

INTERACTIONS OF ATP AND DNA WITH THE MUTL PROTEIN

CHARACTERIZING THE INTERACTIONS OF ATP AND DNA WITH THE MUTL
MISMATCH REPAIR PROTEIN

By MARY CARMEN ORTIZ CASTRO, B. Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the
Requirements for the Degree Master of Science

McMaster University © Copyright by Mary Carmen Ortiz Castro, July 2016

MASTER OF SCIENCE (2016)
(Biochemistry and Biomedical Sciences)

McMaster University
Hamilton, Ontario

TITLE: Characterizing the interactions of ATP and DNA with the
MutL Mismatch Repair protein

AUTHOR: Mary Carmen Ortiz Castro, B.Sc. (Universidad Nacional de
Colombia-Medellín)

SUPERVISOR: Dr. Alba Guarné

NUMBER OF PAGES: xiii, 116

ABSTRACT

The fidelity of DNA replication prevents mutations that may lead to cancer predisposition or neurodegenerative diseases. One mechanism that enhances DNA replication fidelity is DNA mismatch repair, which corrects mismatches and small insertion/deletion loops that have escaped polymerase proofreading. In all eukaryotes and most prokaryotes, MutL (a key mismatch repair protein) has an intrinsic endonuclease activity that nicks the newly synthesized strand and recruits downstream factors to remove and correct errors. It has been proposed that ATP binding promotes a series of conformational changes that induce structural order within MutL and stimulates its endonuclease activity. The C-terminal domain of MutL, which harbors the endonuclease site, does not bind to DNA. This has prevented the molecular characterization of its endonuclease activity. In this thesis, we first show that MutL in *B. subtilis* exhibits asymmetric conformations similar to yeast and human MutL homologs. We also devise a novel approach to bypass the binding defect of the C-terminal domain by using fusion proteins. We find that these fusions bind to DNA specifically and, in the presence of the processivity clamp, can nick DNA. One of these fusion proteins in particular stimulates the nicking activity much more efficiently than the C-terminal domain alone. This work lays the foundation for the mechanistic characterization of the MutL endonuclease and provides a method to stabilize transient protein-DNA interactions.

ACKNOWLEDGMENTS

I would first like to thank God for being my strength and company and for giving me a lifetime of learning experiences in a new country and culture.

Many thanks to my supervisor, Dr. Alba Guarné, for her guidance, encouragement, and for inspiring me to become a leader in science. I would like to thank my committee members, Dr. Lori Burrows and Dr. Marie Elliot, for helpful discussions and valuable advice in my project and for installing confidence in me to continue my graduate career. I also want to acknowledge past and present members of the Guarné Lab that have become not only my colleagues but also friends for life.

I would also like to thank my family for their support. To my parents, Waldo Ortiz and Sara Luz Castro, thank you for always being there for me and teaching me the values of discipline and hard work that were fundamental in completing this degree. To my sisters Maria Claudia and Mariana Carolina, thank you for bringing joy and laughter to my life. Thank you to my entire family- grandparents, cousins, aunts and uncles- who have always celebrated my academic accomplishments. Thank you to Bryan Sullivan for his support and for not doubting for a second that I could do this, even when I did. This degree is as much yours as it is mine.

TABLE OF CONTENTS

Abstract.....	iii
Acknowledgements.....	iv
List of Figures.....	viii
List of Tables.....	x
List of Abbreviations and Symbols.....	xi
Declaration of Academic Achievement.....	xiii
Chapter 1-Introduction.....	1
1.1 DNA replication.....	1
1.2 There are several ways to ensure replication fidelity.....	4
1.3 DNA Mismatch Repair deficiency can lead to cancer.....	5
1.3.1 Mutations in MMR genes have different effects on Lynch Syndrome..	7
1.4 Molecular basis of DNA Mismatch Repair.....	10
1.4.1 MutS recognizes mismatch base pairs and insertion/deletion loops.....	11
1.4.2 Discrimination and removal of the erroneous nascent strand.....	14
1.4.3 MutL plays a central role in DNA Mismatch Repair.....	17
1.4.3.1 The N-terminal domain of MutL is an ATPase.....	19
1.4.3.2 The C-terminal domain of MutL harbors the endonuclease site...	22
1.4.3.3 MutL is a highly regulated endonuclease.....	25
1.5 Thesis objectives.....	27
Chapter 2-Conformational change of MutL upon binding of ATP.....	28
2.1 Materials and Methods.....	29
2.1.1 Expression and protein purification of MutL.....	29
2.1.2 Gel filtration assays.....	30
2.1.3 Dynamic Light Scattering.....	31
2.1.4 Thermal stability of MutL (Differential scanning fluorimetry).....	31
2.1.5 Time course experiments for <i>B. subtilis</i> and <i>E. coli</i> MutL.....	32
2.1.6 Limited trypsin proteolysis of MutL.....	32
2.1.7 Sample preparation of MutL for Atomic Force Microscopy analysis.....	33
2.1.8 Atomic Force Microscopy.....	33
2.2 Results.....	34
2.2.1 MutL conformational change in <i>E. coli</i>	34

2.1.2 Experimental conditions for <i>E. coli</i> MutL are not optimal for <i>B. subtilis</i> MutL.....	36
2.1.3 Optimization of gel filtration for <i>B. subtilis</i> MutL.....	38
2.1.4 Influence of different parameters on the thermal stability of <i>B. subtilis</i> MutL.....	40
2.1.4.1 Effect of pH.....	41
2.1.4.2 Effect of KCl concentration.....	44
2.1.4.3 Effect of imidazole concentration.....	45
2.1.4.4 Influence of ATP and MgCl ₂	46
2.2.5 <i>B. subtilis</i> MutL does not undergo a conformational change similar to <i>E. coli</i> MutL upon binding of ATP.....	48
2.2.6 ATP binding protect MutL from protease degradation.....	50
2.2.7 Optimization of the purification conditions of MutL for AFM analysis.....	53
2.2.8 <i>B. subtilis</i> MutL undergoes several conformational changes in the presence of ATP.....	59
2.3 Discussion.....	62

Chapters 3-A novel approach to study the endonuclease mechanism of the MutL mismatch repair protein.....	66
3.1 Materials and Methods.....	67
3.1.1 Cloning of SKN1-MutL _{CTD} fusions with varying linker lengths, and lacking endonuclease activity.....	67
3.1.2 Expression and protein purification of SKN1-MutL _{CTD} fusions and variants lacking endonuclease activity.....	69
3.1.3 Expression and purification of <i>B. subtilis</i> MutL _{CTD}	70
3.1.4 Expression and purification of <i>B. subtilis</i> β -sliding clamp.....	70
3.1.5 Expression and protein purification of SKN1-L8-MutL _{CTD} heterodimer.....	71
3.1.6 Cloning, expression and purification of SKN1-L8-MutL _{CTD} variants... ..	73
3.1.7 Generation of linear 195 bp substrates.....	73
3.1.8 Protein-DNA binding assays.....	75
3.1.9 Mismatch independent endonuclease assays for SKN1-MutL _{CTD} fusions.....	76
3.1.10 Design of fluorescently labeled molecular weight marker.....	76
3.2 Results.....	77
3.2.1 Characterization of the SKN1-MutL _{CTD} fusions.....	77

3.2.2 The length of the linker affects the stability of the fusion.....	80
3.2.3 Functional studies of L32 fusion in complex with linear 195 bp substrates.....	82
3.2.3.1 L32 fusion binds DNA specifically.....	83
3.2.3.2 L32 fusion has β -dependent endonuclease activity.....	84
3.2.3.3 L32 fusion has less β -dependent nicking activity than the CTD of MutL.....	86
3.2.4 Characterization of SKN1-MutL _{CTD} fusion variant with shorter linker.	87
3.2.5 Functional studies of the heterodimer fusion L8* in complex with linear 195 bp substrates.....	90
3.2.5.1 L8* has more β -dependent endonuclease activity than the L32 fusion and the CTD of MutL.....	90
3.2.5.2 Effect of the position of the SKN1 site and orientation imposed by β -clamp on the endonuclease activity of MutL.....	94
3.2.5.3 Pre-nicked substrate do not stimulate the β -dependent endonuclease activity of MutL.....	96
3.3 Discussion.....	98
Chapter 4-Conclusions and future directions.....	103
References.....	105

LIST OF FIGURES

Figure 1.1 Schematic representation of the DNA replication in the leading and lagging strand.....	3
Figure 1.2 Classification of DNA mutations in MMR genes that cause Lynch Syndrome.....	8
Figure 1.3 InSIGHT classification of DNA variants in MMR genes.....	9
Figure 1.4 Crystal structure of <i>E. coli</i> MutS.....	12
Figure 1.5 Mechanisms of mismatch recognition and strand discrimination in different organisms.....	16
Figure 1.6 Architecture of MutL.....	18
Figure 1.7 MutL conformational changes induced by ATP in different organisms.....	21
Figure 1.8 Structural organization of the C-terminal domain of <i>B. subtilis</i> MutL.....	24
Figure 2.1 Purification of <i>E. coli</i> MutL.....	35
Figure 2.2 Size exclusion chromatography profile of <i>E. coli</i> MutL in the absence or presence of AMPPNP.....	36
Figure 2.3 Purification of <i>B. subtilis</i> MutL.....	37
Figure 2.4 Size exclusion chromatography profile of <i>B. subtilis</i> MutL.....	38
Figure 2.5 Stability of <i>B. subtilis</i> MutL under different conditions using DLS.....	39
Figure 2.6 Size exclusion chromatography of <i>B. subtilis</i> MutL.....	40
Figure 2.7 Effect of pH on the thermostability of <i>B. subtilis</i> MutL.....	44
Figure 2.8 Effect of KCl concentration on the thermostability of <i>B. subtilis</i> MutL....	45
Figure 2.9 Effect of imidazole concentration on the thermostability of <i>B. subtilis</i> MutL.....	46
Figure 2.10 Effect of ATP and MgCl ₂ on the thermal stability of MutL.....	47
Figure 2.11 Size exclusion chromatography profile of <i>B. subtilis</i> and <i>E. coli</i> MutL for 2 h at 4°C in the absence or presence of 5 mM AMPPNP.....	49
Figure 2.12 Trypsin proteolysis of <i>E. coli</i> , <i>B. subtilis</i> and <i>T. aquaticus</i> MutL in the absence or presence of 1 mM ATP.....	51
Figure 2.13 Trypsin proteolysis of <i>E. coli</i> MutL and <i>B. subtilis</i> MutL under different concentrations of ATP.....	52
Figure 2.14 Effect of acetate salts on the stability of <i>B. subtilis</i> MutL in the presence of ATP.....	55
Figure 2.15 Stability of <i>T. aquaticus</i> MutL under different buffers agents and salts...	59
Figure 2.16 AFM images of <i>B. subtilis</i> MutL.....	60
Figure 2.17 <i>B. subtilis</i> MutL conformational states distribution in the absence and presence of 1 mM ADP or 1 mM ATP.....	61

Figure 3.1 Schematic representation of the fusion protein.....	66
Figure 3.2 Small scale induction of L30 fusion.....	78
Figure 3.3 Purification of L30 fusion.....	79
Figure 3.4 Purification of L32 fusion.....	81
Figure 3.5 Electrophoretic mobility shift assays of L32 fusion with 195 bp substrate with the SKN1 site in the middle.....	84
Figure 3.6 Endonuclease activity of L32 fusion and variants with 195 bp linear substrate with SKN1 site in the middle.....	85
Figure 3.7 Endonuclease activity of CTD and L32 fusion with 195 bp linear substrates.....	87
Figure 3.8 Purification of L8*.....	89
Figure 3.9 Endonuclease activity of L8* using 195 bp linear substrate with the SKN1 site in the middle.....	90
Figure 3.10 Endonuclease activity of MutL full length, CTD, and fusion proteins using 195bp linear substrate with SKN1 site in the middle.....	91
Figure 3.11 Endonuclease activity of L8* with 195 bp linear substrates.....	92
Figure 3.12 Endonuclease activity of L8* and variants using 195 bp linear substrate with SKN1 site in the middle.....	94
Figure 3.13 Endonuclease activity of L8* using 195 bp linear substrates single or double labeled.....	96
Figure 3.14 Effect of the nicking enzyme Nt.AlwI on the endonuclease activity of L8*.....	98

LIST OF TABLES

Table 1.1 Components of the replisome in different systems.....	1
Table 1.2 Lifetime risk of cancer reported in Lynch Syndrome families	6
Table 2.1 Conditions for thermofluor of <i>B. subtilis</i> MutL	42
Table 2.2 Conditions for thermofluor of <i>B. subtilis</i> MutL using acetate salts.....	57
Table 2.3 Buffer screen to test the stability of <i>T. aquaticus</i> MutL	58
Table 3.1 Fusion proteins with varying linker length.....	68
Table 3.2 Primers used to generate SKN1-MutL _{CTD} fusions and dead variants.....	68
Table 3.3 Primers used to generate different versions of pUC19 and molecular weight marker.....	75

LIST OF ABBREVIATIONS AND SYMBOLS

α	Alpha
ADP	Adenosine diphosphate
AMPPNP	5'-adenylyl- β - γ -imidodiphosphate
ATP	Adenosine triphosphate
AFM	Atomic force microscopy
bp	Base pair
BM(PEO) ₃	1,11-bismaleimido-triethyleneglycol
β	Beta
BSA	Bovine serum albumin
6-FAM	6-Carboxyfluorescein
CV	Column volume
CMCB	Centre for Microbial Chemical Biology
CTD	C-terminal domain
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
°	Degree
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside-triphosphate
DtxR	Diphtheria toxin repressor
DTT	Dithiothreitol
DLS	Dynamic light scattering
EDTA	Ethylenediaminetetraacetic acid
V _e	Elution volume
EMSA	Electrophoretic mobility shift assay
EXO I	Exonuclease I
γ	Gamma
GHKL	Gyrase, Hsp90, Histidine Kinase, MutL ATPase family
HNPCC	Hereditary Non-polyposis Colorectal Cancer
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
h	Hour
IDLs	Insertion/deletion loops
IPTG	Isopropyl- β -D-thiogalactopyranoside
MagAc	Magnesium acetate
β -Me	β -Mercaptoethanol
MES	2-(N-morpholino) ethanesulfonic acid
MLH1	MutL homolog 1

MLH3	MutL homolog 3
MMR	DNA mismatch repair
min	Minute
MOBIX	McMaster Institute for Molecular Biology and Biotechnology
MWCO	Molecular weight cut-off
M	Molecular weight marker
MOPS	3-(N-morpholino) propanesulfonic acid
MSH2	MutS protein homolog 2
MSH3	MutS protein homolog 3
MSH6	MutS protein homolog 6
NIH	National Institutes of Health
NTD	N-terminal domain
NgvrB	DNA Gyrase B
OD ₆₀₀	Optical density measured at 600 nm
PotAc	Potassium acetate
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PMS2	Postmeiotic segregation increased 2
PMSF	Phenylmethylsulfonyl fluoride
pol	Polymerase
RT	Room temperature
SSB	Single strand binding protein
SAXS	Small angle X-ray scattering
s	Seconds
SodAc	Sodium acetate
SDS	Sodium dodecyl sulphate
PAGE	Polyacrylamide gel electrophoresis
TCEP	Tris (2-carboxyethyl) phosphine
UV	Ultraviolet
v/v	Volume per volume
V	Volts

DECLARATION OF ACADEMIC ACHIEVEMENT

I purified the MutL proteins from *Escherichia coli* and *Bacillus subtilis* according to the purifications protocols of Guarné et al. (2004) and Pillon et al. (2010), respectively. I devised the protocol for the over-expression and purification of *Thermus aquaticus* MutL. Atomic Force Microscopy was conducted by Hunter Wilkins using protein that I purified. Julia Cai cloned, over-expressed, and devised the initial purification protocol of the fusion proteins with different linker lengths. I tested new fusion proteins and optimized the purification protocol further with the most stable fusion. I designed and conducted the electrophoretic mobility shift assays and endonuclease assays using fluorescently labeled substrates. Linda Liu and I designed and purified the heterodimer fusion protein. All the experiments and findings were interpreted by myself and Dr. Alba Guarné.

CHAPTER 1

INTRODUCTION

1.1) DNA Replication

DNA replication is necessary every time a cell divides in order to make a copy of the original DNA molecule and pass it to the next generation. To duplicate the genetic material, cells use a multi-protein complex, known as the replisome, which accurately synthesizes a DNA strand that is complementary to the template strand (Meselson & Stahl, 1958). This replisome is highly conserved across viruses, bacteria, archaea and eukaryotes (Table 1.1) (Yao & O'Donnell, 2010).

Table 1.1: Components of the replisome in different systems*

Component	T7 phage (virus)	<i>Escherichia coli</i> (bacteria)	Eukaryotes	Archaea
Helicase	gp4	DnaB	MCM2-7/CMG	MCM
Clamp loader	-	γ/τ complex	RFC	RFC
Processivity Clamp	-	β	PCNA	PCNA
Polymerase	gp5/TRX	Pol III	Pol $\alpha/\delta/\epsilon$	Pol
Primase	gp4	DnaG	Pol α -Primase	Primase
Single strand binding protein (SSB)	gp2.5	SSB	RPA	RPA
Other	Unknown	Unknown	GIN5, Cdc45	Unknown

*Table extracted from Yao & O'Donnell (2010).

The replisome assembles at the replication fork, where the double-stranded DNA is separated into two individual strands by the replicative helicase. Then, single strand binding proteins (SSB) stabilize the exposed single-strand DNA by removing DNA

secondary structures that would impede replication (Yao & O'Donnell, 2010). The DNA polymerase uses the single-strand DNA as a template to synthesize complementary DNA in the 5' to 3' direction (Yao & O'Donnell, 2010). This directionality, in addition to the antiparallel nature of the DNA strands, requires that one strand (the leading strand) is copied continuously and the other one (the lagging strand) discontinuously in a series of 1-2 kilobase pieces for prokaryotes, or 100-200 bases for eukaryotes (Yao & O'Donnell, 2010).

In *Escherichia coli*, the major replicative polymerase of both strands, Pol III, is composed of α , ϵ , and θ subunits. The α subunit has the polymerase activity, the ϵ subunit has 3' to 5' exonuclease activity which is responsible for removing replication errors, and the θ subunit stimulates the ϵ subunit's proofreading activity (Kelman & O'Donnell, 1995). The situation is more complex in eukaryotes, in which Pol ϵ and Pol δ are the primary leading- and lagging-strand replicases, respectively (Kunkel & Erie, 2015). Although Pols ϵ and δ differ in structure, subunit composition, protein partnerships, processivity and fidelity, they both contain the 3'-exonuclease activity that can proofread replication errors (Johansson & Dixon, 2013).

The leading strand is copied thanks to the coordinated action of the primase, clamp loader, polymerase, and the processivity clamp. The primase synthesizes an RNA primer that serves as initial sequence to start the replication. The clamp loader opens the processivity clamp and subsequently closes it around the template DNA. The processivity clamp then ensures that the polymerase stays attached to DNA, conferring high processivity to the reaction (Kong et al., 1992).

Lagging strand replication proceeds by the formation of Okazaki fragments (Okazaki et al., 1968). In this case, the primase synthesizes an RNA (prokaryotes) or RNA/DNA hybrid (eukaryotes) primer of 10-12 nucleotides (Frick & Richardson, 2001). Then, the clamp loader loads the processivity clamp onto the primer, and the polymerase associated with the processivity factor forms the first Okazaki fragment. Once the Okazaki fragment has been completed, the polymerase releases the clamp and is ready to bind to another clamp on a newly synthesized primer upstream of the sequence (Stukenberg et al., 1994) and, in that way, it begins the next Okazaki fragment. This process continues until the entire lagging strand is copied. Next, the RNA primers are removed and replaced with DNA during Okazaki fragment maturation. Finally, the DNA ligase joins the fragments into a continuous DNA chain (Yao & O'Donnell, 2010) (Figure 1.1).

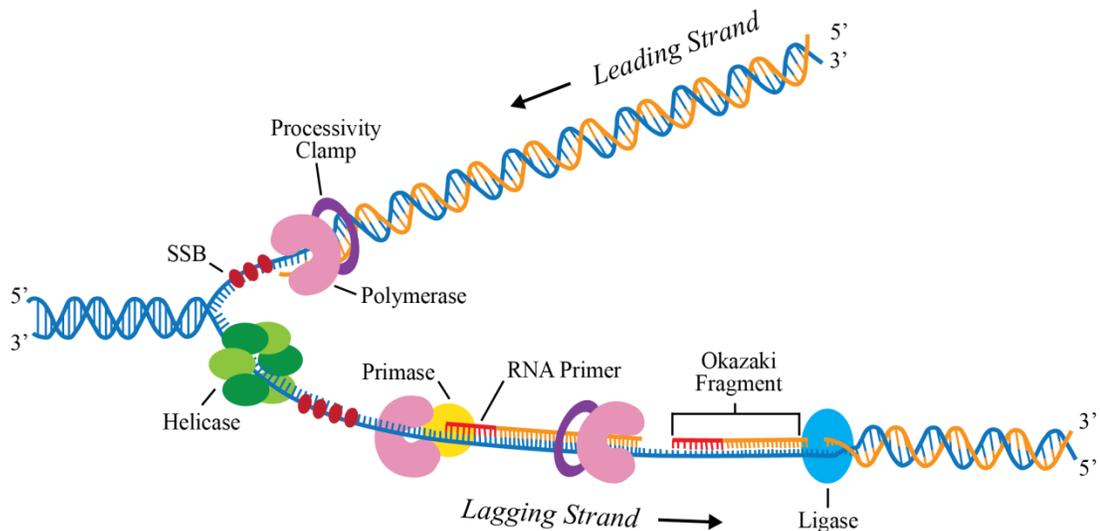


Figure 1.1: Schematic representation of the DNA replication in the leading and lagging strands. Arrows indicate the direction of strand extension.

1.2) There are several ways to ensure replication fidelity

DNA replication requires high fidelity because the genetic information has to be preserved over many generations while preventing the accumulation of mutations that may lead to cancer predisposition or neurodegenerative diseases.

Replicative DNA polymerases incorporate non-complementary nucleotides with a frequency of approximately 1:10,000 to 1:100,000 (Arana & Kunkel, 2010). Polymerases' accuracy is enhanced by the 3'-exonuclease activity encoded in a separate domain of their catalytic subunits allowing a first attempt for repair (Reha-Krantz, 2010). Proofreading occurs when the abnormal geometry of mismatches slows polymerization, promotes fraying, and allows excision of the incorrect bases (Kunkel & Erie, 2015). Although the proofreading process is very efficient (increasing fidelity by approximately two orders of magnitude (Jiricny, 2013)), it is affected by sequence context. A clear example is the little, if any, proofreading of 8-oxoG-dA mismatches whose geometry mimics that of correct base pairs (Kunkel & Erie, 2015) or the inefficient proofreading of insertion and deletion mismatches generated by strand slippage in long repetitive sequences (Kroutil et al., 1996).

Fortunately, there is an additional mechanism that enhances replication fidelity. This mechanism is known as DNA Mismatch Repair (MMR), and is considered the major guardian of genome stability against mismatches and small insertion/deletion loops (IDLs) during DNA replication. MMR increases replication fidelity by 50-1000-fold (Hsieh & Yamane, 2008).

Mismatches represent a unique type of “DNA damage” because they consist entirely of undamaged DNA that is not paired correctly. If the duplex containing a mismatch is unwound, neither strand will contain repairable “damage”, which will lead to a 50% increase in progeny DNA mutations after replication (Jiricny, 2013). This implies that DNA replication and MMR are two well-coordinated processes because the mismatches must be corrected before the next round of replication. Indeed, both processes use several common proteins. For example, the processivity sliding clamp has multiple and important roles in both MMR and replication (Kunkel & Erie, 2015).

1.3) DNA Mismatch Repair deficiency can lead to cancer

The current model of cancer initiation, in which a random unrepaired point mutation eventually results in an alteration of the coding sequence of a key oncogene or tumor suppressor, has been reinforced by the relationship of deficient DNA repair mechanisms and proliferation of mutagenic events in precancerous cells (Jeggo et al., 2016).

One example of a defective DNA repair pathway responsible for cancer initiation is the one related to Lynch Syndrome. This syndrome (previously known as Hereditary Non-Polyposis Colorectal Cancer or HNPCC) is an autosomal dominant disorder caused by a defect in one of the MMR genes. Individuals carrying some of the MMR mutagenic genes are at much higher risk of developing colorectal cancer, endometrial cancer, and other associated cancers than the average risk population at an early age (<45 years) (Table 1.2) (Peltomäki, 2003; Li, 2008; Vasen et al., 2013; Vasen et al., 2007; Siegel et al., 2015).

Lynch Syndrome patients account for about 3% of all cases of colorectal cancer (De la Chapelle, 2004; Giardiello et al., 2014). This is significant because colorectal cancer was the second leading cause of cancer-related deaths in the United States in 2015 with more than 49,700 deaths (Tiwari et al., 2015). Moreover, endometrial cancer is the second most common cancer in Lynch Syndrome patients, but in affected women, it represents a higher risk than colorectal cancer (Umar et al., 2004).

Table 1.2: Lifetime risk of cancer reported in Lynch Syndrome families

Colorectal cancer	80%
Endometrial cancer	20-60%
Ovarian cancer	0.3-20%
Gastric cancer	5-10%
Small bowel cancer	0.4-12%
Urinary tract cancer	0.2-25%
Central Nervous System tumors	1-4%

Although Lynch Syndrome was first described in 1913 (Warthin, 1913), its implications were not evaluated until 1966 when Henry Lynch and his colleagues reported two large families with hereditary colorectal cancer from the Midwest (Lynch et al., 1966). In the early 1990s, the gene defects discovered in Lynch Syndrome patients were mainly caused by germline mutations in the MMR genes —first in MSH2 (Fishel et al., 1993; Leach et al., 1993), MLH1 (Papadopoulos et al., 1994; Bronner et al., 1994), MSH6 (Edelmann et al., 1997), and more recently in MLH3 (Wu et al., 2001) and PMS2 (Worthley et al., 2005). Since the proteins derived from these genes are directly involved in the MMR pathway, it is not surprising that an abnormality in any of them leads to an increase in the genomic mutation rate that results in rapid carcinogenesis (Li, 2008).

MMR deficiency can also be caused by the hypermethylation of the MLH1 promoter (Mensenkamp et al., 2014). This is a common feature in sporadic colorectal cancers that accounts for ~12% of all colorectal cancers. For patients with this type of cancer, it is important to determine whether the cause is sporadic or hereditary. This is because diagnosing a patient with Lynch Syndrome may mean that the patient's family members are at an increased risk of developing colorectal cancer due to their genetics and not by random mutations.

Clinically, the identification of Lynch Syndrome families has become crucial in order to apply efficient screening programs that allow early detection and preventive measures to delay and/or reduce the chance of developing the malignant disease (Bartuma et al., 2012). Additionally, several research projects have studied the molecular mechanism underlying the MMR system and its connections to other biological pathways to better understand the cancer development process and to identify possible targets for preventive and therapeutic interventions (Peltomäki, 2003).

1.3.1) Mutations in MMR genes have different effects on Lynch Syndrome

Pathogenic MMR-gene mutations are found in up to 70-80% of families reported with Lynch Syndrome (Peltomäki, 2003). Their nature is extremely variable: non-sense, missense, frameshift mutations, insertions, or complete deletions of DNA portions can all be responsible for defective MMR systems (Xie et al., 2010). To date, the International Society for Gastrointestinal Hereditary Tumors Incorporated (InSIGHT, <http://www.insight-group.org>) has reported more than 14,000 cases of variants in MMR

genes causing Lynch Syndrome. From the total number of variants, 47% affect MLH1, 36% MSH2, 11% MSH6, 5% PMS2 and 1% MLH3 (Figure 1.2).

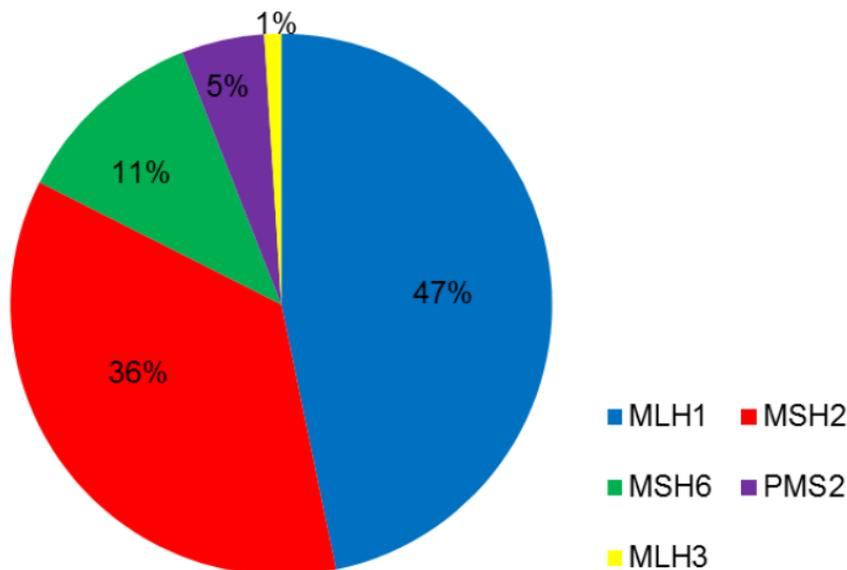


Figure 1.2: *Classification of DNA mutations in MMR genes that cause Lynch Syndrome*

Additionally, the InSIGHT group has described five standardized classes to allow quantitative assessment of variant pathogenicity in the Colon Cancer Family Registry (Plon et al., 2008; Spurdle et al., 2011). These classes range from pathogenic (Class 5) if they are clinically relevant in a genetic counseling setting to non-pathogenic (Class 1) if they do not show evidence of being a dominant high-risk mutation for the patients (Spurdle et al., 2011). Mutations pertaining to Class 5 are usually nonsense interrupting the functional protein domains, large deletions, or large duplications. Classes 4, 3, 2, and 1 represent variants that are likely pathogenic, uncertain, likely not pathogenic, and not pathogenic, respectively.

The InSIGHT classification according the degree of pathogenicity in MMR genes shows that despite MLH1 having the most reported cases; it is also the gene with more non-pathogenic mutations. On the contrary, the MSH2 gene has the majority of pathogenic mutations. Since both of these genes have the highest penetrance, it is expected that class 5 is predominant among them. PMS2 and MSH6 genes do not overpass MLH1 and MSH2 genes, but they have significant amounts of variants that are pathogenic, uncertain, or non-pathogenic (Class 5, 3, and 1) describing their increasing relevance in a medical setting. This InSIGHT classification also includes unclassified variants that do not have any description due to a lack of biochemical and clinical evidence (Figure 1.3).

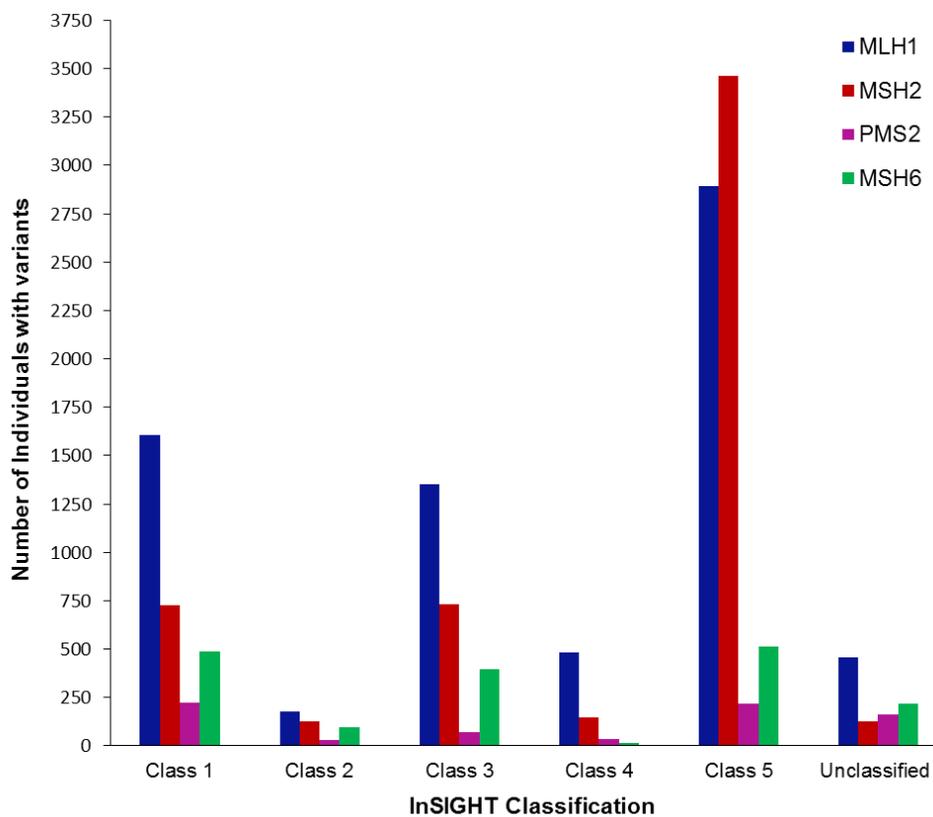


Figure 1.3: *InSIGHT* classification of DNA variants in MMR genes.

Although the InSIGHT organization has provided a very extensive distribution of the mutations in MMR genes that are associated with Lynch Syndrome, there are still several issues at the medical and research levels in assessing the functional implications of the variations with unknown significance (Martín-López & Fishel, 2013). This can be explained because there are no good predictions of the consequences of newly detected mutations. Clearly, this represents a difficult task for the clinical field because genetic testing is currently only authorized for individuals who already have suspected hereditary colorectal cancer, making early detection methods almost impossible.

Several questions need to be addressed related to the detection of new mutations. For example: How can the detection methods improve? How can the clinical and molecular diagnosis technologies coordinate to give accurate predictions of these new mutations where significant family history is lacking? Answers to these questions may only be resolved after identifying and understanding the biophysical functions of the domains in the MMR proteins in order to establish a relationship with the severity of the disease and elucidate the role of the MMR pathway in human tumorigenesis of uncharacterized and characterized variants (Martín-López & Fishel, 2013).

1.4) Molecular basis of DNA Mismatch Repair

The MMR process is present in all living organisms (Jiricny, 2013). The high conservation of the pathway has allowed the use of model organisms such as *Escherichia coli* or *Saccharomyces cerevisiae* to understand the molecular insights of the MMR reaction. Other roles in cellular responses to environmental stress, meiotic recombination,

cell cycle checkpoint control, apoptosis, somatic hypermutation of immunoglobulin genes, and repair of aberrant triplet-repeat expansion strengthen the importance of studying the MMR process in the human cancer field (Kadyrov et al., 2006; Reyes et al., 2015; Kunkel & Erie, 2015).

The basic steps of the MMR system include recognition of the mismatch itself, distinction of the newly synthesized strand from the parental strand, excision and removal of the strand with the wrong nucleotide, and re-synthesis of the erroneous fragment. While MutS and MutL proteins play a central role in the initial steps of the process, the late steps are coordinated by exonucleases (Schmutte et al., 2001; Kadyrov et al., 2009) as well as the replication machinery (Table 1.1). Since the factors involved in the excision and resynthesis are common and well-studied in many other repair pathways (Paull & Gellert, 1998; Viswanathan & Lovett, 1998; Schmutte et al., 1998; Schmutte et al., 2001; Nimonkar et al., 2008; Lee & Wilson, 1999; Lenain et al., 2006), research has been focused on the early steps in order to understand the proper coordination between DNA replication and mismatch repair.

1.4.1) MutS recognizes mismatch base pairs and insertion/ deletion loops

MutS is a modular protein that binds to mismatched DNA (Lamers et al., 2000; Obmolova et al., 2000). In prokaryotes, MutS forms homodimers, while in eukaryotes it functions as heterodimers (Reyes et al., 2015). The eukaryotic heterodimer complexes, MSH2-MSH6 (MutS α), and MSH2-MSH3 (MutS β), have been described based on the nature of the mismatch they recognize. MutS α recognizes base-base mismatches and

small IDLs of 1-2 nucleotides, while MutS β can distinguish IDLs containing up to 16 extra nucleotides in one strand (Kunkel & Erie, 2005).

The general architecture of MutS is conserved from bacteria to eukaryotes (Obmolova et al., 2000; Lamers et al., 2000; Warren et al., 2007; Gupta et al., 2012). Each MutS protomer contains six structural domains with distinct functions (N-terminal mismatch-recognition domain, connector domain, core domain, clamp domain, ATPase domain, and helix-turn helix (HTH) domain) (Obmolova et al., 2000; Lamers et al., 2000; Mendillo et al., 2007) (Figure 1.4).

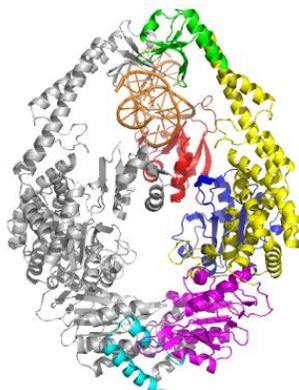


Figure 1.4: *Crystal structure of E. coli MutS (PDB: 1E3M, (Lamers et al., 2000)). Description of the domains of one MutS protomer (N-terminal mismatch-recognition domain (red), connector domain (blue), core domain (yellow), clamp domain (green), ATPase domain (magenta), and helix-turn helix domain (cyan)) while bound to a G/T mismatch (orange).*

One important factor for the recruitment of MutS to the sites of damage is the processivity clamp (Kunkel & Erie, 2005). MutS interacts with this factor (β clamp in bacteria or PCNA in eukaryotes) through the exposed MutS- β clamp binding surface (Simmons et al., 2008). It has been proposed that the processivity sliding clamp stabilizes

MutS at base mispairs by repetitive loading of MutS at the mismatch site in the early steps of MMR (Flores-Rozas et al., 2000; Simmons et al., 2008).

Bacterial MutS and its eukaryotic homolog (MSH6 subunit of MutS α) have a Phe-X-Glu motif that mediates binding to the mismatch (Lamers et al., 2000; Obmolova et al., 2000). The protein anchors itself on the substrate by inserting the phenylalanine residue of the conserved motif into the minor groove of the helix at the mismatch site (Jiricny, 2006). This induces a well-defined kink in the DNA at the mismatch by $\sim 60^\circ$ (Gupta et al., 2012; Warren et al., 2007). Although DNA bending serves important roles in mismatch identification and specificity (Lamers et al., 2000; Obmolova et al., 2000; Natrajan et al., 2003; Wang et al., 2003; Warren et al., 2007; Kunkel & Erie, 2005), binding alone is not sufficient to induce repair (Su et al., 1988). In fact, MutS contains two ATPase active sites that are essential for MMR (Alani et al., 1997; Groothuizen et al., 2015; Hingorani, 2016).

There is a general understanding that after mismatch recognition, MutS undergoes an ATP-dependent conformational change (or changes) (Junop et al., 2001; Jacobs-Palmer & Hingorani, 2007; Jeong et al., 2011; Qiu et al., 2012; Sharma et al., 2013; Sharma & Hingorani, 2013; Hingorani, 2016) to a mobile-clamp state that can move along the DNA (Lyer et al., 2006; Lee et al., 2014). Even though the ATPase activity of MutS is required for its interaction with MutL to initiate repair (Friedhoff et al., 2016), there is not a complete understanding of how mismatch recognition by MutS results in the ATP-dependent recruitment of MutL that enables strand specific nicking on the newly synthesized strand.

There are several models to explain the MutS-MutL mismatch complex formation and the subsequent signaling of repair (Qiu et al., 2015). In one of the models, MutL joins MutS to form MutS-MutL sliding clamps that diffuse along the DNA to interact with the strand-discrimination signal (Gorman et al., 2012). Other models include MutL trapping of MutS clamps near the mismatch followed by DNA looping (Qiu et al., 2015) or MutS-induced polymerization of MutL along the DNA to reach the strand discrimination signal (Lyer et al., 2006; Kunkel & Erie, 2005; Hombauer et al., 2011; Elez et al., 2012).

Single molecule fluorescence studies of *Thermus aquaticus* MutS (Qiu et al., 2012) indicated that the protein is conformationally dynamic when scanning homoduplex DNA, but that its conformation is restricted upon mismatch binding. Using the same technique, Qiu and colleagues demonstrated that in the case of *T. aquaticus*, MutL traps MutS at the mismatch after its ATP-induced activation, but before its conversion into a sliding clamp (Qiu et al., 2015). Therefore, rather than a sliding MutS-MutL clamp model, their findings support a model in which MutL flanks MutS at a mismatch. These findings are consistent with what was first suggested by the Modrich lab (Lyer et al., 2006) and more recently by other investigators (Hombauer et al., 2011; Elez et al., 2012).

1.4.2) Discrimination and removal of the erroneous nascent strand

After mismatch recognition, the strand specificity that ensures a mismatch is corrected on the daughter strand, but not on the parental strand, is the most critical and least understood step of the pathway (Guarné & Charbonnier, 2015). Without strand discrimination, MMR would be a mutagenic process because it could result in the

removal of a mismatch on the parental strand, converting the replication error into a permanent mutation in the genome (Kadyrova & Kadyrov, 2015).

Distinction between the nascent strand and the parental strand occurs by two different mechanisms. The first one is restricted to a subset of γ -proteobacteria including *E. coli* where the strand discrimination is carried out by the MutH protein. This protein recognizes and cleaves the unmethylated strand at the hemi-methylated d(GATC) sites transiently generated during replication. By nicking the DNA, MutH provides an entry point for the subsequent factors to remove the erroneous strand and resynthesis a new one (Figure 1.5) (Lyer et al., 2006).

The second type of MMR mechanism is used by the majority of prokaryotes and all eukaryotes which lack a MutH homolog. This mechanism remained obscure for many years until a decade ago when Kadyrov et al. (2006) showed for the first time that MutL possesses endonuclease activity. This activity confers the ability to nick the daughter strand via interactions with other components of the MMR machinery. As in the methyl-directed process, the nicks provide the entry point for downstream factors to remove the erroneous strand and synthesize a new one (Figure 1.5).

In vitro, the endonuclease activity of MutL targets the nascent strand due to the presence of a pre-existing nick in the DNA (Kadyrov et al., 2006). In vivo, this strand discrimination signal was not so clear. Initially, it was proposed that MMR may have been directed to the newly synthesized DNA by strand discontinuities, such as gaps between Okazaki fragments in the lagging strand during the course of DNA replication (Claverys & Lacks, 1986). This statement was strengthened by observations that the

lagging strand is repaired more efficiently than the leading strand (Pavlov et al., 2003). Recently, it has been shown that transient nicks in the leading strand are dependent on the activity of RNase H2 during ribonucleotide excision repair. It is likely that these strand breaks, which are generated during the removal of the ribonucleotides in the repair reaction, are used by the MMR system as initiation sites for the action of the MutL protein (Lujan et al., 2013; Ghodgaonkar et al., 2013).

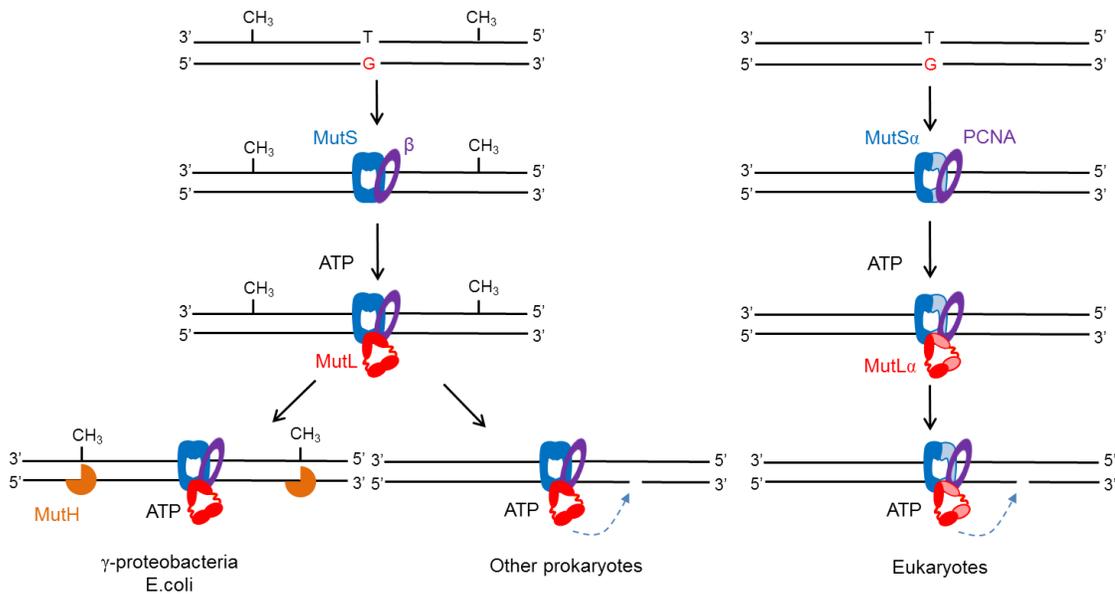


Figure 1.5: Mechanisms of mismatch recognition and strand discrimination in different organisms. Methyl directed MMR process occurs in a subset of γ -proteobacteria such as *E. coli*. Nick-directed mismatch MMR repair occurs in the majority of prokaryotes and all eukaryotes.

To date, several MutL homologs with endonuclease activity have been found. Some of them are from prokaryotic organisms such as *Bacillus subtilis* (Pillon et al., 2010) *Neisseria gonorrhoeae* (Duppatla et al., 2009), *Aquifex aeolicus* (Fukui et al., 2008), *Thermus thermophilus* (Mauris and Evans, 2009), *Pseudomonas aeruginosa* (Correa et al., 2013), *Thermus aquaticus* (Qiu et al., 2015), as well as eukaryotic organisms such as

the yeast *Saccharomyces cerevisiae* (Gueneau et al., 2013) and human (Kadyrov et al., 2006).

After the introduction of the nicks in the nascent strand, an exonuclease (in prokaryotes: ExoI or ExoX (5' to 3' excision), RecJ or ExoVII (3' to 5' excision), and in eukaryotes: Exonuclease 1 (EXO1 (5' to 3' excision))) will remove the strand containing the mismatch, leaving the space for the replication machinery proteins (SSB, β -clamp, DNA ligase, Pol III (or their eukaryotic counterparts)) to perform a second attempt for the resynthesis of the daughter strand (Lyer et al., 2006).

1.4.3) MutL plays a central role in DNA Mismatch Repair

By targeting the erroneous, newly synthesized strand for repair, MutL plays a prominent role throughout the MMR reaction. This is because MutL links the early steps of mismatch recognition and strand discrimination to the later stages of mismatch excision.

In prokaryotes, MutL forms homodimers, whereas in eukaryotes it forms heterodimers (Guarné, 2012 and references therein). The eukaryotic heterodimers are formed from the association of four MutL paralogs: MLH1, PMS1, MLH3, and PMS2 (Kunkel & Erie, 2005). MutL α , the heterodimer formed by association of MLH1 and PMS2 (or MLH1-PMS1 for yeast), has the main mismatch repair function. MutL γ (MLH1-MLH3) is involved in the resolution of recombination intermediates during meiosis (Wang et al., 1999; Zakharyevich et al., 2010) and can partially compensate for

the lack of MutL α in vitro (Cannavo et al., 2005). The function of MutL β (MLH1-PMS1) remains unknown (Räschle et al., 1999).

All MutL homologs are composed of two structurally conserved domains connected by a flexible linker that varies in length and sequence (Figure 1.6) (Guarné et al., 2004). The N-terminal domain (NTD) of MutL has high sequence conservation (~300 residues), and it has been characterized as an ATPase domain belonging to the GHKL (Gyrase, Hsp90, Histidine Kinase and MutL) ATPase kinase superfamily (Ban & Yang, 1998).

The C-terminal domain (CTD) of MutL (~200 residues) has low sequence conservation among all species, but recent structural studies have revealed that it has a conserved structure (Guarné et al., 2004; Pillon et al., 2010; Gueneau et al., 2013; Namadurai et al., 2010). Initially, it was presumed that the CTD was only essential for the dimerization of the protein (Guarné et al., 2004), but work in the past decade has demonstrated its diverse roles, including the endonuclease activity and the modulation of critical protein-protein interactions.

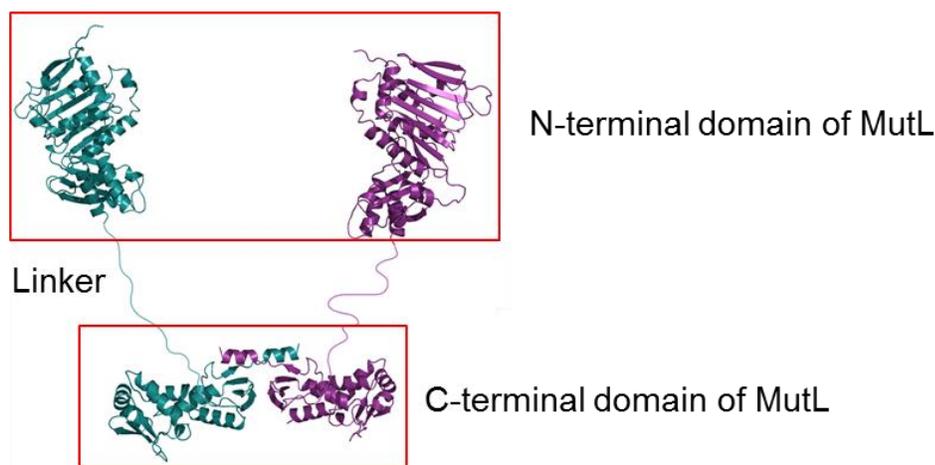


Figure 1.6: Architecture of MutL. Model of full length *E. coli* MutL showing the N- and C-terminal domains (PDB: 1B63 and 1X9Z, respectively) joined by a flexible linker. *Figure extracted from Guarné (2012).

1.4.3.1) The N-terminal domain of MutL is an ATPase

Studies of the crystal structures of the N-terminal regions of *E. coli* MutL, human MLH1, human PMS2 and yeast PMS1 (Ban & Yang, 1998; Guarné et al., 2001; Hall et al., 2002), have described two distinct subdomains of this portion of the protein. The first one contains the four conserved motifs characteristic of the GHKL ATPase superfamily (Bergerat et al., 1997; Ban & Yang, 1998), while the second one presumably mediates DNA binding (Ban et al., 1999).

Although the ATPase activity of MutL is weak (*E. coli* MutL $K_m=90 \mu\text{M}$, $K_{cat}=0.4 \text{ min}^{-1}$) (Ban et al., 1999), it can be stimulated by the presence of DNA in order to coordinate proper MMR mechanism (Ban & Yang, 1998; Guarné et al., 2001; Hall et al., 2002). This activity has been a significant subject of analysis because ATP binding has been suggested to induce different conformational changes on the protein that in conjunction with its hydrolysis can modulate various protein interactions during the repair process (Ban et al., 1999; Guarné et al., 2001; Räschle et al., 2002; Sacho et al., 2008).

Binding of ATP to *E. coli* MutL triggers the self-association of the two NTDs in the dimer. This promotes structural ordering within the domain at around 60 amino acids, and allows interactions between MutL and other proteins, such as MutH, UvrD, MutS, and DNA (Ban & Yang, 1998). MutL dimerization was first shown by X-ray crystallographic in the absence or presence of the adenine nucleotide (Ban & Yang 1998; Ban et al., 1999). Later, this conformational change was explored by size exclusion chromatography where the full-length protein adopted a more compact size in the

presence of a nonhydrolysable ATP analog, AMPPNP (Figure 1.7A) (Ban et al., 1999; Guarné et al., 2004).

The differences in conformations of yeast and human MutL α were confirmed through the use of Atomic Force Microscopy (AFM) (Sacho et al., 2008). In that study, MutL α underwent asymmetric conformational changes in four distinct shapes that change in proportion upon binding of two concentrations of ATP. In the absence of nucleotide cofactors, MutL α existed predominantly in an open and extended conformation, with a large compact dimerization domain connected by flexible arms to two smaller N-terminal domains. In addition to the extended shape, three other minor populations were seen: one-armed, semi-condensed, and condensed conformations. The one-armed shape had a large domain connected to a smaller one, but in this conformation, the central structure was larger than the dimerization domain. The semi-condensed state presented two domains of similar size, and the condensed shape was similar to a compact structure with no protrusions. Upon addition of 0.1 mM ATP, the one-armed conformation became predominant, and when using 5 mM ATP, the condensed conformation was the main shape in the sample (Figure 1.7B) (Sacho et al., 2008).

Not much is known about the mechanistic role of ATP on prokaryotic organisms that use the nick-dependent MMR. Despite several studies on the conformations of the NTD in the absence or presence of the adenine cofactor in some organisms such as *P. aeruginosa* (Miguel et al., 2013) or *A. aeolicus* (Yamamoto et al., 2011; Fukui et al., 2008), there is not a complete picture of the conformational changes the protein may adopt in the presence of ATP. This is intriguing because it is not known whether its

homodimeric nature will make it more similar to *E. coli* MutL or, on the contrary, if the presence of the endonuclease activity will have an effect on different asymmetric conformations as in eukaryotic MutL homologs.

The NTD of *B. subtilis* MutL has only been analyzed using a mutational approach (Bolz et al., 2012). Bolz and colleagues (2012) found that mutations in or near the ATP binding, ATPase activity, and DNA binding motifs drastically compromised repair in vivo. Moreover, impairment of the ATP hydrolysis showed, on average, the highest increase in mutation rate relative to missense mutations in the other two motifs. Even though ATP binding, but not hydrolysis, stimulated the endonuclease activity of *B. subtilis* MutL (Pillon et al., 2010), the form the entire protein takes upon binding to nucleotide has not been established (Figure 1.7C).

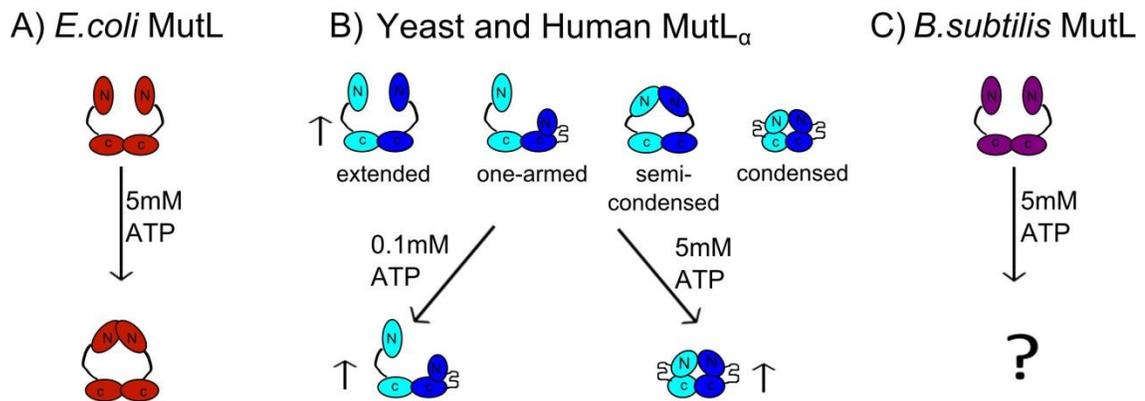


Figure 1.7: MutL conformational changes induced by ATP in different organisms. (A) *E. coli* MutL, (B) Yeast and Human MutL $_{\alpha}$, and (C) *B. subtilis* MutL (arrows pointing up indicate an increase in the population of that form).

Apart from the nucleotide-induced conformational changes, the ATPase domain of MutL also binds to single- and double-strand DNA in a sequence unspecific, cooperative,

and length-dependent manner (Guarné et al., 2001; Hall et al., 2001; Pillon et al., 2010; Iino et al., 2011). In *E. coli* MutL, mutation of arginine 266 abrogates DNA binding; therefore, the groove where this residue is embedded has been proposed to be responsible for binding to the nucleic acid (Ban et al., 1999; Junop et al., 2003). Although this residue is not conserved among MutL homologs, positive potential was reportedly conserved in a homologous region of yeast MutL α (lysine 328 of yeast PMS1 and arginines 273-274 of MLH1) (Hall et al., 2003). The study by Hall et al. (2003) showed that mutations in the MLH1 residues strongly impaired DNA binding while mutations in PMS1 only had a minor effect. This implies that the intact yeast MutL α (MLH1-PMS1) has two independent DNA binding sites that can bring two distant regions of the duplex together (Hall et al., 2003; Kunkel & Erie, 2005).

The NTD of MutL is also responsible for the interaction with MutS (Lenhart et al., 2013). ATP and a mismatch-containing duplex DNA promote this interaction (Winkler et al., 2011). Recently, Groothuizen et al. (2015) trapped the *E. coli* MutS/MutL complex by chemical crosslinking of single cysteine variants with a flexible BM(PEO)₃ cross-linker. Although this research described how MutS positions MutL onto DNA through the formation of two interfaces which drive a novel conformation of MutS, the specific residues involved in the MutL-MutS interaction are still not clear.

1.4.3.2) The C-terminal domain of MutL harbors the endonuclease site

The CTD of MutL is essential for the dimerization of the protein and harbors the endonuclease site (Guarné et al., 2004; Kadyrov et al., 2006). The CTD can be divided

into two subdomains: the regulatory and dimerization subdomains which are connected by a helix (Guarné et al., 2004; Pillon et al., 2010). The N- and C-terminal ends of the domain mediate dimerization. The dimerization interface consists of a four-stranded β -sheet and two helices including the last MutL helix of the protein. The intervening region defines the regulatory subdomain that is exposed to the solvent (Figure 1.8) (Guarné, 2012).

The endonuclease site has been associated with a metal-binding motif that resides at the helix connecting the dimerization and external subdomains (Kosinski et al., 2008). In vitro, there are different ions that allow observing endonuclease activity. For example, *B. subtilis* MutL and eukaryotic MutL α are both strictly Mn²⁺-dependent (Pillon et al., 2010; Gueneau et al., 2013) while *T. thermophilus* and *A. aeolicus* MutL can use Ni²⁺ and Co²⁺ as catalytic metals (Fukui et al., 2008).

Structurally, the endonuclease site is composed of 4 conserved motifs (Figure 1.8): the first three motifs, DQHA(X₂)E(X₄)E, [A/S]C[K/R] and CPHGRP, define a zinc metal binding site (Kosinski et al., 2008) and are conserved for example in *B. subtilis* MutL and yeast MutL α . The sequence of the fourth motif (FXR) is the main difference between bacterial MutL and the eukaryotic MutL homologs. This motif has the consensus sequence FERC in the yeast MutL α , and it allows the coordination of two zinc metal ions in the endonuclease site of PMS1 (Gueneau et al., 2013). On the other hand, *B. subtilis* MutL lacks the last cysteine residue within the motif, and this portion was disordered in the crystal structure (Pillon et al., 2010). As a result, in the structure of *B. subtilis* MutL, the zinc atoms were not fully coordinated. Even though the two zinc-metal ions were

found in both structures, the differences in the crystal structures with or without the metal have led to two different interpretations of the endonuclease site of MutL.

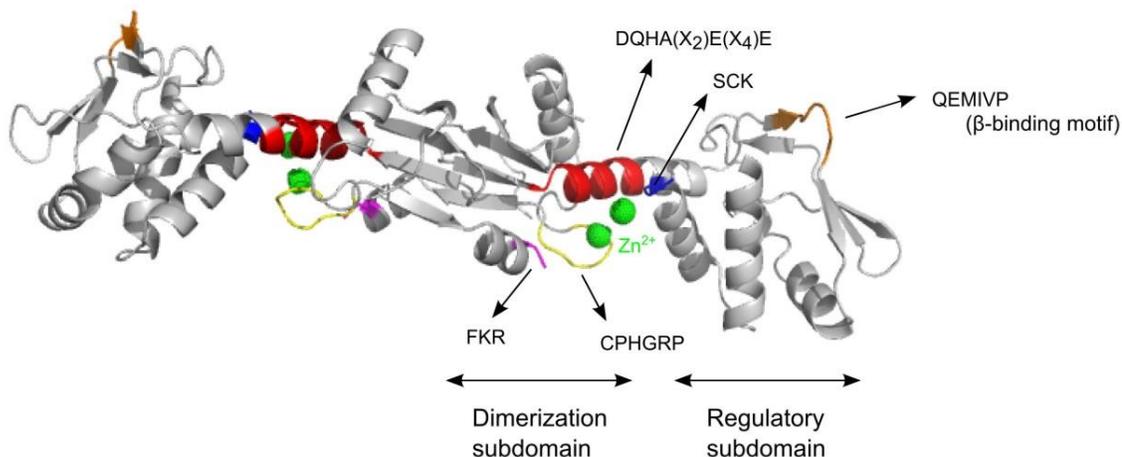


Figure 1.8: Structural organization of the C-terminal domain of *B. subtilis* MutL (3KDK, (Pillon et al., 2010). Arrows show the conserved motifs (DQHA(X₂)E(X₄)E (red), SCK (blue), CPHGRP (yellow), FXR (magenta)) of the endonuclease site and the β -binding motif (QEMIVP (orange)). The endonuclease site clusters to coordinate the zinc ions (green spheres). Double headed arrows represent the C-terminal subdomains.

In the first hypothesis (Pillon et al., 2010), the zinc ion has a structural role with the first glutamate in the motif DQHA(X₂)E(X₄)E involved in its binding (E468 for *B. subtilis* MutL). In addition, the aspartate located at the beginning of the DQHA(X₂)E(X₄)E motif (D462 for *B. subtilis* MutL) plays a catalytic role. In the second hypothesis (Gueneau et al., 2013), the two zinc metal binding sites play a catalytic role rather than a structural one and the aspartate was proposed to have a stabilizing effect on the N-terminus of the helix containing the motif through an N-capping mechanism. Since the catalytic or structural roles of the residues that coordinate the metal binding ions of the endonuclease site of MutL are not clear, there is a need to understand how the endonuclease domain of MutL will behave in the presence of DNA.

1.4.3.3) MutL is a highly regulated endonuclease

The molecular mechanism of the endonuclease activity of MutL remains unclear because MutL is not related to any known nucleases. In fact, the endonuclease domain does not look like any other nuclease, and the only structural similarity is with iron-dependent repressor proteins from the DtxR family which do not possess nuclease activity (Kosinski et al., 2008).

An interesting feature of the CTD of MutL is that it does not bind DNA (Pillon et al., 2010; Gueneau et al., 2013). This DNA binding defect has been proposed as a powerful regulatory mechanism to avoid undesired nicking of the newly synthesized strand in the replication fork, and it has prevented establishing the direct interaction of endonuclease domain of MutL with DNA.

The regulation of the endonuclease activity of MutL is one of the main subjects in the MMR field. One of the factors involved in this coordination is ATP. Since the endonuclease activity of human and yeast MutL α as well as *B. subtilis* MutL is stimulated by ATP binding, it has been suggested that the conformational change imposed by nucleotide binding favors the DNA to be encircled in the protein. This would allow the nucleic acid to reach the endonuclease site once a mismatch has been encountered (Kadyrov et al., 2006; Kadyrov et al., 2007; Sacho et al., 2008; Pillon et al., 2010). In this sense, the NTD and CTD of MutL will communicate to each other in order to progress the repair reaction from the recognition to the strand discrimination step (Guarné & Charbonnier, 2015).

Another important factor that stimulates the endonuclease activity of MutL is the processivity clamp (Pluciennik et al., 2010; Pillon et al., 2015). Pluciennik et al. (2010) showed that the MutL α -PCNA interaction is required for the endonuclease activation of MutL α , but the molecular basis of this interaction has not yet been established.

On the other hand, the interaction of the CTD of MutL and the β -clamp has been characterized in *B. subtilis* (Pillon et al., 2011; Pillon et al., 2015). Our laboratory showed that this interaction is dynamic, yet specific, and that disruption of the motif QX₂(L/I)XP at the external surface of the endonuclease domain (Figure 1.8), best known as the β -binding motif, causes a severe mismatch repair defect in organisms lacking the MutH protein (Pillon et al., 2011). Recently, Pillon et al. (2015) also proposed that the processivity clamp threads the DNA onto the endonuclease site of MutL resulting in the stimulation of the nicking activity in both the full length protein and the CTD, which would otherwise have defective activity due to its inability to bind to DNA. Additionally, it has been suggested that the processivity clamp targets MutL nicking activity towards the nascent strand by virtue of its loading orientation (Pluciennik et al., 2010), but the molecular mechanism of this directionality still remains undefined.

1.5) Thesis objectives

During my thesis project, my first goal was to characterize the ATP-dependent conformational change of *B. subtilis* MutL and establish if this change was similar to the *E. coli* or eukaryotic MutL homologs (Chapter 2).

My second goal was to determine the best conditions to stabilize the transient complex between the endonuclease domain of MutL and DNA. To do this, I biochemically characterized fusion proteins in complex with linear DNA substrates. First, I evaluated the stability of the fusion protein, then I examined the DNA binding capability, endonuclease activity, and factors that may stimulate this activity (Chapter 3).

CHAPTER 2

CONFORMATIONAL CHANGE OF MUTL UPON BINDING OF ATP

E. coli MutL is a homodimeric protein which uses a methyl-dependent mechanism to discriminate the nascent strand from the parental strand in the MMR process. It has been well established that *E. coli* MutL undergoes a conformational change from an open to a more compact form upon binding to ATP (Ban et al., 1999). On the other hand, yeast and human MutL α homologs have a heterodimeric nature and possess endonuclease activity that acts on the nascent strand in the MMR process. Eukaryotic MutL α homologs are present in four conformational states that change in proportion after binding to ATP (Sacho et al., 2008). It is not known whether the asymmetric forms are unique to heterodimers or if the endonuclease activity influences any of them. This chapter is focused on describing the ATP-dependent conformational change of *B. subtilis* MutL, which is a homodimeric protein with endonuclease activity.

My main objective was to identify if the conformational change of homodimeric *B. subtilis* MutL was similar to the well-characterized ATP-dependent conformational change of *E. coli* MutL, or if on the contrary, it had certain similarities with the asymmetric conformations of eukaryotic MutL α . From my findings, I hypothesized that the possession of the nicking activity in MutL homologs is directly correlated to the asymmetric conformations of the protein. In order to confirm this I included a third homodimeric-active endonuclease MutL homolog from *T. aquaticus* in my analysis.

2.1) MATERIALS AND METHODS

2.1.1) Expression and protein purification of MutL

E. coli MutL:

E. coli MutL (pAG 8311) was produced in *E. coli* BL21 (DE3) cells. Cells were grown to OD₆₀₀ of 0.7 at 37°C and protein expression was induced by adding IPTG to a final concentration of 1 mM, followed by incubation with orbital agitation at 37°C for 3 h. Protein was purified as described by Guarné et al. (2004) and stored in buffer containing 20 mM Tris pH 8.5, 5 mM DTT, 1 mM EDTA, 150 mM KCl, and 25% glycerol (v/v) for downstream experiments.

B. subtilis MutL:

B. subtilis MutL (pAG 8220) was produced in *E. coli* BL21 (DE3) Star cells. Cells were grown to OD₆₀₀ of 0.7 at 37°C and protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM, followed by orbital agitation at 25°C for 5 h. Protein was purified as described by Pillon et al. (2010) with nickel affinity chromatography and ion exchange chromatography (MonoQ (5/50) (GE Healthcare)). The salt concentration of the purified protein was diluted to 150 mM KCl for subsequent experiments.

T. aquaticus MutL:

The plasmid from *T. aquaticus* MutL was a kind gift from Dr. Hsieh at National Institutes of Health (NIH). *T. aquaticus* MutL (pAG 9022) was produced in *E. coli* Star (DE3) pRARE cells. Cells were grown to OD₆₀₀ of 0.7 at 37°C and protein expression

was induced by the addition of IPTG to a final concentration of 0.5 mM, followed by incubation with orbital agitation at 37°C for 3 h. Cells were harvested by centrifugation and stored at -80°C. To purify the protein, cell pellets (2 L) were resuspended in 20 mL of lysis buffer (20 mM Tris pH 8.0, 1.4 mM β -mercaptoethanol, 0.5 M NaCl, 0.2 mM PMSF, 30 mM imidazole, and 5% glycerol (v/v)). After the addition of protease inhibitors (PMSF, Leupeptin, Benzamidine, Pepsatin A), cells were lysed by sonication. Protease inhibitors were added again and the lysate was clarified by centrifugation at 39,000 x g. The supernatant was loaded onto a 5 mL HiTrap Chelating HP Nickel column (GE healthcare) equilibrated with lysis buffer. After two washes with 45 mM and 75 mM imidazole, the protein was eluted with buffer containing 20 mM Tris pH 8.0, 1.4 mM β -mercaptoethanol, 0.5 M NaCl; 0.2 mM PMSF, 240 mM imidazole, and 5% glycerol (v/v). Purity was evaluated by 9% SDS-PAGE (75 min at 150 V).

2.1.2) Gel filtration assays

MutL (14.3 μ M) from both *E. coli* and *B. subtilis* was incubated in the absence or presence of 2 mM AMPPNP (5'-adenylyl- β - γ -imidodiphosphate) using nucleotide binding buffer A (20 mM Tris pH 8.4, 150 mM KCl, 0.1 mM EDTA, 5 mM DTT, 5 mM MgCl₂, and 5% glycerol (v/v)) for *E. coli* MutL and nucleotide binding buffer B (20 mM Tris pH 8.0, 150 mM KCl, 0.1 mM EDTA, 5 mM DTT, 5 mM MgCl₂, and 5% (v/v) glycerol) for *B. subtilis* MutL. The reaction in the presence of 2 mM AMPPNP was incubated for 1 hour at room temperature followed by an overnight incubation at 4°C (Guarné et al., 2004; Pillon et al., 2013). Samples (100 μ L) were injected into a

Superdex-200 (GE Healthcare) equilibrated with the respective nucleotide- binding buffer.

2.1.3) Dynamic Light Scattering (DLS)

Dynamic light scattering was carried out using a Zetasizer Nano S (Malvern Instruments). All measurements were taken using a 12 μ L quartz cell (ZEN2112) at 4 °C. Size distribution of the samples was calculated based on the correlation function provided by the Zetasizer Nano S software.

2.1.4) Thermal stability of MutL (Differential scanning fluorimetry)

Differential scanning fluorimetry, commonly known as ThermoFluor, was used to assess thermostability of MutL in a systematic way. For this experiment, a BioRad CFX96 RT-PCR instrument was used with a SYBR-Green filter compatible with the fluorescence excitation/emission maximum spectra of 470 nm and 569 nm of SYPRO-Orange (Invitrogen). According to the protocol of Boivin et al. (2013), the following components were added to a 96-well thin-wall PCR plate (Bio-Rad): (1) 16 μ L deionized water to make a final volume of 25 μ L, (2) 5 μ L of 5 \times buffer, (3) 5 μ L of 5 \times salt (optional), (4) 5 μ L of 5 \times additive (optional), (5) 2 μ L of protein at 66.3 μ M, and 2 μ L of SYPRO-orange 62 \times . Once the microplate had been filled with samples and buffers, it was sealed with optical-clear quality sealing tape (Bio-Rad) and centrifuged at 4°C, 2500 x g for 30 s. The plate was quickly transferred to the RT-PCR equilibrated at 4°C and then heated to 95°C in increments of 1°C per min.

Two other thermofluor assays were performed: one testing specific conditions for *E. coli* MutL, and another one under different concentrations of acetate salts for *B. subtilis* MutL. For these two experiments the FRET channel, instead of the SYBR-Green filter was used. The final protein concentrations in all thermofluor assays were 2.65 μM .

2.1.5) Time course experiments for *B. subtilis* and *E. coli* MutL

MutL (14.3 μM) from *B. subtilis* and *E. coli* were incubated with nucleotide binding buffer (20 mM Tris pH 8.0, 150 mM KCl, 0.1 mM EDTA, 5 mM DTT, 5 mM MgCl_2 , and 5% glycerol (v/v)) in the absence and presence of 5 mM AMPPNP for 2, 4, and 6 h at 4°C. At each time point, 100 μL were loaded into a Superdex-200 (GE Healthcare) equilibrated with the nucleotide binding buffer.

2.1.6) Limited trypsin proteolysis of MutL

Proteolysis reactions were performed in buffer containing 20 mM Tris pH 8.0, 0.5 mM EDTA, 5 mM DTT, 150 mM KCl, 5 mM MgCl_2 , and 5% glycerol (v/v). MutL (2.65 μM) was incubated with increasing concentrations of trypsin (1.56×10^{-3} mg/mL to 0.2 mg/mL) in the absence or presence of ATP for 30 min at room temperature. Different concentrations of ATP (0 mM, 0.1 mM and 1 mM) were used to test the protective effect of the adenine nucleotide. Reactions (10 μL) were stopped by the addition of 2 \times SDS loading buffer and were incubated at 95°C for 10 min. Digestion products were resolved by 11% SDS-PAGE and bands were visualized by staining with Coomassie Brilliant Blue R-250. All experiments were done in triplicate.

2.1.7) Sample preparation of MutL for Atomic Force Microscopy analysis

B. subtilis and *E. coli* MutL (500 μ L) were injected into a Superdex-200 (GE Healthcare) equilibrated in buffer containing 25 mM HEPES pH 7.3, 100 mM potassium acetate, 1.4 mM β -mercaptoethanol and 5% glycerol (v/v). Eluted proteins were divided into 3 μ L aliquots at 2 μ M and frozen in 20% glycerol (v/v).

T. aquaticus MutL (5 mL) was injected into a Desalting column (GE Healthcare) equilibrated in buffer containing 25 mM HEPES pH 7.3, 1.4 mM β -mercaptoethanol, 0.2 mM PMSF, 150 mM KCl, and 5% glycerol (v/v). Eluted protein was concentrated in a 20 mL, 100 kDa MWCO centricon (GE Healthcare). Protein was divided into 3 μ L aliquots at 3 μ M and frozen in 20% glycerol (v/v). The purity of the proteins was evaluated by 9% SDS-PAGE (75 min at 150 V).

2.1.8) Atomic Force Microscopy (AFM)

For AFM imaging, *B. subtilis* MutL was diluted to a final concentration of 30 nM in imaging buffer (20 mM HEPES pH 7.3, 5 mM magnesium acetate, 25 mM potassium acetate) at room temperature. Diluted protein was deposited onto freshly cleaved ruby mica (Spruce Pine Mica Company, Spruce Pine, NC) following the same protocol as in Sacho et al. (2008). Samples in the presence of ATP were mixed and then diluted in imaging buffer to a final concentration of 1 mM ATP. AFM images were collected and analyzed by Hunter Wilkins in the laboratory of Dr. Erie at the University of North Carolina at Chapel Hill, USA.

2.2) RESULTS

2.2.1) MutL conformational change in *E. coli*

E. coli MutL undergoes a conformational change upon nucleotide binding that can be readily monitored with size exclusion chromatography (Ban et al., 1999; Guarné et al., 2004; Pillon et al., 2013). To establish a reference point for the experiments with *B. subtilis* MutL, I repeated the previously published experiments with *E. coli* MutL.

Pure *E. coli* MutL (56.4 μ M) was obtained after a two-step purification consisting of nickel affinity chromatography followed by anion exchange chromatography (Q-Sepharose) (Figure 2.1). Taking advantage of the separation of molecules based on the size and shape of a size exclusion chromatography column (where bigger particles elute earlier in the column), the nucleotide-dependent conformational change of *E. coli* MutL was monitored in the absence or presence of a non-hydrolysable form of ATP, AMPPNP.

In the absence of AMPPNP, MutL eluted at a volume of 11.01 mL. In contrast, after incubating with AMPPNP, MutL eluted later at 11.72 mL (Figure 2.2). This result suggests that *E. coli* MutL changed from an extended conformation to a more compact form in the presence of the adenine nucleotide. The increase of the UV absorption 260 nm over 280 nm ($A_{260}/280$) ratio from 0.53 for MutL alone to 0.77 for the complex (protein and nucleotide) provided evidence of the nucleotide binding. These findings were comparable to the ratios obtained by Ban & Yang (1998), Guarné et al. (2004), and Pillon et al. (2013).

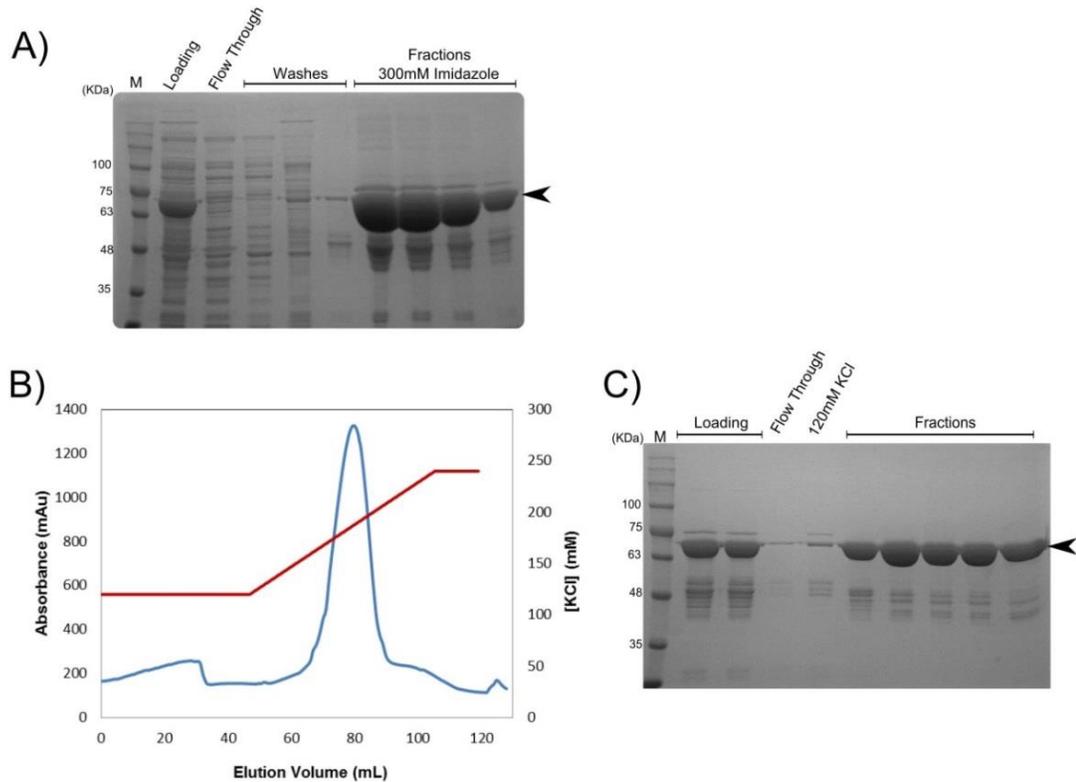


Figure 2.1: Purification of *E. coli* MutL. (A) 9% SDS-PAGE showing the different fractions eluted from a nickel column. The gel shows (from left to right): molecular weight marker (M), lysate loaded onto the column (loading), proteins that flow through from the column (flow through), proteins eluted after two washes (0 mM and 54 mM imidazole (washes)), and fractions of the MutL protein (fractions 300mM Imidazole). (B) Elution profile of MutL from *Q*-sepharose column with a KCl gradient (Blue line represents the absorbance (mAu) and the red line is the concentration of KCl (mM)). (C) 9% SDS-PAGE showing the different fractions eluted from the *Q*-sepharose column. The gel shows (from left to right): molecular weight marker (M), pooled fractions from the nickel column, before and after filtering (loading), proteins that flow through from the column (flow through), protein after a wash with 120 mM KCl (120 mM KCl), and fractions from the peak during linear salt gradient (fractions).

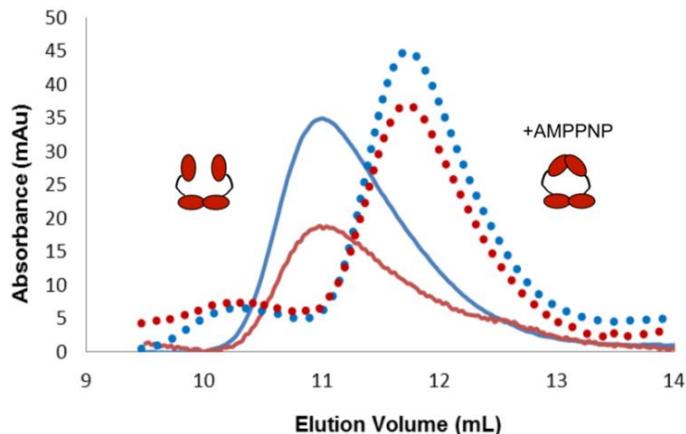


Figure 2.2: Size exclusion chromatography profile of *E. coli* MutL in the absence (continuous line) or presence (dotted line) of AMPPNP. The conformational change was monitored at 280 nm (blue lines) and 260 nm (red lines).

2.2.2) Experimental conditions for *E. coli* MutL are not optimal for *B. subtilis* MutL

Pure *B. subtilis* MutL (68.4 μ M) was obtained after a nickel affinity chromatography and ion exchange chromatography (MonoQ (5/50)) (Figure 2.3). The experimental conditions for the *E. coli* MutL nucleotide-dependent conformational change were replicated for *B. subtilis* MutL. In the reaction without AMPPNP, the majority of the protein eluted in the void volume around 8.07 mL (Peak 1 in Figure 2.4).

The void volume for this column represented the elution volume of molecules that were excluded from the gel filtration medium because they were larger than the largest pores in the matrix and pass straight through the packed bed (Ai, 2006). Considering that the majority of the sample eluted at this volume, this was a sign of aggregation meaning that unlike *E. coli* MutL, *B. subtilis* MutL was not stable when left overnight. From experiments done previously in the Guarné Lab, it was known that this protein had an

elution volume of around 10.7 mL which, in fact, is the smaller peak in Figure 2.4 (Peak 2).

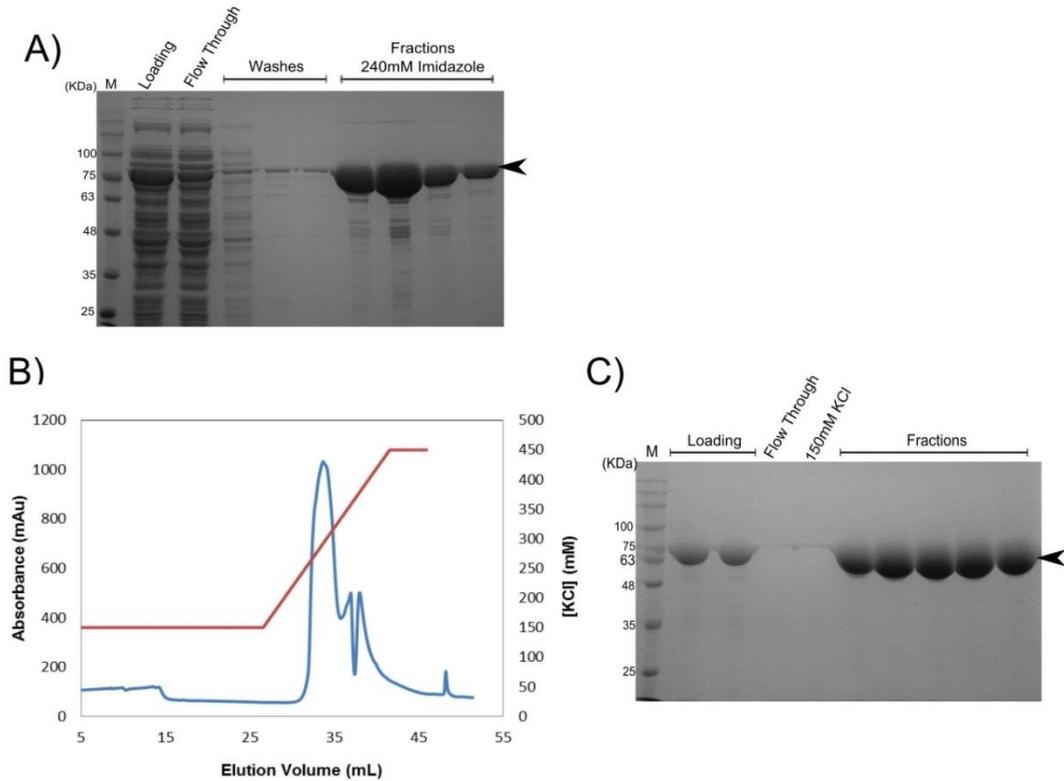


Figure 2.3: Purification of *B. subtilis* MutL. (A) 9% SDS-PAGE showing the fractions eluted from a nickel column. The gel shows (from left to right): molecular weight marker (M), clarified lysate loaded into the column (loading), proteins that flow through from the column (flow through), proteins eluted during two washes (45 mM and 75 mM imidazole (washes)), and fractions of the MutL protein (fractions 240 mM imidazole). (B) Elution profile of MutL from MonoQ (5/50) column with a KCl gradient (Blue line represents the absorbance (mAu) and the red line is the concentration of KCl (mM)). (C) 9% SDS-PAGE showing the different fractions eluted from the MonoQ column (from left to right): molecular weight marker (M), pooled fractions from the nickel column, before and after filter (loading), portion that flows through from the column (flow through), fractions after a wash with 150 mM KCl (150 mM KCl), and fractions from the main peak during linear salt gradient (fractions).

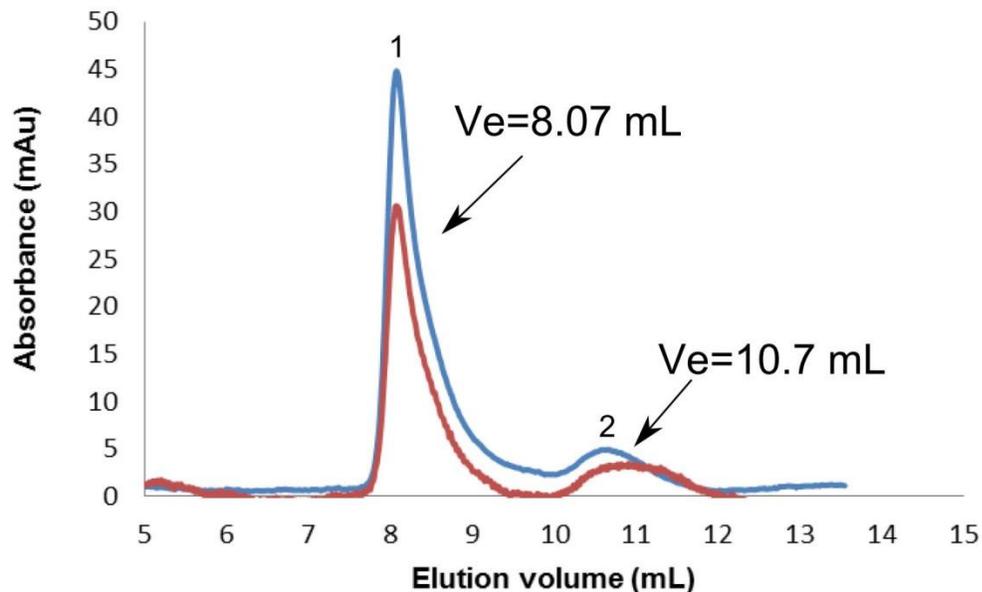


Figure 2.4: Size exclusion chromatography profile of *B. subtilis* MutL. The profile was monitored at 280 nm (blue line) and 260 nm (red line).

2.2.3) Optimization of gel filtration assay for *B. subtilis* MutL

Different parameters (time, pH, KCl concentration, MgCl₂ concentration, and concentrated protein) were tested to assess the quality of the protein using DLS (Figure 2.5). DLS is a tool that determines the distribution of differently sized particles present within a sample. This technique allowed us to discriminate between the presence of properly folded protein (small hydrodynamic radius) and larger oligomers or protein aggregates (large hydrodynamic radius) under different conditions (Malvern, 2008).

The incubation of the sample for one hour at room temperature (RT) compared to an overnight at 4°C did not present major differences in the size distribution by volume (Figure 2.5A). As shorter times usually come with less aggregation, it was expected that the sample at one hour incubation had better behavior than the sample incubated

overnight. In a buffer containing 285 mM KCl, the sample presented two populations of different particle size, while 150 mM KCl showed a homogeneous population (Figure 2.5B). Qualitatively, the concentrated protein had a bigger particle diameter than the sample with the 150 mM KCl. The change in pH was not as significant as with the salt concentration but it seemed at pH 8.0 protein had a smaller particle size compared to pH 8.4 (Figure 2.5C). Finally, in the highest concentration of $MgCl_2$ (5 mM), the sample was more homogeneous than when using 2 mM $MgCl_2$.

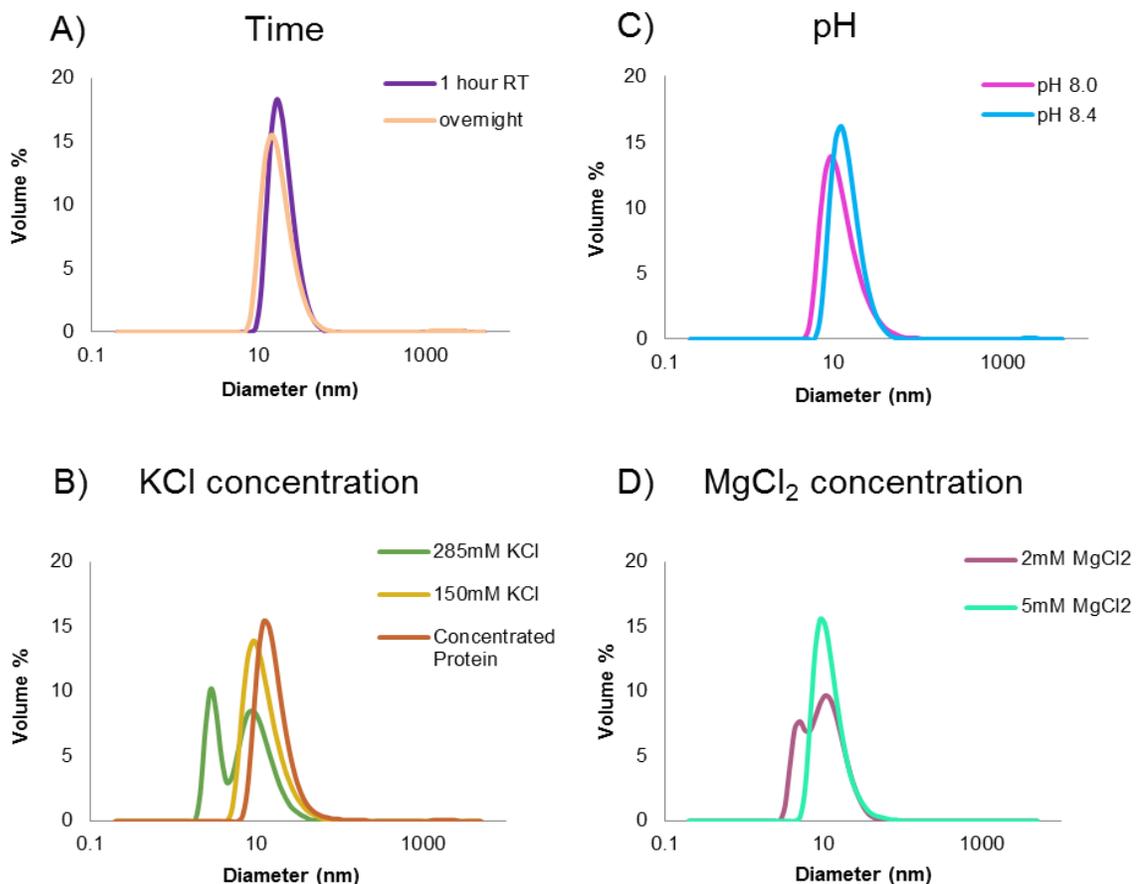


Figure 2.5: Stability of *B. subtilis* MutL under different conditions using DLS. Volume distribution curves of *B. subtilis* MutL testing: (A) time, (B) pH, (C) KCl concentration, and (D) $MgCl_2$ concentration.

Based on these results a new gel filtration assay was performed for one hour incubation at room temperature in a buffer that mainly changed the pH of the reaction (20 mM Tris pH 8.0, 1 mM EDTA, 1.4 mM β -mercaptoethanol, 150 mM KCl, 5 mM $MgCl_2$, and 5% glycerol (v/v)) (Figure 2.6). Even though there was still a portion of MutL in the void volume (peak 1), it was possible to recover more protein in the peak at 10.7 mL (peak 2).

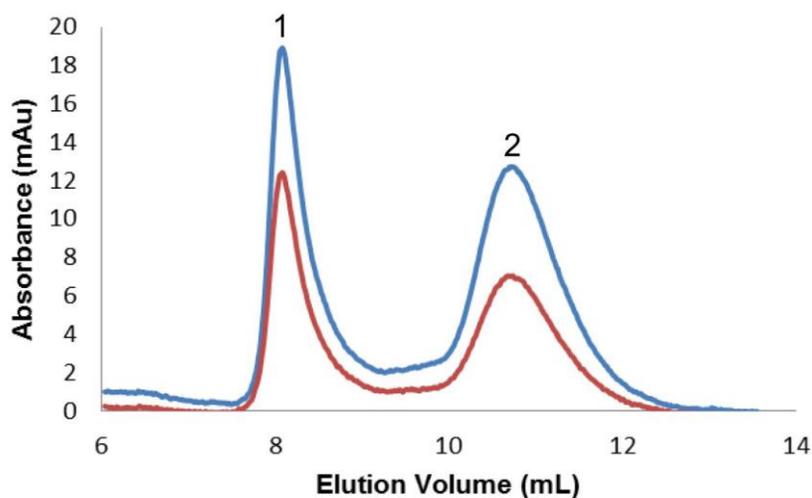


Figure 2.6: Size exclusion chromatography of *B. subtilis* MutL. The profile was monitored at 280 nm (blue line) and 260 nm (red line).

2.2.4) Influence of different parameters on the thermal stability of *B. subtilis* MutL

Despite the improvements in the stability of *B. subtilis* MutL, some portion was still aggregated and it was necessary to analyze its behavior in a more systematic way. To improve the *B. subtilis* MutL sample, differential scanning fluorimetry was performed to assess the thermostability of the protein under different buffer conditions. In this assay, the protein was subjected to a temperature gradient (from 4°C to 95°C) and its thermal denaturation was monitored using a fluorescent dye. The increase in absorbance reflects

how the dye, in this case SYPRO-Orange, binds to hydrophobic pockets that become exposed during unfolding of the protein. We could therefore generate a melting curve for the protein and determine its melting point, which is the temperature at which 50% of the protein was unfolded. The melting temperature could be obtained by integrating the denaturation curve (Boivin et al., 2013).

Using the thermofluor experiment described by Boivin and co-workers, multiple buffer conditions were tested simultaneously to assess reagents that stabilize the protein. A shift towards higher temperatures in the melting temperature curve indicates that the protein is more resistant to heat denaturation and, hence, more stable. On the contrary a shift towards lower temperatures is observed for conditions that destabilize the molecule.

For *B. subtilis* MutL, I tested pH, buffer type, salt and buffer concentration, as well as specific additives such as ATP, EDTA, glycerol and reducing agents (TCEP, DTT, β -Me) using a generic previously published screen (Table 2.1, Boivin et al., 2013). This approach was crucial to guide the optimization of the purification protocol and the conditions for the gel filtration assay. Four of the conditions that had major effects on the thermal stability of *B. subtilis* MutL are described below.

2.2.4.1) Effect of pH

The pH screen indicated that MutL was most stable at pH 8.0 (Figure 2.7). This result was consistent with the DLS measurement (Figure 2.5C) and confirmed this pH as the best condition to purify the protein. Apart from this type of buffer, pH 7 or 7.3 also showed a stabilizing effect but not as significant as with pH 8.0.

Table 2.1: Conditions for thermofluor of *B. subtilis* MutL

	1	2	3	4	5	6	7	8	9	10	11	12
A	water	citric acid pH 4.0	Na Acetate pH 4.5	citric acid pH 5.0	MES pH 6.0	K ₂ PO ₄ pH 6.0	citric acid pH 6.0	Bis Tris pH 6.5	NaCaco dylate pH 6.5	NaH ₂ PO ₄ pH 7.0	KH ₂ PO ₄ pH 7.0	HEPES pH 7.0
B	MOPS pH 7.0	Am Acetate pH 7.3	Tris HCl pH 7.5	NaH ₂ PO ₄ pH 7.5	HEPES pH 8.0	Tris HCl pH 8.0	Tris HCl pH 8.5	CHES pH 9.0	10 mM Tris HCl pH 8.0	50 mM Tris HCl pH 8.0	100 mM Tris HCl pH 8.0	250 mM Tris HCl pH 8.0
C	150 mM KCl	citric acid pH 4.0/ KCl	Na Acetate pH 4.5/ KCl	citric acid pH 5.0/ KCl	MES pH 6.0/ KCl	K ₂ PO ₄ pH 6.0/ KCl	citric acid pH 6.0/ KCl	Bis Tris pH 6.5/ KCl	NaCaco dylate pH 6.5/ KCl	NaH ₂ PO ₄ pH 7.0/ KCl	KH ₂ PO ₄ pH 7.0/ KCl	HEPES pH 7.0/ KCl
D	MOPS pH 7.0/ KCl	Am Acetate pH 7.3/ KCl	Tris HCl pH 7.5/ KCl	NaH ₂ PO ₄ pH 7.5/ KCl	HEPES pH 8.0/ KCl	Tris HCl pH 8.0/ KCl	Tris HCl pH 8.5/ KCl	CHES pH 9.0/ KCl	Buffer “A”/ 2 mM ATP	Buffer “A”/ 5 mM ATP	Buffer “A”/ 2 mM ATP/ 5 mM MgCl ₂	Buffer “A”/ 2 mM ATP/ 5 mM MgCl ₂
E	MES pH 6.0/ 100 mM KCl	MES pH 6.0/ 200 mM KCl	MES pH 6.0/ 350 mM KCl	MES pH 6.0/ 500 mM KCl	MES pH 6.0/ 700 mM KCl	MES pH 6.0/ 1 M KCl	Tris HCl pH 8.0/10 mM KCl	Tris HCl pH 8.0/20 mM KCl	Tris HCl pH 8.0/350 mM KCl	Tris HCl pH 8.0/500 mM KCl	Tris HCl pH 8.0/700 mM KCl	Tris HCl pH 8.0/1 M KCl
F	Buffer “A”/ 50 mM Imidazole	Buffer “A”/ 100 mM Imidazole	Buffer “A”/ 250 mM Imidazole	Buffer “A”/ 500 mM Imidazole	Buffer “A”/ 50 mM Arginine	Buffer “A”/ 50 mM Glu-Arg	Buffer “A”/ 200 mM Glu-Arg	Buffer “A”/ 50 mM Glutama te	Buffer “A”/ 1% glycerol	Buffer “A”/ 5% glycerol	Buffer “A”/ 10% glycerol	Buffer “A”/ 20% glycerol
G	Buffer “A”/ 1 mM TCEP	Buffer “A”/ 5 mM TCEP	Buffer “A”/ 10 mM TCEP	Buffer “A”/ 1 mM β-Me	Buffer “A”/ 5 mM β-Me	Buffer “A”/ 10 mM β-Me	Buffer “A”/ 1 mM DTT	Buffer “A”/ 5 mM DTT	Buffer “A”/ 10 mM DTT	Buffer “A”/ 0.1 mM EDTA	Buffer “A”/ 1 mM EDTA	Buffer “A”/ 5 mM EDTA

H	Buffer "A"/5 mM MgCl ₂	Buffer "A"/10 mM MgCl ₂	Buffer "A"/10 mM CaCl ₂	Buffer "A"/10 mM MnCl ₂	Buffer "A"/ 10 mM NiCl ₂	Buffer "A"/ 10 mM ZnCl ₂	Buffer "A"/ 10 mM CoCl ₂	Buffer "A"/ 100 mM LiCl	Buffer "A"/ 100 mM NaCl	Buffer "A"/ 100 mM KCl	Buffer "A"/ 100 mM NH ₄ Cl	Buffer "A"
----------	---	---	---	---	--	--	--	----------------------------------	----------------------------------	------------------------------	--	---------------

Buffers were used at concentration of 50 mM unless indicated. KCl was used at concentration of 150 mM, unless indicated. Buffer "A" composition: 50 mM Tris HCl pH 8.0, 150 mM KCl.

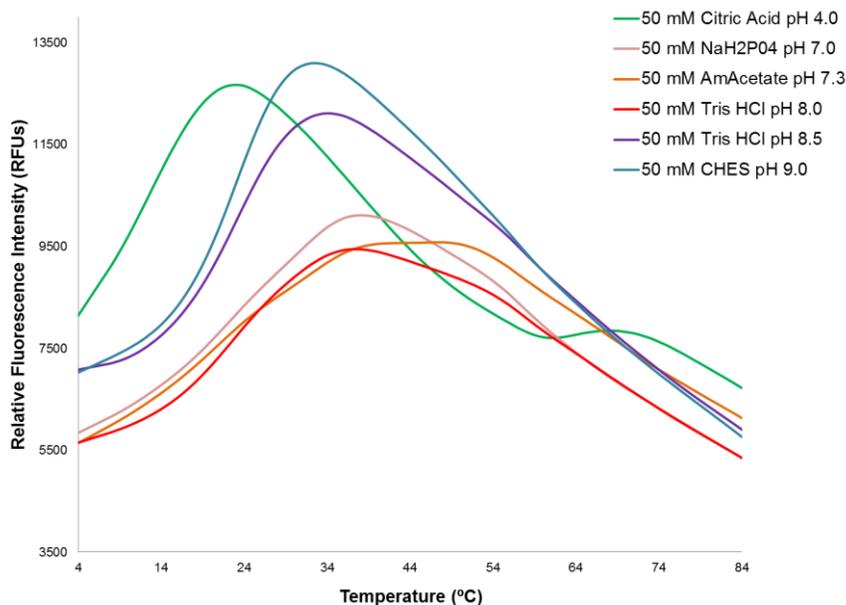


Figure 2.7: Effect of pH on the thermostability of *B. subtilis* MutL.

2.2.4.2) Effect of KCl concentration

The ionic strength was an important factor for the stabilization of MutL. KCl concentrations above 700 mM and below 350 mM decreased the thermal stability of the protein (Figure 2.8). Even though the DLS measurement (Figure 2.5B) showed one population for 150 mM KCl, it seems this was not the most stable form of MutL. It is likely that the first peak for the concentration at 285 mM (peak ~5 nm) may be more stable and that is why in the thermofluor assay higher salt concentrations (500 mM, 700 mM) led to a higher melting temperature, increasing it by approximately 6°C. Although high salt concentrations stabilized the protein, under these conditions the denaturation profile had two peaks. Since MutL has two distinct domains (Ban & Yang, 1998; Guarné et al., 2004), the two peaks may be the result of different melting temperatures among them (one domain unfolding at a lower temperature than the other one).

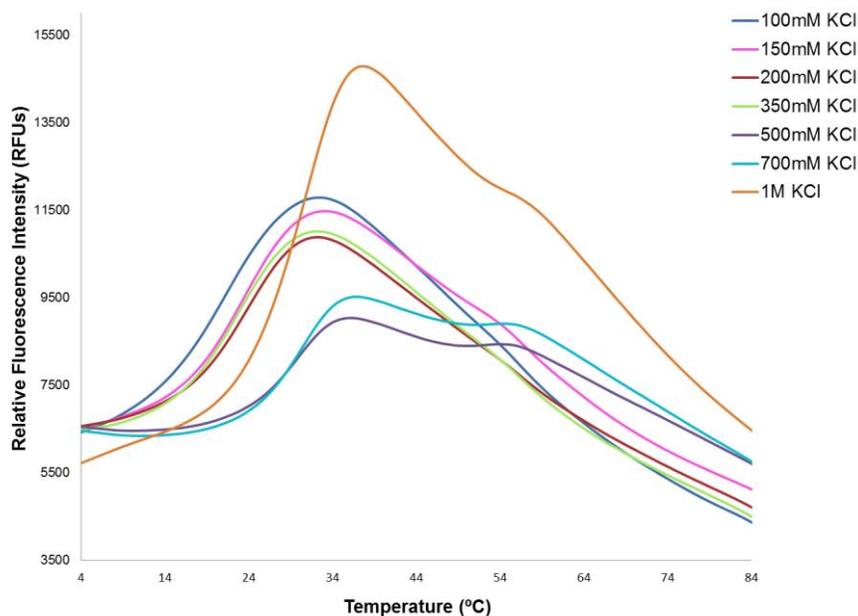


Figure 2.8: *Effect of KCl concentration on the thermostability of B. subtilis MutL. All samples were in 50 mM Tris HCl pH 8.0*

2.2.4.3) Effect of imidazole concentration

We also tested the effect of the imidazole concentration on MutL since this compound was used to elute the protein from the metal affinity chromatography column and it is known to destabilize proteins (Figure 2.9) (Boivin et al., 2013). The concentration of imidazole used for the purification was 240 mM. This is approximately the concentration of imidazole (250 mM) where we see a significant destabilizing effect (a decrease in melting temperature of about 5°C) compared to the protein in 50 mM of the same compound.

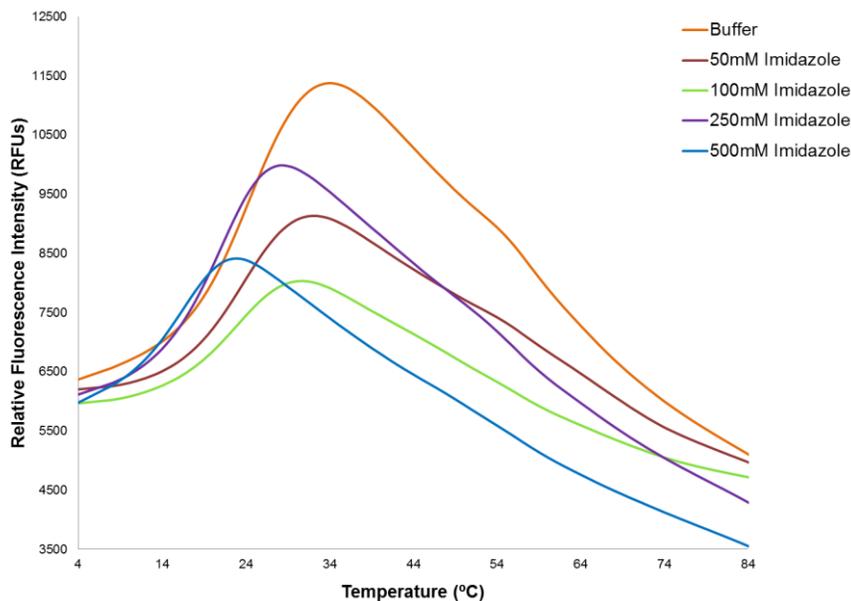


Figure 2.9: Effect of imidazole concentration on the thermostability of *B. subtilis* MutL. All samples were in 50 mM Tris HCl pH 8.0 and 150 mM KCl (buffer).

2.2.4.4) Influence of ATP and MgCl₂

The most remarkable stabilizing effect on MutL was observed in the presence of ATP and MgCl₂. When *B. subtilis* MutL was in buffer (50 mM Tris HCl pH 8.0, 150 mM KCl), and either 5 mM ATP, or 5 mM MgCl₂, there was no effect on the melting temperature. However, once the nucleotide and MgCl₂ were added simultaneously (5 mM ATP/5 mM MgCl₂), there was a significant increase in the melting temperature profile of approximately 11°C (Figure 2.10A). A thermofluor experiment using *E. coli* MutL instead of *B. subtilis* MutL confirmed the stabilization induced by nucleotide binding (Figure 2.10B).

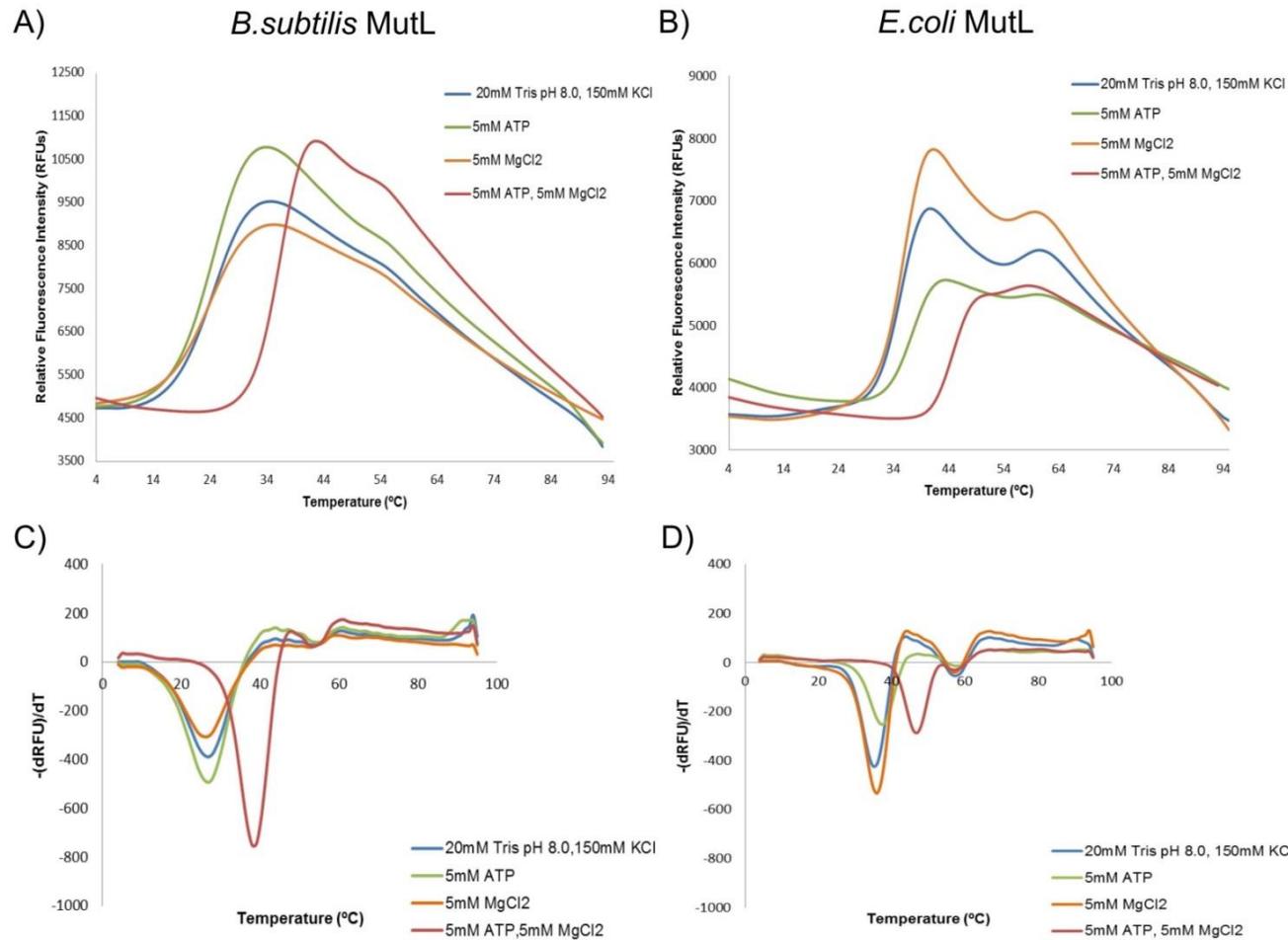


Figure 2.10: Effect of ATP and MgCl₂ on the thermal stability of MutL. Melting temperature profile for (A) *B. subtilis* MutL and (B) *E. coli* MutL in buffer containing 20 mM Tris HCl pH 8.0, 150 mM KCl and the respective additives. Alternative representation of the melting temperature profile using the first negative derivatives for (C) *B. subtilis* MutL and (D) *E. coli* MutL.

The first negative derivatives for these two proteins allowed determining a more accurate melting temperature (Figure 2.10C-D). *B. subtilis* MutL had a melting temperature of 27°C when buffer, ATP or MgCl₂ were present individually. However, the melting temperature increased to 38°C when 5 mM ATP and 5 mM MgCl₂ were both present. The same difference in temperature was observed for the *E. coli* MutL from 36°C to 47°C. Considering that changes above 5°C are significant (Boivin et al., 2013), these positive shifts reflect a more stable form of the MutL protein from both organisms. In addition, since the ATP-bound form of *E. coli* MutL is more stable than the apo conformation (Ban et al., 1999), we inferred the higher stability in both proteins comes from the binding of ATP.

2.2.5) *B. subtilis* MutL does not undergo a conformational change similar to *E. coli* MutL upon binding of ATP

After the optimization of the conditions for the reaction with ATP and realizing *B. subtilis* MutL started to unfold at room temperature, I set reactions with 5 mM AMPPNP for 2, 4 and 6 h at 4°C. This time course experiment monitored the effect of time and the aggregation state of the protein. The reaction in the absence and presence of 5 mM AMPPNP for all time points showed the same elution profile (Figure 2.11A).

In order to determine the shortest amount of time in which MutL undergoes the conformational change, the time course experiment was then repeated for *E. coli* MutL (Figure 2.11B). Even after two h, *E. coli* MutL experienced the conformational change to a more compact form in the presence of the adenine nucleotide. Based on the literature,

this conformational change is attributed to the self-association of the NTD of the protein upon binding with ATP (Ban et al., 1999).

Since *B. subtilis* MutL did not show the same behavior, it is likely that this protein is either not undergoing the same conformational change as *E. coli* MutL or that the association of its N-terminal domains is more labile. The latter would imply that the size exclusion chromatography could not resolve the conformations because the monomer and dimer forms may be in equilibrium.

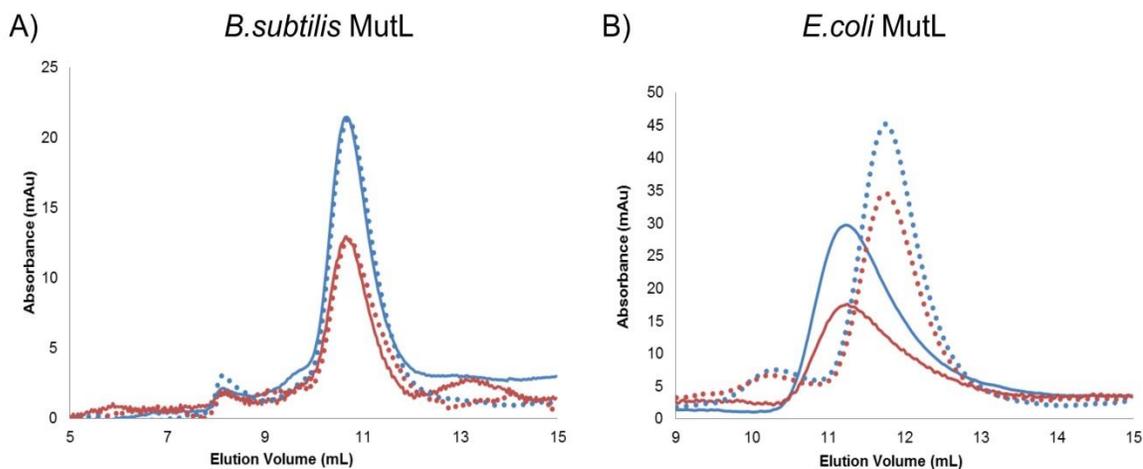


Figure 2.11: Size exclusion chromatography profile of (A) *B. subtilis* and (B) *E. coli* MutL for 2 h at 4°C in the absence (continuous line) or presence (dotted line) of 5mM AMPPNP. The profile was monitored at 280 nm (blue line) and 260 nm (red line) for both cases.

In either of these scenarios, the behavior of *B. subtilis* MutL is different than the one observed in *E. coli* MutL. We believe it is because these proteins use a different mechanism to discriminate the nascent strand during the MMR process. To strength the evidence that *B. subtilis* MutL undergoes a different conformational change than its *E.*

coli MutL homolog, we decided to explore the behavior of another prokaryotic MutL homolog that possesses endonuclease activity: *T. aquaticus* MutL.

2.2.6) ATP binding protects MutL from protease degradation

We performed limited proteolysis with trypsin to determine if ATP binding provides some level of protection for the three MutL homologs from proteolysis. In experiments done by Sacho et al. (2008) there was a direct correlation between a condensed state of yeast MutL α and the highest protection from protease degradation.

The limited proteolysis experiments were conducted for the three MutL homologs at a final concentration of 2.65 μ M (as the thermofluor assay). The MutL homologs were incubated with varying concentrations of trypsin in the absence (Figure 2.12 A, C, E) or presence (Figure 2.12 B, D, F) of 1 mM ATP. *E. coli* MutL demonstrated a greater level of protection from proteolysis than the other two MutL homologs in the presence of ATP, as full length *E. coli* MutL (~70 kDa) persisted to much higher concentrations of trypsin in the presence of the nucleotide (Figure 2.12B) than in its absence (Figure 2.12A). In addition, there were fewer digestion products of *E. coli* MutL in the presence of ATP (disappearance of the ~30 kDa band).

On the contrary, the protective effect was not as significant for *B. subtilis* MutL since, even though the degradation product at ~75 kDa was reduced, the overall persistence of full length MutL was unaffected by the presence of ATP (Figure 2.12 C, D). In the case of *T. aquaticus* MutL there was no observable difference between the digestion profiles of protein incubated with or without ATP (Figure 2.12 F and E, respectively).

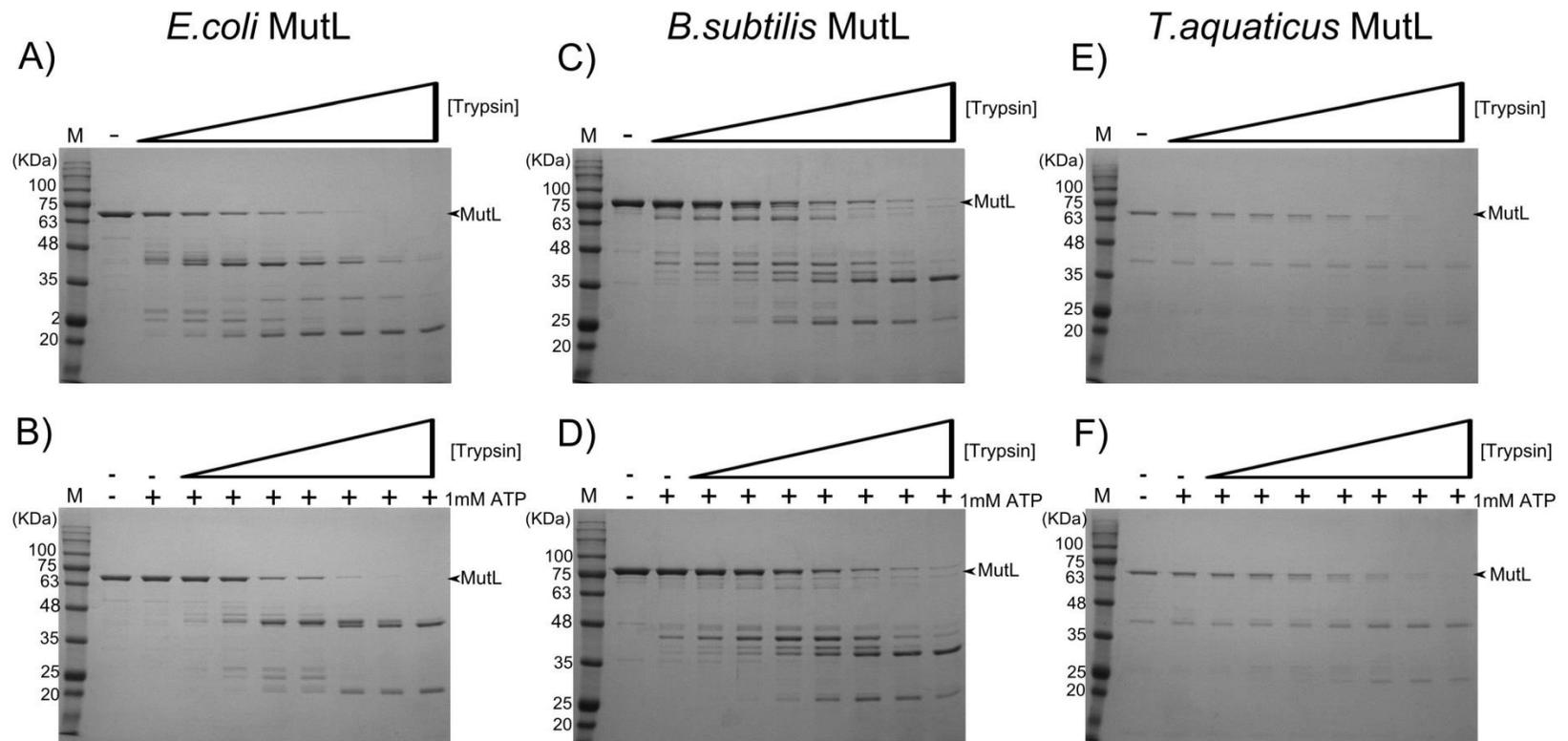


Figure 2.12: Trypsin proteolysis of *E. coli*, *B. subtilis* and *T. aquaticus* MutL in the absence (A, C, E) or presence (B, D, F) of 1 mM ATP.

While higher concentrations of nucleotide (5 mM) did not produce any changes on the protection profile, lower concentrations (0.1 mM) did have an effect (Figure 2.13). Although for *E. coli* MutL the disappearance of the band at ~30 kDa happened for concentrations at 0.1 mM and 1 mM of ATP, the bands close to 40 kDa changed their pattern and intensity with the higher concentration of nucleotide (Figure 2.13A). For *B. subtilis* MutL at 1 mM ATP the intensity of the fragments at ~65 kDa and ~25 kDa were reduced compared to 0.1 mM ATP and even more than in the absence of nucleotide. This means that at this high amount of ATP, higher concentrations of trypsin are needed to produce the same digestion products (Figure 2.13B).

These results support my previous findings that *B. subtilis* MutL behaves differently in the presence of ATP than its homodimeric *E. coli* homolog. The similarities obtained with *B. subtilis* and *T. aquaticus* MutL may suggest that the presence of the endonuclease site leads to a different behavior of the MutL protein where they do not share the characteristic *E. coli* MutL ATP-induced conformational change.

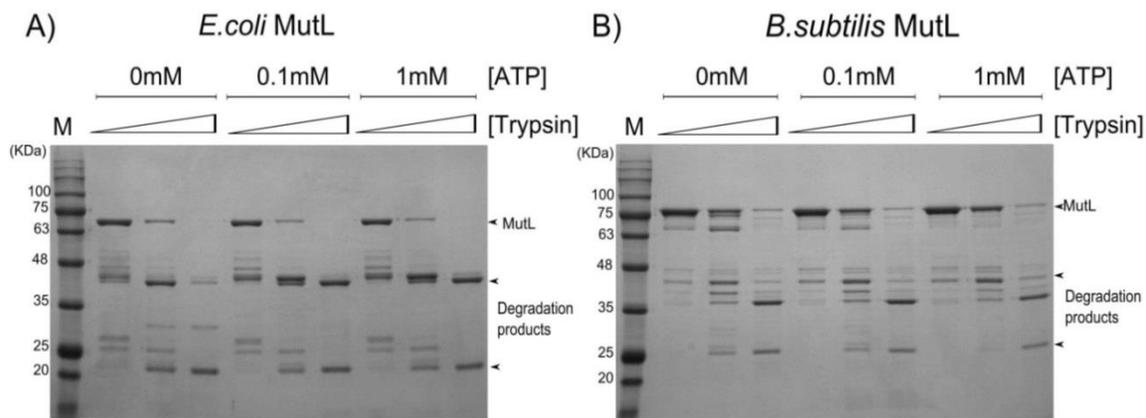


Figure 2.13: Trypsin proteolysis of *E. coli* MutL (A) and *B. subtilis* MutL (B) under different concentrations of ATP.

2.2.7) Optimization of the purification conditions of MutL for AFM analysis

Although we have obtained some information related to the behavior of MutL in the presence of ATP in organisms lacking the MutH protein, the conformation (or conformations) the protein adopts remained unclear. Therefore, it was necessary to use a more direct technique, in this case AFM, to visualize the distinct conformations of MutL in the absence or presence of ATP. These experiments were done by Hunter Wilkins in the laboratory of Dr. Dorothy Erie at the University of North Carolina at Chapel Hill, USA.

Preliminary images were taken for *B. subtilis* MutL. Unfortunately, these images were not of good quality because the protein was aggregated. Hence, it was imperative to improve the preparation of the protein to minimize the aggregated assemblies.

Initially, the buffer composition of the *B. subtilis* MutL purification (20 mM Tris pH 8.0, 278 mM KCl, 5 mM EDTA, 2.8 mM β -mercaptoethanol, 25% glycerol (v/v)) was compared to the imaging buffer (25 mM HEPES pH 7.3, 50 or 150 mM sodium acetate, 5 mM magnesium acetate) since it was possible that one of the components of the purification buffer had a destabilizing effect on the protein.

Some in vitro studies (Leirimo et al., 1987; Richeys et al., 1987) have analyzed the effects of different chloride salt concentrations on protein-nucleic acid interactions. Proteins that interact with nucleic acids, such as MutL, are extremely sensitive to the concentrations and types of electrolyte ions in solution. In fact, KCl concentrations should do not exceed 150 mM, which is below the physiological range of K^+ concentrations and above the physiological range of Cl^- within the cells (Richeys et al.,

1987). This means that using higher concentrations of KCl, as in the case of *B. subtilis* MutL (278 mM KCl), may disrupt protein-nucleic acid interactions.

Based on this analysis, the thermal stability of *B. subtilis* MutL was tested in different concentrations of acetate salts (sodium or potassium acetate, Table 2.2). Acetate salts are in higher abundance on the cells meaning they may have a less negative impact than chloride salts over proteins interacting with DNA (Leirmo et al., 1987). In addition to the acetate salts, buffers (20 mM Tris pH 8.0 or 25 mM HEPES pH 7.3), magnesium chloride or magnesium acetate, and ATP were tested in order to purify the protein in a compatible buffer allowing the maximum stability during AFM analysis.

As previously mentioned the presence of ATP and Mg^{2+} had a positive effect on the thermal stability of MutL. With any type of salt, acetates and chlorides, an increase in the melting temperature was observed. Both pHs, either 20 mM Tris pH 8.0 (red line in Figure 2.14A) or 25 mM HEPES pH 7.3 (blue line in Figure 2.14A) had a similar effect on the stability of the protein. The main difference in this thermofluor screen became apparent when sodium acetate or potassium acetate were used as additives.

The conditions with potassium acetate (continuous lines, red or blue in Figure 2.14A) were always more stable than sodium acetate (dotted line, red or blue in Figure 2.14A), and even more stable than KCl (green line in Figure 2.14A). This supported the idea that chloride salts may have a more destabilizing effect on proteins. Additionally, magnesium acetate could be used in the imaging buffer, since it has the same effect as $MgCl_2$.

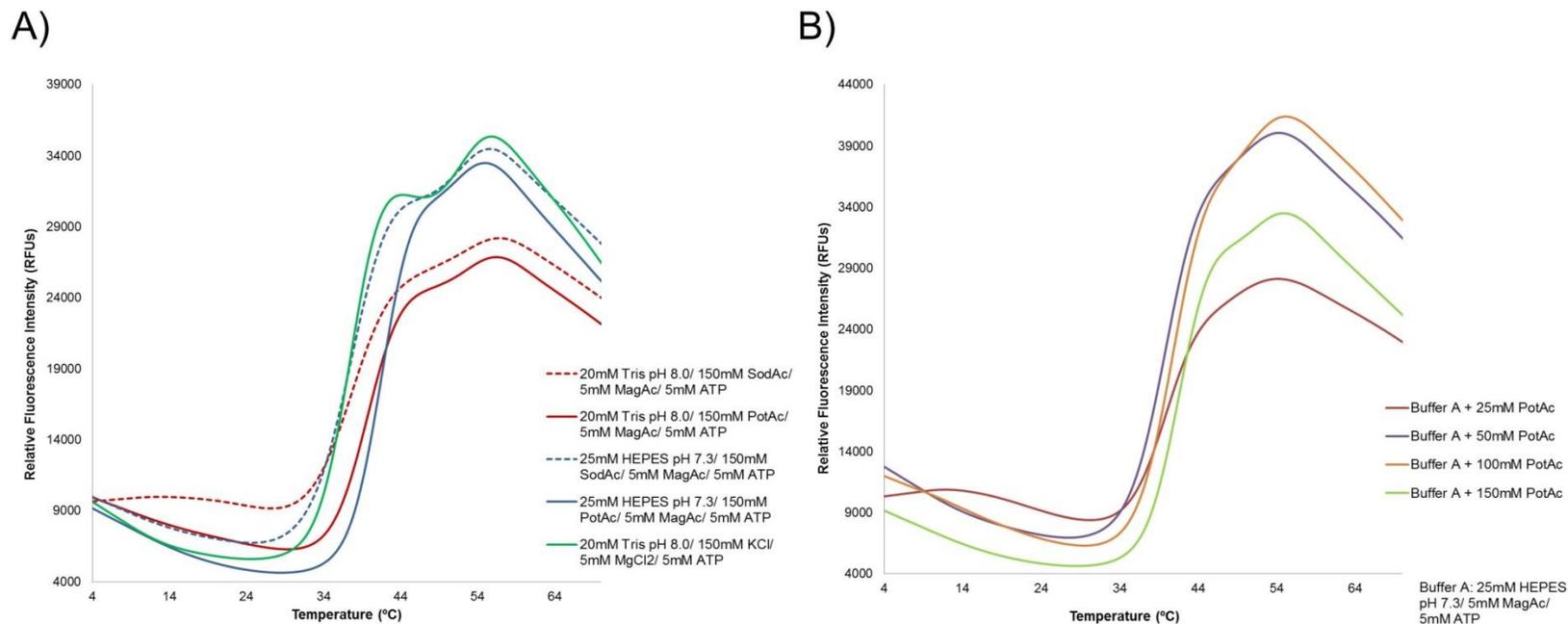


Figure 2.14: Effect of acetate salts on the stability of *B. subtilis* MutL in the presence of ATP. (A) Thermostability curves of *B. subtilis* MutL in 20 mM Tris pH 8.0 (red line) vs. 25 mM HEPES pH 7.3 (blue line) with 150 mM sodium acetate (red or blue dotted line) or potassium acetate (red or blue continuous line). Green line represents the thermostability of MutL with 20 mM Tris pH 8.0 and 150 mM KCl. (B) Thermostability curves of *B. subtilis* MutL in 25 mM HEPES pH 7.3 and different potassium acetate concentrations (25 mM -150 mM).

Lastly, low potassium acetate concentrations (below 100 mM) destabilized the protein, but the negative effect was less dramatic when combined with HEPES pH 7.3 instead of Tris pH 8.0 (Figure 2.14B).

Based on the thermoflour results, new purification conditions were used. *B. subtilis* MutL was obtained with a higher stability in solution after the standard purification protocol (Tris pH 8.0 and KCl) and the addition of a third step consisting of size exclusion chromatography (Superdex-200) to eliminate any aggregates and exchange the buffer to 25 mM HEPES pH 7.3, 100 mM potassium acetate, 1.4 mM β -mercaptoethanol, and 5% glycerol (v/v). The same treatment and similar stability was obtained for *E. coli* MutL.

Interestingly, when following the same strategy for *T. aquaticus* MutL, this protein was not stable in solution and, therefore, it was not possible to obtain any preliminary samples for AFM. DLS measurements were performed under different buffers changing one component at the time (Table 2.3 and Figure 2.15). Although *T. aquaticus* MutL was stable in buffers containing Tris pH 8.0 and KCl (Figure 2.15A), it aggregated in buffers that had potassium acetate. The best condition was buffer E in which HEPES pH 7.3 and KCl had the greatest stabilizing effect for the protein (Figure 2.15B). A desalting column followed by a concentration step were implemented to exchange the buffer and obtain enough *T. aquaticus* MutL for the AFM analysis.

Table 2.2: Conditions for thermofluor of *B. subtilis* MutL using acetate salts

	1	2	3	4	5	6	7	8	9	10	11	12
A	water	Buffer “A”	Buffer “A”/ KCl	Buffer “A”/ KCl/ 2 mM ATP	Buffer “A”/ KCl/ 5 mM ATP	Buffer “A”/ KCl/ 2 mM ATP/ 5 mM MgCl ₂	Buffer “A”/KCl/ 5 mM ATP/ 5 mM MgCl ₂	Buffer “A”/ 5 mM MgCl ₂	Buffer “A”/ KCl/ 5 mM MgCl ₂	Buffer “A”/ 25 mM SodAc	Buffer “A”/ 50 mM SodAc	Buffer “A”/ 100 mM SodAc
B	Buffer “A”/ 150 mM SodAc	Buffer “B”	Buffer “B”/ 25 mM SodAc	Buffer “B”/ 50 mM SodAc	Buffer “B”/ 100 mM SodAc	Buffer “B”/ 150 mM SodAc	Buffer “A”/ 25 mM PotAc	Buffer “A”/ 50 mM PotAc	Buffer “A”/ 100 mM PotAc	Buffer “A”/ 150 mM PotAc	Buffer “B”/ 25 mM PotAc	Buffer “B”/ 50 mM PotAc
C	Buffer “B”/ 100 mM PotAc	Buffer “B”/ 150 mM PotAc	Buffer “A”/ 5 mM MagAc	Buffer “A”/ 25 mM SodAc/ 5 mM MagAc	Buffer “A”/ 50 mM SodAc/ 5 mM MagAc	Buffer “A”/ 100 mM SodAc/ 5 mM MagAc	Buffer “A”/ 150 mM SodAc/ 5 mM MagAc	Buffer “B”/ 5 mM MagAc	Buffer “B”/ 25 mM SodAc/ 5 mM MagAc	Buffer “B”/ 50 mM SodAc/ 5 mM MagAc	Buffer “B”/ 100 mM SodAc/ 5 mM MagAc	Buffer “B”/ 150 mM SodAc/ 5 mM MagAc
D	Buffer “A”/ 25 mM PotAc/ 5 mM MagAc	Buffer “A”/ 50 mM PotAc/ 5 mM MagAc	Buffer “A”/ 100 mM PotAc/ 5 mM MagAc	Buffer “A”/ 150 mM PotAc/ 5 mM MagAc	Buffer “B”/ 25 mM PotAc/ 5 mM MagAc	Buffer “B”/ 50 mM PotAc/ 5 mM MagAc	Buffer “B”/ 100 mM PotAc/ 5 mM MagAc	Buffer “B”/ 150 mM PotAc/ 5 mM MagAc	Buffer “A”/ 25 mM SodAc/ mM MagAc/ 1 mM ATP	Buffer “A”/ 25 mM SodAc/ 5 mM MagAc/ 5 mM ATP	Buffer “A”/ 50 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “A”/ 50 mM SodAc/ 5 mM MagAc/ 5 mM ATP
E	Buffer “A”/ 100 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “A”/ 100 mM SodAc/ 5 mM MagAc/ 5 mM ATP	Buffer “A”/ 150 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “A”/ 150 mM SodAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 25 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 25 mM SodAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 50 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 50 mM SodAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 100 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 100 mM SodAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 150 mM SodAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 150 mM SodAc/ 5 mM MagAc/ 5 mM ATP

F	Buffer “A”/ 25 mM PotAc/5 mM MagAc/ 1 mM ATP	Buffer “A”/ 25 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “A”/ 50 mM PotAc/ 5 mM MagAc/ 1 mM ATP	Buffer “A”/ 50 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “A”/ 100 mM PotAc/ 5 mM MagAc/ 1 mM ATP	Buffer “A”/ 100 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “A”/ 150 mM PotAc/ 5 mM MagAc/ 1 mM ATP	Buffer “A”/ 150 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 25 mM PotAc/5 mM MagAc/ 1 mM ATP	Buffer “B”/ 25 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 50 mM PotAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 50 mM PotAc/ 5 mM MagAc/ 5 mM ATP
G	Buffer “B”/ 100 mM PotAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 100 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “B”/ 150 mM PotAc/ 5 mM MagAc/ 1 mM ATP	Buffer “B”/ 150 mM PotAc/ 5 mM MagAc/ 5 mM ATP	Buffer “A”/ 25 mM SodAc/ 1 mM ATP	Buffer “A”/ 50 mM SodAc/ 1 mM ATP	Buffer “A”/ 100 mM SodAc/ 1 mM ATP	Buffer “A”/ 150 mM SodAc/ 1 mM ATP	Buffer “B”/ 25 mM SodAc/ 1 mM ATP	Buffer “B”/ 50 mM SodAc/ 1 mM ATP	Buffer “B”/ 100 mM SodAc/ 1 mM ATP	Buffer “B”/ 150 mM SodAc/ 1 mM ATP
H	Buffer “A”/ 25 mM PotAc/ 1 mM ATP	Buffer “A”/ 50 mM PotAc/ 1 mM ATP	Buffer “A”/ 100 mM PotAc/ 1 mM ATP	Buffer “A”/ 150 mM PotAc/ 1 mM ATP	Buffer “B”/ 25 mM PotAc/ 1 mM ATP	Buffer “B”/ 50 mM PotAc/ 1 mM ATP	Buffer “B”/ 100 mM PotAc/ 1 mM ATP	Buffer “B”/ 150 mM PotAc/ 1 mM ATP	Buffer “A”/ 50 mM PotAc/ 5 mM ATP	Buffer “B”/ 50 mM PotAc/ 5 mM ATP	Buffer “A”/ 5 mM NiCl ₂	Buffer “A”/ 5 mM NiAcetate

Buffer “A”: 20 mM Tris pH 8.0, Buffer “B”: 25 mM HEPES pH 7.3. KCl was used at concentration of 150 mM.

Table 2.3: Buffer screen to test the stability of *T. aquaticus* MutL

Buffer Fraction	Buffer A	Buffer B	Buffer C	Buffer D	Buffer E	Buffer F	Buffer G
20 mM Tris pH 8.0	20 mM Tris pH 8.0	20 mM Tris pH 8.0	20 mM Tris pH 8.0	20 mM Tris pH 8.0	25 mM HEPES pH 7.3	25 mM HEPES pH 7.3	25 mM HEPES pH 7.3
1.4 mM β -Me	1.4 mM β -Me	1 mM EDTA	1 mM EDTA	1.4 mM β - Me	1.4 mM β -Me	1.4 mM β -Me	1.4 mM β -Me
0.2 mM PMSF	0.2 mM PMSF	5 mM DTT	5 mM DTT	150 mM PotAc	0.2 mM PMSF	150 mM PotAc	0.2 mM PMSF
500 mM NaCl	150 mM NaCl	150 mM KCl	150 mM KCl	5% glycerol	150 mM KCl	5% glycerol	150 mM PotAc
5% glycerol	5% glycerol	5% glycerol	5 mM MgCl ₂	-	5% glycerol	-	5% glycerol
-	-	-	5% glycerol	-	-	-	-

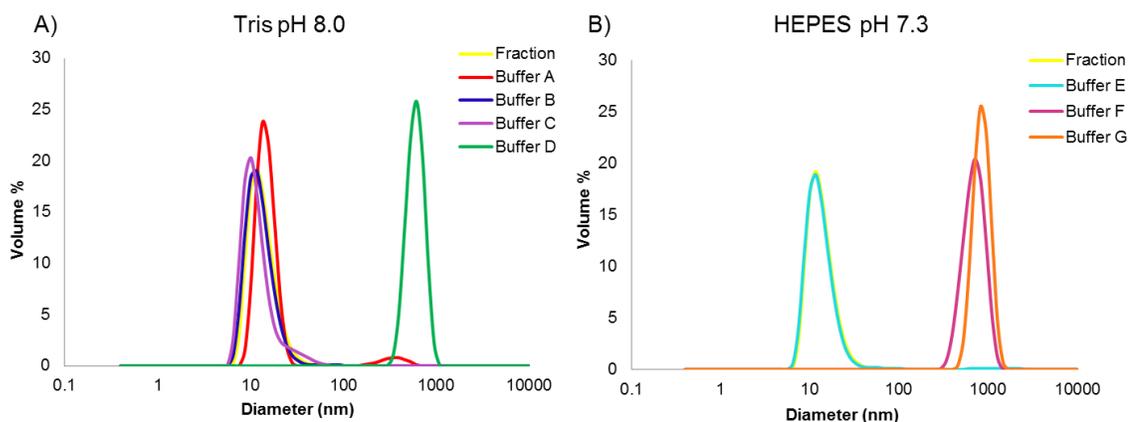


Figure 2.15: Stability of *T. aquaticus* MutL under different buffers agents and salts. (A) DLS measurements of the buffer screen using (A) Tris pH 8.0 or (B) HEPES pH 7.3 with potassium chloride or potassium acetate salts. Fraction (yellow) refers to the most concentrated fraction of the purification of *T. aquaticus* MutL. The other conditions (Buffer A, B, C...) show the behavior of this fraction under the respective buffer screen. Table 2.3 describes the individual components of each buffer.

2.2.8) *B. subtilis* MutL undergoes several conformational changes in the presence of ATP

AFM images of *B. subtilis* MutL revealed the four conformational states that are also present in eukaryotic MutL α (extended, one-armed, semi-condensed and condensed conformations) in the absence of the adenine nucleotide (apo form). The same types of conformations were also observed in the presence of 1 mM ADP (Figure 2.16A) and 1 mM ATP (Figure 2.16B).

In general, for the apo condition, the fraction of the extended state (33%) was lower than with the eukaryotic MutL homologs (50-60%, (Sacho et al., 2008)). Meanwhile, there was an increase of the condensed form (23%) in this condition compare to the same state in human and yeast MutL α (10-15%, (Sacho et al., 2008)) (Figure 2.17, apo).

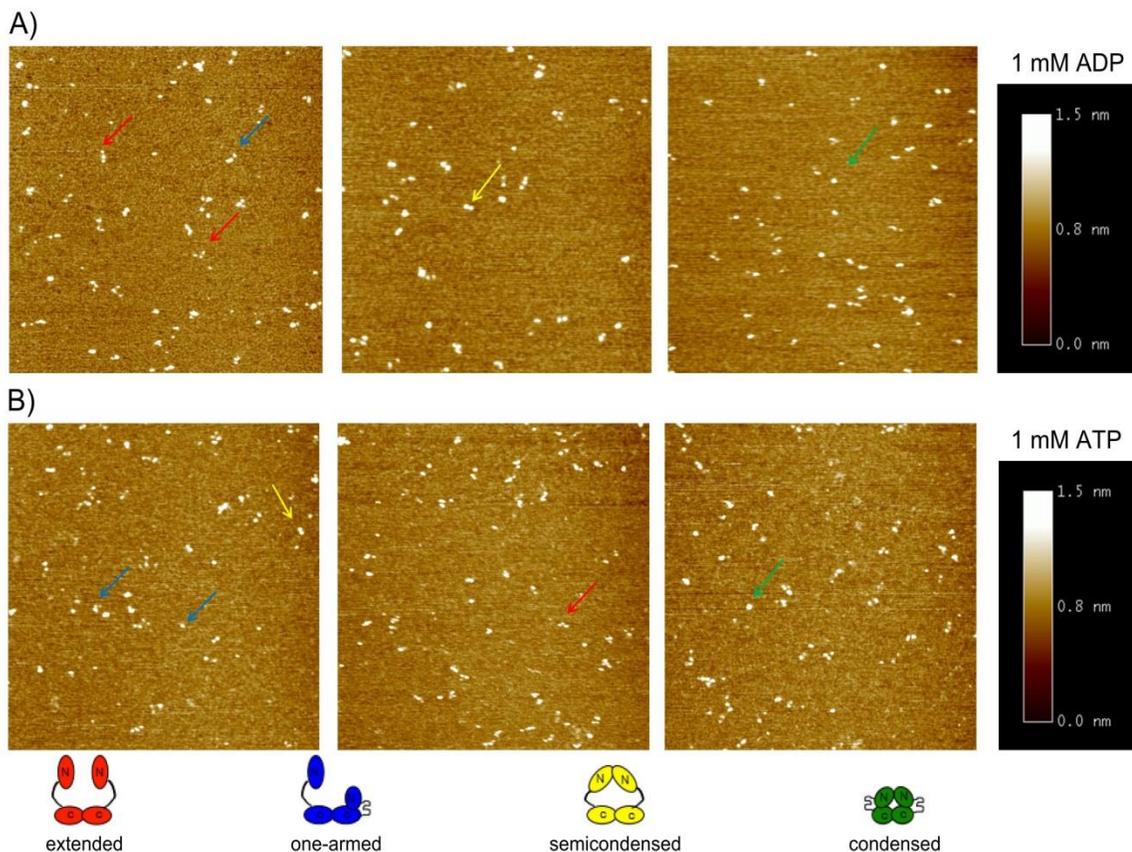


Figure 2.16: AFM images of *B. subtilis* MutL. Images deposited in the presence of (A) 1 mM ADP and (B) 1 mM ATP. Arrows indicate some examples of the four different conformational states of MutL: extended (red), one-armed (blue), semicondensed (yellow), and condensed (green).

Even though there was an increase in the proportion of *B. subtilis* MutL in the condensed state in the presence of ATP, it was not as significant as with the eukaryotic MutL α (from 23% to 43% for *B. subtilis* MutL versus 16% to 70% of yeast MutL α). Interestingly, there were higher proportions of *B. subtilis* MutL in the intermediate states (one-armed (23%) and semicondensed (16%)) than in the case of MutL α (15% and 12% respectively) (Figure 2.17, 1 mM ATP).

Furthermore, for *B. subtilis* MutL in the presence of 1 mM ADP, 46% of the molecules were condensed and only 13% were extended (Figure 2.17, 1 mM ADP). This shows a larger population of condensed MutL compared to the 5 mM ADP data for yeast MutL α (34% condensed).

Contrary to eukaryotic MutL α where the proportions of ADP compared to ATP were significantly different (for example the condensed form in 1 mM ADP was 34% while the same conformation with 1 mM ATP was 70%), there were no major differences in *B. subtilis* MutL incubated with either adenine nucleotide (the percentage change between the two cofactors was always about 3-4 points). Therefore, we concluded that *B. subtilis* MutL is less sensitive to the nature of the nucleotide than yeast MutL α .

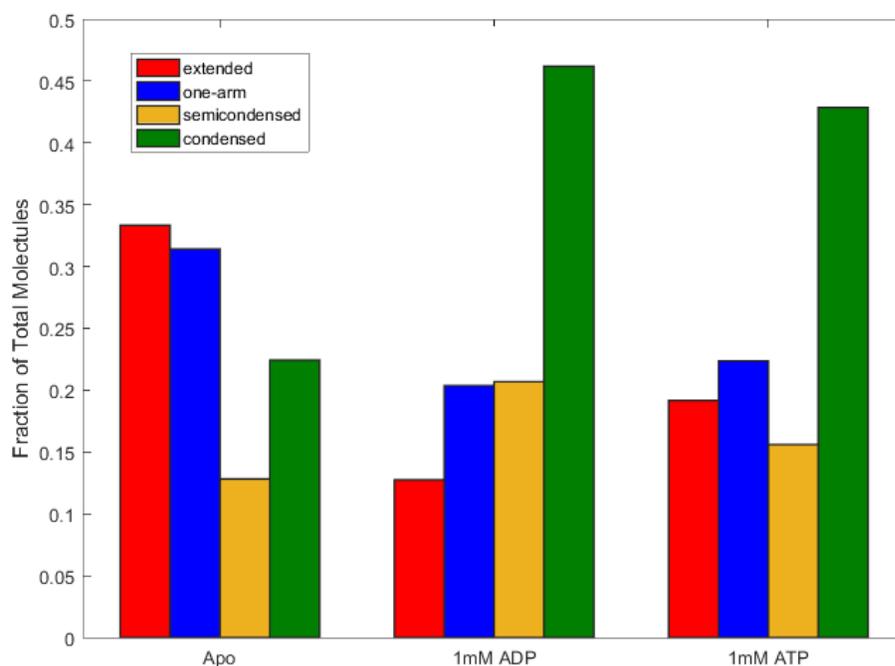


Figure 2.17: *B. subtilis* MutL conformational states distribution in the absence (apo) and presence of 1 mM ADP or 1 mM ATP. (Courtesy of H. Wilkins).

2.3) DISCUSSION

MutL undergoes a series of conformational changes in the presence of ATP. It has been proposed that these conformational changes mediate the regulation and interaction of the protein with downstream factors of the MMR pathway (Ban et al., 1999; Sacho et al., 2008). However, not all MutL homologs undergo identical conformational changes and this variation may be due to the mechanism of strand discrimination specific to the organism. For example, *E. coli* MutL, which uses the methyl dependent mechanism, transforms from an open to a compact conformation after binding of ATP (Ban et al., 1999). On the other hand, yeast or human MutL α , which both use the nick-dependent process, have four major asymmetric conformations in the absence of ATP (extended, one-armed, semicondensed, and condensed). The proportion of molecules in these conformational states changes upon binding of ATP (Sacho et al., 2008).

It is unknown whether these asymmetric conformations are unique to heterodimers and endonuclease active MutL homologs, or if homodimeric MutL homologs with endonuclease activity could adopt them as well. To this end, in this chapter I studied the behavior of the endonuclease active homodimeric *B. subtilis* MutL upon binding to ATP.

First, we analyzed whether or not *B. subtilis* MutL could undergo the same conformational change as homodimeric *E. coli* MutL. Unlike *E. coli* MutL, where a distinct shift to a smaller shape was observed upon nucleotide binding, the shape of *B. subtilis* MutL did not change when incubated with ATP. Based on this, we concluded that *B. subtilis* MutL has a different behavior in the presence of the adenine nucleotide.

One limitation of the experiments with *B. subtilis* MutL was that the ratio of absorbance at 260 nm over 280 nm (A₂₆₀/A₂₈₀) did not increase as it did for *E. coli* MutL. Although we could not confirm ATP-binding with the size exclusion chromatography method, we inferred it through the thermal stability experiments since the protein was highly stabilized in the presence of both ATP and Mg²⁺ (with no effect if either ATP or Mg²⁺ were added independently).

These results were consistent with previously published data that found that ATP stabilizes MutL and that its binding cannot take place without Mg²⁺ ions (Ban et al., 1999). In this case, the octahedral coordination of the Mg²⁺ ions allows the linking of all phosphates to the protein (Ban et al., 1999). This means that each phosphate contributes to chelating the Mg²⁺ ion and that the interaction of Mg²⁺ with the nucleotide and the protein is crucial for the binding and stabilization of MutL to ATP. Even though the Mg²⁺ ions coordination is unusual, it has been observed in all the crystal structures of MutL homologs solved to date (human PMS2 (Guarné et al., 2001), yeast PMS1 (Arana et al., 2010), and human MLH1 (Wu et al., 2015)) and in the structure of the NgyrB (DNA Gyrase B) protein which belongs to the same family of MutL (Wigley et al., 1991).

When using the same methodology as with eukaryotic MutL α homologs, we found that, similar to its MutL α homologs, *B. subtilis* MutL has four conformational states in the absence of the adenine nucleotide. Conversely, when bound to ATP the prokaryotic protein underwent a more subtle conformational change towards condensation than the yeast MutL α . While there was a 54% increase in the condensed state of yeast MutL α bound to ATP, there was only a 19% increase under the same conditions for *B. subtilis*

MutL. Since less than half of the population of *B. subtilis* MutL was in the condensed state upon ATP binding, the gel filtration experiments for this protein were consistent in that there is an equilibrium between the condensed form with the other three states which cannot be resolved using this technique.

Additionally, there was a significant increase of the intermediate states of one-armed and semicondensed for *B. subtilis* MutL compared with the eukaryotic MutL α homologs. Under the ATP concentrations of the AFM experiments done by Sacho and colleagues (1 mM and 5 mM), yeast MutL α had less than 20% of intermediate conformations. These states only increased when using 0.1 mM ATP. This was attributed to the hydrolysis of one ATP molecule in the MLH1 monomer of the eukaryotic MutL homolog (Sacho et al., 2008). Considering that *B. subtilis* MutL is a much weaker ATPase (K_m = 400 μ M) (Pillon et al., 2010) than the MLH1 monomer of yeast MutL α (K_m = 69 μ M) (Hall et al., 2002), it is likely that there is more residual ATP in *B. subtilis* MutL that is not bound to the protein, and this drives the intermediate conformations to be in a higher proportion.

To complement the previous findings, the ATP-dependent conformational changes should be evaluated in another organism with endonuclease activity such as *T. aquaticus* MutL. Comparing the different conformations and proportions of *B. subtilis* and *T. aquaticus* MutL homologs to the shape of *E. coli* MutL will show commonalities between homodimeric MutL homologs that use nick-dependent MMR, not found in those that use the methyl-dependent process (*E. coli* MutL). This will determine if the presence of the endonuclease site confers new characteristics to the behavior of the MutL protein as well

as the connections between the N- and C-terminal domains of MutL in the presence of ATP.

To our knowledge, this is the first time that the different conformations of MutL in a prokaryotic organism such as *Bacillus subtilis* have been shown. Although the condensation of *B. subtilis* MutL upon ATP binding is not as drastic as with the eukaryotic MutL α homologs, we can clearly see similarities between the ATP-dependent conformational changes in MutL homologs belonging to the nick-dependent MMR pathway. These features are not seen in the MutL homologs belonging to the methylation-dependent pathway. Our work has demonstrated that the asymmetric conformations of MutL do not necessarily stem from differences in prokaryotes and eukaryotes (homodimers versus heterodimers). This further justifies the use of prokaryotic MutL homologs as models for understanding the MMR process. In fact, we can use these common features to further investigate a long-standing question in this field: What is the mechanistic role of ATP in the methyl-independent strand discrimination process?

CHAPTER 3

A NOVEL APPROACH TO STUDY THE ENDONUCLEASE MECHANISM OF THE MUTL MISMATCH REPAIR PROTEIN

Despite significant progress over the past decade in defining the role of MutL in the early steps of the MMR process (Kadyrov et al., 2006; Kosinski et al., 2008; Pillon et al., 2010; Pluciennik et al., 2010; Gueneau et al., 2013), more research is needed to elucidate the molecular mechanism of the interaction between the endonuclease domain of MutL and DNA. This has been a challenging task because this domain can cleave the DNA but it does not bind to it.

This chapter is focused on the characterization of a novel approach to bypass the DNA binding defect of the endonuclease domain of MutL. We designed fusion proteins comprising a sequence specific, high affinity DNA-binding domain connected by a linker to the C-terminal domain of MutL (Figure 3.1). By anchoring the DNA to a single position through the DNA binding domain, we expected the nicking activity to be stimulated, allowing us to study its endonuclease mechanism.

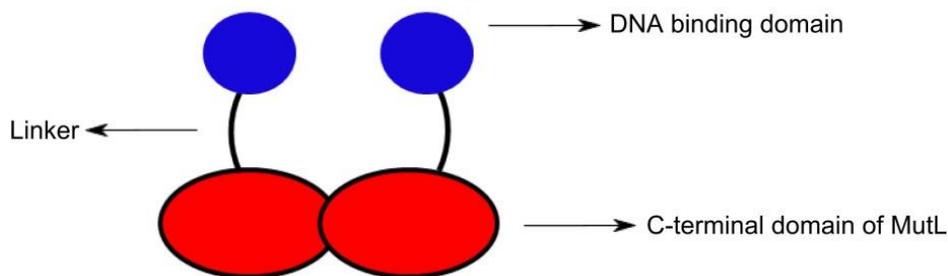


Figure 3.1: *Schematic representation of the fusion protein.*

To biochemically characterize the fusion protein, I asked four main questions: 1) Does the fusion bind to linear DNA substrates? 2) Does the fusion have nicking activity? 3) If so, is MutL C-terminal domain responsible for this nicking? 4) Is the nicking activity stimulated by the presence of the DNA binding domain? The first three questions were answered using a fusion protein with a 32 amino acid linker. The last question was answered using a modified fusion with an 8 amino acid linker.

3.1) MATERIALS AND METHODS

3.1.1) Cloning of SKN1-MutL_{CTD} fusions with varying linker lengths, and lacking endonuclease activity

The constructs consisting of the DNA binding domain of the SKN1 protein ((herein referred to as the SKN1 domain) amino acids 450-533) from *Caenorhabditis elegans* connected to the C-terminal domain of *Bacillus subtilis* MutL (amino acids 433-627) by linkers of different lengths were cloned into pProEx Hta vector (Invitrogen) using NcoI and XhoI restriction sites to flank the fusion. The linkers connecting both domains were flanked by BamHI and EcoRI restriction sites which allowed for simple alteration of the linker region. Initially, five constructs were cloned with linkers of 8, 15, 21, 25, and 30 amino acids (L8, L15, L21, L25, and L30 fusions respectively (Table 3.1)) as part of the undergraduate thesis project of Julia Cai in the Guarné Lab. Using the fusion with the 30 amino acid linker as a template, two more fusions with 28 and 32 amino acids (Table 3.1) were generated using the Q5-Site Directed Mutagenesis Kit (New England Biolabs).

Table 3.1: Fusion proteins with varying linker length*

Linker length (amino acid)	Amino acid sequence of the linker	Fusion protein	Plasmid name
8	GSASKSEF	L8 fusion	pAG 8887
15	GSKGEASGSGSKSEF	L15 fusion	pAG 8886
21	GSGSGSESKSASKGEASKSEF	L21 fusion	pAG 8896
25	GSGSGSESKSASKGEASGSGSKSEF	L25 fusion	pAG 8897
30	GSEGKSSGSGSESKSASKGESSGSGS KSEF	L30 fusion	pAG 8882
28	GSEGKSSGSGSESKSASKGESSGSGS KS	L28 fusion	pAG 8924
32	GSEGKSSGSGSESKSASKGESSGSGS KSEFGS	L32 fusion	pAG 8923

*Amino acid names are in one letter nomenclature.

Following the same methodology, two variants lacking endonuclease activity (Pillon et al., 2010) of the L32 fusion were generated. The first variant substituted the aspartate 462 for asparagine (D462N), and the second one substituted the glutamate 468 for lysine (E468K). All primers were designed according to the software NEBaseChanger.neb.com (Table 3.2). Constructs were verified by DNA sequencing (MOBIX, McMaster University).

Table 3.2: Primers used to generate SKN1-MutL_{CTD} fusions and dead variants

Name	Primer sequence	Plasmid name	Description
ag1953	5' CGCCGCCCAAAAACGTATTAA 3'	pAG 8907	E468K
ag1954	5' TGCTGGTCGATAATATATAGGC 3'		
ag1955	5' ATATATTATCAACCAGCACGCC 3'	pAG 8906	D462N
ag1956	5' AGGCCGTTTTCGTTTTGTG 3'		
ag1975	5' GGCAGCTTCTCTGATCGGGTTCCAATT ATG 3'	pAG 8923	L32 fusion
ag1976	5' TTCGCTCTTACTTCCTGAG 3'		
ag1977	5' TCTGATCGGGTTCCAATTATG 3'	pAG 8924	L28 fusion
ag1978	5' GCTCTTACTTCCTGAGCC 3'		

3.1.2) Expression and protein purification of SKN1-MutL_{CTD} fusions and variants lacking endonuclease activity

The plasmid encoding the L30 fusion (pAG 8882) was transformed in *E. coli* BL21 (DE3) cells. Cells were grown to OD₆₀₀ of 0.7 at 37°C and protein expression was induced by adding IPTG to a final concentration of 0.5 mM, followed by incubation with orbital agitation at 25°C for 5 h. Cells were harvested by centrifugation and stored at -80°C.

To purify the protein, one cell pellet (1 L) was resuspended in 20 mL of lysis buffer (25 mM Tris pH 7.5, 1 M NaCl, 2.8 mM β-mercaptoethanol, 0.2 mM PMSF, and 5% glycerol (v/v)). After the addition of protease inhibitors (PMSF, Leupeptin, Benzamidine, Pepsatin A), cells were lysed by sonication and protease inhibitors were added again. The lysate was clarified by centrifugation at 39,000 x g and loaded into a 1 mL HiTrap Chelating HP Nickel column (GE healthcare) equilibrated with buffer containing 25 mM Tris pH 7.5, 0.5 M NaCl, 1.4 mM β-mercaptoethanol, 0.2 mM PMSF, 30 mM imidazole, and 5% glycerol (v/v). After two washes with 30 mM and 60 mM imidazole, the protein was eluted with a linear gradient to 300 mM imidazole for 16 CV at 0.4 mL/min. Fractions containing the fusion protein were pooled and further purified by ion exchange chromatography over a MonoS column (5/50) (GE Healthcare) equilibrated with 25 mM Tris pH 7.5, 200 mM KCl, 5 mM EDTA, 1.4 mM β-mercaptoethanol, and 5% glycerol (v/v). The protein was eluted in a linear salt gradient to 800 mM KCl for 15 CV at 0.5 mL/min. Two peaks corresponding with different isoforms of the protein were concentrated independently using 30 kDa, 2 mL concentrator (MILLIPORE) and frozen

at a final concentration of 12.6 μM and 29.5 μM , respectively. Purity was evaluated by 12% SDS-PAGE (75 min at 150 V).

Fusion proteins with linkers of 28 and 32 amino acids and lacking endonuclease activity (D462N, E468K) were grown in BL21 (DE3) cells to OD_{600} of 0.7. Expression and purification of these proteins were conducted using the same protocol as described above.

3.1.3) Expression and purification of *B. subtilis* MutL_{CTD}

C-terminal domain (CTD) of *B. subtilis* MutL (pAG 8188) was grown in *E. coli* BL21 (DE3) cells to OD_{600} of 0.7 at 37°C, and protein expression was induced by the addition of IPTG to a final concentration of 1 mM, followed by incubation with orbital agitation at 37°C for 3 h. Protein was purified as described by Pillon et al. (2010) with minor modifications. The Histidine-tag cleavage was not carried out - instead, protein was concentrated using a 10 kDa MWCO, 20 mL concentrator (Vivaspin) in storage buffer (20 mM Tris pH 8.0, 5 mM DTT, 100 mM KCl, and 5% glycerol (v/v)), and frozen in 5 μL aliquots at a final concentration of 1.13 mM. Purity was evaluated by 15% SDS-PAGE (75 min at 150 V).

3.1.4) Expression and purification of *B. subtilis* β -sliding clamp

β -sliding clamp from *B. subtilis* MutL (β -clamp) (pAG 8337) was expressed in *E. coli* BL21 (DE3) *recA*⁻ (BLR) cells to OD_{600} of 0.7 at 37°C. Protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM, followed by incubation with

orbital agitation at 37°C for 2 h (O'Donnell et al., 1993; Simmons et al., 2008). Protein was purified as described by Pillon et al. (2011) with minor modifications. A final step was added to the purification protocol consisting of size exclusion chromatography where the protein (500 µL) was loaded into a Superdex-200 (GE Healthcare) in buffer containing 20 mM Tris pH 7.6, 5 mM DTT, 100 mM KCl, and 5% glycerol (v/v). Fractions from the apex of the peak were concentrated to 177 µM using a 30 kDa MWCO, 2 mL concentrator (Vivaspin). Purity was evaluated by 15% SDS-PAGE (75 min at 150 V).

3.1.5) Expression and protein purification of SKN1-L8-MutL_{CTD} heterodimer

Plasmids encoding the C-terminal domain (CTD) of *B. subtilis* MutL (pAG 8188) and L8 fusion (pAG 8887) were co-transformed into *E. coli* BL21 (DE3) cells and these cells were grown to OD₆₀₀ of 0.7 at 37°C. Protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM, followed by incubation with orbital agitation at 25°C for 5 h. Cells were harvested by centrifugation and stored at -80°C. To purify the protein, one cell pellet (1 L) was resuspended in 20 mL of lysis buffer (25 mM Tris pH 7.5, 0.5 M NaCl, 1.4 mM β-mercaptoethanol, 0.2 mM PMSF, 30 mM imidazole and 5% glycerol (v/v)). Then, protease inhibitors (PMSF, pepstatin A, leupeptin, benzamide) were added to the solution. Cells were lysed by sonication and protease inhibitors were added again. The lysate was clarified by centrifugation at 39,000 x g for 40 min and the supernatant was filtered with 0.2 µm filter (PALL Corporation) and stored at 4°C overnight.

The following day, the cell lysate was spun down at 39,000 x g for 40 min and the supernatant was filtered with 0.2 µm filter (PALL Corporation). Filtered sample was loaded into a 5 mL HiTrap Chelating HP Nickel column (GE healthcare) equilibrated with buffer containing 25 mM Tris pH 7.5, 0.5 M NaCl, 1.4 mM β-mercaptoethanol, 0.2 mM PMSF, 30 mM Imidazole, and 5% glycerol (v/v). After two washes with 30 mM and 45 mM imidazole, the protein was eluted with a linear gradient to 300 mM imidazole for 10 CV at 1 mL/min. Protein was further purified by ion exchange chromatography with a S-Sepharose column (GE Healthcare) equilibrated with 25 mM Tris pH 7.5, 200 mM KCl, 1 mM EDTA, 5 mM DTT, 0.2 mM PMSF, and 5% glycerol (v/v). One wash at 200 mM KCl allowed the CTD of MutL to flow through the column.

The heterodimer was finally eluted in a linear salt gradient to 800 mM KCl for 8 CV at 1 mL/min. Three peaks were part of the elution profile: the first one corresponded to the expected SKN1-L8-MutL_{CTD} heterodimer, the second and third peaks corresponded to two isoforms of the L8 fusion homodimer that were unstable in solution. CTD and the heterodimer were concentrated independently using 10 kDa MWCO, 20 mL concentrator (Vivaspin) and 30 kDa MWCO, 4 mL concentrator (Vivaspin Turbo) respectively. Proteins were frozen in 5 µL aliquots at final concentrations of 96.9 µM and 7.6 µM respectively. Purity was evaluated by 12% SDS-PAGE (75 min at 150 V). This protocol was optimized in collaboration with another graduate student (Linda Liu) in the Guarné Lab.

3.1.6) Cloning, expression and purification of SKN1-L8-MutL_{CTD} heterodimer variants

Catalytically inactive variants of the heterodimer were generated using the Q5-Site Directed Mutagenesis Kit (New England Biolabs (NEB)). Since the heterodimer consists of two subunits with the endonuclease domain of MutL, active site point mutations were made in either the L8 subunit (E468K^{L8*}), the MutL_{CTD} subunit (E468K^{CTD}), or in both endonuclease sites (E468K). An additional variant with a mutation at a different residue (D462N) in both endonuclease sites was also generated. Point mutations were verified by DNA sequencing (MOBIX, McMaster University). Expression and purification of these variants were carried out as described in section 3.1.5.

3.1.7) Generation of linear 195 bp substrates

A 195 bp substrate (+SKN1M) was generated through PCR amplification using 5' fluorescently labeled primers with 6-Carboxyfluorescein (6-FAM) (BioBasic Inc). Primers P195MF (5'TGTAAAACGACGGCCAGTGAATTCGAGCTCGG3') and P195MR (5'AGTTAGCTCACTCATTAGGCACCCCAGGC3') amplified the 378-572 region of pUC19 (Invitrogen) containing an SKN1 site (GTCAT) at position 465. PCR reactions (50 µL) contained 300 ng pUC19 template, 0.5 µM forward primer (P195MF), 0.5 µM reverse primer (P195MR), 0.4 mM dNTPs, 1× PFU buffer, and 1 unit/µL PFU enzyme. The cycling conditions were as follows: initial denaturation at 95°C for 5 min, following 20 cycles of 95°C for 30 s, 55°C for 1 min and 72°C for 35 s, and the final extension period at 72°C for 20 min. PCR fragments were loaded into a 2% agarose gel

and ran for 45 min at 100 V. Bands were excised under UV light (320 nm) and gel extracted using the QIAEX II Agarose gel extraction kit (QIAGEN).

Other substrates with and without the SKN1 site were generated using modified versions of pUC19 templates. These modified forms were produced using the Q5-Site Directed Mutagenesis Kit (NEB). All primers were designed according to the software from NEBaseChanger.neb.com (Table 3.3). I first modified the original pUC19 vector by mutating the SKN1 site (GTCAT) at position 465-469 bp to CAGTT. PCR amplification (as described above) of this modified pUC19 generated a 195 bp substrate lacking the SKN1 site (-SKN1). This new version of pUC19 underwent further modifications (Q5-Site Directed Mutagenesis Kit) to insert an SKN1 site at either position 410-414 bp or at position 536-540 bp. PCR amplifications of the new pUC19 variants generated 195 bp substrates with an SKN1 site in the beginning (+SKN1B) and at the end (+SKN1E) of the duplex DNA substrates. Mutations were verified by DNA sequencing (MOBIX, McMaster University).

In all, four different templates (original pUC19, modified pUC19 I with no SKN1 site, modified pUC19 II with SKN1 site in position 410-414 bp, and modified pUC19 III with SKN1 site in position 536-540 bp) were used to create the following 195 bp substrates: +SKN1M (middle), -SKN1 (no SKN1), +SKN1B (beginning), and +SKN1E (end), respectively. All substrates were generated using the same PCR conditions as described above.

Table 3.3: Primers used to generate different versions of pUC19 and molecular weight marker

Name	Primer sequence	Description
ag2028	5' GTAATCATGCAGTTAGCTGTTTCCTGTG TGAAATTGTTATC ^{3'}	Modified pUC19 I without SKN1 site
ag2029	5' CCAAGCTTGCATGCCTGC ^{3'}	
ag2052	5' TCGAGCTCGGGTCATGGGGATCCTCTAG AG ^{3'}	Modified pUC19 II with SKN1 site in position 410-414 bp
ag2053	5' ATTCACTGGCCGTCGTTTTAC ^{3'}	
ag2054	5' GGCAGCTTCTCTGATCGGGTTCCAATTA TG ^{3'}	Modified pUC19 III with SKN1 site in position 536-540 bp
ag2055	5' GGAAGCATAAGTCATAAAGCCTGGGGT GCC ^{3'}	
puc195R	5' AGTTAGCTCACTCATTAGGCACCCCAGG C ^{3'}	195 bp substrate
puc150R	5' CCGGCTCGTATGTTGTGTGGAATTGTG ^{3'}	150 bp substrate
puc100R	5' AAACAGCTAACTGCATGATTACGCCAA GC ^{3'}	100 bp substrate
puc80R	5' ACGCCAAGCTTGCATGCCTGC ^{3'}	80 bp substrate
puc54R	5' GACTCTAGAGGATCCCCGGGTA ^{3'}	54 bp substrate
puc32R	5' CCGAGCTCGAATTCCTGGCCGTCG ^{3'}	32 bp substrate

3.1.8) Protein-DNA binding assays

Electrophoretic mobility shift assays (EMSAs) were performed with the doubly fluorescent labeled 195 bp substrate with the SKN1 site in the middle (+SKN1M, 10 nM) and increasing concentrations of protein from 10 nM to 640 nM. Reactions (10 μ L) were incubated at room temperature for 30 min in reaction buffer (20 mM Tris pH 8.0, 90 mM KCl, 1 mM DTT, 5 mM EDTA, 0.1 mg/mL BSA, and 15% glycerol (v/v)). Then, samples were loaded into a 5% Tris-glycine polyacrylamide native gel and separated for 50 min at 80 V in 1X Tris-glycine buffer. Native fluorescent gels were visualized using the Typhoon Trio+ (GE Healthcare, CMCB McMaster University). EMSAs were also performed with the addition of poly dIdC as competitor DNA (0.65-20.8 ng/ μ L).

3.1.9) Mismatch independent endonuclease assays for SKN1-MutL_{CTD} fusions

Mismatch-independent MutL endonuclease assays were performed as described in Pluciennik et al. (2010) and Pillon et al. (2015) with some modifications. SKN1-MutL_{CTD} fusions were incubated with linear 195 bp fluorescent labeled substrates (10 nM) in the absence and presence of equimolar amounts of *B. subtilis* β -clamp protein. Protein concentrations ranged from 1 to 1.3 μ M for the CTD of MutL and for fusions of 32 amino acid linkers. Concentrations for heterodimer fusions with 8 amino acid linkers ranged from 240 nM to 960 nM. Reactions (10 μ L) were incubated at 37°C for 1 hour in reaction buffer containing 20 mM Tris pH 7.6, 30 mM KCl, 1 mM MnCl₂, 1 mM MgCl₂, 0.05 mg/mL BSA, 152 μ M zinc acetate dehydrate (Zn(CH₃COO)₂·2H₂O), and 4% glycerol (v/v). Reactions were stopped by adding 25 mM EDTA and 1 mg/mL Proteinase K and incubated at 55°C for 20 min. Immediately afterward, 2 \times -loading dye (90% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, 0.5 mM EDTA, 10% glycerol) was added, and the reaction mixtures were incubated at 95°C for 5 min. Boiled samples were loaded onto a pre-warmed 8% urea polyacrylamide gel (8 M Urea) in 0.5 \times Tris-Borate-EDTA buffer and bands were resolved for 23 min at 200 V. Gels were visualized using the Typhoon Trio+ (GE Healthcare, CMCB McMaster University).

3.1.10) Design of fluorescently labeled molecular weight marker

A molecular weight marker for the endonuclease assays was prepared by amplifying different fragments (195, 150, 100, 80, 54, 32 bp) from the modified pUC19 vector lacking the SKN1 site. In all cases, fluorescently labeled primer P195MF was used as

forward sequence while reverse primers were not fluorescent (Table 3.3 (puc195R, puc150R, puc100R, puc80R, puc54R, puc32R). PCR reactions were carried out as described in section 3.1.7. PCR products were mixed at the same concentrations, and this was used as molecular weight marker in the endonuclease assays (10 nM).

3.2) RESULTS

3.2.1) Characterization of the SKN1-MutL_{CTD} fusion

A former student in our laboratory (Julia Cai) generated variants of the C-terminal region of MutL including a sequence-specific DNA binding domain at the N-terminus of the endonuclease domain. The DNA binding domain of the SKN1 transcription factor from *C. elegans* was chosen because it is a monomer and binds with high affinity (1 nM K_d) to a non-palindromic DNA sequence ((G/A)TCAT) (Rupert et al., 1998). Initially, the two domains were connected by linkers of different lengths. Linkers were rich in glycine and serine, to give them flexibility, as well as lysine and glutamate to increase the solubility of the fusion proteins (Chen et al., 2013).

Fusion proteins were assayed for expression and solubility including those with 8, 15, 21, 25, and 30 amino acid linker lengths. From here, fusions with 8 and 15 amino acid linkers were insoluble, while fusions with 21 and 25 amino acid linkers were unstable after cell lysis. The construct with the 30 amino acid linker presented the best ratio between expressed over soluble protein in BL21 (DE3) Star cells at 25°C for 5 h (Figure 3.2).

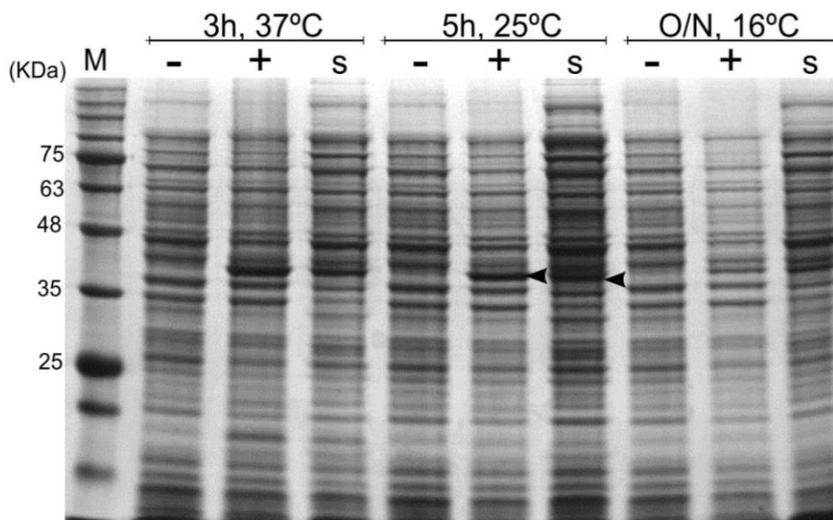


Figure 3.2: *Small scale induction of L30 fusion. SDS-PAGE showing (from left to right): uninduced sample (-), induced sample with 0.5 mM IPTG (+), and soluble fraction of the lysate after induction (s) under three different conditions: 37°C for 3 h, 25°C for 5 h, and 16°C overnight.*

The L30 fusion was purified by nickel affinity chromatography followed by ionic exchange (S-sepharose) chromatography. During the second step, the protein eluted as two distinct peaks at 448 mM and 498 mM KCl. Both peaks corresponded to the same fusion protein as judged by SDS-PAGE. The first of the two peaks was the major isoform. However, dynamic light scattering revealed that this population was aggregated. To improve the separation of the two isoforms, we substituted the S-sepharose for a MonoS (5/50) column, which has a smaller bed size (Amersham, 2002) and could lead to an increase in resolution of the peaks.

The same two isoforms were resolved, but this time, the intensity of the second peak was significantly higher than the first isoform. Since the behavior of the protein in this peak was not known, both forms of the protein were concentrated and frozen

independently. DLS measurements were performed throughout the purification to check the quality of each isoform. After thawing, the first isoform again became aggregated and was rejected for downstream experiments. The second one was monodisperse and this form was stored for future biochemical assays (Figure 3.3).

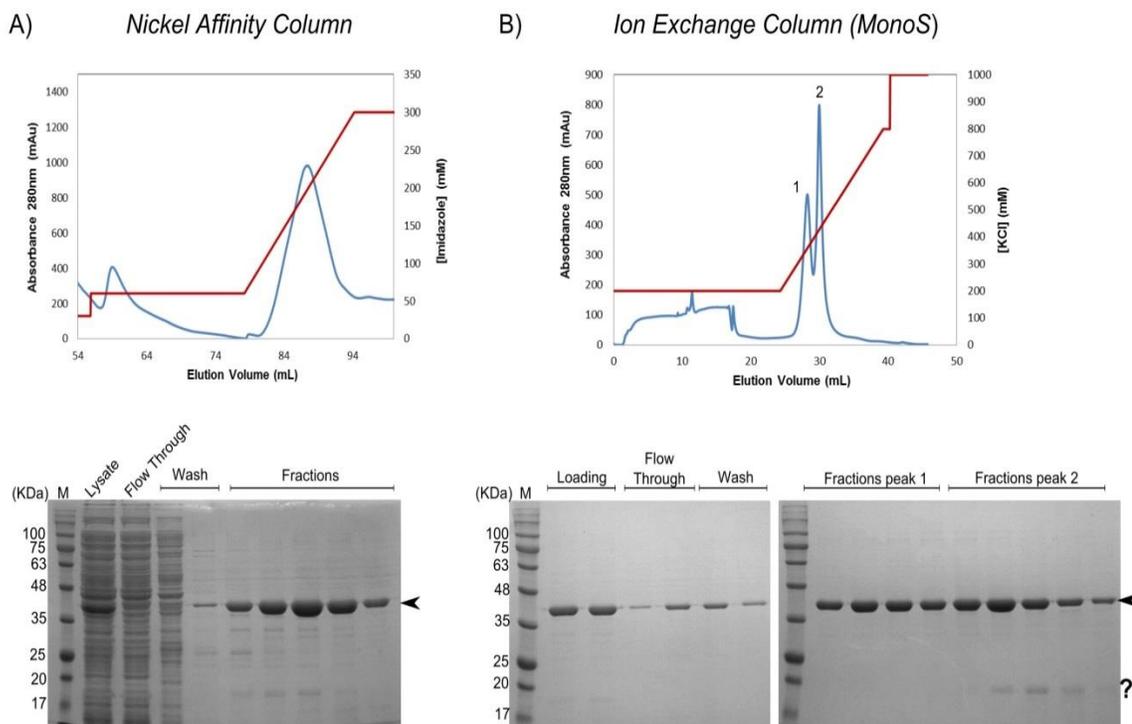


Figure 3.3: Purification of L30 fusion. (A) Nickel Affinity Chromatography profile and 12% SDS-PAG showing (from left to right): molecular weight marker (M), lysate loaded into the column (lysate), proteins that flow through the column (flow through), proteins from two washes (washes), and fractions under the apex of the peak (fractions). (B) Elution profile of fusion protein from MonoS column and 12% SDS-PAG showing (from left to right): molecular weight marker (M), pooled fractions from the nickel column, before and after filter (loading), portion of protein that flowed through (flow through), protein after wash (wash), protein in peak 1 (fractions peak 1) and peak 2 (fractions peak 2). Blue line represents the absorbance at 280 nm (mAu), red line represents either (A) the concentration of imidazole (mM) or (B) the concentration of KCl (mM).

3.2.2) The length of the linker affects the stability of the fusion

Although the L30 fusion presented moderate stability, there were still two main issues: loss of approximately 50% of the protein during the concentration step and the presence of a possible contaminant around 20 kDa that co-eluted on the ion exchange column (Figure 3.3B). There are two possibilities related to the identity of this contaminant. According to the approximate molecular weight, this could be a degradation product of the fusion, possibly the SKN1 domain that co-eluted during the nickel affinity chromatography due to the presence of the Histidine tag in the N-terminal portion of the domain. If this is the case, the contaminant may decrease the activity of the fusion protein by sequestering the DNA. On the other hand, this band could be a contaminant per se. We favor the latter because its concentration did not increase after thawing the sample as would be expected if it was generated by residual proteases present in our preparations.

The instability and possible contaminant in this sample led us to develop a new construct with better behavior in solution. Following the same methodology as with the shorter linker fusions, two fusions were generated with 28 and 32 amino acids (referred to as L28 and L32 fusions, respectively). The level of expression and solubility was tested for both of them and showed a good ratio of expressed over soluble protein when using BL21 (DE3) cell line at 25°C for 5 h.

During the purification step, the L28 fusion precipitated at a high imidazole concentration after being eluted from the nickel affinity column. Therefore, this fusion was discarded. On the other hand, the fusion with the 32 amino acid linker could be readily purified (Figure 3.4). This construct was more stable than the L30 fusion during

concentration, and the 20 kDa contaminant was less prominent (less than 5% of the total sample as judged by the quantification of the gel using ImageJ, Figure 3.4C).

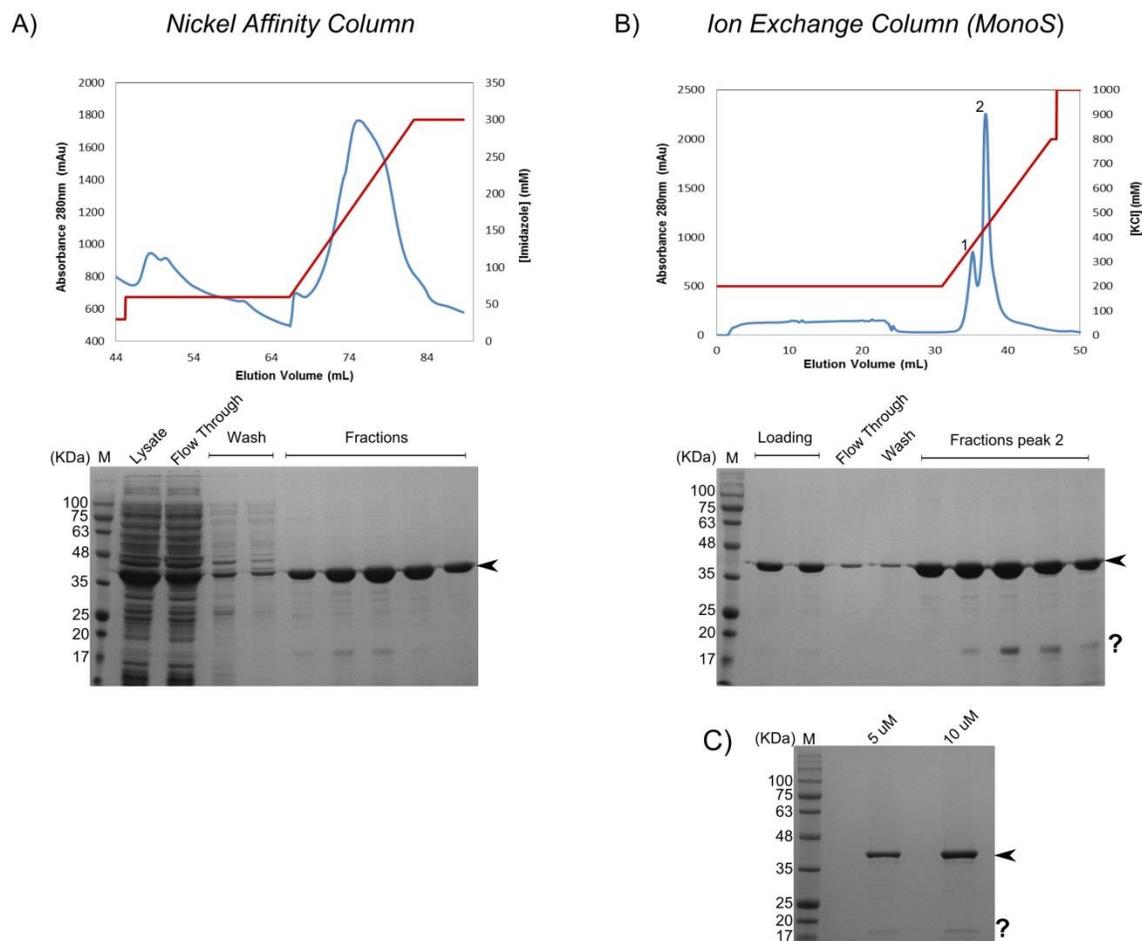


Figure 3.4: Purification of L32 fusion. (A) Nickel Affinity Chromatography profile and 12% SDS-PAG showing (from left to right): molecular weight marker (M), lysate loaded into the column (lysate), proteins that flow through the column (flow through), proteins from two washes (washes), and fractions under the apex of the peak (fractions). (B) Elution profile of fusion protein from MonoS column and 12% SDS-PAG showing (from left to right): molecular weight marker (M), pooled fractions from the nickel column, before and after filter (loading); portion of protein that flowed through (flow through), protein after wash (wash), and protein in peak 2 (fractions peak 2). (C) SDS-PAG showing the L32 fusion (5, 10 μ M) after the concentration step. Blue line represents the absorbance at 280 nm (mAu), red line represents either (A) the concentration of imidazole (mM) or (B) the concentration of KCl (mM).

3.2.3) Functional studies of L32 fusion in complex with linear 195 bp substrates

The first step in the biochemical characterization of the L32 fusion was the selection of the suitable DNA substrate. As shown before (Pillon et al., 2015), when using a 200 bp linear substrate it was possible to observe endonuclease activity of the CTD of MutL in the presence of stoichiometric amounts of the β -clamp. Based on this study, I generated substrates (187 bp, 263 bp and 195 bp) through PCR amplification of different fragments of pUC19 containing an SKN1 site (GTCAT) in the middle of the sequence. From here, the 195 bp substrate was chosen because it resulted in the highest amount of the specific fragment and was similar in length to the substrate used by Pillon et al. (2015).

The main difference in my methodology compared to the assays done by Dr. Pillon and colleagues was the use of fluorescent primers labeled at 5'-end with 6-carboxyfluorescein (6-FAM) instead of radioisotope γ -³²P-ATP. Although the radiolabeled substrates usually give higher sensitivity without introducing artificial structures that might affect binding (Hellman and Fried, 2009), I used fluorescently labeled substrates to avoid diffusion of the bands in the exposure step, labeling reactions, short half-lives, and inherent safety issues related to the use of radioisotopes.

The fluorescent molecule attached to the primers is one of the most commonly used fluorescent dyes because of its low cost and compatibility with most detection instruments. Moreover, it is stable under high temperatures, making it ideal for generating substrates through PCR. Following the generation of the 195 bp substrate, functional studies were performed with the L32 fusion to establish the best conditions for the MutL-DNA interaction.

3.2.3.1) L32 fusion binds DNA specifically

EMSA was performed to detect the formation of the protein-DNA complex in a rapid and sensitive manner (Hellman and Fried, 2009). Specific binding of the L32 fusion to the 195 bp fluorescently labeled substrate with the SKN1 site in the middle (+SKN1M) was detected, even at a 1:1 ratio of protein to DNA (see change in mobility at 10 nM in Figure 3.5A). The specific complex persisted even with the addition of polydIdC which was used as competitor DNA to disrupt non-specific interactions that occur in the assay. Even at a relatively high concentration of polydIdC (20.8 ng/ μ L), the band representing this specific binding mode remained (Figure 3.5B).

When comparing DNA binding of the L32 fusion and the CTD of MutL, it is clear that the latter could not form a complex with the substrate (Figure 3.5C). This shows that the L32 fusion is bypassing the DNA binding defect of the CTD of MutL, and therefore seemed to have potential to facilitate our understanding of the endonuclease mechanism of MutL. I then looked to characterize the endonuclease activity of the L32 fusion and the effect of the SKN1 domain on this activity.

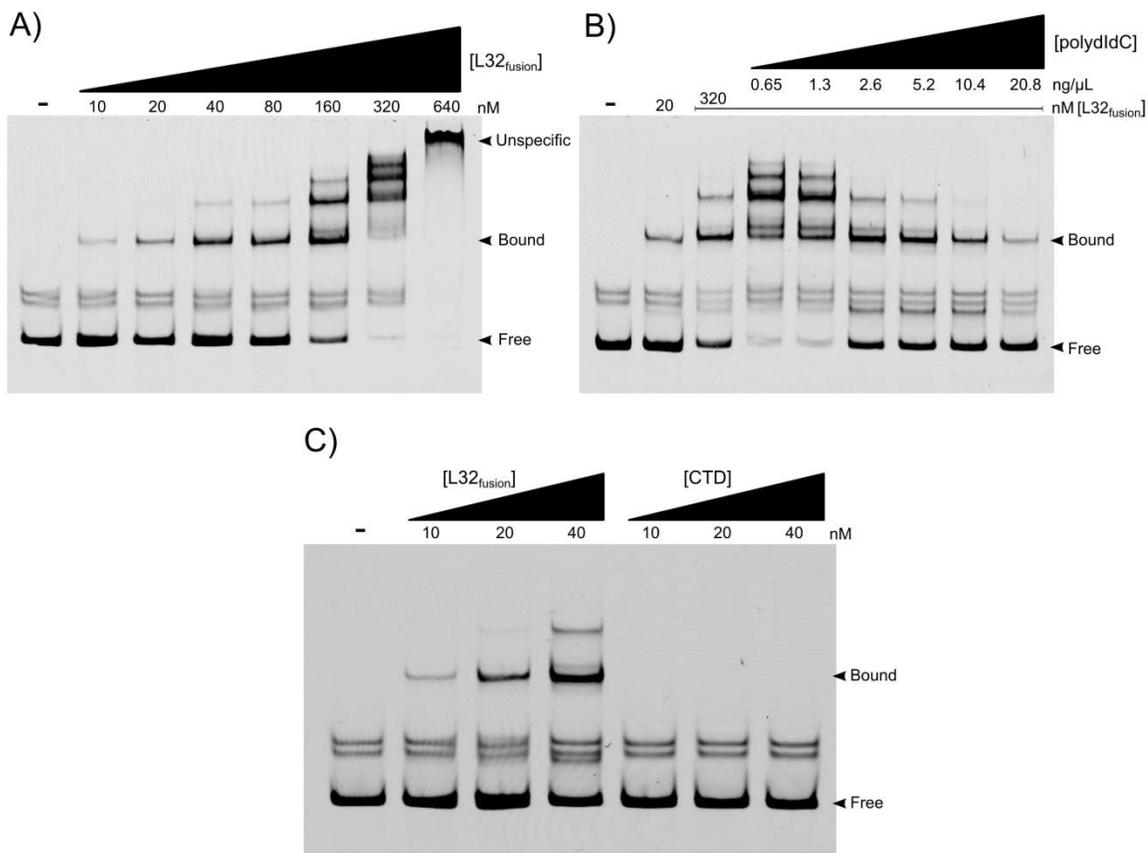


Figure 3.5: Electrophoretic mobility shift assays of L32 fusion with 195 bp substrate with the SKN1 site in the middle. (A) L32 fusion (10-640 nM) was incubated with 10 nM fluorescent labeled substrate (B) PolydIdC competitor (0.65-20.8 ng/μL) was incubated with 10 nM fluorescent substrate and 320 nM of L32 fusion. (C) L32 fusion and CTD of MutL (10, 20 and 40 nM) were incubated with 10 nM fluorescent substrate.

3.2.3.2) L32 fusion has β-dependent endonuclease activity

Preliminary mismatch independent endonuclease assays were performed by increasing the concentration of the L32 fusion from 1 μM to 7.5 μM in the presence of equal amounts of the β-sliding clamp and 10 nM of +SKN1M substrate. In these assays, only the lowest concentrations of the L32 fusion-β-clamp showed nuclease activity (1, 1.5, 2 and 2.5 μM). After 3 μM, there was inhibition of the endonuclease activity which

was most likely due to the increasing volumes of high-salt protein storage buffer used in the reactions. This was expected because DNA affinity decreased in higher ionic strength environments (Kadyrov et al., 2006).

The endonuclease assay of the L32 fusion from 1 μM to 1.3 μM in the presence of the processivity clamp showed no inhibition of the nicking activity (Figure 3.6A). This result was consistent with the findings presented by Pillon et al. (2015) where the activity of the CTD of MutL was stimulated by its interaction with the processivity factor at the same range of concentrations.

MutL endonuclease activity was validated using two L32 variants: aspartate-462 to asparagine (D462N) and glutamate-468 to lysine (E468K). These two residues are located in the endonuclease site, and mutations of them abrogate the nicking function in vitro and the MMR response in vivo (Pillon et al., 2010). The lack of endonuclease activity for both variants denoted that the activity of the L32 fusion was specific to the endonuclease sites present within the protein (Figure 3.6B-C).

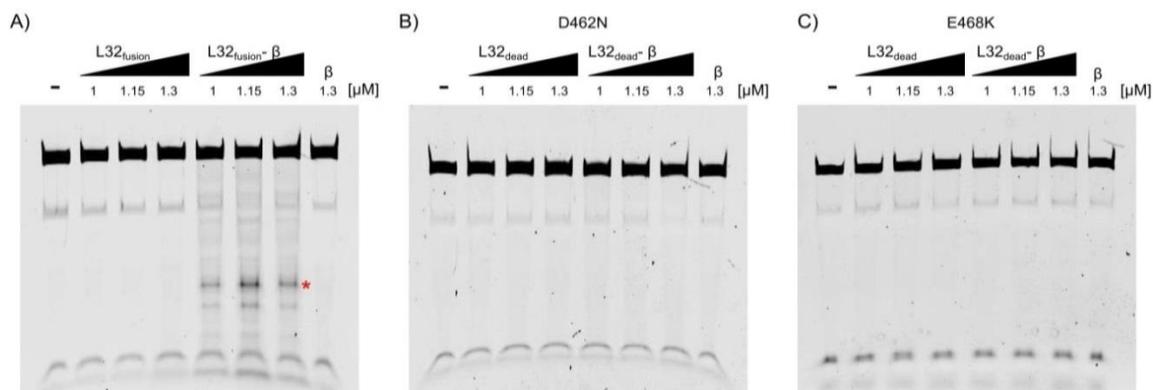


Figure 3.6: Endonuclease activity of L32 fusion and variants with 195 bp linear substrate with SKN1 site in the middle. (A) L32 fusion, (B) D462N variant and (C) E468K variant (1, 1.15, 1.3 μM) were incubated with a 195 bp fluorescent labeled DNA substrate (10 nM) in the absence (L32 or L32_{dead}) and presence (L32- β , L32_{dead}- β) of equimolar amounts of β -clamp.

3.2.3.3) L32 fusion has less β -dependent nicking activity than the CTD of MutL

The level of endonuclease activity of the L32 fusion was compared to the CTD of MutL (Figure 3.7). First, the endonuclease activities of both proteins were tested with a substrate that did not contain any SKN1 site. Since this DNA substrate should not bind to the SKN1 domain, the same level of endonuclease activity was predicted. This assumption was not true; in fact, the nicking activity of the L32 fusion was significantly lower than the CTD when both proteins were interacting with the β -clamp (Figure 3.7A). We believed this phenomenon occurred because the two extra SKN1 domains of the L32 fusion imposed a steric hindrance effect that decreased the interaction with the processivity clamp. Hence, the spatial organization of the L32 fusion was preventing the transient interaction of MutL with the β -sliding clamp, resulting in a lower endonuclease activity.

When the proteins mentioned above interacted with the +SKN1M substrate, the L32 fusion still had less activity than the CTD of MutL (Figure 3.7B). Although it was expected that the presence of the SKN1 domains would have a stimulatory effect by anchoring the DNA to a single position, the opposite happened. This absence of stimulatory effect was attributed to the DNA binding capability of the SKN1 domain in the fusion. Since this domain has a high DNA binding affinity, the sequestration of the nucleic acid away from the endonuclease site may be possible once the DNA is bound to it. The long nature of the linker connecting both domains (32 amino acids) supported this

statement because the DNA would be anchored at the SKN1 portion, being unable to reach the endonuclease site in an efficient way as the initial hypothesis suggested.

These two reasons indicate that the β -clamp cannot tether the DNA onto the endonuclease domain of the L32 fusion to the same extent as with the CTD alone. Therefore, this fusion was not suitable for studying the interaction of the MutL-DNA complex and a different approach was necessary.

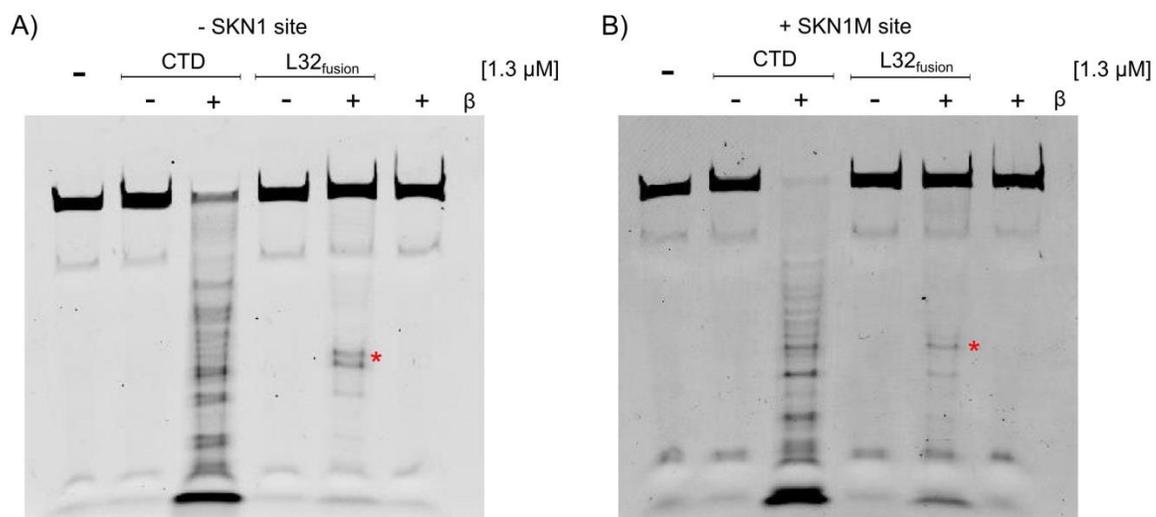


Figure 3.7: Endonuclease activity of CTD and L32 fusion with 195 bp linear substrates. Proteins (1.3 μ M) were incubated with 10 nM 195 bp fluorescent labeled DNA substrate (A) without the SKN1 site (-SKN1) and (B) with the SKN1 site (+SKN1M) in the absence and presence of equimolar amounts of β -clamp (β).

3.2.4) Characterization of SKN1-MutL fusion variant with shorter linker

Although fusion proteins with shorter linker lengths were not stable, they seemed to be the only option to anchor the DNA in close proximity to the endonuclease site. We hypothesized that the presence of two SKN1 domains, which are highly positively charged (with a theoretical isoelectric point of 10), on the N-terminus of each monomeric subunit of the CTD dimer would cause an electrostatic repulsion phenomenon. This

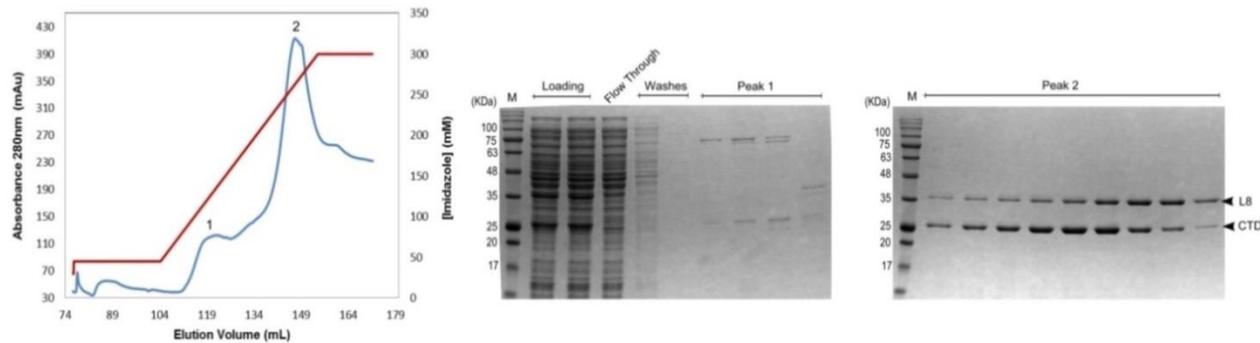
would explain why all the fusion proteins with shorter linker lengths were insoluble or unstable. We believed the SKN1 domains require enough separation to not interfere with the stability of the dimer which, in our case, required a linker length of at least 30 amino acids.

Based on this analysis, the option of eliminating one of the SKN1 domains in the dimer could solve the stability issue and, in that way, fusions with shorter linkers could be possible. An SKN1-MutL_{CTD} heterodimer comprising the CTD of MutL and only one copy of the SKN1 domain was then designed.

The characterization of this heterodimer (herein referred to as L8*) was done with another graduate student in the lab (Linda Liu). This was based on the co-expression of both the homodimer of the CTD of MutL with the homodimer fusion of the shortest linker (L8 fusion). Surprisingly, when expressing both plasmids, some portion of the L8 fusion became soluble. This was assigned to the formation of the heterodimer fusion between the two proteins in a small fraction of the cells.

At the cell lysis step during the purification protocol, the L8 fusion precipitated out of solution. The remaining soluble fraction was purified by nickel affinity chromatography and ion exchange chromatography (S-Sepharose) (Figure 3.8). Based on the charge of the second column, the CTD (which is negatively charged) did not bind to the column, and passed through it as a secondary product of this purification. During the KCl gradient, three peaks were observed: the first one corresponded to the heterodimer L8* at 270 mM KCl, the second and third peaks were two isoforms of the L8 fusion (Figure 3.8B). Although the third peak had the highest absorbance, this form was not stable in solution.

A) Nickel Affinity Chromatography



B) Ion Exchange Chromatography (S-sepharose)

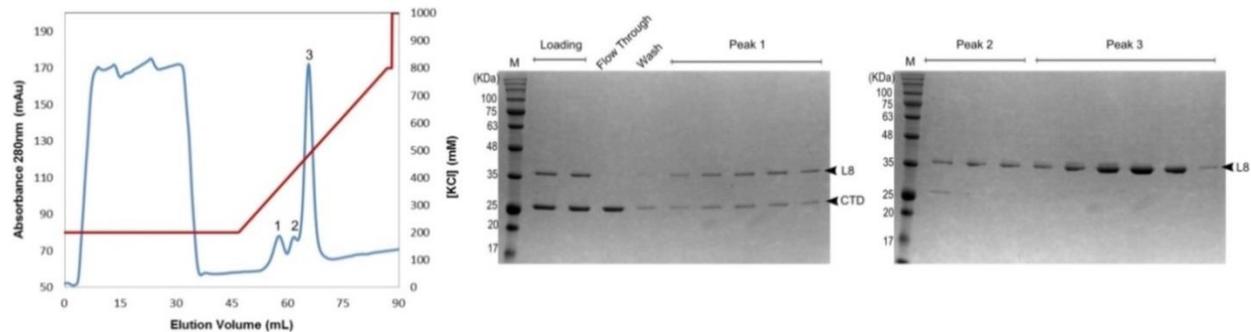


Figure 3.8: Purification of L8*. (A) Nickel Affinity Chromatography profile and 12% SDS-PAGs showing the fractions of proteins from the nickel column. The first gel shows (from left to right): molecular weight marker (M), lysate before and after filtered (loading), proteins that flowed through the column (flow through), two washes (washes), fractions from the peak 1 (peak 1). The second gel shows the molecular weight marker (M), and the fractions under the second peak (peak 2). (B) Ion exchange chromatography profile and 12 % SDS-PAGs showing (from left to right): molecular weight markers (M), pool fractions from nickel, before and after (loading), protein from the flow through (flow through), wash with 200 mM KCl (wash), and fractions under the three peaks (Peak 1, 2, and 3). Blue line represents the absorbance at 280 nm (mAu), and red line represents either (A) the concentration of imidazole (mM) or (B) the concentration of KCl (mM).

3.2.5) Functional studies of the heterodimer fusion L8* in complex with linear 195 bp DNA substrates

Even though the purification protocol of L8* did not have a high yield, the amount of protein was enough to perform biochemical assays. Consistent with previous results, the L8* had endonuclease activity in the presence of the processivity clamp (Figure 3.9).

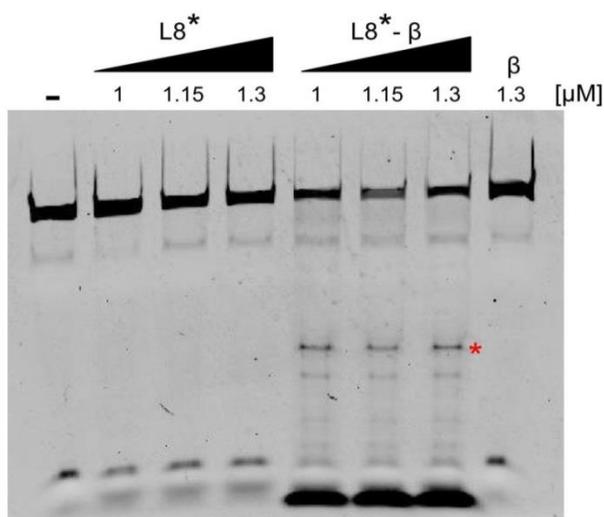


Figure 3.9: Endonuclease activity of L8* using 195 bp linear substrate with the SKNI site in the middle. L8* (1, 1.15 and 1.3 μM) was incubated with DNA substrate (10 nM) in the absence (L8*) and presence of equimolar amounts of β -clamp (L8*- β).

3.2.5.1) L8* has more β -dependent endonuclease activity than the L32 fusion and the CTD of MutL.

To evaluate the level of nicking activity and the difference in patterns between the CTD, L8*, and the L32 fusion, an endonuclease assay was performed at 1.30 μM for all proteins (Figure 3.10A). Based on the intensity of the cleavage products, it was initially thought that the endonuclease activity in the L8* was lower. However, after analyzing

this gel further and comparing it to previous assays, the endonuclease activity of the L8* was actually higher.

The first observation supporting this conclusion was that the intensity of the DNA substrate of the L8*- β sample was significantly lower than the L32 fusion- β reaction. Second, there were short size DNA fragments that accumulated at the bottom of the gel (<32 bp). We therefore conclude that the amount of the initial substrate was decreasing because the nicking activity of the L8* was much higher, leading to the accumulation of a high amount of short DNA fragments.

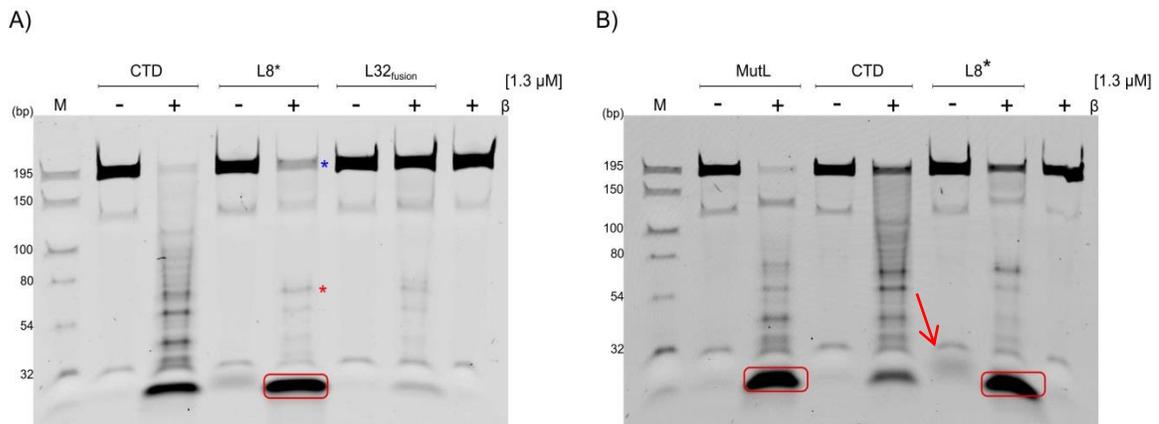


Figure 3.10: Endonuclease activity of MutL full length, CTD, and fusion proteins using 195bp linear substrate with SKN1 site in the middle. Proteins (1.3 μ M) were incubated with DNA substrate (10 nM) in the absence and presence of equimolar amounts of β -clamp (β). (A) Endonuclease activity of CTD, L8*, and L32 fusion. (B) Endonuclease activity of MutL, CTD, and L8*. Arrow indicates the accumulation of short DNA fragments for the L8* in the absence of β -clamp.

Furthermore, when comparing the nicking function of the L8*, CTD against the full length MutL protein, it was evident that full length protein had a different pattern of cleavage products than the other two proteins (Figure 3.10B). While the endonuclease activities of the full length MutL and L8* were higher because they had contributions

from the DNA binding and interaction with the β -clamp, the activity of the CTD produced the least amount of products due to its sole interaction with the processivity clamp. It was not possible to conclude if the activity of the L8* was higher than the full length MutL because of the differences in the pattern of cleavage products.

We confirmed that the endonuclease activity of the L8* was higher than the CTD after obtaining a similar pattern of CTD-degradation products with five times less L8*- β complex (Figure 3.11A). Therefore, the presence of the SKN1 domain in this fusion had a stimulatory effect on the endonucleolytic activity as our initial hypothesis suggested.

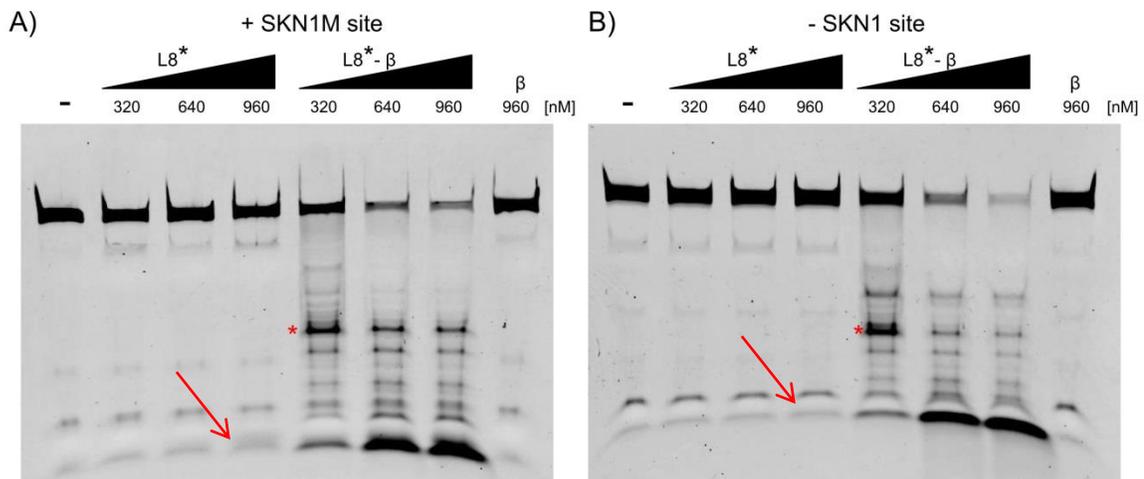


Figure 3.11: Endonuclease activity of L8* with 195 bp linear substrates. L8* (320, 640 and 960 nM) was incubated with 195 bp fluorescent labeled DNA (10 nM) with (A) the SKN1 site in the middle (+SKN1M) or (B) without the SKN1 site (-SKN1) in the absence and presence of equimolar amounts of β -clamp (β). Arrows indicate the accumulation of short DNA fragments with the +SKN1M substrate or lack thereof (-SKN1 substrate) in the absence of β -clamp.

Additionally, it is worth noting that the L8*- β with the -SKN1 substrate (Figure 3.11B) also had endonuclease activity at the lowest concentration of the interaction between L8* and β -clamp. This was explained by the nonspecific binding of the reaction

that could not be avoided at this range of concentrations. At the lowest concentration at which the endonuclease activity of L8* with the β -clamp can be reported (240 nM), there was unspecific binding of the SKN1 domain with the DNA.

As was shown by the DNA binding assays (Figure 3.5), there is specific binding only at the lowest concentrations of proteins (10-80 nM), and the range at which the nicking was observed is much higher than the level of the specific binding to the SKN1 domain. This phenomenon could not be corrected because the transient interaction between MutL and the processivity clamp requires a certain amount of both proteins to ensure the complex formation and to report the endonuclease activity (Pillon et al., 2015).

Interestingly, when using the +SKN1M substrate, there was also a slight accumulation of short DNA fragments for the L8* even in the absence of the processivity clamp (960 nM). The fact that this accumulation was not observed using -SKN1 substrate, confirmed the stimulatory effect of the specific DNA-binding domain on the nicking activity of MutL.

On the other hand, a set of variants of the L8* were generated to validate that the endonuclease activity of this fusion was responsible for the accumulation of short DNA fragments. These variants were named based on the endonuclease site that was mutated: mutation on the CTD portion (E468K^{CTD}), mutation on the L8* portion (E468K^{L8*}), or both sites mutated (E468K or D462N). As expected, mutating both active sites of the dimer caused a disruption of the endonuclease activity (Figure 3.12A). Subsequently, when mutating one monomer at a time (E468K^{CTD}, E468K^{L8*} (Figure 3.12B)), nicking

function was detected for both cases. This suggests that the presence of at least one functional active site is enough to cleave a given DNA substrate.

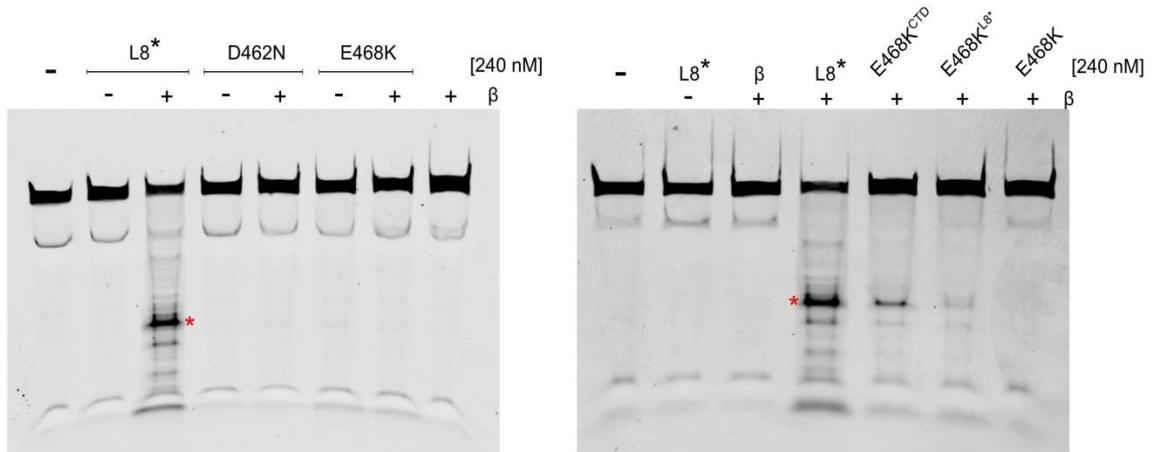


Figure 3.12: Endonuclease activity of L8* and variants using 195 bp linear substrate with SKN1 site in the middle. Proteins (240 nM) were incubated with fluorescently labeled DNA (10 nM) in the absence or presence of the β -clamp (β). (A) L8* and L8* variants D462N and E468K. (B) L8*, and L8* variants E468K^{CTD}, E468K^{L8*} and E468K.

3.2.5.2) Effect of the position of the SKN1 site and orientation imposed by β -clamp on the endonuclease activity of MutL

We next tested whether the position of the SKN1 site on the DNA substrate had an effect on the endonuclease activity of the L8*. Since this site is not palindromic, it was not known how it could affect the interaction with the endonuclease activity of MutL. We hypothesized that depending on where the SKN1 site was located, the substrate may have more flexibility to bind to the SKN1 domain, and in that way, it would leave more space to be cleaved by the endonuclease domain. This would result in an additional stimulation of the nicking activity of the fusion protein.

In addition, it is worth noting that in all endonuclease assays there was one predominant product of approximately 70 bp (Figures 3.9-3.12). The accumulation of this particular length of DNA may be related to the way the β -clamp orients the DNA towards the endonuclease site.

To investigate the effect of the position of the SKN1 site and if the β -clamp favors nicking one strand at a time, I performed endonuclease assays by labeling each strand of the duplex individually. These included the top strand (*), the bottom (*) strand, and both strands (double (**)) of the duplex for the substrates with the SKN1 site in the beginning (+SKN1B, position 33-37 bp), middle (+SKN1M, position 88-92 bp), and end (+SKN1E, 159-162 bp) of the 195 bp sequence (Figure 3.13).

The first assay showed the same pattern of products for both the top single and double labeled strands for all substrates (Figure 3.13A). Since the single labeled substrates only had one strand of fluorescent molecules, the intensity appeared lower in the gel. Moreover, the major product at about 70 bp was still present in all cases, and there were more unspecific products above and below this main product for the double labeled substrates. When comparing the single bottom and the double labeled strands, a different pattern emerged. The 70 bp product was still the major output of the assay, but the single bottom strand had fewer degradation products than the double or single top labeled strands in all substrates (Figure 3.13B). In these assays, there was no difference between any of the single or double labeled substrates regarding the position of the SKN1 site.

The accumulation of the 70 bp product in all of the substrates suggests that MutL can only nick one strand at a time. This idea was reinforced by the fact that when both strands

are fluorescently labeled, the 70 bp product is more prominent than its complementary sequence at 125 bp (Figure 3.13 double labeled (**)). If both strands were cut simultaneously, the intensity of the products at 70 bp and 125 bp would have been the same. Further when labeling one strand (either top or bottom); there is accumulation of the product at 70 bp, while the complement is never seen. Because the 125 bp product is either at lower concentrations or absent, this proves that the β -clamp orients MutL to nick only one strand.

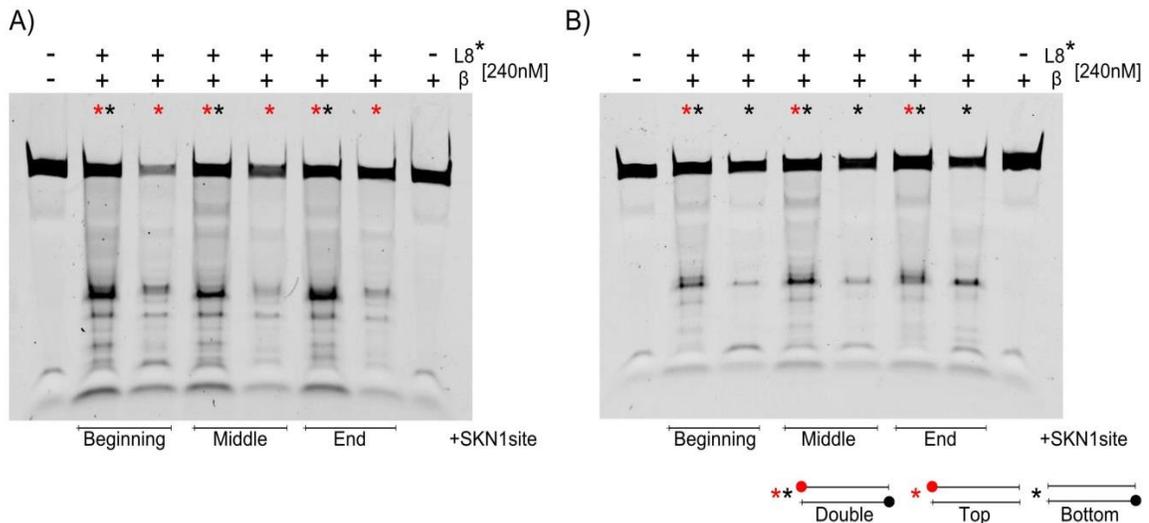


Figure 3.13: Endonuclease activity of $L8^*$ using 195 bp linear substrates single or double labeled. $L8^*$ (240 nM) and equal amount of β -clamp (β) were incubated with substrates (10 nM) with the SKN1 site in the beginning, middle and end of the sequence. (A) Double (**) and single top (*) labeled substrates. (B) Double (**) and single bottom (*) labeled substrates.

3.2.5.3) Pre-nicked substrates do not stimulate the β -dependent endonuclease activity of MutL

As was explained by Kadyrov et al. (2006), MutL endonuclease activity is activated in the presence of a mismatch, MutSa, RFC, PCNA, ATP, and a pre-existing nick in one

heteroduplex strand. From here, the MutL incision is directed to the strand that contains the pre-existing break. Based on these findings, we analyzed whether or not a pre-nicked substrate could enhance the nicking activity towards one specific strand on a determined substrate.

A nicking enzyme was selected from the pool of commercial enzymes from NEB. This search was limited to site specific endonucleases that hydrolyze only one strand of a DNA duplex. After a comprehensive comparison between the available nicking enzymes and the 195 bp DNA substrate, one enzyme, Nt.AlwI, fulfilled the requirements to form nicked substrate. This enzyme is a derivative form of the restriction enzyme AlwI, and it has been engineered to catalyze a single strand break four bases beyond the 3'-end of the recognition sequence on the top strand of duplex DNA. Its recognition sequence ($5'$ GGATCNNNN \uparrow N $3'$) is part of the 195 bp substrate at the 40 bp position. According to this, pre-nicking the DNA would generate a product of approximately 48 bp when using the double labeled or single top strand substrates. For the single bottom strand substrates, Nt.AlwI would not produce any observable nicked products because the enzyme only cuts at the top strand that, in this case, had not been fluorescent labeled. The concentration of the Nt.AlwI enzyme was optimized through a titration from 10 units/ μ L to 0.039 units/ μ L. The concentration at 0.039 units/ μ L was the optimal amount to avoid over-nicking of the reaction at 37°C for 1 hour.

The endonuclease assay of the L8* in the absence and presence of the β -clamp for the differently labeled +SKN1M substrates showed that the intensity of the pre-nicked product (see blue asterisk in Figure 3.14) did not increase, and nor new degradation

products were formed (Figure 3.14A). Therefore, the pre-existing nick did not stimulate the nicking activity of the L8* either by itself or in the presence of the processivity clamp.

Finally, when comparing the absence or presence of the nicking enzyme with the processivity clamp in all reactions, the intensity of the pre-nicked sequence did not increase either, and the same types of products were generated for both cases (Figure 3.14B). The predominant sequence was still the one at 70 bp for the differently labeled DNA sequences.

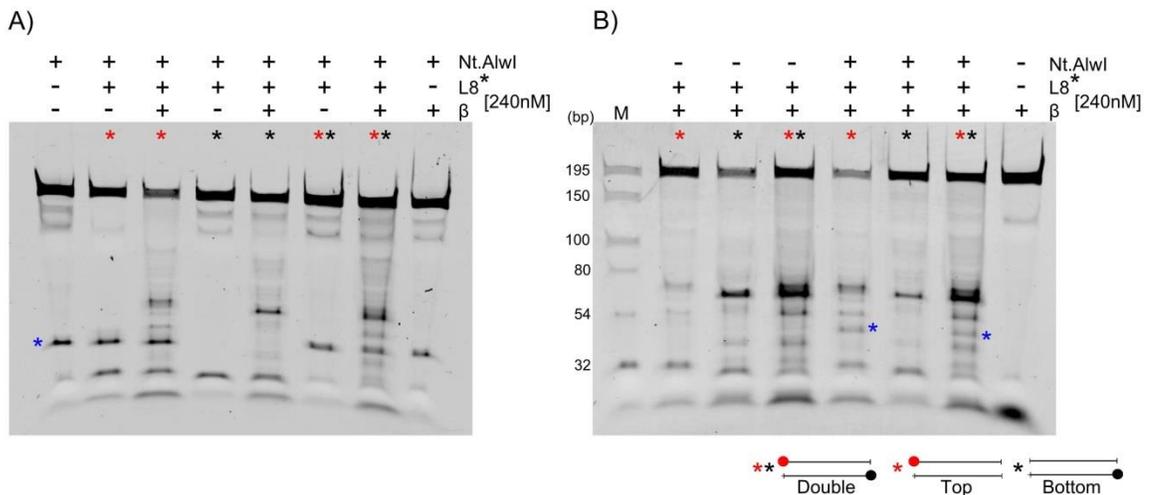


Figure 3.14: Effect of the nicking enzyme *Nt.AlwI* on the endonuclease activity of L8*. (A) L8* (240 nM) was incubated with the nicking enzyme and the absence or presence of equal amounts of β -clamp (β) for the single top (*), single bottom (*), and double (***) substrates which had the SKN1 site in middle of the sequence. (B) L8* and β -clamp were incubated in the absence or presence of the *Nt.AlwI* nicking enzyme with the substrates mentioned in panel A.

3.3) DISCUSSION

In this research, I focused on the endonuclease activity of the MutL protein from *Bacillus subtilis*. Despite the overall low sequence similarity with eukaryotes, structural

information of the CTD of *B. subtilis* MutL (Pillon et al., 2010) revealed the presence of an identical endonuclease site and the same ion requirements as in yeast and human MutL α (Kadyrov et al., 2006; Gueneau et al., 2013). Although the endonuclease site has been described (Pillon et al., 2010; Gueneau et al., 2013), the molecular mechanism of how MutL nicks the DNA remains unclear. This is because the CTD of MutL does not bind to DNA, and therefore, the direct contacts between this portion of the protein and the nucleic acid have not yet been established.

By bypassing the DNA binding defect of the CTD of MutL, we stabilized the transient interaction between this domain and a given DNA substrate. Our approach replaced the N-terminal domain of MutL, which coordinates DNA binding in a sequence unspecific and length-dependent manner, for a domain with sequence specificity and high DNA-binding affinity. We hypothesized that anchoring the DNA to a single position through the DNA binding portion would lead to the stimulation of the nicking activity of MutL.

An important characteristic in the design of the fusion protein was to join the two domains together without affecting their respective activities (Chen et al., 2013). To this end, we biochemically characterized the functions of each domain and the possible cooperation between them using linear DNA substrates.

One limitation of the endonuclease assays was the use of the processivity clamp to report the nicking activity. Even though the interaction of MutL with the β -clamp is weak (Pillon et al., 2015; Pillon et al., 2011; Pluciennik et al., 2013; Pluciennik et al., 2010), this interaction drove the endonuclease activity of MutL. As only minor differences were

observed in the accumulation of short DNA fragments when the β -clamp was not present (see red arrows in Figure 3.10, 3.11), the major effects of the DNA binding domain on endonuclease activity could not be established. These differences were not enough to characterize the nicking activity of the fusion protein and DNA on its own.

Another limitation of the endonuclease assays was that at the concentration range where nicking activity was observable, there were contributions from both specific and unspecific DNA binding. The unspecific binding could not be removed because the interaction of MutL with the β -clamp required a certain amount of both proteins (at least 240 nM) to obtain MutL incision. At this concentration, EMSAs showed that SKN1 had extensive non-specific DNA-binding activity. This feature made it impossible to establish differences between substrates with or without the DNA binding site.

One of our findings from the endonuclease assays was that either active site in the dimer can nick the DNA when the processivity clamp is present in the reaction. We also verified that prokaryotic MutL only needs one functional site to nick a determined substrate, despite having the potential to utilize two. This behaviour is similar to eukaryotic MutL homologs (MLH1-PMS2 or MLH1-PMS1), which use (and possess) only one active site (Kadyrov et al., 2006; Kosinski et al., 2008; Gueneau et al., 2013).

There are several factors to consider related to the orientation the β -clamp may impose on the nicking function. First, since the endonuclease assays were clamp loader independent, the β -clamp was not loaded properly onto the DNA at each end of the sequence. We believe that it is instead sliding along the nucleic acid for a certain number of nucleotides until it encounters the fusion protein bound to the DNA. At this point, the

β -clamp pauses and interacts with the CTD of MutL, causing the activation of the endonuclease activity. Given the pattern of cleavage products, this hypothesis holds true for the substrates with the DNA binding site at the beginning or end of the sequence where the β -clamp slides from any orientation. However, it does not explain why the endonuclease assays on the substrate with a central SKN1 site still gave similar types of cleavage of products.

Since the 70 bp degradation product accumulated regardless of the position and presence of the SKN1 site, it seems that, for this product, the participation of the DNA binding domain was insignificant. Conversely, we believe the DNA binding domain may have a more prominent role in the accumulation of short DNA fragments. Indeed, the lower concentration used in the heterodimer fusion to obtain endonuclease activity confirmed this statement. In other words, the unspecific and short DNA degradation products depend on the stimulation of the DNA binding domain, but the specific one does not.

Our results support the idea that the processivity clamp orients MutL in such a way that its intrinsic endonuclease activity preferentially nicks one strand at a time. This is relevant in the MMR context because the nicked strand must be the nascent strand generated during replication. Although the strand-specific orientation could not be determined because the endonuclease assays were mismatch-, MutS, ATP, and clamp-loader independent, our findings were consistent with the ones presented by Pluciennik et al. (2010).

Additionally, the presence of a pre-existing nick did not stimulate the nicking activity of MutL using linear substrates. The previous study done by Pluciennik et al. (2010) demonstrated that the initial nick serves as a loading site for the processivity clamp and not for the endonuclease activation per se. This supports our results because the β -clamp was not being loaded to one specific strand, and the discontinuity therefore did not have any effect on the nicking activation. Consequently, our findings reinforce the premise that the function of the pre-existing nick is to provide an entry point for loading of the β -clamp and not for MutL action.

In all, the design of the fusion protein composed of the C-terminal domain of MutL and a specific, high-affinity DNA binding protein has allowed us to overcome the DNA-binding defect of the MutL endonuclease domain. Using this fusion, we have established a system for characterizing the binding and nicking of 195 bp linear substrates at much more effective concentrations than have been previously reported. This work leads to more questions for future research related to the mechanistic aspects of the strand discrimination step in the MMR process. These include: Which endonuclease site does the β -clamp recognize (the proximal or the distal from the β -binding motif)? What is the minimum length of the DNA substrate to obtain endonuclease activity? Does the structure of the substrate (resected ends compared to blunt ends) have an effect on the way the β -clamp orients MutL incision? By combining our results with the answers to these questions, we will set the basis of the mechanistic characterization of MutL endonuclease in organisms that use the nick-dependent mismatch repair process.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

We have shown that ATP promotes a series of conformational changes in *B. subtilis* MutL that are reminiscent of the nucleotide-induced changes in eukaryotic MutL homologs. Therefore, my results suggest that *B. subtilis* MutL is more similar to eukaryotic MutL α homologs, not only in the presence of the endonuclease domain, but also in the behavior of the full length protein upon ATP binding. We believe that, in organisms that use the nick-directed MMR process, the conformational changes alter protein contacts that are important for the subsequent endonuclease cleavage in the repair reaction. Further studies of the ATP-dependent conformational changes of another endonuclease active MutL homolog will reinforce the hypothesis that the endonuclease activity confers new characteristics to the conformations of the MutL protein.

We also stabilized a transient protein-DNA complex between the C-terminal domain of MutL and DNA. This work is important because it extends beyond the mismatch repair field, showing how a fusion protein may be used to study other transitory protein-DNA complexes. The biochemical characterization of the fusion proteins with linear 195 bp substrates allowed us to answer several questions related to the molecular mechanism of the endonuclease site of MutL. We concluded that the position of the DNA binding site does not affect the β -dependent endonuclease activity, that homodimeric MutL requires at least one functional active site to nick the DNA, and that a pre-existing nick

does not stimulate the endonuclease activity on linear substrates. Although we bypassed the DNA binding defect of the C-terminal domain, we did not find a major contribution to the activation of the MutL endonuclease.

The long term goal of this project is to determine the molecular mechanism of MutL-DNA interaction through structural studies. We believe that the spatial requirements of the MutL-DNA complex should be analyzed in the context of the β -sliding clamp, which is one of the major regulators of the nicking activity. Recent advances in combining site-specific cross linking with X-ray crystallography or SAXS (Groothuizen et al., 2015; Pillon et al., 2015), have shown the power of these methods to investigate weak protein interactions. In this sense, we could make use of the characterization of the β -clamp-MutL interaction using cross-linked cysteine variants (Pillon et al., 2015), and then add the DNA to the complex. Similarly, we could use this concept to crosslink the DNA to the β -clamp and then add MutL. Answering how the MutL- β clamp complex positions itself around the DNA and how the catalytic metals are coordinated in relation to the DNA backbone will help develop our understanding of the role of these weak transient interactions in the mismatch repair response. My work has focussed on developing the tools and the approach to overcome this long-standing hurdle in the field of MMR. Although there is much more to be done, we have set the basis for unveiling the fundamental mechanism behind one of the most important anti-cancer processes found in all organisms.

REFERENCES

- Ai, E. (2006). Gel-filtration chromatography. *Nature Methods*, 3, 411–412.
- Alani, E., Sokolsky, T., Studamire, B., Miret, J. J., & Lahue, R. S. (1997). Genetic and biochemical analysis of MSH2p-MSH6p: role of ATP hydrolysis and MSH2p-MSH6p subunit interactions in mismatch base pair recognition. *Molecular and Cellular Biology*, 17(5), 2436–2447.
- Amersham. (2002). Ion Exchange Chromatography & Chromatofocusing.
- Arana, M. E., & Kunkel, T.A. (2010). Mutator phenotypes due to DNA replication infidelity. *Seminars in Cancer Biology*, 20(5), 304–311.
- Arana, M. E., Holmes, S. F., Fortune, J. M., Moon, A. F., Pedersen, L. C., & Kunkel, T. A. (2010). Functional residues on the surface of the N-terminal domain of yeast PMS1. *DNA Repair*, 9(4), 448–457.
- Ban, C., Junop, M., & Yang, W. (1999). Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, 97, 85–97.
- Ban, C., & Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell*, 95, 541–552.
- Bartuma, K., Nilbert, M., & Carlsson, C. (2012). Family perspectives in Lynch Syndrome becoming a family at risk, patterns of communication and influence on relations. *Hereditary Cancer in Clinical Practice*, 10(1), 6-10.
- Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, a, & Forterre, P. (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature*, 386 (6623), 414-417.
- Boivin, S., Kozak, S., & Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification*, 91(2), 192–206.
- Bolz, N. J., Lenhart, J. S., Weindorf, S. C., & Simmons, L. A. (2012). Residues in the N-terminal domain of MutL required for mismatch repair in *Bacillus subtilis*. *Journal of Bacteriology*, 194(19), 5361–5367.

- Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., ... Liskay, R. M. (1994). Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. *Nature*, *368*, 258–261.
- Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S. M., Fischer, F., ... Jiricny, J. (2005). Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Research*, *65*(23), 10759–10766.
- Chen, X., Zaro, J. L., & Shen, W.-C. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, *65*(10), 1357–1369.
- Claverys, J. P., & Lacks, S. A. (1986). Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiological Reviews*, *50*(2), 133–65.
- Correa, E. M. E., De Tullio, L., Velez, P. S., Martina, M. a., Argarana, C. E., & Barra, J. L. (2013). Analysis of DNA structure and sequence requirements for *Pseudomonas aeruginosa* MutL endonuclease activity. *Journal of Biochemistry*, *154*(6), 505–511.
- De la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. *Nature Reviews. Cancer*, *4*, 769–780.
- Duppatla, V., Bodda, C., Urbanke, C., Friedhoff, P., & Rao, D. N. (2009). The C-terminal domain is sufficient for endonuclease activity of *Neisseria gonorrhoeae* MutL. *Biochemical Journal*, *423*, 265–277.
- Edelmann, W., Yang, K., Umar, a, Heyer, J., Lau, K., Fan, K., ... Kucherlapati, R. (1997). Mutation in the mismatch repair gene MSH6 causes cancer susceptibility. *Cell*, *91*, 467–477.
- Elez, M., Radman, M., & Matic, I. (2012). Stoichiometry of MutS and MutL at unrepaired mismatches in vivo suggests a mechanism of repair. *Nucleic Acids Research*, *40*, 3929–3938.
- Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. a., Garber, J., ... Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell*, *75*, 1027–1038.
- Flores-Rozas, H., Clark, D., & Kolodner, R. D. (2000). Proliferating cell nuclear antigen and MSH2p-MSH6p interact to form an active mispair recognition complex. *Nature Genetics*, *26*, 375–378.
- Frick, D. N., & Richardson, C. C. (2001). DNA primases. *Annual Review of Biochemistry*, *70*, 39–80.

- Friedhoff, P., Li, P., & Gotthardt, J. (2016). Protein-protein interactions in DNA mismatch repair. *DNA Repair*, *38*, 50–57.
- Fukui, K., Nishida, M., Nakagawa, N., Masui, R., & Kuramitsu, S. (2008). Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. *Journal of Biological Chemistry*, *283*(18), 12136–45.
- Ghodgaonkar, M. M., Lazzaro, F., Olivera-Pimentel, M., Artola-Borán, M., Cejka, P., Reijns, M. A., ... Jiricny, J. (2013). Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Molecular Cell*, *50*, 323–332.
- Giardiello, F. M., Allen, J. I., Axilbund, J. E., Boland, C. R., Burke, C. a., Burt, R. W., ... Rex, D. K. (2014). Guidelines on genetic evaluation and management of Lynch Syndrome: A consensus statement by the us multi-society task force on colorectal cancer. *Gastroenterology*, *147*(2), 502–526.
- Gorman, J., Wang, F., Redding, S., Plys, A. J., Fazio, T., Wind, S., ... Greene, E. C. (2012). Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(45), E3074–83.
- Groothuizen, F. S., Winkler, I., Cristóvão, M., Fish, A., Winterwerp, H. H. K., Reumer, A., ... Sixma, T. K. (2015). MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *eLife*, *4*, 1–24.
- Guarné, A., Junop, M. S., & Yang, W. (2001). Structure and function of the N-terminal 40 kDa fragment of human PMS2: A monomeric GH1 ATPase. *EMBO Journal*, *20*(19), 5521–5531.
- Guarné, A. (2012). The functions of MutL in mismatch repair: The power of multitasking. *Progress in Molecular Biology and Translational Science*, *110*, 41-70.
- Guarné, A., & Charbonnier, J.-B. (2015). Insights from a decade of biophysical studies on MutL: Roles in strand discrimination and mismatch removal. *Progress in Biophysics and Molecular Biology*, *117*, 149–156.
- Guarné, A., Ramon-Maiques, S., Wolff, E. M., Ghirlando, R., Hu, X., Miller, J. H., & Yang, W. (2004). Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO Journal*, *23*(21), 4134–4145.

- Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., ... Charbonnier, J.-B. (2013). Structure of the MutLa C-terminal domain reveals how MLH1 contributes to PMS1 endonuclease site. *Nature Structural & Molecular Biology*, *20*(4), 461–8.
- Gupta, S., Gellert, M., & Yang, W. (2012). Mechanism of mismatch recognition revealed by human MutS β bound to unpaired DNA loops. *Nature Structural & Molecular Biology*, *19*(1), 72–8.
- Hall, M. C., Shcherbakova, P. V., Fortune, J. M., Borchers, C. H., Dial, J. M., Tomer, K. B., & Kunkel, T. A. (2003). DNA binding by yeast MLH1 and PMS1: Implications for DNA mismatch repair. *Nucleic Acids Research*, *31*(8), 2025–2034.
- Hall, M. C., Shcherbakova, P. V., & Kunkel, T. A. (2002). Differential ATP binding and intrinsic ATP hydrolysis by amino-terminal domains of the yeast MLH1 and PMS1 proteins. *Journal of Biological Chemistry*, *277*(5), 3673–3679.
- Hall, M. C., Wang, H., Erie, D. A., & Kunkel, T. A. (2001). High affinity cooperative DNA binding by the yeast MLH1-PMS1 heterodimer. *Journal of Molecular Biology*, *312*, 637–647.
- Hellman, L. M., & Fried, M. G. (2009). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, *2*(8), 1849–1861.
- Hingorani, M. M. (2016). Mismatch binding, ADP-ATP exchange and intramolecular signaling during mismatch repair. *DNA Repair*, *38*, 24–31.
- Hombauer, H., Campbell, C. S., Smith, C. E., Desai, A., & Kolodner, R. D. (2011). Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell*, *147*(5), 1040–1053.
- Hsieh, P., & Yamane, K. (2008). DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mechanisms of Ageing and Development*, *129*, 391–407.
- Iino, H., Kim, K., Shimada, A., Masui, R., Kuramitsu, S., & Fukui, K. (2011). 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. *Bioscience Reports*, *31*, 309–322.
- Jacobs-Palmer, E., & Hingorani, M. M. (2007). The effects of nucleotides on MutS-DNA binding kinetics clarify the role of MutS ATPase activity in mismatch repair. *Journal of Molecular Biology*, *366*(4), 1087–1098.

- Jeggo, P. A., Pearl, L. H., & Carr, A. M. (2016). DNA repair, genome stability and cancer: a historical perspective. *Nature Reviews. Cancer*, *16*(1), 35–42.
- Jeong, C., Cho, W.-K., Song, K.-M., Cook, C., Yoon, T.-Y., Ban, C., ... Lee, J.B. (2011). MutS switches between two fundamentally distinct clamps during mismatch repair. *Nature Structural & Molecular Biology*, *18*(3), 379–385.
- Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology*, *7*, 335–346.
- Jiricny, J. (2013). Postreplicative mismatch repair. *Cold Spring Harbor Perspectives in Biology*, *5*, 1–23.
- Johansson, E., & Dixon, N. (2013). Replicative DNA polymerases. *Cold Spring Harbor Perspectives in Biology*, *5*, 3–6.
- Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P., & Yang, W. (2001). Composite Active Site of an ABC ATPase. *Molecular Cell*, *7*, 1–12.
- Junop, M. S., Yang, W., Funchain, P., Clendenin, W., & Miller, J. H. (2003). In vitro and in vivo studies of MutS, MutL and MutH mutants: Correlation of mismatch repair and DNA recombination. *DNA Repair*, *2*, 387–405.
- Kadyrov, F. A., Genschel, J., Fang, Y., Penland, E., Edelman, W., & Modrich, P. (2009). A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8495–8500.
- Kadyrov, F. A., Dzantiev, L., Constantin, N., & Modrich, P. (2006). Endonucleolytic Function of MutL α in Human Mismatch Repair. *Cell*, *126*, 297–308.
- Kadyrov, F. A., Holmes, S. F., Arana, M. E., Lukianova, O. A., O'Donnell, M., Kunkel, T. A., & Modrich, P. (2007). *Saccharomyces cerevisiae* MutL α is a mismatch repair endonuclease. *Journal of Biological Chemistry*, *282*(51), 37181–37190.
- Kadyrova, L. Y., & Kadyrov, F. A. (2015). Endonuclease activities of MutL α and its homologs in DNA mismatch repair. *DNA Repair*, *38*, 42–49.
- Kelman, Z., & O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annual Review of Biochemistry*, *64*, 171–200.

- Kong, X. P., Onrust, R., O'Donnell, M., & Kuriyan, J. (1992). Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell*, *69*, 425–437.
- Kosinski, J., Plotz, G., Guarné, A., Bujnicki, J. M., & Friedhoff, P. (2008). The PMS2 subunit of human MutL α contains a metal ion binding domain of the iron-dependent repressor protein family. *Journal of Molecular Biology*, *382*, 610–627.
- Kroutil, L. C., Register, K., Bebenek, K., & Kunkel, T. A. (1996). Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry*, *35*(1994), 1046–1053.
- Kunkel, T. A., & Erie, D. A. (2005). DNA mismatch repair. *Annual Review of Biochemistry*, *74*, 681–710.
- Kunkel, T. A., & Erie, D. A. (2015). Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annual Review of Genetics*, *49*, 291–313.
- Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N., & Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature*, *407*, 711–7.
- Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., ... Nyström-Lahti, M. (1993). Mutations of a MutS homolog in hereditary nonpolyposis colorectal cancer. *Cell*, *75*, 1215–1225.
- Lee, B. I., & Wilson, D. M. (1999). The RAD2 domain of human exonuclease 1 exhibits 5' to 3' exonuclease and flap structure-specific endonuclease activities. *Journal of Biological Chemistry*, *274*(53), 37763–37769.
- Lee, J. B., Cho, W. K., Park, J., Jeon, Y., Kim, D., Lee, S. H., & Fishel, R. (2014). Single-molecule views of MutS on mismatched DNA. *DNA Repair*, *20*, 82–93.
- Leirmo, S., Harrison, C., Cayley, D. S., Burgess, R. R., & Record, M. T. (1987). Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions in vitro. *Biochemistry*, *26*(1985), 2095–2101.
- Lenain, C., Bauwens, S., Amiard, S., Brunori, M., Giraud-Panis, M. J., & Gilson, E. (2006). The Apollo 5' Exonuclease Functions Together with TRF2 to Protect Telomeres from DNA Repair. *Current Biology*, *16*, 1303–1310.
- Lenhart, J. S., Pillon, M. C., Guarné, A., & Simmons, L. A. (2013). Trapping and visualizing intermediate steps in the mismatch repair pathway in vivo. *Molecular Microbiology*, *90*, 680–698.

- Li, G.-M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Research*, *18*, 85–98.
- Lyer, R. R., Pluciennik, A., Burdett, V., & Modrich, P. L. (2006). DNA mismatch repair: Functions and mechanisms. *Chemical Reviews*, *106*, 302–323.
- Lynch, H., Shaw, M. ., Magnuson, C. ., Larsen, A. ., & Krush, A. (1966). Hereditary factors in cancer. *Archive Internal Medicine*, *117*, 545–563.
- Malvern. (2008). Assessing the quality of proteins using light scattering techniques.
- Martín-López, J. V., & Fishel, R. (2013). The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. *Familial Cancer*, *12*, 159–168.
- Mauris, J., & Evans, T. C. (2009). Adenosine triphosphate stimulates *Aquifex aeolicus* MutL endonuclease activity. *PLoS ONE*, *4*(9), e7175.
- Mendillo, M. L., Putnam, C. D., & Kolodner, R. D. (2007). *Escherichia coli* MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo. *Journal of Biological Chemistry*, *282*(22), 16345–16354.
- Mensenkamp, A. R., Vogelaar, I. P., Van Zelst-Stams, W. A. G., Goossens, M., Ouchene, H., Hendriks-Cornelissen, S. J. B., ... Ligtenberg, M. J. L. (2014). Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. *Gastroenterology*, *146*(3), 643–646.
- Meselson, M., & Stahl, F. W. (1958). The Replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *44*, 671–682.
- Miguel, V., Correa, E. M. E., De Tullio, L., Barra, J. L., Argaraña, C. E., & Villarreal, M. A. (2013). Analysis of the interaction interfaces of the N-terminal domain from *Pseudomonas aeruginosa* MutL. *PLoS ONE*, *8*(7), e69907.
- Namadurai, S., Jain, D., Kulkarni, D. S., Tabib, C. R., Friedhoff, P., Rao, D. N., & Nair, D. T. (2010). The C-terminal domain of the MutL homolog from *Neisseria gonorrhoeae* forms an inverted homodimer. *PLoS ONE*, *5*(10).
- Natrajan, G., Lamers, M. H., Enzlin, J. H., Winterwerp, H. H. K., Perrakis, A., & Sixma, T. K. (2003). Structures of *Escherichia coli* DNA mismatch repair enzyme MutS in complex with different mismatches: A common recognition mode for diverse substrates. *Nucleic Acids Research*, *31*(16), 4814–4821.

- Nimonkar, A. V., Ozsoy, a Z., Genschel, J., Modrich, P., & Kowalczykowski, S. C. (2008). Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 16906–16911.
- O'Donnell, M., Kuriyan, J., Kong, X.-P., Stukenberg, P. T., & Onrust, R. (1993). Sliding clamps of DNA polymerases. *Journal of Molecular Biology*, *234*, 915–25.
- Obmolova, G., Ban, C., Hsieh, P., & Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, *407*, 703–10.
- Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., & Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, *59*, 598–605.
- Papadopoulos, N., Nicolaides, N., Wei, Y., Ruben, S., Carter, K., Rosen, C., ..., Adams M.D. (1994). Mutation of a MutL homolog in hereditary colon cancer. *Science*, *263*(5153), 1625–1629.
- Paull, T. T., & Gellert, M. (1998). The 3' to 5' exonuclease activity of MRE11 facilitates repair of DNA double-strand breaks. *Molecular Cell*, *1*, 969–979.
- Pavlov, Y., Mian, I., & Kunkel, T. A. (2003). Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Current Biology*, *13*, 744–748.
- Peltomäki, P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. *Journal of Clinical Oncology*, *21*(6), 1174–1179.
- Pillon, M. C., Babu, V. M. P., Randall, J. R., Cai, J., Simmons, L. A., Sutton, M. D., & Guarné, A. (2015). The sliding clamp tethers the endonuclease domain of MutL to DNA. *Nucleic Acids Research*, *9140*(9), 1-14.
- Pillon, M. C., Dubinsky, M., Johnston, R. N., Liu, S. L., & Guarné, A. (2013). Characterization of the defects in the ATP lid of *E. coli* MutL that cause transient hypermutability. *DNA Repair*, *12*(10), 864–869.
- Pillon, M. C., Lorenowicz, J. J., Uckelmann, M., Klocko, A. D., Mitchell, R. R., Chung, Y. S., ... Guarné, A. (2010). Structure of the endonuclease domain of MutL: unlicensed to cut. *Molecular Cell*, *39*, 145–151.
- Pillon, M. C., Miller, J. H., & Guarné, A. (2011). The endonuclease domain of MutL interacts with the β sliding clamp. *DNA Repair*, *10*(1), 87–93.

- Plon, S. E., Eccles, D. M., Easton, D., Foulkes, W. D., Greenblatt, M. S., Hogervorst, F. B. L., & Hoogerbrugge, N. (2008). Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human Mutations*, *29*(11), 1282–1291.
- Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., & Modrich, P. (2010). PCNA function in the activation and strand direction of MutL α endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 16066–16071.
- Qiu, R., DeRocco, V. C., Harris, C., Sharma, A., Hingorani, M. M., Erie, D. A., & Wenginger, K. R. (2012). Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. *EMBO Journal*, *31*(11), 2528–40.
- Qiu, R., Sakato, M., Sacho, E. J., Wilkins, H., Zhang, X., Modrich, P., ... Wenginger, K. R. (2015). MutL traps MutS at a DNA mismatch. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 10914–9.
- Räschle, M., Dufner, P., Marra, G., & Jiricny, J. (2002). Mutations within the hMLH1 and hPMS2 subunits of the human MutL α mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutS α . *Journal of Biological Chemistry*, *277*(24), 21810–21820.
- Räschle, M., Marra, G., Nyström-Lahti, M., Schär, P., & Jiricny, J. (1999). Identification of hMutL β , a heterodimer of hMLH1 and hPMS1. *Journal of Biological Chemistry*, *274*(45), 32368–32375.
- Reha-Krantz, L. J. (2010). DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1804*(5), 1049–1063.
- Reyes, G. X., Schmidt, T. T., Kolodner, R. D., & Hombauer, H. (2015). New insights into the mechanism of DNA mismatch repair. *Chromosoma*, *124* (4), 443-462.
- Richeys, B., Cayley, D. S., Mossing, M. C., Kolka, C., Anderson, C. F., Farrar, T. C., & Record, M. T. (1987). Variability of the intracellular ionic environment of *Escherichia coli*. *Cell*, *262*(15), 7157–7164.
- Rupert, P., Daughdrill, G., Bowerman, B., & Matthews, B. (1998). A new DNA-binding motif in the SKN-1 binding domain-DNA complex. *Nature Structural Biology*, 400–406.

- Sacho, E. J., Kadyrov, F. A., Modrich, P., Kunkel, T. A., & Erie, D. A. (2008). Direct visualization of asymmetric adenine nucleotide-induced conformational changes in MutLa. *Molecular Cell*, 29, 112–121.
- Schmutte, C., Marinescu, R. C., Sadoff, M. M., Guerrette, S., Overhauser, J., & Fishel, R. (1998). Exonuclease I interacts with the mismatch repair protein hMSH2. *Cancer Research*, 58 (215), 4537–4542.
- Schmutte, C., Sadoff, M. M., Shim, K. S., Acharya, S., & Fishel, R. (2001). The interaction of DNA mismatch repair proteins with human exonuclease I. *Journal of Biological Chemistry*, 276(35), 33011–33018.
- Sharma, A., Doucette, C., Biro, F. N., & Hingorani, M. M. (2013). Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair. *Journal of Molecular Biology*, 425(22), 4192–4205.
- Sharma, A., & Hingorani, M. (2013). Understanding the kinetic mechanism of MutS during mismatch recognition and initiation of repair. *Biophysical Journal*, 104(2), 76.
- Siegel, R., Miller, K., & Jemal, A. (2015). Cancer statistics , 2015 . CA: *Cancer Journal for Clinicians*, 65(1), 29.
- Simmons, L. A., Davies, B. W., Grossman, A. D., & Walker, G. C. (2008). β -clamp directs localization of mismatch repair in *Bacillus subtilis*. *Molecular Cell*, 29, 291–301.
- Spurdle, A. B., Thompson, B., Tavtigian, S., & Greenblatt, M. S. (2011). InSiGHT variant interpretation committee: Mismatch repair gene variant classification criteria 2013-08_InSiGHT_VIC_v1.9, From <http://insight-group.org/variants/database>.
- Stukenberg, P. T., Turner, J., & O'Donnell, M. (1994). An explanation for lagging strand replication: Polymerase hopping among DNA sliding clamps. *Cell*, 78, 877–887.
- Su, S. S., Lahue, R. S., Au, K. G., & Modrich, P. (1988). Mismatch specificity of methyl-directed DNA mismatch correction in vitro. *Journal of Biological Chemistry*, 263, 6829–6835.
- Tiwari, a. K., Roy, H. K., & Lynch, H. T. (2015). Lynch syndrome in the 21st century: clinical perspectives. *Qjm:An International Journal of Medicine*, 0, 1-8.
- Umar, A, Risinger, J. I., Hawk, E. T., & Barrett, J. C. (2004). Testing guidelines for hereditary non-polyposis colorectal cancer. *Nature Reviews Cancer*, 4, 153–158.

- Vasen, H., Blanco, I., Aktan-Collan, K., Gopie, J., Alonso, a, Aretz, S., ... Möslin, G. (2013). Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut*, *62*, 812–23.
- Vasen, H. F. a, Möslin, G., Alonso, a, Bernstein, I., Bertario, L., Blanco, I., ... Wijnen, J. (2007). Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *Journal of Medical Genetics*, *44*, 353–62.
- Viswanathan, M., & Lovett, S. T. (1998). Single-strand DNA-specific exonucleases in *Escherichia coli*: Roles in repair and mutation avoidance. *Genetics*, *149*, 7–16.
- Wang, H., Yang, Y., Schofield, M. J., Du, C., Fridman, Y., Lee, S. D., ... Erie, D. A. (2003). DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 14822–7.
- Wang, T.-F., Kleckner, N., & Hunter, N. (1999). Functional specificity of MutL homologs in yeast: Evidence for three MLH1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proceedings of the National Academy of Sciences*, *96*, 13914–13919.
- Warren, J. J., Pohlhaus, T. J., Changela, A., Iyer, R. R., Modrich, P. L., & Beese, L. (2007). Structure of the human MutS α DNA lesion recognition complex. *Molecular Cell*, *26*, 579–592.
- Warthin, A. (1913). Heredity with reference to carcinoma. *Archive Internal Medicine*, *12* (5), 546-555.
- Wigley, D. B., Davies, G. J., Dodson, E. J., Maxwell, A., & Dodson, G. (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature*, *351*, 624–629.
- Winkler, I., Marx, A. D., Lariviere, D., Heinze, R. J., Cristovao, M., Reumer, A., ... Friedhoff, P. (2011). Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in *Escherichia coli*. *Journal of Biological Chemistry*, *286*(19), 17326–17337.
- Worthley, D. L., Walsh, M. D., Barker, M., Ruszkiewicz, A., Bennett, G., Phillips, K., & Suthers, G. (2005). Familial mutations in PMS2 can cause autosomal dominant hereditary nonpolyposis colorectal cancer. *Gastroenterology*, *128*, 1431–1436.
- Wu, Y., Berends, M. J., Sijmons, R. H., Mensink, R. G., Verlind, E., Kooi, K. A., ... Hofstra, R. M. (2001). A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nature Genetics*, *29*, 137–138.

- Wu, H., Zeng, H., Lam, R., Tempel, W., Kerr, I. D., & Min, J. (2015). Structure of the human MLH1 N-terminus: Implications for predisposition to Lynch syndrome. *Acta Crystallographica Section:F Structural Biology Communications*, *71*, 981–985.
- Xie, J., Guillemette, S., Peng, M., Gilbert, C., Buermeier, A., & Cantor, S. B. (2010). An MLH1 mutation links BACH1/FANCD1 to colon cancer, signaling, and insight toward directed therapy. *Cancer Prevention Research*, *3*, 1409–1416.
- Yamamoto, T., Iino, H., Kim, K., Kuramitsu, S., & Fukui, K. (2011). Evidence for ATP-dependent structural rearrangement of nuclease catalytic site in DNA mismatch repair endonuclease MutL. *Journal of Biological Chemistry*, *286*(49), 42337–42348.
- Yao, N. Y., & O'Donnell, M. (2010). SnapShot: The Replisome. *Cell*, *141*(6), 1088–1088.
- Zakharyevich, K., Ma, Y., Tang, S., Hwang, P. Y. H., Boiteux, S., & Hunter, N. (2010). Temporally and biochemically distinct activities of Exo1 during Meiosis: Double-strand break resection and resolution of double Holliday Junctions. *Molecular Cell*, *40*(6), 1001–1015.