A Dissertation Presented

By

Tamalette A. Loh

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 29, 2000

Pharmacology and Molecular Toxicology
MUTATIONAL ANALYSIS OF THE MUTH PROTEIN FROM

ESCHERICHIA COLI

A Dissertation Presented

By

Tamalette Loh

Approved as to style and content by:

Anthony Poteete, Chair of Committee

Emil Smith, Member of Committee

Kai Lin, Member of Committee

William Royer, Member of Committee

Donald Nelson, Member of Committee

Martin G. Marinus, Dissertation Mentor

Thomas B. Miller, Jr., Dean of the Graduate School of Biomedical Sciences

Pharmacology and Molecular Toxicology

September 29, 2000
Acknowledgments

I would like to thank Martin Marinus for all of his support during the many times I have been in and out of his lab. I would also like to thank Larry Hardy for accepting me into his lab in an effort to tackle x-ray crystallography. Ellen Nalivaika deserves special mention for all her help in Larry's lab (especially reaching those top shelves). I would like to thank Te-Hui Wu not only for helping me in the lab, but also for being an excellent listener and a good friend. I have been at the university a long time and will miss everyone (this includes Judy who kept my glassware spotless and unbroken).

I would like my parents to know that all of their visits were helpful in keeping my spirits up. I know it wasn't easy for them to drive all the way out here from Ohio. My dear husband helped also (the lunches always tasted good). He let me sleep many times when the children woke up crying at night. I am glad to have had that support.

Finally, one last thanks to Dr. John Wilson, PhD who suggested that I start this adventure in the first place.
Abstract

DNA mismatch repair is one process in the preservation of genomic integrity. It has been found in Archeae, bacteria, plants, yeast and mammals. The mismatch repair system is highly conserved among species and allows the strand-specific elimination of base-base mispairs, chemical base modifications, as well as short insertion/deletion loops following DNA replication. The repair system also has important effects on homeologous recombination, contributing to the frequency of reciprocal exchanges. In humans, defects in the repair system have been found to be associated with tumorigenesis.

In *Escherichia coli*, this pathway was originally called long patch repair before being renamed the methyl-directed mismatch repair system. It is unique in that it utilizes a DNA methylation pattern to discriminate between the parental DNA strand and the newly synthesized daughter DNA strand. The current model for the initiation of methyl-directed mismatch repair is that the mispaired bases are recognized and bound by the MutS protein with MutL as a helper protein for binding. MutL also assists the MutH protein to bind, thereby forming the completed initiation complex of MutS, MutL and MutH. In the presence of ATP, there is evidence for translocation of the complex along the DNA forming alpha loops. At a d(GATC) site the MutH protein binds and nicks the unmethylated daughter DNA strand 5' to the d(G) (by recognizing the N6-d(A))
methylation of the parental DNA strand which it is unable to cut). This completes the initiation of the repair system and allows the hydrolysis and resynthesis of the daughter DNA strand.

MutH is a monomer of 25.5 kD in solution and contains a latent Mg\(^{2+}\)-dependent endonuclease activity. Unmethylated DNA is nicked without any discrimination on one of the two strands and fully methylated DNA is resistant to cleavage by MutH even though the protein is able to bind the d(GATC) site. The structure of MutH was recently solved and compared to a group of restriction endonucleases that share a structural common core domain with similarly placed catalytic residues. The MutH protein is comprised of two major domains that are able to pivot and rotate with respect to one another. The cleft between the two domains is large enough for double-strand DNA to bind.

This research started with the determination of the MutH structure before it was known. After crystallizing the protein and collecting several heavy atom data sets, it was found that the electron density maps were too discontinuous to trace the structure of the protein. Following that work, site-directed mutagenesis was performed on several areas of MutH based on the similarity of MutH and PvuII structural models. The aims were to identify DNA binding residues (in two flexible loop regions), to determine if MutH has the same mechanism for DNA binding and catalysis as PvuII (MutH histidines 112 and 115), and to localize the residues responsible for MutH stimulation by MutL (MutH C-terminal tail region). An in-vivo screen based on the mutator phenotype was used to
select for functionally defective MutH mutants. These bacteria accumulate mutations at a
greater frequency than wild-type and this was monitored by selection on plates with rifampicin. Three MutH mutants were identified from this screen (K48A, G49A, and Δ214). They were purified and assayed for total activity and binding ability. Four other mutants with wild-type phenotypic screen results were also chosen to confirm they were not involved in any MutH function (D47A, H112A, H115A, and Δ224).

No DNA binding residues (such as D47A) were identified in the two flexible loop regions of MutH, although similar loops in PvuII are involved in DNA binding. The purified D47A MutH protein showed wild-type biochemical activity. Instead, the lysine residue (K48) in the first flexible loop was found to function in catalysis together with the three presumed catalytic amino acids (Asp70, Glu77, and Lys79). This purified MutH protein (K48A) had wild-type binding ability but no endonuclease activity without MutL. In the presence of MutL, the K48A protein had only a three-fold reduction in endonuclease activity. This research has shown that MutL stimulates the wild-type MutH activity by 1000-fold. The wild-type MutH stimulation by MutL for binding was only shown to be 16-fold. The G49A MutH mutant interferes with the proper functioning of the protein but is not informative about the mechanism of action. The binding ability of this mutant was the same as wild-type and the endonuclease activity was down 30-fold with a 10-fold stimulation by MutL. The extra methyl group of the alanine may cause slight structural changes in the lysine 48 side chain that slows catalysis.
The two histidines (H112 and H115) in MutH that are in a similar position as the two histidines (H84 and H85) in PvuII (that signal for DNA binding and catalysis) were changed to alanines, but had wild-type activity both in-vivo and in-vitro. These results indicate that the MutH signal for DNA binding and catalysis remains unknown.

The two deletion mutations (MutHΔ224 and MutHΔ214) in the C-terminal end of the protein, localized the MutL stimulation region to five amino acids (Ala220, Leu221, Leu222, Ala223, and Arg224). Mutant MutHΔ224 had wild-type MutL stimulation activity, while MutHΔ214 showed no MutL stimulation. Another deletion mutant, MutHΔ119, from another laboratory was shown to have wild-type MutL stimulation also. This leaves one (or more) of the remaining five residues as important for MutL stimulation.
Table of Contents

Title Page ................................................................. i
Signature Page ............................................................. ii
Acknowledgments ......................................................... iii
Abstract ................................................................. iv
Table of Contents ..................................................... viii
List of Tables ........................................................... x
List of Figures ......................................................... xi
CHAPTER I ...................................................................... 1
INTRODUCTION .......................................................... 1
Methyl-Directed Mismatch Repair ................................... 1
CHAPTER II ..................................................................... 18
CRYSTAL STRUCTURE OF MUTH ................................. 18
Introduction ............................................................... 18
Materials and Methods ............................................... 20
Results ................................................................. 27
CHAPTER III .......................................................... 47

IN-VIVO MUTATIONAL ANALYSIS OF MUTH ........................................ 47

Introduction ................................................................. 47
Materials and Methods .................................................. 49
Results ........................................................................... 57

CHAPTER IV .............................................................. 71

BIOCHEMICAL ANALYSIS OF MUTH .............................................. 71

Introduction ................................................................. 71
Materials and Methods .................................................. 73
Results ........................................................................... 83

CHAPTER V ................................................................. 106

DISCUSSION ............................................................... 106

Structural Analysis of MutH .............................................. 106
Mutational Analysis of MutH ............................................. 108

CHAPTER VI ................................................................. 118

REFERENCES .................................................................. 118
List of Tables

Chapter II
Table 2-1. Statistics of X-ray data reduction. .............................................................44
Table 2-2. Phases from heavy atom derivatives. ............................................................46

Chapter III
Table 3-1. In-vivo complementation screen (mutH471::Tn5) for mutant MutH proteins. .............................................................65
Table 3-2. In-vivo complementation screen (ΔmutH471::cat) for mutant MutH proteins. .............................................................67

Chapter IV
Table 4-1. Apparent equilibrium dissociation constants for wild-type and mutant MutH proteins. .............................................................105
List of Figures

Chapter I
Figure 1-1. Mismatch repair in *Escherichia coli* ............................................................ 6
Figure 1-2. Mismatch repair in human cells. ................................................................. 11

Chapter II
Figure 2-1. *Escherichia coli* *mutH* sequence and translation. .................................. 29
Figure 2-2. Sequence alignment of MutH from *H. influ.* and *E. coli.* ......................... 31
Figure 2-3. MutH overexpression time course. ............................................................... 34
Figure 2-4. Purification of native MutH protein................................................................. 36
Figure 2-5. Size estimation of the MutH protein in solution. .......................................... 38
Figure 2-6. MutH crystal ................................................................................................. 41
Figure 2-7. Diffraction pattern of MutH crystal. ............................................................... 43

Chapter III
Figure 3-1. Two stage PCR mutagenesis. ...................................................................... 53
Figure 3-2. Comparison of MutH and *PvuII* structures for the construction
of MutH mutants ............................................................................................................ 61
Figure 3-3. Western blotting of the wild-type and mutant MutH proteins. ...................... 70

Chapter IV
Figure 4-1. Activity of native MutH protein .................................................................. 86
Figure 4-2. Wild-type MutH activity .............................................................................. 88
Figure 4-3. Control experiments for the native MutH protein..........................90

Figure 4-4. Endonuclease activity of the wild-type (WT) and mutant
MutH proteins..................................................................................................................94

Figure 4-5. Comparison of wild-type and mutant MutH endonuclease activity.........96

Figure 4-6. Circular dichroism for wild-type and K48A MutH proteins................98

Figure 4-7. Binding abilities of the wild-type and mutant MutH proteins. ..........101

Figure 4-8. Dependence of complex formation on the concentration of the
MutH protein..................................................................................................................104

Chapter V

Figure 5-1. Location of the mutated residues in the MutH structure...................113
CHAPTER I

INTRODUCTION

Methyl-Directed Mismatch Repair

**History.** DNA mismatch repair is one process in the preservation of genomic integrity. Errors from DNA replication or damage are corrected by at least three systems in *Escherichia coli*\textsuperscript{142}. The first system is the 3'-5' exonucleolytic activity of the DNA polymerase III during replication which lowers error frequencies by $10^{-7}$ to $10^{-6}$\textsuperscript{48}. The second system has been called short-patch mismatch repair with excision of 1 to 10 nucleotides\textsuperscript{112}. The third system was called long-patch repair or the methyl-directed mismatch repair system, which causes a further reduction in error frequency by $10^{-2}$ to $10^{-3}$\textsuperscript{33}. These mechanisms bring the spontaneous mutation frequency in bacteria down to $< 10^{-9}$ per nucleotide per generation.

Studies of 5-bromouracil-induced mutagenesis in bacteria first led to the hypothesis of mismatch repair by Witkin in 1964\textsuperscript{167}. She postulated that the mutagenic effect was due to the misincorporation of 5-bromouracil opposite guanine followed by misrepair of the guanine to an adenine, thereby restoring normal base pairing (between
the 5-bromouracil and adenine). In that same year working on homologous recombination, Holliday postulated a similar idea to explain patterns of non-Mendelian segregation of genetic markers\textsuperscript{71}. His hypothesis stated that there were differential corrections of mismatches arising from the recombination intermediates. Starting in 1974, a series of genetic studies showed that a spontaneous mutator phenotype was associated with mutations that inactivated the DNA adenine methylase (\textit{dam}) gene of \textit{E. coli}\textsuperscript{105-107}. This DNA methyltransferase methylates the N6-position of adenine in d(GATC) sequences. Further genetic studies found an association between the \textit{dam}\textsuperscript{-} mutator phenotype and the \textit{mutS} and \textit{mutL} genes when \textit{dam} revertants or cells resistant to killing by DNA chemical modification were examined\textsuperscript{55,107}. By 1976, Wagner and Meselson suggested the existence of the methyl-directed mismatch repair pathway\textsuperscript{164}. They also suggested the possibility of strand discrimination based on the transient undermethylation of d(A) in d(GATC) sites within the newly synthesized \textit{E. coli} DNA. In this way, the mismatches in the new daughter strand would be excised and resynthesized, maintaining genomic fidelity.

The existence of a strand-specific mismatch repair system was established through subsequent experiments with constructed DNA heteroduplexes\textsuperscript{103,134}. These heteroduplexes containing a mismatch and defined states of adenine methylation in d(GATC) sites were transfected into \textit{E. coli}. Strand specificity of repair was checked through sensitivity/resistance to restriction endonucleases. Besides showing that repair had taken place, it was also shown that the level of DNA adenine methylase activity affected the ability of the repair system. Through this \textit{in vitro} mismatch repair assay, the
biochemical dissection of the mismatch repair pathway and the proteins involved in it were achieved\textsuperscript{102,152,164}. Three gene products (\textit{mutS}, \textit{mutL}, and \textit{mutH}) were shown to initiate the repair system and that the correction involves an excision tract of several thousand nucleotides, which are resynthesized. The repair system was successfully reconstituted \textit{in vitro} with purified proteins, which has yielded a comprehensive understanding of the repair mechanism\textsuperscript{90}.

During DNA synthesis, misincorporation of a non-complementary base in the newly synthesized DNA strand causes a base-base mismatch. Purine-pyrimidine mispairs generate transition mutations and pyrimidine-pyrimidine or purine-purine mispairs lead to transversions. Mispairs that are the most frequently generated by DNA polymerase III have the highest mismatch repair efficiency\textsuperscript{88}. Mismatches that are well corrected include G-T, A-C, G-G, and A-A. Moderately repaired substrates for the mismatch system are T-T, C-T, and G-A. The C-C transversion mismatch is poorly repaired by this system. In addition to base mispairs, the mismatch repair system also corrects small (1 to 4 nucleotides) insertions and deletions within the chromosome\textsuperscript{37,50,124}. Slippage of one DNA strand along the other gives rise to insertion/deletion errors. Misalignments, with insertions or deletions, produce frameshift mutations.

Base mispairs can also occur in heteroduplex DNA during recombination between similar but non-identical (homeologous) DNA sequences. There is evidence that the mismatch repair system processes homologous recombinational duplexes when erroneous base pairing occurs\textsuperscript{46}. It also appears to affect recombination between homeologous DNA
sequences and maintains a genetic barrier between species that are closely related.

Finally, base-pairing abnormalities can occur as a result of base damage. A consequence of chemical base modification, such as O6-methylguanine generated from alkylating agents, has been reported to be a substrate for the mismatch repair system\textsuperscript{22}. A mismatch repair system has been identified in other bacteria, yeast and mammalian cells. In \textit{Streptococcus pneumoniae}, this pathway is called the Hex system. It was found before the \textit{E. coli} system but until recently not as much information was available due to the absence of a chromosomal genetic map for \textit{S. pneumoniae} and a genetic complementation assay. Two genes have been identified (\textit{hexA} [=mutS] and \textit{hexB} [=mutL]) that function in this pathway\textsuperscript{12;130;131} and a strand break is thought to be the discriminatory signal\textsuperscript{64}. A complete model of this system has yet to be developed. In yeast and mammalian cells, this pathway is called the mismatch repair system. Proteins that have been identified as participating in the repair process have been named homologs of their corresponding \textit{E. coli} counterparts. A full model has not been completed for either one, but is patterned after the \textit{E. coli} system.

\textbf{Current model in \textit{Escherichia coli}.} The current model for methyl-directed mismatch repair in \textit{E. coli} is shown in Figure 1-1\textsuperscript{113}. The purified MutS protein exists in solution as oligomers of a 95 kD polypeptide\textsuperscript{153}. After a mismatch is formed, MutS efficiently and specifically recognizes and binds the mispaired bases\textsuperscript{124;153}. This interaction leads to a DNA substrate footprint of about 20 base pairs around the mismatch site.
Figure 1-1. **Mismatch repair in *Escherichia coli***. Panel A, the top half of the diagram, illustrates the initiation of mismatch repair. Mispaired bases may arise due to slippage of the DNA polymerase or misincorporation of a base during replication. Therefore, it is the newly synthesized daughter strand that needs correction. MutS recognizes the mispaired bases and binds the DNA at this point. MutL and MutH also bind to form the initiation complex. In the presence of ATP, DNA alpha loops are formed by the complex until the nearest d(GATC) site is reached. This site may be 5' or 3' to the mispaired bases. MutH recognizes the methylation at the d(A) in the d(GATC) site and nicks the unmethylated DNA strand (daughter strand) which transiently exists after replication. Panel B at the bottom half of the diagram, depicts the excision and resynthesis steps in the repair pathway. The left side shows a nick to the 5' side of the mispaired bases. After helicase II unwinds the two DNA strands, Exonuclease VII (ExoVII) or RecJ excise the single-strand daughter DNA. DNA polymerase III holoenzyme, in the presence of single-strand binding protein (SSB), resynthesizes the missing DNA strand. DNA ligase joins the last synthesized base to the DNA strand, completing the repair process. The Dam protein methylates the newly synthesized daughter strand, which then excludes corrections from the mismatch repair system. The right side of the diagram shows the nick to the 3' side of the mispaired bases. In this scenario, it is Exonuclease I that excises the single-strand daughter DNA. The subsequent steps in the process proceed as described for ExoVII and RecJ. A detailed explanation with references are provided in the text.
Figure 1-1.
The affinity of the protein for mismatches varies with the nature of the mismatch and correlates with the efficiency of correction. ATP hydrolysis is not required for MutS to bind to DNA even though MutS contains an ATP-binding site and exhibits a weak ATPase activity. Experimental evidence suggests that ATP hydrolysis is necessary for conformational changes in the MutS-DNA complex that allow the association of the MutL protein and the translocation of DNA through or along the recognition complex. The formation of α-shaped double-stranded DNA loops with MutS and MutL co-localizing at the base of the α-loop has been visualized through electron microscopy.

The purified MutL protein exists in solution as a 70 kD polypeptide. Until recently, no activity had been attributed to MutL. In the presence of ATP it was known that the MutS-MutL complex protects a DNA substrate region of about 100 base pairs which is much larger than the footprint with MutS alone. MutL has now been shown to be a helper protein that binds ATP. It serves to couple mismatch recognition to the downstream events. Besides forming a complex with MutS and enhancing translocation, MutL stimulates MutH endonuclease activity in an ATP-dependent manner. Finally, MutL is required to load UvrD (MutU/Helicase II) onto the DNA at the site of the MutH-induced nick.

MutH is a monomer of 25.5 kD in solution and is the final protein in the initiation complex for mismatch repair. It contains a latent Mg-dependent endonuclease.
activity that cuts the unmethylated strand of hemimethylated DNA 5' to the dG of d(GATC) sites producing 3'-OH and 5'-PO$_4$ termini$^{165}$. Further research has shown that a nick in the heteroduplex substrate can substitute for the d(GATC) sequence and MutH function, thereby defining the activity of MutH in the repair mechanism$^{92}$. Unmethylated DNA is nicked without any discrimination on one of the two strands and fully methylated DNA is resistant to cleavage by MutH$^{165}$. A single d(GATC) site is sufficient to initiate mismatch repair and may be a kilobase or more away from the mismatch, although repair is greatly reduced when the site exceeds two kilobases$^{24;24;31;31;92}$. Nicking of the unmethylated DNA strand leads to the excision process.

Helicase II (UvrD) is loaded onto the DNA at the site of the strand break$^{35}$. After displacement of the incised strand by UvrD, the repair reaction proceeds by exonucleolytic degradation from the GATC site towards and past the mismatch terminating at discrete sites$^{10;24;101}$. The GATC signal may reside on either side of the mismatch allowing for a bi-directional repair system$^{10;32;60}$. Distinct exonucleases are then required. Excision from the 5' side of the mismatch relies on RecJ exonuclease or exonuclease VII whereas excision from the 3' side relies on exonuclease I and exonuclease X$^{28;100;136;162;163}$. Each one of these identified exonucleases is specific for single-strand DNA except for exonuclease X, which also has the capability of degrading duplex DNA.

The last step of the process is gap repair. The polymerase responsible for the resynthesis of the daughter strand is DNA polymerase III holoenzyme$^{90}$. The
requirement for polymerase III holoenzyme is specific as DNA polymerase I, T7 DNA polymerase, and T4 DNA polymerase are not able to substitute for this polymerase\textsuperscript{147}. Single-strand binding protein (SSB) is also required at this stage of the reaction\textsuperscript{90}. DNA ligase completes the repair.

**Current model in eukaryotic cells.** Homologs of MutS and MutL have been found in Archeae, plants, yeast and mammals. *Saccharomyces cerevisiae* has been used more extensively to characterize the mismatch repair pathway in yeast, although it has also been studied in *Schizosaccharomyces pombe*\textsuperscript{51,140,143}. Mice lacking one or more of the mismatch DNA repair proteins (knockout mice) have provided the model system to study human cancer associated with mismatch repair defects\textsuperscript{11,36,132,138}. Extracts from human cells,\textsuperscript{156} as well as *Xenopus laevis* eggs\textsuperscript{21} and *Drosophila melanogaster*\textsuperscript{73} have been used to study mismatch correction in cell free systems. A current model of the mismatch repair system in human cells is shown in Figure 1-2.
Figure 1-2. **Mismatch repair model in human cells.** Panel A illustrates the repair pathway for misincorporated bases. The hMutSα complex (hMSH6 and hMSH2) recognizes the mispaired bases and binds the DNA at this point. The hMutLα complex (hMLH1 and hPMS1) binds to the hMSH proteins and may complete the initiation of mismatch repair. The reaction proceeds in the presence of ATP, proliferating cell nuclear antigen (PCNA), exonucleases, replication protein A (RPA), and an aphidicolin-sensitive DNA polymerase (α, δ, or ε). A DNA ligase has not yet been identified. Panel B illustrates the pathway for insertions and deletions that may occur during replication. The hMutSβ (hMSH3 and hMSH2) complex recognizes these types of mispaired bases and binds the DNA at the mispair. The hMutLα complex binds these hMSH proteins and the reaction proceeds as described above for misincorporated bases. A detailed explanation with references are provided in the text.
Figure 1-2.

A

B

ATP
PCNA
Excision
Resynthesis
Six MutS homologs (MSH) have been identified from *S. cerevisiae*\textsuperscript{72,110,137,148}. Msh1 protects against base mispairs and frameshifts in the mitochondrial genome\textsuperscript{30}. Msh4 and Msh5 have roles in meiotic recombination and do not appear to act following DNA replication\textsuperscript{139}. Msh2, Msh3, and Msh6 recognize mispairs in DNA and have the major role in mismatch repair\textsuperscript{5}. They form two heterodimers, MutS\(\alpha\) (Msh2 and Msh6) and MutS\(\beta\) (Msh2 and Msh3). MutS\(\alpha\) binds base-base mispairs and one base insertions/deletions while MutS\(\beta\) binds 2-4 base insertions/deletions\textsuperscript{86}. However, the binding properties of the purified Msh2 protein alone are different than those of the bacterial MutS protein. Msh2 binds with high affinity to heteroduplexes with 12-14 base inserts and moderate affinity to those with 8-12 base inserts in contrast to the bacterial protein which doesn’t appear to bind to inserts greater than 4 bases\textsuperscript{5}. Both Msh2 and Msh6 have intrinsic ATP binding and hydrolysis abilities supporting the hypothesis that they act in a similar manner as the bacterial MutS protein in mismatch repair\textsuperscript{6,150}. Functionally, there appears to be an asymmetry between Msh2 and Msh6. Mutations in Msh6 are more deleterious for MutS\(\alpha\) activity than the corresponding mutations in Msh2\textsuperscript{77,150}. In mammalian cells, no MSH1 has been identified yet. The other five MSH genes have been identified and the protein products were found to function similarly as those in the yeast. One distinction between the ability of the mammalian and the yeast Msh proteins is that the mammalian system supports an efficient repair of C-C mispairs\textsuperscript{22}. Multiple homologs of MutL have also been identified for the yeast and mammalian repair systems. In *S. cerevisiae* there are four MutL homologs (Mlh1, Mlh2,
MLH3, and Pms1). MLH2 does not function in mismatch repair. The remaining three proteins interact to form two heterodimers, MutLα (MLH1 and PMS1) and MutLβ (MLH1 and MLH3). The MutLα heterodimer provides the primary activity in mismatch repair with MutLβ providing a minor role. MutLα is able to interact with MutSα, MutSβ, and proliferating cell nuclear antigen (PCNA, a polymerase processivity factor), suggesting a complex formation as in *E. coli* mismatch repair. By analogy with bacteria, MutLα may recruit additional factors to the site of the repair and couple mismatch recognition and strand discrimination with excision and resynthesis. ATP binding and/or hydrolysis activity is predicted for MLH1 and PMS2 based on the highly conserved N termini with the bacterial MutL. In mammalian cells, the MutL homologs are MLH1, PMS1, and PMS2. These three proteins also form two heterodimers MutLα (MLH1 and PMS2) and MutLβ (MLH1 and PMS1). The two heterodimers function as in the yeast system.

There have been no MutH homologs found in eukaryotic cells. The mechanism for strand discrimination remains unknown at this time, although the evidence points toward breaks in the newly replicated strand. Relevant to the mechanism of strand discrimination is the implication of PCNA in mismatch repair. This polymerase processivity factor has been shown to be required for mismatch repair prior to the excision step. Mutations in PCNA were found to increase frameshifts in yeast through the disruption of mismatch repair and MSH2 mutations can suppress the lethality of the *pol30-104* PCNA allele. A model has been suggested that PCNA links the
mismatch repair complex to the polymerase at the replication fork, which facilitates the recognition and repair of the newly synthesized strand\textsuperscript{160}.

In yeast, there are three exonucleases that have been implicated in mismatch repair: exonuclease I (ExoI) a 5'→3' exonuclease, and the proofreading exonucleases of DNA polymerases δ and ε\textsuperscript{157;158}. In mammalian cells, a homolog of ExoI (EXO1/HEX1) that interacts with Msh2 has been identified as well as the two proofreading exonucleases of polymerases δ and ε\textsuperscript{135;144;157;158;166}. Resynthesis of the excised strand appears to proceed by an aphidicolin-sensitive DNA polymerase (α, δ, or ε), PCNA, and (in humans) replication protein A (RPA)\textsuperscript{45;73;114;156}. The DNA ligase that completes the reaction remains unknown.

Consequences of deficiencies in the repair system. There are several ways that a mutator gene product may act. The loss of mutS, mutL, or the mutH gene causes a loss in the capacity to repair mismatches that normally occur following DNA replication in E. coli\textsuperscript{55}. This results in a large increase in spontaneous mutability. The mut genes therefore allow a switch between high and low mutation rates depending on the adversity of the environment\textsuperscript{57}. Inactivation of the mismatch repair system results in a decreased sensitivity to the effects of alkylating agents and other base analog mismatches\textsuperscript{17;80;82}. It has also been found that a few of the pathological Salmonella typhimurium and E. coli species have mutations in the mut genes\textsuperscript{95}. The more common mutation among these species is seen in the mutS gene and, only rarely, in the mutH gene\textsuperscript{95}. 
In mammalian cells, the inactivation of mismatch repair has been shown to lead to the development of tumors, methylation-tolerance, and resistance to a wide spectrum of DNA damaging agents. There are several methylation-tolerant mammalian cell lines. The MT1 cell line, derived from human lymphoblastoid TK6 cells, is deficient in hMutSα activity and is a hundred times more resistant to N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) than the parental line. These cells are also hypermutable in the absence of MNNG resulting in mainly A→G transitions and single nucleotide frameshifts. Loss of hMlh1 activity is also associated with methylation tolerance. HCT116 cells are defective in both alleles of hMLH1 and are resistant to killing by MNNG. Resistance to cis-Platin (cis-diaminedichloroplatinum), which reacts with DNA to form intrastrand and interstrand cross-links, monoadducts, and protein-DNA adducts, is a clinically relevant example of the consequences when mismatch repair is inactivated. Cis-Platin is in use as a chemotherapeutic agent, especially with ovarian and testicular tumors. Treatment of the tumors with cisplatin leads to the development of resistance at the same time as the development of mismatch repair deficiencies. The tumors that are the most responsive to cisplatin are those that overexpress hMsh2. Biochemical assays indicate that the human mismatch proteins (hMsh2 and hMutSα) can interact with cisplatin-DNA adducts. Cis-platin adducts though, are mainly repaired by the NER (nucleotide excision repair) system. The accumulation of these DNA adducts block replication, trigger G2 cell-cycle arrest, and lead to apoptosis. One hypothesis for the cell death is that the mismatch repair system
physically competes with the NER system and excludes it from becoming active\textsuperscript{23,34,76,129}. More studies will have to be done to address this issue.

Finally, deficiencies in mismatch repair have been shown to be involved in human cancer. Frequent alterations in the length of microsatellites (microsatellite instability or MSI) were discovered in sporadic and hereditary colorectal cancers\textsuperscript{1,125,128,155}. Microsatellites are repeating tracts of mono-, di-, tri-, or tetrarnucleotides and are susceptible to frameshift mutations due to slipped mispairing during replication. In yeast and bacteria, failure to correct these mispairs through the mismatch repair system results in elevations of MSI\textsuperscript{96,149}. Hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome is a cancer predisposition syndrome that is characterized by the frequent mutations in repeat sequences (A), (GGC), or (CA)\textsuperscript{n}. Identification of defects in mismatch repair genes from hereditary nonpolyposis colorectal cancer cells, which have elevated MSI, confirmed that inactivation of the mismatch repair system in mammalian cells could result in tumors\textsuperscript{20,49,94,123,126}. Since the identification of microsatellite instability in HPNCC tumors, MSI has been found for a significant number of other sporadic tumors such as endometrial, stomach, ovarian, cervical, pancreatic, squamous cell skin, small-cell lung, and prostatic tumors\textsuperscript{43,44,159}. Mutations in five mismatch repair genes, MSH2, MSH6, MLH1, PMS1, and PMS2, have been associated with HNPCC\textsuperscript{4,98,111,127}. The majority of clinical cases have mutations in MSH2 and MLH1 while the other mutations are rare\textsuperscript{98,161}. This association between inactivation of the mismatch repair pathway and human cancer has supported a hypothesis that a mutator phenotype can promote tumorigenesis\textsuperscript{99}. 
The mismatch repair system is highly conserved among species and the main function appears to be mutation avoidance through the strand-specific elimination of base-base mispairs, chemical base modifications, as well as short insertion/deletion loops following DNA replication. The repair system also has important effects on homeologous recombination, contributing to the frequency of reciprocal exchanges.
CHAPTER II

CRYSTAL STRUCTURE OF MUTH

Introduction

A structural model of MutH was required for rational mutagenesis. At the time, no structure of MutH was available. There were also no other homologous mutH sequences known. MutH has no amino acid sequence similarity to restriction endonucleases, so structural or functional predictions could not be made. Random mutagenesis to determine functionally important amino acids had not generated any candidate mutants through our in vivo screen (described in chapter 3). The MutH structure, when known, could be compared to restriction endonuclease structures to determine functional areas within the MutH protein. This would generate a list of residues to be targeted for site-directed mutagenesis. Those mutants (as purified proteins) would be used in biochemical assays to determine the residues' role in MutH function.

In order to determine the structure of MutH by X-ray crystallography, some procedures needed to be optimized. The published overexpression of MutH did not
produce enough protein to grow crystals\textsuperscript{165}. Different vectors and strains for the expression were looked at in combination for maximum production of the MutH protein. Also, the chromosomal copy of \textit{mutH} present in the lab on plasmid pRH71-17, (obtained from Dr. Robert Grafstrom, Frederick Cancer Research Facility, Frederick, MD) was sequenced to determine if it was correct. This copy of \textit{mutH} was identified from a specialized \textlambda{} transducing phage (\textlambda{}pthyA)\textsuperscript{145} and cloned into a pBR322 plasmid (pRH71)\textsuperscript{58}. 
Materials and Methods

Strains, plasmids, and media. TG1 host cells (Amersham Pharmacia Biotech) were used for the recovery of the sequencing plasmids and single-strand DNA preparation. Strain BL21 (DE3) pLysS\(^{151}\) was used for the overexpression of the MutH protein. RF M13mp18 and M13mp19 (New England BioLabs, Inc.) backbones were used to construct the plasmids for sequencing\(^{109}\). The 1 kb DNA fragment containing \textit{mutH} was removed from plasmid pRH71-17\(^{59}\) by \textit{XhoI} and \textit{HindIII} restriction digest and ligated to the \textit{SalI} and \textit{HindIII} digested RF M13 backbone. The pTL2 plasmid used for overexpression of the MutH protein was constructed from a three-piece ligation, with two fragments of \textit{mutH} removed from pRH71-17 by restriction digest with \textit{AflII} and \textit{BamHI} and a pET3d (Novagen) backbone cut with the restriction enzymes \textit{NcoI} and \textit{BamHI}. Luria-Bertani (LB) medium was made according to Sambrook et. al\(^{141}\). Ampicillin (Amp) and chloramphenicol (Cam) were added to media at 100 \text{ug/ml} and 30 \text{ug/ml} respectively. Isopropyl-β-D-thiogalactoside (IPTG) was made as a 1 M stock solution in water.

Single-strand DNA preparation. The single-stranded DNA for sequencing was prepared according to the protocol from the Oligo-directed mutagenesis version 2.0 kit (Amersham). Plaques containing the M13 phage inserted with the \textit{mutH} sequence were prepared from TG1 cells (from the kit). A single TG1 colony picked from a glucose/minimal medium plate was grown overnight in 10 ml of LB medium, shaken at
37° C. 50 ul of the overnight culture was added to 5 ml of fresh medium and a single plaque was transferred as a plug into the medium. This was incubated with shaking at 37° C for 5 hours, poured into five 1.5 ml microcentrifuge tubes, centrifuged for 5 minutes at room temperature, and the supernatant transferred to five fresh tubes. 200 ul of PEG/NaCl (20 % w/v polyethylene glycol 6000, 2.5 mM sodium chloride) was added to the supernatant, shaken and left standing for 15 minutes at room temperature, centrifuged for 5 minutes in a microfuge, and the supernatant discarded. The pellet was centrifuged again for 2 minutes and all traces of the PEG were removed with a Pasteur pipet. 100 ul of TE buffer was added to each tube and vortexed for 30 seconds to resuspend the pellet. The ss DNA was extracted with 50 ul of phenol, centrifuged for 3 minutes, the aqueous layer transferred to a clean tube, extracted again with 500 ul of ether (which was removed as the top layer) and precipitated with 10 ul of 3 M sodium acetate and 250 ul of ethanol. This was incubated at -80° C for 10 minutes, centrifuged for 15 minutes, washed with 500 ul of cold (4° C) ethanol, and dried at room temperature. The ss DNA was resuspended in TE buffer with a total volume of 100 ul. The concentration of the ss DNA was determined by the OD at 260 nm (1 OD = 40 ug/ml).

Overexpression and purification of native MutH. A 10 ml LB-Amp culture was grown for 8 hours at 37° C from a freshly transformed single colony. A 400 ml LB-Amp culture was inoculated with the 10 ml culture and grown overnight. Twelve liters of LB-Amp-Cam medium was inoculated with the 400 ml overnight culture and allowed to grow with aeration at 37° C to an OD600 of 1.0. The culture was induced by the addition of IPTG to a final concentration of 0.1 mM and the culture was allowed to grow for 2
hours at 37\(^0\) C. Cells were harvested, concentrated with a Pellicon 300,000 filter, centrifuged for 20 minutes at 5,000 rpm, and stored frozen (-80\(^0\) C). The MutH protein was purified as described\(^{165}\). The frozen pellet (15 g) was thawed on ice, resuspended in 40 ml of 50 mM KPO\(_4\) (pH7.4), 0.1 mM EDTA, 0.5 mM DTT, and 1 mM phenylmethylsulfonyl floride (PMSF), disrupted by sonication with 15 30-second bursts, centrifuged at 13,000 x g for 1 hour, and the supernatant (Fraction I) was transferred to a clean tube. 1.6 ml of 3 M KCl was added to the supernatant for a final concentration of 0.1 M KCl. This solution was treated with 12 ml of 25 % w/v streptomycin sulfate, incubated on ice for 30 minutes, and centrifuged for 10 minutes at 13,000 x g. The supernatant (Fraction II) was treated with 15.7 g of solid ammonium sulfate while being stirred, incubated for 30 minutes at 4\(^\circ\) C, and centrifuged for 10 minutes as above. The pellet was resuspended in 10 ml of 20 mM KPO\(_4\) (pH 7.4), 100 mM KCl, 0.1 mM EDTA, 0.5 mM DTT, and 1 mM PMSF and dialyzed against two changes of 500 ml of this buffer (Fraction III). Fraction III was diluted by the addition of 64 mL of 39 mM KPO\(_4\) (pH7.4), 38 mM KCl, 0.1 mM EDTA, 0.5 mM DTT, 1 mM PMSF and applied to a column of Bio-Rex 70 (BIO-RAD) (200 ml) equilibrated with 35 mM KPO\(_4\) (pH7.4), 50 mM KCl, and 0.1 mM EDTA. After washing with 150 ml of equilibration buffer containing 0.5 mM DTT and 1 mM PMSF, the column was eluted with a linear gradient of KCl (1000 ml, 50-600 mM) in wash buffer. Fractions containing MutH (followed by absorbance at 280 nm) were pooled (Fraction IV) and loaded onto a hydroxylapatite Bio-Gel HTP column (BIO-RAD) (100 ml) equilibrated with 35 mM KPO\(_4\) (pH7.4), 200 mM KCl and washed with 100 ml of the equilibrium buffer containing 0.5 mM DTT and 1
mM PMSF. MutH, which passes through the column, was then dialyzed against two changes of 20 mM KPO_4 (pH 7.4), 50 mM KCl, 0.1 mM EDTA, 0.5 mM DTT, and 1 mM PMSF (Fraction V). Fraction V was loaded onto a 40 ml heparin agarose (BIO-RAD) column equilibrated and washed (40 ml) with the dialysis buffer. MutH was eluted with a linear gradient of KCl (100 ml, 50-400 mM) in dialysis buffer. Fractions containing MutH (followed by absorbance at 280 nm) were pooled (Fraction VI), frozen in 1 ml aliquots, and stored at -80°C.

**Sequencing.** The native wild-type mutH gene was sequenced using the T7 Sequenase kit (Amersham) by the dideoxy method. The gels were dried, put on Kodak X-ray film overnight, and the film developed in an Xomat processor (Kodak). The autoradiographs were read manually.

**MutH size estimation.** To determine the state of the purified MutH protein in solution, a HiLoad Superdex 26/60 (Pharmacia LKB Biotechnology) column was run according to the manufacturers' instructions. The calibration standards at 10 mg/ml were aldolase from rabbit muscle (M.W. 160,000), Glucose-6-Phosphate dehydrogenase from yeast tortula (M.W. 130,000), bovine serum albumin (M.W. 80,000), ovalbumin (M.W. 43,000), β-lactoglobulin (M.W. 35,000), and myoglobin (M.W. 17,600). A standard curve was calculated (log of M.W. versus retention time) and the molecular weight of the MutH protein in solution was estimated.
Crystallization. The first crystallization trials involved a systematic screen of 50 solutions. One set of sitting drop solutions was done with the MutH protein in Tris-HCl (pH 7.6) and one set of sitting drops was done with the protein in KPO₄ (pH 7.4). One pair (of these sets of solutions) was incubated at 4°C and another pair were incubated at room temperature. The second crystallization trials involved 48 solutions from Crystal Screen II (Hampton Research), set up in the same manner as the earlier screens. MutH crystals of x-ray diffractable quality did not come from the trial solutions, but rather as the protein was being concentrated in a Centricon tube (Amicon) following the last column (with the KCl concentration falling from 400 mM to 123 mM (final)). The concentration buffer was 20 mM KPO₄ (pH 7.4), 0.1 mM EDTA, 1 mM DTT, and 123 mM KCl.

Heavy atom derivatives. The stabilization buffer for the MutH crystals was found to be the concentration buffer with the addition of 100 mM KCl. After the crystals were in the stabilization buffer for 48 hours, stock solutions of the heavy atom compounds were added. Each heavy atom was initially screened at 3 mM and 10 mM and the crystal integrity was checked 24 hours later. The heavy atom concentration was optimized to be the highest concentration that maintained crystal integrity after 24 hours. The heavy atom compounds (and optimized concentrations) that were tried are listed below:

- Platinum diamine dichloride (10 mM)
- P-hydroxymercuri benzoic acid (5 mM)
- Samarium sulfate (1.9 mM)
- Potassium tetrachloroplatinate (II) (7.5 mM)
- Eurolupium trichloride (1 mM)
- Gadolinium chloride (3.75 mM)
- Holmium chloride (3.75 mM)
- Trimethyl lead acetate (3.75 mM)
Mercury chloride (6 mM)  
Gold chloride (5 mM)  
Ethylmercurithiosalicylic acid, Na salt (3.75 mM)

**X-ray Data collection, reduction and electron density maps.** Diffraction data were collected at -150°C using an R-AXIS IIC imaging plate system mounted on a Rigaku RU 200 generator (Molecular Structure Corp., The Woodlands, TX). Crystals were put into a solution containing 30% Polyethylene glycol (PEG) 400, 15% glycerol and mother liquor for 10 minutes before being picked up and put directly into the nitrogen stream. The native and gold chloride datasets were each collected from one crystal using the oscillation method. The crystal was rotated a total of 120° along the φ angle, with each imaging frame containing data from a Δφ range of 2° and exposed to X-rays for 10 minutes. Diffraction data was observed with the native and gold derivative to 2.0 and 3.5 Å resolution respectively. The MutH/Hg⁺ crystal x-ray data was collected at wavelengths of 1.01 and 0.97 and were measured at the National Synchrotron Light Source at Brookhaven National Laboratory on beamline X12C using the MarResearch imaging plate system at a distance of 199.74 mm. The dataset was collected at -180°C from one crystal using the oscillation method. The crystal was rotated a total of 120° along the φ angle, with each imaging frame containing data from a Δφ range of 2° which was exposed to X-rays for 2 minutes. Both heavy atom crystal unit cell dimensions were isomorphous to the native. The collected data were indexed, integrated, and scaled using the programs DENZO and SCALEPACK. CCP4 programs were used for data reduction, phasing, and electron density map calculations. One Hg site and three Au
sites were found using Patterson maps and refined using MLPHARE (CCP4, 1994).
Upon cross-checking (using difference fourier maps) six minor Au sites were found in
addition to the first three sites. An electron density map, calculated to 3.0 Å after solvent
flattening, was used to trace the peptide chain in O81.
Results

**New sequence.** Single strand DNA was used for sequencing the native mutH gene due to the inability to read the sequence from a double-stranded template. Both strands of the gene were sequenced and compared to the sequence in GenBank (Accession number Y00113). Six nucleotide changes were found that resulted in four amino acid changes within the protein (Fig. 2-1). Nucleotide 247 changed from a d(C) to a d(G) resulting in leucine 83 changing to a valine. Nucleotide 390 changed from a d(A) to a d(C) causing arginine 130 to change to a serine. Nucleotide 391 changed from a d(C) to a d(A) with the leucine 131 changing to an isoleucine. Nucleotides 406 and 407 changed from a d(G) to a d(C) and a d(C) to a d(G) respectively changing only one amino acid, an alanine 136 to an arginine. Nucleotide 457 changed from a d(C) to a d(G) with no amino acid changes. These differences were confirmed by a second independent sequencing of the gene from the E. coli chromosome by Dr. M Winkler (University of Texas Medical School, Houston TX). The revised sequence was deposited with GenBank, accession number U16361. This sequence also does not agree with that of the E. coli genome project. The MutH protein from Haemophilus influenzae has also been sequenced and both proteins are well conserved with a sequence identity of 59% and a similarity of 76% (Fig. 2-2). Along with the structure of the E. coli MutH protein, my sequence appears to be correct and conserved areas between the two MutH proteins could be identified for site-directed mutagenesis.
Figure 2-1. *Escherichia coli* mutH sequence and translation.

The full, corrected nucleotide sequence of mutH is shown with the amino acid translation shown beneath. The changes from the original sequence are shown in bold. The revised sequence was deposited in GenBank, accession number U16361.

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>347 (C to G)</td>
<td>leu83 to val</td>
</tr>
<tr>
<td>490 (A to C)</td>
<td>arg130 to ser</td>
</tr>
<tr>
<td>491 (C to A)</td>
<td>leu131 to ile</td>
</tr>
<tr>
<td>506 (G to C)</td>
<td>ala136 to arg</td>
</tr>
<tr>
<td>507 (C to G)</td>
<td>ala136 to arg</td>
</tr>
<tr>
<td>556 (C to G)</td>
<td>None</td>
</tr>
</tbody>
</table>
Figure 2-1.

ATGTCCCAACCTGCACACTGCTCTCTCTCTCCGAAACTGAAGAACAGTTGTTAGCGCAA
1 60
TACAGGGTTGGAGCGGGTGACGAGAGAGGAGGGCTTTGACTTCTTGTCAACAA
121 180
GCGTGTCGTTGAAGACCAA TATGTAACCCTCTTGACCGCCGTGAACAGCCCGACCAA
181 240
CGTGTCGCTCCCTCTTGGACCTCTCTTGAATTTTGA
241 300
GGTCTCTTAATTTTCGCTATTTCCGACCTAACCGCACAATGACCTCTAGACCGATCCA
301 360
GGTCCCGTTA TCGCCCCACTGGACCCTTTGGTCGGTGCACGCGGTGTTTGAGTTTGCGCA
361 420
CCGCTTGCCTAAGACTGCGACGGTGCGCCGAA TAACTTCTTCTT
421 480
GACACCTATGGCCAACTTCCGCTCGCGTCGTAGGGCGACCGCGTCGCAGCGCAACCTAGT
481 540
CCGCTTGCCTAAGACTGCGACGGTGCGCCGAA TAACTTCTTCTT
541 600

AA TTACCTA TACTAACAAGAGCCAGTCCAACTCGCCTAATGGCGAGCAGTGCCCCTCA T A
581
AA TGTCTA TGCTGGCTTTCGTCGCTT ACGCTTTCGCGAA TGGCTTCGGTAACCACGGGCC
601

CCGGAAGATTAAAAAGCAGATAAAGGCTGATTGAGTTGATAGTTGAGATTGATAGTTGAGA
641

CGCCTCGCTAAGACTGCCAGGCGCTCGCGTCGTAGGGCGACCGCGTCGCAGCGCAACCTAGT
681

GATGACCGGCGCAGTAAAGAGCTAGGCTAC
721

LL AR HF L I Q *
Figure 2-2. Sequence alignment of MutH from *H. infl.* and *E. coli.*

The top row is the MutH sequence from *Haemophilus influenzae* and the bottom row is from *Escherichia coli.* The "|" represents identical amino acids and the ":" represents similar ones. There is 59% identity between the two sequences and 76% similarity.
Figure 2-2.

H. infl. MIPQTLEQLSGAQSIAGLTFGELADELHIPVPIDLKRDHGIG 44
1 MSQPRPLLSPETEBQQLAQQLSGLAALVGLVTPENLKRDKGIG 52

45 MLLEALIAGATAGSKEQDFTGLGVEKTLPLNAEGPGLSFFEFVSLAQLVQNS 96
53 VLLLELWAGSGKPEQDFAPALGVEKTPDVLKPELETTFVCVAPLTQNS 104

97 GVKEWNSHVRHKLCVLWMPDGSRIHIGAPIFKPTABGPOQLK 147
105 GVTEWTSVHRKHKLRLMTPVGERSIPLAQKQVRGSLPLWSNEEDQRKL 155

148 QDWEELMDLIVLGKLDQITARIGEVMQLRPKANSRAVTKGICNEIDTE 198
156 EDWEELMDIVLQVERITARHGYIQIRPKAANAKALTEAILGERILIT 206

199 LPLGFYLKKEPTAQILNAFLKTSL 223
207 LPRGFYLTNSALLARHFLIQ.. 229
Purified MutH protein. The plasmid (pTL2) constructed for the overexpression of MutH was tested for the amount of protein production. A time course was done to determine the optimum time for induction of the protein. A 100 mL culture was grown and induced with IPTG in the same manner as the 12 L culture described in the Materials and Methods. Aliquots of 10 mLs taken at various time points show that a 2½ hour induction contains the most protein (Fig. 2-3). A 12 L culture takes approximately 1/2 hour to concentrate, therefore a 2 hour induction was chosen for the larger preparation. Aliquots were saved after each step in the purification of MutH and electrophoresed. The resulting gel follows the protein as it is purified (Fig 2-4). The 12 L culture yielded 800 mg of purified native MutH protein.

Monomeric MutH in solution. A sizing column was utilized in order to determine the state of the purified protein in solution. Proteins of known molecular weight were used to establish a standard curve and the approximate molecular weight of MutH taken from the curve. The relative elution volume of MutH from the sizing column (42.3) gave an apparent molecular weight of 32 kD (Fig. 2-5). The molecular weight of MutH calculated from the amino acid sequence is 25.5 kD. Therefore, MutH exists as a monomer in the purified solution.
Figure 2-3. **MutH overexpression time course.**

Plasmid pTL2 in strain BL21 (DE3) plysS was overexpressed with IPTG. 10 ml aliquots were taken at the time points after the addition of IPTG indicated above the gel. The crude extracts were lysed, centrifuged, and electrophoresed (as described in the Materials and Methods) in a 12% SDS-PAGE gel to determine the amount of time for maximum protein production. The 2½ hour time point shows maximum protein production. The 2 hour time point was chosen because concentration for the 12 L fermentation takes ½ hour.
Figure 2-3.
At the end of each step in the purification process, an aliquot of solution was taken for electrophoresis to demonstrate that the protein was being retained and purified. The total protein concentration of each aliquot was measured using the Bradford assay, therefore, each lane in the 12% SDS-PAGE gel has the same total protein concentration. The steps from which the aliquots were saved are labelled at the top of each lane in the figure. Molecular weight size markers are indicated to the right of the gel. The MutH protein is indicated by an arrow on the left side of the gel. The gel was stained with Coomassie Brilliant Blue.
Figure 2-4.

- W. Standards
- Cell Extract
- Streptomycin sulfate
- Ammonium sulfate
- Bio-Rex 70
- Hydroxylapatite
- Heparin agarose

M.W. Standards

- 45,000
- 31,000
- 21,500
- 14,400

MutH
A Superdex column was used to separate the proteins by size. Proteins of known molecular weight were used to establish the standard curve (listed in Materials and Methods). The relative elution volume of MutH was 42.3, indicated by an arrow in the graph. This indicates that the MutH protein in solution has a molecular weight of 32 kD.
Figure 2-5.
MutH Crystals. The crystals grown were rectangular plates in shape, approximately 0.5 to 1.0 mm along the long axis, and 0.2 to 0.5 mm along the short axis (Fig. 2-6). They were always thin crystals of approximately 0.1 mm. The crystal solvent content was quite low at 37%, therefore, they were very stable. A typical diffraction pattern for MutH is shown in Figure 2-7. This figure contains all of the reflections from a 2° oscillation out to 2 Å.

Statistics of X-ray data. The data reduction of the native and heavy atom derivatives (Hg and Au) used for generating the phases of the MutH protein structure are shown in Table 2-1. The space group for the all of the MutH crystals was found to be P2_12_12_1. The unit cell parameters of the heavy atom derivative MutH crystals were isomorphous to the native MutH crystal. To allow for better detection of the differences due to the presence of just one mercury atom in the unit cell, the data was collected at two wavelengths (before and at the L-III edge of mercury).

Patterson maps. In the space group P2_12_12_1, the MutH protein exists as a monomer in the asymmetric unit. It contains one cysteine residue, and one mercury (Hg) site was found using Patterson maps. Patterson maps were also used to find the three major gold (Au) sites. The sites from the two derivatives were checked by using the phases from one derivative in a difference Fourier to find the second derivative's site(s) in FFT (CCP4, 1994). Six minor sites were found for the Au derivative from the difference Fourier and used in the electron map calculation.
Figure 2-6. **MutH crystal.**

This MutH crystal is rectangular in shape, approximately 0.50 x 0.25 x 0.10 mm. It was grown by concentrating a 2 mg/ml solution of the purified protein in a Centricon unit. A similar intact crystal was mounted with the long axis of the crystal at 90° to the incident X-ray beam to collect the native data set. This crystal was accidentally gouged and split during manipulation with the top moving to one side.
Figure 2-6.
Figure 2-7. **Diffraction pattern for MutH crystal.**

The reflections seen in this figure are from one of the forty $2^\circ$ oscillations collected for the native data set. Cu Kα x-rays (1.54 Å) were provided by a RU200 Rigaku generator set at 50 V, 100 mAmp. X-ray diffraction data was collected on an R-AXIS II imaging plate system placed at 100 mm from the crystal. The outermost reflections represent a resolution of 2.0 Å.
Figure 2-7.
Table 2-1. Statistics of X-ray data reduction.

<table>
<thead>
<tr>
<th>Space group</th>
<th>P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal content</td>
<td>Native</td>
</tr>
<tr>
<td>Soaking time and concentrations</td>
<td>24 hrs., 6 mM</td>
</tr>
<tr>
<td>Unit cell (a, b, c)</td>
<td>42.6, 73.6, 74.4</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>64,087</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>16,709</td>
</tr>
<tr>
<td>⟨I/σ⟩</td>
<td>17.2</td>
</tr>
<tr>
<td>Diffraction data</td>
<td>50.0-2.0 Å</td>
</tr>
<tr>
<td>Last resolution bin</td>
<td>2.07-2.00 Å</td>
</tr>
<tr>
<td>Completeness (%)\textsuperscript{a}</td>
<td>94.0 (79.8)</td>
</tr>
<tr>
<td>R\textsubscript{merge} (%) \textsuperscript{a,b}</td>
<td>5.4 (17.1)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The data from the last resolution bin is shown in parentheses.
\textsuperscript{b}R\textsubscript{merge} = Σ |I - ⟨I⟩| / Σ ⟨I⟩, where I is the measured intensity of each reflection and ⟨I⟩ is the intensity averaged from multiple observations of symmetry related reflections.
Phases and refinement. MLPHARE (CCP4, 1994) was used to refine all of the heavy atom sites found in the two derivatives. The $R_{iso}$ and phasing power for each derivative are shown in Table 2-2. The anomalous signal from the gold derivative, upon refinement, gave average anomalous differences divided by the average closure errors of 0.03. Since this anomalous signal had no phasing power, it was not included in the final refinement to calculate electron density maps.

Electron density maps in O$^{81}$. The initial Fourier map calculated at 3 Å after solvent flattening in DM (CCP4, 1994) revealed discontinuous electron density. Approximately 40 % of the main chain atoms could be traced in O. No problems could be discovered after a reexamination of the calculations and refinement. A few months later, the structure was published by Changill Ban and Wei Yang$^{14}$. 
Table 2-2. Phases from heavy atom derivatives.

<table>
<thead>
<tr>
<th></th>
<th>Hg derivative</th>
<th>Au derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{iso}$ ($F$)$^a$</td>
<td>24.9 (50.0-5.0 Å)</td>
<td>53.6% (50.0-3.51 Å)</td>
</tr>
<tr>
<td>$R_{anom}$$^b$</td>
<td>4.85 (50.0-5.0 Å)</td>
<td>8.17 (50.0-3.51 Å)</td>
</tr>
<tr>
<td>Phasing power$^c$ (isomorphous, anomalous)</td>
<td>1.50 (16.5-2.90 Å)</td>
<td>1.62 (12.58-3.50 Å)</td>
</tr>
<tr>
<td>Figure-of-merit</td>
<td>0.83 (16.5-2.90 Å)</td>
<td>0.44 (12.58-3.50 Å)</td>
</tr>
</tbody>
</table>

$aR_{iso} = \Sigma \left| F_p \right| - \left| F_{ph} \right| / \Sigma \left| F_p \right|.$

$bR_{anom} = \Sigma \left| F_+ - F_- \right| / \Sigma (F_+ + F_-) / 2.$

$c$Phasing power $= \langle \left| F_h \right| \rangle / \epsilon,$ i.e. the mean amplitudes calculated from the heavy-atom model divided by the lack of closure error.
CHAPTER III

IN-VIVO MUTATIONAL ANALYSIS OF MUTH

Introduction

Complementation and dominant negative phenotype. Specific genes that accelerate mutation rates are called mutator \((mut)\) genes\(^{31,146}\). The mutator phenotype occurs when cells lack one or more of the Mut proteins\(^{168}\). The increase in mutation frequency can be monitored by growth on plates containing an antibiotic such as rifampicin or nalidixic acid. Complementation of a recessive allele occurs when the wild-type protein is produced from the chromosome or a plasmid in the \(mut^-\) strain and the mutation frequency decreases to the level of a wild-type strain. A dominant-negative phenotype is apparent with a mutant \(mut\) coding plasmid in a wild-type strain. When the mutant protein is produced at a high level from the plasmid, the mutation frequency of the cell population increases, even though the cells produce a low level of wild-type protein. This is generally found with proteins that form multimers as an active complex. When one of the monomers is defective then the whole complex is usually non-functional.
Plasmid selection for mutagenesis and screening. Finding one *mutH* plasmid to perform the complementation and dominant-negative screening as well as easily produce mutant proteins proved to be difficult. A construct of *mutH* with a N-terminal histidine tag (pTX417) was obtained from Dr. M. Winkler (Univ. of Texas Medical School, Houston TX). The MutH protein could now be purified in one day through commercially available resins. Control of the MutH gene was important for both the screening and the overproduction of the protein. The abundance of nicks generated in DNA from the overexpressed MutH, causes the cells to die or lose the *mutH* gene. Regulation was tried through the pUH (University of Heidelberg) vector\textsuperscript{93}, T7 (pET3d, Novagen), and \( P_{\text{lac}} \) (pSK\textsuperscript{+}, Stratagene) promoters. These constructs were unusable because they produced too much protein without induction. Finally a pBAD\textsuperscript{18}\textsuperscript{65} plasmid was tested. The His-tagged *mutH* gene from pTX417 was ligated into the polylinker region of the pBAD18 plasmid to give pTL7-R and tested for complementation. The *mutH* strain was healthy with this plasmid and was complemented without addition of the inducer, arabinose. As this was the best result obtained from screening all the promoters tested, the pTL7-R construct of *mutH* was used for the screening and overexpression of the mutant MutH proteins.
Materials and Methods

Strains, plasmids and media. Isogenic strains GM1166 (*hsdR17 endA1 thi-1 spoT1 rfbD1 supE44*) and GM3856 (GM1166 *mutH471::Tn5*) were used for the in-vivo screen. Strain CC106 (GM4244) was used for the recovery of the mutant plasmids after site-directed mutagenesis. Strain GM7586 (GM4244 *ΔmutH471::cat*) was constructed through P1 transduction and is otherwise isogenic to strain GM4244. Plasmid pTL7-R is a pBAD18 derivative containing the *mutH* gene with an N-terminal polyhistidine sequence from the plasmid pTX417. The approximately 1 kb *mutH* fragment was obtained by digestion of pTX417 with the restriction enzymes *XbaI* and *HindIII*. The fragment also contains several hundred base pairs of chromosomal DNA distal to the *mutH* gene. pTX417 was kindly provided by Dr. M. Winkler (University of Texas Medical School, Houston TX).

The Luria-Bertani (LB) medium was made with 10 gm of tryptone, 5 gm of yeast extract, 5 gm of NaCl, and 1 ml of 1 M sodium hydroxide per liter of water. Ampicillin (Amp) and rifampicin (Rif) were added to media at 100 ug/ml.

Site-directed mutagenesis. Two stage PCR-mediated mutagenesis was carried out according to Dr. K. Murphy (personal communication) with minor modifications (Figure 3-1). For each mutation, two oligonucleotides of approximately 30 bases (one for each DNA strand) with the changed bases for the new amino acid sequence were
designed. Two oligonucleotides, one for each end of the \textit{mutH} sequence and designated outside flanking primers, were also made. The forward flanking primer contained a \textit{NcoI} site and the reverse flanking primer contained a \textit{HindIII} site. The oligonucleotides were manufactured by Operon Technologies. Each of the two 40.0 ul primary PCR reactions contain; 2.0 ul of a 2.8 uM mutation-causing oligonucleotide, 2.0 ul of the 2.8 uM outside flanking primer to the opposite strand, 0.1 ul of the 200 ug/ml pTL7-R plasmid template, 4.0 ul of the 10X Taq PCR buffer, 3.0 ul of 25 mM MgCl, 0.2 ul of the Taq polymerase (diluted 1:4 with water), 1.0 ul of 2.5 mM each dNTP, and water for the remainder of the volume. The reaction mixture was cycled 30 times in a programmable thermal controller (MJ Research, Inc.). The cycle program contained step 1 at 95°C for 15 seconds, step 2 at 94°C for 1 minute, step 3 at 55°C for 1 minute, step 4 at 72°C for 1 ½ minute, step 5 cycle 29 times to step 2, step 6 at 72°C for 7 minutes, and step 7 at 4°C holding. After each PCR reaction the products were separated by electrophoresis on a 1% agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen). The 40.0 ul secondary reaction contained 5.0 ul of each primary PCR reaction as template, 2.0 ul of each 2.8 uM outside flanking primer, 4.0 ul of 10X Taq PCR buffer, 3.0 ul of 25 mM MgCl, 0.2 ul of the Taq polymerase (diluted 1:4 with water), 1.0 ul of 2.5 mM each dNTP, and water for the remainder of the volume. The reaction was cycled as before and the product was separated and purified as above. The full length product was cut with either \textit{NcoI} or \textit{BglII} (internal \textit{mutH} site) and \textit{HindIII} for one hour at 37°C, separated on a 1% agarose gel, and purified before ligation to the similarly cut pTL7-R backbone. The ligation was carried out overnight at 14°C with 20 ul of the purified mutation containing fragment, 1.0
ul of mini-prep pTL7-R cut backbone, 2.3 ul of T4 ligase 10X buffer, and 1.0 ul of T4 ligase. The entire ligation mixture was transformed into CC106 for recovery and plated onto ampicillin containing LB plates. Colonies that grew from an overnight 37°C incubation were struck out on LB ampicillin plates for isolates. Plasmid DNA from a single colony was isolated using QIAprep Spin Miniprep (Qiagen) and sequenced to identify the mutations. The two outside flanking primers for PCR and sequencing were MM247 (5'-CATCACACAGCGGGCTGG-3') for the forward reactions and MM248 (5'-CAGACCGCTTCTGCG-3') (Operon Technologies) for the reverse reactions. When secondary mutations were present, the sequence containing the correct mutation was removed by restriction enzyme digestion and religated to a similarly cut pTL7-R backbone. The \textit{mutH} gene of the resulting plasmid was then sequenced to confirm the mutational change.

The MutH Δ214 and MutH Δ224 mutants were kindly provided by Dr. K. Murphy (University of Massachusetts Medical School). The MutH Δ224 mutant has the last five amino acids of the C-terminus removed (H, F, L, I, Q). The MutH Δ214 mutant has the last fifteen amino acids of the C-terminus removed (K, N, F, T, S, A, L, L, A, R, H, F, L, I, Q).
Figure 3-1. **Two stage PCR mutagenesis.** There are two primary PCR reactions represented at the top of the figure. Each reaction contains one mutation-causing oligonucleotide and the opposite-strand outside flanking primer. The two primary reaction products are used as templates for the secondary PCR reaction. The same outside flanking primers may be used in the secondary PCR reaction or a second set of outside primers that anneal farther in from the original primers (nested set) may be used. The product from the secondary PCR reaction is shown at the bottom of the figure. This product is then cut with restriction endonucleases (either \textit{NcoI} and \textit{HindIII}, or \textit{BgIII} and \textit{HindIII}) and ligated to the similarly cut plasmid (pTL7-R) for overexpression of the mutant protein.
Figure 3-1.

Primary PCR

Secondary PCR

Ligation to pTL7-R
Sequencing. All DNA sequencing was carried out by the DNA Sequencing Facility at Iowa State University. Both strands of the wild-type and mutant MutH proteins, including the promoter region, were sequenced. The site-directed PCR mutants were identified by sequencing with the outside flanking primers used in the PCR reactions.

Antibodies to MutH. The antibodies to MutH were produced by BAbCO (Berkeley Antibody Company) from our purified native MutH protein. Two rabbits were used for the antibody production with rabbit #1 producing the most sensitive antibodies that were used in the western assays.

Western. A 5 ml overnight culture in LB-Amp broth, inoculated from a single colony, was centrifuged for 2 minutes at 14,000 rpm, resuspended in 100 ul of buffer A (20 mM Hepes pH 7.4, 300 mM NaCl, and 1 mM PMSF), and sonicated for 12 pulses using a Tekmar sonifier to lyse the cells. The cell debris was cleared by centrifugation for 1 minute at 14,000 rpm. 10 ul of the resulting supernatant was boiled for 5 minutes at 95° C and electrophoresed on a 12% SDS-PAGE gel with a 3% SDS-PAGE stacking gel for 2 1/2 hours at 25 mA. Transfer was done from the gel onto Immobilon-P (Millipore) using an OWL electrophoretic system for 15 minutes at 100 mA. The SDS-PAGE gel was stained with Coomassie Brilliant Blue to confirm transfer of the proteins. Detection was performed using the Phototype-HRP Detection Kit for Western Blotting (New England BioLabs). The membrane was incubated in 10 ml of blocking buffer (1X Tris-buffered saline (pH 7.6) (TBS), 0.1% Tween-20, and casein) for 1 hour at room
temperature. The membrane was rinsed in wash buffer (1X TBS and 0.1% Tween-20) and then washed 3 times in fresh changes of wash buffer. The membrane was then incubated with the primary MutH antibody (diluted 1:5000) in blocking buffer with gentle agitation for 2 days at 4°C. The membrane was washed as above and incubated with the secondary rabbit antibody diluted 1:4000 with the blocking buffer for 1 hour at room temperature. The membrane was washed again as above and incubated with 10 ml of LumiGLO (0.4 ml LumiGLO A, 0.4 ml Peroxide B, and 9.2 ml of water) with gentle agitation for 1 minute at room temperature. The membrane was wrapped in Saran wrap and exposed 10 seconds to Fuji x-ray film. The film was processed in a Kodak Xomat processor.

**In vivo screen.** The increase in mutation frequency by the mutant MutH proteins were monitored using rifampicin plates. The vector plasmid (pBAD18) and the wild-type MutH plasmid (pTL7-R) were transformed into strains GM1166 and GM3856 as positive and negative controls. The mutant plasmids were transformed into strain GM3856 for complementation testing and GM1166 for dominant-negative screening. A single colony was inoculated into 1 ml L-Amp broth, grown overnight at 37°C, diluted 10^-6 and 50 ul spread on LB plates with and without ampicillin. The rifampicin plates were spread with 200 ul undiluted wild-type culture (GM1166), 100 ul undiluted mutant culture (GM3856) and 50 ul undiluted mutH plasmid mutant cultures. Two cultures (from two separate colonies) were grown for each test and each culture was plated twice. The plates were incubated at 37°C overnight and colonies counted the next day. The LB plates were used to monitor loss of plasmids due to high cellular amounts of MutH. No arabinose was
included in the plates because this leads to loss of viability. The plasmid produced enough MutH protein to complement the mutH strain without adding arabinose to the plates.
Results

Construction of the mutant plasmids. By comparing the MutH and PvuII structural models, areas in MutH were identified that resembled functional areas of PvuII and these were targeted for mutagenesis (Fig. 3-2). The PvuII restriction endonuclease is a homodimer that recognizes and cleaves the phosphodiester bond in the unmethylated duplex DNA between the central GpC of the sequence 5'-CAGCTG-3' in a Mg$^{2+}$-dependent manner. It produces blunt-ended products with 5'-phosphates. The methyltransferase recognizes the same sequence and modifies the cytosines in the central GpC on each strand generating N4-methyldeoxycytosine. Each monomer of the homodimeric PvuII endonuclease contains three structural regions: a catalytic region, a subunit interface region, and a DNA recognition region. PvuII approaches DNA from the minor groove with one DNA-binding domain reaching around the helix and contacting base pairs in the major groove and another DNA-binding domain making contacts in the minor groove. The major groove contacts are made by residues Asn140, Asn141, Ser81, and the histidine triplets (His83, His84, and His85). The minor groove contacts are made by residues Gln33, Asp34, and Asn35 which make direct contact with the backbone phosphates and the central GC base pair. These minor groove contacting residues are located between the first two helices (Loop L-AB) of the protein (Figure 3-2). The catalytic region resembles structures of catalytic regions in other nucleases. The active site residues are Asp58, Glu68, and Lys70. The three histidines that are
important for DNA binding are also important in the catalytic activity of the protein. Histidine 85 makes an intersubunit hydrogen bond with the corresponding histidine in the other monomer closing off the cleft after DNA has bound\textsuperscript{29}. Histidine 83 and histidine 84 are important for methylation detection and aligning the catalytic residues\textsuperscript{74}. The movement of the histidines in response to the methylated cytosine results in the movement of the catalytic residues away from the proper alignment for catalysis. The movement of the histidines when \textit{PvuII} binds DNA that is unmethylated allows for proper alignment of the catalytic residues and hydrolysis. These histidines provide a signal between DNA binding, state of methylation and the catalytic reaction.

The first area of MutH to be targeted for mutagenesis was the flexible loop between the second and third alpha helices, Loop BC (see Figure 3-2). The second was the C1 loop between the third alpha helix and the first beta sheet. These areas target the analogous flexible loops of \textit{PvuII}, Loop L-AB and Loop 1-2, that make contact with DNA backbone phosphates and recognition bases\textsuperscript{75} (see Figure 3-2)\textsuperscript{118}. MutH could make the same contacts if it approaches DNA from the minor groove. Loop BC residues P41 to G49 and loop C1 residues L59 to Q69 (except for A61 and A63) in MutH were changed to alanines. The third area of MutH contains conserved histidines (H112 and H115). Analogous histidines are present in \textit{PvuII} (H84 and H85) that signal binding of the DNA to allow catalysis to take place, and serve as protein-protein contacts\textsuperscript{74}. The MutH histidines may serve as the same signal therefore, they were also changed to alanines to test this possibility.
Based on the MutH crystal structure, it was suggested that the C-terminal tail region is contacted by MutL to activate the endonuclease activity of MutH\textsuperscript{14,69}. To test this possibility, two deletion mutants were made with the creation of *amber* stop codons. MutH Δ214 and MutH Δ224 (the number of the last residue present in the protein) mutant proteins were made to identify the residues where MutL might stimulate the MutH activity.
Figure 3-2. **Comparison of MutH and PvuII structures for the construction of MutH mutants.** The ribbon diagram of MutH was made from the B monomer with the coordinates submitted by W. Yang (Protein Data Bank code 2AZO) using MIDAS. The ribbon diagram of *PvuII* was made from the B monomer with the coordinates submitted by X. Cheng (Protein Data Bank code 1PVI) using MIDAS. The NH$_2$ and COOH-termini for each protein are labelled on the diagram. Six NH$_2$-termini residues from the *PvuII* monomer have been deleted for clarity. The magenta colored areas of each protein represent structurally similar areas. The two loops and histidines highlighted in green for each model show similar areas that were targeted for mutagenesis in MutH (residues 41-49, residues 59-69, H112, and H115).
Figure 3-2.

MutH

LoopAB

LoopC1

H115

H112

PvuII Monomer

LoopBC

H83, H84, H85

LoopC1
Dominant negative screen. After checking that there were no secondary mutations in each of the mutant MutH plasmids, they were transformed into strain GM1166 (wild-type). The mutant MutH plasmids in GM1166 were tested for a dominant-negative phenotype (mutator phenotype in a wild-type strain) by the same method as the in-vivo screen described in the Materials and Methods. None of the mutant plasmids gave a dominant-negative phenotype (data not shown).

In vivo screen. For the complementation testing in strain GM3856, the controls (GM3856 and pTL7-R in GM3856) show that the wild-type MutH protein causes a 200-fold reduction in the frequency to rifampicin resistance indicating efficient complementation of the chromosomal mutH471::Tn5 allele (Table 3-1). The columns labelled ampicillin and rifampicin in Table 3-1 contain colony counts from the plates and these were used to calculate the mutation frequency. The vector plasmid (pBAD18) causes a small reduction in the mutation frequency and doesn't interfere with the complementation testing. No alterations in mutation frequency were seen in GM1166 with the control plasmids (pBAD18 and pTL7-R). Very few of the MutH site-directed mutants showed a mutator phenotype. Any plasmid giving a mutation frequency value below 3.0 (the total average for the two negative controls) was not considered for further testing (Table 1). Very few of the MutH site-directed mutant plasmids conferred a mutator phenotype, with only plasmid K48A-containing cells showing a strong mutator phenotype and G49A-containing cells showing a 2-fold increase in mutation frequency. These two were chosen for further biochemical studies. The following mutations were
also chosen to validate the screening procedure. The D47A mutant was used because of the D34 analogous position in the \textit{PvuII} structure that recognizes the d(A) in the recognition sequence for \textit{PvuII}\textsuperscript{75}. The H112A and H115A mutants were used because of the signaling properties of the \textit{PvuII} histidines (H84 and H85) described earlier. The Δ214 and Δ224 mutants were used to test the hypothesis that the C-terminal end is where MutL stimulation of MutH activity occurs. The Δ214 mutant had a slightly above normal mutation frequency while Δ224 was normal. There were however, questions concerning the kanamycin insertion into the \textit{mutH} gene. It was predicted from sequencing the insertion area of the chromosome that a shortened version of the MutH protein being produced was possible. It was shown that the \textit{mutH471} strain gives an increased recombination activity, which suggests an aberrant activity (Dr. Kenan Murphy and Dr. Martin Marinus Univ. of Mass. Medical School, personal communication). In this light, another strain was produced that had an in-frame deletion of the \textit{mutH} gene. The \textit{in-vivo} screen was repeated in this strain and the results were clearer (Table 3-2). The original choices for mutants to continue with biochemical studies were not changed but the results with this screen were more pronounced. Greater confidence was placed on these results and the biochemical studies were started.

\textbf{Western blotting.} The steady-state cellular level of MutH produced by the wild-type and mutant constructs was determined as described in the Methods and Materials. The first lane in panels A and B and the second lane in panel C from each of the autoradiographs are the wild-type MutH protein (Fig. 3-3). The different intensities of
each lane are due to the blotting and detection of the protein since they were from the same overnight culture. Each of the mutant plasmids produced comparable amounts of MutH protein to the wild-type. This is shown in lanes 2-9 (3-6 in panel C) of Figure 3-3. Above the MutH protein band, two cross-reacting bands of unknown origin are detected.
Table 3-1. *In-vivo* complementation screen (*mutH471::Tn5*) for mutant MutH proteins.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Colonies on medium containing ampicillin&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Colonies on medium containing rifampicin&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Mutational Frequency&lt;sup&gt;c&lt;/sup&gt; (x10&lt;sup&gt;-8&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM1166</td>
<td>205.3</td>
<td>27.8</td>
<td>3.4 (+/- 1.0)</td>
</tr>
<tr>
<td>GM3856</td>
<td>153.0</td>
<td>394.5</td>
<td>257.8 (+/- 34.9)</td>
</tr>
<tr>
<td>pBAD18/1166</td>
<td>168.0</td>
<td>17.5</td>
<td>2.6 (+/- 1.6)</td>
</tr>
<tr>
<td>pBAD18/3856</td>
<td>94.0</td>
<td>146.8</td>
<td>156.1 (+/- 43.6)</td>
</tr>
<tr>
<td>pTL7-R/1166</td>
<td>216.8</td>
<td>2.8</td>
<td>0.6 (+/- 0.5)</td>
</tr>
<tr>
<td>pTL7-R/3856</td>
<td>263.3</td>
<td>3.9</td>
<td>1.2 (+/- 1.1)</td>
</tr>
<tr>
<td>P41A</td>
<td>111.8</td>
<td>0.3</td>
<td>0.1 (+/- 0.2)</td>
</tr>
<tr>
<td>E42A</td>
<td>106.3</td>
<td>0.0</td>
<td>0.0 (+/- 0.0)</td>
</tr>
<tr>
<td>N43A</td>
<td>93.8</td>
<td>0.3</td>
<td>0.1 (+/- 0.3)</td>
</tr>
<tr>
<td>L44A</td>
<td>77.8</td>
<td>2.0</td>
<td>1.3 (+/- 0.5)</td>
</tr>
<tr>
<td>K45A</td>
<td>61.0</td>
<td>1.0</td>
<td>0.8 (+/- 0.7)</td>
</tr>
<tr>
<td>R46A</td>
<td>84.3</td>
<td>0.5</td>
<td>0.3 (+/- 0.6)</td>
</tr>
<tr>
<td>D47A</td>
<td>84.8</td>
<td>0.8</td>
<td>0.4 (+/- 0.3)</td>
</tr>
<tr>
<td>K48A</td>
<td>135.3</td>
<td>243.8</td>
<td>90.1 (+/- 70.6)</td>
</tr>
<tr>
<td>G49A</td>
<td>86.3</td>
<td>11.8</td>
<td>6.8 (+/- 5.2)</td>
</tr>
<tr>
<td>L59A</td>
<td>61.8</td>
<td>0.8</td>
<td>0.6 (+/- 0.8)</td>
</tr>
<tr>
<td>G60A</td>
<td>82.8</td>
<td>4.8</td>
<td>2.9 (+/- 3.2)</td>
</tr>
<tr>
<td>S62A</td>
<td>98.0</td>
<td>0.5</td>
<td>0.3 (+/- 0.3)</td>
</tr>
<tr>
<td>G64A</td>
<td>124.8</td>
<td>1.0</td>
<td>0.4 (+/- 0.8)</td>
</tr>
<tr>
<td>S65A</td>
<td>89.3</td>
<td>2.3</td>
<td>1.3 (+/- 1.2)</td>
</tr>
<tr>
<td>K66A</td>
<td>52.0</td>
<td>1.5</td>
<td>1.4 (+/- 0.6)</td>
</tr>
<tr>
<td>P67A</td>
<td>67.0</td>
<td>1.0</td>
<td>0.7 (+/- 0.6)</td>
</tr>
<tr>
<td>E68A</td>
<td>76.0</td>
<td>0.5</td>
<td>0.3 (+/- 0.7)</td>
</tr>
<tr>
<td>Q69A</td>
<td>70.5</td>
<td>0.0</td>
<td>0.0 (+/- 0.0)</td>
</tr>
<tr>
<td>H112A</td>
<td>189.3</td>
<td>0.5</td>
<td>0.1 (+/- 0.2)</td>
</tr>
<tr>
<td>H115A</td>
<td>59.8</td>
<td>2.5</td>
<td>2.1 (+/- 1.6)</td>
</tr>
<tr>
<td>Δ214</td>
<td>138.3</td>
<td>24.8</td>
<td>4.5 (+/- 1.4)</td>
</tr>
<tr>
<td>Δ224</td>
<td>173.3</td>
<td>10.3</td>
<td>1.5 (+/- 1.2)</td>
</tr>
</tbody>
</table>
a. The numbers of colonies in this column are from a $10^{-6}$ dilution of the 1 ml overnight culture. These numbers are used to calculate the total number of cells present in the culture: colony# $\times$ (volume of dilution/volume plated) $\times 10^6$. The number of total cells present from an overnight culture is approximately $10^8$ cells.

b. The numbers of colonies in this column are from a percentage plated from the undiluted culture as described in the Experimental Procedures. The total number of rifampicin resistant cells in the culture are then calculated: colony# $\times$ (volume of culture/volume plated).

c. The numbers in this column represent the total rifampicin resistant cells calculated per total number of cells calculated, in the overnight culture (rifampicin total/$10^8$ cells). The numbers in the columns represent the average of three independent experiments. Standard deviations are indicated by the (+/-) numbers following the mutation frequency.
Table 3-2. *In-vivo* complementation screen (ΔmutH471::cat) for mutant MutH proteins.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Colonies on medium containing</th>
<th>Mutation Frequency$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ampicillin$^a$</td>
<td>rifampicin$^b$</td>
</tr>
<tr>
<td>GM4244</td>
<td>279.5</td>
<td>42.5</td>
</tr>
<tr>
<td>GM7586</td>
<td>234.8</td>
<td>1559.0</td>
</tr>
<tr>
<td>pBAD18/4244</td>
<td>261.3</td>
<td>16.5</td>
</tr>
<tr>
<td>pBAD18/7586</td>
<td>362.0</td>
<td>1447.0</td>
</tr>
<tr>
<td>pTL7/4244</td>
<td>108.0</td>
<td>0.0</td>
</tr>
<tr>
<td>pTL7/7586</td>
<td>231.5</td>
<td>41.0</td>
</tr>
<tr>
<td>P41A</td>
<td>215.0</td>
<td>24.0</td>
</tr>
<tr>
<td>E42A</td>
<td>203.0</td>
<td>10.5</td>
</tr>
<tr>
<td>N43A</td>
<td>217.0</td>
<td>13.5</td>
</tr>
<tr>
<td>L44A</td>
<td>239.0</td>
<td>32.5</td>
</tr>
<tr>
<td>K45A</td>
<td>213.0</td>
<td>9.0</td>
</tr>
<tr>
<td>R46A</td>
<td>223.0</td>
<td>6.0</td>
</tr>
<tr>
<td>D47A</td>
<td>244.0</td>
<td>9.0</td>
</tr>
<tr>
<td>K48A</td>
<td>242.0</td>
<td>2178.0</td>
</tr>
<tr>
<td>G49A</td>
<td>269.0</td>
<td>1502.0</td>
</tr>
<tr>
<td>L59A</td>
<td>303.0</td>
<td>19.5</td>
</tr>
<tr>
<td>G60A</td>
<td>246.0</td>
<td>15.0</td>
</tr>
<tr>
<td>S62A</td>
<td>347.0</td>
<td>5.0</td>
</tr>
<tr>
<td>G64A</td>
<td>266.0</td>
<td>18.0</td>
</tr>
<tr>
<td>S65A</td>
<td>214.0</td>
<td>14.5</td>
</tr>
<tr>
<td>K66A</td>
<td>214.0</td>
<td>22.5</td>
</tr>
<tr>
<td>P67A</td>
<td>252.0</td>
<td>27.0</td>
</tr>
<tr>
<td>E68A</td>
<td>280.0</td>
<td>17.5</td>
</tr>
<tr>
<td>Q69A</td>
<td>239.0</td>
<td>15.5</td>
</tr>
<tr>
<td>H112A</td>
<td>171.0</td>
<td>10.5</td>
</tr>
<tr>
<td>H115A</td>
<td>207.0</td>
<td>21.0</td>
</tr>
<tr>
<td>Δ214</td>
<td>199.0</td>
<td>2234.0</td>
</tr>
<tr>
<td>Δ224</td>
<td>175.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>
a. The numbers of colonies in this column are from a $10^{-6}$ dilution of the 1 ml overnight culture. These numbers are used to calculate the total number of cells present in the culture: colony# x (volume of dilution/volume plated) x $10^6$. The number of total cells present from an overnight culture is usually $10^8$.
b. The numbers of colonies in this column are from a percentage plated from the undiluted culture as described in the Experimental Procedures. The total number of rifampicin resistant cells in the culture are then calculated: colony# x (volume of culture/volume plated).
c. The numbers in this column represent the total rifampicin resistant cells calculated per total number of cells calculated, in the overnight culture (rifampicin total/$10^8$ cells). The numbers in the columns represent the average of three independent experiments. Standard deviations are indicated by the (+/-) numbers following the mutation frequency.
Figure 3-3. **Western blotting of the wild-type and mutant MutH proteins.** The western assays were performed as described in the Materials and Methods. The band of interest is labelled 'MutH' between the autoradiographs. Panel A, lane 1 is the wild-type MutH protein followed by the mutant proteins P41A, E42A, N43A, L44A, K45A, R46A, D47A, and K48A (lanes 2-9). Panel B, lane 1 is also the wild-type MutH protein followed by the mutant proteins L59A, G60A, S62A, G64A, S65A, K66A, P67A, and E68A (lanes 2-9). Panel C, lane 1 is 5 ul of the wild-type overnight MutH culture. Lane 2 is 10 ul of the wild-type overnight MutH culture followed by the mutant proteins G49A, Q69A, H112A, and H115A (lanes 3-6). 10 ul from an overnight culture was used in each of the lanes except were noted.
CHAPTER IV

BIOCHEMICAL ANALYSIS OF MUTH

Introduction

The mismatch repair assay was first developed to analyze DNA mismatch repair in cell-free extracts. The results from these first experiments indicated that the repair mechanism was dependent on the state of DNA methylation, ATP, MutH, MutL, MutS, and the products of the $uvrE$ ($uvrD$, $mutU$, and $recI$) gene. In vivo findings confirmed that the four mutator genes $mutH$, $mutL$, $mutS$, and $uvrD$ ($mutU$) are required for mismatch repair, as well as single-strand DNA binding protein. The requirement for a d(GATC) site and MutH was shown using this in vitro mismatch repair assay. The reaction was reconstituted in vitro with all of the required components in a defined system; it consists of an f1 bacteriophage-derived closed, circular heteroduplex with at least one d(GATC) site, MutH, MutL, MutS, DNA polymerase III holoenzyme, single-strand DNA binding protein, Exonuclease VII or RecJ, Exonuclease I, DNA helicase II, DNA ligase/NAD$^+$, MgCl$_2$, ATP, and all four dNTP's. Modification of the basic mismatch repair assay resulted in the endonuclease assay that is employed.
here\textsuperscript{165}. A band shift assay was also developed to study the activity of MutH binding to DNA.
Materials and Methods

Strains, plasmids, and media. Strain GM3856 (GM1166 mutH471::Tn5) was used for the overexpression of the wild type and the mutant MutH proteins. Strain GM5862\(^9\) was used for the overexpression of wild-type MutL. Strain CC106\(^{168}\) was used for the recovery of the mutant plasmids after site-directed mutagenesis. Strain GM2807 (Hfr dam-13::Tn9) was used as the host to prepare the phage MR1 RF homoduplex substrate for the endonuclease assay.

Plasmid pTL7 was described in the Materials and Methods of Chapter 3 and contains wild-type His-tagged mutH in a pBAD18 backbone. Plasmid pMQ393 contains the wild-type His-tagged MutL gene in a pACYC184 backbone\(^{38}\).

The Luria-Bertani (LB) medium was made with 10 gm of tryptone, 5 gm of yeast extract, 5 gm of NaCl, and 1 ml of 1M sodium hydroxide per liter of water. The YT medium and TAE buffer were made according to Sambrook et. al.\(^{141}\). Ampicillin (Amp), chloramphenicol (Cam), and rifampicin (Rif) were added to media at 100 ug/ml, 30 ug/ml, and 100 ug/ml respectively. TE was made with 50 mM Tris-HCl (pH8.0) and 50 mM Na\(_2\)EDTA in water.

Overexpression and purification of MutH. The overexpression and purification of the native wild-type protein was described previously in Chapter 2. The His-tagged wild type and all mutant His-tagged proteins were purified in the same manner. For each
strain, a 20 ml LB-Amp culture was grown overnight from a single colony and used to inoculate one liter of the same medium; growth continued with shaking at 37°C until an OD$_{600}$ of 1.0 was reached. Arabinose (Difco) was added to a final concentration of 0.2% and the culture induced for 2 hours at 37°C. Cells were harvested by centrifugation for 20 minutes at 5,000 rpm, washed with water, and stored frozen (-80°C). The pellet was thawed, resuspended in 4 ml of reconstitution buffer (20 mM Hepes (pH7.4), 300 mM NaCl, and 1mM phenylmethanesulfonylfluoride (PMSF)), and the cells lysed using a French pressure cell. The extract was then sonicated (20 pulses from a Tekmar sonifier) and centrifuged for 30 minutes at 15,000 rpm to remove cell debris. The supernatant was applied to a 4 ml Fast Flow Chelating Sepharose (Pharmacia Biotech) column charged with 100 mM NiCl$_2$ and equilibrated with the reconstitution buffer. Step fractions of 100 mM, 150 mM, 250 mM, and 400 mM imidazole (10 ml each) were used to elute the MutH protein. The 250 mM fraction contained the MutH protein and was dialyzed against two changes of 20 mM Hepes (pH7.4), 300 mM NaCl, and 0.1 mM ethylenediaminetetraacetic acid (EDTA). The MutH concentration was determined by measuring the OD$_{280}$ (1 OD = 0.67 mg/ml). MutH was at least 95% pure as determined by a SDS-PAGE Coomassie Brilliant Blue stained gel.

**Overexpression and purification of MutL.** The His-tagged MutL protein was purified as described$^{38}$, except that the transformation mixture of GM5862 with pMQ393 was incubated for 90 minutes after heat shock and then added to 50 ml YT-Cam medium for overnight incubation at 37°C. The 50 ml overnight culture was used to inoculate one liter of the same medium; growth continued overnight with slow shaking at room
temperature until an OD600 of 1.0 was reached. Isopropyl-β-D-thiogalactoside (IPTG) (US Biological) was added to a final concentration of 0.5 mM and the culture induced for 2 hours at room temperature. Cells were harvested by centrifugation for 20 minutes at 5,000 rpm, washed with water, and resuspended in 4 ml of reconstitution buffer (20 mM Hepes (pH7.4), 300 mM NaCl, and 1mM phenylmethylsulfonylfloride (PMSF)). The cells were lysed using a French pressure cell. The extract was then sonicated (20 pulses from a Tekmar sonifier) and centrifuged for 30 minutes at 15,000 rpm to remove cell debris. The supernatant was applied to a 4 ml Fast Flow Chelating Sepharose (Pharmacia Biotech) column charged with 100 mM NiCl2 and equilibrated with the reconstitution buffer. The step fractions used for elution from the Sepharose column were one 100 mM, two 150 mM, and one 400 mM Imidazole (10 ml each). The 400 mM fraction contained the MutL protein, which was dialyzed against two changes of 20 mM Hepes (pH7.4), 300 mM NaCl, and 0.1 mM EDTA. The concentration of MutL was determined by measuring the OD280 (1 OD = 1.24 mg/ml). The protein was at least 90% pure as measured on a Coomassie Brilliant Blue-stained SDS-PAGE gel. Native MutL was graciously provided by Dr. Paul Modrich and Dr. Claudia Spampinato (Duke University).

Mismatch repair assay. The mismatch repair assay was performed by Dr. Breck Parker as described using the native MutH protein. The reaction (10 ul) consists of 0.1 ug (24 fmol) of hemimethylated heteroduplex substrate, 0.54 ng of MutH, 17 ng of MutL, and 35 ng of MutS protein, 10 ng of DNA helicase II, 200 ng of single-strand DNA binding protein, 95 ng of DNA polymerase III holoenzyme, 1 ng of exonuclease I, 20 ng of E. coli DNA ligase, 2 mM adenosine triphosphate (ATP), and 100 uM each of the four
deoxynucleoside triphosphates (dATP, dCTP, dGTP, and dTTP) in a buffer of 50 mM Hepes (pH 8.0), 20 mM KCl, 6 mM MgCl₂, 1 mM DTT, and 25 uM β-NAD⁺ (nicotinamide adenine dinucleotide). The reactions were incubated at 37°C for 20 minutes, quenched at 55°C for 10 minutes, chilled on ice, and then digested with XhoI or HindIII with ClaI for 60 minutes at 37°C. The reaction products were separated by electrophoresis on a 1% agarose gel containing 0.2% ethidium bromide at 70 V for 2-3 hours.

Bacteriophages MR1 and MR3 that are derivatives of phage f1R229 containing only one d(GATC) site are used to construct the heteroduplex. The MR3 bacteriophage containing an XhoI site, 3'-TTCGAAGCTC-5' (mismatched base in bold), was used to make a large preparation of RF DNA (see endonuclease assay below) and the MR1 bacteriophage containing a HindIII site, 5'-AAGCTTTCGAG-3' (mismatched base in bold), was used to make a large preparation of single-strand DNA from a dam⁻ (GM2807) strain (and therefore was unmethylated). The single-strand DNA was prepared from 0.2 ml aliquots of an overnight culture (20 ml) of the dam⁻ strain with an added 10⁸ bacteriophage to each aliquot. This infected culture was incubated at room temperature for 10 minutes without shaking. It was plated with 3.5 ml of top agar (at 47°C) onto LB plates and incubated overnight at 37°C. The top agar (with infected cells) was then scraped into GSA bottles by rinsing with 2 x 4 ml LB. 1 ml of chloroform was added to the bottles (to kill the bacteria) per 100 ml volume of scraped agar. This was shaken at room temperature for 30 minutes to elute the bacteriophage from the agar, spun for 10 minutes at 10,000 rpm, and the supernatant poured off into a clean bottle. Solid NaCl (32
g/L) and polyethylene glycol 8,000 (55 g/L) was added to the supernatant, stirred at room temperature until dissolved, stored overnight at 4°C, and spun for 60 minutes at 4°C and 10,000 rpm. The supernatant was poured off, the pellet drained well, and resuspended in 4 ml of TE. The resuspended pellet was shaken at room temperature for 2 hours and spun 10 minutes at 10,000 rpm. The supernatant was extracted with 5 ml of phenol at 65°C, back extracted with 1 ml of TE, extracted again with 5 ml of phenol at room temperature, back extracted with 1 ml of TE and extracted with ether until the aqueous phase was clear (approximately 4 times). The extracted single-strand (ss) DNA is dialyzed 3 times with 0.5 liter of TE and quantitated by measuring the A_{260} (1.0 A_{260} = 33 ug/ml of ss DNA).

The heteroduplex was constructed as described\textsuperscript{103}. RF DNA (100 ug) was cleaved by \textit{HincII} (60 U for 2.5 hours), EDTA was added to 10 mM, and the DNA was extracted once with phenol, twice with ether and precipitated with ethanol (chilled at -80°C for 10 minutes, spun for 10 minutes at 10,000 rpm, rinsed with ethanol and spun as before, and finally allowed to dry at room temperature). The pellet was resuspended in 0.5 ml of TE with 500 ug of the ss DNA, NaCl (final concentration 10 mM), and Tris-HCl (pH7.6) (final concentration 50 mM) added to a total volume of 2 ml. At room temperature, 60 ul of 10 N NaOH (freshly prepared) was added to the 2 ml above and incubated for 5 minutes. It was neutralized with 200 ul of 2.9 N glacial acidic acid, 90 ul of 3 M KCl, and 248 ul of 1 M potassium phosphate (pH7.4). This was incubated for 30 minutes at 65°C, 30 minutes at 37°C and stored on ice. The double-strand and single-strand DNA were separated with a 5 ml hydroxylapatite column equilibrated with 30 mM potassium phosphate buffer (pH 6.9). The column was washed with the
equilibration buffer and the single-strand DNA was eluted off the column with 140 mM potassium phosphate buffer (pH 6.9) in 1.0 ml fractions. The elution of the single-strand DNA was monitored by measuring the A_{260}. When the A_{260} returns to baseline the double-strand DNA was eluted off of the column with 420 mM potassium phosphate buffer (pH 6.9) in 1.0 ml fractions. All fractions were checked by electrophoresis in a 1 % agarose gel. The fractions containing only double-strand DNA were pooled with EDTA added to 10 mM and the sample concentrated 4-fold with butanol. The concentrated sample was then dialyzed with 2 changes in 500 ml of TE (one overnight and the other for four hours). The double-strand DNA was then quantitated by A_{260} with the following solutions added per ml of DNA; 11 ul of 1 M Tris-HCl (pH 7.6), 1.1 ul of 0.1 M β-NAD^{+}, 11 ul of 1 M ammonium sulfate, 5.5 ul of 10 mg/ml bovine serum albumin (BSA), 1.37 ul of 0.1 M DTT, and 5.5 ul of 1 M MgCl\textsubscript{2}. For every 100 ug of DNA, 29 nmole of ethidium bromide was added to the reaction with 50 units of E. coli ligase. The reaction was incubated for 2 hours at room temperature, stored on ice, and an aliquot checked by electrophoresis in a 1 % agarose gel (60 minutes at 70 V). EDTA was added to 10 mM before separating the DNA (banding) by cesium chloride (CsCl). For each ml of ligation mixture, 1.05 g of CsCl and 50 ul of 10 mg/ml ethidium chloride was added. The solution was banded by centrifugation at 45,000 rpm overnight at 20° C. The bottom band was pulled from the centrifuged tube with a needle and syringe and extracted with butanol. The DNA solution was dialyzed overnight with TE and extracted once with phenol and once with ether. It was then dialyzed with 2 changes of TE and
quantitated by measuring the $A_{260}$. The DNA was checked by restriction digest with
$XhoI$, $Hind$III, and $MboI$ (for hemi-methylation).

Circular Dichroism. The circular dichroism (CD) experiments were carried out
with Dr. Robert Talanian (BASF, Worcester MA). The CD measurements were taken on
a spectrophotometer by recording spectra at 1 nm increments between 206 and 270 nm
using a 2 second integration time. The sample temperatures of $20^\circ C$ and $75^\circ C$ were
controlled by a thermoelectric cell holder. The concentration of the wild-type MutH
protein was 0.11 mg/ml while the K48A mutant protein was 0.94 mg/ml in the
experiments. The sample temperatures for $\Delta 214$ and the wild-type MutH protein were
$37^\circ C$ and $75^\circ C$. The spectra were recorded at 1 nm increments between 190 and 260 nm
using a 1 second integration time.

Endonuclease assay. The hemi-methylated heteroduplex was constructed as
described above$^{103}$. The covalently-closed replicative form (RF) of the MR1 phage was
purified as well, from GM2807 cells, generating an unmethylated homoduplex. The
GM2807 cells are grown to the logarithmic stage at $37^\circ C$ with shaking and infected with
MR1 phage (ratio of 10 phage: 1 cell). The infected culture is allowed to rest for 5
minutes and then shaken for 15 minutes before chloramphenicol is added to 30 ug/ml.
The infected culture is then shaken for an additional 60 minutes before the cells were
collected by centrifugation. The RF DNA was prepared from the Qiagen midi prep kit
and quantitated by measuring an $A_{260}$ ($1.0 \times A_{260} = 50$ ug/ml). The MR1 and MR3 phage
were provided by Dr. R. Lahue (University of Nebraska).
Enzyme activity (amounts are indicated in the figures) was measured at 37°C in a 10 ul reaction containing 20 mM Tris HCl, (pH 7.7), 5 mM MgCl₂, and 25 fmoles of MR1 RF for 1 hr. MutL and ATP were added at 71 pmoles and 1.25 mM respectively in the MutH endonuclease stimulation assays. The assays were carried out with native MutL or His-tagged MutL. The reaction was stopped with 5 ul of a 50% glycerol solution containing 50 mM EDTA, 1% SDS, 0.05% Bromphenol Blue, and 0.05% Xylene Cyanol. The reaction products were separated by electrophoresis on a 1% agarose gel at 5 V/cm for 3 hrs. The gels were stained with Vistra Green (Amersham) following the manufacturer's instructions (1:10,000 dilution) for 1 hr with shaking. The gels were then scanned on a Storm (Molecular Dynamics PhosphorImager) and the amount of nicking was quantitated using the ImageQuant (Molecular Dynamics) software. Any nicked substrate detected in the negative control was subtracted from the amount seen with the proteins added.

Preparation of labelled oligonucleotide duplex. A 36 base pair (bp) DNA oligonucleotide MM40 5'-GCATACGGAAGTTAAAGTGCCTGATCATCTCTAGCCAG-3' (Operon Technologies) containing a single, centrally located d(GATC) site, was labelled using T4 polynucleotide kinase (PNK) (New England BioLabs). 2 uM of the oligonucleotide was used in a 12 ul reaction with 1.2 ul of PNK buffer, 1 ul of PNK enzyme, 1 ul of gamma P³² ATP at 10 mCi/ml (Amersham Pharmacia Biotech), and water to a final concentration of 0.5 uM oligonucleotide. The reaction was incubated for 1 hour at 37°C and 10 minutes at 75°C. 2 uM of cold complementary oligonucleotide, MM41 or LH65 (containing the N6-methyl adenine in the d(GATC) site) was added with
water to a final volume of 100 ul resulting in an unmethylated or hemimethylated duplex respectively. This was incubated for 5 minutes at 95° C, 30 minutes at 37° C, and 30 minutes at room temperature. The unincorporated $^{32}$P and ss DNA were removed using a MicroSpin G-25 column (Pharmacia-Biotech) and the duplex DNA dissolved in a final volume of 100 ul of water. The fully methylated homoduplex was constructed with the labelled oligonucleotide LH65 and cold complementary oligonucleotide MM122. The control homoduplex, with a d(GGTC) site in place of the d(GATC) site underlined in the oligonucleotide sequence above, was constructed with oligonucleotides MM120 (labelled) and MM121 (cold). A 5X reaction mixture with 50 ul of the labelled homoduplex and 50 ul of 10X AR buffer (200 mM Tris-HCl (pH7.6), 50 mM MgCl$_2$, 1 mM dithiothreitol (DTT), and 0.1 mM EDTA) was used in the band shift assay.

**Band shift assay.** MutH binding to DNA was assayed using the labelled homoduplex prepared as described above. A 5 ul reaction volume contained 22.5 fmol of the labelled homoduplex (1 ul of the 5X reaction mixture) and MutH, which was added in varying concentrations (water makes up the remainder of the reaction volume). MutL and ATP were added at 9 pmoles and 1.25 mM respectively in the stimulation assays. The mixture was incubated for 30 minutes on ice, followed by the addition of 1.6 ul of a 50% sucrose solution and loaded on a pre-run (10 minutes) 6% native polyacrylamide gel in TAE buffer, (pH 7.5) and electrophoresed at 50 mA for 2½ hrs. The gel was then laid on a PhosphorImager screen overnight for scanning on the Storm (Molecular Dynamics). The band intensities were quantitated to determine binding ability using ImageQuant software. In experiments using the wild-type protein, the gels were laid on Fuji film
overnight for development by an Xomat processor (Kodak). In experiments where cutting of the labelled homoduplex was necessary, MgCl₂ was added to the AR buffer in the 5X-reaction mixture. The assay was then performed as described above.

Apparent dissociation constants (K_d) were determined by fitting the experimental data to the equation $1/r = 1 + K_d/[C]_{total}$ if $f = 1 - r$, where $f$ is the degree of dissociation or unbound DNA/total DNA) with the program KaleidaGraph (Synergy Software). The degree of binding (bound DNA/total DNA) is $r$ and $[C]_{total}$ is the protein concentration in the reaction.
Results

Wild-type activity. After the first purification of native MutH, the activity of the protein was determined and compared to reported levels in the literature\textsuperscript{90,165}. The native MutH protein in the repair assay left approximately 30-50\% of the substrate DNA unrepaired (Fig. 4-1A). This assay involves the complete repair pathway with the endonuclease XhoI (as well as ClaI to generate separable pieces) to check for repair. The G-T mismatch that is contained within the heteroduplex may be repaired to generate a XhoI site or a HindIII site depending on which DNA strand is unmethylated and therefore changed. In the heteroduplex constructed for these assays, the mismatch should be repaired to generate the XhoI site (see Materials and Methods for specific details). The endonuclease assay shows that at a molar ratio of 20:1 (protein monomer/heteroduplex) the reaction is almost at completion (Fig 4-1B). In this assay MutH is present alone, therefore the hemi-methylated heteroduplex is only nicked. At higher concentrations of MutH, the nicking is complete. These results are similar to those reported in the literature cited above.

The band shift assays in figure 4-2 compare the wild-type MutH activity upon DNA substrates with different methylation states. The \textit{in vivo} substrate is hemimethylated DNA and represented by the first autoradiograph (Fig. 4-2A). The substrate DNA is fully complexed with the MutH protein somewhere between 80 and 160 pmoles of MutH. The other two methylation states, unmethylated and fully methylated,
are fully complexed at similar concentrations of MutH (Fig. 4-2B, C). To allow MutH to cut the DNA substrate, magnesium at 5 mM (1.25 mM final concentration) was present in the stock buffer. The hemimethylated DNA is cut as soon as it is bound (Fig. 4-2A and D). The unmethylated DNA is bound but not cut as readily, indicating a less ideal substrate (Fig. 4-2B and E). The fully methylated DNA was also bound but cut very little and only at the highest concentration (160 pmoles) of MutH (Fig. 4-2C and F).

Two control experiments were performed using the native wild-type MutH protein. The first experiment tested for specificity by changing the DNA substrate at one base in the sequence. There is a d(GGTC) site in place of the d(GATC) site in the DNA substrate. The MutH protein did not bind to the altered DNA (Fig. 4-3A). The second experiment had the unlabelled DNA substrate added back to the reaction mixture as a cold competitor (after incubation with the labelled substrate). The cold d(GATC) DNA substrate was able to compete for the MutH protein and shift the labelled DNA back to its original position in the gel (Fig. 4-3B).
Figure 4-1. **Activity of native MutH protein.** The mismatch repair assay was performed as described in the Materials and Methods. Panel A, the 0 lane contains a *mutH* cell extract in place of the purified MutH protein. The next lane contains 52 pmoles of purified MutH protein, along with the other components of the assay. The repair products, labelled on the side of the gel, are the substrate DNA after restriction digest with *Xho*I and *Cla*I. The endonuclease assay was performed as described\textsuperscript{165}. Panel B, the 0 lane contains only the substrate DNA and buffer. The next three lanes are labelled as the molar ratio of MutH protein to DNA substrate at 20, 40, and 80.
Figure 4-1.

(A) [MutH] pmoles

- 0 pmoles
- 52.0 pmoles

Unrepaired DNA
Repair Products

(B) mole MutH / mole duplex

- 0
- 20
- 40
- 80

Nicked DNA
cc DNA
Figure 4-2. **Wild-type MutH activity.** The band shift assays were performed as described in the Materials and Methods. These gels were laid on film overnight; therefore no quantitation was done. Native mutH protein was used for all the assays. Panel A, shows the MutH binding ability with a hemimethylated DNA substrate (the *in vivo* substrate). Panel B, shows the amount of binding with an unmethylated DNA substrate. Panel C, shows the binding with a fully methylated DNA substrate. Panel D, shows the cut substrate product with hemimethylated DNA substrate. Panel E, shows the amount of cut product with an unmethylated DNA substrate. Panel F, shows a small amount of cut product with a fully methylated DNA substrate. Panels A, B, and C are without MgCl$_2$ added to the reaction while panels D, E, and F have 1.25 mM MgCl$_2$ (final concentration) in each reaction.
Figure 4-2.

A Hemimethylated  
[MutH] pmoles

0  20  40  80  160

Bound DNA

Free DNA

B Unmethylated  
[MutH] pmoles

0  20  40  80  160

Bound DNA

Free DNA

C Fully methylated  
[MutH] pmoles

0  20  40  80  160

Bound DNA

Free DNA

D Hemimethylated  
[MutH] pmoles

0  20  40  80  160

Linear DNA

Cut DNA

E Unmethylated  
[MutH] pmoles

0  20  40  80  160

Linear DNA

Cut DNA

F Fully methylated  
[MutH] pmoles

0  20  40  80  160

Linear DNA

Cut DNA
Figure 4-3. **Control experiments for the native MutH protein.** The band shift assays were performed as described in the Materials and Methods. Panel A, the DNA substrate contains a d(GGTC) in place of the d(GATC) site. The amount of MutH added to the reactions is shown at the top of each lane. Panel B, the hemimethylated DNA substrate was used in these reaction. The first lane is the DNA substrate alone. MutH was added at 160 pmoles in each of the other lanes. At the top of each lane, the cold competitor concentration is indicated as an excess of the labelled DNA substrate. The cold competitor is the unlabelled hemimethylated d(GATC) DNA duplex.
Figure 4-3

A

[MutH] pmoles

0 20 40 80 160

B

Duplex Cold Competitor

0 2X 22X 220X
**Endonuclease assay.** The product of the reaction is a nick in the MRI covalently closed circular molecule, which has only one d(GATC) site and no mismatch. It has been shown that the MutH activity is mismatch independent and therefore the same results are achieved with a heteroduplex as well as a homoduplex substrate. In the absence of MutL, increasing amounts of the wild-type MutH protein were added to the unmethylated homoduplex until complete nicking was achieved (Fig. 4-4A and Fig. 4-5A). The addition of 2.0 or 3.0 pmoles of wild-type protein resulted in complete nicking of the substrate. The results of the endonuclease assay for the remaining three mutants (K48A, G49A, and Δ214) are shown below the wild-type in Figure 4-4A. K48A has no detectable endonuclease activity and G49A activity is decreased approximately 30-fold (Fig. 4-5A). D47A, H112A, H115A, and Δ224 MutH proteins gave the same result as wild-type and were not tested further (data not shown). These results were in agreement with the in-vivo screen because the mutation frequencies in cells with these proteins (D47A, H112A, and H115A) were similar to the mutation frequency of the wild-type MutH protein in strain GM3856 (mutH471::Kan) and GM7586 (ΔmutH471::cat).

In the presence of MutL and ATP, one thousand-fold less MutH is needed to achieve the same results as without MutL (Fig. 4-4B and Fig. 4-5B & C). The cleavage of mutants K48A and G49A are shown below the wild-type in Figure 4-4. Although G49A is partially stimulated by MutL (it is increased 10-fold relative to the unstimulated activity), its activity is reduced about 200-fold compared to the stimulated wild-type protein (Fig. 4-5B). K48A shows the most dramatic increase and is reduced only 3-fold.
relative to the wild-type protein (Fig. 4-5C). MutL is unable to increase the activity of the G49A mutant as much as that for K48A. There was no MutL stimulation seen for Δ214 in this assay and the activity is decreased about 15-fold from the unstimulated wild-type MutH protein (Fig. 4-5D). It would appear then, to have a 15,000-fold decrease in activity compared to the stimulated MutH protein. Since the reduction in endonuclease activity of the mutant proteins could be due to decreased binding ability, this was tested as described later.

Circular dichroism. A circular dichroism experiment was performed on the K48A mutant protein because of the lack of activity without MutL present in the endonuclease assay. The results show that at 20°C the K48A mutant protein is folded similarly to the wild-type MutH protein, with a mixture of alpha helices and beta sheets (Fig. 4-6). At 75°C the wild-type and K48A MutH proteins show a typical pattern for denatured proteins (Fig. 4-6). A circular dichroism experiment was also performed on the Δ214 mutant protein because of its lack of activity both in binding and endonuclease ability. The results show that at 37°C the Δ214 mutant protein folds similarly to the wild-type protein. There is a shift upwards in the curve toward a larger beta sheet content which is consistent with the loss of the large last helix (Helix F) in the mutant protein (Fig. 4-6). At 75°C the Δ214 mutant protein shows the pattern for a denatured protein, similar to the wild-type protein (Fig. 4-6).
Figure 4-4. **Endonuclease activity of the wild-type (WT) and mutant MutH proteins.**

The endonuclease assays were performed as described in Materials and Methods. Increasing amounts of MutH (indicated at the top of each lane) were incubated with the unmethylated MR1 DNA substrate. The DNA products were separated in a 1% agarose gel before staining and scanning for quantitation. Panel A, all reactions are without MutL and contained 25 fmole of homoduplex DNA substrate. The 0 lane represents unreacted homoduplex DNA substrate. The (+) control for mutants K48A and Δ214 contain 2.5 fmoles of wild-type MutH protein in the reaction. Panel B, all reactions contain 71 picomoles MutL, 1.25mM ATP and 25 fmole of homoduplex DNA substrate. The 0 lane represents unreacted homoduplex DNA. The ccDNA is the closed circular substrate used in the reaction.
Figure 4-4

<table>
<thead>
<tr>
<th>[MutH] (pmoles)</th>
<th>WT</th>
<th>G49A</th>
<th>K48A</th>
<th>Δ214</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A - MutL

- Nicked DNA
- Linear DNA
- ccDNA

B + MutL

- Nicked DNA
- Linear DNA
- ccDNA
Figure 4-5. **Comparison of wild-type and mutant MutH endonuclease activity.**

These graphs represent the quantitation of the data from the endonuclease assays. Panel A, shows the activity of the wild-type and mutant MutH proteins without MutL added. The wild-type MutH protein (■), the G49A mutant (♦), and the K48A mutant (●) are shown together in this graph. Panel B, the activity of the wild-type (■) and G49A (♦) mutant MutH proteins with 9 picomoles MutL added in the reaction. Panel C, the activity of the wild-type (■) and K48A (●) mutant MutH proteins with 9 picomoles MutL added in the reaction. Panel D, represents the Δ214 mutant data. The Δ214 mutant with 9 picomoles MutL added to the reaction (■) and the Δ214 mutant without MutL in the reaction (●) are plotted in the final graph. This shows the lack of stimulation by MutL for this mutant. Each point on all of these graphs represents the average of at least three independent measurements.
Figure 4-5

A

B

C

D

% Nicked

% Nicked

% Nicked

% Nicked

120
100
80
60
40
20
0
10
100
1000
10^nM

120
100
80
60
40
20
0
0.01
0.1
1
10
100
1000
nM

120
100
80
60
40
20
0
0
1
2
3
4
5
6
7
nM

120
100
80
60
40
20
0
0
100
1000
10^nM

% Nicked

% Nicked

% Nicked

% Nicked

0
5
10
15
20
25
30
35
40
100
1000
10^nM
Figure 4-6. **Circular dichroism for wild-type and K48A MutH proteins.** Upper panel, the circular dichroism at $20^\circ C$ shows the proteins folded with mixed alpha helices and beta sheets. The His-tagged wild-type MutH protein at 0.11 mg/ml (■) and His-tagged K48A mutant protein at 0.94 mg/ml (●) are indicated on the graph. Middle panel, the circular dichroism at $75^\circ C$ shows the denatured proteins without the indication of alpha helices or beta sheets. The wild-type MutH protein (■), K48A mutant protein (●), and the Δ214 mutant protein (♦) are shown. Lower panel, the circular dichroism at $37^\circ C$ shows the proteins folded with mixed alpha helices and beta sheets with the Δ214 mutant exhibiting a shift toward more beta sheet content as expected (the last helix had been deleted containing 10 residues). The His-tagged wild-type MutH protein at 0.10 mg/ml (■) and His-tagged Δ214 mutant protein at 0.17 mg/ml (●) are indicated on the graph.

These graphs have not been normalized for concentration differences.
Figure 4-6

**Circular Dichroism 20°C**

- Graph showing the circular dichroism at 20°C with data points and a trend line.

**Circular Dichroism 75°C**

- Graph showing the circular dichroism at 75°C with data points and a trend line.

**Circular Dichroism 37°C**

- Graph showing the circular dichroism at 37°C with data points and a trend line.
**Band shift assay.** The substrate is a linear 36-mer duplex of DNA with one d(GATC) site. Without MutL, the wild-type protein gives a complete band shift with the appearance of a transient band between the free DNA and bound DNA (Fig. 4-7A). Both mutant proteins K48A and G49A gave results similar to the wild-type (data not shown). The Δ214 mutant protein initiated binding like wild-type but failed to achieve a final shifted position (Fig. 4-7A).

The band shift assay for MutH in the presence of MutL was complicated because MutL shows non-specific DNA binding at concentrations above 9 pmoles to approximately the same position in the gel as the MutH band. The amount of MutL, therefore, was lowered until it no longer shifted the DNA substrate. The assay with MutL and wild-type MutH caused the band shift to start a little sooner (compare at 38 pmoles) but it finished at the same concentration as without MutL (Fig. 4-7B). The K48A mutant was like wild-type and the G49A mutant bound a little tighter than wild-type. The Δ214 mutant was not stimulated by MutL and gave the same result as without the added protein (not shown).
Figure 4-7. **Binding abilities of the wild-type and mutant MutH proteins.** The binding ability of the wild-type and mutant MutH proteins were measured by a band shift assay using a $^{32}$P-labeled oligonucleotide as described in the Materials and Methods. The hemimethylated DNA was incubated with increasing amounts of the MutH protein (indicated above each lane) in the reaction and separated by a 6% polyacrylamide gel followed by scanning (Storm) for quantitation. Panel A, all reactions contain 22 fmoles of homoduplex DNA, 1.25mM ATP and do not contain MutL. The 0 lanes are without MutH and represent 22 fmoles of unreacted homoduplex DNA. Panel B, all reactions contain 22 fmoles of homoduplex DNA, 1.25 mM ATP and 9 picomoles of MutL (except the (+) and 0 lanes). The 0 lane represents 22 fmoles of unreacted homoduplex DNA. The (+) control lane contains 135.0 pmoles MutH, 1.25 mM ATP, and 22 fmoles of homoduplex DNA.
Figure 4-7

A: - MutL

B: + MutL

[MutH] (pmoles)

WT

Bound DNA

Free DNA

Δ214

G49A

K48A

WT

(pmoles)

(pmoles)

(pmoles)
From these experiments, the apparent dissociation constants ($K_d$) for the proteins and the substrate were calculated (Fig. 4-8). The binding affinity for the wild-type MutH protein increased at least 16-fold when the MutL protein was present (Fig. 4-8B). K48A was as proficient as the wild-type protein in binding ability, but the activity in the endonuclease assay was reduced, so it has a catalytic defect that MutL is able to overcome. G49A bound 2 to 4-fold tighter than wild-type, which may or may not be a significant contribution to the overall ability of the protein to function. Since Δ214 did not achieve the full band shift position, the apparent dissociation constant was not calculated. While MutL has an effect on MutH binding to the DNA (about 16-fold), the greatest effect is on catalysis (one thousand-fold stimulation). The apparent $K_d$ values of the wild-type and mutant proteins and their substrate are given in Table 4-1.
Figure 4-8. **Dependence of complex formation on the concentration of the MutH protein.** The graphs were plotted by fitting the experimental data from the band shift assay to the equation described in the Materials and Methods. This allows the calculation of an apparent $K_d$. Panel A, the data plotted from the band shift assays without MutL present. The wild-type protein (■) and the G49A mutant (●) are shown in the graph. Panel B, the data plotted from the band shift assay containing MutL. The wild-type protein (■) and the G49A mutant (●) are shown in the graph. $r$ is the degree of binding and [C] the MutH concentration in the reaction. Each point represents the average of at least three independent measurements.
Figure 4-8

A  - MutL  

B  + MutL  

Graph showing the effect of MutL on reaction rates.
Table 4-1. Apparent equilibrium dissociation constants for wild-type and mutant MutH proteins to hemimethylated DNA.

<table>
<thead>
<tr>
<th>MutL</th>
<th>$K_d$ (uM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+)</td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>4.2 (±0.7)</td>
</tr>
<tr>
<td>K48A</td>
<td>9.8 (±0.7)</td>
</tr>
<tr>
<td>G49A</td>
<td>1.0 (±0.5)</td>
</tr>
<tr>
<td>(-)</td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>68.2 (±0.6)</td>
</tr>
<tr>
<td>K48A</td>
<td>68.7 (±10)</td>
</tr>
<tr>
<td>G49A</td>
<td>27.4 (±8)</td>
</tr>
</tbody>
</table>

(±) represents the standard error for the $K_d$'s.
Structural Analysis of MutH

Structural model. The calculated electron density map showed one long helix with several smaller helices and beta sheets comprising 40% of the peptide chain. These structures could not be connected because the density was not continuous. After reexamining and refining the calculations, no apparent problems were revealed. With the strength of the heavy atom derivatives, the problem was thought to be with the crystal form.

A few months after I had done the peptide tracing the structural model of MutH was published\textsuperscript{14}. The basic secondary structures were the same as the ones seen in my electron density map. Crystals with a different space group were used to solve the problem of discontinuous electron density. Ban and Yang had grown MutH crystals in two space groups, P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} and P2\textsubscript{1}. The P2\textsubscript{1} space group allowed 80% of the peptide chain to be traced before further refinement. It appears that the MutH protein was too flexible when crystallizing in the P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} space group to calculate a continuous electron
density map. Ban and Yang used molecular replacement to solve the structure in the P2_12_12 space group. The published structures were without DNA bound to the protein. Mg^{2+} was also not observed, even though it was included in the crystallization solution. Ban and Yang proposed an open and closed conformation of the MutH protein from their published structures showing how the protein is able to rotate and pivot about its' two domains.

The MutH structure was compared to a number of restriction endonucleases with which MutH shares a common core motif. This comparison led to a number of predictions about functional areas of the protein. It contains a large cleft of dimensions (15-18 Å wide by 12-14 Å deep) that would allow DNA to bind. Presumed catalytic amino acids were identified (D70, E77, and K79) as well as possible DNA binding residues (loop BC, loop67, D91, and F94). Three different positions of the MutH protein were found and the three structures were presented in the reference. These three positions highlight the flexibility of the protein and reveal a COOH-terminal tail that protrudes from the body of the protein in one position and is retracted in another. This raises the question of a possible stimulation site for MutL. Therefore, the focus of my project changed to the mutational analysis of the MutH protein.
Mutational Analysis of MutH

**Wild-type activity.** The native MutH protein that was purified was able to complement a mutH deficient strain in the mismatch repair assay. The protein also demonstrated similar endonuclease activity to the previously published activity for MutH. The purified His-tagged MutH protein gave the same results.

The band shift assays with the native MutH protein reveal that all methylation states of the d(A) in the d(GATC) site of the DNA substrate are bound equally well. *In vivo*, the fully methylated state of DNA isn't processed by the mismatch repair system. The MutH protein at cellular concentrations may not be able to nick the fully methylated DNA, but it has the ability to bind. Some restriction endonucleases have been shown to exhibit this ability\(^7\). At very high concentrations, the MutH protein is able to nick the fully methylated DNA homoduplex as with some restriction endonucleases that cut a noncleavable site at a very high protein concentration\(^1\).

**In vivo screen.** Methyl-directed mismatch repair is initiated by a mismatch in the DNA followed by formation of a ternary complex containing the MutS (with bound ATP), MutL, and MutH proteins. A defect in any one of these three proteins causes a mutator phenotype. In this work, that phenotype was used as a screen to identify defective MutH proteins after site-directed mutagenesis. Since the mutated residues are found in flexible loops not many mutants would be expected and the *in vivo* results
indicated that only three of the site-directed mutations, K48A, G49A, and Δ214, exhibited a mutant phenotype in *mutH* *E. coli*. Subsequent biochemical assays confirmed a defect in MutH activity in these mutant proteins. Four site-directed mutants having a wild-type mutation frequency to rifampicin resistance (D47A, H112A, H115A, and Δ224) displayed normal MutH activity *in vitro*. These results indicate that the *in vivo* screen was sensitive enough to detect defective MutH proteins. However, given that the mutant proteins were expressed from a high copy number plasmid, an estimated ten-fold reduction in activity of MutH would probably not have been detected.

**Endonuclease assay for mutants.** The qualitative results between the biochemical and *in vivo* assays agree fairly well. Mutant Δ214 had the highest mutation frequency of the mutants tested in the *in vivo* screen and had the highest reduction in endonuclease activity in the presence of MutL. Mutants G49A and K48A somewhat follow this pattern with G49A having the lower mutation frequency and achieving complete nicking in the endonuclease assay (with increased protein amounts). However, the K48A mutant (reduced 3-fold) had a higher mutation frequency while the G49A (reduced 5-fold) mutant maintained a lower frequency. These discrepancies could be due to an absence of both the mismatch in the MR1 substrate and MutS protein in the endonuclease assay (to fully initiate the repair system) as there would be in the *in vivo* screen. How well MutL is able to stimulate the MutH mutants *in vivo* versus *in vitro* may not correlate well and also contribute to the differences seen. A final aspect is the concentration of MutH present in the assays. It takes very few copies of the protein (estimated at 34 monomers per cell) to
achieve full repair activity\textsuperscript{70}. The copies of protein from the plasmid are probably much higher than those normally in the cell.

**Band shift assay for mutants.** Although the stimulation of MutH activity by MutL has been shown at the endonuclease level\textsuperscript{13,69}, the ability of MutL to stimulate the DNA binding ability of MutH has not been reported. The amount of MutL (9 picomoles) in the band shift assay appears to be adequate for binding all of the MutH (0.1 picomoles) monomers present. Yet the amounts of MutH needed to achieve a full band shift remain the same (Chapter 4 Fig. 6). One conclusion from the data is that the MutL contribution to MutH binding is marginal. The differences in $K_d$'s were about 16-fold. This is significant, but adds very little compared to the one thousand-fold difference MutL makes in catalysis. MutL's presence makes it a little easier for MutH to start binding to the DNA, but overall binding is not affected very much.

**DNA binding residues.** The mutational changes made in MutH were based upon a comparison with the $Pvu$II enzyme where detailed structure-function data are available. Ban and Yang have reported that excluding the first 33 amino acids of $Pvu$II, MutH and $Pvu$II can be superimposed with an r.m.s. deviation of 2.3 over 83 pairs of Cα atoms\textsuperscript{14}. Mutational analysis of the amino acids in the flexible arms would show whether or not MutH has any functional similarity to $Pvu$II. In $Pvu$II there are a few residues present that are involved in DNA binding or the recognition of the binding site; such as an aspartate (D34) that would be comparable to D47 in MutH\textsuperscript{75}. No residues in the flexible
arms of MutH were found to be involved in DNA binding. The amino acids chosen for biochemical studies are shown in Fig. 6-1.

Histidine residues. The conserved histidines found in both PvuII (H84 and H85) and MutH (H112 and H115) do not appear to have the same function. Pvu II is a restriction endonuclease from *Proteus vulgaris* that hydrolyzes the phosphodiester bond in unmethylated DNA. Changing the histidines (H84 and H85) in PvuII results in a loss of catalysis but not binding. Their displacement (seen from the structure) upon unmethylated DNA binding brings the catalytic residues into proper alignment for catalysis. When the DNA is fully methylated, the excess displacement of these histidines (again from a structure) causes the catalytic residues to move further out of alignment and catalysis cannot take place. No loss of binding or catalysis resulted when the histidines were substituted with alanines in MutH. The signal (for DNA binding and state of DNA methylation) for MutH, remains unknown. The DNA binding domain for MutH is clearly not similar to the one in PvuII.
Figure 5-1. **Location of the mutated residues in the MutH structure.** The ribbon diagram of MutH was made from the B monomer with the coordinates submitted by W. Yang (Protein Data Bank code 2AZO) using MIDAS. The NH$_2$ and COOH-termini are labelled on the diagram. The residues tested from the site-directed mutagenesis (D47, K48, G49, H112, and H115) are displayed in red. Also the last residue remaining from each of the deletion mutants are shown in red (K214 and R224).
Lysine 48. A lysine (K48) was found in MutH that functions in catalysis. The lysine to alanine mutation (K48A) in MutH resulted in a protein devoid of activity in the absence of MutL. MutL may have a lysine or other residue that is able to give the ternary complex partial function. The function of lysine 48 in MutH is unclear. It is positioned within the BC loop approximately 13.9 Å from the lysine presumed to be the catalytic amino acid. Upon the subsequent movement of MutH associated with DNA binding and activation of the protein, lysine 48 could be positioned within the catalytic core of the protein at the time of catalysis. This lysine may coordinate a critical water molecule for catalysis, be structurally important for the placement of the catalytic amino acids, or may be the catalytic lysine. The last possibility is very remote and a crystal structure with bound DNA and Mg^{2+} would be needed to confirm one of these hypotheses.

Glycine 49. The G49A mutant may have defects in catalysis because of its proximity to lysine 48. The PHI and PSI values for G49 (obtained from the 2azo.pdb crystal structure by Dr. Donald Nelson, Clark University Worcester, MA) are as follows: PHI dihedral = -114.424 and PSI dihedral = +25.289. These values place the phi-psi coordinate in the upper left hand quadrant of the Ramanchandran diagram, a region of low energy for most residues. The extra methyl group of G49A may cause slight structural changes in the lysine 48 side chain when the protein is in a functional conformation that slows catalysis. When MutL is present, the mutant is able to achieve full nicking of the substrate with larger amounts of protein than the wild-type MutH. The DNA binding of this glycine mutant is slightly tighter than the wild-type. The methyl
group of G49A may be able to reach the phosphates of the DNA. If the lysine 48 next to this glycine is working in catalysis, then the DNA is very close to these residues. Molecular dynamic simulations may be able to yield some information in regard to the functioning of this mutant. This mutational change interferes with the proper functioning of the protein but is not informative about the mechanism of action.

**C-terminal deletion mutants.** Two C-terminal deletion mutations (Δ214 and Δ224) were made to help identify the residues that contact and cause the stimulation of MutH by MutL. This idea is based on predictions from the structural model of MutH\(^\text{14}\). The Δ224 functions similarly to the wild-type protein *in vivo* as well as *in vitro* indicating that residues 225 to 229 at the C-terminal end of the protein are not required for MutL stimulation. The mutation frequency of the Δ214 mutant is elevated in the *mutH*-deleted strain *in vivo* screen and shows decreased cleavage activity with some initial binding capabilities *in vitro*. However, it is unable to be stimulated by MutL. This places the residues between 215 and 224 as being involved in activation by MutL. It has been shown that a deletion of 10 amino acids in the C-terminal region of MutH (Δ219) prevents MutL stimulation\(^\text{69}\). That leaves the amino acids 220 through 224 (sequence: A,L,L,A,R) as contacting MutL and stimulating MutH activity. Comparing the amino acid sequence of *E. coli* and *Haemophilus influenza* MutH, there is one conserved amino acid. It is the leucine at position 222. Site-directed mutagenesis at each of the five positions (220-224) could reveal the most important residues for stimulation.
Future directions. There remains a lot of missing information about the functioning of MutH. A co-crystal of MutH, Mg\(^{2+}\), and cognate DNA would reveal the most immediate information. The importance of the K47 residue in catalysis could be confirmed as well as other DNA binding residues. Some of those residues would be base specific for site recognition and some methylation sensitive for strand discrimination. MutH is unique in that it prefers hemimethylated DNA as a substrate for catalysis. Differences between the DNA binding contacts as compared to similarly folded restriction endonucleases might show where the basis for strand discrimination may be found. Site-directed mutagenesis would need to be done for those residues to confirm any predictions. It could also be shown that the presumed catalytic amino acids (D70, E77, and K79) are correctly positioned for catalysis. Looking farther ahead, a crystal complex of MutH and MutL with Mg\(^{2+}\) and DNA would be able to show much more information. The specific amino acids responsible for MutH stimulation by MutL have not been identified. This crystal would be able to identify all of the contact sites between the two proteins. Site-directed mutagenesis could then determine which residues are necessary for the stimulation. Until then, site-directed mutagenesis needs to be done on the five residues (220-224; A,L,L,A,R) that the research has implicated in stimulation. Molecular modelling and molecular dynamic simulations might also help elicit information.

Understanding how MutH functions in the complex for the initiation of mismatch repair is another goal. MutH binds DNA very poorly and is a slow endonuclease alone. It could not support the rates of mismatch repair that are found either in vivo or in vitro.
MutL appears to help catalysis tremendously but only increases binding ability slightly. MutH may have trouble getting onto the DNA but is able to bind the d(GATC) site efficiently. In the complex, MutS and MutL may help MutH onto the DNA and then when a d(GATC) site is found MutH binds. This assumes that the complete complex is formed at the mismatch and translocates along the DNA until the MutH binding site is found. Some evidence has been found that loops are formed through a translocation mechanism by MutS and both MutS and MutL can be found at the base of the loop. The addition of MutH to this loop experiment has not been performed as yet. The fact that they form a complex in the cell has not been proven. The use of fluorescent proteins may be able to show that the three proteins (MutS, MutL, and MutH) actually co-localize in the cell. There is room for much research to be done.
CHAPTER VI

REFERENCES


