STRUCTURE AND FUNCTION OF *B. SUBTILIS* MUTL

By

JESSICA LORENOWICZ, B.Sc

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

McMaster University

MASTER OF SCIENCE (2008) McMaster University
(Biochemistry) Hamilton, Ontario

TITLE:	Structure and Function of *B. subtilis* MutL
AUTHOR:	Jessica Lorenowicz, B.Sc. (McMaster University)
SUPERVISOR:	Dr. Alba Guarné
NUMBER OF PAGES:	xiii, 66

ABSTRACT

Maintaining genomic integrity is important for any organism. DNA mismatch repair (MMR) serves to correct errors that occur during DNA replication and recombination, such as unpaired bases or mismatched bases. MutL is a key player and serves to coordinate protein-protein interactions. Recently it has been shown that human MutL functions as an endonuclease and that this activity is imperative for functioning MMR. In this work, the X-ray crystal structure of the C-terminal endonuclease domain of *Bacillus subtilis* MutL (BsMutL-CTD) is presented. Diffraction quality crystals of BsMutL-CTD were grown using vapor diffusion. The crystal structure of BsMutL-CTD was solved using multi-wavelength anomalous diffraction. The structure reveals a putative metal binding site which clusters closely in space with endonuclease motif. Using the structure and sequence homology, several mutations were made and an investigation into the endonuclease activity of BsMutL was performed. BsMutL was confirmed to be a manganese-dependent endonuclease and key residues which contribute to endonuclease function were identified.

ACKNOWLEDGEMENTS

Firstly, thanks to Dr. Alba Guarné for giving me such a great project. Your time and commitment to the project was outstanding and I can't wait to see the great things that come out of your lab in the future! I would also like to thank both my committee members for their guidance and help throughout my studies. Thanks to Dr. Ray Truant for great solutions to difficult problems, and showing me just how stable DNA really is. Thanks to Dr. Murray Junop for giving very inspirational talks, helpful advice, equipment and reagents and of course, the most friendly smile just when you need it.

This project would have not have happened without several members of the department. Thanks to Michael D'Elia of the Brown lab for providing the genome of *Bacillus subtilis* which started the ball rolling and the crystals growing. Thanks to the entire Junop lab for graciously sharing equipment and reagents, as well as helpful discussions. Thanks to Sean Jackson for all his encouragment, helpful discussions and teaching me all the wonderful things Pymol can do. Thanks to Robin Hallett for assistance with ImageJ and quantification.

Of course, my special thanks to all the Guarné lab members. You don't know how much I will miss you! All of your help and support have made this project possible. In particular, Yu Seon Chung for data collection of native BsMutL at the synchrotron and hours of help on the computer, Ryan Mitchell for cloning and purification of chelating mutants, Monica Pillon for carrying on this awesome project, Melanie Gloyd for encouragement at the synchrotron, Lindsay

Matthews for keeping me safe at all times and Stu Menzies for opening tightly closed bottles and editing my thesis.

TABLE OF CONTENTS

LIST OF FIGURES

Figure 21: Putative metal binding site of BsMutL with zinc and magnesium placed in.

LIST OF TABLES

LIST OF ABBREVIATIONS

ABC – ATP binding cassette

ADP – adenosine diphosphate

AMPPnP - 5'-adenylyl beta,gamma-imidodiphosphate

ATP – adenosine triphosphate

BER – base excision repair

BSA – bovine serum albumin

BsMutL – *Bacillus subtilis* MutL

BsMutL-CTD – *Bacillus subtilis* MutL C-terminal domain

CTD – C-terminal domain

DNA – deoxyribonucleic acid

dsDNA – double stranded DNA

DtxR – Diphtheria toxin repressor

E. coli – *Escherchia coli*

EDTA - ethylenediaminetetraacetic acid

GHKL – DNA gyrase, Hsp90, histidine kinase, MutL ATPase family

HNPCC – hereditary non-polyposis colorectal cancer

IdeR – Iron-dependent regulator

IDLs – Insertion/deletion loops

IPTG - Isopropyl β-D-1-thiogalactopyranoside

MAD – multi-wavelength anomalous diffraction

MLH1 – MutL homolog 1

MMR – mismatch repair

NER – nucleotide excision repair

NTD – N-terminal domain

OD – optical density

PBS – phosphate buffered saline

PCNA – proliferating cell nuclear antigen

PCR – polymerase chain reaction

PEG – polyethylene glycol

PMS2 – post meiotic segregation 2

RFC – replication factor C

Rmsd – root mean square deviation

SDS PAG - sodium dodecyl sulphate polyacrylamide gel

SeMet – selenomethionine-substituted BsMutL-CTD

ssDNA – single stranded DNA

TAE – tris acetic acid EDTA

TEV – Tobacco etch virus

CHAPTER 1: INTRODUCTION

1.1 GENOME STABILITY

One of the greatest challenges facing living cells is the maintenance of genomic integrity. A cell is constantly exposed to an assortment of damaging agents that threaten the fidelity of its DNA. Not surprisingly, organisms have devised several DNA repair mechanisms to protect and defend against DNA damage and thus safeguard the genome. The role of any DNA repair pathway is to ensure that the DNA is passed on error-free to progeny. There are three key DNA repair pathways: nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR). Each of these pathways have become highly specialized to repair a particular type of damage. Damage resulting in bulky lesions is repaired via NER, whereas DNA bases which have been modified but do not overtly distort the DNA helix are repaired by BER (1). Each of these pathways contains a base that has been physically damaged. However, damage does not always involve modifying a base. Errors which occur during DNA replication, namely mismatched bases or small insertion or deletion loops (IDLs) consist of non-damaged bases which have been mispaired and have escaped the proof-reading capabilities of the replication machinery, for example a G-T. These types of errors are corrected by MMR (2).

Mismatch repair proteins also participate in other genome-stabilizing pathways such as meiotic and mitotic recombination, DNA damage signaling and apoptosis and repressing recombination between similar but non-identical sequences (homeologous recombination). In mammalian cells, mismatch repair proteins are involved in triggering G2 cell cycle arrest in the presence of certain cytotoxic chemicals (3). The broad roles of MMR proteins in a cell explains how the system is able to increase the stability of the genome 100-1000 fold (4).

There are marked increases to the spontaneous mutation rate in systems that have defects in MMR. Mutations in essential MMR proteins in humans manifests itself in several forms of sporadic cancers, most notably, hereditary non-polyposis colorectal cancer (HNPCC) (5,6). HNPCC is a predisposition to colorectal cancer at much earlier ages in a lifetime than the normal population. As well, defects in MMR are implicated in several other related cancers, such as endometrial, stomach, ovarian, and renal. Sequences of repeating DNA termed microsatellite regions are particularly prone to errors due to slippage of the replication machinery, resulting in IDLs. As a consequence of defective MMR, these regions become unstable. This "microsatellite instability" (MSI) is a hallmark characteristic of HNPCC. Over 90% of HNPCC tumors due to MMR deficiencies are classified as MSI (7). The field of DNA mismatch repair has been rapidly expanding since the link to HNPCC was established.

1.2 MISMATCH REPAIR IN *Escherichia coli*

MMR is very well conserved process and homologues to MMR proteins are found in almost all organisms. This suggests that the mode of repair is similar. MMR has been best characterized in *E. coli* and provides a useful paradigm for understanding human mismatch repair.

Mismatch repair requires three distinct stages to execute its function properly: recognition of the mismatched base, excision of the incorrect base, and repair of the lesion created. A key component of this process is the ability to distinguish between the correct base (which lies on the template strand) and the incorrect base, lying on the newly synthesized strand. In *Escherichia coli (E. coli)* and other gram negative bacteria this is accomplished by DNA adenine methylase (Dam methylase). Dam methylase is responsible for methylating the adenine in a GATC sequence of DNA. During DNA replication, the newly synthesized strand is transiently unmethylated. MMR is able to exploit this hemi-methylated state of DNA and thus a distinction between the new strand (which contains the erroneous base) and the template strand (which contains the correct information) can be made. In gram-positive bacteria and in humans, which lack dam methylation, the strand discrimination mechanism is unclear.

MMR is a concerted action of several proteins. The system has been reconstituted *in vitro* as a minimal system and required: MutS, MutL, MutH, UvrD, DNA pol III holoenzyme, single stranded binding protein (SSB), DNA ligase, and

strand-specific exonucleases (ExoI, ExoX, ExoVII, RecJ) (8). MutS, MutL, and MutH play integral roles in the initiation of repair. Briefly, MutS recognizes the mismatch or IDL and recruits MutL, which functions to coordinate protein-protein interactions with other mismatch repair proteins. Together, MutS and MutL recruit the latent endonuclease MutH, which nicks the daughter strand, providing an entry point for helicase UvrD to unwind towards the mismatch. Single-stranded polarity-specific exonucleases are then able to excise the newly synthesized strand beyond the mismatch. *E. coli* utilizes four exonucleases: ExoI and ExoX have 3' to 5' directionality, while RecJ and ExoVII possess 5' to 3'. This redundancy allows for mismatch repair to proceed bidirectionally; the nick made by MutH can be either 5' or 3' to the mismatch, as seen in Figure 1 (9,10). After the removal of the mismatched bases, DNA polymerase III holoenzyme is then able to resynthesize the DNA, followed by ligation by DNA ligase.

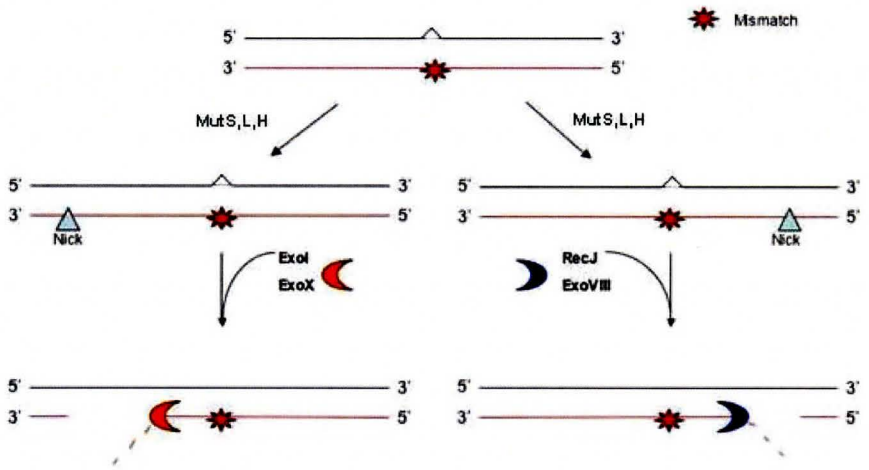


Figure 1: Mismatch repair is bidirectional. Schematic representation of bidirectionality of prokaryotic mismatch repair with template strand DNA shown in black, newly synthesized DNA in red.

In order to understand how a relatively small set of proteins are able to coordinate a task of such magnitude as genome maintenance, each protein has been studied at the molecular level. A combination of structural and biochemical studies have lead to a much better understanding of the role of each protein in MMR initiation and from this, several mechanisms have been proposed.

MutS

Recognition of a mismatched or unpaired base is accomplished by dimer MutS. In *E. coli* and other prokaryotes, MutS functions as a homodimer. MutS forms a sliding clamp and follows the replication machinery in search of mismatches. MutS has been shown to interact with β sliding clamp, which serves as a processivity factor for DNA pol III (11,12). MutS is able to recognize mismatches because mismatched DNA has less stable base stacking, and in turn is more easily deformed. MutS forms an energetically stable interaction with mismatched DNA, thus allowing a distinction to be made from homoduplex DNA. The crystal structure of *Thermus aquaticus* MutS bound to an unpaired thymidine showed that MutS forms a large oval with two distinct cavities large enough to fit DNA, as shown in Figure 2 (13). A similar structure was seen with *E. coli* MutS bound to a GT mismatch (14).

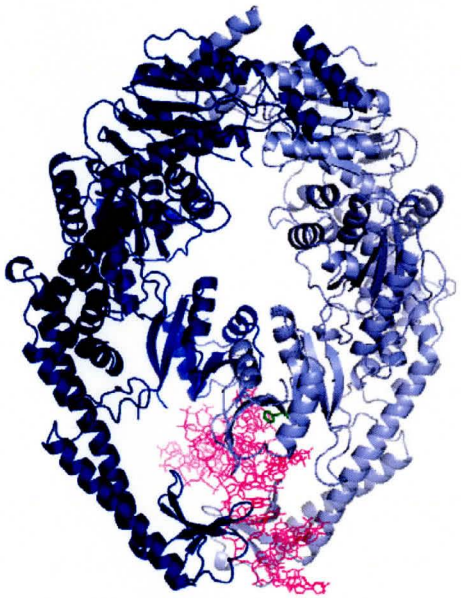


Figure 2: Crystal structure of *T. aquaticus* MutS bound to heteroduplex DNA. (13). Each monomer shown in blue and gray, with DNA in pink and conserved phenylalanine shown in green (PDB 1ewq).

Although a homodimer in *E. coli* and *T. aquaticus*, the structure reveals MutS binds DNA in an asymmetric manner with each monomer contributing differently to stabilize the complex. One monomer contributes a highly conserved phenylalanine which stacks with the unmatched base, causing a kink in the DNA. The other monomer is only involved in stabilizing the heteroduplex complex, via sugar phosphate interactions with the DNA.

Separate from the mismatch binding domain, MutS contains an ATPase domain. MutS belongs to the ATP binding cassette (ABC) ATPase superfamily and the ATPase activity is essential to repair (15). The ATPase activity of MutS brings a very interesting level of regulation to the initiation of repair. DNA binding of MutS causes increased affinity for binding ATP, while binding of ATP reduces the affinity for mismatched DNA. At first the role of ATP seems contradictory,

however, the difference in affinity for homoduplex and heteroduplex DNA allows for a very elegant mechanism allowing for discrimination between the two. ATP also functions to decrease the affinity of MutS for homoduplex DNA even more than heteroduplex DNA. This ensures that MutS only binds to true mismatches and provides a proofreading mechanism when checking for mismatches (16). It is essential that MutS bind both a mismatch and ATP in order to activate downstream mismatch repair proteins, MutH, so that mismatch repair is only initiated when a true mismatch is found.

MutL

MutL mediates several key interactions in mismatch repair, including interacting with MutS on a mismatch, activating MutH, and loading helicase UvrD. MutL plays a distinct role in the coordination between mismatch recognition and excision, earning it the moniker "Molecular Matchmaker".

The N-terminus of MutL homologues are very well conserved across many species. The finding that MutL has a weak ATPase activity came after the crystal structure of the N-terminal fragment of *E. coli* MutL (LN40) was solved – MutL shared a very similar fold to the ATPase domain of DNA gyrase (17). This created a new superfamily of proteins termed GHKL ATPase family (DNA Gyrase, Hsp90, histidine kinase, MutL) which all share a very similar ATP binding domain (18). In *E. coli,* LN40 exists as a monomer however, the addition of AMPPnP caused dimerization of the domains (19). A comparison of the apo-

protein structure with AMPPnP-bound form reveals a more ordered structure with several loops forming the catalytic centre. This ATPase activity is essential for MMR as mutants that are unable to hydrolyze ATP are not able to activate MutH in a mismatch-dependent assay (20).

Full-length MutL has been shown to bind DNA in a sequence and mismatch independent manner. The N-terminal domain is also capable of binding DNA, albeit at a reduced level. Several positively charge conserved residues were found to line the inside of the groove formed when the domains associate (17). A mutant defective in DNA binding was found to be completely inactive in MMR *in vivo*, which highlights the role DNA binding by MutL as being critical (20).

In contrast to the highly conserved N-terminal domain of MutL, the C-termini of the MutL homologues are quite divergent and maintain very little sequence identity with each other. This domain has been credited with mediating dimerization and also aiding the N-terminal domain in DNA binding. In contrast to the N-terminal domain, the C-terminal domain is not capable of binding DNA itself (21,22).

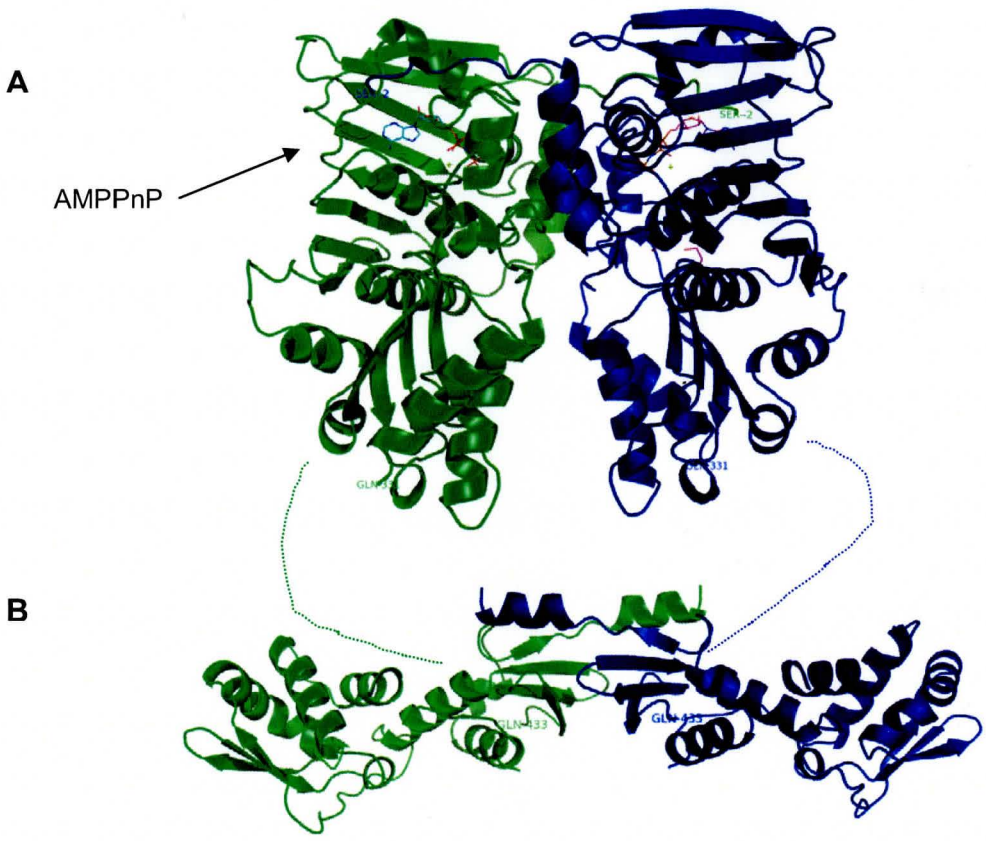


Figure 3: Model of full-length MutL. A – N-terminal domain of MutL bound to AMPPnP. B – C-terminal domain of MutL. Each monomer shown in green and blue (19,21). (PDB: 1b63, 1x9z).

The truncated *E. coli* MutL C-terminus exists as a dimer in solution as expected and crystallized as a dimer as well. A full length model of MutL is shown in Figure 3 (23). It shows the N-terminal ATPase domains as solved bound to AMPPnP connected to the C-terminal domain via a flexible linker, shown as dashed lines.

MutH

Once MutS has bound a mismatch, MutL is recruited in an ATP-dependent manner and together they activate the latent endonuclease MutH. MutH is both a

sequence- and methylation-specific endonuclease: MutH nicks 5' to unmethylated GATC sites of hemi-methylated DNA. Although MutH is able to bind fully methylated GATC sites, it is unable to nick them effectively (24). This provides a method of strand discrimination in *E. coli* as the newly synthesized strand is transiently unmethylated during replication. MutH only nicks hemi-methylated GATC sites, and the activity is increased nearly 30-fold in the presence of a mismatch, MutS, MutL and ATP (25). This provides an important level of regulation, so that the excision event is not ongoing in the absence of a mismatch. Interestingly, the distance between a mismatch and the GATC site can extend up to 1000 base pairs (25). How MutS and MutL manage to communicate with MutH is still under debate. From the apo-protein structure, MutH shares structural similarity with Type II restrictions enzymes, such as PvuII (26). However, MutH functions as a monomer, while PvuII functions as a dimer, explaining how MutH nicks only one strand, while PvuII cleaves double stranded DNA. The active site of MutH is a conserved DEK motif found in a cleft which is responsible for catalysis (26). The structure of *Haemophilus influenzae* MutH bound to hemi-methylated DNA revealed that the active site is flanked on both sides by the N- and C-terminal domains which form two arms that can clamp against double stranded DNA, as seen in Figure 4 (27).

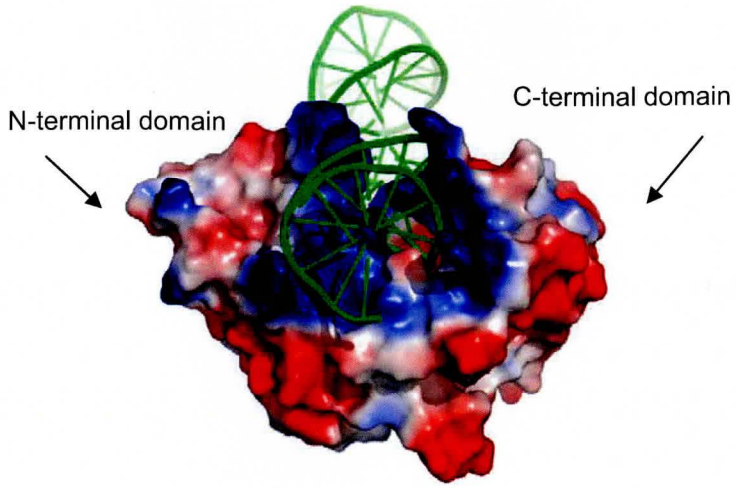


Figure 4: Crystal structure of *H. influenzae* MutH bound to DNA. Electrostatic view of MutH with positively charged residues shown in blue and negatively charged residues shown in red. DNA shown in green (27). (PDB 2aoR).

This closed conformation bound to DNA is stimulated by the presence of other mismatch repair proteins, allowing for MutH to nick. This nick creates an entry point for the helicase UvrD which is able to unwind towards the mismatch (28).

1.3 MISMATCH REPAIR IN EUKARYOTES

Orthologues to MMR proteins are found in nearly all organisms, suggesting that the mode of repair is also well conserved. In eukaryotes, MutS and MutL homologues have evolved to form heterodimers as opposed to homodimers. However, the eukaryotic counterparts still retain common architecture to their prokaryotic orthologues. MutS forms two distinct complexes: MutSα, comprised of MSH2 and MSH6 which plays the primary role in base-base

mismatch recognition and MutSβ, comprised of MSH2 and MSH3, which recognizes larger insertion and deletion loops. Similarly, MutL has three distinct homologues in humans: MutLα (MLH1 and PMS2), the key player in mismatch repair; MutLβ (MLH1 and PMS1), which as of yet has no biological function and MutLγ (MLH1 and MLH3) which is thought to play a role in meiosis. In yeast, the nomenclature is slightly different, with PMS1 being the equivalent to PMS2 in humans, and thus MLH1 and PMS1 comprise MutLα, the major player in repair.

E. coli	Humans
MutS	MutSα (MSH2, MSH6)
	MutSβ (MSH2, MSH3)
MutL	MutLα (MLH1, PMS2)
	MutLβ (MLH1, PMS1)
	MutLγ (MLH1, MLH3)
MutH	?
UvrD	?
ExoI, ExoX	EXOI
ExoVII, RecJ	3' excision from polε, polδ respectively
Pol III	PCNA, Pol δ
SSB	RPA
DNA ligase	DNA ligase

Table 1: Summary of DNA replication and mismatch repair orthologues found in humans (Adapted from (29)).

Despite the strong conservation seen between the two systems, there are marked differences as well. Eukaryotes have no known MutH homologue, leaving the question of strand discrimination in eukaryotes unanswered (see section 1.5).

1.4 PUTATIVE MECHANISMS OF REPAIR

Each protein in the mismatch repair pathway has been characterized biochemically, allowing for several mechanisms of mismatch repair to be proposed. These mechanisms attempt to reconcile how MutS and MutL can signal to MutH up to 1000 base pairs away from a mismatch. They can be broken into two categories: moving *cis* and stationary *trans* models.

There are two distinct moving models: molecular switch and translocation. In the translocation model (30,31), electron microscopy and increasing concentrations of ATP were used to show alpha-loops of DNA being formed. They proposed that bacterial MutS moves away from a mismatch in an ATP-dependent reaction and mediates the formation of alpha-loops of DNA. In this model, ATP hydrolysis reduces the affinity of MutS for a mismatch and thus initiates translocation away from the mismatch. This looping of the DNA can then bring MutS into contact with MutH bound to a hemi-methylated GATC site and induce nicking.

The other moving model is termed the molecular switch model (32). In this model, similar to G-protein signaling, human MutSα is thought of being in two distinct conformations, bound to ADP (on) or bound to ATP (off). In the presence of a mismatch, the exchange of ADP for ATP is performed at an increased rate, while MutSα does not require ATP hydrolysis to leave a mismatch. This would

allow MutSα to then hydrolyze ATP and once again be capable to binding to a mismatch in a cyclic manner. While both of these models (one bacterial and one eukaryotic) imply that MutS leaves the mismatch after initial recognition, the role of the ATPase activity differs. Recent evidence supports the moving model. Using an inactive endonuclease situated between the mismatch site and the hemi-methylated GATC, a decrease in activation of MutH was shown (33). This seems to suggest that some sort of signaling occurs along the DNA between MutS and MutH. Interestingly, the activity was not completely abolished, which the authors attribute to diffusion of the endonuclease away from its binding site. It has been proposed that the sliding action of MutS-MutL may serve to deliver MutH to a GATC site, and initiate nicking (34).

The stationary or trans model hypothesizes that MutS remains bound to a mismatch while activating MutH. The evidence for this came from a *trans* activation assay in which activation of MutH occurred even when two different separate DNA substrates were used. One substrate contained a mismatch and the other contained a GATC site onto which MutH bound. Together, they were used to show that activation of MutH can occur without direct interaction. This contrasts the previous two models by showing that MutS does not need to leave a mismatch in order to effect downstream processes of repair (16). This has further been supported by DNA footprinting data showing that MutL prevents MutS from leaving a mismatch in the presence of ATP (35).

There are strengths and weaknesses to each model. In both of the moving models, MutS leaves the mismatch, thus removing a possible signal to trigger the end of repair. It is puzzling to comprehend why MutS would leave a mismatch after so searching so diligently for it. The models do, however, account for the orientation-specific loading of the strand excision machinery. Conversely, the trans model does not account for data which demonstrates MutS being able to move along a helix even in the presence of MutL (36). This still leaves the debate open as to how a mismatch site and nicking site are able to signal to each other.

1.5 MUTL AS AN ENDONUCLEASE

One of the most fascinating properties of mismatch repair is that it is bidirectional (Figure 1). In contrast to the *E. coli* system, eukaryotes only contain one known exonuclease involved in mismatch repair, EXOI, which has 5' to 3' directionality (Table 1). This quandary was a key question in the field of mismatch repair: how does 3' to 5' excision work? It is difficult to explain how MMR proceeded when the nick was located 3' to a mismatch. In fact, it was even suggested that EXOI contained a cryptic 3' to 5' exonucleolytic function (37). Several recent works have shed light on how mismatch repair occurs in humans bidirectionally. In a reconstituted system comprised of purified proteins, it was found that only MutSα, EXOI, and RPA were needed to support 5' to 3' excision, whereas the addition of MutLα, Proliferating Cell Nuclear Antigen (PCNA) and

Replication Factor C (RFC) allowed for excision in both directions (38). This finding was supported by the ground breaking discovery that MutLα is a latent endonuclease (39). MutLα is able to make single-stranded breaks in DNA in a nick-dependent manner either 5' or 3' to a mismatch. This would allow for exonuclease EXOI a point of entry regardless of orientation to the mismatch. Furthermore, the endonuclease activity was shown to be stimulated in a MutSα-, mismatch-, PCNA-, RFC- dependent manner. However, in the presence of non-physiological quantities of manganese, MutLα was able to excise supercoiled DNA without these cofactors. This suggests that the intrinsic endonuclease activity is quite low (39).

The endonuclease activity of MutL is thought to be conferred by a highly conserved metal binding sequence site, as with most nucleases, a metal binding site is required for catalysis. The conserved metal binding motif, $DQHA(X)_2E(X)_4E$, located in the C-terminus of PMS2 in humans (Figure 5). The other half of the dimer MutLα, MLH1, does not contain this conserved motif, indicating that the endonuclease motif is asymmetrical in humans.

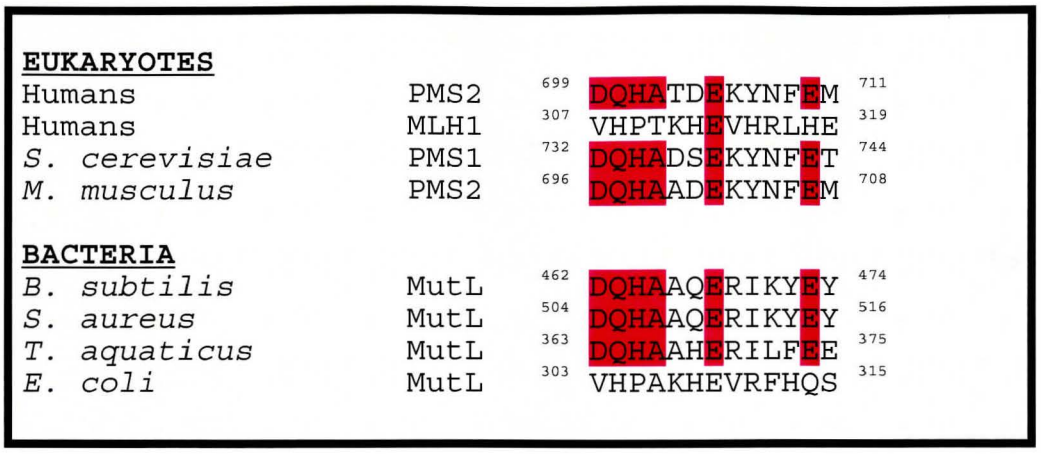


Figure 5: Conservation of putative endonuclease site (highlighted in pink) found in MutL orthologues.

After the unveiling of MutL as an endonuclease, there has been a recent flurry of activity, which has lead to some advancement in the literature. The endonuclease activity has been shown in another eukaryotic system *S. cerevisiae* (40) as well as prokaryotic *Thermus thermophilus* MutL (41). The regulation of this newly discovered endonuclease remains to be elucidated. Indeed, strand discrimination in eukaryotes still remains a question mark in mismatch repair. It is not clear how MutLα is able to distinguish the template and daughter strand in the absence of dam methylation.

The conservation of the endonuclease motif suggests a similar mode of activity amongst organisms lacking using dam methylation as a method of strand discrimination, such as *E. coli* (Figure 5). This suggests that the endonuclease activity in humans, yeast and *T. thermophilus* is conserved in organisms like *Bacillus subtilis*. This conservation allows *B. subtilis* MutL to be a good model for studying human MMR.

1.6 THESIS OBJECTIVE

The objective of this work is to structurally and biochemically evaluate the C-terminal domain of *Bacillus subtilis* MutL that harbors the conserved endonuclease motif.

CHAPTER 2: EXPERIMENTAL PROCEDURES

2.1 Cloning *B. subtilis* MutL constructs

2.1.1 Polymerase Chain Reaction

PCR was used in order to amplify the C-terminal domain of MutL (BsMutL-CTD), from residue 433 to 627. PCR was performed as follows: initial denaturation for 5 minutes at 95°C, 25 cycles of 30 seconds at 95°C, 1 minute at 55°C, 2 minutes at 68°C, and holding at 4°C.

Similarly, PCR was used in order to amplify full-length MutL (BsMutL) encompassing residues 1 to 627. PCR was performed as follows: initial denaturation for 5 minutes at 95°C, 25 cycles of 30 seconds at 95°C, 1 minute at 55°C, 4 minutes at 68°C, and holding at 4°C. For both constructs, the *B. subtilis* genome was used as a template and was a kind gift from the Brown lab, McMaster University. Primers are found in the Appendix.

2.1.2 Ligations

The PCR products contained two unique restriction sites (NcoI and XhoI, New England Biolabs) flanking the gene of interest and were ligated into pCR-Blunt vector as per manufacturers instructions (Invitrogen). The constructs and destination vector were then digested using the NcoI and XhoI restriction sites and gel extracted (Qiagen) as per manufacturers instructions. They were then ligated into pProExHtA (Invitrogen) a 1:5 vector to insert ratio together using T4 ligase (New England Biolabs) overnight at 16°C.

2.2 Production of BsMutL-CTD

2.2.1 Expression of BsMutL-CTD

BsMutL-CTD was transformed into chemically competent *E. coli* BL21(DE3) cells using heat shock method. Cells were grown until OD ~0.7 was reached and induced using 1 mM IPTG. Cells were grown for 3 hours at 37°C and then harvested at 5180 x g for 10 minutes before being resuspended in phosphate-buffered saline (PBS). Cells were then centrifuged again at 3300 x g after which pellets were frozen at -80°C.

Seleno-methionine derived BsMutL-CTD (SeMet BsMutL-CTD) was transformed into chemically competent *E. coli* methionine-auxotrophic B834(DE3) cells using heat shock method. Cells were grown until OD ~1.2 was reached and induced using 1 mM IPTG. Cells were grown for 3 hours at 37°C and then harvested as wild-type BsMutL-CTD.

2.2.2 Purification of BsMutL-CTD

Both BsMutL-CTD and SeMet BsMutL-CTD were purified identically. Pellets from 500 mL cultures were resuspended in 20 mL Nickel Buffer A (20 mM TRIS pH 8, 0.5 M NaCl, 1.4 mM 2-mercaptoethanol, 5% glycerol) with 30 mM imidazole added. Protease inhibitors were added (100mM Phenylmethanesulphonylfluoride, 100 mM Benzamidine, 0.5 mg/mL Leupeptin, 70 mg/mL Pepstatin A). Cells were sonicated and spun down at 36000 x g for 40 minutes. Protease inhibitors were added again.

Cell lysate was loaded onto a HiTrap Ni-chelating HP column (GE Healthcare Inc.) equilibrated with the resuspension buffer as above. The column was washed with 50 mL Buffer A with 30 mM imidazole, 25 mL Buffer A with 36 mM imidazole, 25 mL Buffer A with 45 mM imidazole. Proteins were eluted using Buffer A with 300 mM imidazole added. Protein fractions were combined and diluted 1:4 in Q Sepharose Buffer A (20 mM TRIS pH 8, 0.1 M EDTA, 5 mM DTT, 5% glycerol) in order to lower imidazole and sodium chloride before injecting into a Q Sepharose column (GE Healthcare Inc.) equilibrated with 20 mM TRIS pH 8, 150 mM KCl, 5 mM DTT, 5% glycerol. Proteins were purified over a linear gradient of 150 mM KCl to 400 mM KCl. Fractions were collected and concentrated to 20 mg/mL. The buffer was exchanged for a storage buffer (20 mM TRIS pH 8, 100 mM KCl, 1 mM DTT, 5% glycerol).

2.2.3 Histidine-Tag Removal

Tobacco Etch Virus (Tev) protease was used to remove the N-terminal histidine tag. A two-fold serial dilution of TEV (ranging from ~3 mg/mL to ~ 0.09375 mg/mL) was incubated with ~ 1 mg/mL of protein for 1.5 hours at room temperature in order to determine the best concentration of Tev protease for cleavage. After the histidine tag was removed, another Q Sepharose column was run identical to the first in order to remove both Tev protease and the cleaved histidine tag.

2.2.4 Oligomeric state of BsMutL-CTD

After histidine tag removal and subsequent anion exchange, BsMutL-CTD was spun at 15521 x g at 4°C for 10 minutes before being loaded onto the Superdex 75 (GE Healthcare Inc.) equilibrated at 20 mM TRIS pH 8, 100 mM KCl, 1 mM DTT, 5% glycerol. The column was equilibrated previously with a series of proteins with known molecular weights, thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), albumin (67 kDa), ovoalbumin (43 kDa), chymotrypsinogen A (25 kDa), and ribonuclease A (13.7 kDa).

2.3 Structure Determination of BsMutL-CTD

2.3.1 Crystallization

Crystallization of BsMutL-CTD and SeMet BsMutL-CTD was performed using the hanging drop method. Briefly, this technique requires the protein to be mixed with a precipitant solution, which is then suspended against a reservoir containing precipitant at a higher concentration than in the drop. In order to obtain initial crystallization conditions, a wide range of compounds known to facilitate crystallization of macromolecules were used in the form of commercial screens. Initial screens were set at a ratio of 1:1 of 5 mg/mL protein, both with and without a histidine tag at 4°C. Both Index (NeXtal), and Classics (Qiagen) were attempted and were set against a reservoir of 500 µl of 0.5 M ammonium sulphate. After obtaining initial crystal growing conditions, optimization was performed using an additive screen (Hampton Research). Further optimization

occurred, by varying protein concentration, precipitant concentration, temperature and pH until diffraction quality crystals were grown. Initial diffraction quality crystals were grown at 22°C in 15% PEG 3350, 100 mM $MgCl_2$, 100 mM TRIS pH 8.5. For flash freezing in liquid nitrogen 35% PEG 550 MME was added to the mother liquor. Final diffraction-quality crystals used in refinement were grown at 22°C in 26% PEG 550 MME, 90 mM $MgCl_2$, 100 mM TRIS pH 9 with 10% PEG 400 added to the protein prior to setting the drop. For flash freezing in liquid nitrogen 5% PEG 400 was added to the mother liquor.

SeMet BsMutL-CTD crystals were grown in a similar manner to wild-type crystals. Final diffraction-quality crystals were grown at 22°C in 25% PEG 550 MME, 25 mM $MgCl_2$, 100 mM TRIS pH 9 with 10% PEG 400 added to the protein prior to setting the drop. Crystals were frozen in the mother liquor with 35% PEG 550 MME before being flash frozen in liquid nitrogen.

2.3.2 Data Collection

The native data set of BsMutL-CTD was collected at the home source (using generator Rigaku RAXIS-4++). Crystals were flash frozen in a stream of liquid nitrogen and exposed for 45 minutes per image. A data set was collected at a wavelength of 1.54 Å in one degree increments at a detector distance of 210 mm.

The selenomethionine-derived BsMutL-CTD data was collected at beam line X29 of the Brookhaven National Laboratory, National Synchrotron Light

Source (Long Island, NY). Crystals were flash frozen in a stream of liquid nitrogen and exposed for one second exposure time. Data was collected at three different wavelengths in one degree increments for a total of 180 degrees at a detector distance of 375 mm. Data was processed using HKL2000 (42).

The second native data set of BsMutL-CTD was collected at beam line X12 of the Brookhaven National Laboratory, National Synchrotron Light Source (Long Island, NY). Crystals were flash frozen in a stream of liquid nitrogen and exposed for one second exposure time. Data was collected at a wavelength 1.0809 Å in one degree increments for a total of 180 degrees at a detector distance of 375 mm. Data collection was performed by Yu Seon Chung. Data was processed using HKL2000 (42).

2.3.3 Phasing

In order to gain phasing information, molecular replacement was attempted using the *E. coli* MutL C-terminal domain (PDB 1x9z) as a model. This technique was not successful and no solutions were found.

Using the SeMet crystal data set, multi-wavelength anomalous dispersion (MAD) was used. Selenomethionine positions were found using SOLVE (43). Density modification and solvent flattening were accomplished using Resolve.

2.4 Production of full-length BsMutL

2.4.1 Expression of BsMutL

Full-length MutL (BsMutL) was expressed at 1 mM IPTG in *E. coli* BL21(DE3) Star cells (Invitrogen) for 5 hours at 25°C and then harvested at 5180 x g for 10 minutes before being resuspended in phosphate-buffered saline. Cells were then spun again at 3300 x g after which pellets were frozen at -80°C.

2.4.2 Purification of BsMutL

Pellets were resuspended in 20 mL Nickel Buffer A (20 mM TRIS pH 8, 1.0 M NaCl, 1.4 mM 2-mercaptoethanol, 0.2 mM PMSF, 2 μg/mL leupeptin, 0.7 μg/mL pepstatin, 5% glycerol) with 30 mM imidazole added. Protease inhibitors were added (100 mM PMSF, 100 mM Benzamidine, 0.5 mg/mL Leupeptin, 70 mg/mL Pepstatin A). Cells were sonicated and spun down at 36000 x g for 40 minutes. Protease inhibitors were added again.

Cell lysate was loaded onto a HiTrap Ni-chelating HP column (GE Healthcare Inc.) equilibrated with the resuspension buffer as above. The column was washed with 50 mL Buffer A with 30 mM imidazole, 50 mL Buffer A with 45 mM imidazole, 50 mL Buffer A with 0.5 M NaCl and 45 mM imidazole. Proteins were eluted in Buffer A with 0.5 M NaCl and 240 mM imidazole. Fractions were met with 2 μL of 0.5 M EDTA as they eluted. Protein fractions were combined and diluted to 250 mM in Mono Q Buffer A (20 mM TRIS pH 8, 5 mM EDTA, 2.8 mM 2-mercaptoethanol, 5% glycerol) in order to lower imidazole and sodium

chloride before injecting into a MonoQ (5/5) column (GE Healthcare Inc.) equilibrated with 20% Buffer B (20 mM TRIS pH 8, 200 mM KCl, 5 mM EDTA, 2.8 mM 2-mercaptoethanol, 5% glycerol). Proteins were purified over a gradient of 200 mM KCl to 800 mM KCl. Fractions were collected and concentrated to 1 mg/mL. The buffer was exchanged for a storage buffer (20 mM TRIS pH 8, 100 mM KCl, 1 mM EDTA, 1.4 mM 2-mercaptoethanol, 5% glycerol).

2.5 Endonuclease Assay

BsMutL was assessed for endonuclease activity using a protocol previously described (39). The endonuclease activity was determined on supercoiled DNA in 40 μL reaction volumes containing 20 mM TRIS pH 8, 23 mM KCl, 2% glycerol, 0.5 mg/mL BSA, 8 nM DNA (pUC19, a kind gift from Dr. Murray Junop, McMaster University) and 80 nM BsMutL. Reaction was incubated for 30 minutes at 37°C and reactions were ended by addition of 0.1 mg/mL Proteinase K. After further incubation for 15 minutes at 55°C reactions were run on 1% agarose gel in TAE. As a control, a comparable amount of DNA was digested using nicking restriction enzyme NbBrsDI (New England Biolabs) for 8 minutes at 65°C. Gels were stained with ethidium bromide and visualized under ultra violet light (λ = 280 nm). The amount of nicked and supercoiled DNA was determined using ImageJ software, which is freely available from National Institutes of Health (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2007).

CHAPTER 3: Results

3.1 Purification of BsMutL-CTD

3.1.1 IMAC and Anion exchange

Purification of BsMutL-CTD and seleno-methionine derived (SeMet BsMutL-CTD) was accomplished in two steps: immobilized metal affinity chromatography (IMAC) and anion exchange. The protein contained several contaminants after IMAC and so anion exchange was performed subsequently. After anion exchange, BsMutL-CTD appears pure as judged by coomassie blue stained SDS PAG (Figure 6, lane "After Q").

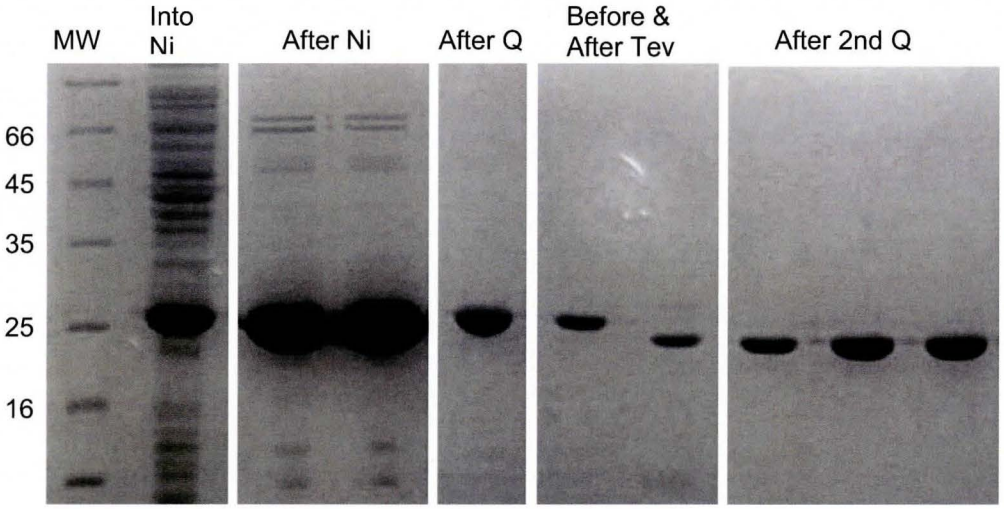


Figure 6 : Purification of BsMutL-CTD. 12% SDS PAG stained with coomassie blue showing from left to right: molecular weight marker in kDa (MW), BsMutL-CTD prior to nickel column (Into Ni), BsMutL-CTD after nickel column (After Ni), BsMutL-CTD after Q Sepharose column (After Q), BsMutL-CTD before and after Tev protease digestion (Before & After Tev), BsMutL-CTD after histidine tag removal and subsequent Q Sepharose column (After 2nd Q).

For crystallization, the N-terminal histidine tag was removed by a site-specific protease, Tobacco Etch Virus (Tev). Subsequently, an anion exchange column was run to remove Tev protease as well the cleaved tag as both do not bind to the column. The fractions used for crystallization are shown to be free of contaminants when stained with coomassie blue (Figure 6, lanes "After 2nd Q").

3.1.2 Oligomeric state of BsMutL-CTD

Previously it has been shown that *E. coli* MutL-CTD exists as a dimer in solution (21). Using size exclusion chromatography, BsMutL-CTD eluted in a single peak off a Superdex-75 column (GE Healthcare Inc.), at an approximate molecular weight of a globular protein of 45 kDa, suggesting that the *Bacillus subtilis* protein also exists as a dimer in solution (Figure 7).

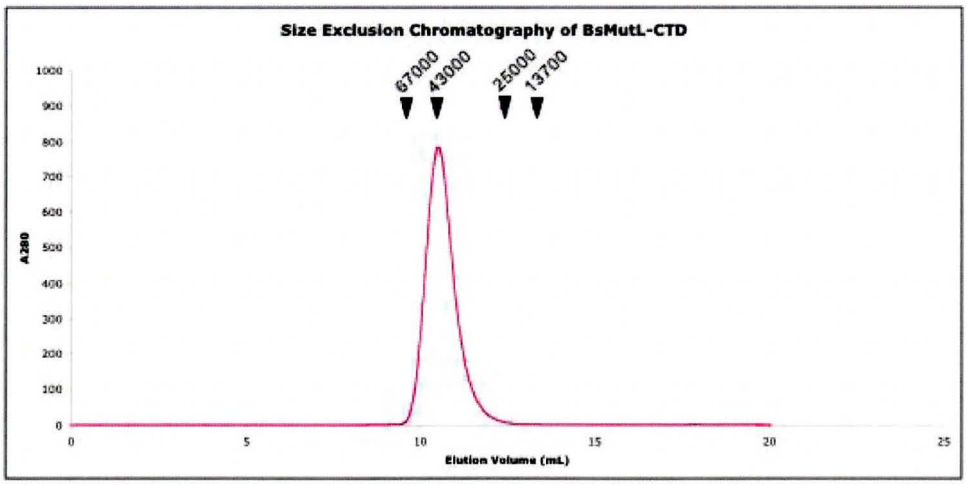


Figure 7: Size exclusion chromatography profile of BsMutL-CTD over a Superdex-75 column (GE Healthcare Inc.). Elution volumes of molecular weight standards (kDa) are indicated by arrows.

3.2 Data Collection of Native and SeMet BsMutL-CTD

3.2.1 Data Collection of SeMet BsMutL-CTD

Native crystals diffracted but phases could not be solved using molecular replacement with *E. coli* MutL (PDB 1x9z) as a model. Thus, in order to solve the structure of BsMutL-CTD using X-ray crystallography, crystals of selenomethionine-derived (SeMet) BsMutL-CTD were grown.

SeMet BsMutL-CTD crystals diffracted weakly to 2.8 Å as seen in Figure 8B. Data collection statistics are shown in Table 2. Data was collected at a total of four different wavelengths, three of which were used for phase determination. The fourth data set was excluded from phase determination due to crystal decay, which resulted in an incomplete data set. Once a complete data set was collected from these crystals, Multi-wavelength Anomalous Dispersion (MAD) was used to solve the structure (44). The first data set was collected at the absorption peak of Selenium K edge (λ=0.9792, E=12661.78, f'=4.68, f"=–7.82). The second data set was collected at the inflection point (λ=0.9794, E=12659.65 eV, f'=2.86, f"=–9.12). The third data set was collected at high energy remote wavelength (λ=0.9686, E=12800 eV). All data sets were collected with an exposure time of one second per image.

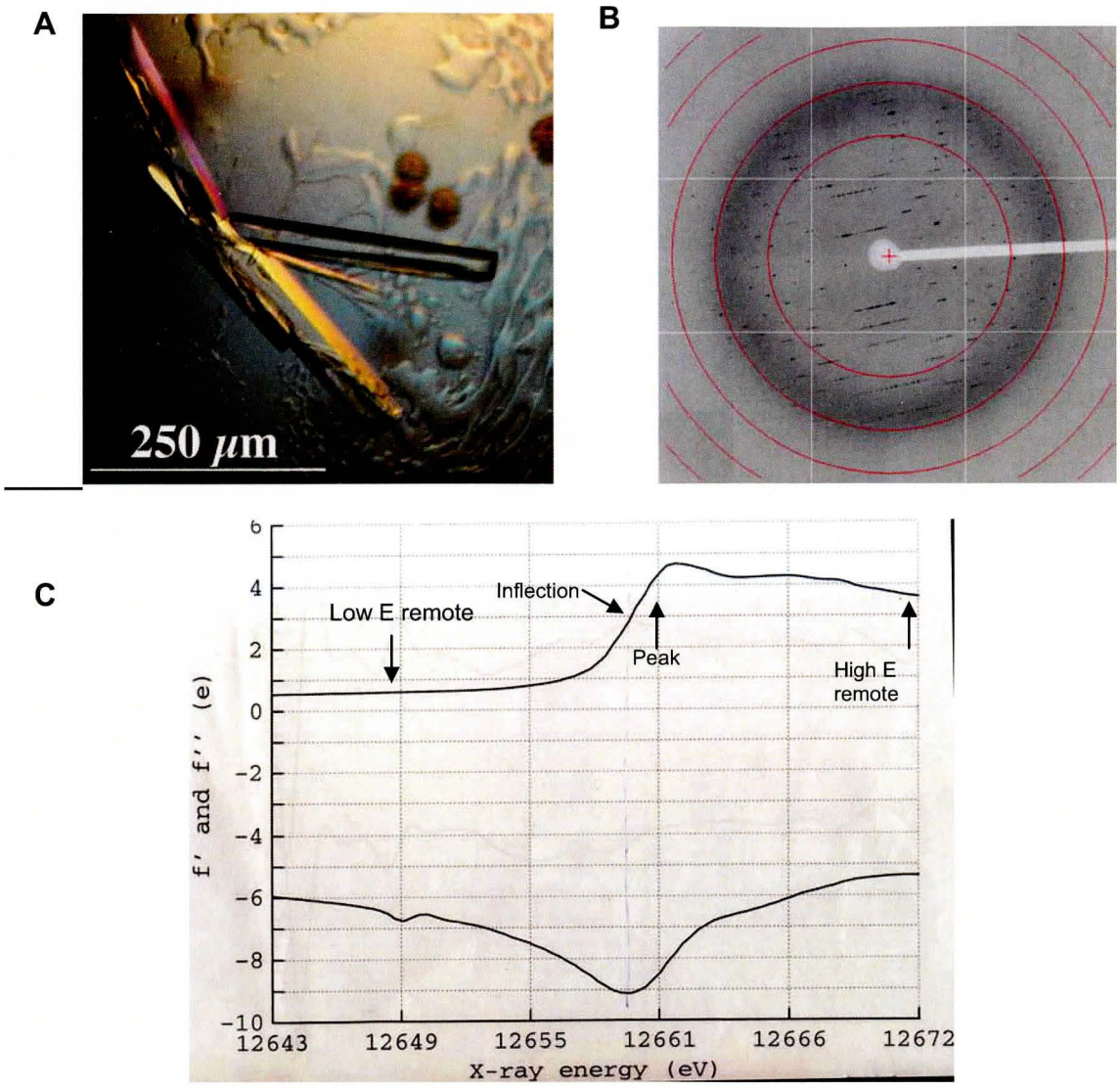


Figure 8: SeMet BsMutL-CTD crystals diffract. A - SeMet BsMutL-CTD crystals grown at 22°C in 25 mM $MgCl_2$, 25% PEG 550 MME, 100 mM TRIS pH 9. B – Sample diffraction pattern of SeMet BsMutL-CTD crystals. Concentric resolution arcs (red) correspond to resolution of 4.5 Å, 3.2 Å, 2.6 Å, 2.3 Å, 2.2 Å respectively from the centre outwards. C – Selenium scan of SeMet BsMutL-CTD crystals between 12643 eV and 12672 eV. Wavelengths data was collected at highlighted by arrows.

Table 2: Data collection statistics for SeMet BsMutL-CTD (values in brackets correspond to highest resolution shell). $*R_{merge} = \sum_h \sum_i | I_{hi} - \langle I_h \rangle | / \sum \langle I_h \rangle$ where I_{hi} is the intensity of the *i*th observation of reflection h, and $\langle I_h \rangle$ is the average intensity of redundant measurements of the h reflections.

Space Group: $C222_1$	4 Molecules in Asymmetric Unit			
Cell Dimensions	$\alpha = \beta = \gamma = 90°$	a = 87.862 Å, b = 94.967 Å, c = 218.978 Å		
Wavelength (Å)	0.9794 (Edge)	0.9792 (Peak)	0.9686 (High remote)	0.9879 (Low remote)
Resolution (Å)	35.0-2.8 (2.9-2.8)	35.0-2.8 (2.9-2.8)	35.0-2.8 (2.9-2.8)	35.0-3.0 (3.1-3.0)
Redundancy	6.2 (5.0)	6.3 (5.4)	6.0 (3.9)	6.3 (5.4)
Completeness (%)	98.2 (88.3)	98.6 (91.4)	91.0 (50.2)	98.4 (90.5)
$I/\sigma(i)$	20.2 (6.9)	18.5 (5.6)	17.8 (4.9)	34.7 (1.9)
$*R_{merge}$	0.065 (0.210)	0.081 (0.213)	0.072 (0.213)	0.053 (0.236)

In total, 26 selenium sites out of a potential 36 were found using SOLVE (45) which allowed for initial phase information to be obtained. After density modification and solvent flattening, the quality of the experimental map was outstanding and allowed for the entire structure to be traced into the density.

The limited resolution, however, hindered the refinement process. In order to have a complete data set, the resolution was cut at 3.0 Å due to crystal decay. At this resolution, the electron density was well defined for most of the main chain, however, it was weak for many side chains and the exact conformation could not be determined (Figure 9). In order to facilitate the refinement process, data of a greater resolution was required.

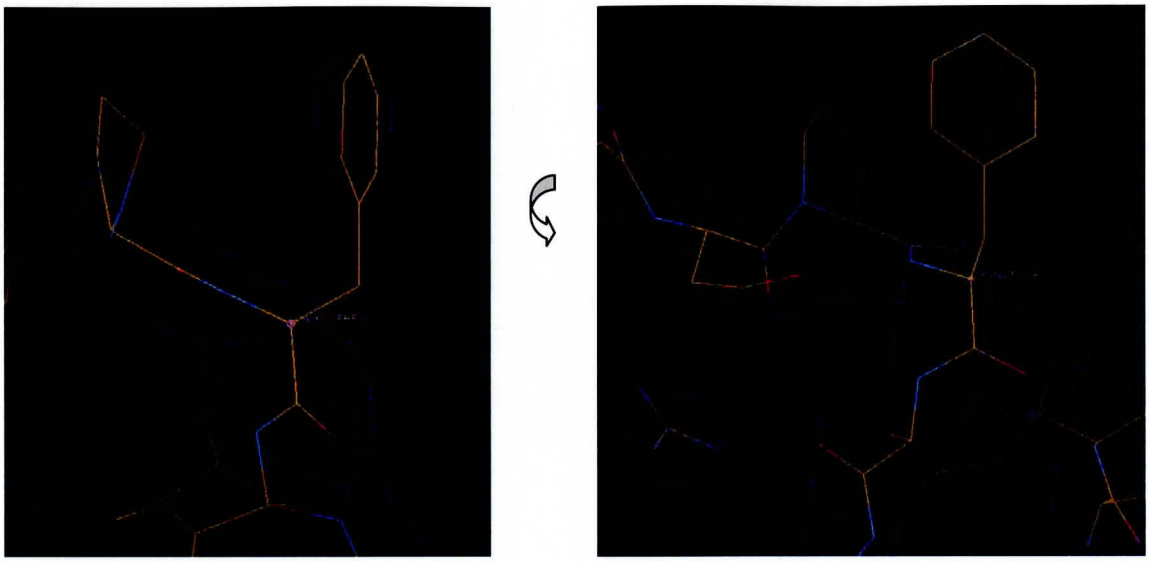


Figure 9: Electron density map after initial phasing. Example of electron density after initially solving the structure using SeMet BsMutL-CTD shown in blue mesh and model of BsMutL-CTD shown in lines. Electron density was contoured at 1σ.

3.2.1 Data Collection of Native BsMutL-CTD

Higher resolution data was obtained by growing larger, native BsMutL-CTD crystals which gave a complete data set at 2.45 Å (Table 3). This data was then used to produce a better electron density map on to which the model was placed, using rigid body refinement.

Table 3: Data collection and refinement statistics for BsMutL-CTD (values in brackets correspond to highest resolution shell). *R_{merge} = $\sum_h \sum_i | I_{hi} - \langle I_h \rangle | / \sum \langle I_h \rangle$ where I_{hi} is the intensity of the *i*th observation of reflection h, and $\langle I_h \rangle$ is the average intensity of redundant measurements of the h reflections. ~R = $\sum |F_o| - |F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} is calculated with 1709 reflections excluded from refinement.

Space Group: $C222_1$		4 Molecules in Asymmetric Unit
Cell Dimensions	$\alpha = \beta = \gamma = 90°$	a = 87.862 Å, b = 94.967 Å, c = 218.978 Å
Wavelength (Å)	1.0809	
Resolution (Å)	50-2.38 (2.47-2.38)	
Redundancy	5.5 (4.5)	
Completeness (%)	95.9 (76.0)	
$I/\sigma(i)$	30.2 (3.7)	
*R_{merge}	0.040 (0.225)	
Refinement		
Resolution (Å)	32.33-2.45	
Completeness (%)	97.9	
Reflections	31745	
~R, R_{free}	27.1%, 32.3%	
Mean B value (Å)	45.791	
Rmsd in bonds (Å)	0.016	
Rmsd in angles (°)	1.68	

3.3 Structure of BsMutL-CTD

3.3.1 Architecture of BsMutL-CTD

Four molecules of BsMutL-CTD were found in the asymmetric unit. Each monomer (residues 436-624 visible in the structure) is identical in domain architecture, but slightly different in terms of subdomain orientation and so are termed A, B, C, D for convenience.

Each monomer can be divided up into two subdomains, connected by an alpha helix. The Outer subdomain, (shown in Figure 10 as orange), is comprised of residues 484-574, while the Dimerization subdomain consists of residues 436-462 and 574-623 (shown in blue). The alpha helix (shown in yellow, residues 463-483) acts a lever between the two subdomains.

The Dimerization subdomain is comprised of three anti-parallel β strands (β1,2,3) which are connected to the lever. The lever (αA) contains the conserved endonuclease sequence and bridges both subdomains. The Outer subdomain consists of four anti-parallel β strands (β4,5,6,7) with a layer of helices on top (αB,C,D). The remaining motifs (αE, β8) then rejoin the Dimerization domain, with αF extending onto the other monomer.

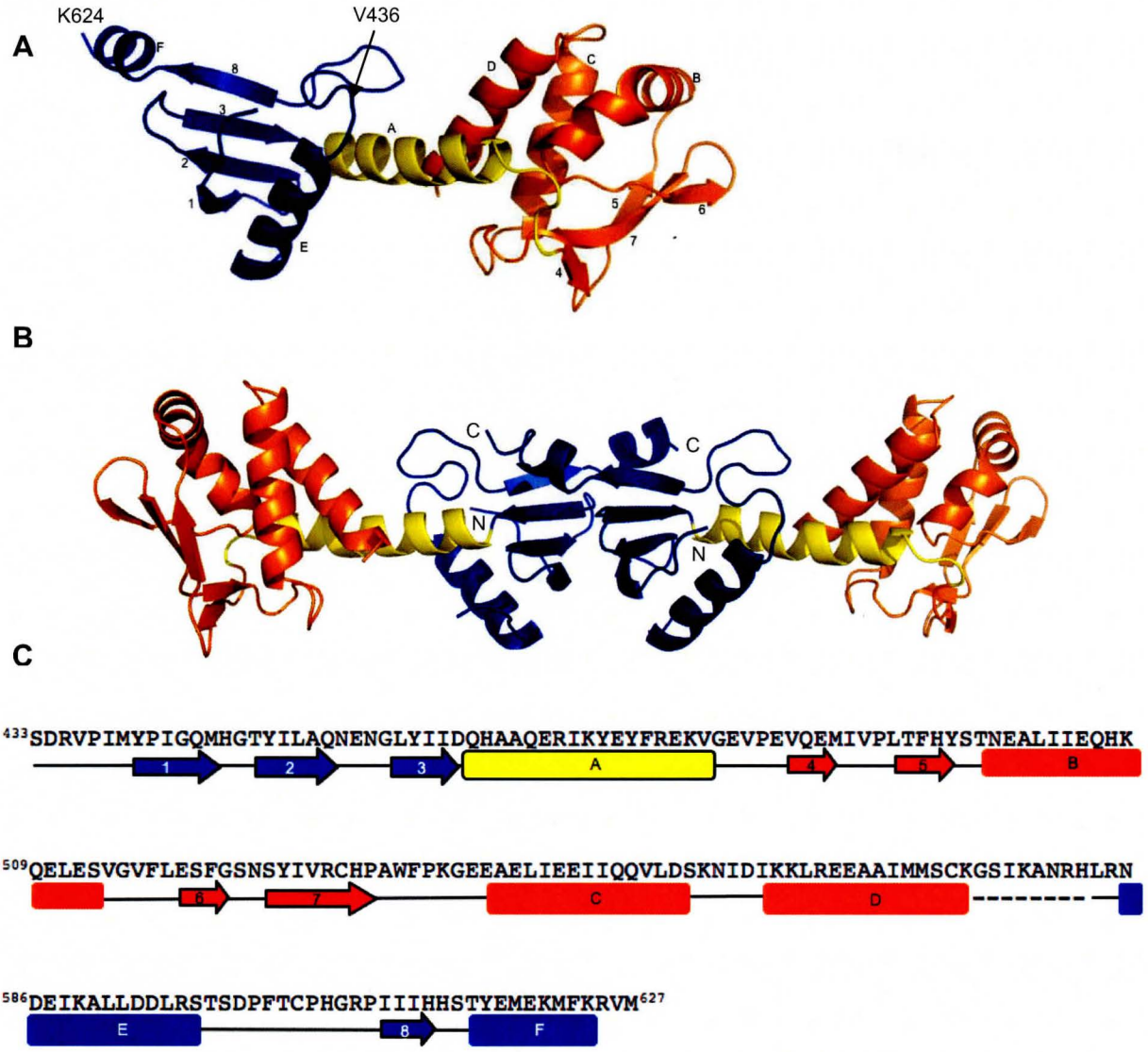


Figure 10: A – Structure of BsMutL-CTD with Outer subdomain (shown in orange) and Dimerization subdomain (shown in blue). B – Dimerization of BsMutL-CTD as seen through symmetry. C - Alignment view of BsMutL-CTD structure with strands shown as arrows and helices shown as cylinders. Subdomains are coloured as in A.

3.3.2 BsMutL-CTD is a dimer

The dimer of BsMutL-CTD is revealed through symmetry. The dimer interface is formed through the Dimerization subdomain with β strands β1-3 (residues 440-460) as well as β8 (610-614) and αF (617-624) mediating the interaction. Together, the dimer has an occluded surface area of 3414 Å, suggesting a stable interface. Interestingly, the four monomers form three distinct dimers, each with a slightly different conformation as judged by superimposition (Figure 11). The first dimer, comprised of two A monomers, called the Open dimer, represents the most open or extended conformation with respect to the two subdomains. The second dimer, comprised of two D monomers, called the Closed dimer and involves the Outer subdomain lifting and coming closer to the Dimerization subdomain. The third dimer comprised of one B monomer and one C monomer, called the Intermediate dimer, represents an intermediate of the two conformations. In order to judge these conformational changes, each monomer was superimposed onto the Open dimer, using residues which form the dimerization interface (440-460), located in the Dimerization subdomain. The visual impression of near identity of the dimerization domains is seen by the near identical root mean square deviation (rmsd) between dimers, however larger differences can be attributed to flexibility contributed by the lever to the outer subdomain. The orientation of the Outer subdomain varies between dimers. This is illustrated by the increase in the rms values when using all residues to superimpose instead of just those located in the Dimerization domain (Figure 11).

A

Residues used	Monomer	rmsd
440-460	Closed → Open	0.563 Å
436-624	Closed → Open	1.586 Å
440-460	Intermediate (B) → Open	0.654 Å
440-460	Intermediate (C) → Open	0.574 Å

B

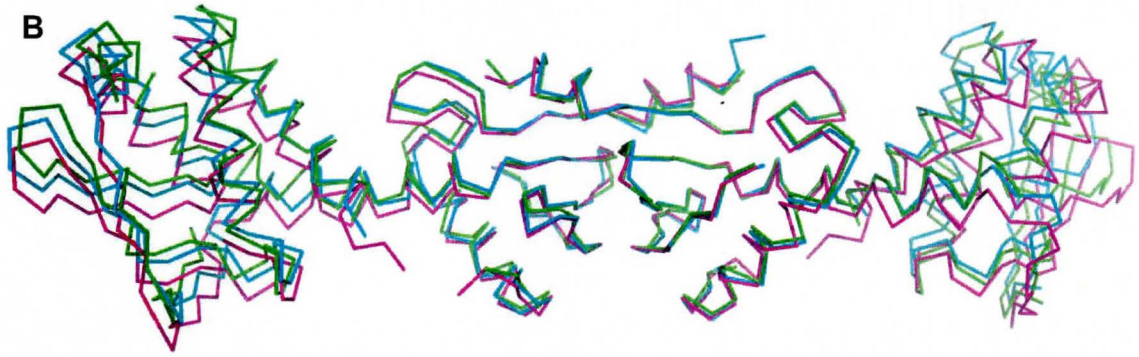


Figure 11: Superimposition of three dimers found in BsMutL-CTD structure. A – Chart of rmsd between C-α trace for different conformations of BsMutL dimers. B - Superimposition of three dimers found in BsMutL-CTD structure with Closed dimer (DD) shown in purple, Open dimer shown in green (AA) and Intermediate (BC) dimer shown in blue.

3.3.3 BsMutL-CTD binds zinc

Recently, several conserved motifs have been found in MutL homologues which lack MutH, including BsMutL, but not in *E. coli* (46). Three of these motifs ($DQHA(X)_2E(X)_4E$, FKR, CPHGRP) are located in the Dimerization subdomain, while the other is found in the Outer subdomain (SCK). Interestingly, looking at the placement of these motifs on the BsMutL-CTD structure, they all are close in space and collectively, the spatial arrangement of these motifs suggests a common activity related to the endonuclease motif (Figure 12). Several point mutants of putative metal binding residues were created by Ryan Mitchell in order to test which residues were critical for ion binding as this activity is necessary for

endonuclease function. These mutants were sent to our collaborator Dr. Peter Friedhoff (Institut für Biochemie, Justus-Liebig-Universität, Germany), and revealed that BsMutL does indeed possess metal ion binding capabilities, specifically of the divalent ion zinc. Using a zinc-specific fluorescent indicator dye, fluorescence was measured by combining the protein with zinc and comparing it to a known amount of zinc. If a protein is stably binding a zinc ion, then the fluorescence will be much lower as the dye will not be able to bind it. Using the curve generated from the known standards, an estimate of the amount of zinc binding to the protein can be made. Upon mutation, several conserved residues were found to impair the ability of the protein to bind zinc, both C604, and H606, found in the conserved CPHGRP motif (Figure 12). The double mutant of C604 and H606 showed the largest effect, which is expected since the CPHGRP motif has been previously proposed to be involved in metal binding activity (23). D462 and E473 were not found to be involved in the binding activity and showed similar binding to wild-type.

A closer examination of the orientation of these conserved residues on the structure reveals two putative metal binding sites, although no metal ion was found in the structure (Figure 12).

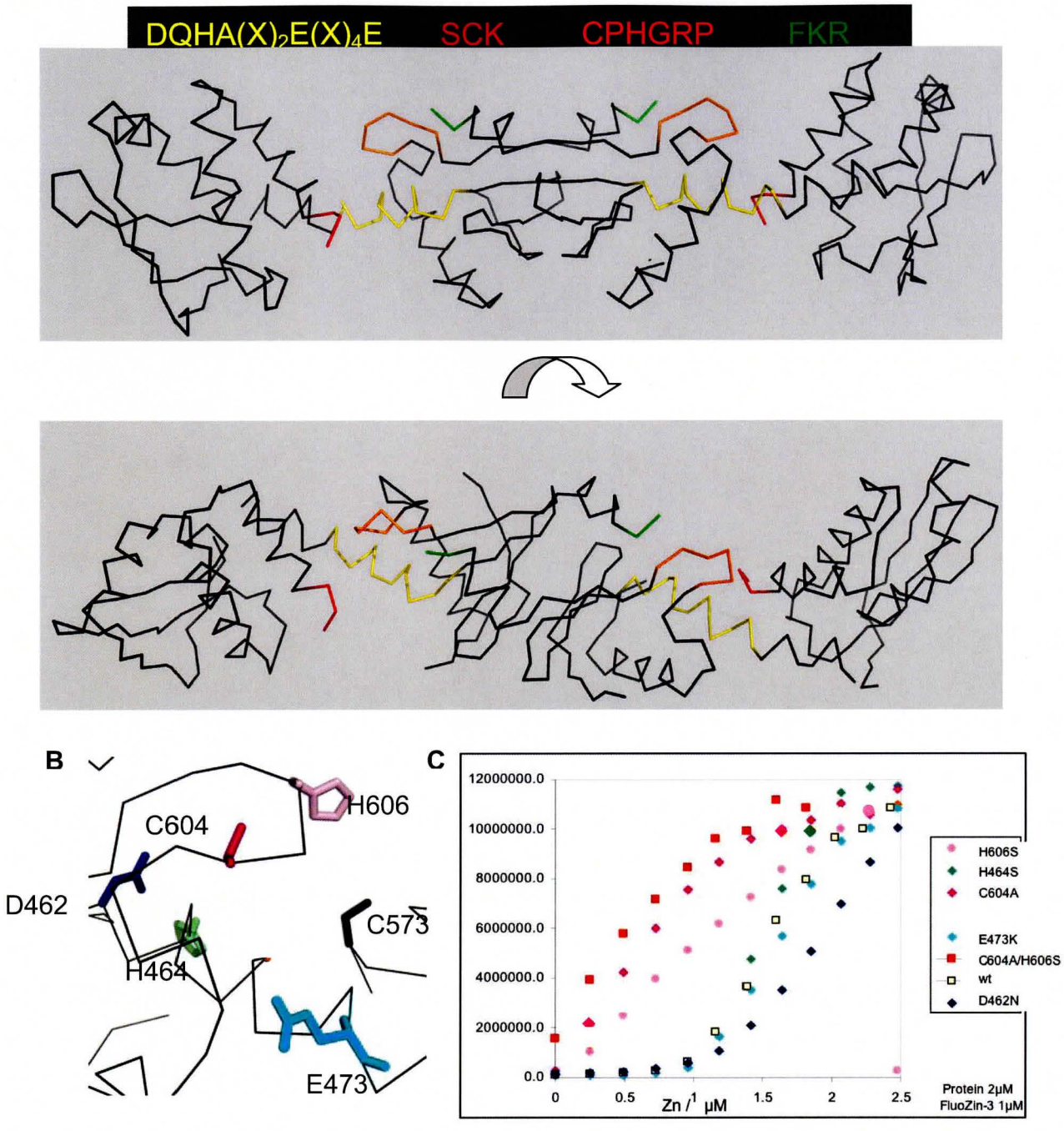


Figure 12: Conserved motifs cluster in space together. A – BsMutL-CTD dimer shown as C-α in black, with conserved sequences shown in yellow, orange, red and green. B – Putative metal binding site of BsMutL. C - Zinc release data for BsMutL mutants. Experiment performed by Dr. P. Friedhoff.

3.3.4 Comparison to E. coli MutL

The overall architecture of the dimer is quite similar to *E. coli* MutL, as predicted (21,47). In *E. coli,* the dimer interface of MutL has been shown to be formed via interactions of residues 437-459 (β1-3) and 598-612 (β8, αG) (47). In order to understand the similarities of the dimer interface between the two structures, BsMutL-CTD (residues 440-460) were superimposed with *E. coli* MutL (residues 437-457) (Figure 13). The root mean square deviation (rmsd) is 1.072 Å, indicating a similar interface.

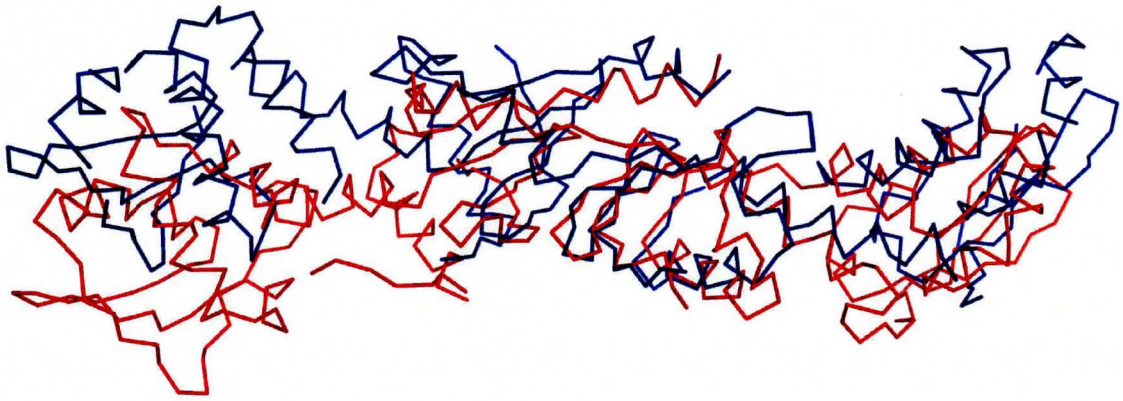


Figure 13: Superimposition of Cα trace of *E. coli* MutL (shown in red) and BsMutL-CTD (shown in blue).

A superimposition of the Outer subdomain from residues 540-570 of BsMutL with residues 534-564 of *E. coli* MutL reveals an rmsd of 2.093 Å. This shows that the Outer subdomains are more divergent than the Dimerization subdomains. Taking these two results together, it appears that the two domains individually are similar but the orientation between them causes greater differences in the overall structure.

3.4 Endonuclease Activity of BsMutL-CTD

In order to determine whether the putative endonuclease motif in BsMutL-CTD functioned similarly to the human and yeast MutLα, an endonuclease assay was performed. MutL has been shown to be a metal-ion dependent endonuclease. BsMutL-CTD was shown to not be an endonuclease as shown in Figure 14. In the presence of several divalent ions, BsMutL-CTD was unable to nick a supercoiled substrate.

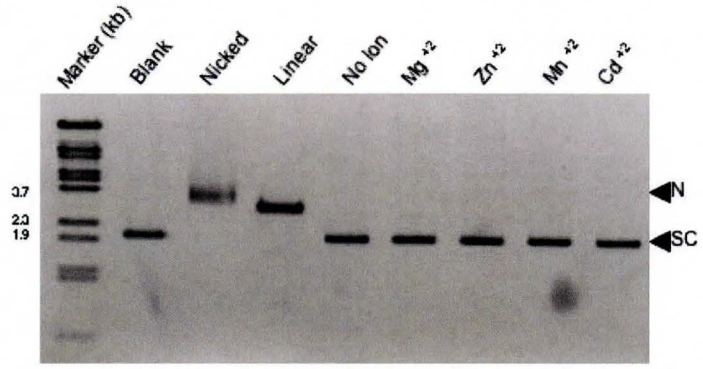


Figure 14: Endonuclease assay of BsMutL-CTD. Endonuclease activity of BsMutL-CTD on supercoiled DNA as described in Experimental Procedures. Where indicated, divalent ion present at 5 mM and BsMutL-CTD present at 80 nM. Arrows indicate the presence of supercoiled (SC) or nicked (N) DNA.

3.5 Purification of BsMutL

3.5.1 IMAC and Anion Exchange

Purification of BsMutL was accomplished using three purification steps of IMAC, anion exchange and size exclusion chromatography (Figure 15). The purification of BsMutL was much more challenging, and the addition of protease inhibitors to all buffers was essential in order to obtain pure, homogenous preparations of BsMutL.

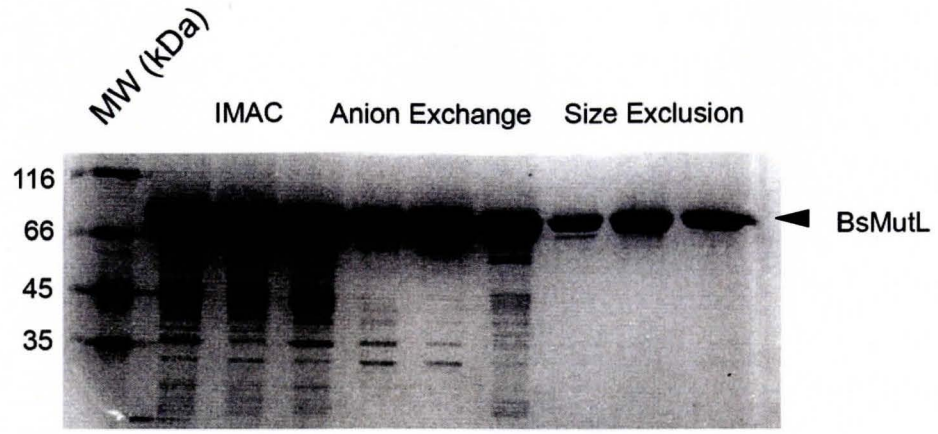


Figure 15: Purification of BsMutL. 12% SDS PAG showing the product after each chromatography step: IMAC, anion exchange and size exclusion.

3.4.2 Oligomeric state of full-length BsMutL

In order to further purify and determine the oligomeric state of BsMutL, size exclusion chromatography was used. Although BsMutL as a dimer has a molecular weight of 140 kDa, it eluted from a size exclusion column as a globular protein of approximately 316 kDa (Figure 16). Previously it has been shown that *E. coli* MutL exists as a dimer in solution with a similar elution profile (17). It is not clear from this data alone if BsMutL exists as a dimer or tetramer.

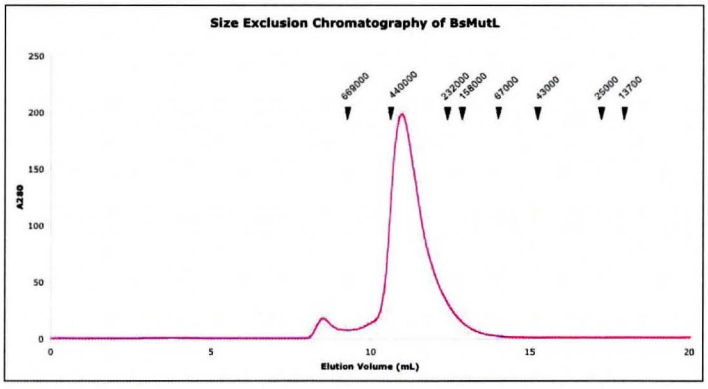


Figure 16: Size exclusion chromatography profile of BsMutL over a Superdex-200 column (GE Healthcare Inc). Elution volumes of molecular weight markers (kDa) are indicated by arrows.

3.6 Endonuclease Activity of BsMutL

BsMutL was assessed for endonuclease activity in the presence of a variety of divalent cations. The endonuclease activity of BsMutL was found to be dependent on several divalent ions (Figure 17). Manganese showed the largest effect, which has been demonstrated previously in several MutL orthologues (39-41).

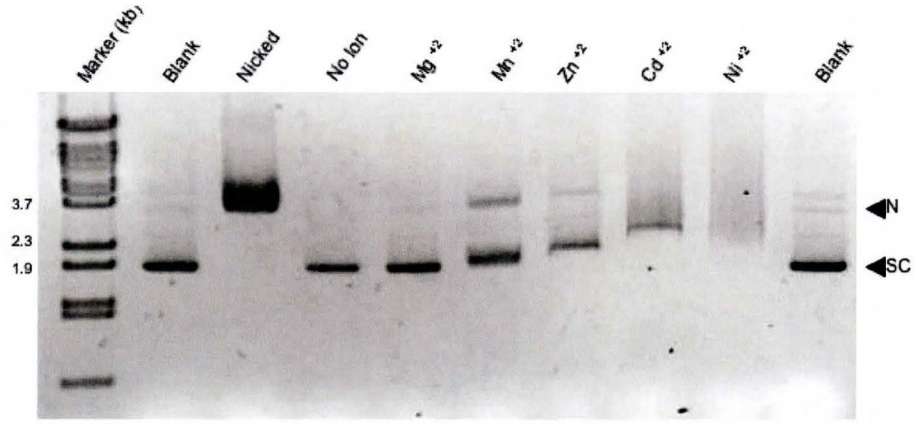


Figure 17: Divalent ion screen of BsMutL. Endonuclease activity of BsMutL on supercoiled DNA in the presence of several divalent ions as described in Experimental Procedures. Where indicated, divalent ion present at 5 mM and BsMutL present at 80 nM. Arrows indicate the presence of supercoiled (SC) or nicked (N) DNA.

When a manganese titration was conducted, a proportional increase in nicking activity is seen. Approximately 15% of the supercoiled plasmid is nicked at concentrations of 5 mM manganese (Figure 18).

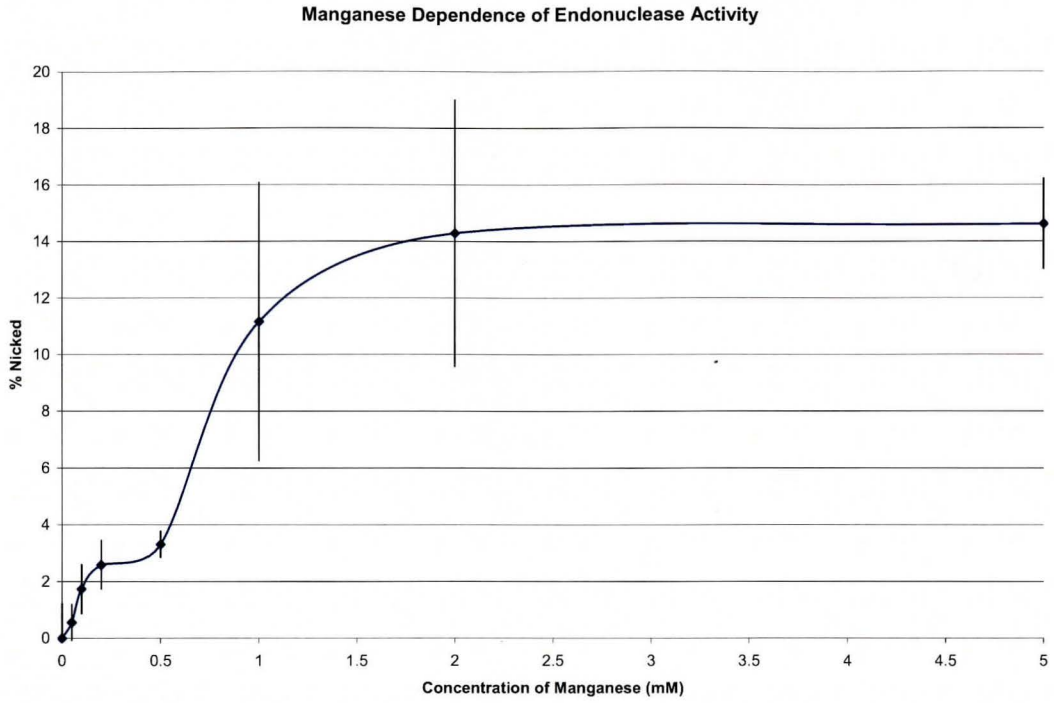


Figure 18: Manganese-dependence of BsMutL endonuclease activity. The percentage of nicked DNA to all supercoiled plasmid DNA was determined using ImageJ software (NIH) and plotted against manganese concentration. Error bars represent variation between three separate independent experiments.

Mutational Analysis

Mutations were made to the putative endonuclease motif (see Figure 5, D462N, E473K) in order to determine which residues were responsible for the endonuclease activity of BsMutL.

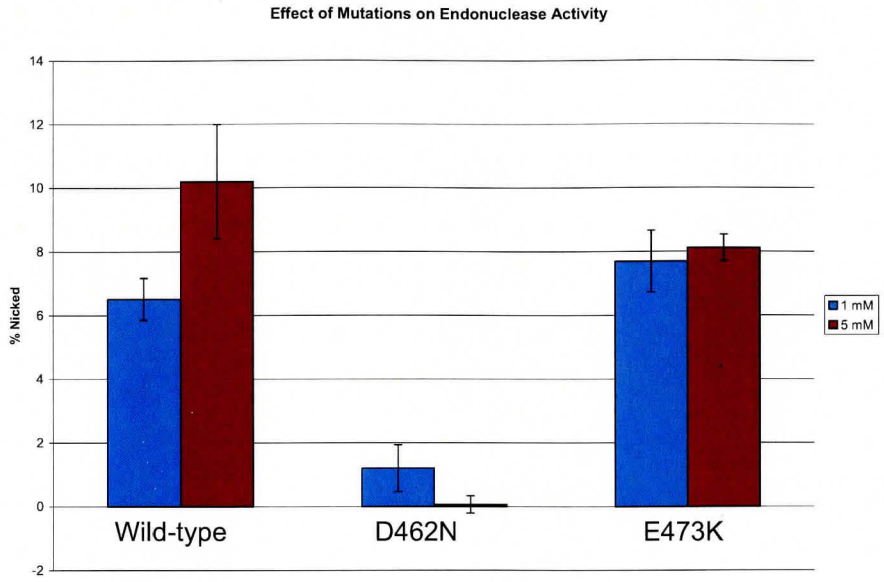


Figure 19: Mutational analysis of endonuclease BsMutL. Endonuclease activity of wild-type MutL, D462N and E473K at 1 mM $MnCl_2$ (in blue bars), and 5 mM $MnCl_2$ (in purple bars). The percentage of nicked DNA to all supercoiled plasmid DNA was determined using ImageJ software (NIH) and plotted against each protein. Error bars represent variation between three separate independent experiments.

Our results indicate that D462 is not involved in zinc ion binding, however, upon mutation to glutamine, it is completely inactive in nicking compared to wild-type (Figure 19). In contrast, E473K showed nicking activity similar to wild-type.

Nucleotide Effect on Endonuclease Activity

MutL belongs to a family of well-conserved ATPases (17). The effect of nucleotides on the regulation of endonuclease activity has been controversial. While work on *Thermus thermophilus* MutL has shown that ATP inhibits endonuclease activity, the opposite has been found for human MutLα where ATP has been shown to stimulate the activity (39, 41). In order to rectify this, the effect of the nucleotide on the endonuclease activity of BsMutL was evaluated. In

order to chelate ATP, magnesium is required. However, in the absence of magnesium, it is possible that other divalent ions may be able to substitute such as manganese. Initially, in the presence of ATP or AMPPnP, but no magnesium, nicking appeared to be effectively inhibited (Figure 20). However, while maintaining constant concentrations of manganese and nucleotide, but increasing the concentration of magnesium to equal that of ATP, nicking is restored to wild-type levels. The presence of ADP was not affected by the concentration of magnesium and retained near wild-type levels. It seems that ATP, ADP and AMPPnP had no clear inhibition or activation of endonuclease activity.

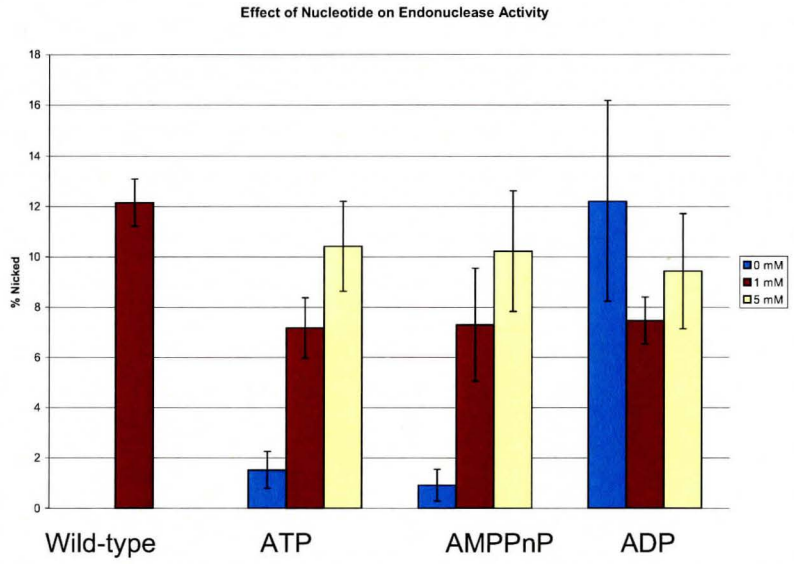


Figure 20: Effect of nucleotides on endonuclease activity of BsMutL. Endonuclease activity of wild-type MutL at 5 mM $MnCl_2$ shown as first purple bar. For the remainder, manganese and nucleotide were held constant at 5 mM. Blue bars indicate 0 mM $MgCl_2$, maroon bars indicate 1 mM $MgCl_2$, yellow bars indicate 5 mM $MgCl_2$. The percentage of nicked DNA to all supercoiled plasmid DNA was determined using ImageJ (NIH) and plotted against each protein. Error bars represent variation between three separate independent experiments.

CHAPTER 3: DISCUSSION

In this work, we have confirmed the endonuclease activity of *Bacillus subtilis* MutL and using X-ray crystallography, solved the structure of the dimerization domain containing the highly conserved endonuclease motif.

There are three different dimers found in the BsMutL-CTD structure (Open, closed or intermediate, Figure 11). We propose that the three conformations offer a snapshot of the protein in the stages leading up to metal ion coordination. Most endonucleases require a divalent ion to function properly (48). No divalent ion is seen bound in the structure, as even the closed dimer does not have the proper distances to allow for divalent ion coordination. Similarly, this has also been shown in other endonucleases, such as MutH, where the divalent ions are only coordinated in the presence of the substrate DNA (26,27). We have shown that BsMutL is capable of binding zinc (Figure 12). Using mutants, the key residues involved in this binding were identified as C604 and H606. These residues are clustered in space together and should be able to form a bona fide metal binding site. Additionally, D462, H464, and E473 are also close in space, but were not shown to be involved in zinc binding. Interestingly, C573 is also located in this region, but was not evaluated for zinc binding activity, as a mutation of this residue could not be made. The distances between key residues suggested to bind a metal ion is approximately 5.5 Å, which is too far apart for most metal binding which is normally 2-2.4 Å (49). This suggests there

may be an even further closed conformation that would allow for coordination to occur. The lever could confer this flexibility to allow for this conformation. Indeed, the presence of the three dimers in the structure confirms that there is variation between the orientation of the Dimerization and the Outer subdomain and so another conformation is plausible.

From a sequence comparison, MutL homologues were found to have similar metal binding sequence to Diphtheria toxin Repressor (DtxR) family of proteins (46). This family of metal-sensing proteins is responsible for binding metal ions and acting as transcriptional repressors on metal ion uptake genes. Recent work on the metal binding activity of the C-terminal domain of human PMS2 suggest that MutL homologues possess a two-metal ion system (46). The prevalence of a two-ion system is not novel; several other nucleases require both a "catalytic" and "structural" ion in order to function. The structural ion is thought to stabilize the structure or cause a conformational change which would allow for the second metal binding site to bind another ion, which is usually catalytic (50). In the case of MutL, both manganese and magnesium have been shown to be necessary for activity *in vitro* and *in vivo* respectively (39). Therefore, the binding of a manganese or zinc ion could allow for the binding of a second divalent ion, which would then allow for endonucleolytic activity. IdeR is capable of binding several divalent ions *in vitro* although the function *in vivo* is attributed to binding iron (51). Similarly, we have shown that BsMutL endonucleolytic activity is dependent on the presence of a variety of divalent ions, such as zinc but excluding magnesium (Figure 17). This indicates the tolerance of the binding site

for a range of divalent ions. The region containing the metal binding motif is also quite flexible, as shown by different orientations of side chains as well as allowing for different angles of the Outer subdomain with respect to the Dimerization subdomain. This flexibility may allow for the coordination of different ions. It is unclear which divalent ion is bound *in vivo;* the concentration of divalent ions in the cell may play a role as to the occupancy of the site. While manganese is virtually absent in a cell, however, magnesium and zinc are more abundant (52).

The excessive manganese may be able to induce a conformational change that is able to mimic the active conformation of MutL *in vivo*. The MutL homologues shown to have endonuclease activity are able to nick DNA non-specifically, as well as in the presence of a series of co-factors. This suggests that MutL is probably being repressed *in vivo* until a mismatch and the proper co-factors are recruited in order to avoid non-specific nicking activity (23). It is not known which proteins or co-factors are responsible for inhibiting MutL, or whether MutL autoinhibits, and requires activation.

Mutational Analysis

The endonuclease motif itself confers several interesting properties. Across several organisms with the conserved motif, when Asp or the first Glu (D462 in BsMutL) are mutated, repair is impaired both *in vitro* and *in vivo* (39,40,53,54). Similar to other studies, our results indicate that D462 is a critical residue for endonuclease activity as they are unable to nick a supercoiled substrate in the presence of manganese when mutated (Figure 19). Mutations to

E473 showed no decrease in nicking activity. This inactivation could initially be attributed to a deficiency in divalent ion binding. Interestingly, our data indicates both D462 and E473 are not involved in zinc binding (Figure 12).

It is interesting to note that both D462 and E473 were not involved in binding zinc. Contrastingly, D462 showed no nicking activity, while E473 retained activity similar to wild-type (Figure 19). From the crystal structure of BsMutL-CTD, D462 is the furthest of the conserved residues from the first putative metal binding site, in close contact with H464. It is possible that both are involved in the second binding site, potentially of a divalent ion (Figure 21); this is not yet clear. The structure of BsMutL bound to a divalent ion is an absolute necessity to elucidate which residues are actually involved in metal ion coordination and to confirm that BsMutL actually employs a two metal ion system similar to IdeR. With no ion bound it is only possible to model where a putative divalent ion would be, however no claim can be made as to exactly how binding is occurring.

Despite being highly conserved, E473 retained its endonuclease activity upon mutation. This suggests that the residue is not critical for endonuclease activity *in vitro.* The conservation of this residue, however, seems to suggest a critical role for MutL homologues. It may be involved in another facet of the roles of MutL, such as meiosis or initiating DNA damage response and triggering cell cycle arrest. From the structure, it appears that E473 is solvent exposed and facing the opposite direction of the proposed metal coordination site. It is however, hydrogen bonded to neighbouring residues (S599) and may be involved in providing integrity to the structure.

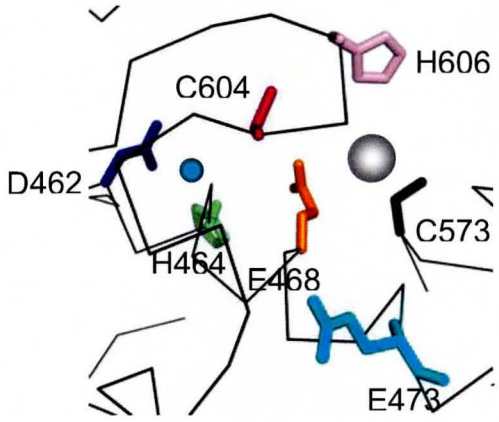


Figure 21: Putative metal binding site with metal 1 (shown in white) and metal 2 (shown in blue) placed in.

Not only is mismatch repair abolished upon mutation of certain residues in the endonuclease motif, but other processes MutL is involved in are effected (53). This implies that the endonuclease site is not only required for protecting against mutations in mismatch repair, but also functions in DNA damage response pathways and suppressing homeologous recombination. Indeed, this is also the case for another MutL homologue, MLH3, where mutations made to the endonuclease site confers defects in meiosis in yeast (55).

It should be noted that the location of the endonuclease motif is in the C-terminal domain of MutL homologues. It has been proposed that mutations in this region may also affect interactions with other proteins and this may explain why the defect is seen in other pathways (53). We do not see this as a likely possibility, as steric hindrance and lack of space around the motif make it an unlikely candidate for protein-protein interactions.

The role of nucleotides has been implicated in the regulation of the endonuclease activity of MutL (41). It has been suggested that ATP is used to suppress the non-specific nicking of MutL *in vivo.* This claim is based on experiments which show that increasing concentrations of ATP abolished the nicking of *T. thermophilus* MutL (ttMutL), while ADP does not (41). In contrast, it has been shown that ATP stimulates the nicking ability of MutLα in both yeast and humans (39,40). In order to address this nucleotide effect, we performed endonuclease assays in the presence of ATP, AMPPnP and ADP. Initially, similar results to ttMutL were obtained for BsMutL, showing that ATP appeared to be reducing the activity of BsMutL in the presence of manganese (data not shown). It is interesting to note that these experiments with ttMutL were performed in the presence of manganese, but not in the presence of magnesium, which is required for ATP to be properly bound. It is possible that ATP is able to chelate the manganese necessary for *in vitro* function away from MutL and thus appears to be inhibiting the activity. Contrastingly, we find that in the presence of excess magnesium, that nicking activity is restored to wild-type levels for BsMutL (Figure 20). The presence of ADP does not appear to have any effect on the activity. However, an increase in the endonuclease activity is not seen in the presence of ATP or AMPPnP. This suggests that ATP is neither inhibiting nor stimulating the endonuclease activity of BsMutL.

A link between the metal ion binding motif of MutL and the ATPase activity has been suggested. When a well-conserved cysteine of ttMutL is modified

(C496, corresponding to C604 in BsMutL), the endonuclease activity is retained. However, the mutant was no longer sensitive to the presence of AMPPnP, which suggested that the binding of ATP is somehow regulated by the presence of a metal binding motif, particularly the conserved cysteine. In order to discover the relationship between the ATPase activity and its regulatory effect on the endonuclease activity, Dr. Alba Guarné created several mutants hypothesized to cause defects in ATPase activity based on mutations previously made to *E. coli* MutL. These mutants should be assessed for affinity for nucleotides, as well as endonuclease activity, which will shed some light onto whether the regulation system described in *T. thermophilus* applies to *Bacillus subtilis* as well.

The structure of the N-terminus of *E. coli* MutL revealed a unique fold shared by members of an ATPase family, which lead to the discovery of its function as a weak ATPase. In *E. coli* MutL the N-terminal domains (NTDs) are able to associate in the presence of ATP or a non-hydrolyzable analog. In both yeast and human MutL homologues, the NTDs exist as monomers even in the presence of ATP (56,57). Indeed, human PMS2 is able to hydrolyze ATP as a monomer (56). In yeast MutLα, MLH1 has been shown to have a higher affinity for ATP than PMS1 and that large conformational changes are induced by the presence of nucleotides (58). This change in conformations is shown very nicely with atomic force microscopy (AFM) (59). MutLα was shown to exist in four distinct conformations, ranging from extended in the absence of nucleotide to condensed in the presence of high concentrations of ATP. A similar effect is also

seen in the presence of ADP suggesting that ATP hydrolysis is not necessary for it to occur. However, mutations to ATP hydrolysis have been shown to abolish the ability of MutL to function in 3' break-directed incision, suggesting that hydrolysis is necessary for function (39). Combining the results of the AFM work with the increased endonuclease activity seen in the presence of ATP, it is possible to suggest that the ATPase cycle is necessary to stimulate conformational changes which can be linked to both DNA binding and subsequently endonuclease activity. In the case of BsMutL, preliminary data has shown that it is indeed an ATPase as expected (experiments performed by Dr. Alba Guarné and Ryan Mitchell, data not shown).

MutL is able to bind DNA in a sequence and mismatch independent manner, with a preference for ssDNA (21). The *E. coli* MutL N-terminal domain structure with bound AMPPnP reveals a patch of positively charged residues inside the two dimers and shown to be important for DNA binding, including R266, R177 and K159 (17) (Figure 4). Interestingly, these residues are not conserved in BsMutL. There is however a patch of three positively charged residues (K302, R306, K309) which when mapped on to the *E. coli* MutL NTD structure they would lie on an alpha helix, posed to bind DNA on the outside of the structure instead of inside, as proposed in *E. coli*. This difference in DNA binding may shed some light on why BsMutL is able to function as an endonuclease, while *E. coli* MutL cannot (40).

While the residues necessary for DNA binding are not conserved from *E. coli* to humans or *Bacillus subtilis*, the ability to bind DNA does. Several MutL homologues have been shown to be able to bind both ssDNA and dsDNA. The binding of ssDNA has been suggested to be involved in facilitating strand excision (19) while dsDNA binding may be used to signal the location of the mismatch and marking the termination of strand excision (57). In yeast MutLα (MLH1 and PMS1), each subunit is able to bind DNA independently and with different affinities (57). Similarly, the N-terminus of human PMS2 has been shown to bind DNA as a monomer (56). This DNA binding has been shown to be critical for repair as mutations causing reduced DNA binding of MutL cause mutator phenotypes *in vivo* (57). One model proposes that the two different binding abilities of MutL can be used simultaneously to stay bound near a mismatch as well as provide a signal to the strand discrimination signal, thus joining together distant regions of homoduplex DNA (57). The evaluation of DNA binding of both BsMutL-CTD and BsMutL has not yet been demonstrated, however it is expected that BsMutL-CTD will not be able to bind DNA (Figure 14) since it was not proficient in endonucleolytic activity, as well as the hypothesis that DNA will bind at the N-terminal domain. BsMutL is expected to bind homoduplex DNA in a sequence-independent manner as it is able to nick DNA non-specifically (Figure 18).

In organisms lacking dam methylase, the mechanism of strand discrimination is still unclear. How does the template strand differ from the newly

synthesized strand? Even more so, how MutL is able to differentiate between them *in vivo* is not clear. In humans, where MutLα is a heterodimer (comprised of MLH1 and PMS2, PMS1 in yeast), the asymmetry of the nicking activity is clear as the endonuclease motif is present in PMS2 but not MLH1 (39). In prokaryotic systems, MutL functions as a homodimer and contains two metal binding sites and endonuclease sites. It is not yet known whether MutL is nicking one strand of DNA or both. When *E. coli* or *T. aquaticus* MutS binds DNA it does so asymmetrically, despite functioning as a homodimer (Figure 2) (14) (13). Perhaps the mechanism of MutL endonuclease activity is similar in principle to MutS. The structure of MutL interacting with DNA will provide some evidence as to the asymmetry of the system.

CHAPTER 6: CONCLUSION & FUTURE WORK

Conclusions

Crystal structures are often necessary to fully understand the mechanism behind biological processes. In this work, the crystal structure of BsMutL-CTD provides a snapshot of the protein in an open conformation, waiting to bind a divalent ion. We also showed that BsMutL is functioning endonuclease *in vitro* and mutational analysis unveiled key catalytic residues that are involved in the activity. Taking the structural data in combination with the biochemical data, it can be seen that *Bacillus subtilis* MutL is an endonuclease in a metal ion-dependent manner similar to humans and yeast.

Future directions

It is tempting to assign roles to BsMutL based merely on sequence homology with other MutL homologues. For example, the DNA binding activity of BsMutL to both ssDNA and dsDNA remains to be characterized. The dimerization interface proposed can also be confirmed by mutational analysis (particularly of helix αF, Figure 10) similarly to how the *E. coli* dimerization interface was elucidated (47).

In terms of structural analysis, it will interesting to see the structure with a metal bound is key to understanding structural conformation which allow for endonuclease activity to take place. This would confirm the hypothesis that MutL utilizes a two-ion system similar to IdeR. Also it would provide another dimension

to the three dimers seen in this structure. According to our data, BsMutL-CTD bound to a metal should retain similar domain architecture while adopting a more closed conformation in the region shown in Figure 21.

In order to fully understand the asymmetry of MutL as an endonuclease, a crystal structure of MutL bound to DNA is necessary. This would provide great detail as to how the homodimer containing two endonuclease motifs is able to only cleave one strand of DNA. This is a challenging task, particularly if BsMutL-CTD is found not to bind DNA. However, there are ways to tether a protein to DNA in order to mimic its' native conformation. This can be accomplished by expressing a DNA binding protein in tandem with BsMutL-CTD which will serve to anchor the protein to DNA.

The combination of biochemistry and X-ray crystallographic studies will provide a detailed view of the role of MutL in *Bacillus subtilis* MMR.

REFERENCES

1. Augusto-Pinto, L., da Silva, C. G., Lopes Dde, O., Machado-Silva, A., and Machado, C. R. (2003) *Genet Mol Res* **2**, 77-91
2. Parker, B. O., and Marinus, M. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1730-1734
3. Stojic, L., Brun, R., and Jiricny, J. (2004) *DNA Repair (Amst)* **3**, 1091-1101
4. Modrich, P. (1991) *Annu Rev Genet* **25**, 229-253
5. Bronner, C. (1994) *Nature* **368**, 258-261
6. Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. (1993) *Cell* **75**, 1027-1038
7. Soreide, K. (2007) *Tumour Biol* **28**, 290-300
8. Lahue, R. S., Au, K. G., and Modrich, P. (1989) *Science* **245**, 160-164
9. Grilley, M., Griffith, J., and Modrich, P. (1993) *J Biol Chem* **268**, 11830-11837
10. Cooper, D. L., Lahue, R. S., and Modrich, P. (1993) *J Biol Chem* **268**, 11823-11829
11. Lopez de Saro, F. J., and O'Donnell, M. (2001) *Proc Natl Acad Sci U S A* **98**, 8376-8380
12. Lopez de Saro, F. J., Marinus, M. G., Modrich, P., and O'Donnell, M. (2006) *J Biol Chem* **281**, 14340-14349
13. Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000) *Nature* **407**, 703-710

14. Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N., and Sixma, T. K. (2000) *Nature* **407**, 711-717

15. Modrich, P., and Lahue, R. (1996) *Annu Rev Biochem* **65**, 101-133

16. Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P., and Yang, W. (2001) *Mol Cell* **7**, 1-12

17. Ban, C., Yang, W. (1998) *Cell* **95**, 541-552

18. Dutta, R., and Inouye, M. (2000) *Trends Biochem Sci* **25**, 24-28

19. Ban, C., Junop, M., and Yang, W. (1999) *Cell* **97**, 85-97

20. Junop, M. S., Yang, W., Funchain, P., Clendenin, W., and Miller, J. H. (2003) *DNA Repair (Amst)* **2**, 387-405

21. Guarne, A., Ramon-Maiques, S., Wolff, E. M., Ghirlando, R., Hu, X., Miller, J. H., and Yang, W. (2004) *Embo J* **23**, 4134-4145

22. Bende, S. M., and Grafstrom, R. H. (1991) *Nucleic Acids Res* **19**, 1549-1555

23. Yang, W. (2007) *DNA Repair (Amst)* **6**, 135-139

24. Welsh, K. M., Lu, A. L., Clark, S., and Modrich, P. (1987) *J Biol Chem* **262**, 15624-15629

25. Au, K. G., Welsh, K., and Modrich, P. (1992) *J Biol Chem* **267**, 12142-12148

26. Ban, C., and Yang, W. (1998) *Embo J* **17**, 1526-1534

27. Lee, J. Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D. N., and Yang, W. (2005) *Mol Cell* **20**, 155-166

28. Dao, V., and Modrich, P. (1998) *J Biol Chem* **273**, 9202-9207

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

30. Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., and Griffith, J. D. (1997) *Embo J* **16**, 4467-4476

31. Blackwell, L. J., Martik, D., Bjornson, K. P., Bjornson, E. S., and Modrich, P. (1998) *J Biol Chem* **273**, 32055-32062

32. Gradia, S., Acharya, S., and Fishel, R. (1997) *Cell* **91**, 995-1005

33. Pluciennik, A., and Modrich, P. (2007) *Proc Natl Acad Sci U S A* **104**, 12709-12713

34. Kolodner, R. D., Mendillo, M. L., and Putnam, C. D. (2007) *Proc Natl Acad Sci U S A* **104**, 12953-12954

35. Schofield, M. J., Nayak, S., Scott, T. H., Du, C., and Hsieh, P. (2001) *J Biol Chem* **276**, 28291-28299

36. Blackwell, L. J., Wang, S., and Modrich, P. (2001) *J Biol Chem* **276**, 33233-33240

37. Genschel, J., Bazemore, L. R., and Modrich, P. (2002) *J Biol Chem* **277**, 13302-13311

38. Dzantiev, L., Constantin, N., Genschel, J., Iyer, R. R., Burgers, P. M., and Modrich, P. (2004) *Mol Cell* **15**, 31-41

39. Kadyrov, F. A., Dzantiev, L., Constantin, N., and Modrich, P. (2006) *Cell* **126**, 297-308

40. Kadyrov, F. A., Holmes, S. F., Arana, M. E., Lukianova, O. A., O'Donnell, M., Kunkel, T. A., and Modrich, P. (2007) *J Biol Chem* **282**, 37181-37190

41. Fukui, K., Nishida, M., Nakagawa, N., Masui, R., and Kuramitsu, S. (2008) *J Biol Chem* **283**, 12136-12145

42. Otwinowski, Z., and Minor, W. (1997) *Methods in Enzymology* **276**, 307-326

43. Terwilliger, T. C. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1937-1940

44. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) *Embo J* **9**, 1665-1672

45. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* **55**, 849-861

46. Kosinski, J., Plotz, G., Guarne, A., Bujnicki, J. M., and Friedhoff, P. (2008) *J Mol Biol*

47. Kosinski, J., Steindorf, I., Bujnicki, J. M., Giron-Monzon, L., and Friedhoff, P. (2005) *J Mol Biol* **351**, 895-909

48. Galburt, E. A., and Stoddard, B. L. (2002) *Biochemistry* **41**, 13851-13860

49. Harding, M. M. (2001) *Acta Crystallogr D Biol Crystallogr* **57**, 401-411

50. Chou, C. J., Wisedchaisri, G., Monfeli, R. R., Oram, D. M., Holmes, R. K., Hol, W. G., and Beeson, C. (2004) *J Biol Chem* **279**, 53554-53561

51. Schmitt, M. P., Predich, M., Doukhan, L., Smith, I., and Holmes, R. K. (1995) *Infect Immun* **63**, 4284-4289

52. Ash, D. E., and Schramm, V. L. (1982) *J Biol Chem* **257**, 9261-9264

53. Erdeniz, N., Nguyen, M., Deschenes, S. M., and Liskay, R. M. (2007) *DNA Repair (Amst)* **6**, 1463-1470

54. Deschenes, S. M., Tomer, G., Nguyen, M., Erdeniz, N., Juba, N. C., Sepulveda, N., Pisani, J. E., and Liskay, R. M. (2007) *Cancer Lett* **249**, 148-156

55. Nishant, K. T., Plys, A. J., and Alani, E. (2008) *Genetics* **179**, 747-755

56. Guarne, A., Junop, M. S., and Yang, W. (2001) *Embo J* **20**, 5521-5531

57. Hall, M. C., Shcherbakova, P. V., Fortune, J. M., Borchers, C. H., Dial, J. M., Tomer, K. B., and Kunkel, T. A. (2003) *Nucleic Acids Res* **31**, 2025-2034

58. Hall, M. C., Shcherbakova, P. V., and Kunkel, T. A. (2002) *J Biol Chem* **277**, 3673-3679

59. Sacho, E. J., Kadyrov, F. A., Modrich, P., Kunkel, T. A., and Erie, D. A. (2008) *Mol Cell* **29**, 112-121

APPENDIX

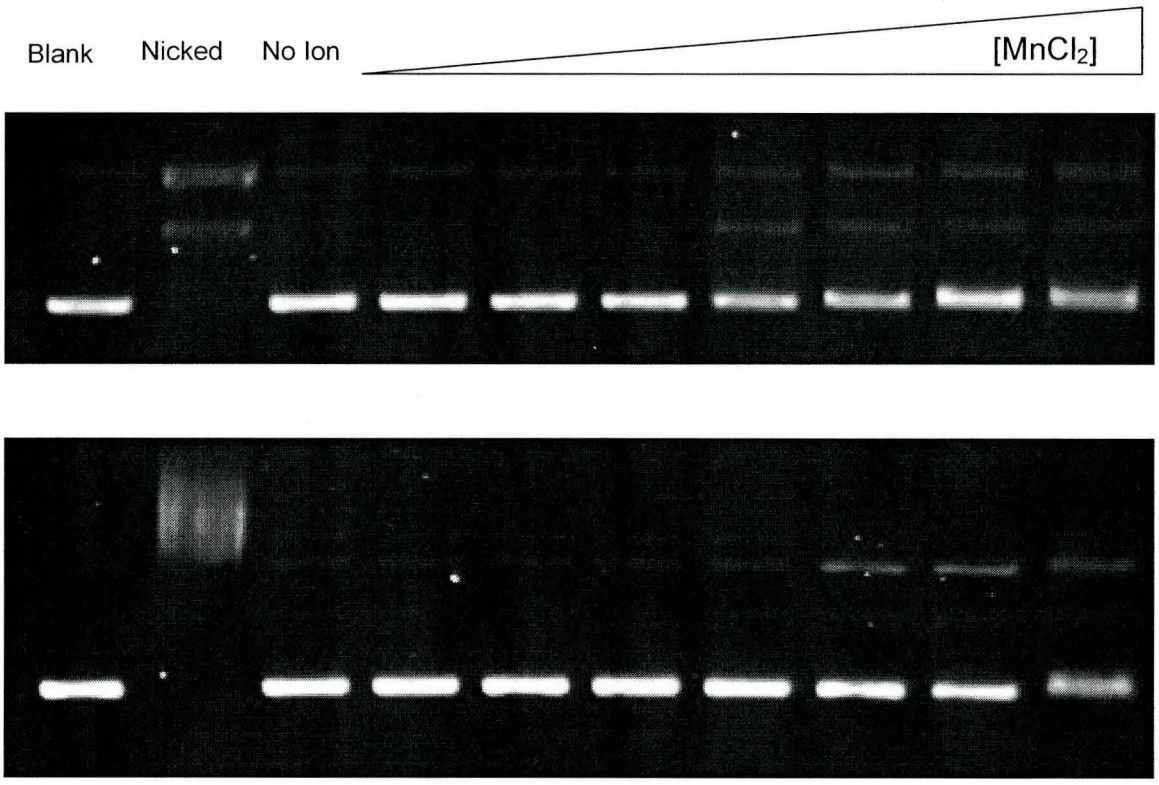
AI: Primers used for cloning BsMutL constructs.

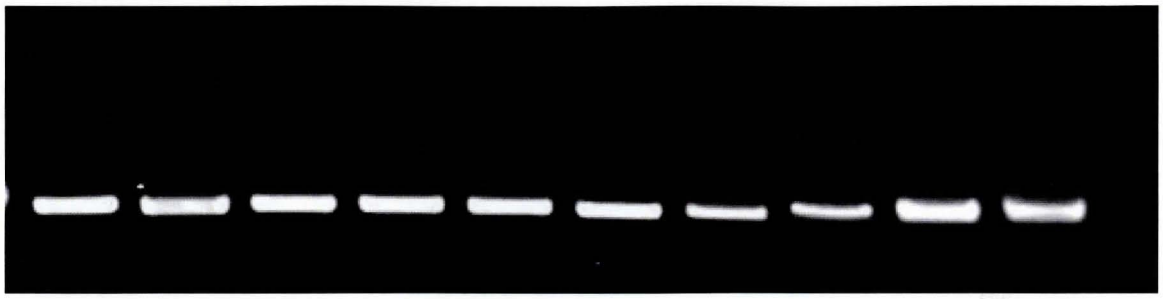
Primers Ag1160/1161 were used to amplify BsMutL-CTD (433-627) and primers Ag1161/1227 were used to amplify BsMutL (1-627). NcoI site shown in blue and XhoI site shown in red.

Primer	Primer Sequence (5' – 3')
Ag1160	GGG GGG **CCA TG G**AT CGG GTT CCA ATT ATG TAC CCG
Ag1161	GGG GGG **CTC GAG** CTA CAT CAC GCG TTT GAA CAT CTT TTC
Ag1227	GGG GG**C CAT GG**C AAA AGT CAT CCA ACT GTC AGA TGA G

A2: Raw data used for quantification of Manganese dependent endonuclease activity (Figure 18).

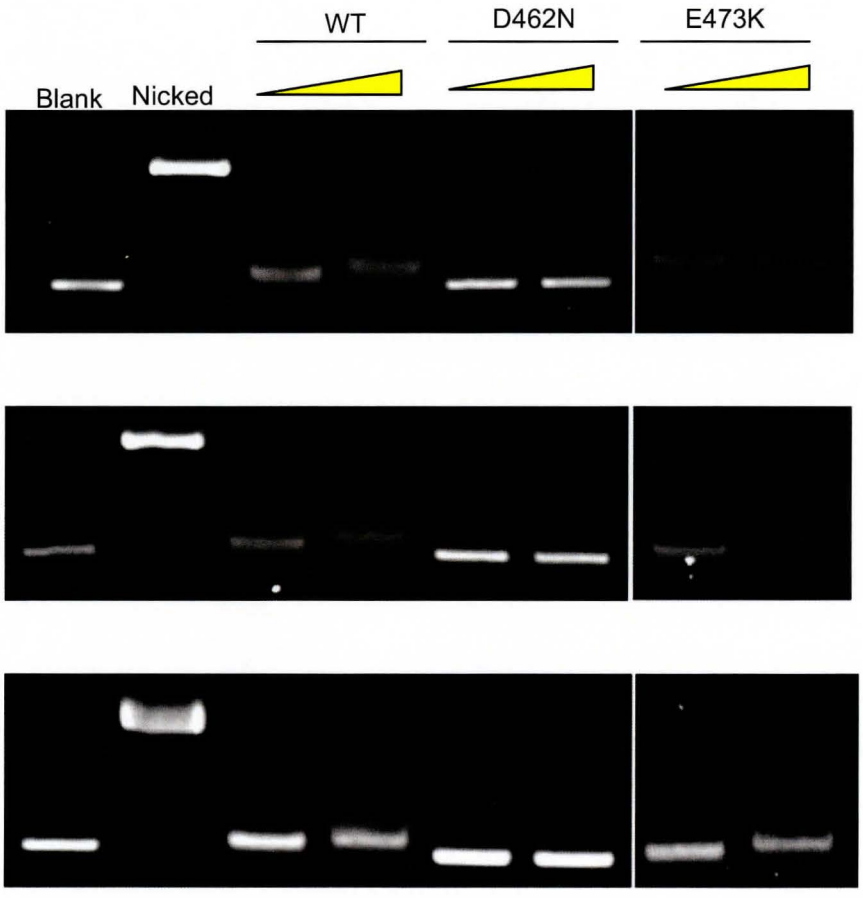
Triplicates of experiments were run on 1% agarose gels, stained with ethidium bromide and destained with water before being visualized under UV light.





A3: Raw data used for quantification of Mutational Analysis on endonuclease activity (Figure 19).

Triplicates of experiments were run on 1% agarose gels, stained with ethidium bromide and destained with water before being visualized under UV light.



A4: Raw data used for quantification of Nucleotide effect on endonuclease activity