CHARACTERIZING HUMAN MISMATCH REPAIR FACTOR MUTLα

CHARACTERIZING THE STABILITY AND MECHANISM OF HUMAN MISMATCH REPAIR FACTOR MUTLα

By JESSICA WONG, B.Sc.H.

A Thesis Submitted to the School of Graduate Studies
In Partial Fulfillment of the Requirements for
The Degree Master of Science

McMaster University © Copyright by Jessica Wong, August 2011

McMaster University MASTER OF SCIENCE (2011) Hamilton, Ontario (Biochemistry)

TITLE: Characterizing the Stability and Mechanism of Human Mismatch Repair Factor MutLα

AUTHOR: Jessica Wong, B.Sc.H. (Queen's University)
SUPERVISOR: Dr. Alba Guarné

NUMBER OF PAGES: xiii, 78

ABSTRACT

DNA mismatch repair (MMR) is a highly conserved process that is responsible for maintaining genome stability where its main role is repairing replication errors generated by DNA polymerase. Dysfunction in MMR leads to microsatellite instability – the hallmark of Lynch Syndrome, also known as hereditary non-polyposis colorectal cancer. One of the essential proteins in MMR is human MutLalpha (hMutLα), which coordinates critical protein-protein interactions during mismatch recognition, strand removal, and DNA synthesis. It has been recently shown that hMutLα is a latent endonuclease, however it is unclear how the activity of hMutLα is regulated to selectively cleave the error-containing strand. All MutL homologs consist of an N-terminal ATPase domain joined to a constitutively dimerized C-terminal domain by a flexible linker. hMutLα is a heterodimer of hMLH1-hPMS2 with the endonuclease active site located in the C-terminal domain (CTD) of hPMS2. Efforts to structurally characterize hMutLα have revealed the unstable nature of hPMS2. This work presents the characterization of hMutLα through limited proteolysis and thermal denaturation experiments in comparison with stable bacterial MutL homologs. The DNA binding capability of the N-terminal portion of the linker is revealed for the first time. Additionally, we show that the C-terminal domain of hMutLα is capable of cleaving DNA in the absence of other factors under low salt conditions.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Dr. Alba Guarné for all her guidance and support through this project. It has been a truly inspirational and educational experience working with you. I have learned valuable skills that I will take away with me. I would also like to thank my committee members for their guidance and suggestions that helped shape this project. To Dr. Cécile Fradin for her insights during the early stages of this project, Dr. Murray Junop for his enthusiasm and encouragement at every step along the way, and Dr. Richard Epand for his helpful discussions about circular dichroism and interpretations of my strange results.

I would like to give a special thanks to Dr. Raquel Epand for all her help in getting me started with circular dichroism and for her helpful suggestions and discussions, and to Dr. Russell Bishop for offering his insights into this project. Thanks to Monica Pillon for all her support and discussions about mismatch repair and for her purified tagless *B. subtilis* MutL protein that was used in this work, and to Anna Zhou who carried out the endonuclease assays.

I would also like to give a big thank you to all members of the Guarné lab. You have all made this a memorable experience, full of laughs and hard work in the lab. Particularly, I would like to thank Lindsay Matthews for all of her support and encouragement and for our chats about life, to Yu Seon, my bench-mate, for keeping

me entertained in the lab, Melanie Gloyd for all her help and good suggestions, and to Tamiza Nanji for her friendly presence in the lab.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Θ_{MRE} – Mean residue ellipticity

*Aa*MutL - *Aquifex aeolicus* MutL

ATP – Adenosine triphosphate

ATPase – Adenosine triphosphatase

BER – Base excision repair

bp – Base pair

BSA – Bovine serum albumin

*Bs*MutL – *Bacillus subtilis* MutL

CD – Circular dichroism

CTD – C-terminal domain

DLS – Dynamic light scattering

DNA – Deoxyribonucleic acid

DTT – Dithiothreitol

DtxR – Diptheria toxin repressor

E. coli – *Escherichia coli*

*Ec*MutL – *E. coli* MutL

EDTA – Ethylenediaminetetraacetic acid

EMSA – Electrophoretic mobility shift assay

GHL – Gyrase B, Hsp90, MutL

GST – Glutathione S-transferase

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hMutLα – Human MutLα

HNPCC – Hereditary non-polyposis colorectal cancer

HR – Homologous recombination

IDL – Insertion deletion loop

IPTG – Isopropyl β-D-1-thiogalactopyranoside

LDAO – Lauryldimethylamine-oxide

MCS – Multiple cloning site

MMR – Mismatch Repair

*Ng*MutL - *Neisseria gonorrhoeae* MutL

NER – Nucleotide excision repair

NHEJ – Non-homologous end joining

NTD – N-terminal domain

PCNA – Proliferating cell nuclear antigen

PCR – Polymerase chain reaction

RFC – Replication factor C

S. cerevisiae – *Saccharomyces cerevisiae*

SDS – Sodium dodecyl sulfate

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SPR – Surface plasmon resonance

Tris – Tris(hydroxymethyl)aminomethane

TAE – Tris-acetic acid-EDTA

TE – Tris-EDTA

TBE – Tris-borate-EDTA

UV – Ultraviolet

yMutLα – yeast MutLα

CHAPTER 1: INTRODUCTION

1.1 MMR and Genome Stability

DNA replication and maintenance of genome stability are necessary to sustain life. Throughout a cell's lifetime, a variety of sources, including exogenous stresses such as UV radiation and chemical agents or endogenous stresses such as metabolic reactive oxygen species, can damage its DNA [1]. These stresses can result in strand breaks or chemical modification of bases, which signal a number of pathways for repair depending on the type of damage and the stage of the cell cycle, among other factors [1, 2]. These pathways include homologous recombination (HR), non-homologous end joining (NHEJ), base excision repair (BER) and nucleotide excision repair (NER). A cell can also incur damage from DNA replication through mispairing of normal bases. DNA polymerases misinsert nucleotides at a rate of 10^{-3} to 10^{-6} mutations per base per replication, and these misinserted bases can occasionally evade the 3'→5' proofreading activity of the polymerase [3]. This results in the generation of mismatched base pairs or insertion/deletion loops (IDLs). To avoid propagation of mutation through cell progeny, cells rely on the system DNA mismatch repair (MMR), which increases replication fidelity by 100-1000 fold [4].

Dysfunctional MMR leads to increased mutation rates and correlates to an increased risk for cancer development. One responsibility of MMR is to safeguard against mutation in locations where DNA polymerases are prone to strand slippage, and one such location are microsatellite regions. These are areas of short repetitive sequences found across the genome. A defective MMR system is thus susceptible to addition or deletion of nucleotides at these regions [5, 6]. Microsatellite instability is used as a biomarker for MMR-deficiency and correspondingly indicates a high predisposition for Lynch Syndrome, also known as hereditary non-polyposis colorectal cancer (HNPCC) [6, 7].

MMR contributes in other ways to maintaining genome stability. Its protein factors have been found to influence other pathways such as genetic recombination [8] and double strand break repair where it coordinates with factors of HR and NEHJ [9]. MMR proteins also act as sensors of DNA damage induced by methylating agents and inter- or intra-strand crosslinking agents [10-12], and they play a role in signaling apoptosis when the damage is irreversible [13]. MMR is evidently involved in the integrity of the genome, however much remains unclear about its mechanism.

1.2 Mechanism of MMR

MMR is conserved from prokaryotes to eukaryotes. In 1964, mismatch repair was postulated to occur during gene conversion events in genetic recombination [14],

which led to the discovery of a number of *E. coli* genes implicated in this repair pathway and allowed for detailed characterization of the mechanism. Corresponding genes were later identified in *Saccharomyces cerevisiae* (*S. cerevisiae*) [15] and eventually in humans through degenerate PCR, which utilized a mix of primers targeted to the most conserved protein sequences of known homologs [16]. Two distinct mechanisms of MMR have now been elucidated, one that occurs in γ-proteobacteria, exemplified by *E. coli,* and one that occurs in other bacterial organisms and all eukaryotes.

In *E. coli*, mismatches or IDLs are specifically identified and bound by the homodimer MutS (Figure 1). MutS is a member of the ABC ATPase family that recruits the homodimer MutL in its ATP-bound form [17]. MutL is also an ATPase and, in coordination with MutS, activates MutH in response to a mismatch/IDL [18-20]. MutH is a strand specific latent endonuclease that selectively cleaves the newly synthesized daughter strand, which contains the error generated during replication. Unlike its complementary parental strand, the daughter strand is unmethylated following replication, and MutH is able to recognize the unmethylated strand at hemimethylated GATC sites [18, 21, 22]. MutL is also responsible for recruiting UvrD helicase to unwind the DNA [23], which exposes single-stranded regions that are immediately bound by single-stranded DNA-binding proteins (SSBs) [24]. An exonuclease is then recruited for nucleotide excision, either ExoI or ExoX if the cleavage site resides 5’ to the mismatch/IDL, or RecJ or ExoVII if the cleavage resides 3’ to the error, [18, 21]. DNA

polymerase III then resynthesizes the daughter strand using the parental strand as a template [21].

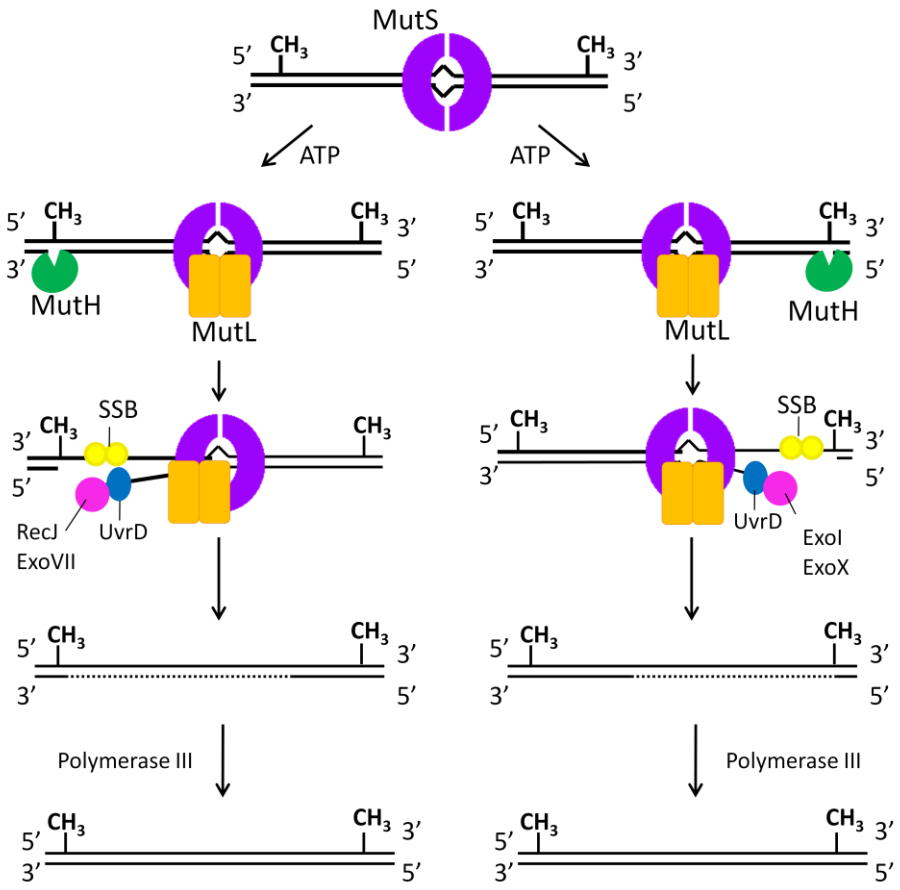

Figure 1. Mismatch repair pathway of *E. coli*.

All other bacterial systems and eukaryotes, which do not exhibit GATC methylation, follow a mechanism with distinct differences from that of the *E. coli* pathway (Figure 2A and B). This mechanism will be described herein using proteins of

the human system. The human system is complicated by having additional homologues of MutS and MutL to form functional heterodimers, as is the case in all eukaryotes, rather than homodimers in the case of all bacteria. hMutSα (heterodimer of hMSH2 and hMSH6) recognizes mismatches and IDLs of 1-2 nucleotides, and hMutSβ (hMSH2 and hMSH3) recognizes larger IDLs up to 16 nucleotides long [18]. hMutLα (heterodimer of hMLH1 and hPMS2) is the main MutL factor involved in MMR [18]. hMutLβ (hMLH1 and hPMS1) does not appear to play a role in MMR [25], while hMutLγ (hMLH1 and hMLH3), has been shown to have a minor role in MMR *in vitro* [26] and has an established role in meiotic recombination [27]. In addition, organisms that follow this pathway do not express MutH or a functional equivalent.

While error recognition by hMutSα or hMutSβ and the recruitment of hMutLα mirror steps in *E. coli* mismatch repair [28, 29], the mechanism of strand discrimination is unknown. In addition, error excision in the human system is reliant on a pre-existing nick on the heteroduplex DNA substrate, while this was not a requirement for mismatch repair in *E. coli* [30, 31]. The nicks are believed to represent the ends of Okazaki fragments *in vivo* and postulated to serve as a strand discrimination signal [32].

Error excision can occur if the pre-existing nick lies 5’ or 3’ to the error, however different proteins are required depending on the position of the nick. hMutSα, the exonuclease ExoI, and single strand binding protein RPA are required when the nick is located 5’ to the error, while hMutLα, the replication processivity clamp proliferating cell nuclear antigen (PCNA), and its clamp loader replication factor C (RFC) are required

in addition when the nick is located 3’ to the error [33, 34]. ExoI is the only exonuclease found to participate in human mismatch repair and has 5’ to 3’ polarity, which poses a problem for excision of error from the 3’ end.

The recent discovery that hMutLα possesses latent endonuclease activity has clarified how excision can proceed from the 3’ end [33]. hMutLα strand cleavage is sequence unspecific and highly biased towards the strand containing the pre-existing nick, and it ensures that ExoI has an entry point for excising the error containing strand [33]. hMutLα has also been shown to interact with ExoI [35] and promote termination of excision beyond the error [36].

The endonuclease activity of hMutLα is dependent on the presence of MutSα, a mismatch, ATP, PCNA, and RFC [33]. It is suspected that PCNA plays a role in directing hMutLα cleavage to the daughter strand by virtue of the specific orientation PCNA is loaded onto DNA [37]. Direct interaction between MutL and the processivity clamp has been shown previously [38], however how strand discrimination is coordinated between PCNA, MutSα, and MutLα remains to be determined .

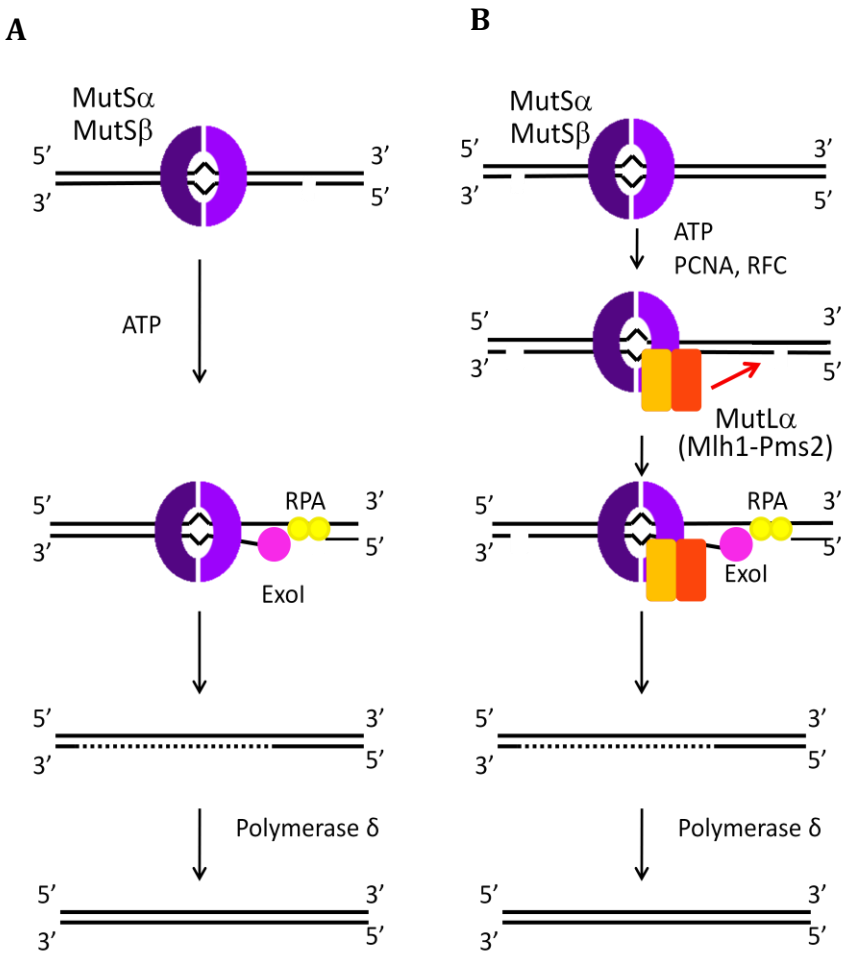

Figure 2. Mismatch repair pathway in humans. A) Repair directed by a 5′ nick, B) repair directed by a 3′ nick.

1.3 Architecture of MutL

To understand the mechanism behind hMutLα activity, many groups have been interested in characterizing the structure of MutL. All MutL homologs consist of a highly conserved N-terminal ATPase domain (NTD), which is a member of the GHL ATPase family [39-41]. Other members of this family include DNA gyrase, Hsp90, type II topoisomerases, and histidine kinases [39]. Proteins in this family share four sequence

motifs that form the ATP binding site [39]. Structures have been solved for the N-terminal domains of *Ec*MutL, hPMS2, hMLH1, hPMS2, and yeast (y) PMS1 (the equivalent of hPMS2) [39-42].

MutL homologs also consist of a constitutively dimerized C-terminal domain (CTD), which is not well conserved in sequence. Alignments of secondary structure predictions reveal a similar order of secondary structure types, with the exception of MLH1 as it contains additional α-helices and unordered regions, which makes the sequences difficult to align (Figure 3) [43-45]. Structures have been solved for the CTD of two MutL homologs that exhibit endonuclease activity, *Bacillus subtilis* (*Bs)* MutL-CTD [44] and *Neisseria gonorrhoeae* (*Ng*) MutL-CTD [46]. The structures of *Ec*MutL-CTD [43] and hMLH1-CTD (PDB: 3RBN), which do not contain endonuclease activity, have also been solved.

The ATPase and dimerization domain are joined by a long, flexible, and variable linker region that is predicted to consist entirely of random coil by secondary structure prediction programs. The linker length can vary among organisms from 100 amino acids in *E. coli* MutL, to approximately 250 amino acids in hPMS2, or to having no linker region at all in *Aquifex aeolicus* (*Aa*) *Aa*MutL. The exact role and purpose of the linker is largely unknown, however it has been shown that mutations within the N-terminal portion of the linker lead to reduced MMR *in vivo* [43, 47].

MutL homologs adopt a large conformational change induced by ATP binding to the N-terminal domain. The conformational change involves N-terminal dimerization,

condensation of the flexible linker region and localization of the N-terminal domains towards the C-terminal domains as depicted in Figure 4 [42, 48]. A gain in secondary structure of hMutLα was reported upon ATP binding as determined by circular dichroism (CD), and this is presumed to take place within the linker region [27].

Figure 3. Secondary structure comparison of MutL homologs. A) Sequence alignment of the C-terminal regions of hPMS2, *Bs*MutL, and *Ec*MutL displayed with secondary structure elements of *Bs*MutL. The conserved motifs found in MutL containing endonuclease activity are highlighted in bright blue, and the conserved motif found in all MutL homologs is heighted in grey. Conserved hydrophobic residues are highlighted in yellow. B) Sequence of hMLH1 displayed with secondary structure elements. (Secondary structure elements of the dimerization subdomain are coloured purple, regulatory subdomain in green, and connecting lever in red.)

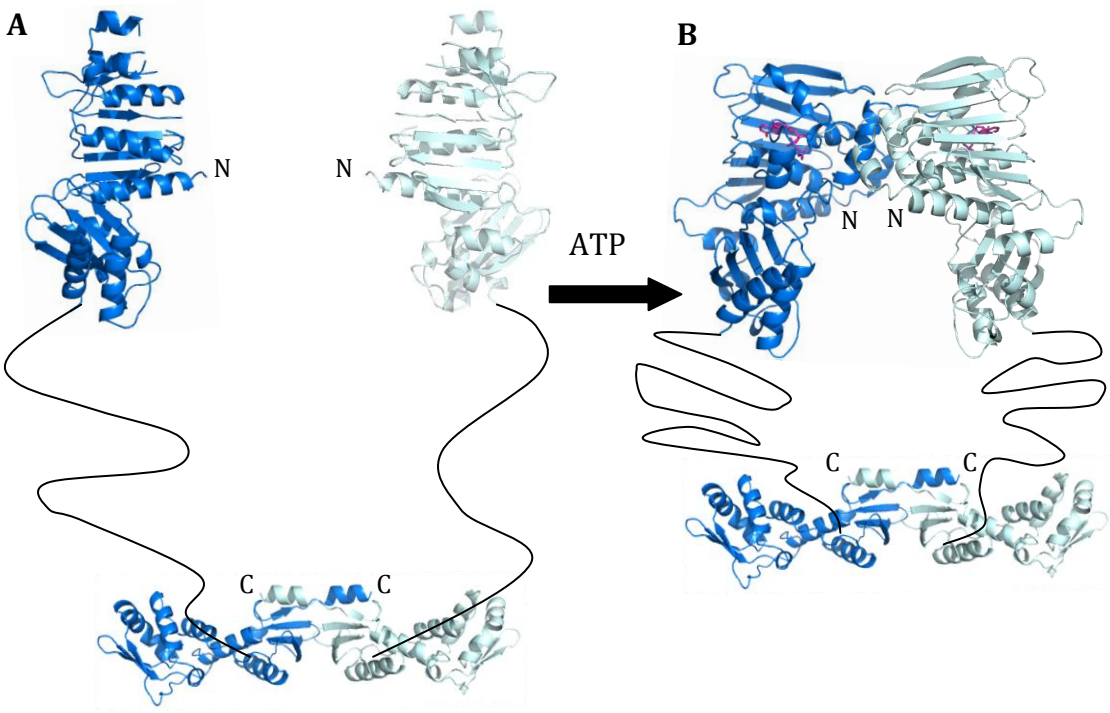

Figure 4. ATP-induced conformational change of MutL displayed using *Ec*MutL domains. A) Apoprotein (PDB ID: 1BKN), B) each monomer is bound to ADPnP coloured in pink (PDB ID: 1B63).

1.4 MutL Endonuclease Active Site

The endonuclease activity of MutL was first discovered in human MutLα where the active site was confined to the C-terminal domain of hPMS2 through the identification of a conserved divalent metal ion binding motif ^{699}DQHA(X)$_2$E(X)$_4$E [33]. Mutation of D699 or E705 within the motif abrogates endonuclease activity and complementation of these mutants into MutLα-deficient nuclear extracts failed to support mismatch repair [33, 45]. E705K is a common mutation associated with Turcot

syndrome, characterized by aggressive types of brain tumors [49], further showing the importance of this motif in hMutLα function for MMR. This motif is found in MutL homologs of organisms that do not express MutH, with the exception of certain eukaryotic MutL paralogs, such as MLH1 and PMS1 in humans.

The C-terminal domain of *Bs*MutL consists of two subdomains, the regulatory subdomain and dimerization subdomain, which are connected by an α-helix termed the lever [44]. The lever is formed by the metal binding motif [44] (indicated in Figure 5A). Four other conserved motifs (ACR, C[P/N]HGRP, FXR, and GQ) were identified in MutL homologs that exhibit endonuclease activity [45]. ACR, C[P/N]HGRP, and FXR are situated around the lever and together form the endonuclease active site [44, 45] (highlighted in Figure 3 and 5). The GQ motif is not situated with the other motifs, but is believed to contribute indirectly to the stability of the active site [44]. A fifth motif has also been identified (QXLLXP) and is conserved in all MutL homologs.

The same organization of motifs is apparent in the C-terminal domain of *Ng*MutL shown in Figure 5B. The C-terminal domain of *Ec*MutL exhibits similar topology to *Bs*MutL and *Ng*MutL, however the organization of secondary structures surrounding the lever prevent formation of the active site [44]. The most notable difference is that the loop formed by the CPHGRP motif in *Bs*MutL contains an additional α-helix in *Ec*MutL and is not localized at the lever [44](Figure 5C). The C-terminal domain of hMLH1 is significantly different from the other homologs. It lacks the division of subdomains and access to the lever is occluded by the N-terminus of the protein (Figure 5D).

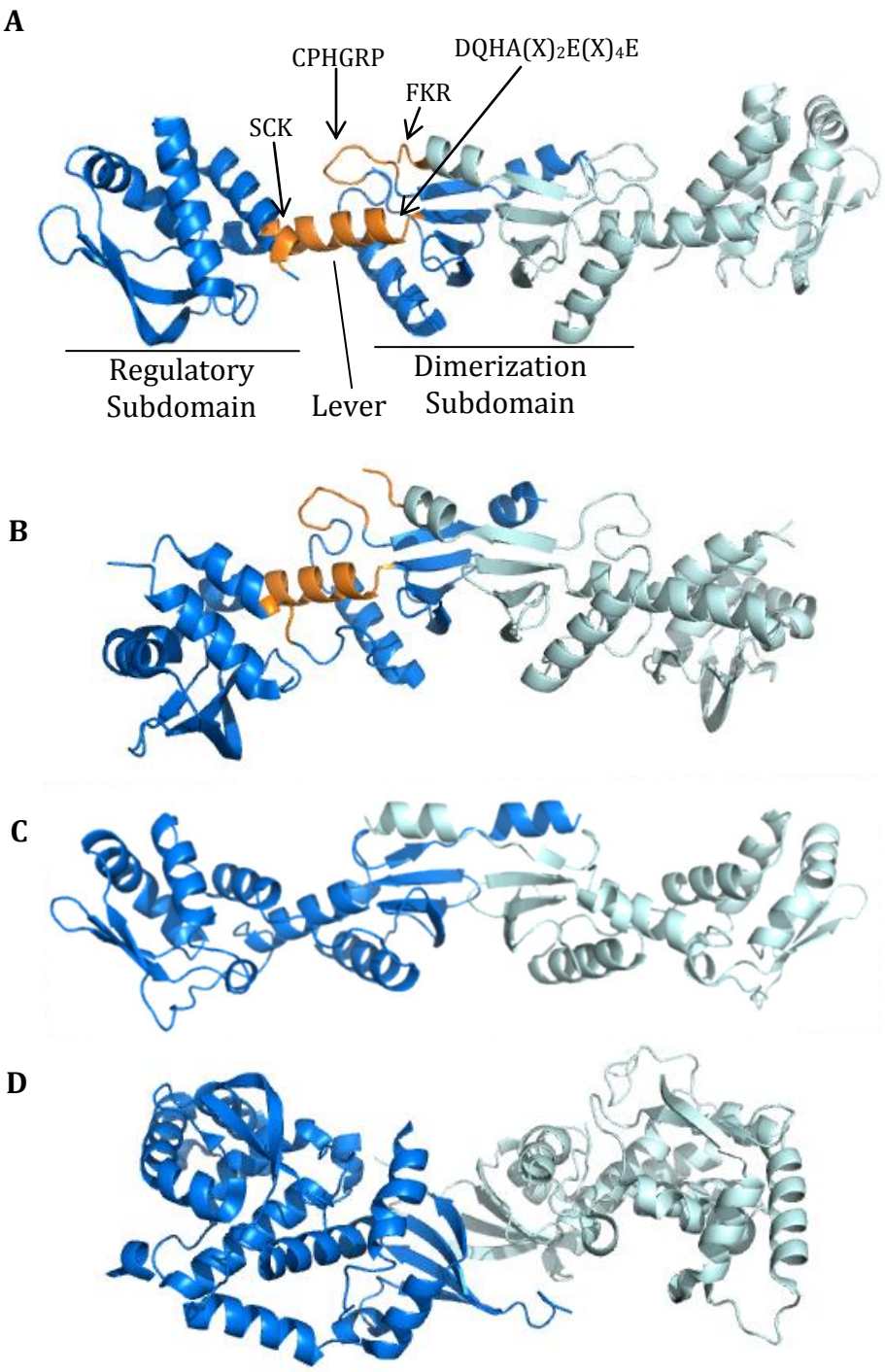

Figure 5. X-ray crystal structures of the C-terminal domain of A) *Bs*MutL (PDB ID: 3KDK), C) *Ng*MutL (PDB ID: 3NCV), C) *Ec*MutL (PDB ID: 1X9Z), and D) hMLH1 (PDB ID: 3RBN). One protomer is shown in blue and the other in light blue. Conserved motifs that define the endonuclease active site are shown in orange.

1.5 Regulatory Role of Zn^{2+} in MutL

Two of the motifs in the endonuclease active site define a Zn^{2+} binding site in *Bs*MutL [44]. Mutation of E468 and H606 from motifs $^{462}DQHA(X)_2E(X)_4E$ and $^{604}CPHGRP$ respectively in *Bs*MutL, prevent Zn^{2+} binding and also increase mutation frequency *in vivo* [44]. C604 from the CPHGRP motif is also important for Zn^{2+} binding, but was not tested for its role *in vivo* [44]. hMutLα was found to bind one zinc metal ion per dimer, and it is believed that hPMS2 and not hMLH1 binds Zn^{2+} [45]. Additionally, mutation of the equivalent Zn^{2+} binding residues in hPMS2 strongly reduce mismatch repair activity [45].

Through computational analyses, the C-terminal domain of hMutLα was predicted to contain a regulatory metal ion binding site similar of that seen in proteins of the Diptheria toxin repressor (DtxR) family [45]. More recently, a member from this family, ScaR, was found to contain a metal ion binding site that differed from other members of the DtxR family, but was identical to the Zn^{2+} binding site of *Bs*MutL [44, 50]. ScaR is a manganese-dependent repressor protein that is allosterically activated to bind DNA by manganese ions, however the function of the regulatory ion is unknown [50].

Zn^{2+} was captured in the crystal structure of *Bs*MutL and shown to lock the conformation between the dimerization and regulatory domain, perhaps a conformation that is optimal for endonucleolytic cleavage [44]. The Zn^{2+}-induced

conformational change has only been shown in *Bs*MutL, but it is likely to have a similar effect on other MutL homologs that contain endonuclease activity.

1.6 MutL Endonuclease Activity

Modrich *et al.* demonstrated that MutLα nicks heteroduplex DNA in the presence of ATP, MutSα, PCNA, RFC, a mismatch, and a pre-existing nick under physiological conditions [33, 51]. In the absence of other factors, MutL is able to nick homoduplex DNA, however the activity is weak and only detectable under low salt conditions [33, 44, 46, 51, 52].

MutL endonuclease activation is dependent on a metal ion, however there are differences in metal ion preference amongst MutL homologs. MutLα and BsMutL are exclusively dependent on Mn^{2+} [33, 44, 51], while *Ng*MutL is activated by Mn^{2+}, Mg^{2+}, and Ca^{2+} [46], and both *Aa*MutL and *Thermus thermophilus* MutL are activated by Mn^{2+}, Ni^{2+}, and Co^{2+} [53, 54]. Although Zn^{2+} is not a catalytic ion, it increases endonuclease activity of *Bs*MutL and *Aa*MutL in endonuclease reactions already containing Mn^{2+} [44, 53].

Activity is stimulated by ATP binding and not hydrolysis, suggesting that the ATP-induced conformational change is important for endonuclease activity [33, 44, 51, 52]. However, weak activity has been detected from the C-terminal domain of *Ng*MutL and *Aa*MutL in the absence of the N-terminal domain. Endonuclease activity was not

detected from the C-terminal domain of *Bs*MutL, though this was attributed to the inability of *Bs*MutL-CTD to bind DNA. The activity of the C-terminal domain of hMutLα has not been tested previously.

1.7 MutL DNA Binding

MutL binds DNA in a sequence unspecific manner [55, 56], and is important for MMR [20, 57, 58]. In *E. coli*, DNA binding of MutL activates UvrD and is proposed to determine how UvrD is loaded onto DNA [43]. DNA binding is also necessary for MutL endonuclease activity.

The structure of the dimerized N-terminal domain of *Ec*MutL revealed a positively charged cleft between the two subunits [42]. R266 is located in this cleft and is crucial for DNA binding and MMR *in vivo* [20, 42]. The positive groove is conserved amongst MutL homologs and equivalent residues in yMutLα have been identified. R274 in yMLH1 is important for DNA binding and mutation in this residue displayed a strong mutator phenotype [57]. The recent structure of the N-terminal domain of yPMS1 has identified K197 and R198 within the positive cleft, and these residues are important for DNA binding [41, 59]

N-terminal dimerization creates a central cavity in the structure of MutL with the C-terminal domain (Figure 4). However, the contribution of the C-terminal domain of MutL to DNA binding remains unclear. The C-terminal domain of *Ec*MutL does not bind

DNA, however full-length protein binds with higher affinity compared to the N-terminal domain alone [43]. A similar observation was seen with yMutLα [57]. On the other hand, the C-terminal domain of *Ng*MutL and *Aa*MutL both bind DNA [53, 60]. The DNA binding capability of the C-terminal domain of hMutLα has not been tested previously. If the C-terminal domain is able to bind DNA, there must be a mechanism to prevent unwanted nicking of DNA. Clarification of the DNA binding ability of the C-terminal domain is necessary.

1.8 Stability of hPMS2

The studies conducted on bacterial MutL have contributed to our knowledge and understanding of hMutLα, however the complexity of the human system and differences between the eukaryotic and bacterial homologs must be taken into consideration. The nature of hMutLα is very different from that of its bacterial homologs. The first piece of evidence supporting this is the fact that bacterial MutL homologs exist as homodimers, and those that have endonuclease activity contain active sites in both protomers, while it is only contained in hPMS2 in the case of hMutLα. Secondly, stable *in vivo* expression of hPMS2 and functional MMR is dependent on hMLH1 [61]. A number of hMLH1 mutations have also been identified in Lynch Syndrome patients, which result in the impairment of dimerization with hPMS2, underscoring the importance of this interaction [61-64].

There is currently contradictory data regarding the dimerization boundaries for hMLH1 and hPMS2 (summarized in Table 1, Figure 6). Guerrette *et al.* identified the dimerization boundary to be hMLH1 506-675 and hPMS2 675-850 using a GST-pull down assay [62]. Using the same assay complemented with yeast two hybrid experiments, Kondo *et al*. later suggested that dimerization occurs between hMLH1 492-742 and hPMS2 612-674 [65]. Based on secondary structure prediction programs, boundaries of the minimal folded region of the CTD are hMLH1 498-756 and hPMS2 672-862.

Table 1. Dimerization boundary of hMLH1 and hPMS2 reported through different methods.

hMLH1	**hPMS2**	**Type of Experiment**	**Reference**
506-675	675-850	GST-*in vitro* transcription and translation pull down assay	Guerrette *et al.*
492-742	612-674	GST-*in vitro* transcription and translation pull down assay, Yeast two hybrid	Kondo *et al*.
498-756	672-862	Secondary structure predictions	PSIPRED
475-756	600-862	Recombinant protein expression	(our unpublished data)

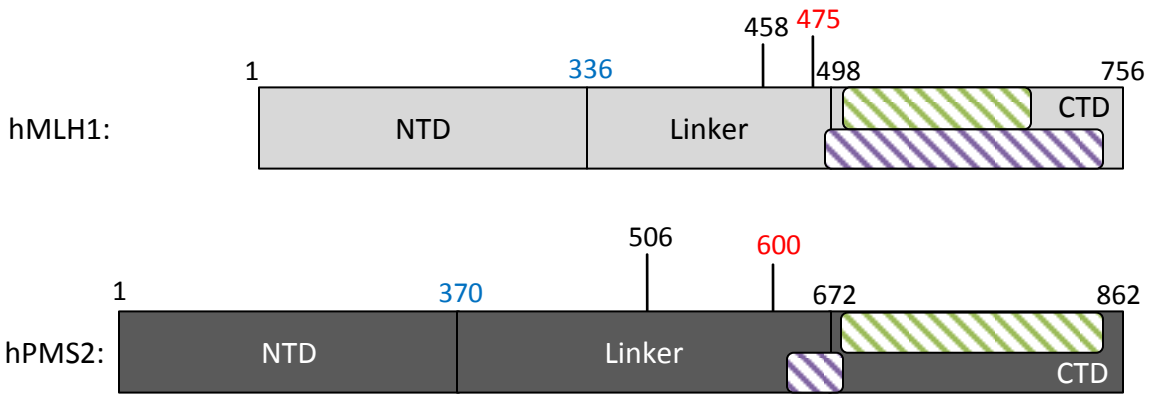

Figure 6. Construct boundaries of hMutLα. Diagonal green regions correspond to dimerization boundaries reported by Guerrette *et al.* [39] and diagonal purple regions correspond to that suggested by Kondo *et al*. [42]. Red text indicates the starting residue for hMutLα-CTD (hMLH1 475-756, hPMS2 600-862), expressing the minimal region required for successful formation of the heterodimer. Blue text indicates the starting residue for hMutLα (hMLH1 336-756, hPMS2 370-862), expressing the CTD with the entire linker region.

This is in contrary to what has been observed previously in our lab. Firstly, recombinant expression of hMutLα-CTD in an *E. coli* expression system requires co-expression of hMLH1 and hPMS2. hMLH1 498-756 was found to be insufficient to support dimerization with any length of hPMS2 and consequently results in formation of hMLH1 homodimers. hMLH1 must be expressed at a minimum length of 475-756 to stabilize the heterodimer and prevent degradation of hPMS2 by cellular proteases. Accordingly, hPMS2 must be expressed at a minimum length of 600-862 to support formation of the heterodimer. The boundaries for successful expression and purification of the heterodimer thus require a significant portion of the presumably

unstructured linker region of both hMLH1 and hPMS2. These observations differ from what has seen from bacterial homologs. The C-terminal domain of *Ec*MutL, *Bs*MutL, and *Ng*MutL were all stably expressed and crystallized with the minimal folded region, without the inclusion of linker [43, 44, 46].

Despite successful formation of the heterodimer, previous work in the lab has shown that hMutLα-CTD (hMLH1 475-756, hPMS2 600-862) is still highly unstable (unpublished data). This led us to hypothesize that the linker region may be functioning as part of the CTD of hMutLα, even though it is predicted to be random coil. This work is aimed at determining whether there is a stabilizing effect when expressing the linker beyond the minimal boundaries for successful heterodimer formation. Characterizing the stabilizing effect of the linker could shine some light on the function of hMutLα activity since these restrictions are not observed in the bacterial homologs.

1.9 Thesis Objective

1) To characterize the stability of the C-terminal domains of hMLH1 and hPMS2 relative to bacterial homologs and establish the most stable boundaries of hMutLα-CTD;
2) To evaluate the DNA binding ability and endonuclease activity of hMutLα-CTD and linker region.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cloning of MutL Variants

*Ec*MutL-CTD (432-615), *Bs*MutL-CTD (433-627), hMLH1 (498-756), and hMLH1 (475-756) were cloned previously in the lab [43, 44][unpublished data]. Variants of hMutLα encoding the C-terminal domain with different lengths of the linker were generated by subcloning fragments of hMLH1 and hPMS2 into a modified pET15b vector (Table 2 and Figure 7). pET15b co-expression vectors encoding hMutLα minimal linker and hMutLα half linker1 were cloned previously in the lab [unpublished data]. These hMutLα constructs encode the hPMS2 gene variant between NdeI and BamHI restriction sites within the expression region of pET15b, downstream of a removable N-terminal His-tag coding region. A second expression region with a separate promoter encodes the hMLH1 gene variant and lacks the His-tag. This expression region is located in the PshAI restriction site as shown in the last step of Figure 7.

Table 2. Residue boundaries of hMLH1 and hPMS2 for to each hMutLα variant.

hMutLα Variant	hMLH1 Residues	hPMS2 Residues	Plasmid ID
minimal linker	475-756	600-862	pAG 8036
half linker1	475-756	506-862	pAG 1348
half linker2	458-756	506-862	pAG 8385
full linker	336-756	370-862	pAG 8519

Co-expression vectors of hMutLα half linker2 and hMutLα full linker were constructed in a number of steps as outlined in Figure 7. First, MLH1 variants were amplified by PCR from a plasmid encoding the complete MLH1 gene using primers 1 and 2 (Table 3), then ligated into a pET15b vector between NcoI and BamHI restriction sites, removing the N-terminal His-tag coding region in the process. A second PCR was performed on the ligated product using primers 3 and 4 (Table 3), which flanked the pET15b expression region and the 3' end of the MLH1 gene. These primers were designed to add PshAI sites at either end. The second PCR product was then subcloned into the co-expression vector of hMutLα-CTD minimal linker in place of the original hMLH1 gene variant using PshAI restriction sites. This was followed by ligation of the desired hPMS2 gene variant in place of the original hPMS2 sequence between NdeI and BamHI restriction sites. All clones were verified by DNA sequencing (MOBIX, McMaster University, Ontario).

Table 3. Sequences of primers used for cloning hMutLα constructs.

	Primer Sequence (5' → 3')	**Restriction Site***	**Primer #**
1	TACCATGGCAGAGAAGAGAGGACCTACTTC	NcoI	ag1408
2	AAGGATCCCTCGAGTTAACACCTCTCAAAGAC	BamHI	ag1410
3	AAGACTATCGTCAGATCTCGATCCCGCGAAATTAATACG	PshAI	ag1411
4	AAGACGATAGTCCTCGAGTTAACACCTCTCAAAG	PshAI	ag1412

*Restriction sites are underlined

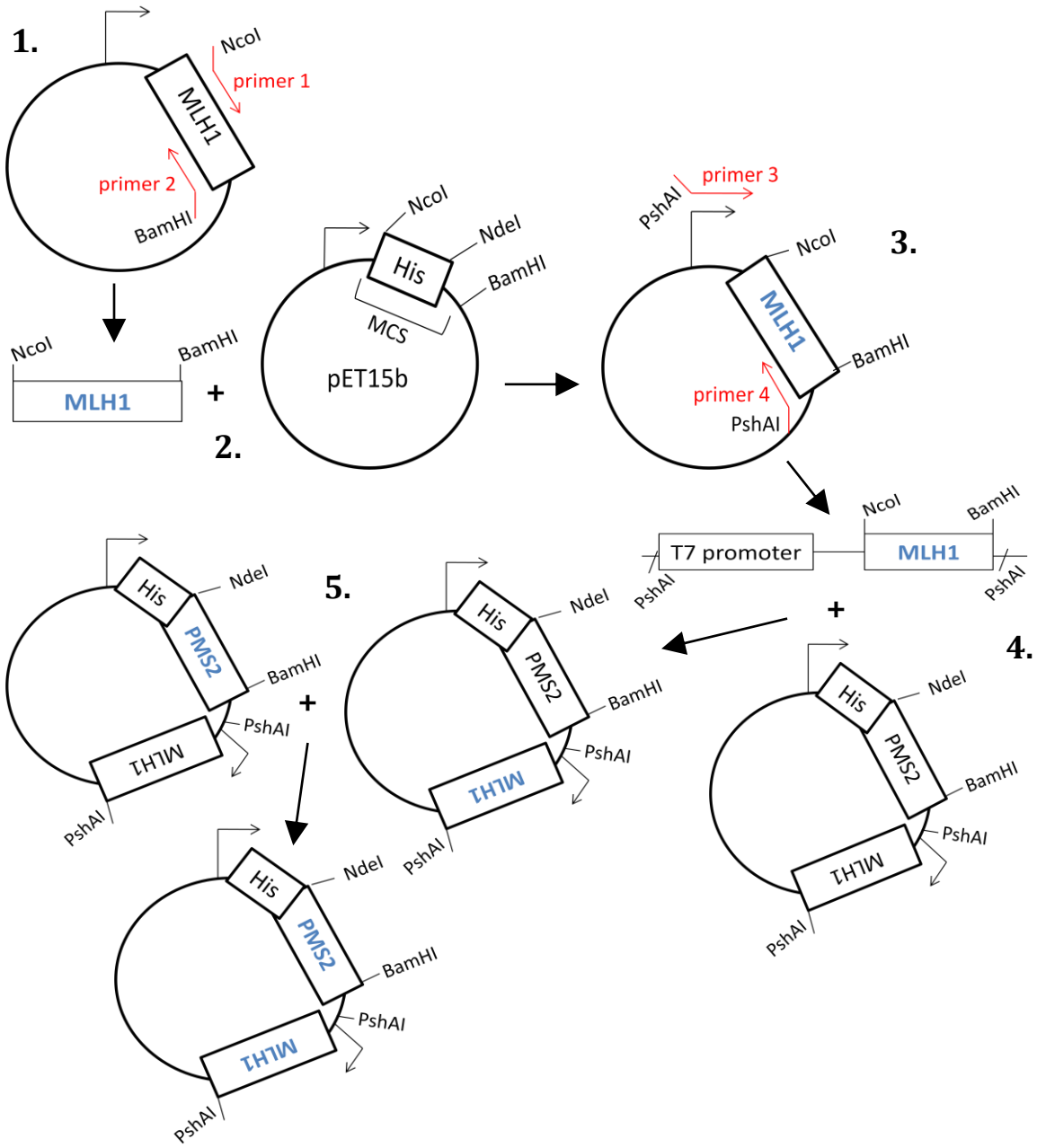

Figure 7. Cloning strategy for co-expression of hMLH1 and hPMS2 gene variants. (desired fragments are coloured blue)

1. Amplify desired fragment of hMLH1 from the complete gene
2. Ligate into a pET15b vector
3. Amplify hMLH1 gene including expression region
4. Ligate into destination vector (previously made) in place of original MLH1 gene
5. Cut and paste desired fragment of PMS2 (previously made) gene into destination vector

2.2 Solubility Assay

hMutLα half linker2 and hMutLα full linker were transformed into BL21(DE3) and BL21 Star (DE3) cells respectively. Both cell lines had been previously transformed with a pRareLysS plasmid. Cells were grown in LB media at 37 °C to an absorbance of 0.7 at 600 nm before being induced with 0.5 mM IPTG, and then expressed at the indicated times and temperatures. Samples were taken at each time point and incubated with 1 mg/mL lysozyme for 30 min, followed by incubation with 90 mM KCl, 10 mM $MgCl_2$, and 0.05% LDAO for 15 min. Cells were frozen and thawed at -80 °C to promote cell lysis, incubated with 10 µL of stock Dnase I (New England Biolabs), and then spun at 13,000 rpm for 20 min to isolate the soluble fraction. Samples were loaded onto 12% sodium dodecyl sulfate (SDS) polyacrylamide gels and subjected to electrophoresis at 150 V for 80 min in Tris-glycine-SDS buffer. Gels were stained with Coomassie Blue.

2.3 Protein Production and Purification

The C-terminal domains of *Ec*MutL and *Bs*MutL were expressed and purified as described previously [43, 44]. hMLH1 constructs and hMutLα half linker2 were transformed into BL21(DE3) pRareLysS cells and all other hMutLα constructs were transformed into BL21(DE3) Star pRareLysS cells. Cells were grown in LB media at 37 °C until an absorbance of 0.7 at 600 nm was reached. Cells were then cold shocked on ice,

induced with 0.5 mM IPTG, and incubated at 25 °C for 5 hours. All hMLH1 and hMutLα constructs were purified using a Ni^{2+}-chelating affinity column (GE Healthcare) equilibrated in buffer A (20 mM Tris pH 8, 0.5 M KCl, 1.4 mM β-mercaptoethanol, 5% glycerol, 100 mM PMSF) with 30 mM imidazole. After loading sample onto the column, the column was washed with 100 mL of buffer A with 30 mM imidazole, then eluted using buffer A with 300 mM imidazole into fractions containing a final concentration of 31 μM PMSF, 0.73 μM leupeptin, 0.16 μM pepstatin A, and 833 μM EDTA. Further purification was performed using an ion-exchange column (Mono-Q 5/50, GE Healthcare) equilibrated with buffer B (20 mM Tris pH 8, 0.5 M EDTA, 10 mM DTT, 5% glycerol, and 125 mM KCl) and eluted using a linear gradient to 400 mM KCl. An additional gel filtration purification step was implemented to exchange the buffer to storage buffer (20 mM Tris pH 8, 1.4 mM β-mercaptoethanol, 10% glycerol, and 100 mM KCl) or buffer for CD experiments (20 mM Tris pH 8, 1.4 mM β-mercaptoethanol, 10% glycerol, and 400 mM KCl).

2.4 Circular Dichroism (CD) Spectroscopy

Far UV spectra were obtained using the AVIV 410 CD spectrometer (AVIV Biomedical Inc.) and a cuvette of 1.0 mm path length. Spectra were recorded at 1 nm intervals at a wavelength range of 195 to 260 nm. Protein concentration was calculated from A_{280} readings using the Beer-Lambert law ($A_{280} = \varepsilon lc$) and diluted to 0.25 mg/mL.

Resulting spectra were corrected for buffer signals by subtracting buffer alone signals and converted to mean residue ellipticity (θ_{MRE}) units following the equation $\theta_{MRE} = (\theta \cdot M_w)/(10 \cdot n_r \cdot d \cdot c)$ where θ is the measured ellipticity, M_w is the molecular weight, n_r is the number of residues, d is the path length in cm, and c is the concentration in g/mL.

Secondary structure composition of MutL constructs were predicted from CD spectra using CD Pro software, which provides estimates from a number of programs, each of which uses a unique computational method. Programs SELCON3 and CONTINLL were used in these experiments and an average was taken from the two [66]. IBasis10 reference set, a compilation of spectra obtained for 43 soluble proteins and 13 membrane proteins of known X-ray crystal structure, was used for program computation and prediction.

Thermal denaturation experiments were carried out at 2 °C temperature steps with a 30 second incubation time at each step and a 3 second averaging time. Readings were collected at 222 nm to assess unfolding of α-helices.

2.5 Limited proteolysis

Purified protein was diluted to 0.5 mg/mL with storage buffer and incubated with a 548 bp oligonucleotide (see section below for preparation of the oligonucleotide) at a 100:1 (protein:DNA) molar ratio or with $ZnCl_2$ at a 1:1 molar ratio in a volume of 9 μL. Protein was then incubated with 1 μL of 50 mM $MgCl_2$ and 1 μL of trypsin. Trypsin

was serially diluted from a 0.5 mg/mL stock in Tris-EDTA (TE) buffer pH 7.4 and concentrations are as indicated in figure legends. The reaction was incubated for 30 minutes at room temperature. Digested products were visualized on both SDS-polyacrylamide gels and Tris-borate-EDTA (TBE) polyacrylamide gels stained with Coomassie blue.

2.6 Electrophoretic Mobility Shift Assay (EMSA)

DNA binding ability was assessed using two DNA substrates. Unmodified 113 bp double-stranded DNA substrate was kindly provided by Yu Seon Chung. 548 bp double-stranded DNA substrate was digested from a plasmid and gel extracted and purified using the Qiagen II Gel Extraction Kit. Sequences can be found in the Appendix. Purified protein was first diluted to 32 μM then serially diluted down to 1 μM in reaction buffer, which consists of 20 mM HEPES pH 7.5, 20 mM KCl, 0.2 mg/mL BSA, and 5% glycerol. Reactions were set with 2 μL of protein and 10 nM of DNA in a total volume of 20 μL. Reactions were incubated at either 22 °C or 30 °C for 30 minutes as indicated in figure legends, then analyzed on a 4.5% or 6% TBE polyacrylamide gel stained with SYBR Green I (Invitrogen) and visualized using the Typhoon 9200 Imager (Molecular Dynamics).

2.7 Endonuclease Assay

Endonuclease Assays were performed as previously described [44]. Protein was diluted to 100 nM in buffer consisting of 20 mM HEPES pH 7.5, 20 mM KCl, 0.2 mg/mL BSA, and 1% glycerol, then incubated with 5 nM of pUC19 supercoiled DNA and either 0, 0.3, 0.6, 0.125, 2.5, or 5 mM of $MnSO_4$. Reactions were incubated at 22 °C for 90 minutes followed by incubation with 0.2 mg/mL of Proteinase K and 5 mM of EDTA at 55 °C for 15 minutes to stop the reaction. Reactions were resolved on 1% Tris-acetic acid-EDTA (TAE) agarose gels and stained with ethidium bromide.

2.8 Dynamic Light Scattering (DLS)

DLS was used to detect aggregation of samples using the Zetasizer Nano (Malvern Instruments). Samples were analyzed in a 12 µL quartz cuvette. Analysis was performed on all samples prior to and after thermal denaturation experiments at the concentration used for experiments (0.25 mg/mL). Samples were also analyzed for the presence of aggregation prior to all other experiments.

CHAPTER 3: RESULTS

3.1 C-Terminus of *Ec*MutL and *Bs*MutL form Stable Domains

Before analyzing the stability of hMutLα, the stability of the bacterial homologs was established as a proof of principle for the techniques used. *Ec*MutL represents MutL homologs that do not exhibit endonuclease activity, while *Bs*MutL represents bacterial homologs that possess endonuclease activity. Both proteins were produced as described in previous work [43, 44], and all experiments were carried out with His-tagged protein unless otherwise specified.

CD analysis was performed to obtain secondary structure information about these MutL homologs. Figure 8 shows CD spectra generated for the C-terminal domains of *Ec*MutL and *Bs*MutL in the far UV range. Both curves display strong negative ellipticity (θ) signals at 208 nm and 222 nm, which are characteristic of high α-helical content [67]. Ellipticity signals are displayed as mean residue ellipticity (θ_{MRE}), which takes into consideration the molecular weight, concentration, and number of residues. The spectral data was used to estimate secondary structure composition using CD Pro software. A comparison between the estimated composition obtained from CD data and that seen in the X-ray crystal structure is shown in Table 4. The proportion of α-helices estimated by CD is similar to that seen in both structures, while the proportion of β-sheets is largely underestimated as expected for the UV range analyzed in these

experiments [68]. This is complemented by a slight overestimation of turns and unordered structure. This information is used as a general comparison against the predicted secondary structure composition of hMutLα variants in the sections to follow.

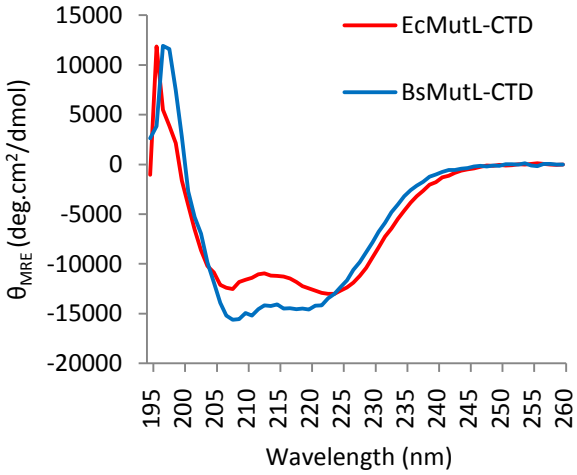

Figure 8. Far UV CD spectra of the C-terminal domains of *Ec*MutL and *Bs*MutL.

Table 4. Comparison of secondary structure composition estimated by CD against X-ray crystal structures of the C-terminal domains of *Ec*MutL and *Bs*MutL.

	α-Helices (%)		**β-Sheets (%)**		**Turns (%)**		**Unordered (%)**	
	CD	X-Ray	CD	X-Ray	CD	X-Ray	CD	X-Ray
*Ec*MutL-CTD	40	43	7	25	23	14	31	18
*Bs*MutL-CTD	39	40	12	24	21	16	29	20

One of the techniques used to analyze the stability of MutL proteins is limited proteolysis. In this experiment, protein is incubated with increasing concentrations of trypsin, followed by separation of degradation products on a SDS-polyacrylamide gel. The C-terminal domain of *Ec*MutL is very stable when incubated with trypsin (Figure 9A).

It is cleaved to a smaller molecular weight product that corresponds to the molecular weight of tagless *Ec*MutL-CTD. Despite having numerous lysines and arginines that trypsin could recognize, this domain remains stable even when exposed to high concentrations of trypsin. This suggests that the domain is well folded and does not have readily exposed regions susceptible to proteolysis as was expected from its structure [43].

The second method used to analyze the stability of MutL proteins was through thermal denaturation. These experiments were carried out using a CD spectrometer. Since MutL homologs contain high α-helical content as shown in Figure 8, experiments were conducted at a wavelength of 222 nm to assess protein unfolding. The melting curve of *Ec*MutL-CTD is shown in Figure 9B, and it displays unfolding in a single step with a melting temperature of approximately 70 °C, indicating high stability.

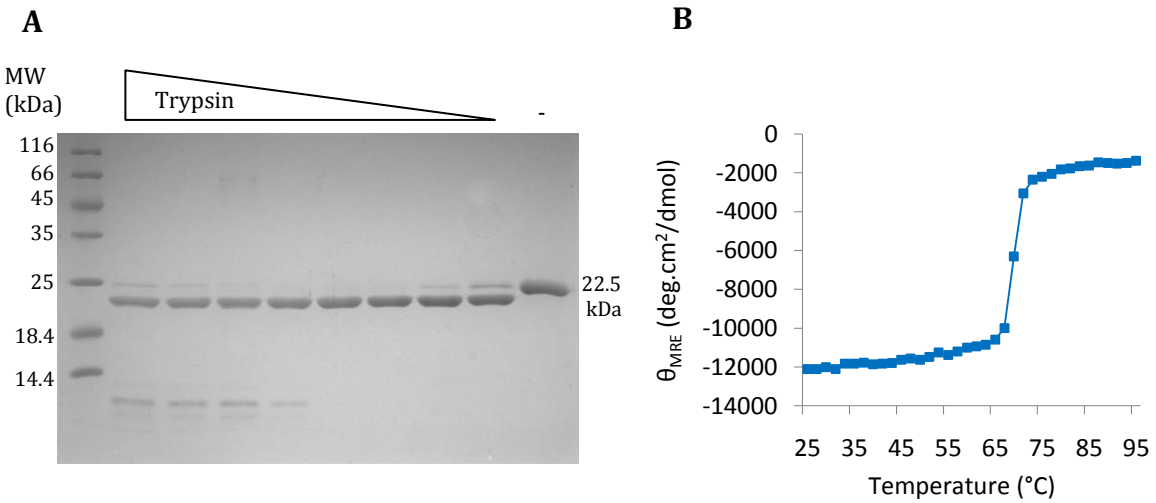

Figure 9. Stability of the C-terminal domain of *Ec*MutL, A) Limited proteolysis (0.5 mg/mL of protein was incubated with 1 µL of 100, 50, 25, 12.5, 6.25, 3.13, 1.56, and 0.78 µg/mL of trypsin in a total volume of 9 µL), B) Thermal denaturation assessed at 222 nm.

Next, the stability of *Bs*MutL-CTD was analyzed using the same experiments as just described. Figure 10A shows limited proteolysis of the C-terminal domain of *Bs*MutL. Like *Ec*MutL-CTD, *Bs*MutL-CTD contains numerous lysines and arginines that trypsin could recognize, but remains stable when incubated with trypsin. However, relative to the C- terminal domain of *Ec*MutL (Figure 9A), the C-terminal domain of BsMutL is more susceptible to degradation when incubated with high concentrations of trypsin.

Unexpectedly, thermal denaturation of *Bs*MutL-CTD generated an atypical melting curve, shown in Figure 10B. The melting curve shows that protein unfolding began at ~55 °C, followed by a gain in negative ellipticity signal, which suggests that refolding occurred. As the temperature is increased further to ~80 °C, the protein undergoes complete unfolding. This experiment was repeated three times from separate sample purifications, therefore the results are unlikely due to an experimental error. When tagless protein was subjected to thermal denaturation, the refolding event no longer occurred (compare Figure 10B and C), therefore the refolding event could be attributed to an anomalous effect of the His-tag.

Unfolding of tagless *Bs*MutL-CTD occurs in two steps, which suggests that the protein may be unfolding as two separate domains. Structures of the C-terminal domain of *Bs*MutL show variation in the orientation between the regulatory subdomain and dimerization subdomain [44], which indicates that the two subdomains could be acting independently of each other. When the protein is denatured under high salt

conditions, the second unfolding step no longer occurs and the protein unfolds in a single step. This suggests that under normal conditions, a particular region of the protein is stabilized by electrostatic interactions, possibly one of the subdomains (see discussion).

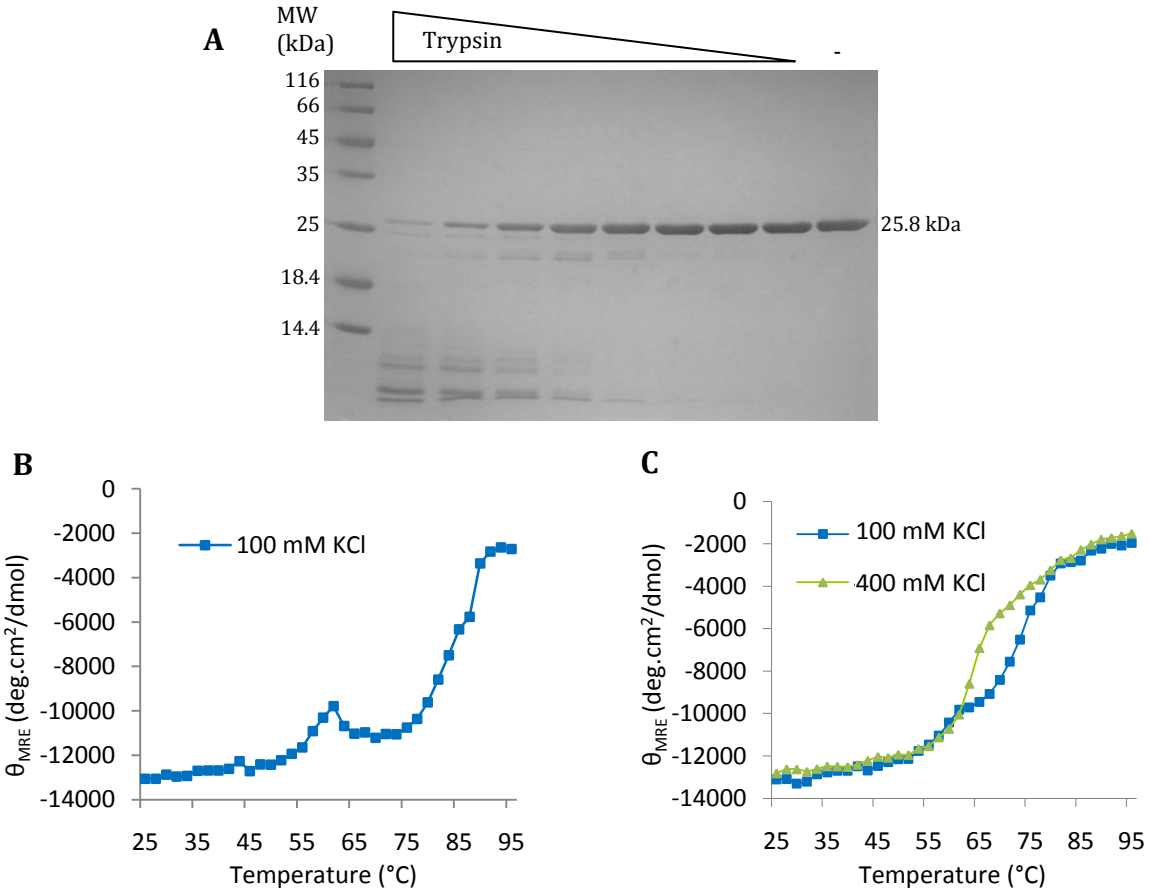

Figure 10. Stability of *Bs*MutL-CTD, A) Limited proteolysis (0.5 mg/mL of protein was incubated with 1 µL of 100, 50, 25, 12.5, 6.25, 3.13, 1.56, and 0.78 µg/mL of trypsin in a total volume of 9 µL), B) Thermal denaturation of *Bs*MutL-CTD with His-Tag, and C) without His-Tag, assessed at 222 nm.

3.2 C-terminus of hMLH1 forms a Stable Domain

Previous work has shown that hMLH1 is required to stabilize hPMS2 to form a heterodimer both *in vivo* and *in vitro* [61]. hMLH1 is able to form homodimers in the absence of hPMS2, whereas hPMS2 is degraded by cellular proteases in the absence of hMLH1 (unpublished data). The stability of two hMLH1 homodimer variants were analyzed as controls. hMLH1 498-756 represents the minimally folded C-terminal domain, while hMLH1 475-756 is the minimal length required for stabilization of hPMS2 (unpublished data).

CD spectra were generated for both hMLH1 constructs and they display nearly identical curves (Figure 11). This data was used to estimate the secondary structure composition, which reveals a high ratio of α-helices to β-sheets (Table 5), that is comparable to the X-ray crystal structure (PDB ID: 3RBN)(Figure 5D).

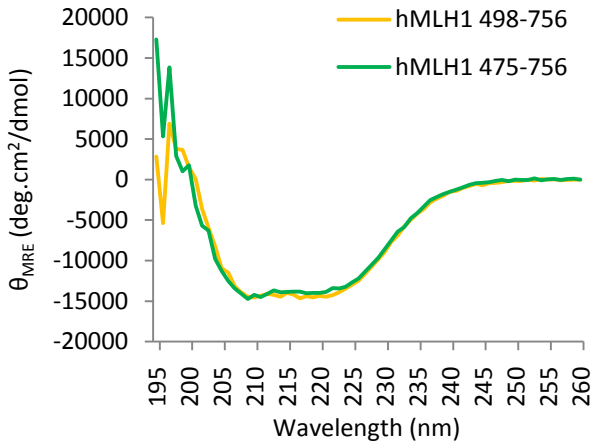

Figure 11. Far UV CD spectra of hMLH1-CTD.

Table 5. Secondary structure composition of hMLH1-CTD estimated by CD compared with X-ray crystal structure data.

	α-Helices (%)		**β-Sheets (%)**		**Turns (%)**		**Unordered (%)**	
	CD	X-Ray	CD	X-Ray	CD	X-Ray	CD	X-Ray
hMLH1-CTD*	39	43	13	16	21	15	28	27

*hMLH1-CTD CD data is based on 498-756 construct

Limited proteolysis of hMLH1 498-756 results in cleavage to a subproduct that corresponds to the molecular weight of tagless protein (29.8 kDa), as well as a lower molecular weight band that is ~27 kDa (indicated in Figure 12A). To identify the boundaries of this band, it was sent for analysis by mass spectrometry at the Bioanalytical and Mass Spectrometry lab at McMaster University. The band was subjected to in-gel trypsin digestion prior to being analyzed by liquid chromatography-tandem mass spectrometry. Once the molecular weight of each peptide fragment was identified, the fragments were queried against protein databases to identify the target protein. Table 6 shows the fragments detected using two proteomics search engines (The Global Proteome Machine (GPM) and Mascot) and their corresponding experimental molecular weight. The collective experimental molecular weight of the fragments is 21 kDa, while the band in the gel appears to be ~27 kDa. Figure 13 highlights the detected fragments on the structure of hMLH1, and they do not form a domain. For a more accurate identification of the domain boundaries, another method should be used, such as *de novo* peptide sequencing. This method involves assigning amino acids based on mass spectrum data rather than relying on protein databases.

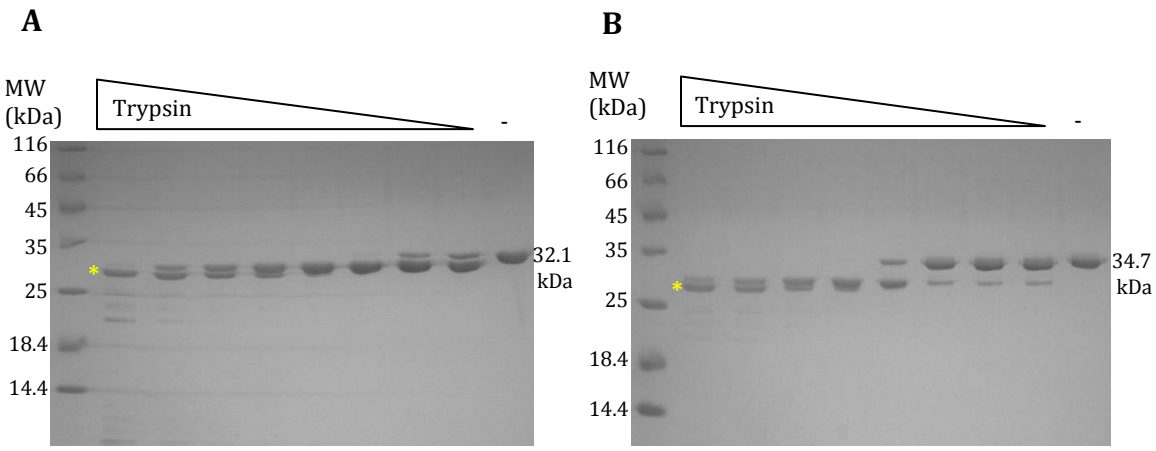

Figure 12. Limited proteolysis of A) His-tagged hMLH1 498-756 and B) His-tagged hMLH1 475-756 (0.5 mg/mL of protein was incubated with 1 µL of 100, 50, 25, 12.5, 6.25, 3.13, 1.56, and 0.78 µg/mL of trypsin in a total volume of 9 µL). Yellow asterisks indicate degradation product that is ~27 kDa.

Table 6. Molecular weight summary of peptide fragments of the hMLH1 498-756 subproduct identified through mass spectrometry.

Peptide Fragment	Molecular Weight	Program
499-522	2789.5069	Mascot
547-554	964.5593	Mascot
565-575	1314.6841	GPM
576-604	3131.4765	GPM
605-616	1410.7515	GPM
642-659	2082.1998	GPM
660-678	2314.0544	GPM
688-713	2879.4058	GPM
714-722	1174.6255	GPM
726-732	790.4693	Mascot
733-751	2201.1237	GPM
Total =	21052.8568	

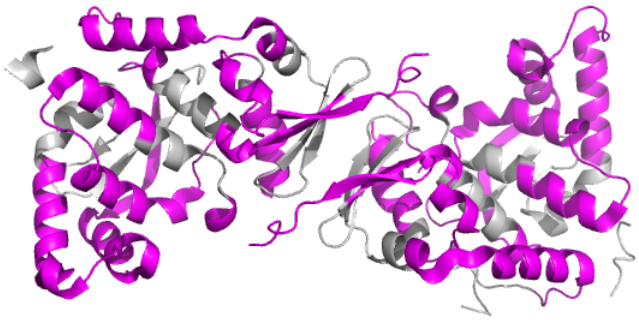

Figure 13. Structure of the C-terminal domain of hMLH1 (PDB ID: 3RBN). Highlighted in pink are the peptide fragments identified from liquid chromatography-tandem mass spectrometry.

Limited proteolysis of hMLH1 475-756 results in cleavage to a subproduct that corresponds to the theoretical molecular weight of hMLH1 498-756 (Figure 12B). This is further cleaved to a subproduct that is ~27 kDa, which is the same size as the band analyzed by mass spectrometry from Figure 12A. The crystal structure of hMLH1 486-751 was recently solved, which is very close to the boundaries of our constructs. The structure shows the first 18 residues to be unstructured and the first secondary structure to begin at residue 504, therefore hMLH1 498-756 is used to represent the minimal folded domain. Although both hMLH1 constructs are cleaved to a subproduct, this subproduct remains stable even when incubated with 1 µL of 100 µg/mL of trypsin.

Thermal denaturation of hMLH1 498-756 and hMLH1 475-756 generated identical melting curves. They each begin to unfold at approximately 45 °C, which is relatively lower than the bacterial homologs (Figure 14 compared to Figure 9B and 10C). After partial unfolding the ellipticity signal tails off rather than continuing on to form a

plateau, an effect that is unaffected by salt. This indicates that further protein unfolding does not occur, likely due to protein aggregation.

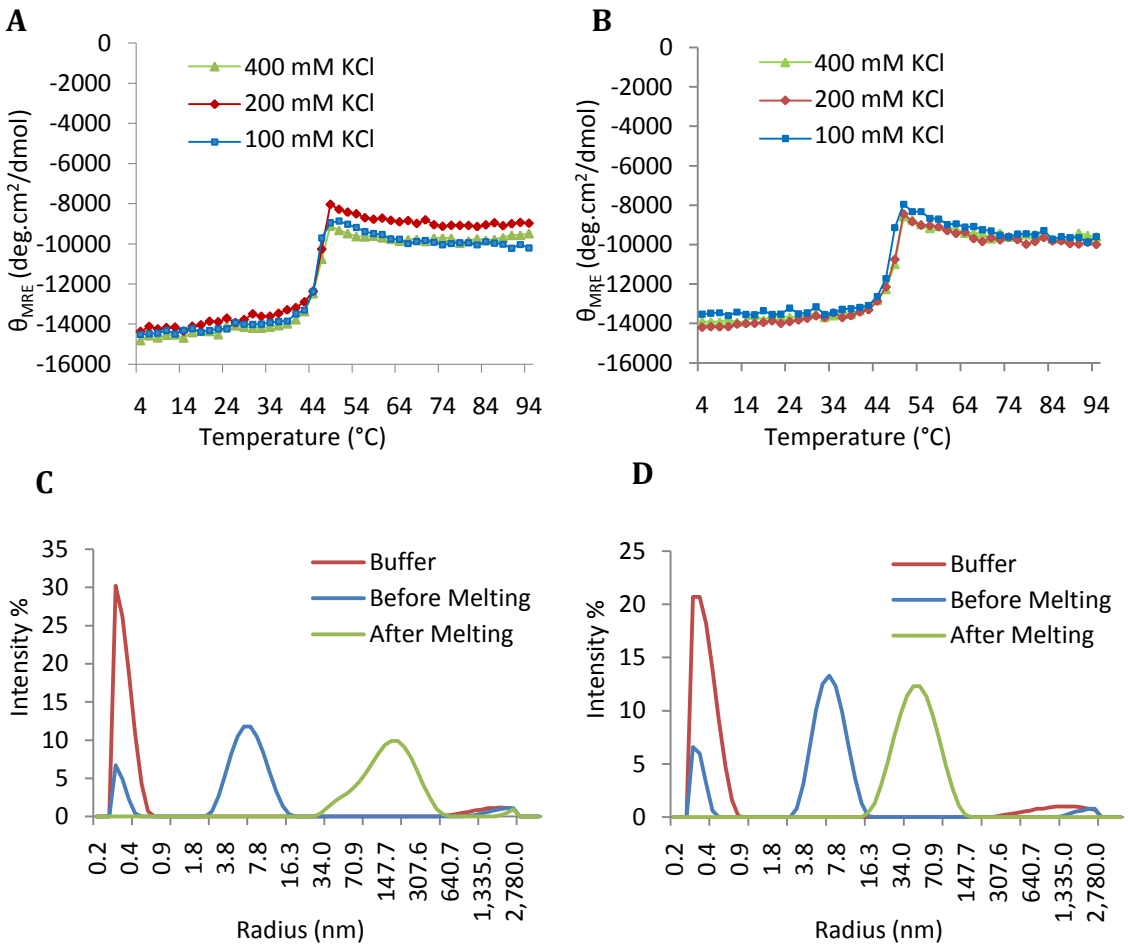

Figure 14. Thermal denaturation of A) hMLH1 498-756 and B) hMLH1 475-756, and dynamic light scattering analysis of the aggregation state of C) hMLH1 498-756 and D) hMLH1 475-756 before and after thermal denaturation.

Dynamic light scattering (DLS) was used to determine whether protein aggregation had taken place. DLS measures the fluctuation of light scattering when a sample is in solution. The intensity of light scattering is affected by the hydrodynamic

size of the particle, i.e. how it diffuses through solution, and can be used to approximate the size of the particle.

Figures 14C and 14D show a larger particle size after thermal denaturation compared to before, which indicates that aggregation has occurred. Despite the tendency to aggregate during unfolding, it appears as though both hMLH1 constructs are unfolding in a single step. Considering that hMLH1-CTD is folded as a single domain as seen in its structure (shown in Figure 5D), this is not surprising.

3.3 Characterization of hMutLα Stability at the C-terminal Domain

3.3.1 Extending the Linker Region of hMutLα

In vivo expression of hPMS2 is dependent on hMLH1 [61, 69]. It has been shown that MLH1-deficient cell lines possess very low steady state levels of hPMS2, however expression of hPMS2 is restored upon complementation with wild type hMLH1 [69]. Additionally, truncation of the conserved terminal eight residues of hMLH1 reduces heterodimer stability and increases spontaneous mutation rates *in vivo* [61].

We have previously shown that recombinant expression of the C-terminal domain of hMutLα can only be obtained by co-expressing hMLH1 and hPMS2 due to the unstable nature of hPMS2. hMLH1 475-756 is able to support formation of the

heterodimer, so long as hPMS2 is expressed at a minimum residue length of 600-862 (unpublished data). According to secondary structure predictions, this includes a significant portion of the presumably unstructured linker region. Although heterodimer formation is supported by these residues, the construct remained highly unstable (unpublished data). This led us to hypothesize that perhaps hMutLα-CTD minimal linker was still lacking a portion of the linker region that may be necessary to further stabilize hPMS2. To test this hypothesis four hMutLα constructs were analyzed (Table 2), each varying in the length of linker region expressed with the C-terminal domain of hMutLα.

Solubility assays were performed on hMutLα half linker2 and hMutLα full linker as described in the Materials and Methods to determine whether the constructs were soluble when expressed in *E. coli* cell lines. Expression was induced with IPTG followed by incubation at 25 °C for 5 hours. Both constructs express after IPTG induction, however the expression level of hPMS2 is significantly lower than that of hMLH1 in hMutLα full linker (Figure 15). After incubation, cells were lysed and the soluble fraction was isolated. As seen in Figure 15, both variants are soluble. hMutLα minimal linker and hMutLα half linker1 were previously established in the lab to be soluble.

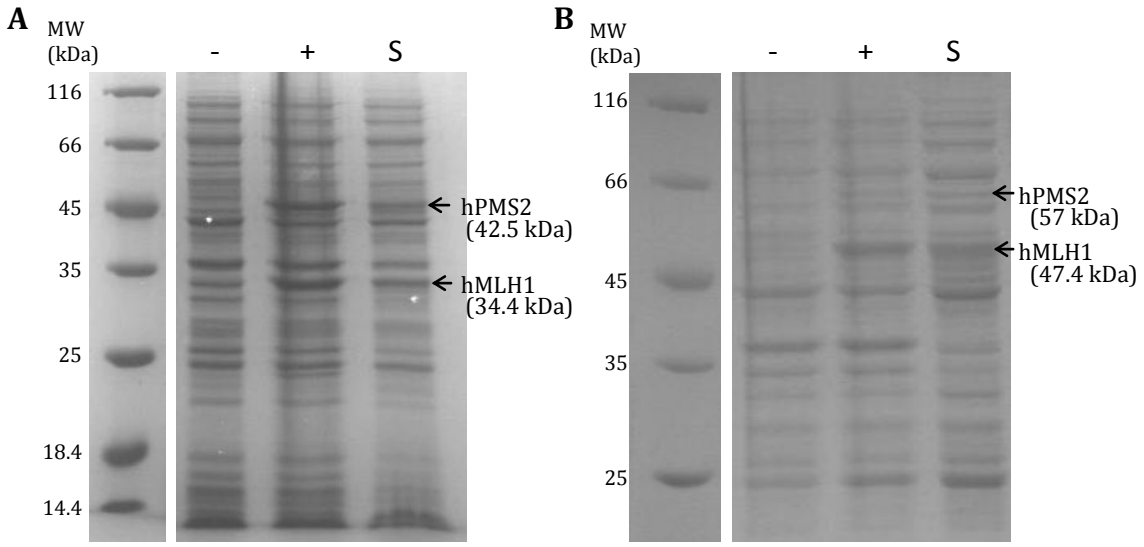

Figure 15. Solubility assays of A) hMutLα half linker2 in *E. coli* BL21(DE3) pRareLysS cells and B) hMutLα full linker in *E. coli* BL21(DE3) Star pRareLysS cells. "-" lanes indicate prior to induction with IPTG, "+" lanes indicate post induction with IPTG, and "S" lanes indicate soluble protein produced after 5 hours at 25 °C.

All hMutLα constructs were purified in the same manner. Purification gels of hMutLα full linker are shown in Figure 16 as an example. Cell lysates were first purified through a Ni^{2+}-chelating affinity column (Figure 16A). hPMS2 is fused to a His-tag and not hMLH1. Since the expression level of hPMS2 is significantly lower than that of hMLH1 (Figure 15B), excess hMLH1 will flow through the column. After elution of protein from the column, an anion exchange column purification step is applied (Figure 16B and C). This step removes remaining contaminants as judged by Coomassie staining on SDS-polyacrylamide gel. A size exclusion chromatography step is implemented as a final buffer exchange.

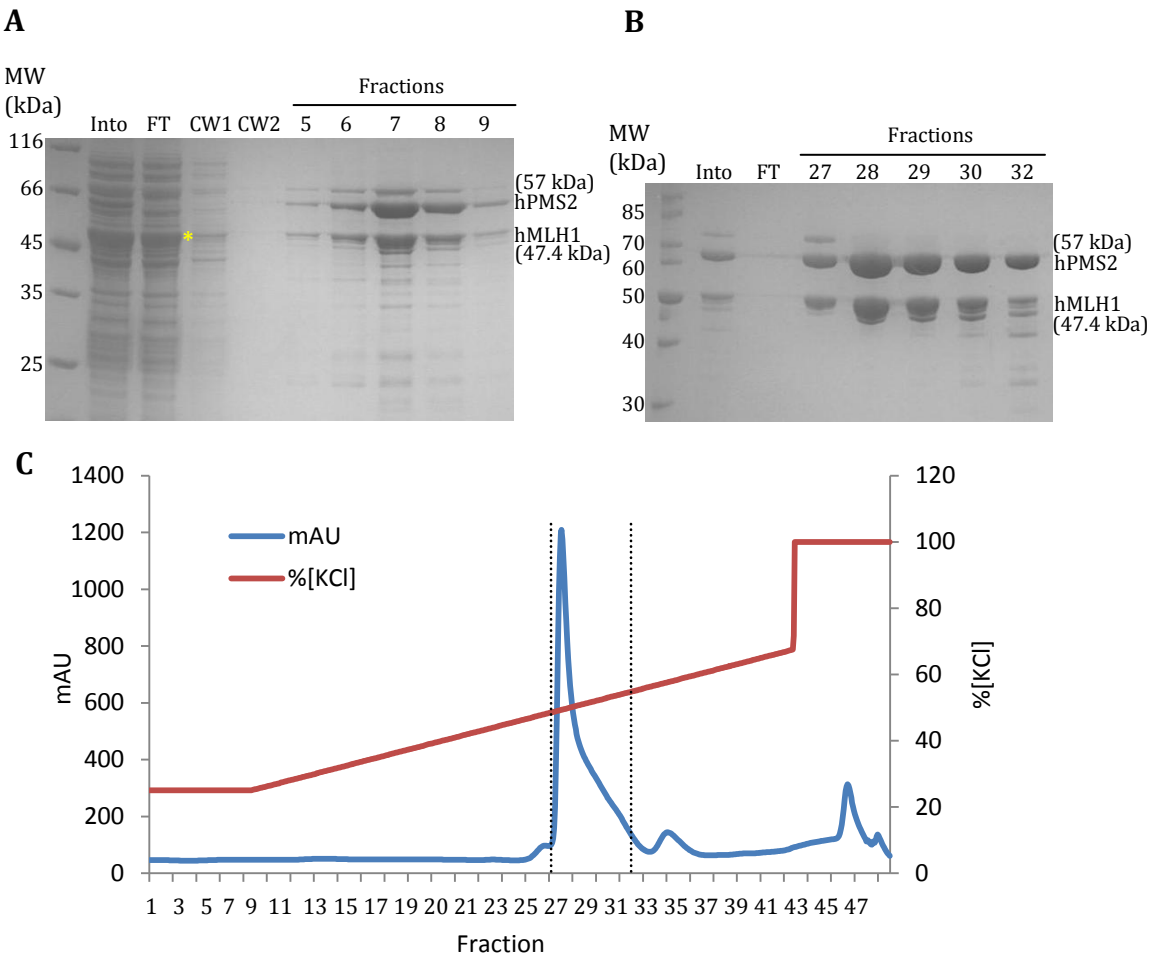

Figure 16. Purification of hMutLα full linker. A) Ni^{2+}-affinity column purification where "Into" represents sample loaded into the column, "FT" represents flow through, and "CW" represents column wash with 45 mM imidazole. Fractions were eluted from column with 300 mM imidazole (yellow asterisk marks hMLH1 homodimer), B) Gel of anion exchange column purification, and C) Elution profile of anion exchange column purification (dotted lines show section corresponding to peak fractions in (B)).

hMLH1 of the hMutLα full linker variant is susceptible to degradation, and proteolysis is likely taking place at the flexible linker region (Figure 16). The addition of protease inhibitors to buffers prior to cell lysis and through Ni^{2+}-affinity column purification did not prevent degradation of the linker. Non-degraded protein elutes

from the anion exchange column as a single peak, while degraded protein elutes in a shoulder following the peak (Figure 16B and C). For subsequent experiments, only fractions containing the least degraded form were used.

3.3.2 Secondary Structure Composition of hMutLα

CD spectra were generated for each hMutLα construct and used to estimate the composition of secondary structure (Figure 17, Table 7). Each hMutLα construct was estimated to be predominantly α-helical, similar to the other MutL homologs. Of the homologs under analysis in this work, hPMS2 shows more similarity to *Bs*MutL than *Ec*MutL through sequence alignments [44] (Figure 3). A comparison of the secondary structure composition between hMutLα minimal linker and *Bs*MutL-CTD shows a slightly lower proportion of α-helices and higher proportion of β-sheets, but is not significantly more unordered. This suggests that the domain is likely folded, although the information provided by this analysis is purely on the level of secondary structures.

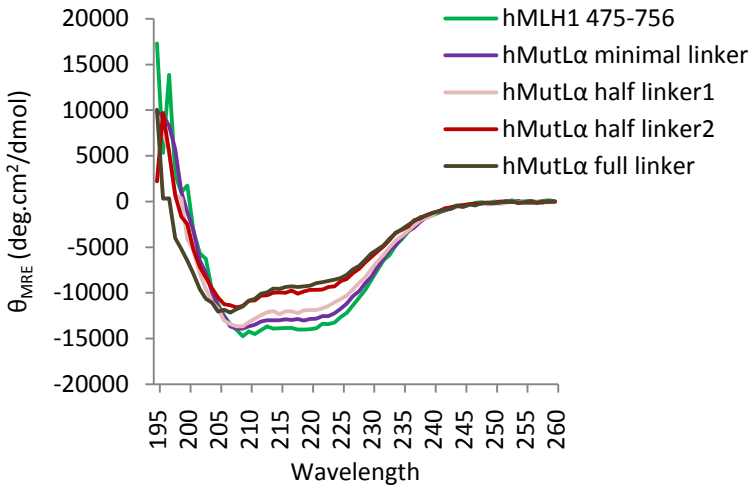

Figure 17. CD spectra of hMutLα constructs.

Table 7. Secondary structure composition of MutL variants as determined by CD.

Construct	Plasmid #	α-Helices (%)	β-Sheets (%)	Turns (%)	Unordered (%)
EcMutL-CTD	pWY 1295	40	7	23	31
BsMutL-CTD	pAG 8188	39	12	21	29
hMLH1 498-756	pWY 1321	39	13	21	28
hMLH1 475-756	pWY 1279	38	11	22	29
hMutLα minimal linker (hMLH1 475-756, hPMS2 600-862)	pAG 8036	35	15	21	30
hMutLα half linker1 (hMLH1 475-756, hPMS2 506-862)	pWY 1348	32	16	21	32
hMutLα half linker2 (hMLH1 458-756, hPMS2 506-862)	pAG 8385	29	18	22	31
hMutLα full linker (hMLH1 336-756, hPMS2 370-862)*	pAG 8519	27	20	21	32

*Sample degrades overtime, therefore calculations based on protein concentration are not exact.

A comparison of each hMutLα-CTD construct shows a trend of having a decreased proportion of α-helices and a concurrent increase in β-sheets and unordered structure as you increase the linker length (Table 7). This would suggest that the linker region is contributing both β-sheet and unordered structures, contrary to secondary structure predictions. Previous work has suggested that the linker region adopts additional α-helices or β-sheets when full-length protein is bound to ATP [48].

3.3.3 Trypsin Digestion of hMutLα

Limited proteolysis was performed on all four hMutLα constructs (Table 2) followed by separation of digested products on a SDS polyacrylamide gel to test the susceptibility to degradation as demonstrated with the bacterial MutL and hMLH1 proteins. As can be seen in Figure 18, hMLH1 follows the same degradation pattern as noted previously in Figure 12 (yellow asterisks indicate degradation products of hMLH1). From the digestion of hMutLα minimal linker (Figure 18A), hPMS2 is completely digested with 1 µL of 0.39 µg/mL of trypsin while hMLH1 remains stable at higher concentrations of trypsin, showing how susceptible hPMS2 is to degradation. From the gels of hMutLα half linker1, half linker2, and full linker (Figure 18B, C, and D), hPMS2 can be seen to be first cleaved to a tagless form before being completely degraded at low concentrations of trypsin as seen with the shorter construct.

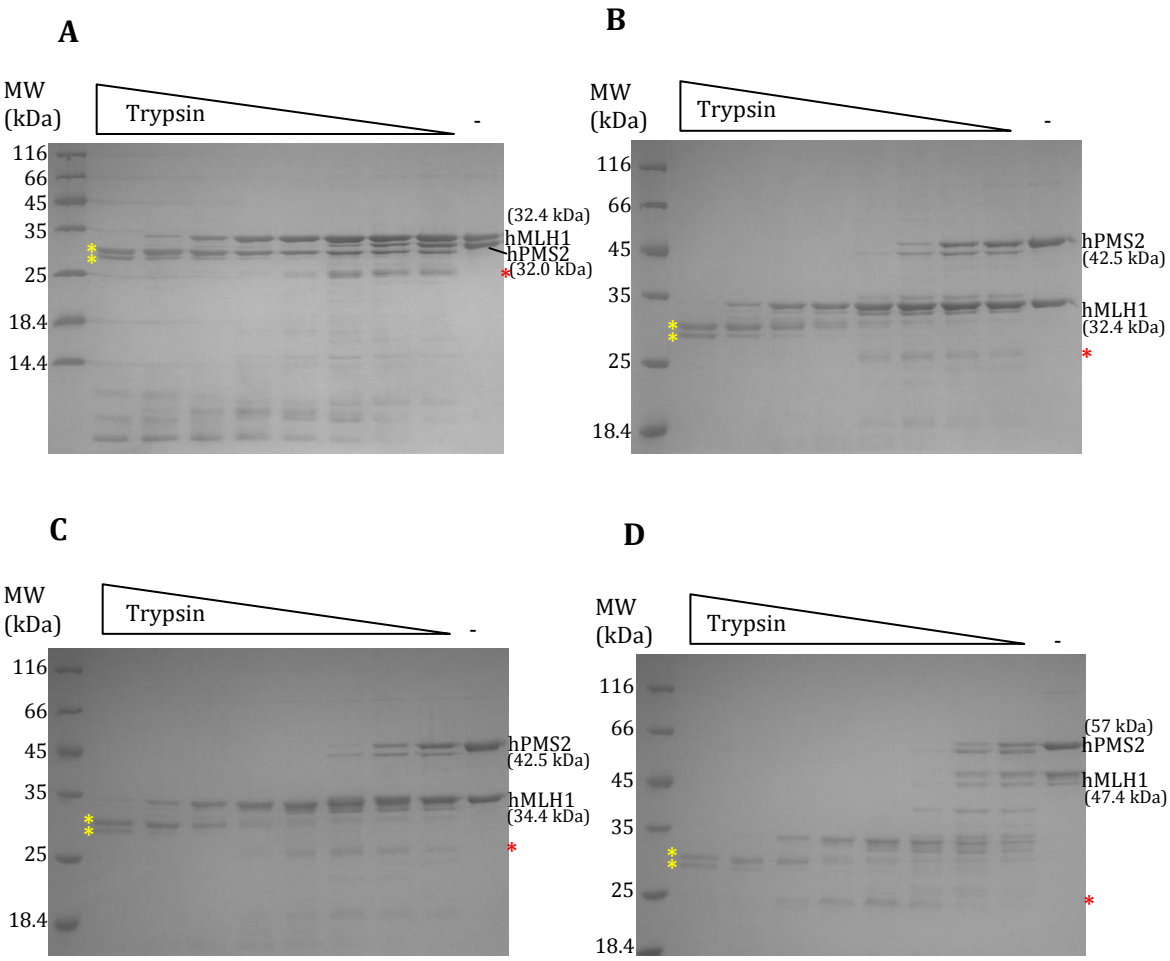

Figure 18. Limited proteolysis of hMutLα constructs A) hMutLα minimal linker, B) hMutLα half linker1, C) hMutLα half linker2, D) hMutLα full linker. 0.5 mg/mL of protein was incubated with 1 µL of 6.25, 3.13, 1.56, 0.78, 0.39, 0.20, 0.10, or 0.05 µg/mL of trypsin in a total volume of 9 µL. Products were separated by SDS-PAGE. Yellow asterisks indicate degradation products of hMLH1, red asterisks indicate the ~25 kDa degradation product.

One minor difference is noted between the digestion pattern of hMutLα full linker and the three shorter hMutLα constructs. Digestion of the three shorter constructs produces a subproduct approximately 25 kDa in size that forms at the lowest concentrations of trypsin used (0.05 µg/mL) and is readily degraded further with

increasing concentrations of trypsin (indicated by red asterisks in Figure 18). When hMutLα full linker is degraded by trypsin, it is cleaved to this subproduct with 0.1-0.2 μg/mL of trypsin and is not further degraded as readily relative to the other constructs. This may suggest that the N-terminal portion of the linker has an influence on a region of the C-terminal domain, though it may not help to stabilize the entire domain, however this idea is speculative.

The susceptibility of hPMS2 to proteolysis could be due to one of two possibilities. One reason could be that the C-terminal domain of hPMS2 contains lysines or arginines on exposed loops, which are susceptible to cleavage by trypsin, and upon denaturation and separation on a SDS polyacrylamide gel, it appears as multiple separated fragments. The second reason could be that the C-terminal domain of hPMS2 consists of a "loose" structure with multiple exposed regions available for cleavage by trypsin. To determine which scenario is taking place, the same trypsin digestion was performed followed by separation on a non-denaturing polyacrylamide gel. As can be seen in Figure 19, hMutLα-CTD is separated into many fragments as opposed to remaining as an intact domain, revealing that the C-terminal domain of hPMS2 is loosely folded. This is again seen in all four hMutLα constructs.

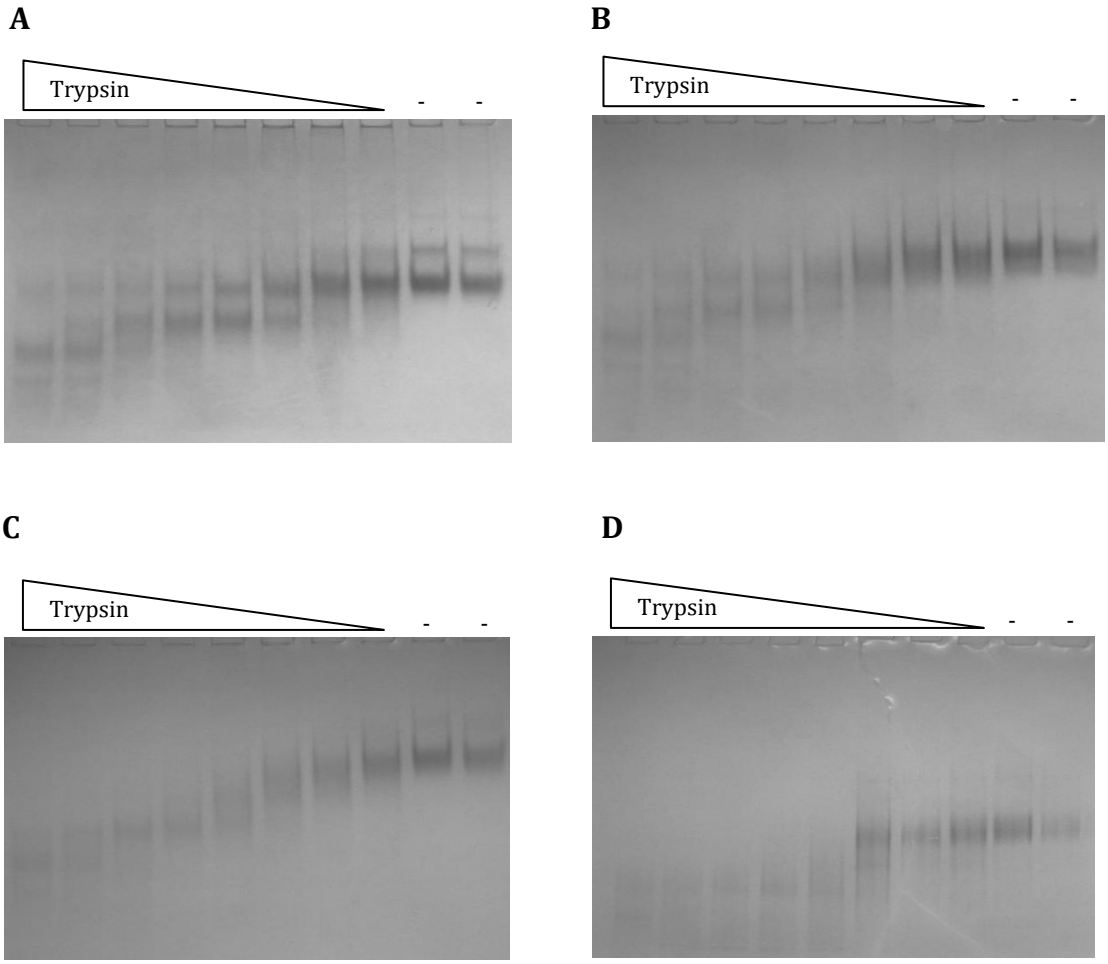

Figure 19. Limited proteolysis of hMutLα constructs A) hMutLα minimal linker, B) hMutLα half linker1, C) hMutLα half linker2, D) hMutLα full linker. 0.5 mg/mL of protein was incubated with 1 µL of 6.25, 3.13, 1.56, 0.78, 0.39, 0.20, 0.10, or 0.05 µg/mL of trypsin in a total volume of 9 µL. Products were separated by non-denaturing PAGE.

Increasing the linker length does not protect the C-terminal domain of hMutLα against proteolysis. That being said, constructs with extended linker lengths are not significantly more susceptible to proteolysis, which would be the expected result if the linker was completely unstructured.

3.3.4 Effect of Zn^{2+} on Trypsin Digestion of hMutLα

hPMS2 was previously predicted to be a Zn^{2+} binding protein through computational analyses [45]. Through Zn^{2+}-release assays performed on purified protein, hMutLα was found to bind one zinc metal ion per dimer, suggesting that only one promoter, presumably hPMS2, binds Zn^{2+} [45]. *Bs*MutL-CTD also binds Zn^{2+}, and the X-ray crystal structures show that Zn^{2+} binding causes a change in orientation between the dimerization and regulatory subdomains, locking the conformation between the two [44]. With this in mind, limited proteolysis analysis of the hMutLα constructs was performed in the presence of Zn^{2+} at a 1:1 molar ratio to determine whether the metal ion exhibits a protective effect against proteolysis.

Limited proteolysis of hMutLα variants produced a subproduct approximately 25 kDa in size (Figure 18). In the presence of Zn^{2+}, this subproduct was not formed readily and persisted at higher concentrations of trypsin, indicating protection of this fragment (Figure 20A, B, and C). Zn^{2+} is likely inducing a local conformational change, limiting the availability of previously exposed regions to trypsin. The presence of Zn^{2+} does not appear to alter the degradation pattern of hMutLα full linker (Figure 20D).

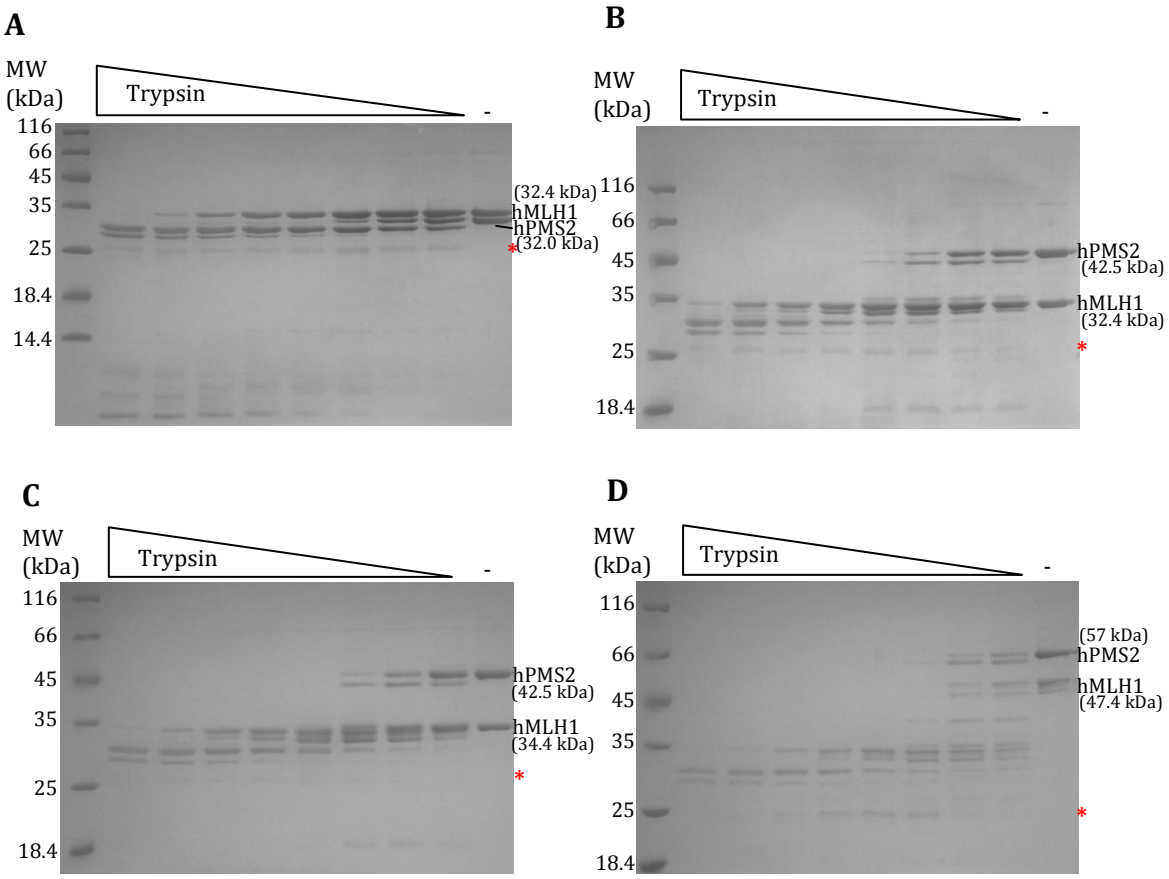

Figure 20. Limited proteolysis of hMutLα constructs in the presence of Zn^{2+}, A) hMutLα minimal linker, B) hMutLα half linker1, C) hMutLα half linker2, D) hMutLα full linker. 0.5 mg/mL of protein was incubated with 1 µL of 6.25, 3.13, 1.56, 0.78, 0.39, 0.20, 0.10, or 0.05 µg/mL of trypsin in a total volume of 9 µL. Products were separated by SDS-PAGE. Red asterisks indicate the ~25 kDa degradation product that is protected.

3.3.5 Thermal Denaturation of hMutLα

To further assess the stability of hMutLα-CTD, melting curves were generated using the CD spectrometer as before, assessing protein unfolding at 222 nm. The unfolding of hMutLα minimal linker appears to occur in multiple steps (Figure 21A). The

first unfolding step begins at ~25 °C, which is a much lower temperature than what was observed for the bacterial homologs and hMLH1 constructs (compare Figure 21A with 9B, 10C, and 14). The second unfolding step begins at ~45 °C and the final step at ~80 °C.

Unfolding was also assessed at increasing concentrations of KCl. Increasing the salt concentration appeared to destabilize the portion of the protein that unfolds last, but had no effect on the other unfolding steps. To test for the ability of Zn^{2+} to promote stability of hMutLα-CTD, unfolding was assessed in the presence of 100 nM of $ZnCl_2$, however there was no significant change in the melting curve in the presence of Zn^{2+}.

Thermal denaturation of the other hMutLα constructs expressing an extended region of the linker reveals the same unfolding profile as hMutLα minimal linker up until 55 °C, although the initial unfolding step is less apparent with hMutLα full linker (Figure 21D). It should be noted that due to the susceptibility of the full linker construct for degradation during sample preparation, concentration readings used for θ_{MRE} calculations may not be exact. Neither increasing the KCl concentration nor including $ZnCl_2$ in the buffer has an effect on the way these constructs unfold.

After 55 °C, the ellipticity signal of the three constructs expressed with extended linker region tails off rather than continuing on to form a plateau. We suspected that the samples were forming aggregates, which prevented further protein unfolding. DLS was used to detect the presence of aggregation after thermal denaturation (Figure 22). Buffer alone displayed a peak of ~ 100 nm. Analysis of hMutLα variants before melting

showed a peak size ~7 nm. After melting, the sample peak disappeared, and only the peak that corresponds to the buffer component can be seen. We suspected that the samples were adhering to the cuvette used during thermal denaturation. Following thermal denaturation experiments, extensive cleaning using strong acid was required to remove residual sample from the cuvette. Therefore the presence of aggregation could not be detected.

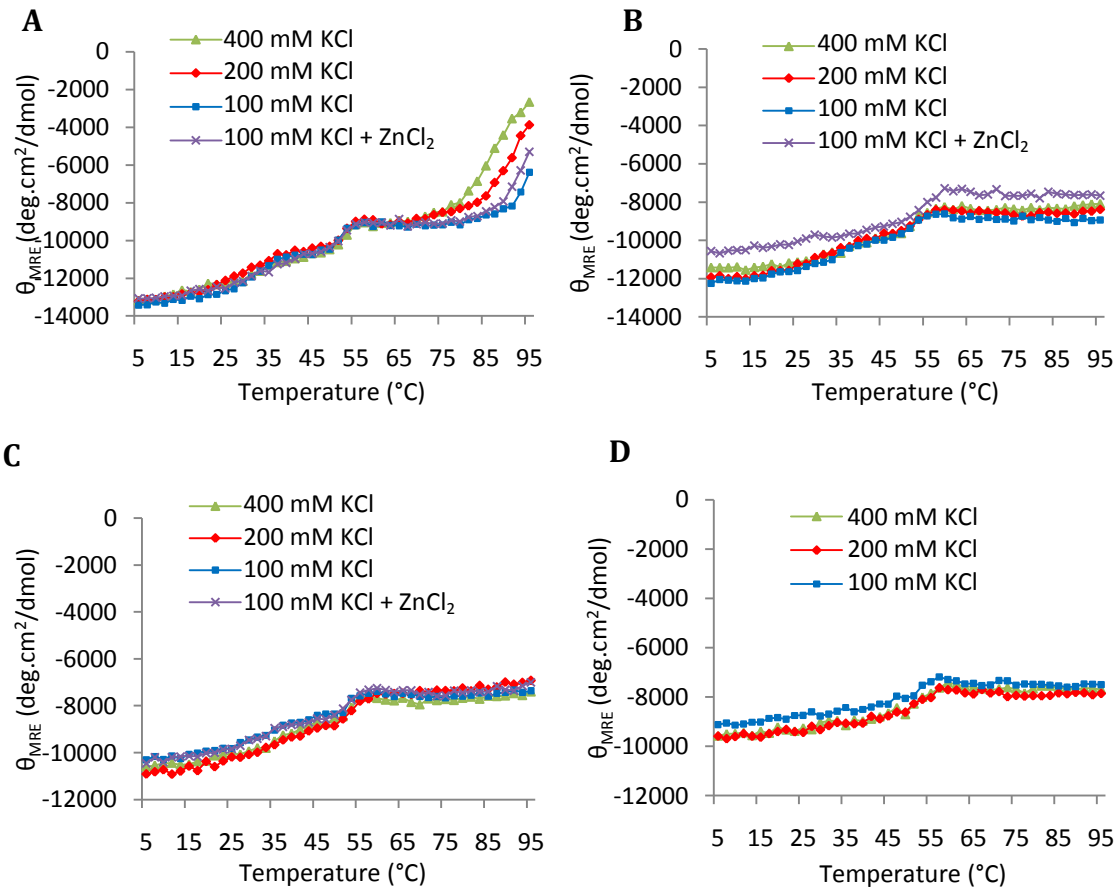

Figure 21. Thermal denaturation curves of A) hMutLα minimal linker, B) hMutLα half linker1, C) hMutLα half linker2, and D) hMutLα full linker, assessed at 222 nm.

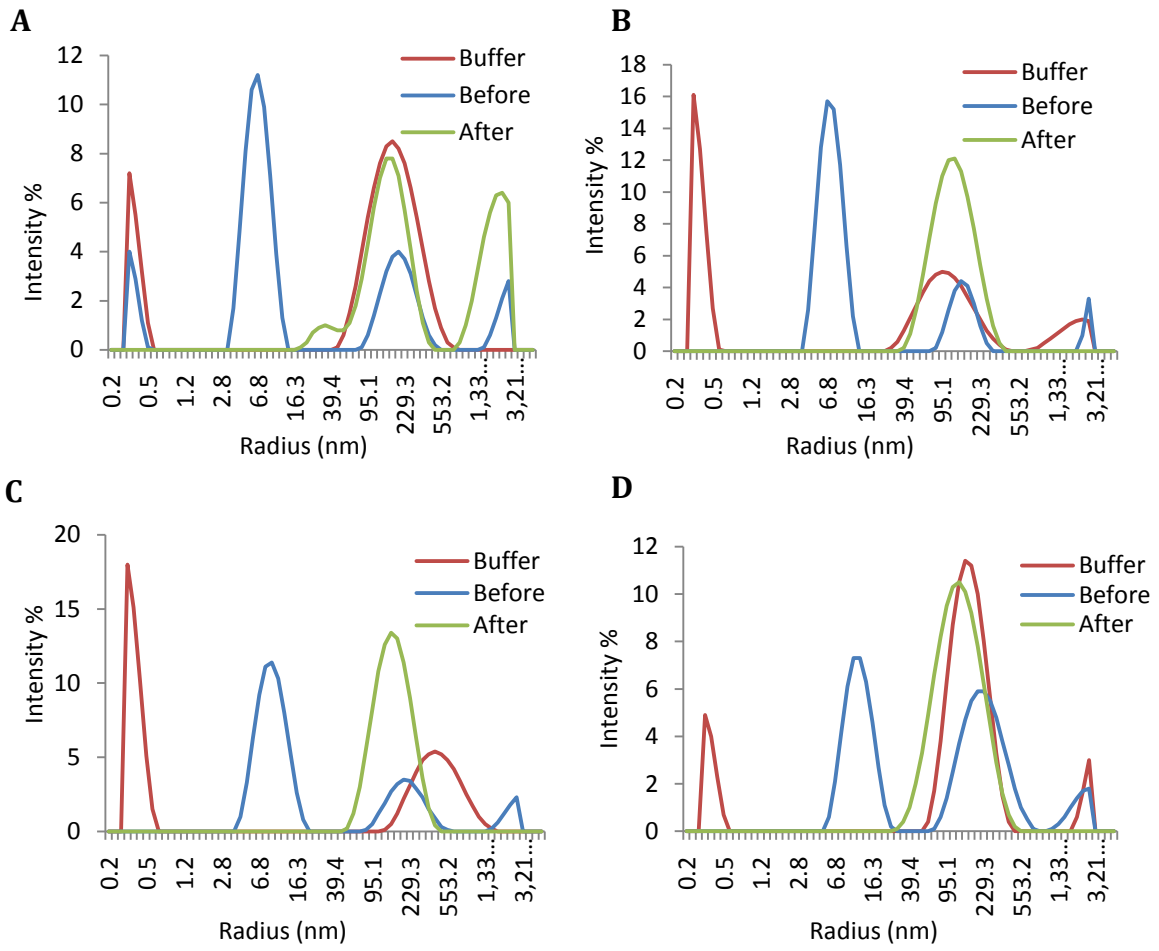

Figure 22. Dynamic light scattering analysis of A) hMutLα minimal linker, B) hMutLα half linker1, C) hMutLα half linker2, and D) hMutLα full linker before and after thermal denaturation.

3.4 DNA Binding Activity of hMutLα

The ability to bind DNA is important for MutL function in mismatch repair [20, 57, 58]. Mutation of conserved residues in the N-terminal domain that are important for DNA binding lead to defective mismatch repair *in vivo* [20, 57, 58]. While key

residues involved in DNA binding in the N-terminal domain have been identified, the contribution of the C-terminal domain to DNA binding is less clear. The N-terminal domain of *Ec*MutL was shown to bind DNA, however full-length protein bound with higher affinity, even though the C-terminal domain could not bind [43]. The DNA binding ability of the C-terminal domain of hMutLα has not been previously shown, and the contribution of the linker region has also never been assessed.

Two double-stranded DNA substrates of different lengths, 113 bp and 548 bp, were used to assess the DNA binding ability of the hMLH1 and hMutLα constructs through electrophoretic mobility shift assays. The DNA binding of full-length MutL proteins has been shown to be length dependent, requiring a minimum of 100 bp to detect binding in *Ec*MutL and greater than 213 bp to detect high affinity binding in yMutLα [55, 56]. Because MutL does not exhibit sequence specificity or recognize DNA modifications [55, 56], these oligonucleotides do not consist of any particular sequence. We suspected that if hMutLα-CTD were to exhibit DNA binding, it would be weak based on the fact that binding was not detected for the C-terminal domain of *Ec*MutL or *Bs*MutL at a 128-fold excess of protein [43, 44]. In the experiments presented here, DNA substrate was incubated with protein up to a 320-fold excess as indicated in the figure legends.

Neither construct of hMLH1 binds to the 113 bp DNA substrate, while all of the hMutLα constructs show extremely minimal binding to the 113 bp, indicated by smearing seen with increasing concentration of protein (Figure 23).

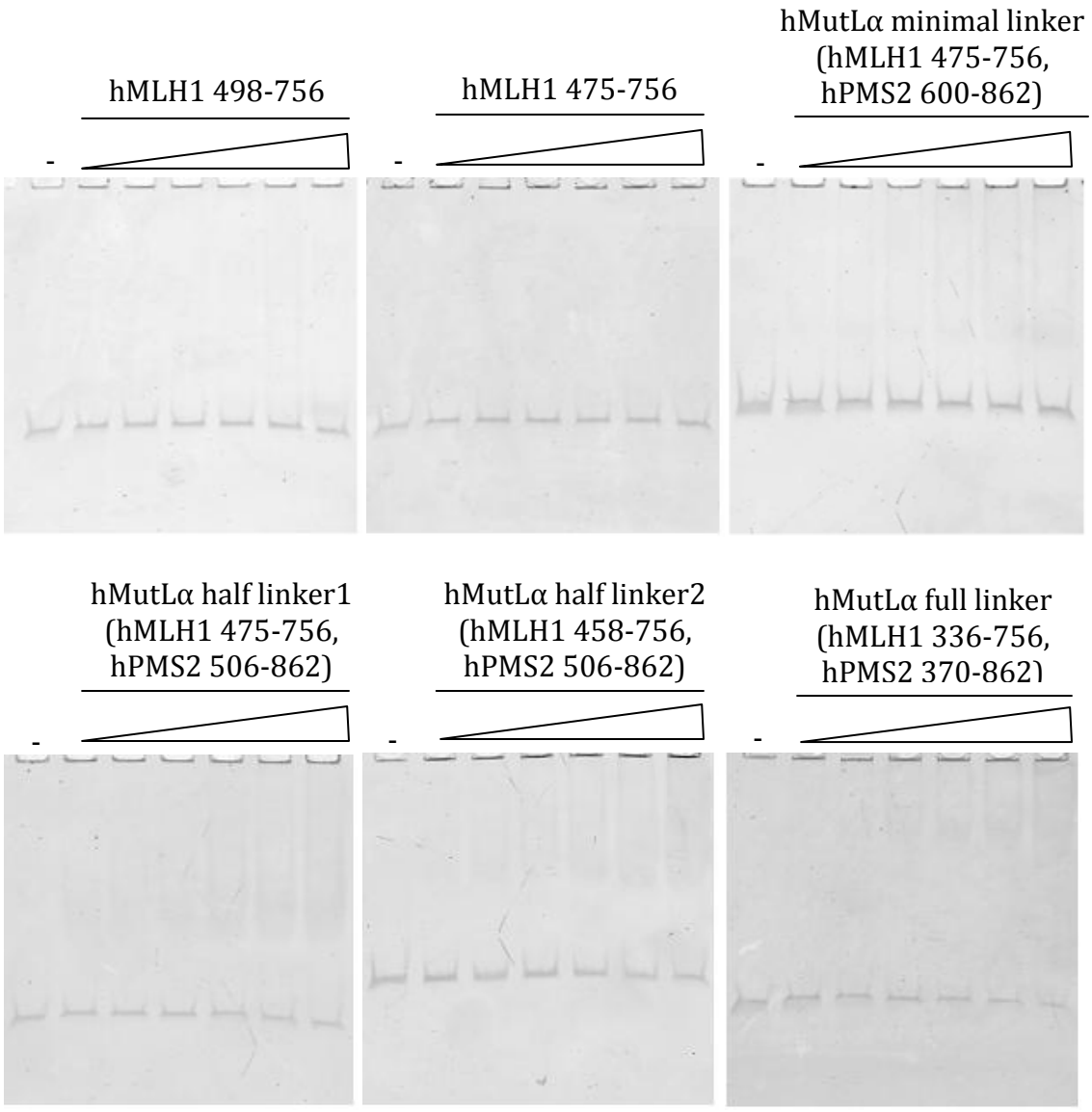

Figure 23. EMSA gels of 113 bp oligonucleotide (10 nM) incubated with increasing concentrations of the indicated hMLH1 or hMutLα construct (0, 100, 200, 400, 800, 1600, 3200 nM) in a total volume of 20 µL. All incubations were carried out at 30 °C for 30 minutes, except for hMutLα full linker construct which was carried out at 22 °C for 30 minutes.

Again, neither construct of hMLH1 was able to bind to the 548 bp DNA substrate, and the three shorter hMutLα constructs – hMutLα minimal linker, half linker1, and half

linker2 – did not show binding (Figure 24). Only hMutLα full linker showed clear binding at a large excess of protein (160:1). This indicates that the N-terminal portion of the linker may be important for DNA binding by hMutLα.

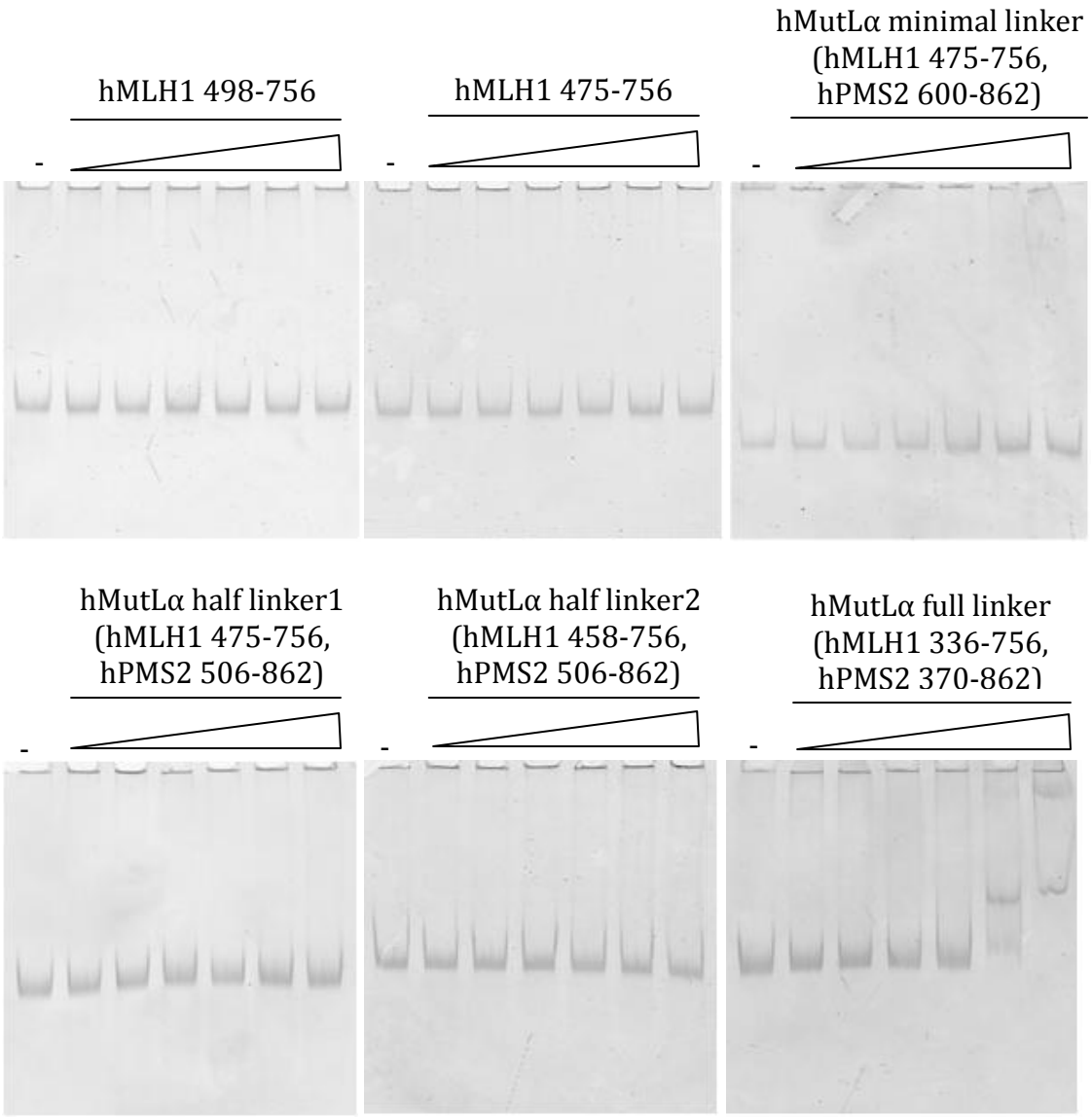

Figure 24. EMSA gels of 548 bp oligonucleotide (10 nM) incubated with increasing concentrations of the indicated hMLH1 or hMutLα construct (0, 100, 200, 400, 800, 1600, 3200 nM) in a total volume of 20 µL. All incubations were carried out at 30 °C for 30 minutes, except for hMutLα full linker construct which was carried out at 22 °C for 30 minutes.

Two distinct shifts can be observed in binding to the 548 bp substrate. The ability of full-length *Ec*MutL to supershift DNA was shown previously through electrophoretic mobility shift assays [55]. This effect was attributed to cooperative binding of protein onto DNA [55], suggesting that this may also be the case in our experiment. This in turn would imply that the cooperative association of MutL is mediated by the linker and/or the C-terminal domain, however this idea awaits validation [70].

3.4.1 Effect of DNA on Trypsin Digestion of hMutLα

Although DNA binding is not observed in the three shorter hMutLα constructs, DNA binding has been reported in the C-terminal domain of *Ng*MutL and *Aa*MutL through surface plasmon resonance experiments [53, 60]. It is possible that the C-terminal domain of hMutLα possesses weak DNA binding activity that is not detectable through electrophoretic mobility shift assays. We wanted to see whether hMutL-CTD binding to DNA would protect the C-terminal domain against proteolysis.

Each hMutLα-CTD construct was incubated with a 548 bp oligonucleotide at a molar ratio of 100:1 (protein:DNA) prior to incubation with trypsin. No change in the digestion pattern was observed in the presence of DNA for any of the constructs compared to the absence of DNA (compare Figure 25 to Figure 18), even with hMutLα full linker, which did have detectable DNA binding (Figure 24).

This may indicate that either the concentration of DNA used was too low to detect any changes in degradation pattern or that DNA does not infer protection against proteolysis of the C-terminal domain of hMutLα.

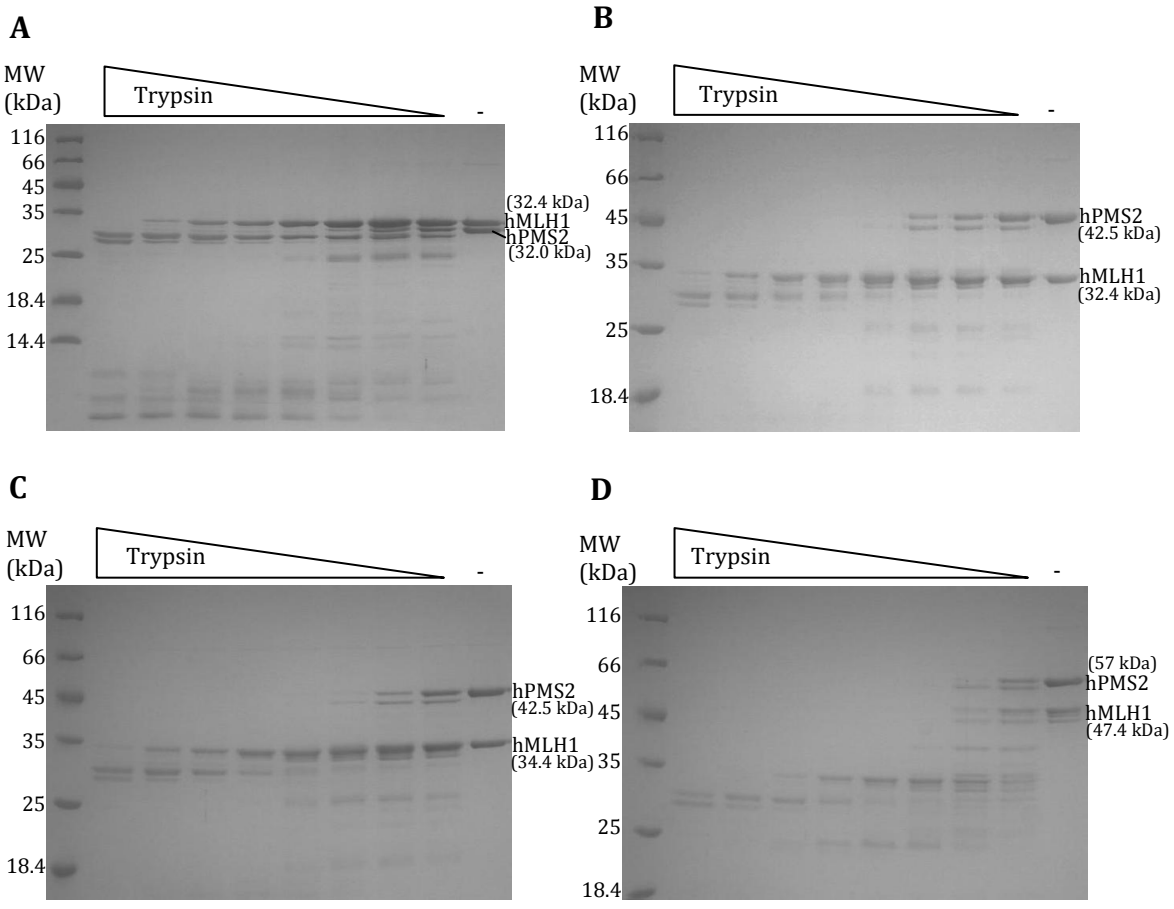

Figure 25. Limited proteolysis of hMutLα constructs in the presence of 548 bp DNA, A) hMutLα minimal linker, B) hMutLα half linker1, C) hMutLα half linker2, and D) hMutLα full linker. 0.5 mg/mL of protein was incubated with 1 µL of 6.25, 3.13, 1.56, 0.78, 0.39, 0.20, 0.10, or 0.05 µg/mL of trypsin in a total volume of 9 µL. Products were separated by SDS-PAGE.

3.5 Endonuclease Activity of hMutLα-CTD

The endonuclease activity of hMutLα full-length protein has been detected previously both in the presence and absence of other MMR factors [33]. In the absence of other factors, hMutLα endonuclease activity was assessed at non-physiological concentrations of KCl (23 mM KCl) [33], because MutL possesses inherently weak endonuclease activity that is difficult to detect at physiological ionic strengths. The activity of hMutLα-CTD in the absence of the NTD has not been assessed previously.

Results from preliminary endonuclease assays of hMutLα-CTD are shown in Figure 26. These experiments were carried out under low salt conditions. Mn^{2+} is used to activate catalysis, since hMutLα endonuclease activity has been previously shown to be activated by this metal ion [33]. pUC19 supercoiled DNA was used as the substrate so that DNA nicked by hMutLα-CTD can be separated from non-nicked supercoiled DNA on an agarose gel.

Endonuclease assays were performed using two constructs, hMutLα minimal linker and hMutLα-CTD full linker (Figure 26). Both constructs exhibit weak endonuclease activity, where activity increases with increasing concentrations of Mn^{2+} (this data was generated by a summer student in the laboratory Anna Zhou). Activity was also assessed in the presence of ATP as a control. ATP is known to stimulate hMutLα endonuclease activity; however, since ATP binding is only exhibited by the N-terminal domain of hMutLα, ATP should not stimulate endonuclease activity of hMutLα-

CTD. As expected, ATP does not stimulate activity (Figure 26). We are currently conducting experiments to determine whether the endonuclease activity of hMutLα-CTD is due to hPMS2 and not a minor contaminant in the protein samples. To that end, Anna Zhou has generated hMutLα-CTD variants encompassing point mutations within the endonuclease site of hPMS2.

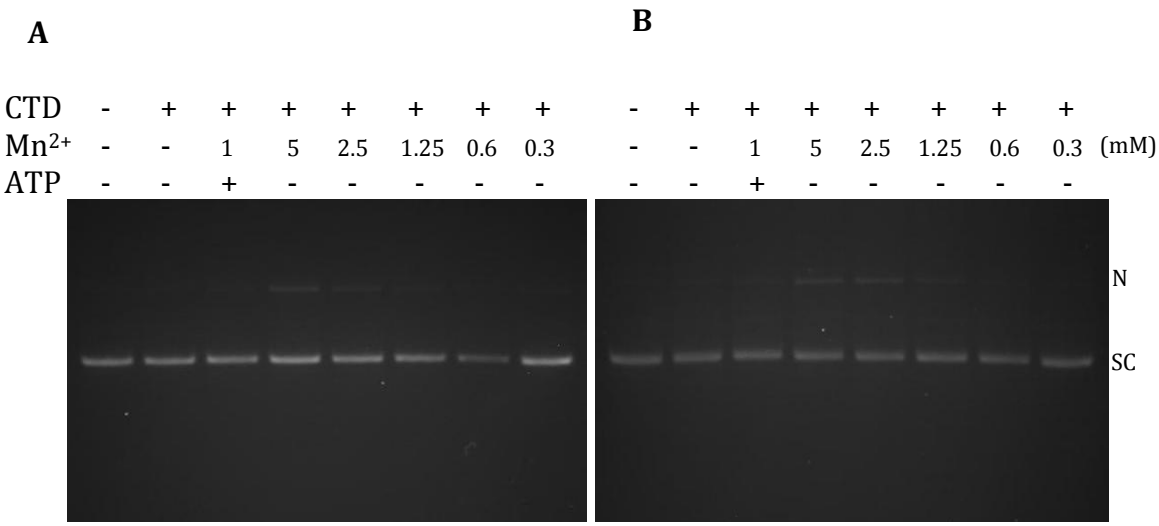

Figure 26. Endonuclease activity of A) MutLα minimal linker and B) MutLα full linker in the presence of increasing concentrations of Mn^{2+}. (N = nicked DNA, SC = supercoiled DNA)

CHAPTER 4: DISCUSSION

In this work, we have established the boundaries for the C-terminal domain of hMutLα and determined that extending the linker region beyond the C-terminal region does not improve the stability of hPMS2. We have also determined that the N-terminal portion of the linker region can bind DNA and that the C-terminal domain of hMutLα contains weak endonuclease activity in the absence of the N-terminal domain or other factors.

4.1 Boundaries of the C-terminal Domain of hMutLα

Characterization of the C-terminal domain of hMutLα is necessary to understand how its endonuclease activity functions in MMR. There is currently discrepancy amongst published data delineating the C-terminal dimerization boundaries of hMutLα. Using GST-pull down assays, Guerrette *et al*. defined the dimerization region to be between hMLH1 506-675 and hPMS2 675-850 [62]. This assay involved expressing either hMLH1 or hPMS2 fused to a GST tag, while deletion mutants of hMLH1 or hPMS2 are produced through *in vitro* transcription and translation. Interaction occurs if both proteins are retained onto GST beads. Using the same assay in combination with yeast two hybrid experiments, Kondo *et al.* reported that dimerization was sufficient between hMLH1 492-742 and hPMS2 612-674 [65] (summarized in Table 1, Figure 6).

In contrast, previous data from our lab reveals that, to form the C-terminal domain of hMutLα, longer regions of both hMLH1 (475-756) and hPMS2 (600-862) are required. In good agreement with our data and that of Kondo *et al*., the homodimerization region of hMLH1 encompasses residues 486-751. However, the structure does not shed light on how hMLH1 may heterodimerize with hPMS2.

It is possible that the techniques used by Guerrette *et al*. and Kondo *et al*. were able to identify regions of interaction between hMLH1 and hPMS2 [62, 65], however those regions are insufficient to support stable dimerization. The ability of hMLH1 and hPMS2 to interact with one another does not necessarily mean that it is a functional interaction. It was shown that a hMLH1 mutant lacking the eight C-terminal residues is able to form a heterodimer with hPMS2, but is still unstable and therefore unable to support MMR *in vivo* [61].

hMutLα variants expressing various linker lengths did not improve the stability of hPMS2 relative to the hMutLα minimal linker construct. Therefore, we conclude that hMutLα minimal linker (hMLH1 475-756, hPMS2 600-862) likely define the boundaries of the C-terminal domain of hMutLα. This domain is still highly unstable and may require external factors for stabilization.

4.2 Differences in Behaviour Between MutL Homologs

The stability of bacterial homologs *Ec*MutL and *Bs*MutL, as well as human homologs hMLH1 and hPMS2 were analyzed in this work. Bacterial MutL homologs exist as homodimers and the results assessing their stability are relatively easier to interpret in comparison to the hMutLα heterodimer. Nevertheless, not all results from the bacterial proteins were as straightforward as originally anticipated.

The C-terminal domain of both *Ec*MutL and *Bs*MutL are stable against trypsin degradation and stable during thermal denaturation, however the results revealed differences in protein unfolding (Figures 9 and 10). *Ec*MutL-CTD unfolded in a single step, while *Bs*MutL-CTD unfolded in two steps, which is indicative of the protein unfolding as two separate domains. We hypothesized that one unfolding step corresponds to the regulatory subdomain and the other to the dimerization subdomain based on the variation in orientation observed in the crystal structure between these two subdomains [44]. The subdomain that unfolds last is stabilized by electrostatic interactions. This led us to search for salt bridges in the dimerization and regulatory subdomains to determine which subdomain was unfolding at which step. The structure of *Ec*MutL-CTD was also analyzed for comparison.

In the structure of *Bs*MutL both bound and unbound to zinc, two salt bridges can be found at the surface of the dimerization subdomain, one on each protomer (R581-E587)(Figure 27A), and one salt bridge was identified in each regulatory subdomain

(E547-K574) (Figure 27B). The higher number of salt bridges located in the dimerization subdomain together with the four-stranded antiparallel β-sheets that make up the dimerization interface and hydrophobic core provide additional stabilizing interactions not present in the regulatory subdomains, therefore the first unfolding step likely corresponds to the regulatory subdomains while the second corresponds to the dimerization subdomain.

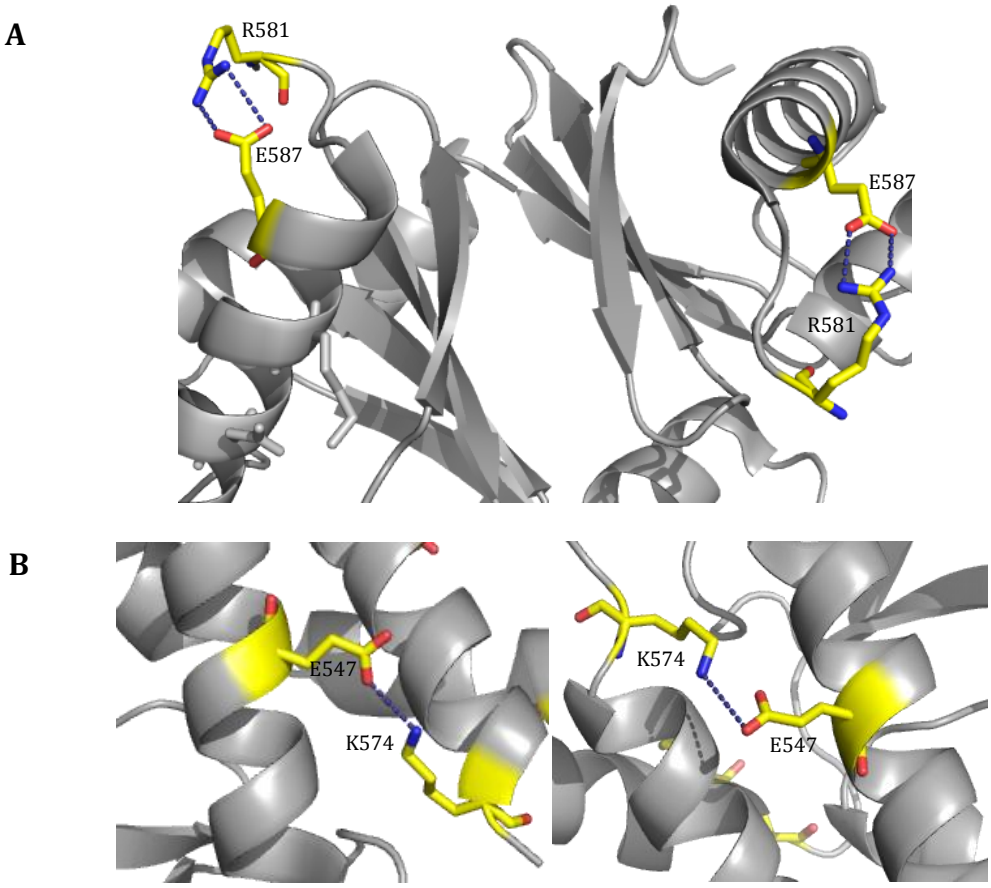

Figure 27. Salt bridges identified in the structure of *Bs*MutL-CTD (displayed from PDB ID: 3KDG, but also found in PDB ID: 3KDK) within A) the dimerization subdomain and B) the regulatory subdomain.

No salt bridges could be identified on the C-terminal domain structure of *Ec*MutL. This supports our results that show *Ec*MutL unfolding as a single domain, as it parallels the melting curve of tagless *Bs*MutL under high salt conditions (Figure 9B and 10C); when the salt bridges are disrupted it unfolds as a single step.

Like the C-terminal domain of *Ec*MutL, hMLH1-CTD unfolds as a single step and no salt bridges could be identified in the structure. The C-terminal domain of hMLH1 also exists as a single domain, which supports our observation of single step unfolding.

hPMS2-CTD on the other hand is more complex and undergoes unfolding in multiple steps, which occur at ~25, 45, and 80 °C (Figure 21A). Drawing from our analysis of the *Bs*MutL-CTD structure, it is likely that the multiple steps correspond to different subdomains of hMutLα. Identification of the region that corresponds to the unfolding step at the lowest temperature could help elucidate methods to stabilize that region. Based on the instability of hPMS2 observed from the limited proteolysis experiments, the region in question is likely contained within hPMS2.

This comparative study between MutL homologs has shown how different hMutLα is from bacterial MutL. Although hPMS2 is unstable, this information may tell us something about how differently hMutLα functions, although we are not at a point of understanding how just yet.

4.3 DNA Binding of the C-terminal Domain of hMutLα and the Linker

The C-terminal domain of hMutLα is unable to bind DNA as assessed by electrophoretic mobility shift assays (Figure 23 and 24). Similarly in *Bs*MutL, DNA binding was not detected for the C-terminal domain [44]. However, DNA binding by the C-terminal domains of *Ng*MutL and *Aa*MutL have been detected through surface plasmon resonance (SPR) analyses, though it was not reported through electrophoretic mobility shift assays [53, 60]. It is possible that the C-terminal domain of hMutLα possesses extremely weak DNA binding activity that is not detectable by electrophoretic mobility shift assays. Analysis using other methods is required to confirm whether or not hMutLα is capable of binding DNA.

hMutLα full linker, which includes the C-terminal domain and the entire linker region, has detectable DNA binding (Figure 24). These results indicate that the N-terminal portion of the linker is important for binding DNA, since other hMutLα variants were unable to bind. This newfound ability of the linker region could in part explain why full-length *Ec*MutL and yMutLα bind DNA with higher affinity than the N-terminal domain alone [43, 57]. However, another explanation for this observation could be that the ATP-induced enclosed conformation of MutL contributes to DNA binding by holding the substrate in place.

4.4 Effect of the Linker on MutL function in MMR

Previous studies have indicated that deletion of the N-terminal portion of the MutL linker or mutation of residues within this region reduces MMR activity [43, 47]. We have established that the N-terminal portion of the linker is able to bind DNA. DNA binding is important for the function of MutL [20, 58], therefore it is possible that this DNA binding capability could improve the efficiency of MMR by helping to secure the DNA substrate. Contrary to this theory, a mutant of *Ec*MutL lacking the 30 N-terminal residues of the linker was shown to exhibit normal DNA binding [43]. Despite the apparent normal activity of this truncated linker mutant, it had a 30-fold increase in mutation frequency compared to wild-type [43]. Although this defect was 10-fold less relative to MutL null mutants, this still indicates that the N-terminal region of the linker plays a role in the function of MutL [43]. The importance of the N-terminal region of the linker is complemented in a yeast study on MLH1 where point mutations of charged residues within the N-terminal region of the linker caused a mutator phenotype, while similar mutations in the intermediate and C-terminal region of MLH1 did not have an effect [47]. The DNA binding ability of the yeast linker mutants was not tested.

One explanation for these observations could be that the DNA binding ability of the linker is more important in MutL homologs that exhibit endonuclease activity. Another possibility could be that the N-terminal portion of the linker possesses the ability to undergo structural changes important for MMR in all organisms. Dimerization

of the N-terminal domains of MutL could result in structural changes of the N-terminal portion of the linker, which may be necessary to cause condensation of the entire linker. Our CD results revealed secondary structure within the linker, and it could have the potential to adopt additional structure, as suggested in previous work [48]. This conformational change in MutL is believed to be important for endonuclease activity as well as roles independent of endonuclease activity, such as mediating interactions with MutH and UvrD in the case of *E. coli* [43, 71, 72]. In this way the linker acts as an extra layer of regulation, requiring first appropriate conditions in the cell before MutL and its associated factors can carry out their function.

4.5 Endonuclease Activity of the C-terminal Domain of hMutLα

Endonuclease activity of the C-terminal domain of hMutLα in the absence of the N-terminal domain was presented for the first time in this work. Our results showed that hMutLα-CTD was able to weakly cleave supercoiled DNA substrates (Figure 26). Endonuclease activity has also been detected in the C-terminal domain of *Ng*MutL and *Aa*MutL [53, 60]. Opposite to these observations, the C-terminal domain of *Bs*MutL was unable to carry out nicking activity on its own, and this defect was attributed to the inability of the domain to bind DNA [44]. Since the C-terminal domain of hMutLα exhibits endonuclease activity, it is likely that hMutLα-CTD does possess minor DNA

binding ability that is not detected using electrophoretic mobility shift assays as discussed.

The ability of the C-terminal domain of hMutLα to nick DNA on its own poses a problem *in vivo*. There would have to be a mechanism of regulation in place to prevent unwanted nicking. Another factor to consider is the conditions under which these experiments were carried out. The reconstituted endonuclease assay presented by Kadyrov *et al.* that initially showed endonuclease nicking was performed under physiological conditions which included 125 mM KCl; an effect which was only seen in the presence of MutLα, ATP, PCNA, RFC, MutSα, and a mismatch [33]. Experiments assessing the endonuclease activity of MutL in the absence of other factors were carried out at non-physiological concentrations of KCl (20-50 mM KCl) because the inherently weak endonuclease activity of MutL is difficult to detect under physiological conditions [33, 37, 44, 60].

Similarly, the experiments reporting DNA binding of the C-terminal domains of NgMutL and AaMutL were conducted at 20 mM and 54 mM KCl respectively [53, 60]. Considering how weak the DNA binding activity of MutL-CTD is, it is likely disrupted by higher salt conditions, and therefore probably does not occur *in vivo*. Experiments that showed the DNA binding ability of the N-terminal domain of MutLα have been carried out under low salt conditions, and full-length yMutLα is unable to bind DNA under physiological salt conditions [40, 57]. Consequently, hMutLα would be unable to cleave

DNA *in vivo* unless the necessary components are in place to activate catalysis; ATP, RFC, PCNA, MutSα, and a mismatch.

RFC functions to load PCNA onto DNA, and its role in stimulating hMutLα endonuclease activity has been restricted to just that, loading PCNA [37]. PCNA is the replication processivity clamp that generally functions to tether proteins to DNA. Interaction between the processivity clamp and MutL has been shown [37, 38, 73], and PCNA has been shown to stimulate endonuclease activity under non-physiological conditions [33]. This interaction is important for MutL function in MMR [44, 73], however it does not offer the complete story of how MutL endonuclease activity is activated *in vivo*. The interaction between MutS and MutL is less well characterized, and the importance of this interaction on MutL endonuclease activity is undetermined.

Evidently the interactions of MutLα with MutSα and PCNA are important for MutLα function in MMR, however there is limited data describing the details of these interactions. How the presence of DNA or ATP affects these interactions are important factors to consider. A more detailed characterization is necessary to determine their role in stimulating the endonuclease activity of MutL.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusion

Understanding the dynamics of hMutLα is important for uncovering how this key MMR protein functions to facilitate repair. This work has examined the stability of hMutLα relative to bacterial MutL homologs and demonstrated the complexity of the human protein. We have revealed the highly unstable nature of the C-terminal domain of hPMS2. It was also shown here that Zn^{2+} binding to hMutLα-CTD induces a local conformational change that we suspect is similar to that seen in the *Bs*MutL structure; however how this ion may influence endonuclease activity remains to be tested.

Exactly how hMutLα carries out endonucleolytic cleavage is still in question, though it is established that the ATP-induced conformational change is important for MutL function. We believe that the linker region helps to facilitate this conformational change across the structure as the linker region possesses very dynamic properties. The C-terminal portion of the linker is required for heterodimer stabilization, while the N-terminal linker region is shown here to possess DNA binding ability.

Our preliminary data reveals that the C-terminal domain of hMutLα is sufficient for endonucleolytic cleavage under low salt conditions. However, these findings need to be further validated.

5.2 Future Directions

Further characterization of hPMS2 is necessary to understand why it so unstable. Independent analyses of the dimerization subdomain and the regulatory subdomain of hPMS2-CTD may help to isolate the problematic area. Identification of a stabilizing binding partner may be one way to answer some questions about how hPMS2 functions and why it behaves so differently from bacterial MutL homologs.

Structural characterization of the C-terminal domain of hMutLα would allow us to visualize the active site, and determine potential residues for DNA interaction. DNA binding of the C-terminal domain of hMutLα should also be examined using other experimental methods. A comparative analysis with structures of bacterial MutL homologs could be performed if the structure of hMutLα-CTD was available, and it would help put into context the work presented here. From there, the interaction surfaces with key binding partners can be more clearly determined, and this may help to uncover their influence on hMutLα activity and regulation.

hMutLα is of course a very dynamic protein and best characterized as a whole. It would be interesting to see whether the ATP-induced conformational change affects structural changes within the CTD as this may affect endonuclease activity. It if is unfeasible to study the intact protein, it would be informative to test the affect of the NTD of both hMLH1 and hPMS2 on endonuclease activity, independently and together.

This would be a systematic way of determining how hMutLα functions and coordinates its multiple activities.

Besides the dynamics of hMutLα, it is important to understand how the changes in hMutLα conformation influences or is influenced by other factors of the MMR system. A clearer characterization of the interactions between hMutSα, hMutLα, and PCNA and how ATP or DNA affects them is crucial for understanding the regulation of hMutLα activity in MMR.

REFERENCES

1. Cline, S.D. and P.C. Hanawalt, *Who's on first in the cellular response to DNA damage?* Nat Rev Mol Cell Biol, 2003. **4**(5): p. 361-72.
2. Marcon, E. and P.B. Moens, *The evolution of meiosis: recruitment and modification of somatic DNA-repair proteins.* Bioessays, 2005. **27**(8): p. 795-808.
3. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity.* Annu Rev Biochem, 2000. **69**: p. 497-529.
4. Modrich, P., *Mechanisms and biological effects of mismatch repair.* Annu Rev Genet, 1991. **25**: p. 229-53.
5. Bhattacharyya, N.P., et al., *Molecular analysis of mutations in mutator colorectal carcinoma cell lines.* Hum Mol Genet, 1995. **4**(11): p. 2057-64.
6. Plotz, G., S. Zeuzem, and J. Raedle, *DNA mismatch repair and Lynch syndrome.* J Mol Histol, 2006. **37**(5-7): p. 271-83.
7. Gologan, A. and A.R. Sepulveda, *Microsatellite instability and DNA mismatch repair deficiency testing in hereditary and sporadic gastrointestinal cancers.* Clin Lab Med, 2005. **25**(1): p. 179-96.
8. Surtees, J.A., J.L. Argueso, and E. Alani, *Mismatch repair proteins: key regulators of genetic recombination.* Cytogenet Genome Res, 2004. **107**(3-4): p. 146-59.
9. Zhang, Y., L.H. Rohde, and H. Wu, *Involvement of nucleotide excision and mismatch repair mechanisms in double strand break repair.* Curr Genomics, 2009. **10**(4): p. 250-8.
10. Jiricny, J., *The multifaceted mismatch-repair system.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.
11. Yoshioka, K., Y. Yoshioka, and P. Hsieh, *ATR kinase activation mediated by MutSalpha and MutLalpha in response to cytotoxic O6-methylguanine adducts.* Mol Cell, 2006. **22**(4): p. 501-10.
12. Liu, Y., et al., *Interactions of human mismatch repair proteins MutSalpha and MutLalpha with proteins of the ATR-Chk1 pathway.* J Biol Chem, 2010. **285**(8): p. 5974-82.
13. Hickman, M.J. and L.D. Samson, *Apoptotic signaling in response to a single type of DNA lesion, O(6)-methylguanine.* Mol Cell, 2004. **14**(1): p. 105-16.
14. Holliday, R.A., *A mechanism for gene conversion in fungi.* Genet Res, 1964. **5**: p. 282-304.
15. Williamson, M.S., J.C. Game, and S. Fogel, *Meiotic gene conversion mutants in Saccharomyces cerevisiae. I. Isolation and characterization of pms1-1 and pms1-2.* Genetics, 1985. **110**(4): p. 609-46.
16. Fishel, R. and R.D. Kolodner, *Identification of mismatch repair genes and their role in the development of cancer.* Curr Opin Genet Dev, 1995. **5**(3): p. 382-95.

17. Acharya, S., et al., *The Coordinated Functions of the E. coli MutS and MutL Proteins in Mismatch Repair.* Mol Cell, 2003. **12**(1): p. 233-246.
18. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.
19. Junop, M.S., et al., *Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair.* Mol Cell, 2001. **7**(1): p. 1-12.
20. Junop, M., *In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination.* DNA Repair (Amst), 2003. **2**(4): p. 387-405.
21. Li, G.M., *Mechanisms and functions of DNA mismatch repair.* Cell Res, 2008. **18**(1): p. 85-98.
22. Lee, J.Y., et al., *MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage.* Mol Cell, 2005. **20**(1): p. 155-66.
23. Yamaguchi, M., V. Dao, and P. Modrich, *MutS and MutL activate DNA helicase II in a mismatch-dependent manner.* J Biol Chem, 1998. **273**(15): p. 9197-201.
24. Ramilo, C., et al., *Partial reconstitution of human DNA mismatch repair in vitro: characterization of the role of human replication protein A.* Mol Cell Biol, 2002. **22**(7): p. 2037-46.
25. Raschle, M., et al., *Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1.* J Biol Chem, 1999. **274**(45): p. 32368-75.
26. Cannavo, E., et al., *Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair.* Cancer Res, 2005. **65**(23): p. 10759-66.
27. Lipkin, S.M., et al., *Meiotic arrest and aneuploidy in MLH3-deficient mice.* Nat Genet, 2002. **31**(4): p. 385-90.
28. Warren, J.J., et al., *Structure of the human MutSalpha DNA lesion recognition complex.* Mol Cell, 2007. **26**(4): p. 579-92.
29. Plotz, G., et al., *hMutSalpha forms an ATP-dependent complex with hMutLalpha and hMutLbeta on DNA.* Nucleic Acids Res, 2002. **30**(3): p. 711-8.
30. Holmes, J., Jr., S. Clark, and P. Modrich, *Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines.* Proc Natl Acad Sci U S A, 1990. **87**(15): p. 5837-41.
31. Thomas, D.C., J.D. Roberts, and T.A. Kunkel, *Heteroduplex repair in extracts of human HeLa cells.* J Biol Chem, 1991. **266**(6): p. 3744-51.
32. Nick McElhinny, S.A., G.E. Kissling, and T.A. Kunkel, *Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}.* Proc Natl Acad Sci U S A, 2010. **107**(49): p. 21070-5.
33. Kadyrov, F.A., et al., *Endonucleolytic function of MutLalpha in human mismatch repair.* Cell, 2006. **126**(2): p. 297-308.
34. Genschel, J. and P. Modrich, *Mechanism of 5'-directed excision in human mismatch repair.* Mol Cell, 2003. **12**(5): p. 1077-86.

35. Dherin, C., et al., *Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair.* Mol Cell Biol, 2009. **29**(3): p. 907-18.
36. Zhang, Y., et al., *Reconstitution of 5'-directed human mismatch repair in a purified system.* Cell, 2005. **122**(5): p. 693-705.
37. Pluciennik, A., et al., *PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair.* Proc Natl Acad Sci U S A, 2010. **107**(37): p. 16066-71.
38. Pillon, M.C., J.H. Miller, and A. Guarne, *The endonuclease domain of MutL interacts with the beta sliding clamp.* DNA Repair (Amst), 2011. **10**(1): p. 87-93.
39. Ban, C. and W. Yang, *Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis.* Cell, 1998. **95**(4): p. 541-52.
40. Guarne, A., M.S. Junop, and W. Yang, *Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase.* EMBO J, 2001. **20**(19): p. 5521-31.
41. Arana, M.E., et al., *Functional residues on the surface of the N-terminal domain of yeast Pms1.* DNA Repair (Amst), 2010. **9**(4): p. 448-57.
42. Ban, C., M. Junop, and W. Yang, *Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair.* Cell, 1999. **97**(1): p. 85-97.
43. Guarne, A., et al., *Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair.* EMBO J, 2004. **23**(21): p. 4134-45.
44. Pillon, M.C., et al., *Structure of the endonuclease domain of MutL: unlicensed to cut.* Mol Cell, 2010. **39**(1): p. 145-51.
45. Kosinski, J., et al., *The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family.* J Mol Biol, 2008. **382**(3): p. 610-27.
46. Namadurai, S., et al., *The C-terminal domain of the MutL homolog from Neisseria gonorrhoeae forms an inverted homodimer.* PLoS One, 2010. **5**(10): p. e13726.
47. Argueso, J.L., et al., *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, 2003. **23**(3): p. 873-886.
48. Sacho, E.J., et al., *Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha.* Mol Cell, 2008. **29**(1): p. 112-21.
49. Deschenes, S.M., et al., *The E705K mutation in hPMS2 exerts recessive, not dominant, effects on mismatch repair.* Cancer Lett, 2007. **249**(2): p. 148-56.
50. Stoll, K.E., et al., *Characterization and structure of the manganese-responsive transcriptional regulator ScaR.* Biochemistry, 2009. **48**(43): p. 10308-20.
51. Kadyrov, F.A., et al., *Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease.* J Biol Chem, 2007. **282**(51): p. 37181-90.

52. Mauris, J. and T.C. Evans, *Adenosine triphosphate stimulates Aquifex aeolicus MutL endonuclease activity.* PLoS One, 2009. **4**(9): p. e7175.
53. Iino, H., et al., *Characterization of C- and N-terminal domains of Aquifex aeolicus MutL endonuclease: N-terminal domain stimulates the endonuclease activity of C-terminal domain in a zinc-dependent manner.* Biosci Rep, 2011. **31**(5): p. 309-22.
54. Fukui, K., et al., *Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL.* J Biol Chem, 2008. **283**(18): p. 12136-45.
55. Bende, S.M. and R.H. Grafstrom, *The DNA binding properties of the MutL protein isolated from Escherichia coli.* Nucleic Acids Res, 1991. **19**(7): p. 1549-55.
56. Hall, M.C., et al., *High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer.* J Mol Biol, 2001. **312**(4): p. 637-47.
57. Hall, M.C., et al., *DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair.* Nucleic Acids Res, 2003. **31**(8): p. 2025-34.
58. Robertson, A., S.R. Pattishall, and S.W. Matson, *The DNA binding activity of MutL is required for methyl-directed mismatch repair in Escherichia coli.* J Biol Chem, 2006. **281**(13): p. 8399-408.
59. Schorzman, A.N., et al., *Modeling of the DNA-binding site of yeast Pms1 by mass spectrometry.* DNA Repair (Amst), 2011. **10**(5): p. 454-65.
60. Duppatla, V., et al., *The C-terminal domain is sufficient for endonuclease activity of Neisseria gonorrhoeae MutL.* Biochem J, 2009. **423**(2): p. 265-77.
61. Mohd, A.B., et al., *Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair.* DNA Repair (Amst), 2006. **5**(3): p. 347-61.
62. Guerrette, S., S. Acharya, and R. Fishel, *The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer.* J Biol Chem, 1999. **274**(10): p. 6336-41.
63. Perera, S. and B. Bapat, *The MLH1 variants p.Arg265Cys and p.Lys618Ala affect protein stability while p.Leu749Gln affects heterodimer formation.* Hum Mutat, 2008. **29**(2): p. 332.
64. Kosinski, J., et al., *Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair.* Hum Mutat, 2010. **31**(8): p. 975-82.
65. Kondo, E., A. Horii, and S. Fukushige, *The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2.* Nucleic Acids Res, 2001. **29**(8): p. 1695-702.
66. Sreerama, N., S.Y. Venyaminov, and R.W. Woody, *Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy.* Protein Sci, 1999. **8**(2): p. 370-80.
67. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochim Biophys Acta, 2005. **1751**(2): p. 119-39.

68. Johnson, W.C., Jr., *Protein secondary structure and circular dichroism: a practical guide.* Proteins, 1990. **7**(3): p. 205-14.
69. Buermeyer, A.B., et al., *The human MLH1 cDNA complements DNA mismatch repair defects in Mlh1-deficient mouse embryonic fibroblasts.* Cancer Res, 1999. **59**(3): p. 538-41.
70. Blackwell, L.J., S. Wang, and P. Modrich, *DNA chain length dependence of formation and dynamics of hMutSalpha.hMutLalpha.heteroduplex complexes.* J Biol Chem, 2001. **276**(35): p. 33233-40.
71. Ahrends, R., et al., *Identifying an interaction site between MutH and the C-terminal domain of MutL by crosslinking, affinity purification, chemical coding and mass spectrometry.* Nucleic Acids Res, 2006. **34**(10): p. 3169-80.
72. Kosinski, J., et al., *Analysis of the quaternary structure of the MutL C-terminal domain.* J Mol Biol, 2005. **351**(4): p. 895-909.
73. Lee, S.D. and E. Alani, *Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA.* J Mol Biol, 2006. **355**(2): p. 175-84.

APPENDIX

A1: Sequence of the 113 bp oligonucleotide

5’-AACATATGAAAACGATTGAAGTTGATGATGAACTCTACAGCTATATTGCCAGCCACACTAAG
CATATCGGCGAGAGCGCATCCGACATTTTACGGCGTATGTTGAAATTTTCC-3’

A2: Sequence of the 548 bp oligonucleotide

5’-TATGAAAACGATTGAAGTTGATGATGAACTCTACAGCTATATTGCCAGCCACACTAAGCATA
TCGGCGAGAGCGCATCCGACATTTTACGGCGTATGTTGAAATTTTCCGCCGCATCACAGCCTGC
TGCTCCGGTGACGAAAGAGGTTCGCGTTGCGTCACCTGCTATCGTCGAAGCGAAGCCGGTCAA
AACGATTAAAGACAAGGTTCGCGCAATGCGTGAACTTCTGCTTTCGGATGAATACGCAGAGCAA
AAGCGAGCGGTCAATCGCTTTATGCTGCTGTTGTCTACACTATATTCTCTTGACGCCCAGGCGTT
TGCCGAAGCAACGGAATCGTTGCACGGTCGTACACGCGTTTACTTTGCGGCAGATGAACAAAC
GCTGCTGAAAAATGGTAATCAGACCAAGCCGAAACATGTGCCAGGCACGCCGTATTGGGTGAT
CACCAACACCAACACCGGCCGTAAATGCAGCATGATCGAACACATCATGCAGTCGATGCAATTC
CCGGCGGAATTGATTGAGAAGGTTTGCGGAACTATCTAAG-3’