examining each homolog and identifying residues in their N-termini that are conserved solely in an individual family, I could ask how mutations of these resides affect binding to a variety of DNA substrates. Such an approach could identify DNA binding pockets for each complex. Another way to investigate this question is to perform protein-DNA crosslinking experiments followed by mass spectrometry to see which residues play a role in this process. One could use oligonucleotides containing a photo-cross linker, such as 5-bromo-dU, combined with purified Mlh1-Pms1 or Mlh1-Mlh3 and induce crosslinking with UV (e.g. Hicke et al., 1994; Manhart & McHenry,

2015). These samples would then be subjected to proteolysis and separated via gel electrophoresis. The resulting bands that show a shift in molecular weight can be excised and examined via mass spectrometry to get the sequence of residues that are crosslinking. In this method, many different substrates with the 5-bromo-dU could be made and analyzed. Any interesting residues can be corroborated in the Multi-Harmony analysis and would be the target for mutagenesis to make sure these are the key amino acids involved in DNA binding. A combination of these two methods (Multi-Harmony and crosslinking) could lead to a better understanding of how Mlh1-Pms1 recognizes and interacts with DNA around a mismatch, as well as how Mlh1-Mlh3 interacts with Holliday junction structures.

Mechanistic approachs to understanding the role of Mlh1-Pms1 linker arms

In Chapter 3, I showed that there are specific sequence and positional requirements for the intrinsically disordered region for Pms1. This mutational analysis also shed light on an interesting combination of deletions, DLD_{null}, that individually conveyed a relatively weak mutator phenotype but in combination conferred a synthetic null defect. This mutant protein was an ideal candidate to ask biochemical mechanistic questions since it seemed that it was on the cusp of functionality. In bulk biochemistry and single molecule experiments in collaboration with the Finkelstein laboratory (Kim et al., 2019), we were able to show that the DLD_{null} lacked the ability to undergo processive rounds of the proposed conformational change that was reported in Sacho et al. (2008). This inability resulted in a reduction of protein motion on DNA curtains, completely obliterating the ability of Mlh1-Pms1 to move on a chromatinized DNA curtain substrate and perform multiple rounds of DNA nicking. While these studies were illuminating, they fell short in addressing the full mechanism that the linker arms play in

mediating conformational changes that occur in MMR, leading to the work performed in Chapter 4.

In Chapter 4, as a follow up to Chapter 3, our goal was to further probe the linker arms role in the conformational change by locking the linkers into a fully closed state using a drug inducible dimerization system with the protein domains FRB and FKBP. We were able to show that limiting the range of motion of either linker arm caused a significant MMR defect. I was also able to show that having these domains on a single linker arm of Mlh1 caused a null phenotype in the presence of rapamycin, further highlighting the idea that the range of motion of these linkers is important for MMR. In order to further investigate these linker arms with the hope of achieving the full mechanism of action they use during MMR, I propose more sophisticated biochemical and structural experiments that could shed light on their role.

One way to achieve more structural information would be to use methods such as small angle x-ray scattering or cryogenic electron microscopy (e.g. Mendillo et al., 2007; R Fernandez-Leiro et al., 2019). While each of these techniques have their challenges, the size of Mlh1-Pms1 is favorable for each approach and both, given the nature of their results, would give potential snap shots of multiple steps in the conformational change at a low, SAXs, and high, resolution that could then be used to assemble a fuller picture of the mode of action for the complex. In each system one could use any of the four FRB-FKBP complexes that I was able to purify in Chapter 4 and perform the experiments in the presence of rapamycin in order to gain information on how the composition and structure of the linker changes when they are locked together. By doing this, one may be forcing the linker into a less flexible or more ordered state that could reveal the structure of the complex when it is actually undergoing its enzymatic activity.

In Chapter 4 we were able to use a reversible cross linker coupled with mass spectrometry (XL-MS) to shed light on where specific lysine residues are interacting with each other at any given time. While we gained important information that showed interesting crosslinks throughout the protein, there remains more to do to understand which sections of the protein are in close proximity to each other in the absence of more complete structural information as the techniques suggested above. One goal in this collaboration, that was not achieved, was to force Mlh1-Pms1 into its fully condensed state in the presence of a non-hydrolyzable nucleotide AMPPNP. The use of this molecule was problematic in our current experimental pipeline. One way to overcome this hurdle, would be to crosslink the proteins in the presence of AMPPNP and then extract the crosslinked peptides from an SDS-PAGE gel, thus removing the nucleotide.

Finally, the last approach I would suggest to further probe how the linker arms are involved in MMR is by the use of *in vitro* reconstitution and single molecule biochemical techniques (e.g. Kim et al., 2019). By using the FRB-FKBP mutant complexes, explained in Chapter 4, in *in vitro* reconstitution assays developed by the Kolodner and Modrich labs (e.g. Bowen and Kolodner, 2017; Constantin et al., 2005) that are dependent on Mlh1-Pms1 function, can allow one to look at the role the linker plays in a more mechanistic approach. The *in vitro* reconstitution assays allow one to add rapamycin along with other MMR component proteins at various steps and ask how the restriction of the linker arm at a given time point affects repair. These assays in concert with the single molecule DNA curtain assays, described in Chapter 3, can highlight the effect of locking the linker arm with rapamycin in terms of protein movement on DNA in general and the ability to bypass single and multiple nucleosomes.

Mutational adaptability using inducible MMR disruption system

In Chapter 4, I describe a mutant allele that has both the FRB and FKBP domains in a single linker arm of Mlh1. Strains bearing this allele are functional for MMR in the absence of rapamycin but become completely defective for MMR in its presence. This observation highlights a potentially useful tool to perform experiments where one can turn on and off MMR at will with a small molecule. One idea is to use this mutant complex in molecular evolution studies similar to those outlined in Bui et al. (2015). In Bui et al. (2015) they examined how fitness in yeast can change over time in incompatible MMR populations compared to control populations when exposed to a stress condition (high salt). They showed that the incompatible population had significant fitness advantage initially that was lost over time. The inducible dimerization of the FRB-FKBP mutant complex would allow one to turn MMR on and off repeatedly and ask similar questions about the fitness of the yeast population. These experiments will be important at understanding further how yeast use mutational modulation in adaptive evolution.

REFERENCES

Argueso JL, Smith D, Yi J, Waase M, Sarin S, Alani E (2002). Analysis of conditional mutations in the Saccharomyces cerevisiae MLH1 gene in mismatch repair and in meiotic crossing over. *Genetics* 160:909-921.

Bolz NJ, Lenhart JS, Weindorf SC and Simmons LA (2012). Residues in the N-terminal domain of MutL required for mismatch repair in Bacillus subtilis. *Journal of bacteriology*, *194*(19), pp.5361-5367.

Bowen N, Kolodner RD (2017). Reconstitution of Saccharomyces cerevisiae DNA polymerase epsilon-dependent mismatch repair with purified proteins. *Proc Natl Acad Sci USA* 114:3607-3612.

Bui DT, Dine E, Anderson JB, Aquadro CF, Alani E (2015). A genetic incompatibility accelerates adaptation in yeast. *PLoS Genetics* 11:e1005407.

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

Fernandez-Leiro, Rafael, et al. (2019). Stepping Through DNA Mismatch Repair Initiation. *Biophysical Journal* 116.3:150a.

Furman C, Elbashir R, Alani E (2020). Expanded roles for the MutL family of DNA mismatch repair proteins. *Yeast* (Special Issue, Exploring the yeast life cycles), In Press.

Guarné A, Junop MS and Yang W (2001). Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *The EMBO journal*, *20*(19), pp.5521-5531.

Gueneau E, Dherin C, Legrand P, Tellier-Lebegue C, Gilquin B, Bonnesoeur P, Londino F, Quemener C, Le Du M H, Márquez J A, Moutiez M, Gondry M, Boiteux S, Charbonnier JB (2013). Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nature Structural and Molecular Biology* 20:461–468.

Gumulya Y, & Gillam EM (2017). Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro'approach to protein engineering. *Biochemical Journal*, *474*(1), 1-19.

Hicke B, Willis M, Koch T and Cech T, (1994). Telomeric Protein-DNA Point Contracts Identified by Photo-Crosslinking Using 5-Bromodeoxyuridine. *Biochemistry*, *33*(11), pp.3364-3373.

Hsieh Y-YP, Makrantoni V, Robertson D, Martson AL, Murray AW (2020). Evolutionary repair: changes in multiple functional modules allow meiotic cohesin to support mitosis. *PLoS Biology* 18:e3000635.

Kim Y, Furman CM, Manhart CM, Alani E, Finkelstein IJ (2019). Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLα mismatch repair complex. *Nucleic Acids Research* 47:1823-1835.

Manhart CM and McHenry CS (2015). Identification of subunit binding positions on a model fork and displacements that occur during sequential assembly of the Escherichia coli primosome. *Journal of Biological Chemistry*, *290*(17), pp.10828-10839.

Mendillo M, Putnam C and Kolodner R, (2007). Escherichia coli MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo. *Journal of Biological Chemistry*, *282*(22), pp.16345-16354.

Merkl R & Sterner R (2016). Ancestral protein reconstruction: techniques and applications. *Biological Chemistry*, *397*(1), 1-21.

Morrison A, Johnson AL, Johnston LH, Sugino A (1993). Pathway correcting DNA replication errors in Saccharomyces cerevisiae. *EMBO Journal* 12:1467-1473.

Sacho EJ, Kadyrov FA, Modrich P, Kunkel TA, Erie DA (2008). Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha. *Molecular Cell* 29:112-121.

Shcherbakova PV, Hall MC, Lewis MS, Bennett SE, Martin KJ, Bushel PR, Afshari CA, Kunkel TA (2001). Inactivation of DNA mismatch repair by increased expression of yeast MLH1. *Molecular and Cellular Biology* 21:940-951.

Thornton JW (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Reviews Genetics*, *5*(5), pp.366-375.