1991-11-01

Repair of DNA Containing Small Heterologous Sequences by Escherichia Coli: a Dissertation

Breck Olland Parker

University of Massachusetts Medical School

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

Part of the Amino Acids, Peptides, and Proteins Commons, Bacteria Commons, and the Genetic Phenomena Commons

Repository Citation


https://escholarship.umassmed.edu/gsbs_diss/123

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Repair of DNA Containing Small Heterologous Sequences by *Escherichia coli*.

A Dissertation Presented
by
Breck Olland Parker

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

**DOCTOR OF PHILOSOPHY IN MEDICAL SCIENCES**

**PHARMACOLOGY**

November 1991
Repair of DNA Containing Small Heterologous Sequences by *Escherichia coli.*

A Dissertation By

Breck Olland Parker

Approved as to style and content by:

Larry W. Hardy  
Chairman of Committee

Eric von Hofe  
Member of Committee

Michael R. Volkert  
Member of Committee

Kendall L. Knight  
Member of Committee

Bruce F. Demple  
Member of Committee

Martin G. Marinus  
Thesis Advisor

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

Department of Pharmacology  
November 1991
ACKNOWLEDGEMENTS

I want to extend my sincere gratitude to my thesis advisor, Martin Marinus, for guiding me through the rigors of graduate school. I’m especially grateful for his patience and encouragement when times were rough, and for sharing in my excitement and enthusiasm when things went well. His guidance and friendship over the last five years is greatly appreciated.

I am also very grateful to the members of my thesis advisory committee: Mike Volkert, Bob Lahue and Larry Hardy, for their constant encouragement, invaluable advice and infectious enthusiasm for science. Thanks Larry, for always providing that different point of view, for always asking the one question that keeps me humble, and for making me work harder for fear that you’ll continue to ask such questions.

I extend my gratitude to Neal Brown and the rest of the Pharmacology department faculty for their support and guidance over the last five years. A special thanks to our departmental statistician, Bob Lew, for those early morning and weekend philosophical interludes while waiting for the coffee to brew.

Finally, my deepest gratitude goes to my family. Their steadfast support and encouragement have helped to make this achievement possible. I especially want to thank my wife, Linda, who has been my partner in everything over the last seven years, for her love, support and patience, my son, Andrew, who has been a constant source of inspiration and motivation, and my Mom and Dad, who have been a stabilizing influence throughout my life and have always supported and encouraged me regardless of what I chose to do.

P.S. Thanks Fran, for constantly reminding me that there is life outside of science, i.e. Walter’s opprobrious legend. Thanks for the Bud.
For his love of life, love of family and love of God,
I dedicate this thesis to

Roland C. Hallen.

(Aug. 16, 1907 to July 1, 1983)
Prostate Cancer
The Dam-dependent mismatch repair system of *Escherichia coli* is part of a large network of DNA surveillance and error avoidance systems that identify and repair DNA damage. In this thesis, I have investigated the Dam-dependent mismatch repair system of *E. coli* and its role in the recognition and repair of DNA substrate molecules containing small insertion/deletion heterologies. This investigation was divided into two parts: the first part utilized genetic techniques to evaluate the specificity of repair and the second part utilized biochemical approaches into the recognition of insertion/deletion heterologies.

I have developed a sensitive *in vivo* transformation system to rapidly evaluate the repair of small insertion/deletion heterologies by Dam-dependent mismatch repair. Heteroduplexes were constructed, for each state of methylation of d(GATC) sequences, by
annealing single strand DNA to the linearized complementary strand of duplex DNA. The unmethylated single strand DNA was isolated from f1 phage (R408) propagated on a strain of E. coli containing the dam-16 allele (Chapter 2) to eliminate the possibility of residual Dam-methylation of d(GATC) sequences. Transformation of E. coli indicator strains with heteroduplexes containing 1, 2, 3, 4 and 5 base insertion/deletion heterologies were scored for repair based on colony color. The results of these experiments show that the Dam-dependent mismatch repair system can recognize and repair 1, 2 and 3 base heterologies as well as repairing G/T mispairs (Chapters 3 and 4). The repair of 4 base heterologies was marginal, while no repair was observed with 5 base heterologies (Chapters 3 and 4). Repair of the 1, 2, 3 and 4 base heterologies proceeded in a Dam-dependent process that required the gene products of mutL, mutS, and 

I have demonstrated that MutS protein from both Salmonella typhimurium and E. coli can recognize and bind in vitro to the same 1, 2, 3 and 4 base heterologies used for the genetic studies above (Chapters 4 and 5). In fact, MutS protein binds to 1, 2 and 3 base heterologies with greater affinity than it binds to a G/T mismatch. The in vitro observation that MutS does not bind to 5 base heterologies is consistent with the in vivo observation that 5 base heterologies are not subject to repair. I have also shown that MutS protein specifically binds to 1, 2 and 3 base heterologies since MutS protects about 25 base pairs of DNA flanking the site of the heterology from DNaseI digestion.

The results of the genetic and biochemical experiments described in this thesis (and summarized above) serve to re-emphasize the importance of the role that methyl-directed mismatch repair plays in mutation avoidance, and hence in the preservation of genetic integrity.

vi
TABLE OF CONTENTS

Title Page............................................................................................................................... i
Signature Page........................................................................................................................ ii
Acknowledgements.............................................................................................................. iii
Dedication............................................................................................................................... iv
Abstract................................................................................................................................... v
Table of Contents.................................................................................................................. vii
List of Tables........................................................................................................................... xii
List of Figures........................................................................................................................ xiii

CHAPTER 1: Introduction........................................................................................................ 1

A. The conservation of genetic information......................................................................... 2
   Systems involved in repairing DNA as a result of endogenous damage.............................. 3
   Systems that protect against environmental and chemical insult...................................... 5
   Initial misincorporation of deoxynucleotides is dependent on base selection by DNA polymerase III holoenzyme and the gene product of mutT.
   The 3’-5’ exonucleolytic proofreading function of DNA polymerase III............................ 7
   Dam-dependent mismatch repair corrects errors after passage of the replication fork. .......... 9
B. d(GATC) Sequences and Dam-dependent Mismatch Repair........................................... 11
   DNA adenine methyltransferase (Dam) is encoded by the dam gene................................. 11
Dam-methylation of adenine in d(GATC) sequences directs the strandedness of repair.

Dam-dependent mismatch correction requires at least one d(GATC) site for efficient correction of a mismatch.

C. Specificity of Dam-dependent Mismatch Repair

Transition mispairs are corrected more efficiently than transversion mispairs.

Repair of small frameshift mutations

D. The Biochemistry of Dam-dependent Mismatch Repair

Isolation and characterization of MutS

Isolation and characterization of MutH

Isolation and characterization of MutL

Components of excision and DNA resynthesis steps

Summary of model for Dam-dependent mismatch repair

E. Mismatch Correction and Genetic Recombination

F. Mismatch Specific Repair Pathways: VSP and mutY

VSP repair pathway

mutY repair pathway

G. Statement of the Problem

CHAPTER 2: A Simple and Rapid Method to Obtain Substitution Mutations in Escherichia coli: Isolation of a dam Deletion/Insertion Mutation.

A. Summary
C. Materials and Methods........................................................................................................ 74
D. Results.................................................................................................................................. 77

The *in vivo* repair assay can determine the repair efficiency of small DNA heterologies.

*In vivo* correction of small DNA heterologies by mismatch repair is Dam-dependent.

*In vivo* correction of unmethylated heteroduplexes proceeds without strand bias.

*In vivo* correction of symmetrically methylated heteroduplexes proceeds independently of Dam-methylation and MutH.

Dam-dependent repair of small heterologies in the *mnt* operator................................. 87

MutS binds to small insertion/deletion heterologies *in vitro*........................................... 88

E. Discussion.............................................................................................................................. 90

F. Materials and Methods Supplement.................................................................................. 99

CHAPTER 5: MutS Protein of *Escherichia coli* Binds to DNA Containing Small Heterologous Sequences.

A. Abstract................................................................................................................................. 107

B. Introduction............................................................................................................................ 108

C. Materials and Methods......................................................................................................... 109

D. Results.................................................................................................................................. 114

MutS protein binds to 1, 2, 3 and 4 base heterologies...................................................... 114

MutS competition binding assays......................................................................................... 116
CHAPTER 6: Final Comments and Future Directions

A. Final Comments

B. Future directions

CHAPTER 7: References
LIST OF TABLES

Table

Chapter 2.
I  Plasmids and strains of *E. coli* K-12 used in this study..............................39
II  Distribution of markers in transductional crosses involving..............................40
dam-16, *aroB*, and *trpS*.

Chapter 3.
I  Transformation of GM4331 and GM4348 with control plasmids and heteroduplexes........61
II  Retransformation of red, white and mixed colonies........................................64
III  Concentration of tetracycline required to detect false white colonies.................67

Chapter 4.
I  Genetic analysis of hemimethylated heteroduplex transformations..........................82
II  Genetic analysis of unmethylated heteroduplex transformations..........................85
III  Genetic analysis of symmetrically methylated heteroduplex transformations........86
LIST OF FIGURES

Figure

Chapter 1.
1-1 Model for methyl-directed mismatch repair......................... 29

Chapter 2.
2-1 Construction of a chromosomal substitution mutation in the........... 39
dam-gene.
2-2 Southern blot analysis of the chromosomal dam gene region........... 40

Chapter 3.
3-1 The plasmid map of pPY97..................................................... 51
3-2 Location and DNA sequence of deletion mutations within the mnt....... 54
operator.
3-3 Repair of heteroduplexes containing either 1, 2, 4, 5, 7 or 15.......... 55
base insertion/deletion mutations within the mnt operator.
3-4 Repair of 1, 2, 4 and 5 base insertion/deletion mutations within........ 57
the mnt operator.
3-5 Description of the mnt structural gene system.......................... 59
3-6 Colonies on MacConkey/lactose are either red, white or "mixed"......... 62

Chapter 4.
4-1 DNA sequence map of the deletion mutations within the mntgene....... 79
Band shift assays using *Salmonella typhimurium* MutS protein............. 89

Scheme for site directed mutagenesis for the construction of............... 100
1, 2, 3, 4 and 5 base deletions within the mnt gene.

Scheme for the construction of heteroduplexes containing 1, 2, 3.............. 104
4 and 5 base heterologies within the mnt structural gene.

Chapter 5.

5-1 Map of deletions used in the mobility shift assays.............................. 111

5-2A Autoradiogram of the mobility shift assay using *E. coli* MutS.............. 115
protein.

5-2B Mobility shift assays using *E. coli* MutS protein and 33 base.............. 115
heteroduplexes containing 1, 2, 3, 4 and 5 base deletions.

5-3 Mobility shift assays: Competition studies with radioactively............... 117
labelled 1 base deletion against 40-fold excess cold competitor DNA.

5-4 Mobility shift assays: Competition studies with radioactively............... 118
labelled 2 base deletion against 40-fold excess cold competitor DNA.

5-5 Mobility shift assays: Competition studies with radioactively............... 119
labelled 3 base deletion against 40-fold excess cold competitor DNA.

5-6 *E. coli* MutS DNA footprints: DNaseI protection assays.......................... 121

5-7 DNA sequence of the MutS footprints in Figure 6............................. 122
CHAPTER 1

Introduction: Dam-Dependent Mismatch Repair.
CHAPTER I.
INTRODUCTION

A. The conservation of genetic information in *Escherichia coli* is mediated by systems that protect its DNA against environmental as well as endogenous insult, and from errors of DNA replication.

In all forms of life, the conservation of genetic information from one generation to the next occurs with remarkable fidelity, thanks in part to an exquisite network of specialized systems that work to identify and repair damaged DNA. Damage to DNA can occur as a result of exposure to ultraviolet radiation from the sun, environmental toxins, spontaneous methylation and deamination of nucleic acids, and errors of replication. Such events are biologically significant since altering the genetic integrity of DNA can precipitate the onset of potentially lethal disease states, i.e. cancer. Exactly how cells guard against such mutational events is an area of intense research. Prokaryotes, such as *Escherichia coli*, have proven to be excellent models for studying systems involved in detecting and correcting mutagenic lesions. The knowledge gained from such studies, and those of other organisms, may help elucidate the complexities of mammalian DNA damage surveillance and error avoidance systems. The Introduction to this thesis will discuss in very general terms, in the context of *Escherichia coli*, a few examples of repair pathways involved in repairing DNA damage as a result of environmental, chemical and
endogenous insult. This will be followed by an in depth review of error avoidance systems that protect DNA against errors of replication. To this end, I will focus primarily on the *E. coli* Dam-dependent mismatch repair system, also known as the methyl-directed mismatch repair system.

**Systems involved in repairing DNA as a result of endogenous damage.** Of the four bases that normally occupy the DNA molecule, three (adenine, guanine, and cytosine) contain exocyclic amino groups. The deamination of these molecules, cytosine being of particular importance since it represents 98% of the observed deaminations (Friedberg, 1985), are biologically significant reactions that occur under normal physiological conditions (Friedberg, 1985). For example, the spontaneous deamination of cytosine to uracil produces G/U base pairs. Since uracil base pairs with adenine during replication, a G/C to A/T transition mutation event can occur (Duncan and Miller, 1980). Such mispairs can be prevented by the action of uracil DNA glycosylase which catalyses the removal of uracil from G/U basepairs through hydrolysis of the N-glycosidic bond producing an abasic site, the processing of which requires AP endonuclease activity (Lindahl, 1982). An analogue of cytosine, 5-methylcytosine, is also subject to deamination resulting in the conversion of 5-MeC to thymine creating a G/T base mispair (Coulondre et al., 1978). This mispair is subject to Very Short Patch (VSP) repair (Lieb, 1983) and will be discussed in greater detail later.

Another form of DNA damage results from the spontaneous hydrolysis of the N-glycosylic bond whereby the base is released from the DNA giving rise to an abasic site. Depurination represents the major form since it occurs spontaneously 20 times more
frequently than depyrimidination (Lindahl and Karlstrom, 1973; Friedberg, 1985). The generation of abasic lesions, in addition to those produced by specific DNA glycosylases (Sakumi and Sekiguchi, 1990), are of particular importance since they may block DNA replication or cause miscoding (Echols, 1991). AP endonucleases are responsible for recognizing and repairing these lesions by making an incision, depending on the AP endonuclease class (I or II), either 3’ to or 5’ to the deoxyribose-sugar moiety (Doetsch and Cunningham, 1990). Further processing is likely since the AP endonuclease nicked site does not always provide an appropriate substrate for DNA polymerase I (Doetsch and Cunningham, 1990).

Naturally occurring non-enzymatic processes within the cell can also produce mutagenic lesions. For example, S-adenosylmethionine (SAM), a common cofactor in reactions that require the presence of a methyl donor, can act as a very weak alkylating agent. Through a spontaneous, non-enzymatic process that occurs under normal physiological conditions, SAM can methylate the N7 and N3 positions of guanine and adenine respectively (Ryberg and Lindahl, 1982). The presence of SAM induced 7-methylguanine moieties in DNA is about 7 fold greater than that observed for 3-methyladenine (Ryberg and Lindahl, 1982). Under normal conditions, the presence of 7-methylguanine in DNA is of little consequence, although the potential for depurination can not be ignored. As discussed above, depurination gives rise to abasic sites which, if not repaired, may promote miscoding (Friedberg, 1985). Conversely, 3-methyladenine poses the more serious threat since these adducts often result in lethal events presumably by blocking the replicative machinery (Friedberg 1985). Therefore, it’s not surprising that 3-methyladenine DNA glycosylase I is constitutively expressed in E. coli specifically to remove
3-methyladenine from DNA (Sakumi and Sekiguchi, 1990).

DNA base modifications can also arise as a result of endogenous oxidative damage. For example, 8-oxoguanine can be produced when dG is attacked by reactive oxygen species formed as a consequence of normal metabolism. 8-oxoguanine adducts are biologically significant since they can base pair with adenine to form a 8-oxoG/A mispairs unless removed prior to replication (Shibutani et al., 1991). The repair of this lesion is presumably carried out by the product of mutM (Michaels et al., 1991), formamidopyridine DNA glycosylase (Fapy-DNA glycosylase), to which 8-oxoguanine serves as a substrate (Tchou et al., 1991). Unlike uracil DNA glycosylase, Fapy-DNA glycosylase possesses dual glycoslyase and AP nicking activity (O'Connor and Laval, 1989) that removes 8-oxoguanine from DNA. Genetic studies performed in E. coli strains carrying mutations in the mutM gene show a marked increase in G/C to T/A transversion mutations in agreement with biochemical observations (Cabrera et al., 1988).

**Systems that protect against environmental and chemical insult.** *Escherichia coli* has evolved a vast network of enzymatic pathways designed to respond to chemically and environmentally induced DNA damage (Friedberg, 1985). An example of this is exposure to ultraviolet radiation which can result in the formation of pyrimidine dimers between two adjacent pyrimidine moieties. Since these adducts interfere with proper base pairing and can promote the premature termination of DNA replication, survival depends on efficient correction of the damaged DNA. Therefore, it is not surprising that multiple pathways exist to process pyrimidine dimers. These include repair by either photolyase (Houten, 1990), which binds to the dimer and catalyzes its separation back
into individual pyrimidine bases, and by nucleotide excision repair which requires the gene products of the uvrABC genes to catalyze a process that recognizes and removes a short dimer containing excision tract (Houten 1990; Grossman and Yeung, 1990).

DNA damage resulting from exposure to reactive chemicals or alkylating agents can be repaired by DNA glycosylases in a process similar to that discussed earlier or by various methyltransferases. For example, O⁶-methyltransferase catalyzes the removal of the methyl group from the 6-position of O⁶-methylguanine to restore guanine to its native state, while 3-methyladenine DNA glycosylase II serves to repair 3-methyladenine, 7-methylguanine and 3-methylguanine moieties. Other forms of chemically induced damage, such as that of acridines and proflavin, promote the formation of +1/-1 frameshift mutations (Skopek and Hutchinson, 1984). The correction of frameshift mutations is carried out by the Dam-dependent mismatch repair system (Skopek and Hutchinson, 1984; Parker and Marinus, 1991; Chp. 4).

**Initial misincorporation of deoxynucleotides is dependent on base selection by DNA polymerase III holoenzyme and the gene product of mutT.**

Proper insertion of deoxyribonucleotides during DNA replication requires the correct Watson/Crick base pairing and nearly equivalent Cl’ bond angles and distances (for a review, see Echols, 1991). Improper geometry may impede the insertion of a dNTP with a non-complementary base by slowing down the formation of phosphodiester bonds and/or promoting dissociation of bases from the polymerase active site (Echols, 1991). The major mispairs formed as a result of misincorporation in the absence of proofreading and mismatch repair are G/T and C/A transitions and G/A transversions (Schaaper, 1988;
Palmer and Marinus, 1991), all of which exhibit Cl' bond angles substantially different than that required for perfect Watson/Crick base pairing. The error frequency of base insertion by DNA Pol III is approximately $10^{-4}$ (Echols, 1991).

Recent observations have suggested that the gene product of mutT, isolated to near homogeneity as a 15 kDa protein (Bhatnagar and Bessman, 1988), assists DNA Pol III in preventing the formation of A/G transversion mispairs by selectively removing the (syn) form of dGTP (Akiyama et al., 1989). Since the (anti) form of dGTP is the proper substrate for G/C base pairing, the selective removal of the (syn) form of dGTP, which can base pair with dA, decreases the probability of A/G mispairing. The importance of mutT to replicational fidelity is strengthened by the observation that some A/G mispairs that are normally prevented by MutT are not substrates for repair by methyl-directed mismatch repair (Schaaper, 1989).

Errors of DNA replication are recognized and corrected by DNA polymerase III holoenzyme 3'-5' exonucleolytic proofreading function.

The proofreading function of DNA Pol III, which increases the fidelity of replication $10^2$ to $10^3$ fold (Echols, 1991), is encoded by the dnaQ (mutD) gene as the $\varepsilon$-subunit (Scheuermann et al., 1983). This subunit displays a 3' to 5' exonuclease proofreading activity that, while independent of the catalytic subunit ($\alpha$) of PolIII (Scheuermann and Echols, 1984), is enhanced 10 to 80 fold when present as an $\alpha-\varepsilon$ complex (Maki and Kornberg, 1987). The mechanism involved in mispair recognition by the $\varepsilon$-subunit is still unclear. However, it has been proposed that the 3'-5' exonuclease proofreading activity has more time to recognize and remove a mispaired base since polymerase III extension
after incorporation of a mismatched base is often slowed by helical instability at the site of the mismatch (Echols, 1991).

Mutations in the mutD allele, e.g. mutD5, result in a hypermutable phenotype (Fowler et al., 1974; Cox, 1976). In general, mutations are of the single base substitution transition/transversion type and an occasional +1/-1 frameshift mutation (Lorenzetti 1983, Fowler et al., 1986, Schaaper 1988, Wu et al., 1990). The type of mutations produced in the lacI reporter gene in mutD5 strains are, in part, a function of growth conditions. The observed mutation frequency of mutD5 cells grown on minimal salts is approximately 480 times that of wild type cells producing a mutation spectrum of which 62% were of the transversion type (Schaaper, 1988). Conversely, mutD5 cells grown on rich media produced mutations, of which 90% were transition mutations, at a frequency 37,000 times that of wild type cells (Schaaper 1988). Schaaper (1988) proposed that mutD5 cells grown in rich media are defective in mutHLS mismatch repair in addition to defective exonucleolytic proofreading. To examine this, mismatch repair was evaluated by transfecting wild type, mutL and mutD5 cells grown in minimal and rich media with M13mp2 hemimethylated heteroduplex DNA containing a G/T mismatch. The results indicated that methyl-directed mismatch repair operated efficiently in mutD5 strains grown on minimal media, but was quite inefficient when grown on rich media.

Schaaper (1988) suggested that either the ε-subunit interacted directly with the mismatch repair machinery or that lack of 3'-5' exonucleolytic proofreading during periods of rapid growth saturates the mismatch repair system with an excess of replication errors. The latter explanation is the best supported by the following observations. First, mutD5 cells were deficient in mismatch repair during log phase growth, while efficient mismatch
repair was restored when the mutD5 cells reached plateau phase of growth (Schaaper and Radman, 1989; Schaaper, 1989). Second, mismatch repair of heteroduplexes in mutD5 cells during early log phase is restored to that of wild type after cells containing a dnaA (Ts) allele are shifted to a non-permissive temperature, thus arresting chromosomal replication (Schaaper and Radman, 1989). Third, the presence of a multicopy plasmid containing either mutL or mutH restored efficient mismatch repair of G/T mispair M13mp2 heteroduplexes during log phase growth in mutD5 cells (Schaaper and Radman, 1989). Finally, a mutD5-dam-16 double mutant produced a spectrum characteristic of mismatch deficient cells similar to that observed when mutD5 cells are grown in rich media (Palmer and Marinus, 1991). In agreement with a recent study by Wu et al. (1990), these observations suggest that mismatch repair functions predominantly to remove transition mutations, while the proofreading exonuclease of DNA polymerase III functions predominantly to remove transversion mutations.

**Dam-dependent mismatch repair corrects errors after passage of the replication fork.**

The following discussion will focus on the Dam-dependent mismatch repair pathway of *Escherichia coli* since it is best characterized in this organism (for reviews see, Radman and Wagner 1986; Modrich, 1987; Meselson, 1988; Modrich, 1991). The existence of a mismatch repair pathway was originally envisioned by Robin Holliday (Holliday, 1964) as an explanation for gene conversion arising as a consequence of genetic recombination. Support for a mismatch repair pathway was provided by λ-heteroduplex transfection experiments using *E. coli* as host (Wildenberg and Meselson, 1975;
Wagner and Meselson, 1976). λ-heteroduplexes were constructed by annealing complementary strands of λ-DNA, with each strand harboring allelic differences, to produce a molecule containing multiple genetic markers. Since each strand contains allelic differences, mismatched base pairs are formed. Repair of the heteroduplexes can be analyzed by determining the genotype of the progeny phage after transfection. Results from such experiments showed a high incidence of multiple repair events involving adjacent mismatched bases residing on the same strand (Wildenberg and Meselson, 1975; Wagner and Meselson, 1976). Their observations led to the proposal that the repair process involved long excision repair tracts spanning the adjacent mismatches on one strand. This gives rise to progeny phage favoring the genotype of the strand opposite the excised strand. From these experiments, it was estimated that excision repair tracts could be as long as 3,000 bases since simultaneous repair of markers spaced in excess of 2,000 base pairs were observed at high frequencies (Wagner and Meselson, 1976).

It was subsequently proposed that mismatch repair could be advantageous if it corrected replication errors within the newly synthesized daughter strand (Wagner and Meselson, 1976). For this to function properly, it required a signal that would direct repair to the newly synthesized strand. Since the newly synthesized daughter strand is transiently undermethylated, it was postulated that post-replicative mismatch repair would be methyl-directed such that repair was directed against the unmethylated strand of hemimethylated DNA (Wagner and Meselson, 1976). I will refer to this system as the Dam-dependent (methyl-directed) mismatch repair system since Dam-methylase supplies the required signal.

Dam-dependent mismatch repair, by virtue of the long excision repair tracts, is
sometimes referred to as Long Patch Mismatch Repair (LPMR). The substrate specificity is quite broad and requires the gene products of \textit{mutH}, \textit{mutL}, \textit{mutS} and \textit{uvrD}, the requirement of which will be discussed below in more detail. This system is functionally quite similar to \textit{mutHLS} of \textit{Salmonella typhimurium} and the \textit{hex} system of \textit{Streptococcus pneumoniae} (for review, Claverys and Lacks, 1986). There also exists a group of specialized mismatch specific repair pathways that produce excision tracts less than 10 bases. These Very Short Patch Mismatch Repair (VSPMR) pathways, VSP repair and MutY, will be discussed later.

\textbf{B. \textit{d(GATC)} sequences and Dam-dependent mismatch repair.}

DNA Adenine Methyltransferase (Dam) is encoded by the \textit{dam} gene.

Two percent of all adenine residues in \textit{E. coli} DNA are present as 6-methyladenine, of which 98\% result from methylation catalyzed by the product of the \textit{dam} gene (for review, see Marinus 1984; Barras and Marinus, 1989). Under normal circumstances, Dam is present at approximately 130 molecules per cell as determined by immunoblotting using a polyclonal antibody to Dam (Boye and Marinus, unpublished data, 1991). Dam has been overexpressed and purified to near homogeneity as a 31 kDa protein which exists as a monomer in solution (Herman and Modrich, 1982). It recognizes the adenine residue in the symmetric sequence 5'-(GATC)-3' and transfers a methyl group from S-adenosyl-L-methionine (SAM) to the 6-position of adenine (Herman and Modrich, 1979). This transfer occurs in a two step process in which the first step involves the binding of SAM to an allosteric site on Dam to create a binary complex (Bergerat et al., 1991). The binding of this binary complex to an unmethylated \textit{d(GATC)} site induces a
conformational change resulting in the formation of a second site (catalytic) that becomes activated subsequent to adenine methylation (Bergerat et al., 1991).

The *dam* gene is not required for viability of *E. coli* strains since deletion of the *dam* gene is not a lethal event (Parker and Marinus, 1988, also see Ch. 2 of this thesis). However, *dam* strains do exhibit pleiotropic effects including sensitivity to 2-aminopurine (AP) (Glickman et al., 1978), hypermutability (Marinus and Morris, 1974, Bale et al., 1979), hyper-recombination (Marinus and Konrad, 1976), increased transposition of Tn10 (Roberts et al., 1985), and sensitivity to ultraviolet light and methyl methanesulfonate (MMS) (Marinus and Morris, 1974, Bale et al., 1979). This wide range of phenotypes, most of which are related to the cell’s inability to correct damaged DNA, suggests that *dam* may be involved in error avoidance. Since the mutation spectrum for *dam* was found to be similar to *mutHL* and *S* in that they all reflect an increased frequency of transition mismatches (Glickman, 1979), it was originally suggested that *dam* may be part of the same error avoidance system as *mutHLS* (Glickman, 1979; Glickman and Radman, 1980).

Of particular interest is the observation that *dam* strains are inviable when present with a mutation in any of the *recABC* alleles (Marinus and Morris, 1974, 1975). Viability of such strains can be restored with second site mutations in either of the *mutL* or *mutS* genes (McGraw and Marinus, 1980). The lethality of the *dam rec* double mutants presumably is due to the generation of double strand breaks as a result of undirected mismatch repair creating overlapping excision tracts on opposite strands (Wang and Smith, 1986). Therefore, the introduction of *mutS* or *mutL* alleles cripples mismatch repair and
restores viability. A similar phenomenon has also been observed in *dam*-3 strains exhibiting resistance to 2-AP (Glickman and Radman, 1980). Second site mutations that mapped to either of the *mutH, mutL* or *mutS* genes rendered *dam* strains resistant to 2-AP (Glickman and Radman, 1980). The mechanism for this resistance is not completely understood, however it is possible that undirected mismatch repair of 2-AP containing DNA may result in double strand breaks at a frequency that exceeds the ability of the *rec* system to rejoin the breaks and restore viability.

**Dam-methylation of adenine in d(GATC) sequences directs the strandedness of repair.**

Meselson had postulated that correction of replication errors by mismatch correction is methyl-directed (Wagner and Meselson, 1976). This implied that mismatch repair must occur after passage of the replication fork but prior to Dam-methylation. The general features of this model predict that repair would be directed against the unmethylated strand of hemimethylated substrates, on either strand of an unmethylated substrate, and not at all on symmetrically methylated substrates. These predictions are in agreement with early studies performed with *E. coli* both *in vivo* (Radman et al., 1980; Pukkila et al., 1983; Wagner et al., 1984) and *in vitro* (Lu et al., 1983). The most striking evidence was demonstrated with an *in vivo* transfection system using *λ* heteroduplexes that represented each of the four possible methylation states (Pukkila et al., 1983). Unlike earlier studies that used partially methylated *λ* DNA, these hemimethylated and symmetrically methylated heteroduplexes were prepared using *λ* DNA that was methylated *in vitro* with Dam methylase. This study clearly demonstrated that mismatch repair
is directed by Dam-methylation since switching the state of hemimethylation produced a simultaneous switch in the strandedness of repair, a result unaffected by transfection into either dam− or dam+ strains indicating that repair occurs prior to any significant methylation of the heteroduplex (Pukkila et al., 1983). Furthermore, repair occurred on either strand of unmethylated heteroduplexes consistent with loss of direction. No repair was observed when the heteroduplexes were symmetrically methylated (Pukkila et al., 1983) which suggests that the hypermutable phenotype observed in dam overproducing strains (Herman and Modrich, 1981; Marinus et al., 1984) can be attributed to rapid methylation of the DNA prior to correction of mutations by mismatch repair. The Wagner-Meselson hypothesis and observations of Pukkila et al., (1983) have been subsequently confirmed using a substrate molecule containing a single d(GATC) site and demonstrating methyl directed repair in vitro by the purified mismatch repair system (Lahue et al., 1989).

**Dam-dependent mismatch correction requires at least one d(GATC) site for efficient correction of a mismatch.**

Initial in vivo and in vitro studies demonstrated that the presence of d(GATC) sites are required for Dam-dependent mismatch repair since efficient repair of covalently closed substrate molecules lacking d(GATC) sequences is not observed (Laengle-Rouault et al., 1986, 1987; Lahue et al., 1987; Bruni et al., 1988; Claverys and Mejean, 1988). In the presence of d(GATC) sites, both in vivo and in vitro studies demonstrated that mismatch repair proceeded according to the state of Dam-methylation. In addition, it was reported that a single d(GATC) site is sufficient to support Dam-dependent mismatch correction (Lahue et al., 1987; Claverys and Mejean, 1988). There has been only one
report to the contrary which suggests that two sites are required for mismatch correction since heteroduplexes containing a single d(GATC) site were not repaired, but the addition of a second d(GATC) sequence to the heteroduplex restored efficient mismatch correction (Laengle-Rouault et al., 1987). Mismatch repair of the heteroduplex was not evaluated for each d(GATC) site individually, therefore it remains a distinct possibility that the single d(GATC) site chosen for the experiment was a poor substrate for mut deficiency dependent mismatch correction while the opposite may be true for the other d(GATC) site.

The requirement for a single d(GATC) site has been confirmed based on experiments utilizing the purified in vitro mismatch repair system (Lahue et al., 1989). Efficient correction of a heteroduplex containing a G/T mismatch and a single d(GATC) site 1024 base pairs from the mismatch was observed to proceed in a Dam-dependent direction. No repair was observed when the single d(GATC) site was symmetrically methylated or when the heteroduplex was devoid of d(GATC) sites. In addition, using the same heteroduplex (G/T, single d(GATC) site) just described, treatment with E. coli cell free extracts under conditions that restrict DNA synthesis produce one single-stranded region or gapped duplex reflecting an intermediate step between repair tract excision and resynthesis (Su et al., 1989). The gaps were mapped by electron microscopy after coating the single-stranded region with single-strand DNA binding (SSB) protein. The excision tracts were not observed in molecules devoid of dGATC sites, but were observed spanning the region in the vicinity of the d(GATC) site and the mismatch in d(GATC) containing molecules (Su et al., 1989). In every case tested, only one single stranded region was observed per molecule and that region always spanned the shorter distance between the mismatch and d(GATC) site.
The relationship between repair efficiency of a mismatch and the distance to the nearest d(GATC) site has not been clearly defined. A recent \textit{in vivo} study using M13 heteroduplexes containing a d(GATC) site placed at six different locations were used to examine this question. Four of the d(GATC) sites were placed within 1,000 base pairs of the mismatch and the other two d(GATC) sites were placed in excess of 2,000 base pairs from the G/T mismatch. The results of the transfection experiments indicate that there is little difference in the efficiency of mismatch correction when the d(GATC) sites are within 1,000 base pairs of the mismatch, but the efficiency of repair decreases as the d(GATC) site exceeds 2,000 base pairs (Bruni et al., 1988). This is in agreement with an \textit{in vitro} study (Lahue et al., 1987) indicating that mismatch correction is more efficient when a single d(GATC) site is 1,024 base pairs from the mismatch than when it is 2,200 base pairs away. In addition, increasing the density of d(GATC) sites in the heteroduplex increased the repair efficiency (Lu, 1987), while decreasing the density had the opposite effect (Lahue et al., 1986; Bruni et al., 1988).

\textbf{C. Specificity of Dam-dependent mismatch correction.}

The Dam-dependent mismatch repair pathway corrects transition mismatches more efficiently than transversion mismatches.

The efficiency of mismatch repair is influenced by various factors, the most important of which involves the nature of the mismatch itself (Kramer et al., 1984; Wagner et al., 1984; Dohet et al., 1985; Wood et al., 1986; Jones et al., 1987; Su et al., 1988; Lahue et al., 1989). The first study designed to evaluate the specificity of mismatch repair was
carried out using M13 heteroduplexes, each containing one of the eight possible mispaired bases within the lacZ gene (Kramer et al., 1984). Following transfection into E. coli strains, the repair efficiency for each class of mispairs was scored based on the percentage of mottled plaques that arise since these indicate the absence of repair. With this system, it was determined that G/T and A/C transition and G/G transversion mispairs are corrected most efficiently, A/A transversions are corrected at an intermediate efficiency, and very little correction was observed with G/A, T/T, C/C and C/T transversions. Dohet et al., (1984, 1985) using unmethylated heteroduplexes of λ phage, observed a similar pattern of repair efficiency with slight variations in the correction of transversion mispairs. Although it is difficult to directly compare M13 transfection experiments to those involving λ, the variation observed in the efficiency of correction of different mispairs suggests that the sequence environment within which the mispair resides can significantly influence its repair.

The role of sequence environment in correction efficiency was addressed in vivo (Jones et al., 1987) and in vitro (Lu and Chang, 1988). Jones et al. (1987) constructed sets of λ heteroduplexes containing base pair mismatches of G/T, A/C, G/A and C/T such that each set was located within an identical sequence of the cl repressor gene. Based on the percentage of mixed infective centers arising after transfection of E. coli, they found that G/T and A/C were efficiently corrected regardless of sequence environment, while the less efficient correction of G/A and C/T mispairs were significantly influenced by the flanking DNA sequence. This is in agreement with observations using a in vitro system of cell free extracts since the correction of transversion mispairs differed markedly with
sequence context (Lu and Chang, 1988). Interestingly, an increasing G-C content surrounding G/A and C/T transversion mispairs correlated with an increase in correction efficiency, while G-C content had little effect on the correction of G/T and A/C transition mispairs (Jones et al., 1987). The significance of this observation is unclear, however, since a similar study, which involved the related mismatch repair system of *S. pneumoniae*, showed no such simple correlation between G/C content and repair efficiency *in vivo* (Gasc et al., 1988).

To obtain a more accurate measure of correction efficiency, *in vitro* studies with cell free extracts and the purified system were performed by evaluating the correction of each of the eight mismatched base pairs placed within the same sequence context (Su et al., 1988; Lahue et al., 1989). The results of such experiments were in general agreement with the *in vivo* studies discussed above. That is, G/T and A/C transitions were corrected with greatest efficiency followed by less efficient repair of transversion mismatches, the poorest of which is the C/C mispair. This also agrees with the observed mutation spectrum for components of mismatch repair since transition mutations predominate in the absence of mismatch repair (Glickman, 1979; Choy and Fowler, 1985; Leong et al., 1986; Carraway et al., 1987; Rewinski and Marinus, 1987; Schaaper and Dunn, 1987; Wu et al., 1990).

**Dam-dependent mismatch repair can correct small frameshift mutations.**

The correction of +1/-1 frameshift mutations by Dam-dependent mismatch repair has been reported using an *in vivo* λ transfection system (Dohet et al., 1986). In addition, Lear and Grafstrom (1989) demonstrated Dam-dependent mismatch correction of 1, 2
and 3 base insertion/deletions in cell free extracts. Correction of heteroduplexes containing these mutations is dependent on the gene products of *mutH* and *S* since the addition of MutH and MutS protein to extracts made from *mutH* and *mutS* cells restored mismatch repair to that observed with wild type extracts. Fishel et al., (1986) have reported that correction of heteroduplexes containing 10 base insertion/deletion heterologies can proceed by both Dam-dependent and Dam-independent mismatch repair pathways. The results of Fishel et al., (1986) are discussed in more detail in Chapter 4 of this thesis. Correction of heteroduplexes containing deletions of 93, 700 and 800 bases by Dam-dependent mismatch repair have not been observed (Kramer et al., 1982; Raposa and Fox., 1986; Dohet et al., 1987). The point at which mismatch repair can no longer recognize and repair insertion deletion mutations has only recently been defined and is the subject of Chapter 4 of this thesis.

**D. The Biochemistry of Dam-dependent Mismatch Correction.**

Genetic studies involving transfection of *E. coli* with λ heteroduplexes led to the identification of genetic loci involved in mismatch repair (Nevers and Spatz, 1975; Rydberg, 1978; Radman et al., 1980; Pukkila et al., 1983). The requirement for these gene products, *mutHLS* and *uvrD*, in Dam-dependent mismatch repair was subsequently demonstrated through *in vitro* analyses using a system of complementation assays and cell free extracts (Lu et al., 1983; Su et al., 1988). Lahue et al. (1989) subsequently reconstituted the mismatch repair reaction *in vitro* consisting of purified MutH, MutL,
and MutS proteins, DNA helicase II, SSB (single-strand DNA binding protein), exonuclease I, DNA polymerase III holoenzyme and DNA ligase. Cofactor requirements included ATP, NAD$^+$ and the four dNTP's (Lahue et al., 1989). With this system, Lahue et al. (1989) demonstrated that mismatch repair supports the correction of 7 of the 8 possible mismatched bases within a heteroduplex containing a single d(GATC) site 1024 base pairs from the mismatch. Of the 8 possible mismatches, C/C was the only mismatch not subject to significant correction (Lahue et al., 1989).

A critical question about the polarity of repair was answered with the purified system. When strand specific correction involved the unmethylated d(GATC) site on the 3' side of the mismatch, separated by 1024 base pairs in the 3'-5' direction, exonuclease I catalyzed efficient correction (Lahue et al., 1989). When the unmethylated d(GATC) site is on the 5' side of the mismatch, 3'-5' exonuclease I would have to excise 5416 base pairs to reach the mismatch in the 3'-5' direction. In this case, the correction efficiency was only 20% of that observed when the unmethylated d(GATC) site and mismatch spanned the shorter distance (Lahue et al., 1989). The generation of a such an excessive excision tract would appear to be rather inefficient and suggested that additional components may be required. The additional requirements have recently been identified as 5'-3' exonuclease VII and the 5'-3' RecJ exonuclease (Modrich, 1991). These components confer bidirectional capability on mismatch repair since the system can support correction of a mispair in both hemimethylated configurations of a linear heteroduplex substrate (Modrich, 1991).
Isolation and characterization of the mutS gene product.

The mutS gene of E. coli has been isolated and cloned into an expression vector placing it under the control of a thermosensitive P_L promoter (Su and Modrich, 1986). Upon thermal induction, MutS protein was overexpressed and purified to near homogeneity (98%) in a biologically active form as a 97,000 Dalton protein (Su and Modrich, 1986). Although present as a dimer in solution, it is not clear what the aggregation state of the biologically active form is since activity associated with MutS sediments on a sucrose gradient as two peaks (Su and Modrich, 1986). The 97 kDa protein binds in vitro to all eight possible mismatched base pairs with varying affinities ranging from 39 nM for the G/T mismatch to 480 nM for the C/C mismatch (Su et al., 1988). The eight mismatches tested, all within the same sequence environment, were placed into three classes based on MutS binding affinities. G/T and A/C represent high affinity substrates for MutS, A/A, T/T, and G/G medium affinity, and A/G, C/T and C/C represent the lowest affinity substrates (Su et al., 1988). Moreover, binding of the heteroduplex substrates by MutS did not require the presence of d(GATC) sites (Su and Modrich, 1986) and did not bind significantly to homoduplex indicating the requirement for the presence of a mismatch (Su and Modrich, 1986; Su et al., 1988). MutS footprints resulting from protection against DNase I cleavage reveal that MutS binds to substrate DNA in an asymmetric pattern that varies with the nature of the mismatch (Su and Modrich, 1986; Su et al., 1988).

In general, the binding affinity of MutS to any of the eight possible mismatches roughly correlates to its repair efficiency in vitro (Su et al., 1988). However, several
lines of evidence indicate that factors in addition to binding affinities influence the efficiency of repair. First, MutS binds to C/A mispairs almost 8-fold greater than it binds to C/T mispairs, yet the correction efficiency in vitro is similar for both (Su et al., 1988). Second, MutS protein bound with high affinity to 16-mer heteroduplexes containing either G/T and C/A mismatched bases, whereas the affinity of MutS to the same 16-mers containing the inverted form of the mismatched bases, T/G and A/C, was significantly lower even though these mismatched bases are efficiently corrected by the mismatch repair system (Jiricny et al., 1988).

The Dam-dependent mismatch repair system of *S. typhimurium* is functionally quite similar to the *E. coli* system in that correction of mismatches is methyl directed and requires the gene products of *mutHLS* and *uvrD* (Pang et al., 1984). Furthermore, Haber et al. (1988) have shown that *S. typhimurium* MutS can complement *E. coli* MutS *in vivo*. The *mutS* gene of *Salmonella typhimurium* has been cloned and sequenced (Pang et al., 1985; Haber et al., 1988), and Mankovich et al. (1989) have stated that the nucleotide sequence shares 94% homology with *E. coli* MutS. One interesting feature arising from the deduced amino acid sequence of MutS involved a region of 14 amino acids of which 13 share homology with the consensus nucleotide binding site (Walker et al., 1982) suggesting that MutS protein may hydrolyze nucleotides (Haber et al., 1988). *S. typhimurium* MutS has since been purified to near homogeneity with subsequent confirmation of the presence of a weak ATPase activity that hydrolyzes ATP to ADP and P$_i$ (Haber and Walker, 1991). A similar finding in *E. coli* MutS has also been reported (Grilley et al., 1990; Modrich, 1991).
E. coli MutS, in the presence of ATP, stimulates the formation of alpha loop structures that contain protein bound at the DNA junction presumably adding support to the structure (Grilley et al., 1990; Modrich, 1991). Its formation is absolutely dependent on the presence of a mismatch and requires the hydrolysis of ATP since the alpha loop is not observed with homoduplex DNA or in the absence of ATP (Modrich, 1991). The binding of DNA by MutS is independent of the presence of ATP since MutS binds to a mismatch in the absence of ATP and protects about 20 base pairs of DNA against DNase I cleavage (Su and Modrich, 1986; Su et al., 1988). However, in the presence of ATP, which promotes alpha loop formation, the area protected against DNase I cleavage decreases by approximately two bases (Grilley et al., 1989). Inclusion of the nonhydrolyzable ATP analog, ATPγS, further reduces the area of DNase I protection (Grilley et al., 1989). A mutation within the ATP binding site of MutS, changing lys-622 to alanine, resulted in a 120 fold reduction in in vitro repair activity while retaining its ability to recognize and bind to mismatch containing substrates (Haber and Walker, 1991). Taken together, these observations suggest that the ATPase function of MutS is involved in promoting the process of repair past the point of mismatch recognition, possibly by fuelling a transport mechanism that searches long stretches of DNA for an unmethylated d(GATC) site.

Isolation and characterization of the mutH gene product:

The E. coli mutH gene has been cloned, sequenced (Grafstrom and Hoess, 1983, 1987), and isolated to near homogeneity (98% pure) in biologically active form as a 25,000 Dalton protein (Welsh et al., 1987). Analysis of MutH function has revealed the presence of a weak Mg²⁺ dependent 5'–d(GATC)-3' specific endonuclease activity which
makes an incision 5' to the dG on the unmethylated strand of a hemimethylated d(GATC) sequence (Welsh et al., 1987). Since a nicked circular heteroduplex, produced by treatment with MutH protein, can be restored to its covalently closed state by DNA ligase, it was reasoned that MutH scission at d(GATC) sites must give rise to 5'-phosphoryl and 3'-hydroxyl termini (Welsh et al., 1987). Substrates containing symmetrically methylated d(GATC) sites are resistant to MutH activity, while unmethylated d(GATC) sequences are nicked on one strand or the other but at much reduced levels than that observed with hemimethylated sites (Welsh et al., 1987). This property of MutH is responsible for determining the strand selectivity or directionality of repair.

The requirement for MutH in mismatch repair can be bypassed by the presence of a nick in either strand of a heteroduplex devoid of d(GATC) sites provided that the nick is not resealed by DNA ligase (Langle-Rouault et al., 1987; Lahue et al., 1987, 1989). Correction of a covalently closed heteroduplex containing a single d(GATC) site required the presence of MutH protein (Lahue et al., 1987, 1989), while the presence of a nick either at the d(GATC) site or elsewhere on the molecule was sufficient to support repair in the absence of MutH and DNA ligase (Langle-Rouault et al., 1987; Lahue et al., 1987, 1989). Furthermore, efficient methyl-directed repair of hemimethylated heteroduplexes containing a single d(GATC) site was observed in the purified system, which included all the components of mismatch repair except MutH and ligase, only after the heteroduplex was pretreated with purified MutH protein (Lahue et al., 1989). In the presence of DNA Ligase, mismatch correction of MutH-pretreated substrates was not observed unless MutH protein was added to the purified system (Lahue et al., 1989). The results of this study confirm that the function of MutH in mismatch repair is to recognize d(GATC)
sites and make an incision on the unmethylated strand.

A curious finding of Welsh et al. (1987) was that MutH exhibited a weak activity in vitro since only 1 d(GATC) site was incised per hour per monomer equivalent. Since high concentrations of MutH had to be used in order to detect d(GATC) strand scission in vitro, it seemed unlikely that the estimated concentration of 20 molecules of MutH per cell would be sufficient to support efficient mismatch repair (Welsh et al., 1987). Welsh et al. (1987) suggested that a latent form of MutH might be activated upon formation of a multi-component complex possibly involving other proteins and/or the DNA substrate, or alternatively that the DNA substrate used in the study did not accurately imitate the native substrate. A recent report has indicated that the latent MutH endonuclease activity can be enhanced 20 to 50 fold when in the presence of MutL, MutS, ATP, and an appropriate heteroduplex substrate (Modrich, 1991). MutH activation is dependent on ATP hydrolysis (Modrich, 1991).

**Isolation and characterization of the mutL gene product:**

Of the gene products of mutHL and S, the least understood is that of mutL. *E. coli* MutL has been isolated and purified to near homogeneity as a 70,000 Dalton protein (Grilley et al., 1989), a value which is in close agreement with the *S. typhimurium* MutL (Pang et al., 1985; Mankovich et al., 1989). MutL is present in solution as a dimer, but its biologically active state has not yet been determined (Grilley et al., 1989). The biological function of MutL in mismatch repair is not clear, however it appears to be involved in at least two processes. First, MutL in combination with MutS, ATP and a mismatch can significantly enhance the activity of MutH (Modrich, 1991). Second,
MutS protects about 20 base pairs of DNA from DNase I cleavage (Su and Modrich, 1986; Su et al., 1988), but in the presence of ATP (or ATPγS) and MutL the area of DNase I protection increases to approximately 100 base pairs (Grilley et al., 1989). This suggests that MutL either interacts with the MutS/mismatch/ATP unit to form a nucleoprotein complex or alternatively that it interacts with MutS prior to mismatch recognition. Formation of this complex with concurrent hydrolysis of ATP is then required for the activation of MutH. In this context, Modrich (1991) has suggested that MutL acts as an interface for communication between the nucleoprotein complex and MutH such that the concurrent hydrolysis of ATP fuels a transport mechanism that allows MutH to search for a d(GATC) sequence. How the signal is transduced between MutL and MutH is not clear. In addition, the nature of the protein:protein and protein:DNA interactions as well as the stoichiometry of the complex have yet to be identified.

Components of the excision and DNA resynthesis steps:

Although the exact role of helicase II in the mismatch repair process is unclear, it likely enters the helix at the MutH induced nick (Runyon et al., 1990) and begins unwinding the DNA duplex driven by the hydrolysis of ATP (Runyon and Lohman, 1989; for review see Matson and Kaiser-Rogers, 1990). Since helicase II translocates in the 3'-5' direction (Matson, 1986), Runyon et al. (1990) have suggested that MutH, after making an incision, may act as a clamp at d(GATC) sites to ensure that helicase II translocates in the proper direction, therefore conferring additional directionality to the repair process. This idea is intriguing from the viewpoint that MutH stabilization of the nicked
d(GATC) site would also discourage religation of the nick by DNA ligase, an idea consistent with recent observations using the purified system in vitro (Lahue et al., 1989). In addition, Modrich (1991) has suggested that the MutS/MutL/ATP/mismatch complex may serve to promote the entry of excision proteins following d(GATC) site incision by MutH.

The unwinding of duplex DNA by helicase II produces regions of single stranded DNA that rapidly renature unless coated with SSB protein (for review see Meyer and Laine, 1990). SSB is an essential component of mismatch repair since repair activity in vitro is severely reduced in extracts made from ssb− cells or by the addition of anti-SSB IgG to ssb+ extracts (Lu et al., 1984). The requirement for SSB in the reaction has since been confirmed by Lahue et al. (1989), using the purified system in vitro.

Three processes involved in repair are markedly enhanced by the presence of SSB (Meyer and Laine, 1990). First, DNA unwinding by helicase II is driven by SSB protein presumably by preventing reannealing. Second, SSB coating single stranded regions of the nick containing strand are sensitized to the single strand specific exonucleases since SSB coated ssDNA prevents renaturation thus exposing SS-terminus to the appropriate exonuclease (Meyer and Laine, 1990). Exonuclease I removes the excision tract in a 3'→5' direction, while either exonuclease VII or the RecJ exonuclease remove the excision tract in 5'→3' direction (Lehman and Nussbaum, 1964; Chase and Richardson 1974, 1977; Lovett and Kolodner, 1989) implying that the excision process possesses the potential to be bidirectional. This is supported by the finding that the repair tract typically spans the shortest distance between the d(GATC) site and the mismatch in either direction (Su et al., 1989; Modrich, 1991). Third, SSB stabilizes the ssDNA template by
removing areas of secondary structure, therefore increasing processivity and replicational fidelity of DNA polymerase III holoenzyme (Meyer and Laine, 1990).

Resynthesis of the repair tract requires DNA polymerase III holoenzyme, and in particular the \( \tau \) and \( \gamma \) subunits (Maki and Kornberg, 1988) since extracts prepared from \( dnaZ \) (Ts) strains fail to support repair of heteroduplexes in vitro at the non-permissive temperature (Lahue et al., 1989). The addition of purified DNA polymerase III holoenzyme to the extracts restored efficient repair of the heteroduplexes (Lahue et al., 1989). DNA ligase, requiring NAD\(^+\) as cofactor, is responsible for restoring the substrate molecules to their covalently closed state (Lahue et al., 1989).

Summary of model for Dam-Dependent Mismatch Correction:

As depicted in Figure 1, initiation of mismatch correction begins with the recognition and binding of an appropriate mismatch by MutS protein. In the presence of ATP, formation of alpha loop structures ensue with structural support provided by protein, presumably MutS, bound to the DNA junction. The association of MutL with the complex and subsequent hydrolysis of ATP activating MutH triggers a search for an unmethylated d(GATC) site. MutH makes an incision 5' to the d(GATC) site on the unmethylated strand of a hemimethylated substrate and presumably stabilizes the nick and facilitates the entry of helicase II. Moving towards the mismatch in a 3'−5' direction, helicase II unwinds the helix which is subsequently bound by SSB protein. The appropriate exonuclease (I, VII or RecJ) removes the excision tract using the shortest distance
Model for Methyl-Directed Mismatch Repair

1. **Initiation**
   - ATP
   - ADP + Pi
   - MutH
   - MutL

2. **MutH incision at unmethylated d(GATC) site**

3. **Excision**
   - Exonuclease VII
   - RecJ exonuclease
   - DNA helicase II
   - SSB
   - exonuclease I
   - 3' → 5'

4. **Resynthesis**
   - DNA polymerase III
   - SSB
   - 3' → 5'

5. **Dam methylation**
   - DNA ligase
between the d(GATC) site and the mismatch. DNA polymerase III holoenzyme resynthesizes the repair tract in a 5'-3' direction, and DNA ligase completes the process by restoring covalent continuity.

E. Mismatch Correction and Genetic Recombination:

The very nature of recombination implies that mismatched base pairs form within a recombination intermediate simply as a consequence of allelic differences in homologous regions between the donor and recipient strains. The processing of recombination intermediates by the mismatch repair system has been demonstrated by Feinstein and Low (1986). Hfr crosses involving two markers (hisF860 X hisF818) in which the recipient strains were also defective in mismatch repair, due to mutHLS or uvrD mutations, were found to significantly increase the frequency of recombination. Since this effect is not observed in mismatch repair proficient recipient strains, it indicated that the mismatch repair system processes recombination intermediates in favor of the host (invaded) strand. This observation is in general agreement with other studies (Jones et al., 1987; Shen and Huang, 1989).

Radman (1989) has referred to this phenomenon as "antirecombination" and proposed that mismatch repair serves to ensure that strand exchange results in homologous sequences such that the genotype of the invading strand is converted to that of the recipient strand. With this in mind, it could be predicted that genetic recombination, in addition to host specific restriction systems, presents a formidable barrier to interspecies recombination. This has proven to be the case since conjugative matings between E. coli Hfr donor and the recipient S. typhimurium F^- strains, which are either mutH, L, S or
uvrD, lead to an increased frequency in intergenic recombination (Rayssiguier et al., 1989). The increased frequency of recombination is approximately 1,000-fold greater than that detected when the recipient strain is mismatch repair proficient (Rayssiguier et al., 1989). In addition, this study showed that the mismatch repair system can efficiently process recombination intermediates with up to 20% divergence, which represents the approximate difference in DNA sequence between E. coli and S. typhimurium (Rayssiguier et al., 1989).

F. Mismatch Specific Repair Pathways: Very Short Patch Repair (VSP) and MutY.

At least five mismatch specific repair pathways have been identified in E. coli. Three of these were discussed earlier and include uracil glycoslyase, and the MutT and MutM pathways. Therefore, the following discussion will focus on the two remaining pathways: VSP repair and the MutY system. The MutY pathway, as is the case for MutT and MutM, is involved in preventing G/A mispairs, albeit by a different mechanism than MutT and MutM. VSP repair, on the other hand, is specific for G/T mispairs located within a specific sequence context. These repair systems produce excision tracts of less than 10 bases and they all promote correction of the mispair back to a G/C base pair.

VSP Repair Pathway: Margaret Lieb (1983) first demonstrated the existence of very short patch mismatch repair by performing genetic crosses involving the am6 allele of the \( \lambda cI \) gene and closely linked markers. The generation of \( cI^+ \) progeny was much greater than expected since the three crossover events required to generate \( cI^+ \) recombinants occurred at a frequency greater than single crossover recombinants (Lieb, 1983). This
high negative interference was the result of localized repair since markers within 20 base pairs of am6 were not co-converted (Lieb, 1983). VSP repair is mismatch and sequence specific since it corrects the T in G/T mispairs within the recognition sequence for DNA cytosine methylase (Dcm), 5'-CCNGG-3' (N = A or T), or the related sequences CNGG and CCNG (Lieb, 1983, 1985; Lieb et al., 1986). The inner most cytosine (in bold print) in these sequences can be converted to thymine either as a result of allelic differences in recombination (i.e. am6), or as a result of deamination of 5-methylcytosine to thymine within the Dcm recognition sequence. In each case, VSP removes the thymine from the G/T mispair and restores the G/C base pair creating an excision tract of 2 to 10 bases (Lieb et al., 1986).

The VSP repair pathway requires DNA polymerase I (Dzidic and Radman, 1989) and the gene products of mutL and mutS (Lieb, 1987). Surprisingly, Lieb (1987) also suggested that Dcm methylase played a direct role in repair based on the observation that a strain of E. coli containing a mutation in the dcm gene, dcm6 (Marinus, 1973), fails to support VSP repair. However, it was subsequently determined that Dcm methylase was not required for VSP repair and that the dcm-6 allele mutates both dcm and vsr (see below) as determined by several observations. First, plasmids carrying the gene coding for EcoRII methylase, which acts on the same sequence as Dcm, failed to restore VSP repair to the dcm-6 mutant while plasmids carrying a wild type copy of dcm restored VSP to normal levels (Lieb and Bhagwat, 1988). Since the cloned region of chromosomal DNA containing dcm was 11 kbp, there existed the possibility that a gene, whose product may function in VSP repair, might be closely linked to dcm. Second, analysis of DNA sequences flanking dcm showed that an ORF overlapped dcm by seven codons, and that
both are transcribed from the same promoter but in different reading frames (Bhagwat et al., 1988; Sohail et al., 1990). Deletion of part of the ORF not affecting the dcm coding region gave rise to a strain that was dcm+ but deficient in VSP repair, while deletion of part of dcm not affecting the ORF was VSP proficient but dcm− (Bhagwat et al., 1988; Sohail et al., 1990). This indicated that Dcm is not required for VSP repair, but that the 18,000 Dalton protein encoded by the ORF, referred to as Vsr, is essential for VSP repair. The Vsr protein has recently been identified as an endonuclease that is both mismatch-dependent and strand-specific in that it makes an incision 5' to the T in G/T base pairs within the VSP sequence (Hennecke et al., 1991).

The biological activity of Vsr protein relative to its involvement with MutS and MutL in this repair process is not clearly defined. In fact, recent observations have raised questions about the requirement of MutL and MutS in VSP repair. First, overexpression of Vsr can overcome the requirements for MutL and MutS (M. Lieb, unpublished data, 1990). Second, Vsr recognizes and makes a strand specific incision at the site of the mismatch in the absence of MutL and MutS (Hennecke et al., 1991). One possible role for MutS and MutL in VSP repair might be as regulatory components. Alternatively, the data that originally implicated MutS and MutL as essential components in VSP repair may have been genetically flawed. A definitive answer must wait until VSP repair is reconstituted in vitro with its purified components.

**MutY Repair Pathway:** The existence of a second small patch repair pathway specific for the correction of G/A transversion mismatches was observed using cell free extracts in vitro with the finding that the correction of G/A mismatches to G/C was largely independent
of Dam-dependent mismatch correction and sequence context (Su et al., 1988; Lu and Chang, 1988). A mutator locus, mutY (also known as micA), has been identified at 64 minutes on the E. coli genetic map which produces G/C to T/A transversion mutations (Ngheim et al., 1988; Radicella et al., 1988; Tsai-Wu et al., 1991). Furthermore, in vivo and in vitro correction of heteroduplexes containing a G/A mispair has been demonstrated using E. coli strains wild type for the mutY locus, but deficient in Dam-dependent mismatch repair indicating that this repair pathway was independent of methyl-direction and the gene products of mutHL and S (Radicella et al., 1988; Au et al., 1988). Conversely, no correction was observed when both Dam-dependent mismatch repair and the mutY gene were inactivated (Radicella et al., 1988; Au et al., 1988). Restoration of the G/A repair pathway to a mutY" strain can be demonstrated by complementation with a plasmid bearing a wild type copy of the mutY gene (Michaels et al., 1990).

Lu and Chang (1988) demonstrated the presence of a binding protein in cell free extracts of E. coli specific for the G/A mismatch since it does not bind to any of the other seven possible mismatched base pairs. Closely associated with this A/G binding protein (AGP) was an endonuclease activity that makes simultaneous incisions at the first phosphodiester bond 3' to and the second phosphodiester bond 5' to the dA of the G/A mismatch (Lu and Chang, 1988). More recently, Au et al., (1989) have reported the isolation and near homogenous purification of a 36,000 Dalton MutY protein that catalyzes the removal of adenine in G/A mismatches by cleaving the glycosidic bond producing an abasic site. However, in contrast to Lu and Chang (1988) no associated endonuclease activity was detected in the purified MutY preparation (Au et al., 1989). In addition, when the MutY repair pathway was reconstituted in vitro, it required the addition of AP
endonuclease II (isolated from HeLa cells) in addition to MutY protein, DNA polymerase I and DNA ligase for successful correction of the G/A mismatch indicating that MutY and the nicking activity of Lu and Chang (1988) are probably separate entities (Au et al., 1989). The processing of the MutY induced abasic site is not presently understood; however it is possible that the nicking activity observed by Lu and Chang (1988) may represent the required activity.

G. Statement of the Problem.

At the time that this thesis work began, very little was known about the repair of insertion/deletion heterologies by the Dam-directed mismatch repair system. Dohet et al. (1986) had provided the first direct evidence for mismatch repair of one base insertion/deletion heterologies, but repair of insertion/deletion heterologies larger than this had not been determined. I chose to investigate the Dam-dependent mismatch repair system and its role in the recognition and repair of insertion/deletion heterologies. This is discussed in more detail in the Introduction to Chapter 3.
CHAPTER 2

A Simple and Rapid Method to Obtain Substitution Mutations in *Escherichia coli*: Isolation of a *dam* Deletion/Insertion Mutation.
A simple and rapid method to obtain substitution mutations in *Escherichia coli*: isolation of a *dam* deletion/insertion mutation

(Methyltransferases; genes; bacteria; recombinant DNA; methylation; conjugation)

Breck Parker and M.G. Marinus

*Department of Pharmacology, University of Massachusetts Medical School, Worcester, MA 01655 (U.S.A.)*

Received 15 July 1988

Revised 18 August 1988

Accepted 21 August 1988

Received by publisher 12 September 1988

SUMMARY

We describe the isolation of a strain of *Escherichia coli* bearing a deletion/insertion (i.e., a substitution mutation) in the *dam* gene (*dam*-16). The mutagenesis protocol used should be applicable to any cloned non-essential gene of *E. coli*. The substitution mutation confers resistance to kanamycin and can easily be transferred to other strains by standard genetic techniques. The amount of Dam methyltransferase (MTase) in *dam*-16 strains as determined either in vitro or in vivo is below the level of detection. We conclude that the Dam MTase is not required for viability of *E. coli*.

INTRODUCTION

The *dam* gene of *Escherichia coli* specifies a DNA adenine methyltransferase (MTase), which modifies adenine residues in 5'-GATC-3' sequences in double-stranded DNA (Marinus, 1984; 1987a,b). The Dam MTase is the only enzyme in *E. coli* known to modify this sequence. The *dam*-3 (Marinus and Morris, 1973) and *dam*-4 (Marinus and Konrad, 1976) mutations have been used extensively and frequency chromosome transfer; kb, kilobase(s) or 1000 bp; Km, kanamycin; mA, N'-methyladenine; MNNG, N-methyl-N'-nitro-N-nitrosoguanidine; MTase, methyltransferase; Mu, defective Mu phage; Oc, ochre; oriC, origin of *E. coli* chromosome replication; PO, point of origin of Hfr transfer; R, resistance; S, sensitivity; SDS, sodium dodecyl sulfate; Sm, streptomycin; SSC, 0.15 M NaCl, 0.0015 M Na3 citrate, pH 7.6; Tn, transposon; ::, novel joint (fusion); [ ], designates plasmid-carrier state.

Correspondence to: Dr. M.G. Marinus, Department of Pharmacology, University of Massachusetts Medical School, 55 Lake Avenue, Worcester, MA 01655 (U.S.A.) Tel. (508)856-3330.

Abbreviations: aa, amino acids; Ap, ampicillin; 2-AP, 2-aminopurine; bp, base pair(s); Cm, chloramphenicol; CGSC, *E. coli* Genetic Stock Center, Department of Biology, Yale University, New Haven CT 06510 USA; dam, gene coding for Dam; Dam, DNA adenine methyltransferase; Del or d, deletion; Hfr, high-frequency chromosome transfer; kb, kilobase(s) or 1000 bp; Km, kanamycin; mA, N'-methyladenine; MNNG, N-methyl-N'-nitro-N-nitrosoguanidine; MTase, methyltransferase; Mu, defective Mu phage; Oc, ochre; oriC, origin of *E. coli* chromosome replication; PO, point of origin of Hfr transfer; R, resistance; S, sensitivity; SDS, sodium dodecyl sulfate; Sm, streptomycin; SSC, 0.15 M NaCl, 0.0015 M Na3 citrate, pH 7.6; Tn, transposon; ::, novel joint (fusion); [ ], designates plasmid-carrier state.

0378-1109/88/$03.50 © 1988 Elsevier Science Publishers B.V. (Biomedical Division)
were isolated after MNNG and ethyl methane sulfonate treatment, respectively. Strains containing these two mutant alleles do not contain detectable DNA adenine MTase activity in vitro (Marinus and Morris, 1973; Marinus and Konrad, 1976) or in vivo (Lacks and Greenberg, 1977; Bale et al., 1979). An initial observation that there was residual enzyme activity in vivo (Marinus and Morris, 1973) was subsequently shown to be due to contaminating m^A from RNA (Bale et al., 1979).

Two other mutant alleles were obtained by insertion of MudII (lac Ap^R) and Tn9 into the coding region of the dam gene to give dam-12 and dam-13, respectively (Marinus et al., 1983). Analysis by DNA hybridization (Southern, 1975) verified that the dam gene had been disrupted by the Mud insertion (Marinus et al., 1983). Cells containing these insertions did not contain detectable DNA adenine MTase activity in vitro or in vivo. Since insertions in the coding region are expected to produce an inactive protein these results led to the conclusion that the dam gene product was not essential for viability (Marinus et al., 1983).

In contrast to the above, Smith et al. (1985) using a different technique detected residual Dam methylation in strains containing all the dam mutant alleles described above. The residual methylation was preferentially located at the oriC region as determined by the ability of various restriction endonucleases (DpnI, which cuts at methylated sites; MboI, which attacks unmethylated sites and Sau3A, which cleaves at both methylated and unmethylated sites) to digest oriC plasmid DNA extracted from Dam^- cells. In contrast, Szyf et al. (1982) showed that there was no detectable DNA adenine methylation in the oriC region of a dam-3 strain as determined by DNA hybridization (Southern, 1975) using an oriC probe after digestion of DNA with DpnI or MboI.

The results described above can be explained by assuming that either another MTase exists in E. coli which preferentially modifies some dam sites at oriC or that there is residual Dam MTase activity in all dam mutants. To clarify the situation, we decided to isolate a dam deletion mutation which if obtained would rule out the possibility of residual Dam activity.

EXPERIMENTAL AND DISCUSSION

(a) Strategy

We have isolated a strain with a substitution mutation in the dam gene in two steps. First, a drug resistance gene was substituted for part of the dam-coding region in a plasmid, and second, the substitution mutation was recombined into the E. coli chromosome. In order to transfer mutations from a plasmid to the chromosome, integration of plasmids with ColE1 origins of replication into the chromosome of polA mutants is often used (Gutterson and Koshland, 1983). This method, however, yields a duplication of the recombinated gene and also because dam polA strains are non-viable (Marinus and Morris, 1974), this approach was not used. In a previous study (Marinus et al., 1983), we circumvented this difficulty by transforming recB recC sbcB multiple mutants with linearized plasmid DNA in order to isolate the dam-13::Tn9 derivative. In the current study, we describe a gene replacement technique which is as simple and can be used for most cloned non-essential genes.

(b) Isolation of the chromosomal substitution mutation

To isolate a deletion in the dam gene on a plasmid, we used pMQ133 which contains the 854-bp dam gene in a 1.85-kb fragment of chromosomal DNA (Fig. 1). The dam gene contains unique EcoRV and HpaI sites. The 500-bp EcoRV-HpaI fragment was removed from pMQ133 and replaced with a 1.1-kb fragment of DNA coding for Km^R derived from Tn903 (Oka et al., 1981; a gift of M. Susskind via A. Poteete) to yield plasmid pMQ192 (Fig. 1). Since the dam promoter region is intact in pMQ192, a truncated polypeptide should be produced containing the 54 N-terminal aa plus two additional out-of-frame aa residues.

Plasmid pMQ192 was introduced into an Hfr strain (KL14) which transfers the dam gene as an early marker. Within this population, recombination should occur such that the dam-16 allele from the plasmid replaces the dam gene on the chromosome (Fig. 1). We detected such recombinants after mating with an appropriate F^- recipient by selecting for Km^R cells which were Dam^- and Cm^S. The fre-
(c) Characterization of the \textit{dam}-16 mutation

If the \textit{dam}-16 allele has been recombined into the chromosome, then the Dam$^-$ and Km$^R$ phenotypes should co-segregate in genetic crosses. A previous study (Marus, 1973) has shown that \textit{dam} is closely linked to \textit{trpS} and \textit{aroB}. We prepared a PI\textit{vir} lysate on GM3819 (\textit{dam}-16) and transduced \textit{trpS}10330- and \textit{aroB}354-containing strains to prototrophy. The results in Table I indicate that the frequency of Km$^R$ linkage to \textit{aroB} (89\%) and \textit{trpS} (91\%) is that expected if the mutation is located in the \textit{dam} gene. To monitor the Dam phenotype of the transductants, we screened the isolates for sensitivity to 2-AP and for susceptibility of chromosomal DNA to \textit{MboI}.

![Diagram](image-url)

**Fig. 1.** Construction of a chromosomal substitution mutation in the \textit{dam}-gene. (a) Plasmid pMQ133 was digested with EcoRV-\textit{HpaI} and the 500-bp fragment so generated replaced with a 1.1-kb fragment conferring Km resistance. The resulting plasmid, pMQ192, was transferred to strain KL14, an Hfr cell which transfers the \textit{dam}-gene as an early marker. (b) Recombination between plasmid and chromosome can generate a recombinant in which the chromosomal allele has been replaced by that on the plasmid. Such recombinants can be detected upon subsequent transfer of the marker to a suitable recipient cell.

**TABLE I**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sex</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB2847</td>
<td>F$^+$</td>
<td>\textit{aroB}351 \textit{mal}-354</td>
<td>CGSC</td>
</tr>
<tr>
<td>KL14</td>
<td>Hfr(PO68)</td>
<td>\textit{thr}-1 \textit{rel}A</td>
<td>K.B. Low</td>
</tr>
<tr>
<td>GM2806</td>
<td>Hfr(PO68)</td>
<td>KL14[pMQ192]</td>
<td>This study</td>
</tr>
<tr>
<td>AB1157</td>
<td>F$^-$</td>
<td>\textit{thr}-1 \textit{ara}-14 \textit{lev}B6</td>
<td>E.A. Adelberg</td>
</tr>
<tr>
<td>GM3819</td>
<td>F$^-$</td>
<td>As AB1157 but \textit{dam}-16(DeI; \textit{Km}$^R$)</td>
<td>This study</td>
</tr>
<tr>
<td>Trp$^+$</td>
<td>F$^-$</td>
<td>\textit{trpS}10330</td>
<td>C.Yanofofsky</td>
</tr>
</tbody>
</table>

Plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMQ133</td>
<td>A 6.8-kb plasmid containing (i) the \textit{dam} gene of \textit{E. coli} on a 1.8-kb \textit{PstI-SaiI} fragment, (ii) a 1.9-kb \textit{PstI} fragment from \textit{TrpS} conferring resistance to \textit{Cm}, and (iii) the larger \textit{SaiI-PstI} fragment of pBR322(Fig. 1a).</td>
</tr>
<tr>
<td>pMQ192</td>
<td>As pMQ133 but deleted for the \textit{EcoRV-HpaI} fragment within the \textit{dam} gene which has been replaced by a 1.1-kb fragment conferring resistance to Km (Fig. 1a).</td>
</tr>
<tr>
<td>pMQ193</td>
<td>A derivative of pBR322 containing a 1.1-kb \textit{PstI} fragment conferring resistance to Km cloned into the \textit{PstI} site of pBR322.</td>
</tr>
<tr>
<td>pMQ195</td>
<td>A derivative of pBR322 containing the \textit{EcoRV-HpaI} fragment of the \textit{dam} gene cloned into the \textit{EcoRV} site of pBR322.</td>
</tr>
</tbody>
</table>
digestion. In every case Km\(^{R}\) recombinants were Dam\(^{-}\).

Extracts of GM3819 (dam-16) contained no detectable Dam DNA MTase activity (less than 1\% of the wild-type activity as determined by the method described previously; Marinus and Morris, 1973) using Dam\(^{+}\) unmethylated pBR322 DNA as substrate. In addition, there was no detectable m\(^{A}\) in DNA in vivo as determined by the method of Bale et al. (1979). Furthermore, chromosomal DNA isolated from GM3819 was digested with Sau3A and MboI but not with DpnI, whereas DNA from AB1157 was digested with Sau3A and DpnI but not MboI (data not shown). In sum, we could detect no Dam methylase activity either in vitro or in vivo in GM3819 (dam-16).

Lack of Dam activity should correlate with loss of the EcoRV-HpaI portion of the dam gene in dam-16 mutants. Fig. 2 shows that the dam EcoRV-HpaI fragment from pMQ195 (Table I) does not hybridize to digested DNA from strain GM3819 (dam-16), but does anneal to digested DNA from AB1157 (dam\(^{+}\)).

**TABLE II**

<table>
<thead>
<tr>
<th>Cross</th>
<th>Selected markers</th>
<th>Unselected markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>dam-16 x araB354</td>
<td>100%</td>
<td>89% 89%</td>
</tr>
<tr>
<td>dam-16 x npS1030</td>
<td>100%</td>
<td>91% 91%</td>
</tr>
</tbody>
</table>

* A Plasmid of GM3819 (dam-16; Km\(^{R}\)) was prepared and used to transduce araB and npS mutants to prototrophy. One hundred prototrophic recombinants from each cross were then scored for Km\(^{R}\) and the Dam phenotype. Dam mutants are more sensitive to 2-AP than wild-type strains and their DNA is digested by MboI.

![Southern blot analysis of the chromosomal dam gene region. Panel A: DNA was isolated from AB1157 (dam\(^{+}\)) (lanes 1, 3, 5) and GM3819 (dam-16) (lanes 2, 4, 6), digested with SphI (1, 2), KpnI (3, 4) or EcoRI (5, 6) and electrophoresed in 1% agarose. After transfer to a nitrocellulose membrane the chromosomal DNA was probed with the purified [\(^{32}\)P]EcoRV-HpaI fragment from pMQ195 for 18 h at 62\(^{\circ}\)C (Southern, 1975). Following hybridization, the membrane was washed several times in 2 x SSC-SDS solution at 42\(^{\circ}\)C prior to autoradiography. Numbers at the left margin indicate (in kb) the positions of end-labeled HindIII fragments of phage \(\lambda\) DNA. Panel B: the membrane used in Fig. 2A was washed in 0.4 M NaOH at 42\(^{\circ}\)C for 30 min followed by an 2 x SSC-SDS wash to remove the probe used in panel A. The membrane was subsequently reprobed with a labeled PstI fragment from pMQ193 which contains the Km\(^{R}\) determinant followed by 2 x SSC-SDS washing and autoradiography. The 0.5 kb SphI fragment in lane 2 is not apparent in this exposure but is shown in panel C in a longer exposure from an identical experiment. The source of this fragment is not clear; there are no internal SphI sites in the gene for Km\(^{R}\) derived from Tn903 (Oka et al., 1981).
Conversely, the DNA fragment conferring Km^R_, from pMQ193 (Table I), hybridizes to digested DNA from GM3819 (dam-16) but not to that from AB1157(dam_+).

We found that the phenotypes we have monitored in dam-16 strains are similar to those described for other dam mutants (Marinus, 1985; 1987a,b). These include: sensitivity to 2-AP and MNNG; increased spontaneous mutation frequency; non-viability of strains or other specialized methods (e.g., Raibaud et al., 1984). We also find that cells bearing the dam-16 allele are devoid of Dam methylase activity which confirms our previous conclusion that Dam activity is dispensable for viability (Marinus et al., 1983). If residual Dam methylation at oriC in dam-16 strains can be demonstrated, it must be due to some other MTase that is specific for Dam recognition sites at oriC.

(d) Conclusions

We have described a useful and rapid method to facilitate gene replacement in E. coli which does not require the use of polA strains or other specialized methods (e.g., Raibaud et al., 1984). We also find that cells bearing the dam-16 allele are devoid of Dam methylase activity which confirms our previous conclusion that Dam activity is dispensable for viability (Marinus et al., 1983). If residual Dam methylation at oriC in dam-16 strains can be demonstrated, it must be due to some other MTase that is specific for Dam recognition sites at oriC.

ACKNOWLEDGEMENTS

This work was supported by USPHS Grant GM30330. We wish to thank Dr. Kimberly Foster for her guidance in the Southern blotting procedure.

REFERENCES


Communicated by S.R. Kushner.
CHAPTER 2

MATERIALS SUPPLEMENT

Southern Blot Procedure. 1.05 µg of GM3819 (dam) and wild type GM550 chromosomal DNA (Smith, 1967) was digested overnight with SaI, KpnI and EcoRI at 37°C. Each digest was electrophoresed in a 1% agarose gel (Sea Kem) in TAE buffer (40 mM Tris-Cl pH 7.8, 1 mM EDTA) at 25°C. Lambda DNA was digested with HindIII and radioactively labelled with 32P-dATP to use as a size marker. After electrophoresis, the gel was incubated for 30 minutes at 25°C in a solution containing 0.4N NaOH/0.6M NaCl to denature the DNA fragments. The gel was neutralized in a solution of 1.5M NaCl/0.5M Tris-HCl (pH 7.5) at room temperature, and the DNA fragments were subsequently transferred to a "Hybond-C extra" nitrocellulose membrane via capillary transfer in 10X SSC (1.5M NaCl, 0.15M Na-Citrate, pH 7.4). The blot (membrane) was removed and washed for 1 minute in 0.4N NaOH to maximize denaturation of immobilized DNA. Excess NaOH solution was rinsed from the blot with 0.2 M Tris-HCl (pH 7.5) and 2X SSC (0.3 M NaCl, 0.03 M Na-Citrate), and the membrane was baked at 80°C for 1.5 hours to permanently fix the DNA to the membrane.

Pre-hybridization: Eighteen ml of prehybridization solution of 50% deionized formamide, 1% SDS, 1M NaCl, 10% dextran sulfate, 1X Denhardt's reagent (100X = 2% Ficoll 400, 2% BSA, 2% polyvinylpyrrolidone) was added to a "Seal a Meal" bag containing the blot and allowed to incubate approximately 17 hours in a water bath at
42°C with moderate agitation.

**Preparation of the radioactively labelled DNA probe:**  
A. *dam probe*. pMQ195 was digested with *EcoRV/HpaI* (Table 1) and electrophoresed on an 0.8% low melt agarose gel (Sea plaque) in TBE (90 mM Tris-borate, pH 8.3, 1 mM EDTA). A 500 base pair restriction fragment, which represents the *dam* gene, was excised from the gel and isolated using "NACS Prepac" mini-columns according to the manufacturer's instructions. The restriction fragment was ethanol precipitated, dried, resuspended in 30 μl 10 mM Tris-HCl (pH 8.0) and labelled with [α-32P]dCTP (3,000 Ci/m mole) according to the manufacturer’s instructions (Oligolabeling Kit, Pharmacia). Briefly, 60 ng of DNA probe (60 ng/μl) was added to 29 μl dH2O, heated to 100°C for 10 minutes and slow cooled for 15 minutes at 37°C. After the addition of reagent mix (dATP, dGTP, dTTP, pd(N)6, Klenow buffer), BSA and 50 μCi [α-32P]dCTP, Klenow fragment (8 units) was added and incubated for 2 hours at 37°C. Unincorporated [α-32P]dCTP was removed using Sephadex G-50 DNA grade mini spin columns (Pharmacia).  

B. *Kanamycin probe*. This follows the same procedure as that outlined above except that pMQ193 (Table I) was digested with *PstI* to yield a 1.1 kb kanamycin restriction fragment.

**Blot-Hybridization:** A 2 ml hybridization solution was prepared by adding the radioactively labelled denatured DNA probe (0.25 ml, typically 1.0 x 10⁶ cpm) to a solution with a final concentration of 50% formamide, 2.5 mg/ml sonicated salmon sperm DNA, and
0.25 ml of the pre-hybridization solution described above. The probe solution was added to the "Seal a Meal" bag containing the pre-hybridized blot and incubated for approximately 18 hours in a 42 °C water bath with moderate shaking.

**Preparation of membrane for autoradiography:** The blot was rinsed with 2X SSC followed by two consecutive 10 minute washings in 2X SSC with gentle agitation at 25°C. The blot was washed twice for 30 minutes in a 500 ml solution of 2X SSC and 1.0% SDS at 62°C, and two additional times for 30 minutes in 500 ml solution of 0.1X SSC at 25°C. The blot was dried on filter paper and placed on 14 X 17 cm X-OMAT film with an intensifying screen for 20 hours.

**NOTE:** In the legend for Figure 2, line 4 should read "for 18 h at 42°C" rather than "for 18 h at 62°C".
CHAPTER 3

The Development of an \textit{in vivo} Transformation System to use in Evaluating the Correction of Insertion/Deletion Heterologies by \textit{Escherichia coli}.
INTRODUCTION

As discussed in Chapter 1, the Dam-dependent mismatch repair system of *Escherichia coli* repairs DNA containing base:base mispairs at efficiencies that vary with both the nature of the mispair and the DNA sequence environment within which the mispair resides (Kramer et al., 1984; Wagner et al., 1984; Dohet et al., 1985; Wood et al., 1986; Jones et al., 1987; Lu and Chang, 1988; Su et al., 1988; Lahue et al., 1989). Unlike base:base mispairs, the repair of small heterologies by Dam-dependent mismatch repair was an area that had received little attention. Of the reports in the literature that had suggested that Dam-dependent mismatch repair could correct +1/-1 frameshift mutations (Siegel and Kamel, 1974; Siegel and Vaccaro, 1978; Skopek and Hutchinson, 1984), only one had provided direct evidence (Dohet et al., 1986). Furthermore, mismatch repair of insertion/deletion mutations larger than one base had not been investigated. Therefore, I decided to investigate the *E. coli* Dam-dependent mismatch repair system and its role in the repair of insertion/deletion mutations. Specifically, I was interested in determining the maximum size of insertion/deletion mutations that could be recognized and repaired.

To investigate the repair of insertion/deletion mutations, heteroduplexes containing deletions of various lengths were constructed by annealing together two strands of DNA that differ only in that one strand contains the complete DNA sequence (wt) and the other strand contains a deletion of defined length. The repair of these heteroduplexes was evaluated after transformation of mismatch proficient and deficient strains of *E. coli*. Two different assays were used to score for repair; the O<sub>mnt</sub> plasmid system and the much superior mnt structural gene system. The mnt operator system scored for repair on the
basis of sensitivity or resistance to tetracycline. However, as a result of numerous experimental and technical problems, this system was subjected to major modifications. These modifications gave rise to the mnt structural gene system (or simply mnt system) which scored for repair on the basis of colony color on MacConkey/lactose agar. A more complete description of the two systems, the problems, and where appropriate, the solutions to those problems, are presented below.

**MATERIALS AND METHODS**

**Plasmid, Phage and Bacterial Strains.** The f1 bacteriophage R408 was a gift from Dr. M. Russel (Russel et al., 1986). λ GM110 contains a mnt::lacZ operon fusion. Plasmid pPY97 has been described by Rewinski and Marinus (1987) and in Figure 1. Plasmids pMQ143, pMQ144, pMQ250-253 are identical to pPY97 except that they contain deletions within the mnt operator (Figure 2.): pMQ143 = 15 bp deletion (-34 to -48), pMQ143 = 1 bp deletion (-43), pMQ250 = 2 bp deletion (-43 to -44), pMQ251 = 4 bp deletion (-47 to -50), pMQ252 = 5 bp deletion (-38 to -42), pMQ253 = 7 bp deletion (-40 to -46). The numbering system is that described in Rewinski and Marinus (1987). Plasmid p4349 is identical to pPY97 except that it contains a 1 bp insertion between nucleotides 62-63 of the mnt gene (mnt). E. coli strain GM1690 (GM3819, but containing F42 (F-lac)) is described in Chapter 2 (Parker and Marinus, 1988). GM856 is AB1157 (described in Chapter 2) but recF143, and was used to prepare plasmid DNA methylated at d(GATC) sequences. GM4159 is pPY97/GM1690 and is used to produce single strand pPY97 DNA that was unmethylated at d(GATC) sequences. GM1166 is F' endA1 thi-1 hsdR17 supE44 (Carraway et al., 1987); GM3856 is as GM1166, but mutH::Tn5,
GM4719 is as GM1166, but *mutL::Tn5*, GM4331 is *Fthr-1 Δ(lac-pro) sx3 txs-33 supE44 galK2 hisG4(Oc) rfb-1 mgl-51 rpsL31 kdgK51 xyl-5 mtl-1 argE3 thi-1 (λ GM110) and GM4348 is as GM4331 but *mutL::Tn5 stu* (Connolly and Winker, 1989).

**Isolation of Single-Stranded DNA.** Five ml of Brain Heart media was inoculated with an isolated colony of *E. coli* strain GM4159 (GM1690 freshly transformed with pPY97) and incubated overnight at 37°C. Each 5 ml overnight culture of GM4159 was added to 1 liter of Brain Heart Infusion Broth (Difco; 20 g per liter), 60 μg/ml ampicillin, 20 μg/ml kanamycin and 1 ml of R408 f1 phage suspension (titer of about 1 x 10^{11} p.f.u./ml). The cultures were incubated at 37°C with aeration (200 rpm) overnight. Following incubation, the cultures were transferred to 1 liter centrifuge bottles and spun at 5,000 rpm for 30 minutes in a Sorvall RC-3 centrifuge. The supernatant was transferred to a clean 1 liter centrifuge tubes containing 36 gm PEG 8000 and 27 gm NaCl, mixed thoroughly, placed at 4°C for 4 hours and spun at 5000 rpm for 30 minutes. The supernatant was discarded and the bottles were allowed to drain upside down for about 15 minutes. The phage pellet was resuspended in 30 ml STE buffer (50 mM NaCl, 10 mM Tris-Cl, pH 8.0, 1 mM EDTA) and transferred to 50 ml centrifuge tubes. 0.6 gm PEG 8000 and 0.9 gm NaCl was added to each suspension, mixed, incubated 1 hour at 4°C and the phage particles were subsequently pelleted at 15000 rpm for 30 minutes in a Sorvall SS-34 rotor. The supernatant was removed and the phage pellet was resuspended in 3 ml STE. The resuspended phage pellet was transferred to 15 ml polypropylene tubes. Three ml of freshly prepared phenol, equilibrated against 10 mM Tris-Cl (pH 8.0)- 0.2 mM EDTA
was added and vortexed extensively followed by centrifugation in a clinical table top centrifuge at 6000 rpm. The aqueous phase was removed to a clean tube and the extraction was repeated with an equal volume of a 1:1 mixture (v:v) of phenol and chloroform. The aqueous phase was extracted 3 times with diethylether. Any remaining ether in the final extraction was removed by evaporation. The single-stranded DNA was ethanol precipitated overnight at -20°C and the yield was quantitated using UV spectroscopy.

**Preparation of Heteroduplexes:** The methodology used initially to construct the heteroduplexes has been described by Kramer et al., (1984). Plasmids pMQ243, pMQ244, and pMQ250-253 were isolated from GM856, and their methylation state confirmed with *Sau3A, DpnI* and *MboI* endonuclease digestions. Each set of plasmids was linearized by digestion with *EcoRI* endonuclease and checked on a 1% agarose minigel. Typically, 4 μg of methylated (at dGATC sites) linear pMQ-DNA was mixed with 2 μg of unmethylated single strand pPY97 in 1X SSC (150 mM NaCl, 15 mM Na Citrate) and heated to 100°C for 7 minutes followed by 30 minutes at 65°C. The annealing mix was transferred to 37°C block for 30 minutes and then placed on ice. The nicked heteroduplexes were isolated from a 0.8% low melting point agarose gel (Sea Plaque) and purified with PrePac NACS columns according to the manufacturers instructions. Four units of T4 DNA ligase was added to the purified heteroduplexes in 50 mM Tris-Cl (pH 7.6), 10 mM MgCl₂, 20 mM DTT, 10 mM ATP and incubated for 18 hours at 14°C. This procedure yields 50 to 100 ng of hemimethylated heteroduplex.
Transformation and Analysis of Repair Procedures. CaCl$_2$ competent cells were prepared by inoculating 10 ml of Brain Heart Infusion with 100 $\mu$L of an overnight culture. The cells were grown at 37°C to a Klett of 85 (mid-log phase, $5 \times 10^8$ cells/ml) and centrifuged for 3 minutes at 12,000 xg. The supernatant was discarded and the pellet was resuspended in 10 ml CP buffer (50 mM CaCl$_2$, 10 mM PIPES pH 6.8). After 20 minutes on ice and subsequent centrifugation, the cell pellet was resuspended in 10 ml CP buffer and refrigerated at 4°C. The CaCl$_2$ competent cells were used within two days of preparation. A transformation mixture of 100 $\mu$L competent cells, 49 $\mu$L PCM buffer (10.5 mM PIPES pH 6.8, 13.6 mM CaCl$_2$2H$_2$O, 10 mM MgSO$_4$) and 1 $\mu$L plasmid was placed on ice for 5 minutes followed by 5 minutes at 37°C. 350 $\mu$L of L-Broth (Difco) was added to the cells followed by a 1 hour incubation at 37°C. The bacteria were spread onto Brain Heart plates containing 40 $\mu$g of ampicillin and placed at 37°C overnight. The Amp$^R$ transformants were replica-plated onto Brain-Heart plates containing 10 $\mu$g/ml tetracycline. The transformation procedure for the mnt structural gene system is the same as that described above, except that transformants are spread onto MacConkey/lactose agar containing 60 $\mu$g/ml ampicillin.

RESULTS AND DISCUSSION

pPY97 mnt operator (O$_{mnt}$) system of mismatch repair analysis: The maximum size of insertion/deletion heterologies that can be recognized and repaired by Dam-directed mismatch repair (DDMR) was initially investigated using the O$_{mnt}$ system. The pPY97 plasmid contains a 500 base pair fragment from bacteriophage P22 cloned into the EcoRI-HindIII site of pBR322 (Fig.1). The 500 base pair fragment contains the mnt
Plasmid pPY97

5.3 kb

M13 Ori

Pvul

Mnt repressor

bla

mnt

O_mnt

H

tet

P

P

E

Plasmid pPY97
**Figure 1. The Plasmid Map of pPY97.** pPY97 is a pBR322 derivative containing the M13 origin of replication and the genes coding for ampicillin (*bla*) and tetracycline resistance. The *mnt* operon, which was isolated from bacteriophage P22, is a 500 base pair region that was cloned into the *EcoRI* (E) and *HindIII* (H) sites of pBR322 to construct pPY97 (M. Susskind). *mnt* codes for a repressor that binds to its operator (*O*) and blocks the promoter controlling *tet* expression. The *tet* gene is expressed in the absence of either a functional Mnt repressor or its operator, i.e. due to mutation. Wild type pPY97 confers Amp<sup>R</sup>/Tet<sup>S</sup>, while the mutant form (*mnt* or *O*) confers Amp<sup>R</sup>/Tet<sup>R</sup>. P= promoter for RNA polymerase and the arrows indicate the direction of transcription.
gene, which encodes the Mnt repressor, the \textit{mnt} operator (O\textsubscript{mnt}), and a P\textsubscript{mnt}-tet gene fusion; RNA polymerase initiates transcription of \textit{tet} from the phage P22 ant promoter. The \textit{tet} gene is negatively regulated by Mnt since Mnt binding to the O\textsubscript{mnt} blocks transcription of the \textit{tet} gene. Conversely, any mutation within the O\textsubscript{mnt} that interferes with the binding of Mnt will result in constitutive expression of \textit{tet} since there is no Mnt mediated blockage of transcription. The expression of the \textit{bla} gene is independent of Mnt and confers ampicillin resistance (Amp\textsuperscript{R}) to cells transformed with either pPY97 or its \textit{mnt} mutant derivatives.

My initial approach to the investigation was to design an experimental system based on the properties of pPY97. I reasoned that repair of a heteroduplex containing an insertion/deletion heterology within the O\textsubscript{mnt} could be determined after transformation of mismatch proficient and deficient strains of \textit{E. coli} by screening the Amp\textsuperscript{R} colonies for \textit{tet} expression. In other words, repair of a heteroduplex using the wild type O\textsubscript{mnt} sequence as template would restore the sequence required for Mnt binding, hence a high percentage of Amp\textsuperscript{R} colonies would be Tet\textsuperscript{R}. Conversely, repair of a heteroduplex using the O\textsubscript{mnt} deletion mutant sequence as template would produce an O\textsubscript{mnt} sequence that Mnt cannot bind, hence Amp\textsuperscript{R} transformants would also be Tet\textsuperscript{R}. No repair would be indicated when 50\% of the Amp\textsuperscript{R} colonies are resistant to tetracycline since replication of an unrepaired heteroduplex should yield plasmids of which half are Tet\textsuperscript{R}.

The Marinus lab had previously isolated deletion mutations located within the O\textsubscript{mnt}. The size of each deletion was determined by sequence analysis creating a collection of mutations that ranged from 1 base to 15 base deletions. The 1 base deletion was used to
construct a 1 base insertion/deletion hemimethylated heteroduplex to serve as the experimental control. To determine the maximum size of the insertion/deletion heterology that could be repaired by mismatch repair, I constructed a group of hemimethylated heteroduplexes containing insertion/deletion heterologies within the O\textsubscript{mut} of 2, 4, 5, 7 and 15 bases (Fig. 2.). These, along with the 1 base control, were constructed by annealing unmethylated single-strand pPY97 to the EcoRI linearized complementary methylated strand (O\textsubscript{mut} deletion mutation) as described in Materials and Methods. GM1166 (repair proficient) and GM3856 (mutH) strains were transformed with these heteroduplexes and plated onto Brain-Heart plates containing 40 μg/ml ampicillin. Amp\textsuperscript{R} colonies that arose were subsequently replica-plated onto Brain Heart plates containing 10 μg/ml of tetracycline to determine the percentage of Amp\textsuperscript{R} colonies that were Tet\textsuperscript{R}. Figure 3 shows a representative sample of data collected with this system. These results suggested that 1 base heterologies were efficiently repaired, while 2, 4, 5, 7 and 15 base heterologies were not repaired. Analysis of the repair products through retransformation experiments and DNA sequencing were in agreement with the results of Figure 3. However, I have since determined that many of the initial transformation experiments contained experimental flaws giving rise to imprecise data interpretation. This placed a good deal of my initial data in question.

There were four main problems with the O\textsubscript{mut} system. First, the results expected after the first heteroduplex transformation of mutH cells (control) should have been close to 90% Amp\textsuperscript{R}/Tet\textsuperscript{R} rather than the observed 40% to 60% Amp\textsuperscript{R}/Tet\textsuperscript{R}. The reason for this is not clear; however one possibility is that the tetracycline concentration used to screen
AGGTCCACCACGTGGACCT

- T  1 bp del
- GT  2 bp del
- TCCA  4 bp del
- GGACC  5 bp del
- ACCGTGG  7 bp del
- CACCGTGACCTATT  15 bp del

EcoRI  mnt  O_mnt  HindIII  tet
Figure 2. The Location and DNA Sequence of Deletion Mutations within the mnt Operator. The operator spans the region from base 37 to base 53. P denotes the position of the promoter region of the mnt and tet genes, and the arrows are pointed in the direction of transcription. The thick lines indicate pBR322 sequence while the thin line between HindIII and EcoRI represents the 500 base pair mnt region of phage P22. The base deletions used in the O_mnt studies are shown directly below the O_mnt DNA sequence. Note that the 15 base deletion involves three bases outside of the O_mnt region (ATT-3').
REPAIR OF INSERTION/DELETION MUTATIONS BY *Escherichia coli*

90-100% Tet resistance indicates repair.
50% Tet resistance indicates no repair.
Figure 3. Repair of heteroduplexes containing either 1, 2, 4, 5, 7 or 15 base insertion/deletion mutations within the mnt operator. Hemimethylated heteroduplexes were methylated at d(GATC) sites on the deletion O\textsubscript{mnt} strand while the wild type pPY97 strand was unmethylated. Repair is indicated when 90% to 100% of the Amp\textsuperscript{R} colonies are also Tet\textsuperscript{R}, while 50% Tet\textsuperscript{R} indicates no repair. The mismatch proficient GM1166 (wild type) and mismatch deficient GM3856 (mutH) E. coli strains were used for this study. The percentages of Tet\textsuperscript{R} colonies were based on the total population of transformants, which ranged from approximately 500 to 800 total colonies per transformation.
for Tet<sup>r</sup> was too high at 10 μg/ml. Alternatively, heteroduplexes isolated from low melting point agarose gels may have contained small amounts of single strand pPY97 thus giving rise to Amp<sup>r</sup>/Tet<sup>r</sup> colonies. Second, preparations of single-strand pPY97 used in the construction of heteroduplexes was contaminated with mutant O<sub>mnt</sub> single-strand DNA as a result of contaminated phage R408 stocks. This created heteroduplexes of unknown definition. To correct this, new phage R408 stocks were cultivated and subsequently used to produce new single strand DNA, which was analyzed for purity by transduction and transformation assays. Third, a major disadvantage of this system was the large number of re-transformations required to collect accurate data. This made the system very labor intensive. Fourth, the mutations within the O<sub>mnt</sub> palindromic sequence were isolated as spontaneous mutations in mismatch repair proficient strains of <i>E. coli</i>. This gives rise to the possibility that these deletion mutations have a pre-existing bias against repair.

Some of the problems described above have since been worked out and the results are presented in Chapter 4 and in Figure 4. Briefly, hemimethylated heteroduplexes were constructed by annealing unmethylated single strand O<sub>mnt</sub> deletion mutants (1, 2, 4 and 5 base deletions) to EcoRI linearized methylated pPY97 (wild type O<sub>mnt</sub>). Transformation of GM1166, GM3856 (<i>mutH</i>) and GM4719 (<i>mutL</i>) <i>E. coli</i> strains with the hemimethylated heteroduplexes produced the data for Figure 4. Since the d(GATC) sequences are methylated on the strand containing the wild type O<sub>mnt</sub> sequence (pPY97), repair should be directed against the strand containing the O<sub>mnt</sub> deletions giving rise to pPY97 plasmids and a Amp<sup>r</sup>/Tet<sup>r</sup> phenotype. The results suggested that 1 and 2 base insertion/deletion heterologies within the O<sub>mnt</sub> were efficiently repaired by Dam-dependent mismatch repair.
REPAIR OF INSERTION/DELETION MUTATIONS BY *Escherichia coli*

0-10% Tet resistance indicates repair.
50% Tet resistance indicates no repair.
Figure 4. Repair of 1, 2, 4 and 5 base insertion/deletion mutations within the mnt operator. Mismatch proficient (GM1166) and mismatch deficient (GM3856 (mutH) and GM4719 (mutL)) strains of E. coli were transformed with hemimethylated heteroduplexes containing the either 1, 2, 4 or 5 base deletions. In this case, the wild type pPY97 strand was methylated at the d(GATC) sites, while the O_mnt deletion strand was unmethylated. Repair of heteroduplexes was indicated when 0% to 10% of the Amp^R colonies were resistant to tetracycline, and no repair was indicated when 50% of the cells were Tet^R. The percentages of Tet^R colonies were based on the total population of transformants, which ranged from approximately 300 to 500 total colonies per transformation.
repair, while 4 and 5 base insertion/deletion heterologies were not (Figure 4.). This was in agreement with initial results generated with the newly developed mnt system described below. Due to the early success of the new system, additional studies with the O_mnt sequence were not pursued further.

The mnt structural gene system. Due to the problems that arose with the O_mnt system, I decided to take a different approach to the investigation of mismatch repair of insertion/deletion mutations. This involved two major alterations to the O_mnt system that ultimately resulted in a highly sensitive assay for the rapid analysis of mismatch repair activity in vivo. The first alteration involved the use of deletion mutations within the mnt structural gene of pPY97 rather than within the O_mnt palindromic sequence. The other alteration involved the use of the E. coli indicator strain GM4331. This strain was originally constructed for use in the testing of various mnt mutations for dominance, however it was immediately apparent that this strain could also serve as a vehicle to evaluate the repair of heteroduplexes.

The indicator strain is a λ-lysogen containing a mnt lacZ fusion operon which places the expression of lacZYA under the negative control of Mnt (Fig. 5.). One of the two key components of the indicator strain is that the chromosomal copy of mnt encodes a non-functional repressor. Therefore, the only potential source of Mnt repressor is that which is plasmid borne, i.e. pPY97 encoded mnt. This sets up the other key component which involves the mnt-lacZYA fusion operon. The expression of lacZYA will cause cells plated on MacConkey/lactose agar to turn red, while blocking the expression of lacZYA will allow them to remain white. From this I reasoned that mismatch repair of pPY97
Transform *E. coli* with pPY97 containing wild type *mnt*  

GM4331  

plate on MacConkey/Lactose/Amp  

White colonies = WT Mnt  
Tet sensitive

Transform *E. coli* with pPY97 containing mutant *mnt*  

GM4331  

plate on MacConkey/Lactose/Amp  

Red colonies = mutant Mnt  
Tet resistant
Figure 5. Description of the mnt structural gene system. As described in the text, the E. coli indicator strain, GM4331, transformed with pPY97 and plated on MacConkey/lactose agar gives rise to white colonies that are Tet\textsuperscript{5}. pPY97 encodes wild type Mnt repressor that binds to the O\textsubscript{mnt} on both the chromosome and the plasmid effectively blocking transcription of lacZY\textsubscript{A} and tet respectively. Conversely, transformation of GM4331 with a plasmid that codes for non-functional Mnt repressor gives rise to Red colonies on MacConkey/lactose agar that are Tet\textsuperscript{R}. Non-functional Mnt repressor fails to bind to its O\textsubscript{mnt} and block transcription of lacZY\textsubscript{A} and tet. O is the O\textsubscript{mnt}, P is the promoter for lacZY\textsubscript{A}, and T signifies the plasmid tet gene.
derived heteroduplexes containing an insertion/deletion heterology within the \textit{mnt} gene could be evaluated on the basis of colony color. In other words, transformants giving rise to colonies that are both red and white ("mixed") is diagnostic of no repair, while transformants that are either red (\textit{mnt'}, constitutive expression of \textit{lacZYA}) or white (\textit{mnt'}, Mnt binds to O\textit{mnt} and blocks transcription of \textit{lacZYA}) is diagnostic of repair. I will refer to this system as the \textit{mnt} system.

\textbf{Evaluation of the \textit{mnt} System. } The \textit{mnt} system was first tested by a series of transformation experiments. As outlined in Figure 5, transformation of GM4331 with pPY97 (\textit{mnt'}) should give rise to white colonies on MacConkey/lactose agar since pPY97 encoded Mnt repressor binds to the O\textit{mnt} on the chromosome blocking transcription of \textit{lacZYA}. Unlike pPY97, transformation of GM4331 with p4349 (\textit{mnf}) should give rise to red colonies when plated on MacConkey/lactose agar. p4349 is identical to pPY97 except that it contains a 1 base insertion (cytosine) within the \textit{mnt} gene that results in the expression of a non-functional Mnt repressor. When GM4331 is transformed with pPY97, only white colonies appear (Table I, Figure 6A), while transformation with p4349 gives rise to only red colonies (Table I, Figure 6B). Since pPY97/GM4331 are white colonies and p4349/GM4331 are red, I predicted that mismatch repair of a hemimethylated heteroduplex containing an unmethylated strand of pPY97 (\textit{mnt'}, white) annealed to a methylated strand of p4349 (\textit{mnf'}, red) would be indicated by the presence of red colonies rather than white colonies. Colonies arising from transformation of a mismatch repair deficient indicator strain, GM4348 (\textit{mutL}), should be both red and white ("mixed"). As predicted, when GM4331 was transformed with the 1 base
Table I. Transformation of GM4331 and GM4348 with Control Plasmids/Heteroduplex.

<table>
<thead>
<tr>
<th>Plasmid/Heteroduplex</th>
<th>GM4331 (WT)</th>
<th>GM4348 (mutL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Colony color and number)</td>
<td>Red</td>
<td>White</td>
</tr>
<tr>
<td>pPY97 (WT)</td>
<td>0</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>p4349 (+C 62-63 mnt)</td>
<td>&gt;1000</td>
<td>0</td>
</tr>
<tr>
<td>p3262 (mnt')</td>
<td>&gt;1000</td>
<td>0</td>
</tr>
<tr>
<td>pPY97 ME/p4349 +ME</td>
<td>316</td>
<td>46</td>
</tr>
</tbody>
</table>

Transformation of GM4331 and GM4348 (mutL) E. coli strains with pPY97 (mnt'), p4349 (mnt'), p3262 (mnt'), and a pPY97/p4349 heteroduplex gave rise to colonies that were either red, white, or mixed when plated onto MacConkey/lactose agar selective for Amp. The number of colonies and their color are indicated to the right of plasmid/heteroduplex description. The heteroduplex represents a 1 base heterology in which unmethylated (-ME) single strand pPY97 was annealed to the complementary methylated (+ME) EcoRI linearized p4349. p4349 contains a +C insertion between 62-63 of the mnt gene.
A. pPY97 control (GM4331)

B. p3459 control (GM4331)

C. Hemimethylated heteroduplex (GM4331)

D. Hemimethylated heteroduplex (GM4348, mutL)
Figure 6. Colonies on MacConkey/lactose are either red, white, or mixed.  
Panel A. Control transformation of GM4331 with pPY97 (mnt). Panel B. Control transformation of GM4331 with p3459 (mnt). Panel C. Transformation of GM4331 (mismatch proficient) with a hemimethylated heteroduplex consisting of an unmethylated pPY97 strand annealed to its complementary methylated mnt mutant strand. Panel D. Transformation of GM4348 (mutL) with the same hemimethylated heteroduplex described in panel C.
insertion/deletion heteroduplex, approximately 84% of the colonies (316/379) were red, 12% were white and 4% were "mixed" (Table I, Figure 6C), while transformation of GM4348 (mutL) resulted in mostly mixed colonies (Table I, Figure 6D).

To determine the accuracy of this system, I picked 20 red, 20 white and 4 mixed colonies from each of the 1 base insertion/deletion heteroduplex transformation experiments, isolated the plasmid DNA and retransformed the indicator strains. If the red colonies truly reflect the presence of only p4349 (mnr'), then retransformation of a freshly prepared indicator strain with the plasmid isolates should give rise to only red colonies. Similarly, if the white colonies reflect the presence of only pPY97 (mnr'), then retransformation of the indicator strain should give rise to only white colonies. Mixed colonies should give rise to both red and white colonies on the same plate stressing the point that individual cells within a colony contain both p4349 and pPY97 plasmids. The results of these experiments are summarized in Table II. Plasmid DNA isolated from red colonies gave rise to only red colonies on retransformation, and mixed colonies gave rise to both red and white colonies as expected. The retransformation experiments involving plasmid DNA isolated from white colonies was quite surprising since it indicated that not all white colonies contain only pPY97 plasmids (Table II). Six out of 20 white colonies isolated from GM4331 and four of 20 colonies isolated from GM4348 heteroduplex transformations gave rise to both red and white colonies indicating that 10 of the 40 white colonies (25%) are actually "mixed" colonies, or "false white". The results of these
**Table II. Retransformation of Red, White and Mixed Colonies.**

**Colony color of Retransformants**

<table>
<thead>
<tr>
<th>#colonies</th>
<th>Color</th>
<th>Strain</th>
<th>R/T</th>
<th>W/T</th>
<th>M/T</th>
<th>Amp$^R$</th>
<th>Tet$^R$</th>
<th>%Tet$^R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Red</td>
<td>GM4331</td>
<td>20/20</td>
<td>0/20</td>
<td>0/20</td>
<td>20</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>20</td>
<td>White</td>
<td>GM4331</td>
<td>0/20</td>
<td>14/20</td>
<td>6/20</td>
<td>20</td>
<td>6</td>
<td>30%</td>
</tr>
<tr>
<td>4</td>
<td>Mixed</td>
<td>GM4331</td>
<td>0/4</td>
<td>0/4</td>
<td>4/4</td>
<td>4</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>20</td>
<td>Red</td>
<td>GM4348</td>
<td>20/20</td>
<td>0/20</td>
<td>0/20</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>20</td>
<td>White</td>
<td>GM4348</td>
<td>0/20</td>
<td>16/20</td>
<td>4/20</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>Mixed</td>
<td>GM4348</td>
<td>0/4</td>
<td>0/4</td>
<td>4/4</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Plasmid DNA was isolated from red, white and mixed colonies after transformation of GM4331 (WT) and GM4348 (mutL) strains with the pPY97/p3459$^{MB}$ heteroduplex. DNA isolated from each colony was used to retransform either GM4331 or GM4348 and plated onto MacConkey/lactose agar selective for Amp$^R$. R/T is the number of total transformations that gave rise to only red colonies. W/T = number of transformations that gave rise to pure white colonies and M/T are the number of transformations that gave rise to both red and white colonies. Amp$^R$ = number of ampicillin resistant colonies, Tet$^R$ = number of Amp$^R$ colonies that were resistant to Tetracycline.
re-transformation experiments are in complete agreement with re-transformation experiments that were recently performed by Greta Carraway (a graduate student in the Mar- nus Lab), who has begun using the mnt system to investigate RecA mediated repair of large insertion/deletion heterologies.

The reason for the presence of "false white" colonies is not completely clear. One possibility deals with the rate of growth of the indicator strain on MacConkey/lactose agar. Qualitative observations of cells containing pPY97 indicate that they appear to grow much better on MacConkey/lactose than cells containing p4348 (and mnt plasmids in general). Since mnt cells do not block the expression of lacZYA, the acid by-products of lactose fermentation pushes the pH in the immediate area of the colony downward resulting in poor growth. Conversely, mnt cells block lacZYA expression and lactose fermentation which favors better growth since these cells grow in a more normal physiologic pH. If the faster growing cells (mnt) produce an alkaline environment that titrates the acidic contribution from the slower growing cells (mnt), then the pH of the MacConkey/lactose agar would favor the appearance of a white colony rather than one that is red and white. I determined the growth curves for both red and white cells in MacConkey/lactose media and found very little difference. However, it turns out that the liquid MacConkey/lactose media (pH 7.3) is substantially different from that used to prepare the agar (pH 7.1), thus making any conclusions about colony growth on MacConkey/lactose plates inappropriate.

I approached the problem of identifying the false white colonies by utilizing a property of the pPY97 plasmid. I reasoned that false white colonies, by virtue of the fact that some of the cells within the colonies contain mnt mutant plasmids conferring resistance
to tetracycline (Fig. 1), should grow on MacConkey/lactose plates containing tetracycline while true white colonies should not (Fig. 3). This being the case, the MacConkey/lactose system has a built-in internal control for identifying false white colonies. In Table II, six of the 20 white colonies picked from the heteroduplex transformation experiment grew on plates containing 10μg/ml tetracycline. To determine the minimum concentration of tetracycline that effectively separates the false white colonies from the true white colonies, I performed another heteroduplex transformation experiment to provide a source of fresh white colonies. From this transformation, I isolated 30 white colonies and patched them onto each of 11 MacConkey/lactose plates containing either 0 or 1 - 10 μg/ml tetracycline. A representative portion of the results are shown in Table III, and indicate that the minimal concentration of tetracycline required to identify false white colonies lies between 2 and 3 μg/ml. I chose to use 3 μg/ml concentration of tetracycline to identify false white colonies in future heteroduplex transformation experiments.

**Improving the Heteroduplex Construction Methodology.** The methodology used in heteroduplex construction also posed a number of problems, some of which may have contributed to the initial experimental inconsistencies. I had three areas of concern regarding heteroduplex construction. The first area of concern involved the low yield of heteroduplexes. Since the concentration of heteroduplexes following extraction from low melting point agarose gels was only about 100 ng, analysis of the covalent state and the methylation state of the molecule was difficult. Southern blots could have been used to analyze the restriction digests required for the determination of methylation state, but it
Table III. Concentration of Tetracycline Required to Detect False White Colonies.

<table>
<thead>
<tr>
<th>µgTet/mlagar</th>
<th>Amp&lt;sup&gt;R&lt;/sup&gt; colonies</th>
<th>Tet&lt;sup&gt;R&lt;/sup&gt; colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

White colonies from a heteroduplex transformation were tested to determine the concentration of tetracycline required to identify the mixed colonies. Individual white colonies were first patched onto MacConkey/lactose plates containing 60 µg Amp/ml and then patched on to MacConkey/lactose plates containing 0, 1, 2, 3, 4, 5, 10 µg tetracycline/ml.
would not have solved the problem of determining the covalent state of the molecule. Second, the heteroduplexes may have contained nicks at the EcoRI site, which was less than 500 base pairs from the O_{mut}. Since the control strain used in these experiments was E. coli mutH, the EcoRI nick may have been sufficient to bypass the requirement for MutH. Even though the heteroduplexes were treated with T4 DNA ligase, the small yield of heteroduplexes prevented accurate analysis of the covalent state of the molecule.

Third, the annealing procedure required that the linear duplex DNA be denatured by boiling the annealing mixture for about 7 minutes. This procedure may have introduced abasic sites into the DNA or created nicks to serve as entry points for mismatch repair.

To address the heteroduplex construction problems, I began using the procedure of Lu et al., (1983), which is discussed in Chapter 4 of this thesis. The major changes are described below. First, plasmid DNA was linearized using PvuII endonuclease rather than EcoRI since PvuII cleaves 2.5 kbp from the mnt region, thus reducing the potential for MutH bypass. I also constructed an isogenic mutL strain (GM4719) to use as an additional control for the O_{mut}/pPY97 plasmid system. Second, the procedure to construct the heteroduplexes involved a gentle alkaline denaturation-renaturation step to avoid the harsh conditions created by 100°C temperatures. Third, hydroxylapatite chromatography was used to separate the heteroduplexes from single strand DNA to reduce contamination problems. Fourth, the nicked heteroduplexes were converted to a covalently closed state by E.coli DNA ligase in the presence of ethidium bromide and subsequently separated from the nicked heteroduplexes on a CsCl/ethidium bromide equilibrium gradient. Finally, the procedure starts with 100 µg of single strand DNA and 500 µg of linearized
duplex DNA producing a typical yield of 10 to 15 μg of covalently closed heteroduplexes. This was more than enough to accurately analyze the state of methylation for each set of heteroduplexes.

Summary. The various approaches taken to investigate the role of Dam-directed mismatch repair in the recognition and correction of insertion/deletion mutations have been discussed. The problems that I encountered with each approach were presented along with the steps that were taken to solve them. Of the approaches taken, the O_mnt system presented the most difficulties. I now believe that the majority of those problems have been worked out as can be observed in Figure 4. However, the initial problems of the O_mnt system forced the investigation of alternative pathways, which ultimately led to the development of a highly sensitive in vivo mnt assay for mismatch repair. I chose to continue the investigation using the mnt system due to its early success, and terminate the work involving the O_mnt system. The data collected using this system have been highly reproducible, and this success is described in Chapter 4 of this thesis.
CHAPTER 4

Repair of DNA Heteroduplexes Containing Small Heterologous Sequences in *Escherichia coli*. 

Page 70
ABSTRACT

Plasmid heteroduplexes were constructed that contain either 1, 2, 3, 4 or 5 unpaired bases within the mnt gene. These were used to assess the efficiency of repair of small heterologous sequences ("heterologies") in DNA by the Escherichia coli Dam-directed repair system. Heteroduplexes in defined states of methylation at d(G-A-T-C) sites were transformed into a repair proficient indicator strain (which has a mnt-lac fusion coding for a non-functional mnt repressor) and its isogenic mutH, L and S derivatives. Using this in vivo transformation system, we scored for repair on the basis of colony color: correction in favor of the strand bearing mnt* coding information gives rise to colonies that are white, whereas correction in favor of the opposite strand (mnt") yields colonies that are red when grown on MacConkey agar. Failure to repair a heterology yields colonies that are both red and white ("mixed"). The correction efficiencies of two heteroduplexes, each containing a single G-T mismatch within mnt, were also monitored for purposes of comparison. Our results show that the repair of heteroduplexes containing 1, 2 and 3 base deletions is as highly efficient as the repair of G-T mismatches. Heteroduplexes containing a 4 base deletion are marginally repaired and DNA containing a 5 base deletion is not detectably repaired. Like the G-T mispairs, correction of the 1, 2, 3 and 4 base heterologies is Dam-dependent and requires MutH, L and S function. In addition, we show that purified MutS protein from Salmonella typhimurium, which can substitute for E. coli MutS in vivo, binds to oligonucleotide duplexes containing 1, 2, 3 and 4 unpaired bases of the identical sequence as that used for the in vivo studies. Specific binding of MutS to homoduplex DNA and to DNA that had undergone a 5 base deletion was not observed.
INTRODUCTION

The mismatch repair systems of Escherichia coli are involved in the correction of DNA containing mismatched bases (for reviews, see ref. 1-3). Mismatched bases can be generated through a number of processes: spontaneous deamination of 5-methylcytosine to thymine giving rise to G-T mispairs, allelic differences in recombination intermediates and biosynthetic errors that arise during DNA replication. Mismatch repair in E. coli is divided into two pathways, long patch ("Dam-dependent") and short patch repair. These pathways are characterized by the length of the excision-repair tract produced during the correction process. In general, short patch repair systems generate small excision-repair tracts up to 10 nucleotides in length during a repair process that corrects specific base:base mismatches (1-4). This pathway is independent of Dam-directed discrimination between the two DNA strands. Conversely, long patch ("Dam-dependent") repair has a broader substrate specificity since it corrects most base:base mismatches in a Dam-directed process characterized by long excision-repair tracts up to and in excess of one kilobase in length (4). This system relies on the methylation of adenine in d(G-A-T-C) sequences to direct the strandedness of repair; this is accomplished by DNA adenine methyltransferase (Dam). The repair efficiency of individual base:base mismatches by Dam-directed repair depends on the nature of the mismatch as well as the DNA sequence environment within which it resides (5, 6). Generally, transition mismatches, i.e. G-T and A-C, are repaired more efficiently than transversion mismatches, while the C-C mismatch is poorly repaired if at all.
The biochemical mechanisms involved in correcting mismatched bases in DNA by the Dam-directed system are still unclear. However, it appears likely that the process involves at least three steps: initiation, repair-tract excision and resynthesis (4). Initiation requires the presence of a mismatch and the gene products of mutH, L and S. MutS is a 97 kDa protein that recognizes and binds mismatched bases in DNA (6, 7) leading to the formation of an alpha loop structure in the presence of a base:base mispair and ATP (4). MutL, a 70 kDa protein, in the presence of ATP has been shown to interact with the MutS/mismatch complex since the level of DNaseI protection is greatly increased relative to protection provided by MutS binding alone (8). Presumably after the formation of a MutS/mismatch complex and in the presence of MutL and ATP, a search for a hemimethylated d(G-C) sequence ensues (4). MutH, containing a latent 25 kDa d(G-A-T-C) specific endonuclease activity, is activated in the presence of this complex and makes an incision 5' to the d(G-A-T-C) sequence on the unmethylated strand (4, 9). It is through this process that the repair system discriminates between the parental strand containing the methylated d(G-A-T-C) sequence from the daughter strand containing the unmethylated sequence (10). Symmetrically methylated sequences are resistant to MutH activity, while unmethylated d(G-A-T-C) sequences can be incised on one side or the other leading to loss of strand bias (9). It should be emphasized that a clearly defined biological function has yet to be assigned to MutL, however it has been suggested that it acts to interface the activity of MutS with that of MutH to coordinate an efficient repair process (4). At some point after MutH makes its incision, excision of the repair tract begins. This step presumably involves DNA helicase II, single stranded binding protein and an
exonuclease to produce excision tracts typically spanning the region between the mismatch and the d(G-A-T-C) sequence (11, 12). Finally, resynthesis is carried out by DNA polymerase III holoenzyme and the DNA molecule is restored to its covalently closed state by DNA ligase (11).

Although the correction of base:base mismatches in DNA by Dam-directed mismatch repair is well documented (1-3), little is known about the specificity of repair of small DNA heterologies. The present study examines the correction of insertion/deletions of 1 to 5 bases to determine the size constraints of the repair pathway both in vivo and in vitro.

**MATERIALS AND METHODS**

**Plasmid, Phage and Bacterial Strains.** Plasmid pPY97 and the numbering of bases in the mnt gene and mnt operator has been described by Rewinski and Marinus (13). The f1 phage R408 was a gift from Dr. M. Russel (14); λ GM110 is λimm$^{21}$ mnt$^{::}$lacZYA nin-5. *E. coli* strain GM4331 (F$^-$ thr-1 Δ(lac-pro)$_{XY}$ tsx-33 supE44 galK2 hisG4(Oc) rfb-1 mgl-51 rpsL31 kdgK51 xyl-5 mtl-1 argE3 thi-1 (λ GM110)); GM4348 (GM4331 but mutL::Km Stu (15)); GM4349 (GM4331 but mutH471::Tn5), GM4388 (GM4331 but mutS201::Tn5 (16)), and GM1690 (F42/dam-16(Del;KmR) thr-1 ara-14 leuB6 Δ(gpt-proA)$^{62}$ lacY1 tsx-33 supE44 galK2 hisG4(Oc) rfb-1 mgl-51 rpsL31 kdgK51 xyl-5 mtl-1 argE3 thi-1) were constructed in this lab. *E. coli* JC9239 (17) is AB1157 (Table 1, Chapter 2) but recF143.

**Preparation of Heteroduplexes.** pPY97 containing deletions of 1, 2, 3, 4 and 5 base pairs within mnt were prepared by site directed mutagenesis using T7 DNA polymerase
The sequences of the mutant constructs were confirmed by DNA sequence analysis (18). Plasmid DNA methylated at d(G-A-T-C) sites was isolated from JC9239, while that devoid of methylation was isolated from GM1690 using Qiagen maxi-columns as recommended by the manufacturer. Single-stranded plasmid DNA devoid of methylation at d(G-A-T-C) sites was extracted and purified from f1 phage R408 virions using plasmid containing GM1690 as host (20).

The construction of heteroduplexes containing 1, 2, 3, 4 or 5 unpaired bases within mnt and the G-T mismatches at position 40 and 41 in mnt were prepared using the method of Lu et al. (21). Briefly, single strand plasmid DNA was isolated from f1 phage R408 virions after infection of GM1690 transformed with the appropriate plasmid. Heteroduplexes were constructed by annealing single-stranded DNA to its complementary alkaline denatured, PvuII (New England Biolabs) digested plasmid DNA at 65°C for 30 minutes. To obtain optimal heteroduplex formation, the annealing mixture contained a 5:1 molar ratio of single stranded DNA to linear duplex DNA. The nicked heteroduplexes were subsequently separated from any excess single strand DNA using hydroxylapatite chromatography (Bio-Rad Laboratories), after which they were converted to a covalently closed supercoiled state by E. coli DNA ligase (β-NAD) (New England Biolabs) in the presence of ethidium bromide. These covalently closed heteroduplexes were separated from nicked heteroduplexes by centrifugation through a cesium chloride/ethidium bromide equilibrium gradient. This study used hemimethylated heteroduplexes involving two different constructs; one in which unmethylated circular single strand wild type pPY97 was annealed to its complementary methylated linear duplex DNA containing deletions of 1-5 base pairs or G-C to A-T transitions, and the other in
which unmethylated circular single strand containing deletions of 1-5 bases was annealed to its complementary methylated linear pPY97 DNA. The hemimethylated heteroduplexes were tested for resistance to MboI and Sau3A digestion.

Fully unmethylated heteroduplexes were formed by annealing unmethylated duplex linear DNA containing deletions to unmethylated circular single strand pPY97 DNA. Digestion by MboI, but not DpnI indicated that these heteroduplexes lack methylation at d(G-A-T-C) sequences. To prepare fully methylated heteroduplexes, in vitro methylation of hemimethylated heteroduplexes, constructed by annealing methylated mutant DNA to unmethylated single strand pPY97, was achieved using Dam methylase in the presence of S-adenosylmethionine (New England Biolabs). Digestion by Sau3A and DpnI indicated that greater than 95% of the d(G-A-T-C) sequences were fully methylated as determined by gel electrophoresis.

Transformation and Analysis of Repair Procedures. Competent cells of strains GM4331, GM4348, GM4349 and GM4388 were prepared freshly when needed as described by Caraway et al., (22). Approximately 1 ng of heteroduplex was used for each transformation by the CaCl2 procedure. The transformants were spread onto Mac-Conkey agar plates containing 60 μg ampicillin/ml and placed at 37 oC for 18 hours. Colonies were scored for repair based on their color. White colonies were patched onto plates containing 3 μg/ml tetracycline to separate the true white colonies from the false white or mixed colonies since true white colonies are sensitive to growth on plates containing tetracycline.
MutS Band Shift Assay. The [γ-32P]ATP (Amersham) end labelling of the bottom strand by T4 polynucleotide kinase (New England Biolabs), annealing of the oligonucleotides (Operon Technologies, Alameda, CA), MutS binding reactions and the band shift assays were carried out as outlined by Jiricny et. al. (23) with the following exceptions. Purified Salmonella typhimurium MutS protein (24), a generous gift from Dr. Graham C. Walker (M.I.T.), was used in place of E. coli MutS. The MutS binding reaction consisted of 0.6 pmol of 36 base oligonucleotide duplex with either 2, 4, 6, 8, 10 or 12 pmols MutS added to the mixture and subsequently incubated at 0°C for 30 minutes. The 36 base oligonucleotides used to make the duplexes were as follows:

5'-GCATACGGAAGTTAAAGTGCGGATCATCTCTAGCCA-3' was the bottom strand and represents the complementary sequence to the top strand,

5'-TGGCTAGAGATGATCCGCACTTAACTCCGTATGC-3'. The bold sequence in the top strand corresponds to Figure 1 which shows the base deletions used to create the 1, 2, 3, 4 and 5 base deletion duplexes. The G-T₄₀ mismatch was constructed by annealing 5'-TGGCTAGAGATGATCCGCTACTTTAACTCCGTATGC-3' to the bottom strand.

RESULTS

The in vivo repair assay can determine the repair efficiency of small DNA heterologies. The plasmid, pPY97, has been described in detail elsewhere (13). Briefly, pPY97 is a pBR322 derivative with an M13 origin of replication so that single strand DNA can be prepared. It confers ampicillin resistance and contains an mnt-tet operon fusion that
places tet expression under the control of Mnt. Site directed mutagenesis was used to create 1, 2, 3, 4 and 5 base deletions within the mnt structural gene (Figure 1). To insure that a non-functional repressor was created, the mutagenesis was directed against the His6 residue of mnt, a residue that has been shown to be critical for proper binding to Omnt (25). The mnt gene was sequenced for each mutant in order to confirm that both the correct mutant mnt sequence had been isolated and that no other mutations were present.

A system has been developed to detect accurately the correction of DNA containing insertion/deletion heterologies as well as base:base mismatches (Fig. 1). Correction of plasmid DNA heteroduplexes is scored on the basis of color changes in colonies after an indicator strain has been transformed with heteroduplex and incubated on MacConkey/ampicillin agar. The indicator strain, GM4331 and its isogenic mutH, L and S derivatives, is a λ lysogen containing lacZYA, and the mnt gene and its operator of bacteriophage P22. The promoter for lacZYA is negatively controlled by Mnt repressor; however in GM4331 the mnt gene codes for a non-functional repressor allowing lacZYA to be constitutively expressed. GM4331 forms red colonies on MacConkey agar. Transformation of the indicator strain with a plasmid coding for a non-functional Mnt repressor also results in lacZYA expression thus forming red colonies on MacConkey agar, and in addition, confers plasmid encoded tetracyline resistance (Figure 1). Conversely, plasmid containing a functional Mnt repressor which binds to the mnt operator (Omnt) on both the chromosome and the plasmid effectively block expression of lacZYA and tetracycline resistance respectively, and thus colonies are white on MacConkey agar (Figure 1).
Phenotype

Chromosome

<table>
<thead>
<tr>
<th>O_{mnt}</th>
<th>lacZ/YA</th>
</tr>
</thead>
</table>

Plasmid

<table>
<thead>
<tr>
<th>O_{mnt}</th>
<th>tet</th>
</tr>
</thead>
</table>

Mnt repressor

Phenotype

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>White</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>mnt^{+}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mnt^{-}</td>
<td>tet^S</td>
<td>tet^R</td>
</tr>
</tbody>
</table>

His^{6}

3'-CTT CAA TTT CAC GCC TAG TAG-5'

\[
\begin{align*}
\text{1 base del} & : & -A \\
\text{2 base del} & : & -C G \\
\text{3 base del} & : & -AC G \\
\text{4 base del} & : & -AC GC \\
\text{5 base del} & : & -T CAC G 
\end{align*}
\]
Figure 1. The indicator system used to evaluate efficiency of repair is based on two major components, plasmid encoded Mnt and chromosomal mnt-lacZYA fusion operon. Plasmid encoded Mnt repressor negatively regulates the expression of both the plasmid encoded tet gene and the chromosomal lacZYA operon. The chromosomal mnt gene codes for a nonfunctional repressor, and thus fails to bind O_mnt. The plasmid encoded bla gene confers resistance to ampicillin. The DNA sequence of mnt shows the base(s) that were deleted in the strand complementary to single strand pPY97 as well as the surrounding sequence. These mutations were constructed so that the His^6 residue within mnt would be eliminated, thus producing a nonfunctional Mnt repressor. The sequence is oriented 3' to 5' to indicate the direction of transcription of mnt. P represents the promoter region and the arrow indicates the direction of transcription for tet and lacZYA genes.
When a heteroduplex of a defined methylation state consisting of a strand containing the wild type mnt sequence and its complementary strand containing an mnt deletion mutation is introduced into an indicator strain, we observe three possible outcomes; transformants giving rise to either red, white or red plus white ("mixed") colonies. Correction of a small DNA heterology is indicated by the presence of either white or red colonies after transformation of the indicator strain with hemimethylated heteroduplexes in which the methylated strand d(G-A-T-C) is wild type or mutant respectively. No repair is indicated by the presence of mixed colonies since replication of the heteroduplex that is refractory to repair should yield colonies that harbor both wild type and mutant forms of pPY97.

To determine the reliability of this method, we transformed the indicator strains with heteroduplex DNA and spread the transformants onto MacConkey/ampicillin agar. The purity of the transformants was evaluated by isolating plasmid DNA from red, white and mixed colonies and retransforming the indicator strain GM4331. The color of the retransformed colonies reflects the plasmid purity. Plasmid DNA isolated from 20 red colonies gave rise to only red colonies after retransformation indicating that these transformants contained mutant pPY97. Plasmid DNA isolated from mixed colonies gave rise to either red or white colonies after retransformation indicating that replication of a heteroduplex proceeded in the absence of repair. All mixed colonies grew well on plates containing 10 μg/ml tetracycline. Plasmid DNA isolated from most white colonies also gave rise to white colonies after retransformation; however DNA from some white colonies gave rise to both white and red colonies indicative of a mixed colony. Therefore, since true white colonies are sensitive to tetracycline and mixed colonies are resistant, we
determined the concentration of tetracycline required to separate true white colonies from "false white" or mixed colonies. True white colonies will not grow on plates that contain 3 μg/ml tetracycline and retransformation using plasmid DNA isolated from these colonies gives rise to white colonies only. Conversely, false white colonies are resistant to tetracycline and since they give both red and white colonies on retransformation, they are scored as mixed colonies.

Transformations of GM4331 (wild type) and the mut strains with homoduplex (pPY97), the various heteroduplexes (and states of methylation), and a 1:10 mixture of homoduplex (pPY97) to heteroduplex revealed no significant loss of heteroduplex viability that might occur as a result of lethal processing of deletion mutations.

In vivo correction of small DNA heterologies by mismatch repair is Dam-directed. The mismatch repair proficient indicator strain, GM4331, and its isogenic mutH, L and S strains were transformed with hemimethylated heteroduplexes, in which the mutant strand contains methylated d(G-A-T-C) sequences, containing either 1, 2, 3, 4 or 5 base heterologies. A G-T mismatch at position 40 and a T-G mismatch at position 41 of mnt were used as positive controls. We chose to compare the efficiency of correction of DNA containing small heterologies to that of the G-T mismatches since the Dam-directed repair system has been shown to correct them with high efficiency. In addition, the G-T mismatches do not lie within sequences that favor VSP repair. Table I shows that 1, 2, and 3 base deletion heteroduplexes are repaired by wild-type mismatch repair proficient cells in a Dam-dependent direction with an efficiency similar to that of the G-T mismatches since 97% to 99% of the wild type transformants give rise to red colonies. The 4
<table>
<thead>
<tr>
<th>Heteroduplex (Mutant/WT)</th>
<th>Wild Type mutL::Km</th>
<th>mutH471::Tn5</th>
<th>mutS201::Tn5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(++)</td>
<td>G-T 40</td>
<td>T-G 41</td>
<td></td>
</tr>
<tr>
<td>(+/-)</td>
<td>98</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(+/-)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(+/-)</td>
<td>98</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(+/-)</td>
<td>99</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>(+/-)</td>
<td>99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(+/-)</td>
<td>97</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(+/-)</td>
<td>98</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(+/-)</td>
<td>98</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(+/-)</td>
<td>97</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(+/-)</td>
<td>39</td>
<td>4</td>
<td>57</td>
</tr>
<tr>
<td>(+/-)</td>
<td>7</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>(+/-)</td>
<td>3</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>(+/-)</td>
<td>7</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>(+/-)</td>
<td>5</td>
<td>3</td>
<td>92</td>
</tr>
<tr>
<td>(+/-)</td>
<td>1</td>
<td>18</td>
<td>79</td>
</tr>
<tr>
<td>(+/-)</td>
<td>3</td>
<td>4</td>
<td>93</td>
</tr>
<tr>
<td>(+/-)</td>
<td>5</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>(+/-)</td>
<td>3</td>
<td>94</td>
<td>1</td>
</tr>
<tr>
<td>(+/-)</td>
<td>5</td>
<td>3</td>
<td>92</td>
</tr>
<tr>
<td>(+/-)</td>
<td>3</td>
<td>4</td>
<td>93</td>
</tr>
<tr>
<td>(+/-)</td>
<td>5</td>
<td>3</td>
<td>92</td>
</tr>
<tr>
<td>(+/-)</td>
<td>3</td>
<td>4</td>
<td>93</td>
</tr>
<tr>
<td>(+/-)</td>
<td>5</td>
<td>3</td>
<td>92</td>
</tr>
<tr>
<td>(+/-)</td>
<td>3</td>
<td>4</td>
<td>93</td>
</tr>
<tr>
<td>(+/-)</td>
<td>5</td>
<td>3</td>
<td>92</td>
</tr>
</tbody>
</table>
Table I. Correction of hemimethylated heteroduplexes. The percentage of Red (%R), White (%W) and Mixed (%M) colonies was calculated from the total population of transformants, which ranged from 400 to greater than 1,300 total colonies/ transformation. The state of adenine methylation is indicated by (+/-); methylation of the deletion containing strand annealed to unmethylated circular wild type single strand of pPY97. (-/+ ) indicates methylated wild type pPY97 strand annealed to unmethylated circular deletion containing mutant single strand.
base heterology is marginally corrected since 39% of the transformants are red colonies and only 4% are white colonies, whereas five base heterologies are refractory to repair since 90% of the transformants give rise to mixed colonies. This repair process is dependent on the gene products of mutH, L and S as the majority of transformants of these strains give rise to mixed colonies. Interestingly, our system detected a low level of methyl-independent correction of 1, 2 and 3 base heterologies and the G-T mispairs after transformation of the mutH (GM4349) indicator strain. This will be addressed in more detail below.

To demonstrate the dependence of this system on methylated d(G-A-T-C) sequences to direct the strandedness of repair, we switched the methylation state so that the wild type strand is methylated rather than the mutant strand. If switching the methylation state reverses the direction of repair, then transformations with these heteroduplexes should mirror the results of the opposite hemimethylated state described above (Table I). In this case, 90% to 95% of the wild type transformations with hemimethylated heteroduplexes containing 1, 2 and 3 base heterologies gave rise to white colonies indicative of repair using the methylated strand as template. The results also show that the 4 base deletion is poorly corrected while the 5 base deletion is not corrected.

We isolated plasmid DNA from 20 red colonies from the transformations in Table I and sequenced the mnt gene to confirm that the red colonies were the result of directed repair of the deletion mutation present in the heteroduplex. In every case, plasmids isolated from red colonies that arose as a result of transformation with heteroduplexes containing 1-5 base insertion/deletion heterologies contained the appropriate deletion mutation.
In vivo correction of unmethylated heteroduplexes proceeds without strand bias.

The Dam-directed repair pathway, while maintaining the ability to correct DNA, loses its ability to discriminate the wild type strand from the deletion containing strand when DNA is devoid of adenine methylation in d(G-A-T-C) sequences (9). Table II shows the results of transformations with unmethylated heteroduplexes as described in Materials and Methods. The G-T and T-G mismatches, 1, 2 and 3 base deletion heteroduplexes show repair consistent with the loss of strand bias since transformation of the mismatch proficient strain gave rise to nearly equivalent numbers of either red or white colonies and a very low percentage of mixed colonies. The data obtained for the 4 base deletion heteroduplex suggests that it is weak substrate for repair while the 5 base deletion heteroduplex is not repaired. This is in agreement with the results of the hemimethylated heteroduplex transformations from table I.

In vivo correction of symmetrically methylated heteroduplexes proceeds independently of Dam-methylation and MutH. Table III shows the data from transformations of the indicator strains with heteroduplexes containing symmetrically methylated d(G-A-T-C) sequences. Since these methylated sequences are resistant to nicking by MutH endonuclease, correction of mismatched bases and insertion/deletion heterologies should be greatly reduced (9, 10). The results in Table III do not support this prediction. Transformation of the mismatch proficient indicator strain with the symmetrically methylated heteroduplexes revealed a pattern of methylation independent repair similar to that observed with mutH transformations in Tables I, II and III. This repair process requires the gene products of mutL and S since greater than 80% of the transformants of these
Table II. Genetic analysis of unmethylated heteroduplex transformations.

<table>
<thead>
<tr>
<th>Heteroduplex (Mutant/WT)</th>
<th>Wild Type</th>
<th>mutL::Km Stu</th>
<th>mutH471::Tn5</th>
<th>mutS201::Tn5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%R %W %M</td>
<td>%R %W %M</td>
<td>%R %W %M</td>
<td>%R %W %M</td>
</tr>
<tr>
<td>(-/-) G-T₄₀</td>
<td>42 53 5</td>
<td>7 6 87</td>
<td>23 21 56</td>
<td>7 7 86</td>
</tr>
<tr>
<td>(-/-) T-G₄₁</td>
<td>39 47 14</td>
<td>8 7 85</td>
<td>27 21 52</td>
<td>6 9 85</td>
</tr>
<tr>
<td>(-/-) △ 1</td>
<td>47 49 4</td>
<td>5 7 88</td>
<td>26 27 47</td>
<td>5 4 91</td>
</tr>
<tr>
<td>(-/-) △ 2</td>
<td>49 47 4</td>
<td>3 6 91</td>
<td>23 19 58</td>
<td>7 6 87</td>
</tr>
<tr>
<td>(-/-) △ 3</td>
<td>47 42 11</td>
<td>5 8 87</td>
<td>12 13 76</td>
<td>4 5 91</td>
</tr>
<tr>
<td>(-/-) △ 4</td>
<td>18 31 51</td>
<td>7 9 84</td>
<td>9 11 80</td>
<td>9 10 81</td>
</tr>
<tr>
<td>(-/-) △ 5</td>
<td>6 10 84</td>
<td>7 8 85</td>
<td>6 7 87</td>
<td>5 8 87</td>
</tr>
</tbody>
</table>
Table II. Correction of unmethylated heteroduplexes. The percentage of Red (\%R), White (\%W) and Mixed (\%M) colonies was calculated from the total population of transformants, which ranged from 450 to greater than 1,000 total colonies/transformation. (-/-) indicates that both the deletion containing strand and the complementary wild type strand are devoid of adenine methylation.
Table III. Genetic analysis of symmetrically methylated heteroduplex transformations.

<table>
<thead>
<tr>
<th>Heteroduplex (Mutant/WT)</th>
<th>Wild Type</th>
<th>mutL::Km Stu</th>
<th>mutH471::Tn5</th>
<th>mutS201::Tn5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%R %W %M</td>
<td>%R %W %M</td>
<td>%R %W %M</td>
<td>%R %W %M</td>
</tr>
<tr>
<td>(+/+) G-T(_{40})</td>
<td>35 27 38</td>
<td>9 8 83</td>
<td>30 31 39</td>
<td>9 5 86</td>
</tr>
<tr>
<td>(+/+) T-G(_{41})</td>
<td>31 25 44</td>
<td>6 5 89</td>
<td>24 30 46</td>
<td>7 13 80</td>
</tr>
<tr>
<td>(+/+) Δ 1</td>
<td>31 30 39</td>
<td>6 10 84</td>
<td>26 29 45</td>
<td>8 6 86</td>
</tr>
<tr>
<td>(+/+) Δ 2</td>
<td>29 23 48</td>
<td>5 3 92</td>
<td>25 19 56</td>
<td>7 5 88</td>
</tr>
<tr>
<td>(+/+) Δ 3</td>
<td>28 19 53</td>
<td>6 5 89</td>
<td>19 25 56</td>
<td>6 7 87</td>
</tr>
<tr>
<td>(+/+) Δ 4</td>
<td>8 12 80</td>
<td>8 6 86</td>
<td>8 9 83</td>
<td>7 7 86</td>
</tr>
<tr>
<td>(+/+) Δ 5</td>
<td>6 8 86</td>
<td>4 5 88</td>
<td>7 6 87</td>
<td>7 10 83</td>
</tr>
</tbody>
</table>
**Table III.** Correction of symmetrically methylated heteroduplexes. The percentage of Red (%R), White (%W) and Mixed (%M) colonies was calculated from the total population of transformants, which ranged from 300 to 800 total colonies/ transformation. (+/+) indicates that both the deletion containing strand and the complementary wild type strand are fully methylated at greater than 95% of the d(G-A-T-C) sequences.
strains give rise to mixed colonies. Lack of correction of 4 and 5 base insertion/deletion heteroduplexes in wild type and mutH strains are also consistent with the data observed in Tables I and II. Therefore, Dam-independent repair follows the same specificity as the Dam-dependent system, albeit at reduced efficiencies. There exists the possibility that some d(G-A-T-C) sites in the heteroduplexes may lack symmetric modification after in vitro Dam methylation; however it is highly unlikely that this effect could account for all of the repair observed in the wild type strain since the results are within 5 to 7 percent of those observed in the mutH transformations.

The mnt gene in plasmid DNA isolated from 16 red colonies of mutH transformations was sequenced to determine if mutations in addition to, or in place of the original mutation (Fig. 1), were generated due to this methyl-independent repair pathway. In every case, the only mutations found in the mnt gene were those involved in the original heteroduplex construction (Fig. 1).

**Dam-directed repair of small heterologies in the mnt operator (O_{mnt}).** Using the identical plasmid system described above, but containing deletion mutations within the mnt operator (O_{mnt}) rather than in the mnt gene, we evaluated the correction of naturally occurring 1, 2, 4 and 5 base insertion/deletion heterologies by Dam-directed repair. Hemimethylated heteroduplexes consisting of unmethylated circular single strand containing deletion mutations within the O_{mnt} and the methylated complementary wild type pPY97 strand were transformed into repair proficient and isogenic mutL and mutH strains. The strains did not contain \( \lambda \) 110, thus the efficiency of correction was not based
on colony color. Rather, this method evaluated repair based on the percentage of ampicillin resistant transformants that were also tetracycline resistant. The results of these transformations indicated that 1 (unpaired A at nucleotide -43 (13)) and 2 (unpaired AC at nucleotides -44,-43) base heterologies were efficiently corrected in mismatch proficient cells since 97% and 93% of the transformants were tetracycline sensitive respectively (data not shown). Heteroduplexes containing 4 (unpaired TGGA at nucleotides -47 to -50) and 5 (unpaired GGTCC at nucleotides -42 to -38) base heterologies were refractory to repair. Repair of 1 and 2 base heterologies was Dam-directed and dependent on the gene products of mutL and H (mutS was not tested in this system).

MutS binds to small insertion/deletion heterologies in vitro. E. coli MutS is a 97 Kd protein that recognizes and binds to DNA containing mismatched bases (7), therefore performing an essential role in Dam-directed mismatch correction. Since MutS binds to mismatched bases and is required for correction of small insertion/deletion heterologies in vivo (Tables I-III), we investigated the possibility that MutS might also play a part in the recognition of insertion/deletion heterologies. Figure 2 displays the results of band shift assays involving 36 base heteroduplexes containing 1 (unpaired T), 2 (unpaired GC), 3 (unpaired TGC), 4 (unpaired TGCG) or 5 (unpaired AGTGC) base insertion/deletion heterologies or the G-T mismatch and Salmonella typhimurium MutS protein. S. typhimurium MutS, a 96 Kd protein, has been shown to complement E. coli MutS in vivo (24; Parker and Marinus, unpublished data). The band shift assays of Figure 2 show that MutS recognizes and binds to 1, 2, 3 and 4 base heterologies as well as to the G-T mismatch, while binding to the 5 base deletion and homoduplex (data not shown) was not
A. G-T mismatch

B. 1 base del.

C. 2 base del.

D. 3 base del.

E. 4 base del.

F. 5 base del.

MutS (pmol)
Figure 2. MutS band shift assays. Increasing concentrations of *Salmonella typhimurium* MutS were added to 0.6 pmols $[^{32}\text{P}]$-labelled 36 base duplexes containing either 1, 2, 3, 4 or 5 base heterologies or a G-T$_{40}$ mismatch and the binding reaction was placed on ice for 30 minutes. The binding reaction was analyzed on a 6% non-denaturing polyacrylamide gel followed by autoradiography. B (bound) represents MutS protein bound to heteroduplex. F (free) represents unbound 36 base duplex.
observed within the constraints of this assay. Densitometric scans of the autoradiographs reveal that the hierarchy for MutS binding to the above mutations followed a surprising pattern; 1 base del > 2 base del > 3 base del > G/T mismatch > 4 base deletions. The binding of MutS to 36 base heteroduplexes is a reversible process as MutS-heteroduplex complexes can be dissociated by SDS (not shown). Furthermore, MutS does not bind to 36 base single strand DNA within the parameters of our experimental system.

**DISCUSSION**

One of the first studies to provide evidence supporting the correction of small heterologies by mismatch repair involved the reversion of frameshift mutations in *lacZ* and *trpA* in *E. coli* mutU4, mutS3 or mutL25 strains (26, 27). These mutator alleles increased significantly the frequency of reversion above those levels observed in wild type strains. Similarly, the frequency of occurrence of 9-aminoacridine induced -1 and +1 frameshift mutations is greatly increased in *E. coli* mismatch deficient *mutL* host cells as compared to mismatch proficient wild type cells (28). The frequency of occurrence of small frameshift mutations in strains deficient in mismatch correction relative to wild type strains is also increased substantially (29). Moreover, Dohet et al. (30) demonstrated that 1 base insertions and deletions are efficiently repaired by mismatch correction using artificially constructed heteroduplexes and an *in vivo* transfection system. Lear and Grafstrom (31), using an *in vitro* system of cell free extracts, have demonstrated that 1, 2 and 3 base deletions can be repaired as efficiently as G-T mismatches in a process directed by the state of adenine methylation of d(G-A-T-C) sequences and requiring the gene products of *mutH, L* and *S*. Mismatch repair systems in other cell types have also been implicated in
the repair of small insertion/deletion mutations. The Hex system of *Streptococcus pneumoniae* efficiently corrects 1 and 2 base heterologies, marginally corrects 3 base heterologies and is refractory to repair of 4 and 5 base heterologies (32), while mismatch repair in *Saccharomyces cerevisiae* corrects single base frameshift mutations (33, 34).

In this study, we have shown that 1, 2 and 3 base insertion/deletion heterologies are repaired *in vivo* as efficiently as G-T and T-G mismatches. This is in agreement with earlier data obtained *in vitro* using cell free extracts (31). In addition, we show that 4 base deletions are repaired at reduced efficiencies while 5 base deletions lack efficient repair. Interestingly, the 4 base insertion/deletion heterologies within *mnt* were subject to moderate correction while a similar heterology located within the O_{mnt} (*mnt* operator) region was refractory to repair. The DNA sequences flanking these heterologies are different and since mismatched bases are corrected with efficiencies that depend in part on their sequence environment, it appears quite probable that the efficiency of correction of insertion/deletion heterologies may be influenced in a similar way.

An earlier report presented evidence for Dam-directed correction of heteroduplexes containing two 10 base insertion/deletion heterologies (35). We have made a number of observations that are in direct contradiction to this. Furthermore, although we can not completely rule out the possibility for correction of 5 base heterologies based on sequence context, we believe that the *E. coli* Dam-directed repair system cannot repair insertion/deletion heterologies larger than 4 bases since the correction of 4 base heterologies is marginal at best and MutS fails to recognize and specifically bind to 5 base heterologies, an essential step in the repair process. In support of this are the following observations. First, we have found that MutS fails to bind specifically to 7 and 9 base
heterologies in vitro (Parker and Marinus, unpublished observations) making it highly unlikely that MutS can recognize and bind to 10 base heterologies unless a smaller, recognizable mismatch is present. Secondly, in vivo correction of heteroduplexes containing 7, 9, and 11 base heterologies within the mnt gene is not observed (Carraway and Marinus, unpublished observations). In addition, deletion heterologies as large as 700 base pairs also escape Dam-dependent mismatch correction (36). Therefore, the most likely explanation for this discrepancy is that the 10 base heterologies, present as inverted repeats in the form of XhoI linkers, induced the formation of secondary structure containing either a mismatched base pair or a 1-4 base insertion/deletion heterology. In other words, the ten base deletion heterologies were most likely repaired as a consequence of MutS recognition of a suitable ten base deletion induced mismatch rather than direct correction of the ten base deletion itself.

A surprising finding in our work was the occurrence of methylation independent, but mutL and mutS dependent correction of 1, 2 and 3 base insertion/deletion heterologies and the G-T mismatches. Correction was independent of mutH function for heteroduplexes in all four methylation states as well as the symmetrically methylated wild type transformations. The latter was unexpected since Dam-directed repair was not observed in fully methylated lambda phage heteroduplexes in vivo (10) or in phage f1 heteroduplexes in vitro (6, 11, 21). The possibility for a Dam-independent repair pathway is interesting since a similar phenomenon has been observed in which symmetrically methylated heteroduplexes containing two small deletion mutations were also repaired in a Dam-independent manner except that it did not require the mutL gene product (35). Furthermore, transfection experiments using heteroduplexes constructed from bacteriophage
lambda DNA containing mutant cI and Pam alleles showed Dam-independent correction that required the gene products of mutL and S, but not mutH or uvrD (36). In this case, the VSP (very short patch) system of repair has since been identified as the repair pathway responsible for correcting the mutant alleles (2) since both mutations are G-T mismatches and each is located within a VSP recognition sequence, CTAG (37). This cannot be the explanation for our results since we observed Dam-independent repair in a dcm-6 indicator strain which is deficient in VSP repair (Parker and Marinus, unpublished data). In addition, the sequence context within which the insertion/deletion heterologies as well as the G-T mismatches lie do not conform with the sequence requirements for VSP repair.

The methyl-independent repair is probably best explained by the presence of strand breaks or nicks that might have been introduced into a closed circular heteroduplex either before, during or following the transformation process. Single strand breaks are sufficient to direct the strandedness of repair for mismatch correction in the absence of either MutH or d(G-A-T-C) sequences (11, 38). Assuming that heteroduplexes can be nicked on either strand with an approximately equal probability and that DNA ligase is unable to convert all nicked heteroduplexes back to a covalently closed state after transformation, correction of insertion/deletion heterologies by mismatch repair, in the absence of MutH or presence of symmetrically methylated d(G-A-T-C) sequences, should be directed according to the nick containing strand. This would be consistent with our observations and we are currently investigating this possibility. An alternative, albeit less likely, explanation might be that components of other pathways involved in localized repair are assisting MutL and S in correcting small heterologies. Since various AP endonucleases
are involved in localized repair of apurinic and apyrimidinic sites which may resemble an unpaired base (39), we constructed an indicator strain of the triple mutant \textit{xthA, rpo, nth} to determine the role of AP-endonuclease III, IV, VI, and exonuclease III in Dam-independent repair. The absence of these gene products did not block Dam-independent repair (data not shown). In addition, mutations in \textit{recF} and \textit{recA} were without effect on Dam-independent repair (data not shown). Work is currently underway to look at the effect of \textit{uvrA} on Dam-independent repair since it recognizes bulky lesions in DNA (39).

Our results show that MutS recognizes and binds to 1, 2, 3 and 4 base, but lacks specific binding to 5 base insertion/deletion heterologies \textit{in vitro}. The affinity of MutS for these mismatches correlates roughly to their efficiency of repair \textit{in vivo}. MutS protein recognizes and binds to all eight base:base mismatches at affinities that roughly correlate to their repair efficiency (6). However, the finding that MutS has an affinity for a G-T mismatch that is higher than for an equally efficiently repaired A-C mismatch suggested that MutS binding alone does not determine the efficiency of correction (23). Interestingly, densitometric scans of the autoradiographs in Figure 2 reveal that MutS binds to 1 and 2 base deletions with higher affinity than to the G-T mismatch or 3 base deletion. Competition studies between 1, 2 and 3 base deletions and a G-T mismatch agree with these results (data not shown). This supports the notion that the efficiency of correction of any given mismatch is determined by multiple factors since 1, 2 and 3 base deletions are repaired as efficiently as G-T mismatches even though the binding affinity of MutS varies considerably. However, it should be recalled from Table I that close to 99% of the 1, 2 and 3 base heterologies and G-T mispairs were corrected \textit{in vivo}. This raises the possibility that the repair system may already have corrected as many mispairs as possible.
prior to plasmid replication, thus making it difficult to clearly define the relationship between MutS binding and correction efficiency. Examining the kinetics of this repair process should help clarify this issue.

A recent study into the structure of DNA containing 1, 2, 3 or 4 base deletions concluded that these molecules contain kinks or bends that increase in proportion to the size of the deletion, while G-G, C-T and C-A mismatches failed to produce any detectable helical distortion (40). Moreover, the sequence flanking a single base deletion markedly influences the extent to which the helix can be kinked (41). Since the G-T mismatch is intrahelical producing minimal distortion (1) and insertion/deletion mutations favor the formation of helical kinks, the binding of MutS to these structurally dissimilar mutations raises a number of questions. Does MutS protein have two separate, but closely related or overlapping DNA binding sites; one for mismatches and the other for deletions? Is the kinked molecule, such as that found with insertion/deletion mutations, a favorable substrate for MutS binding? Does MutS binding to mismatches bend the DNA into a more favorable conformation? How much influence does sequence context have in MutS binding to insertion/deletion heterologies? Work is currently underway to examine some of these possibilities.

ACKNOWLEDGMENTS

We thank Dr. Graham C. Walker for his generous gift of MutS protein and mutH::Tn5 allele, and Drs. Eli C. Siegel and Malcolm E. Winkler for strains used in this work. We
gratefully acknowledge Dr. Robert Lahue for his excellent technical advice, helpful discussions and critical reading of the manuscript. We also thank Dr. Michael Volkert for helpful discussions and critical reading of the manuscript. This work was supported by a grant from the National Institutes of Health (GM33233).

REFERENCES

SECTION I.

Site directed mutagenesis: The oligonucleotides used to create the 1, 2, 3, 4 and 5 base deletion mutations in the mnt gene of pPY97 have been described in the "Materials and Methods" section and Figure 1 of this chapter. An outline of this procedure is shown in Figure S3.

Primer/Template annealing: 25.6 pmols of each of the oligonucleotides were treated with 10 units T4 Polynucleotide Kinase (70 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 5 mM DTT, 1 mM ATP) for 30 minutes at 37°C followed by 10 minutes at 70°C. Single stranded pPY97 was mixed with 5' phosphorylated oligonucleotide at a 1:20 ratio respectively in 10 mM Tris-HCl buffer, pH 8.0 and 1 mM EDTA. The annealing mixture was heated to 70°C for 5 minutes and slow cooled to 25°C over a 30 minute period. The primed template was placed on ice until ready for use.

Extension of primed template: This procedure has been described by Bebenek and Kunkel (1989) using native T7 DNA polymerase. The primed template was elongated in vitro with native T7 DNA polymerase under the following conditions; 10 mM Tris-HCl (pH 7.5), 0.5 mM of each of the 4 dNTP’s, 2 mM DTT, 7 mM MgCl₂, 1 mM ATP, 5 Units T4 DNA ligase, 4 Units T7 DNA polymerase. The mixture was incubated at 37°C for 30 minutes and then placed on ice. Samples were analyzed on a 1% agarose mini-gel to determine the completeness of the reaction.
Figure S3. Scheme for site directed mutagenesis for the construction of 1, 2, 3, 4 and 5 base deletions within the mnt structural gene. As described in the text, the oligonucleotide containing the appropriate deletion was annealed to the single stranded pPY97 template and extended using Native T7 DNA polymerase. The extended products were transformed into GM4331, spread onto MacConkey/lactose agar, and red/Tet$^R$ colonies were isolated and tested for the appropriate deletion mutation. The 500 base pair EcoRI (E)/HindIII (H) restriction fragment was isolated and cloned back into the pPY97 vector (4.8 kb EcoRI/HindIII restriction fragment). A= position of the bla gene conferring ampicillin resistance and T= position of the tet gene.
anneal mutant oligo

A mnt T

sspPY97 template

Extend: Native T7 DNA polymerase
dNTP’s, ATP, T4 DNA ligase, MgCl₂
DTT, Tris buffer pH 7.5
1 hour at 37°

Transform GM4331 plate on MacConkey/Lactose/Tet
isolate Red/Tet colonies

culture and miniprep DNA
Retransform and isolate Red/Tet colonies
Miniprep DNA and check for:
1. Amp resistance
2. EcoRI/HindIII digestion
3. Sequence

Isolate and Purify 500 bp
EcoRI/HindIII Fragment

Clone 500 bp frag into pPY97 vector

mnt mutation

Tranform and Confirm
1. Isolate Red/Tet/Amp colonies
2. Screen for EcoRI/HindIII sensitivity

E H

mnt mutant
Screening procedure: The extended products are essentially unmethylated heteroduplexes since the template is unmethylated and the extension reaction was carried out in the absence of Dam methylase. Transformation of CaCl₂ competent GM4331 (wild type, mismatch proficient) with the extended products will give rise to red, white and red/white ("mixed") colonies when plated onto MacConkey/lactose/ampicillin plates. Red colonies from each transformation were tested for resistance to tetracycline since the deletion mutants should be red and Tet⁺. The plasmid DNA of three red/Tet⁺ colonies from each group was isolated on "Qiagen mini columns" according to the instructions of the manufacturer. The mnt gene was sequenced (dideoxy method) for each miniprep of each group to identify the colonies that harbor the appropriate mutations (1, 2, 3, 4 or 5 base deletions).

Cloning the deletion containing mnt gene into the EcoRI/HindIII pPY97 vector: The deletion plasmids, as determined from sequence analysis, were digested with EcoRI and HindIII. 500 base pair EcoRI/HindIII restriction fragments containing the appropriate deletion mutation in mnt was isolated from a 0.8% low melting point agarose gel (Sea Plaque) and subsequently purified using PrePac (NACS) columns according to manufacturers instructions. The 4.8 kb EcoRI/HindIII restriction fragment of wild type pPY97 was isolated and purified similarly. Approximately equimolar concentrations of the 500 base pair and 4.8 kb vector restriction fragments were mixed in a DNA ligase buffer, 1 mM ATP, 40 Units T4 DNA ligase and incubated 18 hours at 14°C. The clones were transformed into GM856 (recF143) and incubated on MacConkey/lactose plates containing ampicillin and tetracycline. Three sets of plasmid preps for each group were isolated.
from red/Amp\(^R\)/Tet\(^R\) colonies using PrePac (NACS) columns. This cloning step significantly reduced the possibility of a second site mutation in the plasmid vector due to misincorporation of a dNTP during the extension reaction.

**Restriction enzyme analysis:** To determine which plasmids in each set of deletion mutants contain only a single vector and 500 base pair \(mnt\) fragment, each set of plasmid isolates were digested with the restriction endonuclease enzymes \(EcoRI\), \(HindIII\) and \(EcoRI/HindIII\). pPY97 served as a control digest. Analysis of these restriction digests were performed on 1% agarose minigels.

**SECTION II.**

**Isolation of single stranded DNA.** This has been described in Chapter 3 with the following modification. After isolation of the single strand DNA, it was further purified using hydroxylapatite chromatography (BioRad). This step was necessary to remove any contaminating duplex DNA that might have co-purified with the ssDNA. The ssDNA preparations were loaded onto a freshly prepared hydroxylapatite columns buffered with 20 mM KPO\(_4\) (pH 6.9) and washed with 3 volumes of 20 mM KPO\(_4\) (pH 6.9). The ssDNA was eluted with 3 volumes 100 mM KPO\(_4\) (pH 6.9) collected in half volume fractions, (dsDNA eluted at 200 mM KPO\(_4\) (pH 6.9)). An aliquot from each fraction was checked for purity on a 1% agarose minigel. The appropriate fractions were pooled, the volume was concentrated four fold with n-butanol washings, and dialyzed over night.
against 2 liters of TE buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA) at 4°C (TE buffer was changed twice during dialysis). The final yield was again checked for purity on 1% agarose gels and quantitated using UV spectroscopy at $A_{260}$.

SECTION III.

Preparation of Heteroduplexes. The methodology is discussed in the text and outlined in Figure S4 with the following clarifications.

1. The annealing reaction required a mixture of 100 μg ssDNA, 500 μg linear duplex DNA (1:5 molar ratio), NaCl to 10 mM, and Tris-HCl pH 7.6 to 50 mM in a final volume of 2 ml. Freshly prepared 10 N NaOH was added to 0.3 N and the mixture incubated at room temperature for 5 minutes. The mixture was neutralized with HOAc to 0.225 N, KCl to 100 mM and KPO$_4$ pH 7.4 to 90 mM. The mix was placed in ice following a 30 minute incubation at 65°C subsequent to 30 minutes at 37°C. The presence of nicked heteroduplexes were checked on a 1% agarose minigel.

2. The heteroduplexes were separated from ssDNA with the aid of hydroxylapatite (HA) chromatography (BioRad). The heteroduplex mixture was loaded onto a HA column equilibrated with 0.02 mM KP$_4$, pH 6.9 and washed with 3 volumes of the equilibration buffer. Single-strand DNA was eluted with excess washings of 100 mM KP$_4$, pH 6.9 followed by an additional washing of 1 volume 140 mM KP$_4$, pH 6.9. The heteroduplexes were eluted with 3 volumes of 200 mM KP$_4$, pH 6.9 and collected in half volume fractions. The fractions were analyzed on 1% agarose gels for heteroduplex purity, and the appropriate heteroduplex fractions were pooled. EDTA was added to
Heteroduplex Construction

WT

or

mnt

Mutant

Pvull digested

Anneal strands

Alkaline Denaturation
Neutralize, 65 °C 30 min
37 °C 30 min, ice

Hemimethylated heteroduplex
(ndDNA)

Test ME State
(Sau3a, DpnI, MboI)

Extract ccDNA and wash with butanol.
Dialyze against TE

Fully Methylated

Hemimethylated
Unmethylated

in vitro
Dam Methylase
SAM

Hydroxyapatite Chromatography

To remove ssDNA from dsDNA (ndDNA)

Seal Nicked Heteroduplex
E. coli Ligase, NAD⁺
Figure S-4. Scheme for the construction of heteroduplexes containing 1, 2, 3, 4 and 5 base heterologies within the mnt structural gene. As described in the text, the heteroduplex construction involves annealing the unmethylated single strand DNA to its complementary linearized strand to produce nicked heteroduplexes (ndDNA). The ndDNA was separated from single strand DNA by hydroxylapatite chromatography and subsequently converted to the covalently closed state with E. coli ligase. The covalently closed heteroduplexes (ccDNA) were separated from the ndDNA on a cesium chloride/ethidium bromide (CsCl/EtBr) equilibrium gradient. The methylation state for each set of ccDNA was analysed by restriction analysis. Fully methylated heteroduplexes were produced in vitro by treating the heteroduplex with Dam methylase.
10 mM and the solution volume subsequently reduced with n-butanol washings. The heteroduplexes were dialyzed (2X) against 2 liters of TE buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA) at 4°C. The heteroduplex concentration was determined with the aid of UV spectroscopy (A$_{260}$).

3. The nicked heteroduplexes were sealed with 4 units of E. coli DNA ligase in 26 μM β-NAD$, 30 mM$ Tris-Cl pH 7.5, 4 mM MgCl$_2$, 1 mM DTT, 50 μg/ml BSA, 10 mM ammonium sulfate, and 29 nmol of ethidium bromide per 100 μg of heteroduplex. The tubes were wrapped in foil after the addition of DNA ligase and placed at room temperature for 2 hours. The tubes were transferred to ice, an aliquot was tested for ligation on a 1% agarose minigel, and after confirmation of ligation, EDTA was added to 10 mM.

4. The covalently closed heteroduplexes were separated from the open circular species on a CsCl/ethidium bromide equilibrium gradient using 1.02 grams of CsCl per ml of TE buffer. 50 μl of 10 mg/ml ethidium bromide was added for each ml of solution and loaded into 4 ml centrifuge tubes (Beckman). The tubes were placed into a Beckman VTi 65 rotor and centrifuged over night at 45,000 rpms. The lower of the two bands was carefully removed using a 21-gauge needle and syringe. The ethidium bromide was removed with n-butanol washings. The heteroduplexes were dialyzed against 2 liters of TE and checked for purity on 1% agarose minigels. The covalently closed heteroduplexes were analyzed to confirm methylation state and subsequently stored at -20°C.
CHAPTER 5

MutS Protein of *Escherichia coli* Binds to DNA Containing Small Heterologous Sequences.
ABSTRACT

We have previously shown, using an in vivo transformation system, that the Escherichia coli Dam-dependent mismatch repair system can recognize and correct heteroduplexes containing 1, 2 and 3 base heterologies with efficiencies similar to that observed with G/T mismatches. The correction of a 4 base heterology was marginal and no detectable correction was observed with a 5 base heterology. Efficient correction of these heterologies required the gene products of mutH, mutL and mutS. In addition, Salmonella typhimurium MutS protein recognized and bound to 1, 2, 3 and 4 base heterologies and a G/T mispair as determined by band shift assays. A surprising result that arose from these assays was that MutS bound to 1 and 2 base heterologies with a higher affinity than it bound to the G/T mismatch. The present study utilized purified E. coli MutS protein to examine the relationship between MutS binding to small heterologies and the G/T mispair. The results of these experiments confirm that E. coli MutS binds to 1, 2 and 3 base heterologies with greater affinity than it binds to a G/T mismatch. The relative affinities were 1 base heterology > 2 base heterology > 3 base heterology > G/T mispair > 4 base heterology. DNase I protection experiments using a 215 base pair restriction fragment showed that MutS protects approximately 23 to 25 bases of the 1, 2 or 3 base heterologies while protecting approximately 22 base pairs surrounding the G/T40 mispair. Binding of MutS to the substrate molecules was asymmetric, an observation that is in agreement with other studies.
INTRODUCTION

The Dam-dependent mismatch repair system of *Escherichia coli* corrects DNA containing base:base mispairs and small heterologies at efficiencies that depend on both the nature of the mismatch and on the DNA sequence within which the mismatch lies (Radman and Wagner, 1986; Modrich, 1987; Meselson, 1988; Modrich, 1991). Repair of a mismatch by this system is directed according to the state on Dam (DNA adenine methyltransferase) methylation of 5'-d(GATC)-3' sequences: repair is directed against the unmethylated strand of a hemimethylated d(GATC) sequence, either strand of an unmethylated d(GATC) sequence while in contrast symmetrically methylated d(GATC) sites are refractory to repair (Pukkila et al., 1983; Lahue et al., 1989). Dam-dependent mismatch repair can utilize d(GATC) sites in excess of 2,000 base pairs from the site of the mismatch giving rise to long excision repair tracts (Wagner and Meselson, 1976; Su et al., 1989). This mismatch repair system has been reconstituted *in vitro* as a purified system requiring the gene products of *mutH*, *mutL* and *mutS*, helicase II, SSB, exonuclease I, exonuclease VII or RecJ, DNA polymerase III holoenzyme and DNA ligase (Lahue et al., 1989; Modrich, 1991).

Although the biochemistry of mismatch repair is not completely clear, the process can be divided into three steps: initiation, repair tract excision, and repair tract resynthesis. Initiation of repair begins with the recognition and binding of a mismatch by MutS protein. In the presence of MutS/mismatch and ATP, a DNA alpha loop structure forms that is stabilized at the DNA junction with bound protein, presumably MutS (Modrich, 1991). MutL protein associates with the MutS/ATP/DNA complex and increases the number of nucleotides protected against DNaseI digestion (Grilley et al., 1989). MutH, a
d(GATC) specific endonuclease, is activated in the presence of this complex which presumably begins searching for an unmethylated d(GATC) sequence (Modrich, 1991). MutH makes an incision on the unmethylated strand 5' to a hemimethylated d(GATC) site (Welsh et al., 1987). Repair tract excision involves DNA helicase II to unwind the duplex 3' to 5' (Matson, 1986) in a process that requires the presence of SSB (single-strand DNA binding protein). Repair tract excision is directed either 3' to 5' by exonuclease I or 5' to 3' by exonuclease VII or the RecJ exonuclease (Su et al., 1989; Modrich, 1991). DNA polymerase III holoenzyme resynthesizes the repair tract and DNA ligase restores the DNA to its covalently closed state.

The E. coli mutS gene product is a 97 kDa protein that recognizes and binds to all eight possible base:base mispairs in vitro with varying affinities (Su et al., 1988). We have recently demonstrated that Salmonella typhimurium MutS, which complements E. coli MutS in vivo, recognizes and binds to heteroduplexes containing 1, 2, 3 and 4 base heterologies in vitro (Parker and Marinus, in press). The results of that study suggested that MutS protein binds to 1 and 2 base heterologies with greater affinity than it binds to the G/T mismatch. In this study, we have confirmed and extended the previous observations to include competition studies and DNase I protection data.

MATERIALS AND METHODS
Materials: The E. coli MutS protein was a very generous gift from Dr. Paul Modrich (Duke Univ. Medical Center, N. Carolina; Su and Modrich, 1986). MutS was stored at -70°C as 3.1 pmols/µl in 0.1 M KPO₄ (pH 7.4), 0.1 mM EDTA, 0.2 M KCl, 1 mM phenylmethylsulfonyl fluoride and 10 mM 2-mercaptoethanol. Bovine pancreatic DNaseI
was purchased from USB. T4 polynucleotide kinase, *EcoRV, HindIII, AvaII* restriction endonucleases and the large fragment of DNA polymerase I (Klenow fragment) were purchased from New England Biolabs. \( \gamma^{32}\text{P}\)ATP (3,000 Ci/mmol), \( \alpha^{32}\text{P}\)dGTP (3,000 Ci/mmol), and \( \alpha^{32}\text{P}\)TTP (3,000 Ci/mmol) were purchased through Amersham.

**BAND SHIFT ASSAY I. Oligonucleotides.** The oligonucleotides used in this study were synthesized and HPLC purified by Operon Technologies (Alameda, CA). A 33 base oligonucleotide with the following sequence, 5'-'ATACGGAAAGTTAAAGTGGATCATCTCTAGCC-3', served as the wild type sequence (top strand) to which complementary sequences containing 1, 2, 3, 4 or 5 base deletions (Figure 1) were annealed to form heteroduplex substrates for band shift assays. The homoduplex control and G/T\(_{40}\) heteroduplexes were constructed from 36 base oligonucleotides in which the wild type sequence (top strand) of 5'-GCATACGGAAAGTTAAAGTGGATCATCTCTAGCCA-3' is annealed to the complementary sequence to form the 36 base homoduplex and to 5'-TGGCTAGAGATGATCCGTACTTAACTTCCGTATGC-3' to form the 36 base G/T\(_{40}\) heteroduplex.

**Kinasing of Oligonucleotides.** The top strand oligonucleotides were 5' end labelled using T4 polynucleotide kinase and \( \gamma^{32}\text{P}\)ATP (3,000 Ci/mmol). 10 Units of T4 polynucleotide kinase was added to 20 pmols of oligonucleotide in 70 mM Tris-Cl (pH 7.6),
Figure 1.

<table>
<thead>
<tr>
<th>Top Strand-33 mer</th>
<th>Bottom Strand containing the following deletions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-ATACGGAAGTTAA_{37}G_{38}T_{39}G_{40}C_{41}G_{42}GATCATCTCTAGCC-3'</td>
<td>-A_{39} 1 base Δ</td>
</tr>
<tr>
<td></td>
<td>-C_{40}G_{41} 2 base Δ</td>
</tr>
<tr>
<td></td>
<td>-A_{39}C_{40}G_{41} 3 base Δ</td>
</tr>
<tr>
<td></td>
<td>-A_{39}C_{40}G_{41}C_{42} 4 base Δ</td>
</tr>
<tr>
<td></td>
<td>-T_{37}C_{38}A_{39}C_{40}G_{41} 5 base Δ</td>
</tr>
</tbody>
</table>

Map of the deletions used in the mobility shift assays. The sequence maps show the location of the 1, 2, 3, 4 and 5 base deletions and the position of the G/T mismatch. Each of the deletion oligonucleotides were annealed to the complementary 33-mer to create the heteroduplexes used in the mobility shift assays.
10 mM MgCl₂, 5 mM DTT and 20 pmols [γ³₂P]ATP (60 μCi) and incubated for 30 minutes at 37°C. The reaction was terminated by incubating the mix at 70 °C for 10 min. The radioactively labelled oligonucleotides were stored at -20°C and were used within 4 days.

**Annealing Reaction.** 2 pmols of the ³²P-top strand was mixed with 2.5 pmols of the complementary oligonucleotide in 20 mM Tris-HCl (pH 7.6), 0.01 mM EDTA, 5 mM MgCl₂, 0.1 mM DTT. Annealing was carried out by heating the mixture to 70°C for 10 minutes, slow cooled to 27°C over a 30 minute period and placed on ice for 10 minutes prior to use.

**MutS Binding Reaction.** 2.5 pmols MutS protein was added to 40 fmols radioactively labelled heteroduplex (or homoduplex) in a final volume of 10 μl. The binding reaction was placed on ice for 30 minutes followed by the addition of 2 μl 50% w/v sucrose. 5 μl was loaded onto a 6% non-denaturing polyacrylamide gel and electrophoresed in 40 mM Tris-acetate (pH 7.5), 1 mM EDTA at 95 volts (7.5 volts/cm) until bromophenyl blue marker runs approximately 7 cm from the top of the gel, approximately 2 to 2.5 hours. The gel was dried at 80°C for 45 minutes followed by autoradiography on XOMAT film (Kodak).

**BAND SHIFT ASSAY II. Competition binding study.** Same as above except that 1.6 pmols of cold competing duplex (1, 2, 3 and 4 base heterologies, G/T mispair and
homoduplex) was added to 40 fmols radioactively labelled duplex followed by the addition of 2.5 pmols MutS for 1 base Δ heteroduplex, 3.0 pmols MutS for the 2 base heteroduplex and 3.5 pmols of MutS for the 3 base heteroduplex. The mixture was incubated on ice for 30 minutes prior to electrophoresis and autoradiography.

**DNase I protection assay:** The procedure used for DNaseI protection assays has been described by Su and Modrich (1988) with the following modifications. pPY97 heteroduplexes containing 1, 2 and 3 base heterologies and G/T mismatch are described in detail elsewhere (Parker and Marinus, In Press), and the DNA sequences are outlined in Figure 1. These heteroduplexes were digested with *EcoRV* and *HindIII* and loaded on a 4% low melting point gel (Nu Sieve, DNA Grade, FMC products, ME) in 90 mM Tris-borate (pH 8.3), 1 mM EDTA. A 378 base pair *EcoRV/HindIII* restriction fragment containing the mismatch was excised from the gel and purified with PrePac NACS columns according to the manufacturer’s protocol. This 378 base pair fragment was further digested with *AvaII* restriction endonuclease to produce a 215 base pair fragment with a 5’ overhang of 3’-CAG-5’. The Klenow fragment of DNA polymerase I was used to fill in the *AvaII* 3’ end with [α-32P]dGTP (3,000 Ci/mmole), [α-32P]TTP (3,000 Ci/mmole) and dideoxyCTP. The 215 base pair labelled restriction fragment was separated from the 153 and 10 base pair fragments as described above for the 378 base pair fragment. Approximately 50 fmols of 215 base pair labeled fragment was dissolved in 14 μl DNase I buffer (20 mM Tris-Cl, pH 7.5, 5 mM MgCl₂, 0.5 mM CaCl₂, 1 mM DTT, 0.1 mM EDTA). 1 μl *E. coli* MutS (3.1 pmols) was added and the binding reaction was carried out at room temperature for 20 minutes. 2 ng DNase I (2 ng/μl) was added to both MutS bound and unbound
(control) complexes. The reaction was quenched after 15 seconds by the addition of 48 μl of 200 proof ethanol, 3.75 μl 5 M ammonium acetate and 1 μl glycogen (20 mg/ml) (Boehringer Mannheim). The products were analyzed on 8% denaturing polyacrylamide sequencing gels. Maxam and Gilbert G-lane sequencing of the $^{32}$P-labeled 215 base restriction fragment was used as marker lanes for each DNase I protection assay. Briefly, 5 μl 10% dimethyl sulfate was added to 195 μl DMS buffer (50 mM sodium cacodylate, pH 7.0, 1 mM EDTA) containing 50 fmols $^{32}$P-labeled 215 base restriction fragment and incubated for 5 minutes at 20 °C. The reaction was quenched with 50 μl DMS stop solution (1.5 M sodium acetate pH 7.0, 1 M β-mercaptoethanol) and 750 μl 100% ethanol. The DNA was precipitated at -70 °C for 30 minutes and subsequently centrifuged. The DNA pellet was washed, lyophilized, resuspended in 100 μl of 1 M piperdine, incubated at 90 °C for 30 minutes, and lyophilyzed. Washed twice with dH$_2$O, lyophilized and resuspended in sequencing gel-loading buffer (98% deionized formamide, 10 mM EDTA, 0.025% xylene cyanol FF, 0.025% bromophenol blue).

RESULTS

MutS Protein Binds to 1, 2, 3 and 4 Base Heterologies. A mobility shift DNA binding assay was used to evaluate MutS binding to insertion/deletion heterologies (Figure 1.). E. coli MutS protein (Su et al., 1986) binds to 33 base heteroduplexes containing 1, 2, 3 and 4 base heterologies and to a 36 base heteroduplex containing a G/T mismatch control (Figure 2A). The binding of a heteroduplex containing a 5 base deletion and the 36 base homoduplex was not observed. Densitometric scans of the mobility shift assay
# MutS Binding Assay

<table>
<thead>
<tr>
<th>1del</th>
<th>2del</th>
<th>3del</th>
<th>4del</th>
<th>5del</th>
<th>G:T</th>
<th>homo</th>
</tr>
</thead>
</table>

- **B**
- **F**
Figure 2. A. Autoradiogram of the mobility shift assay using E. coli MutS protein. A representative result of a mobility shift assay involving 40 fmols of $^{32}$P-labelled 1, 2, 3, 4 and 5 base deletion 33-mer duplexes and 2.5 pmol MutS protein. The homoduplex and G/T mismatch duplexes are 36-mers. The binding reaction was carried out on ice in 10 µl volume as described in Materials and Methods. Lane 1, 1 base deletion duplex; lane 2, 2 base deletion duplex; lane 3, 3 base deletion duplex; lane 4, 4 base deletion duplex; lane 5, 5 base deletion duplex; lane 6, G/T mismatch duplex; and lane 7, homoduplex. F= free duplex, B= bound duplex. B. Results of densitometric scan of assay A (Figure 2A) and assay B. Assay A (above) and B represent duplicate experiments of mobility shift assays.
autoradiograms were used to determine the relative amounts of free and bound heteroduplex. The results are depicted in Figure 2B, and suggests that MutS binds to the 1 base heterology with a higher affinity than it binds to the remaining heteroduplexes. These results suggested a hierarchy of binding that places the affinity of MutS for 1, 2 and 3 base heterologies higher than that of the G/T control heteroduplex and the 4 base heterology.

MutS Competition Binding Assays. The hierarchy of MutS binding to insertion/deletion heterologies and the G/T mispair was investigated by measuring the binding of MutS protein to a radioactively labelled substrate in the presence of cold competitor DNA. When 250 nM MutS protein was incubated with 4 nM of the $^{32}$P-labelled 1 base heterology heteroduplex in the presence of 40 fold excess cold competitor DNA (1, 2, 3, 4 base heterologies, G/T mismatch and homoduplex), a clear hierarchy of binding was observed (Figure 3). This set of results indicated that MutS binds with high affinity to 1 base heterology > 2 base heterology > 3 base heterology > G/T mismatch. The 4 base heterology and homoduplex failed to compete with the 1 base heterology for MutS binding. The same set of experiments was performed using radioactively labelled 2 base heterology (Figure 4) and 3 base heterology (Figure 5). The results of these experiments are in general agreement with those of Figure 3, but also suggest that the 4 base heterology can compete against the 2 and 3 base heterologies for MutS binding. Interestingly, a 40-fold excess of homoduplex competed, albeit weakly, against the 2 and 3 base heterologies for MutS binding. This observation was not unexpected since a similar study also observed low non-specific binding to a 16-mer homoduplex (Jiricny et al., 1988).
Competition Binding Assays (Gel Shift)

1 base deletion heteroduplexes

% Bound by MutS

Competitor DNA

Assay A  Assay B

None  1 del  2 del  3 del  4 del  G/T  Homo
Figure 3. Mobility Shift Assay; Competition studies with 1 base deletion duplex. The binding of 2.5 pmol MutS to 40 fmol of radioactively labelled 1 base deletion duplex was evaluated in the absence (None) and in the presence of a 40-fold excess (1.6 pmol) of competitor DNA duplex as described in Materials and Methods. The binding reactions were evaluated by mobility shift assays with subsequent densitometric analysis, the results of which are shown above. Assays A and B represent duplicate experiments.
Competition Binding Assays (Gel Shift)

2 base deletion heteroduplex

% Bound by MutS

Competitor DNA

Assay A  Assay B

None  1 del  2 del  3 del  4 del  G/T  Homo
Figure 4. Mobility Shift Assay; Competition studies with 2 base deletion duplex. The binding of 3.0 pmol MutS to 40 fmol of radioactively labelled 2 base deletion duplex was evaluated in the absence (None) and in the presence of a 40-fold excess (1.6 pmol) of competitor DNA duplex as described in Materials and Methods. The binding reactions were evaluated by mobility shift assays with subsequent densitometric analysis, the results of which are shown above. Assays A and B represent duplicate experiments.
Competition Binding Assays (Gel Shift)

3 base deletion heteroduplex
Figure 5. Mobility Shift Assay; Competition studies with 3 base deletion duplex. The binding of 3.5 pmol MutS to 40 fmol of radioactively labelled 3 base deletion duplex was evaluated in the absence (None) and in the presence of a 40-fold excess (1.6 pmol) of competitor DNA duplex as described in Materials and Methods. The binding reactions were evaluated by mobility shift assays with subsequent densitometric analysis, the results of which are shown above. Assays A and B represent duplicate experiments.
**DNase I Protection Assay.** To identify the DNA sequence bound by MutS protein, DNA restriction fragments (215 base pairs) containing either a 1, 2, 3 base heterology or G/T mismatch were incubated with MutS and subsequently treated with DNaseI. The results are shown in Figure 6, and indicate that MutS specifically recognizes 1, 2 and 3 base heterologies since they are located within the region that is protected against DNaseI cleavage. For each heterology, only one area of DNaseI protection was observed within the 215 base pair restriction fragments. In the absence of MutS protein, the DNaseI cleavage patterns within the region normally protected by bound MutS protein vary slightly, while the cleavage patterns both 3' to and 5' to the protected regions are virtually indistinguishable from duplex to duplex. This is consistent with the observations of Lu et al., (1988), and presumably reflect the structural/helical variations that are induced as a result of the different heterologies. These variations in the vicinity of the mismatch/heterology make specific areas of the phosphodiester backbone more susceptible to attack by DNaseI.

Figure 7 shows the DNA sequences that were protected against DNaseI digestion. The heterology/mismatch was asymmetrically positioned within the DNaseI protected region suggesting that MutS binds asymmetrically to both the G/T mismatch and the 1, 2 and 3 base heterologies. MutS protein protected about 22 base pairs surrounding the G/T mismatch, and 23 to 25 base pairs surrounding the 1, 2 and 3 base heterologies.
MutS  Footprints - DNasel

Mismatch 1del 2del 3del G:T
MutS  - + - + - + - +
**Figure 6.** *E. coli* MutS DNA footprints: DNaseI protection assays.

Approximately 50 fmol of 3'-end labeled (*Avai* II site) 215 base pair restriction fragments containing either a 1, 2 or 3 base heterology or a G/T mismatch (see Figure 1) were suspended in 14 μl DNaseI buffer to which 1 μl (3.1 pmol) of MutS protein was added. The binding reaction was incubated for 20 minutes at room temperature followed by the addition of 2 ng DNaseI. The reaction was quenched after 15 seconds. The *minus* symbols indicate no MutS protein and the *plus* symbols indicate the presence of MutS protein. The position of the G/T mismatch and 1, 2 and 3 base heterologies within the restriction fragments are marked by the arrows and brackets. The DNA sequence runs 5' (top) to 3' (bottom).
Figure 7. DNA sequence of the MutS footprints in Figure 6. The arrows mark the position of the G/T mismatch and the 1, 2 and 3 base deletions. The DNA sequence in large bold print represents that which is protected against DNaseI cleavage by bound MutS protein.

G/T mismatch

\[ 5'{-\text{G}}\text{G-CAT} \quad \text{ACGGAAGTTAAAGTGC}-\text{CGGATCA}^{\text{TCTCTAGCCATGCCAT}-3'} \]

22 bases

1 base deletion

\[ 5'{-\text{G}}\text{GCATACG} \quad \text{GAAGTTAAAGTGC}-\text{CGGATCATCTCTA}^{\text{GCCATGCCAT}-3'} \]

unpaired T

2 base deletion

\[ 5'{-\text{G}}\text{GCATACG} \quad \text{GAAGTTAAAGTGC}-\text{CGGATCATCTCTA}^{\text{GCCATGCCAT}-3'} \]

unpaired GC

3 base deletion

\[ 5'{-\text{G}}\text{GCATACG} \quad \text{GAAGTTAAAGTGC}-\text{CGGATCATCTCTA}^{\text{GCCATGCCAT}-3'} \]

unpaired TGC

25 bases
DISCUSSION

The *E. coli* mutS gene product has been isolated and purified to near homogeneity as a 97 kDa protein (Su and Modrich, 1986). MutS recognizes and binds to all eight possible mismatched bases in DNA with an affinity that roughly correlates to their *in vitro* repair efficiency (Su et al., 1988). Using DNaseI protection assays, Su et al. (1988) calculated the apparent equilibrium dissociation constants for these mismatched base pairs based on the assumptions that the MutS protein bound as a monomer and was 100% active in mismatch recognition. The highest affinity of MutS was for the G/T and A/C mismatches at 39 nM and 53 nM respectively, and the lowest affinity was for the C/C mispair at about 480 nM (Su et al., 1988). Interestingly, MutS protein binds to the C/C mispair even though it is not corrected by mismatch repair (Su et al., 1988). Furthermore, MutS binds to the A/C mispair with high affinity (53 nM), but repairs it at moderate frequencies (Su et al., 1988). These observations serve to underscore the point that, while MutS recognition and binding is required for repair, it is not the sole determinant of mismatch repair efficiency. MutS does not bind to 5 base heterologies, which are not repaired *in vivo* (Parker and Marinus, in press).

Studies performed both *in vivo* (Parker and Marinus, in press) and *in vitro* (Learns and Grafstrom, 1989) have indicated that heteroduplexes containing 1, 2 and 3 base heterologies are repaired as efficiently as a G/T mismatch. Since previous binding studies have indicated that MutS binds with highest affinity to the G/T mismatch *in vitro* (Su and Modrich, 1986; Su et al., 1988; Jiricny et al., 1988), we expected to observe similar MutS binding affinities for 1, 2 and 3 base heterologies. Rather, our results indicate that MutS binds with higher affinity to 1, 2 and 3 base heterologies than to the G/T mismatch. This
is in agreement with the calculated apparent equilibrium dissociation constants for 1, 2
and 3 base heterologies and the G/T mispair as 10.4 nM, 21 nM, 165 nM and 202 nM
respectively. The $K_{D_{app}}$ was calculated according to the method of Su et al., (1988)
except that data from mobility shift assays were used rather than from the more accurate
DNaseI protection method. This method has a number of imperfections that can lead to
error in determining the $K_{D_{app}}$. First, the binding equilibrium is constantly changing dur-
ing electrophoresis, thus increasing the ratio of free to bound ligand. Second, the $K_{D_{app}}$ of
MutS for 1, 2 and 3 base heterologies and G/T mismatch were calculated using the data
in Figure 2B. Since MutS binding to 1 and 2 base heterologies was approaching the pla-
teau phase in excess of 90%, the $K_{D_{app}}$s are likely less accurate than if they had been cal-
culated from values closer to 50% binding such as was the case for the 3 base heterology
and the G/T mismatch. Finally, as described by Su et al., (1988) it was assumed, since
neither the aggregation state nor the biologically active form of MutS had been defined,
that MutS bound as a monomer and that it was 100% active in its ability to recognize and
bind to mismatched DNA. Although the limitations described above allow only a rough
estimate of the $K_{D_{app}}$ for MutS binding, these values do provide important information
about the relative affinities of MutS to the DNA heterologies.

The $K_{D_{app}}$ for the G/T in this study was about 5 fold greater than that calculated by
Su et al., (1988). The difference in the affinities between the G/T mispairs used in this
study (202 nM) and that of Su et al., (39 nM) may be due in part to: a) the limitations of
the mobility shift assays used in this study, and b) the DNA sequence context differences
between the G/T mispairs used in each study since they were positioned in different DNA
sequence environments. It is likely that our $K_{D_{app}}$ values, which were calculated using
data from mobility shift assays, are higher than if they had been calculated using the more accurate DNaseI protection assays of Su et al., (1988). If this is the case, then the \( K_{D_{app}} \) for MutS binding to the 1 base heterology would be even lower than the 10.4 nM value reported here. The reported \( K_{D_{app}} \) of MutS for the G/T mispair (39 nM) (Su et al., 1988) is higher than that of the 1 base heterology making this the first report of a substrate for which MutS has a higher affinity than the G/T mispair.

DNaseI footprinting experiments have shown that MutS protein protects about 20 base pairs surrounding a G/T mismatch from DNaseI cleavage (Su and Modrich, 1986; Su et al., 1988; Grilley et al., 1989). Moreover, in the presence of MutL and ATP, MutS protects about 100 base pairs (Grilley et al., 1989). In the present study, we have found that MutS protein asymmetrically protects about 22 base pairs surrounding the G/T mismatch, and 23 to 25 base pairs surrounding the 1, 2 and 3 base heterologies. The asymmetry of MutS binding is in agreement with a related study that demonstrated a similar asymmetry by methidiumpropyl-EDTAFe(II) footprinting (Su and Modrich, 1986). The asymmetries observed with the 1, 2 and 3 base heterologies are the same, but opposite of that observed with the G/T mispair. This asymmetry suggests that MutS may recognize a specific face of the helix surrounding the heterology such that changing the heterology from one strand to the other may provoke a 180° change in MutS binding asymmetry. However, it is quite possible that changing a heterology from one strand to the other may produce a concurrent change in MutS binding affinity since the sequence environment also changes. This has been observed by Jiricny et al., (1988) since MutS affinity for G/T and C/A mispairs decreased when they were inverted to T/G and A/C mispairs.
Experiments are currently being planned to investigate the strandedness of small heterologies relevant to their influence on the asymmetry of MutS binding. In addition, since DNaseI protection assays tend to over-estimate the region protected by bound protein, a more precise investigation aimed at identifying the nucleotide residues involved in MutS binding using the more sensitive methodology of hydroxyl radical footprinting are currently underway.

ACKNOWLEDGMENTS

We thank Dr. Paul Modrich for his generous gift of MutS protein, and Drs. Robert Lahue, Michael Volkert and Larry Hardy for technical advise and helpful discussions. This work was supported by a grant from the National Institutes of Health (GM33233).

REFERENCES


CHAPTER 6

Final Comments and Future Directions
FINAL COMMENTS

The recognition and repair of insertion/deletion mutations is a vitally important cellular process since unrepaired mutations of this type can have devastating consequences to viability. These types of mutations can result in the production of non-functional proteins by altering the genetic reading frame, or they can affect the level of genetic expression of a protein by changing regulatory DNA sequences. In this thesis, I have presented evidence in support of the recognition and repair of small heterologies by the Dam-dependent mismatch repair system of *Escherichia coli*.

The experimental approaches used in this work have led to the development of a sensitive *in vivo* transformation system for the rapid analysis of DNA repair activity. This system circumvents many of the problems experienced with other *in vivo* systems used to evaluate DNA mismatch repair. For example, the "strand-loss" phenomenon that has been observed with phage dependent systems was not observed with the *mnt* system. Strand loss complicates the mismatch repair analysis in phage based systems since as many as 70% of the infective centers reflect pure phage of the genotype of either strand of a heteroduplex in the absence of repair. In other words, lack of repair of an appropriate heteroduplex molecule would be indicated by the presence of mixed infective centers, of which the frequency of occurrence does not exceed 30% (Reviewed in Meselson, 1988). Conversely, efficient mismatch repair of a heteroduplex is indicated when the frequency of occurrence of mixed infective centers drops to about 5%. This makes the analysis "window" for repair very small thus decreasing the sensitivity of the system. In the *mnt* system the absence of mismatch repair is indicated by the presence of about 90% mixed colonies and a small percentage of pure red or white colonies (the presence of the
pure colonies may be the result of localized repair due to nick translation, or due to low efficiency repair by other DNA repair systems). This provides a large analysis "window" which serves to increase the sensitivity of the assay. The marginal repair of 4 base heterologies was detected by the mnt system, whereas the repair of this heterology in phage based systems might not have been detected.

The methods used to analyze for mismatch repair using the phage based systems are quite laborious in comparison to the MacConkey/lactose system. In most cases, transfected cells must be plated prior to lysis in order to detect mixed infective centers. Phage progeny isolated from these infective centers must be used to reinfect the host strain in order to identify the genotype. Unlike the phage systems, the mnt system can rapidly evaluate mismatch repair of heteroduplexes after a single transformation on the basis of colony color. This system also contains the added benefit of an internal control, based on the tetracycline phenotype of the transformant, to detect the presence of false white colonies.

In summary, I have presented both genetic and biochemical evidence supporting a key role for the repair of small heterologies by Dam-dependent mismatch repair. Using the mnt in vivo system, I have shown that: 1. DNA containing 1, 2 and 3 base insertion/deletion heterologies are repaired with efficiencies similar to G/T mispairs in a methyl-directed process that required the gene products of mutH, L and S, 2. 4 base heterologies are marginally repaired, and 3. no repair is observed when the DNA heteroduplex contains a 5 base heterology. In addition, in vitro studies with MutS protein have shown that: 1. MutS protein binds to DNA containing 1, 2, 3 and 4 base heterologies, which are
subject to repair by mismatch repair, but not to 5 base heterologies, which are not substrates for mismatch repair. 2. The hierarchy of MutS binding to heterologies relative to a G/T mismatch is 1 base Δ > 2 base Δ > 3 base Δ > G/T > 4 base Δ, 3. MutS recognizes and specifically binds to the G/T mismatch and 1, 2 and 3 base heterologies as determined by DNaseI protection assays.

**FUTURE DIRECTIONS**

The potential for future applications for the mnt system is significant and a few of these are outlined below. In addition, the work presented in this thesis has given rise to a new set of questions. The following section introduces a list of questions for future consideration followed by a brief outline of the experimental approach that could be used in the investigation.

1. Is mutH independent repair of 1, 2 and 3 base heterologies due to nicking activity from the gene product of endA (endonuclease I)? Construct a mutH indicator strain that is endA', transform with heteroduplexes containing 1, 2 and 3 base heterologies and evaluate for the presence of mutH independent repair on MacConkey/lactose indicator agar. Use GM4331 (wild type) as a control.

2. If given enough time prior to plasmid replication, will 4 and 5 base heterologies be repaired? This can be investigated using an indicator strain containing a dnaA (Ts) allele.
to block plasmid replication at 42°C for various time periods after transformation, followed by a shift to the permissive temperature. Screen for repair on MacConkey/lactose indicator plates.

3. Is the efficiency of repair the same for each of the four (A, T, G, C) 1 base heterologies? Evaluate the efficiency of repair for each of the single base insertion/deletion heterologies within the same sequence context, but at different locations within the mnt gene. Assay for repair efficiency using mnt system.

4. Is the rate of repair of 1, 2 and 3 base heterologies similar to the G/T mispair? Use E. coli cell free extracts to evaluate the rate of repair of heteroduplexes containing 1, 2 and 3 base heterologies and the G/T mispairs in vitro. Transform into mutL or mutS indicator strains. Score for the relative rate of repair.

5. What are the on and off rates for binding of MutS protein to the 1 base heterology and the G/T mispair? Set up filter binding assays for kinetic analysis.

6. What are the apparent equilibrium dissociation constants for 1, 2, 3 and 4 base heterologies and how do they compare to the G/T mismatch? Prepare heteroduplexes, isolate and radioactively label appropriate restriction fragment, titrate a fixed concentration of heteroduplex with increasing concentrations of MutS protein, and treat bound DNA fragments with DNaseI. Calculate $K_{Dapp}$ using the point at which MutS provides 50% protection from DNaseI digestion.
7. Does the asymmetry of MutS binding change if the insertion/deletion mutation is moved from one DNA strand to the other? Prepare two DNA substrate molecules containing the same insertion/deletion at the same site, but on opposite strands. Evaluate using DNaseI protection assays.

8. Which nucleotide residues are involved in MutS binding? Hydroxy radical footprinting of each strand of a 1 base heterology in combination with NMR studies of the 1 base heterology duplex may provide some useful information. This study is currently being planned.
CHAPTER 7

REFERENCES
REFERENCES


