MUTL-MEDIATED PROTEIN INTERACTIONS IN DNA MISMATCH REPAIR

CHARACTERIZATION OF MUTL-MEDIATED PROTEIN INTERACTIONS IN DNA MISMATCH REPAIR

By MONICA PILLON, B. Sc. (Hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the

Requirement for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2014) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: Characterization of MutL-mediated Protein Interactions in DNA Mismatch Repair

AUTHOR: Monica Pillon, B. Sc. (McMaster University)

SUPERVISOR: Dr. Alba Guarné

NUMBER OF PAGES: xix, 236

Abstract

DNA encodes the genetic information of the cell, therefore, every single living organism has a precise DNA damage response mechanism to safeguard DNA integrity. Base mismatches are endogenous DNA lesions introduced by the replicative polymerase during DNA replication. The conserved DNA mismatch repair pathway corrects these base mismatches. Mismatch repair initiation is orchestrated by two proteins, MutS and MutL. MutS recognizes and binds to base mismatches and relays the presence of the lesion to MutL. MutL, in turn, interacts with downstream factors to coordinate mismatch excision. The β clamp, typically known for its role in tethering the DNA polymerase to DNA during replication, is also involved in several steps of this repair process including MutL endonuclease activation and strand resynthesis. The dynamics of the MutS-MutL and MutL-β clamp interactions present one of the bottlenecks to uncovering the spatial and time organization of these protein assemblies. Therefore, little is known about the interactions that orchestrate the early steps of mismatch repair. The biochemical and structural work included in this thesis outlines a precise series of molecular cues that activate MutL.

Acknowledgements

I would like to thank my supervisor, Dr. Alba Guarné, for her years of guidance and encouragement. My personal growth throughout my graduate career is a true testament of your dedication. Thank you for challenging me, guiding me, and your unwavering confidence in me. The DNA mismatch repair project required both creativity and stamina. I will always remember fondly our late nights preparing for the synchrotron trip that lead to the structure of MutL bound to zinc.

I would like to thank my committee members Dr. Murray Junop and Dr. Paul Berti for providing valuable advice throughout my graduate studies. As well as, thank you Dr. Lyle Simmons for our productive collaboration.

I would like to acknowledge past and present members of the Guarné and Junop laboratories. Sharing this experience with friends made this journey all the more memorable. I enjoyed our fruitful scientific discussions and silly moments.

I would like to thank my family for their constant support throughout these six years. My parents, Lilianna and David Pillon, always show an interest in my project and are eager to celebrate my academic accomplishments. Thank you to Sylvia and Samantha Pillon for your constant encouragement and advice. I would also like to thank Bartosz Orzel for his enthusiasm towards my academic accomplishments and research.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

2-AP	2-aminopurine
β	Sliding β clamp
β-ME	2-mercaptoethanol
f	Frequency
μ	Mutation rate
σ	Standard deviation
ADP	Adenosine diphosphate
AMPPNP	Adenosine 5′-(β,γ-imido) triphosphate
ATP	Adenosine triphosphate
ATPgS	Adenosine 5'-O-(3-thio) triphosphate
BNL	Brookhaven National Laboratory
BS^3	Bis (sulfosuccinimidyl) suberate
BSA	Bovine serum albumin
Cam	Chloramphenicol
CI	Confidence intervals
COOT	Crystallographic Object-Oriented Toolkit
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside-triphosphate
DPS	Dithiobis [succinimidyl propionate]
DTT	Dithiothreitol
DtxR	Diphtheria toxin repressor
EDTA	Ethylenediaminetetraacetic acid
EXOI	Exonulcease I
FEN-1	Flap endonuclease 1
GFP	Green fluorescent protein
GHKL	Gyrase, Hsp90, Histidine Kinase, MutL
IdeR	Iron-dependent regulator
IDLs	Insertion/deletion loops
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
K_D	Dissociation constant
kDa	kilodaltons
LB	Luria-Bertani
LC MS/MS	Liquid chromatography-tandem mass spectrometry
MAD	Multi-wavelength anomalous diffraction
MIP BOX	Mlh1 interacting protein box
MIS90	90 base-pair G/T mismatch DNA
MLH1	MutL homolog 1
MLH3	MutL homolog 3

Mls	Lincomycin
MMR	Mismatch repair
MMTS	Methyl methanethiosulfonate
MntR	Manganese transport regulator
MSH	MutS homolog
MutH	Mutagenic H
MutL	Mutagenic L
MutS	Mutagenic S
MW	Molecular weight
NSLS	National Synchrotron Light Source
NTD	N-terminal domain
OD_{600}	Optical density (600 nm)
ORI	Origin of replication
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PIP Box	PCNA-interacting motif
PMS2	Postmeiotic segregation increased 2
PMSF	Phenylmethylsulfonyl fluoride
Pol	Polymerase
PXRR	Protein Crystallization Research Resource
SAXS	Small angle X-ray scattering
ScaR	Streptococcal coaggregation regulator
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sel-Met	Selenomethionine
SEM	Standard error of the mean
Spc	Spectinomycin
SSB	Single-strand binding protein
RFC	Replication factor C
RGD	Regulatory domain
Rif	Rifampin
RMR	Relative mutation rate
RMSD	Root-mean-square deviation
RPA	Replication protein A
rNTP	Ribonucleoside-triphosphate
TBS-T	Tris-buffered saline + Tween 20
Tet	Tetracycline
TEV	Tobacco etch virus
tRNA	Transfer ribonucleic acid
V_0	Void volume

Declaration of Academic Achievement

The experiments outlined in this thesis were designed, conducted, and interpreted by myself and Dr. A. Guarné unless specified otherwise in the author preface. *In vivo* mismatch repair assays in *Bacillus subtilis* were performed by Andrew Klocko, Justin Lenhart, and Dr. Lyle Simmons (Chapters 2 and 3). As well as, mismatch repair assays in *Escherichia coli* were conducted by Dr. Jeffrey Miller, Vignesh Babu, and Dr. Mark Sutton (Chapters 4 and 5). The crystal structures of the *Bacillus subtilis* MutL endonuclease domain were solved by Jessica Lorenowicz and myself (Chapter 3). Additionally, small angle X-ray scattering data collection and data processing was conducted by myself (Chapter 5).

Chapter 1

Introduction

1.1 The Discovery of DNA

DNA (or deoxyribonucleic acid) is the hereditary material of the cell and is often referred to as 'the molecule of life.' The discovery of DNA was founded by transforming work that spanned almost 150 years. Gregor Mendel's famous pea plant experiments outlined the laws of heredity in 1865 and marked a shift in biology that exhibited a growing interest in understanding the components inside the cell (Orel and Wood, 2000). Four years later, Friedrich Miescher isolated 'nuclein' (now known as DNA) from the nuclei of leucocytes (Dahm, 2005). It wasn't until 1929 that Phoebus Levene identified the four building blocks (adenine, cytosine, guanine, and thymine) of DNA (Simoni et al., 2002). This marked the beginning of a controversy founded on the notion that a four-component molecule was too simple to encode the heredity material of the cell. It took another fifteen years before Oswald T. Avery, Colin MacLeod and Maclyn McCarty demonstrated that DNA is the genetic material of the cell (Avery et al., 1944).

The interest in DNA instigated a race to solve its structure. The three-dimensional model of the DNA double-strand helix was published by James Watson and Francis Crick in 1953 (Watson and Crick, 1953). This work was accompanied by the X-ray analysis of DNA by Rosalind Franklin (Franklin and Gosling, 1953) and Maurice Wilkins (Wilkins et al., 1953) who confirmed the repetitive helical structure of DNA. The discovery of the DNA structure along with the first DNA polymerase in 1956 (Lehman et al., 1958) paved the way for Matthew Meselson and Franklin Stahl to propose a model for semi-conservative DNA replication (Meselson and Stahl, 1958). Innovations soon followed and included techniques for DNA sequencing (Maxam and Gilbert, 1977; Sanger et al., 1977) and DNA amplification (Mullis et al., 1986). These advances provided the necessary tools for the international Human Genome Project (1990-2001), which successfully sequenced the entire human genome (Lander et al., 2001). These breakthroughs contribute to the foundation for our current understanding of life at the molecular level.

1.2 The Structure of DNA

1.2.1 B-form DNA

DNA is a double-strand helix comprised of phosphates, pentose sugars, and nitrogenous bases. The antiparallel helix is predominantly found as B-form DNA inside the cell. B-form DNA has a helix diameter of 20 Å and requires 10 bases to complete a single turn (Voet and Voet, 2004). Base pair stacking induces a 34˚ pitch (rise per turn) forming periodic major and minor grooves along the DNA helix. The minor groove

exposes the edge of the glycosidic bond between carbon-1 of the pentose sugar and the nitrogenous base (Voet and Voet, 2004). The major groove exposes the opposite edge of the glycosidic bond and is wider and deeper than the minor groove. The nitrogenous bases found at the center of the helix base pair through hydrogen bonds. These base pairs are planar and lie perpendicular to the DNA axis. Therefore, the DNA helix is stabilized by two forces, base pair stacking (π stacking) and, to a lesser extent, hydrogen bonds between complementary base pairs (Voet and Voet, 2004).

1.2.2 Watson-Crick Base Pairing

The genetic information of the cell is encoded in the order of nitrogenous bases. There are four nitrogenous bases in DNA: adenine (A), thymine (T), guanine (G), and cytosine (C) (Levene, 1919). These bases can be subdivided into two groups known as purines and pyrimidines. Purines are large bases comprised of a pyrimidine ring fused to an imidazole ring (adenine and guanine) and pyrimidines are smaller bases with a single carbon-nitrogen aromatic ring (thymine and cytosine) (Figure 1.1). In 1952, Erwin Chargaff uncovered two key rules for DNA base pairing. Independent of the organism, DNA has a consistent 1:1 ratio of purines to pyrimidines (Vischer and Chargaff, 1948). Furthermore, there are proportional amounts of adenine to thymine nucleotides and guanine to cytosine nucleotides (Vischer and Chargaff, 1948). This trend was termed Chargaff's Rule and was the original evidence suggesting that purines strictly pair with pyrimidines to form A-T and G-C base pairs.

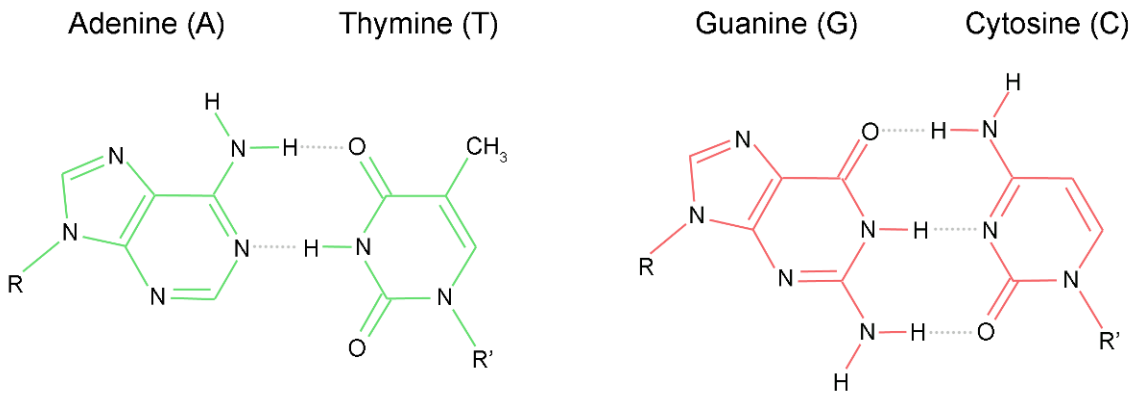


Figure 1.1. Watson-Crick base pairs. Adenine pairs with thymine (green) and guanine pairs with cytosine (red). Hydrogen bonds are shown as grey dotted lines. R and R' mark carbon-1 from the pentose sugar.

The specificity of base pairing is influenced by geometric and chemical constraints. Purine-pyrimidine base pairs are favorable because they can span the diameter of the DNA helix to form specific hydrogen bonds described as Watson-Crick base pairing (Figure 1.1). Purine-purine base pairing would exceed the diameter of the helix causing DNA distortions (Voet and Voet, 2004). Conversely, pyrimidine-pyrimidine base pairs are too small to span the diameter of the helix. Chemical constraints strictly limit purine-pyrimidine pairing to A-T and G-C. Watson-Crick base pairing between adenine and thymine form two hydrogen bonds while guanine-cytosine pairs form three hydrogen bonds (Figure 1.1). Base mispairing is unfavorable because it would negate hydrogen bonds (Voet and Voet, 2004).

1.3 DNA Replication

Life is dependent on the cell's ability to grow and divide. Moreover, cell proliferation requires accurate duplication of the parental cell's genetic material so that the resulting daughter cells each have an identical copy of the genome. The DNA replication machinery, known as the replisome, is largely conserved from prokaryotes to eukaryotes (Yao and O'Donnell, 2010). The replisome is a multi-protein complex (Table 1.1) found at the replication fork and has the enzymatic capability to accurately synthesize single-strand DNA that is complementary to the template strand (Meselson and Stahl, 1958; Yao and O'Donnell, 2010).

Table 1.1 Components of the Replisome

	E. coli	**Human**
Helicase	DnaB	MCM2-7
Clamp loader	γ/τ complex	RFC
Processivity clamp	β clamp	PCNA
Replicating polymerase	pol III	pol δ and pol ε
Primase	DnaG	pol α
Single strand binding protein	SSB	RPA

The replisome orchestrates a series of events at the replication fork to ensure successful DNA replication. Replication initiation occurs at replication origins. DNA helicases load at the replication origin to unwind the DNA duplex and form a double-strand/single-strand DNA junction known as the replication fork (Figure 1.2). Single-strand binding proteins stabilize the exposed single-strand DNA generated by the moving

helicase (Wold and Kelly, 1988). The DNA polymerase holoenzyme is part of the replisome and has the capability to synthesize complementary DNA in a 5' to 3' direction. In *E. coli*, the core polymerase is comprised of three subunits. The α subunit harbors 5' polymerase activity, the ε subunit encodes 3' exonuclease activity that is responsible for removing replication errors, and the Θ subunit enhances ε proofreading activity (Kelman and O'Donnell, 1995). The antiparallel nature of the DNA duplex and the strict 5' directionality of the replicating polymerase requires different mechanisms for leading strand and lagging strand synthesis.

The leading strand is replicated continuously by the coordinated action of the clamp loader, the replicative polymerase, and the processivity clamp. The clamp loader is bound to the moving helicase and loads the ring-shaped processivity clamp onto the template strand (Kelch et al., 2012). The processivity clamp tethers the replicative polymerase to the template strand during DNA extension and significantly increases its processivity (Prelich et al., 1987). The replicative polymerase translocates continuously towards the moving helicase as it synthesizes the leading strand.

The lagging strand is copied discontinuously and requires additional components of the replisome. Discontinuous replication occurs in sections known as Okazaki fragments (Okazaki et al., 1968). The length of Okazaki fragments are significantly longer in bacteria (1-2 kilobases) compared to eukaryotes (100-200 bases), however, the

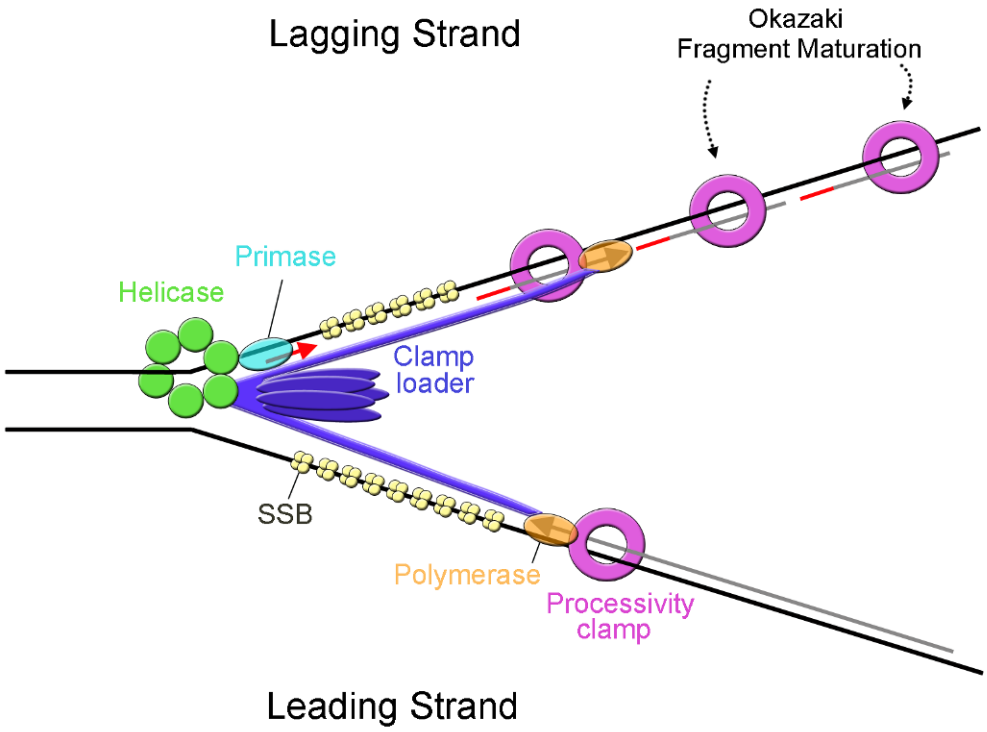


Figure 1.2. Model of the replication fork. The DNA helix is unwound by a helicase (green) and single strand binding protein (SSB, yellow) coats the single-strand DNA. The clamp loader (blue) is bound to the moving helicase and bridges the replicating polymerases (orange). The processivity clamp (purple) tethers the polymerase to the template DNA to improve polymerase processivity. On the lagging strand, primase (cyan) is anchored at the helicase and synthesizes RNA or RNA/DNA primers (red). The replicating polymerase extends the primers to synthesize Okazaki fragments (grey). The processivity clamp recruits Okazaki fragment maturation factors to complete replication. Arrows mark the direction of strand extension.

overall mechanism to synthesize these fragments are similar (Blumenthal and Clark, 1977; Wu et al., 1992). A primase first synthesizes a RNA (bacterial) or RNA/DNA hydrid (eukaryotic) primer of 10-12 nucleotides (Figure 1.2) (Wu et al., 1992; Zechner et al., 1992). The clamp loader loads the processivity clamp onto the primer and the replicative polymerase associates with the processivity clamp to synthesize the Okazaki

fragment (Tsurimoto and Stillman, 1991). Upon completion of the Okazaki fragment, the polymerase releases the processivity clamp and 'hops' to a processivity clamp waiting at a newly synthesized primer found upstream (Stukenberg et al., 1994). The replisome repeats this process until the entire lagging strand is copied. The processivity clamps left on Okazaki fragments recruits Okazaki fragment maturation factors which complete replication by replacing the RNA primers with DNA and DNA ligase seals the gaps (Figure 1.2) (Beattie and Bell, 2011; Waga et al., 1994). Eukaryotes also contain additional replisome proteins, such as GINS and CDC45, which are necessary for the proper coordination of replication, however, their roles are not well understood (Stillman, 2008).

Replication of damaged DNA requires specialized polymerases. The processivity clamp is essential for recruiting these DNA processing enzymes to active replication (Essers et al., 2005). Bulky DNA lesions, such as cyclobutane pyrimidine dimers caused by UV light, stall the progression of the replication fork threatening cell survival. To overcome a stalled replication fork, a polymerase switch on the processivity clamp exchanges the replicative polymerase for a translesion synthesis polymerase that has a low stringency in nucleotide selectivity and lacks 3' proofreading activity (Heltzel et al., 2012; Lehmann, 2003). While the translesion polymerase is inherently mutagenic, the open active site allows the polymerase to extend past the bulky lesion and prevents replication fork collapse (Ling et al., 2001; Yang, 2005). Misincorporated ribonucleotides and deoxyribonucleotides, as well as, DNA loops from polymerase strand

slippage are rare, but natural events in DNA replication that threaten genomic integrity (Nick McElhinny et al., 2010a; Nick McElhinny et al., 2010b; Pray, 2008). In addition to specialized polymerases, the processivity clamp also recruits DNA repair factors to these endogenous lesions (Maga and Hubscher, 2003).

1.4 Replication Errors can lead to Cancer

The need for a link between replication and repair is underscored by the obvious dangers associated with errors caused by the replication machinery. The replicative polymerase accurately duplicates the genome, however, on rare occasion (1 in every 10 billion nucleotides) the wrong nucleotide will be inserted into the growing nascent strand (Kunkel, 1992). The strict nucleotide selectivity of the replicative polymerase often dissuades the formation of base mismatches and largely contributes to the low polymerase error rate (Kunkel, 2009). Furthermore, the replicative polymerase encodes 3' exonuclease activity that can remove base mismatches. When the base mismatch escapes the polymerase proofreading activity, organisms support a conserved DNA repair pathway known as DNA Mismatch Repair (MMR). Processivity clamps found at the DNA replication fork recruit MMR factors to areas of active replication and, in turn, decreases the error rate by 20-400 fold (Maga and Hubscher, 2003; Schaaper, 1993; Simmons et al., 2008).

Defects in MMR genes result in mutator phenotypes (Cox et al., 1972). In addition to base mismatches, there is a higher frequency for insertion and deletions loops (IDLs) in repetitive DNA sequences known as microsatellites (Markowitz, 2000). The accumulation of replication errors are particularly detrimental due to the potential onset of tumor development. Lynch Syndrome is a hereditary cancer caused by mutations in several MMR genes (Lynch et al., 2009; Markowitz, 2000). Lynch Syndrome families have a high risk of developing cancers, such as colorectal (~24-75%), endometrial (27-71% in women), ovarian (3-13% in women), gastric (2-13%), urinary tract (1-12%), small bowel (2-7%), and brain (1-4%) cancer (Vasen et al., 2007). MMR defects have also been attributed to 15-25% of sporadic cancers due to spontaneous transcriptional silencing of a MMR gene (Kane et al., 1997; Peltomaki, 2003). The tight link between defects in MMR and cancer highlights the pivotal role of MMR in safeguarding genome integrity.

1.5 DNA Mismatch Repair Corrects Replication Errors

Many MMR proteins encode the consensus motif responsible for supporting an interaction with the processivity clamp (β clamp in bacteria and PCNA in eukaryotes) (Lopez de Saro et al., 2006; Lopez de Saro and O'Donnell, 2001; Pillon et al., 2011). In most cases, these processivity clamp binding motifs are essential for MMR, thereby suggesting an exquisite coordination between replication and repair that safeguards the integrity of the genome (Iyer et al., 2008; Pillon et al., 2010; Simmons et al., 2008).

The MMR pathway was first characterized in *E. coli*. The early steps include mismatch recognition and strand discrimination, which require the coordinated action of specialized MMR factors. MMR is bidirectional, thus, the removal and resynthesis of the strand including the error can occur either 5' or 3' to the mismatch (Constantin et al., 2005; Cooper et al., 1993; Dzantiev et al., 2004; Grilley et al., 1993; Zhang et al., 2005). Therefore, repair requires the collective action of exonucleases and the replication machinery which are common in the final steps of many DNA repair pathways. Consequently, the early steps of the reaction have gripped most of the attention because they define a specialized mechanism to coordinate DNA replication and mismatch surveillance.

1.6 Mismatch Recognition

MutS is responsible for detecting replication errors near the replication fork and triggers a signaling cascade that will lead to repair (Acharya et al., 2003). Structural and biochemical characterization of MutS homologs reveal the extensive dynamics of this protein. The overall structure of MutS is conserved from bacteria to eukaryotes (Gupta et al., 2012; Lamers et al., 2000; Obmolova et al., 2000; Warren et al., 2007). MutS is an oval shaped dimer comprised of five domains (mismatch binding domain, connector domain, core domain, clamp domain, and ATPase domain) (Figure 1.3). Bacterial MutS homologs are homodimers that recognize both base mismatches and small IDLs. Conversely, eukaryotes have three MutS paralogs MSH2, MSH3, and MSH6 that form two specialized heterodimers. MutSα (the dimer formed by the association of MSH2

with MSH6) identifies base mismatches and small IDLs while MutSβ (MSH2/MSH3) recognizes large IDLs (up to 16 nucleotides) (Kunkel and Erie, 2005).

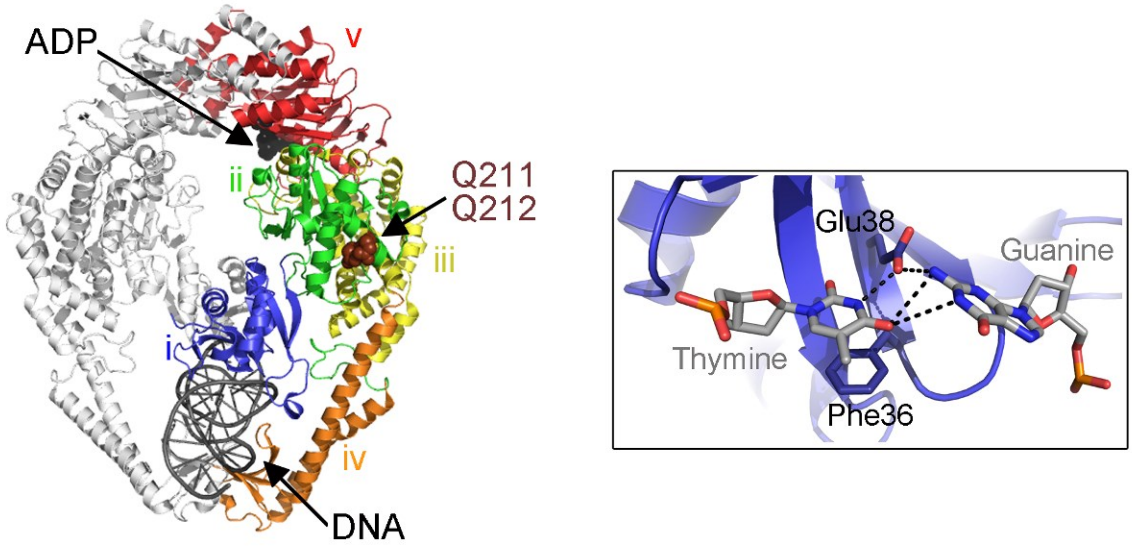


Figure 1.3. Structural organization of *E. coli* MutS. Ribbon diagram of *E. coli* MutS (PDB 1E3M) bound to ADP (black) and a G/T DNA substrate (grey). The MutS protomer can be divided into the mismatch binding domain (i, blue), connector domain (ii, green), core domain (iii, yellow), clamp domain (iv, orange), and ATPase domain (v, red). MutS residues Q211 and Q212 (brown spheres) are part of the connector domain and are implicated in MutL binding (Mendillo et al., 2009). The inset is a zoom highlighting the MutS Phe-X-Glu motif bound to a G/T mispair.

MutS adopts multiple conformations prior to activating MMR. The MutS mismatch binding domain and clamp domain are dynamic in the absence of DNA (Cho et al., 2007; Obmolova et al., 2000). Upon DNA binding, the MutS clamp domains become ordered and wrap around the DNA duplex by making sequence unspecific contacts which bend the DNA backbone (Cho et al., 2007; Obmolova et al., 2000). This MutS 'DNA scanning mode' diffuses along the DNA in search of replication errors while the

mismatch binding domains continue to undergo dynamic conformational changes (Figure 1.4) (DeRocco et al., 2014; Qiu et al., 2012; Winkler et al., 2011). Upon mismatch detection, MutS forms the mismatch recognition complex where one mismatch binding domain contacts the DNA backbone and the other directly interacts with the mispair through a conserved Phe-X-Glu motif (Lamers et al., 2000; Obmolova et al., 2000). The conserved MutS phenylalanine residue forms π stacking with the mispair while the conserved glutamic acid discriminates between Watson-Crick base pairs and base mismatches (Figure 1.3; see inset) (Lamers et al., 2000; Obmolova et al., 2000; Schofield et al., 2001a; Warren et al., 2007).

The β clamp has been implicated in stabilizing MutS at base mispairs (Flores-Rozas et al., 2000; Simmons et al., 2008). The mismatch recognition conformation exposes the MutS β clamp binding surface and triggers an interaction with the β clamp (Simmons et al., 2008). While the location of the β clamp binding surface is controversial, this interaction is important for MMR (Flores-Rozas et al., 2000; Iyer et al., 2008; Lopez de Saro et al., 2006; Shell et al., 2007; Simmons et al., 2008). The proposed functional importance of the *B. subtilis* MutS-β clamp complex is repetitive loading of MutS at the mismatch during early MMR (Simmons et al., 2008). Moreover, the human MutSα-PCNA complex has also been implicated in strand resynthesis by an unknown mechanism (Iyer et al., 2008).

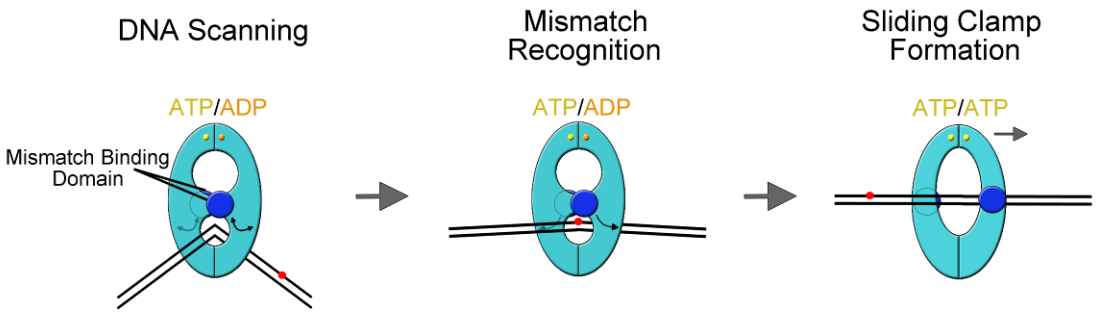


Figure 1.4. Model for MutS Sliding Clamp Formation. MutS (cyan) bends DNA while it scans for base mismatches (red dot). The dynamics of the mismatch binding domains (blue) during DNA scanning are shown as black double-headed arrows. Upon mismatch detection, one of the mismatch binding domains makes specific contacts with the mispair and the DNA is unbent at the mismatch. The MutS ATP/ADP state is proficient in transforming into a mobile sliding clamp which diffuses away from the mismatch.

There are three working models that describe how MutS initiates MMR following the detection of a base mismatch. In the 'Stationary' model MutS binds ATP to identify base mismatches and recruits downstream MMR proteins such as MutL (Junop et al., 2001; Schofield et al., 2001b). The immobile MutS-MutL complex remains at the mismatch until mismatch excision. In the 'Moving' model MutS binds to a base mismatch and undergoes an ADP to ATP nucleotide exchange that triggers a large conformational change (Gradia et al., 1997; Gradia et al., 1999). MutS converts into a sliding clamp that diffuses away from the mismatch in an ATP hydrolysis-independent manner to search for MMR factors (Gradia et al., 1997; Gradia et al., 1999). Once MutS has left the base mismatch, additional MutS proteins can load at the mismatch resulting in a high concentration of MutS at the replication error (Acharya et al., 2003). The alternative 'Translocation' model also proposes that MutS leaves the mismatch in search of downstream MMR factors, however, diffusion along the DNA requires ATP hydrolysis

(Allen et al., 1997; Blackwell et al., 1998). While there is ambiguity associated with the order of events during MMR initiation, there is growing evidence supporting the nucleotide-induced transformation of MutS into a sliding clamp (Gorman et al., 2007; Groothuizen et al., 2013; Jeong et al., 2011; Lee et al., 2014; Qiu et al., 2012; Sharma et al., 2013).

The transition from the MutS mismatch recognition complex into a sliding clamp is proposed to trigger the repair signaling cascade. This dramatic conformational change is heavily influenced by the nucleotides occupying the MutS ATPase sites. The ATPase activity of MutS is essential for MMR (Figure 1.3) (Alani et al., 1997). The two ATPase sites of the dimer behave asymmetrically where one ATPase site has high affinity for adenine nucleotides and is primed for ATP hydrolysis while the adjacent protomer demonstrates low binding affinity (Lamers et al., 2003; Studamire et al., 1998). The ATP-bound state of MutS causes a 'tightening' of the ATPase sites and decreases the overall radius of the protein while the ADP-bound state is correlated with a 'relaxation' of the ATPase domains (Kato et al., 2001; Lamers et al., 2004).

Several combinations of MutS nucleotide-bound states can scan DNA, however, only the asymmetric ADP/ATP MutS conformation is proficient in transforming into a sliding clamp (Antony and Hingorani, 2004; Monti et al., 2011; Qiu et al., 2012). The MutS sliding clamp forms when the bridging mismatch binding domains move away from the center of the protein to form a single cavity (Figure 1.4) (Antony and Hingorani,

2004; Monti et al., 2011; Qiu et al., 2012). An ADP to ATP nucleotide exchange has also been proposed to accompany this conformational change and allows for ATP hydrolysis independent diffusion along DNA (Gradia et al., 1997; Gradia et al., 1999; Qiu et al., 2012). Together with DNA unbending at the mismatch (DeRocco et al., 2014; Tessmer et al., 2008; Wang et al., 2003), MutS transforms into a sliding clamp and translocates away from the mismatch in search of MMR factors responsible for strand discrimination.

1.7 Strand Discrimination and Removal

Through these conformational changes, MutS communicates the presence of a mismatch to MutL and the MutS-MutL complex coordinates the series of events that mark the nascent strand for repair (Acharya et al., 2003). Strand discrimination is the most critical step of this pathway because only the daughter strand requires repair, yet DNA mismatches lack the distinguishing features that often mark a DNA lesion.

In *E. coli*, strand discrimination takes advantage of the methylation state of DNA (Pukkila et al., 1983). Immediately behind the replication fork, the nascent strand is transiently unmethylated forming hemi-methylated GATC sites (Langle-Rouault et al., 1987). The *E. coli* MutS-MutL complex activates the sequence specific MutH endonuclease which nicks the unmethylated strand 5' to the GATC sequence (Figure 1.5) (Giron-Monzon et al., 2004; Joseph et al., 2004; Welsh et al., 1987). The MutH incision

provides the necessary entry point for downstream factors to load and remove the mismatch.

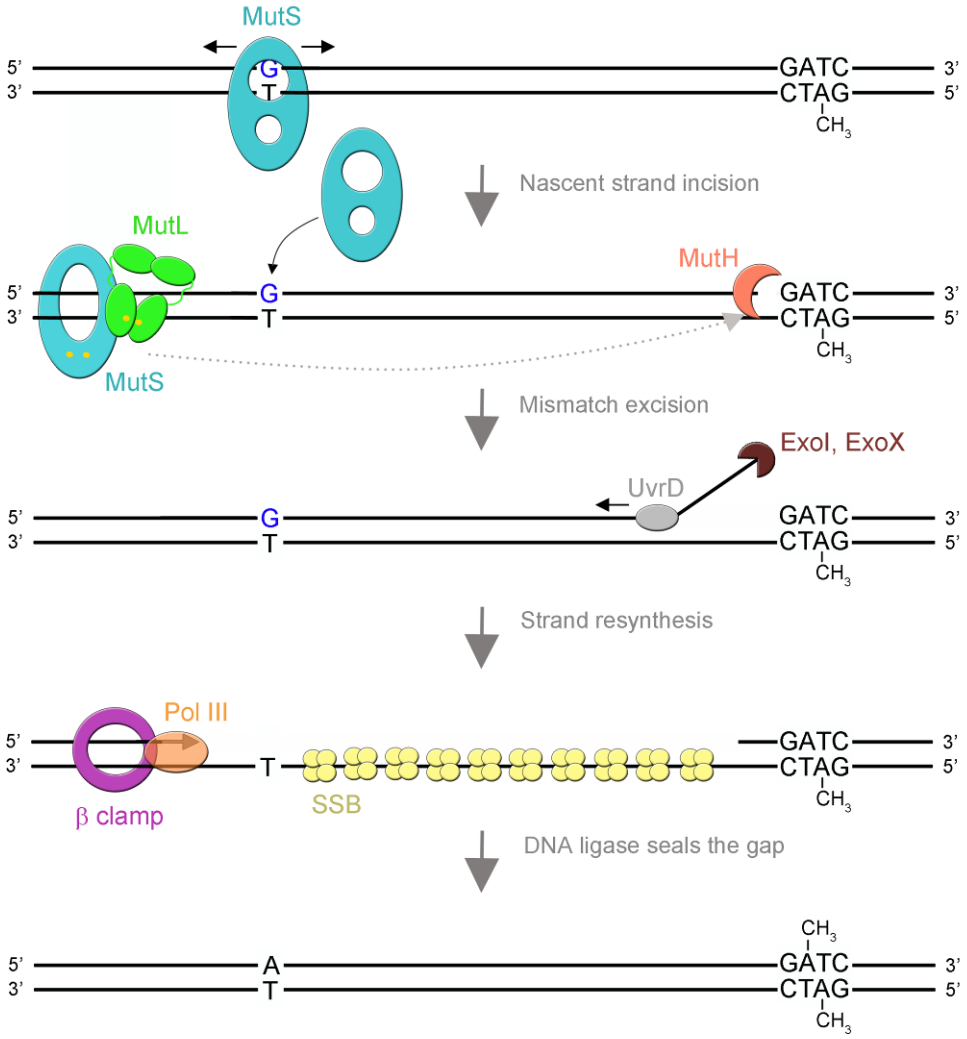


Figure 1.5. Model of *E. coli* methyl-directed mismatch repair. MutS detects replication errors (G/T) and underdoes a nucleotide (yellow dot) exchange to transform into a sliding clamp. Together MutS-MutL activate MutH endonuclease activity. MutH nicks 5' to the hemimethylated GATC site. UvrD (grey) unwinds the DNA and an exonuclease (brown) excises the error. Replication machinery (replicative polymerase (Pol III, orange), single strand binding protein (SSB, yellow), and β–clamp (purple)) resynthesize the strand.

Mismatch removal is orchestrated by MutL and the UvrD helicase. MutL repetitively loads UvrD at the MutH incision to efficiently unwind the stretch of DNA spanning the MutH incision and the mismatch (Figure 1.5) (Dao and Modrich, 1998; Guarné et al., 2004; Matson and Robertson, 2006). Single-strand binding proteins (SSB) coat the exposed single strand DNA generated by the moving helicase to prevent reannealing of the DNA. Finally, mismatch excision is supported by a series of 5' and 3' exonucleases (RecJ, ExoI, ExoVII, ExoX) that degrade the nascent strand (Figure 1.5) (Burdett et al., 2001).

While elegant work spanning the mid-1990s to early 2000s has highlighted the molecular cues for strand discrimination in *E. coli*, there are several caveats to this model organism. First, the *E. coli* MutS residues (Q211, Q212) responsible for MutL binding are not conserved in other MutS homologs, therefore, it remains unknown how most MutS homologs communicate with MutL (Figure 1.3) (Mendillo et al., 2009). Second, only a subset of γ-proteobacteria encode the *mutH* gene. Consequently, it was unclear how most organisms (both bacterial and eukaryotic) mark the nascent strand for repair.

Seminal contributions over the last ten years have succeeded in uncovering the mechanism of strand discrimination in organisms lacking MutH and has redefined our understanding of the field (Ghodgaonkar et al., 2013; Kadyrov et al., 2006; Kadyrov et al., 2007; Lujan et al., 2013; Pluciennik et al., 2010). Gaps between Okazaki fragments were first identified as MMR strand discriminating signals, however, these gaps are

limited to the lagging strand (Pavlov et al., 2003). Nascent strand discontinuities also arise from the repair of misincorporated ribonucleoside-triphosphates (rNTPs) introduced during DNA replication (Ghodgaonkar et al., 2013; Lujan et al., 2013). In *Saccharomyces cerevisiae* (*S. cerevisiae*), the imperfect selectivity of the replicative polymerase results in the misincorporation of rNTPs into both the leading (l in every 1250 dNTPs) and lagging (1 in every 5000 dNTPs) strands (Nick McElhinny et al., 2010a; Nick McElhinny et al., 2010b). These rNTPs are targeted for repair by ribonucleotide excision repair which replaces the rNTP with dNTP through a series of steps including RNase H2-mediated DNA incision, strand displacement, FEN1-mediated flap processing, and DNA ligation (Sparks et al., 2012). Eukaryotic MMR uses the transient gap introduced by RNase H2 as a strand discriminating signal (Figure 1.6) (Eder and Walder, 1991; Ghodgaonkar et al., 2013; Lujan et al., 2013).

Eukaryotes also support bidirectional mismatch repair, but they only encode a single 5' exonuclease (EXOI) (Constantin et al., 2005; Dzantiev et al., 2004; Genschel et al., 2002). Consequently, eukaryotes have alternative mechanisms for 5'- and 3'-directed mismatch excision. *In vitro* reconstitution of 5'-directed mismatch excision requires the mismatch sensing protein (MutSα), the 5' exonuclease (EXOI), and single strand binding protein (RPA) (Dzantiev et al., 2004). Conversely, 3'-directed mismatch excision additionally requires MutLα, the replication factor C (RFC), and the eukaryotic processivity clamp (PCNA) (Dzantiev et al., 2004). For successful 3'-directed mismatch excision, RFC loads PCNA in an orientation-specific manner at the RNase H2 incision

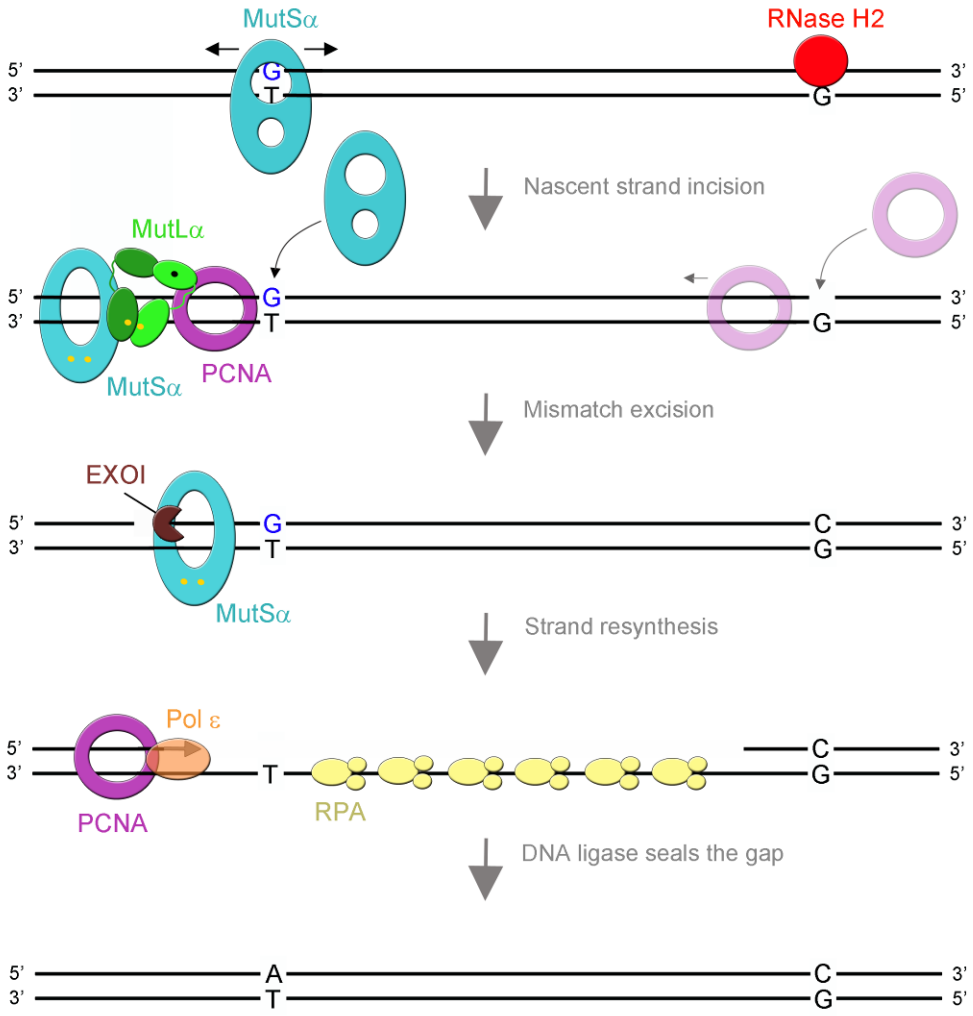


Figure 1.6. Model of eukaryotic 3'-directed mismatch repair. Misincorporated ribonucleotides act as strand discriminating signal in MMR. MutSα detects a G/T mismatch and undergoes an ATP-dependent conformational change. MutSα moves away from the mismatch in search of MutLα (green). PCNA (purple) is loaded at the transient gap produced by RNase H2. PCNA activates the MutLα endonuclease (black dot) to provide a 5' nick relative to the mismatch. EXOI (brown) loads onto the DNA and removes the error. Replication machinery (PCNA, Pol ε (orange), and RPA (yellow)) resynthesize the strand.

and PCNA diffuses along the DNA towards MutLα (Figure 1.6). PCNA activates MutLα and directs its endonuclease activity towards the nascent strand creating a 5' nick relative to the mismatch (Ghodgaonkar et al., 2013; Kadyrov et al., 2006; Kadyrov et al., 2007; Pluciennik et al., 2010). MutLα incisions are somewhat reminiscent of the MutH incisions in *E. coli* MMR because MutLα provides the necessary entry points for EXOI-mediated mismatch removal.

EXOI induces sharp DNA bends using a protruding helix known as the 'hydrophobic wedge' to identify the 5' DNA incision (Orans et al., 2011). Once EXOI is stably bound at the nick, it remains inactive until MutSα relieves its autoinhibition and increases EXOI processivity (Figure 1.6) (Genschel and Modrich, 2003; Orans et al., 2011). Following excision of the mismatch, RPA reduces EXOI processivity to terminate strand removal (Genschel and Modrich, 2009). Together MutSα and MutLα also suppress further EXOI nuclease activity in the absence of a mismatch to prevent unnecessary DNA degradation (Genschel and Modrich, 2009).

MutL homologs with endonuclease activity have also been found in several *mutH*-less Gram-positive (*B. subtilis*) and Gram-negative (*N. gonorrhoeae*, *A. aeolicus*, *T. thermophilus, and P. aeruginosa*) bacteria (Correa et al., 2013; Duppatla et al., 2009; Fukui et al., 2008; Mauris and Evans, 2009; Pillon et al., 2010). The growing list of MutL homologs harboring endonuclease activity emphasizes the central role of MutL in the early steps of MMR.

1.8 MutL is a Molecular Matchmaker

MutL is known as the 'molecular matchmaker' because it coordinates many interactions during MMR. The ATPase domain alone interacts with DNA, MutS, MutH, and the processivity clamp (Ban et al., 1999; Giron-Monzon et al., 2004; Guarné et al., 2001; Hall et al., 2003; Lopez de Saro et al., 2006; Plotz et al., 2006; Winkler et al., 2011).

The domain organization of MutL is integral to its functions. The MutL protomer is comprised of structurally conserved N- and C-terminal domains connected by a flexible linker of variable length. While the architecture of both domains are conserved, the sequence of the C-terminal domain is variable allowing for the formation of both homo- and heterodimers. Bacterial MutL are homodimers, whereas the eukaryotic homologs are heterodimers formed from the MutL paralogs MLH1, PMS1 (MLH2 in yeast), MLH3, and PMS2 (PMS1 in yeast) (Kunkel and Erie, 2005). In humans, MutLα (the dimer formed by the association of MLH1 and PMS2) has a primary role in MMR, MutLβ (MLH1/PMS1) has an unknown function, and MutLγ (MLH1/MLH3) is involved in meiosis and suppresses IDLs (Fukui, 2010; Kunkel and Erie, 2005; Lipkin et al., 2002).

1.8.1 MutL N-terminal Domain is an ATPase

The MutL N-terminal domain harbors ATPase activity that is essential for proper mismatch repair (Junop et al., 2003; Spampinato and Modrich, 2000). This domain belongs to the Gyrase/Hsp 90/Histidine Kinase/MutL (GHKL) ATPase superfamily (Ban

and Yang, 1998). Members of this superfamily encode four conserved motifs defining a Bergerat-fold (Ban and Yang, 1998; Bergerat et al., 1997; Dutta and Inouye, 2000; Mushegian et al., 1997). The ATPase domains of the MutL dimer associate upon nucleotide binding inducing a large conformational change that drives MutL towards a compact state. This conformational change was first identified in bacterial homologs (Ban et al., 1999; Ban and Yang, 1998; Fukui et al., 2008; Guarné et al., 2004) and later visualized in human and yeast MutLα (Figure 1.7) (Sacho et al., 2008; Tran and Liskay, 2000). Cycling between these open and closed forms likely regulate MutL functions (Ban et al., 1999; Sacho et al., 2008).

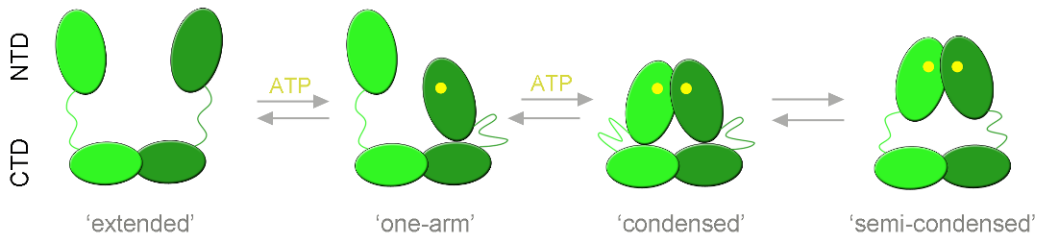


Figure 1.7. Four ATP-dependent conformational states of MutLα. Human and *S. cerevisiae* MutLα undergo reversible conformational changes upon nucleotide binding (yellow dot). NTD and CTD mark the N-and C-terminal domains, respectively.

DNA binding stimulates MutL ATP hydrolysis activity and is essential for MMR (Ban et al., 1999; Junop et al., 2003). The ATPase domain binds to single- and double-strand DNA in a sequence unspecific, but length-dependent manner (Bende and Grafstrom, 1991; Guarné et al., 2001; Hall et al., 2001; Iino et al., 2011; Pillon et al., 2010). In *E. coli*, an arginine residue (R266) has been linked to MutL DNA binding

activity (Figure 1.8) (Ban et al., 1999; Junop et al. 2003). R266 is located in a positively charged crevasse formed when the N-terminal domains associate (Ban et al., 1999). This work has precipitated a model where DNA binds to the central cavity of the MutL dimer. Yet, this model remains controversial since R266 is not conserved. Yeast MutLα does not require the dimerization of the N-terminal domains for DNA binding activity (Hall et al., 2003). Dr. Kunkel and colleagues suggest an alternative model where the N-terminal domains bind distant DNA sites to bridge the base mispair to the strand discriminating signal (Hall et al., 2003).

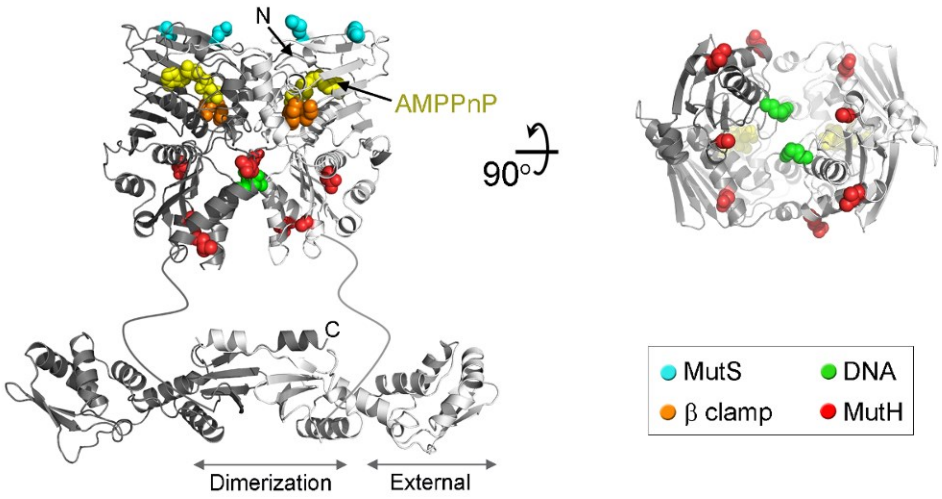


Figure 1.8. Structural organization of *E. coli* MutL. Model of full length *E. coli* MutL using the atomic structures of the MutL N-terminal domain bound to AMPPnP (yellow) (PDB 1B63), and the MutL C-terminal domain (PDB 1X9Z). *E. coli* MutL residues implicated in MutS (cyan), DNA (green), β clamp (orange), and MutH (red) binding are shown as spheres. Double-headed arrows mark the C-terminal subdomains.

The MutL ATPase domain also mediates the interaction with MutS (Acharya et al., 2003; Lenhart et al., 2013; Plotz et al., 2006; Winkler et al., 2011). The ATPase domain is sufficient to interact with MutS, however, the residues responsible for MutS

binding are unclear (Lenhart et al., 2013; Winkler et al., 2011). The binding interface is presumably near the back of the MutL ATP binding site, but the structural organization of this interface remains unknown (Figure 1.8) (Plotz et al., 2006; Winkler et al., 2011).

The processivity clamp has been implicated in the early steps of both methyl-directed and nick-directed MMR (Kadyrov et al., 2006; Lopez de Saro et al., 2006; Umar et al., 1996). PCNA, the eukaryotic processivity clamp, stimulates the endonuclease activity of MutLα (Kadyrov et al., 2006; Kadyrov et al., 2007) and, while the direct interaction between the two proteins has not been tested, the bacterial β clamp interacts with MutL weakly but specifically (Lopez de Saro et al., 2006; Pillon et al., 2011). Conserved residues in the MutL N-terminal domain (^{150}LF) are important for β clamp binding activity (Figure 1.8) (Lopez de Saro et al., 2006). The solvent accessibility of ^{150}LF is likely regulated by ATP since ATPase dimerization conceals this motif (Ban et al., 1999). Despite the biochemical characterization of ^{150}LF, the functional importance of this β clamp binding surface is unclear (Lopez de Saro et al., 2006). Furthermore, disruption of ^{150}LF drastically reduces β clamp binding affinity, but does not abolish the interaction (Lopez de Saro et al., 2006). This suggests that there is another β clamp binding surface in MutL.

In *E. coli*, the N-terminal domain of MutL is also responsible for the activation of MutH which marks the nascent strand for repair (Giron-Monzon et al., 2004). MutH activation relies on MutL ATP binding suggesting that association of the N-terminal

domains is a prerequisite for strand discrimination (Acharya et al., 2003; Ban et al., 1999; Spampinato and Modrich, 2000). Elegant biochemical characterization of the *E. coli* MutL-MutH complex has mapped the MutH binding surface to the edge of the crevasse formed by the ATPase dimer (Figure 1.8) (Giron-Monzon et al., 2004).

The MutL ATPase domain is connected to the C-terminal domain through a variable linker. Eukaryotic MutL homologs typically have longer linkers than the prokaryotic homologs. A subset of bacterial MutL homologs, including the *Thermus* genus, have an extremely short linker (Figure 1.9). MutL homologs with short linkers are functional likely reflecting the pliability of the MutL linker length. Indeed, almost two thirds of the *E. coli* MutL linker can be removed without disrupting MutL ATPase, DNA binding, and MMR activities (Guarné et al., 2004). Nevertheless, the C-terminal region of the yeast MutLα linker plays a direct role in DNA binding activity (Plys et al., 2012), indicating that some parts of the linker might be critical for MutL function.

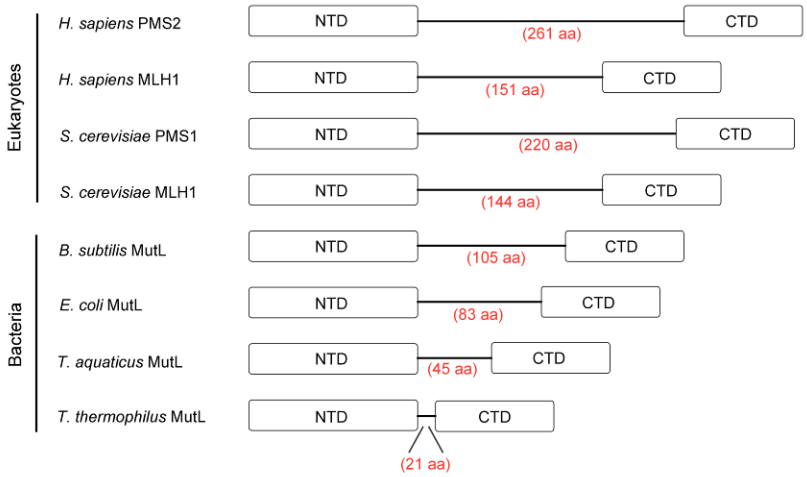


Figure 1.9. The MutL linker is variable in length. The MutL N-terminal domain

Figure 1.9 continued on page 27…

Figure 1.9 continued from page 26…
(NTD) is connected to the C-terminal domain (CTD) through a variable linker. The MutL domain boundaries were assigned based on MutL sequence alignments (Larkin et al., 2007) and secondary structure predictions (Buchan et al., 2013). The number of amino acids (aa) contributing to the linker are indicated in red.

1.8.2 MutL Dimerization Domain

The MutL C-terminal domain harbors the constitutive dimerization interface supporting homodimerization in bacteria and heterodimerization in eukaryotes. The dimerization interface is conserved across MutL homologs. Hydrophobic residues from a 4-strand β-sheet significantly contribute to the hydrophobic surface driving MutL dimerization (Abe and Masuda, 2000; Guarné et al., 2004; Gueneau et al., 2013; Kosinski et al., 2005; Namadurai et al., 2010; Pillon et al., 2010). Yeast MutLα also supports additional contacts creating a dimerization interface twice the size of its bacterial homologs (Gueneau et al., 2013).

Most MutL homologs supporting nick-directed strand discrimination harbor an endonuclease site in the dimerization domain. The endonuclease site is found at the junction between the dimerization and external subdomains (Figure 1.10). The α-helix (αA) connecting the dimerization and external subdomains includes the conserved $DQHA(X)_2E(X)_4E$ motif that was first linked to MutL nicking activity (Kadyrov et al., 2006; Pillon et al., 2010). The other two conserved motifs within the domain, ACR and CPHGRP, cluster with helix αA (Figure 1.10), defining an elongated endonuclease site (Gueneau et al., 2013; Pillon et al., 2010).

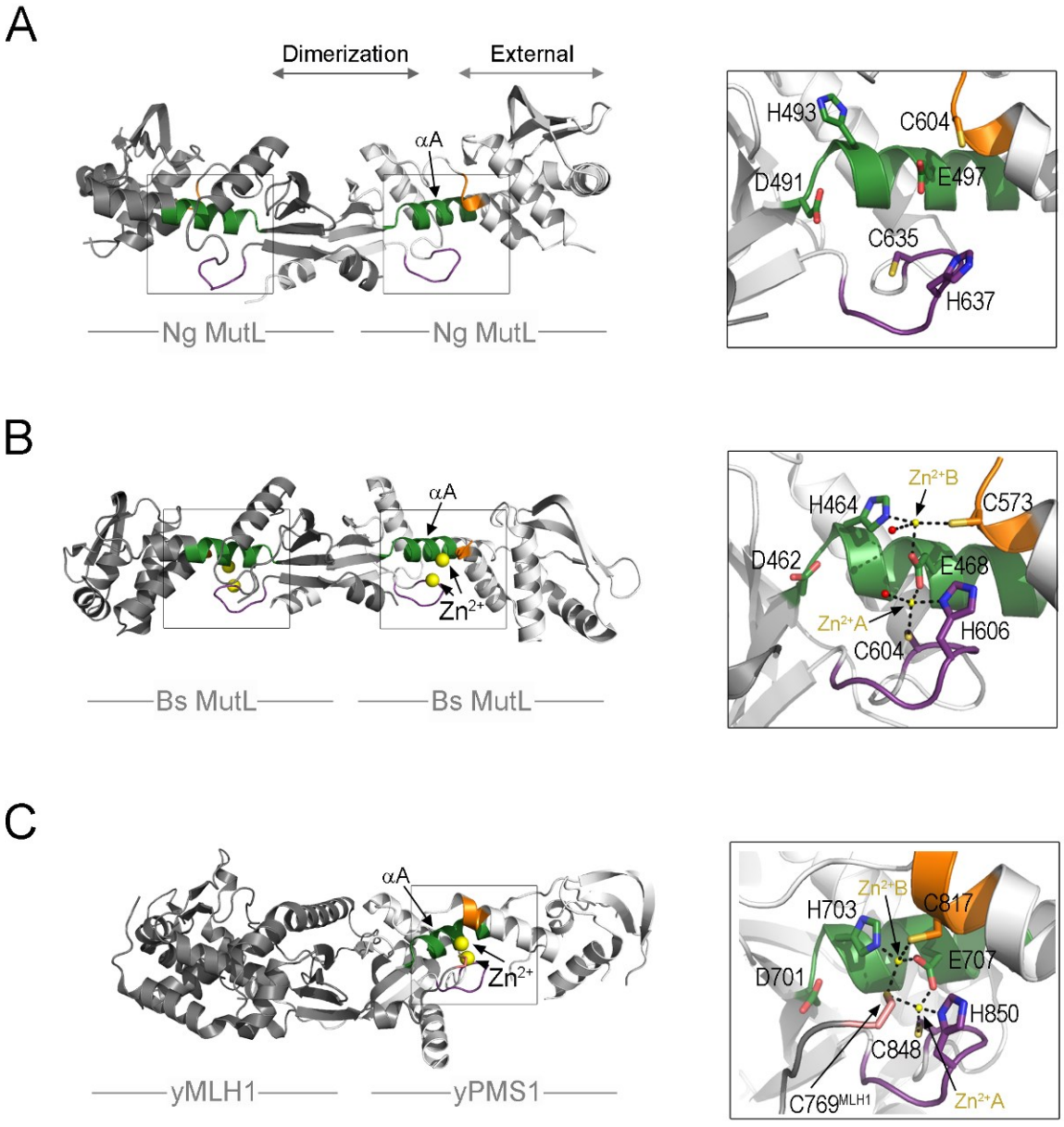


Figure 1.10. Structural organization of the MutL endonuclease active site. Ribbon diagrams of the MutL C-terminal domain from (**A**) *N. gonorrhoeae* (Ng MutL; PDB 3NCV), (**B**) *B. subtilis* (Bs MutL; PDB 3KDK), and (**C**) *S. cerevisiae* (yMLH1/yPMS1; PDB 4FMO). Double-headed arrows marks the C-terminal subdomains. Insets show the endonuclease active site. Conserved motifs $DQHA(X)_2E(X)_4E$ (green), ACR (orange), and CPHGRP (purple) cluster to coordinate zinc metals (yellow spheres). *S. cerevisiae* MLH1 C769 (pink) also contributes to zinc coordination whereas this position is occupied by ordered water molecules (red spheres) in *B. subtilis*.

MutL is a manganese-dependent endonuclease *in vitro* (Kadyrov et al., 2006; Pillon et al., 2010). The metal selectivity can vary among MutL homologs since several bacterial homologs (*A. aeolicus* and *N. gonorrhoeae*) can also substitute the catalytic metal for Mg^{2+}, Ni^{2+}, Ca^{2+}, or Co^{2+} (Duppatla et al., 2009; Iino et al., 2011). This highlights a potential broad metal-binding selectivity *in vivo.* Indeed, manganese is a common surrogate for magnesium, manganese is not readily abundant in the cell, and magnesium and manganese share the same coordination geometry (Harding, 1999; Maguire and Cowan, 2002). While the catalytic binding site is unknown, the conserved aspartate residue from $\underline{D}QHA(X)_2E(X)_4E$ is important for catalysis and, presumably, coordinates the catalytic metal (Guarné, 2012; Kadyrov et al., 2006).

Several MutL homologs also coordinate zinc (Gueneau et al., 2013; Iino et al., 2011; Kosinski et al., 2008; Pillon et al., 2010). There are two zinc binding sites (Zn^{2+}A and Zn^{2+}B) that lay adjacent to the putative catalytic metal binding site (Figure 1.10B and C) (Gueneau et al., 2013; Pillon et al., 2010). The conserved residues involved in zinc coordination are essential for MMR, however, zinc does not support catalysis (Pillon et al., 2010). Zinc stimulates the manganese-dependent MutL endonuclease (Pillon et al., 2010). Zinc binding also induces a local rearrangement of the overall structure suggesting that zinc plays a structural role.

The MutL Zn^{2+}A binding site resembles the coordination geometry of a secondary metal found in *Streptococcus gordonii* ScaR (Guarné, 2012; Stoll et al., 2009). ScaR

belongs to the DtxR/MntR family of metalloregulators and controls the cellular level of manganese by controlling metal uptake (Stoll et al., 2009). ScaR binds manganese at a conserved metal binding site to activate DNA binding activity (Stoll et al., 2009). ScaR also has a novel zinc metal binding site adjacent to its putative catalytic Mn^{2+} binding site that strongly resembles the zinc binding site in *B. subtilis* MutL and *S. cerevisiae* MutLα (Guarné, 2012). The function of Zn^{2+} in ScaR is unknown, however, it is also proposed to be a regulatory metal (Stoll et al., 2009).

1.8.3 Regulation of the Endonuclease Activity of MutL

MutL is an inherently weak endonuclease. This characteristic likely prevents unspecific DNA nicking at the replication fork. Consequently, the activation of MutL must be tightly regulated.

ATP stimulates the endonuclease activity of MutL in humans, yeast, and *B. subtilis* (Kadyrov et al., 2006; Kadyrov et al., 2007; Pillon et al., 2010). ADP does not have an effect on nicking activity, revealing an activation mechanism specific for ATP binding (Pillon et al., 2010). MutL underdoes a well characterized ATP-dependent protein compaction (Figure 1.7), suggesting that only the compact form of MutL is a proficient nuclease (Ban et al., 1999; Ban and Yang, 1998; Fukui et al., 2008; Guarné et al., 2004; Sacho et al., 2008; Tran and Liskay, 2000). In good agreement with this idea, ATP triggers an in trans interaction between the isolated *A. aeolicus* MutL N-terminal and C-terminal domains (Iino et al., 2011; Yamamoto et al., 2011). Furthermore, the presence

of ATP also induces local structural changes to the secondary structure elements encompassing the conserved DQHA$(X)_2$E$(X)_4$E and CPHGRP motifs (Yamamoto et al., 2011).

The eukaryotic MutLα endonuclease is also activated by PCNA (Kadyrov et al., 2006; Kadyrov et al., 2007). The mechanism for PCNA-dependent MutLα activation is currently unclear, however, it is presumably by tethering MutLα to DNA (Kadyrov et al., 2006; Kadyrov et al., 2007; Lopez de Saro et al., 2006; Maga and Hubscher, 2003). Additionally, PCNA orients MutLα nicking activity towards the nascent strand (Pluciennik et al., 2010). The asymmetry of the PCNA ring and its orientation specific loading onto DNA has been proposed to direct MutLα nicking towards the nascent strand (Pluciennik et al., 2010). Nevertheless, the PCNA-dependent regulatory mechanism controlling the MutLα endonuclease remains largely unknown.

1.9 Thesis Objectives

The last two decades have identified the key players involved in the repair of replication errors. Extensive structural and biochemical characterization has uncovered the mechanisms for these proteins at the molecular level. Nevertheless, the spatial and temporal regulation of mismatch repair remains an active area of research.

I began my Ph.D. two years after the discovery of the endonuclease activity of MutLα. This discovery posed new pressing questions. How does the endonuclease of

MutL work? How is this activity regulated by other mismatch repair factors? What is the role of the processivity clamp at this step of mismatch repair? In particular, work included in this thesis addressed the following:

1) Unveiling the interface of the MutS-MutL complex in *Bacillus subtilis* (chapter 2).
2) Determining the crystal structure of the endonuclease domain of *Bacillus subtilis* MutL (chapter 3).
3) Understanding how the β clamp interacts with the endonuclease domain of MutL (chapters 4 and 5).

Chapter 2

Trapping and Visualizing Intermediate Steps in the Mismatch Repair Pathway *in vivo*

Reprinted from Molecular Microbiology, Vol. 90, J.S. Lenhart, M.C. Pillon, A. Guarné, and L.A. Simmons, Trapping and Visualizing Intermediate Steps in the Mismatch Repair Pathway *in vivo*, 680-698, Copyright 2013, with permission from John Wiley & Sons Ltd.

2.1 Author's Preface

This chapter is focused on understanding how MutS communicates with MutL to coordinate downstream MMR events. Characterization of the *B. subtilis* MutS-MutL interaction revealed two exposed phenylalanine residues (F319, F320) on MutS that are essential for MutL binding. Using a peptide array, J.S. Lenhart performed a preliminary search of MutS sequences that could support MutL binding (Figure 2.1 panel B). A. Guarné and I mapped the 15 peptides onto a model of *B. subtilis* MutS and designed a series of MutS variants. I generated these variants and conducted *in vitro* assays to assess MutL binding (Figures 2.2 and 2.3), oligomerization (Figure S2.4), DNA binding (Figure 2.2), and ATP hydrolysis (Figure 2.2). I identified a single MutS variant that was defective for MutL binding. This MutS variant harbored mutations to four tandem residues that J.S. Lenhart further refined to two phenylalanines. J.S. Lenhart performed all the *in vivo* assays monitoring DNA mismatch repair (Tables 2.1 and 2.2), MutS localization (Figure 2.5) and MutL recruitment (Figure 2.5). J.S. Lenhart and I analyzed the data and prepared the figures. L.A. Simmons and A. Guarné wrote the manuscript.

2.2 Abstract

During mismatch repair, MutS is responsible for mismatch detection and the recruitment of MutL to the mismatch through a mechanism that is unknown in most organisms. Here, we identified a discrete site on MutS that is occupied by MutL in *Bacillus subtilis*. The MutL binding site is composed of two adjacent phenylalanine residues located laterally in an exposed loop of MutS. Disruption of this site renders MutS defective in binding MutL *in vitro* and *in vivo,* while also eliminating mismatch repair. Analysis of MutS repair complexes *in vivo* shows that MutS mutants defective in interaction with MutL are "trapped" in a repetitive loading response. Furthermore, these mutant MutS repair complexes persist on DNA away from the DNA polymerase, suggesting that MutS remains loaded on mismatch proximal DNA awaiting arrival of MutL. We also provide evidence that MutS and MutL interact independent of mismatch binding by MutS *in vivo* and *in vitro,* suggesting that MutL can transiently probe MutS to determine if MutS is mismatch bound. Together, these data provide insights into the mechanism that MutS employs to recruit MutL, and the consequences that ensue when MutL recruitment is blocked.

2.3 Introduction

Mismatch repair (MMR) is a highly conserved pathway responsible for identifying and correcting DNA polymerase errors, which substantially improves the overall fidelity of genome replication (Iyer et al., 2006; Kunkel and Erie, 2005; Lenhart et al., 2012; Schofield and Hsieh, 2003). Defects in bacterial MutS or MutL cause a

substantial increase in mutation rate (Cooper et al., 2012; Cox et al., 1972; Davies et al., 2011; Ginetti et al., 1996; Prudhomme et al., 1989), while inactivation of the eukaryotic homologues, MutSα and MutLα, causes an increase in mutation rate and microsatellite instability (Fishel et al., 1994; Umar et al., 1994). In humans, disruption of MMR can lead to the development of sporadic cancers, as well as hereditary cancers such as Lynch and Turcot syndromes (Fishel et al., 1993; Hamilton et al., 1995; Nystrom-Lahti et al., 2002; Peltomaki, 2005). In prokaryotes, disruption of MMR can lead to an increased possibility of generating mutations that confer antibiotic resistance, and has been linked to antibiotic resistant strains of nosocomial human pathogens (Klein et al., 2007; Klevens et al., 2007; Kluytmans et al., 1997; Lowy, 1998).

In bacteria, the pathway and the mechanisms underlying MMR are best understood in the MutH and Dam containing bacterium *Escherichia coli*. In *E. coli*, MMR is initiated upon the recognition of single base mismatches or insertion/deletion loops (IDLs) by the mismatch binding protein MutS [for review (Iyer et al., 2006; Larrea et al., 2010; Lenhart et al., 2012)]. While scanning for replication errors, MutS exists in an ADP bound state (Bjornson et al., 2000; Blackwell et al., 2001). Following mismatch recognition, a prominent model is that MutS exchanges ADP for ATP, converting MutS to a sliding clamp causing MutS to diffuse away from the mismatch along the DNA in search of MutL (Acharya et al., 2003; Winkler et al., 2011). After arrival, MutL then performs several tasks necessary to facilitate removal of the strand bearing the mismatch (Ban et al., 1999; Ban and Yang, 1998; Guarné et al., 2004).

The initial steps of MMR have been thoroughly studied and elucidated in the Gram-positive bacterium *Bacillus subtilis,* an organism lacking the Dam and MutH dependent pathway [for review (Lenhart et al., 2012)]. Preceding mismatch detection, MutS is targeted to newly replicated DNA through interaction with the DNA replication processivity clamp DnaN (Dupes et al., 2010; Klocko et al., 2011; Lenhart et al., 2012; Simmons et al., 2008). DnaN, a critical component of the pathway, is required for 90% of MMR in *B. subtilis* (Lenhart et al., 2013). During Okazaki fragment maturation, DnaN accumulates behind the progressing replication forks forming a transient DnaN clamp zone that facilitates coupling between mismatch detection and concurrent DNA replication (Lenhart et al., 2013). Within this zone, MutS detects mismatches and initiates the downstream steps of repair, which includes MutL recruitment. The mechanism used by MutS to recruit MutL is unknown *in vivo* and *in vitro* for *B. subtilis*.

Even though MutS and MutL have been extensively characterized at the biochemical and genetic level, their binding interface and the mechanism used to recruit MutL is poorly understood in most organisms. In *E. coli*, an effort employing hydrogen/deuterium exchange mass spectrometry identified a MutL docking site on MutS composed of two adjacent glutamines (residues 211 and 212) found within the MutS connector domain (Mendillo et al., 2009). However, this site is not conserved in Gram-positive bacteria (Figure S2.1), suggesting a separate uncharacterized interface that facilitates binding in other organisms. To our knowledge, no other sites have been identified in any bacterium lacking the Dam/MutH-dependent MMR, and the effect of

MutS mutants defective for MutL interaction have not been tested on repair intermediates *in vivo* for any organism. Therefore, very little is known about MutL recruitment, yet this step represents the second step in one of the most important pathways for maintaining high fidelity replication in organisms from bacteria to humans.

Here we define the MutS•MutL interface in *B. subtilis*. We show that MutS binds the N-terminal domain of MutL via two adjacent phenylalanine residues, F319 and F320. Substitution of these phenylalanines to serine eliminates crosslinking of MutS to the N-terminal domain of MutL *in vitro* while also eliminating MMR *in vivo*. Importantly, these substitutions do not seem to affect other biochemical properties of MutS, including dimerization, ATPase activity, and mismatch binding. Furthermore, using single cell fluorescence microscopy, we show that MutS mutants defective in MutL interaction form repair centers that increase in both frequency within the cell population, as well as overall fluorescence intensity. These data provide *in vivo* evidence for *in vitro* models proposing that MutS loads repetitively at a mismatch. Our work also defines a regulatory role for MutL in limiting or preventing additional MutS dimers from loading at a mismatch. We show that repetitive loading of MutS is repressed following excision of the mismatch, which requires not only MutL recruitment but also endonuclease directed nicking of the DNA. We also provide evidence against the paradigm that MutL requires MutS bound to a mismatch to initiate interaction. We show that within living cells and with purified components, we can selectively crosslink MutS to MutL in the absence of a mismatch, suggesting a mechanism where MutL can transiently probe MutS to determine if MutS is

indeed mismatch bound, and if so, license repair. Together, our data provide new insight into the MutL recruitment mechanism, the physiological consequences that result from MutS mutants unable to bind and recruit MutL, and we describe a model where MutL can transiently probe MutS before initiating the second step of MMR.

2.4 Materials and Methods

2.4.1 Bacteriological methods

B. subtilis strains were grown according to established procedures (Dupes et al., 2010). Briefly, strains were grown in Luria-Bertani (LB) medium or defined $S7_{50}$ minimal medium. Unless otherwise stated antibiotics were used when appropriate with the following concentrations: 100 µg/mL spectinomycin (*spc*), 5 µg/mL chloramphenicol (*cam*), 5 µg/mL tetracycline (*tet*), 0.5 µg/ml erythromycin and 12.5 µg/mL lincomycin (*mls*), 5 µg/mL kanamycin (*kan*), 150 µg/mL rifampin (*rif*).

2.4.2 Peptide array analysis

The MutS peptide array was synthesized at the Massachusetts Institute of Technology Biopolymers laboratory, (Cambridge, MA). The synthesized MutS peptides provided 1x coverage spanned the entire amino acid sequence of MutS by overlapping 10 mer peptides offset by 3 residues. The final array consisted of 282 spotted peptides. The peptide array was activated by wetting with 100% ethanol, followed by 3 successive washes in Tris-buffered saline + Tween 20 (TBS-T) (50 mM Tris-Cl pH 7.5, 150 mM NaCl and 0.05% Tween 20) (pH 7.6) for 5 minutes to remove excess ethanol. The array

was then blocked overnight in TBS-T and 10% milk solids at 4˚C. The following day, the array was washed in TBS-T, and then incubated in 56 nM MutL-myc in protein incubating solution (40 mM HEPES-KOH pH 7.6, 23 mM KCl, 1 mM $MgSO_4$, 1 mM DTT, 0.5 mg/mL BSA, 2% glycerol, and 0.5 mM of either AMPPNP or ADP) for 15 hours at 4˚C with gentle rocking. The next day, the array was washed for 30 minutes total with 3 washes each of the following buffers in order: TBS-T, TBS-T with 500 mM NaCl, TBS-T+0.5% Triton X-100, and TBS-T. The array was then incubated with 1:5000 α-myc antibody in TBS-T+5% milk for one hour at room temperature. The wash series was repeated, followed by incubation with 1:2000 anti-mouse in TBST+5% milk for 1 hour at room temperature. After antibody incubation, one more wash series was performed and the array was exposed using Pierce SuperSignal. Exposure time course of 2 minutes, 5 minutes, and overnight were obtained to identify bound peptides.

False-positive peptides were removed by comparing to a negative control (no myc-tagged protein exposure). Finally, peptides that were surface exposed based on a structure-guided sequence alignment of MutS homologs were deemed putative MutL binding peptides.

2.4.3 Strains and plasmid

All *B. subtilis* strains used are derivatives of PY79 and are described in supplemental Table 2.1. Plasmids created for use in this study are as follow:

B. subtilis mutS and *mutL* expressing plasmids MutS (pAG 8483; residues 1-858) was amplified from *B. subtilis* strain 168 genomic DNA and ligated into pET-15b (Novagen) using restriction sites NdeI and BamHI. MutS variants Patch 1 (pAG 8561; E155S, R156S, L157A, E158S), Patch 2 (pAG 8674; E245S, E247S, E248S), Patch 3B (pAG 8635; F320S, E321S, R322S, E323S), Patch 4 (pAG 8634; E392S, E395S, E396S), Patch 5 (pAG 8616; E510S, E512S, E514S), Patch 6A (pAG 8646; Q806A, L807A, F809A, F810A), and Patch 6B (pAG 8535; D811S, E812S, E814S) were generated using overlap PCR and ligated into pET-15b using NdeI and BamHI. *B. subtilis* MutL N-terminal domain (MutL-NTD) (pAG 8286; residues 1-339) was amplified and ligated into pProEX HTa (Invitrogen) using NcoI and XhoI. All mutants were verified by DNA sequencing (MOBIX, McMaster University).

2.4.4 Purification of his_6MutS

B. subtilis MutS variants were overproduced in BL21 (DE3) pRARE or BL21 (DE3) pRARE pLysS cells (Invitrogen) and induced with 1 mM IPTG for 5 hours at 25°C. Cells were resuspended in buffer A (20 mM Tris pH 8.0, 0.5 M NaCl, 30 mM imidazole, 1.4 mM 2-mercaptoethanol, and 5% glycerol), lysed by sonication, and clarified by centrifugation at 39,000 g. The soluble fraction was purified over a nickel-chelating column equilibrated with buffer A and eluted with 240 mM imidazole. MutS was then injected onto an ion exchange column (Q-Sepharose, GE Healthcare) equilibrated with buffer B (20 mM Tris pH 8.0, 5 mM EDTA, 2.8 mM 2-mercaptoethanol, 100 mM NaCl, and 5% glycerol) and eluted using a linear gradient to

400 mM NaCl. MutS was injected into a gel filtration column (Superdex-200, GE Healthcare) equilibrated with crosslinking buffer (20 mM Hepes pH 7.5, 100 mM NaCl, 5 mM DTT, and 5% glycerol). Protein concentration was measured at 280 nm.

2.4.5 Purification of the MutL N-terminal domain

B. subtilis MutL-NTD was overexpressed in BL21 Star (DE3) cells (Invitrogen) with 0.5 mM IPTG for 5 hours at 25°C. MutL-NTD was purified using a nickel chelating column equilibrated with buffer A (pH 9.0) and eluted using 240 mM imidazole. MutL-NTD was then injected into a sizing column (Superdex-200, GE Healthcare) equilibrated with crosslinking buffer. Protein concentration was measured by absorbance at 280 nm.

2.4.6 Spontaneous mutation rate analysis

Fluctuation analysis was performed essentially as described (Bolz et al., 2012; Lenhart et al., 2013). We inoculated 3 mL of LB with a single colony, and grew at 37˚C until an OD_{600} of ~1.2. At that point, 1 mL of culture was pelleted and resuspended in 100 µL of saline. A portion of this resuspension was further diluted to 10^{-6}, and plated onto LB plates in order to enumerate the total viable cells with incubation overnight at 30ºC to ensure the plates with viable cells did not over grow. The original resuspension was plated on LB supplemented with 150 µg/mL rifampin plates overnight at 37ºC in order to determine the number of spontaneous mutations causing rifampin resistance. After performing a minimum of 15 independent cultures, the mutation rate was determined using the MSS Maximum Likelihood Method using the publicly available FALCOR tool

at http://www.mitochondria.org/protocols/FALCOR.html. 95% confidence intervals were determined and percent mismatch repair activity, was determined using the following equation: [(RMR null – RMR strain)/(RMR null – RMR wild type)]•100 where RMR = relative mutation rate (Hall et al., 2009).

2.4.7 Chemical crosslinking

B. subtilis MutS variants (20 μM), 20 μM Mis90, and 20 mM ATP were pre-incubated on ice for 1 hour. MutL-NTD (40 μM) was then added with equal volume to the MutS•ATP•DNA reaction and incubated for 30 minutes at 4°C. Reactions were then incubated with 0.8-1.6 mM bis (sulfosuccinimidyl) suberate (Sigma, BS^3) for 30 minutes at 22°C. Reactions (10 μL) were quenched with 30 mM Tris pH 7.5 for 15 minutes at 22°C and separated on a 4-15% SDS gradient gel (BioRad) and stained with Coomassie Blue.

2.4.8 ATPase

ATP hydrolysis assays were performed as previously described (Junop et al., 2003) with minor modifications. ATPase activity was measured with 0.3 μM MutS and 5 mM $MgCl_2$ in reaction buffer (20 mM Tris pH 8.0, 90 mM KCl, 1 mM DTT, 1 mg/ml BSA, and 5% glycerol). Reactions (15 μL) were initiated by the addition of 1 mM α-^{32}P-labeled ATP and incubated for 1 hour at 22°C. Reactions were stopped with 25 mM EDTA and hydrolyzed product was detected by thin-layer chromatography using 750 mM

KH_2PO_4 for running buffer. ATPase activity was measured in triplicates for each MutS variant.

2.4.9 DNA binding

Mis90 is a 90 base pair DNA substrate harboring a G/T base mismatch (5' gaaaacctgtatttcagggcaggcctattggaattcaacatatgaagtcgacgcagctggcggccgcttctagaggatccctcg agaag 3' annealed to 5' gcttctcgagggatcctctagaagcggccgccagctgcgtcgacttcatatgttgaattccaat aggcctgccctggaaatacaggtttt 3'). MutS (600 pmol) was incubated with equimolar Mis90 in binding buffer (10 mM Hepes pH 7.5, 70 mM KCl, 2 mM DTT, 5 mM $MgCl_2$, 1 mg/ml BSA, and 15% glycerol) for 1 hour on ice. Reactions (15 μL) were resolved on a 6% TBE gel and stained with ethidium bromide. Bands were quantified using ImageJ (http://rsbweb.nih.gov/ij/). DNA binding activity was measured in triplicates for each MutS variant.

2.4.10 Live cell microscopy

Cultures for imaging were prepared as described previously (Dupes et al., 2010; Klocko et al., 2011; Lenhart et al., 2013). Briefly, strains for imaging were inoculated in pre-warmed $S7_{50}$ minimal media supplemented with either 1% L-arabinose or 2% D-glucose at a starting OD_{600} of 0.05. Cells were grown past three doublings to an OD_{600} of 0.4-0.5 and imaged. To treat cultures with the mismatch-forming drug 2-aminopurine, we split the cultures and added a mock treatment to one and 600 μg/mL 2-aminopurine to the other followed by growth for an additional hour. Cell membranes were visualized with

the fluorescent dye TMA-DPH at a working concentration of 10 μM (Lenhart et al., 2013). MutS fluorescent fusions were captured with a 1.2 second exposure. Colocalization experiments were conducted with L-arabinose as the sole carbon source, where all other experiments used D-glucose.

2.4.11 *In vivo* crosslinking/co-immunoprecipitation

B. subtilis cultures were inoculated in LB at a starting OD_{600} of 0.05 and grown at 37°C to an OD_{600} of 0.7. Cells were pelleted, washed twice with crosslinking buffer (40 mM HEPES pH 7.4, 500 mM sucrose, 2 mM $MgCl_2$, 150 mM NaCl, 0.02% Tween-20) to remove LB, and resuspended in 1.75 mL of crosslinking buffer. To crosslink intracellular protein complexes, 0.5 mM of Dithiobis[succinimidyl propionate] (DSP) was added to the growing cells and crosslinking occurred for 30 minutes at room temperature on a rotisserrie. Cultures were quenched by adding Tris-HCl (pH 7.5) to a final concentration of 20 mM, and incubated an additional 30 minutes at room temperature on a rotisserrie. After quenching, cells were lysed via sonication. Lysates were cleared of debris by centrifugation for 30 minutes at 4° C at 14,000 rpm. Lysates were then concentrated to 50 μL, resuspended in crosslinking buffer supplemented with 1X protease inhibitor cocktail and 0.5 mM EDTA to a final volume of 500 μL. A 5% input fraction was pulled from the final volume. The 5% input and the rest of the prepared lysate were incubated overnight on a rotisserie at 4°C. The IP fraction was incubated with 50 μL equilibrated magnetic beads bound with affinity purified α-MutS antisera (MI-1042). Beads were prepared according to protocol. In the morning, the lysates were washed 5X for 5 minutes each

with crosslinking buffer on a room temperature rotisserrie. The antibodies were eluted from the magnetic beads by a 10 minute incubation in 900 µL of antibody stripping buffer (5 mM Glycine pH 2.4, 150 mM NaCl). The IP fraction was concentrated by TCA precipitation, and resuspended in 1X western loading dye. IP and Input fractions were electrophoresed on the same gel (4-15% gradient gel). Quantitative analysis of the resulting bands was conducted in ImageJ. The numbers represent the statistical mean of 3 independent experiments with the background subtracted from the JSL281 strain. Relative numbers were determined relative to JSL364 (PY79 wild type strain).

2.4.12 Western and Far Western Blotting

B. subtilis whole-cell extracts were obtained by centrifuging 25 mL of mid-exponential cultures, followed by resuspension in lysis buffer [10 mM Tris-HCl (pH 7.0), 0.5 mM EDTA, 1 mM AEBSF, 1X Protease Inhibitor cocktail) followed by 3 rounds of sonication (20 Hz, 45 second duration) on ice as described (Lenhart et al., 2013). After sonication, SDS was added to a final concentration of 1% and non-soluble cellular debris and whole cells were removed by centrifugation at 4˚C. The lysate was divided into one-time use samples and stored at -20˚C. Total protein concentration of prepared soluble lysates was determined using Pierce BCA Protein Assay Kit (Thermo Scientific). Equal amounts of total protein were applied to each lane on a 4-15% gradient gel followed by transfer to a nitrocellulose membrane (Simmons et al., 2009; Simmons and Kaguni, 2003). Protein levels were determined by using primary antisera: α-MutS (MI-1042), α-MutL (MI-1044), and α-DnaN (MI-1038).

Immunodot blotting was performed as described (Klocko et al., 2011). Briefly, equal molar amounts of the indicated proteins were immobilized onto a nitrocellulose membrane with the assistance of a Bio-dot microfiltration apparatus (Bio Rad). The membrane was incubated in blocking buffer (5% milk solids, 17.4 mM Na_2HPO_4, 2.6 mM NaH_2PO_4, 150 mM NaCl, 0.05% Tween-20, 0.5 mM ATP, 4 mM $MgSO_4$) at 22˚C for one hour. All subsequent washes and incubations took place in blocking buffer. After blocking, the membrane was incubated with 0.4 μM MutL in blocking buffer for 3 hours at 22˚C. The blot was subsequently washed three times and then incubated in affinity purified α-MutL antisera overnight at 4˚C. In the morning, the blot was removed from primary antibody and washed three times at 22˚C and placed in secondary antisera (1:2000 α-Rabbit) for 2 hours at 22˚C. The blot was washed 3 more times, followed by a wash in PBS (17.4 mM Na_2HPO_4, 2.6 mM NaH_2PO_4, 150 mM NaCl, 0.05% Tween-20) to remove excess milk solids and exposed.

2.5 Results

2.5.1 The *E. coli* MutL binding interface is not conserved in *B. subtilis*

The MutS•MutL interface has previously been characterized in the Gram-negative bacterium *E. coli* (Mendillo et al., 2009). The interface is found within the connector domain of MutS, and centers around a double glutamine motif (Q211 and Q212) (Mendillo et al., 2009). Disruption of this site causes a loss of mismatch repair *in vivo* and has been shown to eliminate interaction with MutL on a mismatched DNA substrate *in vitro* (Mendillo et al., 2009). Initially, we asked if the *E. coli* MutL binding motif was

conserved in the Gram-positive bacterium *B. subtilis*; however, a sequence alignment revealed that the connector domain motif is not conserved, and the surrounding amino acid sequence is highly variable (Figure S2.1A). Further examination of a *B. subtilis* MutS model shows that although the amino acid sequence in the connector domain is not conserved, the secondary and tertiary structure of the connector domain is conserved with that of *E. coli* MutS (Figure S2.1B and C). Therefore, we mutated four residues, ^{205}VTII (*mutS Patch Ec*), which directly align with the *E. coli* ^{211}QQ motif, and occupy the corresponding location in a *B. subtilis* MutS model. Mutation of ^{205}VTII to ^{205}ASAA has no effect on MMR *in vivo*, conferring a mutation rate identical to the wild type control ($2.47X10^{-9}$ mutations /generation [0.95-3.82]) (Table 2.1**,** last row). With this result, we conclude that the MutL binding site on *B. subtilis* MutS is distinct from the site identified for *E. coli* MutS.

Table 2.1. Mutation rate analysis of *mutS* patch variants

Genotype	*mutS* variant	Number of cultures	Mutation rate (10^{-9} mutations/generation) ± [95% CI]	Relative mutation rate (% MMR activity)
Wild-Type (PY79)	$mutS^{WT}$	24	3.30 [1.44-5.00]	1 (100%)
mutL::spec	$mutS^{WT}$	18	159.9 [152.5-167.2]*	48.5 (0%)
mutS Patch 1	*E155S, R156S, L157A, E158S*	19	4.50 [2.23-6.64]	1.36 (99.2%)
mutS Patch 2	*E245S, E247S, E248S*	24	4.28 [2.10-6.34]	1.30 (99.4%)
mutS Patch 3A	*E306S, E307S, E310S*	25	4.19 [2.18-6.11]	1.27 (99.4%)
mutS Patch 3B	*F320S, E321S, R322S, E323S*	26	78.2 [72.2-84.2]*	23.7 (52.1%)

Table 2.1. continued on page 49…

Table 2.1. continued from page 48…

Genotype	*mutS* variant	Number of cultures	Mutation rate (10^{-9} mutations/generation) ± [95% CI]	Relative mutation rate (% MMR activity)
mutS Patch 4	*E392S, E395S, E396S*	20	5.60 [2.59-8.40]	1.70 (98.5%)
mutS Patch 5	*E510S, E512S, E514S*	20	7.24 [4.45-9.94]	2.20 (97.5%)
mutS Patch 6A	*Q806A, L807A, F809A, F810A*	23	8.83 [6.03-11.58]*	2.68 (96.5%)
mutS Patch 6B	*D811S, E812S, E814S*	20	3.03 [1.37-4.57]	0.92 (101%)
mutS Patch Ec	*V205A, T206S, I207A, I208A*	18	2.47 [0.95-3.82]	0.03 (101%)

All *mutS* variants were constructed using allelic replacement (see "Experimental Procedures"), which maintains the *mutS* variant gene at its normal genetic locus and under the control of its native promoter. The downstream *mutL* gene remains intact. Brackets enclose the lower bounds and upper bounds respectively of the 95% confidence limits. Percent MMR activity was determined using the following equation: [(R.M.R.null-R.M.R.strain)/(R.M.R.null-R.M.R.wild type)]•100. RMF=relative mutation rate. Relative mutation rate was obtained by dividing the mutation rate of each strain by that obtained for the wild type control.

2.5.2 MutL binds several surface exposed peptides on MutS

We previously showed that a direct interaction between MutS and MutL can be detected in *B. subtilis* without a DNA substrate using a far Western blot (Klocko et al., 2010). To verify the direct MutS•MutL interaction, we performed a far Western blot to compare the binding of MutL to MutS and another known binding partner, the replication processivity clamp DnaN (Figure 2.1A) (Simmons et al., 2008). We found that MutS retained MutL and DnaN on the nitrocellulose membrane during the binding reaction.

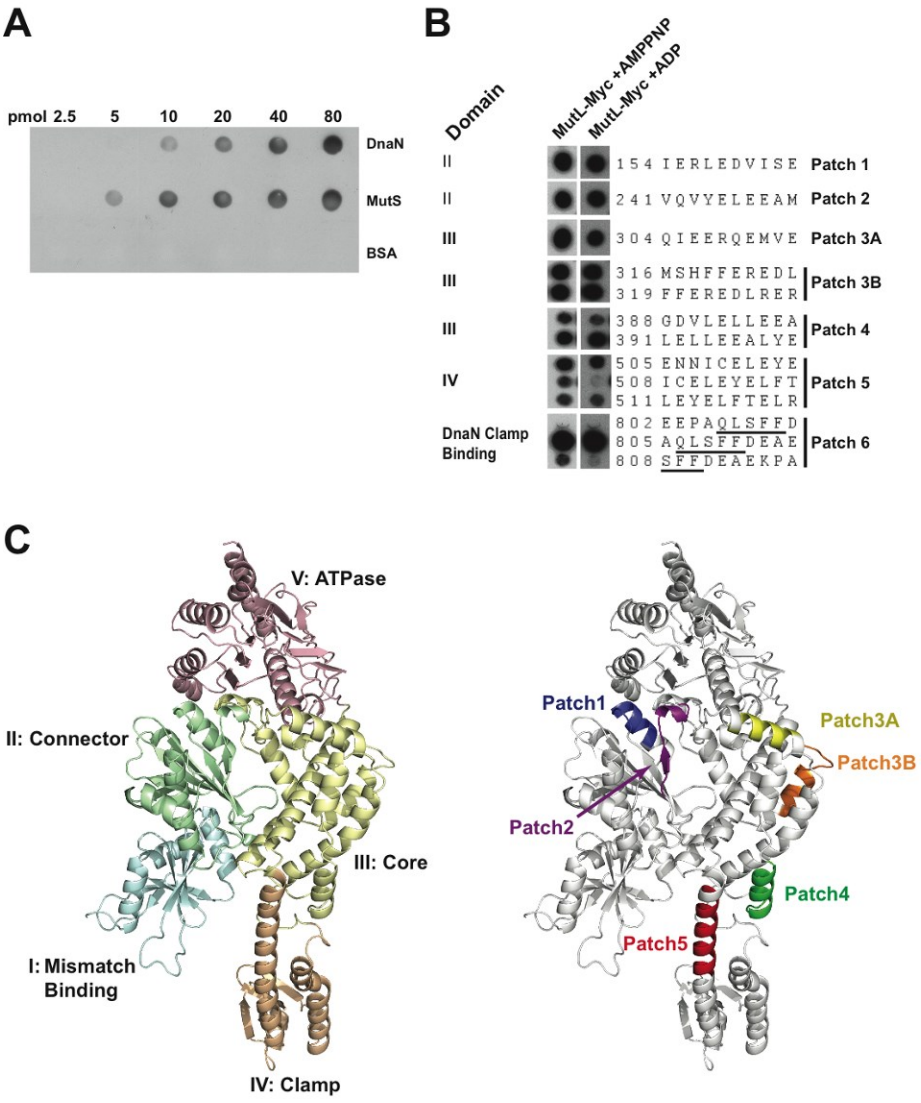


Figure 2.1. ***Bacillus subtilis*** **MutL binds surface exposed peptides in MutS. (A)** A far Western blot using MutL to probe for interaction with purified MutS, DnaN, and BSA. Equal amounts of the indicated protein monomer was applied via a dot blot apparatus and probed with 0.4 μM purified MutL. **(B)** Screening of a MutS peptide array library with MutL-Myc. MutL-Myc was incubated with 0.5 mM of either ADP or AMPPNP during incubation with the peptide array. MutL-Myc bound peptides were detected with α-Myc antibodies. Indicated position of positive peptides on the array, as well as the amino acid sequence, is shown adjacent to the MutL-Myc bound peptides. **(C)** *B. subtilis* MutS was modeled using the SWISS-MODEL server (Arnold et al., 2006). Both monomers of the model are shown as a ribbon diagram, with either the five functional domains of MutS (left panel) or the surface exposed peptides identified in the peptide array (right panel) color-coded and labeled according to their representative patch definition.

To identify candidate residues in MutS that may be important for MutL binding, we employed a peptide array library, which functions analogously to the far Western, using peptides in place of purified proteins. We screened a peptide array library composed of peptides representing the entire MutS primary structure. The MutS peptide library consisted of 10mer peptides offset by 3 residues, providing 3-fold coverage of the entire sequence of MutS. We determined the association of MutL bearing a single C-terminal Myc tag with the MutS peptide array library in the presence of ADP and the non-hydrolysable ATP analog adenosine 5′-(β,γ-imido) triphosphate (AMPPNP). We used ADP and AMPPNP to determine if the nucleotide bound state altered the putative MutL binding sites on MutS, because it has been previously shown that MutL undergoes substantial conformation changes during ATP binding and hydrolysis (Sacho et al., 2008).

We found that MutL•AMPPNP bound to 18 of 292 total peptides screened while MutL•ADP showed a nearly identical pattern and bound to 16 of the 18 peptides identified with MutL•AMPPNP (Figure 2.1B, showing only surface exposed peptides**)**. Peptides 508 and 808 did not retain MutL-ADP binding (Figure 2.1B). The data further shows that MutL-ADP bound at least one peptide in groups of overlapping peptides, suggesting that interaction within these regions occurred regardless of the nucleotide cofactor, and that overall, the nucleotide composition of MutL doesn't affect the specific MutS peptides bound. Further analysis of the amino acid composition of the MutS peptides bound by MutL-Myc revealed an enrichment of glutamic acid and phenylalanine

residues, suggesting a preferred amino acid target on the MutS peptide array (Figure S2.2).

To determine the location of each putative MutS binding peptides, we modeled *B. subtilis* MutS based on the *E. coli* and *T. aquaticus* structures (Lamers et al., 2000; Obmolova et al., 2000). In doing so, we found that most peptides (13 in the AMPPNP group and 11 in the ADP group) were surface exposed and located on the outer rim of the MutS dimer (Figure 2.1B and 2.1C). Based on the location of the surface exposed peptides, we defined six unique regions (identified as patch 1-6) composed of single or multiple MutL-Myc bound peptides, which could facilitate an interaction between MutS and MutL (Figure 2.1C). Since the peptides spanning residues 802-817 (patch 6) are absent from the crystal structures of *E. coli* and *T. aquaticus* MutS (Lamers et al., 2000; Obmolova et al., 2000), we were unable to include them in the model. Interestingly, patch 6, overlaps with the site known to bind DnaN, referred to as the DnaN clamp-binding motif (Figure 2.1B, ^{806}QLSFF) (Dalrymple et al., 2001; Simmons et al., 2008). In *B. subtilis* mutation of this region does reduce MutL recruitment into foci although the mutant *mutS* still retains almost all MMR activity *in vivo* suggesting this region is not critical for binding MutL (Simmons et al., 2008).

2.5.3 Substitution of surface exposed residues within the putative MutL interaction sites on MutS causes defects in MMR

The peptide array analysis identified sites on MutS that could potentially mediate a direct interaction with MutL, thus we began by introducing three to four amino acid substitutions in residues both conserved and surface exposed within each "patch" to

determine the effect on repair (Table S2.1, for summary of substitutions). Each mutant *mutS* patch allele encoding a set of missense mutations was used to replace the wild type allele at the native *mutS* locus by allelic exchange as described (Lenhart et al., 2013). For each mutant allele, we determined the mutation rate by measuring the rate of spontaneous rifampin resistant colony formation as an indicator for mutagenesis and MMR dysfunction [(Bolz et al., 2012; Dupes et al., 2010; Klocko et al., 2011; Lenhart et al., 2013) and "Experimental Procedures"]. Patch mutants 1, 2, 3A, 4, 5 and 6B conferred a statistically equivalent mutation rate to wild type *mutS*, showing no effect on the MMR pathway *in vivo* (Table 2.1). Patch mutant 6A, which contains the DnaN clamp-binding motif, showed a slight but significant increase in mutation rate at $8.83X10^{-9}$ mutations/generation (Table 2.1) as we previously reported (Simmons et al., 2008). Interestingly, we found that the four missense mutations introduced into patch 3B caused a significant increase in mutation rate ($78.2X10^{-9}$ mutations/generation), resulting in this mutant retaining only 50% of MMR activity *in vivo*. With these data, we conclude that patch mutant 3B, which includes the *F320S, E321S, R322S*, and *E323S* missense mutations, causes a significant defect in the MMR pathway in *B. subtilis* (Table 2.1). Hereafter, we refer to the patch mutant 3B as MutS3B.

2.5.4 MutS3B is defective for interaction with MutL

The MutS•MutL interaction has been previously monitored using chemical crosslinking (Winkler et al., 2011). The work by Winkler and co-workers demonstrated that the MutS•MutL interaction only requires the N-terminal domain of MutL (MutL-

NTD) and it is enhanced in the presence of ATP and a heteroduplex. Therefore, we purified N-terminal His_6 tagged variants of *B. subtilis* MutS and MutL-NTD and screened for interaction defects using this approach. Incubation of MutS and the chemical crosslinker bis(sulfosuccinimidyl)suberate (BS3) in the presence of ATP and a 90 base-pair G/T mismatch DNA substrate (Mis90) resulted in the formation of several high molecular weight species. Conversely, incubation of MutL-NTD with BS3 predominantly yielded monomers, as expected due to the absence of the dimerization domain of the proteins. Incubation of MutS with MutL-NTD and BS3 in the presence of ATP and Mis90, yielded a new species that was not present when either protein was incubated with BS3 and corresponded to the molecular weight of the MutS•MutL-NTD complex (Figure 2.2A). We excised this band, and using LC MS/MS, verified the presence of both MutS and MutL-NTD as the sole components of this band (data not shown). Interestingly, we do observe some interaction between MutS and MutL-NTD in the presence of a 90 bp DNA homoduplex in place of Mis90, showing that the MutS•MutL-NTD interaction is not strictly dependent on the presence of a mismatched substrate (Figure S2.3).

We subsequently tested whether any of the MutS patch variants abrogated the interaction with MutL-NTD. We found that all MutS variants formed a MutS•MutL-NTD complex except for the MutS3B variant (F320S, E321S, R322S and E323S) (Figure 2.2A). We note that the MutS5 variant (including the E510S, E512S and R514S mutations) showed a very prominent band of a molecular weight consistent with formation of a MutS tetramer. In fact, this prominent species was present in the

crosslinking reaction of all MutS variants when MutL-NTD was not present (data not shown), but disappeared upon incubation with MutL-NTD. Since our goal was to probe for the formation of a MutS•MutL complex, and MutS5 retained the interaction with MutL-NTD, we did not characterize this variant further. The crosslinking defect of MutS3B agrees well with the mutation rate analysis showing that *patch 3B* lost 50% of

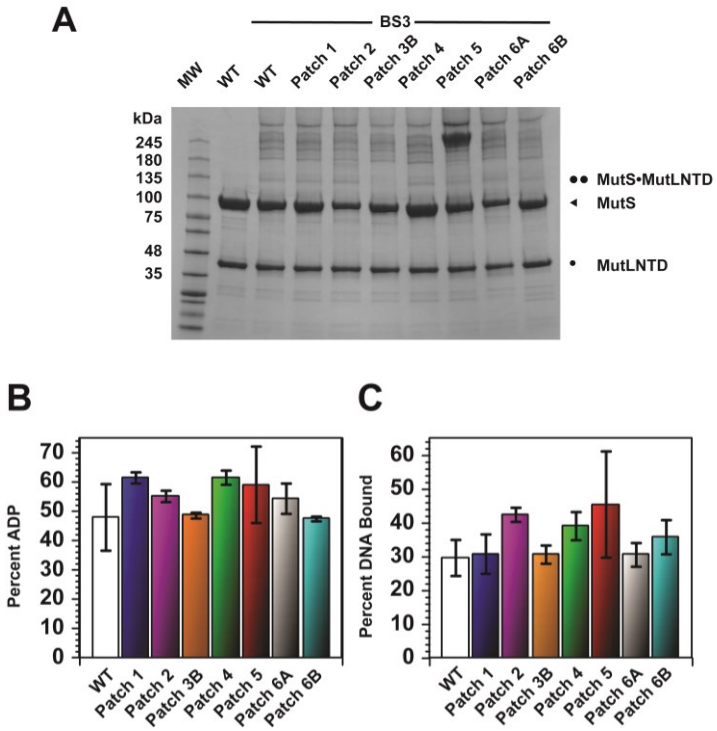


Figure 2.2. Purified MutS3B fails to crosslink with the N-terminal domain of MutL. **(A)** Crosslinking of MutS or MutS variants to the N-terminal domain of MutL with a 90 bp DNA substrate containing a centrally located G/T mismatch (Mis90). Mixtures of each protein, 10 mM ATP, and the G/T DNA substrate were incubated with the crosslinker BS3. Protein complexes were then resolved on a 4-15% gradient SDS polyacrylamide gel. The bands corresponding to the MutS and MutL-NTD monomers, as well as the MutS•MutL-NTD complex are labeled. All MutS variants show similar ATPase activity **(B)** and DNA binding to the G/T mismatched DNA substrate **(C)** to wild-type MutS. Bar diagrams present the average of three independent measurements and the error bars correspond to the standard errors of the mean (SEM=σ/√n, where σ is the standard deviation and n the sample size).

MMR activity *in vivo* (Table 2.1). All "patch" variants of MutS behave similar and have similar mismatch binding and ATPase activities compared to wild-type MutS (Figure 2.2B and 2.2C), implying that the reduced MMR activity of the MutS3B variant is unlikely due to improper folding or attenuation of other critical biochemical activities. Furthermore, all of the MutS variants eluted from a gel filtration column similarly to the wild type protein (data not shown) and formed dimers in solution as measured by dynamic light scattering (Figure S2.4). Collectively, we show that the MMR defect associated with the MutS3B variant is due to the impaired interaction with MutL rather than loss of some other biochemical activity of MutS. Furthermore, we conclude that residues changed in the MutS3B (F320, E321, R322, and E323) variant are important for direct interaction between *B. subtilis* MutS and MutL-NTD.

2.5.5 Residues F319 and F320 define the MutL binding site on MutS

Since *mutS3B* contains four successive missense mutations, we replaced the native *mutS* gene in *B. subtilis* with alleles encoding each of the single missense mutations that comprise *mutS3B* using allelic exchange in order to further define the functional residues important for MutL interaction. We also included amino acids S317, H318, and F319 in this analysis due to their adjacent position in MutS relative to the peptide identified in the array and because each residue is predicted to be surface exposed. We found that mutation of F319S and F320S separately reduced MMR activity below 50% *in vivo* (Table 2.2). In addition, we found that *mutSE323S* had the most striking effect of all the single missense mutations on MMR as this allele supported only 7% of MMR activity *in*

vivo. It should be noted that the effect observed in the *mutSE323S* mutant far exceeds that of the *mutS3B* mutant, which reduced MMR to 50% of wild type level. We suggest that the E321S and R322S substitutions may partially suppress the defect caused by E323S on its own. All other substitutions examined confer a mutation rate indistinguishable from wild type (Table 2.2). We did not pursue E323S for a role in MutL binding because this single mutant blocks MutS localization on its own and may have a folding defect. We describe the effects of this mutant later within this manuscript.

Because *mutSF319S* and *mutSF320S* showed significant and substantial defects in MMR, we combined these missense mutations to measure the effect on MMR *in vivo*. The resulting *mutSF319SF320S* allele showed a mutation rate ($156X10^{-9}$ mutations/generation) indistinguishable from a strain lacking *mutL* (*mutL::spc*) function (Table 2.2). Immunoblot analysis verified that MutSF319SF320S, as well as MutS variants containing each individual mutation, accumulate to the same steady state levels as wild-type MutS *in vivo* (Figure 2.3A). These results show that *mutSF319SF320S* phenocopies loss of *mutL* function in *B. subtilis* supporting the hypothesis that *mutSF319SF320S* is defective in MutL interaction. In addition we asked if overexpression of *mutL* could suppress the increased mutation rate caused by MutSF319SF320S. We expressed *mutL* using an IPTG inducible promoter from and ectopic locus and recovered only ~16 of MMR (Table 2.2, last row). This experiment further supports our conclusion that the MutSF319SF320S variant is substantially impaired for MutL interaction *in vivo*.

Table 2.2. Mutation rate analysis of missense mutations in and near *mutS3B*

Genotype	Mutation rate (10^{-9} mutations/generation) ± [95% CI]	Fold increase in mutation rate	% MMR activity
Wild-Type (PY79)	3.30 [1.44-5.00]	1	100
mutL::spec	159.9 [152.5-167.2]	48.5	0
mutSS317A	2.62 [0.94-4.09]*	0.63	100.4
mutSH318S	2.36 [0.79-3.69]*	0.55	100.6
mutSF319S	105.1 [97.9-112.4]	31.9	35.0
mutSF320S	94.8 [87.3-102.4]	28.8	41.5
mutSE321S	4.02 [1.83-6.06]*	1.22	99.5
mutSR322S	2.46 [0.92-3.78]*	0.75	100.5
mutSE323S	148.7 [140.1-157.2]#	45.1	7.2
mutSF319SF320S	156.0 [148.9-163.1]#	47.3	2.5
mutSF319SF320S, amyE::Pspac mutL	134.0 [127.3-140.5]#	40.6	16.6

All *mutS* variants were constructed using allelic replacement (see "Experimental Procedures"), which maintains the *mutS* variant gene at its normal genetic locus and under the control of its native promoter. The downstream *mutL* gene remains intact. Brackets enclose the lower bounds and upper bounds respectively of the 95% confidence limits. Percent MMR activity was determined using the following equation: [(R.M.R.null-R.M.R.strain)/(R.M.R.null-R.M.R.wild type)]•100 RMF=relative mutation rate. Relative mutation rate was obtained by dividing the mutation rate of the strain by that of wild type. The symbols * and # indicates that the mutation rate is statistically equivalent to that of the wild type and MMR deficient strains respectively. For expression of *mutL*, 1 mM IPTG was added to the media during growth.

In order to determine if residues F319 and F320 of MutS define a MutL binding site, we purified MutSF319SF320S and tested its ability to interact with MutL-NTD using chemical crosslinking (Figure 2.3B**)**. Like MutS3B, MutSF319SF320S fails to crosslink with MutL-NTD, indicating that the mutation of the phenylalanine pair is sufficient to eliminate interaction between MutS and MutL *in vitro*. We also verified that these substitutions were wild type for other biochemical activities of MutS. MutSF319SF320S maintained wild type levels of ATPase activity, binding to mismatched DNA substrate,

and dimer formation (Figure 2.3C and D, and data not shown), suggesting that loss of *in vivo* MMR in the *mutSF319SF320S* background is attributed to loss of binding to MutL.

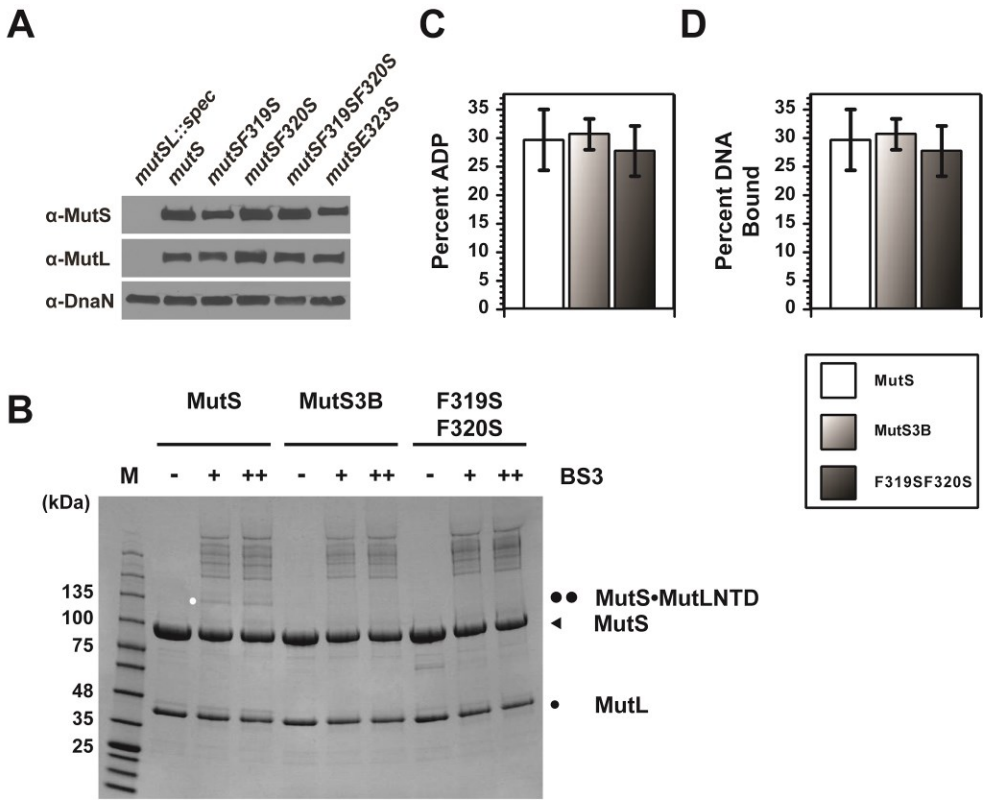


Figure 2.3. A distinct di-phenylalanine binding site within and around MutS3B defines the MutL binding interface. (A) Immunoblot analysis indicated proteins from the soluble fraction of cell lysates. 50 μg of soluble fraction was probed for MutS, MutL and DnaN. **(B)** Complex formation of MutS, MutS3B, and MutSF319SF320S to the N-terminal domain of MutL was assayed on a 90 bp DNA substrate containing a centrally located G/T mismatch using crosslinking analysis. Reactions contained 10 μM MutS variants, 20 μM MutL-NTD, protein, 10 mM ATP, and 10 μM of the G/T DNA substrate were incubated with the hydrophilic crosslinker BS3 (**+**=0.8 mM and **++**=1.6 mM, respectively). The products were then separated on a 4-15% gradient SDS-PAGE. The bands corresponding to the MutS and MutL-NTD monomers, as well as the MutS•MutL-NTD complex are labeled. The biochemical activity of purified MutS, MutS3B, and MutSF319SF320S were tested for **(C)** ATPase activity and **(D)** DNA binding to the G/T DNA substrate. Bar diagrams present the average of three independent measurements and the error bars correspond to the SEM.

2.5.6 MutSF319SF320S defines a highly conserved MutL binding site on MutS in Gram-positive bacteria

We asked if the MutL binding site on MutS is conserved in other organisms. The MutS residues important for MutL binding, F319 and F320, model to the outer rim of MutS and reside in the loop connecting helices α4 and α5 of the core domain (Figure 2.4A). Importantly, both residues appear solvent exposed, and available for interaction with MutL based on our structural model (Figure 2.4A). In human MSH2 and the Gram-negative bacteria *E. coli* and *T. aquaticus* MutS, the site appears to be structurally conserved, despite the limited sequence conservation (Figure 2.4B). Based on previous results (Mendillo et al., 2009), this interface is not the sole binding interface for *E. coli* MutL, but may however function as a secondary site located on the opposite side of the MutS face. Importantly, helix α4 is part of the allosteric transmitter proposed to connect the ATP- and DNA-binding sites of MutS (Obmolova et al., 2000), and hence, F319 and F320 pose an attractive mechanism to relay the nucleotide- and mismatch-bound state of MutS to MutL. Interestingly, we do find that the di-phenylalanine site is conserved in several eukaryotic proteins known to bind MutL homolog Mlh1 (Dherin et al., 2009) including MutSβ (Figure S2.5), as well as in anchoring interaction between mammalian Rev1 and Polκ (Iyer et al., 2010; Wojtaszek et al., 2012a; Wojtaszek et al., 2012b)(see discussion).

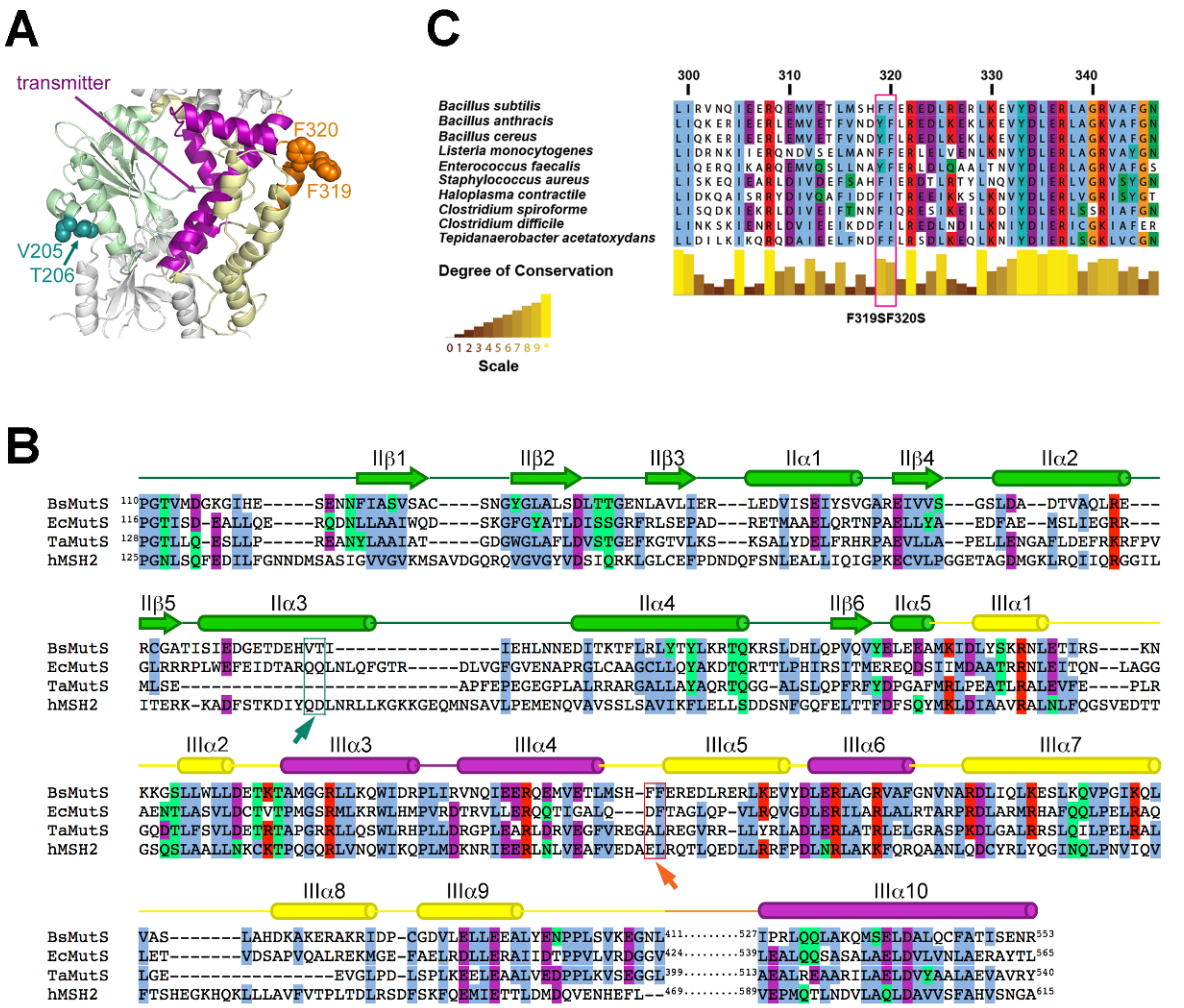


Figure 2.4. The di-phenylalanine site is conserved in MutS homologs. (A) Ribbon diagram of the connector (light green) and core (light yellow) domains of MutS. The side chains of the di-phenylalanine motif are shown in orange, those of the QQ motif are shown in teal (Mendillo et al., 2009) and the structural elements of the transmitter proposed by Obmolova and co-workers belonging to the core domain are colored in purple (Obmolova et al., 2000). **(B)** Structure based sequence alignment of *B. subtilis* MutS and other MutS homologs for which the three-dimensional structure are known. Conserved hydrophobic (blue), polar (green), positive- (purple) and negative-charged (red) residues are highlighted. The secondary structure elements are color-coded for domains II, domain III and transmitter as in **A**. The location of the QQ and FF motifs is indicated with teal and orange carets, respectively. **(C)** Sequence alignment of MutS from Gram-positive bacteria shows conservation of F319 and F320.

When we align MutS sequences from Gram-positive bacteria, many of which cause serious health concerns including *Staphylococcus aureus* and *Listeria monocytogenes*, we find that these residues are highly conserved in *mutS* homologs (Figure 2.4C). In some Gram-positive bacteria, a few accepted substitutions are tolerated at these positions, such as the aromatic residue tyrosine or the hydrophobic residue isoleucine (Figure 2.4C). Based on our results, we suggest that mutation of these conserved residues could eliminate MMR function in related pathogenic bacteria, increasing mutagenesis and altering antibiotic susceptibility and persistence within the host environment.

2.5.7 mutSF319SF320S is defective for recruitment of MutL *in vivo*

It has been previously shown that MutL-GFP forms foci in response to spontaneous or 2-aminopurine (2-AP) formed mismatches detected by MutS, providing an *in vivo* assay to monitor MutL recruitment in response to mismatch detection by MutS (Lenhart et al., 2013; Smith et al., 2001). A caveat with this assay is that the *mutL-gfp* allele is nearly defective for MMR as measured by mutation rate (Smith et al., 2001), however focus formation of MutL-GFP is dependent on *mutS,* providing a single cell assay for MutL-GFP recruitment in live cells (Lenhart et al., 2013; Smith et al., 2001). We asked if MutSF319SF320S was able to recruit MutL-GFP into foci in cells grown with 2-AP. In a background with the native *mutS* gene, we observed MutL-GFP repair centers in ~25% of cells (Figure 2.5A). We found that cells with *mutSF319SF320S* or the Δ*mutS* allele did not support MutL-GFP focus formation, as MutL-GFP repair centers

only formed in ~3% of the cell population in both genetic backgrounds. Furthermore, we also found that MutSF319SF320S is defective for recruitment of MutL-GFP in response to mismatch detection *in vivo*, supporting the *in vitro* experiments showing that *mutSF319SF320S* is defective for interaction with MutL.

2.5.8 MutSF319SF320S forms large repair complexes *in vivo*, supporting a model for persistent loading

With a MutS variant defective in recruitment of MutL, we can now uncouple mismatch binding from functional repair and "trap" repair intermediates that would normally be resolved during repair. To observe mismatch repair intermediates, we fused *mutS* to a monomeric *gfpmut3* variant (*gfpmut3* referred to herein as *gfp*) since *gfpmut3* represents the most monomeric derivative of GFP, providing the least invasive method for observing protein localization in living bacterial cells (Landgraf et al., 2012). We constructed a native locus *mutS-gfp* strain by allelic exchange in order to maintain expression of the downstream gene *mutL* under its native promoter (Figure 2.5B). The *mutS-gfp* background maintained ~85% of MMR activity (mutation rate $2.56X10^{-8}$ [2.0-3.1]), providing a functional fusion to observe active repair in real time. Upon mismatch detection, MutS-GFP forms complexes in response to mismatches in order to orchestrate repair. The *mutS-gfp* strain forms repair centers in ~9% of cells within the population during exponential growth, and repair center formation is stimulated to ~42% of cells by addition of 2-AP to the growth media (Figure 2.5C, D, and E) (Dupes et al., 2010; Klocko et al., 2011; Lenhart et al., 2013; Simmons et al., 2008). Thus, using *B. subtilis*, we can bridge biochemical and genetic data to understand how disruption of MutL recruitment by

MutS alters repair center dynamics *in vivo*, providing important mechanistic insight into intermediate steps.

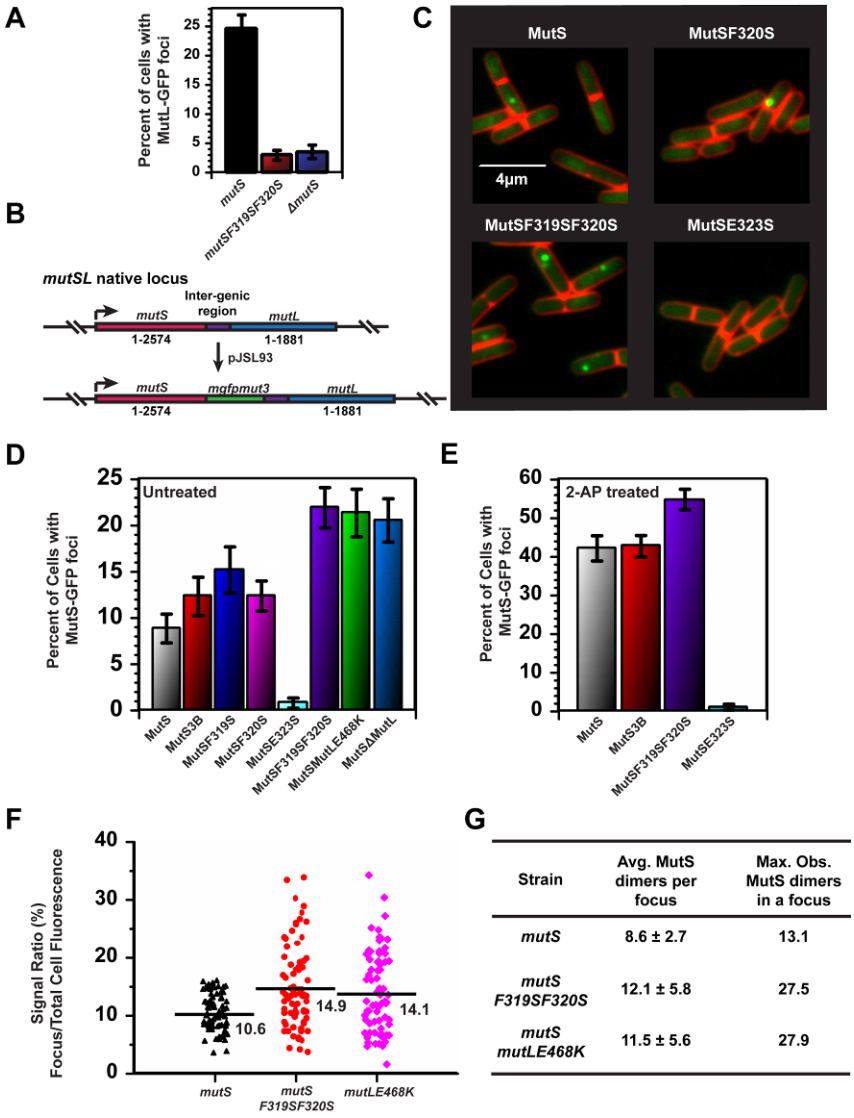


Figure 2.5. MutS mutants defective for MutL interaction form persistent complexes *in vivo*. (A) Fluorescent single cell microscopy of MutL-GFP repair centers responding to mismatch formation in a *mutS, mutSF319SF320S* or *ΔmutS* background (n=1320, 1559, and 988 cells scored). Cells were treated with 600 μg of 2-AP and incubated for 1 hour prior to imaging. 95% confidence intervals are shown. **(B)** Shown is a schematic for cloning an unmarked in frame fusion of *mutS* to *mutS-22-mgfpmut3* (*mutS-gfp*), while *Figure 2.5 continued on page 65…*

Figure 2.5 continued from page 64…
maintaining expression of the *mutL* gene downstream ("Experimental Procedures"). **(C)** Representative micrographs of the indicated MutS-GFP fusion proteins. The cell membrane was imaged using the vital membrane stain TMA-DPH, which was pseudo-colored red. **(D)** Shown is a bar graph of the percent of cells with each MutS-GFP fusion untreated during exponential growth (n=1276, 957, 796, 1568, 1008, 1382, and 1148 respectively refers to the number of cells scored for each strain). Error bars represent 95% confidence intervals (CI). All groups are statistically significant with respect to MutS, including MutS Patch 3B (P= 0.0038) and MutSF320S (P= 0.0013). The strain with *ΔmutL* has a *mutS-mgfpmut2* fusion. **(E)** The percent of cells with the indicated MutS-GFP fusion following challenge with 2-AP (n=879, 1212, 711, and 725 are the number of cells scored respectively) the error bars represent 95% CI. **(F)** Focus intensity was determined by normalizing total signal of the repair centers to the total cell fluorescence. A total of 75 MutS-GFP foci were analyzed for each group. All foci examined were in cells statistically equivalent in regards to area, length, and average intensity of cellular fluorescence. **(G)** Using the number of MutS molecules per cell (Figure S2.6) and the average MutS repair center fluorescent intensity F), we were able to determine the average number of MutS dimers per repair center ± standard deviation, as well as the highest observed number of MutS dimers within repair centers scored.

We subsequently fused *gfp* to *mutS3B*, *mutSF319S*, *mutSF320S*, *mutSF319SF320S*, and *mutSE323S* and found that all strains except for *mutSE323S-gfp* formed repair complexes *in vivo* (Figure 2.5C). Interestingly, MutSE323S-GFP, which was defective for repair *in vivo*, was also completely defective for focus formation suggesting that although this protein accumulates *in vivo* (Figure 2.3A), the E323S mutation appears to cause some defect other than blocking MutL interaction, since it failed to form a repair complex. The MutSE323S variant was not amenable to recombinant expression and purification and therefore we did not further pursue characterization of this variant (data not shown).

MutSF319S-GFP, MutSF320S-GFP, and MutSF319SF320S-GFP all formed foci in a higher percentage of untreated cells than the MutS-GFP control. We hypothesized that the increase in focus formation is a consequence of an increase in the duration of repair center formation due to unproductive repair caused by a failure to properly signal for MutL. Another possibility is that there is an increase in mismatch detection, however, we ruled out this possibility by showing that a deletion of *mutL* downstream of *mutS-gfp* causes the same effect by increasing MutS-GFP repair centers *in vivo* (Figure 2.5D). Furthermore, since the error rate of the replication process in the absence of functional MMR is one mispair every two rounds of replication the likelihood of closely spaced errors is extremely low [(Bolz et al., 2012; Lenhart et al., 2013) and Table 2.1]. Time-lapse imaging of repair center formation and resolution would be preferred to support our hypothesis, but is not feasible due to long exposure times of the MutS-GFP fusions and rapid photobleaching dynamics (data not shown). MutSF319S-GFP, MutSF320S-GFP, and MutS3B-GFP formed repair complexes in a nearly indistinguishable percentage of cells (12-13% of the population) (Figure 2.5D). The double mutant, MutSF319SF320S-GFP, shows an increase in the percentage of cells with MutS-GFP foci above our measurements for each of the single variants (Figure 2.5D). Furthermore, 2-AP treatment elicited an increase in the percentage of cells with MutSF319SF320S-GFP, showing that this variant still binds mismatches and initiates repair, further supporting our *in vitro* results that mismatch binding is unaffected (Figure 2.5E & Figures 2.3D). Ultimately, loss of MutL recruitment by MutS causes a corresponding increase in the percentage of cells with MutS repair complexes.

We also asked if MutS repair center formation is not only affected by MutL recruitment, but also by the next step of repair--incision. To do so, we asked if MutS-GFP repair centers accumulate in cells where MutL endonuclease nicking is prevented using the *mutLE468K* allele (Pillon et al., 2010). Nicking by MutL is a required step for repair and we have previously shown that the E468K substitution eliminates MutL endonuclease activity *in vitro* and MMR activity *in vivo* (Pillon et al., 2010). Indeed, the percentage of cells with MutS-GFP foci increased in the *mutLE468K* background to levels observed in both the *mutSF319SF320S* and the *ΔmutL* backgrounds, indicating that if MutL-directed nicking is prevented, MutS-GFP foci persist when the next step of repair is blocked (Figure 2.5D). With these data we argue that MutL recruitment is not sufficient to halt MutS loading *in vivo* per se, but that timely repair of the mismatch is required to prevent further loading.

We also found that a proportion of repair centers exhibited high fluorescence intensity in backgrounds defective for MutL recruitment and MutL endonuclease activity (Figure 2.5C, F and G). We quantified percent focus intensity relative to whole cell fluorescence intensity. In doing so, we found that many repair centers associated with MutSF319SF320S had elevated focus intensities relative to a MutL recruitment proficient MutS-GFP strain (Figure 2.5F). These data suggest that more MutS protomers are present in a focus for MutS mutants defective in MutL interaction or in strains where MutL function has been eliminated by blocking incision (*mutLE468K*) (Figure 2.5F). We also

analyzed MutSF319SF320S foci in cells where *mutL* expression was induced and observed no difference in percent of cells with foci or focus intensity (data not shown).

We quantified the number of MutS dimers found within *B. subtilis* under the exact conditions used during live cell imaging, and found that in *B. subtilis* steady state levels of MutS are ~80 dimers per cell (100 nM) (Figure S2.6). Using these data, we determined that the mean number of MutS dimers in a repair center was ~8.5 (this corresponds to 17 GFP moieties) (Figure 2.5G). Both the MutL recruitment and endonuclease-deficient backgrounds contained a higher mean number of MutS-GFP dimers per repair center (12.1 and 11.5 respectively) (Figure 2.5G). The increase in repetitively loaded MutS-GFP dimers is more pronounced in the broad distribution of individual intensity measurements of the repair centers, with as many as >3 fold (~30 MutS-GFP dimers; ~35% of cellular MutS) more molecules in the highest intensity MutS complexes observed in repair deficient strains. These observations support a model where MutS can load iteratively at a mismatch, increasing the local concentration of MutS. We propose that iterative MutS loading aids in efficient MutL recruitment to the mismatch, providing *in vivo* support for *in vitro* observations (Acharya et al., 2003). We also find it interesting that we quantify 8.6$\pm$2.7 MutS dimers per focus and in *S. cerevisiae* the number of Msh6 dimers per focus was determined to be 10.8$\pm$4.4 (Hombauer et al., 2011). Therefore, the stoichiometry of MutS within a focus is remarkable similar between these two organisms.

2.5.9 mutSF319SF320S repair centers localize away from the replisome

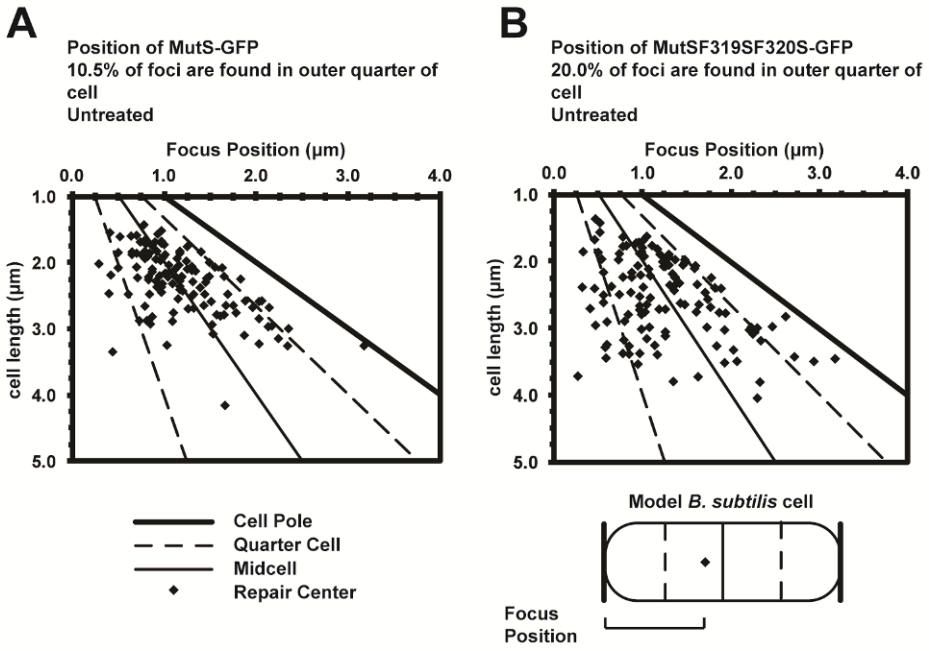
During DNA replication, chromosomal DNA is replicated within an organized replisome (Berkmen and Grossman, 2006; Lemon and Grossman, 1998; Lemon and Grossman, 2000). Here, the replisome is defined as replication associated proteins that localize as discrete foci *in vivo*. Within *B. subtilis*, replisomes maintain a well-characterized subcellular position (Berkmen and Grossman, 2006; Lemon and Grossman, 1998; Lemon and Grossman, 2000). Once replication is initiated from the single origin of replication (*ori*), two sets of replication forks are often contained within a single replisome predominantly found at midcell (Berkmen and Grossman, 2006; Lemon and Grossman, 1998; Lemon and Grossman, 2000). Once replicated, the daughter chromosomes begin to translocate to the cell poles, taking mismatched DNA away from the centrally located replisome. We have previously shown that MutS foci colocalize to the replisome preceding mismatch detection and are released following mismatch binding (Lenhart et al., 2013). Therefore, we asked if localization of MutS repair complexes is altered when MutS is broken for MutL recruitment.

In order to test if the MutSF319SF320S repair centers persist at the site of mismatch identification, we monitored their position during DNA replication in minimal medium under slow growth conditions (~ 123 min. doubling time). Slow growth maintains approximately half of the cell population with a single replisome focus (~52% of cells). We first determined the distance of the MutS and MutSF319SF320S repair centers relative to the cell poles (Figure 2.6A). MutS-GFP repair centers maintain a

mostly midcell position with 48.4% found within the middle 10% of the cell. Only 10.5% of these repair centers occupy a distal position within the outer quarters of the cell. Relative to the distribution of MutS-GFP, MutSF319SF320S-GFP position was more dispersed, as only 27.2% of repair centers were found within the middle 10% of the cell (Figure 2.6B). About 2 fold more MutSF319SF320S repair centers (20%) were found in the distal quarter of the cell. These data support the hypothesis that upon identifying a mismatch at the replisome, the assembled MutS repair center is maintained at the site of the mismatch for extended periods of time, causing migration away from the replisome as DNA synthesis continues, an effect more pronounced when MutL recruitment is blocked.

To further test this hypothesis, we examined colocalization between the MutS-GFP repair centers and replisomes during the same slow growth conditions described above. Colocalization between MutS-GFP and DnaX-mCherry (a component of the processivity clamp loader complex) was performed and scored as described (Lenhart et al., 2013). During exponential growth, MutS-GFP forms repair complexes that colocalize with the replisome in about 51% of cells. When we stimulate mismatch formation by adding 2-AP to the media, we found a decrease in colocalization to ~35% ($p=0.00052$), consistent with previous results (Figure 2.6C) (Lenhart et al., 2013). During exponential growth, MutSF319SF320S repair complexes colocalize with the replisome in 38% of the population; a significant decrease compared to MutS-GFP during exponential growth ($p=0.0090$). Upon treatment with 2-AP, only 29% of repair complexes colocalize with the replisome. With these results we conclude that when MutS-GFP is unable to recruit MutL

to the site of a mismatch, repetitive loading of MutS-GFP at the mismatch will continue, resulting in a brighter and more persistent MutS-mismatch complex, which migrates away from the replisome as replication continues.



C

Colocalization of MutS-GFP foci with the Replisome (DnaX-mCherry)

Fusion	Addition of 2-AP	% colocalization (+/- 95% CI)	No. of cells analyzed	
MutS-GFP	-	51.3% (7.8)	158	*
	+	35.0% (5.8)	260	* #
MutSF319SF320S-GFP	-	38.2% (7.4)	165	**
	+	29.4 % (6.4)	197	**

Figure 2.6. MutSF319SF320S foci persist on DNA away from the replisome in the absence of MutL recruitment. The position of repair centers for **(A)** MutS-GFP and **(B)** MutSF319SF320S-GFP within each cell was plotted by the coordinates (cell length, distance to pole). Solid black line indicates midcell, whereas dashed lines indicates the quarter cell positions. The thick black line indicates the cell end. n=125 **(C)** Table indicating colocalization values for MutS-GFP with DnaX-mCherry. The number of cells scored is indicated (n). p-values: *=0.00052, **=0.040, #=0.0090, and difference between the 2-AP treatment groups =0.105.

2.5.10 MutL crosslinks with MutS independent of mismatch detection *in vivo* and *in vitro*

An outstanding problem in MMR is understanding how MutL senses when MutS is mismatch bound to initiate downstream steps of repair. Previously, we showed that the *mutSF30A* allele, supports formation of more MutL-GFP repair centers than are observed in the *ΔmutS* background (Lenhart et al., 2013). MutSF30A is a variant that is unable to distinguish mismatched DNA from complementary DNA (Lenhart et al., 2013). This observation is interesting because it suggests that MutS can interact with MutL, in the absence of mismatch binding *in vivo*, even though the interaction is reduced (Lenhart et al., 2013). Here, we directly test the hypothesis that MutL can transiently probe MutS for the appropriate conformational change to initiate MMR. To test this hypothesis, we used immunoprecipitation (IP) targeting MutS to co-IP any proteins associated with MutS *in vivo*. Since the MutS•MutL interaction is transient in nature, we employed the use of the thiol-cleavable, membrane permeable crosslinker Dithiobis[succinimidyl propionate] (DSP) to crosslink MutS•MutL complexes formed in growing cells (Figure 2.7).

The IP was accomplished under normal growth conditions in the absence of 2-AP to test for association in the absence of active MMR. Using this procedure, we were able to IP ~10% of the intracellular MutS. Importantly, we were able to capture the MutS•MutL interaction in the wild type strain, yet failed to IP MutL in the *ΔmutS* strain, validating the requirement of MutS for successful co-IP of MutL. We were able to detect a MutL band (0.02% of input) in the IP lane from the wild type strain. The low amount of

MutL recovered in the wild type strain is likely because we are precipitating only 10.0% of intracellular MutS, as well as we expect only 9% of cells to have ongoing MMR as determined by the assembly of active MutS-GFP repair centers. In agreement with our *in vitro* data and the MutL-GFP microscopy (Table 2.2 and Figure 2.5A), we recovered low amounts of MutL in the IP lane from the *mutSF319SF320S* lysate (<0.001% of the input), confirming that MutSF319SF320S is compromised for interaction with MutL *in vivo* (Figure 2.7). In Figure 2.7, we also present error measurement from three independent IP experiments. In the other experiments performed we did not recover any detectable amount of MutL in the MutSF319SF320S lysate further supporting our conclusion that this mutant does not interact with MutL (data not shown).

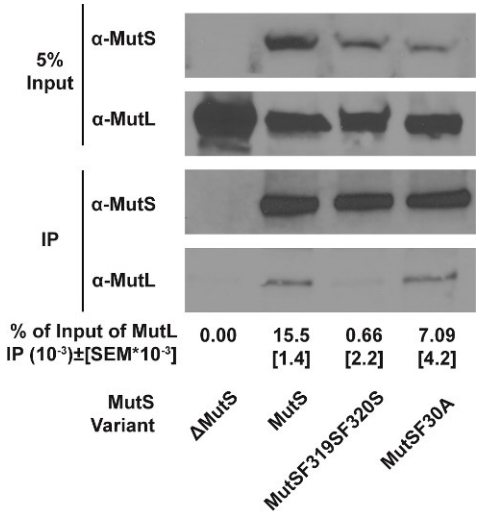


Figure 2.7. MutS crosslinks with MutL in the absence of mismatch detection *in vivo*. Co-immunoprecipitation of MutS and MutL in the indicated backgrounds with affinity purified polyclonal antibodies against MutS. MutS and MutL levels were probed for using antiserum directed against MutS and for MutL. Band intensity was determined by using ImageJ quantitation software (See Experimental Procedures). Relative IP MutL levels reflect absolute band intensity per lane normalized to the wild type MutS lane. The error (SEM) were calculated from 3 independent experiments.

We then tested whether mismatch detection was necessary to facilitate MutS•MutL interaction *in vivo*, speculating that MutL may frequently probe MutS for the

appropriate protein conformation, signaled by mismatch detection. To test this, we used MutSF30A: a MutS variant capable of DNA binding, yet incapable of discriminating mismatched DNA from complementary DNA (Lenhart et al., 2013). When MutSF30A was immunoprecipitated, we found that we successfully captured MutL (~0.007% of input). This result shows that the MutS•MutL interaction may dynamically occur independent from mismatch identification *in vivo*, suggesting that MutL is capable of transiently "checking" to determine if MutS is mismatch bound before licensing downstream repair events. Similar observations have been seen in *S. cerevisiae in vitro* showing that MutSα interaction with MutLα is not entirely mispair dependent (Kijas et al., 2003).

2.6 Discussion

Here, we have identified a conserved MutL binding site on MutS in the Gram-positive bacterium *B. subtilis*. Using peptide array mapping, extensive mutagenesis, single-cell fluorescence microscopy and *in vitro* crosslinking studies, we have identified residues found within the core domain, important for MMR *in vivo* and interaction with MutL-NTD *in vitro*. This site was further refined to a discrete MutL docking site composed of adjacent phenylalanine residues F319 and F320. Substitution of both phenylalanines to serine completely eliminates MMR *in vivo* and is defective for crosslinking to MutL-NTD *in vitro*. We also show that purified MutSF319SF320S is similar to wild type MutS for dimerization, ATPase activity, and binding to mismatched DNA substrates. We can therefore attribute the loss of MMR *in vivo* to a failure in MutL

binding and recruitment. To our knowledge this effort defines the first MutL binding site on MutS in a bacterial organism lacking a methylation-directed MMR pathway.

Importantly, the di-phenylalanine motif that we identified in *B. subtilis* MutS to mediate interaction with MutL appears to be conserved and is part of a larger S[X]FF motif known to mediate MutL interaction with eukaryotic proteins. In *S. cerevisiae,* the S[X]FF motif was shown to be important for interaction between eukaryotic MutL homolog (Mlh1) and several Mlh1 bindings partners including Exo 1, BLM and Sgs1 proteins (Dherin et al., 2009). Furthermore, a S[X]FF motif was also shown to mediate interaction between MutSβ (Msh2-Msh3) and MutLβ (Mlh1-Pms2), for the human proteins (Iyer et al., 2010). In addition, a di-phenylalanine motif has been shown to be critical for interaction between mammalian translesion polymerases Rev1 and pol κ (Wojtaszek et al., 2012a; Wojtaszek et al., 2012b). Our results in consideration with those above, show that adjacent phenylalanine residues play important roles in mediating protein interactions in a wide-variety of organisms.

Previous analysis of the *E. coli* MutS•MutL interaction identified residues in the mismatch recognition and connector domains involved in this interaction (Mendillo et al., 2009; Winkler et al., 2011). However, the connector domain of *E. coli* MutS on its own only weakly interacts with MutL, suggesting that additional surfaces on MutS may be involved in this interaction. While the mismatch recognition and connector domains are in close proximity to the core domain, the residues identified previously in *E. coli* MutS

(Q211 and Q212) and patch 3B (F319 and F320) reside in opposite faces of the monomer and are separated by the allosteric transmitter that connects the mismatch- and ATP-binding domains (Figure 2.8). It is conceivable that the different techniques used in our study and that by Mendillo and co-workers may have revealed distinct anchoring points of the MutS•MutL interface. If true, the MutS•MutL complex could adopt two distinct architectures. MutL could interact with both protomers of the MutS dimer to form a productive complex (arrow ii in Figure 2.8), thereby implying a mechanism to "check" MutS for mismatch binding through contacts with the mismatch-binding domain that would support the distance restraints reported by Winkler and co-workers (Winkler et al., 2011).

Alternatively, MutL could interact with a single protomer of the MutS dimer (arrow i in Figure 2.8). This model poses an attractive mechanism to sense the mismatch and nucleotide binding states of MutS. The mismatch- and the nucleotide-binding domains of MutS are connected by a transmitter helix that runs along the outer rim of the MutS protomer (Figure 2.8) (Obmolova et al., 2000). Therefore, if MutL binds this face of MutS, the transmitter helix is probably a central feature of the interaction interface. This model also supports the established idea that only one of the MutS protomers mediates the interaction with MutL (Habraken et al., 1997; Mendillo et al., 2009; Prolla et al., 1994). Interestingly, mutation of patch 3A, which is part of the MutS transmitter, does not affect mismatch repair *in vivo* (Table 2.1 and Figure 2.1). However, the MutS3A variant could not be over-expressed recombinantly in *E. coli*, implying a stability defect that

could result from mutation of the transmitter. This, in turn, implies that MutL senses a different region of the transmitter, potentially α-helix 10 located in the C-terminus of the core domain (Figures 2.4 and 2.8).

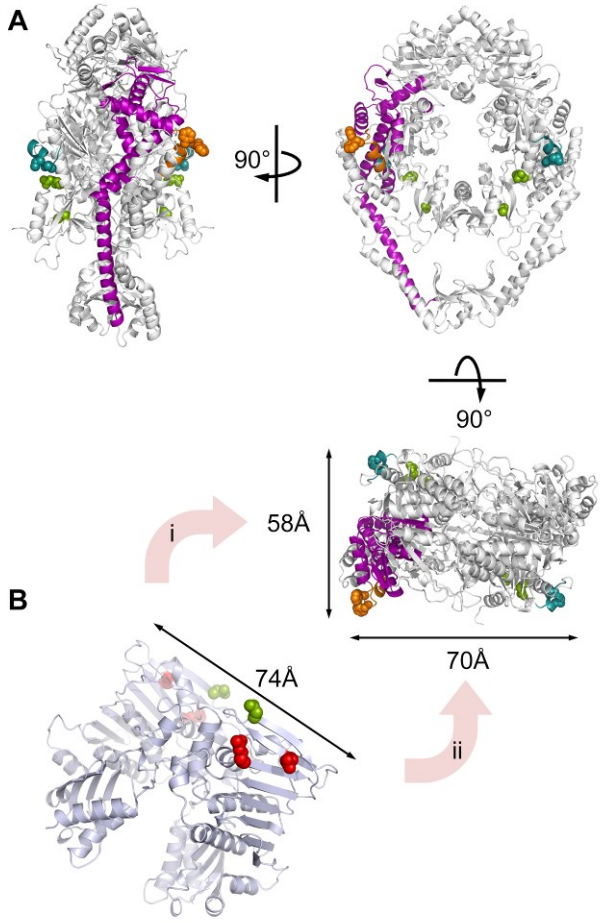


Figure 2.8. Potential interfaces of the MutS•MutL complex. (A) Orthogonal views of the *B. subtilis* MutS dimer shown as a ribbon diagram with the residues deemed important for the interaction with MutL shown in orange (FF motif, this work), teal (QQ motif, (Mendillo et al., 2009)) or green (distance constraints identified by crosslinking, (Winkler et al., 2011)). The transmitter region of MutS is highlighted in purple. **(B)** Ribbon diagram of the *E. coli* MutL-NTD dimer shown as a ribbon diagram with the residues identified in crosslinking studies shown in green (Winkler et al., 2011) and additional residues deemed important for the interaction with MutS shown in red (Plotz et al., 2006). Dimensions of the MutS and MutL dimer surfaces are indicated in angstroms (Å) and the two potential surfaces of MutS that MutL could recognize are indicated with pink arrows and labeled i and ii, respectively.

Upon identifying the binding interface, we extended our study to investigate the dynamic nature of MutS repair complexes *in vivo*. We present data showing that

disruption of MutL recruitment causes repetitive MutS loading in response to mismatch formation. Since MutL recruitment is blocked we interpret this to mean that a MutS intermediate is "trapped" because the downstream step is prevented and we find that MutS foci persist with the number of MutS dimers per focus increased. These results support a model for repetitive loading by MutS in response to mismatch formation. Furthermore, even upon successful recruitment of MutL to a mismatch, we show that loss of function due to disruption of the endonuclease active site phenocopies a *ΔmutL* allele, supporting the hypothesis that not only does MutS loading occur independent from MutL recruitment, but that endonuclease directed nicking, and presumably excision of the mismatch, is a critical feature to disassemble MutS complexes. As more dimers of MutS load onto the mismatch proximal DNA, more MutS is available to recruit MutL. In support of this hypothesis, real-time *in vitro* imaging of MutSα and MutLα on DNA curtains revealed that the interaction requires a mismatch, yet interaction between MutSα and MutLα may occur after MutSα formed the ATP hydrolysis-dependent sliding clamp (Gorman et al., 2012). Therefore, even mismatch-dissociated MutSα dimers can still facilitate a MutLα interaction, in essence amplifying a signal for MutL recruitment and for the advancement of repair. Our experiments represent *in vivo* data supporting repetitive loading of MutS at a mismatch, supporting previous *in vitro* experiments showing repetitive loading using circular mismatch containing substrates (Acharya et al., 2003; Gradia et al., 1999). In addition to providing evidence for repetitive loading *in vivo*, we also provide evidence that mismatch excision is an important step in disassembling MutS repair complexes.

After excision of the mismatch, MutS loading is halted and the already bound MutS dimers will dissociate from the DNA in a timely manner, leading to disassembly of the repair center. *In vitro*, single molecule imaging reveals that after lesion recognition, the newly formed MutSα sliding clamp will remain on DNA with a lifetime of $t_{1/2} \geq 198 \pm 23.4$s (Gorman et al., 2012). If loading is restricted to nascent DNA, then defective repair centers should persist on DNA flanking mismatches for up to 10 minutes after initial mismatch recognition. Furthermore, as DNA replication continues, newly replicated DNA moves farther away from the replisome, taking newly formed mismatches with it. In support of the hypothesis that MutS repair centers defective for MutL recruitment persist longer on DNA surrounding the mismatch, the distribution of these MutS centers are located farther away from the replisome than repair centers engaging in active repair. Moreover, in exponentially growing cells, less repair centers colocalize with predominantly midcell replication centers. These two observations support the hypothesis that unproductive MutS repair centers persist on mismatch proximal DNA.

In *B. subtilis*, a model is emerging for the early steps of mismatch repair *in vivo*. We propose that MutS is positioned at the replisome preceding mismatch detection by a DnaN clamp zone that results from Okazaki fragment maturation (Figure S2.7A) (Lenhart et al., 2013). MutS binds free DnaN clamps via a DnaN-binding motif (^{806}QLSFF) found in the unstructured C–terminal clamp-binding domain. MutS is able to find ~90% of mismatches through a DnaN coupled mechanism. Once MutS detects a mismatch, we

propose that MutS (Figure S2.7B) loads repetitively at the mismatch, producing numerous DNA-bound MutS dimers (Figure S2.7C). We propose that repetitive MutS loading facilitates efficient MutL recruitment by increasing the local concentration of DNA bound MutS dimers surrounding the mismatch. MutS diffusing away from a mismatch with MutL may also help MutL identify strand discontinuities necessary to direct incision to the nascent strand (Figure S2.7D). Finally, the data presented here support the model that MutL-incision is necessary for disassembly of MutS complexes suggesting that mismatch excision is important for preventing further MutS loading.

Overall, this work describes the interaction between the core domain of MutS and MutL both *in vitro* and *in vivo*, and the implications of this interaction for the recruitment and activation of MutL at MutS repair centers, providing insight into the intermediate steps of mismatch repair in live cells.

2.7 Acknowledgements

We thank members of the Simmons and Guarné laboratories for helpful discussions and comments on this manuscript. We would also like to thank Nicholas Bolz for assistance in early cloning and mutagenesis assays. This work was supported by grant MCB1050948 from the National Science Foundation to L.A.S. and by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to A.G. In addition, M.C.P was supported in part by a doctoral scholarship from NSERC

and J.S.L was supported in part by a predoctoral fellowship from the Rackham Graduate School at the University of Michigan.

2.8 Supplementary Procedure

2.8.1 Quantitative Western blotting (LiCOR) analysis.

Whole cell lysates were prepared from independent JSL364 (wild-type) and JSL281 (*ΔmutS*) strains in 6 mL cultures ($S7_{50}$ minimal media supplemented with 2% D-Glucose) grown at 30˚C and harvested at a normalized OD_{600} of 0.5, while simultaneously plated for viables (10^{-6} dilution, see *Spontaneous mutation rate analysis* in the main text*)*. Cells were pelleted and incubated in 1 mL of lysis buffer (10 mM Tris HCl [pH 7.5], 1 mM EDTA, 10 mM $MgCl_2$, 1 mM AEBSF, 0.5 mg/mL lysozyme, and 0.1mg/mL DNase I) and incubated at 37˚C for 10 minutes. After incubation, SDS was added to a final concentration of 1% to lyse cells. Cells were heated for 5 minutes at 100˚C and lysates concentrated to a known final volume in a 10 kDa concentrator column (Amicon Ultra Centrifugal Filters, Millipore).

Immunodot blotting was performed essentially as described (Klocko *et al.*, 2011) and as described in "Material and Methods" in the main text. Briefly, whole cell lysates were immobilized onto a nitrocellulose membrane via wet transfer using the mini Trans-Blot electrophoresis transfer cell in transfer buffer without SDS (Bio Rad). The membrane was incubated in blocking buffer (5% milk solids, 17.4 mM Na_2HPO_4,

2.6 mM NaH_2PO_4, 150 mM NaCl) at 22˚C for one hour. All subsequent washes and incubations took place in blocking buffer. After blocking, the membrane was incubated with primary antisera α-MutS (MI 1042) in blocking buffer (minus tween-20) overnight at 4˚C with constant agitation. The next morning, the blot was washed three times for 15 minutes each in blocking buffer supplemented with 0.05% Tween-20. After washing, the blots were then incubated in the dark in 1:15,000 Odyssey Goat anti-Rabbit IR Dye 800CW (926-32211, LiCOR Biosciences) at 22˚C for 2 hours in blocking buffer. All subsequent steps were performed in the dark. The blot was then washed 3 more times in blocking buffer with 0.05% Tween-20, followed by a wash in PBST (17.4 mM Na_2HPO_4, 2.6 mM NaH_2PO_4, 150 mM NaCl, 0.05% Tween-20) to remove excess milk solids. Membranes were dried for 2 hours followed by exposure using an Odyssey CLx Infrared Imaging System (LiCOR, Lincoln, Nebraska). All data analysis and band quantifications were performed using the Odyssey CLx software.

2.9 Supplementary Figures

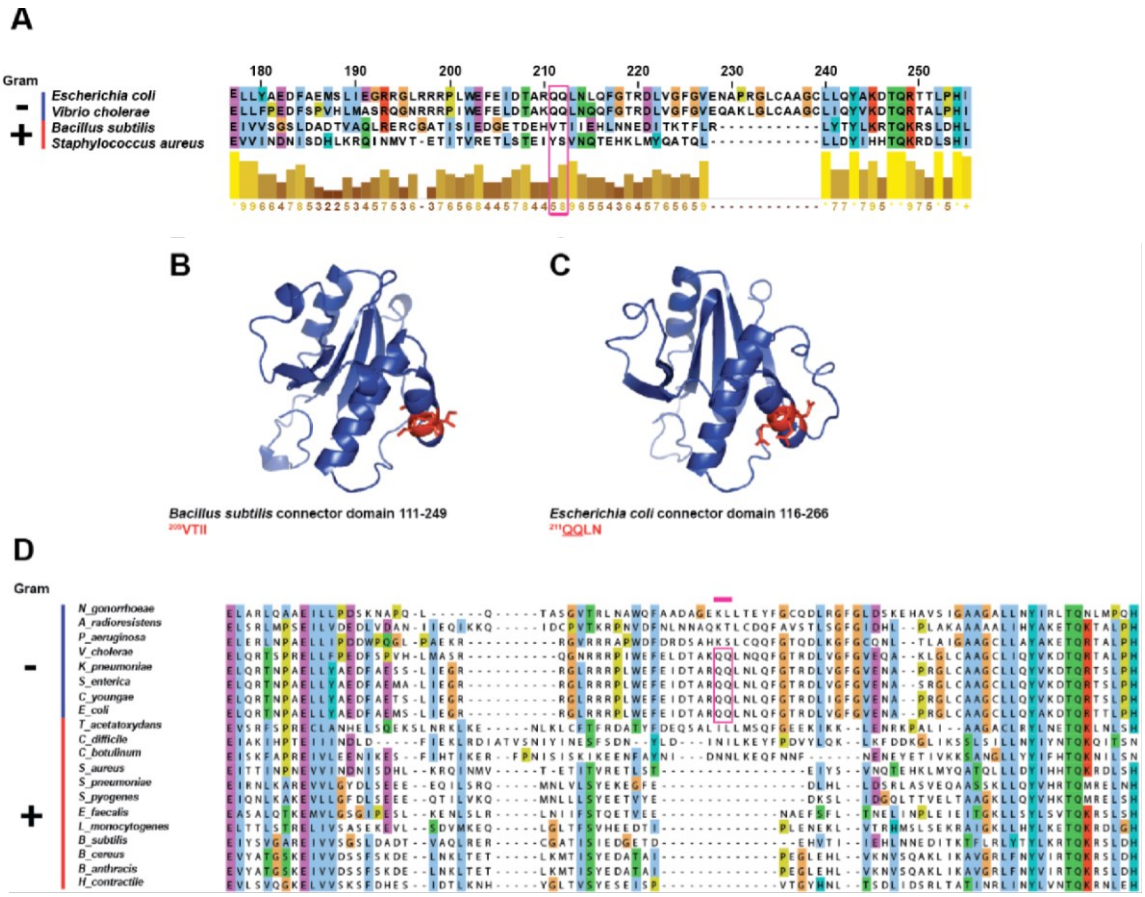


Figure S2.1. The *E. coli* MutS di-glutamine (Q211 and Q212) binding site for MutL is not conserved in *B. subtilis* MutS. (A) A sequence alignment directly comparing the MutS amino acid sequence of the Gram-negative bacteria *E. coli* and *V. cholerae* to the MutS amino acid sequence of the Gram-positive bacteria *B. subtilis* and *S. aureus*. The region surrounding the diglutamine MutL docking site (underlined in pink) is shown. The alignment was generated using the http://www.ebi.ac.uk/Tools/msa/clustalw2/ server. The residue numerical designations shown above the alignment are relative to the *E. coli* amino acid sequence. Protein structure models of the **(B)** *B. subtilis* (Phyre2 server model) and the **(C)** *E. coli* connector domain (PDB file 1E3M). Shown in red are the residues corresponding to either the ^{205}VTII site in *B. subtilis* or the ^{211}QQLN of *E. coli*. **(D)** A sequence alignment generated employing the http://www.ebi.ac.uk/Tools/msa/clustalo/ server to visualize the conservation of the QQ site in *E. coli* shown in both Gram-positive and negative bacteria. The order within the alignment is based on a phylogenetic organization of aligned MutS homologs using a Neighbour-joining tree without distance corrections.

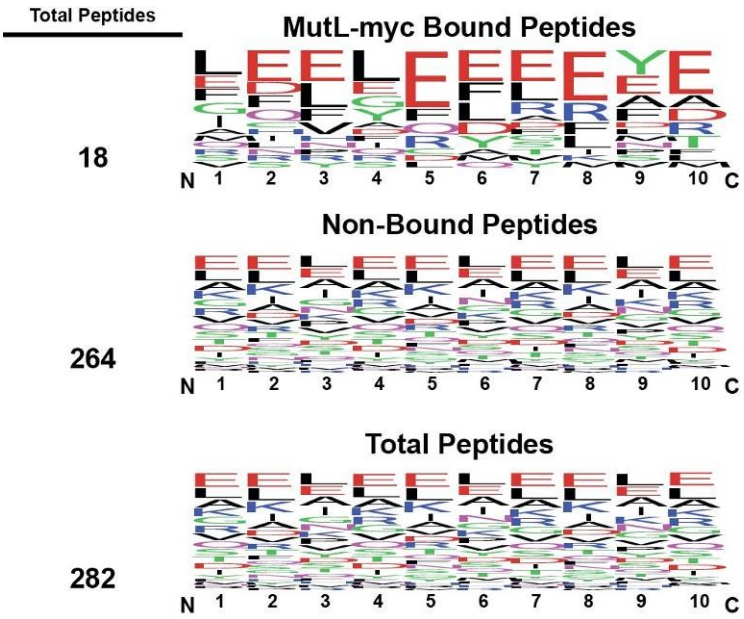


Figure S2.2. Amino acid composition of MutS peptides recognized by MutL. Analysis of the amino acid composition of MutL interaction peptides, MutL non-bound peptides, and total MutS peptides in the peptide array. Results visualized with WebLogo 3.1.

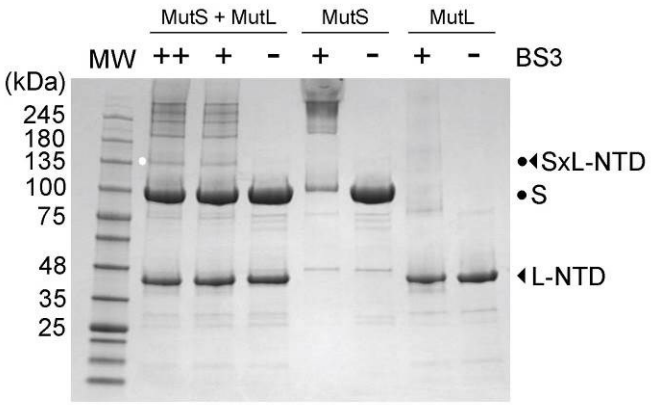


Figure S2.3. Crosslinking of wild type MutS to the N-terminal domain of MutL with homoduplex DNA. Mixtures of each protein, 10 mM ATP, and a 90 base-pair homoduplex DNA were incubated with the chemical crosslinker BS3 (+=0.8 mM and ++=1.6 mM, respectively). Protein complexes were then separated on a 4-15% gradient SDS polyacrylamide gel. The bands corresponding to the MutS and MutL-NTD monomers, as well as the MutS•MutL-NTD complex are labeled.

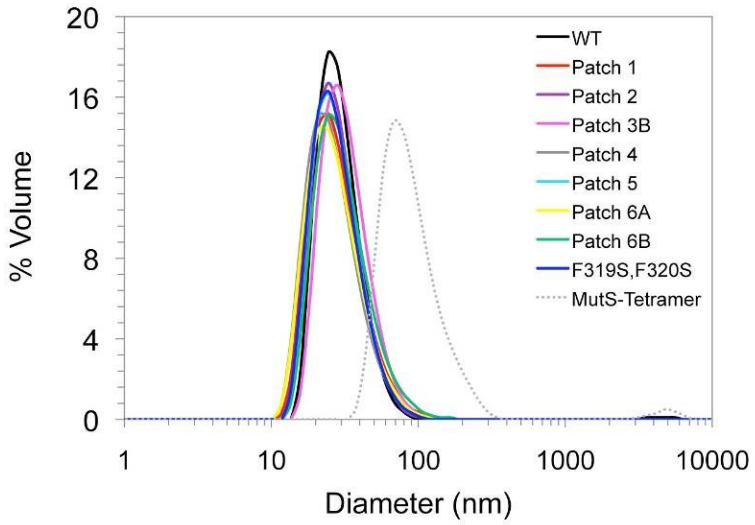


Figure S2.4. Characterization of the particle size distribution of *B. subtilis* MutS variants at 10 µM (dimer) and 220 µM (tetramer) concentrations reveal similar oligomeric status. Samples were centrifuged at 15,700 x g for 10 minutes at 4˚C and measured using a Zetasizer Nano S (Malvern Instruments) with a 4 mW He-Ne laser at 633 nm. All measurements were taken using a 12 µL quartz cell (ZEN2112) at 4°C.

Bacillus subtilis MutS/ 306-330	EERQEMVETLMSHFFEREDLRERLK
Homo sapiens MSH3/ 14-38	SSSAPARQAVLSRFFQSTGSLKSTS
Saccharomyces cerevisiae MSH3/ *26-50*	LLTIMAGQPTISRFFKKAVKSELTH
Homo sapiens BLM/ 1308-1332	AEELDEEIPVSSHYFASKTRNERKR
Homo sapiens EXO1/ 493-517	ESGAVVVPGTRSRFFCSSDSTDCVS
Saccharomyces cerevisiae Ntg2/ 13-37	IPVDIEEVEVRSKYFKKNERTVELV
Saccharomyces cerevisiae Sgs1/ 1372-1396	SNGIAQSTGTKSKFFGANLNEAKEN
Concensus sequence	--S[HKR][YF]F--

Figure S2.5. Overlay of the *B. subtilis* MutL binding site on MutS with the Mlh1 binding site on its binding partners. Shown is an overlay of the *B. subtilis* MutL binding site on MutS with the MIP box (Mlh1 Interacting Protein box_[R/K]-S-[H/R/K]-[Y/F]-F) reveals a conserved serine followed by the di-phenylalanine shared between the *B. subtilis* MutL binding site on MutS and Mlh1 binding partners in *S. cerevisiae*. The blue text in the overlay represents heavily conserved residues found within the MIP box, with the darkest blue representing the most conserved residues based on an eukaryotic Exo I alignment (Dherin *et al.*, 2009).

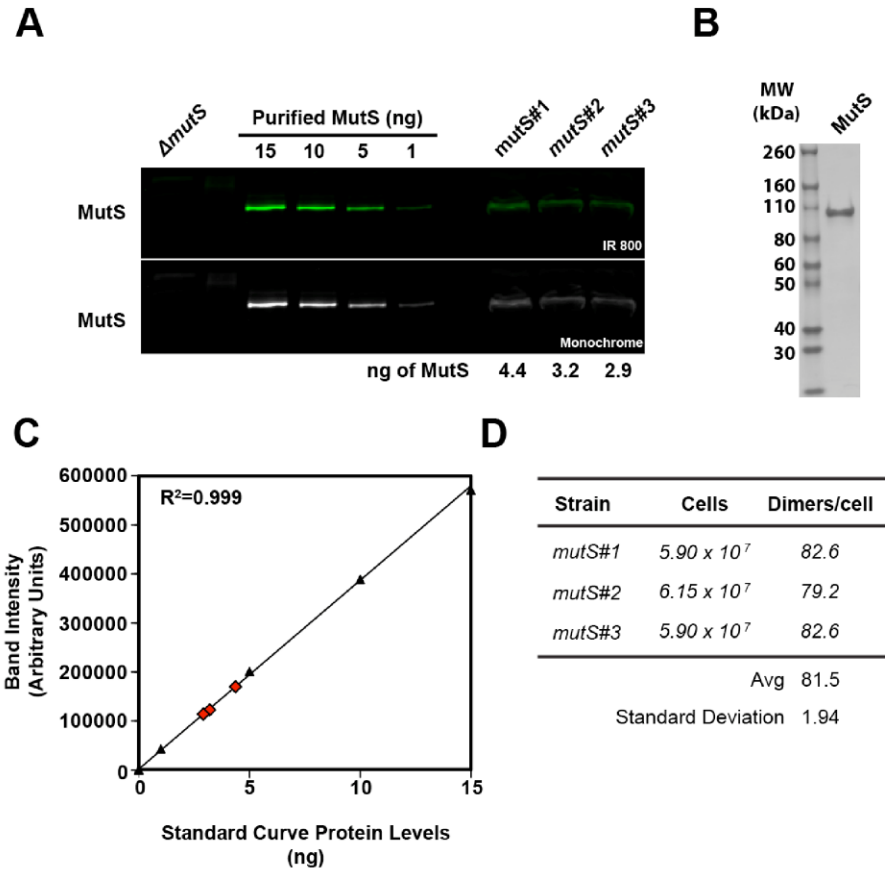


Figure S2.6. Determination of the absolute number of MutS molecules in *B. subtilis*. **(A)** The steady state levels of wild type MutS from whole cell extracts grown in the same conditions as those used for live cell microscopy were compared to a protein standard using purified MutS to determine the number of MutS molecules per cell. Band intensity was determined using LI-COR quantitative Western analysis technology. **(B)** Image of 2 µg of purified MutS used to construct the protein standard curve found in **A** on a 4-20% gradient gel. **(C)** Standard curve of purified MutS protein is pictured in **A**. Red squares indicate the coordinates of total MutS found within the extracts in A. **(D)** A fraction of the culture used to make the whole cell extract was used to determine the number of viable cells by plating on LB agar. The cellular content of MutS within the whole cell extract was determined by adjusting the total ng of MutS identified in **A** and **C** by normalizing the amount of MutS to g mol^{-1} using the molecular weight of MutS (97394 Da), followed by multiplying by Avogadro's number (6.022×10^{23}) to obtain total molecules within the extract. This amount was divided by the total viable count of the culture to obtain the number of MutS monomers $cell^{-1}$. This number was further divided by 2 to obtain the number of dimers per cell. Determination of molarity (M) was based on *B. subtilis* cell lengths and widths under standard imaging conditions (avg. length=2.63 µm, avg. width=0.83 µm, V_{cyl}=~5.32×10^{-15} L, M=113 nM).

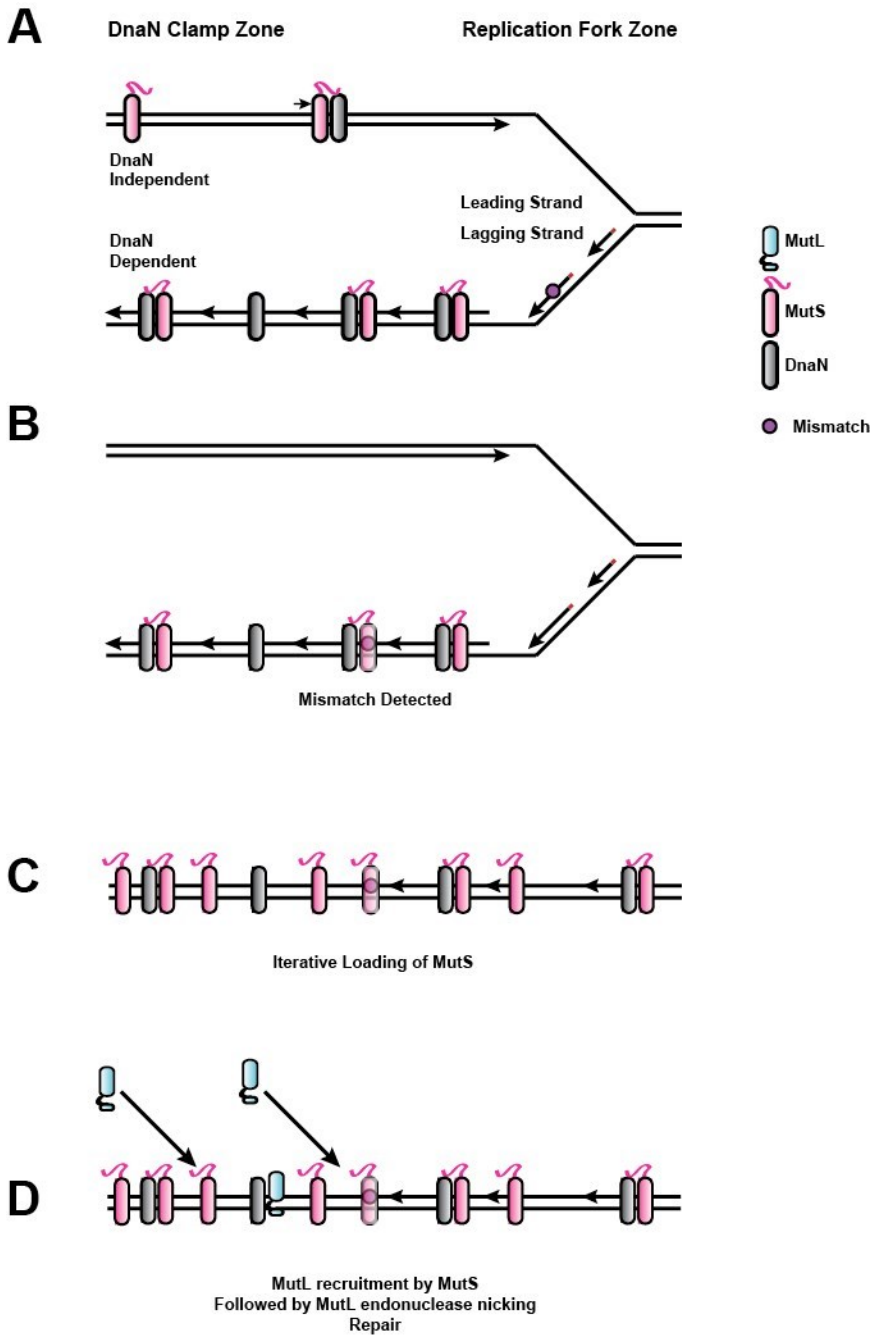


Figure S2.7. A model of the initial steps of *B. subtilis* mismatch repair. (A) DnaN clamp zones direct MutS to newly replicated DNA to enhance mismatch detection. **(B)** Mismatch detection by MutS. **(C)** Iterative loading of MutS occurs at the site of the mismatch, which **(D)** facilitates recruitment of MutL and endonuclease activation.

Table S2.1. Individual amino acid substitutions that comprise each MutS patch variant.

Patch designation	Residue substitutions	Domain
WT	None	
Patch 1	E155S, R156S, L157A, E158S	Connector
Patch 2	E245S, E247S, E248S	Connector
Patch 3a	E306S, E307S, E310S	Core
Patch 3b	F320S, E321S, R322S, E323S	Core
Patch 4	E392S, E395S, E396S	Core
Patch 5	E510S, E512S, E514S	Clamp
Patch 6a	Q806A, L807A, F809A, F810A	DnaN clamp binding
Patch 6b	D811S, E812S, E814S	DnaN clamp binding

Each amino acid substitution tested in the MutS patch mutants are listed and the domain location for each patch mutant is indicated based on the structural determination of bacterial MutS homologs (Lamers *et al.*, 2000; Obmolova *et al.*, 2000).

Table S2.2. *B. subtilis* strains used in this study

Strain	Relevant Genotype	Source or Reference
JSL364	PY79 Prototroph, SP β˚	(Youngman et al., 1984)
LAS393	*mutL::spec*	(Smith et al., 2001)
LAS435	*mutS::mutS-23-mgfpmut2-spec, mutL*$^{-}$	
JSL161	*mutL::mutL-23-mgfpmut2-spec*	
JSL305	*ΔmutSmutL-23-mgfpmut2-spec*	
JSL342	*mutS Patch 3B-1*	
JSL345	*mutS Patch 4-1*	
JSL346	*mutS Patch 5-6*	
JSL355	*mutS Patch 1-14*	

Table S2.2 continued on page 89…

Table S2.2 continued from page 88…

Strain	Relevant Genotype	Source or Reference
JSL372	*mutS Patch 3a*	
JSL377	*mutS Patch 2*	
JSL380	*mutSF320S (2-1)*	
JSL382	*mutS E321S*	
JSL386	*mutS Patch 6a-C2*	
JSL395	*mutS Patch 6b-3*	
JSL400	*mutS-22-mgfpmut3mutL*$^{+}$	
JSL402	*mutSF319SF320S*	
JSL414	*mutSF320S-22-mgfpmut3mutL*$^{+}$	
JSL416	*mutS R322S-1*	
JSL419	*mutSE323S-1*	
JSL424	*mutSF319S*	
JSL425	*mutSF319SF320S-22-mgfpmut3mutL*$^{+}$	
JSL438	*mutSF319SF320SmutL::mutL-23mgfpmut2-spec*	
JSL440	*mutS-22mgfpmut3mutL::mutLE468K-cm*	
JSL450	*mutSF319S-22mgfpmut3mutL*$^{+}$	
JSL453	*mutSE323S-22mgfomut3mutL*$^{+}$	
JSL455	*mutSF319SF320S-22mgfpmut3mutL*$^{+}$*, dnaX::dnaX-23mCherry-spec*	
JSL460	*mutS-22mgfpmut3mutL*$^{+}$*, dnaX::dnaX-23mCherryspec*	
SL467	*mutS Patch 3B-22mgfpmut3mutL*$^{+}$	
JSL469	*mutSV206A, T207S, I208A, I209AmutL*$^{+}$	
JSL471	*mutSS317A*	
JSL473	*mutSH318S*	

All strains listed are derivatives of PY79.

2.10 References

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

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics (Oxford, England) *22*, 195-201.

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

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

Berkmen, M.B., and Grossman, A.D. (2006). Spatial and temporal organization of the Bacillus subtilis replication cycle. Molecular microbiology *62*, 57-71.

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

Blackwell, L.J., Bjornson, K.P., Allen, D.J., and Modrich, P. (2001). Distinct MutS DNA-binding modes that are differentially modulated by ATP binding and hydrolysis. The Journal of biological chemistry *276*, 34339-34347.

Bolz, N.J., Lenhart, J.S., Weindorf, S.C., and Simmons, L.A. (2012). Residues in the N-terminal domain of MutL required for mismatch repair in Bacillus subtilis. Journal of bacteriology *194*, 5361-5367.

Cooper, L.A., Simmons, L.A., and Mobley, H.L. (2012). Involvement of mismatch repair in the reciprocal control of motility and adherence of uropathogenic Escherichia coli. Infection and immunity *80*, 1969-1979.

Cox, E.C., Degnen, G.E., and Scheppe, M.L. (1972). Mutator gene studies in Escherichia coli: the mutS gene. Genetics *72*, 551-567.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proceedings of the National Academy of Sciences of the United States of America *98*, 11627-11632.

Davies, B.W., Bogard, R.W., Dupes, N.M., Gerstenfeld, T.A., Simmons, L.A., and Mekalanos, J.J. (2011). DNA damage and reactive nitrogen species are barriers to Vibrio cholerae colonization of the infant mouse intestine. PLoS pathogens *7*, e1001295.

Dherin, C., Gueneau, E., Francin, M., Nunez, M., Miron, S., Liberti, S.E., Rasmussen, L.J., Zinn-Justin, S., Gilquin, B., Charbonnier, J.B.*, et al.* (2009). Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair. Mol Cell Biol *29*, 907-918.

Dupes, N.M., Walsh, B.W., Klocko, A.D., Lenhart, J.S., Peterson, H.L., Gessert, D.A., Pavlick, C.E., and Simmons, L.A. (2010). Mutations in the Bacillus subtilis beta clamp that separate its roles in DNA replication from mismatch repair. Journal of bacteriology *192*, 3452-3463.

Fishel, R., Ewel, A., Lee, S., Lescoe, M.K., and Griffith, J. (1994). Binding of mismatched microsatellite DNA sequences by the human MSH2 protein. Science (New York, N.Y.) *266*, 1403-1405.

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

Ginetti, F., Perego, M., Albertini, A.M., and Galizzi, A. (1996). Bacillus subtilis mutS mutL operon: identification, nucleotide sequence and mutagenesis. Microbiology *142 (Pt 8)*, 2021-2029.

Gorman, J., Wang, F., Redding, S., Plys, A.J., Fazio, T., Wind, S., Alani, E.E., and Greene, E.C. (2012). Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. Proceedings of the National Academy of Sciences of the United States of America *109*, E3074-3083.

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

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

Hall, B.M., Ma, C.X., Liang, P., and Singh, K.K. (2009). Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics (Oxford, England) *25*, 1564-1565.

Hamilton, S.R., Liu, B., Parsons, R.E., Papadopoulos, N., Jen, J., Powell, S.M., Krush, A.J., Berk, T., Cohen, Z., Tetu, B.*, et al.* (1995). The molecular basis of Turcot's syndrome. The New England journal of medicine *332*, 839-847.

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

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

Iyer, R.R., Pluciennik, A., Genschel, J., Tsai, M.S., Beese, L.S., and Modrich, P. (2010). MutLalpha and proliferating cell nuclear antigen share binding sites on MutSbeta. The Journal of biological chemistry *285*, 11730-11739.

Junop, M.S., Yang, W., Funchain, P., Clendenin, W., and 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 (Amst) *2*, 387-405.

Kijas, A.W., Studamire, B., and Alani, E. (2003). Msh2 separation of function mutations confer defects in the initiation steps of mismatch repair. Journal of molecular biology *331*, 123-138.

Klein, E., Smith, D.L., and Laxminarayan, R. (2007). Hospitalizations and deaths caused by methicillin-resistant Staphylococcus aureus, United States, 1999-2005. Emerging infectious diseases *13*, 1840-1846.

Klevens, R.M., Morrison, M.A., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L.H., Lynfield, R., Dumyati, G., Townes, J.M.*, et al.* (2007). Invasive methicillin-resistant Staphylococcus aureus infections in the United States. JAMA : the journal of the American Medical Association *298*, 1763-1771.

Klocko, A.D., Crafton, K.M., Walsh, B.W., Lenhart, J.S., and Simmons, L.A. (2010). Imaging mismatch repair and cellular responses to DNA damage in Bacillus subtilis. Journal of visualized experiments : JoVE.

Klocko, A.D., Schroeder, J.W., Walsh, B.W., Lenhart, J.S., Evans, M.L., and Simmons, L.A. (2011). Mismatch repair causes the dynamic release of an essential DNA polymerase from the replication fork. Molecular microbiology *82*, 648-663.

Kluytmans, J., van Belkum, A., and Verbrugh, H. (1997). Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. Clinical microbiology reviews *10*, 505-520.

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

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

Landgraf, D., Okumus, B., Chien, P., Baker, T.A., and Paulsson, J. (2012). Segregation of molecules at cell division reveals native protein localization. Nature methods *9*, 480-482.

Larrea, A.A., Lujan, S.A., and Kunkel, T.A. (2010). SnapShot: DNA mismatch repair. Cell *141*, 730.e731.

Lemon, K.P., and Grossman, A.D. (1998). Localization of bacterial DNA polymerase: evidence for a factory model of replication. Science (New York, N.Y.) *282*, 1516-1519.

Lemon, K.P., and Grossman, A.D. (2000). Movement of replicating DNA through a stationary replisome. Molecular cell *6*, 1321-1330.

Lenhart, J.S., Schroeder, J.W., Walsh, B.W., and Simmons, L.A. (2012). DNA repair and genome maintenance in Bacillus subtilis. Microbiology and molecular biology reviews : MMBR *76*, 530-564.

Lenhart, J.S., Sharma, A., Hingorani, M.M., and Simmons, L.A. (2013). DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. Molecular microbiology *87*, 553-568.

Lowy, F.D. (1998). Staphylococcus aureus infections. The New England journal of medicine *339*, 520-532.

Mendillo, M.L., Hargreaves, V.V., Jamison, J.W., Mo, A.O., Li, S., Putnam, C.D., Woods, V.L., Jr., and Kolodner, R.D. (2009). A conserved MutS homolog connector domain interface interacts with MutL homologs. Proceedings of the National Academy of Sciences of the United States of America *106*, 22223-22228.

Nystrom-Lahti, M., Perrera, C., Raschle, M., Panyushkina-Seiler, E., Marra, G., Curci, A., Quaresima, B., Costanzo, F., D'Urso, M., Venuta, S.*, et al.* (2002). Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. Genes Chromosomes Cancer *33*, 160-167.

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

Peltomaki, P. (2005). Lynch syndrome genes. Fam Cancer *4*, 227-232.

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

Plotz, G., Welsch, C., Giron-Monzon, L., Friedhoff, P., Albrecht, M., Piiper, A., Biondi, R.M., Lengauer, T., Zeuzem, S., and Raedle, J. (2006). Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic Acids Res *34*, 6574-6586.

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

Prudhomme, M., Martin, B., Mejean, V., and Claverys, J.P. (1989). Nucleotide sequence of the Streptococcus pneumoniae hexB mismatch repair gene: homology of HexB to MutL of Salmonella typhimurium and to PMS1 of Saccharomyces cerevisiae. Journal of bacteriology *171*, 5332-5338.

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

Schofield, M.J., and Hsieh, P. (2003). DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol *57*, 579-608.

Simmons, L.A., Davies, B.W., Grossman, A.D., and Walker, G.C. (2008). Beta clamp directs localization of mismatch repair in Bacillus subtilis. Molecular cell *29*, 291-301.

Simmons, L.A., Goranov, A.I., Kobayashi, H., Davies, B.W., Yuan, D.S., Grossman, A.D., and Walker, G.C. (2009). Comparison of responses to double-strand breaks between Escherichia coli and Bacillus subtilis reveals different requirements for SOS induction. Journal of bacteriology *191*, 1152-1161.

Simmons, L.A., and Kaguni, J.M. (2003). The DnaAcos allele of Escherichia coli: hyperactive initiation is caused by substitution of A184V and Y271H, resulting in defective ATP binding and aberrant DNA replication control. Molecular microbiology *47*, 755-765.

Smith, B.T., Grossman, A.D., and Walker, G.C. (2001). Visualization of mismatch repair in bacterial cells. Molecular cell *8*, 1197-1206.

Umar, A., Boyer, J.C., Thomas, D.C., Nguyen, D.C., Risinger, J.I., Boyd, J., Ionov, Y., Perucho, M., and Kunkel, T.A. (1994). Defective mismatch repair in extracts of colorectal

and endometrial cancer cell lines exhibiting microsatellite instability. The Journal of biological chemistry *269*, 14367-14370.

Winkler, I., Marx, A.D., Lariviere, D., Heinze, R.J., Cristovao, M., Reumer, A., Curth, U., Sixma, T.K., and Friedhoff, P. (2011). Chemical Trapping of the Dynamic MutS-MutL Complex Formed in DNA Mismatch Repair in Escherichia coli. The Journal of biological chemistry *286*, 17326-17337.

Wojtaszek, J., Lee, C.J., D'Souza, S., Minesinger, B., Kim, H., D'Andrea, A.D., Walker, G.C., and Zhou, P. (2012a). Structural basis of Rev1-mediated assembly of a quaternary vertebrate translesion polymerase complex consisting of Rev1, heterodimeric polymerase (Pol) zeta, and Pol kappa. The Journal of biological chemistry *287*, 33836-33846.

Wojtaszek, J., Liu, J., D'Souza, S., Wang, S., Xue, Y., Walker, G.C., and Zhou, P. (2012b). Multifaceted recognition of vertebrate Rev1 by translesion polymerases zeta and kappa. The Journal of biological chemistry *287*, 26400-26408.

Chapter 3

Structure of the Endonuclease Domain of MutL: Unlicensed to Cut

Reprinted from Molecular Cell, Vol. 39, M.C. Pillon, J.J. Lorenowicz, M. Uckelmann, A.D. Klocko, R.R. Mitchell, Y.S. Chung, P. Modrich, G.C. Walker, L.A. Simmons, P. Friedhoff, and A. Guarné, Structure of the Endonuclease Domain of MutL: Unlicensed to Cut, 145-151, Copyright 2010, with permission from Elsevier.

3.1 Author's Preface

In 2006 Dr. Modrich's group made the discovery that human MutLα is an endonuclease (Kadyrov et al., 2006). The following year they showed that the MutLα homolog from *Saccharomyces cerevisiae* also supports nicking activity (Kadyrov et al., 2007). This chapter presents the first crystal structures of the MutL endonuclease domain. The *B. subtilis* MutL endonuclease motif (^{462}DQHAAQERIKYE) along with two other conserved motifs (^{572}SCK and ^{604}CPHGRP) coordinate two zinc metals. The residues responsible for coordinating zinc are essential for MMR, however, endonuclease assays exclude zinc as the catalytic metal. Alternatively, the zinc ions likely play a regulatory role in activating the MutL endonuclease. J.J. Lorenowicz solved the crystal structure of the MutL endonuclease domain in the absence of metal. I solved the crystal structures of the MutL endonuclease domain in an alternative conformation, as well as, bound to zinc. R.R. Mitchel and I purified proteins for functional assays. J.J. Lorenowicz and A. Guarné conducted endonuclease assays and I characterized the DNA binding activity of MutL. M. Uckelmann (from the laboratory of Dr. P. Friedhoff) performed zinc release assays (Figure 3.3 panel B) and A. Klocko (from the laboratory of Dr. L.A. Simmons) performed *in vivo* DNA mismatch repair assays (Figure 3.3 panel C). Y.S. Chung prepared the movie (Movie S3.1) describing the multiple MutL conformations. A. Guarné and I designed and interpreted the experiments and A. Guarné wrote the manuscript.

3.2 Abstract

DNA mismatch repair corrects errors that have escaped polymerase proofreading, increasing replication fidelity 100- to 1000-fold in organisms ranging from bacteria to humans. The MutL protein plays a central role in mismatch repair by coordinating multiple protein-protein interactions that signal strand removal upon mismatch recognition by MutS. Here we report the crystal structure of the endonuclease domain of *Bacillus subtilis* MutL. The structure is organized in dimerization and regulatory subdomains connected by a helical lever spanning the conserved endonuclease motif. Additional conserved motifs cluster around the lever and define a Zn^{2+}-binding site that is critical for MutL function *in vivo*. The structure unveils a powerful inhibitory mechanism to prevent undesired nicking of newly replicated DNA and allows us to propose a model describing how the interaction with MutS and the processivity clamp could license the endonuclease activity of MutL. The structure also provides a molecular framework to propose and test additional roles of MutL in mismatch repair.

3.3 Introduction

DNA mismatch repair (MMR) maintains genomic stability by correcting errors that have escaped polymerase proofreading (Kunkel and Erie, 2005). MMR proteins are also implicated in a variety of other cellular processes such as DNA damage signalling, apoptosis, meiotic and mitotic recombination, and somatic hypermutation (Modrich, 2006). Mutations in mismatch repair genes are associated with an increased mutation rate

and microsatellite instability, the hallmark of human non-polyposis colorectal cancer (Peltomaki, 2005).

Initiation of MMR depends on the coordinated action of three proteins. MutS recognizes a mismatched base pair or a small insertion/deletion loop and recruits MutL in an ATP-dependent manner. Subsequently, the newly synthesized strand is marked for repair. In *Escherichia coli*, strand discrimination is achieved by mismatch-provoked activation of the MutH endonuclease, which cleaves the unmethylated DNA strand at hemimethylated GATC sites transiently generated during DNA replication. Although most bacteria and all eukaryotes do not encode a MutH homolog, a pre-existing nick is sufficient to activate mismatch repair in a system reconstituted from purified proteins (Zhang et al., 2005). It has been shown that MutL homologs from species lacking a MutH endonuclease harbor an intrinsic latent nicking endonuclease activity that is vital for its function in mismatch repair (Erdeniz et al., 2007; Kadyrov et al., 2006; Kadyrov et al., 2007; Kosinski et al., 2008).

MutL is composed of two structurally conserved domains connected by a variable flexible linker (Guarné et al., 2004). The N-terminal region encompasses an ATPase domain of the GHL ATPase superfamily that is conserved from bacteria to humans (Ban et al., 1999; Guarné et al., 2001). Conversely, the sequence conservation in the C-terminal dimerization region of MutL is low. The structure of the C-terminal domain of

E. coli MutL reveals that this region is organized into two distinct subdomains (Guarné et al., 2004; Kosinski et al., 2005).

While prokaryotic MutL homologs form homodimers, their eukaryotic counterparts form heterodimers. In humans, there are four paralogs of MutL (hMLH1, hPMS2, hPMS1 and hMLH3) that form three heterodimers by association of hMLH1 with hPMS2 (hMutLα), hPMS1 (hMutLβ) and hMLH3 (hMutLγ) (Li and Modrich, 1995; Raschle et al., 1999; Lipkin et al., 2000). hMutLα is necessary for mismatch repair function and hMutLγ has a role in meiotic recombination, however the function of hMutLβ is unknown (Kunkel and Erie, 2005). The C-terminal regions of hPMS2 and hMLH3 encompass a conserved $DQHA(X)_2E(X)_4E$ motif that is required for endonuclease activity. Based on sequence analysis and molecular modelling, three additional conserved motifs (ACR, C(P/N)HGRP and FXR) have been predicted to form a single active site with the endonuclease motif (Kosinski et al., 2008). Analysis of the reconstituted human MMR system indicates that the endonuclease activity of MutLα provides a loading site for MutSα-activated exonuclease I (Kadyrov et al., 2006).

Here we present the structure of the C-terminal dimerization domain of *Bacillus subtilis* MutL (BsMutL) harboring the endonuclease activity of the protein. The structure reveals the conserved three-dimensional organization of the endonuclease site of MutL and exposes the presence of a regulatory Zn^{2+}-binding site that is important for the mismatch repair function of BsMutL *in vivo*. The structure allows for us to propose a

model describing how the association of MutS and the DNA polymerase III processivity clamp (β-clamp), with MutL could license nicking of a newly synthesized DNA strand.

3.4 Materials and Methods

3.4.1 Cloning, purification and crystallization

Full-length BsMutL was amplified from genomic DNA and cloned into the pProEXHTa expression vector (Life Technologies). His-tagged BsMutL was purified using a Ni^{2+}-chelating affinity column equilibrated with 20 mM TRIS pH 8, 0.5 M NaCl, 1.4 mM β-mercaptoethanol, 5% glycerol and 100 mM PMSF. BsMutL was eluted using 300 mM imidazole and subsequently injected on a hydrophobic column equilibrated with 20 mM TRIS pH 8, 1 M KCl, 1 mM DTT and 5% glycerol. BsMutL was further purified by ionic exchange and size exclusion chromatography (MonoQ 5/50 and Superdex-S200, GE Healthcare) equilibrated with 20 mM TRIS pH 8, 100 mM KCl, 1 mM DTT and 5% glycerol (storage buffer). Guided by a structure-based sequence alignment, we subcloned the C-terminal fragment of BsMutL (BsMutL-CTD, residues 433-627). BsMutL-CTD was purified similarly to BsMutL with an additional ionic exchange purification step after His-tag cleavage with TEV-protease. Mutants of BsMutL and BsMutL-CTD were generated by QuikChange (Stratagene) and verified by DNA sequencing (MOBIX, McMaster University).

Crystal form I was grown in 25% PEG-monomethyl ether 550, 0.1 M $MgCl_2$, 0.1 M TRIS pH 9 and 5% PEG 400. Two additional crystal forms were obtained when the

protein was supplemented with 50 nM $ZnCl_2$ and 50 nM $CoCl_2$. Crystal form II was grown in 25 % PEG 3,350, 0.15-0.2 M NaCl and 0.1 M TRIS pH 7. Addition of 0.5 mM $ZnCl_2$ to this crystallization solution yielded crystal form III. 10 % PEG 400 was added to all crystallization conditions prior to flash freezing in liquid nitrogen.

3.4.2 Data collection and structure determination

For crystal form I, a three-wavelength MAD data set was collected at X29B in NSLS, Brookhaven National Laboratory (Upton, NY). Data were indexed, processed and merged using HKL2000 (Otwinowski and Minor, 1997). Twenty-seven out of thirty-six Selenium sites were found and refined using SOLVE (Terwilliger and Berendzen, 1999). A native data set to 2.5 Å was used for subsequent manual building and refinement, which was done using standard protocols in phenix.refine and COOT (Afonine et al., 2005; Emsley and Cowtan, 2004). Complete data sets of crystal forms II and III, were collected at the X25 beamline in NSLS. Data were collected at a wavelength corresponding to the Zn^{2+}-absorption edge (Table 2.1), as measured using fluorescence scans. All final models have over 92% of the residues within the most favored regions in the Ramachandran plot and none in disallowed regions. Figures depicting molecular structures were generated using PyMol (DeLano, 2002).

3.4.3 Endonuclease and DNA-binding assays

BsMutL nicking activity was assayed as previously described (Kadyrov et al., 2006), with minor modifications described in the Supplemental Information. To assess DNA binding by BsMutL, supercoiled DNA (5 nM) was incubated with BsMutL variants (100 nM, dimer) in endonuclease buffer for 90 minutes at 37 ˚C. Reaction mixtures (20 μL) were resolved on 1% TAE agarose gels and quantified using ImageJ (http://rsbweb.nih.gov/ij/).

3.4.4 Zinc-affinity fluorescence assay

Wild type and variants of BsMutL-CTD (2 μM) were incubated with 1 μM FluoZin-3 (Invitrogen) in 10 mM HEPES pH 6 and 200 mM KCl buffer treated with 1% Chelex-100 (Fluka). A calibration fluorescence curve was generated using buffer including increasing concentrations of $ZnCl_2$ (0.25-3 μM) in the absence of protein. Spectra (500-600 nm) were recorded at an excitation wavelength of 494 nm (corrected for buffer effects). Inner-filter effects were neglected because a linear fluorescence intensity response up to stoichiometric amounts of Zn^{2+} was seen when using 4 μM FluoZin-3.

3.4.5 Mismatch repair assays of BsMutL variants

Mismatch repair assays were performed largely as described (Simmons et al., 2008). See the Supplemental Information for a more complete description of the methods.

3.5 Results and Discussion

3.5.1 Crystal Structure of BsMutL-CTD

Three crystal forms of the C-terminal domain of BsMutL (BsMutL-CTD) were obtained. Crystal form I was used to determine the structure of BsMutL-CTD by multiwavelength anomalous diffraction using crystals grown with Sel-Met substituted protein (Table 2.1). This crystal includes four independent monomers (molecules A to D) in the asymmetric unit that associate through crystal symmetry to form the functional BsMutL-CTD dimer. Crystal forms II and III contained a single dimer in the asymmetric unit. In the three structures, the N- and C-terminal ends of BsMutL-CTD (residues 433-461/580-627) define the dimerization subdomain, while residues 474-573 define an external subdomain that protrudes to the solvent, herein referred to as the regulatory subdomain (Figure 3.1A-B). The subdomains are connected by helix αA (residues 463-473), encompassing the conserved endonuclease motif, and the linker connecting helices αD-αE (residues 575-581), which is disordered in our structures (Figure 3.1A). However, the relative orientation between subdomains varies from one crystal form to another (Movie S3.1).

The three complimentary conserved motifs associate with the endonuclease motif cluster around helix αA to delineate a single catalytic site with the conserved endonuclease motif (^{462}DQHA(X)$_2$E(X)$_4$E) (Figure 3.1B-C). ^{604}CPHGRP resides in the αE-β8 loop, ^{572}SCK (consensus sequence ACR) is the last turn of the αD helix and ^{623}FKR, at the C-terminus of the protein, reaches the active site of the other protomer

(Figure 3.1B). Except for the ^{572}SCK motif, contributed by the regulatory subdomain, all motifs reside in the dimerization subdomain.

Table 3.1. Data Collection and Refinement

Data Collection	Crystal Form I			Crystal Form II	Crystal Form III
Space Group	$C222_1$			$P2_12_12_1$	$P2_12_12_1$
Unit Cell (Å)	a= 88.3, b= 94.9, c= 218.7			a=33.2, b=74.6, c=182.2	a=33.2, b=74.6, c=182.2
Wavelength (Å)	0.9794 (Se-edge)	0.9792 (Se-peak)	0.9686 (Se-remote)	1.0809 (native)	1.282 (Zn-peak)	1.2796 (Zn-peak)
Resolution (Å)[a]	35-2.8 (2.9-2.8)	35-2.8 (2.9-2.8)	35-2.8 (2.9-2.8)	50-2.38 (2.47-2.38)	50-2.0 (2.03-2.00)	50-2.15 (2.19-2.15)
Completeness (%)[a]	98.2 (88.3)	98.6 (91.4)	92.5 (55.9)	95.9 (76.0)	100 (100)	99.3 (93.3)
Redundancy[a]	6.2 (5.0)	6.3 (5.4)	6.3 (3.9)	5.5 (4.5)	7.5 (7.4)	10.3 (6.3)
Rmerge (%)[a]	6.5 (21.0)	8.1 (21.3)	7.2 (23.1)	4.0 (22.5)	8.3 (67.6)	8.9 (70.7)
I/σ(I)[a]	20.2 (6.9)	18.5 (5.6)	17.8 (4.9)	41.5 (4.5)	25.9 (3.1)	22.4 (2.3)
ASU content	4 monomers			4 monomers	1 dimer	1 dimer
Data Refinement						
Resolution (Å)				30-2.5	38.9-2.0	32.8-2.3
Reflections (work)				31,772	60,020	41,004
Reflections (test)				1,631	3,049	2,113
Atoms refined				6,122	3,294	3,120
Solvent atoms				105	190	74
Zn atoms				--	--	4
R (R_{free}) (%)				21.7 (26.8)	19.1 (22.8)	21.4 (26.9)
Rmsd in bonds (Å)				0.003	0.005	0.004
Rmsd in angles (°)				0.609	0.875	0.745
Mean B values (Å)				44.1	45.6	43.9

a. Data in the highest resolution shell is shown in parentheses.

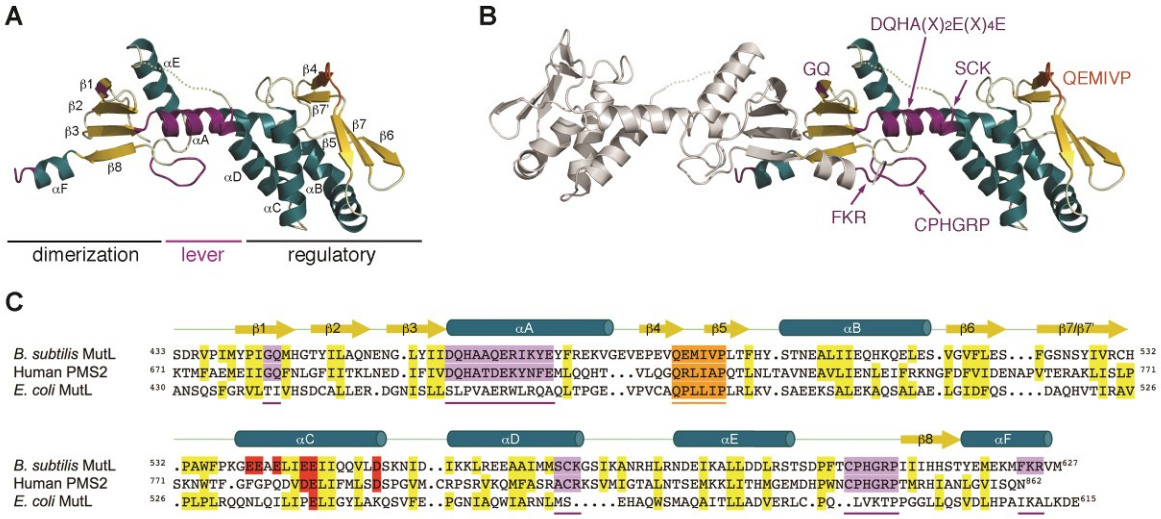


Figure 3.1. Crystal structure of BsMutL-CTD. (A) Ribbon diagram of the BsMutL-CTD monomer. Secondary structure motifs are labeled and colored blue (helices) and yellow (strands) with the connecting loops in light green. The endonuclease and the endonuclease-associated motifs are shown in purple, while the additional conserved motifs are shown in orange. **(B)** Ribbon diagram of the BsMutL-CTD dimer with one protomer shown as in (a) and the other one as grey ribbons. **(C)** Sequence alignment of the C-terminal regions of BsMutL, hPMS2 and EcMutL. Secondary structure elements of BsMutL-CTD are shown as arrows (strands) and cylinders (helices). The five conserved motifs are highlighted in purple and underlined. Conserved hydrophobic residues are highlighted in yellow. The conserved ^{487}QEMIVP motif is highlighted in orange. See also Figure S3.1.

Two additional conserved motifs have been identified within the C-terminal region of MutLα (Kosinski et al., 2008). The ^{443}GQ motif resides on the β1 strand of the dimerization subdomain. This strand is in the vicinity of the αD-αE loop and hence it may indirectly contribute to the overall stability of the active site. The ^{487}QEMIVP motif (consensus sequence QXLLXP) is on the surface of the regulatory subdomain and conspicuously exposed to the solvent (Figure 3.1B). Conservation of the QXLLXP motif

is not correlated with the endonuclease activity of MutL, suggesting that it could mediate the interaction with other repair factors.

Even though the C-terminal regions of EcMutL and BsMutL have very low sequence similarity, their structures have nearly identical topologies (Figure S3.1). However, key differences exist. Superimposition of the EcMutL-CTD monomer onto the BsMutL-CTD monomer returned root mean squared deviations of only 1 Å for the dimerization subdomains but > 2 Å for the regulatory subdomains, reflecting the increased divergence of this region. This is intriguing because the regulatory subdomain contributes minimally to the endonuclease site. The most striking difference between the two structures is the organization of the secondary structure elements surrounding helix αA, which would preclude the formation of a functional endonuclease site even if EcMutL had the conserved $DQHA(X)_2E(X)_4E$ motif (Figure S3.1). Notably, the extended αE-β8 loop in BsMutL, rather than the additional helix seen in EcMutL, brings the ^{604}CPHGRP motif closer to helix αA and secludes the endonuclease site. The dimerization interfaces are also remarkably different. While the BsMutL dimer buries 1,065 $Å^2$, the EcMutL dimer only conceals 910 $Å^2$. Interestingly, the reorientation of the apposing β-sheets in the BsMutL-CTD dimer allows the ^{623}FKR motif to reach the adjacent endonuclease site.

3.5.2 BsMutL has weak endonuclease activity

Similarly to other MutL orthologues, BsMutL had a weak endonuclease activity dependent on Mn^{2+} (Figure 3.2A). Both a point mutation in the endonuclease motif (D462N) or deletion of the ATPase domain virtually abolished the endonuclease activity of BsMutL (Figure 3.2A). This finding is interesting in light of the recent experiments revealing that the C-terminal domains of *Neisseria ghonorrhoeae* and *Aquifex aeolicus* have endonuclease activity (Duppatla et al., 2009; Mauris and Evans, 2009). However, the specific activity of AaeMutL-CTD is much lower than that of the full-length protein. We suspected that the lack of nicking activity by BsMutL-CTD could be due to a DNA-binding defect, since EcMutL-CTD does not bind DNA stably (Guarné et al., 2004). Indeed, BsMutL-CTD did not bind supercoiled DNA while other variants of BsMutL did (Figure 3.2C).

Addition of 0.5 mM ATP stimulated the nicking activity of BsMutL, but higher concentrations of ATP (5 mM) inhibited the nicking activity, presumably due to excess nucleotide chelating Mn^{2+} ions away (Figure 3.2B, lanes 5 and 6). Unexpectedly, addition of ATP and/or Mg^{2+} stimulated a second cut on the nicked DNA to yield a linear product. The cut of the two strands at nearby points could be due to the presence of two endonuclease sites in the BsMutL homodimer or a consequence of the high-ion concentrations used in the experiment. We favor the former because incubation with 10 mM Mn^{2+} did not cause nicking of the two strands (data not shown), but addition of only 1 mM of a second metal ion such as Zn^{2+} or Co^{2+} yielded a linear product (Figure 3.2D).

Interestingly, Mg^{2+} did not support double nicking under these conditions, suggesting that BsMutL may have higher affinity for Zn^{2+} or Co^{2+} than Mg^{2+}.

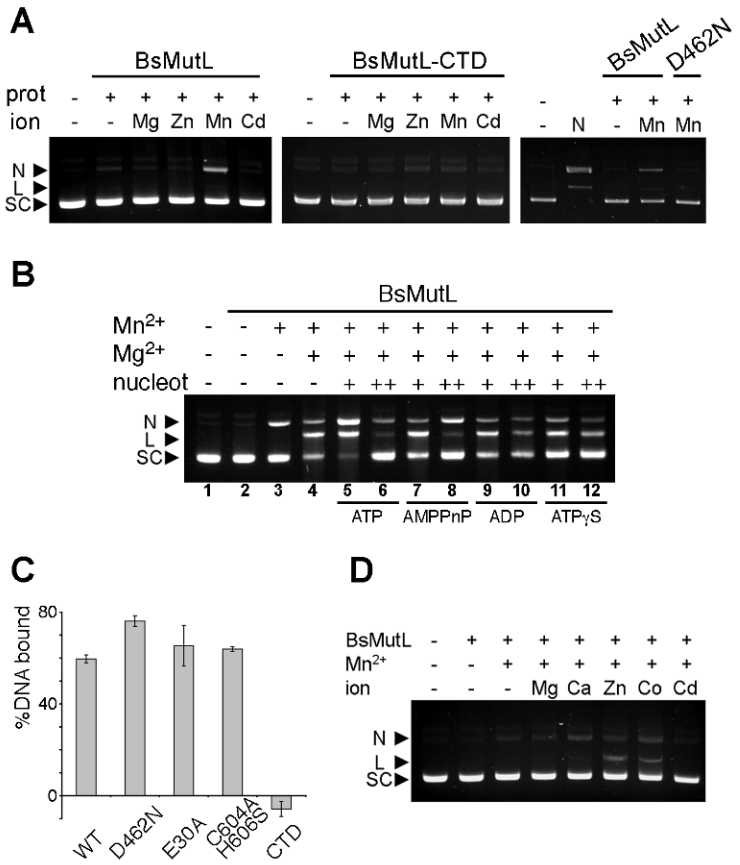


Figure 3.2. Endonuclease activity of BsMutL. (A) Nicking activity of BsMutL (left) and BsMutL-CTD (center) in the presence of Mg^{2+}, Zn^{2+}, Mn^{2+} or Cd^{2+} as indicated. Comparison of the nicking activity of BsMutL and BsMutL-D462N in the presence of Mn^{2+} (right). Migration of supercoiled (SC), nicked (N) and linear (L) DNA is indicated. **(B)** Endonuclease activity of BsMutL in the presence of 0.5 mM (+) and 5 mM (++) nucleotide. **(C)** DNA binding by BsMutL (WT), BsMutL-CTD (CTD) and BsMutL variants as indicated. Data are presented as the mean of three independent measurements and the error bars correspond to the standard errors of the mean (SEM=$\sigma/\sqrt{n}$, where σ is the average and n the sample size). **(D)** Stimulation of the endonuclease activity of BsMutL (1 mM Mn^{2+}) by a second divalent metal ion (1 mM).

We then characterized the ATPase activity of BsMutL (K_m= 0.4 mM and k_{cat}= 0.3 min^{-1}) and found that it is a weaker ATPase than other MutL homologues (Guarné et al., 2001; Hall et al., 2002; Ban et al., 1999). Given the slow ATP-hydrolysis rate, the stimulation of the endonuclease activity of BsMutL was likely due to ATP-binding rather than ATP-hydrolysis. In good agreement with this idea, ADP did not stimulate the endonuclease activity of BsMutL (Figure 3.2B, lanes 9 and 10). However, two known non-hydrolyzable analogues of ATP, AMPPnP and ATPγS, did not stimulate the endonuclease activity of BsMutL beyond the levels observed when both Mn^{2+} and Mg^{2+} were present either (Figure 3.2D, compare lanes 4, 7, 9 and 11). Conceivably, the ATP-dependent stimulation of the endonuclease activity of BsMutL could be due to a conformational change induced by ATP binding as seen in other MutL orthologues (Ban et al., 1999; Sacho et al., 2008), that would bring DNA bound at the ATPase domain in close proximity to the endonuclease site. If this is the case, our results suggest that only ATP can induce efficiently such conformational change.

3.5.3 BsMutL has a regulatory Zn^{2+}-binding site

Although the endonuclease activity of BsMutL was metal-dependent and the conserved motifs around helix αA define a putative Zn^{2+}-binding site (Kosinski et al., 2008), no metal ions were found in the BsMutL-CTD structure (crystal form I). However, different crystal forms were obtained when the protein storage buffer (crystal form II) and the crystallization solution (crystal form III) were supplemented with $ZnCl_2$. While the ^{462}DQHA(X)$_2$E(X)$_4$E, ^{572}SCK and ^{604}C(P/N)HGRP motifs were much closer in crystal

form II than crystal form I, no metal was found in this crystal form either. Conversely, the BsMutL-CTD dimer found in the asymmetric unit of crystal form III contained two Zn^{2+} ions bound to each protomer; a fully-occupied Zn^{2+} ion (Zn^{2+}A) was coordinated by the side chains of residues Glu468, Cys604, His606 and a well-ordered water molecule, and a partly occupied site (Zn^{2+}B) coordinated by the side chains of residues His464, Glu468, Cys573 and a water molecule (Figure 3.3A). The nature of the metal ion was confirmed on the anomalous difference electron density maps from diffraction data collected at the Zn^{2+} absorption edge (Table 2.1).

It had been previously reported that the putative Zn^{2+}-binding site in hPMS2 could be related to the regulatory metal-binding site found in the iron-dependent repressors from the DtxR/MntR family (Kosinski et al., 2008). A structural comparison revealed that, while motifs ^{462}DQHAX$_2$EX$_4$E, ^{572}SCK and ^{604}CPHGRP from BsMutL could be superimposed to the regulatory metal-binding sites of IdeR or MntR, the residues coordinating the metal ion differed. However, another Mn^{2+}-dependent repressor from the same family (ScaR, PDB: 3HRU) encompasses a regulatory metal-binding site identical to the fully occupied Zn^{2+} site in BsMutL-CTD.

To probe whether the two metal-bound sites found in our structure were true Zn^{2+}-binding sites, we measured the affinity of purified BsMutL-CTD and various BsMutL-CTD variants for zinc. To this end, we measured the fluorescence of increasing concentrations of Zn^{2+} bound to the fluorescence indicator FluoZin-3. In the absence of

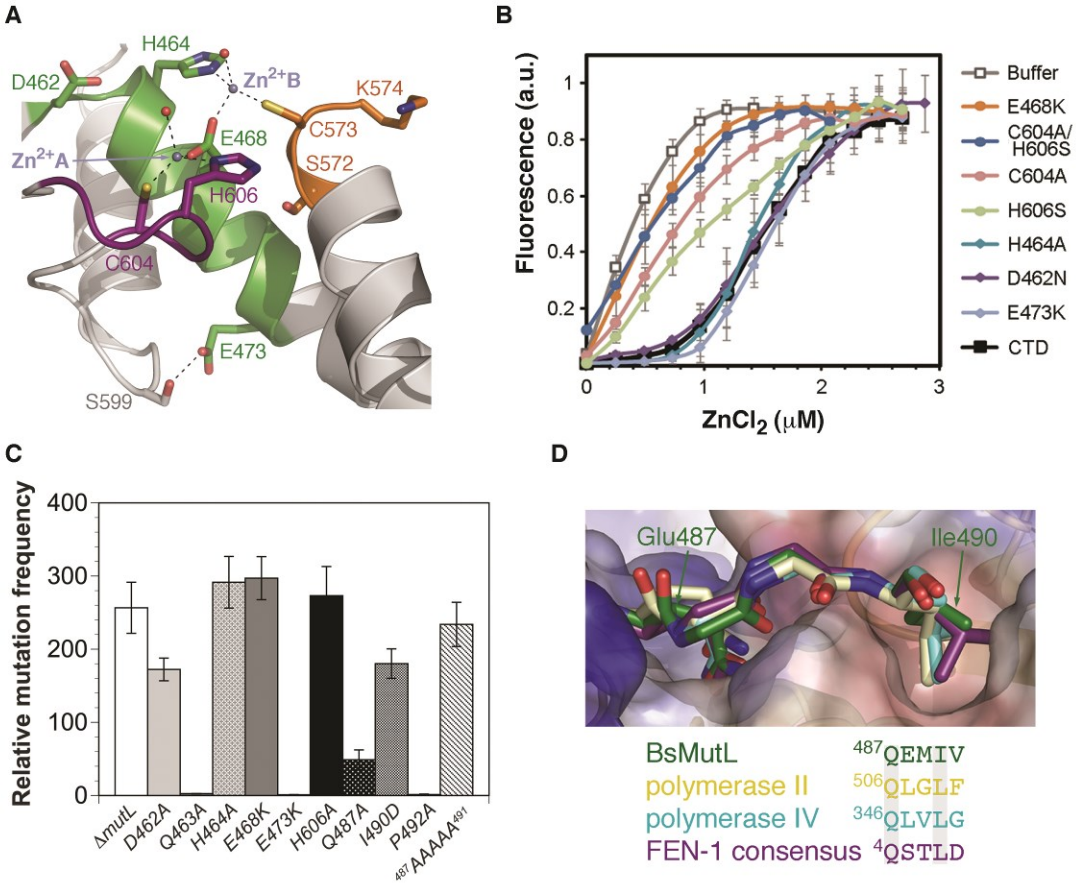


Figure 3.3. Regulatory Zn^{2+}-binding site in BsMutL-CTD. (A) Organization of the endonuclease site of BsMutL-CTD bound to Zn^{2+} (crystal form III). Hydrogen bonds are shown as black dashed lines with the water molecules and Zn^{2+} ions shown as red and lilac spheres, respectively. Conserved motifs are color-coded green (^{462}DQHAX$_2$EX$_4$E), purple (^{604}CPHGRP) and orange (^{572}SCK). **(B)** Zinc-affinity profiles of BsMutL-CTD and point mutants of BsMutL-CTD as indicated. **(C)** Bar diagram showing the relative mutation frequency of the indicated *mutL* variants altered in the Zn^{2+}-binding, the endonuclease site and the putative β-binding motif. Data are presented as the mean of 4 independent cultures +/- SEM. **(D)** Superimposition of the β-binding motif in BsMutL-CTD (green) onto those of pol II (yellow, PDB 3D1E), pol IV (cyan, PDB 1UNN) and FEN-1 (purple, PDB 1RXM) shown as a main-chain trace with the β-clamp structure (PDB 3D1E) presented as a semi-transparent electrostatic potential surface. See also Figures S3.2-S3.3 and Movie S3.1.

protein, fluorescence increased exponentially reaching maximum values at around 1.5-2 μM $ZnCl_2$. The sigmoidal response observed for wild type and the D462N, H464S and E473K variants was characteristic of zinc binding by the protein (Figure 3.3B), suggesting that these BsMutL-CTD variants still retained the ability to bind zinc. Addition of the sulfhydryl-modifying agent methyl methanethiosulfonate (MMTS) resulted in an increase of fluorescence to the level detected in the absence of protein, indicating that one or more cysteine residues within the C-terminal domain of BsMutL were important for Zn^{2+}-binding (data not shown). Conversely, fluorescence profiles of E468K, C604A, H606S and C604A/H606S mutants did not have a sigmoidal response (Figure 3.3B), revealing that these BsMutL-CTD variants had lost the ability to bind zinc. These results confirmed that Glu468, Cys604 and His606 define the Zn^{2+}-binding site in BsMutL (Figure 3.3A).

3.5.4 Integrity of the conserved motifs is important for mismatch repair *in vivo*

We presumed that the integrity of the Zn^{2+}-binding site in MutL would be important for proper mismatch repair. Therefore, we measured the MMR efficiency of BsMutL variants encompassing point mutations in the conserved residues involved in the endonuclease or Zn^{2+}-binding sites. Mutation of D462A, H464A, E468K or H606A completely inactivated MMR *in vivo* (Figure 3.3C), underscoring the importance of these residues. Similarly, the equivalent mutations in hPMS2 also conferred a strong mutator phenotype (Kosinski et al., 2008). Conversely, the BsMutL-Q463A and BsMutL-E473K

variants had similar mismatch repair efficiency to wild-type BsMutL, suggesting that not all the conserved residues within these motifs play essential roles in mismatch repair.

We next analyzed the importance of other conserved motifs found in the C-terminal region of MutL, namely the ^{487}QEMIVP motif. Mutation of Ile490 almost completely inactivated MMR *in vivo*, whereas mutation of Pro492 was without effect (Figure 3.3C). A BsMutL-Q487A variant conferred approximately a 50-fold mutator phenotype about 5-fold lower than a *mutL* null strain. Replacement of residues ^{487}QEMIV with five alanine residues also abrogated MMR *in vivo* (Figure 3.3C). This conserved loop is conspicuously exposed and loosely resembles the consensus β-binding motif (Dalrymple et al., 2001). Most notably, its conformation is nearly identical to that seen in the structures of other peptides bound to the β-clamp (Figure 3.3D). Superimposition of the ^{486}VQEMIVPL sequence from BsMutL onto the structures of β-clamp bound to peptides from polymerase II, FEN-1 and polymerase IV returned r.m.s. deviations smaller than 0.5 Å. Accordingly, the regulatory subdomains of both BsMutL and EcMutL could be directly docked onto the structure of the *E. coli* β-clamp. Interestingly, docking of the MutL-CTD dimer suggests that the interaction of one protomer would prevent the interaction of the other due to steric hindrance (Figure S3.2), suggesting a possible regulatory role for this interaction.

3.5.5 A model for the activation of the endonuclease activity

Based on the structures of BsMutL-CTD, we suggest that the conserved motifs in MutL define an endonuclease active site with two distinct subsites. A structural Zn^{2+}-binding site defined by the side chains of Glu468, Cys604 and His606; and a catalytic site likely defined by Asp462 and His464. Metal binding at the structural site locks the orientation between the dimerization and regulatory subdomains, which is highly variable in the absence of metal (Movie S3.1). Based on the three-dimensional organization of Asp462 and His464, the catalytic subsite could coordinate one or two metal ions to nick DNA (Yang, 2008). Supporting this idea, double-stranded DNA could be modeled onto the structure of BsMutL-CTD with the scissile bond at a distance compatible with catalysis activated by Asp462 and the adjacent 3' phosphate providing the fourth coordination ligand of the regulatory Zn^{2+} ion (Figure S3.3).

In the context of the replication fork, the endonuclease activity of MutL ought to be repressed until a mismatch is encountered. A look at the electrostatic potential surface of BsMutL-CTD reveals a powerful suppression mechanism of DNA-binding. The regulatory subdomain is covered with negative charges that guard the endonuclease site (Figure 3.4A). This could be a widely spread repression mechanism since some of the negatively charged residues in helix αC are conserved in other MutL homologues harboring the endonuclease motif (Figure 3.1C and (Kosinski et al., 2008)). Licensing the endonuclease activity of MutL would thus require a significant conformational change or the interaction with other repair factors to overcome the DNA repulsion in the vicinity of

the endonuclease site. We presume that a conformational change like that induced in MutLα upon nucleotide binding could allow DNA bound at the ATPase domain to reach the endonuclease site (Sacho et al., 2008). However, additional repair factors are likely required to mask helix αC.

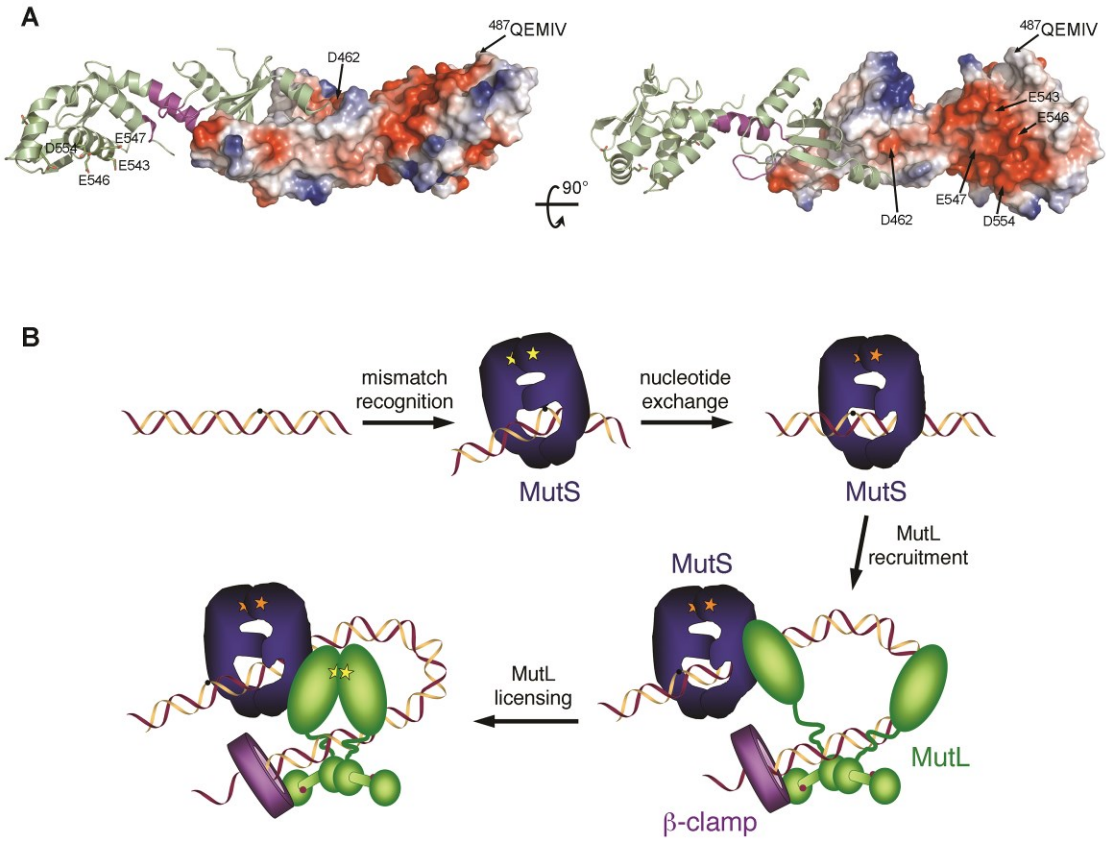


Figure 3.4. Model of activation of the endonuclease activity of MutL. (A) Orthogonal views of the electrostatic surface potential of the BsMutL-CTD protomer. The second protomer is shown as a ribbon diagram with the endonuclease motifs in purple. **(B)** Upon mismatch binding, MutS (blue) undergoes a nucleotide-dependent conformational change that triggers recruitment of MutL (green) to the mismatch site, likely aided by the β-clamp (purple). ATP binding by MutL then promotes the association of its two ATPase subunits and brings the ATPase in close proximity to the dimerization domain of the protein. Coordinated interaction of MutS and β-clamp bound to DNA (ribbon diagram) with ATP-bound MutL could thus license the latent endonuclease activity of MutL. ATP and ADP are shown as yellow and orange stars, respectively.

The endonuclease activity of MutLα is greatly stimulated by the presence of PCNA and RFC (Kadyrov et al., 2006; Kadyrov et al., 2007), the eukaryotic homologues of the β-clamp and the clamp loader. Additionally, MutSα and PCNA form a stable complex (Iyer et al., 2008). Human MutLα interacts with MutSα through its ATPase domain (Plotz et al., 2006), but the region of MutSα that interacts with MutLα is not known. Conceivably, the three proteins could form a ternary complex involved in strand discrimination, however, whether MutSα and MutLα can interact simultaneously with PCNA is controversial (Lee and Alani, 2006; Dzantiev et al., 2004). Bacterial MutS has two binding sites for the β-clamp (Lopez de Saro et al., 2006; Simmons et al., 2008). In *B. subtilis*, the C-terminal site is necessary to recruit MutL to mismatches and to activate the MMR response (Simmons et al., 2008). We presume that the ATPase domain of MutL could interact with MutS, while its C-terminal domain interacts with the β-clamp (Figure 3.4B). This model is supported by the presence of the β-binding like motif (^{487}QEMIV) within the C-terminal domain of MutL and the fact that a PCNA binding sequence has been identified in the dimerization region of yeast MLH1 (Lee and Alani, 2006).

Collectively, our data pose an attractive model where the endonuclease activity of MutL is repressed by impaired DNA binding. Based on this data the simplest mechanism would harness MutL and β-clamp allowing for DNA binding and licensing of the endonuclease activity. Consequently, the structure provides a platform for future mechanistic studies of MutL-MutS-β at the early steps of mismatch repair.

3.6 Acknowledgements

We thank Michael d'Elia, Pablo Romero and the PXRR staff at NSLS-BNL for technical assistance. We are grateful to Drs. Junop and Ortega for critical reading of the manuscript. This work was supported by the National Sciences and Engineering Research Council of Canada (288295 to A.G.), the German Science Foundation (FR-1495/4-1 to P.F.), start-up funds from the University of Michigan (to L.A.S.) and the National Institutes of Health (CA21615 to G.C.W. and GM45190 to P.M.). M.C.P. is supported by an NSERC scholarship and G.C.W. is an American Cancer Society Research Professor.

3.7 Supplementary Figures

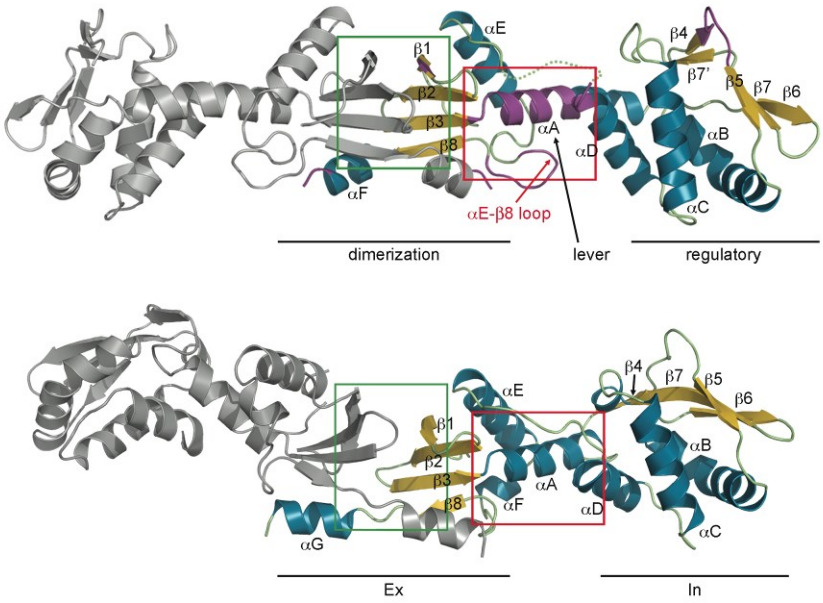


Figure S3.1 (related to Figure 3.1). Comparison of the C-terminal domains of *B. subtilis* and *E. coli* MutL. Ribbon diagrams of the BsMutL-CTD (top) and EcMutL-CTD (bottom) dimers. The dimerization and regulatory subdomains, as well as the
Figure S3.1 continued on page 119…

Figure S3.1 continued from page 118…
connecting lever, are labeled. Original names of the EcMutL-CTD subdomains are indicated (Guarné et al., 2004). The most striking differences between the two structures are the dimerization interface (green box) and the organization of the secondary structure elements surrounding helix αA (red box), that would preclude the formation of functional endonuclease site even if EcMutL had the conserved $DQHA(X)_2E(X)_4E$, ACR, C(P/N)HGRP and FXR motifs. Note the presence of the extended αE-β8 loop in BsMutL -absent in EcMutL- that secludes the endonuclease site.

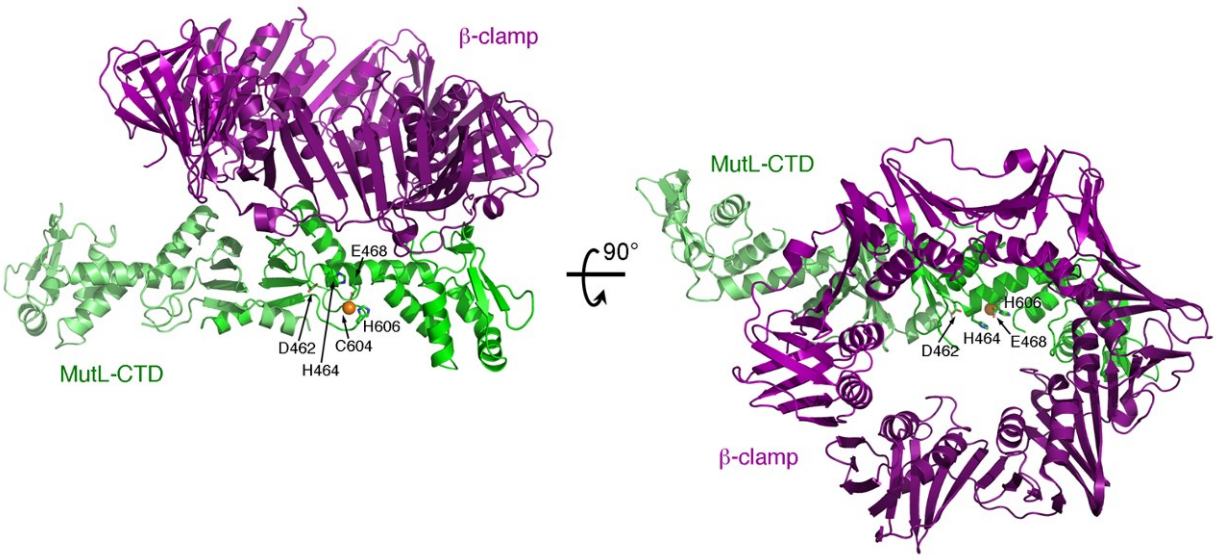


Figure S3.2 (related to Figure 3.3). Structural model for the interaction of BsMutL-CTD and β-clamp. Orthogonal views of the structure of the Zn^{2+}-bound form of BsMutL-CTD (crystal form III) shown in green modeled onto the structure of the β-clamp from *E. coli* (purple) in complex with a peptide from polymerase II (PDB ID: 3D1E). To model the complex, residues ^{487}QEMIV for BsMutL-CTD were superimposed onto the polymerase peptide ^{506}QLGLF. The proteins are shown as ribbon diagrams with the side chains involved in catalysis or Zn^{2+}-binding shown as sticks. The Zn^{2+} metal ion is shown as an orange sphere. The model reveals that the interaction of one protomer of the BsMutL dimer with the β-clamp precludes the interaction of the other one due to steric hindrance.

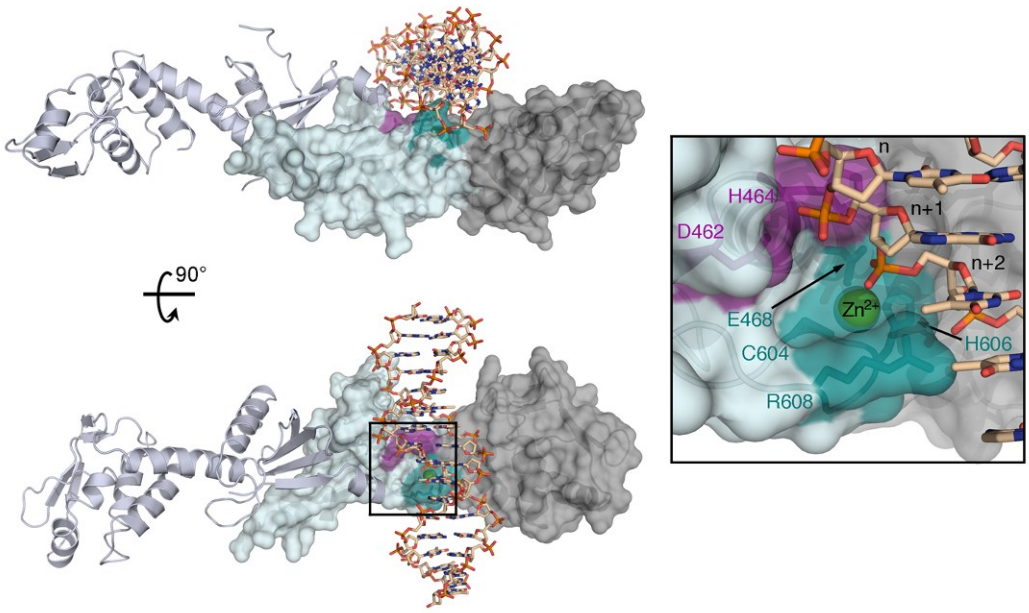


Figure S3.3 (related to Figure 3.3). Structural model for the BsMutL-CTD/DNA interaction. A standard B-DNA duplex (PDB ID: 3BSE) was modeled onto the BsMutL-CTD structure (crystal form III) based on four criteria: 1) avoidance of steric hindrance between DNA and protein; 2) distance between Asp462 and the scissile bond compatible with catalysis mediated by one or two metal ions; 3) putative role of the adjacent 3' phosphate group in coordinating the Zn^{2+} metal ion; and 4) electrostatic potential compatibility between the dimerization subdomain and the DNA duplex. The structure of BsMutL-CTD is shown as a ribbon diagram (one protomer) and a surface with the endonuclease site highlighted in purple and the Zn^{2+}-binding site shown in teal (the second protomer). The Zn^{2+} ion is shown as a green sphere and labeled. The DNA structure is shown as color-coded sticks.

Movie S3.1. Rearrangement of the endonuclease site of BsMutL-CTD in the different crystal forms. Ribbon diagram of the superimposed dimerization subdomains of the four monomers found in crystal form I and the two monomers found in crystal forms II and III. The endonuclease motifs are colored green (^{462}DQHAX$_2$EX$_4$E), purple (^{604}CPHRGP) and orange (^{572}SCK). Hydrogen bonds are shown as black dashed lines and three reference distances (His464 to Glu468, Cys573 to His606 and Glu473 to Ser599) are shown as red dashed lines and labeled. Water molecules are depicted as red spheres and Zn^{2+} ions as lilac spheres.

3.8 Supplemental Experimental Procedures

3.8.1 Cloning, purification and crystallization

Full-length BsMutL was amplified from genomic DNA and cloned into the pProEXHTa expression vector (Life Technologies). His-tagged BsMutL was purified using a Ni^{2+}-chelating affinity column equilibrated with 20 mM TRIS pH 8, 0.5 M NaCl, 1.4 mM β-mercaptoethanol, 5% glycerol and 100 mM PMSF. BsMutL was eluted using 300 mM imidazole and subsequently injected on a hydrophobic column equilibrated with 20 mM TRIS pH 8, 1 M KCl, 1 mM DTT and 5% glycerol. BsMutL was further purified by ionic exchange and size exclusion chromatography (MonoQ 5/50 and Superdex-S200, GE Healthcare) equilibrated with 20 mM TRIS pH 8, 100 mM KCl, 1 mM DTT and 5% glycerol (storage buffer). A structure-based sequence alignment was used to subclone the C-terminal fragment of BsMutL (BsMutL-CTD, residues 433-627). BsMutL-CTD was purified similarly to BsMutL with an additional ionic exchange purification step after His-tag cleavage with TEV-protease. Mutants of BsMutL and BsMutL-CTD were generated by QuikChange (Stratagene) and verified by DNA sequencing (MOBIX, McMaster University).

Crystal form I was grown in 25% PEG-monomethyl ether 550, 0.1 M $MgCl_2$, 0.1 M TRIS pH 9 and 5% PEG 400. Two additional crystal forms were obtained when the protein was supplemented with 50 nM $ZnCl_2$ and 50 nM $CoCl_2$. Crystal form II was grown in 25 % PEG 3,350, 0.15-0.2 M NaCl and 0.1 M TRIS pH 7. Addition of 0.1 mM

$ZnCl_2$ to this crystallization solution yielded crystal form III. 10 % PEG 400 was added to all crystallization conditions prior to flash freezing in liquid nitrogen.

3.8.2 Data collection and structure determination

For crystal form I, a three-wavelength MAD data set was collected at X29B in NSLS, Brookhaven National Laboratory (Upton, NY). Data were indexed, processed and merged using HKL2000 (Otwinowski and Minor, 1997). Twenty-seven out of thirty-six Selenium sites were found and refined using SOLVE (Terwilliger and Berendzen, 1999). A native data set to 2.5 Å was used for subsequent manual building and refinement, which was done using standard protocols in phenix.refine and COOT (Afonine et al., 2005; Emsley and Cowtan, 2004). Complete data sets of crystal forms II and III, were collected at the X25 beamline in NSLS. Data were collected at a wavelength corresponding to the Zn^{2+}-absorption edge (Table 1), as measured using fluorescence scans. All final models have over 92% of the residues within the most favored regions in the Ramachandran plot and none in disallowed regions. Figures depicting molecular structures were generated using PyMol (DeLano, 2002).

3.8.3 Endonuclease and DNA-binding assays

BsMutL nicking activity was assayed on supercoiled pUC19 (Fermentas). Digestion with the nicking enzyme NbBsrDI (New England Biolabs) was used as a positive control. DNA (5 nM) was incubated with MutL (250 nM, dimer) in 20 mM HEPES pH 7.5, 20 mM KCl, 0.2 mg/mL bovine serum albumin, 1% glycerol

(endonuclease buffer) and 5 mM divalent ion for 90 minutes at 37 ˚C. Nicking reactions were stopped with 0.5 mM EDTA and proteinase K (for 15 minutes at 55 ˚C).

To assess DNA binding by BsMutL, supercoiled DNA (5 nM) was incubated with BsMutL variants (100 nM, dimer) in endonuclease buffer for 90 minutes at 37 ˚C. Reaction mixtures (20 μL) were resolved on 1% TAE agarose gels and quantified using a UVP BioDoc-It^{TM} System and ImageJ (http://rsbweb.nih.gov/ij/).

3.8.4 Zinc-affinity fluorescence assay

Wild type and variants of BsMutL-CTD (2 μM) were incubated with 1 μM FluoZin-3 (Invitrogen) in 10 mM HEPES pH 6 and 200 mM KCl buffer treated with 1% Chelex-100 (Fluka). A calibration fluorescence curve was generated using buffer including increasing concentrations of $ZnCl_2$ (0.25-3 μM) in the absence of protein. Spectra (500-600 nm) were recorded at an excitation wavelength of 494 nm (corrected for buffer effects). Inner-filter effects were neglected because a linear fluorescence intensity response up to stoichiometric amounts of Zn^{2+} was seen when using 4 μM FluoZin-3.

3.8.5 Mismatch repair assays of BsMutL variants

Relative mutation frequency was determined as described earlier (Simmons et al., 2008). *B. subtilis* cells were grown at 37 ^{o}C to O.D.$_{600}$ of ~1.0, collected by centrifugation and resuspended in 0.85% saline. A portion of the cells was serial diluted into 0.85% saline followed by plating on LB agar for determining the number of viable cells per mL

of culture. The remaining portion of cells was plated on LB agar plates supplemented with 100 μg/mL rifampicin. The mutation frequency was determined for each strain by comparing the number of rifampicin resistant cells relative to the number of viable cells. The mutation frequency was then normalized to the mutation frequency obtained for the wild type control strain LAS284 (relevant genotype: *amyE*::P_{spank}*mutL*$^{+}$, *mutL*::*spc*) (see table below). LAS284 was used because each *mutL* variant was expressed in single copy from the *amyE* locus under control of the P_{spank} IPTG inducible promoter. The native *mutL* gene was inactivated by insertion of a spectinomycin resistance cassette (Simmons et al., 2008). The mutation frequency of LAS284 is 1.8-fold higher than the parent strain PY79. The observed mutation frequency for wild type PY79 is $2.77X10^{-9} \pm 9.10X10^{-10}$.

Table S3.1. List of *B. subtilis* strains

Strain	Relevant Genotype	Reference
PY79	Prototroph, SPβ^{o}	(Youngman et al., 1984)
LAS284	*amyE*::P_{spank}*mutL*$^{+}$, *mutL*::*spc*	This work
AK31	*amyE*::P_{spank}*mutL (D462A), mutL*::*spc*	This work
AK32	*amyE*::P_{spank}*mutL (Q463A), mutL*::*spc*	This work
AK33	*amyE*::P_{spank}*mutL (H464A), mutL*::*spc*	This work
AK34	*amyE*::P_{spank}*mutL (Q487A), mutL*::*spc*	This work
AK35	*amyE*::P_{spank}*mutL (I490D), mutL*::*spc*	This work
AK36	*amyE*::P_{spank}*mutL (P492A), mutL*::*spc*	This work
AK37	*amyE*::P_{spank}*mutL (H606A), mutL*::*spc*	This work
AK38	*amyE*::P_{spank}*mutL-(^{487}AAAAA^{491}), mutL*::*spc*	This work
LAS286	*amyE*::P_{spank}*mutL (E468K), mutL*::*spc*	This work
LAS287	*amyE*::P_{spank}*mutL (E473K), mutL*::*spc*	This work

3.9 References

Afonine, P.V., Grosse-Kunstleve, R.W., and Adams, P.D. (2005). phenix.refine. CCP4 Newsletter *42*, contribution 8.

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

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proc Natl Acad Sci U S A *98*, 11627-11632.

DeLano, W.L. (2002). The PyMOL Molecular Graphic Systems (Palo Alto, CA, USA: DeLano Scientific).

Duppatla, V., Bodda, C., Urbanke, C., Friedhoff, P., and Rao, D.N. (2009). The carboxy-terminal domain is sufficient for endonuclease activity of Neisseria gonorrhoeae MutL. Biochem J.

Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M., and Modrich, P. (2004). A defined human system that supports bidirectional mismatch-provoked excision. Mol Cell *15*, 31-41.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

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

Guarné, A., Junop, M.S., and Yang, W. (2001). Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. Embo J *20*, 5521-5531.

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

Hall, M.C., Shcherbakova, P.V., and Kunkel, T.A. (2002). Differential ATP binding and intrinsic ATP hydrolysis by amino-terminal domains of the yeast Mlh1 and Pms1 proteins. J Biol Chem *277*, 3673-3679.

Iyer, R.R., Pohlhaus, T.J., Chen, S., Hura, G.L., Dzantiev, L., Beese, L.S., and Modrich, P. (2008). The MutSalpha-proliferating cell nuclear antigen interaction in human DNA mismatch repair. J Biol Chem *283*, 13310-13319.

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

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

Kosinski, J., Plotz, G., Guarné, A., Bujnicki, J.M., and Friedhoff, P. (2008). The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family. J Mol Biol *382*, 610-627.

Kosinski, J., Steindorf, I., Bujnicki, J.M., Giron-Monzon, L., and Friedhoff, P. (2005). Analysis of the quaternary structure of the MutL C-terminal domain. J Mol Biol *351*, 895-909.

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

Lee, S.D., and Alani, E. (2006). Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA. J Mol Biol *355*, 175-184.

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

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

Lopez de Saro, F.J., Marinus, M.G., Modrich, P., and O'Donnell, M. (2006). The beta sliding clamp binds to multiple sites within MutL and MutS. J Biol Chem *281*, 14340-14349.

Mauris, J., and Evans, T.C. (2009). Adenosine triphosphate stimulates Aquifex aeolicus MutL endonuclease activity. PLoS One *4*, e7175.

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

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, C.W. Carter, and R.M. Sweeet, eds. (New York: Academic Press), pp. 307-326.

Peltomaki, P. (2005). Lynch syndrome genes. Fam Cancer *4*, 227-232.

Plotz, G., Welsch, C., Giron-Monzon, L., Friedhoff, P., Albrecht, M., Piiper, A., Biondi, R.M., Lengauer, T., Zeuzem, S., and Raedle, J. (2006). Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic Acids Res *34*, 6574-6586.

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

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

Simmons, L.A., Davies, B.W., Grossman, A.D., and Walker, G.C. (2008). Beta clamp directs localization of mismatch repair in Bacillus subtilis. Mol Cell *29*, 291-301.

Terwilliger, T.C., and Berendzen, J. (1999). Automated structure solution for MIR and MAD. Acta Crystallo. D *55*, 849-861.

Yang, W. (2008). An equivalent metal ion in one- and two-metal-ion catalysis. Nat Struct Mol Biol *15*, 1228-1231.

Youngman, P., Perkins, J.B., and Losick, R. (1984). Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene. Plasmid *12*, 1-9.

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

Chapter 4

The endonuclease domain of MutL interacts with the β sliding clamp

Reprinted from DNA Repair (Amst), Vol. 10, M.C. Pillon, J.H. Miller, and A. Guarné, The endonuclease domain of MutL interacts with the β sliding clamp, 87-93, Copyright 2011, with permission from Elsevier.

4.1 Author's Preface

Our structural and functional characterization of the *B. subtilis* MutL endonuclease domain (chapter 3) suggests there is an interaction with the β clamp that is important for MMR. Similar to the MutS-MutL complex characterized in chapter 2, the MutL-β clamp interaction likely coordinates MMR events. In this chapter we identify a direct, but weak interaction between MutL and the β clamp. This weak interaction is dependent on a conserved motif located in the regulatory subdomain of MutL. The MutL-β clamp interaction is conserved in *B. subtilis* and *E. coli* suggesting there is another putative role for this complex that extends beyond activating the MutL endonuclease. I generated the MutL variants and monitored the MutL-β clamp interaction using a chemical crosslinker. J.H. Miller conducted *in vivo* DNA mismatch repair assays (Table 4.1) to complement the *in vitro* work. A. Guarné and I designed and interpreted the experiments. I prepared the figures and A. Guarné and I wrote the manuscript.

4.2 Abstract

Mismatch repair corrects errors that have escaped polymerase proofreading enhancing replication fidelity by at least two orders of magnitude. The β and PCNA sliding clamps increase the polymerase processivity during DNA replication and are important at several stages of mismatch repair. Both MutS and MutL, the two proteins that initiate the mismatch repair response, interact with β. Binding of MutS to β is important to recruit MutS and MutL to foci. Moreover, the endonuclease activity of human and yeast MutLα is stimulated by PCNA. However, the concrete functions of the processivity clamp in the repair steps preceding DNA resynthesis remain obscure. Here, we demonstrate that the C-terminal domain of MutL encompasses a bona fide β-binding motif that mediates a weak, yet specific, interaction between the two proteins. Mutation of this conserved motif correlates with defects in mismatch repair, demonstrating that the direct interaction with β is important for MutL function. The interaction between the C-terminal domain of MutL and β is conserved in both *B. subtilis* and *E. coli*, but the repair defects associated with mutation of this β-binding motif are more severe in the former, suggesting that this interaction may have a more prominent role in methyl-independent than methyl-directed mismatch repair systems. Together with previously published data, our work strongly suggests that β may stimulate the endonuclease activity of MutL through its direct interaction with the C-terminal domain of MutL.

4.3 Introduction

Mismatch repair (MMR) corrects errors that have escaped polymerase proofreading, thereby enhancing replication fidelity (Kunkel and Erie, 2005). Mutations inactivating mismatch repair proteins lead to genome instability and, in humans, a predisposition to sporadic and hereditary cancers (Peltomaki, 2005). In the prototypical Gram-negative bacterium *Escherichia coli*, three proteins work together to identify the mismatch and target the newly synthesized strand for repair. MutS recognizes a mismatched base pair or a small insertion/deletion loop and recruits a molecular matchmaker, MutL, which in turn activates the latent endonuclease MutH. MutH cleaves the unmethylated DNA strand at hemimethylated GATC sites transiently generated during DNA replication. Subsequently, helicase II (UvrD), exonucleases (ExoI and ExoX or RecJ and ExoVII, depending whether the nick is generated 3' or 5' to the mismatch), single-strand binding protein (SSB), DNA polymerase III holoenzyme and DNA ligase are recruited to excise the newly synthesized strand slightly past the mismatch and re-synthesize it correctly (reviewed in (Iyer et al., 2006; Kunkel and Erie, 2005)).

Mismatch repair proteins are evolutionary conserved and homologues of MutS and MutL have been identified in all organisms. Although most bacteria and all eukaryotes do not encode a MutH homolog, a pre-existing nick is sufficient to activate mismatch repair in a system reconstituted from purified proteins (Zhang et al., 2005). It has been shown that certain human and yeast homologues of MutL (hPMS2, hMLH3, yPMS1 and yMLH3) encompass a PCNA-dependent endonuclease activity (Kadyrov et

al., 2006; Kadyrov et al., 2007). PCNA, and its bacterial counterpart the β subunit of DNA polymerase III (β), are ring-shaped structures that enhance polymerase processivity by creating a topological link with the DNA template and enabling sliding during chain elongation (O'Donnell et al., 1992), hence they are often referred to as sliding clamps. Beyond replication, sliding clamps are also important for exchanging polymerases when the replication fork encounters damaged DNA and in orchestrating post-translational repair (Friedberg et al., 2005; Kunkel and Erie, 2005). Most proteins that interact with PCNA and β do so through extended sequences at their N- or C-termini. However, while the PCNA-interacting motif (PIP box, QxxLxxFF) has a strictly defined consensus sequence, the β-interacting motif (QLxLF) is poorly conserved and absent in a number of β-binding proteins.

PCNA interacts with eukaryotic MutSα (hMSH2-hMSH6) and MutLα (hMLH1-hPMS2) (Gu et al., 1998; Lee and Alani, 2006; Umar et al., 1996). It has also been shown to interact with exonuclease I (EXOI) and to co-localize with EXOI at DNA replication foci (Nielsen et al., 2004). PCNA plays a clear role in DNA re-synthesis and it is presumed to function at earlier steps of mismatch repair, however its role remains unclear. Similarly, bacterial MutS has two binding sites for β, a weak site at the N-terminus, within the mismatch binding domain, and a stronger site at its C-terminus (Lopez de Saro et al., 2006; Simmons et al., 2008). In the Gram-positive bacterium *Bacillus subtilis*, the C-terminal site is necessary to recruit MutS to mismatches and to activate the MMR response (Simmons et al., 2008). One putative β-interacting site has

been identified within the ATPase domain of *E. coli* MutL (Lopez de Saro et al., 2006). Interaction with β mediated through this β-binding site is regulated by conformational changes induced by nucleotide- and single-stranded DNA binding to MutL (Lopez de Saro et al., 2006). However, mutation of this motif only reduces the interaction with β, suggesting that additional β-binding sites may be present.

The recent crystal structure of the endonuclease domain of *B. subtilis* MutL (BsMutL-CTD) has revealed that three conserved motifs (ACR, C(P/N)HGRP and FXR) cluster around the endonuclease motif ($DQHA(X)_2E(X)_4E$) to define a unique active site (Pillon et al., 2010). Sequence analysis also revealed the presence of an additional conserved motif within the C-terminal domain of MutL that is unrelated to the endonuclease activity of the protein (Kosinski et al., 2008). The consensus sequence of this motif loosely resembles that of the β-binding motif and its location in the structure of BsMutL-CTD suggests that it could mediate protein-protein interactions (Pillon et al., 2010). Here we present evidence demonstrating that this conserved motif is a bona fide β-binding motif and that it indeed mediates the interaction between the C-terminal domain of MutL and β. Disruption of this motif abrogates the interaction between MutL and β in both *E. coli* and *B. subtilis*, suggesting that this interaction is conserved in both methyl-directed and methyl-independent mismatch repair systems. However, the mutator phenotype associated with the disruption of this motif is milder in *E. coli* than in *B. subtilis*, implying that the interaction between MutL and β may have a more prominent role in methyl-independent mismatch repair systems.

4.4 Materials and Methods

4.4.1 Cloning of MutL variants

BsMutL-CTD (pAG8188, residues 433 to 627) was cloned as described earlier (Pillon et al., 2010). The BsMutL-CTD* variant (pAG8350, encompassing a ^{487}QEMIVP mutated to ^{487}AEMAAP) was generated by QuikChange (Stratagene) using pAG8188 as template and the regulatory subdomain of BsMutL, BsMutL-RGD (pAG8313, residues 471 to 574), was subcloned into the pProExHTa vector using the NcoI and XhoI restriction sites. The pET15b plasmids encoding full-length (EcMutL, pTX418) and the C-terminal domain of *E. coli* MutL (pWY1295, residues 432 to 615) were a kind gift from Dr. Wei Yang. The EcMutL-CTD* variant (pAG8417, ^{482}QPLLIP to ^{482}ASAAAP) was generated by overlap PCR and subcloned in pET15b between the NdeI and BamHI restriction sites. The regulatory subdomain of EcMutL, EcMutL-RGD (pAG8442, residues 466 to 569), was subcloned into the pProExHTa vector using the NcoI and XhoI restriction sites. For the *in vivo* mismatch repair assays, three variants of full-length EcMutL were generated by overlap PCR and subcloned in pET15b: pAG8472 (EcMutL-Q482A), pAG8480 (EcMutL-L485A) and pAG8477 (EcMutL*, ^{482}QPLLIP to ^{482}ASAAAP). All plasmids were verified by DNA sequencing (MOBIX, McMaster University).

4.4.2 Protein expression and purification

All MutL variants were overproduced and purified as described earlier with minor modifications (Ban and Yang, 1998; Guarné et al., 2004; Pillon et al., 2010). Purified

proteins were stored in 20 mM TRIS pH 8, 100 mM KCl, 1 mM DTT and 5% glycerol (storage buffer). The over-expression plasmids encoding *E. coli* β (residues 1-366) and *B. subtilis* β (residues 1-378) were generous gifts from Dr. Mike O'Donnell and Dr. Lyle A. Simmons. Ecβ was overproduced in *E. coli* BL21 (DE3) Star cells and Bsβ in BL21 (DE3) recA$^-$ cells as described earlier (Kong et al., 1992; Simmons et al., 2008). Cell pellets containing Ecβ were resuspended in buffer A (20 mM Tris pH 8.0, 1 mM EDTA, 5 mM DTT, 0.05 M NaCl, and 5% glycerol), incubated with 0.6 mg/ml lysozyme for 30 minutes on ice, and cells disrupted by sonication. Lysates were clarified by centrifugation at 39,000 g and loaded onto a heparin column connected in tandem to a Q-sepharose column equilibrated with buffer A. Purified Ecβ was eluted from the Q-sepharose column using a salt gradient to 0.5 M NaCl. The sample was further purified over a MonoQ 5/50 column (GE Healthcare).

Cell pellets containing Bsβ were resuspended in 50 mM Tris pH 7.0 and 10% sucrose and lysed by freeze-thaw in 0.5 M NaCl, 20 mM $SpCl_3$, and 0.45 mg/ml lysozyme. Cell lysates were clarified by centrifugation as above and the soluble fraction containing Bsβ was loaded onto a HiTrap nickel-chelating column equilibrated with 20 mM Tris pH 7.6, 0.5 M NaCl, 1.4 mM 2-mercaptoethanol, 45 mM imidazole, and 15% glycerol and eluted with 0.24 M imidazole. The sample was subsequently purified by ionic exchange using a Q-sepharose column pre-equilibrated with 20 mM Tris pH 7.6, 1 mM EDTA, 5 mM DTT, 0.15 M KCl, and 15% glycerol using a salt gradient to 0.4 M KCl. All subsequent experiments were performed with His-tagged Bsβ.

4.4.3 Protein Cross-linking with BS^3

Frozen protein samples were thawed and loaded onto a Superdex-S200 (GE Healthcare) equilibrated in 20 mM Hepes pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM DTT, and 5% glycerol (reaction buffer) to exchange the buffer, as well as, to remove protein aggregates caused during freezing/thawing. MutL variants (20 μM) or β (10 μM) were incubated with 0.4, 0.8, 1.0, and 1.2 mM Bis(sulfosuccinimidyl)suberate (BS^3, Sigma) in reaction buffer for 30 minutes at 22°C. Crosslinking reactions (10 μL) were quenched with 30 mM Tris pH 7.5 for 15 minutes at 22°C and resolved on 10% SDS polyacrylamide gels stained with Coomassie blue. Optimal crosslinking for both proteins was observed at BS^3 concentrations of 0.8 mM and hence all subsequent controls using the individual proteins were done at this crosslinker concentration. MutL and β were incubated at 2:1 molar ratios to account for the possibility that two MutL dimers were bound to a dimer of β. Mixtures of MutL:β (2:1) were incubated with 0.4, 0.8, 1.0, and 1.2 mM BS^3 and the reaction was quenched after 30 minutes as described above. The reaction products were resolved on 4-15% SDS polyacrylamide gels (BioRad) stained with Coomassie blue. The identity of the bands was confirmed by mass spectrometry (Bioanalytical and Spectrometry Laboratory, McMaster University).

4.4.4 Determination of Mutation rates

The CC107 *mutL::miniTn10* strain was constructed by transducing CC107 (Cupples et al., 1990) to Tet^r with a P1 *vir* lysate grown on strains carrying a mini*Tn10* insertion in the *mutL* gene (Miller et al, unpublished data). CC107 *mutL::miniTn10* cells

were then transformed with either pET15b, pTX418, pAG8472, pAG8477 or pAG8480. 1 or 2 ml Luria-Bertani broth cultures containing 100 μg/ml ampicillin were seeded with 100-1,000 transformed cells and grown overnight. Samples were plated on LB plates containing 100 μg/ml rifampicin, and dilutions were plated on LB plates with ampicillin to determine the titer. The frequencies of rifampicin resistant (Rif^r) mutants were determined, and the median frequency (f) from a set of cultures was used to calculate the mutation rate per replication by the method of Drake, using the formula $\mu = f/\ln N\mu$, where N is the number of cells in the culture (Drake, 1991). Ninety-five percent confidence limits were determined according to Dixon and Massey (Dixon and Massey, 1969).

4.5 Results

4.5.1 Putative β-binding motif in MutL-CTD

Sequence analysis has identified a conserved motif with a consensus sequence of QxΦ[L/I]xP (where Φ is a hydrophobic residue) within the C-terminal domain of MutL. This motif is found in prokaryotic and eukaryotic MutL homologues harboring an endonuclease activity, but also in those that do not (Figure 4.1A and (Kosinski et al., 2008)). The structures of the C-terminal domains of *E. coli* and *B. subtilis* MutL reveal that this chiefly hydrophobic motif resides in the regulatory subdomain of the C-terminal region and it is conspicuously exposed to the solvent (Figure 4.1B-C and (Guarné et al., 2004; Pillon et al., 2010)). Superimposition of the regulatory subdomains of BsMutL and EcMutL returns root mean square deviations of 2 Å for 89 Cα-atoms (Figure 4.2A). This

value is twice that obtained when superimposing the dimerization subdomain of both proteins. However, the conformation of the conserved QxΦ[L/I]xP motif (^{487}QEMIVP in BsMutL and ^{482}QPLLIP in EcMutL) is virtually identical in both structures (r.m.s.d. of 0.27 Å), suggesting that it could mediate specific interactions between MutL and other repair factors.

Since the endonuclease activity of human and yeast MutLα is highly stimulated by PCNA (Kadyrov et al., 2006), we suspected that this ubiquitous motif could mediate the interaction between MutLα and PCNA, or between bacterial MutL and β (Pillon et al., 2010). Indeed, the QxΦ[L/I]xP motif loosely resembles the consensus β-binding motif (QLxLF) (Dalrymple et al., 2001). In all the β structures bound to peptides or small protein domains encompassing a β-binding motif, the motif adopts an extended conformation that lines the C-terminal end of β. Two critical residues, glutamine at position one and leucine at position four, anchor this interaction. Additional hydrophobic and bulky residues at positions five or six strengthen the interaction, but they are less conserved and adopt variable conformations (Figure 4.2B). The putative β-binding motifs in BsMutL-CTD (^{487}QEMIVP) and EcMutL-CTD (^{482}QPLLIP) contain the conserved Gln and Leu/Ile at positions one and four. Superimposition of these motifs onto several β-binding motifs bound to β (or PIP motifs bound to PCNA) returned root mean square deviations smaller than 0.4 Å, strongly suggesting that this loop could mediate the interaction between the C-terminal domain of MutL and β. Indeed, the first

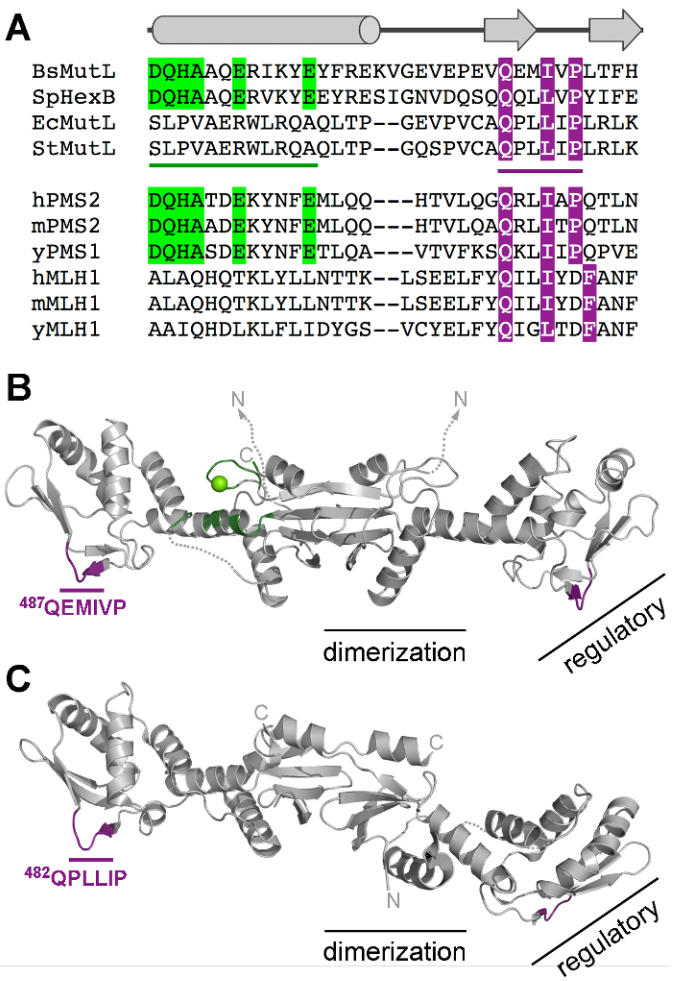


Figure 4.1. Conserved putative β-binding motif in MutL. (A) Sequence alignment of the putative β-binding motif (highlighted in purple) found in MutL and its location with respect to the endonuclease motif when present (highlighted in green). The top group includes *Bacillus subtilis* MutL (BsMutL) and *Streptococcus pneumoniae* HexB (SpHexB) that contain the conserved endonuclease motif, as well as, *Escherichia coli* MutL (EcMutL) and *Salmonella typhimurium* MutL (StMutL) that do not have endonuclease activity. The bottom group includes the eukaryotic *Homo sapiens* PMS2 (hPMS2), *Mus musculus* PMS2 (mPMS2), *Saccharomyces cerevisiae* PMS1 (yPMS1), as well as, eukaryotic MutL homologues that do not encompass an endonuclease motif: *Homo sapiens* MLH1 (hMLH1), *Mus musculus* MLH1 (mMLH1), *Saccharomyces cerevisiae* MLH1 (yMLH1). **(B)** Ribbon diagram of the C-terminal domain of *Bacillus subtilis* MutL (PDB ID: 3KDK) with the endonuclease motif shown in green and the β-binding motif shown in purple. The structural Zn^{2+} metal ion found at the endonuclease site is depicted as a green sphere. The N- and C-termini, the β-binding motif and the dimerization and regulatory subdomains are labeled for clarity. **(C)** Ribbon diagram of the C-terminal domain of *Escherichia coli* MutL (PDB ID: 1X9Z) shown as in (B).

four residues of the motif adopt the same extended conformation seen in other bona fide β-binding motifs with the side chains of the conserved Gln and Leu/Ile occupying the major binding pockets on β and PCNA (Figure 4.2C). At position five, the main chain changes direction, in a similar manner to the consensus PIP motif in FEN-1 (PDB 1RXM), and it is followed by a sharp kink at the following proline that projects the MutL away from the ancillary binding pockets in the β-clamp (Figure 4.2C).

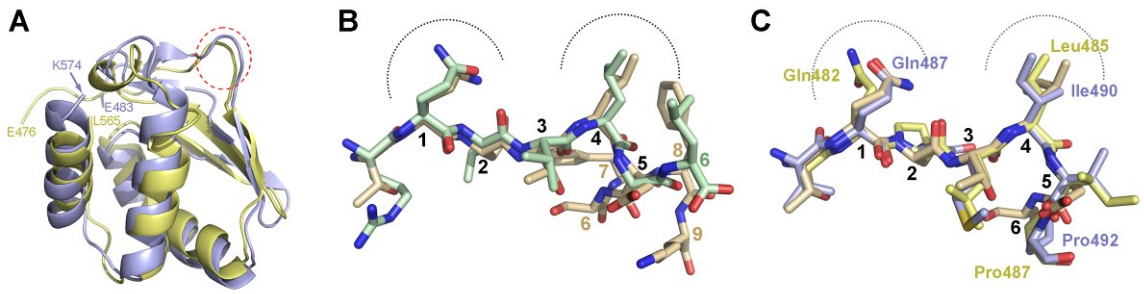


Figure 4.2. Structural comparison of the β-binding motifs found in MutL-CTD and other clamp-binding proteins. (A) Superimposition of the regulatory subdomains of *B. subtilis* MutL (blue) and *E. coli* MutL (yellow) shown as ribbon diagram. The N- and C-terminal domain boundaries are labeled. **(B)** Superimposition of the β-binding motif from polIV/DinB (green, PDB ID: 1UNN) and the PIP-box from FEN-1 (tan, PDB ID: 1RXM). Dotted lines depict the polar (left) and hydrophobic (right) pockets occupied by the conserved Gln and Leu residues. **(C)** Superimposition of the β-binding motifs of *B. subtilis* MutL (blue) and *E. coli* MutL (yellow) onto the structure of FEN-1 bound the PCNA (tan, PDB ID: 1RXM). Conserved residues are labeled.

4.5.2 *Bacillus subtilis* MutL-CTD interacts specifically with β

We wanted to probe whether this conserved motif in MutL could indeed mediate the interaction with β. However, we could not detect the interaction by either size

exclusion chromatography or fluorescence anisotropy (data not shown), suggesting that the interaction between BsMutL-CTD and Bsβ is probably transient. Since transient interactions can often be detected using chemical crosslinkers, we treated BsMutL-CTD and Bsβ with Bis(Sulfosuccinimidyl) suberate (BS3), a crosslinker that contains two amine reactive NHS-esters separated by an 11.4 Å spacer. Treatment of BsMutL-CTD (22,664 Da) with increasing concentrations of BS3 resulted in two crosslinked products (Figure 4.3). The first one migrated on SDS-polyacrylamide gels faster than the BsMutL-CTD monomer and could result from the intramolecular crosslinking of the dimerization and regulatory subdomains. The second product migrated at a molecular weight consistent with the formation of BsMutL-CTD dimers (45,328 Da). Similarly, treatment of the Bsβ (42,103 Da) with increasing concentrations of BS3 resulted in the formation of intra- and inter-molecular crosslinks, consistent with the formation of β dimers (84,206 Da).

When BsMutL-CTD and Bsβ (2:1) were pre-incubated prior to treatment with BS3, a new product appeared with an apparent molecular weight consistent with a BsMutL-CTD monomer bound to a Bsβ monomer (Figure 4.3, top). The presence of the two proteins in this crosslinked product was confirmed by mass spectroscopy (data not shown). The equivalent product was also seen when the regulatory subdomain of BsMutL (BsMutL-RGD, 12,136 Da) was preincubated with Bsβ (Figure 4.3, center). This region of MutL lacks the dimerization subdomain and, therefore, only monomers of BsMutL-RGD were found when the protein was incubated with BS3. The presence of an

additional crosslinked product in the samples containing both BsMutL-RGD and Bsβ indicates that the regulatory subdomain of MutL mediates the interaction with β. Since the conserved ^{487}QEMIVP motif resides in this region of MutL, we subsequently tested whether mutation of this sequence abrogated their interaction. When Bsβ and the BsMutL-CTD* variant (^{487}QEMIVP-^{487}AEMAAP, 22,537 Da) were treated with BS3, only crosslinking products corresponding to MutL-MutL or β-β interactions were detected (Figure 4.3, bottom), demonstrating that the ^{487}QEMIVP motif mediates the interaction between the C-terminal domain of BsMutL and Bsβ.

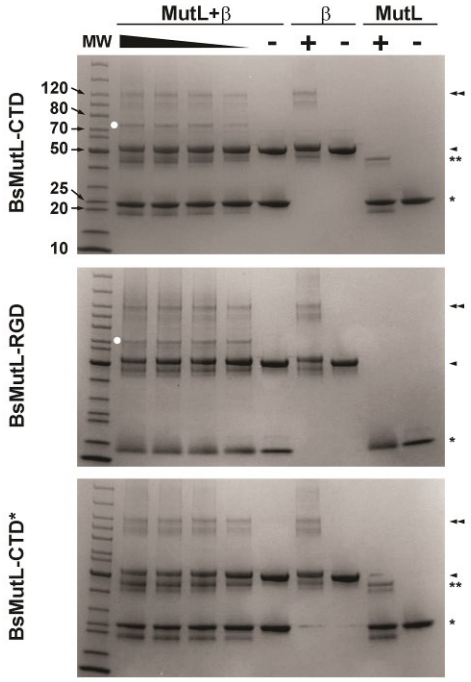


Figure 4.3. Complex formation between the endonuclease domain of *B. subtilis* MutL and β. Interaction between the C-terminal domain of BsMutL (BsMutL-CTD, top), the regulatory domain of BsMutL (BsMutL-RGD, center) or the C-terminal domain of BsMutL encompassing a mutated β-binding motif (BsMutL-CTD*, ^{487}QEMIVP-AEMAAP, bottom) with the *B. subtilis* β (β). The proteins were incubated in the presence/absence of BS3 and the reaction products were resolved by SDS-PAGE. From *Figure 4.3 continued on page 143…*

Figure 4.3 continued from page 142…
left to right, the gels show molecular weight markers (MW), mixtures of MutL (0.02 mM) and β (0.01 mM) incubated with decreasing concentrations of BS3, Bsβ incubated in the presence (+) or absence (-) of BS3, BsMutL variant (as indicated) incubated in the presence (+) or absence (-) of BS3. Monomers and dimers of BsMutL are indicated with one or two asterisks, and monomers and dimers of the β are indicated with one of two arrowheads, respectively. The presence of crosslinked products corresponding to the interaction of BsMutL with β is indicated with a white dot. Incubation of BsMutL-CTD* and β in the presence of BS3 does not result in the formation of this crosslinked product, indicating that the integrity of the β-binding motif is necessary to maintain the interaction.

4.5.3 *Escherichia coli* MutL-CTD interacts specifically with β

Since this motif is highly conserved, we next asked whether the interaction between β and MutL was conserved in other species. To this end, we assayed for a specific interaction between *E. coli* MutL-CTD (EcMutL-CTD, 20,158 Da) and *E. coli* β (Ecβ, 40,587 Da). Similarly to Bsβ, we found that crosslinking of Ecβ only resulted in the formation of dimers (81,174 Da). However, EcMutL-CTD formed dimers, as well as, larger oligomers (Figure 4.4A). This finding was not unexpected since the crystal structure of EcMutL-CTD had initially suggested that dimerization occurred by the association of the regulatory subdomains (Figure 4.4B (Guarné et al., 2004)), but ensuing analysis of the quaternary structure of EcMutL-CTD proved that dimers formed through the dimerization subdomain of the protein as depicted in Figure 4.1C and (Kosinski et al., 2005). Given the presence of multiple interaction surfaces, dimers (40,316 Da), trimers (60,474 Da), tetramers (80,632 Da) and larger oligomers were indeed expected following crosslinking with BS3 (Figure 4.4A).

Similarly to the *B. subtilis* proteins, a new crosslinking product also appeared when EcMutL-CTD and Ecβ (2:1) were pre-incubated prior to treatment with BS3 (Figure 4.4C). This product migrated on SDS-polyacrylamide gels at a molecular weight consistent with the formation of a complex between MutL-CTD and β (60,745 Da, Figure 4.4C). However, this product co-migrated in SDS-polyacrylamide gels with the EcMutL-CTD trimer (60,474 Da). Crosslinking experiments conducted at higher MutL:β ratios revealed that the band migrating at this molecular weight was indeed a doublet, suggesting the presence of both EcMutL-CTD trimers and EcMutL-CTD/Ecβ complexes (see bands labeled 1 and 2 on the Figure 4.4C inset). We verified the presence of both, MutL and β, in this crosslinked product using mass spectrometry (data not shown). As we had seen with the proteins from *B. subtilis*, the regulatory subdomain of EcMutL also supported the interaction with Ecβ (Figure 4.4D). We then assayed whether a variant of EcMutL-CTD encompassing a mutated ^{482}QPLLIP motif (EcMutL-CTD*, ^{482}QPLLIP-^{482}ASAAAP, 19,965 Da) could support the interaction and found that this variant of the protein did not form a complex with Ecβ (Figure 4.4E). Therefore, we concluded that the ^{482}QPLLIP motif mediates the interaction between the C-terminal domain of EcMutL and β.

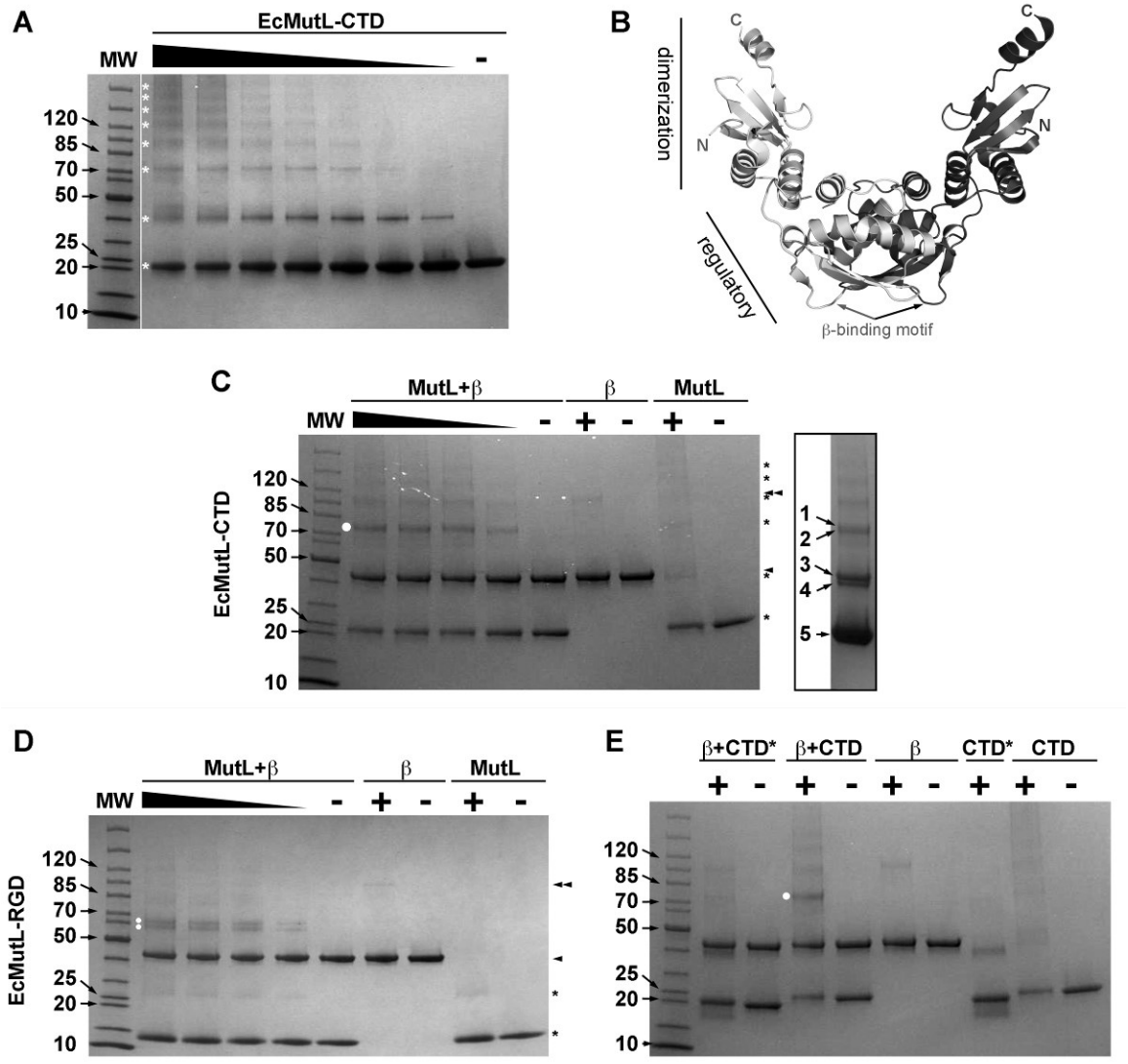


Figure 4.4. Complex formation between the C-terminal region of *E. coli* MutL and β. (A) 4-15% SDS-PAG showing the BS3-crosslinked products of the C-terminal domain of EcMutL (EcMutL-CTD). White asterisks indicate the migration of EcMutL-CTD oligomers. From left to right, the gel shows molecular weight markers (MW), EcMutL-CTD (0.04 mM) incubated with decreasing concentrations of BS3 (1.2, 1.0, 0.8, 0.5, 0.4, 0.3 0.2 and 0.1 mM) and EcMutL-CTD alone (-). **(B)** Alternative dimer interface found in the EcMutL structure (PDB ID: 1X9Z). The crystallographic dimer is maintained by the interaction of the regulatory subdomains, whereas the physiological dimer shown in Figure 4.1 is maintained by the interaction of the dimerization subdomains. **(C)** Mixtures of EcMutL-CTD and *E. coli* β were incubated in the presence or absence of BS3 and the reaction products were resolved by gradient SDS-PAGE. The protein and crosslinker concentrations, the gel layout and the labels of the bands are the same as Figure 4.3. The *Figure 4.4 continued on page 146…*

Figure 4.4 continued from page 145…
inset corresponds to the same experiment performed with excess EcMutL-CTD (0.2 mM, #5) and shows that some of the oligomeric forms of EcMutL-CTD (trimers #2 and dimers #4) migrate similarly to the β monomer (#3) and the crosslinked EcMutL-CTD/β product (#1). **(D)** Crosslinking products between the regulatory subdomain of EcMutL (EcMutL-RGD) and Ecβ obtained in the presence of BS3. The gel is shown and labeled as panel (C). **(E)** Comparison of the crosslinked products obtained with EcMutL-CTD and EcMutL-CTD* encompassing a mutated β-binding motif (^{482}QPLLIP-^{482}ASAAAP) and β. From left to right, the gels show molecular weight markers (MW), mixtures of EcMutL-CTD* and Ecβ, EcMutL-CTD and Ecβ, Ecβ, EcMutL-CTD* and EcMutL-CTD incubated in the presence (+) or absence (-) of BS3. For all reactions containing BS3 a final concentration of 1.2 mM was added.

It has been previously shown that the polC subunits from *Staphylococcus aureus*, *Streptococcus pyogenes* and *B. subtilis* can use *E. coli* β as their processivity subunit, while *E. coli* DnaE cannot use β from other species (Bruck and O'Donnell, 2000; Klemperer et al., 2000; Low et al., 1976). Therefore, we assayed whether the interaction between heterologous MutL and β could be detected. Indeed, we detected specific interactions between BsMutL-CTD and Ecβ, as well as, between EcMutL-CTD and Bsβ (Figure 4.5). These results reinforce the idea that the conserved motif in the C-terminal domain of MutL likely binds to the conserved pocket in β in a species-independent manner.

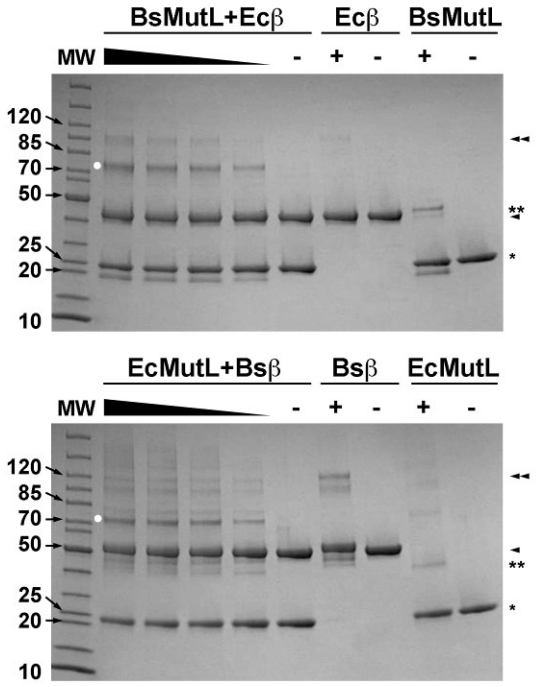


Figure 4.5. Interaction between heterologous MutL-CTD and β. Interaction between BsMutL-CTD and Ecβ (top) and EcMutL-CTD and Bsβ (bottom). From left to right, the gels show: molecular weight markers (MW), mixtures of MutL-CTD and β as indicated incubated with decreasing concentrations of BS3, β and MutL-CTD in the presence (+) or absence (-) of BS3.

4.5.4 The MutL-CTD/β interaction is important for efficient mismatch repair

We have previously shown that mutations on the conserved ^{487}QEMIVP motif of BsMutL impair mismatch repair in *B. subtilis* (Pillon et al., 2010). Therefore, we next assayed whether the integrity of the ^{482}QPLLIP motif in EcMutL was also important for mismatch repair in *E. coli*. To this end, we generated the EcMutL-Q482A, EcMutL-L485A and EcMutL-(^{482}QPLLIP-^{482}ASAAAP) variants and measured both the frequencies (*f*) and mutation rates (μ) per cell per replication (Table 4.1). As a control, we measured the mutation rates obtained for a mismatch repair deficient strain, as well as, this strain complemented with a plasmid encoding wild-type MutL, which were virtually identical to our previously published values (Garibyan et al., 2003; Guarné et al., 2004). Mutation of the entire ^{482}QPLLIP motif clearly impaired MutL function, but not to the

same extend as the *mutL*-deficient strain (Table 4.1). This finding was interesting, because disrupting the ^{487}QEMIVP motif in BsMutL impairs mismatch repair to a level comparable to the *mutL*-deficient strain (Pillon et al., 2010). Accordingly, the EcMutL-Q482A and EcMutL-L485A variants also had milder phenotypes than the corresponding BsMutL variants (Table 4.1 and (Pillon et al., 2010)). In fact, mutation of Gln482 resulted in a MutL variant that behaved similar to wild-type MutL by this particular assay. Collectively, this data suggests that the interaction between MutL and β is important for MutL function across species, but it has a more prominent role in methyl-independent than methyl-directed mismatch repair systems.

Table 4.1 Frequencies (*f*) and rates (μ) of mutations in *rpoB* of a *mutL::miniTn*10 strain transformed with plasmids encoding variants of EcMutL.

	Empty vector (pET15b)	EcMutL (pTX418)	EcMutL-Q482A (pAG8472)	EcMutL-L485A (pAG8480)	EcMutL-^{482}ASAAAP (pAG8477)
f x10^{8} *	910 (600-1,400)	3.3 (2.1-8.9)	2.7 (0.9-7.2)	272 (210-380)	66 (34-88)
μ x10^{8} **	110 (77-160)	0.9 (0.64-2.0)	0.9 (0.41-1.9)	39 (30-52)	11 (6.1-14)

* The *rpoB* mutation frequency was calculated by dividing the median number of mutants by the average number of cells in a series of cultures. Values in parentheses are the 95% confidence limits (Dixon and Massey, 1969).

** The mutation rate per cell per replication (μ) was determined by the method of Drake (Drake, 1991). Values in parentheses are 95% confidence limits.

4.6 Discussion

We have shown that the C-terminal domains of *B. subtilis* and *E. coli* MutL interact specifically with β through a conserved motif that resides within the regulatory subdomain of the protein. The motif that mediates this interaction (QxΦ[L/I]xP) partially resembles the consensus sequences of the β-binding motif (QLxLF) and the PIP-box (QxxLxxFF). These three motifs have two strictly conserved residues at positions one (Gln) and four (Ile/Leu). In the structures of β and PCNA bound to peptides encompassing β-binding motifs or PIP-boxes, respectively, these conserved residues occupy two well-defined binding pockets (Figure 4.2 and (Bunting et al., 2003; Chapados et al., 2004)). Accordingly, mutation of the conserved Gln and Ile/Leu residues in either BsMutL-CTD or EcMutL-CTD abrogated binding to β and affected mismatch repair function.

Most proteins that interact with either β or PCNA do so through extended sequences at their C-terminus. However, the β-binding motif in MutL is embedded in the regulatory subdomain and thus its conformation is restricted by the tertiary structure of the protein. The conserved proline at the C-terminus of the motif likely restrains the loop on an exposed conformation, indirectly enhancing the interaction to β. In the BsMutL-CTD* (^{487}QEMIVP to ^{487}AEMAAP) and EcMutL-CTD* (^{482}QPLLIP to ^{482}ASAAAP) variants this proline residue is intact and yet they have lost the ability to interact with β (Figures 4.3 and 4.4). Furthermore, mutation of this Pro492 in BsMutL to alanine does

not affect mismatch repair *in vivo*, whereas mutation of the conserved Gln487 and Ile490 does (Pillon et al., 2010), reinforcing the idea that this proline may help present the interacting motif but does not interact directly with the sliding clamp. Interestingly, Gly and Pro residues are often found immediately following internal β-binding motifs (Dalrymple et al., 2001), suggesting that an increased flexibility or a restrained conformation may assist at presenting these internal binding sites to β.

In vitro, PCNA stimulates the endonuclease activity of MutLα in a mismatch independent manner (Kadyrov et al., 2006; Kadyrov et al., 2007) and, by virtue of its loading orientation, seems to determine the strand direction of MutLα incision in a mismatch-dependent manner (Pluciennik et al., 2010). Therefore, this interaction must be tightly regulated to ensure that the endonuclease activity of MutLα is only stimulated when a mismatch has been encountered. The lack of aromatic residues at the C-terminus of the MutL β-binding motif could result in a weaker interaction with β than other β-binding factors. A suboptimal β-binding motif could potentially be out competed by other β-binding proteins encompassing canonical β-binding sites, thereby providing an additional level of regulation. Indeed, the interaction between yeast MutLα and PCNA appears to be mediated by yMLH1 rather than yPMS1 supporting the idea that the interaction with a canonical PIP-box (as the one found in yMLH1) is stronger than that with a suboptimal site such as the motif present in MutL homologues encompassing an endonuclease activity (Figure 4.1 and (Lee and Alani, 2006)).

Both MutS and MutL have multiple β-binding sites. MutS has a strong β-binding site located at its C-terminal domain that mediates the MutS-β interaction in solution and aids the localization of MutS to mismatches in *B. subtilis*. The integrity of this site is not essential for mismatch repair in *E. coli*, although it is critical for mismatch repair in *B. subtilis*. Additionally, a weaker β-binding site is found in the mismatch-binding domain of MutS. Mutation of this site does not disrupt the MutS-β interaction in solution, but it confers a mutator phenotype (Lopez de Saro et al., 2006; Pluciennik et al., 2009). Initially, one β-binding site was identified in MutL, located in the ATPase domain of the protein. Interaction with β mediated by this site only occurs in the presence of single stranded DNA and is weakened by nucleotide binding, presumably due to the dimerization of the ATPase domains of MutL. Mutation of this β-binding site weakens, yet does not abolish, the MutL-β interaction suggesting that other β-binding sites may be present (Lopez de Saro et al., 2006). The β-binding site of MutL characterized in the present work also mediates a specific interaction with β.

It is plausible that the two β-binding sites found in MutL orchestrate its multiple functions in mismatch repair. For instance, binding of β to the N-terminal site could aid in localizing MutL to mismatches, while binding of β to the C-terminal site could tether the endonuclease domain of MutL to DNA. The latter scenario would suggest that β, or its eukaryotic analog PCNA, stimulate the endonuclease activity of MutL by bypassing the inability of its C-terminal domain to bind DNA (Pillon et al., 2010), thereby drawing a

parallel with the way that PCNA simulates the activity of FEN-1 (Tom et al., 2000). This idea is in agreement with a recent study by Modrich and co-workers proposing that the interaction with PCNA determines the strand bias and the strand direction of MutLα incision (Pluciennik et al., 2010). Defects in this motif confer a strong mutator phenotype in *B. subtilis* (Pillon et al., 2010), but only a mild mutator phenotype in *E. coli* (Table 4.1), likely reflecting the different roles of this interaction in methyl-directed and methyl-independent mismatch repair systems. Since MutL homologues that do not encode an endonuclease activity also interact with β through this C-terminal β-binding motif, tethering of the C-terminal domain of MutL to DNA may also enhance other steps of the repair process. Indeed, both the N- and C-terminal domains of MutL are necessary to interact and activate the helicase activity of UvrD in a DNA-dependent manner (Guarné et al., 2004). Therefore, the MutL-β interaction could potentially influence activation of UvrD in *E. coli*. While this idea needs further testing, it suggests that the interaction between the C-terminal domain of MutL and β could regulate several steps in the mismatch repair process, or help MutL recruit additional repair factors to the damaged site. This, in turn, would explain why mutation of this β-binding site is more deleterious in some organisms than others and reinforces the idea that β, and its eukaryotic analog PCNA, orchestrate the sequence of events that lead to mismatch repair in the cell.

4.7 Acknowledgements

We thank Faraz Ahmadpour for subcloning the regulatory subdomain of BsMutL and Dr. Lyle Simmons and members of the Guarné laboratory for helpful discussions and critical reading of the manuscript.

4.8 References

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

Bruck, I., and O'Donnell, M. (2000). The DNA replication machine of a gram-positive organism. The Journal of biological chemistry *275*, 28971-28983.

Bunting, K.A., Roe, S.M., and Pearl, L.H. (2003). Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. EMBO J *22*, 5883-5892.

Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J.A. (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell *116*, 39-50.

Cupples, C.G., Cabrera, M., Cruz, C., and Miller, J.H. (1990). A set of lacZ mutations in Escherichia coli that allow rapid detection of specific frameshift mutations. Genetics *125*, 275-280.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proceedings of the National Academy of Sciences of the United States of America *98*, 11627-11632.

Dixon, W.J., and Massey, F.J. (1969). Introduction to Statistical Analysis (New York: McGraw-Hill).

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proceedings of the National Academy of Sciences of the United States of America *88*, 7160-7164.

Friedberg, E.C., Lehmann, A.R., and Fuchs, R.P. (2005). Trading places: how do DNA polymerases switch during translesion DNA synthesis? Molecular cell *18*, 499-505.

Garibyan, L., Huang, T., Kim, M., Wolff, E., Nguyen, A., Nguyen, T., Diep, A., Hu, K., Iverson, A., Yang, H.*, et al.* (2003). Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. DNA Repair (Amst) *2*, 593-608.

Gu, L., Hong, Y., McCulloch, S., Watanabe, H., and Li, G.M. (1998). ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. Nucleic Acids Res *26*, 1173-1178.

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

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

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

Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., and Modrich, P. (2007). Saccharomyces cerevisiae MutLa is a mismatch repair endonuclease. The Journal of biological chemistry.

Klemperer, N., Zhang, D., Skangalis, M., and O'Donnell, M. (2000). Cross-utilization of the beta sliding clamp by replicative polymerases of evolutionary divergent organisms. The Journal of biological chemistry *275*, 26136-26143.

Kong, X.P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992). Three-dimensional structure of the beta 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., and Friedhoff, P. (2008). The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family. Journal of molecular biology *382*, 610-627.

Kosinski, J., Steindorf, I., Bujnicki, J.M., Giron-Monzon, L., and Friedhoff, P. (2005). Analysis of the quaternary structure of the MutL C-terminal domain. Journal of molecular biology *351*, 895-909.

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

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

Lopez de Saro, F.J., Marinus, M.G., Modrich, P., and O'Donnell, M. (2006). The beta sliding clamp binds to multiple sites within MutL and MutS. The Journal of biological chemistry *281*, 14340-14349.

Low, R.L., Rashbaum, S.A., and Cozzarelli, N.R. (1976). Purification and characterization of DNA polymerase III from Bacillus subtilis. The Journal of biological chemistry *251*, 1311-1325.

Nielsen, F.C., Jager, A.C., Lutzen, A., Bundgaard, J.R., and Rasmussen, L.J. (2004). Characterization of human exonuclease 1 in complex with mismatch repair proteins, subcellular localization and association with PCNA. Oncogene *23*, 1457-1468.

O'Donnell, M., Kuriyan, J., Kong, X.P., Stukenberg, P.T., and Onrust, R. (1992). The sliding clamp of DNA polymerase III holoenzyme encircles DNA. Mol Biol Cell *3*, 953-957.

Peltomaki, P. (2005). Lynch syndrome genes. Fam Cancer *4*, 227-232.

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

Pluciennik, A., Burdett, V., Lukianova, O., O'Donnell, M., and Modrich, P. (2009). Involvement of the beta clamp in methyl-directed mismatch repair in vitro. The Journal of biological chemistry *284*, 32782-32791.

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

Simmons, L.A., Davies, B.W., Grossman, A.D., and Walker, G.C. (2008). Beta clamp directs localization of mismatch repair in Bacillus subtilis. Molecular cell *29*, 291-301.

Tom, S., Henricksen, L.A., and Bambara, R.A. (2000). Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. The Journal of biological chemistry *275*, 10498-10505.

Umar, A., Buermeyer, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M., and Kunkel, T.A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. Cell *87*, 65-73.

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

Chapter 5

A β-clamp variant for the stabilization of weak β complexes

Manuscript submitted to Nucleic Acids Research

Authors: M.C. Pillon, V.M.P. Babu, M.D. Sutton, and A. Guarné

5.1 Author's Preface

This chapter investigates the structural architecture of the MutL-β clamp complex described in Chapter 4. To overcome the weak affinity of this complex, we engineered *E. coli* MutL and β clamp cysteine variants that form a disulfide bridge at the known interaction interface (chapter 4) when the proteins associate. Using this stable complex, we identified two MutL-β clamp complexes in solution. Functional characterization of the MutL-β clamp complex also suggests it has a minor role in stimulating UvrD unwinding activity. I designed the MutL and β clamp cysteine variants. V.M.P. Babu (from the laboratory of Dr. M.D. Sutton) performed *in vivo* assays to ensure the β clamp variant could complement the wild type protein (Tables 5.1 and 5.2). I structurally characterized the stable MutL-β clamp complex using small angle X-ray scattering and performed *in vitro* functional assays. A. Guarné and I designed and interpreted the experiments. I prepared the figures and wrote the manuscript.

5.2 Abstract

The sliding β-clamp enhances polymerase processivity and coordinates DNA replication with other critical DNA processing events including translesion synthesis, Okazaki fragment maturation and DNA repair. The relative binding affinity of the sliding β-clamp for its partners determines how these processes are orchestrated and is essential to ensure the correct processing of newly replicated DNA. However, while stable clamp interactions have been extensively studied; dynamic interactions mediated by the sliding β-clamp remain poorly understood. Here, we characterize the interaction between the sliding β-clamp (β) and one of its weak-binding partners, the DNA mismatch repair protein MutL. Disruption of this interaction causes a mutator phenotype, confirming its importance in DNA mismatch repair. We stabilize the MutL-β interaction by engineering two cysteine residues at variable positions of the interface. Disulfide bridge crosslinking resulted in the formation of a specific complex between MutL and β and we characterized the structure of the MutL-β complex using small angle X-ray scattering. Similar to other β-binding partners, MutL interacts with a single protomer of the β dimer. The MutL-β complex modulates mismatch repair by stimulating the helicase activity of UvrD. Furthermore, the presence of the additional cysteine residue at the C-terminal end of β does not alter its functions *in vivo*, thereby providing a tool to study how the β-clamp interacts and modulates the activity of other weak-binding partners.

5.3 Introduction

Protein-protein and protein-nucleic acid interactions regulate virtually every cellular transaction. However, our understanding of these regulatory processes is biased by our ability to study these complexes. Most well characterized protein-protein complexes form stable, high-affinity interactions ($K_D < 10^{-6}$ M). Weak-affinity ($K_D > 10^{-4}$ M) and transient interactions are equally important in the regulation of many cellular pathways (Qin and Gronenborn, 2014), but they are poorly understood. Weak and transient interactions are especially difficult to study in proteins that coordinate multiple processes, as they interact with many binding partners often using a common interface. Allostery and conformational malleability are defining aspects of the hierarchy of these interactions (Nussinov et al., 2013; Qin and Gronenborn, 2014), however most structural biology techniques fail to provide this type of information.

The sliding β-clamp (β) and its eukaryotic counterpart (PCNA) are a paradigm for proteins coordinating multiple interactions using a common binding site. They were first identified as processivity factors that tether the replicative polymerase to DNA during DNA replication. However, they also play critical roles coordinating DNA replication with other key cellular functions including Okazaki fragment maturation, polymerase switching during lesion bypass, DNA repair and DNA transposition (Gomez et al., 2014; Lopez de Saro and O'Donnell, 2001; Moldovan et al., 2007; Parks et al., 2009). Sliding β-clamps form ring-shaped structures that are conserved at the structural, but not sequence, level. Their central cavity threads DNA and creates a topological link between their

binding partners and DNA (Georgescu et al., 2008; Stukenberg et al., 1991). They interact with their binding partners through a conserved hydrophobic groove located at the C-terminus of the protein (Figure 5.1). Reciprocally, clamp-binding partners contain a conserved linear motif known as PIP box (PCNA-interacting protein box) or β-binding motif that binds the hydrophobic groove of the clamp. β-binding motifs are not as well conserved as PIP boxes and, hence, β-clamp binding partners are difficult to identify (Dalrymple et al., 2001). The sequence variability of this binding motif has been correlated to binding affinity, thereby allowing clamps to mediate both weak and strong interactions (Bruning and Shamoo, 2004; Maga and Hubscher, 2003; Pillon et al., 2011; Rolef Ben-Shahar et al., 2009; Yin et al., 2013).

One of the repair pathways orchestrated by the sliding β-clamp is DNA mismatch repair, a conserved post-replicative repair pathway that corrects replication errors introduced by DNA polymerase. The initiation of the mismatch repair response in *Escherichia coli* (*E. coli*) requires the coordinated action of three proteins: MutS, MutL and MutH. MutS recognizes base mismatches and small insertion or deletion loops and, once bound to the mismatch, it recruits MutL in an ATP-dependent manner (Acharya et al., 2003; Junop et al., 2001). Together, MutS and MutL activate the MutH endonuclease that nicks the newly synthesized strand at a nearby hemi-methylated GATC site (Acharya et al., 2003; Giron-Monzon et al., 2004; Joseph et al., 2004). This nick acts as an entry point for the UvrD helicase and exonucleases (ExoI, ExoVII, ExoX, or RecJ) to remove the erroneous strand (Iyer et al., 2006). The replication machinery subsequently fills the

gap with the correct sequence and completes repair. Beyond, its role at the DNA synthesis step, the sliding β-clamp plays critical roles at earlier steps of the mismatch repair process. It interacts with MutS and recruits it to sites of damage (Lopez de Saro et al., 2006; Simmons et al., 2008). The β-clamp also interacts with MutL (Lopez de Saro et al., 2006; Pillon et al., 2011), however the role of this interaction in *E. coli* remains unclear.

DNA mismatch repair is evolutionary conserved, but MutH is only found in a subset of gamma-proteobacteria including *E. coli*. Organisms lacking a *mutH* gene encode MutL homologs harbouring endonuclease activity (Duppatla et al., 2009; Fukui et al., 2008; Kadyrov et al., 2006; Kadyrov et al., 2007; Mauris and Evans, 2009; Pillon et al., 2010). MutL consists of two structurally conserved domains connected by a flexible linker (Kunkel and Erie, 2005). The N-terminal domain supports DNA binding and ATPase activity, whereas the C-terminal domain (CTD) mediates protein dimerization and harbours a metal binding motif associated to the endonuclease activity and a conserved β-binding motif (Ban et al., 1999; Ban and Yang, 1998; Duppatla et al., 2009; Fukui et al., 2008; Guarné et al., 2004; Kadyrov et al., 2006; Kadyrov et al., 2007; Mauris and Evans, 2009; Pillon et al., 2010; Pillon et al., 2011). PCNA stimulates the endonuclease activity of human MutLα and, due to its loading orientation, determines the strand direction of the incision (Pluciennik et al., 2010). Accordingly, disruption of the β-binding motif of *Bacillus subtilis* MutL completely abrogates mismatch repair (Pillon et al., 2010).

Interestingly, disruption of the β-binding motif in *E. coli* MutL also causes a mutator phenotype (Pillon et al., 2011). This DNA mismatch repair defect is weak, but it suggests that the MutL-β complex has additional roles unrelated to the endonuclease activity of MutL. In order to study this role of the β-clamp, we have devised a mechanism to stabilize the *E. coli* MutL-β clamp complex. We have characterized this variant of the MutL-β clamp complex to reveal the structural organization of this complex and its potential role in *E. coli* DNA mismatch repair.

5.4 Experimental Procedures

5.4.1 Cloning MutL and β clamp cysteine variants

The expression plasmids encoding *E. coli* β-clamp and MutL were kind gifts from Dr. Michael O'Donnell and Dr. Wei Yang, respectively. Variants of both proteins, where surface exposed cysteines had been replaced by serines, were generated by overlap PCR. *E. coli* β^{Cys} (pAG 8769; residues 1-367) harboring mutations C260S, C333S, L366S, and C367 was subcloned into a pET15b vector lacking the sequence for the poly-histidine tag using the NdeI and BamHI restriction sites. *E. coli* $MutL^{Cys}$ (pAG 8814; residues 1-615) and its C-terminal domain CTD^{Cys} (pAG 8768; residues 431-615) including mutations C61S, C446S, and C588S were also subcloned in pET15b using NdeI and BamHI. In the β-binding variants of $MutL^{Cys}$ ($MutL^{Cys}$*; pAG 8824) and CTD^{Cys} (CTD^{Cys}*; pAG 8772) the conserved ^{482}QPLLI motif was replaced by ^{482}ASAAA. All mutations were verified by DNA sequencing (MOBIX, McMaster University).

5.4.2 Protein expression and purification

CTD^{Cys} (20.6 kDa) and β^{Cys} (40.6 kDa) were transformed in BL21 (DE3) and BL21 Star (DE3) (Invitrogen) cells, respectively, supplemented with a plasmid encoding for uncommon tRNAs. Cells were grown at 37ºC to an OD_{600} ~0.7 and protein production was induced with 1 mM IPTG. Cells were harvested by centrifugation after 3 hours at 37ºC. $MutL^{Cys}$ (70.1 kDa) was grown in BL21 (DE3) (Invitrogen) cells supplemented with a plasmid encoding for uncommon tRNAs. Protein expression was induced with 0.5 mM IPTG and the cells were incubated for 3 hours at 37ºC. β^{Cys} and all $MutL^{Cys}$ variants were purified as described previously (Ban and Yang, 1998; Guarné et al., 2004; Pillon et al., 2010; Pillon et al., 2011).

5.4.3 Analysis of the β^{Cys} function *in vivo*

The ability of β^{Cys} to support *E. coli* viability and MMR *in vivo* was measured using a plasmid shuffle assay (Babu and Sutton, 2014). Briefly, strain MS201 lacks a functional β-clamp gene (*dnaN*), due to a frameshift mutation ($dnaN^{-1FS}$). Viability of strain MS201 relies on the ampicillin resistant plasmid pAMP$dnaN^{+}$, which expresses physiological levels of the wild type β-clamp (Sutton et al., 2005). The plasmid shuffle assay was performed by transforming MS201 bearing pAMP$dnaN^{+}$ with chloramphenicol resistant pACM (negative control), pACM$dnaN^{+}$ (wild type β-clamp) or pACMβ^{Cys} (β^{Cys}). Twenty randomly selected transformants were passaged ~100 generations in LB medium (10 g/l Difco tryptone, 5 g/l Difco yeast extract, 10 g/l NaCl) supplemented with

40 μg/ml chloramphenicol. The frequency of pAMP*dnaN*$^{+}$ retention was measured by patching cells onto agar plates supplemented with 150 μg/ml ampicillin (to score for pAMP*dnaN*$^{+}$) or 40 μg/ml chloramphenicol (as a control). Representative clones bearing pACM*dnaN*$^{+}$ (VB100) or pACMβ^{Cys} (VB101) were analyzed further to confirm they lacked pAMP*dnaN*$^{+}$ and contained the *dnaN*$^{-1FS}$ allele using diagnostic PCR and automated nucleotide sequence analysis as described previously (Babu and Sutton, 2014). Spontaneous mutation frequencies of strains VB100 and VB101 (Sanders et al., 2011) and 95% confidence limits were measured as described (Dixon and Massey, 1969).

5.4.4 MutL^{Cys}-β^{Cys} complex formation

To form the MutL^{Cys}-β^{Cys} and CTD^{Cys}-β^{Cys} complexes, β^{Cys} was incubated with either MutL^{Cys} or CTD^{Cys} at a 1:1 ratio to a final concentration of 20 μM. The samples (1-2 mL) were dialyzed against 1 L of dialysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 10 mM DTT, 5% glycerol) over 2 hours at 4˚C. The mixture were transferred into dialysis buffer with 5 mM DTT for 1 hour followed by 1 hour in dialysis buffer without DTT. The sample was then left overnight in dialysis buffer (without DTT). Complex formation was monitored over time by resolving the samples on either 11% (MutL^{Cys}-β^{Cys}) or 15% (CTD^{Cys}-β^{Cys}) denaturing gels stained with Coomassie Brilliant Blue. Prior to forming the full length MutL^{Cys}–β^{Cys} complex, MutL^{Cys} (43 μM) was pre-incubated with 2 mM AMPPNP (Sigma) for 1 hour at room temperature followed by an overnight incubation at

4˚C. Association of the N-terminal domains of $MutL^{Cys}$ was monitored as previously described (Ban et al., 1999; Guarné et al., 2004).

5.4.5 SAXS data collection and analysis

Protein samples were resolved over a Superdex-200 (GE Healthcare) gel filtration column and sample homogeneity was confirmed by dynamic light scattering (Malvern Instruments). Samples (35 μL) were spun at 15,700 x g for 10 minutes and scattering data was collected on a Rigaku BioSAXS-1000 at 10˚C. Consecutive scans of 10, 30, 60 and/or 180 minutes were collected over a range of protein concentrations (β^{Cys}: 47-186 μM; CTD^{Cys}: 91-364 μM; CTD^{Cys}-β^{Cys} (Day1): 2.4-4.0 μM; CTD^{Cys}-β^{Cys} (Day 2): 31-123 μM). SAXSLab 3.0.0r1 (Rigaku) was used to generate scattering curves. Data quality was assessed by comparing scattering curves over a range of protein concentrations and exposure times using the ATSAS 2.5.1 program suite (Konarev et al., 2006). Radius of gyration and pair-distance distribution functions were determined using Primus and Gnom, respectively (Table 5.3) (Konarev et al., 2003). Twenty *ab initio* models were independently generated for each sample using DAMMIN, DAMMIF, or GASBOR (Franke and Svergun, 2009; Svergun, 1999; Svergun et al., 2001). Average filtered models for β^{Cys} and CTD^{Cys} were generated using DAMAVER (Volkov and Svergun, 2003). The CTD^{Cys}-β^{Cys} models were clustered based on the normalized spatial discrepancy using DAMCLUST (Petoukhov et al., 2012). The reported molecular weights were calculated based on the volume of correlation (Rambo and Tainer, 2013).

5.4.6 ATP Hydrolysis Assay

ATP hydrolysis activity was performed as previously described (Guarné et al., 2004) with minor modifications. $MutL^{Cys}$ and $MutL^{Cys}$-β^{Cys} (1 μM) were incubated with $MgCl_2$ (5 mM) and 3.33 μM α-^{32}P ATP (competed with 62.5‑2000 μM cold ATP) in reaction buffer (20 mM Tris pH 8.0, 90 mM KCl, 0.1 mg/ml BSA, 5% glycerol). Reactions (15 μl) were incubated at room temperature for two hours and resolved by thin-layer chromatography in 0.75 M KH_2PO_4. ADP accumulation was visualized using the Typhoon Trio+ (GE Healthcare) and quantified using ImageJ (http://rsbweb.nih.gov/ij/). All experiments were performed in triplicates and error bars represent standard deviation.

5.4.7 DNA Binding Assay

The DNA substrate (250 base pairs) was amplified from the pUC19 vector (Invitrogen) using primers 5'-d(GCTTAATCAGTGAGGCACCTATCTCAGCG) and ^{32}P 5′ end-labeled 5′-P-d(CGGCAACAATTAATAGACTGGATGGAGGCG). DNA (5 nM) was incubated with $MutL^{Cys}$ and $MutL^{Cys}$-β^{Cys} from 40-320 nM in reaction buffer (20 mM Tris pH 8.0, 90 mM KCl, 15% glycerol). Protein-DNA mixtures were incubated at 22˚C for 10 minutes followed by 30 minutes on ice. Samples (15 μl) were resolved on 4.5% tris-borate-EDTA gels and visualized using the Typhoon Trio+ (GE Healthcare). All experiments were performed in triplicate.

5.4.8 Helicase Assay

Helicase assays were performed as described previously (Guarné et al., 2004) with minor modifications. The UvrD expression vector (pWY 1365) was a kind gift from Dr. Wei Yang. A 250 base-pair substrate was generated as described above and nicked near the center of the substrate using Nb.BsrDI (New England Biolabs). The nicked DNA substrate (5 nM) was pre-incubated with increasing amounts of either $MutL^{Cys}$ or $MutL^{Cys}$-β^{Cys} (5-80 nM) for 20 minutes at 22˚C in reaction buffer (20 mM Tris pH 7.5, 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mg/ml BSA). Unwinding was initiated by addition of 5 nM UvrD. Reactions (10 μL) were incubated for 35 minutes at 37°C and resolved on 4.5% tris-borate-EDTA gels. Gels were visualized using the Typhoon Trio+ (GE Healthcare) and quantified using ImageJ (http://rsbweb.nih.gov/ij/). All experiments were performed in triplicates and error bars represent the standard deviation.

5.5 Results and Discussion

5.5.1 Cysteine variants of MutL and β are functional and stable in solution

We have previously shown that the conserved ^{482}QXLLXP motif found in the dimerization domain of *E. coli* MutL is a genuine clamp binding motif (Pillon et al., 2011). However, the complex between MutL and β is weak and, thus, difficult to study using structural biology techniques. To stabilize the complex, we exploited the presence of a naturally occurring cysteine in *E. coli* MutL (Cys480), located immediately upstream to the β-binding motif. This residue is not conserved in other MutL homologues,

indicating that it is dispensable for the interaction with β. The C-terminal residue of β (Leu366) is less than 10 Å away from the hydrophobic groove where β-binding proteins interact with the clamp (Figure 5.1), therefore we added a cysteine at the C-terminus (Cys367) to promote the formation of a disulfide bridge with Cys480 found in MutL.

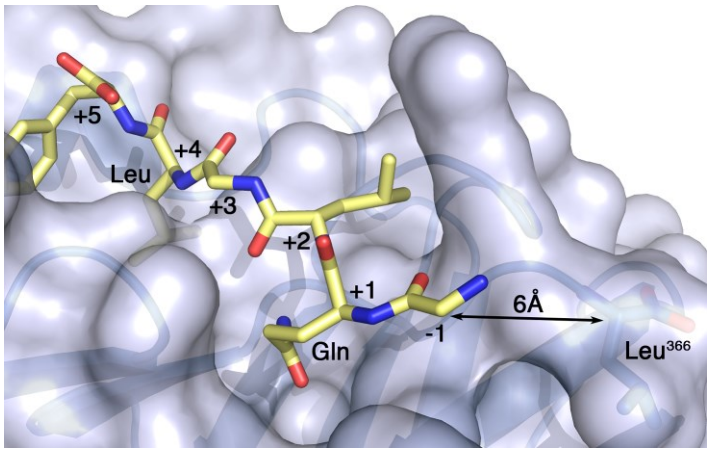


Figure 5.1. Crystal structure of the β-clamp bound to a polymerase II peptide. Surface representation of the β-clamp bound to the β-binding motif of polymerase II (color-coded sticks, PDB ID 3D1E). β has a well-defined groove near its C-terminus where partners interact. The conserved Gln and Leu residues at positions +1 and +4 of the β-binding motif are labeled. The C-terminal residue of β is also labeled and the distance to the N-terminal amino acid of the peptide is indicated.

To favour the formation of a specific complex, we used variants of MutL and the β-clamp where known surface-exposed and reactive cysteines had been replaced by serines to prevent unspecific binding of the two proteins. Kosinski *et al*. had previously shown that the single cysteine variant of full length *E. coli* MutL (L^{480C}) retains wild type DNA mismatch repair activity *in vivo* (Kosinski et al., 2005). We also wanted to ensure that β^{Cys} retained normal function *in vivo*. The β-clamp plays a critical role in DNA

replication, an essential cellular process. We used a β-clamp plasmid shuffle assay to investigate β^{Cys} function *in vivo*. Briefly, this assay measures the ability of a plasmid expressing physiological levels of wild type or mutant β-clamp to support viability of an *E. coli* strain harboring a frameshift in the endogenous β-clamp gene (Babu and Sutton, 2014). β^{Cys} supported *E. coli* viability as efficiently as did the wild type β clamp, indicating it was functional for replication (Table 5.1). To ensure β^{Cys} also supported DNA mismatch repair activity *in vivo*, we measured spontaneous mutation frequencies of isogenic *E. coli* strains expressing either wild type β or β^{Cys}. Both strains displayed similar mutation frequencies (Table 5.2), indicating β^{Cys} supports wild type mismatch repair function *in vivo*.

Table 5.1. Ability of β^{Cys} to support *E. coli* viability

Transforming plasmid*	β-clamp protein being assayed	Frequency of pAMP*dnaN*$^+$ retention&	Ability of assayed β-clamp to support *E. coli* viability#
pACM	None (negative control)	20/20 (100%)	–
pACM*dnaN*$^+$	Wild type β (positive control)	2/20 (10%)	+
pACMβ^{Cys}	β^{Cys}	2/20 (10%)	+

* Strain MS201 bearing plasmid pAMP*dnaN*$^+$ was transformed with the incompatible plasmids pACM, pACM*dnaN*$^+$ or pACMβ^{Cys}, as indicated.
& The frequencies of pAMP*dnaN*$^+$ retention in 20 randomly selected pACM, pACM*dnaN*$^+$ or pACMβ^{Cys} transformants following ~100 generations are indicated. Representative clones lacking pAMP*dnaN*$^+$ were characterized further to verify they expressed the appropriate *dnaN* allele.
# Viability refers to the ability of pACM, pACM*dnaN*$^+$ or pACMβ^{Cys} to support growth of *E. coli* in the absence of pAMP*dnaN*$^+$. Symbols are as follows: –, plasmid is unable to support *E. coli* viability, meaning 100% of the clones still contain pAMP*dnaN*$^+$ after ~100 generations of growth in the absence of its selection; +, plasmid is able to support *E. coli* viability.

Table 5.2. Ability of β^{Cys} to support mismatch repair function *in vivo*

Strain*	β-clamp protein	Spontaneous mutation frequency&
VB100	Wild type β	1.79×10^{-9} ($\leq 8.40 \times 10^{-10} - 4.76 \times 10^{-9}$)
VB101	β^{Cys}	1.68×10^{-9} ($8.40 \times 10^{-10} - 3.36 \times 10^{-9}$)

* Strains are derivatives of MS201 and express physiological levels of either wild type β clamp or β^{Cys} as the sole clamp protein.
& Mutation frequencies represent the median value from 14 (VB100) or 15 (VB101) independent cultures; 95% confidence intervals are in parentheses.

5.5.2 CTD^{Cys} forms a stable and specific complex with the β^{Cys}

The C-terminal domain of $MutL^{Cys}$ (CTD^{Cys}) was incubated with β^{Cys} in the absence of reducing agents to promote the formation of a disulfide bridge when the two proteins interact. A new product immediately appeared and it accumulated at the same rate as free CTD^{Cys} (20.6 kDa) and β^{Cys} (40.6 kDa) disappear. This new species (~60 kDa) had a molecular weight consistent with the presence of a monomer of each protein and it was sensitive to the presence of reducing agent, indicating that a disulfide bridge mediates the interaction (Figure 5.2A). To assess the stoichiometry of this new species, we resolved the sample using size exclusion chromatography and isolated a single peak with an apparent molecular weight of 166 kDa, consistent with the calculated molecular weight of the complex consisting of one β-clamp dimer and one CTD dimer (122.4 kDa) (Figure 5.2B).

To determine whether the CTD^{Cys}-β^{Cys} interaction was specific, we tried to form the CTD-β complex with a variant of CTD^{Cys} (CTD^{Cys}*) lacking the β-binding motif (Pillon et al., 2011). Incubation of CTD^{Cys}* with β^{Cys} resulted in the formation of two new species of molecular weights ~60 and ~80 kDa in a denaturing polyacrylamide gel (Figure 5.2A). Importantly, mutagenesis of the β-binding motif resulted in a significant delay in the accumulation of the 60 kDa species suggesting that formation of this complex depends on the integrity of the β-binding motif (Figure 5.2A). The higher molecular weight species (80 kDa) could be caused by the interaction of Cys480 in another CTD^{Cys}* molecule to a partially exposed cysteine (Cys180) present in the β^{Cys} variant. Since only the 60 kDa species is dependent on the presence of the β-binding motif of MutL, we decided to continue our analysis with this crosslinked product.

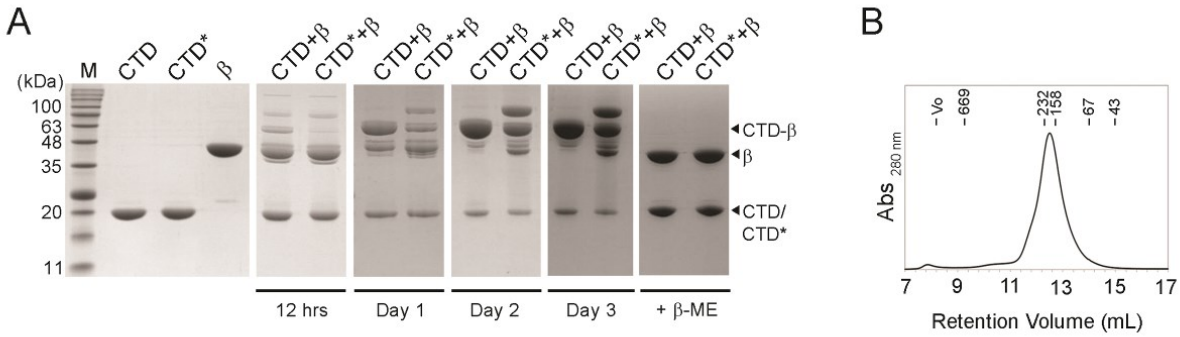


Figure 5.2. The C-terminal domain of $MutL^{Cys}$ and β^{Cys} form a specific complex in solution. **(A)** CTD^{Cys}, CTD^{Cys}* and β^{Cys} were purified (left panel) and equimolar mixtures of either CTD-β or CTD*-β were incubated in the absence of reducing agent. Samples withdrawn from the reaction at the indicated times were resolved on denaturing gels in the absence of β-mercaptoethanol (β-ME). The right panel is a control gel run in the presence of β-ME. **(B)** Elution profile of the CTD^{Cys}-β^{Cys} complex over a Superdex-200 size exclusion chromatography column. The void volume (V_0) and the molecular weight standards (kDa) are indicated.

5.5.3 CTD^{Cys} and β^{Cys} form two distinct complexes in solution

We were intrigued by the relatively slow formation of the CTD-β complex (more than one day after removing the reducing agent from the solution) and considered whether complex formation may be a multi-step process. Therefore, we decided to study this new species at two different time points, herein referred to as 'Day 1' and 'Day 2' (see the third and fourth panels in Figure 5.2A). To this end, we used small angle X-ray scattering (SAXS), a structural technique where data is collected quickly and in solution, thereby facilitating the analysis of samples at different time points.

First, we collected scattering data of the C-terminal domain of $MutL^{Cys}$ (CTD^{Cys}) and β^{Cys} at a range of concentrations and found that both samples were monodisperse and dimeric (Table 5.3). The scattering curves for CTD^{Cys} and β^{Cys} were also similar to the theoretical scattering profiles derived from the crystal structures of MutL-CTD (PDB ID 1X9Z) and β-clamp (PDB ID 1MMI). The discrepancy between the theoretical and experimental scattering curves resulted in chi values of 0.96 (MutL-CTD) and 2.22 (β-clamp), respectively (Figure 5.3A, compare black lines to the experimental scattering curves). Accordingly, the pair-distance distribution functions of the β^{Cys} and CTD^{Cys} were indicative of toroidal (β^{Cys}) and elongated (CTD^{Cys}) particles (Figure 5.3B) and the crystal structures of β and MutL-CTD could be readily superimposed onto the *ab initio* models generated from the solution scattering curves of β^{Cys} and CTD^{Cys} (Figure 5.3C and Movies S5.1 and S5.2).

Table 5.3. SAXS data-collection and scattering-derived parameters

	β^{Cys}	CTD^{Cys}	CTD^{Cys}-β^{Cys}	
			Day 1	**Day 2**
Data-collection parameters				
Exposure time (min)	60	10	180	180
Concentration (μM)	93	91	2	38.5
Structural parameters				
I_0 (cm^{-1}) [from Guinier]	1.738 ± 0.014	0.381 ± 0.010	0.077 ± 0.003	1.261 ± 0.004
R_g (Å) [from Guinier]	34.0 ± 0.5	31.1 ± 0.2	39.4 ± 2.1	35.1 ± 0.2
Fidelity of Guinier	0.874	0.830	0.956	0.882
I_0 (cm^{-1}) [from P(r)]	1.720 ± 0.004	0.375 ± 0.000	0.073 ± 0.001	1.274 ± 0.003
R_g (Å) [from P(r)]	32.9 ± 0.1	31.9 ± 0.0	36.3 ± 0.4	35.3 ± 0.1
D_{max} (Å)	90	102	104	101
Experimental MW [from Q_R]*	85,855 Da	50,052 Da	152,490 Da	143,778 Da
Calculated MW	81,263 Da	41,108 Da	122,371 Da	122,371 Da
Ab initio analysis	*DAMMIN*	*DAMMIF*	*GASBOR*	*GASBOR*
Chi^2 of *Ab Initio* model	1.087-1.089	0.981	1.00	1.23-1.28

* MW determined using ScÅtter (Rambo and Tainer, 2013)

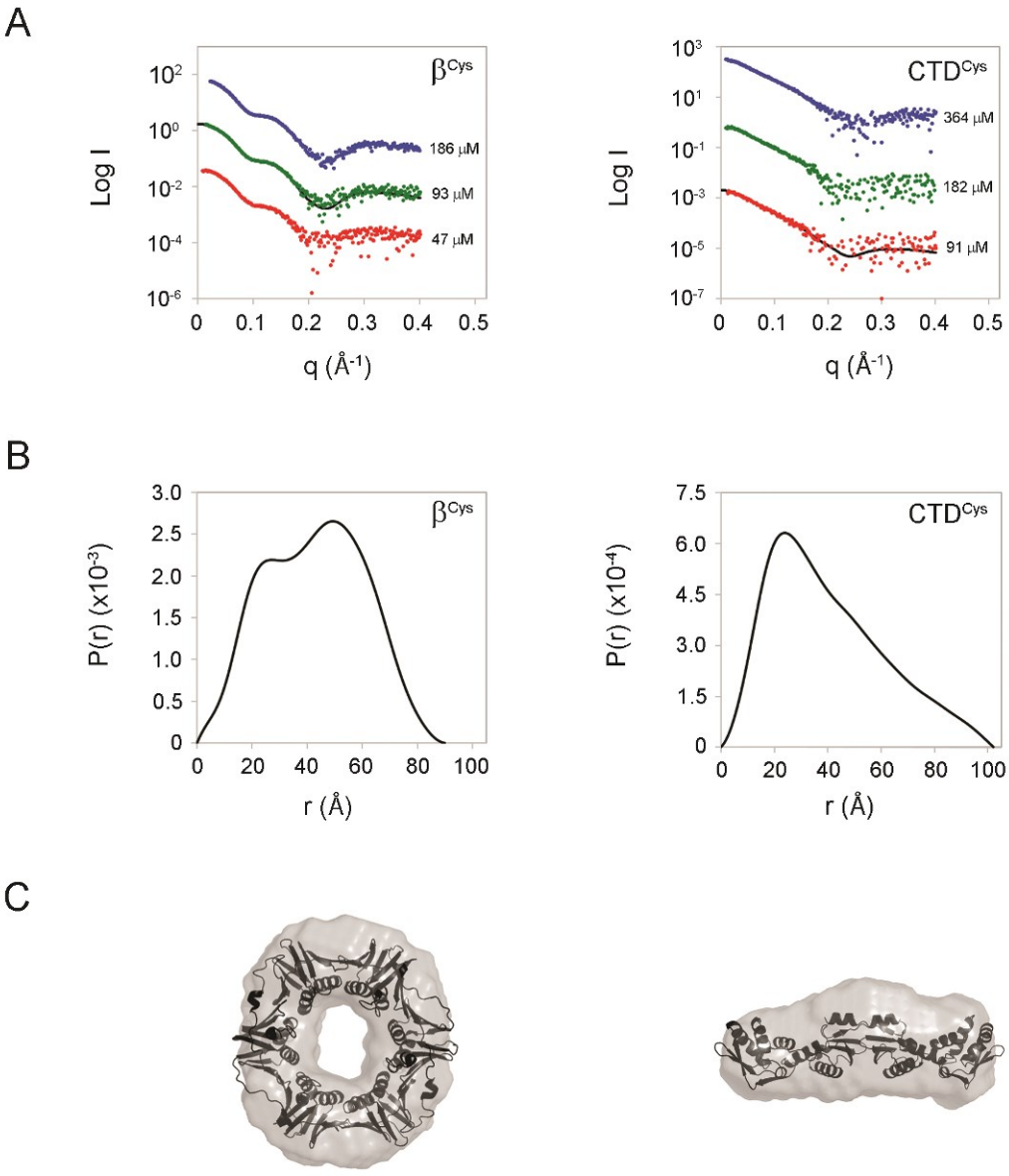


Figure 5.3. Solution structures of β^{Cys} and CTD^{Cys}. (A) Experimental scattering profiles of β^{Cys} (left panel) and the C-terminal domain of $MutL^{Cys}$ (right panel). Scattering data for β^{Cys} was collected at 47 (red), 93 (green), and 186 (blue) μM, whereas data for CTD^{Cys} was collected at 91 (red), 182 (green), and 364 (blue) μM. The theoretical scattering profiles derived from the crystal structures of β (PDB ID 1MMI) and MutL-CTD (PDB ID 1X9Z) were generated using Crysol (Svergun et al., 1995) and are shown as solid black lines. (**B**) Pair-distance distribution functions for β^{Cys} at 93 μM (left panel) and CTD^{Cys} at 91 μM (right panel). (**C**) Surface rendering of the averaged and filtered model generated from ten independent *ab initio* models. The ribbon diagrams of the crystal structures of β and MutL-CTD are superimposed.

We then incubated CTD^{Cys} and β^{Cys} for either one ('Day 1') or two ('Day 2') days, resolved the mixtures by size exclusion chromatography and collected scattering data for both samples (Figure 5.4A-B, top panels). The samples showed no signs of protein aggregation and were properly folded. They had similar pair-distance distribution functions and their estimated molecular weights were consistent with the calculated molecular weight of the CTD-β complex at a 1:1 ratio (Table 5.3 and Figure 5.4A-B, bottom panels). We generated twenty independent *ab initio* models for each sample and clustered them based on their normalized spatial discrepancy using DAMCLUST (Petoukhov et al., 2012). We identified four different clusters for the sample at 'Day 1' and five different clusters for the sample at 'Day 2' (Figure S5.1). However, the number of models assigned to each cluster was markedly different for each sample (Figure 5.4C and Figure S5.1).

Cluster A is the most populated cluster for the sample at 'Day 1' (Figure 5.4C, white bars). The representative model for this cluster forms a ring-shaped structure with a handle that resembles a perforated curling stone (Figure 5.4D and Movie S5.3). The toroidal moiety of the model is closely related to the SAXS model of β^{Cys} and the 'handle' has the same shape and dimensions as the SAXS model of CTD^{Cys} (Figure 5.4D and 5.3C). The β-binding motifs are found at both ends of the MutL-CTD dimer, therefore cluster A represents a conformation of the CTD-β complex where only one protomer of the CTD dimer is bound to the β ring. Conversely, cluster E was the most populated cluster for the sample at 'Day 2' (Figure 5.4C, black bars). The β^{Cys} moiety of

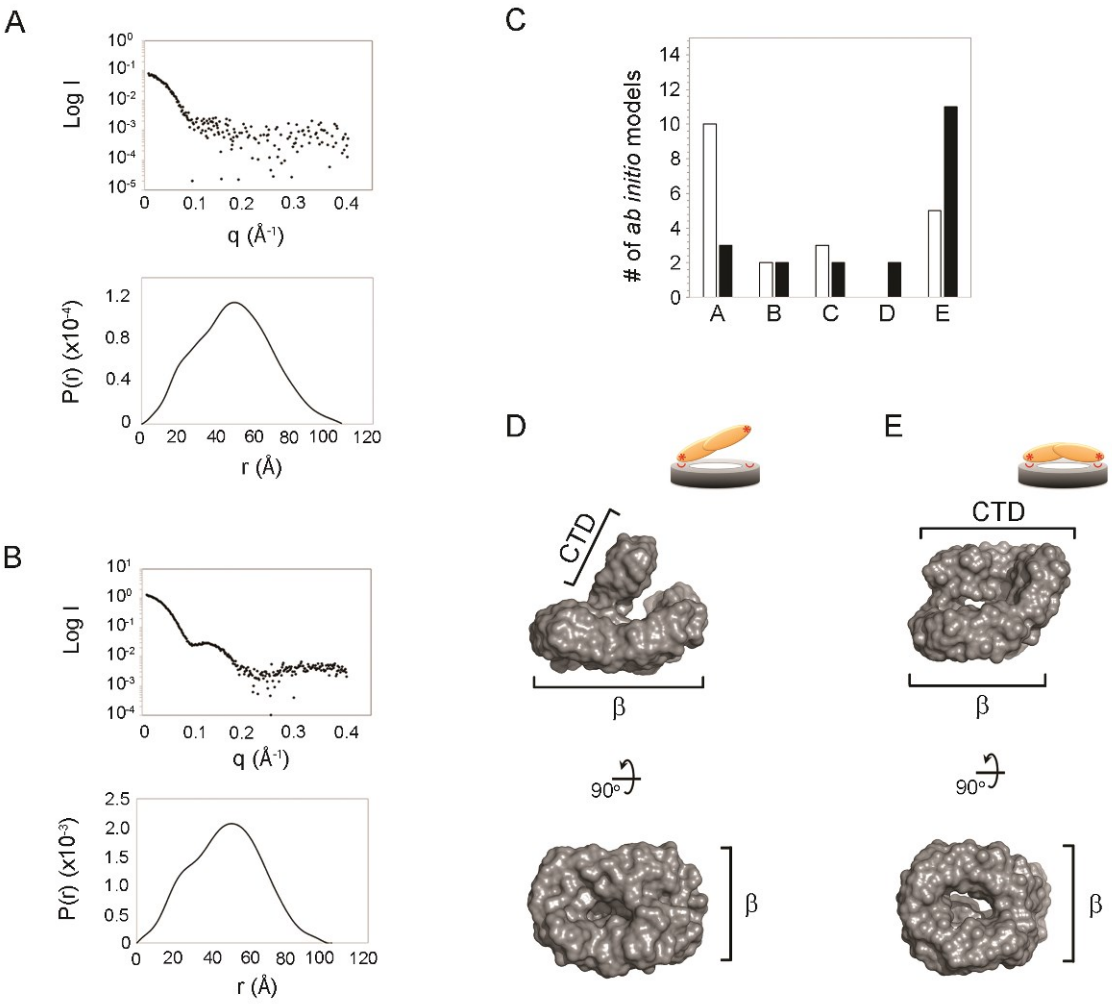


Figure 5.4. Solution structures of the L^{Cys}-β^{Cys} complexes. Experimental scattering profiles (top) and pair-distance distribution functions (bottom) for the L^{Cys}-β^{Cys} samples at 'Day 1' **(A)** and 'Day 2' **(B)**. **(C)** Clustering of twenty independently generated *ab initio* models for 'Day 1' (white) and 'Day 2' (black). Orthogonal views of the representative model for the predominant clusters at 'Day 1' **(D)** and at 'Day 2' **(E)** are shown as surface renderings with the CTD and β moieties of the models labeled.

the complex is also clearly visible on this cluster, but the 'handle' has collapsed on top the β^{Cys} ring, indicating that the two β-binding motifs of the CTD^{Cys} dimer are bound to the β^{Cys} (Figure 5.4D and Movie S5.3). The lack of free CTD^{Cys} and β^{Cys} after two days of

incubation (Figure 5.2) supports the idea that both subunits of CTD^{Cys} are bound to the β^{Cys} ring.

The relative cluster populations at 'Day 1' and 'Day 2' indicates a time-dependent shift from one to two protomers of the CTD dimer bound to β (Figure 5.4C). Cluster B resembles cluster A, whereas clusters C and D appear to be distorted conformations of cluster E (Figure S5.1). However, we cannot rule out the possibility that these species are the result of unspecific binding of Cys480 (MutL) to partially exposed cysteines still present in β^{Cys} (namely Cys180).

Binding partners of the β sliding clamp typically bind a single cleft on the ring (Bubeck et al., 2011; Heltzel et al., 2009; Sakurai et al., 2005; Sutton et al., 2005). The complex represented by cluster A may be the functional form of the L-β complex, but cluster E occurs because two reactive cysteines are exposed at the other end of the complex. Alternatively, cluster E may represent the functional form of the L-β complex and cluster A is merely an intermediate step necessary to align the second protomer of MutL to the second binding cleft of β. One limitation of the SAXS experiments is that they were done in the absence of any additional factors (MutS, DNA, MutH, UvrD, etc) normally present during the repair of a mismatch. Moreover, these experiments were done with only the C-terminal domain of MutL. MutL undergoes a large conformational change upon ATP binding (Ban et al., 1999; Fukui et al., 2008; Guarné et al., 2004; Sacho et al., 2008; Tran and Liskay, 2000). Therefore, it is possible that the presence of the N-

terminal domains of MutL or additional mismatch repair factors favour the formation of only one of the complexes with the β-clamp.

5.5.4 MutL^{Cys} binds to a single cleft on β^{Cys}

We generated a variant of full length MutL (L^{Cys}) where known exposed cysteine residues, except for Cys480, were replaced by serines. A similar variant (L^{C480}) is functional in a mismatch repair assay (Kosinski et al., 2005), and L^{Cys} undergoes the characteristic nucleotide-dependent conformational change when purified by size exclusion chromatography in the presence/absence of a non-hydrolyzable analog of ATP (Figure 5.5A). To test if L^{Cys} favours the singly- or doubly-bound form of the L-β complex, we incubated L^{Cys} with equimolar amounts of β^{Cys} in the absence or presence of AMPPNP and monitored complex formation (Figure 5.5B). In good agreement with previous experiments using the C-terminal domain of MutL, a new species consistent with the formation of a 1:1 complex readily appeared and accumulated over time (Figure 5.5B). However, this new species stopped accumulating after 'Day 1' and free L^{Cys} and β^{Cys} did not completely disappear (Figure 5.5B), thereby suggesting that MutL only binds to one of the binding clefts present on the β-clamp ring.

This trend was even more significant when the complex was assembled with a variant of MutL (L^{Cys}*) where the β-binding motif (^{482}QPLLI) had been replaced by ^{482}ASAAA (Figure 5.5C). Incubation of L^{Cys}* with β^{Cys} resulted in the formation of two new species of molecular weights ~130 and ~180 kDa in a denaturing polyacrylamide gel

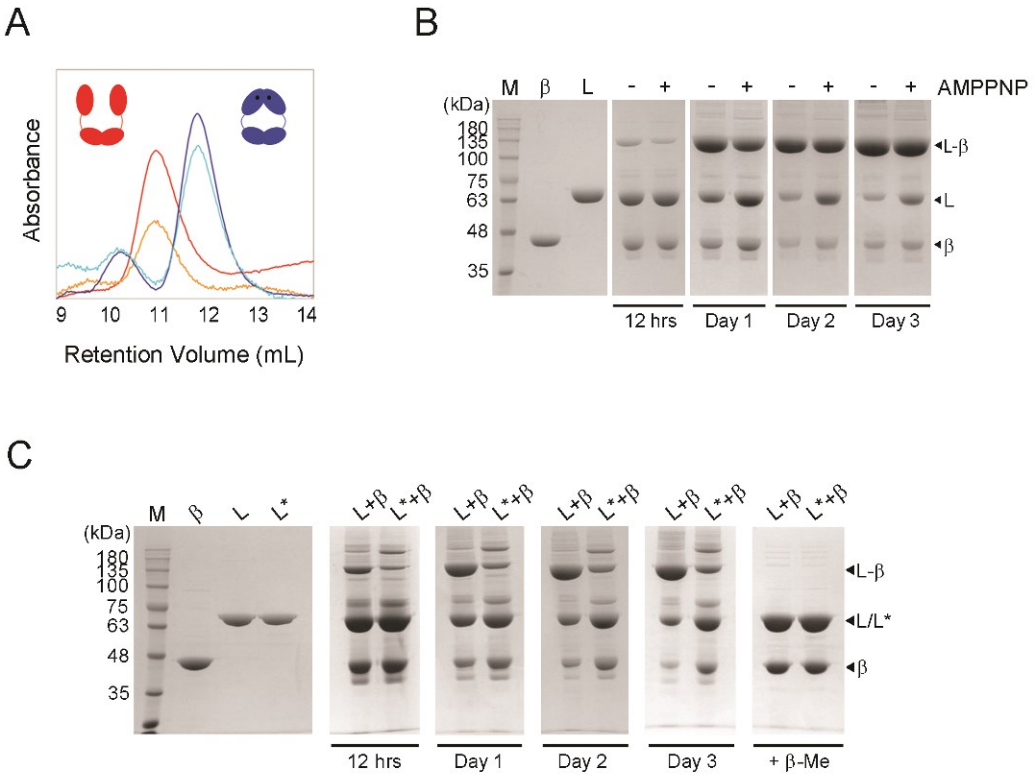


Figure 5.5. MutL^{Cys} binds to a single cleft of β^{Cys}. **(A)** Size exclusion chromatography profile of MutL^{Cys} (14.3 μM) in the absence (red/orange) or presence (blue/cyan) of AMPPNP. The conformational change of MutL^{Cys} and nucleotide binding are monitored by comparing the peak absorbance at 280 nm (red/blue) and 260 nm (orange/cyan). **(B)** Purified MutL^{Cys} (labeled L) was mixed with equimolar amounts of β^{Cys} (labeled β) and incubated in the absence/presence of AMPPNP. Samples withdrawn from the reaction at the indicated times were resolved on denaturing gels in the absence of β-mercaptoethanol (β-ME). After three days of incubation a control gel was run in the presence of β-ME and it is shown on the right panel. **(C)** L^{Cys} (L), a variant of L^{Cys} missing the β-binding motif (L*) and β^{Cys} were purified (left panel) and equimolar mixtures of L-β and L*-β were incubated in the absence of reducing agents. Samples withdrawn from the reaction were resolved on denaturing gels as described in **(B)**.

(Figure 5.5C). These two products were sensitive to the presence of reducing agent, indicating that disulfide bridges mediate these interactions (Figure 5.5C). The lower molecular weight species is consistent with the presence of one monomer of MutL and

one monomer of β and it can presumably form because the exposed Cys480 (L) and Cys367 (β) can partially react even in the absence of the β-binding motif. The higher molecular weight species is consistent with the presence of two monomers of MutL and one monomer of β (180.7 kDa). The β^{Cys} variant has a partially exposed cysteine (Cys180) that could mediate the interaction with the second MutL monomer. It is worth noting that Cys180 (β) does not mediate the interaction with a second monomer of MutL unless the β-binding motif has been disrupted, suggesting that the conformation of the 130 kDa species formed with L^{Cys}* is not identical to that formed with L^{Cys} (Figure 5.5C).

5.5.5 Functional implications of the MutL interaction with the β-clamp

We have previously shown that disruption of the β-binding motif found in MutL causes a severe mismatch repair defect in organisms lacking MutH, but only a moderate defect in those encoding a *mutH* gene (Pillon et al., 2010; Pillon et al., 2011). Mismatch recognition by MutS and the MutS-dependent activation of MutL are common steps in MutH-dependent and MutH-independent mismatch repair pathways, suggesting the β-clamp affects a function of MutL at or following nicking of the newly synthesized strand. Indeed, the human counterpart of the β-clamp, PCNA, stimulates the endonuclease activity of human MutLα (Kadyrov et al., 2006; Pluciennik et al., 2010). Since MutH, rather than MutL, is responsible for nicking the newly synthesized strand in *E. coli* mismatch repair, we wanted to explore whether β alters a function of MutL following DNA nicking.

MutL has a role in repetitively loading UvrD onto DNA to facilitate unwinding of the nascent strand towards the mismatch (Matson and Robertson, 2006) and this function has been linked to the C-terminal domain of MutL (Guarné et al., 2004). Therefore, we tested whether MutL stimulated the helicase activity of UvrD differently when bound to β^{Cys}. In good agreement with previously published data (Guarné et al., 2004), we found that $MutL^{Cys}$ stimulated the unwinding activity of UvrD when added in equimolar amounts to the reaction (Figure 5.6A). The stimulatory effect was exacerbated when MutL was added in excess to the reaction (note that MutL concentration ranges from 5-80 nM which corresponds to 1-16 molar excess with respect to UvrD-DNA). We observed a small, yet significant, increase in UvrD unwinding when $MutL^{Cys}$-β^{Cys} complex replaced $MutL^{Cys}$ (Figure 5.6A). We did not see, however, a similar helicase stimulation when using shorter nicked DNA duplexes (data not shown). Both L^{Cys} and L^{Cys}-β^{Cys} had similar ATPase activity (Figure 5.6B) and bound comparably to a 250 base-pair duplex DNA (Figure 5.6C), indicating that binding of MutL to β is the actual cause of the enhanced stimulation of the helicase activity of UvrD. These results suggest that MutL binding to the β-clamp may only be significant when UvrD has to unwind long stretches of DNA before reaching the mismatch. This, in turn, would explain why disruption of the β-binding motif in *E. coli* MutL only causes a weak mutator phenotype.

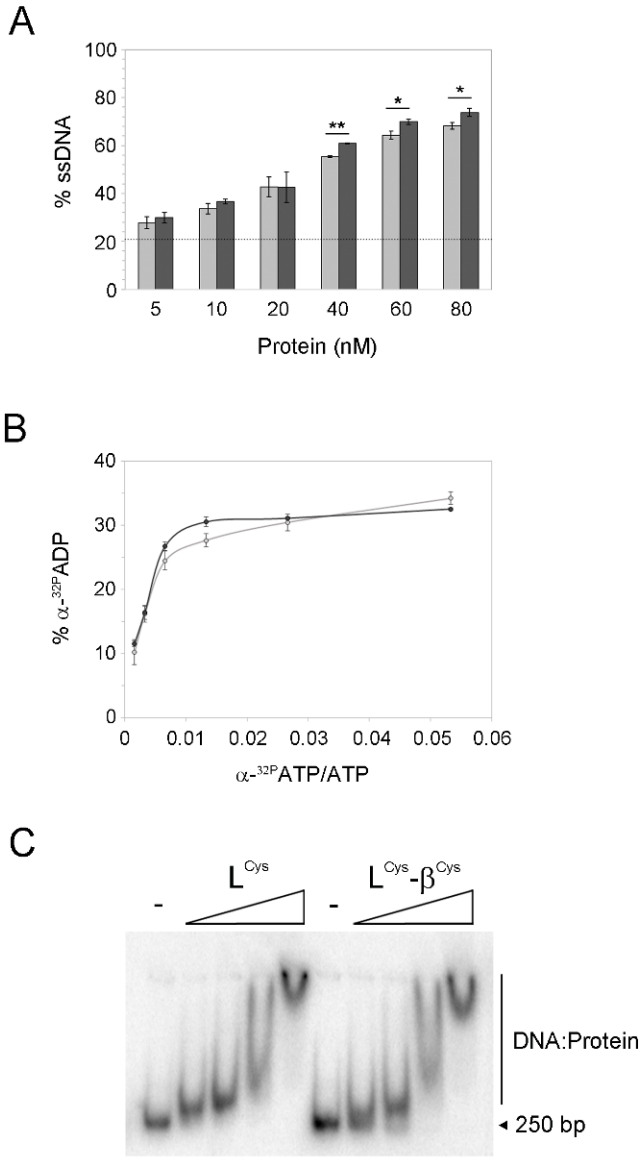


Figure 5.6. The L-β complex stimulates the helicase activity of UvrD. (A) UvrD helicase activity in the presence of L^{Cys} (light grey) and L^{Cys}-β^{Cys} (dark grey) on a 250 base-pair nicked DNA (see methods). The dotted line denotes the unwinding activity of UvrD in the absence of L or L-β. Error bars represent the standard deviation of three independent experiments. The p-value was calculated using a two-sample unequal variance t-test. *, $p < 0.013$ and **, $p < 1.4 \times 10^{-4}$. **(B)** ATP hydrolysis activity of L^{Cys} (grey line) and L^{Cys}-β^{Cys} (black line). Error bars represent the standard deviation of three independent experiments. **(C)** DNA binding of L^{Cys} and L^{Cys}-β^{Cys} measured using a 250 base-pair DNA.

5.6 Conclusions

It is well documented how β binds and tethers its binding partners to DNA (Burnouf et al., 2004; Chapados et al., 2004; Jergic et al., 2013; Moldovan et al., 2007; Sutton, 2010; Xing et al., 2009). However, the affinities for different β-containing complexes vary by several orders of magnitude. In this work, we engineered a β-clamp variant that covalently binds to one of its weak interacting partners, MutL. This variant allowed us to study the structural organization of the MutL-β complex and unveiled one of the minor, albeit important, roles of the β-clamp in DNA mismatch repair. Given that most binding partners interact with the sliding β-clamp using the same molecular determinants, this approach can be easily translated to study the roles of the sliding β-clamp in other cellular processes.

5.7 Supplementary Figure

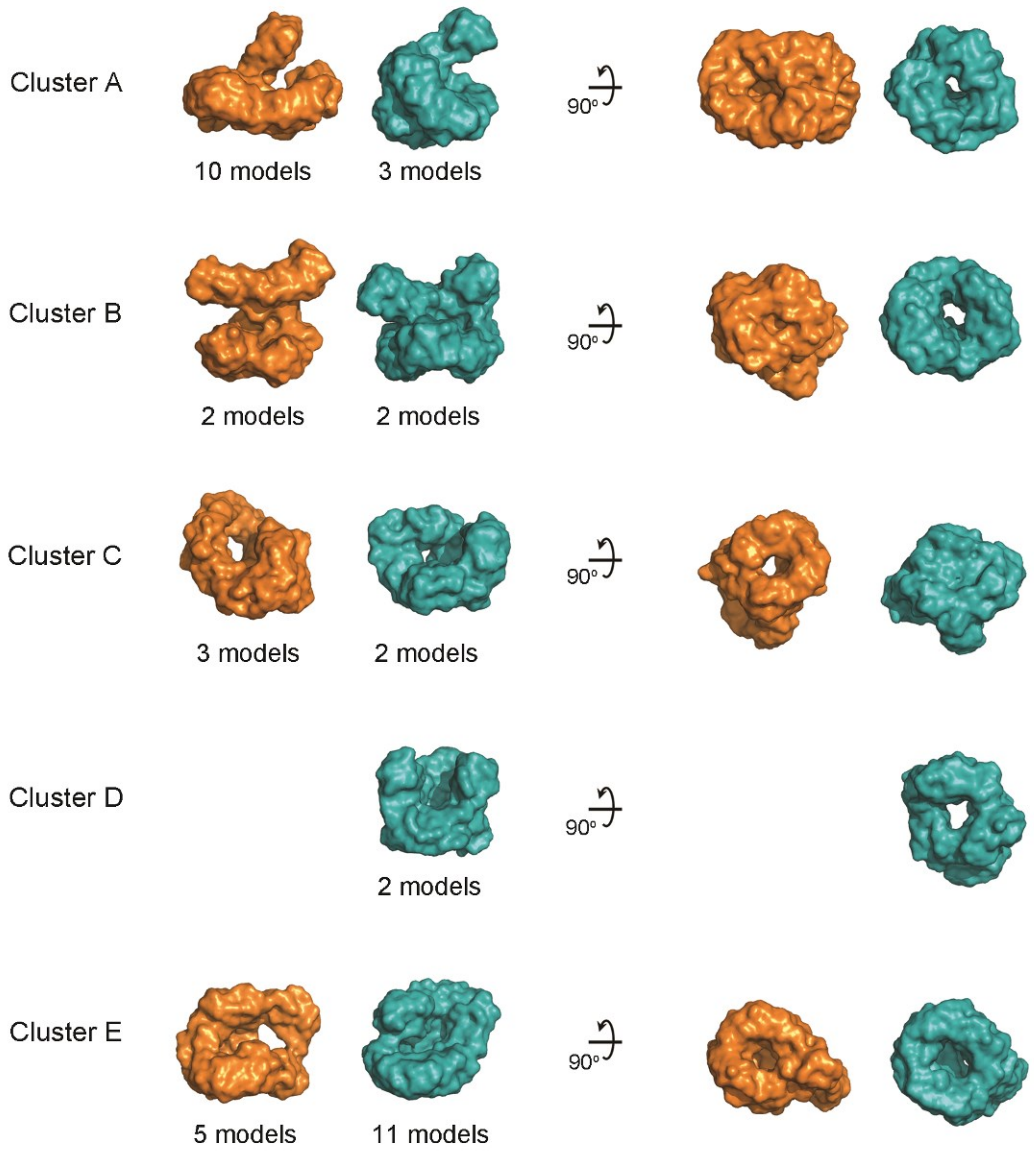


Figure S5.1. Clusters of the CTD^{Cys}-β^{Cys} SAXS *ab initio* models. The representative model for each cluster identified by DAMCLUST is shown as orange ('Day 1') or blue ('Day 2') surface renderings. The number of models in each cluster is indicated. The two major clusters (A for 'Day 1' and E for 'Day 2') represent the singly and doubly bound conformations of the CTD-β complex.

5.8 Acknowledgements

We are grateful to members of the Guarné laboratory and Dr. M. Junop for helpful discussions. We also thank Dr. M. O'Donnell for providing the sliding β-clamp expression vector and Dr. W. Yang for providing the MutL and UvrD expression vectors and helpful discussions.

5.9 References

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

Babu, V.M.P., and Sutton, M.D. (2014). A dnaN Plasmid Shuffle Strain for Rapid In Vivo Analysis of Mutant Escherichia coli beta Clamps Provides Insight Into the Role of Clamp in umuDC-Mediated Cold Sensitivity. PloS one *9*, e98791.

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

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

Bruning, J.B., and Shamoo, Y. (2004). Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. Structure (London, England : 1993) *12*, 2209-2219.

Bubeck, D., Reijns, M.A., Graham, S.C., Astell, K.R., Jones, E.Y., and Jackson, A.P. (2011). PCNA directs type 2 RNase H activity on DNA replication and repair substrates. Nucleic acids research *39*, 3652-3666.

Burnouf, D.Y., Olieric, V., Wagner, J., Fujii, S., Reinbolt, J., Fuchs, R.P., and Dumas, P. (2004). Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases. Journal of molecular biology *335*, 1187-1197.

Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J.A. (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell *116*, 39-50.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proceedings of the National Academy of Sciences of the United States of America *98*, 11627-11632.

Dixon, W.J., and Massey, F.J. (1969). Introduction to Statistical Analysis (New York: McGraw-Hill).

Duppatla, V., Bodda, C., Urbanke, C., Friedhoff, P., and Rao, D.N. (2009). The carboxy-terminal domain is sufficient for endonuclease activity of Neisseria gonorrhoeae MutL. Biochem J.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Cryst *42*, 342-346.

Fukui, K., Nishida, M., Nakagawa, N., Masui, R., and Kuramitsu, S. (2008). Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. The Journal of biological chemistry *283*, 12136-12145.

Georgescu, R.E., Kim, S.S., Yurieva, O., Kuriyan, J., Kong, X.P., and O'Donnell, M. (2008). Structure of a sliding clamp on DNA. Cell *132*, 43-54.

Giron-Monzon, L., Manelyte, L., Ahrends, R., Kirsch, D., Spengler, B., and Friedhoff, P. (2004). Mapping protein-protein interactions between MutL and MutH by cross-linking. The Journal of biological chemistry *279*, 49338-49345.

Gomez, M.J., Diaz-Maldonado, H., Gonzalez-Tortuero, E., and Lopez de Saro, F.J. (2014). Chromosomal replication dynamics and interaction with the beta sliding clamp determine orientation of bacterial transposable elements. Genome biology and evolution *6*, 727-740.

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

Heltzel, J.M., Maul, R.W., Scouten Ponticelli, S.K., and Sutton, M.D. (2009). A model for DNA polymerase switching involving a single cleft and the rim of the sliding clamp. Proceedings of the National Academy of Sciences of the United States of America *106*, 12664-12669.

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

Jergic, S., Horan, N.P., Elshenawy, M.M., Mason, C.E., Urathamakul, T., Ozawa, K., Robinson, A., Goudsmits, J.M., Wang, Y., Pan, X.*, et al.* (2013). A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. The EMBO journal *32*, 1322-1333.

Joseph, N., Sawarkar, R., and Rao, D.N. (2004). DNA mismatch correction in Haemophilus influenzae: characterization of MutL, MutH and their interaction. DNA repair *3*, 1561-1577.

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

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

Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., and Modrich, P. (2007). Saccharomyces cerevisiae MutLa is a mismatch repair endonuclease. The Journal of biological chemistry.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. J Appl Cryst *39*, 277-286.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Cryst *36*, 1277-1282.

Kosinski, J., Steindorf, I., Bujnicki, J.M., Giron-Monzon, L., and Friedhoff, P. (2005). Analysis of the quaternary structure of the MutL C-terminal domain. Journal of molecular biology *351*, 895-909.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annual review of biochemistry *74*, 681-710.

Lopez de Saro, F.J., Marinus, M.G., Modrich, P., and O'Donnell, M. (2006). The beta sliding clamp binds to multiple sites within MutL and MutS. The Journal of biological chemistry *281*, 14340-14349.

Lopez de Saro, F.J., and O'Donnell, M. (2001). Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I. Proceedings of the National Academy of Sciences of the United States of America *98*, 8376-8380.

Maga, G., and Hubscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. Journal of cell science *116*, 3051-3060.

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

Mauris, J., and Evans, T.C. (2009). Adenosine triphosphate stimulates Aquifex aeolicus MutL endonuclease activity. PloS one *4*, e7175.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665-679.

Nussinov, R., Tsai, C.J., and Ma, B. (2013). The underappreciated role of allostery in the cellular network. Annual review of biophysics *42*, 169-189.

Parks, A.R., Li, Z., Shi, Q., Owens, R.M., Jin, M.M., and Peters, J.E. (2009). Transposition into replicating DNA occurs through interaction with the processivity factor. Cell *138*, 685-695.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Cryst *45*, 9.

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

Pillon, M.C., Miller, J.H., and Guarne, A. (2011). The endonuclease domain of MutL interacts with the beta sliding clamp. DNA repair *10*, 87-93.

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

Qin, J., and Gronenborn, A.M. (2014). Weak protein complexes: challenges to study but essential for life. FEBS Journal *281*, 1948-1949.

Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477-481.

Rolef Ben-Shahar, T., Castillo, A.G., Osborne, M.J., Borden, K.L., Kornblatt, J., and Verreault, A. (2009). Two fundamentally distinct PCNA interaction peptides contribute to chromatin assembly factor 1 function. Molecular and cellular biology *29*, 6353-6365.

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

Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., and Hakoshima, T. (2005). Structural basis for recruitment of human flap endonuclease 1 to PCNA. The EMBO journal *24*, 683-693.

Sanders, L.H., Devadoss, B., Raja, G.V., O'Connor, J., Su, S., Wozniak, D.J., Hassett, D.J., Berdis, A.J., and Sutton, M.D. (2011). Epistatic roles for Pseudomonas aeruginosa MutS and DinB (DNA Pol IV) in coping with reactive oxygen species-induced DNA damage. PloS one *6*, e18824.

Simmons, L.A., Davies, B.W., Grossman, A.D., and Walker, G.C. (2008). Beta clamp directs localization of mismatch repair in Bacillus subtilis. Molecular cell *29*, 291-301.

Stukenberg, P.T., Studwell-Vaughan, P.S., and O'Donnell, M. (1991). Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. The Journal of biological chemistry *266*, 11328-11334.

Sutton, M.D. (2010). Coordinating DNA polymerase traffic during high and low fidelity synthesis. Biochimica et biophysica acta *1804*, 1167-1179.

Sutton, M.D., Duzen, J.M., and Maul, R.W. (2005). Mutant forms of the Escherichia colibeta sliding clamp that distinguish between its roles in replication and DNA polymerase V-dependent translesion DNA synthesis. Molecular microbiology *55*, 1751-1766.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal *76*, 2879-2886.

Svergun, D.I., Barberato, C., and Koch, M.H. (1995). CRYSOL-a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates J Appl Cryst *28*, 6.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal *80*, 2946-2953.

Tran, P.T., and Liskay, R.M. (2000). Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha. Molecular and cellular biology *20*, 6390-6398.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. J Appl Cryst *36*, 860-864.

Xing, G., Kirouac, K., Shin, Y.J., Bell, S.D., and Ling, H. (2009). Structural insight into recruitment of translesion DNA polymerase Dpo4 to sliding clamp PCNA. Molecular microbiology *71*, 678-691.

Yin, Z., Kelso, M.J., Beck, J.L., and Oakley, A.J. (2013). Structural and thermodynamic dissection of linear motif recognition by the E. coli sliding clamp. Journal of medicinal chemistry *56*, 8665-8673.

Chapter 6

Conclusions

This thesis focusses on the regulatory requirements for MutL activation. Two years prior to beginning my graduate work, Dr. Modrich's group made the groundbreaking discovery that human MutLα is a metal-dependent endonuclease (Kadyrov et al., 2006). This work revolutionized our understanding of eukaryotic MMR and precipitated the model for nick-directed MMR. This knowledge gave rise to new and pressing questions such as "What is the structural organization of the MutL endonuclease?" and "What are the regulatory mechanisms controlling MutL nicking activity?" The work described in this dissertation addresses these questions and contributes to the growing understanding for the coordinated events initiating MMR.

How MutS communicates the presence of a mismatch to MutL has been an outstanding question in the MMR field. MutS is a dynamic protein and has proven to be a bottleneck for uncovering the structural organization of the MutS-MutL complex. Work from Dr. Kolodner's group identified tandem glutamine residues in *E. coli* MutS that are necessary for MutL binding (Mendillo et al., 2009). However, these tandem

glutamines are not conserved in other MutS homologs. Our biochemical characterization of the *B. subtilis* MutS-MutL complex (chapter 2) builds onto Dr. Kolodner's work by describing the MutS-MutL interaction in the context of nick-directed MMR. We identified a distinct site in *B. subtilis* MutS that is necessary for MutL binding. This body of work also includes *in vivo* work that supports the 'moving' MutS sliding clamp model and suggests that MutS repetitively loads at the mismatch until MutL recruitment (Lenhart et al., 2013).

Following mismatch detection, a series of events activate MutL endonuclease activity. In response to the discovery that human MutLα is an endonuclease, there was a rush to characterize other MutL homologs from *mutH*-less organisms. To date, endonuclease activity has been detected in MutL homologs from eukaryotes (human and *S. cerevisiae)*, Gram-positive bacteria (*B. subtilis)*, and Gram-negative bacteria (*N. gonorrhoeae*, *A. aeolicus*, *T. thermophilus, and P. aeruginosa)* (Correa et al., 2013; Duppatla et al., 2009; Fukui et al., 2008; Kadyrov et al., 2006; Kadyrov et al., 2007; Mauris and Evans, 2009; Pillon et al., 2010). Despite the growing list of biochemically characterized MutL homologs, the structural organization of the MutL endonuclease remained elusive for several years.

Our atomic structures of *B. subtilis* MutL in its apo-form and zinc-bound state were the first structures to reveal the spatial organization of the endonuclease active site (chapter 3). Three previously identified conserved MutL motifs ($DQHA(X)_2E(X)_4E$,

ACR, and CPHGRP) cluster together to form the MutL active site (Kadyrov et al., 2006; Kosinski et al., 2008; Pillon et al., 2010). In the metal-bound structure of *B. subtilis* MutL, two zinc metals are coordinated through essential residues from these motifs (Pillon et al., 2010). The atomic structure also revealed the MutL Zn^{2+}A binding site resembles a novel metal binding site first identified in the metalloregulator, ScaR (Kosinski et al., 2008; Stoll et al., 2009). Zinc is proposed to act as a regulatory metal in ScaR and suggests Zn^{2+}A may play a similar role in MutL (Guarné et al., 2004). Indeed, zinc stimulates the manganese-dependent MutL endonuclease (Pillon et al., 2010). The apo-form of *N. gonorrhoeae* MutL (Namadurai et al., 2010) and the zinc-bound state of *S. cerevisiae* MutLα (Gueneau et al., 2013) followed and corroborated our work describing *B. subtilis* MutL.

Dr. Friedhoff's group conducted an extensive sequence analysis to identify conserved motifs in the MutL C-terminal domain. Along with the conserved motifs that form the endonuclease active site, the QXLLXP motif is also conserved in both methyl-directed and nick-directed MutL homologs (Kosinski et al., 2008). QXLLXP is surface exposed in all the structures of the MutL C-terminal domain (*E. coli*, *B. subtilis*, *N. gonorrhoeae*, and *S. cerevisiae*) and loosely resembles the consensus sliding clamp binding motif (QLsLF) (Dalrymple et al., 2001). We determined that the QXLLXP motif in both *E. coli* and *B. subtilis* MutL are important for methyl-directed and nick-directed MMR, respectively (chapters 3 and 4). Furthermore, we used chemical crosslinking to show that QXLLXP is a genuine sliding clamp binding motif in both *E. coli* and *B.*

subtilis MutL (chapter 4). The role for this weak, yet specific interaction was unclear for methyl-directed MMR and was presumed to stimulate MutL endonuclease activity in nick-directed MMR.

We structurally characterized the *E. coli* MutL-sliding clamp interaction to uncover the functional importance of this complex in methyl-directed MMR (chapter 5). To stabilize the weak MutL-sliding clamp complex for structural analysis, we designed cysteine variants that were indistinguishable from wild type and could form a disulfide bridge when the proteins interact. Using small angle X-ray scattering, we structurally characterized one of the weakest known sliding clamp binding partners. Functional assays of the stable MutL-sliding clamp complex also uncovered that it is important for UvrD-mediated DNA unwinding in methyl-directed MMR.

The importance of this work extends beyond mismatch repair since the sliding clamp cysteine variant may be used to study other weak and transient sliding clamp binding partners. Indeed, the sliding clamp binds to a series of binding partners with a broad range of cellular functions and binding affinities (Bruning and Shamoo, 2004; Maga and Hubscher, 2003; Pillon et al., 2011; Rolef Ben-Shahar et al., 2009; Yin et al., 2013). The sliding clamp cysteine variant is a feasible approach to stabilizing weak and transient sliding clamp binding partners that are otherwise refractory to structural analysis.

Weak and transient protein interactions have recently gained the attention of the scientific community due to the revelation that these unstable interactions have a significant influence on many protein networks (Perkins et al., 2010; Qin and Gronenborn, 2014). This newfound interest in unstable protein complexes is pushing the limitations of current structure determination methods. Small angle X-ray scattering is advantageous to study weak and transient protein complexes because it is possible to generate *ab initio* models of low abundant protein complexes from sample mixtures contaminated with unbound protein (Tuullanen and Svergun, 2014). Furthermore, the suitable range in protein size is broad (kDa to GDa). The structural analysis of the MutL-sliding clamp complex identified a time-dependent conformational change. The distribution of MutL-sliding clamp conformations shifted over time and ranged from 10-50% of the total population. This work highlights the power of studying protein dynamics in solution and exceeds a previous report that protein conformations as low as 20% can be extracted from mix populations (Blobel et al., 2009).

The long term aim for this project will be to determine the mechanism for MutL-mediated DNA hydrolysis. Extensive biochemical and structural characterization of MutL homologs has provided the foundation to understand MutL endonuclease activity, however, there are lingering questions. There is currently no structural information describing the MutL-DNA interaction and how the catalytic metal is coordinated relative to the DNA backbone. To understand the spatial requirements that activate the MutL endonuclease, the MutL-DNA complex should be analyzed in context with its regulators,

MutS and the sliding clamp. This aim will certainly address the current void in the mismatch repair field by revealing how these weak and transient interactions orchestrate the dynamic mismatch repair response.

Chapter 7 - Appendix

Characterization of the defects in the ATP lid of *E. coli* MutL that cause transient hypermutability

7.1 Author's Preface

This chapter describes the functional defect associated with naturally occurring MutL variants (MutL-2LA and MutL-4LA) expressed in hypermutator bacteria. M. Dubinsky and I generated the *E. coli* MutL variants. I purified the MutL variants and assessed their stability using dynamic light scattering. I performed *in vitro* assays measuring DNA binding, MutS binding, ATP binding and ATP hydrolysis. Our work reveals that incomplete repair in hypermutator strains is due to a defect in MutL ATP binding. A. Guarné and I designed and interpreted the experiments. I prepared the figures and wrote the manuscript.

7.2 Abstract

Mutator strains spontaneously arise in bacterial populations under stress in an attempt to increase evolutionary adaptation. Inactivation of the ubiquitous DNA mismatch repair pathway, whose normal function is to correct replication errors and hence increase replication fidelity, is often the cause of the mutator phenotype. One of the essential genes in this pathway, *mutL*, includes a short tandem repeat that is prone to polymerase slippage during replication. While extensive work has established that this repetitive sequence is a genuine genetic switch, the mechanism of MutL inactivation remains unclear. This short tandem repeat is translated into a LALALA motif that resides near the ATPase active site of MutL. Therefore, changes in the length of this motif are presumed to alter the ATPase activity of MutL. We have engineered variants of *E. coli* MutL with shorter/longer LALALA motifs and characterized their ATPase and DNA binding functions. We have found that the deletion or insertion of a single LA repeat did not compromise the structural integrity of the protein, nor did it affect MutS- or DNA-binding activity. However, it severely compromised ATP binding and, consequently, engagement of the N-terminal domains; both essential activities for proper DNA mismatch repair. These results are discussed in the context of the structure of MutL.

7.3 Introduction

Bacteria have evolved towards a low DNA mutation frequency in stable environments because mutations often have deleterious effects (Sniegowski et al., 2000; Taddei et al., 1997). However, bacterial populations in a new environment are confronted

with the need to adapt. Since asexual populations rely on mutagenesis to increase genetic variation, bacteria tend to have hotspots in antimutator genes, such as DNA repair genes (Jolivet-Gougeon et al., 2011). To compensate for the inevitable deleterious impact of maintaining a high mutation rate, bacterial mutator phenotypes are often transient so the population can spontaneously revert back to a low mutation frequency (Taddei et al., 1997).

Defects in the conserved post-replicative DNA mismatch repair (MMR) pathway are often the cause of reversible hypermutability in bacteria. MMR increases fidelity of DNA replication by repairing mismatched bases and small insertion/deletion loops introduced during DNA synthesis. Three key proteins initiate the repair response (reviewed in (Iyer et al., 2006; Kunkel and Erie, 2005)). MutS recognizes and binds to a mismatch and recruits the molecular matchmaker MutL in an ATP-dependent manner (Junop et al., 2001). The ATP-bound form of MutL can then activate the latent endonuclease MutH which, in turn, nicks the newly synthesized strand at the nearest hemimethylated GATC site (Giron-Monzon et al., 2004; Joseph et al., 2004). The nick marks the nascent strand for repair by providing an entry point for the UvrD helicase and exonucleases to unwind and degrade the erroneous strand, thereby providing the replication machinery with a second chance to re-synthesize the strand correctly (Iyer et al., 2006). While MutS and MutL are evolutionary conserved, MutH is only present in a subset of gamma-proteobacteria including *Escherichia coli (E. coli)*. In organisms lacking MutH, MutL harbours a latent endonuclease activity that uses the gap left by the

removal of misincorporated ribonucleotides as the strand discrimination signal (Ghodgaonkar et al., 2013; Lujan et al., 2013) and provides entry sites for downstream MMR factors (Duppatla et al., 2009; Fukui et al., 2008; Kadyrov et al., 2006; Kadyrov et al., 2007; Mauris and Evans, 2009; Pillon et al., 2010).

Several independent studies have linked bacterial hypermutability to mutations in a conserved short tandem repeat at the beginning of the *mutL* gene (Chen et al., 2010; Gong et al., 2007; Le Bars et al., 2013; Shaver and Sniegowski, 2003; Wang et al., 2010). Sequencing of the *E. coli mutL* gene from hypermutator strains revealed either a deletion or an insertion in a triple six base pairs tandem repeat ($^{5'}$CTGGCGCTGGCGCTGGCG$^{3'}$) between nucleotides 213-230 (Shaver and Sniegowski, 2003). Alterations to this repeat are likely the result of strand slippage, making the mutator phenotype reversible (Shaver and Sniegowski, 2003). This hexanucleotide repeat is strictly conserved among *mutL* homologues found in gamma-proteobacteria encoding a *mutH* gene, loosely conserved in gamma-proteobacteria that do not encode a *mutH* homolog (Figure 7.1A), and absent outside the proteobacteria phylum (Wang et al., 2010).

At the protein level, this repeat results in a LALALA motif that resides in the N-terminal domain of MutL. MutL homologues consist of two structurally conserved domains connected by a flexible linker (Kunkel and Erie, 2005). The N-terminal domain supports DNA-binding and ATPase activity, while the C-terminal domain harbours a constitutive dimerization interface (Ban et al., 1999; Ban and Yang, 1998; Guarné et al.,

2004). The N-terminal domain of MutL belongs to the GHKL superfamily of phosphoryl transferases. This family is characterized by the presence of four conserved motifs defining a Bergerat fold (Bergerat et al., 1997; Dutta and Inouye, 2000; Mushegian et al., 1997). In the case of MutL, the loops surrounding the ATP binding site are disordered in the absence of nucleotide, but they become ordered upon ATP binding (Ban et al., 1999; Ban and Yang, 1998). The LALALA motif precedes one of these loops, the ATP lid that on its ordered –closed– conformation covers the nucleotide binding site (Figure 7.1B). The closed conformation of the lid is stabilized by the coordination of a monovalent cation (Hu et al., 2003). A second metal binding site harbours the catalytic magnesium ion (Ban et al., 1999). Similar to other phosphoryl transferases, both metal ions are required for optimal ATP binding and, in turn, for proper mismatch repair (Hu et al., 2003; Spampinato and Modrich, 2000). ATP binding triggers the association of the N-terminal domains resulting in a closed conformation (Ban et al., 1999; Fukui et al., 2008; Guarné et al., 2004; Sacho et al., 2008; Tran and Liskay, 2000). Upon ATP hydrolysis, the N-terminal domains dissociate to revert back to an open conformation. Cycling between the open and closed conformation regulates downstream MMR events. In particular, ATP binding has been shown to have a stimulatory effect on MutL endonuclease activity in organisms lacking MutH (Duppatla et al., 2009; Kadyrov et al., 2006; Kadyrov et al., 2007; Mauris and Evans, 2009; Pillon et al., 2010).

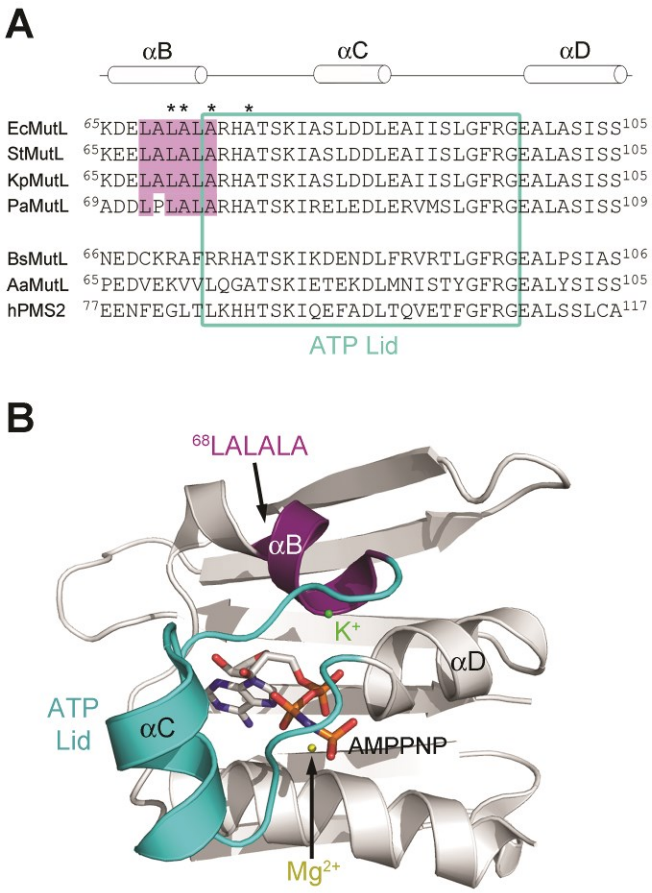


Figure 7.1: The LALALA motif is part of the ATP binding pocket. (A) Sequence alignment of the MutL ATP lid from organisms in the gamma-proteobacteria class (*E. coli* K12 (Ec), *Salmonella typhimurium* LT7 (St), *Klebsiella pneumoniae* (Kp), and *Pseudomonas aeruginosa* (Pa)) and from organisms outside this class (*Bacillus subtilis* (Bs), *Aquifex aeolicus* (Aa), and human MutL homolog PMS2 (hPMS2)). The secondary structure corresponds to the *E. coli* MutL crystal structure (PDB ID 1B63). The lid region is indicated by a blue box and the LALALA motif is highlighted in purple. *E. coli* MutL residues involved in chelating the monovalent metal ion are indicated by black asterisks. **(B)** Ribbon diagram of the *E. coli* MutL nucleotide binding pocket (PDB ID 1NHI). The LALALA motif is shown in purple and the ATP lid is in blue. The AMPPNP molecule is shown as color-coded sticks and the Mg^{2+} and K^{+} ions are shown as yellow and green spheres, respectively.

The role of this short tandem repeat as a genetic switch is well established, however the mechanistic defects associated with variations in the LALALA motif are

unknown. In this work, we have produced MutL variants with a deletion (MutL-2LA) or an insertion (MutL-4LA) in the LALALA motif and characterized their ability to bind MutS, DNA, as well as to bind and hydrolyze ATP, to reveal the functional defects that cause the acquisition of the strong mutator phenotype.

7.4 Materials and Methods

7.4.1 Cloning MutL variants

E. coli MutL (pTX418, residues 1-615) was a kind gift from Dr. Wei Yang (LMB, NIDDK). *E. coli* MutL-2LA (pAG8529, ^{66}DELALALARH^{75}–^{66}DELALARH^{73}) and MutL-4LA (pAG8705, ^{66}DELALALARH^{75}–^{66}DELALALALARH^{77}) were generated by overlap PCR, cloned into the pET15b expression vector (Novagen), and verified by DNA sequencing (MOBIX, McMaster University).

7.4.2 Protein expression and purification

MutL, MutL-2LA and MutL-4LA were produced in *E. coli* BL21(DE3) cells (Invitrogen), protein expression was induced with 1 mM IPTG for 3 hours at 37°C (MutL and MutL-4LA) or 5 hours at 25°C (MutL-2LA). Cells were harvest by centrifugation, the cell pellets resuspended in buffer A (20 mM Tris pH 8.0, 1.4 mM β-mercaptoethanol, and 0.5 M NaCl) and lysed by sonication. Lysates were clarified by centrifugation at 39,000 g, loaded onto a HiTrap nickel-chelating column (GE Healthcare) equilibrated with buffer A, and eluted with 0.3 M imidazole. The sample was further purified using a Q-sepharose column (GE Healthcare) equilibrated using buffer B (20 mM Tris pH 8.5, 90

mM KCl, 5 mM DTT, 1 mM EDTA, and 10% glycerol) and MutL was eluted with a linear salt gradient to 240 mM KCl. MutL was loaded onto a Superdex-200 (GE Healthcare) equilibrated with storage buffer (20 mM Tris pH 8.5, 5 mM DTT, 1 mM EDTA, 150 mM KCl, and 10% glycerol).

7.4.3 Dynamic light scattering

Dynamic light scattering was performed using a Zetasizer Nano S (Malvern Instruments). All measurements were taken using a 12 μL 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.

7.4.4 DNA binding assay

DNA binding activity was evaluated using supercoiled DNA as described previously (Pillon et al., 2010), with minor changes. MutL variants (final concentration, 833 nM) were incubated with supercoiled pUC19 plasmid DNA (5 nM) in DNA binding buffer to a final reaction volume of 15 μL. Reaction mixtures were then incubated for 90 minutes at 37°C and resolved on 1% Tris-Acetate-EDTA agarose gels. Binding to linear DNA was assessed using 68 base-pair substrates as described earlier (Guarné et al., 2004). In brief, single- or double-stranded DNA (final concentration, 5 nM) were mixed with either MutL, MutL-2LA or MutL-4LA (at two different final concentrations: 240 nM and 160 nM) and incubated in DNA binding buffer as described earlier.

7.4.5 MutL N-terminal domain association

MutL variants (14.3 μM) were incubated in the absence or presence of 2 mM AMPPNP (Sigma) using nucleotide binding buffer (20 mM Tris pH 8.4, 150 mM KCl, 0.1 mM EDTA, 5 mM $MgCl_2$, 1 mM DTT, and 5% glycerol). Reactions were incubated for 1 hour at room temperature followed by an overnight incubation at 4°C. Reactions (100 μL) were loaded onto a Superdex-200 column equilibrated using nucleotide binding buffer.

7.4.6 ATP binding and hydrolysis assay

ATP binding was detected by incubating MutL variants (serial dilution from 25 μM) with 1 mM α-^{32}P-labeled ATP (Perkin Elmer) in nucleotide binding buffer. Reactions were incubated for 2 hours at room temperature; 5 μL was spotted on a 0.2 μm nitrocellulose membrane (Bio-Rad) and washed in nucleotide binding buffer for 6 minutes. ATP hydrolysis assays were performed as described previously (Guarné et al., 2001), with minor modifications. MutL variants (1 μM) were incubated with $MgCl_2$ (5 mM) and α-^{32}P-labeled ATP (62.5-1000 μM) in the absence or presence of 3 μM single- or double-stranded DNA (68 bp long) for 2 hour at room temperature. Reactions were resolved by thin-layer chromatography in 0.75 M KH_2PO_4 running buffer. All experiments were performed in triplicates. Error bars represent the standard deviation.

7.4.7 Chemical crosslinking with bis(sulfosuccinimidyl)suberate (BS^3)

The plasmids encoding *E. coli* MutS (pWY1426, residues 1-800) and *E. coli* MutL were a kind gift from Wei Yang (NIDDK, NIH). MutS and the ATPase domain of MutL were produced as described elsewhere with minor modifications (Ban et al., 1999; Junop et al., 2003). The crosslinking reactions were performed as described earlier (Pillon et al., 2011), with the following modifications. MutS (10 μM) was pre-incubated with 10 μM 90 bp G/T mismatch DNA in crosslinking buffer (20 mM Hepes pH 7.5, 1 mM EDTA, 1 mM DTT, 100 mM NaCl, 10% glycerol) for 1 hour on ice. Equal volumes of either MutL, MutL-2LA or MutL-4LA (10 μM), as well as 10 mM ATP, were subsequently added and the reactions were incubated for another 30 minutes. 1.2 mM bis(sulfosuccinimidyl)suberate (BS3, Sigma) was then added and the reactions were allowed to proceed for 30 minutes at 22°C. Crosslinked products were subsequently resolved on 4-15% SDS polyacrylamide gels (Bio-Rad) and stained with Coomassie blue. The presence of MutL and MutS in the novel crosslinked product appearing when both protein were present in the reaction was confirmed by LC-MS/MS (BMSL, UWO Biological Mass Spectrometry Laboratory).

7.5 Results

7.5.1 The length of the LALALA motif does not alter the overall folding of MutL

To investigate the role of the LALALA motif, we generated *E. coli* MutL variants with either a deletion or an insertion of one LA repeat (MutL-2LA and MutL-4LA,

respectively). Both variants were purified similarly to the wild type protein and dynamic light scattering indicated that they had comparable oligomerization to wild-type MutL (**Figure 7.2A**). Accordingly, the estimated mean diameters of MutL-2LA (17.7 nm) and MutL-4LA (17.4 nm) were similar to the size of MutL (17.9 nm), indicating that changes on the length of the motif did not cause major conformational changes to MutL.

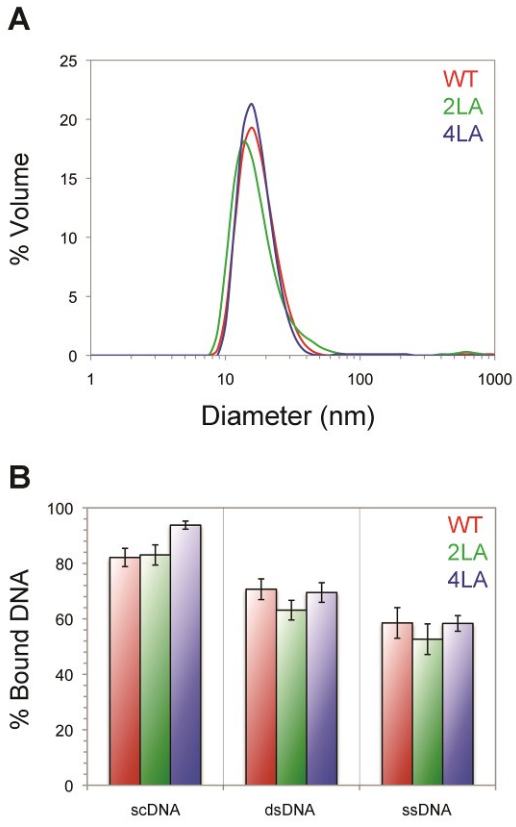


Figure 7.2: MutL-2LA and MutL-4LA are well behaved in solution. (A) Characterization of the particle size distribution of MutL WT (red), MutL-2LA (green), and MutL-4LA (blue) using dynamic light scattering. **(B)** DNA binding abilities of MutL WT (red), MutL-2LA (green), and MutL-4LA (blue) to supercoiled DNA (scDNA, left), linear double-stranded DNA (dsDNA, center) and linear single-stranded DNA (ssDNA, right).

Residues responsible for DNA-binding in *E. coli* MutL are located in the ATPase domain (Ban et al., 1999), though distant from the LALALA motif. We measured DNA-binding activity to assess the effect of changing the length of the motif on this function of MutL. Electrophoretic mobility shift assays indicated that changes on the length of the motif did not affect the ability of MutL to bind supercoiled or linear DNA (Figure 7.2B), thereby confirming that variations to the LALALA motif did not compromise additional functions of the ATPase domain.

It has been previously shown that MutS interacts with the N-terminal domain of MutL (Winkler et al., 2011). Therefore, we next tested whether variations in the LALALA motif affected the interaction with MutS. The N-terminal domain of MutL, lacking the dimerization domain of the protein, only self-associates slowly in the presence of ATP providing a way to assay the interaction between MutS and the open conformation of MutL (Ban et al., 1999). We tested MutS•MutL interaction by incubating MutS with the chemical crosslinker bis(sulfosuccinimidyl)suberate (BS3) in the presence of ATP, a 90 base-pair G/T mismatch DNA substrate and the N-terminal domain of either MutL, MutL-2LA or MutL-4LA. When both MutS and one of the MutL variants were present in the reaction, a new crosslinked species that was not present when either protein was incubated with BS3 appeared at a molecular weight consistent with the formation of the MutS•MutL-NTD complex in the three reactions (Figure 7.3). The formation of this new crosslinked product was not dependent on the presence of either

mismatched DNA or ATP (Figure S7.1). Using LC-MS/MS, we verified the presence of both MutS and MutL-NTD as the sole components of this band.

Collectively, these results indicate that variations on the length of the LALALA motif do not alter the overall architecture of MutL and suggest that the mismatch repair defects associated with the MutL-2LA and MutL-4LA variants are not caused by impaired recruitment of MutL to sites of damage.

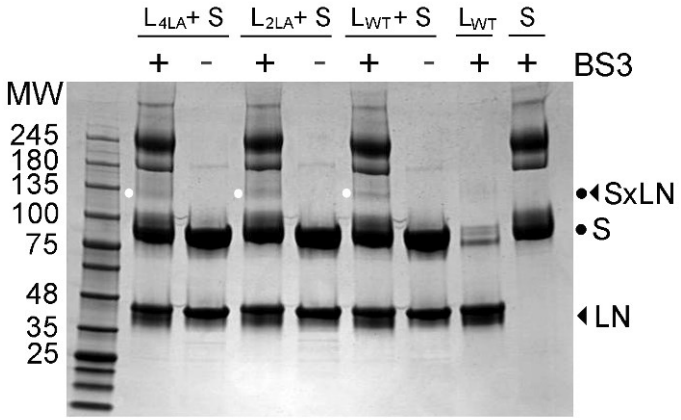


Figure 7.3: MutL, MutL-2LA and MutL-4LA interact with MutS. 4-15% SDS-polyacrylamide gel showing the interaction between the N-terminal ATPase domains (L-N) of MutL (L_{WT}), MutL-2LA (L_{2LA}), and MutL-4LA (L_{4LA}) with MutS (S) by chemical crosslinking with bis(sulfosuccinimidyl)suberate (BS^3, Sigma). From left to right the gel shows: molecular weight markers; equimolar mixtures of MutS and the ATPase domain of MutL-4LA in the presence (+) and absence (-) of BS^3 (1.2 mM); equimolar mixtures of MutS and the ATPase domain of MutL-2LA in the presence (+) and absence (-) of BS^3; equimolar mixtures of MutS and the ATPase domain of MutL in the presence (+) and absence (-) of BS^3; the ATPase domain of MutL in the presence of BS^3; and MutS in the presence of BS^3. Crosslinked products containing both MutS and the ATPase domain of MutL are indicated with a white dot.

7.5.2 The length of the LALALA motif is important for ATP binding

The close proximity of the LALALA motif to the ATP lid led to predict that impaired ATPase activity was the ultimate cause of the DNA mismatch repair defects observed in strains encoding MutL variants with longer or shorter LALALA motifs (Gong et al., 2007). We measured the ATPase activity of MutL-2LA and MutL-4LA and found that both variants had significantly reduced activity (Figure 7.4A). Single- and, to a lesser extent, double-strand DNA have been previously shown to stimulate the ATPase activity of MutL (Ban et al., 1999). However, neither single- nor double-stranded DNA stimulated the residual ATPase activity of the MutL-2LA and MutL-4LA variants (Figure 7.4A and Figure S7.2), suggesting that the ATPase defect was likely due to defective ATP binding.

The ATPase domain of MutL self-associates upon ATP binding, in turn causing a series of conformational changes that are essential for proper mismatch repair (Ban et al., 1999; Fukui et al., 2008; Guarné et al., 2004; Sacho et al., 2008; Tran and Liskay, 2000). To test if the ATP hydrolysis defect was caused by an ATP-binding defect, we tested whether the conformation of MutL-2LA and MutL-4LA changed in the presence of nucleotide. To this end, we pre-incubated full-length MutL, MutL-2LA and MutL-4LA with AMPPNP and the mixtures were resolved by size exclusion chromatography. The presence/absence of nucleotide was monitored by changes in the Abs_{260}:Abs_{280} ratio. As expected, wild type MutL had a smaller apparent volume in the presence of AMPPNP than in its absence due to the association of its N-terminal ATPase domains (Figure 7.4B,

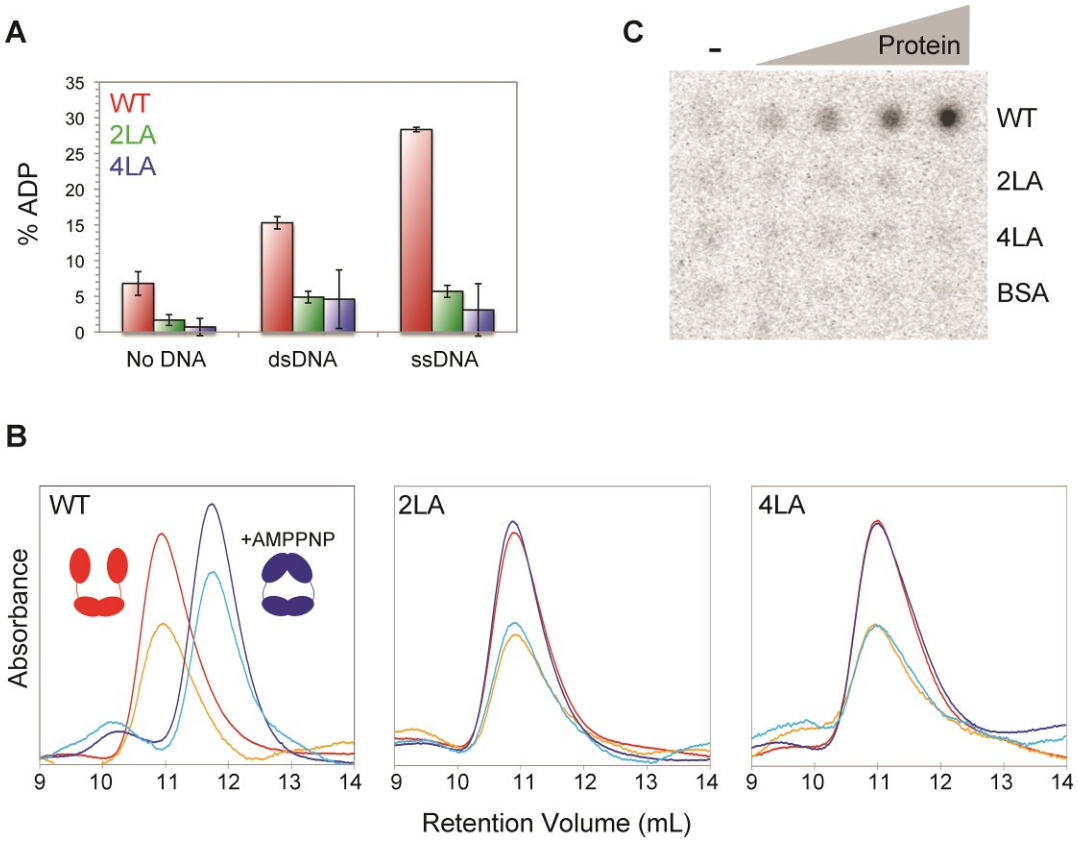


Figure 7.4: MutL-2LA and MutL-4LA are defective for ATP-binding. (A) ATPase activity assay of MutL, MutL-2LA and MutL-4LA in the absence and presence of either double- or single-stranded DNA. Error bars represent the standard deviation of three independent experiments. **(B)** Size exclusion chromatography profiles of MutL (left), MutL-2LA (middle), and MutL-4LA (right) in the absence (red/orange lines) or presence (blue/cyan lines) of AMPPNP. The conformational change of the three variants (14.3 μM) was monitored at 280 nm (red and blue lines) and 260 nm (orange and cyan lines). **(C)** Filter binding assay of α-^{32}P-labeled ATP (1 mM) incubated with either MutL, MutL-2LA or MutL-4LA (serial dilution from 25 μM). Bovine serum albumin (BSA) was used as a negative control.

left panel) (Ban et al., 1999; Ban and Yang, 1998; Guarné et al., 2004). Conversely, neither MutL-2LA nor MutL-4LA underwent an analogous conformational change (Figure 7.4B, middle and right panels). The presence of AMPPNP did not affect their

Abs_{260}:Abs_{280} ratio either, indicating that the MutL-2LA and MutL-4LA variants do not bind AMPPNP stably. We confirmed these results with a filter binding assay and found that MutL-2LA and MutL-4LA do not bind ATP (Figure 7.4C). Collectively, these results indicate that the length of the LALALA motif is critical for the ATP binding of MutL and, in turn, for its proper function in mismatch repair.

7.5.3 Structural basis for the ATP-binding defect of the MutL-2LA and MutL-4LA variants

Since MutL must bind ATP to activate downstream mismatch repair factors such as MutH and UvrD (Ban and Yang, 1998; Giron-Monzon et al., 2004; Joseph et al., 2004; Matson and Robertson, 2006; Robertson et al., 2006), defects on ATP binding will necessarily result in a mutator phenotype. Our data demonstrates that altering the length of the LALALA motif abrogates MutL's ability to bind ATP and thus explains the hypermutability of strains containing variants on the LALALA motif. Interestingly, the LALALA motif does not contact the nucleotide directly. However, it precedes the ATP lid and three residues of the motif coordinate a conserved K^+ metal ion (Figure 7.5). A K^+ ion at this position has also been found in other GHKL phosphotransferases and it has been proposed to assist in forming the ATP lid (Hu et al., 2003). The K^+-Mg^{2+} tandem has also been observed in phosphoryl transfer enzymes beyond the GHKL superfamily, where it provides optimal docking for the phosphate moiety of the substrate (Page and Di Cera, 2006). Indeed, the structure of human MutL homolog PMS2, that lacks the K^+ metal ion and has a disordered ATP lid, hydrolyzes ATP extremely weakly (Guarné et al.,

2001), reinforcing the idea that MutL homologues probably use a K^+-Mg^{2+} tandem to reach its fully active conformation. In turn, this explains the residual ATPase activity for MutL-2LA and MutL-4LA (Figure 7.4A).

In MutL, the K^+ ion is coordinated by carbonyl oxygens of Leu70, Ala71 and Ala73 in the LALALA motif, Ala76 in the tight turn following helix αB, Gly96 at the end of the ATP lid, and a water molecule that bridges the K^+ with both the α-phosphate moiety of AMPPNP and helix αD at the C-terminus of the ATP lid (Figure 7.5). Therefore, binding of this metal ion stabilizes the ends of the ATP lid. Since the interaction with the K^+ metal ion is exclusively mediated by main chain carbonyl groups, the relative location of helix αB –rather than its specific sequence– emerges as the critical feature for K^+ binding. Given the lack of sequence specificity, the interaction could potentially be maintained in the MutL-2LA and MutL-4LA variants. However, shortening the motif by one LA repeat would bring the conserved Arg74-His75 close to the hydrophobic core of the protein, while lengthening the motif would expose the side chains of the additional LA repeat to the solvent (Figure 7.5). Therefore, both MutL-2LA and MutL-4LA variants will have a distorted geometry at the N-terminus of the ATP lid, thereby preventing K^+ binding and, in turn, nucleotide binding.

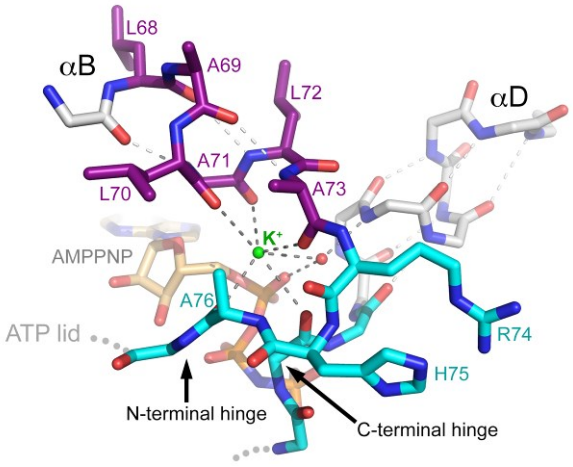


Figure 7.5: The LALALA motif is involved in coordinating a potassium ion. The structure of MutL (PDB ID 1NHI) includes a well ordered K^+ metal ion (green sphere) coordinated by the carbonyl oxygens of residues Leu70, Ala71, Ala73, Ala76 and Gly96, as well as a water molecule (red sphere) that bridges the K^+ metal ion to helix αD at the C-terminus of the ATP lid. The LALALA motif is shown in purple, the N- and C-terminal ends (often referred as hinges) are shown in light blue and the AMPPNP nucleotide in light orange.

7.6 Conclusion

The K^+-Mg^{2+} tandem is likely a conserved feature in all MutL homologues. Although the length of helix αB is conserved across species, only gamma-proteobacteria include the LALALA motif. At the nucleotide level, its repetitive nature makes this sequence prone to strand slippage causing *mutL* inactivation. This genetic switch allows gamma-proteobacteria to spontaneously gain a mutator phenotype, thereby allowing them to adapt to hostile environments. While this provides a great avenue for adaption, spontaneous conversion of bacterial strains into hypermutators has become a critical problem in hospitals trying to control the emergence of multidrug resistant microbes (Wang et al., 2010). Around 1% of pathogenic *E. coli* strains and up to 30% of

pathogenic *Salmonella typhimurium* strains have strong mutator phenotypes (Baquero et al., 2004; Oliver et al., 2000). Similar rates have been found in multidrug-resistant clinical strains of *Pseudomonas aeruginosa*, and it has been suggested that variations in the imperfect ^{72}LPLALA motif could be the source of hypermutability (Wang et al., 2010). Therefore, understanding the molecular mechanisms that regulate this and other genetic switches leading to bacterial hypermutability is important to grasp the emergence of drug resistant pathogens.

7.7 Acknowledgements

We thank members of the Guarné laboratory for helpful discussions and critical reading of the manuscript.

7.8 Supplementary Information

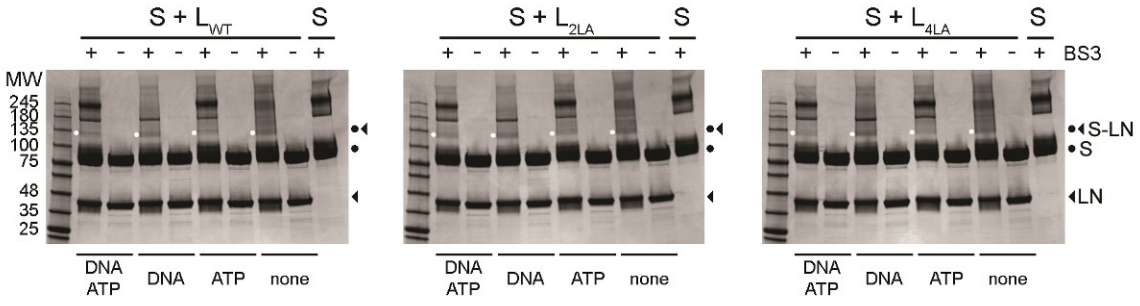


Figure S7.1 Crosslinking of MutS to MutL does not require DNA or ATP. 4-15% SDS-polyacrylamide gel showing the interaction between the N-terminal ATPase domains (LN) of MutL (L_{WT}, left), MutL-2LA (L_{2LA}, center), and MutL-4LA (L_{4LA}, right) with MutS (S) by chemical crosslinking with bis(sulfosuccinimidyl)suberate (BS3). From left to right the gels show: molecular weight markers; equimolar mixtures of MutS and the ATPase domain of each MutL variant in the presence (+) and absence (-) of BS3
Figure S7.1 continued on page 218…

Figure S7.1 continued from page 217…
(1.2 mM) when ATP and DNA are present, when only DNA or ATP is present, or when neither DNA nor ATP are present in the reaction; the last lane of each gel shows the effect of BS3 on MutS when neither DNA or ATP are present. Crosslinked species containing both MutS and the ATPase domain of MutL are indicated with a white dot.

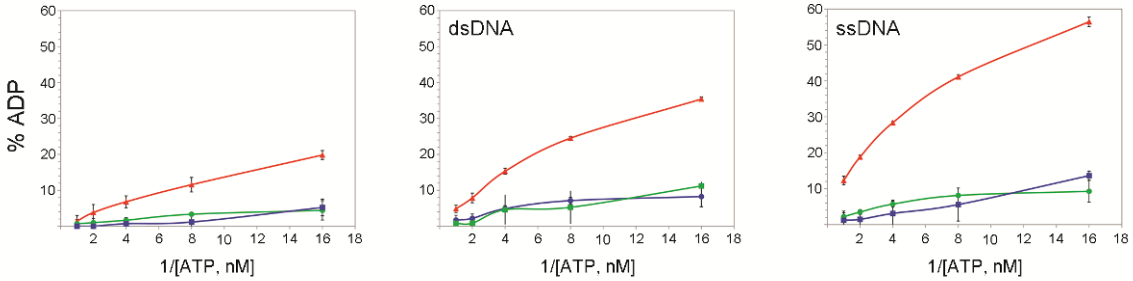


Figure S7.2 MutL-2LA and MutL-4LA do not retain DNA-dependent ATPase activity. Plot of the percentage of hydrolyzed ATP for MutL (red), MutL-2LA (green) and MutL-4LA (blue) at increasing concentrations of ATP. The effect of DNA on the ATPase activity of the three variants was assessed in the presence of a 68 base pairs duplex DNA (middle panel) and a 68 nucleotides single-stranded DNA (right panel). Error bars represent the standard deviation of three independent measurements.

7.9 References

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

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

Baquero, M.R., Nilsson, A.I., Turrientes Mdel, C., Sandvang, D., Galan, J.C., Martinez, J.L., Frimodt-Moller, N., Baquero, F., and Andersson, D.I. (2004). Polymorphic mutation frequencies in Escherichia coli: emergence of weak mutators in clinical isolates. Journal of bacteriology *186*, 5538-5542.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature *386*, 414-417.

Chen, F., Liu, W.Q., Liu, Z.H., Zou, Q.H., Wang, Y., Li, Y.G., Zhou, J., Eisenstark, A., Johnston, R.N., Liu, G.R.*, et al.* (2010). mutL as a genetic switch of bacterial mutability: turned on or off through repeat copy number changes. FEMS Microbiol Lett *312*, 126-132.

Duppatla, V., Bodda, C., Urbanke, C., Friedhoff, P., and Rao, D.N. (2009). The carboxy-terminal domain is sufficient for endonuclease activity of Neisseria gonorrhoeae MutL. Biochem J *423*, 265-277.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci *25*, 24-28.

Fukui, K., Nishida, M., Nakagawa, N., Masui, R., and Kuramitsu, S. (2008). Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. The Journal of biological chemistry *283*, 12136-12145.

Ghodgaonkar, M.M., Lazzaro, F., Olivera-Pimentel, M., Artola-Boran, M., Cejka, P., Reijns, M.A., Jackson, A.P., Plevani, P., Muzi-Falconi, M., and Jiricny, J. (2013). Ribonucleotides Misincorporated into DNA Act as Strand-Discrimination Signals in Eukaryotic Mismatch Repair. Molecular cell *50*, 323-332.

Giron-Monzon, L., Manelyte, L., Ahrends, R., Kirsch, D., Spengler, B., and Friedhoff, P. (2004). Mapping protein-protein interactions between MutL and MutH by cross-linking. The Journal of biological chemistry *279*, 49338-49345.

Gong, J., Liu, W.Q., Liu, G.R., Chen, F., Li, J.Q., Xu, G.M., Wang, L., Johnston, R.N., Eisenstark, A., and Liu, S.L. (2007). Spontaneous conversion between mutL and 6 bpDeltamutL in Salmonella typhimurium LT7: association with genome diversification and possible roles in bacterial adaptation. Genomics *90*, 542-549.

Guarné, A., Junop, M.S., and Yang, W. (2001). Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. Embo J *20*, 5521-5531.

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

Hu, X., Machius, M., and Yang, W. (2003). Monovalent cation dependence and preference of GHKL ATPases and kinases. FEBS Lett *544*, 268-273.

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

Jolivet-Gougeon, A., Kovacs, B., Le Gall-David, S., Le Bars, H., Bousarghin, L., Bonnaure-Mallet, M., Lobel, B., Guille, F., Soussy, C.J., and Tenke, P. (2011). Bacterial hypermutation: clinical implications. J Med Microbiol *60*, 563-573.

Joseph, N., Sawarkar, R., and Rao, D.N. (2004). DNA mismatch correction in Haemophilus influenzae: characterization of MutL, MutH and their interaction. DNA Repair (Amst) *3*, 1561-1577.

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

Junop, M.S., Yang, W., Funchain, P., Clendenin, W., and 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 (Amst) *2*, 387-405.

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

Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., and Modrich, P. (2007). Saccharomyces cerevisiae MutLa is a mismatch repair endonuclease. The Journal of biological chemistry *282*, 37181-37190.

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

Le Bars, H., Bousarghin, L., Bonnaure-Mallet, M., and Jolivet-Gougeon, A. (2013). Role of a short tandem leucine/arginine repeat in strong mutator phenotype acquisition in a clinical isolate of Salmonella Typhimurium. FEMS Microbiol Lett *338*, 101-106.

Lujan, S.A., Williams, J.S., Clausen, A.R., Clark, A.B., and Kunkel, T.A. (2013). Ribonucleotides Are Signals for Mismatch Repair of Leading-Strand Replication Errors. Molecular cell *50*, 437-443.

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

Mauris, J., and Evans, T.C. (2009). Adenosine triphosphate stimulates Aquifex aeolicus MutL endonuclease activity. PLoS One *4*, e7175.

Mushegian, A.R., Bassett, D.E., Jr., Boguski, M.S., Bork, P., and Koonin, E.V. (1997). Positionally cloned human disease genes: patterns of evolutionary conservation and functional motifs. Proceedings of the National Academy of Sciences of the United States of America *94*, 5831-5836.

Oliver, A., Canton, R., Campo, P., Baquero, F., and Blazquez, J. (2000). High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science (New York, N.Y.) *288*, 1251-1254.

Page, M.J., and Di Cera, E. (2006). Role of Na+ and K+ in enzyme function. Physiol Rev *86*, 1049-1092.

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

Pillon, M.C., Miller, J.H., and Guarne, A. (2011). The endonuclease domain of MutL interacts with the beta sliding clamp. DNA Repair (Amst) *10*, 87-93.

Robertson, A.B., Pattishall, S.R., Gibbons, E.A., and Matson, S.W. (2006). MutL-catalyzed ATP hydrolysis is required at a post-UvrD loading step in methyl-directed mismatch repair. The Journal of biological chemistry *281*, 19949-19959.

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

Shaver, A.C., and Sniegowski, P.D. (2003). Spontaneously arising mutL mutators in evolving Escherichia coli populations are the result of changes in repeat length. Journal of bacteriology *185*, 6076-6082.

Sniegowski, P.D., Gerrish, P.J., Johnson, T., and Shaver, A. (2000). The evolution of mutation rates: separating causes from consequences. Bioessays *22*, 1057-1066.

Spampinato, C., and Modrich, P. (2000). The MutL ATPase is required for mismatch repair. The Journal of biological chemistry *275*, 9863-9869.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997). Role of mutator alleles in adaptive evolution. Nature *387*, 700-702.

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

Wang, Y.M., Wan, Y.L., Wang, X., Wang, P.Y., Wang, Q.S., Cao, X.L., and Liu, Y.C. (2010). The mutL mutation in Pseudomonas aeruginosa isolates reveals multidrug-resistant traits and possible evolutionary trends. J Int Med Res *38*, 2011-2022.

Winkler, I., Marx, A.D., Lariviere, D., Heinze, R.J., Cristovao, M., Reumer, A., Curth, U., Sixma, T.K., and Friedhoff, P. (2011). Chemical Trapping of the Dynamic MutS-MutL Complex Formed in DNA Mismatch Repair in Escherichia coli. The Journal of biological chemistry *286*, 17326-17337.

References

Abe, Y., and Masuda, H. (2000). Genetic alterations of sporadic colorectal cancer with microsatellite instability, especially characteristics of primary multiple colorectal cancers. J Surg Oncol *74*, 249-256.

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

Alani, E., Sokolsky, T., Studamire, B., Miret, J.J., and 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*, 2436-2447.

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

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

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. The Journal of Experimental Medicine *79*, 137-158.

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

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

Beattie, T.R., and Bell, S.D. (2011). The role of the DNA sliding clamp in Okazaki fragment maturation in archaea and eukaryotes. Biochemical Society transactions *39*, 70-76.

Bende, S.M., and Grafstrom, R.H. (1991). The DNA binding properties of the MutL protein isolated from Escherichia coli. Nucleic acids research *19*, 1549-1555.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature *386*, 414-417.

Blackwell, L.J., Martik, D., Bjornson, K.P., Bjornson, E.S., and Modrich, P. (1998). Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. The Journal of biological chemistry *273*, 32055-32062.

Blobel, J., Bernado, P., Svergun, D.I., Tauler, R., and Pons, M. (2009). Low-resolution structures of transient protein-protein complexes using small-angle X-ray scattering. Journal of the American Chemical Society *131*, 4378-4386.

Blumenthal, A.B., and Clark, E.J. (1977). Discrete sizes of replication intermediates in Drosophila cells. Cell *12*, 183-189.

Bruning, J.B., and Shamoo, Y. (2004). Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. Structure (London, England : 1993) *12*, 2209-2219.

Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic acids research *41*, W349-357.

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

Cho, M., Chung, S., Heo, S.D., Ku, J., and Ban, C. (2007). A simple fluorescent method for detecting mismatched DNAs using a MutS-fluorophore conjugate. Biosensors & bioelectronics *22*, 1376-1381.

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

Cooper, D.L., Lahue, R.S., and Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. Journal of biological chemistry *268*, 11823-11829.

Correa, E.M., De Tullio, L., Velez, P.S., Martina, M.A., Argarana, C.E., and Barra, J.L. (2013). Analysis of DNA structure and sequence requirements for Pseudomonas aeruginosa MutL endonuclease activity. Journal of biochemistry *154*, 505-511.

Cox, E.C., Degnen, G.E., and Scheppe, M.L. (1972). Mutator gene studies in Escherichia coli: the mutS gene. Genetics *72*, 551-567.

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. Developmental biology *278*, 274-288.

Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E., and Jennings, P.A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proceedings of the National Academy of Sciences of the United States of America *98*, 11627-11632.

Dao, V., and Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. The Journal of biological chemistry *273*, 9202-9207.

DeRocco, V.C., Sass, L.E., Qiu, R., Weninger, K.R., and Erie, D.A. (2014). Dynamics of MutS-mismatched DNA complexes are predictive of their repair phenotypes. Biochemistry *53*, 2043-2052.

Duppatla, V., Bodda, C., Urbanke, C., Friedhoff, P., and Rao, D.N. (2009). The carboxy-terminal domain is sufficient for endonuclease activity of Neisseria gonorrhoeae MutL. Biochem J.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. Trends in biochemical sciences *25*, 24-28.

Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M., and Modrich, P. (2004). A defined human system that supports bidirectional mismatch-provoked excision. Molecular cell *15*, 31-41.

Eder, P.S., and Walder, J.A. (1991). Ribonuclease H from K562 human erythroleukemia cells. Purification, characterization, and substrate specificity. The Journal of biological chemistry *266*, 6472-6479.

Essers, J., Theil, A.F., Baldeyron, C., van Cappellen, W.A., Houtsmuller, A.B., Kanaar, R., and Vermeulen, W. (2005). Nuclear dynamics of PCNA in DNA replication and repair. Molecular and cellular biology *25*, 9350-9359.

Flores-Rozas, H., Clark, D., and Kolodner, R.D. (2000). Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. Nature genetics *26*, 375-378.

Franklin, R.E., and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740-741.

Fukui, K. (2010). DNA mismatch repair in eukaryotes and bacteria. Journal of nucleic acids *2010*.

Fukui, K., Nishida, M., Nakagawa, N., Masui, R., and Kuramitsu, S. (2008). Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. The Journal of biological chemistry *283*, 12136-12145.

Genschel, J., Bazemore, L.R., and Modrich, P. (2002). Human exonuclease I is required for 5' and 3' mismatch repair. The Journal of biological chemistry *277*, 13302-13311.

Genschel, J., and Modrich, P. (2003). Mechanism of 5'-directed excision in human mismatch repair. Molecular cell *12*, 1077-1086.

Genschel, J., and Modrich, P. (2009). Functions of MutLalpha, replication protein A (RPA), and HMGB1 in 5'-directed mismatch repair. The Journal of biological chemistry *284*, 21536-21544.

Ghodgaonkar, M.M., Lazzaro, F., Olivera-Pimentel, M., Artola-Boran, M., Cejka, P., Reijns, M.A., Jackson, A.P., Plevani, P., Muzi-Falconi, M., and Jiricny, J. (2013). Ribonucleotides Misincorporated into DNA Act as Strand-Discrimination Signals in Eukaryotic Mismatch Repair. Molecular cell *50*, 323-332.

Giron-Monzon, L., Manelyte, L., Ahrends, R., Kirsch, D., Spengler, B., and Friedhoff, P. (2004). Mapping protein-protein interactions between MutL and MutH by cross-linking. The Journal of biological chemistry *279*, 49338-49345.

Gorman, J., Chowdhury, A., Surtees, J.A., Shimada, J., Reichman, D.R., Alani, E., and Greene, E.C. (2007). Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. Molecular cell *28*, 359-370.

Gradia, S., Acharya, S., and Fishel, R. (1997). The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. Cell *91*, 995-1005.

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

Grilley, M., Griffith, J., and Modrich, P. (1993). Bidirectional excision in methyl-directed mismatch repair. Journal of biological chemistry *268*, 11830-11837.

Groothuizen, F.S., Fish, A., Petoukhov, M.V., Reumer, A., Manelyte, L., Winterwerp, H.H., Marinus, M.G., Lebbink, J.H., Svergun, D.I., Friedhoff, P.*, et al.* (2013). Using stable MutS dimers and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation. Nucleic acids research *41*, 8166-8181.

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

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

Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., Londino, F., Quemener, C., Le Du, M.H., Marquez, J.A.*, et al.* (2013). Structure of the MutLalpha C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. Nature structural & molecular biology *20*, 461-468.

Gupta, S., Gellert, M., and Yang, W. (2012). Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops. Nature structural & molecular biology *19*, 72-78.

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

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

Harding, M.M. (1999). The geometry of metal-ligand interactions relevant to proteins. Acta crystallographica. Section D, Biological crystallography *55*, 1432-1443.

Heltzel, J.M., Maul, R.W., Wolff, D.W., and Sutton, M.D. (2012). Escherichia coli DNA polymerase IV (Pol IV), but not Pol II, dynamically switches with a stalled Pol III* replicase. Journal of bacteriology *194*, 3589-3600.

Iino, H., Kim, K., Shimada, A., Masui, R., Kuramitsu, S., and 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. Biosci Rep *31*, 309-322.

Iyer, R.R., Pohlhaus, T.J., Chen, S., Hura, G.L., Dzantiev, L., Beese, L.S., and Modrich, P. (2008). The MutSalpha-proliferating cell nuclear antigen interaction in human DNA mismatch repair. The Journal of biological chemistry *283*, 13310-13319.

Jeong, C., Cho, W.K., Song, K.M., Cook, C., Yoon, T.Y., Ban, C., Fishel, R., and Lee, J.B. (2011). MutS switches between two fundamentally distinct clamps during mismatch repair. Nature structural & molecular biology *18*, 379-385.

Joseph, N., Sawarkar, R., and Rao, D.N. (2004). DNA mismatch correction in Haemophilus influenzae: characterization of MutL, MutH and their interaction. DNA repair *3*, 1561-1577.

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

Junop, M.S., Yang, W., Funchain, P., Clendenin, W., and 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., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. Cell *126*, 297-308.

Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., and Modrich, P. (2007). Saccharomyces cerevisiae MutLa is a mismatch repair endonuclease. The Journal of biological chemistry.

Kane, M.F., Loda, M., Gaida, G.M., Lipman, J., Mishra, R., Goldman, H., Jessup, J.M., and Kolodner, R. (1997). Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res *57*, 808-811.

Kato, R., Kataoka, M., Kamikubo, H., and Kuramitsu, S. (2001). Direct observation of three conformations of MutS protein regulated by adenine nucleotides. Journal of molecular biology *309*, 227-238.

Kelch, B.A., Makino, D.L., O'Donnell, M., and Kuriyan, J. (2012). Clamp loader ATPases and the evolution of DNA replication machinery. BMC biology *10*, 34.

Kelman, Z., and O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. Annual review of biochemistry *64*, 171-200.

Kosinski, J., Plotz, G., Guarné, A., Bujnicki, J.M., and Friedhoff, P. (2008). The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family. Journal of molecular biology *382*, 610-627.

Kosinski, J., Steindorf, I., Bujnicki, J.M., Giron-Monzon, L., and Friedhoff, P. (2005). Analysis of the quaternary structure of the MutL C-terminal domain. Journal of molecular biology *351*, 895-909.

Kunkel, T.A. (1992). DNA replication fidelity. The Journal of biological chemistry *267*, 18251-18254.

Kunkel, T.A. (2009). Evolving views of DNA replication (in)fidelity. Cold Spring Harbor symposia on quantitative biology *74*, 91-101.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annual review of biochemistry *74*, 681-710.

Lamers, M.H., Georgijevic, D., Lebbink, J.H., Winterwerp, H.H., Agianian, B., de Wind, N., and Sixma, T.K. (2004). ATP increases the affinity between MutS ATPase domains. Implications for ATP hydrolysis and conformational changes. The Journal of biological chemistry *279*, 43879-43885.

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

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

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langle-Rouault, F., Maenhaut-Michel, G., and Radman, M. (1987). GATC sequences, DNA nicks and the MutH function in Escherichia coli mismatch repair. The EMBO journal *6*, 1121-1127.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) *23*, 2947-2948.

Lee, J.B., Cho, W.K., Park, J., Jeon, Y., Kim, D., Lee, S.H., and Fishel, R. (2014). Single-molecule views of MutS on mismatched DNA. DNA repair.

Lehman, I.R., Bessman, M.J., Simms, E.S., and Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. The Journal of biological chemistry *233*, 163-170.

Lehmann, A.R. (2003). Replication of damaged DNA. Cell cycle (Georgetown, Tex.) *2*, 300-302.

Lenhart, J.S., Pillon, M.C., Guarne, A., and Simmons, L.A. (2013). Trapping and visualizing intermediate steps in the mismatch repair pathway in vivo. Molecular microbiology *90*, 680-698.

Levene, P. (1919). The structure of yeast nucleic acid. The Journal of biological chemistry *40*, 10.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. Cell *107*, 91-102.

Lipkin, S.M., Moens, P.B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., Thomas, J., Cheng, J., Touchman, J.W., Green, E.D*., et al.* (2002). Meiotic arrest and aneuploidy in MLH3-deficient mice. Nature genetics *31*, 385-390.

Lopez de Saro, F.J., Marinus, M.G., Modrich, P., and O'Donnell, M. (2006). The beta sliding clamp binds to multiple sites within MutL and MutS. The Journal of biological chemistry *281*, 14340-14349.

Lopez de Saro, F.J., and O'Donnell, M. (2001). Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I. Proceedings of the National Academy of Sciences of the United States of America *98*, 8376-8380.

Lujan, S.A., Williams, J.S., Clausen, A.R., Clark, A.B., and Kunkel, T.A. (2013). Ribonucleotides Are Signals for Mismatch Repair of Leading-Strand Replication Errors. Molecular cell.

Lynch, H.T., Lynch, J.F., and Attard, T.A. (2009). Diagnosis and management of hereditary colorectal cancer syndromes: Lynch syndrome as a model. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne *181*, 273-280.

Maga, G., and Hubscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. Journal of cell science *116*, 3051-3060.

Maguire, M.E., and Cowan, J.A. (2002). Magnesium chemistry and biochemistry. Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine *15*, 203-210.

Markowitz, S. (2000). DNA repair defects inactivate tumor suppressor genes and induce hereditary and sporadic colon cancers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *18*, 75s-80s.

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

Mauris, J., and Evans, T.C. (2009). Adenosine triphosphate stimulates Aquifex aeolicus MutL endonuclease activity. PloS one *4*, e7175.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America *74*, 560-564.

Mendillo, M.L., Hargreaves, V.V., Jamison, J.W., Mo, A.O., Li, S., Putnam, C.D., Woods, V.L. Jr., and Kolodner, R.D. (2009). A conserved MutS homolog connector domain interface interacts with MutL homologs. Proceedings of the National Academy of Sciences of the United States of America *106*, 22223-22228.

Meselson, M., and 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.

Monti, M.C., Cohen, S.X., Fish, A., Winterwerp, H.H., Barendregt, A., Friedhoff, P., Perrakis, A., Heck, A.J., Sixma, T.K., van den Heuvel, R.H.*, et al.* (2011). Native mass spectrometry provides direct evidence for DNA mismatch-induced regulation of asymmetric nucleotide binding in mismatch repair protein MutS. Nucleic acids research *39*, 8052-8064.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology *51 Pt 1*, 263-273.

Mushegian, A.R., Bassett, D.E., Jr., Boguski, M.S., Bork, P., and Koonin, E.V. (1997). Positionally cloned human disease genes: patterns of evolutionary conservation and functional motifs. Proceedings of the National Academy of Sciences of the United States of America *94*, 5831-5836.

Namadurai, S., Jain, D., Kulkarni, D.S., Tabib, C.R., Friedhoff, P., Rao, D.N., and Nair, D.T. (2010). The C-terminal domain of the MutL homolog from Neisseria gonorrhoeae forms an inverted homodimer. PloS one *5*, e13726.

Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundstrom, E.B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010a). Genome instability due to ribonucleotide incorporation into DNA. Nature chemical biology *6*, 774-781.

Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundstrom, E.B., Burgers, P.M., Johansson, E., Chabes, A., and Kunkel, T.A. (2010b). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proceedings of the National Academy of Sciences of the United States of America *107*, 4949-4954.

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

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and 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.

Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P., and Beese, L.S. (2011). Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. Cell *145*, 212-223.

Orel, V., and Wood, R.J. (2000). Essence and origin of Mendel's discovery. Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie *323*, 1037-1041.

Pavlov, Y.I., Mian, I.M., and Kunkel, T.A. (2003). Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. Current biology : CB *13*, 744-748.

Peltomaki, P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *21*, 1174-1179.

Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. Structure (London, England : 1993) *18*, 1233-1243.

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

Pillon, M.C., Miller, J.H., and Guarne, A. (2011). The endonuclease domain of MutL interacts with the beta sliding clamp. DNA repair *10*, 87-93.

Plotz, G., Welsch, C., Giron-Monzon, L., Friedhoff, P., Albrecht, M., Piiper, A., Biondi, R.M., Lengauer, T., Zeuzem, S., and Raedle, J. (2006). Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic acids research *34*, 6574-6586.

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

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

Pray, L.A. (2008). DNA replication and causes of mutation. Nature Education *1*, 214.

Prelich, G., Tan, C.K., Kostura, M., Mathews, M.B., So, A.G., Downey, K.M., and Stillman, B. (1987). Functional identity of proliferating cell nuclear antigen and a DNA polymerase-delta auxiliary protein. Nature *326*, 517-520.

Pukkila, P.J., Peterson, J., Herman, G., Modrich, P., and Meselson, M. (1983). Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. Genetics *104*, 571-582.

Qin, J., and Gronenborn, A.M. (2014). Weak protein complexes: challenges to study but essential for life. FEBS Journal *281*, 1948-1949.

Qiu, R., DeRocco, V.C., Harris, C., Sharma, A., Hingorani, M.M., Erie, D.A., and Weninger, K.R. (2012). Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. The EMBO journal *31*, 2528-2540.

Rolef Ben-Shahar, T., Castillo, A.G., Osborne, M.J., Borden, K.L., Kornblatt, J., and Verreault, A. (2009). Two fundamentally distinct PCNA interaction peptides contribute to chromatin assembly factor 1 function. Molecular and cellular biology *29*, 6353-6365.

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

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America *74*, 5463-5467.

Schaaper, R.M. (1993). Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. The Journal of biological chemistry *268*, 23762-23765.

Schofield, M.J., Brownewell, F.E., Nayak, S., Du, C., Kool, E.T., and Hsieh, P. (2001a). The Phe-X-Glu DNA binding motif of MutS. The role of hydrogen bonding in mismatch recognition. The Journal of biological chemistry *276*, 45505-45508.

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

Sharma, A., Doucette, C., Biro, F.N., and 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*, 4192-4205.

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

Simmons, L.A., Davies, B.W., Grossman, A.D., and Walker, G.C. (2008). Beta clamp directs localization of mismatch repair in Bacillus subtilis. Molecular cell *29*, 291-301.

Simoni, R.D., Hill, R.L., and Vaughan, M. (2002). The structure of nucleic acids and many other natural products: Phoebus Aaron Levene. The Journal of biological chemistry *277*.

Spampinato, C., and Modrich, P. (2000). The MutL ATPase is required for mismatch repair. The Journal of biological chemistry *275*, 9863-9869.

Sparks, J.L., Chon, H., Cerritelli, S.M., Kunkel, T.A., Johansson, E., Crouch, R.J., and Burgers, P.M. (2012). RNase H2-initiated ribonucleotide excision repair. Molecular cell *47*, 980-986.

Stillman, B. (2008). DNA polymerases at the replication fork in eukaryotes. Molecular cell *30*, 259-260.

Stoll, K.E., Draper, W.E., Kliegman, J.I., Golynskiy, M.V., Brew-Appiah, R.A., Phillips, R.K., Brown, H.K., Breyer, W.A., Jakubovics, N.S., Jenkinson, H.F.*, et al.* (2009). Characterization and Structure of the Manganese-Responsive Transcriptional Regulator ScaR. Biochemistry *48*, 10308-10320.

Studamire, B., Quach, T., and Alani, E. (1998). Saccharomyces cerevisiae Msh2p and Msh6p ATPase activities are both required during mismatch repair. Molecular and cellular biology *18*, 7590-7601.

Stukenberg, P.T., Turner, J., and O'Donnell, M. (1994). An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. Cell *78*, 877-887.

Tessmer, I., Yang, Y., Zhai, J., Du, C., Hsieh, P., Hingorani, M., and Erie, D.A. (2008). Mechanism of MutS searching for DNA mismatches and signaling repair. Journal of biological chemistry *283*, 36646-36654.

Tran, P.T., and Liskay, R.M. (2000). Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha. Molecular and cellular biology *20*, 6390-6398.

Tsurimoto, T., and Stillman, B. (1991). Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. The Journal of biological chemistry *266*, 1950-1960.

Tuullanen, A.T., and Svergun, D. (2014). Weak protein-ligand interactions studied by small angle X-ray scattering. The FEBS journal *281*, 1974-1987.

Umar, A., Buermeyer, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M., and Kunkel, T.A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. Cell *87*, 65-73.

Vasen, H.F., Moslein, G., Alonso, A., Bernstein, I., Bertario, L., Blanco, I., Burn, J., Capella, G., Engel, C., Frayling, I.*, et al.* (2007). Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). Journal of medical genetics *44*, 353-362.

Vischer, E., and Chargaff, E. (1948). The separation and quantitative estimation of purines and pyrimidines in minute amounts. The Journal of biological chemistry *176*, 703-714.

Voet, D., and Voet, J.G. (2004). Biochemistry, Vol One, 3rd edn (USA: John Wiley & Sons, Inc.).

Waga, S., Bauer, G., and Stillman, B. (1994). Reconstitution of complete SV40 DNA replication with purified replication factors. The Journal of biological chemistry *269*, 10923-10934.

Wang, H., Yang, Y., Schofield, M.J., Du, C., Fridman, Y., Lee, S.D., Larson, E.D., Drummond, J.T., Alani, E., Hsieh, P.*, et al.* (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*, 14822-14827.

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

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

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

Wilkins, M.H., Stokes, A.R., and Wilson, H.R. (1953). Molecular structure of deoxypentose nucleic acids. Nature *171*, 738-740.

Winkler, I., Marx, A.D., Lariviere, D., Heinze, R.J., Cristovao, M., Reumer, A., Curth, U., Sixma, T.K., and Friedhoff, P. (2011). Chemical Trapping of the Dynamic MutS-MutL Complex Formed in DNA Mismatch Repair in Escherichia coli. The Journal of biological chemistry *286*, 17326-17337.

Wold, M.S., and Kelly, T. (1988). Purification and characterization of replication protein A, a cellular protein required for in vitro replication of simian virus 40 DNA. Proceedings of the National Academy of Sciences of the United States of America *85*, 2523-2527.

Wu, C.A., Zechner, E.L., and Marians, K.J. (1992). Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. The Journal of biological chemistry *267*, 4030-4044.

Yamamoto, T., Iino, H., Kim, K., Kuramitsu, S., and Fukui, K. (2011). Evidence for ATP-dependent structural rearrangement of nuclease catalytic site in DNA mismatch repair endonuclease MutL. The Journal of biological chemistry *286*, 42337-42348.

Yang, W. (2005). Portraits of a Y-family DNA polymerase. FEBS letters *579*, 868-872.
Yao, N.Y., and O'Donnell, M. (2010). SnapShot: The replisome. Cell *141*, 1088, 1088.e1081.

Yin, Z., Kelso, M.J., Beck, J.L., and Oakley, A.J. (2013). Structural and thermodynamic dissection of linear motif recognition by the E. coli sliding clamp. Journal of medicinal chemistry *56*, 8665-8673.

Zechner, E.L., Wu, C.A., and Marians, K.J. (1992). Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. III. A polymerase-primase interaction governs primer size. The Journal of biological chemistry *267*, 4054-4063.

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